Výskum a štatistika v spoločenských vedách „s humorom a láskou“

Ľubomír Oláh, Michal Oláh

Výskum a štatistika v spoločenských vedách (s humorom a láskou)

Ľubomír Oláh, Michal Oláh

Bratislava

2023

Čiernobiele elektronické koláže a obrázky bez uvedenia autora: Ľubomír Oláh

Ilustračné obrázky: Daniel Péter

Vydavateľ: Vysoká škola zdravotníctva a sociálnej práce sv. Alžbety, Bratislava

Rok: 2023

ISBN: 978-80-8132-273-0

Obsah

Str.


Zaujímavé je, že štatistiku považujú za zaujímavú hlavne štatistici...

Ú v o d

Byť, či nebyť? To je otázka!
Wiliam Shakespeare: Hamlet

Aké otázky si kladú humanitné vedy? Aká je motivácia pomáhajúceho profesionála v tzv. pomáhajúcich profesiách ako sociálna práca, ošetrovateľstvo či verejné zdravotníctvo, psychológia alebo pedagogika? Má zmysel, aby si kládli nejaké otázky, alebo sa majú venovať terénu a tam robiť, čo sa dá? Traduje sa, že veriace židovské matky sa nepýtajú synov, ktorí sa vrátili zo školy, ješivy, či vedeli odpovedať, ale či vedeli položiť otázku.

Keď odhliadneme od okrajových javov, väčšina študentov a sociálnych pracovníkov sa rozhodla pomáhať ľuďom. Sociálna práca je z istého pohľadu zastrešením čiastkových snáh mnohých iných disciplín, napr. sociológie, klasickej i evolučnej psychológie, sociálnej antropológie, ale aj právnych vied, ekonómie, verejného zdravotníctva a mnohých iných odvetví. Ale konkrétna práca s konkrétnymi ľuďmi v konkrétnej situácii jej vždy zabezpečuje, že sa neodtrhne od reality, od praxe. Väčšina pomáhajúcich profesionálov alebo iných absolventov humanitných vied bude vždy vykonávať mravčiu terénnu prácu s individuálnymi klientmi alebo v skupine v zabehnutých koľajach a veríme, že sú skvelými interpretmi. V hudbe si ich veľmi vážime. Niektorí však dostali do daru istú tvorivosť a predovšetkým zvedavosť v tom najlepšom slova zmysle, ako silný náboj ľudského dôvtipu. Ak napríklad sociálny pracovník zistí pri práci v nejakej komunite, že v šiestich z desiatich možných prípadov dochádza k domácemu násiliu a pritom v skupine, s ktorou pracoval nedávno to boli len tri prípady z desiatich a dokonca má vo svojej praxi aj takú societu, kde to bol len jeden jediný prípad, začne rozmýšľať, ako a s čím to súvisí. S veľkou pravdepodobnosťou si neuvedomí, že práve urobil jednu veľmi dôležitú činnosť, ktorou sa zaoberá matematika: porovnanie veľkosti nejakej charakteristiky. Nájsť súvislosť, pričom tuší, že asi existuje, a môcť pretransformovať uvedenú vlastnosť prvej skupiny na tretiu, by bol veľmi dobrý a užitočný výsledok. Podobné otázky, mnohé kontroverzné, sa môžu objaviť pri práci na rôznych závislostiach, s utečencami, s nezamestnanými, seniormi atď. Aktuálna je napr. problematika stability rodiny, čo si pod tým vlastne predstavujeme, ale aj problematika pedofílie, legalizácie či nelegalizácie drog a mnoho iných oblastí, kde je možné hľadať a nájsť často netriviálne a intuitívne neočakávané súvislosti a závislosti. Dá sa urobiť nejaké zovšeobecnenie a preniesť pomoc jednotlivcovi na väčšiu skupinu?

V predchádzajúcom odseku sme vyslovili jedno hrozné slovo – matematika. Predpokladáme, že väčšina čitateľov má k nej bežný vzťah, získaný pôsobením nášho školstva, na ktoré sme z nejakého dôvodu hrdí, teda, že sa po desiatkach rokov od skončenia akejkoľvek povinnej školskej dochádzky budia spotení z hrozného sna, v ktorom musia riešiť akúsi matematickú úlohu, a celý deň potom majú úzkosť okolo žalúdka. Takže zo všetkej krásy, ktorú by im mohlo ponúknuť logické myslenie im zostala neuróza a keďže sa s tým musia nejako vysporiadať, tak z toho urobia zásluhu: byť zadobre s matematikou je niečo nepatričné, v slušnej spoločnosti čudácke. Určite ich poteší bonmot, ktorý sme dávnejšie našli na jednej matfyzáckej webovej stránke:

Dôvodom, prečo si každá väčšia univerzita udržuje matematický ústav je fakt, že to príde lacnejšie, ako všetkých tých ľudí hospitalizovať.

Ďalšia veľká skupina ľudí sú tí, ktorí sa rozhodli postaviť čelom k úlohám, ktoré prináša sám život, či chceme alebo nie. Ich najväčšou zaťažkávacou skúškou je, keď ich ratolesť začne navštevovať základnú školu a už na prvom stupni prinesie domov zákernú slovnú domácu úlohu o tom, ako zistiť z určitých zadaných podmienok vek starého otca. Po prebdenom veľmi ťažkom víkende, keď to dcérka vzdala hneď na začiatku, že je tam zlé zadanie, aby mohla ísť von, mamička, doktorka filozofie príde k záveru, že úloha nemá riešenie. Otec, racionálnejší typ, lekár, dosiahne výsledok: vek starého otca je -1 (slovom mínus jeden) rokov. Slúži im ku cti, že výsledok nepovažujú s veľkou pravdepodobnosťou za úplne správny a zavolajú niekomu, o ktorom vedia, že by to mohol vyriešiť a keď sa za tridsať sekúnd dozvedia správne riešenie, teda ani neminuli príliš veľa peňazí na spojenie, utvrdí ich to v tom, čo už dávno tušili, že ten na druhom konci rozhovoru musí byť jednoducho génius.

To, samozrejme, nie je pravda. Ale o tom, ako pracuje matematik, si povieme neskôr. Absolvent humanitných vied sa niekedy musí orientovať vo veľmi neprehľadných vzťahoch a javoch sociálneho systému. S istým láskavým preháňaním možno povedať, že to má omnoho ťažšie, ako napr. niektorí prírodovedci, ktorí si môžu s atómom robiť čo chcú, dokonca ho aj rozbiť a on neodvráva, nemusia chrániť jeho osobné údaje, neuteká v nevhodnej chvíli, neklame, nemanipuluje a ani inak nespôsobuje problémy. Keď chce niekto pozorovať galaxiu, napr. Mliečnu dráhu, má na to dostatok času, veď sa otočí okolo svojej osi raz asi za 250 miliónov rokov. Je to podobné, ako keď sa vojaka Švejka na žandárskej stanici v Putimi, kde ho odchytili pri jeho ceste do Budějovíc, pýtali, či je ľahké

fotografovať železničné nádražie. Odpoveď bola, že áno, pretože sa nehýbe. Nech nám je istým prvým varovaním na našej ceste, že bol z toho vyvodený nesprávny záver, že je ruským špiónom.

Na našej spoločnej ceste sa budeme snažiť vidieť alebo aspoň cítiť, že javy okolo nás by mohli mať svoju zákonitosť, poriadok, rád, že Boží svet nie je úplný chaos. Nebude to ľahké, ale môže nám to dať potešenie niektoré zákonitosti a poriadok stvorenia odhaľovať. A to všetko môže byť pre sociálneho pracovníka a hlavne pre ľudí, ktorým sa snaží pomôcť, veľmi užitočné. Pevne verím, že na konci dospeje štatisticky významná časť čitateľov k prekvapujúcej pravde, že matematika a jej dcéra s mimoriadne zlou povesťou – štatistika – nie sú tu preto, aby nám život komplikovali, ale práve naopak, jeho zložitosť zjednodušovali.

Vybrali sme sa cestou ponúknuť študentom a pracovníkom humanitných vied vhodný pracovný nástroj pre prax. Pretože svet je, ako hovoria fyzici, pravdepodobnostný, a my nemáme dôvod im neveriť. Teda nepísať odbornú publikáciu pre odborníkov, ktorú aj tak nikto nečíta, pretože odborníci čítajú len svoje výtvory, ale sústrediť sa na aplikácie a zrozumiteľnosť pri predpokladanej matematickej zručnosti čitateľov. Aby sme sa dostali na spoločnú pojmovú bázu, budú začiatky často až triviálne. Prednášajúci a autori publikácií, ktorí popíšu mnoho tabúľ a strán nezrozumiteľnými znakmi a rovnicami dosiahnu akurát to, že s istou pravdepodobnosťou sa dá povedať, že sú v matematickej štatistike lepší. Nič viac!

Aby sme preplávali medzi smrteľnými obludami Scyllou a Charybdou, pustili sme sa na jednej strane do veľkého a neodpustiteľného zjednodušenia matematických základov štatistiky, ktoré by mnohým matematicky vzdelaným odborníkom prinieslo stav hlbokej frustrácie. Rôzne vzťahy a tvrdenia budú naozaj len pracovným návodom bez potrebného logického dôkazu a mnohokrát aj bez precíznosti a jasnosti, čím čitateľov ochudobňujeme o krásu ich hĺbky, ale keby po nej túžili, zvolili by si asi iné štúdium. Sme presvedčení, že keď si niekto chce večer doma zapnúť svetlo, nepotrebuje predtým vyriešiť Maxwellove rovnice elektromagnetizmu v integrálnej forme, ani preštudovať dopady Hirošimy, Nagasaki a Černobyľu na všetky úrovne biologických funkcií človeka pred návštevou zubného rtg či byť konštruktérom, dizajnérom a dopravným inžinierom, keď si chce sadnúť za volant svojho auta, ale vstupuje do prirodzenej spolupráce s príslušnými odborníkmi.

Na druhej strane popularizácia nie je populizmus. Chcieť robiť sám bez spolupráce a bez štatistiky nejaký kvantitatívny alebo integrovaný výskum v sociálnej práci s dopadom

na terén, ale možno aj na legislatívu či na potrebné financovanie nevyhnutných projektov, je asi to isté, ako chcieť zahrať úplne sám Mahlerovu alebo Beethovenovu 9. symfóniu bez nôt a bez nástrojov a možno aj bez rúk, nôh a uší.

Dúfame, že sa nám tento kompromis podarí a že na jeho konci bude neochota použiť štatistické metódy limitované nie nedostatkom potrebného vzdelania, ale len intelektuálnou lenivosťou. Jedna rada na záver úvodu: Neponáhľajte sa! Nikto múdry z neba nespadol, chce to určite istú intelektuálnu námahu a čas, ale sľubujeme vám, že ak vydržíte, štatistika sa vám odvďačí.

Bratislava 1. januára 2023


Ja vám verím, že štatistika je krásna. Ale prečo potom vždy lietajú facky?

I. Útek spod Babylonskej veže alebo hľadanie spoločného jazyka

Dôležitejšie ako fakty je, ako ich pomenuješ.
Jeden z Murphyho zákonov

Možno si pamätáte na drobnú hereckú etudu Zdeňka Svěráka vo filme Kolja, keď si obrúskom prikryl ľavú ruku a svoju partnerku (hrala ju Libuše Šafránková) požiadal, aby povedala číslo od jedna do päť.

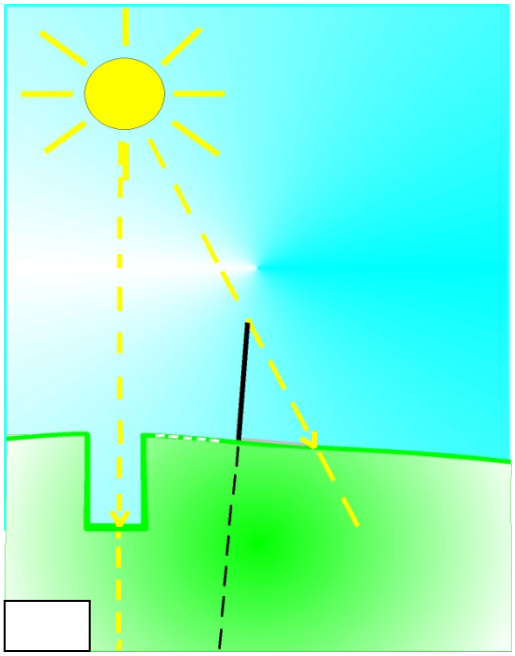
„Tri," - znela zároveň odpoveď i otázka.

Nasledovala zámerne naťahovaná dlhá chvíľa napínavého sústredenia a očakávania, zakončená trochu previnilým a trochu rozpačitým Svěrákovým úsmevom, keď spoza obrúska ukázal tri prsty.

Táto rozkošná infantilita nás priviedla k jednému zo základných pojmov elementárnej matematiky, k číslu. Jeho objav bol obrovským skokom v myslení a poznávaní sveta. Keď sa pračlovek vrátil do jaskyne, životne dôležitou informáciou pre celú tlupu bolo, koľko ulovil mamutov. Nosiť ukázať ulovené kusy cez úzky vchod do jaskyne a potom späť na spracovanie bolo trochu nepohodlné, preto dlho a zádumčivo sedával pri osamelom ohni a mechanicky rýpal pazúrikom do palice, až mu jedna iskierka osvetlila riešenie: Na palici alebo na kosti nejakého zvieraťa videl toľko vrypov, koľko práve v ten deň ulovil mamutov. S údivom zdvihol ruku a roztiahol presne toľko prstov. Neviem, či sa práve vtedy zrodila matematika, ale paleontológovia a historici predpokladajú, že tento objav sa uskutočnil v priebehu geologicky krátkej doby niekoľkých tisícročí na viacerých miestach našej planéty. Zaujímavejšie je štiepenie následnej evolúcie. Predstaviteľ jednej odnože, pre zjednodušenie ich budem možno nepresne volať neandertálci, pochopil význam nového poznania vlastným spôsobom. Presvedčil ostatných, že bude najlepšie, keď on bude robiť evidenciu ulovených mamutov a ich spravodlivé delenie, teda vlastne všetko riadiť, čím mu, samozrejme, už nezostane čas na samotný lov, ale zato, keď všetko pôjde tak, ako to naplánuje a všetci sa budú viac snažiť, sľubuje svetlú budúcnosť s omnoho väčším množstvom úlovkov a všeobecným blahobytom. Stačí, keď si ho zvolia za vodcu tlupy. Z nejakého záhadného, historikmi doteraz neobjasneného dôvodu väčšina z nich však do dnešnej doby nevyhynula. Druhá skupina vypozorovala, že v niektorých obdobiach mohli zaznamenať viac vrypov o úlovkoch a v iných,

keď však tiež potrebovali jesť, ani jeden a že sa to viac-menej pravidelne opakovalo. Preto si v čase dobrého lovu robili zásoby. A prežili tiež.

Ťažko usúdiť, či sa v tomto momente zrodili základy vedy, ale aj keď to mnohí tvrdia, veda určite nezačala až Galileom Galileim. Skôr by sme mohli jej začiatok posunúť do Antiky. Veď čo by sme si počali bez starých Grékov! Predstavujeme si Eratosthéna z Kyrénie (276-194 pred Kr.), ako meria rýchlosť pohybu karavány tiav, aby pomocou tejto znalosti dospel k výpočtu obvodu a polomeru zemegule, pretože už v 3. storočí pred Kristom bol na základe pozorovaní presvedčený o jej tvare [1], [2]. Pokúsme sa skoro po dvoch a pol tisícročiach chvíľu putovať s jeho karavánou medzi starovekou Alexandriou a oázou Syene, ležiacou južne od nej na obratníku Raka. Eratosthénes zo svojich ciest vedel, že na pravé poludnie počas letného slnovratu dopadá slnečné svetlo v oáze Syene priamo na dno hlbokej studne, teda Slnko musí byť priamo nad ňou. V tom istom čase zmeral dĺžku tieňa pekne rovno sa týčiaceho najštíhlejšieho z troch obeliskov v Alexandrii a zistil, že pri jeho výške 80,5 starovekých attických stôp je dĺžka tieňa 10,17 stopy. [11]



Obr.I.1. Ilustračný obrázok, Daniel Péter

Pozrime sa na obr. I.1: Trojuholník so stranami obelisk, slnečný lúč a tieň obelisku môžeme s istým prijateľným zjednodušením považovať za pravouhlý. Keď vydelíme dĺžku tieňa výškou obelisku, dostaneme veličinu, ktorú nám na základnej škole nazvali tangens uhla (tg alebo tgn) dopadu slnečných lúčov v Alexandrii, označme ho α. Keď použijeme kalkulačku, dostaneme $\mathrm{tg}\ \alpha = 10{,}17/80{,}5$ a pre samotný uhol $\alpha = 7{,}20°$, teda $7°12'$.

Počas putovania karavány máme dosť času. Prisadnime si k odpočívajúcim pútnikom do tieňa paliem počas siesty. Bokom od ostatných sedí drobný mladý muž s hustou kučeravou bradou, ale aj s plešinou neprimeranou jeho veku, okolo ktorého sa vznáša ako keby jemný opar myšlienok. Ešte nevie, že o nejakých 11 rokov sa stane riaditeľom Alexandrijskej knižnice, vychýrenej vedeckej ustanovizne s nesmiernym pokladom písomností vtedajšieho sveta.

„Buďte pozdravený premúdry Eratosthénes z prekrásnej Kyrénie, nech vás ochraňujú olympskí bohovia, hlavne matka všetkej múdrosti Glaukópis Aténa!“ – prihovárame sa kvetnato podľa dobového zvyku osamelému učencovi. Vzhliadne od svojich papyrusov, neprítomne odvetí:

„Tak, tak...“

Nebola to odpoveď hodná potomka Homéra, tak to trpezlivo skúšame znovu:

„Už dlhšie vás pozorujeme a chceme sa spýtať...“

„Špehujete ma?!“ – vybuchol podráždene, až mu ohryzok v krku behá prudko hore-dolu.

„Ak pracujete pre toho hlupáka Euklida, tak zmiznite a nech vaše kosti vyblednú v piesku púšte. Odkážte tomu ozembuchovi, ktorý sa ani poriadne nevedel matematiku naučiť, že s ním spolupracovať nebudem ani keby sa Orión teraz v lete na oblohe zjavil!“

Začíname mať pocit, že sa konverzácia nevyvíja úplne podľa našich prianí.

„Nie, to je omyl, s pánom Euklidom nemáme nič spoločného, sme len jeho obdivovatelia...“

„Somár obdivuje somára!“ – vybuchol znovu, ale v poludňajšej horúčave musel nabrať dych, tak sme to využili:

„ ... ale vás a hlavne vašu prácu obdivujeme omnoho viac.“

Podozrievavo na nás pozrel, ako to myslíme. Musíme uznať, že v logike je naozaj dobre podkutý.

„A o akú moju prácu máte záujem?“ – dodal po chvíli s rafinovaným úsmevom a pozrel priamo na nás, aby videl naše rozpaky, ako nás nachytal.

„Videli sme vás pri alexandrijskom obelisku ako aj pri studni v Syene a myslíme si, že sa pokúšate odmerať obvod zemegule, len nevieme, ako chcete pokračovať.“

Niekedy je premena počas rozhovoru taká hlboká, že vás to samotných prekvapí. Najprv mu oťažela brada a s otvorenými ústami a rozšírenými očami dlho a bez pohnutia hľadel na nás, ako keby sa pred nim zjavila hlava Medúzy.

„Vy naozaj veríte, že Zem je guľatá?“

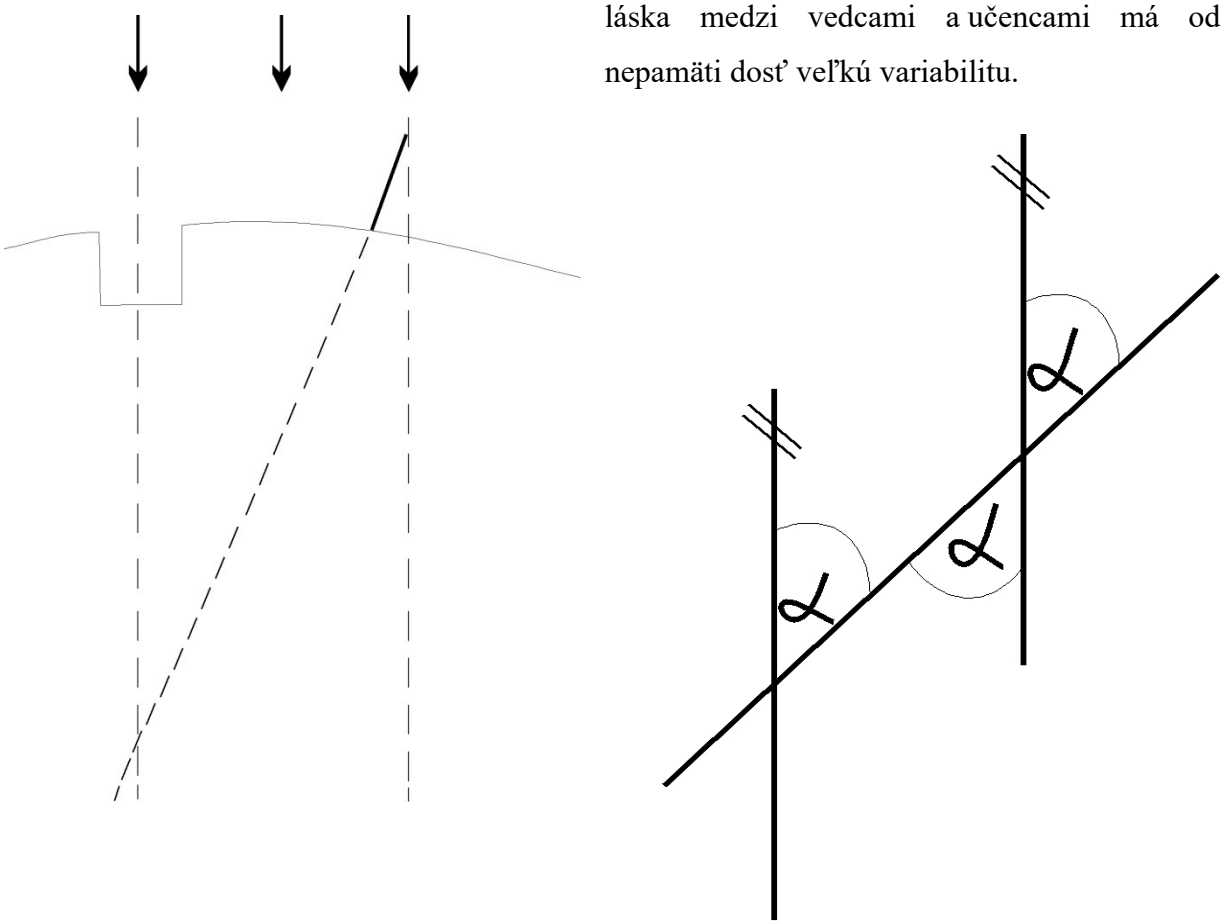
„Noooó..., my to vieme.“

„Tak to asi naozaj nebudete pätolízači toho nedouka Euklida,“ – viditeľne sa pozbieral z prvého šoku.

„Máme tu niekoľko predpokladov,“ – pokračoval živo a začal kresliť palicou do piesku – „predovšetkým si musíme myslieť, že boh Apolón vo svojom zlatom kočiari sa po nebeskej klenbe preháňa ešte o niečo vyššie ako je najvyšší vrch Olymp. Teda jeho lúče dopadajúce na Zem sú všetky rovnobežné. Ak je Zem guľa a predĺžime studňu v Seyne až do jej stredu, stretne sa tu s predlžením alexandrijského obelisku. A v deň letného slnovratu tieto dve priamky zvierajú ten istý uhol, pod akým dopadajú Héliové lúče na vrchol obelisku v Alexandrii. Rozumiete?“

„Hm, myslím, že áno, použili ste druhý a piaty Euklidov postulát...“

„Nespomínajte mi tú škriekajúcu opicu!“ – zaškriekal tak divoko, že spiace ťavy zodvihli hlavy, začali hlasno fŕkať okolo seba a vydávať híkavé zvuky. Pravdepodobne chcel ukázať, že láska medzi vedcami a učencami má od nepamäti dosť veľkú variabilitu.



Obr.I.2: Pokus o rekonštrukciu Eratosthénových náčrtkov. © Daniel Péter

„V Kyrénii každé malé dieťa vie, že dve rôznobežky v rovine sa pretnú, že keď priamku pretnú dve rovnobežky, tak s ňou zvierajú rovnaké uhly! Aj stará ženská na trhu vám prezradí, že vrcholové uhly sú zhodné, na to nepotrebuje *Základy*!“ – vyskočil na krátke krivé nohy a v rituálnom tanci dlhou pútnickou palicou kreslil do piesku krivky a rovné čiary, ktoré sa bohužiaľ do súčasnosti nezachovali, ale viac-menej ich obsah možno pochopiť z obrázku I.2.

Palica lieta hore-dolu nad našimi hlavami, tak v strachu rýchlo prikyvujeme.

„Vidím, že to chápete, preto nechcem nosiť sovy do Atén. Uhol, ktorý som zmeral v Alexandrii je presne 1/50 celého uhla. A je to ten istý uhol v strede Zeme, ktorý vytína na obvodovej kružnici vzdialenosť Alexandrie a Syeny. Tieto lenivé ťavy mi pomáhajú pri 40-dňovej púti ju zmerať. Za prvých desať dní som zaznamenal, že sme prešli každý deň nasledovné úseky,“ – rýchlo ukázal na papyrus s tabuľkou I.1:

Deň	attické štádia
1.	135
2.	140
3.	125
4.	127
5.	110
6.	99
7.	162
8.	125
9.	150
10.	77

Tabuľka I.1. Eratosthénove záznamy putovania ťavej karavány

„Ale ešte si to idem overiť...,“ – zohýba sa a v hlbokom zamyslení zbiera rozhádzané papyrusy. Už ho nezaujímame.

Nám, štatisticky intuitívne naladeným, to však úplne stačí. S grandióznou ľahkosťou zistíme, že karavána tiav sa s veľkou pravdepodobnosťou pohybuje priemernou rýchlosťou 125 štádií za deň, teda za 40 dní urobí 5000 olympijských štádií. To je teda vzdialenosť Alexandrie a Syeny. Takže keď to vynásobíme 50, dostaneme obvod Zeme 250 000 štádií. Opusťme teraz karavánu, nám stačí odskočiť si do starovekej Olympie a zmerať dĺžku pretekárskeho štadióna. Náš presný laserový diaľkomer ukazuje 176,40 metra. Jednoduchým násobením dostávame pre vzdialenosť Alexandrie a Syeny 882 km, čo v nás nevyvoláva žiadne emócie. Ale keď to vynásobíme ešte 50, dostávame pre obvod Zeme hodnotu 44 100 km, čo sa od skutočnosti líši len o nejakých 10%, a to je skvelý výkon! Sme len na začiatku, ale už len vlastným umom,

pomocou tužky, papiera a niekoľko málo vstupných údajov dokážete zistiť obvod a teda aj polomer našej materskej planéty Zeme! A verím, že dokážete ešte omnoho viac. Pre hlbšie pochopenie je namieste, aby si každý, pri použití všetkých možných výdobytkov súčasnej vedy a techniky s výnimkou nahliadnutia do astronomických tabuliek a encyklopédií, navrhol a uskutočnil dva nezávislé pokusy, pomocou ktorých by zmeral obvod a polomer našej planéty a porovnal ich s Eratosthénovým (a naším) výsledkom.

Vráťme sa k číslam. Najjednoduchšie sú prirodzené čísla, teda celé čísla od nuly do nekonečna. Spočítať ich by trvalo trochu pridlho, tak ich množinu označujeme

$$\mathbf{N = \{0,1,2, ..., \infty\}} \qquad \mathbf{[I.1.]}$$

Ak nejaké číslo **x** patrí do množiny prirodzených čísel, tak to zapíšeme $\mathbf{x \in N}$. Rozšírením tejto množiny o záporné celé čísla dostaneme množinu celých čísel **Z**:

$$\mathbf{Z = \{-\infty, ..., -2,-1, 0,1,2, ..., \infty\}} \qquad \mathbf{[I.2.]}$$

Množina všetkých zlomkov **a/b**, teda pomerov ľubovoľných čísel **a** aj $\mathbf{b \in N}$, sa nazýva množina racionálnych čísel **Q**.

Okrem nich existujú aj iracionálne čísla, ktoré sa nedajú vyjadriť v tvare zlomku, napr. odmocnina z 2, Ludolfovo číslo **π**, ktoré udáva pomer obvodu a priemeru kružnice a jeho hodnota je približne

$$\mathbf{\pi \cong 3,14159} \qquad \mathbf{[I.3.]}$$

alebo Eulerovo číslo **e**

$$\mathbf{e \cong 2,71828} \qquad \mathbf{[I.4.]}$$

Znak ≅ budeme používať vo význame približne, približný ale dostatočný zaokrúhlený odhad presnej hodnoty. Čísla racionálne a iracionálne spolu tvoria množinu reálnych čísel **R**. Z iných čísel, ktoré už nebudeme potrebovať, uveďme ešte zábavné číslo **i**, s vlastnosťou, že keď ho vynásobíme samým sebou, dostaneme **-1,** a voláme ho imaginárna jednotka.

Platí teda:

$$\mathbf{i \,.\, i = i^2 = -1}$$

(Nebojte sa, v tejto súvislosti ho budem používať len výnimočne. Nemusíte si to pamätať, vždy na to upozorním.)

S číslami môžeme kadečo robiť, všelijako nimi manipulovať, ale napodiv má to svoju logiku. Hovoríme, že existujú medzi číslami matematické operácie. Najjednoduchšia operácia je operácia porovnania veľkosti, napr. číslo **a** je väčšie ako číslo **b**, alebo $\mathbf{N_0}$ je menšie nanajvýš rovné **N**, resp. tri poloviny sú menšie ako osem štvrtín, potom píšeme:

$$\mathbf{a > b} \text{ alebo } \mathbf{N_0 \leq N} \text{ resp. } \frac{3}{2} < \frac{8}{4} \qquad \textbf{[I.5.]}$$

V matematike často výsledky alebo nejaké tvrdenia vyjadrujeme vo forme intervalu. Čaká to aj nás, preto si vyjadrenia intervalov a čo to znamená zhrnieme v tab. I.2:

Tabuľka I.2. Vyjadrenie intervalov.

Vyjadrenie pomocou nerovností	Interval	Pozn.:
$a < x < b$	$x \in (a; b)$	Otvorený interval, bez **a** a **b**
$a < x \leq b$	$x \in (a; b\rangle$	Polouzavretý interval sprava
$a \leq x < b$	$x \in \langle a; b)$	Polouzavretý interval zľava
$a \leq x \leq b$	$x \in \langle a; b\rangle$	Uzavretý, s **a** a **b** hranicami

Niekoľko poznámok k tabuľke I.2: Čísla **a** a **b** sa nazývajú hranice intervalu. Ak je hranicou intervalu $-\infty$ alebo $+\infty$, je to na strane ich výskytu vždy otvorený interval. Analogicky sa vyjadria intervaly s opačným smerovaním nerovnosti, len pozor na to, čo je ľavou a pravou hranicou intervalu. Asi to chce nejaké konkrétne príklady. Niektoré mladšie humanitné vedné disciplíny, ako napr. sociálna práca, urobila veľký pokrok, keď zistila to, čo sa doteraz nepodarilo ani sociológii, ani psychológii, že jednotlivci akejkoľvek vzdelanostnej úrovne vynikajúco rozumejú aj najzložitejším matematickým operáciám, pokiaľ sa týkajú ich osobných financií. Preto občas uvedieme praktické vysvetlenia z tejto oblasti.

Pr.I.1. Ak váš osobný príjem v eurách označíme písmenom **u** a výdaj písmenom **v**, s veľkou pravdepodobnosťou platí nerovnica

$$\mathbf{u < v}$$

Ak s údivom zistíte, že to isté platí aj pre štátny rozpočet, môžete urobiť záver, že sa správate rovnako ako ekonomicky vyspelý moderný štát, čo môže byť v niektorých prípadoch isté uspokojenie. Pokiaľ zistíte, že obsah vašej peňaženky alebo účtu sa pohybuje sprava doľava v intervale **⟨0; u⟩** len prvý týždeň po mesačnej výplate, je už ťažšie hľadať nejaké upokojujúce analógie.

Ak napr. sociálny pracovník, riešiaci problematiku seniorov, zistí, že interval, v ktorom sa pohybujú starobné dôchodky je **⟨200 €; 1600 €⟩**, ale priemerný starobný dôchodok u nás je len niečo cez 300 € miesto intuitívne predpokladaných 900 €, tak na objasnenie tejto nezrovnalosti bude potrebovať trochu hlbšiu štatistickú analýzu, ktorú sa naučíme neskôr.

Známou matematickou operáciou je sčítanie. Dá sa robiť vo všetkých číselných množinách bez väčších problémov, ktorými by sme sa museli zaoberať. Budeme sa stretávať s rôznymi postupnosťami čísel a ich súčtami, nazývanými aj (aritmetické) číselné rady:
Postupnosť: $\mathbf{a_1}$, $\mathbf{a_2}$, **...,** $\mathbf{a_n}$ má **n** členov (napr. vek respondentov pri prieskume, príjem, čas venovaný športovým aktivitám a pod.), kde index **n** určuje poradie člena v postupnosti. Znak **Σ** (sigma, veľké písmeno gréckej abecedy) sa v matematike používa vo význame „suma“, teda súčet všetkých členov postupnosti s indexmi od dolnej hranice po hornú. Jej súčet značíme potom nasledovne:

$$\sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n \qquad [\mathbf{I.6.}]$$

pričom indexy **k** a **n** sú celé nezáporné čísla (používajú sa aj iné písmena napr. **i,j,k** atď,) členovia postupnosti môžu byť vo všeobecnosti reálne čísla. V úspornej matematickej symbolike, na ktorú si začíname zvykať je to:

$\mathbf{k, n \in N, a_n \in R.}$

Pr.I.2: Počet zachytených ilegálnych utečencov po prekročení východnej hranice republiky za posledných desať dní je v tabuľke:

n	1	2	3	4	5	6	7	8	9	10
$\mathbf{a_n}$	3	0	5	1	1	3	3	3	1	0

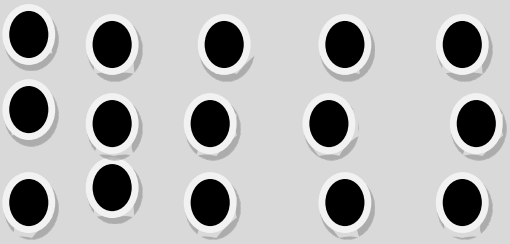
Na zistenie, koľko osôb pribudlo v azylantskom tábore, kam boli všetci presunutí, urobíme súčet tejto číselnej rady v zmysle vzťahu **[I.6.]** a dostaneme výsledok**:**

$$\sum_{n=1}^{10} a_n = 20$$

Matematici sa niekedy radi múdro vyjadrujú. Napríklad opačnú operáciu k sčítaniu, teda odčítanie, nazývajú, že je to k sčítaniu **inverzná operácia**.

Pr.I.3. Vráťme sa k príkladu I.1, kde sme zistili, že vaša osobná mikroekonomika je dostatočne popísaná nerovnicou **u < v**, kde **u** je váš príjem a **v** vaše výdavky zaokrúhlené na celé čísla. Pohľad do vašej peňaženky vám neodhalí v plnej kráse stav vašich osobných financií, pretože jej obsah sa dá vyjadriť len nezápornými číslami, teda po zaokrúhlení na eurá len v množine N. Rozdiel **u – v** vám poskytne výpis z vášho bankového účtu. Pokiaľ môžete ísť bez obmedzenia do debetu, tak bude pracovať v množine Z a vy dostanete omnoho pravdivejšiu a o dosť smutnejšiu informáciu. Teda na uskutočnenie operácie odčítania v každom prípade potrebujeme množinu všetkých celých čísel **Z**. V množine **Q** a **R** je to taktiež bez problémov.

Pr. I.4. Jaskynný človek je po dobrom love spokojný, každý z piatich lovcov kmeňa ulovil rovnaký počet 3 mamutov. V dobrej nálade a trochu pod vplyvom kvaseného nápoja, ktorý mu ženy priniesli, vyrezáva do kosti postupne 15 zárezov, čo mu zaberie celú noc. Nadránom prenikol do jaskyne prvý slnečný lúč a nášho lovca zaujali rozhádzané kamienky pri ohnisku. Postupne ich začal ukladať pod seba. Tri pre prvého lovca, ďalšie tri pre druhého atď, až mu vznikol asi takýto obrazec:



Jasnozrivo si uvedomil, že mu stačí zobrať 5 krát po 3 mamuty, aby sa jedinou operáciou dostal k tomu istému výsledku ako pomocou 14 operácií sčítania:

1+1+1+1+1+1+1+1+1+1+1+1+1+1+1 = 5 x 3

Toto je jeden malý dôkaz tvrdenia, že matematika sa snaží život zjednodušiť, zľahčiť. Spomeňme si prosím na to aj pri trochu ťažších problémoch a výrazoch. Ich zložitosť je zrkadlom zložitosti sveta, nie zlomyseľnosťou matematiky.

Dostali sme sa k ďalšej zaujímavej operácii, k násobeniu. Výsledok **c** násobenia činiteľov **a** a **b** zapisujeme:

c = a x b, alebo **a . b**, alebo jednoducho **ab** **[I.7.]**

Operácia násobenia sa dá bez pre nás dôležitých obmedzení uskutočniť v každej spomenutej číselnej množine. Niekedy potrebujeme aj súčin nenulových členov číselnej postupnosti, ktorý značíme pomocou veľkého písmena gréckej abecedy pí Π:

$$\prod_{k=1}^{n} a_k = a_1.a_2.\ \dots\ .a_n \qquad [\mathbf{I.8.}]$$

Pr. I.5. Zoberme tabuľku z príkladu I.2. a usporiadajme $\mathbf{a_n}$ podľa veľkosti (usporiadaný číselný rad):

N	1	2	3	4	5	6	7	8	9	10
$\mathbf{a_n}$	0	0	1	1	1	3	3	3	3	5

Násobením nulou dostávame vždy nulu, čiže ako keby sme zmazali všetky informácie o číselnej postupnosti, preto súčin môžeme robiť až od tretieho člena. Výsledok je:

$$\prod_{n=3}^{10} a_n = 405$$

Zvláštne miesto v operácii súčinu má súčin prvých **n** členov postupnosti kladných celých čísel, nazvaný **n faktoriál** so symbolom **n!**:

$$\mathbf{n! = n.(n-1).(n-2).\ \dots\ .3.2.1} \qquad [\mathbf{I.9.}]$$

Pr.I.6.

3! = 3.2.1 = 6

4! = 4.3.2.1 = 24

10! = 3628800

Vidíme že operácia súčinu, nazvaná faktoriál, dáva so zvyšovaním **n** rýchlo rastúce hodnoty.

Inverzná operácia k násobeniu je delenie. Dá sa bez veľkých problémov uskutočniť až v množine racionálnych čísel **Q** a v množine reálnych čísel **R**. Píšeme

$$c = \frac{a}{b} = a : b \qquad [\mathbf{I.10.}]$$

kde vo všeobecnosti môžu $\mathbf{a,b,c \in R}$. Nulou nikdy nedelíme, teda $\mathbf{b \neq 0}$.

Pokiaľ je **c ∈ Q**, hovoríme o zlomkoch, potom **a** je čitateľ zlomku, **b** menovateľ a medzi nimi je zlomková čiara. Majú svoju aritmetiku, teda trochu náročnejšie pravidlá počítania s nimi, napríklad dva zlomky možno sčítať resp. odčítať, keď majú spoločného menovateľa. Napr.

Pr.I.7. Váš mesačný príjem je 375 €. Jedna pätina sú náklady na školu, dve tretiny náklady na cestovanie, bývanie, mobil, internet a iné priame poplatky. Koľko vám zostane na stravu?
Výdaje okrem stravy sú:

$$\frac{1}{5}+\frac{2}{3}=\frac{3}{15}+\frac{10}{15}=\frac{13}{15}$$

Na stravu zostane:

$$1-\frac{13}{15}=\frac{15}{15}-\frac{13}{15}=\frac{2}{15}$$

2/15 z 375 € dostaneme tak, že 375 vynásobíme 2 a celé vydelíme 15. Nemusíte to robiť z hlavy. Pri použití kalkulačky by ste sa mali prepracovať k výsledku 50 € na stravu na mesiac a môžete si každý druhý deň dať v skromnejšom stravovacom zariadení menu.

Pre odľahčenie môžeme upozorniť, že matematika sa zaoberá podmienkami, za ktorých sa dá realizovať delenie aj v množine celých čísel Z, teda ich deliteľnosťou. Z tohto pohľadu člení potom celé čísla na prvočísla a zložené čísla. Prvočísla sú čísla, ktoré sú bezo zvyšku deliteľné len 1 a sebou samým. Zložené čísla sa dajú rozložiť na súčin prvočísel. Napr. 3 je prvočíslo, 4 = 2.2 je zložené číslo, taktiež 9 = 3.3, ale 11, 17 a i. sú prvočísla. Ak sa vám chce hľadať nejaké doteraz skryté vnútorné súvislosti, baví vás to a nechce sa vám chodiť do práce, či do školy, dáme vám tip: Vezmite si najjednoduchšiu číselnú množinu, množinu všetkých celých nezáporných čísel N. Skladá sa, ako sme uviedli, z čísel, ktoré nazývame prvočísla, teda bezo zvyšku sú deliteľné len jednotkou alebo sebou samým, a čísla zložené, ktoré sa dajú vyjadriť ako súčin nejakých prvočísel. Skúste objaviť nejaký vzťah, vzorec, postup alebo niečo podobné, ktorý by vyjadril, ako sú tieto stavebné kamene - prvočísla v množine celých čísel rozložené. Prezradíme vám, že je na to vypísaná odmena 1 milión amerických dolárov. [3], [4]. To už stojí za to trochu sa na tú matematiku pozrieť. Skôr než si pôjdete po ten milión, pozrime sa veľmi stručne ešte na niektoré potrebné operácie.

Ďalšia operácia, s ktorou sa často budeme stretávať súvisí s násobením rovnakého čísla medzi sebou samým a nazýva sa umocňovanie. Píšeme pre **n**-krát násobenie **x**:

$$x \cdot x \cdot x \cdot \ldots x = x^n \qquad \textbf{[I.11]}$$

a čítame x na entú. Napr. keď exponent n=2, čítame x na druhú, niekedy aj štvorec x z geometrickej interpretácie. Pre **n = 0**

$$x^0 = 1 \qquad [I.12]$$

Môžeme mať exponent záporný, n < 0, ale pre to existuje jednoduché pravidlo. Mocninu so záporným exponentom preklopíme do opačnej polohy zlomku a rátame ako s kladným exponentom:

$$x^{-n} = \frac{1}{x^n}; \quad \frac{1}{y^{-m}} = y^m \qquad [I.13]$$

Pr. I.8. Práca s mocninami nie je príliš zložitá. Sčítavať a odčítavať môžeme rovnaké mocniny rovnakého základu ako iné objekty: $\mathbf{a^2 + a^2 = 2\,a^2}$; $\mathbf{3b^5 - b^5 = 2b^5}$ atď. Pri násobení a delení mocnín rovnakého základu exponenty sčítame resp. odčítame: $\mathbf{a^2 \,.\, a^3 = a^5}$; $\mathbf{b^7/b^3 = b^4}$
Pokiaľ máme rôzne exponenty a rôzne základy, tak sa už moc nedá robiť.
Konkrétne umocňovanie:

$$\mathbf{2^0 = 1,\ 2^1 = 2,\ 2^2 = 4,\ 2^3 = 8,\ ...,\ 2^{10} = 1024,\ atď.}$$

Inverznou operáciou k umocňovaniu je **odmocňovanie**. Entú odmocninu z **x** zapisujeme:

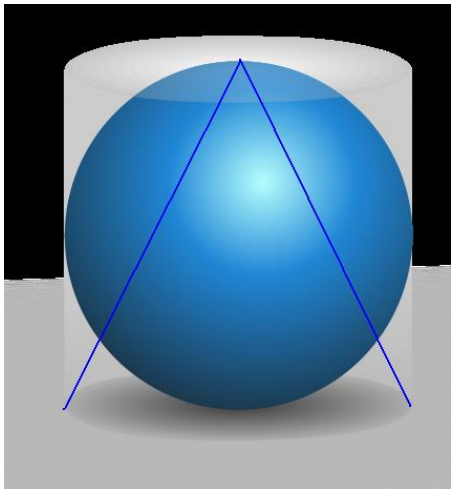
$$\sqrt[n]{x}\ alebo\ x^{\frac{1}{n}} \qquad [I.14]$$

Pre najčastejšiu druhú odmocninu nepíšeme v znaku 2: $\sqrt{x}$. Ako sa počíta? Keď máme $3^2 = 9$, tak $\sqrt{9} = 3$. Na výpočet odmocnín používame často úspešne kalkulačku. Keďže odmocninu môžeme písať ako mocniny so zlomkovým exponentom, teda z množiny **Q**, ich aritmetika (sčítanie, odčítanie, násobenie, delenie) je podobná ako u mocnín s prihliadnutím na prácu so zlomkami (pozri príklad I.7.).

Uviedli sme veľmi stručne základy elementárnych matematických operácií s číslami. Pre naše spoločné potešenie ich môžeme rôzne miešať, kombinovať, čím vzniknú omnoho zaujímavejšie konštrukcie, vzorce a výrazy, ktoré si občas vyžiadajú intenzívnejšie zapojenie neurónových okruhov v našich šedých kôrach mozgových. Uveďme si aspoň 3 známe výrazy o umocňovaní dvojčlenov, ktoré uviazli niekde zasunuté v pamäti zo základného vzdelania:

$$(a+b)^2 = a^2 + 2ab + b^2$$
$$(a-b)^2 = a^2 - 2ab + b^2 \qquad \textbf{[I.15]}$$
$$a^2 - b^2 = (a-b)(a+b)$$

V tejto kapitole sa spoločne veľmi jemne a zľahka (aby nám moc neublížila) dotýkame matematiky, ktorú pre jej úsporný spôsob vyjadrenia maximálneho obsahu môžeme nazvať poéziou vedy. Trpezlivý čitateľ vie, že každá snaha vyžaduje určitú námahu, vo vede to platí mnohonásobne viac. Už Platón vo svojej *Academii*, ktorá je predobrazom dnešných vysokých škôl, výskumných centier a iných inštitúcií, mal svoje ústredné heslo "Bez znalosti geometrie vstup zakázaný!" Platón vedel, že nie každý, kto vstupuje na jej pôdu, bude matematik. Ale vedel aj to, že keď sa diskusii, dialógu medzi ľuďmi nedajú základy logiky a vzájomnej úcty, bude to vyzerať ako v našom parlamente. V Antike si kráľ Ptolemaios povolal na svoj dvor slávneho starogréckeho učenca Euklida a požiadal ho, aby ho rýchlo, stručne a jasne, bez zbytočnej straty cenného kráľovského času, do tej populárnej a zaujímavej matematiky zasvätil. Iný extrém bol a doteraz je, že matematické vedomosti vzbudzujú v najrôznejších formách strach, niekedy až hrôzu a poverčivosť v mnohých ľuďoch, tak ako sa to stalo vojakovi rímskej légie Marka Marcellusa pri dobytí Syrakúz, keď objavil Archimeda riešiaceho nejakú zložitú a určite veľmi zaujímavú geometrickú konštrukciu, ako si maľuje do piesku geometrické obrazce. "Nedotýkajte sa mojich kruhov!" - boli posledné Archimedove slová predtým, než ho legionár napriek výslovnému rozkazu veliteľa zabil mečom.



Obr. I.3. Obrazec z Archimedovho náhrobku

Hrôza z človeka, ktorého mozog a jeho vynálezy niekoľko rokov umožnili Syrakúzam brániť sa obrovskej prevahe rímskych légií, mala silu imaginárneho strachu pred kúzlami a mágiou. V tom sa ľudia do dnešných čias príliš nezmenili. Je zaujímavé, že keď sa v detstve dievča Sophie Germainová dozvedela o tejto príhode, veľmi ju zaujalo, aká musí byť matematika úchvatná vec, keď zanietenie ňou je silnejšie ako hrozba smrti! A stala sa z nej vynikajúca matematička začiatku 19. storočia, napriek tomu, že jej okolie ako žene v tomto príliš naklonené nebolo. [5]

Pr. I.9. Na Archimedovom náhrobku bol vraj geometrický obrazec, podľa ktorého jeho hrob objavili po niekoľkých storočiach, podobný ako na obr. I.3. Čím je zaujímavý? Obal je valec, ktorého výška **v** je rovná priemeru podstavy **d = 2r**, kde **r** je polomer kružnice podstavy. Vo vnútri je kužeľ s rovnakou podstavou ako valec a výškou **d = 2r** a guľa s polomerom **r**. Keď si dáme do pomeru ich objemy, ktoré nájdeme medzi základnými geometrickými vzorcami (valec $\mathbf{V_v}$, guľa $\mathbf{V_g}$, kužeľ $\mathbf{V_k}$) a trochu si započítali (môžete sa pokúsiť!) dostali by sme:

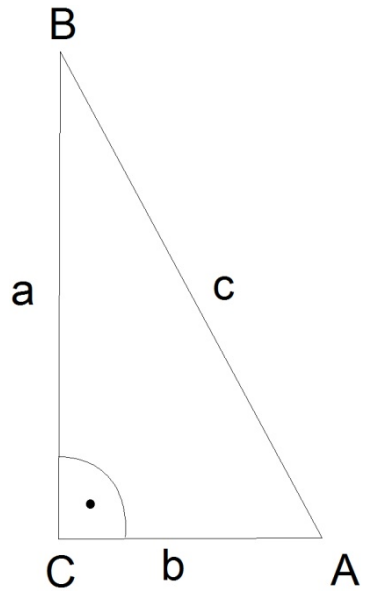
$$V_v : V_g : V_k = 3 : 2 : 1$$

To už je celkom pekný objav aj pekný inotaj pre náhrobok!

Čo teda matematik vlastne robí? Keď povieme, že hľadá vzťahy medzi rôznymi štruktúrami za podmienok extrémnej abstrakcie, príliš sme nezmúdreli. Matematici sú väčšinou normálni ľudia, s istou schopnosťou abstrakcie, estetického cítenia, a možno podľa rôznych psychotestov majú aj niektoré schopnosti, mnohí však so sklonom k lenivosti, preto vymýšľajú všelijaké postupy, počtárske algoritmy a neviem čo ešte, aby si prácu podstatne zľahčili. Či tomu veríme alebo nie, naozaj to nie sú ľudia, ktorí sa snažia druhým skomplikovať život, práve naopak. A že z toho dostaneme niečo dosť zložitého, nuž čo už s tým narobíme! Ale hlavne, čím sa matematik zaoberá nie je počítanie. Je to hľadanie nejakých vzťahov medzi pre niekoho dosť abstraktnými objektmi, ako sú napríklad aj čísla. Takýto vzťah sa snažia veľmi presne a úsporne vyjadriť a potom príde to najkrajšie, logicky to dokázať. Hľadanie dôkazu nejakého tvrdenia, matematickej vety, je hlavnou pracovnou náplňou matematika. Keď ho nájde, môže sa toto tvrdenie potom použiť v práci aj pri iných ako matematických problémoch. Uveďme si pre zábavu (začíname mať zvláštne záujmy, že?) príklad dôkazu tvrdenia jednej notoricky známej vety, ktorá je vtĺkaná do hláv žiakov už takmer dve a pol tisícročia a potešme sa prekrásnym Pythagorovým geometrickým dôkazom jeho slávnej vety:

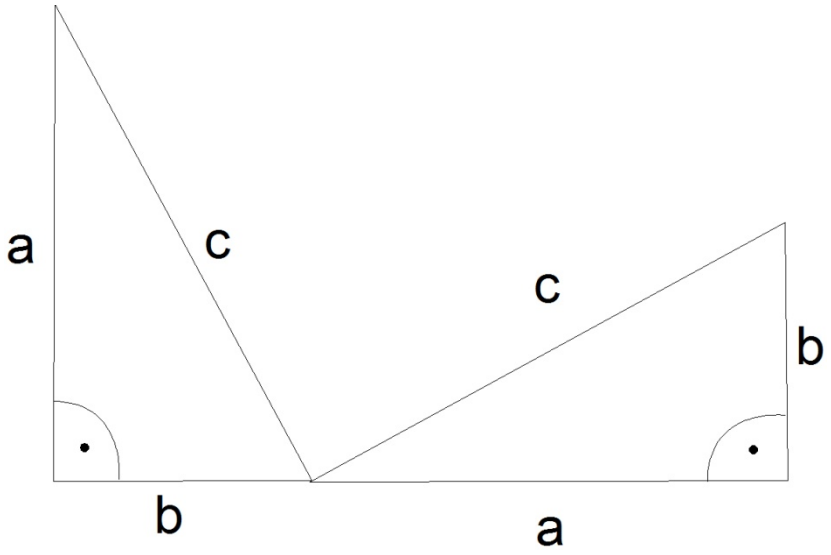
Pr. I.10.

1. Majme ľubovoľný pravouhlý trojuholník (obr. I.4), s odvesnami **a, b** a preponou **c**.



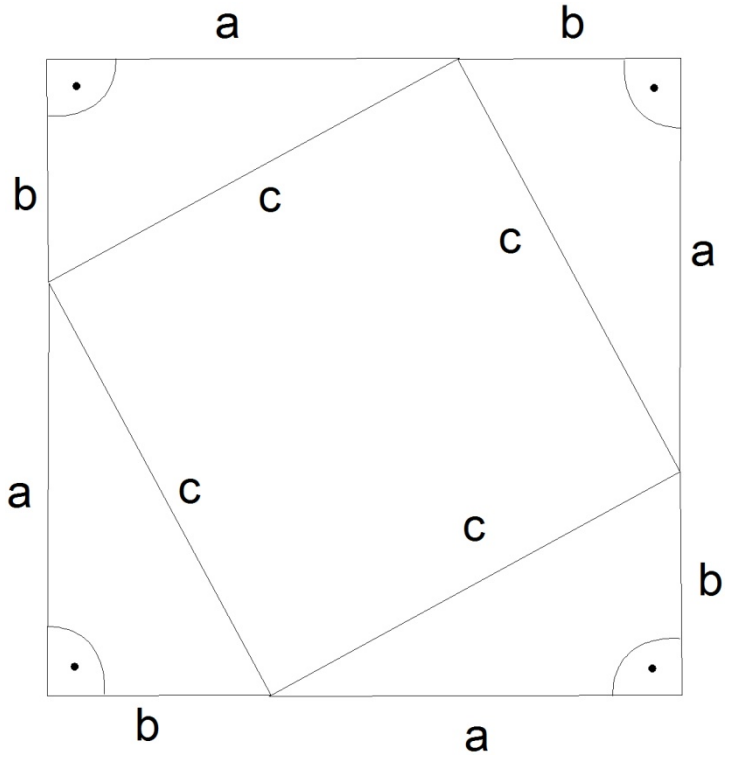
Obr. I.4.

2. Priložme k nemu rovnaký trojuholník tak, aby odvesna **b** druhého trojuholníka bola pokračovaním odvesny **a** prvého trojuholníka, teda vytvoria spoločnú úsečku dĺžky **a** + **b** (obr. I.5).



Obr. I.5

3. Ak rovnakým spôsobom pridáme rovnaký pravouhlý trojuholník ešte dvakrát, dostaneme obrazec, načrtnutý na obr. I.6:



Obr. I.6.

Vznikli vlastne dva štvorce, jeden veľký so stranou **a+b**, druhý menší so stranou **c**. Predstavme si na chvíľu obdĺžnik so stranami **a** a **b**, rozdelený uhlopriečkou **c**, čím vzniknú dva „naše“ pravouhlé trojuholníky. Obsah obdĺžnika je **a.b**, (objavili sme to už v príklade I.4.) teda obsah pravouhlého trojuholníka je $\frac{a.b}{2}$. Obsah veľkého štvorca možno vyjadriť ako súčet obsahu malého štvorca a obsahov štvorcov 4 pravouhlých trojuholníkov:

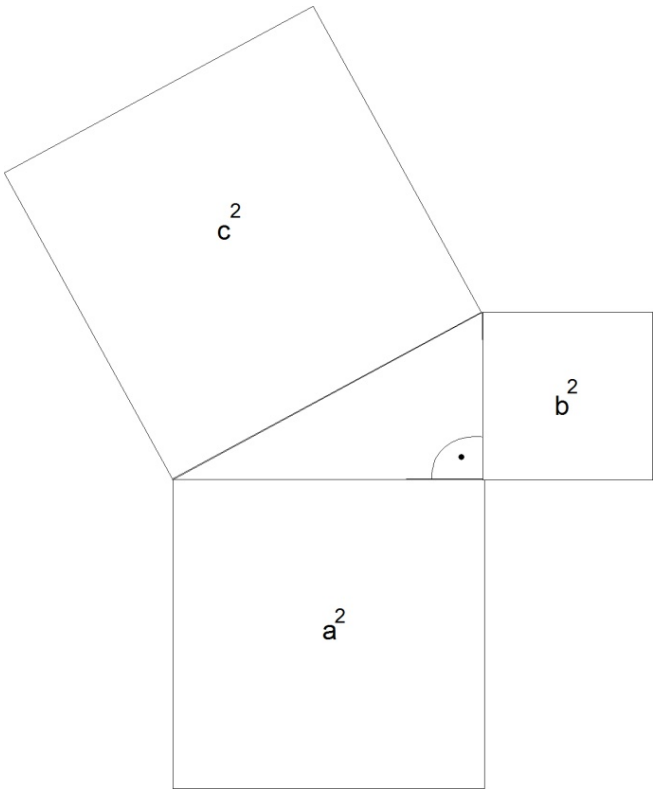
$$(a+b)^2 = c^2 + 4.\frac{ab}{2}$$

Umocnením ľavej strany pomocou 1.vzorca vzťahov [I.15.] a vykrátením 4 a 2 na pravej strane dostaneme:

$$\mathbf{a^2 + 2ab + b^2 = c^2 + 2ab}$$

Jednoduchou úpravou, odčítaním výrazu 2ab od oboch strán rovnice dostaneme:

$$\mathbf{a^2 + b^2 = c^2}$$



Obr. I.7. Pythagorova veta

Slovne: **Obsah štvorca nad preponou pravouhlého trojuholníka sa rovná súčtu obsahov štvorcov nad jeho odvesnami.** Q.E.D.

Bolo to ľahké, že? Ale malo to nesmierny význam:

1. Asi najväčším prínosom je, že ste sa naučili sami urobiť matematický dôkaz, o čom sa vám na začiatku kapitoly ani nesnívalo.

2. Dôkaz bol robený pre ľubovoľný, teda obecný pravouhlý trojuholník a je nástrojom k tomu, že teraz už viete, že Pytagorova veta naozaj platí pre všetky pravouhlé trojuholníky v rovine v celom vesmíre. Nemusíte tomu už len veriť. Je to absolútna pravda odvodená logickým postupom, nevyvrátiteľná, platná dve a pol tisícročia, ale aj navždy, pre všetky možné prípady a pritom vyjadrená tak úsporne, minimalisticky.

3. Dôkaz Pytagorovej vety spojuje abstraktný svet matematiky s reálnymi objektmi sveta. Na jej základe bolo vyriešených nespočítateľné množstvo matematických, geometrických, vedeckých a technických i architektonických problémov.

Traduje sa, že keď Pythagoras objavil dôkaz tejto vety, ilustrovanej na obr.I.7, obetoval gréckym bohom ako bývalo zvykom 100 vykŕmených volov. Odvtedy, ako dodal Jan Werich, sa vždy, keď múdreho človeka napadne myšlienka, voly trasú.

Zastavme sa ešte na chvíľu pri pojme, ktorý sme použili o niečo vyššie a ktorý každý intuitívne pozná – množina. Zhruba pred storočím sa im venovala zvýšená pozornosť, bolo potrebné nanovo definovať ich základy, zaviedli sa pojmy ako spočítateľnosť množiny, jej mohutnosť označovaná prvým písmenom hebrejskej abecedy alef $\aleph$, napr. mohutnosť množiny celých čísel je $\aleph_0$ (čítaj alef nula). Tak ako v matematike bolo 19. a 20. storočie dobou

velikánov matematiky, hlavne ich prelom, potom prišiel čas velikánov z oblasti úradníctva. Boli to bohatierske časy. Matematici, ktorí naliehavo ako vždy žiadajú urýchlenú zmenu, v tomto prípade chceli modernizáciu vyučovacieho procesu v oblasti matematiky. Asi oprávnene. Poukazovali na rôzne výdobytky a pokroky v oblasti matematiky, ktoré sa vôbec nepremietli do výučby. Výsledkom procesu prechodu tejto požiadavky cez ministerské štruktúry bola najslávnejšia školská reforma v našich lúhoch, hájoch a hvozdoch, oproti ktorej sú tie nasledujúce už len slabým čajovým odvarom. Zmenu logického myslenia požadoval ministerský výnos. Teda politickým rozhodnutím za podmienok, keď moc slúžila hlavne sama sebe, bola celá spoločnosť hodená do rybníka množinovej matematiky, aby sa každý naučil v nej plávať alebo sa utopil. Prebehli najprv rýchlokurzy pre tých, ktorí mali celý proces začať odovzdávať ďalej, potom nasledovali školenia, prednášky a zácviky pre ostatnú populáciu. Boli kurzy pre učiteľov, ale aj pre rodičov, pre opatrovateľky v predškolských zariadeniach i pre stredoškolských profesorov, povinné školenia pre štátnu správu a záujmové prednášky pre umelcov a intelektuálov, ktorí chcú byť vždy na tepe doby. V procese ďalšieho vzdelávania zdravotníckych pracovníkov sa tiež našiel priestor so zvláštnym dôrazom pre zdravotné sestry; na svoje si mohli prísť aj seniori. Nikto nemohol byť k tejto reforme ľahostajný. Vonku zúrila už dlhšie beatlesománia a deti kvetov sa schádzali v San Franciscu alebo vo Woodstocku. K nám keď vtedy zavítal nezainteresovaný pozorovateľ, nevychádzal z údivu. Všade sa debatovalo len o množinách. Rozhovor v kaderníctve:

„Pani Nováková, vaša dcéra je taká šikovná, zvládla už zobrazenie na množinu?“

„Samozrejme!“ – dcéra pani Novákovej vždy všetko zvládla. Hluk sušiacej prilby jej trochu skreslil otázku, tak odpovedala na základe pojmov, ktoré sama vstrebala:

„Filoménka robí už zobrazenia nielen nad množinu, ale aj pod množinu, dokonca včera sa priznala, že už sa pokúsila aj o niečo vedľa množiny.“

Milenci na lavičke si do srdiečka vyrývali nie **J love you**, ale **Alef you**. Náročnejšie debaty sa často zvrtli od mohutnosti a spočítateľnosti množín na logické paradoxy typu:
„Klamem.“
A následnej dlhej debaty, či klamem alebo hovorím pravdu. Populárnejší bol holičský paradox: „V jednom meste holí holič všetkých ľudí, ktorí sa neholia sami. Holí holič sám seba?“ A vždy dodali, že toto predstavuje revolúciu v matematike, logickom myslení a vo svete vôbec, pričom každý takýto intelektuál sa vyhrieval v žiari svojich obdivovateliek ako Che Guevara modernej

matematiky, chýbal mu už len kalašnikov. Z doteraz neznámeho dôvodu nedopadala vtedy na nás spŕška Fieldsových medailí udeľovaných za výnimočné pokroky v matematike ako paralela Nobelovej ceny, ktorá sa za matematiku nedáva. Traduje sa, že toto bolo obdobie, kedy mnohí matematici v pude sebazáchovy začali pracovať trochu spoločensky izolovane, väčšinou v jaskyniach, nedostupných stržiach a na pustých ostrovoch.

A ministerskí úradníci na svojich pracovných výjazdoch prísne kontrolovali, či sa nová množinová logika uchytila. Neočakávane prepadávali aj bezbranné osamotené vidiecke školy, kde na základe toho, do čoho stúpali cestou medzi ministerskými limuzínami a školou, zadávali deťom rafinovane koncipované úlohy typu:

Množinu husí, ale aj kačíc, moriek, sliepok a šijacích strojov, aby deti mohli z množiny vylúčiť prvky, ktoré do nej logicky nepatria, uzavreli na miestnom futbalovom ihrisku. Deti, v tom čase už dostatočne ponorené do netušených zákutí množinovej matematiky, ako sa ten proces z nejakého dôvodu nazýval, postupne vylúčili najprv husi, potom morky a postupne všetku hydinu, aby im zostalo mierne surrealistické zátišie opustených šijacích strojov na prázdnom dedinskom futbalovom štadióne. Prízemnejší jedinci s menšou imagináciou a nerozvinutým estetickým cítením nakoniec vyčiarkli aj štadión.

Vráťme sa teraz k nášmu rýchlokurzu. Výlet do krajiny čísel a elementárnych operácii s nimi, zavŕšime ešte trochu menej známym pojmom kombinačného čísla, ktorý nám bude tiež užitočný. Zaoberá sa nimi časť matematiky, nazvaná kombinatorika. Pre kombinačné číslo $\binom{n}{k}$, čítame **n nad k** a pozor, nie je to omyl, že tam chýba zlomková čiara, platí, že je to múdro povedané kombinácia k-tej triedy z n-prvkov bez opakovania:

$$C_k(n) = \binom{n}{k} = \frac{n!}{(n-k)!\,.k!} \qquad [I.16]$$

kde vystupujú nám už známe faktoriály celých čísel $\mathbf{n,k} \in \mathbf{N}$, pričom $\mathbf{k} \leq \mathbf{n}$.
Platia vzťahy:

$$\binom{n}{1} = n; \quad \binom{n}{0} = \binom{n}{n} = 1 \qquad [I.17]$$

Pr.I.11. Výpočet hodnoty konkrétnych kombinačných čísel podľa [I.16]:

$$\binom{10}{1} = 10; \quad \binom{5}{2} = 10; \quad \binom{40}{5} = 658008$$

Pokiaľ by ste si chceli vsadiť v lotérii, kde sa náhodne ťahá 5 čísel zo 40 na istotu, teda aby ste určite vyhrali 1.cenu, museli by ste podať $\binom{40}{5}$ tiketov, teda 658008. Pokiaľ by ste za jeden tiket museli zaplatiť 3 €, tak by ste museli vložiť takmer 2 milióny na istú výhru 1.ceny, ktorá by bola napr. 10 000 €. Nevýhodné, že? Veľkým uspokojením by nebolo, že by ste ešte vyhrali niekoľko druhých, konkrétne $\binom{35}{1}$ a niekoľko tretích cien $\binom{35}{2}$, s rádovo nižšou výhrou. (To si už ľahko dopočítate a porozmýšľajte, ako sme na to prišli). Teda pravdepodobnosť výhry 1. ceny v tejto lotérii je 1 : 658008, dosť malá. Pokúste sa uživiť niečím iným. Ale je to príklad, ako vypočítať pravdepodobnosť nejakého javu z jednoduchej definície pravdepodobnosti ako pomeru priaznivých výsledkov ku všetkým možným.

Pr.I.12. V komunite 11 alkoholikov sleduje napr. psychoterapeut po štvoriciach vzájomné interakcie, aby vytvoril optimálne 4-členné bunky pri zbavovaní sa závislosti. Koľko štvoríc je možné vytvoriť? Je to tiež kombinácia 4. triedy z 11 prvkov (bez opakovania), teda

$$\binom{11}{4} = 330$$

Kombinatorika sa zaoberá ešte mnohými ďalšími vzťahmi, pre niekoho užitočnými. Sú to napr. kombinácie s opakovaním, variácie, permutácie, atď. V prípade, že na to na našej ceste narazíme, objasníme ich ad hoc.

Nie je možné, aby sme obišli ešte jeden, po čísle asi druhý najdôležitejší pojem v matematike, pojem funkcie, ktorý predstavuje popis nejakých závislostí a javov, pôvodne pozorovaných v bežnom živote. Funkciou popisujeme napríklad závislosť dráhy nejakého telesa od času pri danej rýchlosti, závislosť teploty od množstva dodanej energie, závislosť rastú ceny výrobku od nákladov a miezd pracovníkov, závislosť rastu alkoholizmu od spotreby alkoholu, závislosť výskytu narkománie od veku a mnoho iných.

Historicky sa vráťme opäť do antického Grécka k eleatskej filozofickej škole: Zenónové apórie boli vyjadrením rozporu logického myslenia vrcholiaceho v matematickom popise, ktorý neumožňoval žiaden pohyb a pozorovania svojho okolia. Veď ako môže šíp, vystrelený z luku zasiahnuť svoj cieľ, keď musí najprv preletieť polovicu potrebnej vzdialenosti? A aby preletel

polovicu, musí najprv preletieť polovicu polovice, a tak ďalej, až do ustrnutia pohybu. Alebo ako môže rýchlonohý Achilles dobehnúť desaťkrát pomalšiu korytnačku, ktorá má pred ním náskok? Ak má dobehnúť na miesto, kde sa nachádza korytnačka, tá bude už o niečo ďalej, opäť do nekonečna. I keď sa zdá, že je to komický problém, trvalo ľudstvu niekoľko tisícročí, kým ho vyriešilo. Pracovali na tom mnohí, ale najmä nemecký filozof a matematik Gottfried Wilhelm Leibniz a jeho súčasník Isaac Newton zaviedli v matematike tzv. kalkulus, ktorý nás bezprostredne nezaujíma, ale museli preštudovať logické základy pojmu funkcia. Opäť veľmi múdro povedané, v prípade funkcie jednej reálnej premennej, je to istá množina usporiadaných dvojíc, ktorá priraďuje prvkom podmnožiny definičného oboru (D) prvky podmnožiny oboru hodnôt (H). Všeobecný vzťah pre funkciu (diskrétnu čiže bodovú, alebo spojitú) je:

$$\mathbf{y = f(x)} \qquad \mathbf{[I.18.]}$$

Hodnoty premennej **x**, nazývanej nezávisle premenná, sú z definičného odboru funkcie D a k nej sú funkčným predpisom priraďované vždy hodnoty závisle premennej **y** z oboru hodnôt funkcie H. Konkrétna i-tá hodnota **x** a konkrétne z nej vypočítaná i-tá hodnota **y** tvoria spolu usporiadanú dvojicu $[x_i, y_i]$, kde index **i** naznačuje, že ide o i-tú dvojicu z danej množiny. Dá sa to uskutočniť nielen výpočtom, ale napr. aj graficky.

Predpokladám, že teraz tomu už nerozumieme nikto, preto si radšej uveďme nejaký príklad.

Pr.I.13. Predstavme si, že máme cyklistu, ktorý sa pohybuje rovnomerne priamočiaro tak, že za každú hodinu prejde dráhu **s = 10 km**. Závislosť jeho dráhy od času si môžeme zobraziť graficky, ako je na obr. I.8.

Obr. I.8: Závislosť dráhy od času pri rovnomernom priamočiarom pohybe

Je to slušný cyklista. Ľahko popíšeme jeho jazdu a všetko, čo sa dialo počas nej. Jeho okamžitá rýchlosť **v** v každom bode je rovnaká ako jeho priemerná rýchlosť za celý čas bicyklovania:

$$v = \frac{s}{t}$$

Keď si to chceme napísať ako funkciu (funkčnú závislosť), vyjadríme si ju ako závislosť prejdenej dráhy od času, matematicky:

$$\mathbf{s = f(t)}$$

a konkrétne v našom prípade

$$s = v.t$$

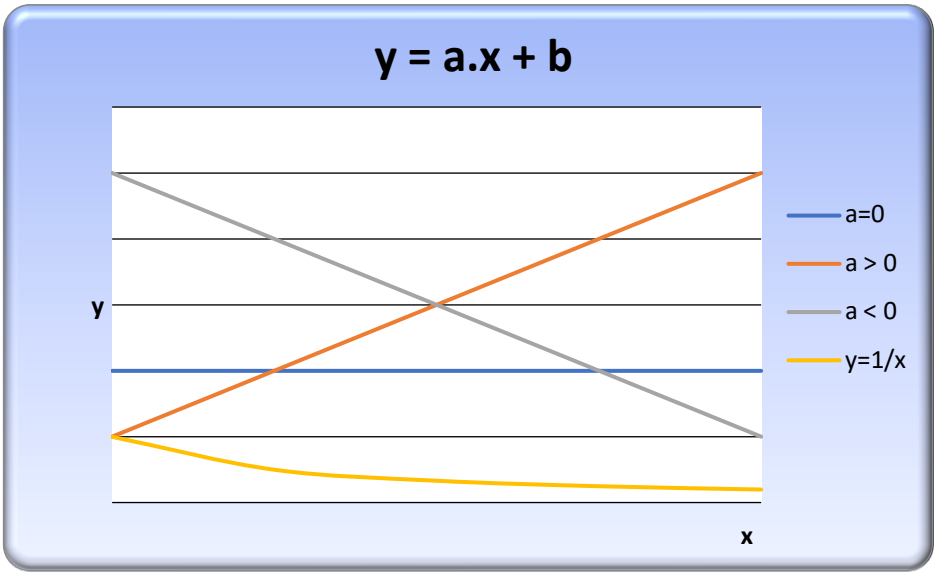
Funkčný predpis je v tomto prípade taký, že každej hodnote času bicyklovania **t** priradíme hodnotu prejdenej dráhy **s** za tento čas tak, že ju vynásobíme konštantou **v**, ktorá predstavuje rýchlosť pohybu cyklistu.

Leibniz a Newton dokázali tento postup zovšeobecniť aj pre cyklistu, ktorý sa rozhodol robiť nám problémy, jazdí si ako sa mu zachce, raz rýchlejšie, inokedy pomalšie, zastavuje, zrýchľuje a vôbec, jeho jazda je tak neznesiteľne komplikovaná, že sa už pomaly a isto blíži reálnemu pohybu. Priemernú rýchlosť $\mathbf{v_p}$ môžeme vypočítať z celkového času jazdy **t** a celkovej dráhy **s**, ktorú prešiel

$$v_p = \frac{s}{t}$$

To nám ale nič nehovorí o procese jazdy. Prešiel celú dráhu za prvých desať minút a potom už len oddychoval? Alebo išiel veľkou rýchlosťou trikrát tam a späť? Nevieme. Jej grafický záznam by bol omnoho komplikovanejší, nie taký pekný lineárny ako na funkcii na obr. I.8. Uvedení matematici zistili, že Zenón a eleátska antická škola, keď delila dráhu Achillesa a korytnačky na nekonečne malé úseky, trochu pozabudla deliť na takéto malé úseky aj čas, za ktorý ich museli prejsť. A vniesli do toho pojem rýchlosti, vypočítanej pomocou sčítania nekonečne malých úsekov (matematici tomu hovoria integrovanie) a tak môžeme prisúdiť každému pohybu (aj Achillesovi aj korytnačke) jeho vlastnú rýchlosť a nemáme problém vypočítať, kde a kedy sa dve telesá s rozdielnymi rýchlosťami stretnú. Podobne pre dráhu šípu vystreleného z luku vieme spočítať okamžitú rýchlosť v každom bode a potom celú dráhu aj keď ju rozdelíme na nekonečne malé úseky. Navyše, do fyziky, matematiky a celej vedy vniesli pohyb, dynamiku, ale to len na okraj.

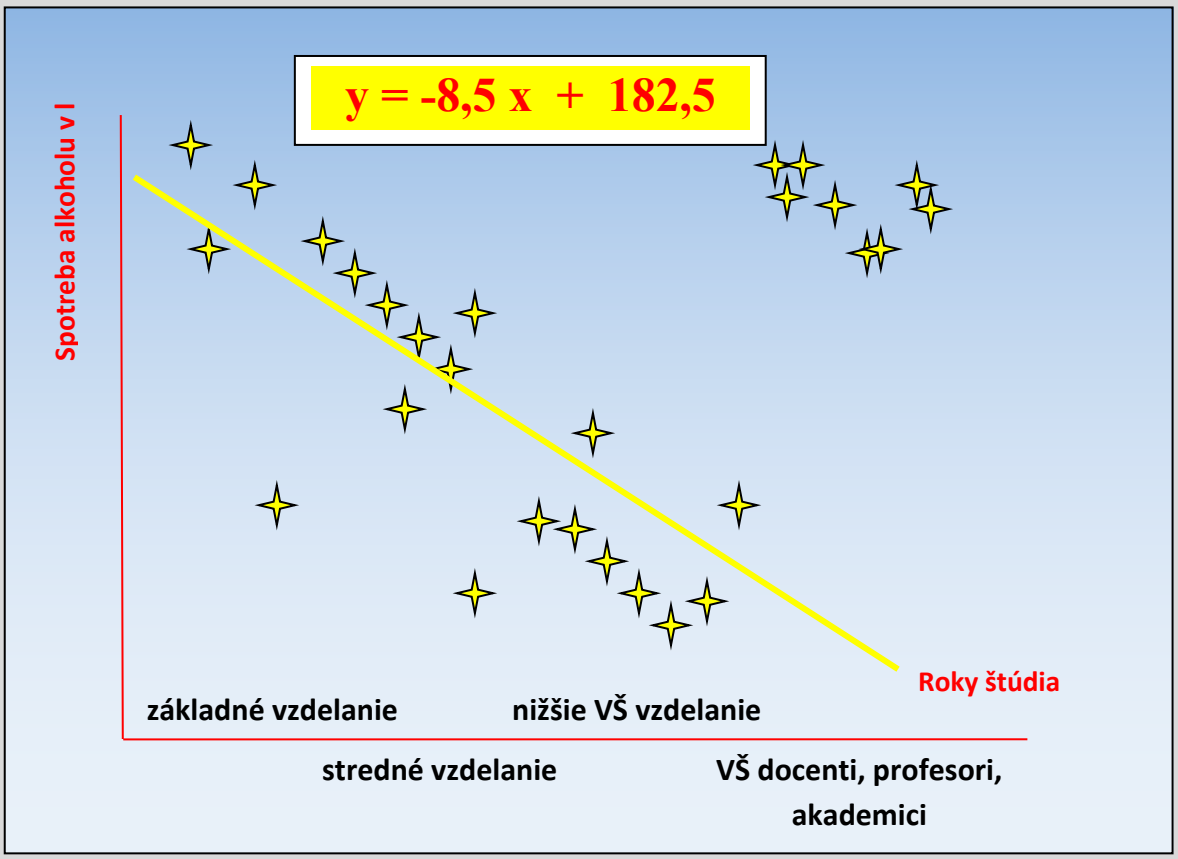
V pr.I.13 sme sa stretli s lineárnou funkciou, kde **y** získame tak, že **x** násobíme nejakou konštantou **a**, pričom nemusí začínať od **0**, ale od nejakej hodnoty **b**. Hovoríme, že **y** je úmerné 1.mocnine **x**. Jej obraz je priamka, ako vidíme vo všeobecnosti na obr. I.9:



Obr.I.9: Obraz (graf) lineárnej funkcie pre rôzne hodnoty smernice a a koeficientu b, pre zaujímavosť aj y=1/x (nepriama úmernosť), teda nelineárny vzťah aj keď je x v 1.mocnine

Pr. I.14: Študent niektorej tzv. pomáhajúcej profesie chce riešiť problematiku alkoholizmu cez vzdelanie a podložiť to argumentmi pre poslancov parlamentu. Budeme sa na chvíľu pohybovať v rovine sci-fi literatúry: Žijeme v ideálnom štáte, ktorý nechce okamžite vyberať

nemalé dane z výroby a spotreby alkoholu, ale myslí do budúcnosti, kde ušetrí ešte väčšiu možno mnohonásobnú sumu na zdravotnej starostlivosti. Ale zákonodarný zbor potrebuje pre svoje racionálne rozhodnutia kvalitné podklady a kvantitatívne výsledky. Veď ide o to, či investovať do vzdelania, alebo neinvestovať. Rozsiahlym a viacnásobne overeným prieskumom sa zistili údaje, ktoré sú zobrazené na obr. I.10, ktorý predstavuje závislosť spotreby alkoholu, prepočítanej na l čistého etanolu za rok, od vzdelania, vyjadreného počtom rokov úspešne ukončeného štúdia (t.j. zo súboru boli vylúčené niektoré dáta, keď respondent má za sebou 18 rokov vzdelávacieho procesu, ale ukončil ho na 1.stupni základnej školy).



Obr. I.10. Závislosť spotreby alkoholu od vzdelania

Ako preložiť v teréne získané údaje optimálnou priamkou sa naučíme neskôr. Ak vás zarazí koncovka, môžete ju zatiaľ dať do odľahlých výsledkov a vylúčiť ich, alebo prijať inú napr. nelineárnu hypotézu a iné závery a odporúčania, podľa toho či máte ešte pred obhajobou diplomovej práce, alebo už po nej. Je to na vás.

Trochu náročnejšia je kvadratická funkcia, kde **y** závisí od 2.mocniny **x**:

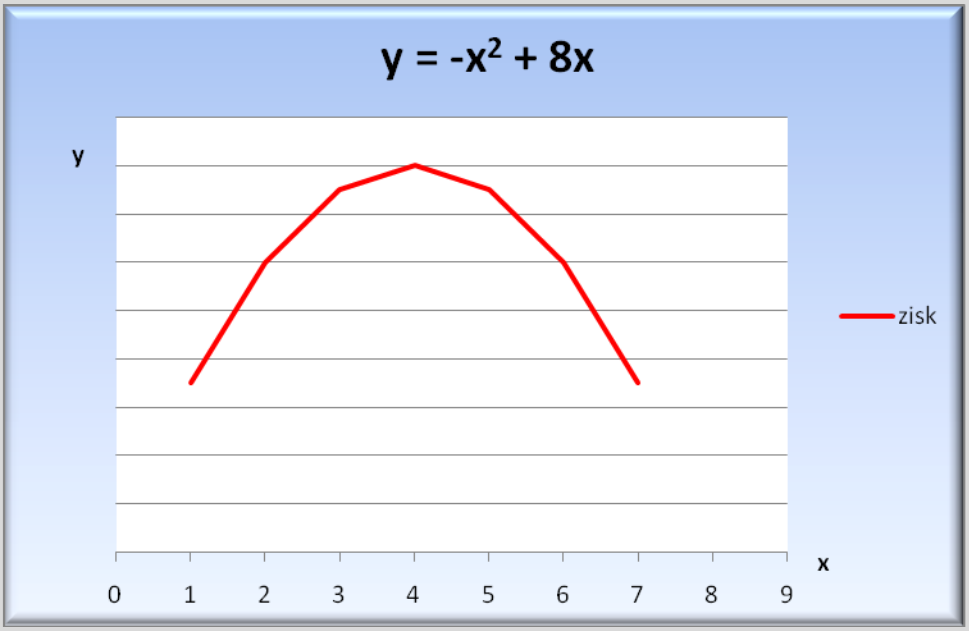
$$y = f(x^2) \qquad \textbf{[I.19.]}$$

Napr. $y = x^2 + 1$, alebo $y = 5x^2 - 3x + 2{,}1$ resp. $y = x^2 + 6x$. Kvadratická funkcia býva často nástrojom optimalizačných problémov, teda hľadania najvhodnejšej hodnoty alebo intervalu hodnôt. Ale samozrejme použitie má omnoho širšie.

Pr.I.15. Centrum pre zdravotne znevýhodnených žiakov trpí nedostatkom finančných prostriedkov na svoj rozvoj. Napr. vychovávateľ sa rozhodne pre aktívny prístup predajom výrobkov žiakov špeciálnopedagogického centra. Zvyšovaním ceny predávaných výrobkov spočiatku zvyšuje zisk, ale ďalším rastom ceny predaj začne klesať, jednoducho začínajú byť príliš drahé. Aká je optimálna cena, teda maximálny zisk? Analýzou závislosti zisku od ceny dospeje k vzťahu

$$y = -x^2 + 8x$$

ktorej priebeh je na obr.I.11.



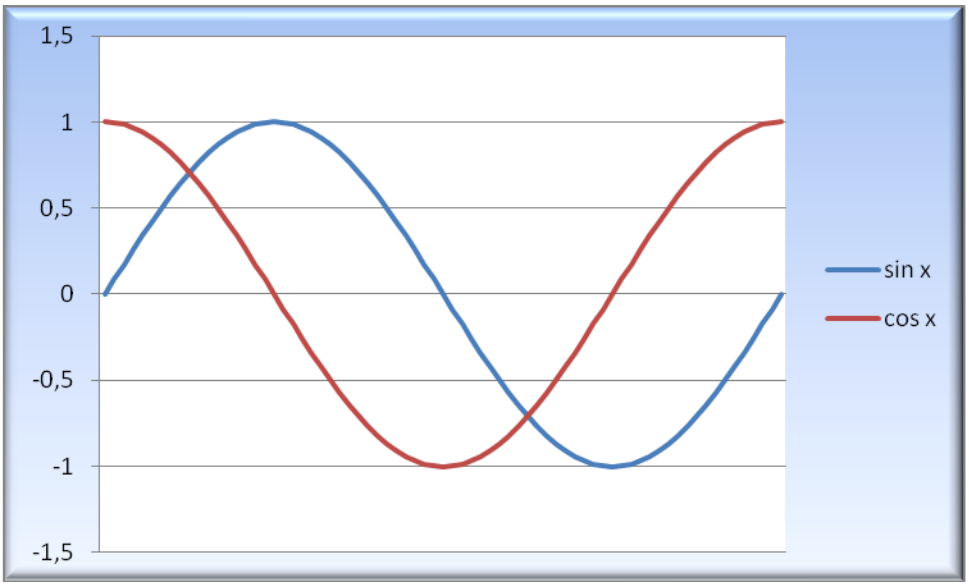
Obr. I.11: Priebeh kvadratickej funkcie vyjadrujúcej závislosť zisku od ceny výrobku.

V tomto prípade optimálnu cenu predávaných výrobkov nájde jednoduchým dosadzovaním. Inokedy je to náročnejšie, ale aj na to existujú elegantné matematické nástroje. Inokedy môže mať funkcia priebeh s minimom, napr. keď príde psychológ pôsobiť do prostredia s vysokou spotrebou alkoholu. Po nejakom čase a nepatrnom poklese tohto negatívneho faktora si vyžiada

pomoc. Zvyšujúcim sa počtom psychológov klesá spotreba alkoholu až do nejakého minima, potom začne opäť rásť. V dôsledku spotreby samotnými psychológmi.

Niektoré cyklické javy, pri ktorých dochádza k pravidelnému nárastu a poklesu nejakej veličiny sa dajú popísať goniometrickými funkciami

$$y = \sin x; \qquad y = \cos x \qquad [I.20.]$$



Obr. I.12: Priebeh goniometrických funkcií sin x a cos x

Exponenciálna funkcia má tvar:

$$y = a^x \qquad [I.21.]$$

Nezávisle premenná **x** sa nachádza v exponente pri nejakom základe **a>0**, ktorý má nasledujúci vplyv na jej priebeh:

0<a<1 ... funkcia je klesajúca

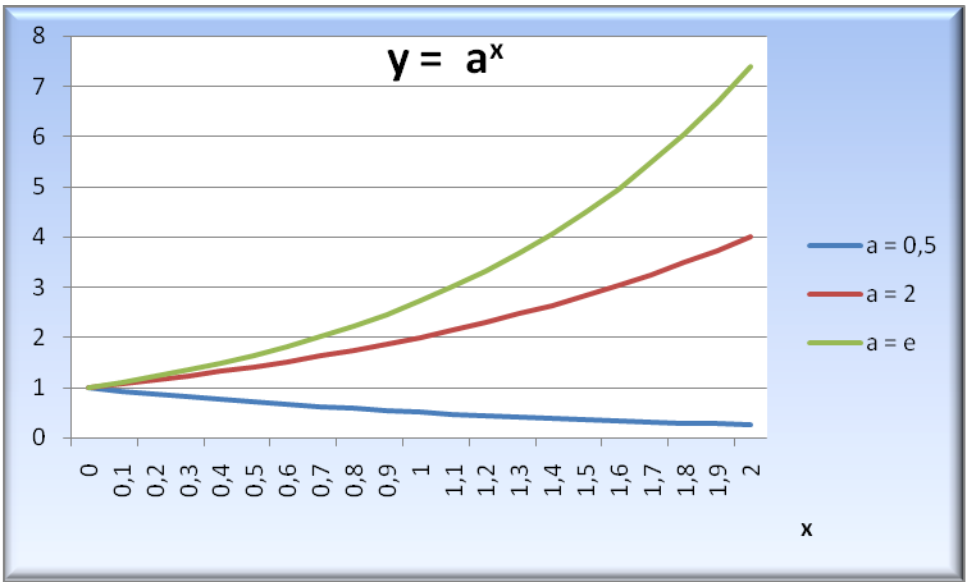
a = 1 ... funkcia je konštanta, y = 1 v celom rozsahu

a> 1 ... funkcia je rastúca

Význačné postavenie má základ exponenciálnej funkcie **a = e**, kde **e** je **Eulerovo číslo**, ktorého hodnotu sme uviedli vo vzťahu **[I.4.]**. Z nejakého dôvodu sa číslo **e** často vyskytuje v mnohých vedeckých, technických, ekonomických a iných výrazoch a vzťahoch, určite sa s ním ešte stretneme:

$$y = e^x \qquad [I.22.]$$

Názornejší ako veľa slov je obrázok:



Obr. I.13: Priebeh exponenciálnej funkcie $y = a^x$ pre vybrané základy a

Na tejto funkcii si môžeme ukázať, ako nás intuícia často klame, využijeme na to jednu starú legendu:

Pr.I.16.: Veľkého padišácha navštívil starý potulný mudrc a predviedol mu zaujímavú bojovú hru, ktorá sa hrala s figúrkami na drevenom štvorcovom podklade s 8x8=64 políčkami. Na počesť vladára ju nazval šach. Padišáchovi sa hra zapáčila a ako odmenu ponúkol starcovi, že si môže priať čokoľvek, čo si jeho srdce zažiada. On ho poprosil, aby mu dal na prvé políčko jedno zrnko ryže a na každé ďalšie z celej šachovnice dvojnásobok predchádzajúceho; teda na druhé dve zrnká, na tretie políčko štyri zrnká ryže, atď.

Padišách s hnevom odvrkol:

„Dajte mu to vrecko ryže, ktoré požaduje a nech ho už nevidím, keď pohŕda mojou veľkorysosťou!“ – asi mal na mysli, že by ho bol býval zlatom vyvážil, keby o to požiadal. Po nejakom čase sa sluhovia vrátili a v rozpakoch padli pred panovníkom na tvár, že nesplnili rozkaz. Bol to osvietený monarcha, dal ich popraviť bez zbytočného mučenia a aby sme boli úprimní, aj mu to trochu vyhovovalo, pretože jeho strážca pokladne mu už šiel riadne na nervy, keď ako papagáj stále opakoval niečo o šetrení a fiskálnej zodpovednosti. Ale, keď bola koňmi roztrhaná už tretia várka sluhov nesplniacich jeho rozkaz, začal sa trochu zamýšľať, čo za tým bude a nechal si to vyložiť dvornými matematikmi. Je to poučné, ako nás môžu mystifikovať

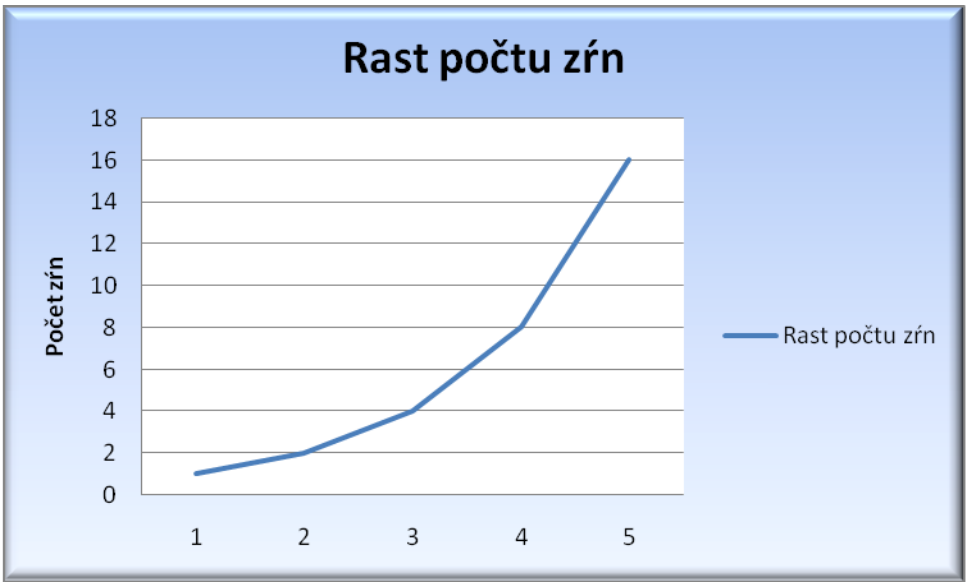
naše myšlienkové stereotypy. Urobme spolu výpočet požadovaného množstva ryže. Môžeme ísť na to dvomi cestami. Prvú môžeme nazvať metóda hrubej sily (skrátene MHS): Budeme počítať od políčka k políčku a sčítavať zrnka až do omrzenia. Ba ešte omnoho ďalej, až do konečného výsledku.

Druhú metódu pomenujeme metóda lenivého čitateľa (skrátene MLČ), ktorá sa bude napodiv blížiť k tomu, ako by postupoval matematik. Z niekoľkých počiatočných výsledkov sa pokúsime nájsť nejakú zákonitosť, vzťah, zovšeobecniť ho a dokázať jeho platnosť a potom urobíme konečný výpočet. Poďme na to, urobme si tabuľku, počítajme spolu:

Číslo políčka n	1	2	3	4	5
Počet zrniek ryže na n-tom políčku	1	2	4	8	16
Počet zrniek ryže na prvých n políčkach	1	3	7	15	31

Zatiaľ boli metóda MHS a MLČ totožné. Metóda MHS pokračuje v rozširovaní tabuľky až na všetkých 64 políčok a postupnom vypĺňaní jej stĺpcov. MLČ sa tu zastaví, zamyslí a položí si otázku: Nie je tu nejaká zákonitosť?

V 1.riadku nie je nijaká záludnosť, **n** sú len poradové čísla políčiek idúcich za sebou. Trochu zaujímavejší je druhý riadok: Keď sme na 1. políčko dali jedno zrnko, na každom ďalšom políčku sa ich počet zdvojnásobí. Zdvojnásobenie obsahu predchádzajúceho políčka už musí byť nejaká zákonitosť. Skúsme si to vyjadriť graficky (obr. I.14.)



Obr.I.14: Grafické vyjadrenie závislosti rastu počtu zŕn od poradového čísla šachového políčka

Takže, keď obsah políčka vynásobíme 2, dostanem obsah nasledujúceho políčka. Ako by sme si vyjadrili obsah n-tého políčka? Vo vzťahu by mala vystupovať **2** aj **n**. Keď skúsime napríklad:

Počet zŕn na n-tom políčku = 2.n

alebo

Obsah n-tého políčka = n^2

skoro zistíme, že nám to nesedí. Nesedí ani pripočítavanie 2, ani jej odčítanie, ani delenie 2. Ani kombinácie týchto operácií. Ako by sme mohli teda pokračovať? Skúsme si postupnosť v druhom riadku tabuľky vyjadriť ako násobky 2:

Miesto	1	2	4	8	16
máme	1	2	2.2	2.2.2	2.2.2.2

a to si môžeme zapísať pomocou mocnín 2 nasledovne:

	2^0	2^1	2^2	2^3	2^4

Na prvom políčku máme $\mathbf{2^0}$ ryžových zrniek, pretože vieme (povedali sme si to vyššie), že akékoľvek číslo umocnené na 0 je vždy 1, na druhom políčku máme $\mathbf{2^1}$ zrniek, na treťom $\mathbf{2^2}$ a vo všeobecnosti na n-tom políčku máme $\mathbf{2^{n-1}}$ zrniek ryže. To je už celkom dobrý výsledok,

použijeme ho v treťom riadku tabuľky, v ktorom je vyjadrený súčet všetkých zrniek na prvom až n-tom políčku. Zapíšme si postupnosť tretieho riadku:

1 3 7 15 31

Opäť si môžete, ak chcete, pomôcť grafom, aby bolo jasné, že nejaká zákonitosť tu je. Zdá sa ale, že sme opäť v koncoch, veď čo s tým? Sú to samé nepárne čísla, takže s 2 nemajú veľa spoločného. A predsa: Skúsme ako by vyzerala táto postupnosť, keby sme ju vyjadrili výrazom 2^n:

n	1	2	3	4	5
2^n	2	4	8	16	32
Naša postupnosť	1	3	7	15	31

Teraz už jasne vidíme, že skutočnosť je vždy o jedno ryžové zrnko menšia ako v predchádzajúcom riadku, ktorý sme odhadli pomocou výrazu 2^n. Tak si môžeme postaviť hypotézu, že súčet zrniek $\mathbf{S_n}$ na prvých **n** políčkach bude:

$$\mathbf{S_n = 2^n - 1}$$

Vyzerá to dobre, múdro, ale platí to naozaj? A pre všetky možné políčka? Či sa nám to páči alebo nie, o tom už nerozhodujeme my, naša intuícia, ale matematický dôkaz, ktorý však radšej prenechajme na hranie matematikom. My mu budeme veriť, veď sme si ho pekne, logicky odvodili. Vieme, že tým ochudobňujeme hlavne pomáhajúcich profesionálov o veľkú časť hĺbky a krásy matematickej kreativity, ale zároveň ich azda udržujeme pri zmysloch.

Počet zrniek ryže, ktoré požadoval potulný mudrc od padišácha je jednoducho $2^{64} – 1$. Kalkulačka nám pomôže, je to 18 446 744 073 709 551 615 ryžových zŕn. Pre také veľké čísla nemáme vrodený zmysel, nemáme to s čím porovnávať, preto ešte neodkladajte kalkulačku a trochu si započítajte. Môžeme urobiť odhad, že na 1 gram ryže potrebujeme 30 zrniek, vypočítajte akú hmotnosť ryže kázal padišách vydať potulnému mudrcovi? Môžete začať v kilogramoch, potom prejdite na tony, prípadne na celé nákladné vlaky s 50 dvadsaťpäťtonovými vagónmi. Ak si urobíme tiež zaokrúhlenie, že hustota ryže sa príliš nelíši od hustoty vody, môžeme si povedať, že 1 kg ryže predstavuje objem asi 1 liter, teda 1 dm^3 a môžeme vypočítať, aký objem by všetka tá ryža zaberala. Niekto bude schopný z hodnoty

rozlohy SR (49 035 km^2 aj s vodami) vypočítať akú jej časť a do akej výšky by ryža zaberala. Zaujímavé, že?

Inverzná funkcia k exponenciálnej funkcii je logaritmická funkcia so zápisom:

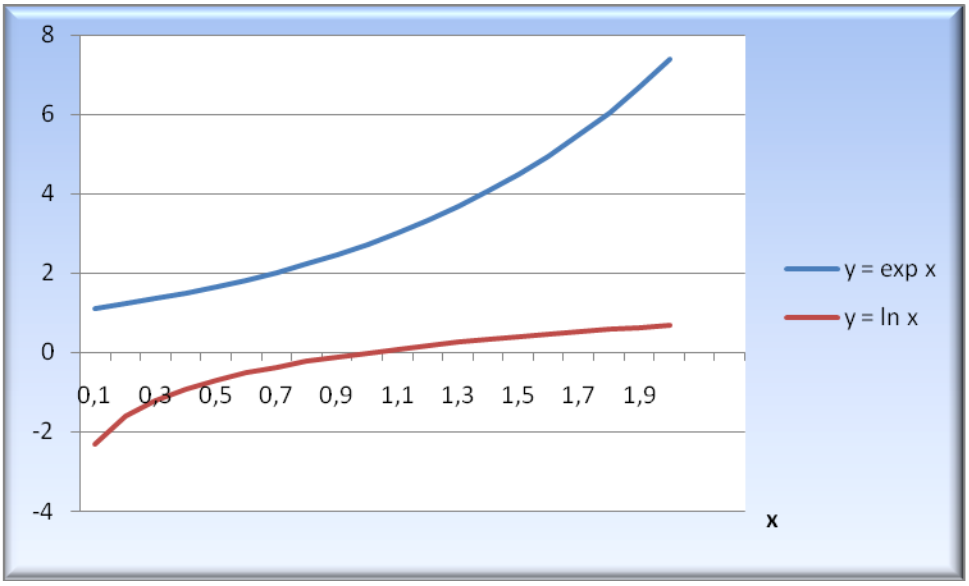
$$y = \log_a x \qquad [I.23.]$$

Základ logaritmu býva často **10,** alebo **e**, vtedy zapisujeme:

$$y = \log x \qquad [I.24.]$$

$$y = \ln x \qquad [I.25.]$$

Ale obrázok, ukazujúci vzťah medzi exponenciálnou a logaritmickou funkciou pre základ **e** bude asi užitočnejší, pre samotné výpočty nám určite bude stačiť trochu lepšia kalkulačka, resp. počítač:



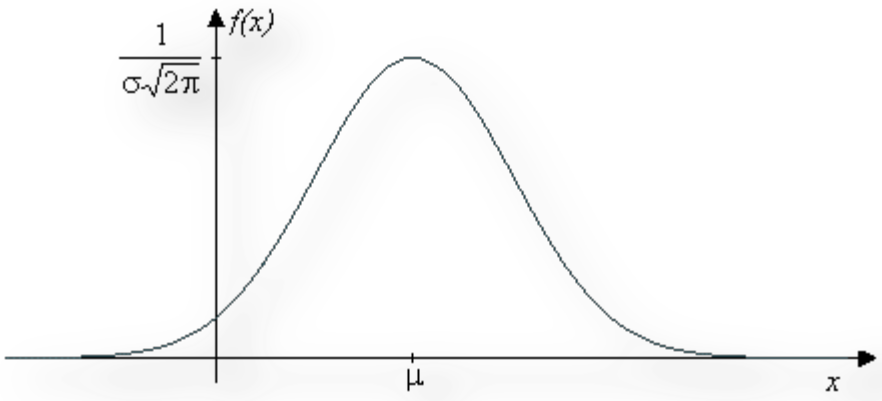
Obr.I.15: Porovnanie priebehu exponenciálnej a logaritmickej funkcie pri spoločnom základe e.

Všetky uvedené funkcie, ktoré sme spomenuli vyššie patria do množiny tzv. elementárnych funkcií. Rôznym skladaním elementárnych funkcií dostávame zložené funkcie, ktoré často vyzerajú veľmi zložito a teda aj veľmi neprívetivo, ale netreba sa ich zľaknúť. Všetko, čo sme doteraz popísali, ste už niekde niekedy počuli, určite ste sa s tým stretli, niekto

viac, niekto menej úspešne na základnej a strednej škole. Trochu zábavy a možnosti hlbšieho zopakovania si učiva ponúka [6] až [10]. Ale aj zložitejšie výrazy, ako napr.:

$$y = \frac{1}{\sigma\sqrt{2\pi}} . e^{\frac{(x-\mu)^2}{2\sigma^2}} \qquad [I.26.]$$

s priebehom



alebo

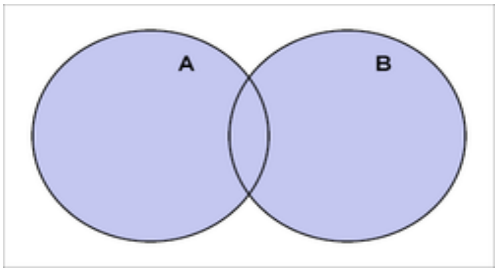
$$y = \binom{n}{x} p^x (1-p)^{n-x} \qquad [I.27.]$$

sú tiež zaujímavé, ale netreba si preto, že im teraz nerozumieme, kaziť náladu. Celá kapitola slúžila na to, aby ste si oprášili základné pojmy a je načase si povedať, že väčšinu výpočtov, aj tých zložitých, teda väčšinu „manuálnej“ práce, dnes už dokáže urobiť osobný počítač a vhodný dostupný software. Nám postačí, keď budeme problému rozumieť. Aby sme si udobrili množinových matematikov, uveďme si malú ukážku, ako jednoducho sa s nimi pracuje:

Majme dve množiny, označme ich **A** a **B**. Ich prvky môžu byť veci, ľudia, kadečo, ale aj rôzne čísla, s nejakými vlastnosťami. Aj s množinami môžeme robiť všelijaké operácie, uveďme si pre zaujímavosť niektoré.

a) Zjednotenie množín A a B. Je to niečo ako ich sčítanie. Vznikne nová množina **M**, ktorá má všetky prvky z množín **A** aj **B**, ak mali nejaké prvky spoločné, zarátavajú sa do spoločnej zjednotenej množiny len raz. Zjednotiť môžeme aj viac množín. Píšeme:

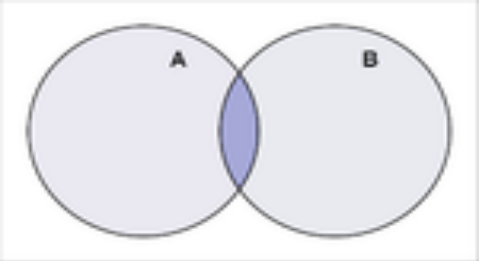
$$\mathbf{M = A \cup B} \qquad \mathbf{[I.28.]}$$



Obr.I.16: Zjednotenie množín A a B do jednej množiny M

b) Prienik množín A a B. V novej množine **N** sa nachádzajú len tie prvky, ktoré sú aj v množine **A** a súčasne aj v množine **B**. Prienik možno nájsť aj pre viac množín. Vzťah sa píše:

$$\mathbf{M = A \cap B} \qquad \mathbf{[I.29.]}$$



Obr.I.17: Prienik množín A a B do jednej množiny

Pr.I.17.: Nech množina **A** obsahuje ako prvky čísla 1, 2, 3, 4 a 5.

A = {1;2;3;4;5} a nech

B = {4;5;11;23;28}

Zjednotenie množín A a B je

M = {1;2;3;4;5;11;23;28}

a prienik množín A a B je

N = {4;5}

To je však dosť abstraktný príklad, uveďme si na čo nám to môže byť dobré:

Pr.I.18.: Sociológ rozbieha projekt vzdelávania vo vylúčených rómskych komunitách a potrebuje na to spolupracovníkov. V prvej osade našiel 7 Rómov s maturitou, v druhej 11.

Množiny si označme **A = {7}** a **B = {11}**. Ich zjednotením **M = A∪B = {18}** získame množinu všetkých potenciálnych spolupracovníkov.

Podrobnejším skúmaním zistil, že v prvej osade sú dvaja Rómovia s maturitou, ktorí majú potrebnú dostatočnú znalosť rómskeho, slovenského a anglického jazyka, v druhej osade traja. Prienikom množín **A** a **B** v oblasti znalosti jazykov si vytypoval vhodných spolupracovníkov: **N = A∩B = {5}**.

Pr.I.19.: Potrebujete sa urýchlene dostať v meste na železničnú stanicu. Ochotný pracovník Dopravného podniku vám vysvetlí, že sa to dá buď električkou č.3, ktorá chodí ako jedna z troch možných, alebo autobusom X, ktorý chodí vo frekvencii ako každý štvrtý. Pravdepodobnosť, že najbližšia električka je pre mňa vhodná, je potom 1 ku 3, t.j. jeden priaznivý prípad z troch možných, teda napíšme si **p(el) = 1/3**. Pravdepodobnosť, že najbližší autobus ma zavezie na stanicu je **p(aut) = ¼.** Akú mám celkovú pravdepodobnosť dostať sa najbližším spojom na železničnú stanicu? S týmito pravdepodobnosťami môžeme pracovať ako s množinami, v tomto prípade zjednotiť ich:

$$p(celk) = p(el) \cup p(aut) = 1/3 + 1/4 = 7/12$$

teda viac ako 58%.

Ale už stačilo. Dostali sme sa ďaleko, ale skončíme nateraz túto našu púť, určite nie ľahkú, odpoveďou Euklida kráľovi Ptolemaiovi na jeho pyšnú a nafúkanú požiadavku:

"V matematike neexistuje kráľovská cesta."

Literatúra k I. kapitole:

[1] Bober, J.: Malá encyklopédia bádateľov a vynálezcov, Obzor, Bratislava, 1973

[2] Zamarovský, Vojtěch: Grécky zázrak, Bratislava, Perfekt, 2002

[3] Derbyshiere, J.: Posedlost prvočísly, Academia 2007

[4] Devlin, K.: Problémy pro 3. tisíciletí, Dokořán 2005

[5] Singh, S.: Velká Fermatová veta, Academia, Praha 2007

[6] Aczel, A.D.: Náhoda–příručka pro hazardní hráče, zamilované, obchodníky s cennými papíry a ostatní, Dokořán 2008

[7] Acheson, D.: 1089 a další parádní čísla, Dokořán 2006

[8] http://www.fch.vutbr.cz/~polcerova/mat1/texty/zaklady.pdf

[9] http://www.studopory.vsb.cz/studijnimaterialy/Zaklady_matematiky/

[10] Rektorys, K. a i.: Přehled užité matematiky, SNTL, Praha 1981

[11] Oláh, Ľ.: Ochrana zdravia pred žiarením, skriptá VŠZaSP sv.Alžbety, Bratislava 2012


Pokiaľ rozumiem vašej kombinatorike, pane, práve ste ma požiadali o ruku.

II. Nadhľad alebo ako sa orientovať bez závratu

Je veľmi nepravdepodobné myslieť si, že všetko je pravdepodobnosť.
Anonym

Stará múdrosť na situáciu, keď nám detaily zatienia celok, vraví: *„Pre stromy nevideli les“*. Často nezostáva nič iné, len vyliezť do vysokej koruny a pozrieť sa na húštinu „zhora“. Lewis Thomas (1913-1993), americký biomedicínsky výskumník a známy popularizátor vedy vo svojej eseji *Spoločnosti ako organizmy* [1] sa zaoberá pohľadom z výšky na správanie nejakej učenej spoločnosti počas svojej výročnej konferencie, na ktorú sa zbiehajú jedinci z celého sveta. Tvoria rôzne dynamické vibrujúce zhluky s nejasnými obrysmi pre neustály pohyb, prerušovaný hyperaktívnymi členmi pobehujúcimi sem a tam, ako keby sa dotýkali, odovzdávali si útržky informácií, neustále splývali a štiepili sa. Občas sa z masy vyčlení bočný výhonok alebo dlhé vlákno smerujúce k bufetu alebo ku konferenčnej sále.

Keď si pozrieme správanie sociálneho hmyzu, napr. niektorých druhov mravcov (ilustračné foto autor), vidíme veľmi podobné nutkavé sociálne správanie. Navyše pestujú huby, chovajú a doja svoj dobytok (vošky), vykazujú silný altruizmus ale aj bojujú vo vojnách, berú a zotročujú zajatcov, pričom používajú chemické zbrane a mnoho iných užitočných a nám blízkych činností. Nevieme, neboli sme v mravenisku, ale máme pocit, že po večeroch potichu pozerajú nejakú mravčiu televíziu. Pozorovanie osamelého mravca (včely, osy a i.) nedáva nejaké zmysluplné výsledky. Pohybuje sa chaoticky, náhodne, ako keby niečo hľadal, možno sám seba. Niekoľko mravcov už dokáže niesť nejaké to sústo, mŕtveho motýľa, cik-cak, predsa ale nakoniec smerom k mravenisku. Od nejakej kritickej hodnoty zoskupenia mravcov, termitov a pod. dochádza skokom k novej kvalite. Začne zmysluplná komunikácia, deľba práce

a jej organizácia, začínajú sa budovať stavby ako úžasné gotické katedrály s neobyčajnými klenbami, akustikou, vetraním, s vysokým stupňom organizácie spoločnosti.

Takéto zoskupenie pôsobí veľmi inteligentným dojmom. Možno sa zamyslieť, čo je vlastne sociálna jednotka, mravec alebo mravenisko?

Obr.II.1. Ilustračný obrázok ľudí okolo obelisku vo Washingtone, USA [2]

Uvádzame toto kontroverzné porovnanie samozrejme úmyselne. Je to užitočné pre každého, kto sa opája extrémnym individualizmom, pocitom absolútnej slobody a nezávislosti od ostatných ľudí, teda v konečnom dôsledku k pocitu silnej nadradenosti a egoizmu, čo z hľadiska humanitných, najmä pomáhajúcich profesií nie je práve tá najvhodnejšia vlastnosť; a dúfame, že ich porovnanie so sociálnym hmyzom dostatočne uráža. Ale priznajme si, že nás to všetkých znepokojuje a nechceme s tým mať nič spoločného, akoby nás to do istej miery zbavovalo vlastnej slobodnej vôle. V slušnej spoločnosti aj vedcov sa takéto úvahy radšej „nenosia“.

Ďalším zámerom je ukázať, ako sa tvoria predsudky. Predpokladom je fakt, že čím menej niečomu rozumieme, tým viac sa k tomu vyjadrujeme. Ako vo filmovej rozprávke *Soľ nad zlato* kráľov radca (Vlasta Burian) odpovedá kráľovi (Janovi Werichovi) na nejakú poznámku: „To si vyprosujem! Jednak tomu nerozumiem, čo si kráľ povedal, a jednak ma to uráža!“ – a zabuchol dvere, teda skončil dialóg [3]. Urazené pobúrenie je druhým potrebným predpokladom. Potom je už ľahké vložiť autorovi do úst čo nikdy nepovedal a bojovne sa proti tomu ohradiť. Napr.: „Čo sme my nejakí faraóni, šváby, či ploštice?“ A toho sa už autor výskumu alebo publikácie tak ľahko nezbaví. Predsudok si žije vlastným životom a každé nové dementi, že autor nepovažuje ľudí za sociálny hmyz, ho len posilňuje. Pretože ako hovorí jeden anonymný bonmot z oblasti pedagogiky, vysvetliť sa dá všetko, ale nie všetkým.

My sme sa spolu rozhodli používať kritické racionálne myslenie okorenené štipkou štatistiky. Preto nás môže potešiť zaujímavý jav, že pri sociálnej interakcii veľkého počtu indivíduí, ktorá nie je vždy len racionálna, ale je založená na vysokom stupni náhodilosti,

v dôsledku najrôznejších vplyvov, dochádza k štatisticky podobným javom a zákonitostiam. Bolo by možné, že takto je stvorený náš Boží svet?

Náhoda v ňom hrá väčšiu úlohu ako si väčšinou pripúšťame. To nie je, prosím, zasahovanie filozofom do ich remesla. Keď sa niečo okolo nás deje, pôsobí na to množstvo často skrytých faktorov a výsledok je neistý. Napríklad to s kým sa dnes stretneme nezávisí len od nášho presného diára, ale od mnohých iných okolností, na ktoré najčastejšie ani nemáme vplyv. A potom niekto z nás zvolá: "No to je ale náhoda, že som ťa dnes stretol! Už tri dni na teba neustále myslím..." Chceme tým naznačiť, že pravdepodobnosť stretnutia s dotyčnou osobou, aj keď sme na ňu mysleli, bola veľmi malá. A predsa sa stala. Pokiaľ by bolo toto zvolanie adresované manželke, s ktorou máme spoločnú domácnosť, malo by asi inú dohru. Nám postačuje, zatiaľ na kvalitatívnej úrovni, že stretnutie s manželkou, s ktorou žijeme viac rokov pod jednou strechou má podstatne vyššiu pravdepodobnosť. A matematik zajasá, pretože sa to dá asi previesť na úroveň čísel, teda kvantifikovať. Repetitio est mater studiorum: Nebudeme sa zaoberať úvahami, či nám manželku dnes do cesty zoslala vyššia moc, ani či je svet deterministický alebo chaotický, to ponechávame povolanejším, ktorí tomu rozumejú. Keď rozhodca pomocou mince losuje, ktoré futbalové mužstvo bude mať výkop, alebo ktorý šachista začína bielymi figúrkami, nekladieme si tieto otázky. Úlohou tejto publikácie nie je vysvetliť, aký svet je, ale dotknúť sa toho, čo dokážeme o niektorých jeho javoch povedať. Prijmeme preto ako fakt, že svet je pravdepodobnostný.

Fyzici o tom už dávnejšie nepochybujú. Na úrovni mikrosveta to berú ako principiálny zákon, paradigmu, teda, že je to charakteristika sveta a nielen v dôsledku našich neznalostí všetkých faktorov vplyvu. Potom popisujú vlastnosti rôznych objektov na tejto úrovni hustotou ich pravdepodobnosti. Vybudovali si na to rozkošný systém, nazvaný kvantová teória a najväčšiu radosť majú, keď ich to privedie k nejakému nezmyslu alebo paradoxu, o ktorom sa potom vášnivo hádajú. Je až udivujúce, že tento systém im dáva veľmi presné výsledky.

Iným príkladom je sledovanie správania veľkého počtu entít na mikroskopickej úrovni z vyššieho makroskopického nadhľadu. Keď máme v nejakej uzavretej nádobe alebo v balóne plyn, jeho čiastočky sa pohybujú chaoticky rôznou rýchlosťou, jedny väčšou iné menšou. Keď ho zohrejeme, teda dodáme mu energiu, priemerná rýchlosť čiastočiek sa zvýši a ňou naráža na steny nádoby či balónu. Navonok, makroskopicky, sa to javí, že sa zvýšil tlak plynu. Balón je pružný, tak sa viditeľne zväčší, niekedy až do prasknutia.

Mnohé javy okolo nás i v nás sú dynamické. Prejavujú sa tým, že sa menia v čase, väčšinou sa pohybujú sem a tam okolo nejakej rovnovážnej polohy. Niekoľko príkladov: Kmitanie atómov okolo rovnovážnej polohy, otáčania planét okolo Slnka, kolísanie krvného

tlaku a iné fyziologické funkcie, biorytmy, verejná mienka, pohyb cien na burze, atď. Aj keď nie sme radi, keď nám nálada klesá pod bod mrazu, môžeme si uvedomiť, že táto možnosť pohybovať sa v širokom variačnom rozpätí emócií a psychických stavov je vlastne schopnosť vyrovnávať sa s meniacimi vonkajšími podmienkami. Pri „pevne zadrátovanom hardvéri“ ako hovoria informatici, by sme sa „spálili“ pri najmenšom výkyve. Týmto fluktuáciám podliehajú aj náhodné javy, ktoré sa pohybujú okolo veličiny nazvanej pravdepodobnosť náhodného javu.

Pravdepodobnosťou pri neúplnom súbore informácií o niečom sa zaoberáme veľmi často, sprevádza naše rozhodovanie. Stretávame sa s ňou aj v Starom zákone: Urim a thummim (hebrejsky svetlo a pravda) boli kamenné žreby resp. lósy na poznanie Božej vôle v najstarších časoch Izraelitov. Veľmi zaujímavá je predpoveď proroka Izaiáša o tom, ako si vojaci budú lósom deliť Mesiášove šaty, čo sa splnilo o 700 rokov na Golgote [4]. Tým sme sa dostali do oblasti hazardných hier, o ktorých správy máme už zo starej Babylónie, Mezopotámie, starého Grécka a Ríma. „Kocky sú hodené!“ – je známy Caesarov výrok, po ktorom prekročil rieku Rubikon a zmenil dejiny, ktoré sa už nedali vrátiť.

Novodobú históriu pravdepodobnosti začali vraj hazardní hráči niekedy v 14. a 15. storočí. Známym sa stal taliansky renesančný lekár a závislý hazardný hráč Girolamo Cardano (1501-1576) vďaka svojej práci o pravdepodobnosti, nazvanej *Kniha o náhodných hrách*, ktorá sa stala na dlhšie obdobie príručkou všetkých hazardných hráčov. Jeho viac-menej subjektívna náuka o pravdepodobnosti obsahovala aj potrebu nie úplne racionálne podloženej šťastnej náhody. A tým sa riadia gambleri dodnes. Lepšie podložená subjektívna pravdepodobnosť je napríklad na dostihoch, kde hodnotíme stav koňa i džokeja, jeho posledné výsledky, súperov a i. Určite si často kladiete aj otázky typu: Cena zlata rastie už tri týždne po sebe. Mám predať všetky zlaté tehličky, uložené doma v skrini? Aká je pravdepodobnosť, že porastie ešte týždeň a keď počkám, dosiahnem ďalší obrovský bezpracný zisk a neobyčajné šťastie? Alebo že zajtra príde po dlhodobom raste k prudkému prepadu jeho ceny a ja sa dostanem do stavu hlbokého zúfalstva, že som prepásol príležitosť?

Klasická matematická teória pravdepodobnosti konca 17. storočia bola dosť protivná, pretože z nej vyhnala pojem šťastnej náhody. Figurujú v nej také mená ako Blaise Pascal (1623-1662), Pier de Fermat (1601-1665), Gottfried W. Leibnitz (1646-1716) a iní, ktorí namiesto šťastnej náhody ukázali, že aj v zložitých

náhodných javoch možno zaviesť nejaký typ zákonitosti. V tomto zmysle je pravdepodobnosť výskytu nejakého náhodného javu p(A) rovná

$$p(A) = \frac{n_A}{n} \qquad [II.1]$$

kde $\mathbf{n_A}$ je počet priaznivých prípadov (elementárnych náhodných javov, ktoré už nemožno zjednodušiť) z celkového počtu prípadov **n**.

Pr. II.1.: Najčastejšie uvádzaný príklad pre všetkých hazardných hráčov je pravdepodobnosť hodenia 6 hracou kockou: **1/6.** Pri dlhodobom hádzaní, teda pre dostatočne veľké **n** by sme zistili, že výsledok sa naozaj blíži k **1/6**. Pravdepodobnosť náhodného javu je konečne nejaké číslo, ktoré je mierou očakávania výskytu javu. Náhodným javom rozumieme opakovateľnú činnosť prevádzanú za rovnakých (alebo približne rovnakých) podmienok, ktorých výsledok je neistý a závisí na náhode.

Rovnakú pravdepodobnosť má hodenie aj iných čísel kocky (pokiaľ nie je falošná), teda

$$P(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$$

Aká je pravdepodobnosť, že pri hode kockou padne napr. nepárne číslo, teda buď 1, alebo 3, alebo 5? Pravdepodobnosti sa sčítajú:

$$p(nepár) = p(1) + p(3) + p(5) = \frac{3}{6} = \frac{1}{2}$$

Trochu presnejšie napísané, javy, že padne 1, alebo 3, alebo 5, ktoré sú nezávislé a nemôžu nastať súčasne, ako sme si uviedli už v závere I. kapitoly vo vzťahu **[I.28.]** a v príklade **I.19**:

$$p(1 \cup 3 \cup 5) = p(1) + p(3) + p(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2} \qquad [II.2]$$

kde sa vyskytol znak zjednotenia ∪. Pre to aby nastali nezávislé javy A a B súčasne, používame zase znak ∩.

Doteraz v tom nebol žiaden problém. Máme napríklad v evidencii nezamestnaných 4 stolárov. V blízkej stolárskej dielni sa uvoľnilo 1 pracovné miesto. Pravdepodobnosť zamestnania stolárov z našej evidencie je pre každého ¼. Ak predpokladáme, že sa rodí zhruba rovnaký počet chlapcov a dievčat, je pravdepodobnosť narodenia dievčatka ½, teda **0,5**.

Ak sa vám podarí dosť pravidelne za mesiac (t.j. 25 pracovných dní) 1 krát prísť do práce načas, pravdepodobnosť, že v ďalšom období prídete neskoro je 0,96 teda 96%. Pravdepodobnosť, že stretnete človeka, od ktorého ste si požičali peniaze v najnevhodnejšiu

chvíľu je rovná 1, zatiaľ čo napriek usilovnému hľadaniu toho, kto vám dlhuje a nie a nie ho nájsť, tam sa pravdepodobnosť stretnutia rovná 0.

Zhrnutie vybraných vlastností pravdepodobnosti

Urobme si v tejto chvíli malé upratovanie. Spresnime a zhrňme si, čo sme sa o pravdepodobnosti dozvedeli, alebo sa musíme doučiť, aby sme s ňou mohli trochu pracovať. Postupujte pomaly, pohrajme sa s tým, používajte príklady. Bude to pre nás na ďalšej ceste užitočné. Nie je to žiaden veľký kumšt, len na to nie sme zvyknutí:

a) Pravdepodobnosť elementárneho alebo aj zloženého náhodného javu **A** je z intervalu

$$P \in \langle 0;1 \rangle \qquad \textbf{[II.3]}$$

b) Pravdepodobnosť opačného javu **nA** (niekedy značíme **A´**, **¬A**) k javu **A** dostaneme

$$P(nA) = 1 - P(A) \qquad \textbf{[II.4]}$$

c) Jav, ktorý nemôže nastať za žiadnych okolností (nemožný jav) má pravdepodobnosť

$$P = 0 \qquad \textbf{[II.5]}$$

d) Istý jav, ktorý nastane v každom prípade, má pravdepodobnosť

$$P = 1 \qquad \textbf{[II.6]}$$

e) **Pravidlo sčítania pravdepodobnosti** t.j. pravidlo pre výpočet pravdepodobnosti zjednotenia dvoch alebo viacerých javov.

Všeobecne pravdepodobnosť zjednotenia dvoch javov **A** a **B** sa rovná súčtu pravdepodobností týchto javov zmenšenému o pravdepodobnosť súčasného výskytu týchto javov (platí pre javy, ktoré sa nevylučujú, ale majú spoločný prienik, použili sme **[I.28], [I.29]** a pr.I.19):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad \textbf{[II.7]}$$

Ak javy **A**, **B** – sú navzájom sa vylučujúce javy, potom podľa **[I.28]**:

$$P(A \cup B) = P(A) + P(B) \qquad \textbf{[II.8.]}$$

pretože **A∩B** je javom nemožným a teda **P(A∩B) = 0** (nemá prienik). Tak sme dostali z **[II.7]** vzťah **[II.8]**. Je to pravdepodobnosť, že nastane buď jav A alebo jav B (vzájomne sa vylučujúce javy)

Pravidlo platí v uvedených tvaroch aj pre **n** náhodných javov.

f) **Pravidlo násobenia pravdepodobnosti** t.j. pravidlo pre výpočet pravdepodobnosti prieniku 2 alebo viacerých javov. Máme tu 2 prípady:

Pre nezávislé javy:

Pravdepodobnosť prieniku dvoch nezávislých javov **A** a **B** (súčasný výskyt) sa rovná súčinu pravdepodobností týchto javov, podľa **[I.29]**:

$$P(A \cap B) = P(A).\ P(B) \qquad \textbf{[II.9.]}$$

Teda v dvoch nezávislých pokusoch nastanú javy A aj B. Pravidlo platí aj pre n nezávislých javov.

Pre javy závislé:

Najprv si povieme, že existuje aj podmienená pravdepodobnosť, t.j. že niečo sa stane, keď sa stane predtým alebo súčasne aj niečo iné. Vysvetlím neskôr na príklade, zatiaľ len toľko, že podmienenú pravdepodobnosť značíme **P(A/B)** resp. **P(B/A).** Pre javy závislé pravdepodobnosť prieniku dvoch javov **A** a **B** sa rovná súčinu pravdepodobnosti jedného z týchto javov a podmienenej pravdepodobnosti javu druhého.

$$P(A \cap B) = P(A)\ .\ P(B/A),$$

alebo $$P(A \cap B) = P(B)\ .\ P(A/B) \qquad \textbf{[II.10]}$$

Javy **A, B** sú závislé, ak: $P(B/A) \neq P(B/nA)$ **[II.11]**

Javy **A, B** sú nezávislé, ak platí: $P(B/A) = P(B/nA) = P(B)$ **[II.12]**

kde **nA** je opačný jav k javu **A** s vlastnosťou **A + nA = 1**, keď nastane len jeden z nich.

g) **Veta o úplnej pravdepodobnosti**

Predpokladajme, že jav **A** sa môže uskutočniť len v kombinácii s niektorým z javov $B_1, B_2, \ldots B_n$, ktoré tvoria úplný súbor javov ($\Sigma p(B_i) = 1$) a teda sú navzájom sa vylučujúce:

Môžu sa vyskytnúť možnosti**:** $A \cap B_1$ alebo $A \cap B_2, A \cap B_3, \ldots A \cap B_n$.

Pravdepodobnosť, že jav **A** sa uskutočnil, sa vypočíta ako pravdepodobnosť zloženého javu:

$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots\ldots\ldots \cup P(A \cap B_n)$, čiže

$$\mathbf{P(A)} = P(A \cap B_1) \cup P(A \cap B_2) \cup \ldots\ldots \cup P(A \cap B_n) =$$

$$= P(B_1) \,.\, P(A/B_1) + P(B_2) \,.\, P(A/B_2) \ldots\ldots\ldots = \sum_{i=1}^{n} \mathbf{P(B_i) \,.\, P(A/B_i)} \quad \mathbf{[II.13]}$$

h) **Bayesova veta**

Zaoberá sa tiež podmienenými pravdepodobnosťami javov. V jednoduchej forme máme dva náhodné javy **A** a **B** s pravdepodobnosťami **p(A)** a **p(B)**, pričom **p(B) > 0.** Podmienenú pravdepodobnosť javu **A** za predpokladu výskytu javu **B** značíme **p(A/B)**. **p(A/B)** súvisí s opačnou podmienenou pravdepodobnosťou **p(B/A)** nasledovne:

$$p(A/B) = \frac{p(B/A).p(A)}{p(B)} \qquad \mathbf{[II.14]}$$

Úplný Bayesov vzťah nadväzuje na vetu o úplnej pravdepodobnosti. Použijeme **[II.13]** analogicky pre $\mathbf{p(B) = \Sigma_i\, p(A_i).p(B/A_i)}$ a dosadíme do **[II.14]:**

$$p(A/B) = \frac{p(B/A).p(A)}{\sum_i p(Ai).p(B/Ai)} \qquad \mathbf{[II.15]}$$

i) Nakoniec si zopakujme ešte prvý vzťah tejto kapitoly pre klasickú pravdepodobnosť:

$$p(A) = \frac{n_A}{n} \qquad \mathbf{[II.1]}$$

A teraz si povedzme, čo sme to spolu porobili!

Bod a) so vzťahom **[II.3]** nám hovorí, že pravdepodobnosť vyjadrujeme číslom od 0 do 1 prípadne v percentách od 0 do 100 %. Napríklad pravdepodobnosť úspešnosti skúšky zo štatistiky 0,2 znamená, že ju zloží asi 20% študentov. Iný príklad: Predstavme si hyperkonzervatívnu, netolerantnú, zaostalú a nerozvíjajúcu sa nemodernú spoločnosť, v ktorej sa rodia stále ešte len chlapci a dievčatá, s populačnou pravdepodobnosťou 0,527 pre chlapcov a 0,473 pre dievčatá. To znamená, že pravdepodobnosť narodenia chlapca je v nej 52,7 % - ná. To neznamená, že vieme predpovedať konkrétny najbližší pôrod, ale vieme odhadnúť koľko chlapcov a koľko dievčat sa narodí napr. z najbližšej tisícky úspešných pôrodov. Na toto budeme často upozorňovať.

Bod b), vzorec **[II.4]** nám hovorí o vzťahu dvoch navzájom sa vylučujúcich javoch, z ktorých jeden musí nastať. Napr. Buď žijem, alebo som mŕtvy.

Bod c), vzťah **[II.5]** predstavuje javy s nulovou pravdepodobnosťou. Napríklad, že cez záhradku vašej podhorskej chalupy, kde široko ďaleko nie sú koľaje, prefrčí Orient expres. Alebo, že preteky v rýchlokorčuľovaní, ktorých sa môže zúčastniť niekoľko stoviek národností, národnostných menšín, rôznych kmeňov a rás, nevyhrá Holanďan.

Bod d) a vzťah **[II.6]** naopak predstavuje javy, ktoré sa naozaj musia stať, napr., že zajtra ráno opäť vyjde slnko, alebo že v prostriedku MHD, ktorým sa musíte mimoriadne unavený trmácať cez celé mesto z práce domov budú všetky sedadlá počas celej cesty obsadené.

Bod e) vzťah **[II.8]** sme si popísali pri kockách v príklade II.1.

Čo predstavuje bod f) vzťah **[II.9]**? Je to napríklad pravdepodobnosť javu, že keď hádžeš kockou 2 krát, hodíš 1 a 6; (alebo naraz dvomi kockami), potom: **p = 1/36**. Pri dvoch súčasných pôrodoch vo vyššie uvedenej čudnej spoločnosti je pravdepodobnosť narodenia 2 chlapcov p = 0,527x0,527 = 0,278.

Bod f) a závislé javy s podmienenou pravdepodobnosťou, teda pre vzťah **[II.10]** výslednú pravdepodobnosť určíme ako súčin pravdepodobnosti jedného javu a podmienenej pravdepodobnosti druhého javu vyskytujúceho sa v súčinnosti s výskytom prvého javu. Znie to veľmi zložito, ale je na to veľa príkladov: Napr. pravdepodobnosť, že sa zamestnáte, keď s istou pravdepodobnosťou zložíte skúšku v rekvalifikačnom kurze; pravdepodobnosť, že máte nejakú zhubnú chorobu, keď vám ju s istou pravdepodobnosťou diagnostikovali; pravdepodobnosť, že kúpite poruchový výrobok pri istej pravdepodobnosti, že sa takýto výrobok môže dostať do predaja atď.

Bod g) a vzťah **[II.13]** pomáhajú pri rôznych testoch kvality výroby, účinnosti liekov a pod.

Bod h) a vzťah **[II.14]** resp. **[II.15]** (Bayes) sú často používané v mnohých prípadoch, keď pracujeme s podmienenými pravdepodobnosťami, ale aj v bioštatistike, pri hodnotení pravdepodobnosti diagnózy, možnosti úmrtia a i.

Ukážme si niekoľko príkladov na výpočet pravdepodobnosti:

Pr.II.2.: Aká je pravdepodobnosť hodenia súčtu 5 dvomi kockami, alebo 21 štyrmi kockami, vo všeobecnosti ľubovoľného možného súčtu $\mathbf{S_n}$ **n**-kockami? Aj keď je zadanie veľmi jednoduché, toto je skôr úloha pre matematikov, aby sa aj oni mali čím zabaviť, keď sa sem dostanú. Pre ostatných uvádzam konečný výsledok bez dôkazu ako ukážku, že výpočet pravdepodobnosti nemusí byť vždy tak triviálne jednoduchý:

$$p(\mathbf{Sn}) = \frac{\sum_{k=0}^{n}\left[(-1)^{k}.\binom{n}{k}.\binom{Sn-6.k-1}{n-1}\right]}{6^{n}} \qquad [II.16]$$

Započítajte si ak máte chuť a mali by ste sa dostať k výsledkom $p(5_2) = 1/9$, $p(21_4) = 5/324$.

Môžete využiť vzťahy **[I.6]**, **[I.9]**, **[I.16]** a **[I.17]** z I. kapitoly a samozrejme **[II.1]** . V ďalšom sa už budeme snažiť vyhýbať kockám. Keď si pozriete web, učebnice pravdepodobnosti a štatistiky a ich zbierky, tak sa vám bude o hádzaní kociek, mincí, či zmysluplnom vyťahovaní rôznych guliek ešte dlho snívať.

Ak javy **A** a **B** nie sú nezávislé, teda že napr. jav **B** nastane súčasne, keď nastane jav **A**, pracujeme často s podmienenou pravdepodobnosťou **p(A/B)**. Vo všeobecnosti je výpočet pravdepodobnosti nejakých zložitejších javov dosť náročná práca. Asi aj tu pomôžu najviac príklady (pozri aj napr. [5]):

Pr.II.3.: V meste máme spolu 6 úradov práce(UP) s evidenciou nezamestnaných remeselníkov a štatistikov podľa tabuľky:

Číslo UP	remeselníci	štatistici
1	2	1
2	4	2
3	1	9
4	1	9
5	4	2
6	4	2

Na ich spoločný nadriadený orgán príde naliehavá požiadavka na jednu pracovnú silu, remeselníka, pričom ako vždy „už včera bolo neskoro“. Niet času na hľadanie v neprehľadných záznamoch, tak je úplne náhodne vybraná 1 osoba z náhodne vybraného úradu práce. Aká je pravdepodobnosť úspechu v zamestnaní, teda, že štatistik nebude musieť robiť remeselníka a naopak?

Riešenie:

Elementárny jav výberu úradu práce má pravdepodobnosti:

$p(A_1) = 1/6$ (UP s 2 remeselníkmi a 1 štatistikom v evidencii je jeden zo 6)

$p(A_2) = 2/6 = 1/3$ (UP s 1 remeselníkom a 9 štatistikmi v evidencii sú dva zo 6)

$p(A_3) = 3/6 = ½$ (3 UP zo 6)

$p(A_1) + p(A_2) + p(A_3) = 1$ (úplná pravdepodobnosť).

p(B) vybratia remeselníka vo všeobecnosti nie je známa, poznáme len pravdepodobnosti podmienené evidenciou remeselníkov v jednotlivých úradoch práce. Pre jednotlivé typy UP s rovnakým zastúpením remeselníkov a štatistikov dostaneme:

$p(B/A_1) = 2/3$ (2 remeselníci z 3 nezamestnaných), $p(B/A_2) = 1/10$, $p(B/A_3) = 2/3$
Výber UP a výber remeselníka sú nezávislé javy, ktoré súčasne nastanú, preto pre jednotlivé typy UP máme pravdepodobnosti podľa **[II.10]:**

$p(A_1/B) = p(B/A_1).\ p(A_1)$
$p(A_2/B) = p(B/A_2).\ p(A_2)$
$p(A_3/B) = p(B/A_3).\ p(A_3)$

Pre celkovú pravdepodobnosť (výber sa uskutoční z jedného UP, teda náhodné javy sa navzájom vylučujú) dostávame podľa **[II.8]:**
$p(A/B) = p(B/A_1).\ p(A_1) + p(B/A_2).\ p(A_2) + p(B/A_3).\ p(A_3) = \sum_{i=1}^{3} p(B/Ai).p(Ai) =$
$= 2/3.1/6 + 1/10.1/3 + 2/3.1/2 = 1/9 + 1/30 + 1/3 = (10 + 3 + 30)/90 = 43/90 = 0{,}478$
Záver: Pri náhodnom výbere je odhad pravdepodobnosti úspechu 47,8 %.

Pr.II.4: [lit. 6,7 a i.] Ak je vo vašom študijnom krúžku 60 študentov, prijali by ste stávku, že je medzi vami minimálne jedna dvojica, ktorá oslavuje narodeniny v rovnaký deň? Študentov je 60, počet dní v bežnom roku 365 (berieme rok, ktorý nie je prestupný), zdá sa, že stávku musíte hravo vyhrať. Zamyslíme sa a pozrime sa na to cez počet pravdepodobnosti:
1.krok: Z 365 dní náhodne vyberáme 60 dní, každý jednotlivý výber každého dňa je nezávislý. Koľko máme všetkých možností? Je to trochu problém, pretože sú to dosť veľké čísla, tak si pomôžme náhradným príkladom: Koľko dvojíc môžeme zostaviť z 3 prvkov a, b, c?
Vypíšeme si ich: [aa], [bb], [cc], [ab], [ac], [ba], [bc], [ca], [cb], a keď si to spočítame, je ich 9, teda 3^2. Ľahko si overíte, že výber dvojíc zo 4 prvkov a,b,c,d má 16, teda 4^2 možností. Podobne by sme zistili, že výber všetkých trojíc z 5 prvkov má 5^3, teda 125 možností a už sa to dá zovšeobecniť: Počet všetkých možností vybrať **k** prvkov zo všetkých **n** prvkov sa v kombinatorike, ktorá je peknou časťou matematiky, nazýva **variácia k-tej triedy z n-prvkov s opakovaním** a píše sa:

$$\mathbf{V'_k(n) = n^k} \qquad \mathbf{[II.17]}$$

Počet možností, koľkými možno vybrať 60 dní roka (n-krát po sebe si v kalendári náhodne vyberáme jeden jediný deň, pričom jednotlivé výbery sú na sebe nezávislé a dni sa môžu opakovať) je $\mathbf{365^{60}}$. To je dosť veľké číslo!

2.krok: Teraz chceme vybrať 60 dní z roka tak, aby sa žiaden z nich neopakoval. Postupne vyberáme 1 deň z 365, teda máme 365 možností, potom 1 deň z 364 atď. Sú to tiež nezávislé výbery, preto podľa **[II.9]** dostanem:

365 . 364 . 363. ... = 365 . (365 – 1) . (365 – 2) . ... . (365 – (k – 1))

To sa dá zapísať ako **365!/(365-k)!**

Takýto výber sa v kombinatorike nazýva **variácia k-tej triedy z n-prvkov bez opakovania** a píše sa:

$$V_k(n) = \frac{n!}{(n-k)!} = n.(n-1).(n-2). \dots (n-k+1) \qquad [II.18]$$

3. krok: $V_{60}(365)$ z výrazu **[II.18]** je počet všetkých priaznivých prípadov, že stávku vyhráte, teda, že neexistuje vo vašom krúžku dvojica s rovnakým dňom narodenín.

$V'_{60}(365) = 365^{60}$ zo vzťahu **[II.17]** je potom počet všetkých možností. Ich pomerom v zmysle klasickej definície pravdepodobnosti **[II.1]** dostaneme hľadanú pravdepodobnosť vašej výhry:

$$p(k) = \frac{365.364.363. \dots (365-k+1)}{365^k} = \frac{365!}{(365-k)!.365^k}$$

Pri koncovej úprave vzorca sme sa trochu pohrali s vlastnosťami faktoriálov v zmysle **[II.18]**. Urobili sme to pre ľubovoľné **k**, aby ste si mohli vzťah aktualizovať na reálnu skupinku ľudí.

4.krok: Keď si dosadíme **k = 60**, máme však problém, také veľké čísla vaša kalkulačka odmietne akceptovať, môžete ju prosiť hlboko do noci, neustúpi. Poďme na to teda postupne. Dosaďme za **k** a upravme krok po kroku:

k =1 $\quad p(1) = \frac{365!}{364!.365} = \frac{365!}{365!} = 1$

k = 2 $\quad p(2) = \frac{365!}{363!.365.365} = \frac{365.364.363!}{363!.365.365} = \frac{365.363!}{365.363!}.\frac{364}{365} = 1.\frac{365-1}{365} = p(1).\frac{365-1}{365}$

k = 3 $\quad p(3) = p(2).\frac{365-2}{365}$

atď.

k = n+1 $\quad p(n+1) = p(n).\frac{365-n}{365}$

napr.

k = 60 $\quad p(60) = p(59).\frac{365-59}{365}$

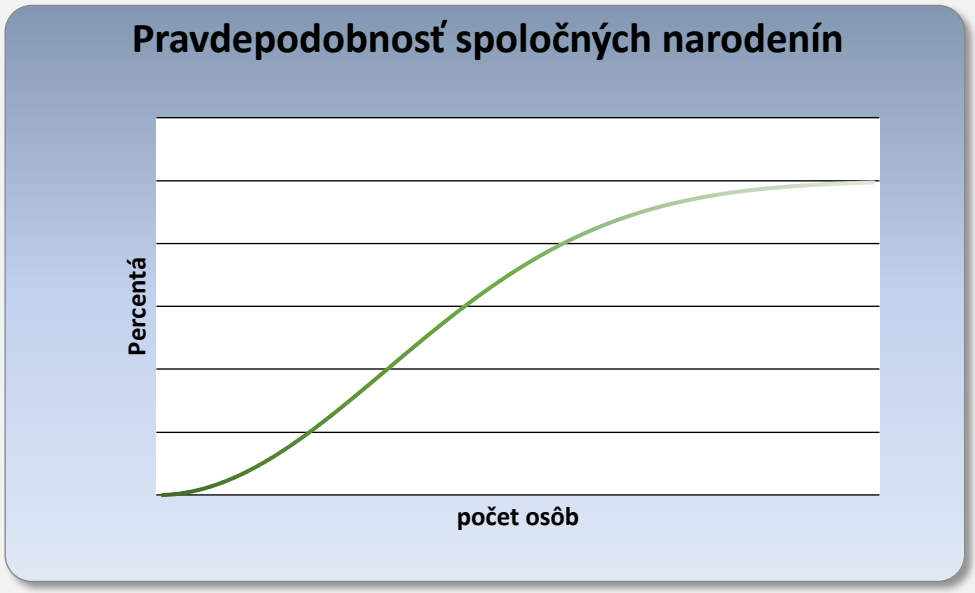
To sa už dá s kalkulačkou za dlhých zimných večerov vypočítať, ale dajme hneď na začiatku priestor pre prácu vášho osobného počítača, on sa s tou protivnou rutinou vyrovná hravo. Napr. pomocou programu MS EXCEL sa môžeme pre prípad vašej výhry dopočítať

$$\mathbf{p(60) = 0{,}005877}$$

Málo, že? Pravdepodobnosť našej výhry v stávke je

$$\mathbf{1 - p(60) \cong 99{,}41\ \%}$$ vyjadrené v percentách.

5. krok: Jeden vhodný obrázok je cennejší ako tisíc slov, preto si budeme často výsledky zobrazovať. Aj v našom príklade je zaujímavá závislosť pravdepodobnosti výskytu viacerých narodenín v jednom dni od počtu osôb v skupine:



Obr. II.2 Grafické zobrazenie závislosti pravdepodobnosti výskytu spoločných narodenín od počtu osôb v skupine - výsledkov príkladu II.4.

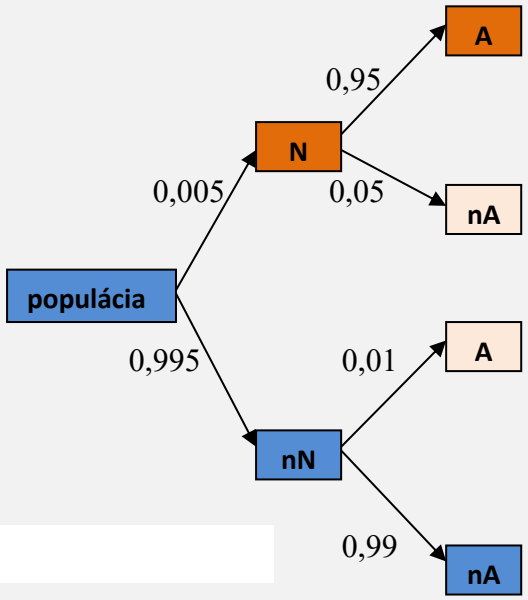
Pr.II.5: Úderná skupina policajtov z *protidrogového* spolu so zohratou partiou streetworkerov mali v miestnom drogovom prostredí dobre zabehnutú spoluprácu. Rešpektovali ich aj narkomani, díleri sa im vyhýbali, kde-tu sa stal nejaký zločin pod vplyvom drog, kde stačilo odobrať krv, poslať do laboratória, odkiaľ za nejaký čas prišli výsledky a všetko bolo jasné. Zdalo sa, že krehkú rovnováhu nič nemôže narušiť. Nanešťastie nejakou zvláštnou, málo pravdepodobnou svojvôľou osudu sa do tímu dostal štatistik. V čom bol problém?

Štatistik tvrdil, že laboratórium potvrdí prítomnosť pervitínu v krvi na hladine významnosti 0,05 a že to je akreditovaný postup. Keď videl hlboko chápavé prikyvovanie kolegov dodal, že to znamená, že je 95%-ná pravdepodobnosť pozitívneho výsledku analýzy, ak zadržaná osoba pervitín v krvnom obehu naozaj mala. Ale zároveň môže metodika stanoviť

pozitívny výsledok asi 1 % osôb, ktoré nikdy pervitín neužívali. Taktiež zdôraznil známy fakt, ktorý potvrdili aj policajti, že 5 ľudí z 1000 starších ako 15 rokov v bežnej populácii má s pervitínom pravidelnú skúsenosť. Aká je teda pravdepodobnosť pri pozitívnom výsledku analýzy, že zadržaný jedinec bol pri čine naozaj pod vplyvom drogy?

Odkedy pribudol v tíme štatistik, v teréne postupne nastalo zvláštne napäté ticho, akoby v oku hurikánu. Preto sa trochu nudili a nočný čas trávili v bufete nižšej cenovej kategórie, takže štatistik mohol začať na zafŕkané papiere kresliť niečo, čo sám nazýval **rozhodovací strom.** Vyzeralo to asi takto:

Označenie: **N** – narkoman, osoba s prítomnosťou pervitínu v organizme; **nN** – čistá osoba
A – výsledok analýzy pozitívny; **nA** – výsledok analýzy negatívny



p(A∩N) = 0,005.0,95 = 0,00475

p(nA∩N) = 0,005.0,05 = 0,00025

p(A∩nN) = 0,995.0,01 = 0,00995

p(nA∩nN) = 0,995.0,99 = 0,98505

Štatistik napriek všetkému, čo si o ňom kto myslel, bol veľkorysý človek a začal od podlahy, celou populáciou. Rozvetvil ju v prvom kroku na *narkomanov* a *čistých* s príslušnou pravdepodobnosťou ich začlenenia: p(N) = 0,005 t.j. 5 z 1000; p(nN) = 0,995, teda väčšina populácie, ktorá si zatiaľ život bez pervitínu bez problémov užíva. Mimochodom, je jasné, že je to úplná pravdepodobnosť: p(N) + p(nN) = 1 (100%), alebo p(nN) = 1 – p(N).

Druhé vetvenie obsahuje už nám známe podmienené pravdepodobnosti, teda, že niečo sa stane, keď sa stane predtým alebo súčasne aj niečo iné. Konkrétne výsledok analýzy pre rôzne začlenenie osôb populácie. Sú tu potom podmienené pravdepodobnosti p(A/N), p(nA/N), p(A/nN) a p(nA/nN), v každej vetve je to úplná pravdepodobnosť kde ich súčet = 1. p(A/N) =0,95 a p(A/nN) = 0,01 ako sme si hovorili na začiatku, teda, že pozitívny výsledok analýzy pre *narkomana* s pervitínom v krvi má 95%-nú pravdepodobnosť a že 1% *čistých* analytická metóda stanoví pozitívny výsledok. Ostatné si dopočítame do 1.

Štatistik si uvedomil, že v bufete všetko stíchlo a začalo sa zamýšľať, kam vlastne smeruje. Donášači odbehli za bossmi podsvetia, ktorí po ich informovaní húfne opúšťali charitatívne dobročinné akcie a nasledovaní pozornými vysokými politikmi spolu zanechávali v žiari reflektorov svoje uslzené trblietavé manželky a ponáhľali sa promtne previesť väčšinu disponibilných finančných prostriedkov do bezpečnejších daňových rajov.

Tak pokračoval zvýšeným hlasom: Ak teraz chceme určiť pravdepodobnosť, že prebehlo dané začlenenie a zároveň sa dostavil výsledok testu, máme tu dva nezávislé náhodné javy, ktoré súčasne nastali, teda podľa vzťahu **[II.10]** nám stačí vynásobiť pravdepodobnosti uvedené na príslušnej vetve. Napr. že narkoman má v krvi pervitín a zároveň mu vyšiel negatívny výsledok analýzy je **p(nA∩N) = p(nA/N) . p(N) = 0,05.0,005 = 0,00025**. Všetky možné výsledky sú v poslednom stĺpci vedľa rozhodovacieho stromu. Elegantne s nimi môžeme pracovať. Aká je napríklad pravdepodobnosť pozitívneho výsledku analýzy **p(A)** ?

Obr.II.3. Anglický reverend a matematik Thomas Bayes (1701-1761)

Jednoducho, použijeme pritom všetko čo vieme zo vzťahu **[II.8]**:

p(A) = p(A∩N) + p(A∩nN) = 0,00475 + 0,00995 = 0,0147

Vráťme sa teda k pôvodnej otázke: Aká je pravdepodobnosť pri pozitívnom výsledku analýzy, že zadržaný jedinec bol pri čine naozaj pod vplyvom drogy? Hľadáme p(N/A), ktorá sa nedá jednoducho z rozhodovacieho stromu odčítať.

Všetci cítili, že sa štatistik blíži do finále a videli, že sa postavil na stoličku, aby ho bolo počuť aj v najodľahlejšom kúte. Našťastie to vyriešil reverend Thomas Bayes svojou slávnou vetou, kde podmienenú pravdepodobnosť nejakého náhodného javu odvodil z opačného javu.

Štatistik zabudol dodať, že sa tak stalo v 1.polovici 18.storočia a nechal donášačov aj policajtov, aby si ukradomky zapisovali neznáme meno. Možno to zapísať vzťahom, pokračoval, podľa **[II.14]** :

$$p(N/A) = \frac{p(N \cap A)}{p(A)}$$

V našom prípade je $p(N \cap A) = 0{,}00475$, $p(A) = 0{,}0147$, teda **$p(N/A) \cong 0{,}32313$**. Slovom: Pravdepodobnosť toho, že osoba, ktorej analýza vyšla pozitívne, naozaj v krvi pervitín mala, je len asi 32,3% !

Rozbúrená drogová scéna sa čoskoro upokojila, pretože mafia si doviezla svojich odborníkov, ktorí hravo vypočítali, že pre biznis je najlepším riešením štatistika jednoducho zastreliť.

Pr.II.6: Dvoch priateľov z jednej firmy (jeden s prezývkou Inžinier pracuje v riadiacej funkcii, druhého prezývka Zvárač vystihuje aj jeho pracovné zaradenie) poslala pracovná zdravotná služba na pravidelnú rutinnú lekársku prehliadku, v rámci ktorej im urobili rtg snímku pľúc. Ich predstava príjemného horúceho odpoludnia pri orosenom zlatistom horkastom moku v známom blízkom pube sa rozplynula, keď im obom oznámila primárka pozitívny rtg test na karcinóm pľúc. Inžinier sa náhodou nedávno dočítal, že tieto testy majú hladinu významnosti 0,95, teda že v 5% prípadov môžu byť pozitívne aj keď pacient rakovinu nemá. Taktiež tam bolo uvedené, že test je v 30 % falošne negatívny, teda že vyhodnotí pacienta, že rakovinu nemá a on ju má. Dlhú chvíľu si mlčky premietali v pamäti film svojho života, rodinu, priateľov, minulosť i nádejnú budúcnosť, ale keďže boli viac mužmi činu ako hlbokých meditácií, príliš dlho ich to nebavilo. Inžinier vytiahol svoj iphone, pripojil sa na web a niekde vydoloval výsledok, že pracovníci v jeho zaradení zomierajú na rakovinu pľúc v priemere asi 1 zo 600 prípadov úmrtí. Pre zváračov typu jeho priateľa bola táto štatistika vyššia, asi 2 prípady zo 100. To už chcelo kus papiera a pero. Najprv si zaznačili v jazyku počtu pravdepodobnosti, čo už vedeli:

T – test pozitívny, **nT** – test negatívny

R – rakovina prítomná, **nR** – rakovina neprítomná

$p(R_1) = 1/600 = 0{,}00167$ pre Inžiniera a **$p(R_2) = 2/100 = 0{,}02$** pre Zvárača. K úplnej pravdepodobnosti sa ľahko dopočítali: **$p(nR_1) = 1 - p(R_1) = 0{,}99833$** pre inžiniera a **$p(nR_2) = 1 - p(R_2) = 0{,}98$** pre zvárača.

Podmienená pravdepodobnosť **$p(T/R) = 0{,}7$** (v 70% prípadoch je test pri prítomnosti rakoviny pozitívny) a **$p(T/nR) = 0{,}05$** (v 5% prípadov môžu byť pozitívne aj keď pacient rakovinu nemá).

Inžinier mal Bayesa v malíčku, preto ho bez zaváhania načarbal:

$$p(R/T) = \frac{p(T/R).p(R)}{p(T/R).p(R) + p(T/nR).p(nR)}$$

Pre Inžiniera vyšlo:

$$p(R/T) = \frac{0{,}7.0{,}00167}{0{,}7.0{,}00167 + 0{,}05.0{,}99833} \cong 0{,}0229$$

a pre Zvárača:

$$p(R/T) = \frac{0{,}7.0{,}02}{0{,}7.0{,}02 + 0{,}05.0{,}98} \cong 0{,}2222$$

Zistili, že pravdepodobnosť Inžiniera, že má rakovinu pľúc je len asi 2,29 % a Zvárača asi 22,22%. Zhodnotili situáciu, že v živote sú omnoho väčšie nebezpečenstvá, taktiež pripustili fakt, že Inžinier je na tom lepšie, veď je vedúci. A spokojne išli na pivo.

V posledných dvoch príkladoch II.5. a II.6. sme využívali vzťahy dôstojného pána reverenda Thomasa Bayesa **[II.14]**; **[II.15]**. Keďže v mnohých výpočtoch a prognózach majú svoje nezastupiteľné miesto, skúsme sa na ne pozrieť obecnejšie, aby sme s nimi dokázali ľahšie pracovať. Vzťahy **[II.14]** a **[II.15]** sú totožné, matematicky v menovateli je vždy súčet, ktorý dáva úplnú pravdepodobnosť. V jednoduchom prípade máme nejaký alternatívny (dichotomický) jav napr. „muž - žena“, „živý - mŕtvy“, zamestnaný – nezamestnaný“ a pod. S každým stavom je spojený nejaký alternatívny znak, napr. „fajčiar – nefajčiar“, „zdravý – chorý“ „áno – nie“ a pod. Aby sme si sprehľadnili údaje, dá sa to usporiadať do tzv. kontigenčnej resp. asociačnej tabuľky 2x2:

B / A	B^+	B^-	spolu
A^+	a	b	a+b
A^-	c	d	c+d
spolu	a+c	b+d	n=a+b+c+d

Tabuľka II.1: Kontigenčná tabuľka 2x2

Pr.II.7: Na škole máme 100 maturantov, z toho 60% chlapcov, pričom fajčí z nich každý desiaty. Medzi dievčatami je 80 % fajčiarok.
Ako by vypadala kontigenčná tabuľka 2x2 ?

fajčiar / rod	áno	nie	spolu
Dievča	32	8	40
Chlapec	6	54	60
spolu	38	62	n=100

Môžeme si zaviesť označenie A_1 – dievčatá, A_2 – chlapci, $n = A_1 + A_2 = 100$
$p(A_1)$ – pravdepodobnosť nájdenia dievčaťa pri náhodnom výbere spomedzi maturantov = 0,4

p(A_2) – pravdepodobnosť nájdenia chlapca pri náhodnom výbere spomedzi maturantov = 0,6

p(B/A_1) – podmienená pravdepodobnosť, že fajčiar je dievča = 0,8

p(B/A_2) – podmienená pravdepodobnosť, že fajčiar je chlapec = 0,1

Dopočítaním do 1 pomocou úplnej pravdepodobnosti si môžeme zapísať

p(B/nA_1) – podmienená pravdepodobnosť, že nefajčiar je dievča = 0,2

p(B/nA_2) – podmienená pravdepodobnosť, že nefajčiar je chlapec = 0,9

1.bunka tabuľky je počet dievčat, ktoré fajčia, vypočítali sme to nasledovne

n. p(A_1). p(B/A_1) = 100.0,4.0,8 = 32

2.bunka tabuľky je počet dievčat, ktoré nefajčia, vypočítali sme to nasledovne

n. p(A_1). p(B/nA_1) = 100.0,4.0,2 = 8

3.bunka je súčet všetkých dievčat

4.bunka tabuľky je počet chlapcov, ktorý fajčia, vypočítali sme to nasledovne

n. p(A_2). p(B/A_2) = 100.0,6.0,1 = 6

5.bunka tabuľky je počet chlapcov, ktorý nefajčia, vypočítali sme to nasledovne

n. p(A_2). p(B/nA_2) = 100.0,6.0,9 = 54

6.bunka je súčet všetkých chlapcov

7.bunka je súčet všetkých fajčiarov, 8. súčet všetkých nefajčiarov, 9.bunka je **n**

Dúfam, že je dostatočne jasné, ako sme sa sem prepracovali. Teraz použime paralelne Bayesa aj kontigenčnú tabuľku na otázky, ktoré nás v tejto súvislosti napadnú:

Napr. za kríkom cez prestávku na uzavretom školskom dvore vidno len tenisky a hustý cigaretový dym. Je to dievča?

Sme opäť pri Bayesovi **[II.15]**

$$p(A/B) = \frac{p(B/A).p(A)}{\sum_i p(Ai).p(B/Ai)}$$

kde pre **i=2**

$$p(A1/B) = \frac{p(B/A1).p(A1)}{p(B/A1).p(A1) + p(B/A2).p(A2)} \qquad \textbf{[II.19]}$$

Do čitateľa **[II.19]** dosadíme hodnotu 1.bunky kontigenčnej tabuľky, do menovateľa súčet 1. a 4.bunky, **n** sa v zlomku vykráti, tak je to to isté, ako keby sme dosadzovali len pravdepodobnosti. Výsledok:

P(A_1/B) = 0,8421

Výsledok dáva odpoveď na našu otázku, s akou pravdepodobnosťou je fajčiaci študent za kríkom dievča.

Z kontigenčnej tabuľky II.1. je táto pravdepodobnosť

$$p = \frac{a}{a+c}$$

Alebo nás viac zaujíma výber jedného chlapca do potápačského družstva. Členovia musia mať silné pľúca, takže nesmú fajčiť. Aká je pravdepodobnosť takéhoto výberu?

Vzhľadom na položenú otázku Bayesa **[II.19]** prepíšeme nasledovne

$$p(nA2/B) = \frac{p(B/nA2).p(nA2)}{p(B/nA2).p(nA2) + p(B/A2).p(A2)}$$

Výsledok je ako sa môžete presvedčiť

$$p = \frac{d}{d+c} = 0,9 \quad t.j. 90\%$$

A môžeme zistiť ešte ďalšie veci, napr. náhodný výber nefajčiara (dievča alebo chlapca) do speváckej súťaže, alebo výber chlapca fajčiara na vyšetrenie obštrukčnej bronchitídy v súvislosti s fajčením a i. Vždy si k tomu pripravíme príslušný Bayesov vzťah a odpoveď si ľahko vypočítame.

Pr.II.8: V kolektíve 30 výskumných pracovníkov Likérky so 70% zastúpením žien má tretina z nich problémy s kontrolovanou konzumáciou alkoholu. Absolvent humanitných vied môže mať rôzne otázky na nejakú pravdepodobnosť, ktorá ho zaujíma, a ktorú potrebuje v nejakom rozhodovacom procese. Napr.: Aká je pravdepodobnosť, že keď dovezú z výročnej oslavy založenia Likérky na záchytku jej pracovníka, že to bude muž, ktorý je tam pravidelným hosťom? (33,33 %) Alebo že je to žena, ktorá prvý krát v živote ochutnala liehoviny? (2/3) Alebo že to bude žena? (70%)

Je to obdobný príklad ako predchádzajúci II.7., môžete si urobiť pre každú otázku príslušný Bayesov vzťah a počítať. Výsledky sú uvedené v zátvorke za každou otázkou. Kontigenčná tabuľka 2x2 je nasledovná:

.

závislý / rod	áno	nie	spolu
Žena	7	14	21
Muž	3	6	9

spolu	10	20	n=30

Pr.II.9:: Manažér v spolupráci s inými odborníkmi pomáhajú zamestnaneckej firme zaviesť reklamou odporúčaný nový test na drogy, ktorý má citlivosť aj špecifičnosť 99%, teda, že zachytí 99% všetkých reálnych prípadov užitia drogy a vylúči 99% tých, ktorí drogu neužili. Prevalencia užívania drog vo firme je 0,5% (teda 0,5% všetkých zamestnancov drogy naozaj berie). Aká je pravdepodobnosť, že pracovník firmy, ktorý mal pozitívny test je aj užívateľom drog? Je to dosť dôležitý príklad, urobme si značenie toho, čo potrebujeme:

N – narkoman medzi zamestnancami, **nN** – zamestnanec, ktorý drogy neužíva.

p(N) – pravdepodobnosť, že zamestnanec drogy užíva – prevalencia = 0,005 (0,5%)

p(nN) - pravdepodobnosť, že zamestnanec drogy neužíva **= 1- p(N)** = 0,995 (99,5%)

T – test pozitívny

nT – test negatívny

p(T/N) – podmienená pravdepodobnosť, že test je pozitívny, ak je zamestnanec užívateľom drog = 0,99 (senzitivita testu = 99%)

p(T/nN) – podmienená pravdepodobnosť, že test je pozitívny, ak nie je zamestnanec užívateľom drog = 0,01 (doplnok k špecifickosti testu, ktorá je 99%)

Kontigenčnú tabuľku môžeme zostaviť len z pravdepodobnosti:

test / užívateľ	pozitívny	negatívny	spolu
Áno	0,99.0,005	0,01.0,005	0,005
Nie	0,01.0,995	0,99.0,995	0,995
spolu	0,0149	0,9851	n=1

Sme opäť pri Bayesovi **[II.15]**, prepíšeme si ho:

$$p(N/T) = \frac{p(T/N).p(N)}{p(T/N).p(N) + p(T/nN).p(nN)}$$

a dosadíme, aby nám vyšla pravdepodobnosť, že keď je test pozitívny, pracovník je naozaj užívateľ drog: p(N/T) = 0,00495/0,0149 ≅ 0,3322

Je to len 33,22 % pravdepodobnosť! Mali sme vysokú senzitivitu aj špecifikosť testu, napriek tomu pravdepodobnosť jeho úspechu, teda využiteľnosť v prípade potreby je prekvapujúco malá.

V Bayesovom vzťahu okrem technických charakteristík testu vystupuje ako dôležitý faktor aj prevalencia, t.j. výskyt v populácii. Používa sa dosť často v populačných etiologických

štúdiách, pri diagnostickom, terapeutickom a prognostickom rozhodovacom procese. Okrem humanitných vied sa preto často využíva v biomedicínskej štatistike. [8-10].

Literatúra k II. kapitole:

[1] Thomas, L..: Myšlenky pozdě v noci, Mladá fronta, Praha, 1989

[2] https://www.dailymail.co.uk/travel/article-621354/Travel-pictures-week.html

[3] Zeman, B.: Byl jednou jeden král, čs. film, Studio hraných filmů 1955 na motívy rozprávky Soľ nad zlato Boženy Němcovej

[4] Sväté písmo, Jeruzalemská biblia, Dobrá kniha, Trnava 2013

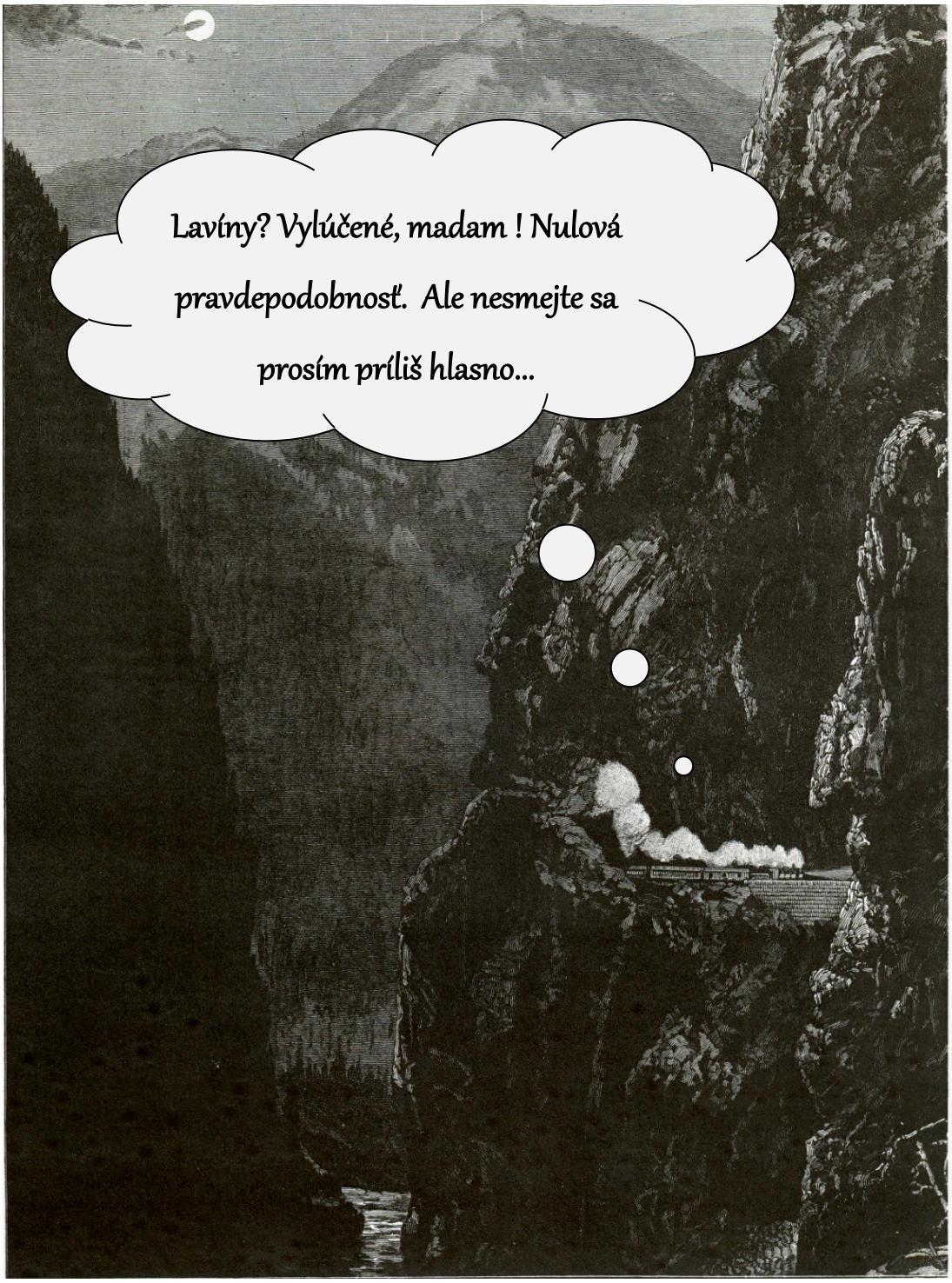
[5] Listchmannová, M.: Vybrané kapitoly z pravděpodobnosti (Interaktivní učební texty a řešené příklady), Vysoká škola báňská TU, Ostrava a Západočeská univerzita Plzeň, 2011. [6] Aczel, A.D.: Náhoda–příručka pro hazardní hráče, zamilované, obchodníky s cennými papíry a ostatní, Dokořán 2008

[7] Feynman, R. P.-Leighton, R. B.-Sands, M.: Feynmanove prednášky z fyziky. Bratislava, Alfa 1986

[8] http://www.algoritmy.net/article/45037/Bayesova-veta

[9] https://www.wikiskripta.eu/w/Bayesova_věta

[10] Rektorys, K. a i.: Přehled užité matematiky, SNTL, Praha 1981


Lavíny? Vylúčené, madam ! Nulová pravdepodobnosť. Ale nesmejte sa prosím príliš hlasno...

III. Kompas v nás[1] alebo kúzlo pravdepodobnosti

Či už veríte, že niečo dokážete, alebo naopak veríte, že to nedokážete, v oboch prípadoch máte s najväčšou pravdepodobnosťou pravdu.

Henry Ford

V priebehu necelých piatich mesiacov na prelome rokov 2013 a 2014 som si robil záznam vždy, keď som v masmediálnom priestore (do ktorého som v tomto prípade zaradil aj prednášky, diskusné fóra a iné verejné akcie) zachytil vetu ozdobenú ornamentom *„s najväčšou pravdepodobnosťou“* alebo veľmi blízkymi vyjadreniami. Možno vzletne a veľmi odborne povedať, že som si stanovil cieľ výskumu zistiť frekvenciu a zmysel použitia tohto slovného spojenia atakujúceho poslucháča v istom období, ale priznám sa, že v skutočnosti som počas etapy zberu údajov takto ešte neuvažoval. Keď si na to dávame pozor, je až udivujúce, ako často sa tento pojem z počtu pravdepodobnosti používa v najrôznejších situáciách a súvislostiach. Výsledky som neskôr zhrnul do tabuľky III.1.

	A	B	C	Spolu
Výskyt	**83**	**14**	**51**	**148**

Tab.III.1: Záchyt termínu *„s najväčšou pravdepodobnosťou“* a blízkych výrazov v masmediálnom priestore v sledovanom období 5 mesiacov.
A – záchyt vety s vyjadrením *„s najväčšou pravdepodobnosťou“*
B – záchyt vety s vyjadrením *„s veľkou pravdepodobnosťou“*
C – záchyt vety s vyjadrením *„s istotou“* alebo *„môžem úplne vylúčiť“*

Je to ilustračný príklad, na ktorom si môžeme niečo ukázať. Výsledky „experimentu“ sú výrazne ovplyvnené subjektívnym správaním spracovateľa, t.j. autora tejto publikácie. Predovšetkým nebol robený systematicky, ale len úplne náhodne, hlavne z toho dôvodu používam termín *záchyt*, miesto *výskyt*. Možno sa to nezdá byť dôležité, ale je dobré častejšie si zopakovať vetu:

Výsledok analýzy nie je nikdy lepší, ako je kvalita vstupných údajov.

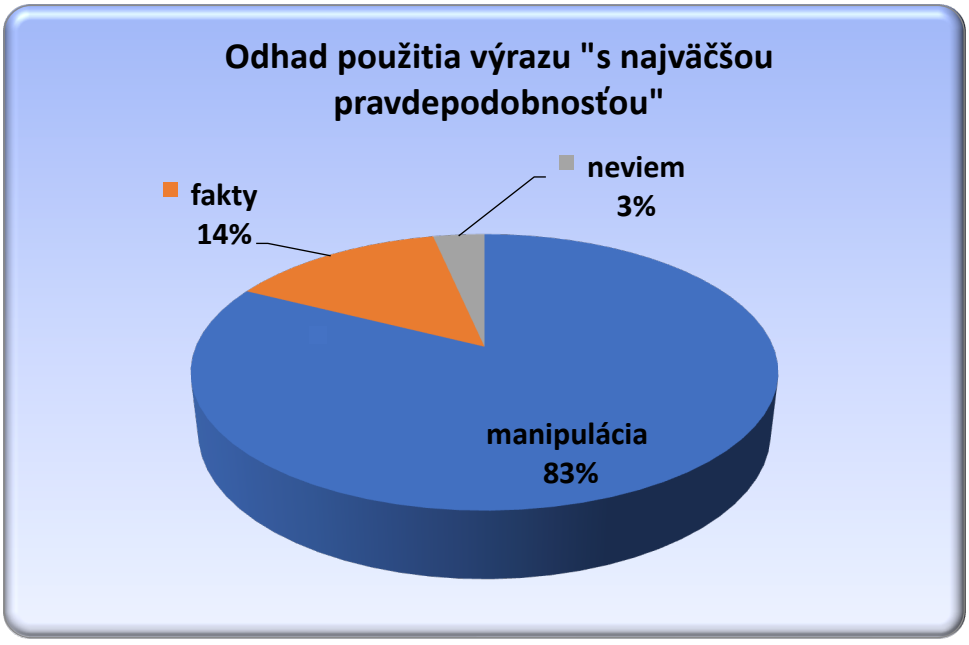
Keďže vyjadrenia typu A, typu B a typu C majú z hľadiska počtu pravdepodobnosti obecne rovnaký obsah, sčítal som ich do sumárneho výskytu za celé obdobie. Istý stupeň nutkavej zvedavosti mi nariadil každému záchytu priradiť odhad vlastnosti, ktorú môžeme nazvať trochu zložitejšie *zmysel použitia slovného spojenia.* Preložené do zrozumiteľného jazyka:

1. Nazdávam sa, že autor resp. užívateľ výroku sa ním pokúšal podporiť nejaké svoje tézy a úmysly bez ich hlbšej alebo lepšie povedané bez akejkoľvek štatistickej analýzy. Použil ich na zvýraznenie iných faktov, zlepšenie imidžu, obhajoby pravdivosti svojich tvrdení tam, kde sa nedali ani dokázať, ani vyvrátiť a možno ani neboli pravdivé. Takéto použitie môžeme spoločne nazvať *manipulácia*.
2. Výrok bol použitý osobou na základe experimentov, pozorovaní, analýzy nejakých javov ako ich výsledok, alebo prinajhoršom ako jeho odhad. Takéto použitie môžeme spoločne nazvať *predloženie faktov* alebo stručnejšie *fakty*.
3. Do poslednej kategórie som zaradil výroky, kde som sa nevedel rozhodnúť a ani žiadnym myšlienkovým postupom zistiť ich cieľ a zmysel. Názov – *neviem*.

Po tomto úvode je možno zaujímavejší údaj predstavujúci rozdelenie početnosti podľa priradených kategórií ako vidno z tabuľky III.2 a grafu III.1:

Kategória	1 - manipulácia	2 - fakty	3 - neviem	Spolu
Výskyt	122	21	5	148

Tabuľka III.2. Rozdelenie početnosti záchytu výrazu „s najväčšou pravdepodobnosťou“ podľa odhadu ich použitia.



Obr. III.1. Rozdelenie početnosti záchytu výrazu „s najväčšou pravdepodobnosťou“ podľa odhadu ich použitia. Grafické znázornenie hodnôt tab. III.2. – koláčový diagram

Zdá sa, že rozdelením javu podľa početnosti výskytu orientovaným na nejakú vlastnosť, získavame nové, cenné informácie. Iný pohľad dáva rozdelenie sumárnej početnosti záchytu výrazu „s najväčšou pravdepodobnosťou“ v časovej škále do 5 tried podľa mesiacov:

mesiac	11/2013	12/2013	1/2014	2/2014	3/2014	spolu
záchyt	11	16	31	63	27	148
%	7,43	10,81	20,95	42,57	18,24	100

Tabuľka III.3. Rozdelenie početnosti záchytu výrazu „s najväčšou pravdepodobnosťou“ podľa času záchytu (v mesiacoch).

Pozrime si výsledky v bodovom spojnicovom grafe:

Obr. III.2. Rozdelenie početnosti záchytu výrazu „s najväčšou pravdepodobnosťou“ podľa mesiaca v období sledovania. Grafické znázornenie hodnôt tab. III.3.

Po tomto najjednoduchšom tabuľkovom a grafickom spracovaní údajov by sme mohli urobiť ich rozbor a prezentáciu, čo je poslednou etapou výskumu. Náhodný jav, metodiku výskumu, vhodnosť a spoľahlivosť experimentátora a výsledné údaje som nijako neoveroval. George Horace Gallup (1901-1984), americký novinár a svetoznámy zakladateľ moderných štatistických metód prieskumov verejnej mienky (áno, je to ten človek, po ktorom sú pomenované renomované ústavy prieskumu verejnej mienky po celom svete), ktoré hýbu politikou, reklamou a mnohými inými rozhodovacími činnosťami tvrdil, že keď kuchár dobre zamieša obsah hrnca, tak mu postačuje nabrať a ochutnať iba za lyžicu polievky, aby zistil, ako je to s celým jeho obsahom.

Ak vezmem ako mieru kvality vyššie uvedeného výskumu pomer jeho reprezentatívnosti a nákladov, tak je to kvalitný výskum. Preto si dovolím uviesť niekoľko záverov, s ktorými samozrejme môže čitateľ polemizovať:

1. Prvotné spracovanie získaných údajov poukazuje na prevažne manipulatívne použitie výrazu *„s najväčšou pravdepodobnosťou“* v masmediálnom priestore. Tab. III.2 a graf III.1.
2. Tabuľka III.3 a graf III.2 môžu poukazovať na skutočnosť, že ľudia, novinári, politici a i. sú v období Adventu, Vianočných sviatkov a novoročných predsavzatí (mesiace 11/13 až 1/14) akýsi lepší, viac duchovnejší a ohľaduplnejší, zbožní a ušľachtilí, menej náchylní klamať a podvádzať a ich správanie naznačuje, že raz bude svet lepší, alebo že ste sa nejakým prazvláštnym omylom ocitli medzi anjelskými bytosťami, ktoré z nejakého dôvodu musia ešte trpieť na tejto zemi. Maximum početnosti sledovaného javu, využívaného hlavne s cieľom manipulovať poslucháčstvo, ktoré sa vyskytlo vo februári 2014, by chcelo podstatne hlbšie štúdium. Niektorí analytici však tvrdia, že to mohlo súvisieť aj s kampaňou pred prezidentskými voľbami.

Tak ďaleko však rozbor výsledkov môjho výskumu nesiaha, možno len vyvodiť cenný záver, že keď budete počuť vetu, obsahujúcu bonmot *„s najväčšou pravdepodobnosťou“*, mali by ste byť okamžite ostražitý, pretože taká veta s najväčšou pravdepodobnosťou klame.

V predchádzajúcom texte sme sa zaoberali náhodným pokusom, ktorého výsledkom je náhodný jav, teda závisí od náhody. Aj keď sa opakuje a možno za rovnakých podmienok, jeho výsledok je neistý, náhodný aj v tej najjednoduchšej, elementárnej forme. Počet pravdepodobnosti sa snaží dávať isté pravidlá, zákonitosti na prácu s náhodnými javmi. Súbory dát, ktoré získavame pri sledovaní takýchto javov, bývajú občas veľmi rozsiahle. Uviesť jeden takýto súbor napr. zo sčítania ľudu by mnohonásobne presiahlo rozsah tejto publikácie a na veľký zármutok mnohých čitateľov by nezostal priestor na iné zaujímavé skvosty pravdepodobnosti a štatistiky. Preto sa v našich ilustračných príkladoch budem veľmi obmedzovať ich rozsahom len na nevyhnutnú mieru pre pochopenie podstaty. Je to len z dôvodov výučby, nie je to prevažujúca vlastnosť reálnych štatistických súborov. Práve naopak.

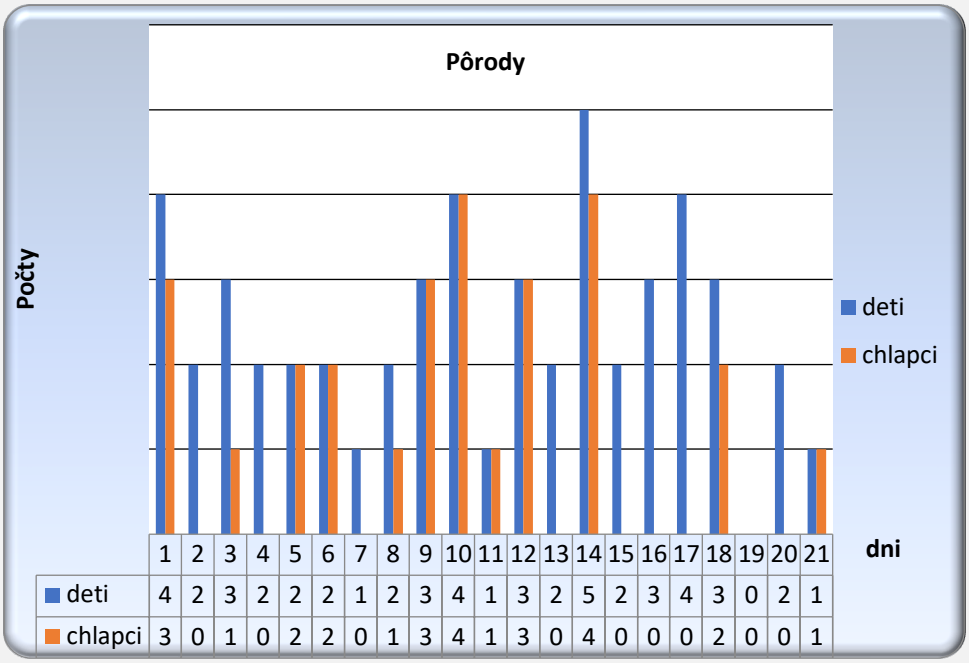
Náhodný jav je teda nejaké tvrdenie o náhodnom pokuse, ktorého výsledkom je jeho pravdivosť, alebo nepravdivosť. Náhodný pokus môže teda nadobúdať rôzne hodnoty o čom rozhoduje náhoda. V najjednoduchšej alternatívnej forme je výsledkom pokusu jedna z dvoch dichotomických možností „áno – nie“, „ nula – jednotka“, „muž - žena“ atď. Napr. pokus o odhad, koho ako prvého stretnete hneď za najbližším rohom. Muža alebo ženu? Osoba,

s ktorou ste sa práve rozprávali o polnoci v blízkosti cintorína bola živá, alebo mŕtva? Klávesnica vášho počítača bude po poliati kávou buď funkčná, alebo nefunkčná, atď.

Inokedy môže náhodná veličina **X** nadobúdať viac hodnôt x_i. Pri mnohonásobnom opakovaní náhodného pokusu sa zaznamená dosť neprehľadná zmes rôznych výsledkov. A v nich sa potrebujeme zorientovať. Prvou možnosťou je uložiť ich postupne do tabuľky.

Pr.III.1: Pôrodnosť za tri týždne mesiaca máj, vyjadrená počtom živonarodených detí na pôrodníckom oddelení jednej fakultnej nemocnice, z toho chlapcov, na obrázku, ktorý je nazývaný aj **stĺpcový graf**:

Obr.III.3. Prehľad pôrodnosti za tri týždne mesiaca máj v pôrodnici



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
deti	4	2	3	2	2	2	1	2	3	4	1	3	2	5	2	3	4	3	0	2	1
chlapci	3	0	1	0	2	2	0	1	3	4	1	3	0	4	0	0	0	2	0	0	1

Zistené údaje sú graficky spracované stĺpcovým grafom. Rad hodnôt, ktorý sme zaznamenali jednoducho z pozorovania postupným zapisovaním v tomto prípade deň za dňom, nazývame **empirický rad hodnôt**.

Prehľadnejšia bude tabuľka počtu narodení v jednotlivých dňoch. Za týmto účelom si urobme z postupnosti narodených detí usporiadaný súbor, dostaneme tzv. **variačný rad** (chlapcov už musíme k príslušným usporiadaným hodnotám priradiť):

Tab.III.4. Usporiadaný súbor pôrodnosti za tri týždne mesiaca mája v pôrodnici

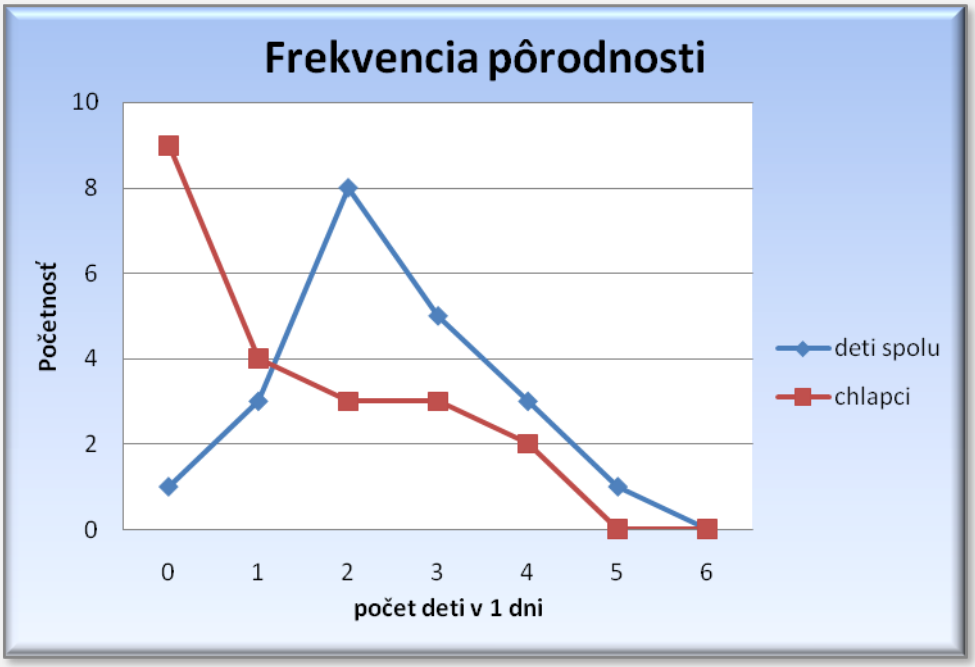
deti	chlapci
0	0
1	0
1	1
1	1
2	0
2	0
2	2
2	2
2	1
2	0
2	0
2	0
3	1
3	3
3	3
3	0
3	2
4	3
4	4
4	0
5	4
51	27

Tabuľka **absolútnej početnosti** vznikne tak, že počtu narodených dní priradíme početnosť dní, v ktorých sa tak udialo. Je to triedenie podľa kvantitatívnych znakov: V sledovanom období je 1 deň, v ktorom sa nenarodilo žiadne dieťa, 3 dni s jedným narodeným dieťaťom atď. Bude potom nasledovná:

Tab.III.5. Početnosť pôrodov za tri týždne mesiaca mája v pôrodnici

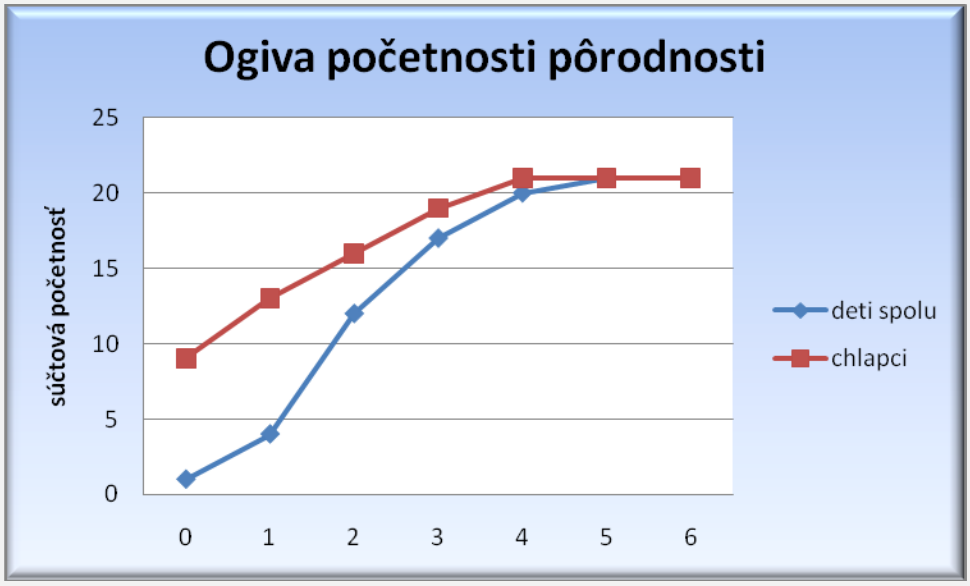
Počet detí (chlapcov) narodených v 1 deň	Početnosť	chlapci	Počet detí v 1 deň	Relatívna početnosť (%)	chlapci (%)
0	1	9	0	4,76	42,85
1	3	4	1	14,29	19,05
2	8	3	2	38,09	14,29
3	5	3	3	23,81	14,29
4	3	2	4	14,29	9,52
5	1	0	5	4,76	0
6	0	0	6	0	0
spolu	21	21	spolu	100	100

Relatívnu početnosť sme vyjadrili v %. Obr. III.4 tohto príkladu je už o niečo prehľadnejší. V takejto frekvenčnej analýze vynášame na x os hodnotu znaku, v našom prípade počet narodených detí v jednom dni, na zvislú os potom početnosť – v tomto prípade absolútnu početnosť. Môžeme vyjadriť spojnicovým grafom ako **polygón rozdelenia početnosti**.



Obr. III.4. Počet detí narodených v 1 deň (frekvencia pôrodnosti)

Iný typ spracovania údajov je **súčtová krivka (ogiva)**, ktorá vznikne ako neklesajúca krivka sčítaním hodnoty znaku, teda početnosti od bodu k bodu, tiež nazvaná **kumulatívna početnosť**:



Obr. III.5. Súčtová krivka detí narodených v 1 deň

Ešte som nič nepovedal o kvantitatívnych charakteristikách súboru dát, to príde neskôr, všetko je to len elementárne spracovanie a prezentácia výsledkov. Pokiaľ som výsledky nejakého šetrenia získal priamo v teréne (pozorovaním, anketou a pod.) sú to **primárne údaje**.

Sekundárne sú tie, ktoré získavam z iných štatistických dostupných zdrojov. Všetky uvedené detaily a termíny budeme v ďalšom rutinne používať.

Rozdelenie početnosti nám väčšinou dáva informáciu o vnútornej štruktúre súboru dát. Je to jeden z prvých krokov hľadania skrytého poriadku v chaose, ktorý nás obklopuje.

Typy rozdelenia početnosti:

A. Podľa toho, či sa početnosti sústreďujú okolo jednej triedy a jej okolia alebo viacerých:

a) Unimodálne – má 1 vrchol (napr. frekvencia pôrodnosti – deti na obr. III.4. príkladu III.1)

b) Bimodálne, polymodálné – má dva resp. viac vrcholov

c) U-rozdelenie; antimodálne – opak unimodálneho rozdelenia, s jedným minimom

B. Podľa symetrie rozdelenia

a) symetrické (v praxi väčšinou približne symetrické)

b) asymetrické - ľavostranné (tiež s kladnou šikmosťou resp. pozitívne) – početnosti sa sústreďujú viac pri nižších hodnotách znaku, vyššie sú chudobnejšie

- pravostranné naopak.

c. Iné, napr. J-rozdelenie – na začiatku maximum, postupne klesá ako po klesajúcej exponenciále; alebo opak J rozdelenia, teda na začiatku má minimum potom exponenciálne stúpa. Konštantné rozdelenie je rovnaké pre všetky hodnoty.

Ako delíme dáta v súbore do tried? Najčastejšie sa nám rozdelenie ponúka prirodzene (na časové úseky, alebo nejaké pevné intervaly, skupiny a pod.). Je viac spôsobov určenia počtu intervalov **k**, bežne používaný spôsob je jeho odhad

$$k = \sqrt{n} \qquad \textbf{[III.1.]}$$

kde **n** je počet hodnôt v súbore; alebo **Sturgesovo pravidlo**

$$k \approx 1 + 3{,}3.log\ n \qquad \textbf{[III.2.]}$$

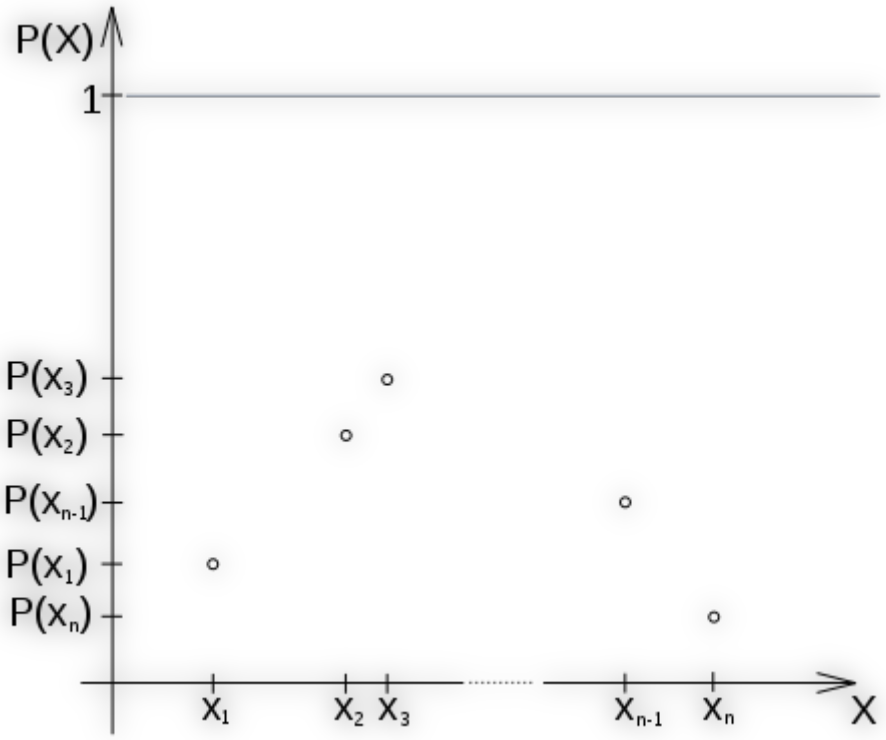
kde **log n** je dekadický logaritmus z **n**, k výsledku vám pomôže vaša kalkulačka. V pr. III.1. sme mali počet detí za celé sledované obdobie n = 51 a podľa **[III.1.]** by sme dostali po zaokrúhlení k ≅ 7. Podľa **[III.2.]** by bol odporúčaný počet intervalov 6 alebo 7. Početnosť sme nakoniec prirodzene rozdelili do 7 tried podľa frekvencie počtu detí narodených v 1 deň.

Rozdelenie početnosti nás privádza k pojmu **funkcia náhodnej premennej P(x).** Doteraz sme si ju vyjadrovali len tabuľkou alebo grafom. V tabuľke aj v grafe môžu byť aj hodnoty pravdepodobnosti. Ako v každej funkcii, (pozri kap. I. vzťah **[I.18.])** máme aj v nej nezávisle a závisle premennú. V tomto prípade však ide o náhodnú premennú, teda môže nadobúdať viac rôznych hodnôt z dôvodu náhodnosti procesu. Na štatistické znaky môžeme pozerať ako na náhodné premenné (napr. počet detí narodených v jeden deň z pr.III.1), ktoré

môžu nadobúdať rôzne hodnoty z pravdepodobnostného oboru hodnôt (z priestoru pravdepodobností). Náhodná premenná je dostatočne popísaná zákonom jej rozdelenia. Zákony rozdelenia sú modelmi, ktorými sa snažíme rozdelenie náhodnej premennej popísať, pretože sú dobre matematicky preštudované a majú známe vzťahy pre ich dôležité charakteristiky, ktoré môžeme potrebovať. S funkciou náhodnej premennej je úzko zviazaný pojem **distribučnej funkcie F(x)**. Distribučná funkcia v pr.III.1 bola znázornená ako súčtová krivka počtu detí narodených v 1 deň na obr. III.5. Je to teda akási „súčtová funkcia“ pravdepodobnosti. Každému reálnemu číslu **x** priraďuje pravdepodobnosť, že náhodná premenná (NP) nadobudne hodnotu menšiu než toto číslo t.j.:

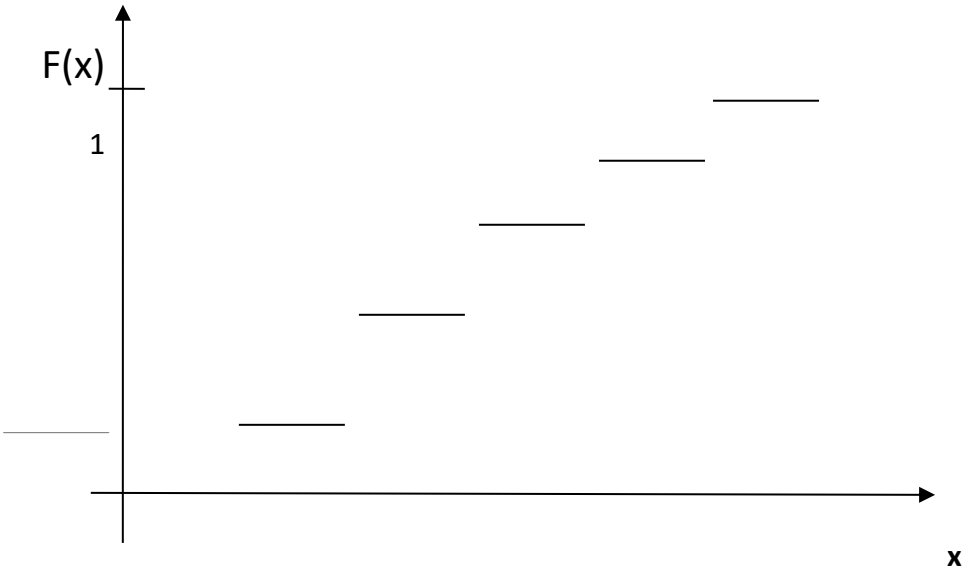
$$F(x) = P(X \leq x) \qquad [III.3.]$$

Rozdelenie početnosti aj pravdepodobnosti môže byť **nespojité (diskrétne)** napr. frekvencia počtu narodených detí, vek a rôzne iné skokom sa meniace údaje. Plynule sa meniace údaje nazývame **spojité (kontinuálne) rozdelenie**.

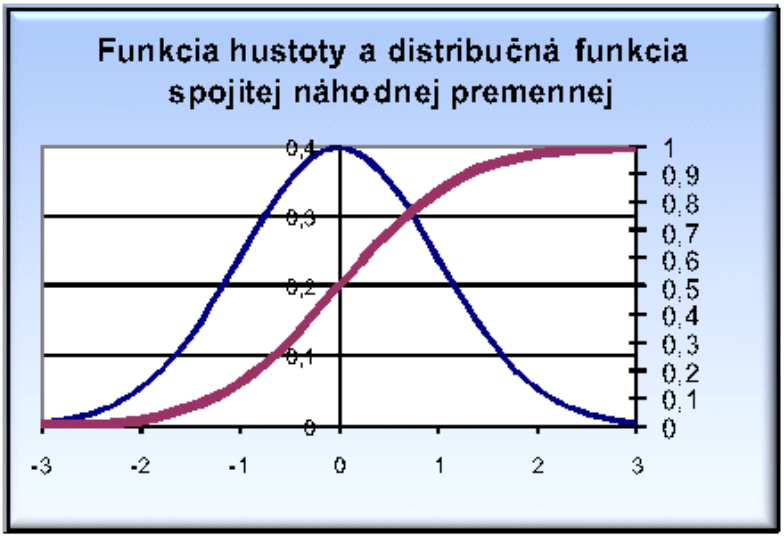


Obr. III.6. Diskrétne rozdelenie pravdepodobnosti [2]

Obr. III.6. až III.8. modelovo zobrazujú pravdepodobnostné a distribučné funkcie pre diskrétnu a pre spojitú náhodnú premennú. Rozdelenie pravdepodobnosti spojitej náhodnej premennej charakterizuje hustota pravdepodobnosti.



Obr. III.7. Distribučná funkcia diskrétnej funkcie náhodnej premennej



Obr. III.8. Tvar funkcie hustoty (modrá) a distribučnej funkcie náhodnej premennej

Sledovanie resp. vyšetrovanie náhodného javu nazývame pokusom, experimentom a pod. Pri jeho n-násobnom opakovaní napr. pre **n** len niekoľko málo jednotiek až desiatok môžeme dostať nejaké diskrétne rozdelenie pravdepodobnosti, ako na obr. III.6. Keď to mnohonásobne opakujeme, tvar diskrétneho rozdelenia sa postupne vypĺňa, zahusťuje; a pre spojitú náhodnú premennú potom hovoríme o hustote pravdepodobnosti. Jej distribučná funkcia začína zľava od 0 a spojito rastie až do 1. Ak sa hodnota náhodnej premennej **x** nachádza v intervale

$$x_1 < x < x_2$$

jej pravdepodobnosť možno z distribučnej funkcie zistiť jednoduchým vzťahom

$$p(x_1<x<x_2) = F(x_2) - F(x_1) \qquad \textbf{[III.4.]}$$

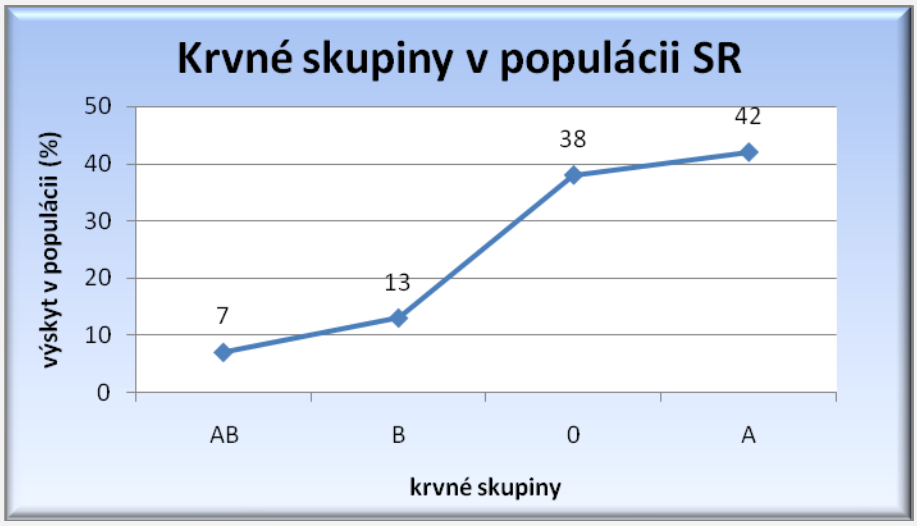
Nenechajte sa odradiť množinou nových pojmov a obrázkov. Uvediem radšej niekoľko príkladov:

Pr. III.2. Pri významnej návšteve z družobného mesta Nesferatu sa jej delegáti mimoriadne zaujímali o náš systém odberu krvi. Pozitívny ohlas u nich vyvolal fakt, že v našej populácii je výskyt krvných skupín nasledovný (tabuľka III.5. podľa [3]), pretože u nich a vo väčšine sveta je najčastejšou krvnou skupinou ľudí typ 0, pre nich z nejakého dôvodu menej obľúbenou. Na Slovensku je to inak:

Krvná skupina	Zastúpenie v populácii [%]
A	42
0	38
B	13
AB	7
Spolu	100

Tab.III.5. Pravdepodobnosť výskytu krvných skupín v populácii SR podľa systému AB0

Pravdepodobnostná funkcia má tvar

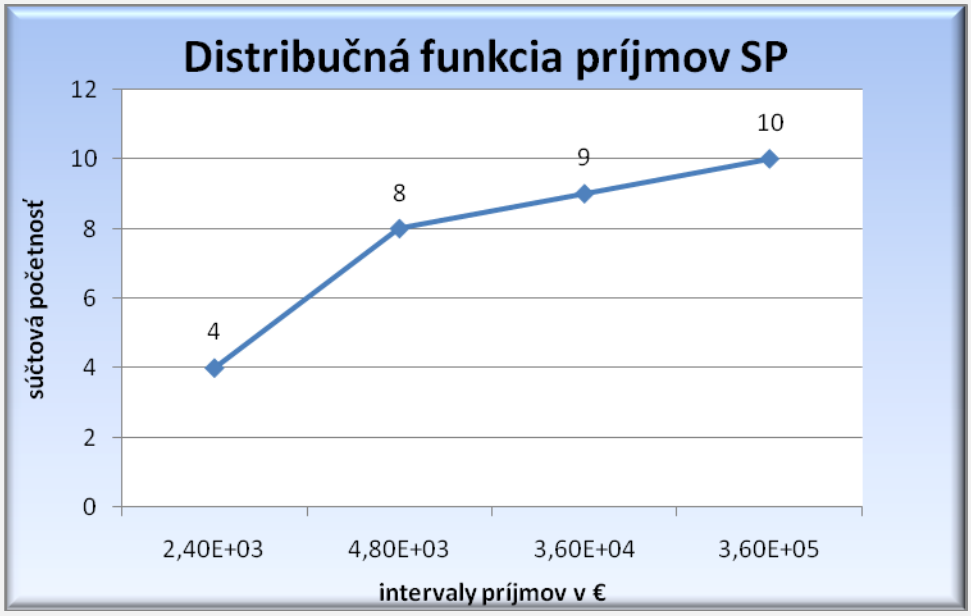


Obr. III.9. Pravdepodobnostná funkcia zastúpenia krvných skupín podľa systému AB0 v súčasnej populácii SR.

Rovnako sa zaujímali aj o distribučnú funkciu. Keď sme im ju zobrazili (nasledujúci obrázok III.10), radostne zatlieskali vychudnutými rukami s neuveriteľne dlhými pestovanými nechtami, požiadali nás o trochu samoty a intimity a rozbehli sa pozorovať nočný život nášho mesta plného mladých ľudí, ako sa hovorí „krv a mlieko“.

Obr. III.10. Distribučná funkcia výskytu krvných skupín v populácii SR

Pr. III.3. V správe o stave istej banánovej republiky sa dočítate, že súhrnné čisté ročné príjmy 10 profesionálnych sociálnych pracovníkov sú vo výške takmer 425 000.- €. Keďže máte v sebe štatistické cítenie, ľahko si porovnáte svoj príjem s priemerným príjmom sociálneho pracovníka v onej krajine. Skôr ako si pobalíte kufre a sťahujete sa, aby ste tam mohli pracovať, zavoláte svojmu priateľovi z tamojšej univerzity, aby vám informáciu potvrdil. Poslal vám o tom podrobnejšiu správu, ale z dôvodov cenzúry prešiel len jeden graf distribučnej funkcie, ktorej domáci cenzori asi nerozumeli a jedna strohá informácia, ktoré uvádzame:



Obr. III.11. Distribučná funkcia ročných príjmov sociálnych pracovníkov v istej krajine

K tomu prišla ešte informatívna tabuľka, ktorá poukazovala na skutočnosť, že väčšina sociálnych pracovníkov pracuje v teréne a len nevyhnutné minimum sa ich zaoberá administratívnou činnosťou, takže krajina patrí medzi najlepšie na svete v boji proti byrokracii:

zaradenie	náplň	[%]
riadenie	Správa fondov, darov, zahraničnej pomoci, dotácií a ich rozdeľovanie, riadiaca a kontrolná činnosť	10
vyhodnocovanie	Zber a prezentácia údajov, vládna štatistika, marketing a PR, sprevádzanie hostí	10
terén	Práca v teréne v štandardných a vládnych mestských štvrtiach, poradenstvo	40
terén	Práca v getách, slumoch a chudobnejších i nedostupných lokalitách	40

Tab.III.6. Sociálni pracovníci rozdelení podľa pracovného zaradenia.

Z obr. III.10. ste videli, že má 4 príjmové intervaly s hranicami súčtovej početnosti výskytu sociálnych pracovníkov v nich zaradených: 1. ⟨0;4), 2. ⟨4;8), 3. ⟨8;9), 4. ⟨9;10). Pravdepodobnosť, v tomto prípade vyjadrenú početnosťou prípadov, možno v každom intervale vyjadriť ako sme uviedli vo vzťahu **[III.4.]** nasledovne:

$$\mathbf{p(x_1<x<x_2) = F(x_2) - F(x_1)}$$

Pre 1.príjmový interval je $\mathbf{F(x_2)=4}$ a $\mathbf{F(x_1)=0}$, teda početnosť výskytu sociálnych pracovníkov v ňom je $\mathbf{F(x_2) - F(x_1) = 4 – 0 = 4.}$

Obdobne pre 2.príjmový interval je $\mathbf{F(x_2)=8}$ a $\mathbf{F(x_1)=4}$, teda početnosť výskytu sociálnych pracovníkov v ňom je $\mathbf{F(x_2) - F(x_1) = 8 – 4 = 4.}$

Pre 3.príjmový interval je $\mathbf{F(x_2)=9}$ a $\mathbf{F(x_1)=8}$, teda početnosť výskytu sociálnych pracovníkov v ňom je $\mathbf{F(x_2) - F(x_1) = 9 – 8 = 1.}$ Rovnako aj vo 4.intervale nám výjde $\mathbf{p = 1.}$

Keďže máte kvalitné vzdelanie a dobrú kalkulačku, viete, že väčšie čísla sa môžu vyjadrovať „vedecky“ v exponenciálnej forme, napr. $100 = 10^2$ a píše sa 1E+02, alebo $3600000 = 3{,}6\text{x}10^6 = 3{,}6\text{E+06}$ atď. Viete dešifrovať x-ovú os a hravo si môžete zostaviť svoju tabuľku:

Č. intervalu	Ročný príjem v €	Počet sociálnych pracovníkov v intervale
1	2400.-	4
2	4800.-	4
3	36000.-	1
4	360000.-	1

Aj keď už niečo tušíte, pre istotu ešte raz zavoláte svojmu priateľovi v tej krajine a on vám vysvetlí, že všetky fondy, zahraničné dotácie a dary do uvedenej oblasti má na starosti generál X.Y., ktorý je námestníkom ministra práce a zhodou okolnosti zaťom vodcu krajiny, najvyššieho generála Y.Z.z.; o štátnu štatistiku a jej prezentáciu sa stará bratranec jeho tety, ktorá je aj jeho krstnou mamou. Keďže to nie je o moc lepšie ako doma, vybaľujete kufre.

Pristavme sa ešte na chvíľu pri jednom z najstarších štatistických pojmov, pri pojme pravdepodobnosť. Je mierou očakávania výskytu určitých vecí resp. javov, reálnych v minulosti, súčasnosti, alebo ich prognózujeme. Mieru očakávania dostaneme tak, že si zistíme pomer počtu priaznivých prípadov (meraním, výpočtom, štatisticky, a pod.) ku všetkým možným prípadom. Nič viac a nič menej.

Pokiaľ sme si čokoľvek jednoznačne neoverili napr. priamou skúsenosťou, dôkazom, odvodením či iným vhodným postupom, ostatné už je vecou viery, teda nie sme si tým úplne istí. A tu na nás číha pasca. Nielen politici a neandertálci, ale často aj každý z nás na niečo takéhoto rád dáva nálepku, že je to „možné“, „pravdepodobné“ alebo naopak. To je asi klamanie pravdepodobnosťou. Jedna voľne parafrázovaná historka vraví, že vo vlaku cez škótsku vysočinu cestovali spolu absolvent humanitných vied (asi filozof), prírodovedec a štatistik, keď cez okno zbadali čiernu ovcu.

„Jéééej, v Škótsku sú ovce čierne!“ – zvolal (asi) filozof.

„Myslím, že v Škótsku sú niektoré ovce čierne,“ – spresnil prírodovedec.

„Dá sa povedať, že v Škótsku existuje minimálne jedna ovca, ktorá je minimálne z jednej strany čierna,“ – uzavrel debatu štatistik.

Častým typom „nedorozumenia“ je napr. keď vám nejaký ezoterický guru povie, že nad každým človekom v každom okamihu sa s 13,5 % pravdepodobnosťou vznáša niekoľko neviditeľných UFO, ktoré majú síce mimozemský pôvod ale v čase a priestore medzipristáli a naberali energiu v starovekej Sumerskej ríši. (Pôvod ezoterických javov je vhodné umiestňovať do prehistórie, kde sa dajú dosť ťažko potvrdiť. Ale aj vyvrátiť a potom sa všetci divia, že na tom teda niečo musí byť.). Pokiaľ takýto výrok nemá byť len ďalší neandrtálsky bluf, jeho autor by musel urobiť veľké množstvo pozorovaní priestoru nad hlavami dostatočne veľkého množstva úplne náhodne vybraných ľudí a nejakým spôsobom potvrdiť v približne 135 prípadoch z 1000 prítomnosť viac ako 1 sumerského UFO. Chcem tým povedať, že žiadne „pravdepodobnostné“ potvrdzovanie nejakého nepotvrdeného javu nás nepribližuje k pravde. To nie je pravdepodobnosť, to je zavádzanie. Že sa to robí v politike a reklame asi nikoho neprekvapuje. Vo Veľkej Británii štatistická komisia chcela dosiahnuť, aby členom vlády bolo zakázané skúmať štatistické údaje pred ich zverejnením a zabránilo sa tak možnosti pomocou nich niečo politicky ovplyvňovať alebo ich zneužívať [4]. Zo skúsenosti z našich zemepisných dĺžok a šírok im môžeme trochu závidieť, že ich politici a neandrtálci si dávajú aspoň tú námahu nazrieť do skutočných štatistických údajov.

Autor sa priznáva, že sám niekedy používa „pravdepodobnostné“ výrazy neprimerane a s istou pravdepodobnosťou ich už aj použil v predchádzajúcom texte. Nuž čo už...

Prejdime teraz k pravdepodobnostným rozdeleniam niektorých funkcií náhodných veličín. Sú to teoretické matematicky dobre preštudované modely, ktoré nám môžu pomôcť pri mnohých štatistických a pravdepodobnostných výpočtoch a analýzach.

Keď rozhodca pred futbalovým zápasom hodí normálnu mincu, aby náhodný pokus určil po dohode, ktoré mužstvo bude mať výkop, sme s tým uzrozumení, že pravdepodobnosť že padne *panna* alebo *orol* je teoreticky rovnaká: 0,5. Keby však chcel robiť štatistickú show a z nejakého dôvodu by sa s kapitánmi dohodol, že bude hádzať desať krát a že vyhráva ten, komu padne viackrát jeho strana, môže sa stať napr., že 9 krát padne *panna* a 1 krát *orol*. Kapitán *orlieho* mužstva by sa mohol cítiť do značnej miery ukrivdený, dokonca psychologicky by mu takáto nepriazeň osudu mohla pokaziť náladu a výkonnosť. Ak by v škole na hodinách štatistiky dával trochu pozor a diváci to dovolili, požiadal by rozhodcu nech hodí mincou aspoň 1000 krát. Ešte lepšie by bolo zopakovať to 100 000 krát alebo hneď milión, pretože výsledný pomer padnutia *panny* a *orla* by sa už zaručene blížil k 0,5. Drobná odchýlka v prospech víťaza by už bola psychologicky akceptovateľná. Tomuto sa vraví **zákon veľkých čísel,** ktorý je v štatistike ústredným zákonom: Čím viac náhodných pokusov nejakého javu máme, tým viac sa experimentálna priemerná hodnota blíži k teoretickej. K popisu takéhoto procesu nám slúžia modelové rozdelenia pravdepodobnosti. Rozdelenie je pravidlo, ktoré každej hodnote resp. každému intervalu hodnôt priradí pravdepodobnosť, že náhodná veličina nadobudne túto hodnotu, resp. hodnotu z tohto intervalu. Ako sme si už vyššie naznačili existujú diskrétne a spojité rozdelenia. Sú zaujímavé a je ich dosť, niektoré sú bežne používané, iné dosť „exotické“. Pozrime sa na niektoré pre nás užitočné:

A: Diskrétne rozdelenia

1. Binomické rozdelenie. (tiež Bernoulliho rozdelenie)

Nech v jednom individuálnom pokuse sledovaný jav **A** môže nastať s pravdepodobnosťou **p** alebo nenastať s pravdepodobnosťou **q = 1 – p**. Keď uskutočníme takýto pokus **n**-krát, pričom pokusy musia byť štatisticky nezávislé a **p = konšt.**, jav **A** môže nastať **k**-krát, kde **k∈ ⟨0;n⟩**. Pravdepodobnosť, že jav **A** nastane v **n** pokusoch presne **k**-krát je vyjadrená vzťahom:

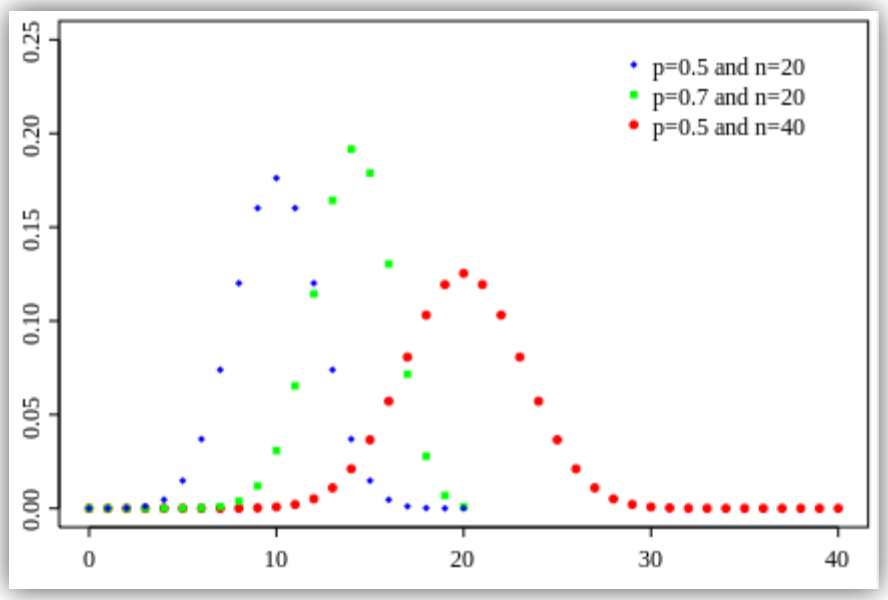
$$p(X=k)=\binom{n}{k}.p^k.(1-p)^{n-k} \qquad \text{[III.5.]}$$

Každé rozdelenie má dve základné charakteristiky: **strednú hodnotu E(X) a rozptyl D(X)**. Stredná hodnota počtu výskytu náhodného javu **A** v **n** pokusoch je v binomickom rozdelení

$$E(X)=n.p \qquad \text{[III.6.]}$$

a rozptyl (disperzia)

$$D(X)=n.p.(1-p) \qquad \text{[III.7.]}$$



Obr. III.12. Tvar binomického rozdelenia pre rôzne parametre p a n [5]

Pr. III.4. Máme 4 rovnakých nezamestnaných remeselníkov. S ich žiadosťami chceme navštíviť dlhodobo sledovaných 10 pracovísk, pričom v každej je rovnaká šanca, že 1 remeselníka do pracovného pomeru zoberú alebo nezoberú. Akú máme pravdepodobnosť umiestnenia evidovaných remeselníkov na tomto trhu práce?

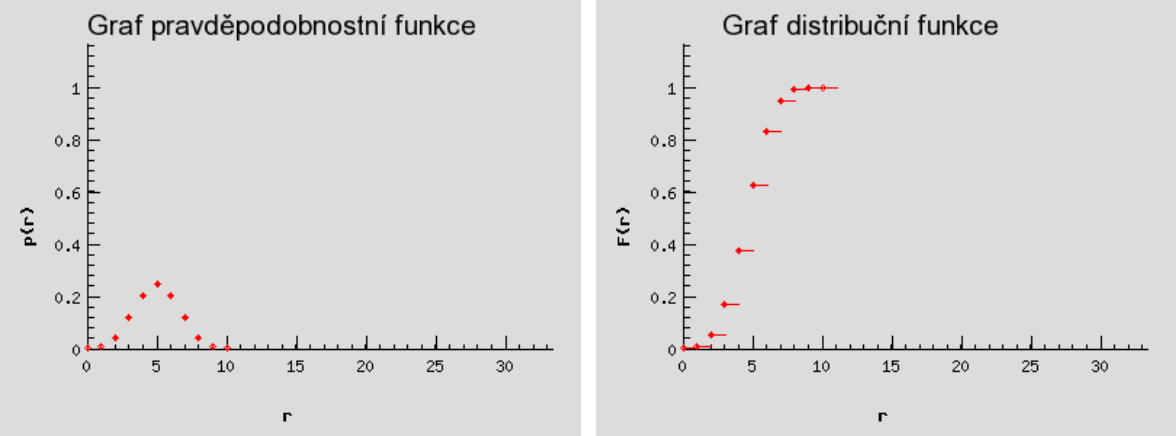
Binomické rozdelenie v tomto príklade má parametre:

n = 10, k = 4, p = ½ = 0,5 t.j. rovnaká šanca, že na 1 pracovisku 1 remeselníka príjmu alebo nie.

Pre pravdepodobnosť umiestnenia všetkých 4 evidovaných remeselníkov na trhu práce, ktorý predstavuje 10 firiem dostaneme z [III.5] pomocou šikovnej kalkulačky

$$p(X=4) = \binom{10}{4}.0{,}5^4.(1-0{,}5)^6 = 0{,}205078$$

kde $\binom{10}{4}$ je kombinácia podľa [I.16] z I. kapitoly a dáva 210 možnosti umiestnenia 4 remeselníkov na 10 pracoviskách. Táto hodnota sa vynásobí príslušnou mocninou pravdepodobnosti prijatia a opačného javu, neprijatia a výsledok je, že za daných podmienok máme asi 20,5 % pravdepodobnosť ich zamestnania. Ak sa vám nechce naprogramovať si to v EXCELI (je to funkcia **=BINOMDIST(k;n;p;0)**) na webe máme napr. v **[6]** niečo ako online kalkulátor binomického rozdelenia. Uvediem výsledky z neho:



Obr. III.13. Graf pravdepodobnostnej a distribučnej funkcie binomického rozdelenia pre parametre p = 0,5 a n = 10 [6]

Podľa očakávania z uvedených obrázkov je stredná hodnota podľa [III.6] E(X) = 5 s rozptylom podľa [III.7] D(X) = 2,5. Program - vypočítané pravdepodobnosti rozdelenia uvádza v tabuľke, v ktorej pre k = 4 vychádza hodnota pravdepodobnosti ako aj nám vyššie p(X=4) = 0,205078.

Pomocou binomického rozdelenia by sme mohli vypočítať pravdepodobnosť napr. ak chce mať rodina **n** detí, aká je pravdepodobnosť, že budú mať 1, alebo 2 vo všeobecnosti **k** chlapcov, keď pravdepodobnosť narodenia chlapca v súčasnej populácii je 52%.

Pr. III.5. Veľký kalif sa mieni oženiť, ale musí zabezpečiť aj pokračovanie vládnuceho rodu, teda dediča. Potenciálne nevesty, ktoré sa na jeho dvor zbehli ako osy na med, upozornil, že za ženu si môže vyvoliť len tú, ktorá mu dá aspoň 5 detí ale hlavne medzi nimi následníka trónu. Takmer všetky záujemkyne o švárneho kalifa odišli, lebo jeho požiadavku mu nemohli zaručiť. Zostala len princezná Binomiálka, v rozprávkach často prezývaná ako *Šeherezáda*, ktorá sa nechala predviesť pred kalifa a vraví mu:
„ Óóó, mocný kalif, najžiarivejšia hviezda všetkých poddaných, dovoľ svojej nehodnej služobnici prehovoriť.“
V istom očakávaní nepatrne prikývol, potom sa už k slovu nedostal.
„Premúdry vládca, len vďaka tebe všetci vieme, že o živote a smrti rozhoduje Najvyšší v nebi a ty ako jeho zástupca. Keďže nie sme pánmi svojho života, nikto z nás okrem Najvyššieho nemôže rozhodnúť, čo sa má narodiť.“
Trochu nespokojne sa zamrvil, ale skôr ako ju mohol dať vyhnať, rýchlo pokračovala:

„Najjasnejší vládca, som len obťažujúcou muchou, ktorú môžeš hocikedy svojou mocnou rukou rozmliaždiť, ale ja Ti môžem dať aspoň veľkú nádej. Keď som sa pripravovala na svoju úlohu dobrej manželky, okrem iného som sa zaoberala rozdelením pravdepodobnosti pána Jacquesa Bernoulliho, v Európe vraj slávneho učenca. Na základe jeho vzťahov Ťa môžem ubezpečiť, že ak budeme mať spolu 5 detí, tak fakt, že by sa nám nenarodil z toho žiaden syn je takmer nemysliteľný, pravdepodobnosť je len asi 2,5%. Pravdepodobnosť narodenia 1 syna z piatich detí je 13,8%, 2 synov už 29,9% a viac ako 32% máme pravdepodobnosť, že sa nám narodia 3 synovia. Ba ešte aj 4 synov môžeme mať so 17,5% pravdepodobnosťou. To, že všetkých 5 deti budú synovia Ti neviem zaručiť, je to mizivá pravdepodobnosť, necelé 4%. Stredná hodnota je medzi 2 a 3 synmi, presnejšie 2,6 s rozptylom 1,25.“

Princezná Binomiálka zabudla povedať, že rovnakú pravdepodobnosť by mal vznešený kalif s hociktorou dievčinou. Ale ani to nebolo potrebné, dôležité bolo sústrediť všetky sily a pozornosť na blížiaci sa svadobný obrad a kalif tuho rozmýšľal ako zabrániť toľkým svojim synom, aby sa o trón pobili.

Ešte malá poznámka: Binomické rozdelenie s **n = 1** je tzv. alternatívne rozdelenie (jav **A** nastal alebo nenastal), často používané v teórii hier a i.

2. Poissonovo rozdelenie alebo aj rozdelenie riedkych javov

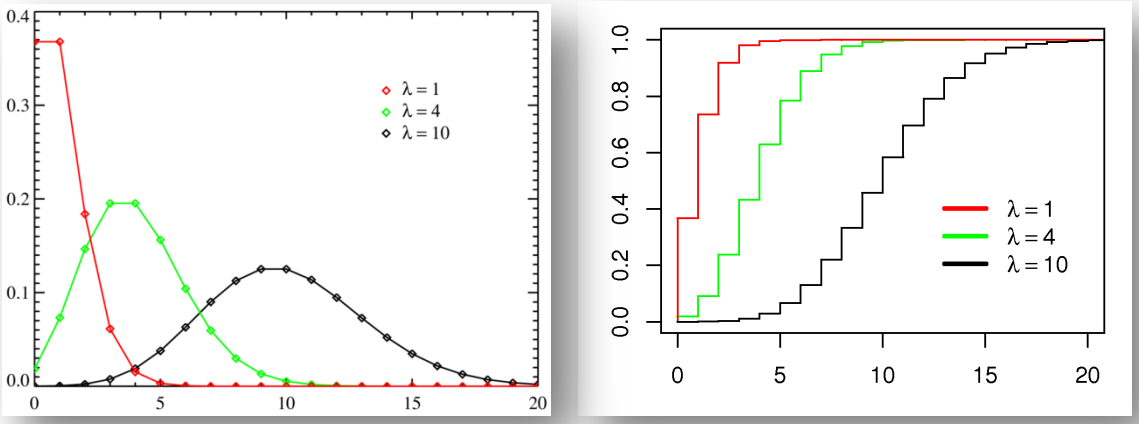
Francúzovi Simeonovi Denisovi Poissonovi (1781–1840) sa vraj ako členovi Francúzskej akadémie Bernoulliho rozdelenie nepozdávalo už len preto, že Jacques Bernoulli (1654 – 1705) bol Švajčiar. Pre veľký počet pokusov javov s malou pravdepodobnosťou bolo zdĺhavé a nevyhovujúce. Navrhol vlastné, ktoré sa ukázalo pre veľké **n>30** a malé **p≤0,1** ako dobrou aproximáciou binomického rozdelenia s jediným parametrom, strednou hodnotou rozdelenia **λ>0**; a pri svojej veselej povahe ho hravo aplikoval na výpočty úmrtnosti za predpokladu, že v populácii k úmrtiam pri väčšine chorôb dochádza nezávisle a náhodne:

$$p(X = k) = e^{-\lambda}.\frac{\lambda^k}{k!} \qquad \textbf{[III.8]}$$

Stredná hodnota a rozptyl je rovná parametru **λ**

$$\mathbf{E(X) = D(X) = \lambda} \qquad \textbf{[III.9]}$$

V programe EXCEL sa výpočet charakteristík Poissonovho rozdelenia nachádza vo funkcii: **POISSON(k, λ,0)**. Web kalkulátor nájdete tiež v **[7] resp**. **[9].**



Obr. III.14. Graf pravdepodobnostnej a distribučnej funkcie Poissonovho rozdelenia pre parametre λ=1; 4 a 10 [8]

Pr. III.6. Ruský ekonóm a štatistik, žijúci na prelome 19. a 20. storočia v Nemecku a prednášajúci na Univerzite v Strassburgu Ladisaus Bortkiewicz (1968-1931) sa venoval závažným úkazom svojej doby. Študoval politickú ekonómiu Karola Marxa. Taktiež vydal v roku 1898 knihu *Das Gesetz der kleinen Zahlen,* ktorá ho preslávila. V nej zverejnil svoje 20 ročné pozorovania v 10 vojenských zboroch kavalérie nemeckej armády. Išlo mu o počty vojakov za rok, ktorí zomreli následkom kopnutia koňa. 20 rokov x 10 zborov kavalérie = 200 nezávislých pokusov, teda **n = 200**. V tomto čase sumárne došlo k 122 smrteľným prípadom, teda očakávaná pravdepodobná hodnota počtu úmrtí za rok v jednom zbore **λ = 122/200 = 0,61** Bortkiewicz použil ako parameter **λ = 0,61.** Odhadnime spolu pomocou Poissonovho rozdelenia koľko sa priemerne za rok v zboroch kavalérie nemeckej armády vyskytlo k smrteľných úrazov, **k = 0,1,2,3,4,5** ? Ak ste zruční v narábaní s vašou kalkulačkou, dosaďte si a po dlhej dobe dostanete (zohľadnite vzťahy [I.17] a Pr.I.8. z I. kapitoly, kde sme uviedli, že každé číslo umocnené na nultú =1)

$$p(X=0) = e^{-0,61}.\frac{0,61^0}{0!} = 0,543351$$

Bolo to namáhavé, preto radšej využite možnosti programu EXCEL, alebo online kalkulátora v **[9]**, ktorý vám vyhodí potrebnú tabuľku, ktorú si doplníme dvomi stĺpcami: vypočítanými počtami úmrtí vojakov VH (zaokrúhlime na celé čísla) a pozorovanými hodnotami PH.

k	p	VH	PH
0	0,543350869	109	109
1	0,331444030	66	65
2	0,101090429	20	22
3	0,020555054	4	3
4	0,003134646	1	1
5	0,000338243	VH →0	0

Tab.III.7. Hodnoty pravdepodobnosti Poissonovho rozdelenia p, teoretických (VH) a pozorovaných (PH) úmrtí vojakov kopnutím koňom v Bortkiewiczovom experimente. [10]

Slušne to sedí, že? Ako povedal jeden šachista z Hané na Morave, potom ako prehral dôležitú partiu, pretože súper mal jazdca navyše: *Kůň je nebezpečné!*

Pr. III.7. Populárny drogový díler *Fifík* mal dobrú náladu. Kšefty išli nad očakávanie dobre, zákazníci sami doliezali a práve dostal väčšiu dodávku, ktorá mu zabezpečí veľmi slušný zisk a teda aj šťastný život. Riedením s trochu nekvalitným materiálom, omietkou, ktorý by sám nikdy nevzal ani v najväčšej núdzi, ako si v duchu pri práci hovoril, urobil 1000 dávok.V dobrom rozmare ale náhodne každú desiatu z nich ponechal v kvalitnom čistom stave, nech má aj nejaký šťastlivec, alebo šťastlivci vydarený deň.
Ak si chceme dnes večer urobiť na intráku dobrú párty a idem na ňu zaobstarať 10 dávok péčka, akú mám pravdepodobnosť, že medzi nimi budú aj super kvalitné? Je to Poissonovo rozdelenie pravdepodobnosti s parametrom **λ=100/1000 = 0,1**. Pomocou vlastných výpočtov, EXCELU alebo najlepšie online kalkulátora dostaneme tabuľku pravdepodobnosti:

K	P(X=k)	P(X≤k)	P(X=k) v [%]
0	0.049787068367864	0.049787068367864	**4,98**
1	0.14936120510359	0.19914827347146	**14,93**
2	0.22404180765539	0.42319008112684	**22,40**
3	0.22404180765539	0.64723188878223	**22,40**
4	0.16803135574154	0.81526324452377	**16,80**
5	0.10081881344492	0.9160820579687	**10,08**
6	0.050409406722462	0.96649146469116	**5,04**
7	0.021604031452484	0.98809549614364	**2,16**
8	0.0081015117946814	0.99619700793832	**0,81**
9	0.0027005039315605	0.99889751186988	**0,27**
10	0.00081015117946814	0.99970766304935	**0,08**

pripojili sme posledný stĺpec s % pravdepodobnosťou, že pri nákupe získam **k** kvalitných speedov. Ako vidno, od *Fifíka*, keď má dobrú náladu, oplatí sa nakupovať!

B: Spojité rozdelenia

1. Normálne (Gaussovo alebo Gauss-Laplaceovo) rozdelenie $N(\mu,\sigma^2)$ a $N(0,1)$

Je najdôležitejšie rozdelenie pravdepodobnosti spojitej náhodnej veličiny, preto sa mu trochu pozrime na zúbky. Popisuje veľmi veľa pravdepodobnostných a štatistických javov nášho sveta. Navyše aj mnohé iné diskrétne i spojité rozdelenia môže za určitých podmienok dostatočne aproximovať. Už sme sa s nim stretli ako s príkladom nejakej funkcie v kap. I, bol to vzťah **[I.26]** a jeho obrázok, tak si ho zopakujme a popíšme. Jeho matematická formula je

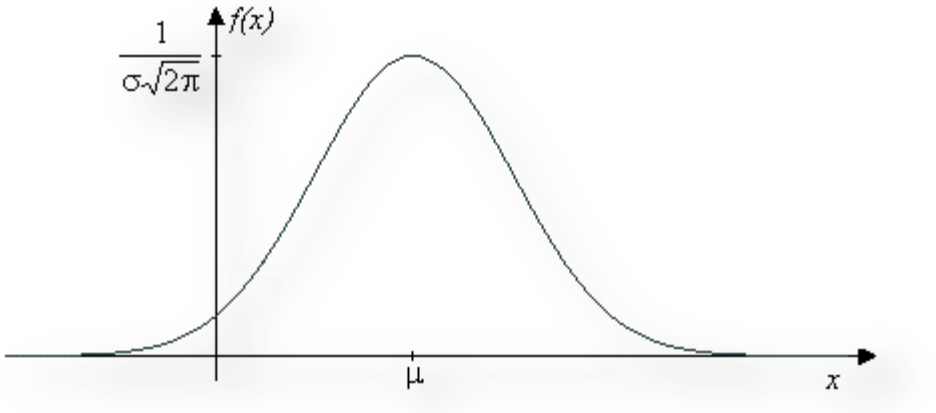
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} . e^{\frac{-(x-\mu)^2}{2.\sigma^2}} \qquad \textbf{[III.10.]}$$

ktorého strednú hodnotu označujeme μ a distribúciu σ^2:

$$\mathbf{E(X) = \mu} \qquad \textbf{[III.11]}$$

$$\mathbf{D(X) = \sigma^2} \qquad \textbf{[III.12]}$$

Jeho obecný tvar je



Obr. III.15.Tvar normálneho rozdelenia funkcie náhodnej premennej–Gaussová krivka

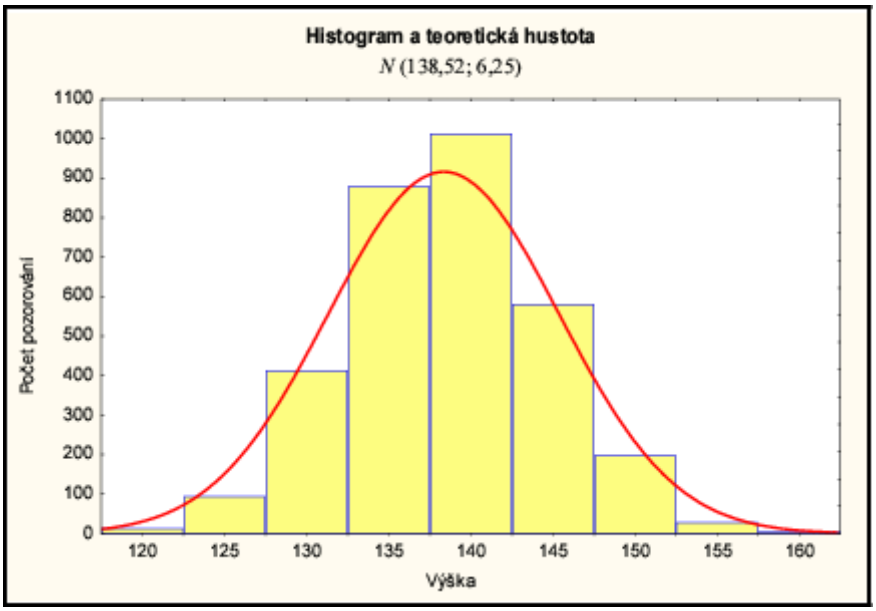
Sledovanie výšky, váhy a pod. celej populácie alebo vybraného súboru, napr. žiakov v triede, ľudí v meste, členov kmeňa Pygmejcov alebo hráčov NBA, najrôznejšie vedecké a technické merania, výskyty a pod. to všetko má pri dostatočne veľkom súbore tvar Gaussovej krivky, teda podlieha normálnemu rozdeleniu. Rozdelenie jednotlivých prípadov pravdepodobnosti je tak husté, že je spojité a hovoríme o rozdelení hustoty pravdepodobnosti, ktorá je úmerná ploche pod krivkou. Na obr. III.8. sme si už znázornili tvar funkcie hustoty a distribučnej funkcie spojitej náhodnej premennej. Než si priblížime vzťah hustoty a plochy pod krivkou, uveďme si jeden príklad [11]:

Pr. III.8. Sledovala sa výška 3231 chlapcov vo veku 9,5 až 10 rokov. Početnosť výskytu bola rozdelená do 9 tried so šírkou intervalu 5 cm. Výsledky boli spracované v tab.III.8. a na obr. III.16.:

Tab. III.8: Rozdelenie chlapcov vo veku 9,5-10 r podľa telesnej výšky (dĺžka triedneho intervalu 5 cm) [11]

Stred triedy x_i	Absolútna početnosť n_i	Kumulatívna absolútna početnosť	Relatívna početnosť n_i/n	Kumulatívna relatívna početnosť
120	13	13	0,0040	0,0040
125	95	108	0,0294	0,0334
130	414	522	0,1281	0,1615
135	880	1402	0,2724	0,4339
140	1013	2415	0,3135	0,7474
145	582	2997	0,1801	0,9275
150	199	3196	0,0616	0,9891
155	29	3225	0,0090	0,9981
160	6	3231	0,0019	1,0000
Spolu	3231	-	1,0000	-

Obr III.16: Histogram výberového rozdelenia telesnej výšky 3231 chlapcov vo veku 9,5-10 r a teoretická hustota normálneho rozdelenia telesnej výšky[11]

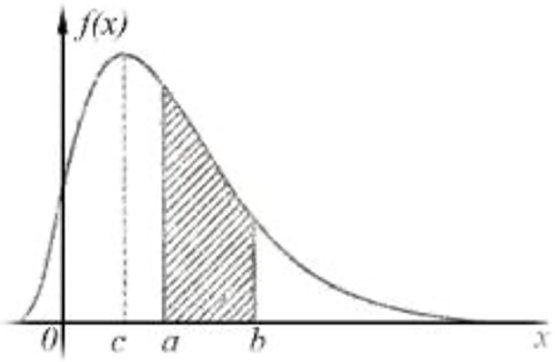


Pokiaľ by sme zväčšovali veľkosť výberu chlapcov nad všetky medze a zároveň skracovali dĺžku triedneho intervalu, obrys histogramu by sa postupne približoval teoretickej Gaussovej spojitej krivke.

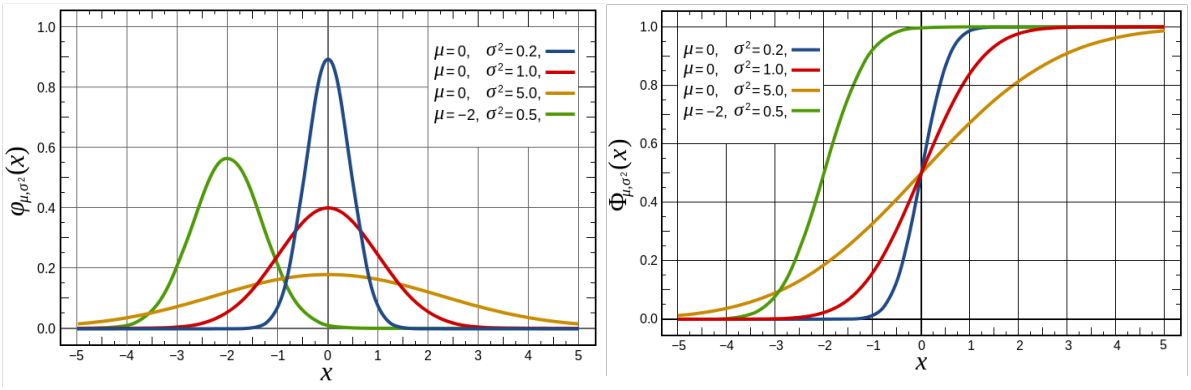
Pod hustotou pravdepodobnosti budeme v nejakom intervale **⟨a;b)** teda rozumieť ako pri diskrétnych rozdeleniach

$$\mathbf{P(a \leq X \leq b) = F(b) - F(a)}$$

čo je rozdiel hodnôt distribučnej funkcie v bode **b** a **a**, t.j. rozdiel plochy pod krivkou od -∞ až po **b** a plochy od -∞ až po **a**. Presne sa to dá urobiť trochu náročnejšou matematickou operáciou nazvanou integrovanie (viď ilustračný obr. III.17. pre všeobecný tvar rozdelenia), nám môže zatiaľ postačiť približná hodnota, predstavujúca stĺpec histogramu v danom intervale ako na obr. III.16. Napr. v intervale 135 až 140 cm bude 880 chlapcov, čo je veľkosť 4. stĺpca a dostatočne sa približuje ploche tohto intervalu pod teoretickou spojitou Gaussovou krivkou.



Obr.III.17. Výpočet hustoty pravdepodobnosti v intervale ako plochy pod teoretickou spojitou krivkou rozdelenia.

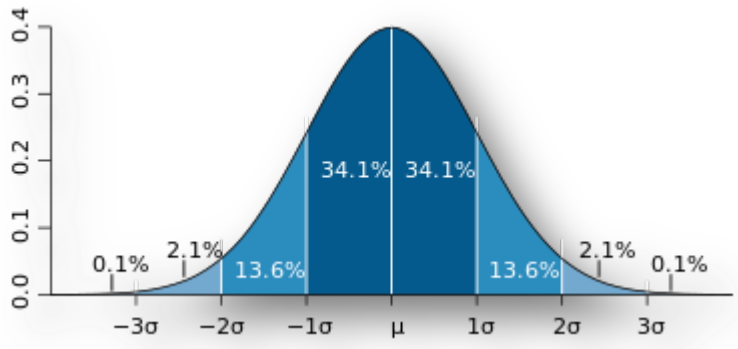


Obr.III.18. Tvar normálneho rozdelenia a jeho distribučnej funkcie pre niektoré hodnoty parametrov μ a σ [12].

Pomocou normálneho rozdelenia budeme interpretovať mnohé štatistické javy a postupy. Mnohé iné rozdelenia budeme s normálnym rozdelením porovnávať. Francúzsky astronóm a matematik Pierre-Simon Laplace (1749-1827) sa v značnej miere podieľal na tom, že počet pravdepodobnosti sa stal nástrojom zmysluplnej redukcie veľkých dátových súborov

a zisťovania ich neistoty. Dokázal, že na výsledky meraní majú vplyv mnohé nezávislé chyby a odvodil matematicky zákon chýb. Údaje, ktoré sú ovplyvňované veľmi veľkým počtom malých a na sebe nezávislých efektov, budú rozdelené viac-menej normálne. A matematicky odvodil **centrálnu limitnú vetu**, ktorá je základom počtu pravdepodobnosti. Výraz „normálny" bol dosť frekventovaný v 19. storočí v medicíne ako protipól k pojmu „patologický". Čoskoro sa začal používať obecnejšie, aj vo vzťahu k ľuďom a ich chovaniu, i k veciam vo význame „aké by asi mali byť". Pre krivky tvaru zvonu sa to prebralo od astronómov, ale tiež vo význame pre javy ako" typické", bežné", „obvyklé", teda niečo so zvyčajnou, priemernou vlastnosťou, kde „norma" predstavovala istý ideál.

A krivka nesklamala ani vo svojej matematickej variante. Pozrime si obr. III.19.



Obr.III.19: Krivka hustoty normálneho rozdelenia pravdepodobnosti s vysvetlením významu parametrov μ,σ.

μ je stredná, priemerná hodnota aj v zmysle **[III.11]** aj ako vidíme, okolo nej sú symetricky rozložené všetky väčšie a všetky menšie hodnoty. S akou pravdepodobnosťou, ukazuje parameter **σ**. V intervale **±σ** sa nachádza prakticky takmer 70% všetkých hodnôt normálneho rozdelenia, teda tie sú najviac „normálne". Vravíme že v **1σ**-intervale dostaneme hodnotu so 68% pravdepodobnosťou, alebo na hladine významnosti 68% resp. 0,68. Širší **2σ**-interval predstavuje už 95% hladinu významnosti a **3σ**-interval dokonca 0,99. Parameter **σ** je ako vyplýva zo vzťahu III.12. odmocnina z rozptylu a nazýva sa smerodajná odchýlka, ktorej veľkosť udáva ako široko sú rozložené hodnoty v sledovanom súbore:

$$\sigma = \sqrt{D(X)} \qquad \text{[III.13.]}$$

Každé normálne rozdelenie **N(μ,σ)** sa dá pretransformovať na normované normálne rozdelenie **N(0,1)**, kde stredná hodnota **μ = 0** a **σ = 1** pomocou transformácie:

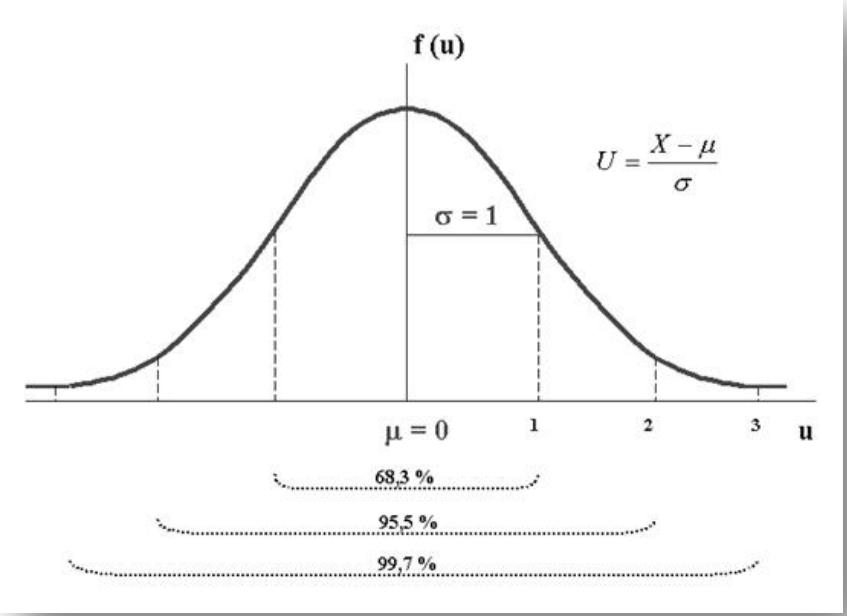
$$u = \frac{X - \mu}{\sigma} \qquad \text{[III.14.]}$$

po ktorej vzťah pre normálne rozdelenie III.10 nadobúda tvar:

$$f(u) = \frac{1}{\sqrt{2\pi}} . e^{\frac{-u^2}{2}} \qquad \text{[III.15.]}$$

Vyzerá to zložito, ale urobili sme len to, že strednú hodnotu sme umiestnili na vodorovnej osi do **0** a pravdepodobnosť na zvislej osi sme prerátali na interval od **0** do **1**.

Graf rozdelenia N(0,1) je na obr. III.20:



Obr.III.20: Krivka hustoty normovaného normálneho rozdelenia pravdepodobnosti s parametrami μ=0, σ=1.

Načo je to dobré? Hodnoty distribučnej funkcie **F** normálneho rozdelenia, teda plochy pod nejakou časťou zvonovej krivky nie je pre normálneho čitateľa najjednoduchšie vypočítať (výraz *normálny* používam, ako ste si všimli, už v novom pravdepodobnostnom a štatistickom význame, teda *najviac sa vyskytujúci, bežný*). Pre normované normálne rozdelenie sú však tieto hodnoty uvádzané vo všetkých dostupných základných štatistických tabuľkách pre rôzne **u**, čím sa výpočet hustoty pravdepodobnosti stáva dostupnou rutinou pre väčšinu čitateľov, minimálne pre **1σ** - interval, teda asi 68% celého ich súboru.

Častou úlohou pri aplikácii normálneho rozdelenia je nájsť pravdepodobnosť toho, že náhodná premenná X nadobudne hodnoty z intervalu x_1 až x_2 . Pri výpočte tejto pravdepodobnosti využívame normovanie takto:

$$P(x_1 < X < x_2) = P\left(\frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_2 - \mu}{\sigma}\right) = P(u_1 < U < u_2)$$

z vlastnosti distribučnej funkcie vyplýva :

$$P(u_1 < U < u_2) = F(u_2) - F(u_1)$$

Podobne postupujeme aj pri výpočte pravdepodobnosti toho, že náhodná premenná *X* je menšia než vopred zvolená konštanta *x*

$$P(X < x) = P\left(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right) = P(U < u) = F(u)$$

Pr. III.9. Architekt Imhotep a pisári faraóna 3.dynastie Starej ríše Džósera mali pri Sakkare veľa práce. K stavbe slávnej faraónovej pyramídy výrobcovia piva museli dodať aspoň 850 hl piva, aby bol zabezpečený jeho dostatok pre všetkých robotníkov, remeselníkov a dozor. Dlhodobo normálne dodávali priemerne 1000 hl denne s rozptylom 100 hl. Faraón Džóser sa však začal venovať svojej obľúbenej kratochvíli, vojenskému ťaženiu k susedom, čo odčerpalo pivovarníkom takmer každého piateho človeka z mužských pracovných síl a tým sa znížila produkcia o 16% a rozptyl sa zvýšil na 225. Pisári museli zistiť, či je dostatočná pravdepodobnosť, že dodávky piva budú aj cez vojnu aspoň 850 hl, inak by sa celý systém stavby najväčšej faraónovej pýchy zrútil resp. by museli prejsť na jednoduchší variant stupňovitej pyramídy podobnej babylonským zikkuratom, ktorý vyžadoval výrazne menšie pracovné nasadenie.

Starí Egypťania nejaké to tisícročie pred Kristom dotiahli celý proces výroby piva, ale aj prepracovaný systém jeho rozdeľovania, takmer k dokonalosti. A niet sa čo diviť. Veď skúste si v najhorúcejšom lete postaviť doma na záhradke takú celkom malú pyramídku, povedzme 50 krát 50 metrov s výškou 25 metrov, s čím medzi pyramídami veľmi nevyniknete, ale uvidíte, respektíve pocítite tú neodolateľnú chuť na pivo.

Máme priemernú dennú dodávku piva μ = 1000 hl. Rozptyl E(X) = 100 teda smerodajná odchýlka σ = 10. Priemer nových denných dodávok sa zníži na 840 hl s rozptylom 225. Aká je pravdepodobnosť, že denné dodávky počas vojny neklesnú pod 850 hl?

μ = 840

σ = 15

p(X≥850) = ?

Normovanie normálneho rozdelenia sa robí cez transformáciu $u = \frac{x-\mu}{\sigma}$.

Distribučná funkcia F(u) v intervale $u_1 = -\infty$ až $u_2 = \frac{x-\mu}{\sigma}$ má hodnoty 0 a (850-840)/15=0,67.

Ich rozdiel podľa [III.4.] je pravdepodobnosť v našom prípade, že veľkosť dodávky piva bude v intervale pod 850 hl denne: $p(-\infty<x<x_2) = F(0,67) - F(-\infty) = F(0,67)$

V štatistických tabuľkách napr. [13] nájdeme pre hodnotu u = 0,67 pravdepodobnosť $p(x\leq850) = 0,7486$. Pravdepodobnosť, že dodávky budú nad 850 hl piva denne je potom $p(X>850) = 1 – 0,7486 = 0,2514$.

Architekt Imhotep s pisármi s hrôzou zistili, že vojna znížila pravdepodobnosť potrebných dodávok piva len málo nad 25%, teda len asi každá 4.dodávka bude postačujúca, inak budú pracovníci na pyramíde hynúť od smädu; a keďže nechceli byť zaživa do pyramídy zamurovaní, ihneď prešli na úspornejší zikkuratový variant. Dodnes ho obdivujú archeológovia a turisti.

Pr. III.10. Učeníci celú noc lovili ryby, ale nič nechytili. Nadránom sedeli smutní, hladní a zamyslení na brehu, keď sa Šimon Peter ozval:

„Ako je to možné? Veď bežne sme mali úlovok okolo 150 rýb plus mínus 10. Aj Pán hovoril, že je to normálne, že tak to zariadil jeho nebeský Otec. Viete, čo? Ja mu verím, skúsim to ešte raz...“

Tomáš pochybovačne prehodil:

„Myslíte, že máme nejakú nádej chytiť 142 až 153 rýb?“

Ale keď videl ako sa všetci, aj keď s ochkaním a stonaním pozbierali a nasledovali Petra, pridal sa k ním.

Skúsme spoločne spočítať Tomášovi nádej na úlovok, teda pravdepodobnosť:

$\mu = 150$ - stredná hodnota všetkých úlovkov v rybolovoch v minulosti

$\sigma = 10$ - smerodajná odchýlka rybolovu

$p(142\leq X\leq153) = ?$

Pravdepodobnosť, že pri uvedených parametroch a normálnom rozdelení lovu chytia ryby v počte medzi 142 až 153 vypočítame nasledovne:

Hranice intervalu sú 142 a 153. Vypočítame pre ne transformovanú hodnotu distribučnej funkcie podľa [III.14]: $u = \frac{x-\mu}{\sigma}$

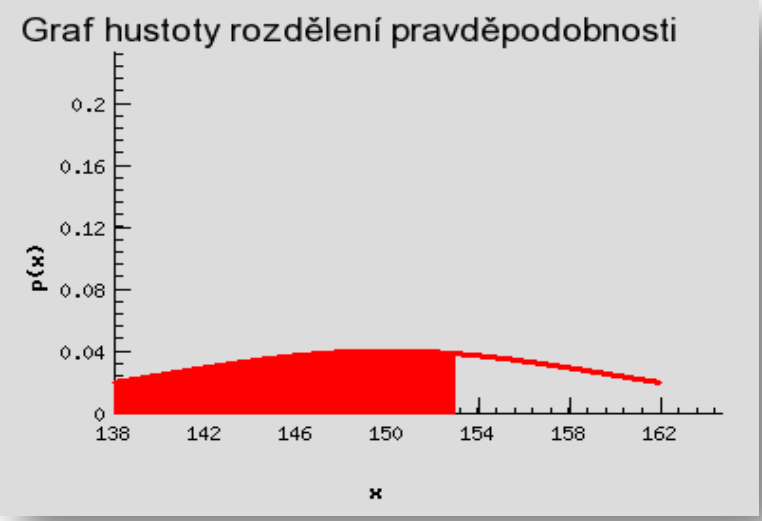
$u_1 = (142-150)/10 = -0,8$; $u_2 = (153-150)/10 = 0,3$

$p(142 \leq X \leq 153) = p[(142-150)/10 \leq (X-150)/10 \leq (153-150)/10] = p(-0,8 \leq u_x \leq 0,3) = F(0,3) - F(-0,8)$.

Vieme, že pravdepodobnosť v tomto intervale zistíme z tabuliek distribučnej funkcie normovaného normálneho rozdelenia, keď najprv vypočítame hodnoty distribučnej funkcie pre pravú hranicu intervalu a pre ľavú hranicu a potom ich odčítame. Výsledok odčítania je hľadaná pravdepodobnosť.

Hodnotu pravdepodobnosti z distribučnej funkcie pravej strany intervalu F(0,3) v tabuľke nájdeme: $p(0,3) = 0,6179$

Predstavuje to plochu pod krivkou hustoty pravdepodobnosti od -∞ do hodnoty 153 ako je na obr. III.21(zobrazuje sa len od hodnoty 138, pretože na to, aby sa zobrazil interval od -∞, by sme potrebovali trochu väčší obrázok, dokonca až nekonečne veľký):



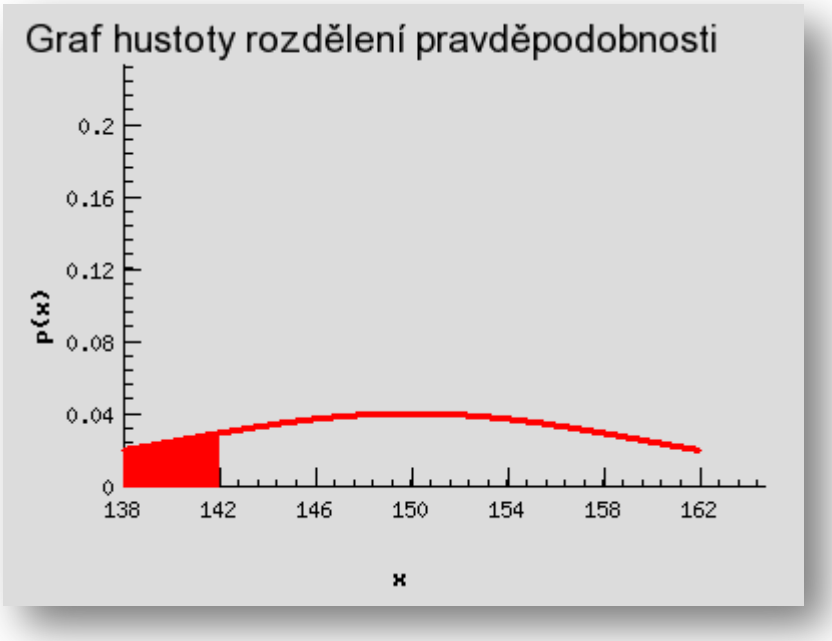
Obr. III.21. Zobrazenie pravdepodobnosti hodnôt úlovkov rýb x ≤ 153.

Pravidlo pre záporný argument pre symetrickú funkciu akou je normálne rozdelenie:

$$F(-x) = 1-F(x) \qquad [III.16]$$

teda v našom prípade $F(u_1) = F(-0,8) = 1 - F(0,8)$, kde v tabuľkách pre u = 0,8 dostaneme 0,7881 a pre pravdepodobnosť dostaneme $p(-0,8) = 1 - 0,7881 = 0,2119$

Predstavuje to plochu pod krivkou hustoty pravdepodobnosti od -∞ do hodnoty 142 ako je na obr. III.22:

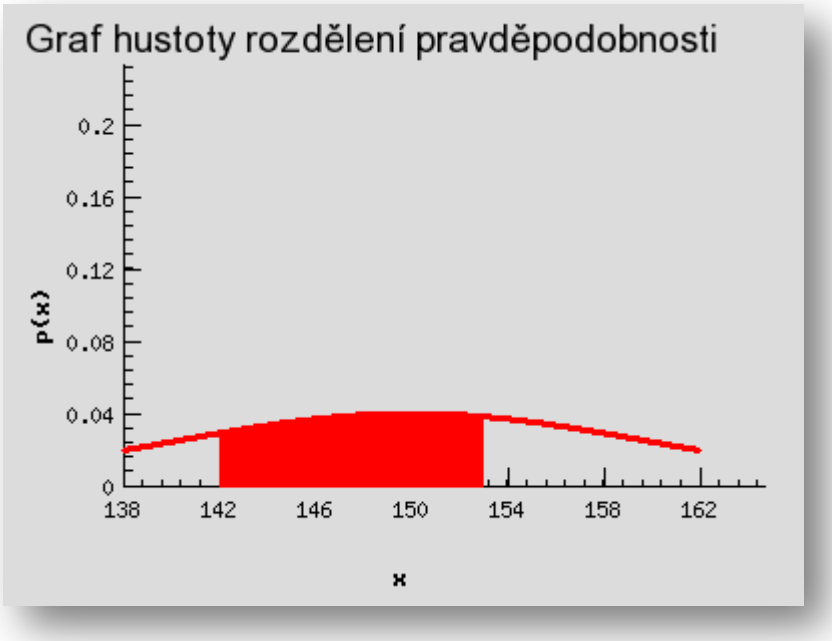


Obr. III.22. Zobrazenie pravdepodobnosti hodnôt úlovkov rýb x≤142.

Výslednú pravdepodobnosť úlovku medzi 142 a 153 rybami dostaneme ich odčítaním:

p(142≤X≤153) = $F(u_2) - F(u_1) = F(0,3) – F(-0,8) = 0,6179 – 0,2119 =$ **0,406**

Graficky:



Obr. III.23. Zobrazenie pravdepodobnosti hodnôt úlovkov rýb v intervale 142≤x≤153.

Pre Tomáša 40,6% pravdepodobnosť nebola nič moc a na lov išiel s pocitom, že aby to ulovili, musel by sa stať zázrak.

Ako to dopadlo, môžete sa dočítať v Jánovom evanjeliu: Jn 21, 1-12. **[15]**

Pr. III.11. V zariadeniach sociálnych služieb pre seniorov v okrese sa nám darí ročne umiestniť približne jedného z desiatich žiadateľov, pričom počty umiestnených sú náhodnou veličinou s normálnym rozdelením a odchýlkou ± 5. Umiestnili sme v tomto roku len 30 žiadateľov. Aký je priemerný počet voľných miest v zariadeniach sociálnych služieb pre seniorov v okrese?

p= 0,1

$\sigma = 5$

X = 30

$\mu = ?$

Pravdepodobnosti p = 0,1 v tabuľkách zodpovedá hodnota distribučnej funkcie pre interval od 0 po 30; F(u)=0,1. Hodnoty distribučnej funkcie sú tabelované až od 0,5 (pre pravdepodobnosti ≥ 0). Pre nižšie hodnoty máme ďalší algoritmus výpočtu (platí len pre F<0,5):

$$\mathbf{F(u) = 1 – F(-u)}$$

Dosadením dostávame $0,1 = 1 – F(-u)$ a z toho

$$F(-u) = 0,9$$

Pre distribučnú funkciu (pravdepodobnosť) 0,9 dostaneme z tabuliek hodnotu 1,28, teda pre našu veličinu **-u = 1,28** alebo po vynásobení (-1): **u = -1,28**. Z normovania [III.14] $u = \frac{X-\mu}{\sigma}$ úpravou dostaneme:

$$\mu = X - \sigma . u \qquad \textbf{[III.17]}$$

Rovnicu, v ktorej neznámou je μ postupne upravíme

$\mu = 30-5.(-1,28) = 30+6,4 \cong 36$.

V požiadavke na nadriadený orgán môžeme uviesť, že ročne priemerne umiestnime v zariadeniach sociálnych služieb 36 seniorov, skutočná potreba je 10-násobná.

••

Pozn.: Súhrn vzťahov, ktoré platia pre **hustotu** (označovanú *f(x)*) a **distribučnú funkciu** (označovanú *F(x)*) **normálneho normovaného rozdelenia**

- *f(x) = f(-x) (vyplýva zo symetrickosti okolo μ)*
- *F(-x) = 1- F(x)*
- P(│X│< x) = P(-x < X < x) = F(x) - F(-x) = 2F(x)-1 *(pravdepodobnosť symetrického intervalu)*
- P(-∞< x < ∞) = 1 *(pravdepodobnosť pod celou zvonovou krivkou je 1)*

Z týchto dôvodov sú tabuľky normálneho normovaného rozdelenia tabelované iba pre hodnoty x ≥ 0.

••

Verím, že to boli trochu náročnejšie výpočty, ale nezúfajte, nie ste nejakí extrémne slabí počtári, ono je to naozaj trochu náročnejšie. Teda ste normálni, už aj v štatistickom ponímaní. Uvedené príklady boli modelové, kde ste sa mohli naučiť, ako s normálnym Gauss-Laplaceoovým rozdelením pracovať. Určite tak ako v celej štatistike, je aj tu potrebné využiť radosť z tvorivého racionálneho myslenia, teda vedieť približne, čo asi tak robíme. Okrem toho ste sa naučili používať štatistické tabuľky a rutinné výpočty za vás hravo zvládne váš počítač. V pr. III.9. a III.10 ste mohli využiť program EXCEL a funkcie

NORMDINST(x,μ,σ,1) pre výpočet hodnoty distribučnej funkcie

NORMDINST(x,μ,σ,0) pre výpočet hustoty rozdelenia

NORMSDINST(x) pre výpočet hodnoty normovanej distribučnej funkcie **(0,1).**

Taktiež úspešne sa dajú využiť online kalkulátory dostupné na webe. Na obr. III.19 a III.20 bolo znázornené, aký % podiel hodnôt sa dostane do bežne používaného **1σ**, **2σ** a **3σ-** intervalu. Hodnoty mimo nich (chvostíky pod zvonom úplne vľavo alebo úplne vpravo) sú extrémne hodnoty ďaleko od **μ**. Napr. pri sledovaní výšky v populácii sú to extrémne malí, alebo extrémne vysokí jedinci, a pod.

Normálne rozdelenie nás odteraz bude sprevádzať už navždy. Neskôr sa dozvieme viac o jeho charakteristikách a jeho vlastnosti budeme využívať v najrôznejších prípadoch.

2. Exponenciálne rozdelenie: Je to zaujímavé spojité rozdelenie pravdepodobnosti náhodnej veličiny s jedným parametrom **λ**, ktorý môžeme charakterizovať napr. ako priemerný počet výskytov náhodnej udalosti za jednotku času:

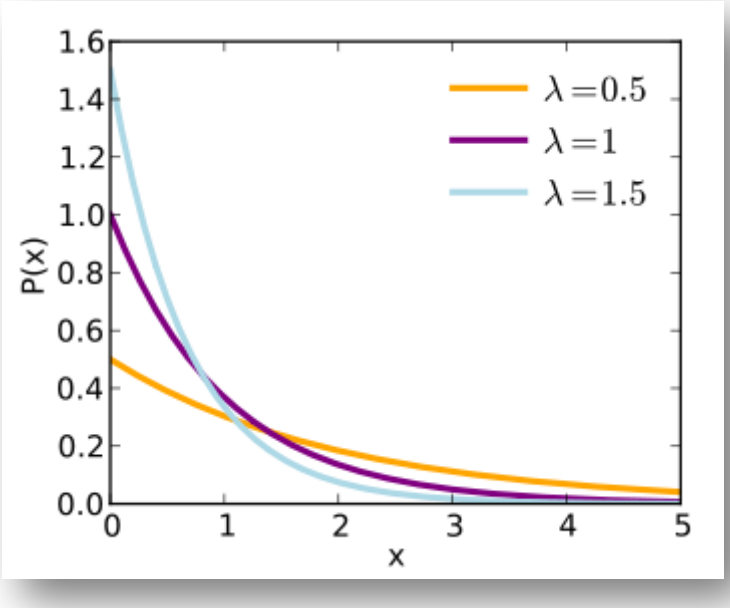
$$f(x) = \lambda . e^{-\lambda x} \quad pre\ x \geq 0 \qquad [III.18]$$

Stredná hodnota a disperzia exponenciálneho rozdelenia je

$$E(x) = 1/\lambda \qquad [III.19]$$

$$D(x) = 1/\lambda^2 \qquad [III.20]$$

Kde sa s ním môžeme stretnúť? Popisuje ako vhodný model viaceré systémy, napr. čakanie u lekára, na zaparkovanie, v systémoch hromadnej obsluhy atď. Dá sa ním často modelovať doba životnosti, čakanie na výskyt nejakej náhodnej udalosti, čas trvania nejakého javu a pod. Vplyv veľkosti parametra **λ** na priebeh rozdelenia hustoty pravdepodobnosti je na obr. III.24. [16].

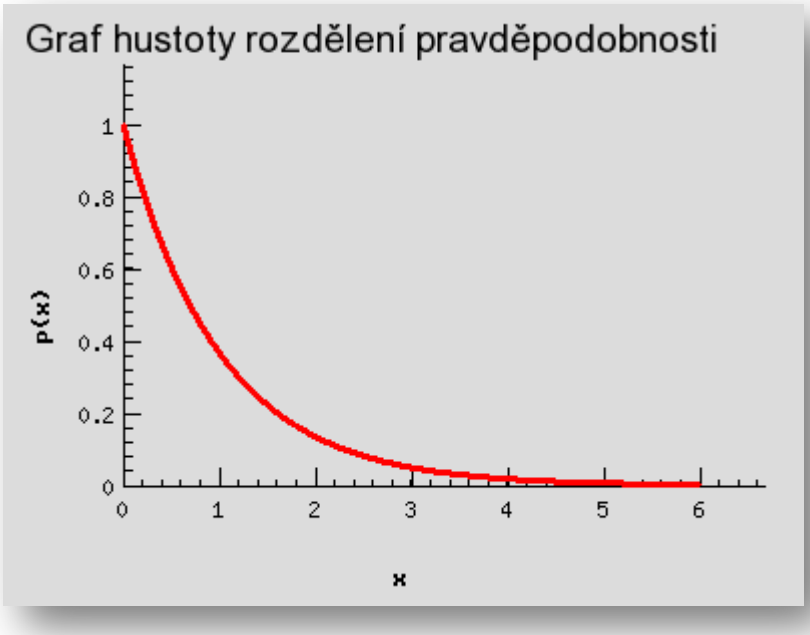


Obr.III.24. Tvar funkcie hustoty pravdepodobnosti exponenciálneho rozdelenia pre rôzne parametre λ.

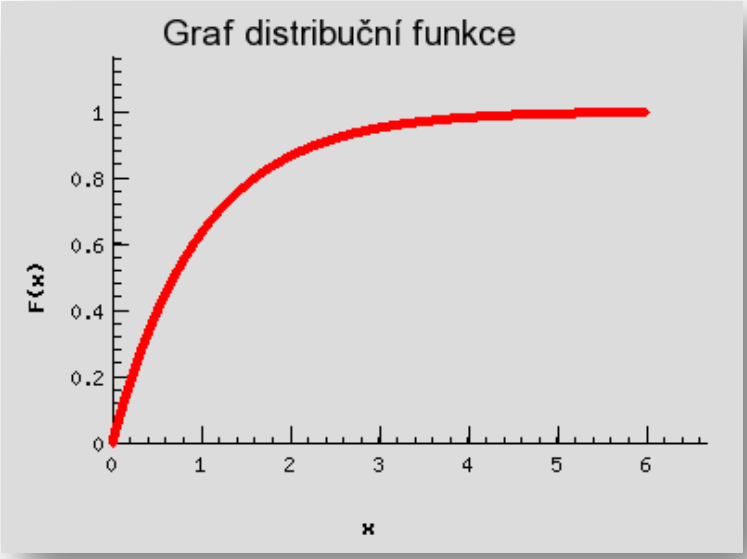
Vzťah pre distribučnú funkciu exponenciálneho rozdelenia:

$$F(x) = 1 - e^{-\lambda x} \quad pre\ x \geq 0 \qquad [III.21]$$

Na obr. III.25. a III.26. je znázornený graf pravdepodobnostnej a graf distribučnej funkcie exponenciálneho rozdelenia pravdepodobnosti



Obr.III.25. Graf pravdepodobnostnej funkcie exponenciálneho rozdelenia



Obr.III.26. Graf distribučnej funkcie exponenciálneho rozdelenia

V EXCELI sú pre výpočet k dispozícii funkcie:

EXPONDIST(x, λ, 0) – funkcia rozdelenia hustoty pravdepodobnosti

EXPONDIST(x, λ, 1) – distribučná funkcia

Exponenciálne rozdelenie nie je symetrické okolo strednej hodnoty ako normálne.

Pr. III.12. Pouličná výdajňa dostáva sterilizované injekčné striekačky v cykle, ktorého dĺžka je náhodná premenná s exponenciálnym rozdelením pravdepodobnosti a so strednou hodnotou E(x) = 8 dní. Do výdajne priviezli zásielku, takže majú teraz zásobu asi na 12,5 dňa. Je pravdepodobné, že im to bude stačiť?

Zo vzťahu [III.19] si vypočítame parameter λ:

λ = 1/E(x) = 1/8 = 0,125

x = 12,5

Pravdepodobnosť, že zásielka vydrží 12,5 dňa je ako vždy rozdiel hodnoty distribučnej funkcie F(12,5) – F(-∞) = F(12,5) – 0 = F(12,5). A že sa zásoba do 12,5 dní vyčerpá bude potom 1 - F(12,5), takže podľa **[III.21]**:

$$1 - F(x) = 1 - F(12,5) = 1 - \left(1 - e^{-\lambda x}\right) = e^{-0,125.12,5} \cong 0,21$$

Dostali sme asi 21% pravdepodobnosť, že sa zásielka injekčných striekačiek vyčerpá skôr, ako je plánovaný čas využitia zásob. Samozrejme, že sa mi nechcelo počítať a využil som webovský online kalkulátor [9].

Pr. III.13. [17] Keď sa ročné dieťa nechá bez dozoru, dokáže rozbiť, alebo pokaziť priemerne 4 hračky za hodinu. Ako dlho ho môže mama nechať bez dozoru, aby s pravdepodobnosťou 0,9 nenastal problém?

Zrejme potrebujeme zistiť čas, za ktorý nedôjde k rozbitiu hračky s pravdepodobnosťou 0.9 a vieme, že ak za časovú jednotku zvolíme t = 1 hod, tak $\lambda = 4$. Pravdepodobnosť nezničenia hračky je $p(X \leq t)$, teda nejaká pravdepodobnosť v čase menšom ako čas **t**, ktorý hľadáme. Pravdepodobnosť zničenia hračky 0,9 v čase **t** je potom:

$p(X > t) = 1 - p(X \leq t) = 1 - F(t) = 0.9$

Pretože $1 - F(t) = 0,9$, tak $F(t) = 0,1$

a dosadíme do [III.21], dostávame

$1 - e^{-4t} = 0.1$

To je síce pekný vzťah, ale ako z neho vydolujeme, teda vypočítame **t** ? V I. kapitole sme si vraveli niečo o operáciách a funkciách a o ich inverzných tvaroch. Povedali sme si, že inverzná operácia k sčítaniu je odčítanie, k umocňovaniu odmocňovanie a inverzná funkcia k exponenciálnej funkcii, akú tu máme, je logaritmická funkcia. Keby sme išli do toho, čo sme si povedali trochu hlbšie, dostali by sme nasledovné pravidlá:

1. Ak máme funkciu e^x, tak jej logaritmus bude

$$\mathbf{\ln e^x = x}$$

2. Pravidlo pre sčítanie logaritmov:

$\ln x + \ln y = \ln (x.y)$; **$\ln x - \ln y = \ln (x/y)$**

Teraz už môžeme vzťah $1- e^{-4t} = 0.1$ upraviť (odčítame 1, vynásobíme (-1) a logaritmujeme):

$-4.t = \ln 0,9 \rightarrow t = \ln (0.9) /-4.= 0.0263$

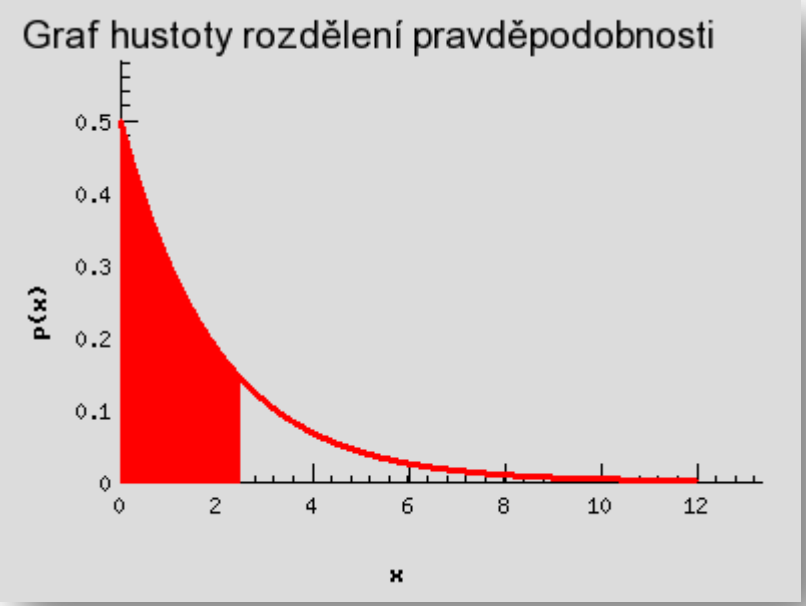
Čas t= 0.0263 hod., čo je približne 1 min 35 s.

Pr. III.14. Ak túžite po veľkom šťastí na skúške zo štatistiky bez prílišnej intelektuálnej námahy, existuje zaručený recept: Ľahnite si 12. augusta večer do trávy a keď bude bezoblačná teplá a vlahá noc a nebudú vás štípať komáre, uvidíte „padať hviezdy“. Astronómovia ich volajú Perzeidy, ale je všeobecne známe, že keď si niečo veľmi prajete keď padá

hviezda, tak sa vám to takmer určite splní. Pokiaľ je frekvencia „padania hviezd“ asi 5 za 10 minút, aká je pravdepodobnosť, že v priebehu najbližších dva a pol minúty zachytíte padajúcu hviezdu a stihnete si počas jej pádu zapriať úspech na skúške?

Hľadáme pravdepodobnosť p(x≤ 2,5), kde pri časovej jednotke [min] máme λ = ½ = 0,5

p(x≤ 2,5) = F(2,5) – F(-∞) = F(2,5) = $1 - e^{-\lambda x} = 1 - e^{-0,5.2,5} = 0,7135$



Obr.III.27. Graf znázorňujúci pravdepodobnosť šťastia na skúške zo štatistiky.

Pokiaľ ste dokázali samostatne vypočítať pravdepodobnosť vášho úspechu na skúške zo štatistiky p = 71,35 %, je dosť veľká pravdepodobnosť, že na nej úspech naozaj budete mať.

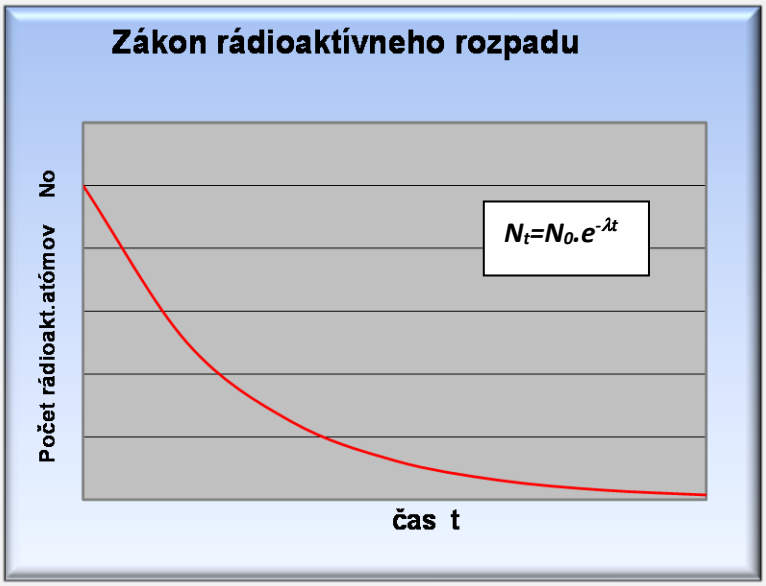
Pr. III.15. Alexander Litvinenko sa snažil. Bol jedným z najaktívnejších agentov KGB a neskôr jej ruskej nástupnickej FSB a za neohrozený boj s organizovaným zločinom dostal vysoké vyznamenania. Ako prenikal stále vyššie, zistil prepojenia medzi vedúcimi špičkami FSB, ale aj celej krajiny vrátane prezidenta Putina a jeho suity s mafiou a organizovaným zločinom. A povedal to. Po niekoľkonásobnom zatknutí a fraškách v podobe súdov mu táto výstraha stačila a v roku 2000 emigroval do Veľkej Británie, kde dostal azyl. V novembri 2006 sa stretol s krajanmi v londýnskom sushi-bare, krátko po tomto stretnutí zomrel. Britská tajná služba MI5 bola prekvapená, prečo zdravý štyridsiatnik na vrchole fyzických a psychických síl tak rýchlo odišiel a začala to vyšetrovať. Po odbere biologických vzoriek veľmi skoro zistili ich

silnú α-rádioaktivitu. Na začiatku namerali detektorom α-žiarenia 10000 impulzov a potom počas 10 dní merali aktivitu vzorky s výsledkami uvedenými v tab. III.9:

Deň merania	Impulzy	Hodnota λ
0	10000	
1	9950	0,005013
2	9900	0,005025
3	9851	0,005004
4	9802	0,005000
5	9753	0,005002
6	9704	0,005008
7	9656	0,005001
8	9607	0,005012
9	9559	0,005011
10	9512	0,005003
	Priemer	**0,005008**

Tab.III.9: Meranie rádioaktivity biologickej vzorky s vyhodnotením parametra λ.

Čo vlastne špecialisti MI5 zisťovali? Vedeli už, že majú do činenia s α-žiaričom. Aby ho mohli identifikovať, museli zistiť jeho dôležitú charakteristiku – polčas rozpadu T, čo predstavuje čas, za ktorý sa náhodne rozpadne polovica atómov rádionuklidu. Zákon rádioaktívneho rozpadu manželov Curie je znázornený na obr.III.28 **[18]**:



Obr.III.28. Zákon rádioaktívneho rozpadu

V tomto zákone $\mathbf{N_0}$ je počet atómov na začiatku merania. Je úmerný počtu impulzov, ktoré zachytí detektor, v našom prípade N_0= 10000. $\mathbf{N_t=N_0.e^{-\lambda t}}$ je potom počet im pulzov v čase **t**. Vidíme, že rádioaktívny rozpad má exponenciálne rozdelenie pravdepodobnosti s parametrom λ, nazvaným v tomto prípade rozpadová konštanta. Keďže vieme už trochu pracovať s exponenciálnymi a logaritmickými funkciami, odvodíme si z tohto zákona vzťah pre výpočet parametra λ

$$\lambda = -\frac{\ln \frac{N_t}{N_0}}{t}$$

a vypočítané hodnoty z neho pre rôzne časy t sme uviedli v poslednom stĺpci tab.III.9. aj s priemernou hodnotou λ_p = 0,005008. Ako zistili pracovníci MI5 u fyzikov, polčas rozpadu je vo vzťahu s rozpadovou konštantou nasledovne:

$$T = \frac{ln\,2}{\lambda}$$

Mimochodom, toto by ste dokázali odvodiť aj vy, keby ste si do zákona rádioaktívneho rozpadu dosadili za $\mathbf{N_t=N_0/2}$ a keby sa vám chcelo niečo pre tajné služby odvodzovať. Agenti MI5 mali tiež celkom dobrú kalkulačku, tak im vyšlo T = 138,4 dňa. Takýto polčas z α-žiaričov má vysokotoxický a nebezpečný rádionuklid polónium-210 (^{210}Po). Jeho stopy našli napodiv aj v dotyčnom sushi-bare. Našťastie ruská strana vydala okamžité dementi, že s tým nemá nič spoločného a je naďalej za mierové spolužitie všetkých ľudí a národov.

V štatistike budeme potrebovať a prakticky využívať ešte jeden typ rozdelení pravdepodobnosti, tzv. výberové rozdelenia, o ktorých si viac prezradíme, keď začnú byť aktuálne:

C: Výberové rozdelenia

1. χ^2 rozdelenie s **k** stupňami voľnosti $\chi^2(k)$

$E[\chi^2(k)] = k$

$D[\chi^2(k)] = 2k$

2. Studentovo t-rozdelenie s **k** stupňami voľnosti

E(T)=0

D(T) = k/(k-2)

3. Fisherovo F-rozdelenie s k1 a k2 stupňami voľnosti

E(F) = k2/(k2-2) pre k2>2

Rozptyl F(k1,k2) je

$D(F) =(2k2^2.(k1+k2-2)/k1.(k2-2)^2.(k2-4)$ pre k2>4

Skončíme rozdelenie rozdelení Caesarovým a Machiavelliho výrokom: **Divide et impera!**

Literatúra k III. kapitole:

[1] Jilemnický, P.: Kompas v nás, Pravda, Bratislava, 1973

[2] http://cs.wikipedia.org/wiki/Rozd%C4%9Blen%C3%AD_pravd%C4%9Bpodobnosti

[3] Sviteková,K.: *Zastúpenie krvných skupín*. Národná transfúzna služba SR, 2.11.2005,

[4] Magnello, E., Van Loon, B.: Seznámte se...statistika, Portál, Praha 2010

[5] http://cs.wikipedia.org/wiki/Binomick%C3%A9_rozd%C4%9Blen%C3%AD

[6] http://www.elektro-energetika.cz/calculations/bi.php

[7] http://www.elektro-energetika.cz/calculations/calculator.php?language=

[8] http://cs.wikipedia.org/wiki/Poissonovo_rozd%C4%9Blen%C3%AD

[9] http://www.elektro-energetika.cz/calculations/po.php

[10] Skuhra, J.M.: Kurz pravdepodobnosti, in http://www.george11.eu/matematika/pst/K1.htm

[11] Zvarová, J.: Základy statistiky pro biomedicínske odbory, EuroMISE centrum, Ústav informatiky AV ČR, Praha 2006

[12] http://cs.wikipedia.org/wiki/Rozd%C4%9Blen%C3%AD_pravd%C4%9Bpodobnosti

[13] http://www.fhi.sk/files/katedry/ks/tabulky.pdf

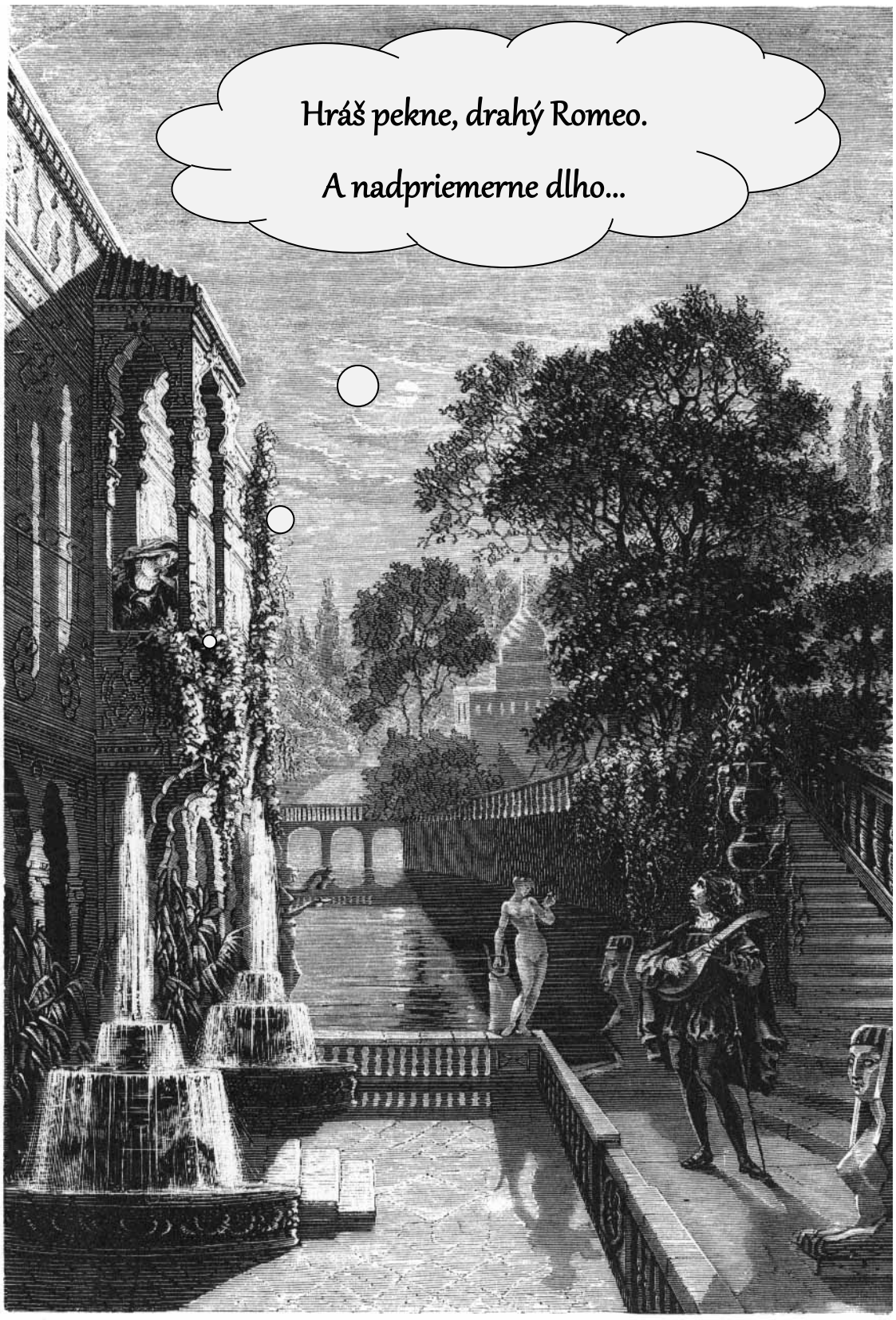
[14] Zamarovský, V.: Za siedmimi divmi sveta, Perfekt, Bratislava 2003

[15] Nový zákon s komentármi Jeruzalemskej Biblie, Dobrá kniha, Trnava, 2008

[16] http://cs.wikipedia.org/wiki/Exponenci%C3%A1ln%C3%AD_rozd%C4%9Blen%C3%AD

[17] http://www.umat.feec.vutbr.cz/~hlinena/INM/prednasky/prednaska11_2008.pdf

[18] S.Usačev J. Chrapan, M. Chudý, J. Vanovič: Experimentálna jadrová fyzika, ALFA Bratislava, SNTL Praha, 1982


Hráš pekne, drahý Romeo.
A nadpriemerne dlho...

IV. Chvála priemeru alebo štatistika v posteli

Väčšina ľudí je vždy na strane väčšiny.
Tomáš Janovic

Podľa jednej starej gréckej povesti žil v nehostinnom divokom kraji starobylej Attiky v údolí rieky Kéfis istý Prokrustes. Prevádzkoval penzión pre pocestných, ktorí do týchto miest náhodou zablúdili alebo si chceli skrátiť cestu. Pozýval ich k sebe, ponúkal všetkým pohostenie a odpočinok. V dobrej nálade pred uložením k spánku Prokrustes pocestného vždy požiadal, aby si vyskúšal lôžko, ktoré nechal vyhotoviť podľa priemernej výšky starovekého Gréka, aby vedel, kde ho má ubytovať. Dramatická zápletka nastala, keď si pocestný na lôžko ľahol. Pokiaľ mu bola krátka, Prokrustes ho ponaťahoval až na jeho dĺžku, čo bývalo dosť fatálne. Pokiaľ cez posteľ prečnieval, boli mu odťaté nohy, hlava alebo podľa nálady oboje, a Prokrustes sa zmocnil mobilného majetku, ktorý mal pri sebe. Túto lásku k priemeru mu pokazil až bájny hrdina Théseus cestou do Atén a neskôr k Minotaurovi. Bol to priemerný Grék, keď si ľahol na lôžko a na veľké prekvapenie mu sedelo ako uliate. Využil moment prekvapenia, na lôžko umiestnil samotného lupiča a okamžite mu odťal prečnievajúcu hlavu čím ukončil jeho živnosť.

Máme nielen radosť z bohatstva výdobytkov, ktoré zanechala našej civilizácii krétsko-mykénska kultúra starého Grécka, ale aj z *priemeru*, s ktorým sme sa práve stretli. Je to jeden zo základných štatistických pojmov, kde intuitívne cítime, o čo asi ide. Pomáha nám zorientovať sa v záplave rôznych čísel, ktoré sa na nás denne valia zo všetkých strán a pekne nám ich zredukuje na jednu hodnotu, ktorú považujeme za reprezentatívnu. Je to aj odraz trochu statickej a zjednodušenej predstavy o svete, že prvky v ňom majú svoju ideálnu povahu, ku ktorej sa všetko ostatné, odchyľujúce sa snaží priblížiť. Rozmanitých štatistík, ktoré sa nás snažia buď o niečom poučiť, alebo za nejakým účelom mystifikovať, sú plné noviny, časopisy, ale aj web a iné elektronické média. Ich prevažujúcou charakteristikou je zneužívanie priemeru a zahmlenie resp. zatajovanie variability. A tak sa okrem úradných štatistík o sobášoch, rozvodoch, narodeniach a úmrtiach stretávame najčastejšie s informáciami o kriminalite,

vraždách, vojnách, hlade, úpadku a pod., ktoré keby neslúžili len na zvýšenie pozornosti a na získavanie reklamy, ale boli naozaj pravdivé, asi by boli prejavom úplne iného sveta, než v akom žijeme. Ťažko si predstaviť, že by reportéri a kameramani komerčnej TV opustili tragický požiar bytu, v ktorom zhorelo niekoľko detí, a vtrhli radšej do domácnosti, v ktorej sa celý deň nič mimoriadneho nestalo, rodina ho prežila s trochou starosti a radosti, povinnosti, učenia a zábavy, aby si zmorená bežnou únavou išla ľahnúť večer pokojne spať, len preto, že je to neporovnateľne častejší jav. Tak začínate mať pocit, ba až presvedčenie, že vo vašom meste, možno priamo v štvrti, je veľmi vysoká kriminalita a ľudia, ktorí ju s vami obývajú sú mimoriadne nebezpeční a úchylní. A po večerných televíznych správach už viete, že by ste radšej ani nemali vyjsť na ulicu.

Takže ako sa v tom vyznať a zorientovať? Nie je to až také ťažké, väčšina ľudí má na to vyvinuté intuitívne štatistické cítenie a vieme, že sa z prívalu negatívnych štatistík nezblázni. Ale ešte lepšie je pozrieť sa, čo štatistika vlastne je. Múdri ľudia vravia, že je to veda. Tým sa nenechajme zastrašiť. Lepšie je, keď nám povedia, že sa zaoberá hromadnými javmi, využíva počet pravdepodobnosti a hľadá v nich nejaké zákonitosti. Má samozrejme svoju terminológiu:
Experiment – v štatistike si pod experimentom predstavujeme zber dát za istým účelom: niečo sa chceme dozvedieť, naučiť, alebo dokonca chceme niečo objaviť. V prírodných vedách, v technike a niekedy aj inde to môžu byť merania. V humanitných vedách, v medicíne a pod. získavame údaje napr. o výške, váhe, veku respondentov, o vzdelaní, príjmoch, stravovacích návykoch, o spôsobe dopravy a o mnohých iných veličinách v závislosti od zamerania štatistickej analýzy. Experiment predstavuje dostatočne presné, správne a reprodukovateľné resp. overené získanie takéhoto údaju.
Premenná – je veličina, ktorá môže nadobúdať nejaké hodnoty, alebo je možné ich podľa určitých pravidiel vypočítať. Premenná v štatistike môže nadobúdať jednu alebo aj viac hodnôt, nazývame ju aj **náhodná premenná**, ktorá môže určité hodnoty nadobúdať s istou pravdepodobnosťou. Delíme ich na **diskrétne**, ktoré môžu nadobúdať len určité presné hodnoty a je ich vždy konečný počet, napr. údaj, ktorú politickú stranu budem preferovať v najbližších voľbách, alebo hodnotenie študentov na skúške. **Spojité premenné** majú možnosť nadobúdať nekonečne mnoho hodnôt z nejakého intervalu a ich zmena je plynulá napr. výška populácie, príjem rodín, hodnota krvného tlaku pre veľké súbory.
Štatistický súbor – určitá množina ľudí resp. položiek, predmetov, úkazov, ktoré sú predmetom skúmania (študenti nejakej školy, nakupujúci v obchodných centrách, všetci občania krajiny).

Základný súbor (populácia) – súbor všetkých možných sledovaných prvkov nazývaných **štatistické jednotky**. Napr. všetkých obyvateľov SR pri sčítaní ľudu.

Výberový súbor – je podmnožinou populácie. Náklady a čas na zber štatistických údajov sú úmerné veľkosti sledovaného štatistického súboru. Preto je najčastejšie nutné nahradiť základný súbor podstatne menším výberovým súborom, ktorý ho v sledovanej vlastnosti reprezentuje. Veľká časť problémov štatistiky je zabezpečenie, aby výberový súbor bol správny, teda reprezentatívny. Výsledky skúmania výberového súboru chceme zovšeobecniť na základný súbor. Napr. náhodný výber respondentov pri prieskume verejnej mienky, stratifikovaný resp. účelový výber z nejakej homogénnej podskupiny – skúmanie trávenia voľného času lekárov budeme skúmať na vzorke medicínsky vzdelaných pracovníkov. Výber, ktorým je celá populácia nazývame aj *cenzus*. Keď nejaký psychiater konštatuje, že 90% ľudí sú hysterickí neurotici, môže vás to na chvíľu zaskočiť. Potom si však uvedomíte, že ľudia, s ktorými sa dotyčný psychiater stretáva sú vo veľkej väčšine pacienti, iní psychiatri a jeho vlastná rodina, teda, že jeho výberový súbor na takéto štatistické hodnotenie nebol v žiadnom prípade reprezentatívny a môžete jeho tvrdenie hystericky či neuroticky odmietnuť.

Štatistický jav (znak) – sledovaná vlastnosť štatistického súboru, napr. výška, príjem, farba očí, spotreba alkoholu a i.

Početnosť – koľkokrát sa nejaký výsledok určitého štatistického javu v štatistickom súbore vyskytol. Môže byť **absolútna** (skutočný počet výskytov), alebo **relatívna** v %.

Rozdelenie – popisuje súbor možných hodnôt, ktoré môže náhodná premenná nadobúdať. Zobrazujeme tabuľkou alebo grafom.

Pod **štatistikou** často rozumieme nielen vedu, ale aj konkrétnu vlastnosť súboru, o ktorú sa zaujímame, napr. úroveň nezamestnanosti v nejakej oblasti, najvyššia spotreba alkoholu, násilie v rodine za určité obdobie a i.

V štatistike sa vždy budeme zaoberať **hromadnými javmi (štatistickými znakmi)**, pričom potichu predpokladáme, že ich štatistické charakteristiky sú v istej miere stabilné. Uvedomujeme si, že sa vyjadrujeme trochu vágne, alebo aspoň nepresne, ale precíznejší spôsob popisu by vyžadoval matematický formalizmus, ktorý by spôsobil kolektívny útek čitateľov do nebezpečných neprebádaných džunglí. Ale netreba sa strachovať, že zostaneme len u populistického nič nehovoriaceho popisu štatistickej činnosti, určite si mnohé potrebné pojmy ešte trochu hlbšie preštudujeme. Dostávame sa však na pôdu, kde okrem niektorých základných metód štatistickej analýzy, ktoré sa môžeme naučiť, sa musíme pokúsiť ku každému tvrdeniu zaujať postoj kritického rozumu. Vraví sa, že vlastný experiment, vlastné overenie alebo

prepočítanie je lepšie ako názor 1000 odborníkov. Niektoré rizikové okruhy aplikovania štatistiky:

1. Štatistické spracovanie je správne, interpretácia výsledkov je nesprávna alebo neúplná. Napr. metadónova substitučná liečba drogovej závislosti a jej vyhodnotenie v niektorom roku zdravotníckymi pracovníkmi je úplne iná štatistika, ako jej interpretácia cez bulvárne média, ktoré tvrdili, že predstavuje nádej na takmer úplnú redukciu závislosti na heroíne.

2. Výberový súbor je malý, neúmyselne. Je to profesionálne zlyhanie. V [1] autor uvádza príklad testovania vakcíny proti detskej obrne. Zaočkovaných bolo 450 detí, 680 detí bol kontrolný súbor. Výsledok, že ani jedno zo zaočkovaných detí nedostalo obrnu sa zdal byť skvelý. Aj štatistické súbory vyzerali dosť veľké. Avšak sa zistilo, že ani v kontrolnom súbore nedošlo ani k jedinému nakazeniu obrnou a prevalencia tejto choroby v populácii bola tak nízka (asi 0,002), že na nejaké zmysluplné štatistické výsledky by bolo potrebné 15 až 20 krát viac detí. Celý test aj jeho štatistické vyhodnotenie, často s presnosťou na mnoho desatinných miest (!) bol od počiatku nezmysel. Veľkosťou výberového súboru sa budeme ešte kvantitatívne zaoberať neskôr.

Iný príklad, keď je výberový súbor malý úmyselne. V malom súbore sa odchýlka od očakávanej strednej hodnoty môže vyskytnúť častejšie. Keď hodím 10 krát mincou, môže sa ľahko stať, že mi 8 krát padne *orol*, pričom očakávaná pravdepodobnosť je 0,5. Keby som hodil 1000 krát, alebo 1000000 krát, stredná hodnota by sa už veľmi blížila k 0,5. Tak potom vznikajú štatistiky, že nový prací prášok má o 30% vyšší prací účinok (ponecháme bokom iné problémy štatistickej analýzy ako je napr. oproti čomu, veľkosť súboru, metodika, atď.), že prípravok na výživu kĺbov zaberá v 95% prípadov, že nový margarín s vyšším obsahom 3-ω mastných kyselín vám zvýši inteligenciu o 20% a pod. To už je klamanie.

3. Za štatistikou sa skrývajú politické a iné záujmy, samotná štatistika je nezmysel. Pred rokom 1989 bývali voľby do Národného frontu, kde kandidátov vybraných a schválených vedúcou a jedinou politickou stranou vždy zvolilo 99,7 až 99,8% oprávnených voličov, čo vraj potvrdzovalo dôveru, jednotu a zomknutosť pracujúceho ľudu s politickým vedením krajiny. Ak sa vám to zdá byť zábavné, pozrime sa na referendum v roku 2014 na polostrove Krym, kde žije asi 60% Rusov, 30% Ukrajincov, 5% Krymských Tatárov a 5% iné. Referendum pod samopalmi za pripojenie k Rusku skončilo vraj s výsledkom: 97,5% bolo za pripojenie.

4. Uvádzaná je priemerná hodnota, v ktorej sa môže skryť variabilita súboru. Takmer vždy ide o klamanie priemerom. Napr. keď si fandíme, že priemerná spotreba alkoholu u nás opäť poklesla o 0,12% oproti minulému obdobiu na 13 litrov čistého alkoholu na osobu za rok. Nič to nehovorí o alkoholizme mužov a žien, neplnoletých, o úmrtnosti na alkohol, o rozvratoch

rodín atď. Ani o tom, ako boli stanovené také pekné presné čísla, o metodike, výberovom štatistickom súbore atď.

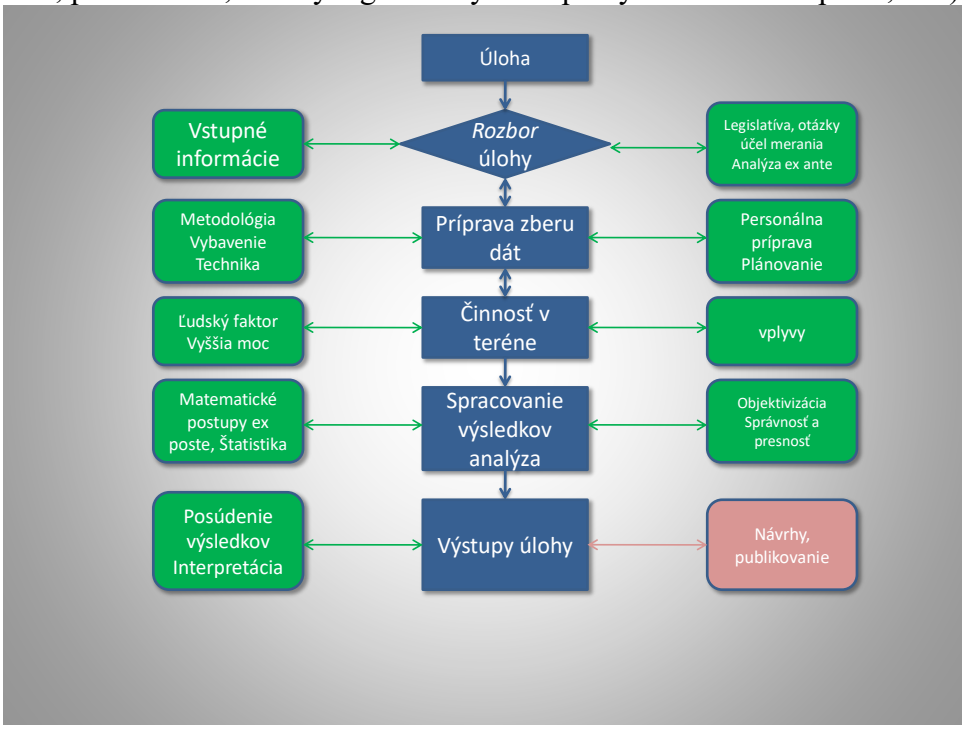
5. Skreslený výber. Veľmi častá chyba aj profesionálnych štatistikov, skreslená je potom aj interpretácia výsledkov. Napr. sledovanie riešenia niektorých javov v problematike spolužitia s rómskou menšinou. Ako urobiť správny a reprezentatívny výberový súbor? Zahrnúť predovšetkým respondentov Rómov, alebo nerómov, ľudí z konfliktnej oblasti, alebo z akademickej obce atď. Asi si uvedomujeme ťažkosti tohto problému. Štatistika, ktorá by aj numericky potvrdila, že u nás majú ženy nadštandardné práva a k žiadnemu domácemu násiliu nedochádza by mohla naznačovať vysokú úroveň našej krajiny v naznačenej oblasti, keby výberový súbor nepozostával z mužov, návštevníkov pohostinských zariadení nižšej cenovej kategórie.

Je ešte viacero oblastí, ktoré by sa dali nazvať mínové polia štatistiky, pri niektorých sa zastavíme neskôr. Opäť si však zdôraznime, že štatistika nerobí výklad sveta, len popisuje nejakou metodikou isté javy. V prípade emocionálne výrazne ladených javov sa ľahko môže stať, že sa bude zamieňať výsledok štatistickej analýzy s postojom jej autora. Je to isté upozornenie na to, že je potrebné dávať si pozor na interpretáciu výsledkov štatistickej analýzy, nepodľahnúť mámeniu skĺznuť do nejakých obľúbených novinárskych slovných zvratov a začať vykladať to, čo štatistická analýza netvrdí. Aj keď si štatistik dáva pozor, musí mať niekedy istú dávku odvahy na svoj výskum. Môže sa stať, že autorka štatistickej práce o homosexuálnom správaní časti študujúcej mládeže, alebo vo väzniciach bude označená za lesbičku napriek tomu, že žije po všetkých stránkach v klasickom manželstve so 6 deťmi bez akéhokoľvek vnútorného odporu alebo premáhania sa. Alebo výskumník, ktorý hľadá postoje väčšinového obyvateľstva krajiny k menšinám bude nakoniec niektorými politikmi velebený, inými subjektmi bude označovaný za rasistu a xenofóba. Ako vidíme, začíname klásť na terénneho pracovníka - výskumníka dosť veľké nároky.

Pri kvantitatívnom výskume v teréne sa budeme zaoberať ľuďmi, ktorí majú rôzne vlastnosti, postoje, sociálne vzťahy, záujmy a motivácie a len niektoré javy je možné spriemerovať. Je na to veľa názorov, ale nám postačí, keď si štatistický prieskum zhrnieme do 4 bodov [2]:

1. Formulácia cieľov, analýza informácií pred začiatkom – *ex ante*: Čo a o kom (o čom) chceme zistiť?
2. Práca v teréne: Zber dát.
3. Štatistická analýza Spracovanie získaných údajov z terénu *ex post*. Výsledkom je potom informácia.

4. Vyhodnotenie informácie. Dostaneme nové poznanie, prezentácia výsledkov (publikovanie, prednášanie, návrhy legislatívnych resp. iných zmien z zlepšení, iné.)



Obr.IV.1. Schéma štatistickej analýzy terénnych dát – vývojový diagram

V rámci experimentov v teréne získame viac-menej komplikovaný súbor údajov. Tieto sa potom pomocou štatistickej analýzy snažíme sprehľadniť. K tomuto účelu slúži časť štatistiky, nazvaná **popisná (deskriptívna) štatistika**.

Druhá časť štatistiky – **induktívna štatistika** – nazývaná v literatúre aj štatistika inferencií sa zaoberá vzťahom výberového a základného súboru, teda podmienkami a pravidlami, aby sme mohli na základe vlastnosti výberového súboru niečo povedať o základnom súbore.

V tejto kapitole sa spolu pozrieme najprv na to ako popisná štatistika pracuje s **charakteristikami polohy**. V ďalšom sa budeme zaoberať **charakteristikami variability**. Tieto charakteristiky patria nerozlučne spolu, výpovedná hodnota výsledkov aj najjednoduchších štatistických analýz má zmysel, keď sú spoločne uvedené. Pokiaľ sa niečo z nich neuvedie, ide buď o chybu pri štatistickom spracovaní a interpretácií výsledkov, alebo o úmyselné zahmlievanie ich skutočnej informácie. Rozdelenie charakteristík do rôznych kapitol bude len preto, aby sa nám to trochu nepoplietlo, aby sme mohli látku radšej po troche ľahšie zvládnuť. Nič nám však nebráni vrátiť sa k ilustračným príkladom a vždy, keď nás premkne neobyčajná radosť zo získania novej štatistickej vedomosti, aplikovať ju spätne, práve naopak, svedčilo by to o vašej nadpriemernej usilovnosti.

Štatistickú analýzu údajov z terénu začíname ich tabuľkovým, pripadne grafickým spracovaním ako sme si naznačili už v predchádzajúcej kapitole. Dostali sme nejaký súbor dát a pokúsime sa z neho vydolovať užitočné štatistické charakteristiky. Charakteristiky polohy sú nasledovné stredné hodnoty E(x) (pozri aj vzťahy **[III.6]**, **[III.9]**, **[III.11]** a **[III.19]**), teda úroveň štatistického znaku, okolo ktorej sú hodnoty znaku viac alebo menej koncentrované:

a) priemery

– aritmetický $\bar{x}$

– geometrický $\bar{x}_g$

– harmonický $\bar{x}_h$

b) medián ***Med(x)***

c) modus ***Mod(x)***

Aritmetický priemer:

S aritmetickým priemerom nebudete mať žiadne problémy. Je to najčastejšie používaná charakteristika polohy a vypočítava sa zo všetkých hodnôt súboru (to platí aj pre ostatné priemery) tak, že sčítame všetky hodnoty štatistického znaku a výsledok vydelíme ich počtom **n**, teda **rozsahom súboru**:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \textbf{[IV. 1]}$$

Pr.IV.1.: Máte k dispozícii 5 vodičov s rovnakými služobnými osobnými autami a viac-menej aj s rovnakou vyťaženosťou (najazdenými km). Sledovali ste počas 20 pracovných dní mesiaca ich spotrebu pohonných hmôt (PHM), aby ste zabránili zbytočnému plytvaniu. Výsledky aj s vypočítanými aritmetickými priemermi sú v tab. IV.1.:

č.	**Spotreba PHM/prac. deň**															$\bar{x}$
1.	18,2	17,3	15,5	19,8	21,2	18,1	14,4	18,0	21,5	20,7	18,6	11,4	21,5	19,9	20,8	**18,5**
2.	16,4	22,2	17,9	17,8	17,8	21,1	20,2	15,6	18,7	19,4	16,5	21,6	21,3	18,1	17,3	**18,8**
3.	21,6	14,6	19,9	17,6	17,0	17,0	22,3	12,8	17,9	15,6	18,7	19,3	17,9	16,5	18,4	**17,8**
4.	31,3	33,3	28,7	32,5	27,4	29,8	26,8	35,6	33,2	33,6	33,8	33,2	22,5	29,3	28,2	**30,6**
5.	20,0	12,5	12,5	22,2	19,8	18,6	21,3	19,9	19,9	21,2	15,3	19,5	18,7	18,2	16,3	**18,4**

Z vašej štatistickej analýzy vyplýva, že budete v 1 prípade musieť urobiť opatrenia: Dať urobiť mimoriadnu servisnú kontrolu automobilu, prípadne ukončiť s jedným vodičom pracovno-právny pomer.

Pr.IV.2.: Máme tri skupiny študentov sociálnej antropológie (s počtami 6, 9 a 4) zameraných na misijnú činnosť, z ktorých sa bude vyberať pracovná skupina pre činnosť v strednej Afrike, kde budú komunikovať v anglickom jazyku. Na misii budú spolupracovať, tak sa dá odborne predpokladať, že je postačujúce, keď priemerný zisk bodov z testu na skúške z angličtiny bude aspoň 45. Výsledky testov boli nasledovne:

Rozsah 1.súboru n = 6; 2. n = 9 a 3. n = 4.

Sledovaný štatistický znak je počet bodov v teste z anglického jazyka.

1. skupina:

č.	1	2	3	4	5	6
body	38	50	43	49	48	55

2.skupina

č.	1	2	3	4	5	6	7	8	9
body	26	31	53	48	47	33	28	36	47

3. skupina:

č.	1	2	3	4
body	47	52	43	59

Pre každú skupinu vypočítame aritmetický priemer bodov z testu z angličtiny podľa **[IV.1]**.

Pre 1.skupinu:

$$\overline{x1} = \frac{x_1 + x_2 + \cdots + x_6}{n} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{283}{6} \cong 47$$

Podobne:

$$\overline{x2} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{349}{9} \cong 39$$

$$\overline{x3} = \frac{x_1 + x_2 + \cdots + x_4}{n} = \frac{\sum_{i=1}^{4} x_i}{n} = \frac{201}{4} \cong 50$$

Keby sme vypočítali teraz jednoduchý aritmetický priemer z výsledkov všetkých 3 skupín, dostali by sme celkom uspokojivý, ale zavádzajúci výsledok

$$\overline{x} = \frac{\overline{x1} + \overline{x2} + \overline{x3}}{3} = \frac{\sum_{i=1}^{3} \overline{xi}}{n} = \frac{137}{3} \cong 45{,}7$$

Keďže skupiny majú nerovnakú veľkosť, prispievajú do spoločného priemeru rôznou váhou. „Váženie“ urobíme tak, že výsledok každej skupiny vynásobíme počtom študentov, sčítame a súčet vydelíme celkovým počtom študentov:

$$\overline{x} = \frac{\overline{x1}.n1 + \overline{x2}.n2 + \overline{x3}.n3}{n1 + n2 + n3} = \frac{\sum_{i=1}^{3} \overline{xi}.ni}{n} = \frac{833}{19} \cong 43{,}8$$

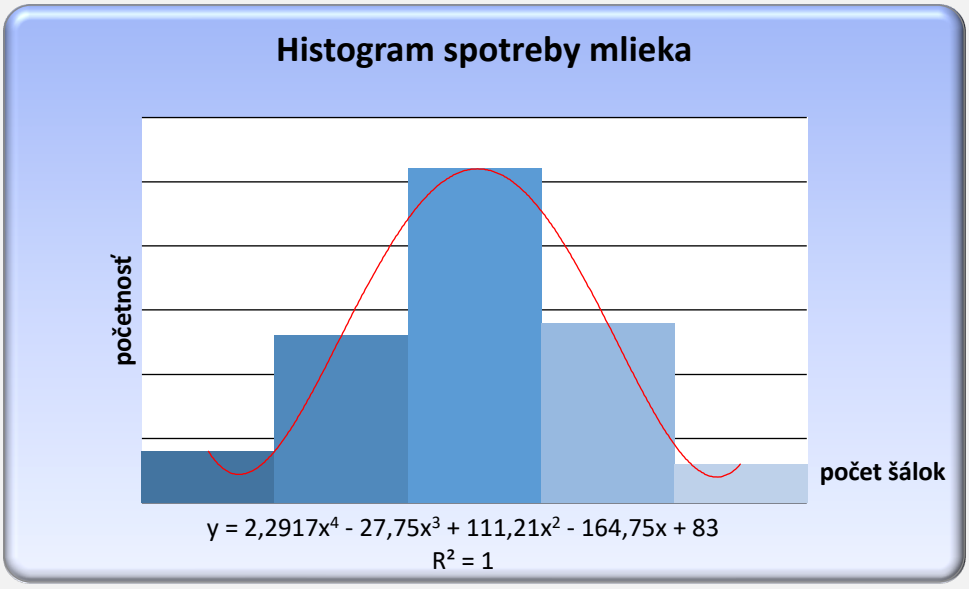
Okrem toho, že celému ročníku odporučíme ešte jeden semester intenzívneho kurzu anglického jazyka, odvodili sme si ďalší dobrý vzťah pre vážený aritmetický priemer. Pre **k** –súborov s počtom prvkov **n1, n2, ..., nk** máme **vážený aritmetický priemer:**

$$\overline{x} = \frac{\overline{x1}.n1 + \overline{x2}.n2 + \cdots + \overline{xk}.nk}{n1 + n2 + n3} \qquad \textbf{[IV.2]}$$

Pr.IV.3.: Do predškolského zariadenia pri imigrantskom tábore so 60 deťmi zabezpečujeme dovoz mlieka z darov okolitých farmárov. S veľkým úsilím sa dá doviezť asi 40 litrov denne. Požiadame rehoľné sestry, ktoré sa o zariadenie a deti v ňom starajú, aby nám podľa možnosti zistili priemernú spotrebu mlieka detí denne. Sestričky sa ako vždy usmiali a na naše prekvapenie nám na ďalší deň odovzdali výsledky svojho šetrenia. Ako štatisticky dobre podkuté terénne pracovníčky to urobili jednoducho: Vedeli, že deti pijú mlieko pod ich dohľadom po celých 2 dl šálkach, poznali ich, tak si všetky rozdelili do skupín podľa počtu vypitých šálok mlieka za deň nasledovne:

Počet šálok x_i	**0**	**1**	**2**	**3**	**4**	**Spolu**
Početnosť n_i	**4**	**13**	**26**	**14**	**3**	**60**

s celkom pekným viac-menej symetrickým histogramom:



Obr.IV.2. Histogram dennej spotreby mlieka

Potom už bez problémov vypočítali vážený aritmetický priemer vypitých šálok mlieka na dieťa za deň:

$$\bar{x} = \frac{x_1.n_1 + x_2.n_2 + \cdots + x_k.n_k}{n_1 + n_2 + \cdots + n_k} \qquad \textbf{[IV.3]}$$

$$\bar{x} = \frac{119}{60} = 1,983333 \cong 2$$

Priemerný počet vypitých šálok na dieťa za deň je 2. Celkovo sa vypije 2 x 60 x 0,2=24 litrov. To sa dalo už bez väčších problémov zabezpečiť.

Aj pri takej jednoduchej štatistike, akou je aritmetický priemer je možné uletieť, preto je vhodné uviesť a zapamätať si **vlastnosti aritmetického priemeru**:

- Jednoduchosť, ľahký výpočet.
- Vypočítava sa zo všetkých hodnôt súboru, aj preto je najčastejšou a najpoužívanejšou charakteristikou polohy.
- Citlivosť voči extrémnym hodnotám, tie ho môžu posunúť mimo oblasť väčšiny dát.
- Vyžaduje unimodálne a symetrické rozdelenie.
- Pri bimodálnom alebo viacvrcholovom rozdelení je potrebné uprednostniť **modus**, o ktorom si niečo povieme o chvíľu.
- Asymetrické rozdelenie - uprednostniť **medián**.
- Stálosť súčtu hodnôt. Ak jednotlivé hodnoty znaku v súbore nahradíme $\bar{x}$, súčet hodnôt zostane nezmenený.
- Ak ku každej hodnote znaku $\mathbf{x_i}$ pripočítame (odpočítame) rovnaké číslo $\mathbf{k}\varepsilon R$, zväčší (zmenší) sa o dané číslo **k** aj aritmetický priemer $\bar{x}$.
- Súčet odchýlok všetkých hodnôt znaku od priemeru sa rovná nule (symetria):

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \cdots + (x_n - \bar{x}) = \sum_{i=1}^{n} (x_i - \bar{x}) = 0 \qquad \textbf{[IV.4]}$$

- súčet druhých mocnín (štvorcov) odchýlok všetkých hodnôt znaku od ich priemeru je najmenšie možné číslo a je menšie ako súčet druhých mocnín (štvorcov) odchýlok všetkých hodnôt znaku od akejkoľvek inej hodnoty $\mathbf{X \neq \bar{x}}$ (najjednoduchšie vysvetlenie a použitie metódy najmenších štvorcov):

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 < \sum_{i=1}^{n} (x_i - X)^2 \qquad \textbf{[IV.5]}$$

Geometrický priemer $\bar{x}_g$

V mnohých výpočtoch ekonomického charakteru, pri hodnotení rastu resp. tempa rastu (poklesu) nejakej veličiny, pri geometrickom raste (poklese) hodnoty a exponenciálnom rozdelení početnosti náhodnej premennej, teda pri výpočte priemernej veľkosti zmeny by aritmetický priemer dával často nezmysly. Používa sa **geometrický priemer** $\bar{x}_g$ definovaný ako n-tá odmocnina zo súčinu všetkých nenulových hodnôt štatistického znaku získaného súboru:

$$\bar{x}_g = \sqrt[n]{x_1 . x_2 \dots . x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} \quad kde \;\; x_i > 0 \qquad [IV.6]$$

Pokiaľ dáme do tabuľky usporiadané hodnoty početnosti f_i, môžeme vážený geometrický priemer vypočítať

$$\bar{x}_g = \sqrt[n]{x_1^{f_1} . x_2^{f_2} \dots . x_k^{f_k}} \; ; \quad pričom \sum_{i=1}^{k} f_i = n \qquad [IV.7]$$

Pr.IV.3.: V Starom zákone sa môžeme dočítať, že do Egypta za čias Jakuba a Jozefa prišla jedna rodina, ktorej výberovú vzorku predstavovali muži v počte 75 (**Gn 46, 8-27**). Podľa úvodu k Piatim knihám Mojžišovým **[3]** ho opúšťali asi po 450 rokoch. Vzorka populácie, ktorá vyšla z otroctva z Egypta je veľmi približne 600 000 mužov (Nm 1, 46). Za týchto okolností, keď sa uspokojíme s počtom mužov ako vhodným výberovým súborom (iný nemáme), môžeme odhadnúť priemerný koeficient rastu židovského národa v Egypte za rok:

Označme priemerný koeficient ročného rastu populácie **k.** Na začiatku bolo Jakubovej rodiny 75 mužov, po 450 rokoch ich bolo v národe asi 600 000.

	$\mathbf{n_0} = 75$ na $\mathbf{n_1} = k.75$
V druhom roku	$\mathbf{n_2} = \mathbf{k} . \mathbf{n_1} = \mathbf{k} . \mathbf{k} . 75 =$
	$\mathbf{k^2}.75$
V treťom roku potom	$\mathbf{n_3} = \mathbf{k^3}.75$, atď.
V 450. roku teda	$\mathbf{n_{450}} = \mathbf{k^{450}}. 75 = 600000$

Úpravou dostaneme

$$k = \sqrt[450]{\frac{600000}{75}} = \sqrt[450]{8000} = 1,020172$$

Koeficient priemerného rastu populácie vyvoleného národa v Egypte bol k = 1,020172.

Pozn.: Použite radšej EXCEL, kalkulačka nemusí vždy postačovať.

Pr.IV.4.: Populárny úžerník *Harpoš* stonal ako keby ho na nože brali. Vykladal takmer s plačom svojej návšteve, aké má nehorázne výdavky, musí platiť ochranku, svoj podiel si pýtajú aj vymáhači dlžôb, aby odviedli svoju prácu poriadne a dlžníka len exemplárne dokaličili a nie zabili; aj sám musí reprezentovať a pritom ho aj jeho ženské niečo stoja, ceny stále rastú, je toho na jedného obetavého jedinca naozaj dosť! Ale aby návštevník nepovedal, že pri všetkých svojich starostiach nevidí problémy druhých, predsa sa mu len rozhodol niečo požičať, ale najviac 100.- €. Pritom nie je ako banka, ktorá pýta 14%! Jemu stačia 2% denne. Návštevník si z hlavy po dlhej dobe spočítal, že 2% zo 100 € sú 2 €, takže keď mu zajtra vráti 102.- €, bude to v poriadku. Neskôr to bude trochu viac, ale tým sa už nezaoberal, vzal pôžičku a odišiel. Harpoš si spokojne mädlil ruky:

Vstupná suma je 100.- €. Úrok je 2% denne, koeficient nárastu je **k = 1,02**.

Po 1.dni narastie suma na $n_1 = k.100 = 1{,}02.100 = 102.-€$

Po druhom dni to bude $n_2 = k.n_1 = k.k.100 = k^2 . 100 = 1{,}0404 . 100 = 104{,}04.-$; nič hrozné sa zatiaľ nestalo.

Po n-tom dní bude dlžná suma $k^n . 100$.

Ak napríklad bude môcť dlžník vrátiť dlžobu po mesiaci, nech je to 30 dní:

Suma $= k^{30} . 100 = 1{,}02^{30}. 100 = 1{,}81136 * 100 = 181{,}14$.- €. Začína to byť pre *Harpoša* zaujímavé. A ak to chce vrátiť až po roku (365 dní):

Dlžná suma $= 1{,}02^{365}.100 = 1377{,}4083 . 100 = 137740{,}83$.- €.

Našťastie to nie je prestupný rok.

Pr.IV.5.: Ešte jeden príklad z oblasti pracovného prostredia a pre tých, čo majú radi trochu silnejšie zbrane: Pri rozštiepení atómu uránu ^{235}U sa okrem veľkého množstva energie uvoľnia tri neutróny, ktoré sú potrebné na štiepenie ďalších jadier. Neutróny, tak ako každé iné častice a žiarenie majú nejaký dolet v danom materiály. Keď sú rozmery štiepneho materiálu menšie ako je priemerný dolet neutrónov v ňom, väčšina uvoľnených neutrónov opustí materiál bez zachytenia a teda bez ďalšieho štiepenia. Takejto hmotnosti hovoríme z hľadiska jadrovej bezpečnosti podkritické množstvo. Hmotnosť, pri ktorej začína spontánne štiepenie je kritické množstvo a akákoľvek väčšia čiže nadkritická hmotnosť už zabezpečuje reťazovú reakciu. Pre tých, ktorí si to chcú vyskúšať doma, takýto jednoduchý je princíp atómovej bomby: Vezmeme dve pologule podkritického množstva štiepneho materiálu a dostatočne rýchlo ich priblížime k sebe, aby vzniklo nadkritické množstvo, v ktorom prebehne reťazová štiepna reakcia v priebehu zlomku sekundy, takže treba dosť rýchlo utekať.

Teoreticky sa dá kadečo vypočítať, ale určiť kritickú hmotnosť štiepneho materiálu pre bombu tak, aby sa dalo s ňou bezpečne narábať a aby explodovala len vtedy, keď na to bola požiadavka bolo náročné a muselo sa to spresniť experimentálne. V americkom stredisku vývoja atómovej bomby v Los Alamos na tom pracoval tím kanadského biofyzika a chemika Louisa Alexandera Slotina (1910 – 1946), veľmi dobrodružne založeného muža, ktorý bol prezývaný aj *hlavný zbrojár Spojených štátov*, keďže pripravil výbušné jadro prvej pokusnej ako aj prvých dvoch v Japonsku, vojensky použitých amerických bômb.

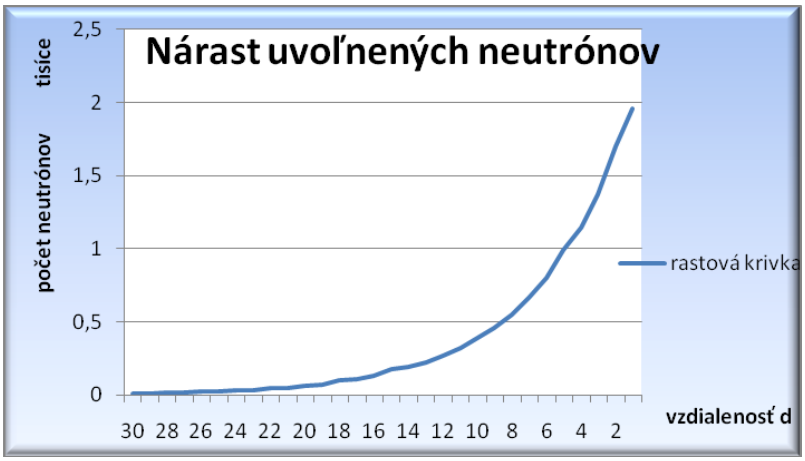
Experiment stanovenia kritického množstva štiepneho materiálu spočíval v tom, že dve oddialené pologule napr. vysoko obohateného uránu, uložené v berýliovom obale (ktorý pôsobil ako neutrónové zrkadlo a vracal utekajúce neutróny späť, čím sa mohlo ušetriť značné množstvo veľmi drahého štiepneho materiálu a zvýšiť účinok) sa veľmi pomaly a veľmi trpezlivo približovali k sebe. Jednoducho ručným priťahovaním skrutiek a šróbov pomocou skrutkovača a zaznamenávala sa pomaly rastúca produkcia neutrónov. Toto flirtovanie s možnosťou spustenia štiepnej jadrovej reakcie nazval fyzik a nobelista Richard Feynman „šteklenie draka pod chvostom“. Keď experiment videl iný fyzik Enrico Fermi, bol zhrozený a ostro kritizoval Slotinovú nedbanlivosť a podľa neho nedostatočný pud sebazáchovy. Tvrdil, že stačí takto pokračovať v experimente a všetci účastníci projektu Manhattan budú do roka na pravde Božej.

Pri experimente sa musel zabezpečiť taký nárast vyžiarených a detektorom zachytených neutrónov v impulzoch, aby koeficient nárastu bol napr. k = 1,2. V tabuľke sú hodnoty experimentu a my môžeme tento koeficient vypočítať:

vzdialenosť	neutróny[imp]	vzdialenosť	neutróny[imp]	vzdialenosť	neutróny[imp]
30	10	20	65	10	393
29	12	19	74	9	460
28	15	18	99	8	552
27	19	17	107	7	662
26	24	16	128	6	799
25	25	15	174	5	994
24	33	14	195	4	1145
23	35	13	222	3	1370
22	46	12	266	2	1698
21	51	11	319	1	1958

Tab.IV.2. Počty detegovaných neutrónov pri rôznych vzdialenostiach podkritických množstiev štiepneho materiálu pri výrobe atómovej bomby.

Z obr. IV.3 vidíme, že je to exponenciálny nárast:



Obr.IV.3. Grafické znázornenie experimentu pri vývoji atómovej bomby

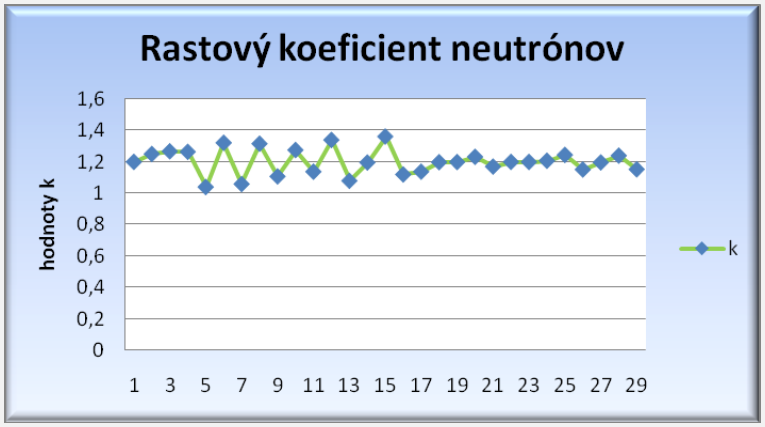
Medzi 1. a 2. hodnotou vzdialenosti, teda po prvom otočení skrutkovača sa zvýšil počet zachytených vyžiarených neutrónov z 10 na 12. Koeficient rastu vypočítame pre každý prípad ako:

$$k = \frac{n_{k+1}}{n_k} = \frac{12}{10} = 1,2$$

Hodnoty z experimentu sú v tab. IV.3:

i	1	2	3	4	5	6	7	8	9
	1,2	1,25	1,266667	1,263158	1,041667	1,32	1,060606	1,314286	1,108696
10	11	12	13	14	15	16	17	18	19
1,27451	1,138462	1,337838	1,080808	1,196262	1,359375	1,12069	1,138462	1,198198	1,199248
20	21	22	23	24	25	26	27	28	29
1,231975	1,170483	1,2	1,199275	1,206949	1,244055	1,151911	1,196507	1,239416	1,153121

Ako sa pohybuje hodnota rastového koeficientu okolo strednej hodnoty ukazuje obr.IV.4:



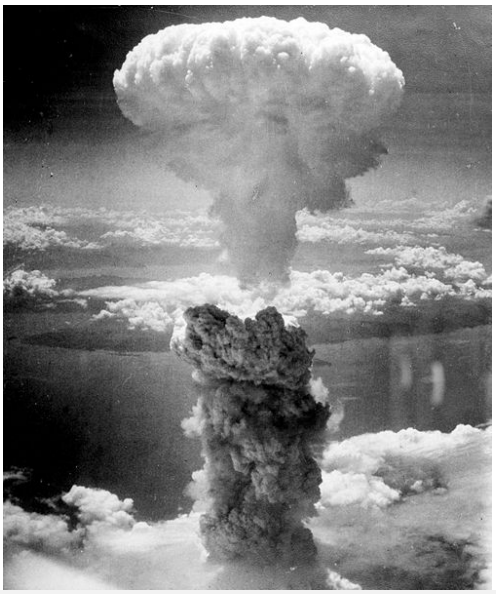
Obr.IV.4. Rozloženie hodnôt rastového koeficientu produkcie neutrónov pri hľadaní kritickej hmotnosti štiepneho materiálu pri vývoji atómovej bomby okolo strednej hodnoty.

Aritmetický priemer: $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} =$ **1,202159**

Geometrický priemer: $\bar{x}_g = \sqrt[n]{x_1 . x_2 \ldots . x_n}$ **=1,199577**

Aritmetický priemer by naznačoval na dosiahnutie kritického množstva, ale geometrický priemer odhalil, že je potrebné urobiť ešte úpravy, pretože rastová konštanta nedosahuje požadovanú hodnotu. Vráťme sa ešte na chvíľu k pracovnému prostrediu:

21. mája 1946 mal Slotin za sebou už viac ako štyridsať takýchto úspešných experimentov. Možno už pristupoval k práci rutinne, menej pozorne. O 15,20 hod. sa mu pošmykol skrutkovač a horná berýliová pologuľa sa prudko priblížila k dolnej, čo spôsobilo krátkodobé dosiahnutie kritickej hmotnosti. Všetci prítomní spozorovali modrý záblesk v miestnosti a závan horúčavy. Slotin inštinktívne rukami oddialil od seba pologule, čím zachránil ostatných spolupracovníkov prítomných pri experimente pred ďalším ožiarením. O svojom osude po masívnej dávke neutrónov a sprevádzajúceho γ-žiarenia nemal najmenšie pochybnosti. Nepodľahol panike, ale ešte pokiaľ mohol, vyzval všetkých, aby zaujali svoje miesta, kde sa nachádzali v čase nehody, nakreslil to na príručnú tabuľu, aby lekári a biofyzici neskôr mohli urobiť výpočet dávky ožiarenia každého z nich. Potom opustil budovu, sadol si na chodník a čakal privolanú zdravotnú pomoc. Začal zvracať, prejavovať zmätenosť a iné príznaky akútnej choroby z ožiarenia. Napriek takmer okamžitej lekárskej pomoci, transfúziám a iným zásahom zomrel po deviatich dňoch 30. mája, v prítomnosti svojich rodičov. Pochovaný je doma vo Winnipegu. [4]



Takže aj keď ste sa práve naučili vyrábať atómovú bombu, správajte sa prosím pri tom trochu opatrne.

Harmonický priemer $\bar{x}_h$

Použijeme ho, keď potrebujeme spriemerovať nejaké výkony.

$$\bar{x}_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \quad kde\ x_i > 0 \qquad [IV.8]$$

Vážený harmonický priemer, t.j. keď sú k dispozícii početnosti výskytu nejakého javu $\mathbf{f_i}$ v

sledovanom súbore vypočítame:

$$\bar{x}_h = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_k}{x_k}} = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}} \quad pričom \sum f_i = n \qquad [IV.9]$$

Pr.IV.6.: Pri vypuknutí epidémie s priemernou dennou úmrtnosťou 12 ľudí potrebujete zabezpečiť okamžité pochovávanie. Máte poruke 3 hrobárov, z ktorých prvý dokáže vykopať hrobové miesto za 3 hod., druhý za 4 hod. a tretí, najšikovnejší za 1,5 hod.. Dokážete s nimi zvládnuť kritickú situáciu pri nepretržitom kopaní 10 hod denne? Z nám už známych vzorcov a vzorca [IV.8] vypočítame priemery výkopu jedného hrobu jedným hrobárom:

$\bar{x}$ = **2,833**

$\bar{x}_g$ = **2,621**

$\bar{x}_h$ = **2,400**

Čas výkopu 12 hrobov 3 hrobármi vypočítame ako súčin (12xpriemerný čas)/3. Pre rôzne priemery dostaneme hodnoty:

E(x)	t[hod]
$\bar{x}$	11,33333
$\bar{x}_g$	10,48297
$\bar{x}_h$	9,6

Aritmetický i geometrický priemer dávajú hodnoty príliš vysoké. Harmonický priemer naznačuje, že by ste pochovávanie s 3 hrobármi mohli zvládnuť.

Medián ***Med(x), med(x)***

Pre súbory s extrémnymi hodnotami by priemery nemuseli byť charakteristické. Aj pri asymetrických rozdeleniach používame **medián Med (x)**, predstavujúci prostrednú hodnotu usporiadaného štatistického súboru, ktorý je potom vhodnejším ukazovateľom prevažujúcej tendencie ako aritmetický priemer. Prostredná hodnota nepárneho počtu prvkov nie je problém, rozdeľuje súbor na dve rovnaké polovice. Napr. pre 9 prvkov je to piata hodnota v usporiadanom rade (4 má vľavo a 4 vpravo od nej). Pri párnom počte prvkov máme dve prostredné hodnoty, tak urobíme ich aritmetický priemer. Štatistický súbor má len jeden medián. 50% hodnôt súboru je menších a 50% väčších ako medián. Patrí medzi tzv. **kvantily**, hodnoty deliace súbor na rovnaké časti. **Kvartily** delia súbor na 4 časti, **decily** na 10, **percentily** na 100. Pre súbory symetricky rozložené okolo aritmetického priemeru je

$$\bar{x} = med(x) = mod(x) \qquad [IV.10]$$

Pr.IV.7.: Kriminalisti sledovali sedem pneumológov z hľadiska vydávania receptov na niektoré druhy liekov. Pri prepočítaní na jednotný počet kapitovaných pacientov im vychádzal priemerný mesačný počet receptov na lieky s obsahom efedrínu nasledovne:

Lekár č.	1	2	3	4	5	6	7
Počet receptov	48	32	55	145	39	347	62

$\bar{x} = 104$

Ich štatistické cítenie im hovorilo, že to nie je charakteristická hodnota súboru. Usporiadali si súbor a charakterizovali ho radšej mediánom:

recepty	32	39	48	55	62	145	347

med(x) = 55

Niekde okolo mediánu by sa mala pohybovať hodnota vydaných receptov. Nebolo ťažké pri dvoch lekároch objaviť prepojenie na drogovú komunitu so sofistikovanou výrobou a distribúciou pervitínu.

Pr.IV.8.: Pri sledovaní príjmu seniorov z hľadiska kvality ich života bolo náhodne vybraných 10 mesačných starobných dôchodcov s nasledujúcimi usporiadanými výsledkami (zaokrúhlené na celé €):

dôchodca	1	2	3	4	5	6	7	8	9	10
dôchodok	136	252	160	311	320	322	340	420	956	1428

$\bar{x} = 454$

med(x) = 321

Oficiálna hranica chudoby: X = 346.- €

x<X: 70%

Vašou úlohou je odborne odhadnúť, ktoré hodnoty boli uvedené ako výsledky sociálnej politiky (štátu, vlády, politickej elity akéhokoľvek zafarbenia atď. dosaďte si podľa vlastného uváženia a potreby sa emočne excitovať) a ktoré údaje boli zamlčané.

Modus ***Mod(x), mod(x)***

Predstavuje najčastejšie sa vyskytujúcu hodnotu znaku v štatistickom súbore, najväčšiu početnosť výskytu, ktorá ho typicky charakterizuje. Napr. typická veľkosť oblekov pre odevný priemysel na výrobu konfekcie, typická cena chleba v nejakom regióne a i. Pri viacvrcholovom rozdelení môžeme dostať viac modusov súboru.

Pr.IV.9.: Máte v utečeneckom tábore 30 rodín, potrebujete ich vybaviť stanmi s určitým počtom lôžok. Rodiny majú rôzny počet členov, početnosti sú v tabuľke:

Počet členov v rodine	1	2	3	4	7	12	16	spolu
početnosť	0	1	1	20	3	2	3	30

Aritmetický priemer $\bar{x} = 5,93 \cong 6$

Modus: ***mod(x)*** = **4**

Podľa aritmetického priemeru by bolo potrebné stavať 6-lôžkové stany. Modus poukazuje, že najvhodnejšie budú 4-lôžkové. Jednoduchá analýza nám povie viac: 1-člennú rodinu nemáme, nepotrebujeme pre ňu žiaden stan. Pre jednu 2-člennú rodinu by sme potrebovali jeden 6-lôžkový stan, zostali by 4 lôžka prázdne, resp. jeden 4-lôžkový stan s 2 nevyužitými lôžkami atď. Výsledky analýzy si môžeme dať do tabuľky:

Počet členov v rodine	Početnosť rodín	6-lôžkové .stany		4-lôžkové stany	
		počet	voľné lôžka	počet	voľné lôžka
1	0	0	0	0	0
2	1	1	4	1	2
3	1	1	3	1	1
4	20	20	40	20	0
7	3	6	15	6	3
12	2	4	0	6	0
16	3	9	6	12	0
Spolu	30	41	68	46	6

Potrebujeme buď 41 6-lôžkových stanov, pričom bude 68 lôžok nevyužitých, alebo 46 4-lôžkových stanov, pričom bude len 6 lôžok nevyužitých. Je to asi omnoho lepšia využiteľnosť lôžok.

Objasnili sme si charakteristiky polohy štatistických súborov aj s niektorými ich plusmi a mínusmi pri použití v štatistickej analýze. K ďalšiemu štúdiu možno nazrieť aj do dostupnej literatúry a zdrojov informácii, napr. [5] - [10].

Urobili sme spolu prvé krôčiky na ceste k štatistickej analýze, je potrebné si uvedomiť, že nie všetko sa dá. Napr., miešať hrušky s jablkami. Alebo je možné sa niekedy stretnúť s udivujúcimi štatistikami, ktoré sa vymykajú doterajšiemu nášmu zadeleniu medzi poriadne, neporiadne a mystifikujúce. Ako príklad uvediem informáciu o návštevnosti milovníkov športu na futbalovom zápase. Platiacich divákov vraj bolo 137. Po nekontrolovateľnej bitke medzi prívržencami oboch táborov bolo 313 zranených divákov, z toho 12 ťažko. Navyše boli 65 fanúšikovia zatknutí. Tu už je ťažšie robiť štatistickú analýzu. Poďme však ďalej.

Literatúra k IV. kapitole:

[1] Huff, D.: Jak lhát se statistikou, Brána, Praha, 2013

[2] GIBILISCO, s.: Statistika bez předchozích znalostí, Computer Press, Brno, 2009

[3] Sväté písmo s komentármi Jeruzalemskej Biblie, Pentateuch, Dobrá kniha, Trnava 2013

[4] Junkg, R.: Jasnější než tisíc sluncí. Osudy atomových vědců, Mladá fronta, Praha, 1963

[5] Chajdiak,J., Komorník, J., Komorníková, M.: Štatistické metódy, STATIS, Bratislava 1999

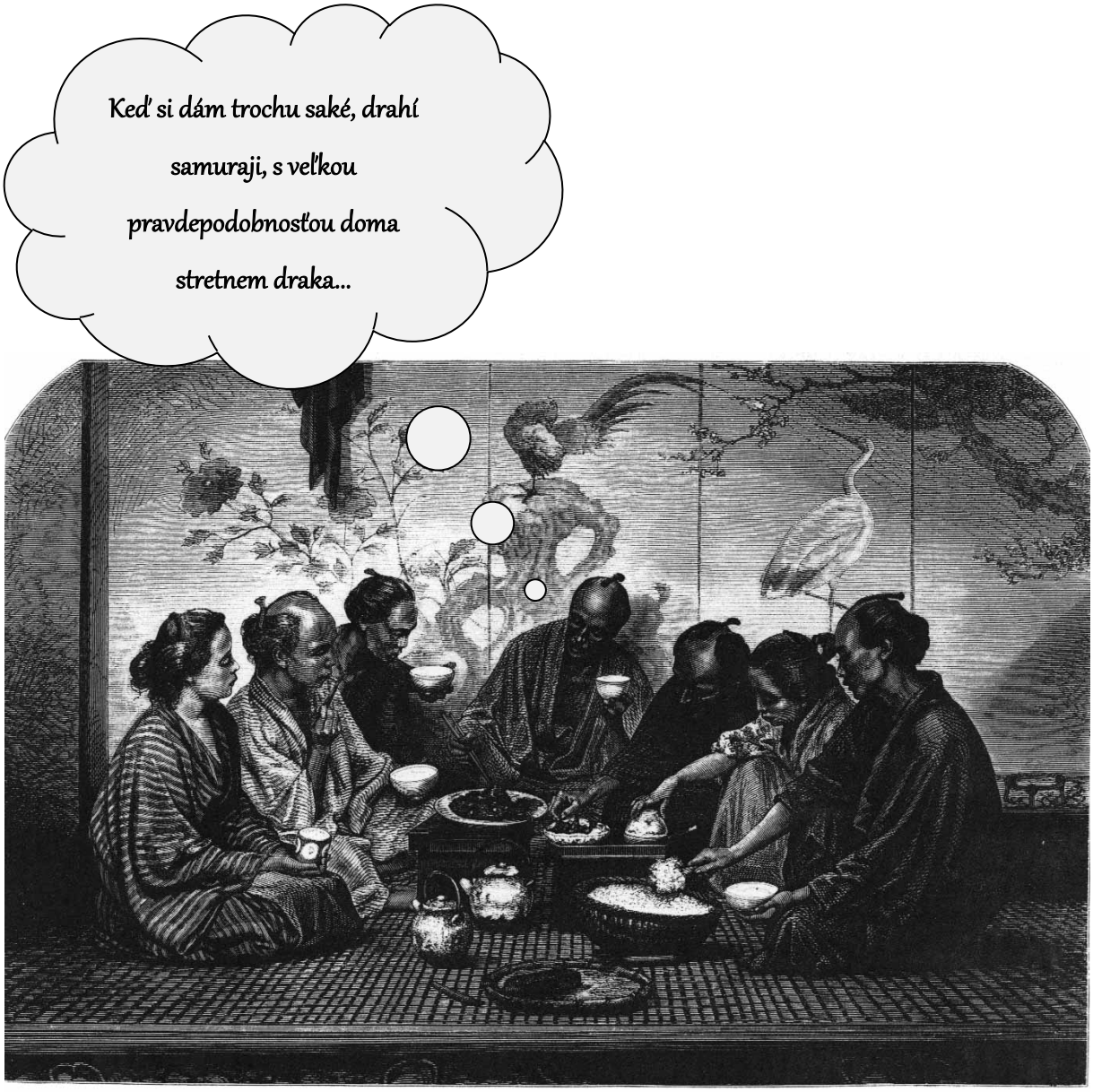
[6] http://sk.wikipedia.org/wiki/Medi%C3%A1n

[7] http://rimarcik.com/navigator/och.html

[8] Bílková, D., Budinský, P., Vohánka, V.: Pravdepodobnost a statistika, Aleš Čeněk, Plzeň, 2009

[9] http://www.priklady.eu/sk/Riesene-priklady-matematika/Pravdepodobnost-a-statistika/Statistika.alej

[10] Magnello, E., Van Loon, B.: Seznámte se...statistika, Portál, Praha 2010


Keď si dám trochu saké, drahí samuraji, s veľkou pravdepodobnosťou doma stretnem draka...

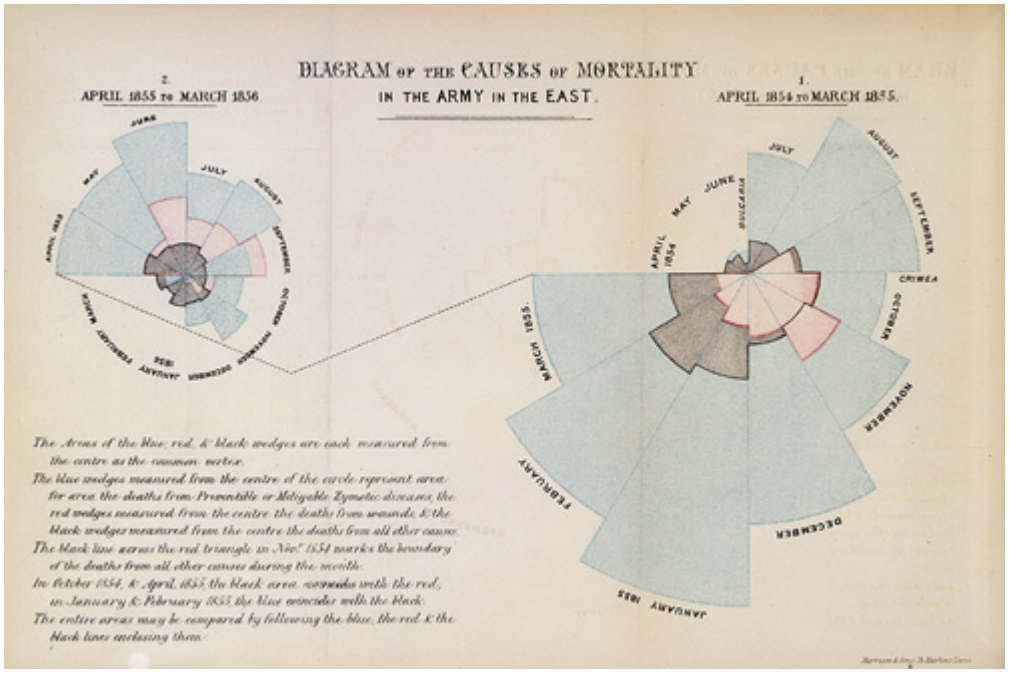
V. Variabilita alebo keď sa čísla rozbehnú

To dievča, čo som prvé pobozkal,
si pamätám dosť presne!
Zavrela oči – ja som zavrel tiež.
Minuli sme sa tesne.

Hilaire Belloc [1]

Mnohí dnešní študenti majú pocit, že štatistika vznikla ako sofistikovaná nadstavba dnešnej zložitej spoločnosti, ktorá si rada robí na všetko ornamenty a kudrlinky, bez naozaj skutočnej potreby, len aby to vyzeralo múdro, učene a v konečnom dôsledku komplikovalo štúdium. Možno majú aj trochu pravdu, ale v štatistike to bolo predsa aj trochu inak. Keď sa pozrieme na jej počiatky a vývoj, zdajú sa nám problémy nie až tak vzdialeného 19. storočia veľmi „súčasné“. Malthusové „ekonomické“ teórie o nelineárnom raste populácie popri lineárnom raste zdrojov a následne prognóza nutného katastrofického vývoja sa odvtedy opakuje s tvrdošijnou pravidelnosťou aj na najvyšších svetových fórach, tváriacich sa, že chcú niečo zlepšiť a scenár je rovnaký: Vždy sa odporúča, ako by sa mali správať chudobní, hlavne skromne, zdržanlivo, viac pracovať, mať menej detí atď. Výsledok? Bohatí by mali väčšiu stabilitu vlastnej existencie, ale nechceme tu robiť revolúcie, len uviesť čitateľa do situácie. V 19. storočí vlády krajín ani nemali prehľad o stave chudoby, ani koľko ľudí dostáva a nedostáva rôzne podpory, ba možno nemali ani predstavu o množstve financií v obehu. (Ste presvedčení, že dnešné vlády sú na tom lepšie? Alebo sa len tak tvária a používajú k tomu často „optimistickú“ neandertálsku štatistiku.) To všetko sú príznaky elementárnej nestability štátnych záležitostí. Belgický štatistik a vedec Adolphe Quetelet (1796-1874), ktorého odborníci na zdravú výživu poznajú ako zakladateľa BMI indexu, a Brit Edwin Chadwick (1800-1890) boli jednými z prvých priekopníkov zavedenia štatistiky do rozhodovacích procesov štátu a do humanitných vied. Ale pristavme sa pri známej postave **Dámy s lampou**, anglickej viktoriánskej ľudovej štatističky Florence Nightingaleovej (1820-1910), ktorá kráčala v stopách sv. Alžbety Uhorskej.

Narodila sa v šľachtickej rodine, kde získala veľmi slušné klasické vzdelanie, ale aj v matematike a iných vedách. Jej neobyčajný záujem ju priviedol do zdravotníctva. Všade kam cestovala sa zaujímala o nemocnice, ich usporiadanie, systém riadenia, rôzne činnosti, úmrtnosť a pod. Navrhovala rôzne opatrenia, pričom sa sústredila hlavne na ošetrovateľstvo, ale v dobe, ktorá nemala pre ženu a jej spoločenské aktivity veľké sympatie, bez väčších úspechov. Ani vo vlastnej rodine nenachádzala pre svoje aktivity pochopenie, ošetrovateľky mali veľmi nízky spoločenský status (stále sme v 19. storočí !). Spolupracovala s Williamom Farrom(1807-1883), priekopníkom medicínskej štatistiky. Keď v polovici 19. storočia vypukla Krymská vojna (jej história je aj dnes veľmi aktuálna, ale nesúvisí s náplňou tejto publikácie), bola požiadaná Ministerstvom obrany o zorganizovanie ošetrovateľskej činnosti v teréne. Keď prišla aj s niekoľkými sestričkami a dievčatami do lazaretov v Turecku, našla tam jemne povedané chaos. Lazarety v otrasnom stave, bez aj tej najzákladnejšej hygieny, pravidelnej stravy, liekov, ošetrovania, ale aj bez záznamov o úmrtiach a ich príčinách, bez vzájomnej spolupráce medzi zdravotníckymi zariadeniami, či bez prikrývok a postelí. Lekári a dôstojníci ich neprijali, ale postupne si vydobyla postavenie, keď prijali mnohé opatrenia v prospech ranených vojakov a pacientov. Vojaci si ju pamätali ako v noci s lampou chodí po lazarete kontrolovať stav pacientov. V správe pre ministerstvo a domáce úrady využila všetky svoje vedomosti zo štatistiky a jej grafického spracovania. Bez preháňania, bola to štatistika, kde šlo o život.



Obr.V.1. Ukážka grafického spracovania Nightingaleovej štatistickej analýzy z vojenských nemocníc počas Krymskej vojny [2].

Vytvorila polárny plošný diagram, rozdelený na 12 rovnakých výsekov podľa mesiacov a na nich znázorňovala plošne aj farebne časové zmeny úmrtí: ružová – úmrtia v dôsledku zranení v boji, šedá – úmrtia z iných príčin, najväčšia, modrá – úmrtia v dôsledku ochorení. Vo vojenských lazaretoch bola úmrtnosť dvojnásobná ako v civilných. Na týfus a choleru zomieralo viac mladých mužov ako za veľkej epidémie moru v Londýne v r.1665-1666. Jej nový systém a neobyčajné nasadenie znížili 60% úmrtnosť do konca Krymskej vojny na 1 až 2%. A jej štatistika bola základom reformy zdravotníctva aj na britských ostrovoch a následne v systéme ošetrovateľstva a jeho profesionalizácii v celom civilizovanom svete. Do akej miery aj u nás doma, musí posúdiť na základe vlastných skúseností čitateľ.

Uviedli sme si, že priemer resp. priemery môžu zredukovať a vhodne charakterizovať štatistický súbor, môže zrovnať, „vyhladiť“ náhodné odchýlky pri získavaní dát, ale aj to, že v priemere sa môže kadečo skryť, že prostredníctvom priemeru sa informácia zhrnula do jedného predstaviteľa všetkých možných prípadov, do jedinej hodnoty. **Variabilita**, ktorou sa ideme teraz zaoberať, vyjadruje informácie o individuálnych odlišnostiach prvkov súboru, ale predstavuje aj vlastnosti chýb zberu dát a celkovo umožňuje spolu s charakteristikami polohy dať ucelenú informáciu o vlastnostiach súboru. Až v takejto forme, **priemer + variabilita**, sa dostávame z často neandertálskej štatistiky k niečomu zmysluplnému (ešte nižšie k nim pribudnú **charakteristiky tvaru**). Všetky tieto charakteristiky popisujú štatistický súbor, voláme ich **opisné charakteristiky**. Popisná štatistika sa snaží výsledky štatistického skúmania vyjadriť prehľadne, hutne, pomocou tabuliek, grafov a popisných charakteristík. Variabilitu popisujeme **charakteristikami variability.** Sú to čísla, udávajúce, akou mierou sa hodnoty znaku odchyľujú od vypočítanej charakteristiky polohy alebo od seba navzájom [3]**:**

a) Variačné rozpätie R

Je to len orientačná charakteristika variability hodnôt znaku. Predstavuje informáciu o rozsahu hodnôt x_1, x_2, ..., x_n sledovaného znaku x štatistického súboru. Vypočíta sa ako rozdiel maximálnej a minimálnej hodnoty:

$$\mathbf{R = x_{max} - x_{min}} \qquad \mathbf{[V.1]}$$

Je to zrozumiteľná charakteristika, ale nedáva dostatočný obraz rozptýlenia údajov v súbore. Závisí od veľkosti výberu, preto nie je možné porovnávať variačné rozpätie rôzne veľkých súborov. Jej závislosť iba od krajných hodnôt súboru spôsobuje mimoriadnu citlivosť na málo charakteristické extrémne hodnoty, ktoré sa v týchto krajných hodnotách môžu zobrazovať. Taktiež aj dva úplne odlišné súbory, ak majú náhodou zhodné extrémne hodnoty môžu mať rovnaké variačné rozpätie.

Variačné rozpätie je veľmi jednoduchá charakteristika. Poďme si to teda trochu skomplikovať, aby sme si mohli život zjednodušiť. Nevýhody citlivosti variačného rozpätia R od extrémnych hodnôt je možné odstrániť iným rozpätím štatistického súboru. V minulej kapitole sme si charakterizovali strednú hodnotu medián **Med(x)** ako **kvantil** rozdeľujúci súbor na dve rovnaké časti. Je to teda jediná hodnota súboru, pre ktorú platí, že 50% všetkých hodnôt je menších a 50% väčších ako medián. Je to hranica polovičného resp. 50%-ného intervalu. Máme aj iné kvantily, napr. percentily (delenie súboru na 100 rovnakých častí, musíme mať teda 99 percentilov), máme 9 decilov, deliacich súbor na 10 častí. Keď si usporiadaný súbor štatistických prvkov rozdelím na 4 rovnaké časti, 3 hranice týchto intervalov voláme kvartily: Q_1 alebo $Q_{25\%}$; Q_2 alebo $Q_{50\%}$ (= medián); Q_3 alebo $Q_{75\%}$. Nepotrebujeme Q_4, pretože to je už maximálna hodnota celého súboru. Po tomto trochu zdĺhavom predslove môžeme uviesť, čo je to medzikvartilové rozpätie R_{Q3-Q1}, ktoré niekedy používame miesto variačného rozpätia R. Medzikvartilové rozpätie je rozdiel hodnôt súboru v ¾ a v ¼ pevných intervaloch rozsahu jeho usporiadaných hodnôt, teda medzi 75. a 25. percentilom, čo reprezentuje oblasť 50% stredných hodnôt súboru. Načo je to dobré? R_{Q3-Q1} nie je ovplyvnený extrémnymi hodnotami, takže je často lepšou charakteristikou variability.

$$\mathbf{R_{Q3-Q1} = Q_3 - Q_1} \qquad \mathbf{[V.2]}$$

Pr.V.1.: V štatistikách ÚPSVaR sa uvádza počet podporených zamestnávateľov príspevkom úradmi PSVR v SR na zriadenie chránenej dielne alebo chráneného pracoviska v r.2013 **[4]**:

Mesto	Počty	Mesto	Počty	Mesto	Počty
Bratislava	201	Dolný Kubín	83	Stropkov	262
Malacky	63	Námestovo	226	Vranov n/Topľou	440
Pezinok	99	Liptovský Mikuláš	138	Košice	522
Dunajská Streda	136	Martin	272	Michalovce	131
Galanta	65	Ružomberok	112	Rožňava	53
Piešťany	154	Žilina	334	Spišská Nová Ves	279
Senica	153	Banská Bystrica	263	Trebišov	397
Trnava	158	Banská Štiavnica	139	Kežmarok	107
Partizánske	210	Brezno	79	**Spolu**	9 936
Nové M.n/Váhom	83	Lučenec	27		
Považská Bystrica	166	Revúca	23		
Prievidza	752	Rimavská Sobota	174		
Trenčín	343	Veľký Krtíš	310		
Komárno	142	Zvolen	243		
Levice	307	Bardejov	283		
Nitra	175	Humenné	64		
Nové Zámky	170	Poprad	207		
Topoľčany	182	Prešov	526		
Čadca	253	Stará Ľubovňa	430		

Usporiadaný súbor hodnôt tejto reálnej štatistiky z r.2013 je potom v tabuľke vľavo.

1	23
2	27
3	53
4	63
5	64
6	65
7	79
8	83
9	83
10	99
11	107
12	112
13	131
14	136
15	138
16	139
17	142
18	153
19	154
20	158
21	166
22	170
23	174
24	175
25	182
26	201
27	207
28	210
29	226
30	243
31	253
32	262
33	263
34	272
35	279
36	283
37	307
38	310
39	334
40	343
41	397
42	430
43	440
44	522
45	526
46	752

Výpočet kvantilov vo všeobecnosti môžeme urobiť zjednodušene nasledovne: Po usporiadaní súboru hodnotu kvantilu získame ako **k**-tú hodnotu súboru, pričom poradie **k** dáva vzťah:

k = (počet hodnôt)x(úroveň kvantilu)/100

Pre 1.kvartil to bude

k1 = 46x25/100 = 11,5 ≅ 12

teda rozsah súboru sme delili 4 a zaokrúhlili nahor; pre 3.kvartil urobíme to isté, len ešte vynásobíme 3:

k3 = 46x75/100 = 34,5 ≅ 35

Potom: $Q_1 = 112$, $Q_2 = 279$.

Pre uvedený súbor dostaneme charakteristiky polohy a vybrané charakteristiky variácie:

Súčet: $\Sigma x_i = 9936$

Aritmetický priemer: $\bar{x} = 216$

Medián: **Med(x) = 174,5**

Variačné rozpätie: R = 729

Medzikvartilové rozpätie: $\mathbf{R_{Q3-Q1} = 279 - 112 = 167}$

Záver: Keby sme chceli zistiť aký je priemerný počet podporovaných subjektov v uvedených mestách SR, pomocou aritmetického priemeru by sme dostali hodnotu skreslenú niektorými extrémnymi číslami (216), strednú hodnotu vystihuje lepšie medián (174,5). Variačné rozpätie súboru 729 je ovplyvnené extrémnymi krajnými hodnotami, lepšie ju vystihuje medzikvartilové rozpätie 167.

Pr.V.2.: Podľa Slovenského platového monitora **[5]** majú napr. terénni sociálni pracovníci priemerný mesačný plat 520.-€. Bohužiaľ viac k tomuto údaju nie je pripojené, ani štatistické charakteristiky, ani či ide o hrubú mzdu a pod. Takže s ním môžeme voľne pracovať za účelom prehĺbiť si svoje vedomosti. Máme dva súbory po 10 terénnych sociálnych pracovníkov z rôznych oblastí a v nasledujúcej tabuľke sú uvedené ich priemerné čisté mesačné platy (2. stĺpec usporiadane) aj s popisnými štatistickými charakteristikami, ktoré sme sa doteraz naučili. Pokúsme sa spolu urobiť analýzu výsledkov.

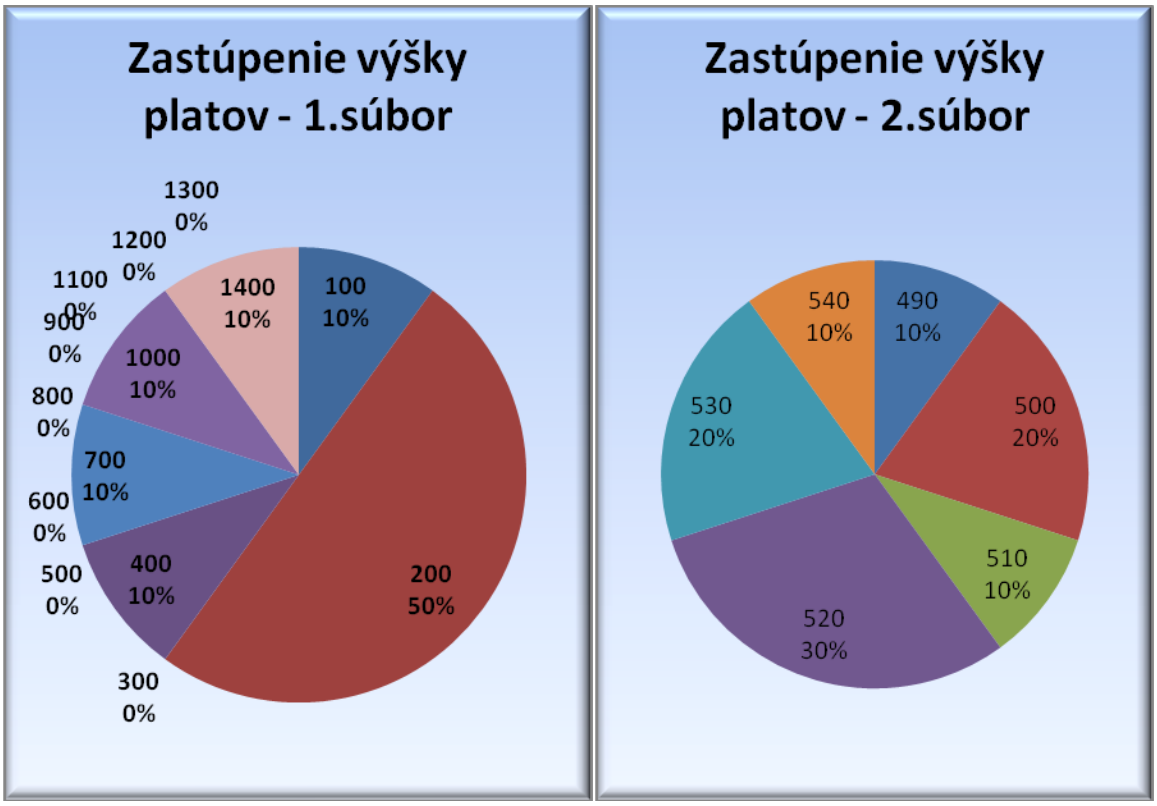
č.	Súbor č.1		Súbor č.2	
1	280	150	495	495
2	1488	230	525	500
3	795	256	519	504
4	230	278	528	519
5	150	280	523	523
6	296	296	540	525
7	1007	420	500	528
8	256	795	536	530
9	420	1007	530	536
10	278	1488	504	540
spolu		5200		5200
Priemer $\bar{x}$		**520**		**520**
medián Med(x)		**288**		**524**
Variačné rozpätie R		**1338**		**45**
Q1		**256**		**504**
Q3		**795**		**540**
R(Q3-Q1)		**539**		**36**

Tab.V.1. Priemerné platy súborov sociálnych pracovníkov v dvoch regiónoch.

Je to zaujímavá tabuľka. V oboch súboroch je priemerný mesačný plat 520.-€ na pracovníka. V 1.súbore sa v tejto hodnote skrýva veľká variabilita a trochu lepšie ho charakterizuje medián Med(x) = 288, ktorý naznačuje, že v súbore musia byť zriedkavé extrémne vysoké hodnoty. Naozaj 7 hodnôt z 10 sa nachádza pod priemerom. Väčšina hodnôt sa pohybuje skôr okolo mediánu, ktorý je hlboko pod priemerom. Nasvedčuje tomu aj vysoké variačné rozpätie R = 1338. Medzikvartilové rozpätie je nižšie R(Q3-Q1) = 539 je však stále ešte dosť veľké. Zastúpenie platov v 100 € intervaloch s uvedenou dolnou hranicou je na obr. V.2.

V 2. súbore sa aritmetický priemer zdá byť naozaj charakteristickou hodnotou. Medián Med(x) = 524 sa od neho líši len bezvýznamne. Variačné aj medzikvartilové rozpätie naznačujú nízku variabilitu súboru. Zastúpenie platov v 5 € intervaloch s uvedenou dolnou hranicou je na obr. V.3.

Môžeme konštatovať, že 2. súbor predstavuje vyrovnanosť platov sociálnych pracovníkov v uvedenom regióne, pričom odchýlky od strednej hodnoty sú natoľko symetrické, že ich je možné považovať za normálne (s normálnym rozdelením hodnôt).



Obr. V.2 a V.3: Zastúpenie výšky platov v [%] dvoch vybraných súborov sociálnych pracovníkov v rôznych regiónoch.

V predchádzajúcom príklade sme videli, že v štatistickom súbore, ktorého znakom je náhodná premenná, hodnoty „kolíšu“ okolo priemeru, (resp. inej strednej hodnoty). Lepšou charakteristikou variability ako sú rozpätia, je priemerná odchýlka Δ_p, pretože jej hodnota závisí od každej hodnoty súboru. Dostaneme ju ako aritmetický priemer absolútnych hodnôt všetkých odchýlok od aritmetického priemeru:

$$\Delta p = \frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - \bar{x}| \qquad [V.3]$$

Ak máme danú tabuľku rozdelenia početnosti (f_i je početnosť znaku v i-tom intervale):

$$\Delta p = \frac{1}{n} \cdot \sum_{i=1}^{k} |x_i - \bar{x}| \cdot f_i \quad pričom \quad \sum_{i=1}^{k} f_i = n \qquad [V.4]$$

Ak si v tab.V.1 ku každej hodnote vypočítate absolútnu odchýlku, spočítate ich a vydelíte 10, dostanete pre 1.súbor hodnotu $\Delta_{p\,1} = 346$; pre 2.súbor podstatne menšiu hodnotu $\Delta_{p\,2} = 12{,}4$, čo nasvedčuje, že dáta 2.súboru majú podstatne menšie (užšie) rozptýlenie okolo priemernej hodnoty.

Súčet odchýlok hodnôt znaku od ich aritmetického priemeru je = 0, tak je aritmetický priemer definovaný. Preto sme vo vzťahu pracovali s absolútnou hodnotou (matematicky to značíme miesto zátvoriek dvomi zvislými čiarami), pre ktorú platí, že z akéhokoľvek čísla, či kladného alebo záporného, vráti vždy jeho kladnú hodnotu.

Teraz si ukážeme ešte lepšiu cestu, ako odstrániť záporné hodnoty z odchýlok. Umocníme každú odchýlku na druhú a vieme, že druhá mocnina každého čísla je kladná, napr. $3x3 = 3^2 = 9$, ale aj $(-3)x(-3) = (-3)^2 = 9$. Ako bonus dostaneme, že extrémne hodnoty v súbore sa zvýraznia, bude väčší rozdiel medzi malými a veľkými odchýlkami. Možno to bolo trochu komplikované, ale takto sme sa dostali k veľmi dôležitým charakteristikám variability a pojmom **rozptyl** a **smerodajná odchýlka**. V predchádzajúcej kapitole sme si spomenuli teoretické rozdelenia pravdepodobnosti. A aj to, že každé rozdelenie má dve základné charakteristiky: **strednú hodnotu E(X)** a **rozptyl D(X)**. Stredné hodnoty sme si popísali ako charakteristiky polohy (priemery a pod.). Teraz si niečo povieme o rozptyle v súvislosti so štatistickým súborom. Keď sme sa dostali až sem, nie je to už žiaden problém. Tak ako sme si charakterizovali priemernú odchýlku ako aritmetický priemer absolútnych hodnôt jednotlivých odchýlok, tak si **rozptyl D(X)**, ktorý je najčastejšou charakteristikou variability, definujeme ako aritmetický priemer druhých mocnín hodnôt jednotlivých odchýlok:

$$D(X) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \textbf{[V.5]}$$

Pri rozdelení súboru do **i** intervalov, kde f_i je početnosť výskytu znaku v **i**-tom intervale, je výpočet:

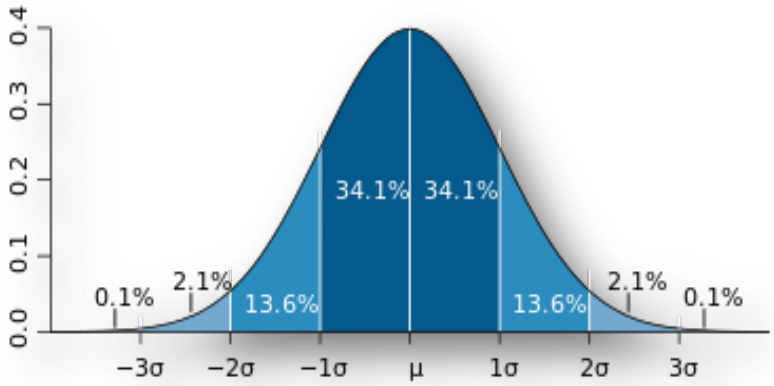
$$D(X) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 . f_i \quad pričom \sum_{i=1}^{k} f_i = n \qquad \textbf{[V.6]}$$

Smerodajná (niekedy aj **štandardná** alebo **stredná kvadratická) odchýlka σ** je potom jednoducho podľa **[III.13.]** kladná odmocnina z rozptylu. Na základe vzťahu [V.5] (obdobne to bude aj pre [V.6]) dostaneme

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \textbf{[V.7]}$$

Častejší zvyk charakterizovať variabilitu štatistického súboru smerodajnou odchýlkou **σ** vyplýva z toho, že **σ** má rovnaký rozmer ako priemer, zatiaľ čo rozptyl je udávaný v druhej mocnine rozmeru priemeru, napr. ak by sme zisťovali výšku príjmu v nejakom výbere

populácie, strednú hodnotu by sme dostali v €, ale rozptyl v €^{2}. Smerodajná odchýlka **σ** bude opäť v €. Pripomeňme si na tomto mieste ešte raz obr. III.19 z III. kapitoly, predstavujúci krivku hustoty normálneho rozdelenia s vysvetlením parametrov **μ** a **σ**:



μ = $\bar{x}$ = Med(x) je stredná hodnota normálneho rozdelenia. **σ** je smerodajná odchýlka v zmysle vyššie uvedených vzťahov pre základný štatistický súbor, teda pre všetky jeho prvky. Potom výsledok štatistickej analýzy vyjadrený intervalom

$$\bar{x} \pm \sigma \qquad \textbf{[V.8]}$$

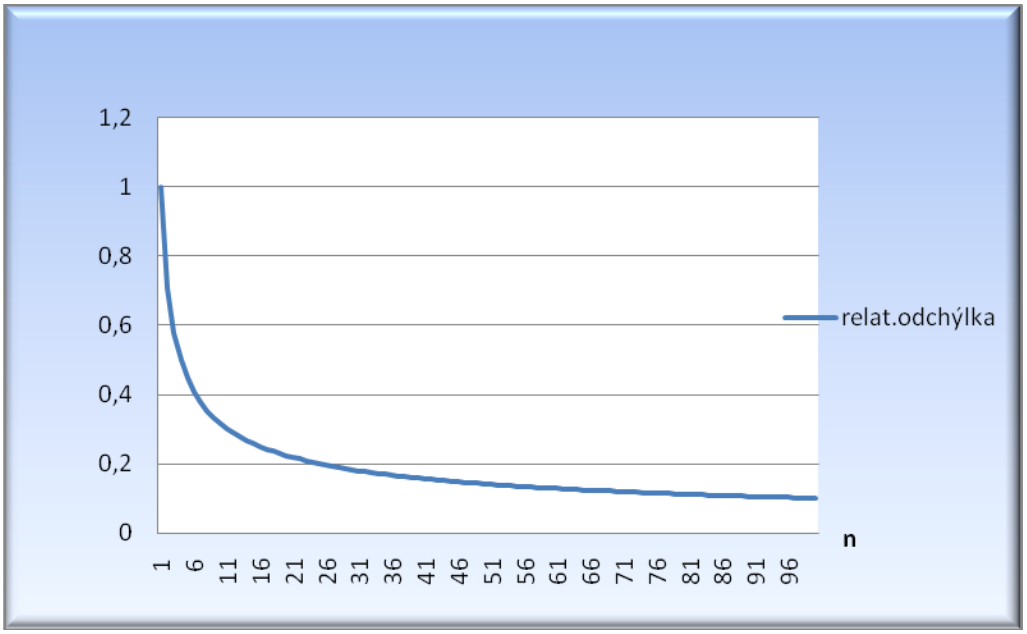
predstavuje interval okolo aritmetického priemeru so 68% všetkých hodnôt súboru, **± 2σ** s 95% všetkých hodnôt, **± 3σ** s 99% všetkých hodnôt. Smerodajná odchýlka je mierou rozptýlenia hodnôt okolo priemeru, teda často napr. pri meraní aj istou mierou presnosti. Pri výberovom súbore máme **odhady** parametrov, napr. odhad strednej hodnoty, odhad aritmetického priemeru, **odhad smerodajnej odchýlky (jedného merania) s**, pre ktorý platí trochu odlišný vzťah ako [V.7] :

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \textbf{[V.9]}$$

Podobne to platí aj pre výberový rozptyl. Výberové charakteristiky majú mať čo najmenší rozptyl (majú byť „čo najvýdatnejšie“) aby boli pre súbor charakteristické. Pre súbory veľkého rozsahu (pre veľké **n**) sa **s** blíži k **σ: s → σ.** Dôležitejšie je poznať **smerodajnú odchýlku výberového aritmetického priemeru** získaných dát $s_{\bar{x}}$: (Pozri aj [3], [6], [7]).

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \textbf{[V.10]}$$

Smerodajná odchýlka výberového aritmetického priemeru $s_{\bar{x}}$ klesá s rastúcim počtom prvkov súboru (rozsahom) **n**. Predstavu o tom dáva obr. V.4:



Obr.V.4.: Závislosť relatívnej veľkosti smerodajnej odchýlky aritmetického priemeru od rozsahu súboru n: $\frac{s_{\bar{x}}}{s} = f(n)$

Výrazný pokles smerodajnej odchýlky aritmetického priemeru, (ktorú budeme nazývať aj **chyba strednej hodnoty** $s_{\bar{x}}$), rastom rozsahu súboru **n** je významný pre n < 10 až 20, zvyšovanie rozsahu výberu z dôvodu jej ďalšieho zmenšovania už nie je efektívne.

Pomocou odchýlok máme základnú predstavu o rozložení dát v súbore, nie však o konkrétnom prvku súboru. Smerodajná odchýlka **σ** alebo jej odhad **s** môžu slúžiť ako prvá orientácia v tom, s akou spoľahlivosťou je možné konkrétny prvok v nejakom intervale nájsť. Tak môžeme interval $\bar{x}\pm\sigma$ nazvať v prvom priblížení **interval 68% spoľahlivosti**, t.j. máme odhad, že nejaký znak má 68% hodnôt, ktoré sa v tomto intervale nachádzajú. Podobne môžeme interval $\bar{x}\pm2\sigma$ nazvať **interval 95% spoľahlivosti** a interval $\bar{x}\pm3\sigma$ nazvať **interval 99% spoľahlivosti**. Poznamenávam, že tieto úvahy sa týkajú normálneho symetrického rozdelenia údajov štatistického súboru. Nie je problém určiť si aj iný interval spoľahlivosti, ale týmto sa budeme podrobnejšie zaoberať neskôr. Dajme si jednoduchý príklad zo základov teórie merania, výsledky sa dajú zovšeobecniť.

Pr.V.3.: Z Ústavu pre znevýhodnených, s ktorým úzko spolupracujete, ste si objednali ich výrobok, pracovný stôl so 160 cm vrchnou doskou. Aby ste si precvičili svoje štatistické poznatky a zároveň urobili istú kontrolu kvality výrobkov, zmerajte dĺžku svojho pracovného

stola 3 spôsobmi. 1.spôsob: požiadajte 10 svojich návštevníkov o ich odhad, výsledky si zaznamenajte do tabuľky. 2.spôsob: Použite na získanie 10 hodnôt dĺžky stola krajčírsky, trochu „pružný“ meter. 3.spôsob: použite na zmeranie najpresnejšie meradlo, aké máte k dispozícii, pravítko, remeselnícke zvinovacie meradlo alebo laserový merač vzdialenosti, opäť si výsledky 10 meraní zaznamenajte. Dostaneme tri nezávislé súbory po 10 nezávislých meraní v [cm], ako je napr. v tabuľke V.2 (získal som aj odhad od 4-ročného dievčatka, ktorý som ako veľmi odľahlý výsledok zaťažený neprimeranou chybou zo súboru sofistikovane vylúčil) a vypočítanú množinu štatistických charakteristík pre každý súbor, pričom jednotlivé odchýlky označujeme $\Delta = (x_i - \bar{x})$:

i	súbor č.1			súbor č.2			súbor č.3		
	x_i	$\|\Delta\|$	Δ^2	x_i	$\|\Delta\|$	Δ^2	x_i	$\|\Delta\|$	Δ^2
1	240	60	3600	161,5	0,5	0,25	159,8	0,2	0,04
2	165	15	225	163	2	4	160,3	0,3	0,09
3	145	35	1225	160	1	1	160,1	0,1	0,01
4	195	15	225	159	2	4	159,9	0,1	0,01
5	135	45	2025	159,5	1,5	2,25	159,9	0,1	0,01
6	165	15	225	158	3	9	160,0	0	0
7	155	25	625	159	2	4	160,2	0,2	0,04
8	225	45	2025	162	1	1	160,1	0,1	0,01
9	195	15	225	163	2	4	159,8	0,2	0,04
10	180	0	0	165	4	16	159,9	0,1	0,01
Σ	1800	270	10400	1610	19	45,5	1600	1,4	0,26
n	10			10			10		
$\bar{x}$	180			161			160		
σ	32,24903			2,133073			0,161245		
s	33,99346			2,248456			0,169967		
$s_{\bar{x}}$	10,74968			0,711024			0,053748		
Med(x)	172,5			160,75			159,95		
Mod(x)	165			163			159,9		
x_{min}	135			158			159,8		
x_{max}	240			165			160,3		
R	105			7			0,5		
D(x)	1155,556			5,055556			0,028889		
V(%)	17,91613			1,32489			0,100778		

Tab.V.2. Výsledky 3 nezávislých súborov meraní dĺžky stola s vypočítanými popisnými štatistickými charakteristikami.

Je to tiež zaujímavá tabuľka, že? A obsahuje kadečo, čo nás na prvý pohľad vystrašilo, ale keď sa na to pozrieme bližšie, dáva nám množstvo informácií, zároveň si popíšeme, čo a ako sme počítali. Vysvetlíme si to krok po kroku, majte trpezlivosť:

1. Máme tri nezávislé štatistické súbory. Sú to výberové súbory, pretože sme uskutočnili v každom z nich 10 meraní z nekonečného množstva všetkých možných. Základný súbor by bol, keby sme urobili nekonečne mnoho meraní, ba ešte viac, keby nás to bavilo a mali sme na to čas.
2. Rozsahy súborov sú rovnaké, každý má rozsah n = 10 (meraní).
3. Sledovaný štatistický znak (jav) je dĺžka pracovného stola v cm.
4. V hornej časti tabuľky sú namerané údaje. 1.sĺpec pod hlavičkou *i* je číslo merania. Každý súbor má pripravené 3 stĺpce údajov, prvý z nich pod hlavičkou $\mathbf{x_i}$ obsahuje vždy namerané hodnoty dĺžky stola v cm.
5. Aby sme mohli vyplniť ďalšie stĺpce, musíme zistiť strednú hodnotu každého súboru. Sú to jednoduché súbory nezávislých meraní, kde dobrým odhadom strednej hodnoty je aritmetický priemer podľa vzťahu **[IV.1]**.
6. Pre každé meranie si vypočítame odchýlku nameranej hodnoty od aritmetického priemeru $\Delta = (x_i - \bar{x})$. Môžu byť kladné aj záporné, podľa toho, či sa výsledok merania nachádza nad, alebo pod priemerom. Keby sme ich všetky sčítali, dostaneme vždy nulu (pozri aj **[IV.4]**. Preto si urobíme absolútnu hodnotu z každej hodnoty odchýlky Δ, a túto zapíšeme do ďalšieho stĺpca. V 3.stĺpci každého súboru je potom druhá mocnina odchýlky Δ^2. Je vždy kladná, preto nepotrebuje absolútnu hodnotu. To je všetko, čo potrebujeme pre základnú popisnú štatistiku, ktorú sme sa doteraz naučili.
7. Začíname s výpočtom štatistických charakteristík jednotlivých štatistických súborov. Sústreďme sa na spodnú časť tabuľky. V jej 1.riadku označenom symbolom Σ máme súčty hodnôt v stĺpcoch nad nimi. Pre každý súbor je to súčet všetkých nameraných hodnôt x_i, súčet absolútnych hodnôt odchýlok jednotlivých meraní od aritmetického priemeru Δ a súčet ich štvorcov (druhých mocnín) Δ^2.
8. Druhý riadok je len rozsah každého súboru $\mathbf{n = 10}$. Tretí riadok je zaujímavejší. Keď vydelíme súčet všetkých hodnôt meraní $\sum x_i$ rozsahom súboru **n**, dostaneme aritmetický priemer $\bar{x}$ podľa **[IV.1].**
9. V 4.riadku máme smerodajnú odchýlku σ vypočítanú podľa vzťahu **[V.7]**, teda vzali sme spod stĺpca pre druhé mocniny odchýlok od aritmetického priemeru ich súčet $\sum \Delta^2$, vydelili sme to rozsahom súboru **n** a z výsledku sme pomocou kalkulačky urobili odmocninu.

10. Keďže ide o výberový súbor vhodnejšia je iná odchýlka. V ďalšom riadku máme odhad smerodajnej odchýlky jedného merania **s** v zmysle vzťahu **[V.9]**, pričom sme postupovali ako v bode 9., ale nedelili sme súčet $\sum \Delta^2$ hodnotou n, ale hodnotou (n-1) t.j. v našom prípade 9.

11. Obdobne v ďalšom riadku máme smerodajnú odchýlku aritmetického priemeru $s_{\bar{x}}$ vypočítanú podľa vzťahu **[V.10]**.

12. V ďalších dvoch riadkoch sú prezentované pre každý súbor stredné hodnoty. Medián vypočítaný ako 50% kvantil, teda hodnota, ktorá rozdeľuje štatistický súbor presne na polovicu; a modus ako najčastejšia hodnota súboru, tak ako sme si to popísali v predchádzajúcej kapitole.

13. V ďalších troch riadkoch máme už nám známe veci zo vzťahu **[V.1]**: minimálnu hodnotu $\mathbf{x_{min}}$, maximálnu hodnotu $\mathbf{x}_{max}$ a variačné rozpätie každého súboru $\mathbf{R = x_{max} - x_{min}}$.

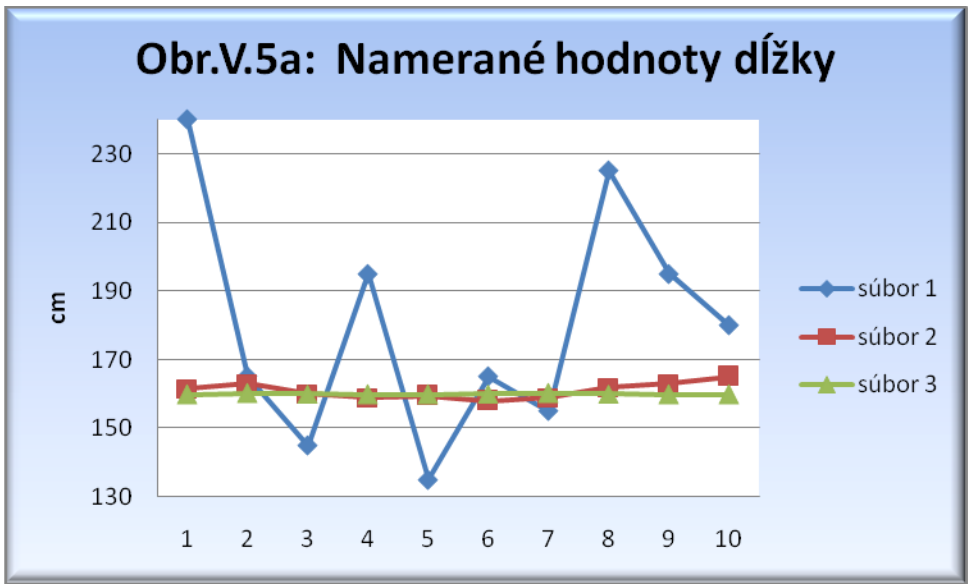
14. V predposlednom riadku je rozptyl (disperzia) **D(X)**, ktorý sa dá vypočítať podľa **[V.5]**, teda ako súčet druhých mocnín odchýlok $\sum \Delta^2$, vydelený rozsahom **n**.

15. Jediná novinka v celej tabuľke je posledný riadok, v ktorom je uvedený **Pearsonov variačný koeficient V**, vyjadrený väčšinou v %:

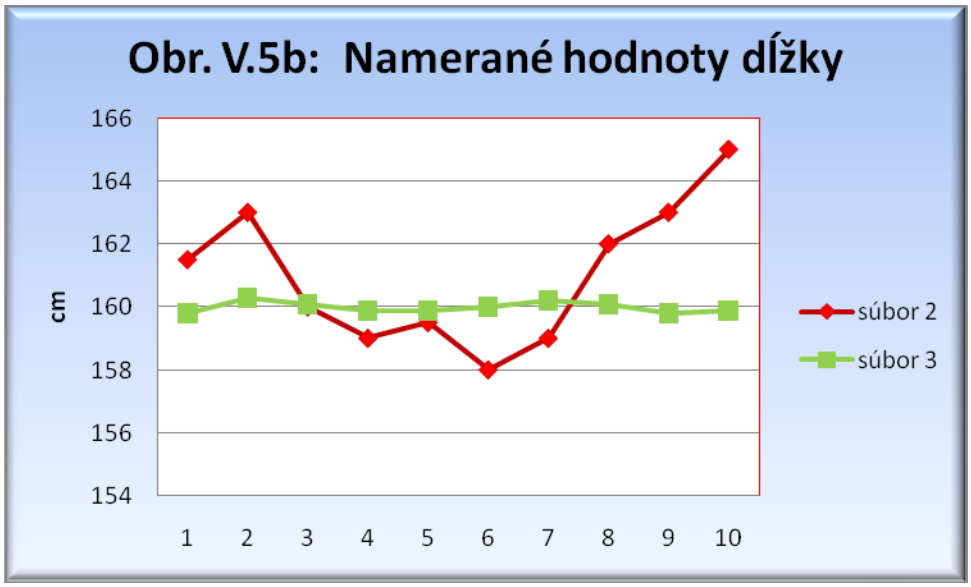
$$V = \frac{\sigma}{\bar{x}}.100 \qquad [V.11]$$

Pearsonov variačný koeficient **V** je obecnejším ukazovateľom pre porovnávanie variability súborov ako aritmetický priemer so smerodajnou odchýlkou.

Na **obr.V.5.a** je grafické znázornenie, ako v jednotlivých súboroch oscilujú hodnoty dĺžky okolo priemeru.



Na prvý pohľad je jasné, že hodnoty 1.súboru majú najväčšiu variabilitu. Veľký rozptyl výsledkov meraní odhadom rôznymi osobami spôsobuje v tomto prípade, že na ich grafické vyjadrenie je potrebný taký rozsah zvislej osi grafu, ktorý potlačí vizuálne porovnanie variability ďalších dvoch súborov údajov, získaných meraním. Na to potrebujeme samostatný obrázok **V.5.b** (všimnite si zmenu mierky zvislej osi):



Vidno, že najmenej sa „pohybujú“ namerané hodnoty v 3.súbore dát, ktoré sme získali najpresnejším meradlom. Kvantitatívnou mierou tejto presnosti je veľkosť odhadu smerodajnej odchýlky jedného merania ***s*** resp. smerodajnej odchýlky aritmetického priemeru $s_{\bar{x}}$, ale hlavne Pearsonov variačný koeficient **V**. Čím sú menšie, tým lepšie. Smerodajná odchýlka je absolútnou mierou variability, to znamená, že keby sme merali rovnakou metódou svoj pracovný stôl a omnoho dlhší pult, dostali by sme v druhom prípade väčšiu smerodajnú odchýlku, čo však neznamená, že sme sa v meraní a jeho presnosti „zhoršili“. Karl Pearson (1857-1936), britský matematik, štatistik a zakladateľ biometriky, ako aj prvej katedry štatistiky, ktorý zaviedol aj používanie pre Gauss-Laplaceové rozdelenie „normálna“ krivka, pre účel porovnávania variability aj medzi súbormi s rôznou veľkosťou a rôznou strednou hodnotou zaviedol relatívnu mieru variability – **variačný koeficient V(%)**. Charakteristiky $\bar{x}$ **, s,** $s_{\bar{x}}$**, V(%)** popisujú výberový štatistický súbor a jeho variabilitu a umožňujú kvantitatívne porovnávanie s inými súbormi. V našom príklade máme súbory z veľkej množiny súborov, kedy sa snažíme o čo najmenšiu variabilitu, teda veľká variabilita a rozbehnutie sa jednotlivých výsledkov od priemeru nie je vítaná. Ideálny stav je stredná hodnota – aritmetický priemer. Samozrejme, že máme aj systémy, súbory, kde sledujeme variabilitu ako pozitívnu

charakteristiku. Na druhej strane ani nízka úroveň variability, t.j. v našom prípade presnosť v niektorých prípadoch nemusí znamenať správnosť výsledkov. To je problematika tzv. systematických chýb meraní. Systematická chyba, pokiaľ nie je včas odhalená a eliminovaná môže spôsobiť, že meriame hodnoty pre nejaký sledovaný znak síce veľmi presne, ale so strednou hodnotou vzdialenou od skutočnej hodnoty. Vo všeobecnosti môže byť systematická chyba skrytá aj v tom, ako uskutočňujeme výber zo základného súboru, teda jeho úmyselným alebo neúmyselným skreslením.

Pearson urobil poriadok v popise empirických dát a ich rozdelenia systémom založeným na **metóde momentov.** Termín „moment“ prevzal z v tej dobe veľmi populárnej fyziky. Definoval:

1.moment – aritmetický priemer,

2.moment – priemer druhej mocniny odchýlky od priemeru (rozptyl),

3.moment – priemer tretej mocniny odchýlky od priemeru (šikmosť),

4.moment – priemer štvrtej mocniny odchýlky od priemeru (špicatosť).

Tieto štyri veľmi úsporné parametre sú nevyhnutné pre interpretáciu akéhokoľvek súboru štatistických dát s ľubovoľným nielen normálnym rozdelením. [8]

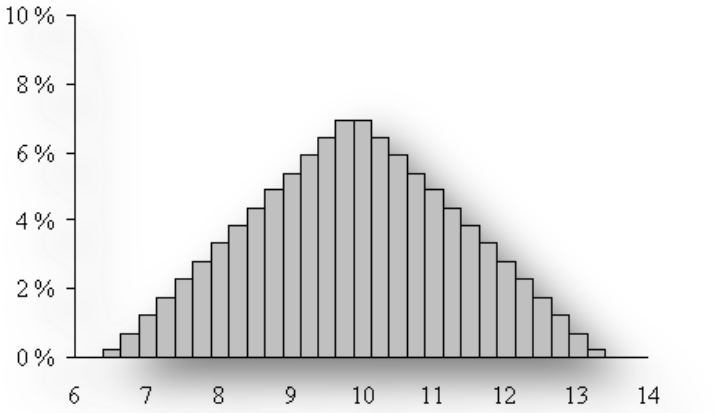
Šikmosť a **špicatosť** charakterizujú tvar rozdelenia súboru a preto sa nazývajú **charakteristiky tvaru:**

Šikmosť je mierou akou sa súbor resp. jeho rozdelenie odlišuje od normálneho rozdelenia. Aritmetický priemer je potom bližšie ku „chvostu“ rozdelenia. Pearson dal ako jednoduchý odhad 1.koeficientu šikmosti použil vzťah

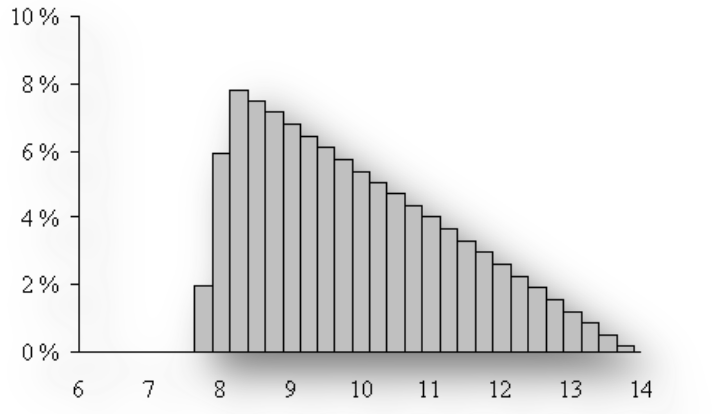
$$\gamma_1 = \frac{\bar{x} - mod(x)}{\sigma} \qquad [V.12]$$

Podľa definície a v modernej štatistike používame presnejší vzťah [9]:

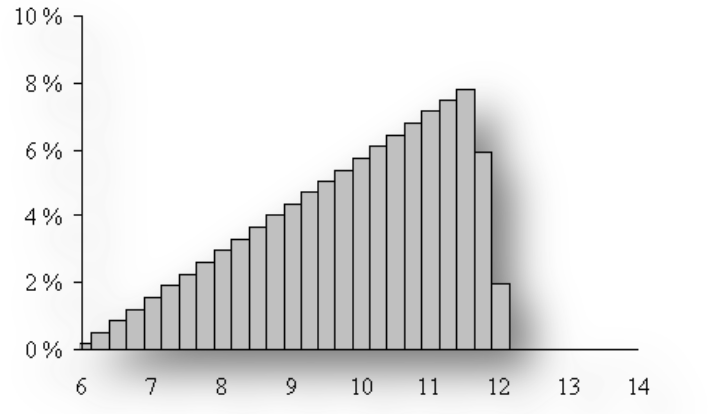
$$\gamma_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\sigma^3} \qquad [V.13]$$



Obr.V.6a. Symetrické rozdelenie



Obr.V.6b. Rozdelenie s kladnou (pozitívnou) šikmosťou



Obr.V.6c. Rozdelenie so zápornou (negatívnou) šikmosťou [9]

Z obr.V.6a až c vidíme, že môžu nastať 3 prípady:

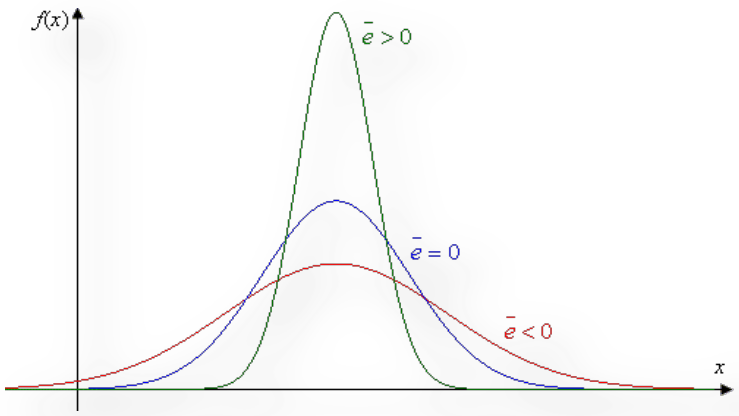
1. $\gamma_1 = 0$, rozdelenie je symetrické $\bar{x} = mod(x) = med(x)$
2. $\gamma_1 > 0$, rozdelenie s kladnou šikmosťou, $\bar{x} > med(x) > mod(x)$
3. $\gamma_1 < 0$, rozdelenie so zápornou šikmosťou, $\bar{x} < med(x) < mod(x)$.

Prvý prípad predstavuje normálne rozdelenie, aj nám už známe a dosť sympatické. Šikmosť (aj špicatosť) poukazujú na to, ako je rozdelenie odlišné od normálneho. Môže to byť spôsobené rôznymi príčinami (malý súbor, nepresné meranie, systematická chyba a pod.), pritom predpokladáme, že rozdelenie by malo byť normálne. Ale sú rozdelenia zošikmené už vo svojej podstate. Rozdelenie s $\gamma_1 > 0$ t.j. s kladnou šikmosťou je napr. rozdelenie rodín podľa počtu detí (skúste si to predstaviť, nakresliť, alebo nájsť nejaké štatistiky a dať ich do histogramu). Podmnožinou kladne zošikmených rozdelení súborov sú tzv. **L-rozdelenia** s maximálnou početnosťou na začiatku a potom najprv strmo, v ďalšom zvoľna klesajú. Populárne sa L-rozdelenie zvykne nazývať **rozdelenie krásnych vecí**. Pretože mu podlieha výskyt krásnych a vzácnych vecí a úkazov. Rozdelenie početnosti vlastníctva krásnych a drahých briliantov v populácii asi nikdy nebude mať tvar Gaussovej krivky. Podobne bohatstvo, reálny finančný príjem jednotlivcov atď. Opačné **J-rozdelenie** je omnoho vzácnejšie, ako príklad si môžeme uviesť náš subjektívny názor, akú mzdu by sme mali za svoju prácu dostávať.

Špicatosť (exces) je druhou charakteristikou tvaru. Charakterizuje výskyt extrémnych hodnôt v súbore. Má vzťah

$$e = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\sigma^4} - 3 \qquad [V.14]$$

Predstavu o špicatosti rozdelenia dáva obr. V.7:



Obr. V.7.: Posúdenie symetrického rozdelenia z hľadiska parametra špicatosti e, pričom e = 0 predstavuje normálne rozdelenie.

Týmto máme popisnú štatistiku takmer kompletnú. Ak sa v duchu pýtate, či toto všetko sme v relatívne jednoduchom príklade V.3. počítali, prezradíme vám tajomstvo. Ani nás nenapadlo! Na to sme príliš leniví. Ak ste podobný typ ako ja, tak potom sa taktiež nezaobídete bez programu EXCEL, v ktorom je to všetko zakomponované, len je potrebné vedieť ako na to; a to vám teraz vysvetlíme:

1. Ak máte menší alebo stredne veľký štatistický súbor s rozsahom n <256, tak vám EXCEL postačí. Pokiaľ pracujete s väčšími súbormi, tak ste buď študent a ten si musí poradiť, alebo ste na ceste stať sa profesionálnym štatistikom, pracujúcim v tíme, kde máte vždy možnosť to na niekoho buď menej skúseného alebo absolventa matfyzu hodiť. Dá sa to aj obísť, rozdelením väčších súborov na menšie a pod., ale to už ponechávam na vašu fantáziu.

2. EXCEL okrem rôznych štatistických funkcií pre jednotlivé výpočty má aj podprogram pre spracovanie štatistického súboru štatistickými metódami nazvaný ***Analýza dát***. K nej sa dá dostať cez príkazy ***Nástroje – Analýza dát***. V prípade, že po zadaní príkazu ***Nástroje*** nenájdete v ponuke ***Analýza dát***, tento balík si musíte doinštalovať. Budete postupovať nasledujúcim spôsobom. Vyberiete príkaz ***Doplnky*** a označíte ***Analytické nástroje***. Po potvrdení tlačidlom ***OK*** sa balík ***Analýza dát*** doinštaluje a môžete začať pracovať.

3. Stačí vám do EXCELu zadať zjednodušenú tabuľku nameraných údajov z nášho príkladu

i	$x1_i$	$x2_i$	$x3_i$
1	240	161,5	159,8
2	165	163	160,3
3	145	160	160,1
4	195	159	159,9
5	135	159,5	159,9
6	165	158	160
7	155	159	160,2
8	225	162	160,1
9	195	163	159,8
10	180	165	159,9

4. V hlavnom menu EXCELU si zvolíte hlavičku ***Údaje*** a vpravo hore ***Analýza dát***. Otvorí sa okienko, v ktorom si vyberiete riadok ***Popisná statistika*** a stlačíte ***OK***. V ďalšom okne je bežec

na kolónke ***Vstupná oblasť*** (v českej verzii ***Vstupní oblast***), ktorú vyplníte jednoducho tak, že ľavou myšou prebehnete prvý stĺpec dát (od hodnoty 240 po 180). V kolónke sa zobrazí rozsah buniek, kde sú dáta uložené a vám stačí stlačiť OK. Skočte a kliknite na malé koliesko pred okienkom ***Výstupní oblast*** a ešte raz kliknite na zviditeľnené okienko vpravo, aby vám tam blikal kurzor. Potom kliknite na niektoré voľné pole mimo tabuľky, jeho koordináty sa zjavia v okienku. Ešte potrebujete zadať, čo vlastne chcete. Kliknite si na prázdny štvorček vľavo od nadpisu ***Celkový přehled***. Myslíme, že to nebolo príliš zložité, keď v tom získate rutinu, je to určite mnohonásobne menej namáhavé, ako vypočítavať všetky odchýlky, ich mocniny, súčty, funkcie a pod. Dá sa to robiť naraz pre všetky tri stĺpce, ukazujem to však po jednom, aby to bolo názornejšie a možno aby ste si to mohli na jednoduchom probléme odskúšať.

5. Stlačte OK. Ako zázrakom sa zjaví nová tabuľka (ponechávame ju tak, ako sme ju dostali, aj v jazykovej mutácii nainštalovanej na našom počítači):

Sloupec1	
Stř. hodnota	180
Chyba stř. hodnoty	10,74968
Medián	172,5
Modus	165
Směr. odchylka	33,99346
Rozptyl výběru	1155,556
Špičatost	-0,53829
Šikmost	0,55688
rozsah	105
Minimum	135
Maximum	240
Součet	1800
Počet	10

6. Ak si to porovnáte s tabuľkou V.2, zistíte, čo všetko sme touto excelovskou analýzou dostali. K výsledkom sme pridali už len výpočet σ a **V(%)** podľa **[V. 17]** a **[V. 11]**. Analogický postupujete pre 2. a 3.súbor (ďalšie stĺpce) a porovnáte výsledky. Hotovo!

Pre príklad V.1. dostanete takýmto postupom parametre popisnej štatistiky, čo je už kompletnejšia štatistická analýza, ktorú môžete slovne analyzovať :

Sloupec1	
Stř. hodnota	216
Chyba stř. hodnoty	21,78383
Medián	174,5
Modus	83
Směr. odchylka	147,7451
Rozptyl výběru	21828,62
Špičatost	2,83989
Šikmost	1,455268
rozsah	729
Minimum	23
Maximum	752
Součet	9936
Počet	46
σ	149,3777
V(%)	69,15635

Podobne si môžeme zanalyzovať aj príklad V.2. pre platy sociálnych pracovníkov.

Ako sa pracuje s intervalovým delením a početnosťami si uvedieme v nasledovnom príklade:

Pr.V.4.: Dotazníkovou metódou a nepriamym overovaním pomocou dlhodobého pozorovania boli v rámci projektu *Fialový nos* zisťované priemerné mesačné výdavky 50 rodín na alkoholické nápoje. Respondenti mali zaškrtnúť svoje výdavky v možnostiach 100 až 700.- €/mesačne po stovkách. Výsledky sú zhrnuté v nasledujúcej tabuľke. Urobme spolu štatistickú analýzu:

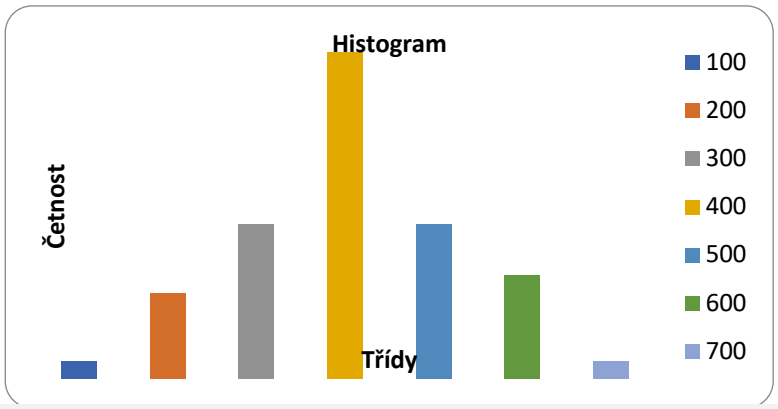
Rozsah súboru n = 50

Sledovaný štatistický znak: Odhad priemerných mesačných výdavkov na alkoholické nápoje v stovkách €.

Priemerné mesačné výdavky vybraných rodín na alkoholické nápoje [€]				
200	400	200	400	300
100	200	500	600	500
400	600	400	400	300
300	400	300	300	300
200	400	500	400	500
400	600	400	500	200
700	500	500	300	400
600	400	400	400	400
400	600	500	300	600
500	400	300	400	400

V Exceli máme v štatistickom podprograme ***Analýza dát*** ďalšiu zaujímavú možnosť spracovania údajov, nazvanú ***Histogram***. Umožní po vložení ***Vstupní oblast*** ako v predchádzajúcom prípade a ***Hranice tříd***, kde môžeme vložiť číslice 100;200;300;400;500;600;700 . ***Výstupní oblast*** vložíme štandardne nejaké voľné políčko kliknutím naň myšou. Ešte dolu klikneme na posledný štvorček ***Vytvořit graf*** a stlačíme ***OK***. MS Excel nám vráti tabuľku rozdelenia do tried a početností v nich z celého výberového súboru:

Třídy	*Četnost*
100	1
200	5
300	9
400	19
500	9
600	6
700	1
nad	0



Ako užívatelia Excelu si to všetko trochu upravíme a doplníme (aj pomocou lit. napr. [10]):

i	Triedy x_i	f_i	$x_i.f_i$	$\|\Delta\|$	Δ^2	$\Delta^2.f_i$
1	100	1	100	304	92416	92416
2	200	5	1000	204	41616	208080
3	300	9	2700	104	10816	97344
4	400	19	7600	4	16	304
5	500	9	4500	96	9216	82944
6	600	6	3600	196	38416	230496
7	700	1	700	296	87616	87616
spolu		50	20200	1204	280112	799200

Histogram nám dáva predstavu o symetrii rozdelenia početnosti súboru. Tabuľka umožňuje vypočítať potrebné charakteristiky popisnej štatistiky, stačí na to kalkulačka. V Exceli už nie je možnosť kompletného spracovania takéhoto súboru údajov, alebo sme to aspoň nenašli.
Strednú hodnotu, v našom prípade **vážený výberový aritmetický priemer**, vypočítame podľa vzťahu **[IV.3]**:

$\bar{x} = 404$

Jednotlivé odchýlky **Δ** podľa **[V.4]**, hodnoty Δ^2 a $\Delta^2.f_i$ nie sú žiaden problém, ako aj ich súčty. Výberový rozptyl **D(X)** = σ^2 dostaneme podľa vzťahu **[V.6]**, odmocnením aritmetického priemeru druhých mocnín odchýlok dostaneme smerodajnú odchýlku **σ**.

D(x) = 16310,2

σ = 126,4278

Výberovú smerodajnú odchýlku jedného merania **s** dostaneme vynásobením **σ** odmocninou z **n/(n-1)** t.j. z 50/49 resp. výberová smerodajná odchýlka strednej hodnoty $s_{\bar{x}}$ sa získa vydelíme odmocninou z **n**:

s = 127,7114

$s_{\bar{x}}$ **= 18,06112**

S tabuľky početnosti ľahko zistíme, že

med(x) = mod(x) = 400

R = 600

x_{min} **= 100**

x_{max} **= 700**

a taktiež vypočítame šikmosť, špicatosť a variačný koeficient **[V.11]**, **[V.13]** a **[V.14]**:

γ_1 **= -0,01626**

e = -0,11027

V(%) = 31,29402

Záver:

Okrem tabuľky rozdelenia početnosti a histogramu máme parametre popisnej štatistiky. Ospravedlňujeme sa za nezaokrúhlené čísla, nikoho sme nechceli ohúriť mimoriadnou

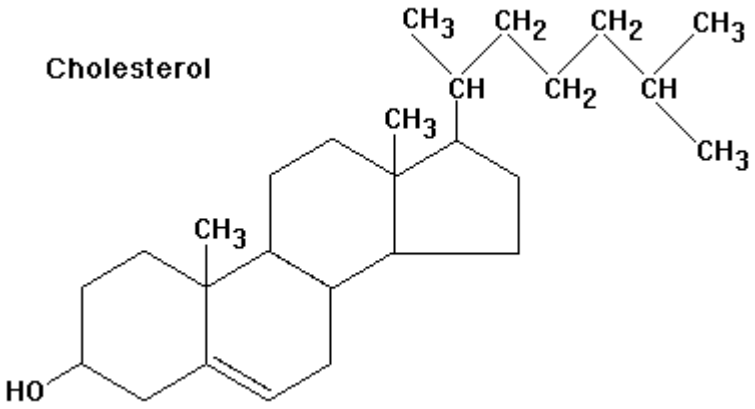
presnosťou a exaktnosťou výpočtov, je to len na to dobré, aby si čitateľ pri vlastných výpočtoch mohol robiť bez problémov kontrolu.

Podľa nášho zadania je priemerný mesačný výdaj na alkoholické nápoje jednej rodiny skúmaného výberového súboru $\bar{x} = 404$ €. Hodnota med(x) = mod(x) = 400 sa líši iba nepatrne od odhadu výberového aritmetického priemeru, preto výberový aritmetický priemer možno považovať za vhodný odhad strednej hodnoty súboru. Jeho smerodajná odchýlka je $s_{\bar{x}}$ **= ±18 €**. Výberový súbor má variačné rozpätie **R = 600, x_{min} = 100** a **x_{max} = 700 €.** Smerodajná odchýlka jedného merania je **s = 128 €** a variačný koeficient **V(%) = 31,3**. Nízke hodnoty charakteristík tvaru (špicatosť a šikmosť) naznačujú na normálne rozdelenie početnosti. Nepresnosť vstupných údajov spôsobená možnosťami dotazníka zaškrtnúť sumu v celých stovkách € je prijateľná. Iné chyby na vstupe sú podmienené tým, že respondenti, hlavne utrácajúci väčšie sumy zvyknú podhodnocovať svoje výdavky. Taktiež výdavky rodín nemusia presne kopírovať spotrebu alkoholu, ktorá je ovplyvnená ešte inými faktormi (kupovanie lacnejšieho alkoholu, vlastná výroba, atď.).

Niekedy je vhodné popisnú štatistiku a histogram doplniť usporiadaním súboru do variačného radu a priradením percentilov ku každému bodu. A to v Exceli, v štatistickom podprograme ***Analýza dát*** analytický nástroj ***Pořadová statistika a percentily***. Spracovaním pôvodnej náhodne zloženej tabuľky terénnych dát sme získali bohatú škálu nových informácií, ktoré nás informujú o vlastnostiach štatistického súboru.

Načrtli sme v predchádzajúcom texte, ako spracovať údaje získané v teréne a zahrnuté do pracovného štatistického súboru. Interpretáciou výsledkov sme sa nezaoberali, alebo len veľmi zľahka, okrajovo. Je to veľký problém štatistiky, pretože niekedy aj poriadna štatistická analýza, keď sa preženie mediálnym priestorom, tak sa nestačíme diviť, čo prinesie. Iným príkladom môže byť, keď profesionálny vedec, odborník, niečo objaví, podloží to kvalitnou štatistickou analýzou, ale interpretácia výsledkov vytrhnutých z omnoho zložitejšieho dynamického systému spôsobí to, že kvalita a presnosť použitých analytických nástrojov sú vlastne zavádzajúce a dostaneme sa napriek ich kvalite na úroveň neandertálskej štatistiky.

Medzi zaujímavé trochu zložitejšie látky, o ktorých už každý počul, patrí napríklad aj známy výživový strašiak cholesterol. Jeho sumárny chemický vzorec je $C_{27}H_{45}OH$, ale lepšiu predstavu nám dáva štruktúrny vzorec na obr.V.8.:



Obr.V.8.: Štruktúrny vzorec molekuly cholesterolu [11]

Ešte krajší má systematický chemický názov podľa Medzinárodnej únie pre čistú a aplikovanú chémiu IUPAC (The International Union of Pure and Applied Chemistry):

(10R,13R)-10,13-dimethyl-17-(6-methylheptan-2-yl) - 2,3,4,7,8,9,11,12,14,15,16,17-dodecahydro-1H-cyclopenta [α]phenanthren-3-ol

To je, prosím, jedno slovo, určite úctyhodné, pred ktorým sa skloní v nemom úžase nejeden čitateľ. Našťastie na jeho dešifrovanie máme profesionálnych chemikov a aj keď si to mnohí nemyslia, nejde tu o žiaden alchymistický tajuplný kód.

Je to príklad molekuly, opradenej množstvom štatisticky podložených a napriek tomu iracionálnych informácií, zapustených ako mémy do obecného povedomia, ktoré majú mimoriadnu stabilitu a životnosť. Zdá sa, že je to podobné ako s malými deťmi. Keď ich chcete niečo užitočné naučiť, je to veľmi namáhavý proces so smutne malou účinnosťou. Ale keď vám v nestráženom okamihu vyletí z úst výraz, po ktorom by ste si najradšej odhrýzli jazyk, máte istotu, že ho vaša ratolesť okamžite prevzala do svojho slovníka a bude ho ešte veľmi dlho hlasno používať v tých najnevhodnejších situáciách.

Všetci vedia, že cholesterol je zlý a že nám hrozí smrteľným nebezpečenstvom. Poniektorí počuli, že existuje „dobrý“ a „zlý“ cholesterol, a už len učené hlavy, alebo možno ešte lúštitelia krížoviek vedia, že ten zlý je LDL a ten dobrý HDL cholesterol. A okrem úzkeho kruhu špecialistov, ktorí odolali tlaku farmaceutických firiem, už naozaj málokto vie, aký je cholesterol v našom organizme dôležitý. Či už pri stavbe nervovej sústavy, mozgu a viacerých dôležitých hormónov, ale aj samotných buniek resp. ich membrán, že si ho organizmus tvorí z veľkej väčšiny sám a len časť dostávame potravou, že len s jeho pomocou môžeme

v organizme spracovávať tuky. Jeho nadbytok v krvi môže byť samozrejme varovným príznakom, ale civilizačné choroby, medzi ktoré patria aj dnes časté ochorenia kardiovaskulárneho systému, sú multifaktoriálnym javom. Nie je možné bez ostatných súvislostí vytrhnúť z kontextu len problematiku cholesterolu, aj to veľmi skreslenú, a zachrániť zbedačenú populáciu, ako to na škodu veci robili viacerí odborníci na problematiku výživy. Rozkošným dojmom pôsobí, keď vám na malom prenosnom stolíku v tržnici medzi kyslou kapustou, zemiakmi a kvetmi ponúkajú pomocou malého ale veľmi učene a teda dôveryhodne sa tváriaceho prístroja expresné vyšetrenie cholesterolu v krvi. Úplne bezplatne! Ak ste vošli do tržnice ako bezstarostný povrchný jedinec, ktorý chcel svojmu zdravím kypiacemu organizmu dopriať zopár kulinárskych rozkoší, po vyšetrení odchádzate ako biedna zruinovaná ľudská troska, ktorá nemá istotu, či sa vôbec dotacká k najbližšej internej alebo kardiologickej ambulancii. Štatisticky i psychologicky zaujímavý je aj jav, že manipulácii pri pultovom vyšetrení cholesterolu podliehajú vo svojej zvýšenej dôverčivosti aj starší občania, ktorí sa dožili už nadpriemerného veku, a hneď ako im obsluhujúci odborný personál prezradí, že pri takých koncentráciách cholesterolu v krvi, aké im práve našťastie odhalili, sa nemusia dožiť vysokého veku, sa ponáhľajú nakúpiť do lekárne vrelo odporúčané výživové doplnky. Veď čo by človek neurobil pre svoje zdravie, že? Hlavne, keď to má z prvej ruky od odborníkov. Táto publikácia má jeden zo svojich cieľov, v čitateľoch vybudovať pokiaľ sa to dá, určitý cit pre takéto štatistiky a ich interpretácie. Ale moje odporúčanie v tomto prípade: Bude užitočnejšie miesto trhového vyšetrenia na cholesterol zájsť o tri kroky ďalej a kúpiť si kyslú kapustu, ktorá je bohatým zdrojom vitamínu C.

Aby sme neboli príliš pesimistickí, uveďme si ešte jeden zaujímavý príbeh, ktorý by sme mohli nazvať *„štatistika pre život“*:

V záhrade augustiánskeho kláštora v Brne v srdci Moravy sedí plachý chlapec a nesmelo sa prizerá bujarým neviazaným hrám svojich vrstovníkov. Niekedy sa mu dostane od nich toľko posmeškov, že by najradšej utiekol, ale nakoniec sa vždy nájde niekto, kto mu pomôže prekonať jeho bojazlivosť a nadviazať kontakt.

Johann Gregor Mendel (1822 - 1884)

Predstavení kláštora však vidia trochu viac. Za chlapcovou utiahnutou povahou sa skrýva neobyčajná hĺbavosť a vytrvalosť, preto ho poslali na štúdia.

A dobre urobili. Mladý Johann Gregor Mendel (1822 - 1884) sa postupne vypracoval až na odborného asistenta známeho fyzika a prírodovedca, profesora Christiana Dopplera. V práci bol veľmi svedomitý a precízny. Avšak v kontakte s ľuďmi sa nikdy nezbavil svojich úzkostí; čo bolo asi jednou z hlavných príčin, prečo štúdiá predčasne opustil a vrátil sa do brnenského kláštora, kde sa nakoniec stal, ako to už paradoxy ľudských osudov prinášajú, jeho významným opátom. Ale to predbiehame, pretože dôležité je aj to, čo robil medzitým, čo jeho rehoľní bratia občas dobromyseľne nazývali *hranie s kytičkami.* Trpezlivo experimentoval s krížením hrachu a systematicky dlhodobo zaznamenával prenos viditeľných vlastností na ďalšie generácie. Pomohla mu prax u profesora Dopplera, alebo vlastná intuícia, že jeho experimenty neskončili len ako dobrá hrachová kaša v kláštornej kuchyni? Svoje výsledky podrobil analýze pomocou matematickej štatistiky a prekvapujúco odhalil zaujímavé zákonitosti. Vďaka tomu mohol postulovať niektoré základné princípy prenosu vlastností ako napríklad farba kvetov a pod. v živých systémoch, zatiaľ bez väčšieho zovšeobecnenia. Vo vede, ktorá sa dovtedy zaoberala takmer výlučne popisom živých organizmov – biológii - boli položené základy dedičnosti, modernej genetiky. Odštartoval tak neobyčajný proces nového nazerania na život. Hľadanie jeho princípov, mechanizmov a základov viedlo stále k hlbším a miniatúrnejším sférám. Bola objavená bunka a postulovaná Wirchovová dogma, že každá bunka na tomto svete mohla vzniknúť len z inej bunky, žiadny iný mechanizmus zatiaľ nepoznáme, ale aká úžasná cesta plná príbehov a lemovaná rôznofarebnými kvetmi hrachu sa otvorila. [12]

Poďme ešte k tak nepatrnej, bez výkonného mikroskopu neviditeľnej veci, akou je

bunečné jadro. V objavovaní jeho tajomstiev je skrytý ďalší zaujímavý príbeh, odohrávajúci sa koncom 40. a začiatkom 50. rokov dvadsiateho storočia v známom Cavendishovom laboratóriu Cambridgskej univerzity, v londýnskej King´s College a v kalifornskom Caltechu na druhej strane Atlantiku. Je to príbeh niekoľkých fyzikov a chemikov, ktorí prenikli na pole molekulárnej biológie a ich spolupracovníkov i konkurentov, pracujúcich na štúdiu komplikovaných priestorových štruktúr základných stavebných prvkov živých organizmov –bielkovín. K Francisovi Crickovi, dynamickej osobnosti tohto výskumu, ktorého postavu a prejavy nebolo možné v Cavendishovom laboratóriu prehliadnuť, ba dokonca pred ktorého pozornosťou

a výpadmi mnohí spolupracovníci utekali do tichších a pokojnejších zákutí, pribudol v r.1951 Američan James D.Watson. Ich tvorivá spolupráca sa čoskoro orientovala výlučne na odhalenie štruktúry deoxyribonukleovej kyseliny DNA, ktorá sa nachádza v bunečnom jadre a ktorá bola podozrivá z prenosu genetických vlastnosti. Ich výskum nadväzoval na mravenčiu prácu Mauricea Wilkinsa z King´s College v oblasti röntgenovej štruktúrnej analýzy. Silným konkurentom v pretekoch na odhalení štruktúry DNA im nebol nikto menší, ako svetoznámy americký chemik Linus Carl Pauling, objaviteľ mnohých komplikovaných štruktúr zložitých biomolekúl a neskorší dvojnásobný držiteľ Nobelovej ceny (za chémiu a za mier), pôsobiaci v tom čase na Kalifornskom technologickom inštitúte. Z tohto vzrušujúceho súboja vyšla víťazne anglická skupina, ktorej bola udelená Nobelova cena v r.1962. Ich zásluhou bola nielen objavená priestorová štruktúra dvojzávitnice DNA, ale odhalením jej tajomstva sa odhalil základ a informačná podstata života a jeho toku, ako aj prenosu dedičných vlastností rodičov na potomstvo a posunulo poznanie v tejto oblasti obrovským skokom vpred. Za všetkým je skrytá aj kvalitná štatistika. V súčasnosti z nej čerpajú napríklad všetky moderné biotechnologické aplikácie. Asi stačí na udelenie jednej Nobelovej ceny. Cesta je otvorená, možno sa ľuďom z mnohých ešte ani netušených možností zatočila hlava, mnohí majú a treba podotknúť, že často oprávnený pocit, že stojíme na okraji priepasti... [13].

Zastavme sa ešte pri jednom významnom míľniku vedy, ktorý určite ovplyvnil ďalšiu históriu ľudskej spoločnosti viac ako si dnes vieme predstaviť. Ide o správu o úspešnom zavŕšení projektu HUGO (HUGO - Human Genome Organization, založená v r.1988), ktorý po istý čas viedol aj vyššie spomínaný laureát Nobelovej ceny za objav štruktúry DNA, profesor James D.Watson. Na prelome tisícročí sa podarilo zmapovať kompletnú sekvenciu ľudského genómu, teda popísať genetický kód, všetky gény, ako základnú absolútnu aj štatistickú informáciu o postupnosti ich zložiek. Jedným z dôsledkov tohto objavu je aj fakt, že ľudia rôznej farby pleti, teda rôznych rás ako sa zvykne hovoriť, sa líšia v genetickej informácii tak nepatrne, že táto odlišnosť nedáva žiaden podklad k hypotézam o rozdieloch medzi ľuďmi. Je to prekvapujúce? [14]

Literatúra k V. kapitole

[1] Koľko jahôd rastie na mori? Slovenský spisovateľ, 1992, preložil Ľubomír Feldek

[2] Kosara, R.: Shining a Light on Data: Florence Nightingale, 2009
in http://eagereyes.org/blog/2009/shining-a-light-on-data-florence-nightingale

[3] Chajdiak, J.: Štatistika jednoducho, STATIS, Bratislava 2010

[4] https://www.upsvr.gov.sk/statistiky/aktivne-opatrenia-tp-statistiky/aktivne-opatrenia-trhu-prace-2013-1.html?page_id=380213

[5] http://www.naseplaty.sk/prehlad-platov/treti-sektor-neziskovky-nadacie.html

[6] GIBILISCO, s.: Statistika bez předchozích znalostí, Computer Press, Brno, 2009

[7] Zvára,K., Štěpán,J.: Pravděpodobnost a matematická statistika, Matfyzpress, Praha 2006

[8] https://www.google.com/#q=Karl+Pearson

[9] http://rimarcik.com/navigator/och.html

[10] Chajdiak, J.: Štatistika v EXCELI, STATIS, Bratislava 2002

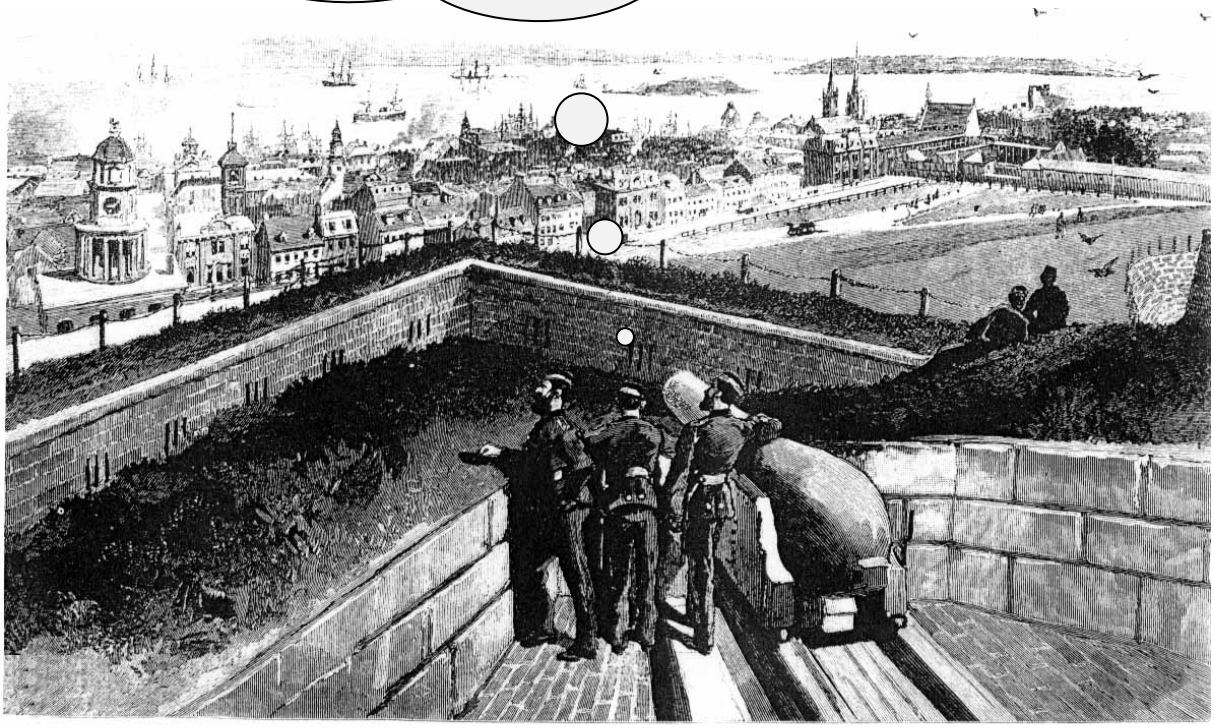
[11] http://sk.wikipedia.org/wiki/Cholesterol

[12] Komárek, S.: Dějiny biologického myšlení, Vesmír, Praha, 1997

[13] Watson, J.D.: Tajemství DNA, Academia, Praha 1995

[14] http://www.hugo-international.org/

Podľa štatistík , ľudia nás tu majú radi. Hlavne v dostrele ...



VI. Ako si dobre vybrať alebo spoľahlivosť veľká cnosť

Máme radi rozprávky, aby aspoň niekedy vyhrala spravodlivosť.
Anonymný štatistik

V 50.rokoch 20.storočia si dala jedna nemecká firma u Svetovej zdravotníckej organizácie WHO patentovať látku pod názvom Thalidomid, na trh sa dostala pod obchodným názvom c*ontergan*. Ako sedatívum a hypnotikum, voľne prístupné. Niekoľko rokov sa zdalo, že je to veľmi dobré sedatívum a bolo vrelo odporúčané aj budúcim mamičkám. Začiatkom 60.rokov sa začali objavovať sťažnosti na niektoré dráždivé a svrbivé pocity v končatinách. Na odporúčanie Spolkového zdravotníckeho úradu v Nemecku bol liek zaradený len na predpis. Avšak už v r.1956 pracovníčka firmy porodila malformované dieťa. Skrylo sa to v štatistike! Ale potom to prepuklo. Pamätníci spomínajú na aféru s conterganovými deťmi, s chýbajúcimi údmi a inými časťami tela, bolo ich niekoľko tisíc. Samozrejme súdne procesy, obrovské ľudské tragédie a zbytočné utrpenie, celé veľmi dramatické. Ale nechceme strašiť, od toho sú masmédia, chceme povedať iné:

V USA dostala *contergan* na starosť Dr. Frances Oldham Kelseyová, mladá a čerstvá pracovníčka Úradu pre potraviny a liečivá FDA, s právomocami schvaľovať liečivá a prípravky. Šesť krát žiadosť o registráciu lieku zamietla. Pritom bol na ňu vyvíjaný neobyčajný lobistický nátlak zo všetkých strán, dokonca od mnohých lekárov, či kolegov. Medzi iným tu bola aj snaha zbaviť ju svojprávnosti. Možno zdôrazniť jednu skutočnosť - táto mladá dáma predstavovala kombináciu osobnej statočnosti a dostatočného odborného a vedeckého poznania, teda toho čo môžeme nazvať „vedecká gramotnosť“. Medziiným aj rozumela štatistike. Dôvod jej odmietnutia lieku bol, stručne povedané, nasledovný: Správa firmy o testoch (klinických aj predklinických) bola príliš oslnivá na to, aby to mohla byť pravda. Spoľahlivosť nebola preukázaná a podložená dostatočnou štatistickou analýzou. Asi je už vedľajšie, že v ďalšom roku po prepuknutí aféry dostala od vtedajšieho prezidenta J.F.Kennedyho zlatú medailu za statočnosť a v rámci FDA sa stala jednou z riadiacich pracovníčok, dodávam to len preto, že máme radi rozprávky s dobrým koncom [1], [2]. Pre tých, ktorí uprednostňujú aj dramatické filmové spracovanie skutočných príbehov [3].

Zdá sa vám, že sa to týkalo ďalekej a dávno prekonanej minulosti? Nie je skutočnosť skôr taká, že poverčivosť ľudí sa preniesla dnes skôr na pole neandertálskej štatistiky? Veď trh s vývojom a distribúciou liekov pre väčšinu chorôb je prakticky nasýtený. Farmaceutické

spoločnosti predstavujúce dnes úplne odosobnený biznis, veľmi vzdialený od reálneho pacienta, museli sa orientovať na zdravých ľudí. Súčasťou je „strašenie“ všetkým, čo potenciálne alebo reálne vo vás drieme a hlodá; a masívna zavádzajúca reklama. Punc objektivity a vedeckosti tomu dodáva štatistika. A môže sa rozbehnúť trh s rôznymi odporúčanými prípravkami, minerálmi, vitamínmi, homeopatikami, preparátmi mimo lekárskych predpisov či receptov a výživovými doplnkami. Rast ich predaja si môžete overiť v rôznych štatistikách, alebo si urobiť vlastný odhad z toho, ako napriek kríze nakupujú vaši blízki a známi.

Urobme si na túto tému vlastný myšlienkový štatistický experiment, ktorý neberte, prosím, ako spracovanie podnikateľského zámeru. Vytvorme najprv Príbeh:

Starý muž, veľký majster Pien Čchüe sedel v trstinovom kresle v *Záhrade mesačného svitu* svojho hostiteľa a mecenáša, kniežaťa prímorskej krajiny Amaravati, ku ktorému sa uchýlil na úteku pred krvavým cisárom Čchin Š'-chuang-ti. Pripravoval lieky pre prvú ženu kniežaťa, do ktorých vkladal ducha múdrosti starých liečiteľov. Podvečerné slnko prenikajúce cez zeleň sviežich jarných stromov prekrásneho kraja nad červeno sa trblietajúcim zálivom vytváralo farebné pohyblivé škvrny na stole a rôznych fľašiach, pohároch a nádobách, ktoré ho zapĺňali. Očakávaná chvíľa, tajomný neopakovateľný čas medzi dňom a nocou mu zbystrili všetky zmysly a tak si všimol nový vzhľad destilátu v čírej sklenenej nádobe, do ktorého padlo niekoľko lupienkov kvetu starej jablone. Elixír vydával striedavo tajomne zelenkastú a ružovkastú opalescenciu, ba žiaril aj po zotmení a mal jemnú vôňu jari. Pien Čchüe bol majster svojho remesla, veľký vedec a liečiteľ, ktorý rozumel tomu, čo je posolstvo starých mudrcov. Po podaní elixíru žene svojho dobrodinca došlo k skorému uzdraveniu. Tak sa na vedeckom základe zrodil neobyčajný liek, ktorý nazval *gió xuân*, po našom *Jarný vánok*.

Príbeh je to „starobylý“, zasadený do histórie a geografie juhovýchodnej Ázie, dostatočne vzdialený v čase a v priestore, aby sa nedal vystopovať, ale na druhej strane poukazuje na vysoké kultúrne hodnoty starých ríš a na dlhú tradíciu toho, čo chceme nakoniec

ponúknuť. Príprava preparátu je jednoduchá: Do čistého liehu vhoďte za hrsť lupeňov jabloňového kvetu a nechajte nejaký čas postáť. Nič vhodnejšieho sme narýchlo nenašli, teda nič rovnako neškodného a pritom ešte nevyužívaného súčasnou medicínou alebo liečiteľstvom. Pokiaľ chcete dosiahnuť silnejší „jablkový" efekt, pridajte jablčné šupky, poprípade prilejte trochu *calvadosu*. A môžete začať ponúkať na trhu svoj univerzálny liečivý, alternatívny, 100% prírodný produkt na mnohé neduhy, ktoré gniavia súčasné ľudstvo. Možno väčšiu údernosť bude mať jeho latinský názov *Vere aura*. V dnešnej dobe, z nejakého dôvodu nazvanej „racionálna", by to chcelo prizdobiť niečím odborným, trochou vedeckosti. Laboratórne testy sú drahé a nemusia hneď vyjsť podľa potreby, ideálna je však štatistika. Urobte si vlastný výskum. Vezmite si nejakú bežnú pohromu, napr. nádchu. Sledovaným ukazovateľom môže byť napr. potrebná doba liečenia, tento čas bude podliehať veľmi pravdepodobne normálnemu rozdeleniu s nejakou strednou hodnotou a smerodajnou odchýlkou. Presvedčte štyroch svojich priateľov resp. známych, že im chcete ponúknuť starobylú dávno odskúšanú medicínu na liečenie ich nachladnutia. (Samozrejme, že to môžete robiť aj s inými neduhmi a ochoreniami, fantázii sa medze nekladú, pretože máte poruke úplne neškodný a neúčinný preparát. Nevstupujte však na pole naozaj zhubných ochorení, aby ste nebránili ich ozajstnej liečbe.) Viac respondentov do pokusu nepripusťte, pretože potrebné efekty môžu nastať len pri veľmi malých rozsahoch štatistického súboru. Jeden z respondentov nech sa lieči klasickým spôsobom, pomocou súčasnej medicíny a farmácie. To je kontrolná vzorka. Ďalším trom podávajte napr. 10 kvapiek do vlažného čaju denne.

Čo nasleduje? Vašich respondentov rozdeľuje predpokladané normálne rozdelenie na tri intervaly 1;2;1. Čím? Napr. časom liečenia nádchy: jeden ju prekoná za podpriemerne krátky čas, dvaja sa budú pohybovať okolo priemernej doby jej liečby, a jednému to potrvá o niečo dlhšie. Toto sa udeje, aj keby sa váš kompletný súbor výrazne kvalitatívne odlišoval od celej ostatnej populácie. V tomto rozdelení môžu nastať tri prípady s rovnakou pravdepodobnosťou:
a) kontrolná vzorka, teda jedinec liečiaci sa klasicky, bude vyliečený za podpriemerne krátky čas, ostatným to bude trvať dlhšie,
b) kontrolná vzorka spadne do priemeru,
c) kontrolnej vzorke liečba potrvá najdlhšie.
Ak nastane prípad a), odporúčam experiment opakovať, pretože štatistická analýza ani pri veľkej snahe neprinesie potrebný efekt.
Prípad b) je možné popísať nasledovne: 2/3 pacientov, ktorí sa podrobili liečbe nachladnutia pomocou prírodného prostriedku *Jarný vánok (Vere Aura)* dosiahlo výsledky liečby zrovnateľné alebo omnoho lepšie ako pri štandardnej medicíne. Štatisticky bolo dokázané, že

až 33,333% (počet desatinných miest je dôležitý, neobmedzujte sa) pacientov užívajúcich váš overený elixír *Vere Aura* malo výsledky liečby zrovnateľné s výsledkami dnešného nákladného a organizmus zaťažujúceho zdravotníctva a ďalších 33,333% ich malo dokonca výrazne lepšie. Ak nastane prípad c) je to pre vašu reklamu najlepšie: Prakticky 75% celej sledovanej populácie, ktorá bola nakazená epidémiou nádchy bolo úspešne vyliečených vašim overeným liečivým doplnkom *Vere Aura*, pričom ich výsledky liečby dosahovali výrazne vyššie parametre ako pri bežnom postupe zaužívanom dnešným inštitucionalizovaným zdravotníctvom. Navyše 100% všetkých osôb, ktoré užili váš preparát malo výrazne kratší čas úspešnej liečby ako pri použití doterajších liekov.

Zdá sa vám to trochu prehnané a pritiahnuté za vlasy? Máte pravdu a je to tak zámerné. Pretože to nie je o nič lepšie ani horšie od ponúk takmer všetkých možných liečiteľských a doplnkových produktov na trhu. Časť z nich má lepšie spracovaný príbeh, dali reklamu do rúk profesionálnej agentúre, tak na ich obrázkoch vidíte múdro sa tváriacich odborníkov v bielom plášti a šťastné vyliečené rodinky na brehu jemne zvlneného jazera prežiareného prekrásnym počasím vrcholiacej jari. V televíznych spotoch sa takéto šťastím naplnené reklamy najlepšie uchytia hneď po správach, alebo uprostred katastrofického filmu pri vrcholiacej zápletke okolo nezadržateľne sa šíriacej epidémie hroziacej vyhubiť ľudstvo. Majú k dispozícii testy renomovaných laboratórií, ktoré opäť potvrdili normálne rozdelenie niečoho, čo si dali do súvisu s vlastnosťami svojho produktu, masívnu mediálnu reklamu a distribučnú sieť. A zamerali sa na nejaký populárny nešvár, dnes sú to hlavne straty kognitívnych funkcií centrálnej nervovej sústavy vekom, teda úplne prirodzený jav zvyšujúci svoju početnosť predlžovaním priemerného veku ľudí. Nenadertálskou štatistikou dosahujú väčšie zisky. Toto nebol návod na podvádzanie. Podvodníci to už dávno a určite dokonca omnoho lepšie poznajú. Je to skôr návod pre ostatných, ako na podobnú reklamu a štatistiku reagovať. Možno ste už prestali veriť, že keď vám prebehla čierna mačka cez cestu, budete mať mrzutosti, ale ešte stále veríte, že ak máte nadváhu a nie ste už úplne najmladší, že nepotrebujete schudnúť, upraviť životosprávu a viac sa športovo pohybovať, ale na boľavé kĺby vám stačí zakúpiť si ich výživu v podobe „štatisticky osvedčeného“ nejakého farmaceutického voľne dostupného doplnku. Navyše väčšina, ktorá si to už kúpila, vás bude presviedčať o tom, aké je to účinné, už len preto, aby nevyzerali ako hlupáci. Pritom sú v tom do značnej miery nevinne; zodpovednosť nesie ten kto klame a zavádza, pre nás zjednodušene neandertálec, nech je v akomkoľvek postavení alebo prestrojení. Je to zvláštny úkaz. Neandertálcom sa môže stať aj normálny človek, dokonca aj profesionálny štatistik, pokiaľ začne komukoľvek za peniaze ponúkať svoje vedomosti cielene

s vopred požadovaným efektom, teda prostituovať. Opačný jav nebol v histórii ešte zaznamenaný.

Dotkli sme sa závažnej problematiky rozsahu a všeobecne výberu štatistického súboru. Je to jeden z vážnych a bazálnych problémov induktívnej štatistiky. Ale pekne po poriadku:

Máte **naformulovaný** nejaký **problém** (napr. rozvodovosť, násilie v rodine, kvalita života seniorov, začlenenie imigrantov do spoločnosti, závislosti, problémy na pracoviskách atď.) a stanovenie cieľov výskumu (tie môžu byť od holého kvantitatívneho zistenia faktov a situácie až po teoretické zovšeobecnenia javov, môžu predstavovať praktické návrhy a odporúčania, návrhy zmeny financovania či legislatívy, dokonca môžu predvídať vývoj). V humanitných a zvlášť pomáhajúcich profesiách i v rôznych blízkych odboroch budete pracovať v teréne, získavať **primárne dáta** (**sekundárne dáta** získate z rôznych dostupných prehľadných štatistík pre vlastné použitie, ak už boli niekde získané a dajú sa pre vašu problematiku použiť) priamo od ľudí, ktorí predstavujú nositeľov rôznych sociálnych charakteristík, vzťahov, vlastností, potrieb, záujmov, postojov a iných javov v rámci kvantitatívneho alebo integrovaného výskumu. Preskúmať celý základný súbor, teda vyšetriť celú populáciu je u nás úlohou predovšetkým Štatistického úradu SR, aj ten si to môže dovoliť len raz za dosť dlhú dobu v rámci sčítania obyvateľstva. Tým chcem povedať, že je to veľmi drahé a časovo mimoriadne náročné. Pokiaľ budete jeho riaditeľom, už túto publikáciu nebudete veľmi potrebovať. V bežnej štatistickej praxi sa musíte zaoberať omnoho menšou vzorkou populácie, ktorá je limitovaná vašimi finančnými a časovými možnosťami. Túto vzorku voláme **výberový súbor**. Definujeme si najprv:

- Základný súbor, štatistickú jednotku, ktorej množina ho tvorí: napr. všetci občania a jednotlivý občan; ale aj domácnosť, firma, inštitúcia atď.
- Spôsob a rozsah výberu: väčšinou náhodný výber, ale aj iný.
- Konkrétna metodika práce v teréne (prieskum, dotazník, pozorovanie, iné).
- Chyby zberu údajov, ako im zabrániť alebo ako ich minimalizovať.

Formulácia problému, ktorý idete riešiť môže byť všelijaká. Pre toho, komu to posielate na schválenie a kto ho bude financovať resp. prideľovať nejaký grant, môže znieť veľmi učene, ale v podstate si to celé musíte zjednodušiť na jednoduchú a hlbokú otázku, ktorej formulácii porozumejú nielen respondenti, ale aj vy a váš riešiteľský kolektív. Otázky typu *Sledovanie multifaktoriálneho vplyvu na medzigeneračný prenos nestability rodín pri začlenení SR do štruktúr otvorenej spoločnosti a jeho dočasné obmedzenia národnou a nadnárodnou legislatívou,* si ponechajte radšej na nejakú veľmi vysoko hodnotenú konferenciu, kde sa treba ukázať, ale kde nikto nikoho v skutočnosti nepočúva. Taktiež výskum typu *Vplyv výchovy na*

stravovacie zvyky a návyky predškolských detí bude omnoho menej finančne aj časovo nákladný, keď sa to priamo spýtate niekoľkých viacnásobných starých mám. Výskum a jemu nápomocná štatistika je pre tých, ktorí sa na nič nehrajú (opakujem, to prenechávame pre neandertálcov) ale chcú niečo urobiť: zlepšiť poznanie, situáciu, zmierniť utrpenie a pod. Jednoducho formulované problémy ako napr. *Vyššia rozvodovosť potomstva z rozvrátených rodín*, alebo *Prvá skúsenosť s návykovou látkou*, umožňujú klásť jednoduché, zrozumiteľné a vyčerpávajúce otázky. Toto samozrejme nie je primárne problematika štatistiky, ale bolo nevyhnutné sa aspoň okrajovo venovať aj týmto otázkam metodiky humanitných vied, pretože majú významný vplyv na priebeh a výsledky štatistickej analýzy, ktorej produkt nemôže byť kvalitnejší, ako boli vstupné dáta. Dosť podrobne sa uvedenou praxou zaoberá napr. [4].

A teraz k **rozsahu výberu**. Má samozrejme svoje ekonomické a časové mantinely, ale aj očakávania. Poddimenzovaný výber neprinesie predpokladané efekty, naddimenzovaný výber bude plytvaním prostriedkov. Prvým krokom je eliminácia systematických chýb a odôvodnený predpoklad, že výber má nejaké náhodné resp. pravdepodobnostné rozdelenie. Taktiež obsah výskumu bude podmieňovať rozsah výberu.

Následne je zámerom štatistickej analýzy prepracovať sa na základe údajov o výberovom súbore ku všeobecným záverom o základnom súbore. Každý základný súbor je charakterizovaný určitými znakmi, ktoré môžeme opísať rozdelením pravdepodobnosti a číselnými charakteristikami, ktoré nazývame parametre znakov základného súboru. Parametrami znakov základného súboru sú napríklad stredná hodnota, rozptyl a pod. V reálnych situáciách nepoznáme rozdelenia pravdepodobnosti znakov na základnom súbore ani jeho parametre. Zo základného súboru urobíme náhodný výber, ktorý tvorí výberový štatistický súbor. Na základe hodnôt pozorovaného znaku na prvkoch výberového súboru robíme závery o rozdelení pravdepodobnosti znaku na základnom súbore resp. o jeho neznámych parametroch. Nedostávame presné výsledky, ale len **odhady** parametrov.

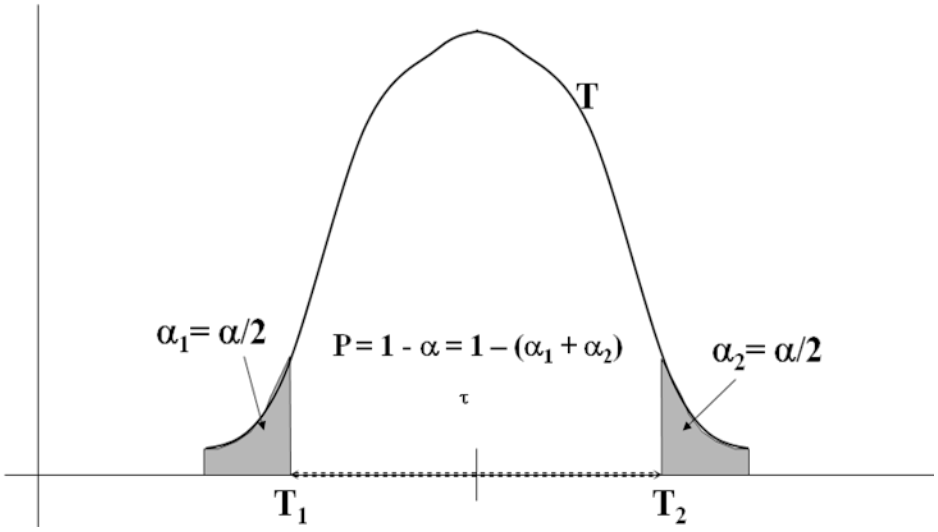
Stredná hodnota $\mu = \mathbf{E(X)}$ a rozptyl $\sigma^2 = \mathbf{D(X)}$ sú najdôležitejšími parametrami základného súboru, ktoré však nepoznáme. Ak sme pre výberový súbor vypočítali aritmetický priemer $\bar{x}$ a rozptyl $\mathbf{s^2}$ podľa **[IV.1]** a **[V.9]**, a ako výsledok sme dostali konkrétne čísla, máme tzv. **bodový odhad** štatistických parametrov, teda výberový odhad aritmetického priemeru $\bar{x}$ je bodovým odhadom strednej hodnoty μ, bodovým odhadom rozptylu σ^2 je odhad výberového rozptylu $\mathbf{s^2}$. Ak je výsledkom odhadov interval (častejší prípad), ktorý s danou pravdepodobnosťou pokrýva skutočnú hodnotu hľadaného parametra, ide o **intervalový odhad**. Šírka intervalu vyjadruje presnosť odhadu (čím užší interval, tým presnejší odhad).

O šírke rozhoduje požiadavka na spoľahlivosť odhadu. Spoľahlivosť odhadu je zvolená pravdepodobnosť

$$P = 1 - \alpha \qquad [VI.1],$$

že sa skutočná hodnota parametra **T** v intervale nachádza. Tieto odhady potrebujeme získať väčšinou s vopred zadanou presnosťou a spoľahlivosťou. Napr. máme zistiť s 95% spoľahlivosťou (P = 0,95, potom α = 0,05) a chybou odhadu menšou ako 15% (s = 0,15) priemerný vek žien odchádzajúcich vo vybranom regióne do predčasného dôchodku a pod.

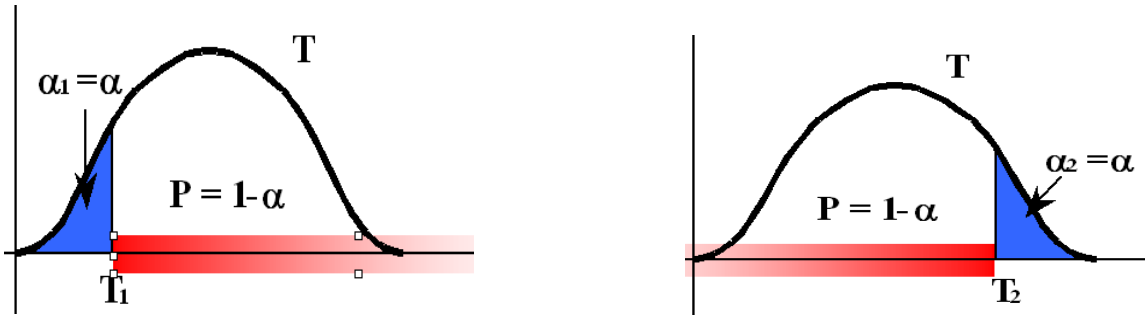
Intervalový odhad umožňuje určiť nielen jeden najlepší odhad, ale celý interval pravdepodobne možných odhadov parametra základného súboru. Interval, v ktorom sa pravdepodobne nachádza parameter základného súboru, sa nazýva **interval spoľahlivosti**.



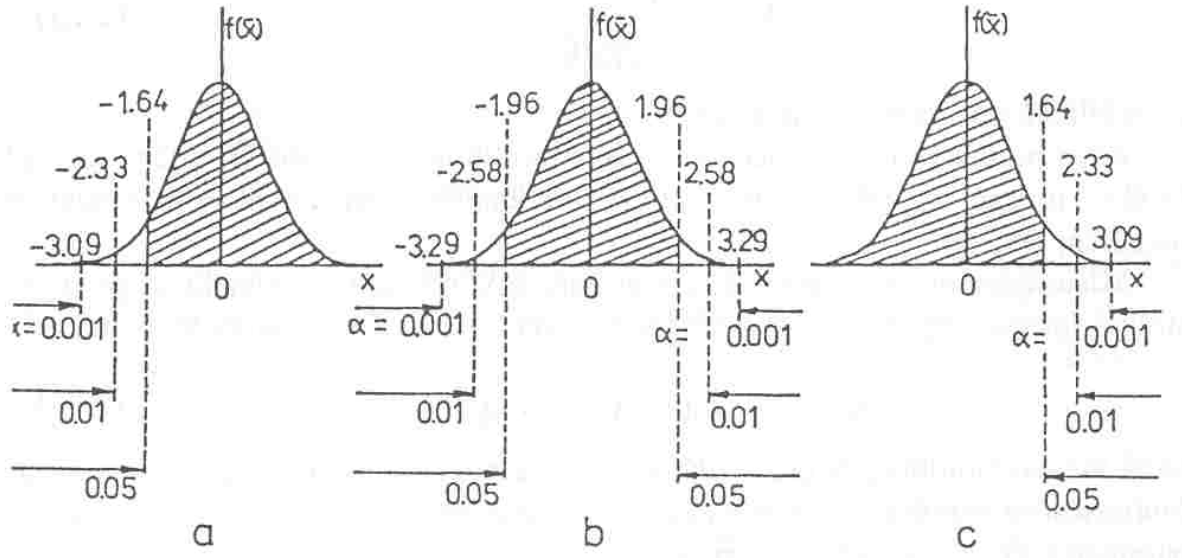
Obr. VI.1. Grafické znázornenie významu obojstranného intervalu spoľahlivosti (T_1;T_2) pri zadanej hladine významnosti α

Obojstranný interval spoľahlivosti používame napr., keď potrebujeme určiť, že s požadovanou presnosťou sa v ňom nachádza stredná hodnota na danej hladine významnosti. Napr. úlohou je určiť na hladine významnosti α = 0,05 priemerný predčasný starobný dôchodok a jeho interval spoľahlivosti s presnosťou **s** = 0,2. Parameter α zaručuje, že maximálne 5% výsledkov bude mimo intervalu spoľahlivosti a rozkladá sa na obe strany intervalu. Pre symetrické rozdelenie pravdepodobnosti je na každej strane $\alpha/2$.

Ak chceme zistiť interval spoľahlivosti odhadu, že parameter dosahuje maximálne (minimálne) nejakú hodnotu, napr. či nameraná hodnota spĺňa požiadavky uvedené v technických normách, vo vyhláškach, vykonávacích predpisoch niektorých zákonov a pod., používame jednostranné intervaly spoľahlivosti:



Obr. VI.2. Grafické znázornenie významu ľavo- a pravostranného intervalu spoľahlivosti $(T_1;\infty)$ resp. $(-\infty;T_2)$ pri zadanej hladine významnosti α



Obr.VI.3. Grafické znázornenie hodnôt hraníc intervalov (kvantilov, kritických hodnôt) normovaného normálneho rozdelenia N(0;1), ktoré odpovedajú rôznym hladinám významnosti α pre ľavostranný (a), obojstranný (b) a pravostranný interval spoľahlivosti (c) [5].

Pre praktické výpočty hľadania intervalu spoľahlivosti, ako si ukážeme o chvíľu, sú kvantily, teda hranice intervalov N(0;1) pre rôzne hladiny významnosti α tabelované. Napr. hľadáte hranice intervalu spoľahlivosti pri hladine významnosti α = 0,05. Hranice intervalu budú mať $P = \langle 1 - \alpha/2; \alpha/2\rangle = \langle 0{,}025; 0{,}975\rangle = \langle -0{,}975; 0{,}975\rangle$. Pre hodnotu P = 0,975 v tabuľkách nájdeme hodnotu kvantilu $\mathbf{u_P = 1{,}960}$ a interval spoľahlivosti bude ⟨-1,96; 1,96⟩. **α** je malá pravdepodobnosť, predstavujúca riziko chybného záveru (odhadu). Čím je menšie **α**, tým je širší interval spoľahlivosti, teda väčšia spoľahlivosť, že odhad je správny, ale tým menšia presnosť odhadu.

Hľadáme teraz **minimálny rozsah výberu**, potrebný pre zaistenie požadovanej presnosti a spoľahlivosti. Podľa charakteristík výberového súboru máme niekoľko prípadov:

1. Základný súbor má normálne rozdelenie **N(μ,σ²)**. Výberový súbor s rozsahom **n>30**, sme získali prostým náhodným výberom s vrátením vybraného prvku vždy späť do súboru.
Bodový odhad strednej hodnoty **μ** je odhad výberového aritmetického priemeru. Smerodajnú chybu odhadu aritmetického priemeru $s_{\bar{x}}$ dostaneme z výberovej smerodajnej odchýlky **s** podľa [V.10]

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad \textbf{[VI.2]}$$

Ak si označíme prípustnú chybu odhadu ***d***, ktorá je stanovená podľa povahy sledovanej úlohy a ktorú je možné akceptovať, tak je daná ako

$$|\bar{x} - \mu| < d \qquad alebo$$
$$\bar{x} - \mu < d \qquad alebo$$
$$d < \bar{x} - \mu \qquad \textbf{[VI.3]}$$

Je to trochu náročnejšie, preto sa vráťte na chvíľu do III. kapitoly, kde sme okrem iného preberali normálne rozdelenie náhodnej premennej N(μ;σ²) aj normované normálne rozdelenie N(0;1), ktoré sa dá získať z normálneho rozdelenia tak, že pre každú nameranú alebo inak získanú hodnotu **X** normálneho rozdelenia vypočítame jej odchýlku od strednej hodnoty **μ**, teda **X - μ** a výsledok vydelíme smerodajnou odchýlkou **σ**, bol to vzťah [III.14]:

$$u = \frac{X - \mu}{\sigma}$$

u je tzv. **normovaná odchýlka** a normované normálne rozdelenie nadobúda tvar [III.15]:

$$f(u) = \frac{1}{\sqrt{2\pi}} . e^{\frac{-u^2}{2}}$$

Pomocou tabuliek s hodnotami **u** sme si potom dokázali jednoducho vypočítať pravdepodobnosti, že náhodná premenná sa bude nachádzať v nejakom intervale, resp. nájsť príslušný interval. Predpokladáme (dá sa to dokonca dokázať, ale nechajme nejakú zábavu aj matematikom), že aj trochu iná štatistická veličina

$$u_p = \frac{\bar{x} - \mu}{s} . \sqrt{n} \qquad \textbf{[VI.4]}$$

má tiež rozdelenie pravdepodobnosti, ktoré sa blíži k normovanému normálnemu rozdeleniu N(0;1). Tento postup teda môžeme s dobrým svedomím použiť na výpočet intervalu spoľahlivosti pre odhad strednej hodnoty základného súboru $\mu \pm d$.

Interval spoľahlivosti odhadu strednej hodnoty μ pomocou výberového aritmetického priemeru $\bar{x}$, kde $d = u_p \cdot \frac{s}{\sqrt{n}}$ bude:

$$\bar{x} - u_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + u_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \qquad \textit{pre obojstranný interval}$$

$$\mu \leq \bar{x} + u_{\alpha} \cdot \frac{s}{\sqrt{n}} \qquad \textit{pre pravostranný interval}$$

$$\bar{x} - u_{\alpha} \cdot \frac{s}{\sqrt{n}} \leq \mu \qquad \textit{pre lavostranný interval} \qquad \textbf{[VI.5]}$$

kde u_p je tabelovaná hodnota kvantilu N(0;1) pre hladinu významnosti α, prípustná chyba odhadu je

$$d = u_p \cdot \frac{s}{\sqrt{n}} \qquad \textbf{[VI.6]}$$

kde $u_p = u_{\frac{\alpha}{2}}$ pre dvojstranný symetrický interval, $u_p = u_{\alpha}$ pre jednostranný interval.

Intervaly [VI.5] potom nazývame **100.(1-α) percentný obojstranný (ľavo, pravostranný) interval spoľahlivosti**. Ako vidíme, chyba, ktorej sa pri odhade dopúšťame je tým menšia, čím je väčší rozsah výberového súboru **n**. Zo vzťahu [VI.5] ľahko určíme minimálny rozsah výberu **n**, potrebný na zaistenie požadovanej presnosti a spoľahlivosti odhadu:

$$n = (u_p \cdot \frac{s}{d})^2 \qquad \textbf{[VI.7]}$$

2. Za rovnakých podmienok bude rozsah **výberu bez vrátenia n_{bv}** o niečo menší ako s vrátením vybraného prvku do súboru n_{sv}, takže **[VI.7]** môže poslúžiť ako dobrý odhad aj v tomto prípade.

Najlepší výber je žiadny výber, teda práca so základným súborom. Ale keďže to nejde, je potrebné si uvedomiť, ako už bolo uvedené, presnosť a spoľahlivosť idú proti sebe. Zvyšovať ich spoločne možno len zvyšovaním rozsahu výberu.

Nesľúbili sme čitateľovi, že to bude vždy ľahké. Na druhej strane, pokiaľ má trochu trpezlivosti a pracovitosti a dokáže vstrebať látku tejto kapitoly a vytvoriť si pre jej použitie v centrálnom nervovom systéme nové obvody, dostáva sa do finále popisnej štatistiky, pretože si dokáže celkom slušne kvantitatívne odhadnúť potrebný rozsah výberového súboru a deskriptívnu štatistiku zavŕšiť výpočtom intervalov spoľahlivosti odhadov sledovaných parametrov. Dajme si však radšej príklad:

Pr.VI.1.: Urobme si odhad, že v hlavnom meste s počtom obyvateľov 600 tis. potrebujú asi 2% populácie v staršom veku istú formu opatrovateľskej služby. Zavedenie opatrovania seniorov v domácom prostredí sa v mnohých smeroch ukazuje ako dostatočná až optimálna služba

sociálnej pomoci. Potrebujeme odhadnúť počet všetkých (dobrovoľných i profesionálnych) opatrovateľov, ktorí by zvládli túto úlohu, preto je nutné na hladine významnosti α=0,05 stanoviť odhad, koľkým seniorom môže priemerne jeden terénny sociálny pracovník denne poskytnúť opateru s rozptylom 1.

Riešenie: Počet seniorov, ktorým je potrebné poskytnúť opatrovateľskú službu je 2% zo 600000 obyvateľov mesta = 12000. Ako prvý nástrel odhadu sme s jedným opatrovateľom strávili v teréne 10 pracovných dní, s výsledkami náhodne pridelených služieb:

Deň	1	2	3	4	5	6	7	8	9	10
výkon	11	6	5	7	6	9	4	12	5	5

Štandardnými postupmi popisnej štatistiky sme vypočítali nasledovné parametre:

Počet (rozsah súboru) n = 10

Súčet výkonov: $\Sigma = 70$

Stredná hodnota: $\bar{x} = 7$

Smerodajná odchýlka: **s = 2,748737**

Minimálna hodnota: $x_{min} = 4$

Maximálna hodnota: $x_{max} = 12$

Rozpätie: R = 8

Pre spresnenie strednej hodnoty priemerných denných výkonov opatrovateľov potrebujeme urobiť prieskum medzi viacerými opatrovateľmi. Pre odhad počtu prvkov výberového štatistického súboru opatrovateľov sme použili nasledujúce hodnoty:

Smerodajná odchýlka **s = 2,748737**, vypočítaná z predchádzajúceho súboru podľa **[V.9]**, požadovaná presnosť **d = 1** (druhá odmocnina z rozptylu v zadaní), požadovaná hladina významnosti **α=0,05**, z toho pre dvojstranný symetrický interval **$\alpha/2$=0,025** a **1-$\alpha/2$=0,975**. V tabuľkách N(1;0) tejto hodnote prislúcha kvantil $\mathbf{u_p = 1,96}$. Potrebný výber sme vypočítali podľa posledného výrazu, ktorý sme sa naučili **[VI.7]**:

$$n = (u_p.\frac{s}{d})^2 = (1,96.\frac{2,748737}{1})^2 = 29,02542 \cong 30$$

Urobili sme náhodný výber 30 mesačných pracovných výkazov terénnych opatrovateľov, ktorým sa denné návštevy a služby prideľujú losovaním. Ich výkony sú náhodnou veličinou s normálnym rozdelením. Z mesačných výkazov sme zistili hodnoty priemerných denných výkonov, ktoré sú v tabuľke:

i	x_i	i	x_i	i	x_i
1	11	11	5	21	7
2	8	12	8	22	5
3	7	13	7	23	8
4	13	14	6	24	15
5	6	15	5	25	9
6	11	16	15	26	12
7	10	17	8	27	9
8	8	18	12	28	10
9	15	19	4	29	7
10	3	20	11	30	10

Tento súbor sme premleli cez štandardnú ***Analýzu dát*** v Exceli, tak ako v minulej kapitole, ale rozšírili sme ju ako vždy o výpočet smerodajnej odchýlky **σ** a variačného koeficientu **V(%)**:

Sloupec1	
Stř. hodnota	9
Chyba stř. hodnoty	0,575356
Medián	8,5
Modus	8
Směr. odchylka	3,151354
Rozptyl výběru	9,931034
Špičatost	-0,40055
Šikmost	0,304574
Rozpätie	12
Minimum	3
Maximum	15
Součet	270
Počet	30
σ	3,098387
V(%)	34,42652

Dvojstranný interval spoľahlivosti dostaneme pomocou vzťahu **[VI. 5]** nasledovne:

$$\bar{x} - u_{\frac{\alpha}{2}} . \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + u_{\frac{\alpha}{2}} . \frac{s}{\sqrt{n}} = 9 \pm 1{,}96 . \frac{3{,}151354}{\sqrt{30}} = 9 \pm 1{,}127698 \cong 9 \pm 1$$

Priemerný denný výkon opatrovateľov je 9 so smerodajnou odchýlkou 3 a intervalom spoľahlivosti na hladine významnosti α=0,05: ⟨8;10⟩. Za týchto podmienok by bolo potrebné pre odkázaných seniorov zabezpečiť priemerne asi 1333 opatrovateľov, pričom s 95% pravdepodobnosťou by potrebné služby malo zabezpečiť 1200 až 1500 opatrovateľov.

Pre výpočet intervalu spoľahlivosti máme k dispozícii funkciu v EXCELI; syntax: **CONFIDENCE(alpha;standard_dev;size)**, kde dosadzujeme za **alpha** kritickú hodnotu, napr. **α=0,05**; za **standard_dev** hodnotu smerodajnej odchýlky **s**; za **size** rozsah, v našom prípade 30.

Ak je prípustná chyba **d** zadaná percentuálne, potom vo vzťahu [VI.7] bude namiesto smerodajnej odchýlky **s** figurovať variačný koeficient **V**:

$$n = (u_p.\frac{V}{d})^2 \qquad [VI.8]$$

Pr.VI.2.: V domove sociálnych služieb sú problémy s vodou. Zdroj, na ktorý je pripojený má výdatnosť 5 litrov/min. Priemerná denná spotreba vody na hlavu v celej populácii sa odhaduje na 90 až 140 litrov, v domove v dôsledku zvýšenej hygieny je to asi 150 ± 30 litrov. Je potrebné odhadnúť priemernú spotrebu vody na hlavu s chybou 15% a zistiť na hladine významnosti 0,1, pre koľko klientov bude zdroj vody postačujúci, alebo či je potrebné sa zamerať na dodatočné zdroje, pretože efektívna prevádzka zariadenia je pri minimálnom množstve 50 klientov a 10% sa musí pripočítať na spotrebu zamestnancov.

Riešenie:

Počet klientov N, t.j základný súbor môže byť teoreticky neobmedzený, resp. aj niekoľko stoviek. Vodný zdroj s výdatnosťou 5 l/min dáva denne 7200 litrov vody. Je potrebné zistiť priemernú dennú spotrebu vody na klienta s chybou max.15% a kritickou hodnotou **α = 0,1**; keď sa uvádza, že denná spotreba na hlavu v takýchto zariadenia býva asi 150 ± 60 litrov, teda **μ = 150, σ=60**. Vyberieme náhodne **n** klientov, pri ktorých budeme nejaký čas sledovať priemernú spotrebu. Prípustnú odchýlku máme v %: **d = 15%**. Použijeme vzťah [VI.8], nato si vopred vypočítame variačný koeficient V(%) podľa vzťahu [V.11]

$$V = \frac{\sigma}{\bar{x}}.100 = \frac{60}{150} * 100 = 40\,\%$$

Na hladine významnosti **α = 0,1** určíme aj pravostranný interval spoľahlivosti priemernej spotreby vody. Pre 1-α = 0,9 máme hodnotu kvantilu N(0,1) z tabuliek: $\mathbf{u_\alpha = 1,282}$. Potom z [VI.8] pre rozsah výberu dostaneme:

$$n = (u_p.\frac{V}{d})^2 = (1,282.\frac{40}{15})^2 = 11,687 \cong 12$$

Pre 12 klientov dostaneme pozorovaním priemernú dennú spotrebu vody:

I	1	2	3	4	5	6	7	8	9	10	11	12
n_i[l/deň]	105	94	121	153	154	86	141	113	107	121	122	135

Tento výberový súbor premelieme excelovskou štatistickou analýzou a dostaneme výsledok:

Sloupec1	
Stř. hodnota	121
Chyba stř. hodnoty	6,264377
Medián	121
Modus	121
Směr. odchylka	21,70044
Rozptyl výběru	470,9091
Špičatost	-0,78381
Šikmost	0,101843
rozpätie	68
Minimum	86
Maximum	154
Součet	1452
Počet	12
σ	22,66537
V(%)	18,73171

Pre pravostranný interval spoľahlivosti by sme použili zo vzťahu [VI.5]

$\mu \leq \bar{x} + u_\alpha . \frac{s}{\sqrt{n}} = 121 + 1{,}282 . \frac{21{,}7}{\sqrt{12}} = 121 + 8,$ a pravostranný interval spoľahlivosti pre strednú hodnotu 121 pri α=0,1 by bol **(-∞; 129⟩**.

Pozor! Je tu však jeden problém: Pre malé súbory s n < 30 je použitie štatistiky u_p a normovaného normálneho rozdelenia N(0;1) nedostatočne presné. Vtedy sa používa iné výberové rozdelenie, ktoré sme si zatiaľ len spomenuli, **Studentovo t-rozdelenie s (n-1) stupňami voľnosti**, ktoré je tiež symetrické a má normovanú výberovú charakteristiku

$$t = \frac{\bar{x} - \mu}{s} . \sqrt{n} \qquad \textbf{[VI.9]}$$

Zaviedol ho britský matematik a chemik W.S.Gosset (1876-1937), keď popíjal a hodnotil kvalitu produktov dublinského pivovaru Guinness. Samozrejme, že po dobrom a ťažkom tmavom pive musel dosiahnuť skvelé matematické výsledky, ale z dôvodov zmluvy o utajení ich nesmel publikovať, tak použil pseudonym Student. Aby odľahčil svedomiu, tak prehovoril majiteľov prosperujúceho pivovaru na kompletnú opravu dublinskej katedrály sv. Patricka.

Interval spoľahlivosti (dvojstranný, podobne aj jednostranný) je potom daný vzťahom

$$\bar{x} - t_{n-1,\frac{\alpha}{2}} . \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1,\frac{\alpha}{2}} . \frac{s}{\sqrt{n}} \qquad \textbf{[VI.10]}$$

kde kritické hodnoty $t_{n-1,p}$ pre dané α a počet stupňov voľnosti **(n-1)** sú dostupné taktiež v štatistických tabuľkách.

Pre náš príklad bude hľadanie pravostranného intervalu spoľahlivosti nasledovné:

$$\mu \leq \bar{x} + t_{n-1,\alpha} \cdot \frac{s}{\sqrt{n}}$$

Pre $\alpha = 0{,}1$ je $P(1-\alpha) = 0{,}9$ a $t_{n-1,\alpha} = t_{11;0.90} = 1{,}363$

Dostaneme interval $\mu \leq \bar{x} + t_{n-1,\alpha} \cdot \frac{s}{\sqrt{n}} = 121 + 1{,}363 . \frac{21{,}7}{\sqrt{12}} = 121 + 8{,}54 \cong 130$ a pravostranný interval spoľahlivosti pre strednú hodnotu 121 pri α=0,1 by bol (-∞; 130⟩, teda o niečo väčší.

Konečný výpočet:

Máme výberový odhad priemernej hodnoty dennej spotreby vody na hlavu $\bar{x}$**=121** s odhadom jej výberovej smerodajnej odchýlky $s_{\bar{x}}$**= 6,3**. Pre minimálny počet 50 klientov je priemerná denná spotreba vody 121x50 = 6050 litrov s odchýlkou ± 50x6,3 = ± 315. Vypočítané hodnoty musíme zvýšiť ešte o 10%, aby sme zarátali spotrebu zamestnancov:

Odhad priemernej dennej spotreby vody v zariadení sociálnych služieb 6050x1,1 = 6655, s odhadom smerodajnej odchýlky priemeru ± 315x1,1 = ±347: $\bar{x}$ **=6655 ±347**.

Pre maximálnu spotrebu vody v zariadení s 50 klientmi použijeme hornú hranicu intervalu spoľahlivosti t.j. **130** l/deň/klienta. Pre 50 klientov to bude 50x130 = 6500 litrov za deň a s 10% pre zamestnancov 6500x1,1 = **7150 litrov/deň**. Pri kritickej hodnote **α = 0,1** máme 90% pravdepodobnosť, že denná spotreba vody zariadenia sociálnych služieb s 50 klientmi nepresiahne hranicu 7150 litrov denne, pričom výdatnosť zdroja vody je 7200 litrov denne. Vodohospodárov by sme mali presvedčiť, že zdroj je dostačujúci, pokiaľ nezvýšime počet klientov.

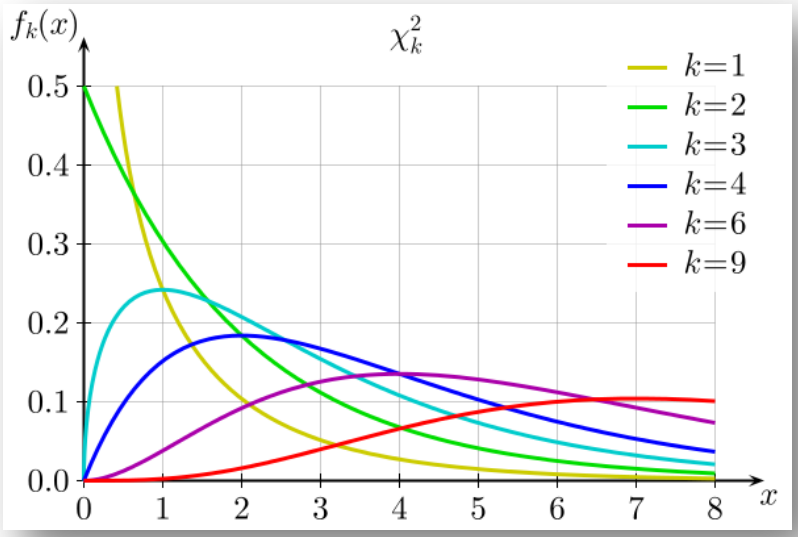
Častou úlohou bude, aký počet prvkov výberového súboru zo základného súboru s normálnym rozdelením **n** je potrebný, aby chyba pri odhade nepresiahla nejakú zadanú hodnotu **d**; pritom prvý orientačný náhodný výber dáva výberové charakteristiky $\bar{x} \pm s$.

Pr.VI.3.: Sledujete výdavky rodín s malými deťmi na hry podobné hazardným (karty, ruleta a pod.), ktoré by mohli v mladej generácii stimulovať záujem o skutočný hazard a prípadne gamblerstvo. Predpokladáme, že výdavky podliehajú normálnemu rozdeleniu. Ak pri orientačnom náhodnom výbere získate strednú hodnotu mesačných výdavkov **45 ± 5** (v €), aký početný musí byť náhodný výber vášho skúmania, aby chyba pri odhade strednej hodnoty nebola väčšia ako 40%, t.j. **d = 0,4**? Vo vzťahu **[VI.7]** pre kritickú hodnotu **α= 0,05** máte už všetko k dispozícii: štatistika $u_{\frac{\alpha}{2}}(0{,}975) = 1{,}960$ z tabuliek:

$$n = (u_p.\frac{s}{d})^2 = (1,96.\frac{5}{0,4})^2 = 600$$

Potrebný rozsah vášho výberu je 600.

3.Prípad: Pri zisťovaní **intervalu spoľahlivosti pre smerodajnú odchýlku σ** resp. rozptyl **σ^2** použijeme ďalšie výberové rozdelenie: **χ^2- rozdelenie** (čítaj chí kvadrát rozdelenie) **s k=(n-1) stupňami voľnosti**. **χ^2**-rozdelenie je asymetrické ako ukazuje obr. VI.4. [6]:



Obr. VI.4. Pravdepodobnostná funkcia χ^2- rozdelenia náhodnej premennej pre k stupňov voľnosti, k = n-1, pri rôznom k.

Pre odhad rozptylu základného súboru s normálnym rozdelením **σ^2** možno použiť rozptyl pre výberový súbor s^2 podľa **[V.9]**:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \textbf{[VI.11]}$$

Zo základného súboru sa náhodne vyberie **n** prvkov a vypočíta sa odhad výberového rozptylu **s^2**. Rozptyl **s^2** nemá symetrické rozdelenie, ale rozdelenie **χ^2**. (1-α) %-ný interval spoľahlivosti pre odhad rozptylu základného súboru je

a) obojstranný interval spoľahlivosti pre disperziu **σ^2**

$$\left\langle \frac{(n-1).s^2}{\chi^2_{\frac{\alpha}{2},k}} ; \frac{(n-1).s^2}{\chi^2_{1-\frac{\alpha}{2},k}} \right\rangle \qquad \textbf{[VI.12]}$$

b) ľavostranný interval spoľahlivosti pre disperziu σ^2

$$\left\langle \frac{(n-1).s^2}{\chi^2_{\frac{\alpha}{2},k}}; \infty \right\rangle \qquad [VI.13]$$

c) pravostranný interval spoľahlivosti pre disperziu σ^2

$$\left\langle 0; \frac{(n-1).s^2}{\chi^2_{1-\frac{\alpha}{2},k}} \right\rangle \qquad [VI.14]$$

Pre určenie intervalu spoľahlivosti vypočítame z **[VI.11]**

$$s^2(n-1) = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

a hodnoty $\chi^2_{p,k}$ sú pre **(1-α)** a stupne voľnosti k=n-1 tabelované.

Pr.VI.4.: Pri sledovaní dlhodobej nezamestnanosti sa malo zistiť 95%-ným intervalovým odhadom rozptyl doby, počas ktorej nezamestnaní napriek vlastným aktivitám i aktivitám úradu práce nemohli zamestnanie nájsť. Bolo náhodne vybraných 20 dlhodobo nezamestnaných s dĺžkou nezamestnanosti v mesiacoch:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
n_i	12	18	13	18	13	16	11	17	20	15	14	14	15	16	17	15	13	13	16	14

Predpokladajme, že doba nezamestnanosti je náhodná premenná s normálnym rozdelením $N(\mu;\sigma^2)$. Spočítame výberový aritmetický priemer $\bar{x} = 15$, odchýlky $|\Delta|$ a ich kvadráty Δ^2

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
n_i	12	18	13	18	13	16	11	17	20	15	14	14	15	16	17	15	13	13	16	14
$\|\Delta\|$	3	3	2	3	2	1	4	2	5	0	1	1	0	1	2	0	2	2	1	1
Δ^2	9	9	4	9	4	1	16	4	25	0	1	1	0	1	4	0	4	4	1	1

$s^2(n-1) = \sum_{i=1}^{n} (x_i - \bar{x})^2 = 98$

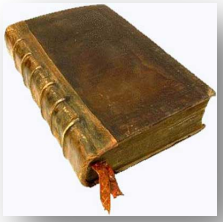
1-α = 0,95; teda **α=0,05** a **α/2=0,025; 1-α/2 = 0,975**; **k = n-1 = 19**. Potom z tabuliek $\chi^2_{p,k}$ – rozdelenia pre (1-p)

$$\chi^2_{1-\frac{\alpha}{2},k} = \chi^2_{0,975;19} = 8,91$$

$$\chi^2_{\frac{\alpha}{2},k} = \chi^2_{0,025;19} = 32,9$$

Vzťah **[VI.12]** dáva interval

$$\left\langle \frac{(n-1).s^2}{\chi^2_{\frac{\alpha}{2},k}}; \frac{(n-1).s^2}{\chi^2_{1-\frac{\alpha}{2},k}} \right\rangle = \left\langle \frac{98}{32,9}; \frac{98}{8,91} \right\rangle \cong \langle 3;11 \rangle$$



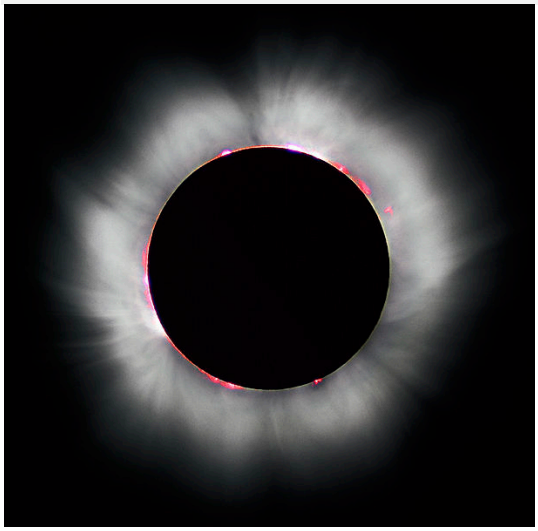
Rozptyl doby dlhodobej nezamestnanosti pokrýva s 95%-nou pravdepodobnosťou interval v mesiacoch^2 ⟨3;11⟩.

χ^2 vyzeral najprv dosť exoticky a odstrašujúco, so zaujímavým „múdrym“ názvom, ale videli ste, ako ľahko sa s ním pomocou tabuliek v prípade hľadania intervalu pracuje. Budeme ho ešte potrebovať aj v iných úlohách induktívnej štatistiky, keď príde ich čas.

Pr.VI.5.: Keď cisár Yao, najvýznamnejší potomok zakladateľa civilizácie Číny *Žltého cisára*, prechádzal krajinou, videl ako strážnici spútavajú muža.

„Čo si urobil?“ – spýtal sa ho.

„Ukradol som jedlo, lebo po veľkom suchu zomierame s rodinou hladom,“ – zúfalo odvetil nešťastník. Cisár zostúpil z voza k nemu a rozkázal, aby ho napriek protestom tiež spútali, pretože ako vládca krajiny je tiež zodpovedný za to, že ľudia pri nepriazni osudu z hladu kradnú. Celé blízke i ďaleké okolie sa prišlo podívať, ako tam spútaný cisár stojí, mnohí boli pohnutí k slzám, k vyznaniam vlastných neprávostí a ochotní prijať spravodlivý trest. V priebehu vlády tohto zakladateľa dynastie Xia sa ľuďom žilo dobre. V krajine vládol pokoj a spravodlivosť, bol však nezmieriteľný voči neporiadnikom a flákačom. Keď mu posielali príbuzných, aby im pridelil nejaký úrad, ako napríklad dvoch vzdialených prasynovcov Hi a Ho, nebol k nim príliš zhovievavý. Keďže sa samoľúbo chválili svojimi vedomosťami z matematiky, astrológie a niečoho, čo by sa dalo nazvať štatistikou, poveril ich pri slávnostnom ceremoniály dvora *strážením štyroch svetových strán*. K tomu patrilo vopred upozorniť cisára na blížiace sa zatmenie slnka, pretože jeho nepriatelia šírili správy, že nebeské sily sú s jeho vládou nespokojné a mohli to zneužiť.



Obr.VI.5: Zatmenie Slnka, Francia 1999 [8]

Niekoľkokrát synovcov upozornil na dôležitosť ich úradu a potrebnú precíznosť služby. V prípade, že by s nimi nebol spokojný, musel pripustiť exemplárne tresty. Na rodinu mal dokonca prísnejšie kritéria, pretože podľa rodiny mohli ľudia posudzovať celú vládu aj panovníka, ale Hi aj Ho to brali len ako vladárske klišé. Už mali možnosť ochutnať skvelé vína podhorskej oblasti nad meandrami Žltej rieky, vynikajúce likéry dvorných dodávateľov a vedeli, že v priľahlých štvrtiach *Zakázaného mesta* sa dajú bez problémov vyhľadať najkrajšie

konkubíny celej krajiny. Peniaze im teraz, keď sú takými dôležitými úradníkmi, nebudú chýbať a navyše môžu predávať astrologické horoskopy, o ktoré je veľký záujem, len sa musia naučiť pri predaji týchto hlúpostí úplne vážne tváriť. Kto by sa v takom krásnom svete zaoberal nejakými nudnými pozorovaniami oblohy alebo zbytočne počítal!

Aby zahrali nejakú aktivitu, dali si pisármi priniesť záznamy o doterajších zatmeniach slnka a vybrali si posledných 12 zápisov. Do tabuľky si zaznamenali počet dní, ktoré uplynuli od posledného zatmenia a medzi zatmeniami podľa cyklu *Saros*:

i	x_i [dni]
1	19740
2	19775
3	19776
4	19738
5	19742
6	19769
7	19743
8	19773
9	19768
10	19747
11	19745
12	19746
$\bar{x}_{11}$	**19756**
$\bar{x}_{11}-x_{12}$	**10**

Aj keď sa im už nechcelo príliš počítať, pod tabuľku si urobili ešte aritmetický priemer počtov dní prvých 11 kompletných medziobdobí medzi zatmeniami $\bar{x}_{11}$ a od tejto hodnoty odčítali počet dní, ktoré ubehlo od posledného zatmenia až do dnešného dňa x_{12}. Týmto jednoduchým spôsobom im vyšlo, že najbližšie zatmenie slnka by malo byť najskôr o 10 dní. Ťapli si bujaro pravými rukami na znak, akí sú fantastickí; a usúdili, že s takou somarinou nepobežia hneď za cisárom. Mohol by ich ešte poveriť ďalšou úlohou a vonku na nich čakajú neobyčajné zážitky. Zavolali služobníctvo a dali pripraviť nosidlá.

Cisár Yao vládol múdro a prezieravo. Neponechával dôležité záležitosti jednostranne len na úradníkov, ktorí nepôsobili úplne dôveryhodným dojmom. Pomyslel na svojich učiteľov, troch starých mudrcov z pustovní v prekrásnej *Tiesňave troch brán*, ktorí sa vzápätí ako vždy ohlásili v jeho pracovni. Predložil im na posúdenie problém, ba aj to, čo počítali jeho prasynovcovia.

Múdri mali schopnosti nahliadnuť do podstaty vecí a problémov a vidieť javy aj cez priepasť času do minulosti i do ďalekej budúcnosti. Keď zbadali dáta, zbledli a chvíľu sa chveli, ale ihneď doplnili tabuľku mládencov o hodnoty odchýlok a ich druhých mocnín, druhé mocniny odchýlok sčítali. Cisára Yao mali radi, bol to ich vydarený žiak, preto sa bez slov dohodli, že pre neho dôležité výpočty, hlavne interval spoľahlivosti rozptylu, odhadnú na vysokej 95% úrovni:

i	x_i	$\|\Delta\|$	Δ^2
1	19740	16	256
2	19775	19	361
3	19776	20	400
4	19738	18	324
5	19742	14	196
6	19769	13	169
7	19743	13	169
8	19773	17	289
9	19768	12	144
10	19747	9	81
11	19745	11	121
12	19746	**spolu**	2510

Čas bol proti cisárovi, preto zjednotili svoje mysle v hlbokej meditácii, aby sa im v duchu z diaľok budúcich vekov, vynorili dve čísla, ktorými potrebovali súčet štvorcov odchýlok vydeliť, aby dostali interval spoľahlivosti rozptylu:

$$\chi^2_{1-\frac{\alpha}{2};k} = \chi^2_{0,975;10} = 3,2$$

$$\chi^2_{\frac{\alpha}{2};k} = \chi^2_{0,025;10} = 20,5$$

Interval spoľahlivosti rozptylu σ^2 dostali teda ľahko

$$\left\langle \frac{(n-1).s^2}{\chi^2_{\frac{\alpha}{2};k}}; \frac{(n-1).s^2}{\chi^2_{1-\frac{\alpha}{2};k}} \right\rangle = \left\langle \frac{2510}{20,5}; \frac{2510}{3,2} \right\rangle \cong \langle 122; 784 \rangle$$

Interval spoľahlivosti smerodajnej odchýlky σ rozptylu dostali ako odmocninu hraníc intervalu pre rozptyl:

$$\left\langle \bar{x}_{11} - \sqrt{\frac{(n-1).s^2}{\chi^2_{\frac{\alpha}{2};k}}}; \bar{x}_{11} + \sqrt{\frac{(n-1).s^2}{\chi^2_{1-\frac{\alpha}{2};k}}} \right\rangle = \left\langle 19756 - \sqrt{\frac{2510}{20,5}}; 19756 + \sqrt{\frac{2510}{3,2}} \right\rangle$$

$$\cong \langle 19756 - 11; 19756 + 28 \rangle$$

Aj trom *múdrym* aj cisárovi bolo jasné, že zatmenie mohlo byť už včera a že s veľkou pravdepodobnosťou, ba takmer istotou, nastane každú chvíľu v priebehu budúcich 27 dní a zostali v mihotavom šere niekoľkých lampiónov zamĺknuto sedieť. Ticho noci prehlušovalo

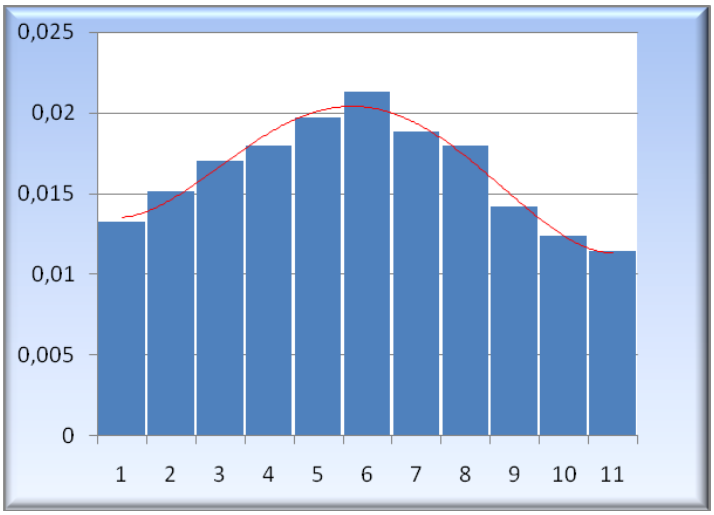
len silné kvákanie žiab v jazierkach palácových záhrad a nepokoj vtáctva, akoby príroda tušila blízke udalosti vesmírneho rozsahu.

Skôr zo zvyku, každú prácu urobiť poriadne a dotiahnuť do konca, spracovali pre prvých 11 hodnôt súboru dní medzi zatmeniami, popisnú štatistiku:

Sloupec1	
Stř. hodnota $\bar{x}$	19756
s($\bar{x}$)	4,776838
Medián	19747
Modus	
Směr. odchylka	15,84298
Rozptyl výběru	251
Špičatost	2,182537
Šikmost	0,208018
rozpätie	38
Minimum	19738
Maximum	19776
Součet	217316
Počet	11

Aby sa ubezpečili o „normalite“ súboru, teda, že pracovali korektne, zobrali odhad smerodajnej odchýlky (je to náš vzťah **[VI.11]** ale aj hodnota v predchádzajúcej tabuľke: s ≅ 16), potom hodnoty usporiadali podľa veľkosti. Následne každej hodnote priradili hodnotu pravdepodobnosti normálneho rozdelenia pomocou excelovskej funkcie **NORMDIST (x, mean, stand-dev, 0)**, kde **x** je postupne hodnota zo stĺpca usporiadaných hodnôt, **mean** je priemer, **stand_dev** je odhad smerodajnej odchýlky, v tomto prípade 16, a nula je nula. Dostali tabuľku a k tomu si zašpicateným uhlíkom, akvamarínom a rumelkou na vlastnoručne vyrobený papier namaľovali graf:

i	x_i	normdist
4	19738	0,013242
1	19740	0,015123
5	19742	0,017003
7	19743	0,017924
11	19745	0,019686
10	19747	0,021285
9	19768	0,018821
6	19769	0,017924
8	19773	0,014179
2	19775	0,012319
3	19776	0,011416



Pri nedostatku času na náročnejšie testy, im dal obrázok určitú orientáciu, že súbor dostatočnú „normalitu“ má. S hlbokou poklonou popísaný a pomaľovaný jemný hodvábny papier so svojou pečaťou podali cisárovi a rozplynuli sa.

Cisár Yao bol aj rozvážny. Svoju rolu v tom však zohrala aj skutočnosť, že nechcel počúvať na rodinných oslavách sústavné nariekavé škriekanie svojich pratiet. Preto sa rozhodol, že dá Hi a Ho ešte príležitosť svoju prácu odviesť poriadne a predložiť mu výsledky. Počká ešte jeden deň.

Hi a Ho sa veľmi dobre zabávali. Dievčatá boli krásne, ale najlepšie sa cítili v hostincoch pre pocestných, kde bola dobrá hudba a kde sa vždy našlo aj zopár zábavných chlapíkov, ktorí rozprávali veselé historky, ako používali svoju neandertálsku štatistiku. Mohli sa kadečomu priučiť. Ani ich nenapadlo, aby sa vrátili do paláca a tak zostali vonku ešte ďalšiu noc.

Keď sa zobudili napoludnie tretieho dňa, napriek silnej opici hneď vytušili, že niečo nie je v poriadku, všade bolo počuť krik, a akoby sa zvečerievalo. Vojaci pobehovali a pátrali po sprisahancoch, ktorí chceli využiť začínajúce zatmenie slnka a zvrhnúť cisára. Ale to už prichádzali zamračení palácoví drábi z cisárskej gardy a Hi a Ho odrazu vedeli, že hlava ich už dlho bolieť nebude [7].

4. prípad: **Odhad relatívnej početnosti** základného súboru **π**. Ak je sledovaný štatistický znak v základnom súbore rozdelený alternatívne (áno – nie; niekoho vo voľbách volím alebo nevolím; muž alebo žena; zamestnaný alebo nezamestnaný a pod.) môžeme pomocou výberového súboru intervalovo odhadnúť podiel, teda relatívnu početnosť jednotiek so sledovanou vlastnosťou v základnom súbore. Podobne ako pri binomickom rozdelení náhodnej premennej, pravdepodobnosť získame ako podiel počtu „priaznivých“ javov **m** ku všetkým možným **n**: $p = \frac{m}{n}$

Niekoľko poznámok: Pre dostatočne veľký rozsah súboru **n** sa binomické rozdelenie blíži k našej spokojnosti k normálnemu. Táto aproximácia (nahradenie binomického rozdelenia normálnym) je dobrá, keď je relatívna početnosť nie príliš vzdialená hodnote 0,5: $\pi \rightarrow \mathbf{0,5}$. Niekedy sa používa podmienka na požadovaný rozsah výberu v tomto prípade: $n > \frac{9}{\pi(1-\pi)}$.

Štatistika na získanie intervalu spoľahlivosti odhadu relatívnej početnosti je

$$u = \frac{p - \pi}{\sqrt{\pi(1-\pi)}} . \sqrt{n} \qquad [\text{VI.15}]$$

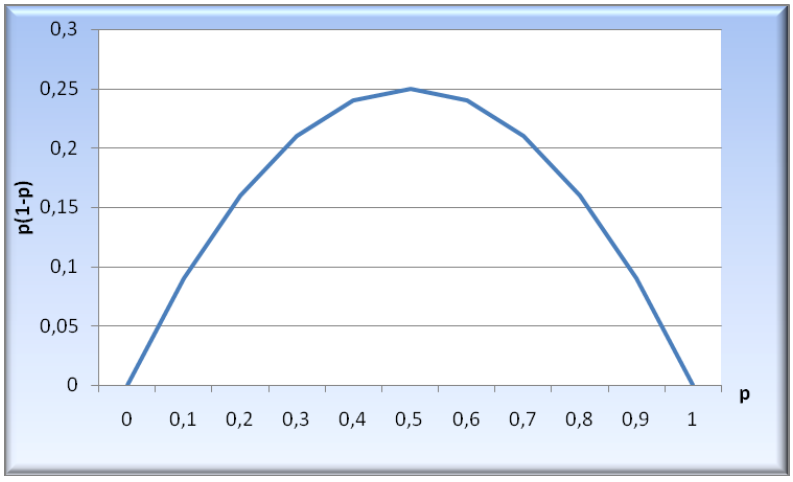
Interval spoľahlivosti odhadu relatívnej početnosti základného súboru bude na hladine významnosti **1-** $\frac{\alpha}{2}$ (resp. 1-α pri jednostranných intervaloch):

$$\langle p - u_{1-\frac{\alpha}{2}} . \sqrt{\frac{\pi(1-\pi)}{n}} ; p + u_{1-\frac{\alpha}{2}} . \sqrt{\frac{\pi(1-\pi)}{n}} \rangle \qquad [\text{VI.16}]$$

kde **u** sú tabelované kvantily N(0,1) pre hladinu významnosti **α**. Pre stanovenie minimálneho potrebného rozsahu výberu (nezávislý výber tak, aby sa neprekročila požadovaná chyba ***d***) platí:

$$n = u_p^2 . \frac{\pi(1-\pi)}{d^2} \qquad [\text{VI.17}]$$

Relatívnu početnosť základného súboru **π** nepoznáme, ale keď si pre jednotlivé pravdepodobnosti **p** urobíme súčiny **p.(1-p)**, vidíme z obr. VI.6, že tento súčin je maximálny pri hodnote **p = 0,5**, preto keď vo vzťahu **[VI.17]** použijeme hodnotu p = 0,5; tak máme pri danej hladine významnosti taký rozsah výberu **n**, ktorý nám zabezpečí nielen požadovanú, ale aj lepšiu presnosť ***d***.



Obr.VI.6: Závislosť súčinu p.(1-p) na pravdepodobnosti p∈⟨0;1⟩.

Pr.VI.6.: Pri sledovaní xenofóbnych postojov obyvateľstva potrebujeme zistiť interval spoľahlivosti odhadu relatívnej početnosti tých, ktorí na otázku, či majú radi cudzincov odpovedajú záporne. Tento odhad chceme realizovať na hladine významnosti α =0,1 tak, aby chyba odhadu bola menšia ako 5% (d=0,05). Aký má byť rozsah výberového súboru **n**?

α = 0,1 $\rightarrow \alpha/2$ = 0,05 a 1 - $\alpha/2$ = 0,95. Z tabuliek N(0;1) zistíme $u_{1-\alpha/2}$ = 1,645. Potom z **[VI.17]** dostaneme

$$n = u_p^2 . \frac{\pi(1-\pi)}{d^2} = 1,645^2 . \frac{0,5.0,5}{0,05^2} \cong 270$$

Pr.VI.7.: V istej krajine vstupujú do parlamentných volieb dve politické strany, RAB (Radosť a Budúcnosť) a ZAP (Zodpovednosť a pracovitosť). Populistická strana RAB, ľudovo nazvaná rabiáti, sľubuje vysoké sociálne istoty, ktoré zabezpečí štát, „jánošíkovskú“ ekonomiku, chlieb a hry, a obranu pred hegemonistickými snahami svojich susedov. Zapáci sľubujú vytvorenie podmienok na tvorivosť, inováciu, neustále vzdelávanie, na spoločnosť, kde každý občan aktívne prevezme zodpovednosť za vlastný život a ochranu dostanú len znevýhodnení a taktiež chce presadiť kultúrne obohacovanie sa vzťahmi so susedmi. Predvolebný prieskum má zistiť preferencie strany RAB v populácii, pričom sa orientačne ukazuje, že by mohla získať až 70% hlasov zúčastnených voličov. Prieskumná agentúra si určila interval, v ktorom sa výsledok bude pohybovať medzi 60 a 80%. Na hladine významnosti α = 0,05 urobila náhodným výberom prieskum vo vzorke 400 respondentov tak, aby chyba odhadu neprekročila 5%.

Pokiaľ je očakávaná hodnota populačnej relatívnej početnosti v nejakom intervale, vo výpočte potrebného rozsahu súboru sa vždy berie hranica, bližšia k 0,5. V tomto prípade bude π = 0,6. Ďalej α = 0,05 $\rightarrow \alpha/2$ = 0,025 a 1 - $\alpha/2$ = 0,975. Z tabuliek N(0;1) zistíme $u_{1-\alpha/2}$ = 1,96. d = 0,05. Potom z **[VI.17]** agentúra dostala

$$n = u_p^2 . \frac{\pi(1-\pi)}{d^2} = 1,96^2 . \frac{0,6.0,4}{0,05^2} \cong 370$$

Predvolebný prieskum agentúra realizovala na vzorke 400 respondentov a zistila výsledok aj s intervalom spoľahlivosti odhadu π:

Stranu RAB by volilo 76,2% opýtaných respondentov, ktorí sa chystajú ísť voliť, s 95%-ným intervalom spoľahlivosti v zmysle **[VI.16]**:

$$\left\langle p - u_{1-\frac{\alpha}{2}} . \sqrt{\frac{\pi(1-\pi)}{n}}; p + u_{1-\frac{\alpha}{2}} . \sqrt{\frac{\pi(1-\pi)}{n}} \right\rangle = \langle 0,762 \mp 1,96 . \sqrt{\frac{0,24}{400}} \rangle \cong \langle 0,714; 0,81 \rangle$$

teda medzi 71,4 až 81,0%.

Pr.VI.8.: Mierne okyselený vodný roztok náhrady cukru, bieleho a červeného umelého farbiva a želatíny dala firma na trh ako nový ovocný prírodný jogurt PROSOLIS s neobyčajnými výživovými a zdravotnými vlastnosťami, odporúčaný hlavne pre moderné ženy a pre deti. Reklamná agentúra musí zabezpečiť jeho predajnosť v krajine a odhadnúť jej úroveň. Po 30 dňovej intenzívnej kampani si reklamná agentúra urobila prieskum na 95% hladine významnosti s 2% chybou, aký podiel žien si bude denne kupovať minimálne 1 jogurt PROSOLIS. Odhad veľkosti základného súboru kúpyschopných žien je 2 000 000.

Vstupné údaje pre zistenie intervalu spoľahlivosti odhadu relatívnej početnosti žien, kupujúcich denne jogurt PROSOLIS:

Prípustná chyba odhadu: 2% $\rightarrow d = 0{,}02$

Hladina významnosti odhadu: $\alpha = 0{,}05 \rightarrow \alpha/2 = 0{,}025$ a $1 - \alpha/2 = 0{,}975$. Z tabuliek N(0;1) zistíme $u_{1-\alpha/2} = 1{,}96$.

Veľkosť základného súboru: N = 2 000 000

Potrebný rozsah výberového súboru pri uvedených podmienkach bol vypočítaný podľa **[VI.17]**, pre tento výpočet bola použitá hodnota $\pi = 0{,}5$:

$$n = u_p^2 . \frac{\pi(1-\pi)}{d^2} = 1{,}96^2 . \frac{0{,}5 . 0{,}5}{0{,}02^2} \cong 2400$$

Na 2400 respondentiek náhodného nezávislého výberu sa realizoval prieskum, pritom sa zistilo, že 65% z nich si rozhodne bude jogurt PROSOLIS kupovať. Interval spoľahlivosti odhadu relatívnej početnosti základného súboru je podľa **[VI.16]**:

$$\langle p - u_{1-\frac{\alpha}{2}} . \sqrt{\frac{\pi(1-\pi)}{n}} ; p + u_{1-\frac{\alpha}{2}} . \sqrt{\frac{\pi(1-\pi)}{n}} \rangle = \langle 0{,}65 \mp 1{,}96 . \sqrt{\frac{0{,}25}{2400}} \rangle \cong \langle 0{,}63;\ 0{,}67 \rangle$$

Ak hranice intervalu vynásobíme veľkosťou základného súboru, dostaneme s 95% pravdepodobnosťou a chybou do 2% odhad absolútnej dennej predajnosti nového propagovaného jogurtu PROSOLIS v krajine v intervale, čo je významná informácia pre výrobcu:

$$\langle 0{,}63 . 2000000;\ 0{,}67 . 2000000 \rangle = \langle 1\,260000; 1\,340000 \rangle$$

Pr.VI.9.: Máte pred sebou dôležité rozhodnutie o smerovaní vášho náročného kvantitatívneho výskumu. Dávate na misky váh všetky *pre* a *proti*, a koncom týždňa chcete kompetentným orgánom predložiť svoje jasne podložené a racionálne odôvodnené rozhodnutie; je tu však malý, ale zásadný problém: Koncom týždňa je piatok trinásteho! A aj keď neveríte na povery, veď ste moderný výskumník, ale čo keď platia, aj keď im neveríte?! Našťastie, máte poruke moderné metódy štatistickej analýzy. Preveríte si, či je piatok trinásteho vhodný deň na veľké

rozhodnutia tým, že zistíte ako sa iným ľuďom v tento deň vodilo. Bude to napr. odhad relatívnej početnosti mrzutosti v piatok trinásteho na hladine významnosti α=0,05 s presnosťou d=10%. Hravo si podľa **[VI.17]** spočítate rozsah výberu, ktorý musíte pre splnenie zadaných podmienok realizovať (v tabuľkách kvantilov N(0,1) je $u_{1-\alpha/2}$ = 1,96; pretože α=0,05 $\rightarrow$1-α/2 = 0,975; d = 0,1):

$$n = u_p^2 . \frac{\pi(1-\pi)}{d^2} = 1,96^2 . \frac{0,5.0,5}{0,1^2} \cong 96$$

Je krásna neskorá noc a vy zabezpečujete náhodný výber tak, že bez rozmýšľania vyťukávate za sebou na mobile čísla. Realizovaný pokus bude pre vás ten, keď sa druhá strana s vami spojí. Priaznivý výsledok budú predstavovať všetky priame i nepriame kladné odpovede na otázku: *Mali ste už niekedy v piatok trinásteho veľké mrzutosti?* Teda po odfiltrovaní deťom, matkám a mládeži neprístupných výrazov, to budú napr. odpovede: „Áno, ty ...“; Samozrejme, čo sa tak idiotsky pýtaš...“; „ ...“; „Ak ešte raz zavoláš, tak aj ty budeš mať veľké mrzutosti, ty ...“ a pod.

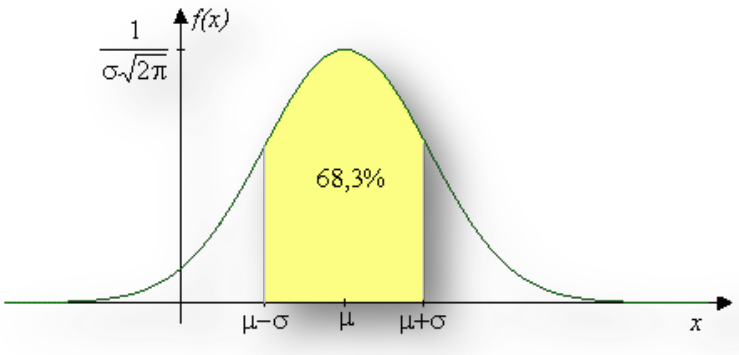
Keď ste realizovali celý výber n = 96 respondentov a dostali ste 81 kladných odpovedí, máte odhad relatívnej početnosti mrzutosti v piatok trinásteho

$$p = \frac{m}{n} = \frac{81}{96} \cong 0,84$$

Zdá sa vám to predsa len trochu odstrašujúce, preto si urobíte ešte podľa **[VI.16]** interval spoľahlivosti odhadu relatívnej početnosti základného súboru

$$\langle p - u_{1-\frac{\alpha}{2}} . \sqrt{\frac{\pi(1-\pi)}{n}} ; p + u_{1-\frac{\alpha}{2}} . \sqrt{\frac{\pi(1-\pi)}{n}} \rangle = \langle 0,84 \mp 1,96 . \sqrt{\frac{0,25}{96}} \rangle \cong \langle 0,74; 0,94 \rangle$$

Zistili ste s chybou menšou ako 10%, na bežnej hladine významnosti α = 0,05, že 74 až 94% populácie malo v piatok trinásteho veľké mrzutosti. A odložíte svoje rozhodnutie na pondelok.



ZHRNUTIE

Dostali sme sa k záveru náročnejšej, ale veľmi užitočnej kapitoly s množstvom nových poznatkov a informácií, pre praktické používanie publikácie si preto urobme malé upratovanie:
Základný súbor – súbor prvkov veľkého rozsahu, ktorý nesie sledovanú charakteristiku a z ktorého sa výber uskutočňuje. Predpokladáme, že výber sa dá uskutočniť a že má náhodný charakter, t.j. že každý prvok základného súboru má rovnakú pravdepodobnosť dostať sa do výberu. Jeho rozsah je **N**. Charakteristiky **μ; σ; σ^2**, rozdelenie **$N(\mu;\sigma^2)$**.
Výberový súbor – súbor náhodne vybraných prvkov zo základného súboru. Rozsah je **n**. Charakteristiky: $\bar{x}$; s; s^2

n>30 – veľký súbor resp. veľký rozsah výberového súboru; **n<30**, malý súbor resp. malý rozsah výberového súboru.
Výbery s opakovaním – vybraný prvok sa po kontrole vráti naspäť do základného súboru, čiže počet prvkov v základnom súbore sa nemení, ide o nezávislé javy.
Výbery bez opakovania – vybraný prvok sa po kontrole nevráti naspäť do základného súboru a teda počet prvkov v ňom sa postupne mení (zmenšuje), ide o závislé javy.
Výberové metódy umožňujú riešiť úlohy typu:

1. **odhad charakteristík základného súboru pomocou výberových charakteristík**
   + určovanie intervalov spoľahlivosti,
2. **úlohy určovania rozsahu výberových súborov.**

Na základe stanovenej chyby, ktorá nemá byť pri odhade charakteristiky základného súboru prekročená, sa určuje najmenší rozsah výberového súboru
Normovaná výberová charakteristika *u* má tvar napr.

$$u = \frac{\bar{x} - \mu}{\sigma} . \sqrt{n} \qquad [VI.18]$$

Vlastnosti výberových charakteristík:
a) **konzistentnosť** – čím viac sa blíži výberový súbor k základnému, tým viac sa blíži výberová charakteristika charakteristike základného súboru. Napr. n→N tak $\bar{x} \to \mu$.

b) **nestrannosť** – bez systematických chýb.

c) **eficientnosť -** výberová charakteristika má mať čo najmenší rozptyl (charakteristiky majú byť čo najvýdatnejšie).

Bodový odhad výberových charakteristík je daný jedným číslom napr.: aritmetický priemer, rozptyl:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{potom } s = \sqrt{s^2}$$

Intervalový odhad výberových charakteristík udáva interval pokrývajúci odhadovaný parameter s určitou vopred predpísanou pravdepodobnosťou a chybou.

Interval spoľahlivosti nazývame interval, ktorý bude obsahovať charakteristiku základného súboru s určitou pravdepodobnosťou. Pravdepodobnosť, že skutočná hodnota parametra sa bude nachádzať v danom intervale, je **p=1- α**. Parameter **α** je malá pravdepodobnosť, ktorá predstavuje riziko mylného záveru. Pod pravdepodobnosťou **α** rozumieme pravdepodobnosť, že skutočná hodnota parametra bude ležať mimo hraníc intervalu spoľahlivosti, nazýva sa aj **hladina významnosti** alebo **kritická hodnota**. Prípustnú chybu označujeme ***d***. ***d*** aj **1-α** sa stanovuje vopred. Riziko mylného záveru sa zmenší, ak sa zväčší 1-α. Čím je α menšie, tým je interval širší, t.j. zase menšia presnosť.

<u>Stanovenie intervalových odhadov:</u>

1. Interval spoľahlivosti strednej hodnoty:

1.1. Základný súbor má normálne rozdelenie **N(μ,σ²)**. Výberový súbor **n>30**, disperzia **σ²** je známa:

Výberový odhad strednej hodnoty: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Normovaná výberová charakteristika má tvar

$$u_p = \frac{\bar{x} - \mu}{\sigma}.\sqrt{n}$$

a rozdelenie N(0;1). u_p je pre rôzne kritické hodnoty **p = 1- α**, resp. **1-α/2** tabelované.

Interval spoľahlivosti odhadu strednej hodnoty **μ** základného súboru pomocou výberového aritmetického priemeru $\bar{x}$, kde $d = u_p.\frac{s}{\sqrt{n}}$ bude:

$$\bar{x} - u_{\frac{\alpha}{2}}.\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + u_{\frac{\alpha}{2}}.\frac{s}{\sqrt{n}} \qquad \textit{pre obojstranný interval}$$

$$\mu \le \bar{x} + u_{\alpha}.\frac{s}{\sqrt{n}} \qquad \textit{pre pravostranný interval}$$

$$\bar{x} - u_{\alpha}.\frac{s}{\sqrt{n}} \le \mu \qquad \textit{pre lavostranný interval}$$

a minimálny rozsah výberu **n**, potrebný na zaistenie požadovanej presnosti a spoľahlivosti odhadu je:

$$n = (u_p.\frac{s}{d})^2$$

1.2. Základný súbor má normálne rozdelenie **N(μ,σ²)**. Disperzia **σ²** je neznáma (častejší prípad), **n<30**:

Normovaná výberová charakteristika má tvar

$$t_a = \frac{\overline{x} - \mu}{\sigma}.\sqrt{n}$$

a **Studentovo t-rozdelenie**. t_a je pre rôzne kritické hodnoty **p = 1- α**, resp. **1-α/2** a pre (n-1) stupňov voľnosti tabelované.

Interval spoľahlivosti:

$\overline{x} - t_{n-1;\frac{\alpha}{2}}.\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{n-1;\frac{\alpha}{2}}.\frac{s}{\sqrt{n}}$ *pre obojstranný interval*

$\mu \leq \overline{x} + t_{n-1;\frac{\alpha}{2}}.\frac{s}{\sqrt{n}}$ *pre pravostranný interval*

$\overline{x} - t_{n-1;\frac{\alpha}{2}}.\frac{s}{\sqrt{n}} \leq \mu$ *pre ľavostranný interval*

2. Interval spoľahlivosti odhadu rozptylu základného súboru:

Interval spoľahlivosti pre smerodajnú odchýlku σ resp. pre rozptyl **σ²** pomocou výberového rozdelenia **χ^2 s k=(n-1) stupňami voľnosti** sa vypočíta nasledovne:

Normovaná výberová charakteristika má tvar

$$\chi 2 = \frac{(n-1).s^2}{\sigma^2}$$

Zo základného súboru sa náhodne vyberie **n** prvkov a vypočíta sa odhad výberového rozptylu **s²**. Rozptyl **s²** nemá symetrické rozdelenie, ale rozdelenie **χ^2**. (1-α) %-ný interval spoľahlivosti pre odhad rozptylu základného súboru je

a) obojstranný interval spoľahlivosti pre disperziu **σ²**

$$\left\langle \frac{(n-1).s^2}{\chi^2_{\frac{\alpha}{2};k}}; \frac{(n-1).s^2}{\chi^2_{1-\frac{\alpha}{2};k}} \right\rangle$$

b) ľavostranný interval spoľahlivosti pre disperziu **σ²**

$$\left\langle \frac{(n-1).s^2}{\chi^2_{\frac{\alpha}{2};k}}; \infty \right\rangle$$

c) pravostranný interval spoľahlivosti pre disperziu σ^2

$$\langle 0; \frac{(n-1).s^2}{\chi^2_{1-\frac{\alpha}{2},k}} \rangle$$

Pre určenie intervalu spoľahlivosti vypočítame z $[VI.11]\, s^2(n-1) = \sum_{i=1}^{n}(x_i - \bar{x})^2$

a hodnoty $\chi^2_{p,k}$ sú pre **(1-α)** a stupne voľnosti **k=(n-1)** tabelované.

3. Interval spoľahlivosti odhadu relatívnej početnosti základného súboru π.

Štatistika na získanie intervalu spoľahlivosti odhadu relatívnej početnosti (alternatívne rozdelenie **p = m/n**) je

$$u = \frac{p - \pi}{\sqrt{\pi(1-\pi)}}.\sqrt{n}$$

Interval spoľahlivosti odhadu relatívnej početnosti základného súboru bude na hladine významnosti **1-** $\frac{\alpha}{2}$ (resp. 1-α pri jednostranných intervaloch):

$$\langle p - u_{1-\frac{\alpha}{2}}.\sqrt{\frac{\pi(1-\pi)}{n}}; p + u_{1-\frac{\alpha}{2}}.\sqrt{\frac{\pi(1-\pi)}{n}} \rangle$$

kde **u** sú tabelované kvantily N(0,1) pre hladinu významnosti α. Pre stanovenie minimálneho potrebného rozsahu výberu (nezávislý výber tak, aby sa neprekročila požadovaná chyba ***d***) platí:

$$n = u_p^2.\frac{\pi(1-\pi)}{d^2}$$

Ak relatívnu početnosť základného súboru π nepoznáme, berieme hodnotu p = 0,5 a máme pri danej hladine významnosti taký rozsah výberu **n**, ktorý nám zabezpečí nielen požadovanú, ale aj lepšiu presnosť ***d***.

Literatúra k VI. kapitole

[1] Biermann, K. a kol.: Kronika medicíny, Fortuna Print, Praha 1994

[2] http://de.wikipedia.org/wiki/Contergan-Skandal

[3] Contergan, TV film, Nemecko, 2007, réžia Adolf Winkelmann

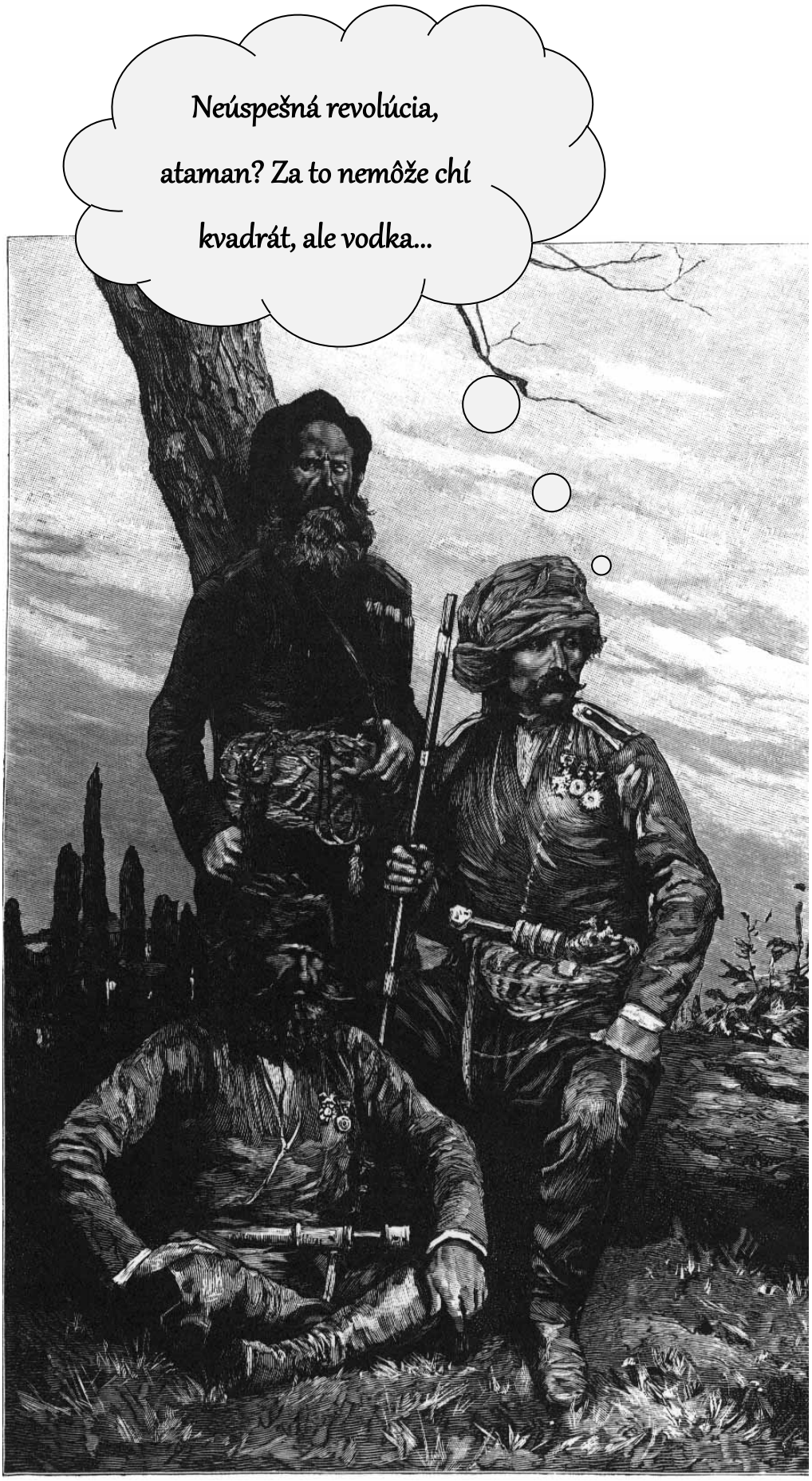
[4] Pecáková,I.: Statistika v terénních průzkumech, Praha, Professional Publishing 2011

[5] http://user.mendelu.cz/drapela/Statisticke_metody/teorie%20text%20I.pdf

[6] http://en.wikipedia.org/wiki/Chi-squared_distribution

[7] Hajduk, A. a kol.: Encyklopédia astronómie, Obzor, Bratislava 1987

[8] http://sk.wikipedia.org/wiki/S%C3%BAbor:Solar_eclips_1999_4.jpg


Neúspešná revolúcia,
ataman? Za to nemôže chí
kvadrát, ale vodka...

VII. Ako si vybrať ešte lepšie alebo hypotetická spokojnosť

V štatistike nejde ani tak o to veci chápať, ako skôr si na ne zvyknúť.
John von Neumann (voľne parafrázované podľa citátu o matematike)

„To nedopadne dobre," – smutne sa pousmial Jára Cimrman v rohu reštaurácie *Pod Vyšehradem*, zameranej na rybie špeciality, keď skladal svoje New York Times, kde sa práve dočítal, že Nór Roald Amundsen otočil svoju loď zo severu na juh ihneď, ako sa dozvedel, že severný pól dobyl Američan Robert Peary; a začal naháňať Brita Roberta F.Scotta. A mal pravdu: Keď Scott s vynaložením všetkých síl a po skonzumovaní posledného polárneho poníka dorazil k najjužnejšiemu bodu zemegule, zistil, že Amundsen ho predbehol o 35 dní. Scott a celá jeho výprava zostali pochovaní v zúfalstve a vo večnom ľade. Amundsenovi na spiatočnej ceste kalilo radosť z víťazstva nielen zlé počasie, ale aj nevysvetliteľná záhada vyrezávanej palice z českej lipy, zapichnutá v ľade presne v bode predĺženia zemskej osi, na ktorej sa v lúčoch nízkeho slnka trblietal zavesený ozdobný medailónik, skrývajúci fotografiu neznámeho muža s dvomi Inuitkami.

Jára Cimrman nemyslel však na túto tragédiu, južný pól ho už nezaujímal. Objednal si porciu arktického tuniaka a na zahriatie horúci grog. Bol rozhodnutý. Aj keď doteraz úspešne skrýval svoje víťazstvá pri dobíjaní pólov, aby to nezneužila propaganda nenávidenej c. a k. monarchie, musela dostať prednosť vedecká pravda pred politikárčením [1], [2].

Mal pri sebe denník výpravy, ktorú začal samozrejme zo *Zeme cisára Františka Jozefa I.*, na rozdiel od Pearyho, ktorý tvrdošijne štartoval z okraja Grónska, z *Ellesmerovho ostrova*. Musel prejsť prakticky 900 km, teda o 140 km viac ako Peary, ale jeho výhodou bola malá početnosť výpravy. Štyria otužilci, ktorých vzal v Prahe do výpravy, sa zostali kúpať v miestnych gejzíroch. V záverečnej fáze tak vyštartoval len s dvomi Inuitkami, ktoré udržiavali teplo rodinného krbu. Po počiatočnej polárnej trudnomyseľnosti povzbudzoval Jára Cimrman vo svojej minivýprave dobrú náladu vymýšľaním rôznych zábavných až

roztopašných odmien pre toho (tú), kto prvý dosiahne vytýčený denný cieľ. Každá z Inuitiek poháňala jeden pretekársky psí záprah, čím sa dosiahla nadpriemerná rýchlosť postupu. Keď sa ho pýtali na smer, tak Cimrman, ktorý v tej dobe ešte neprenikol úplne do tajov inuitského jazyka, kričal (kvôli silnému severáku) stereotypne: „Na sever! Na sever!“

Po treťom grogu vytiahol zápisníky Roberta Pearyho i Fredericka Cooka. Získal ich od natešeného výčapníka v tmavom grónskom pube, ktorý ich dostal do zálohy za sekeru, čo mu zanechali premrznutí návštevníci. Množstvo údajov, ktoré mal k dispozícii musel usporiadať a porovnať. Rozhodol sa aplikovať svoju metódu *CUŠATEHY* (Cimrmanov univerzálny štatistický algoritmus testovania hypotéz). Výsledky mienil predložiť priamo v jame levovej, v kongrese Spojených štátov, kde sa chystal v kuloároch prejednať nové budúce usporiadanie Európy v prípade, že by vypukla svetová vojna, o čom bol presvedčený, že je na spadnutie.

Podstatou Cimrmanovej metódy *CUŠATEHY* bolo, že si stanovil tri objektívne hypotézy, ktoré sa rozhodol otestovať:

H_{-1}: Frederick Cook dosiahol severný pól ako prvý.

H_0: Robert Peary dosiahol severný pól ako prvý.

H_1: Jára Cimrman dosiahol severný pól ako prvý.

Mínus prvú hypotézu Jára Cimrman zamietol okamžite z niekoľkých dôvodov:

a) mal zvlášť vyvinutý a neustálym precvičovaním praktických úloh posilňovaný hlboký štatistický cit,

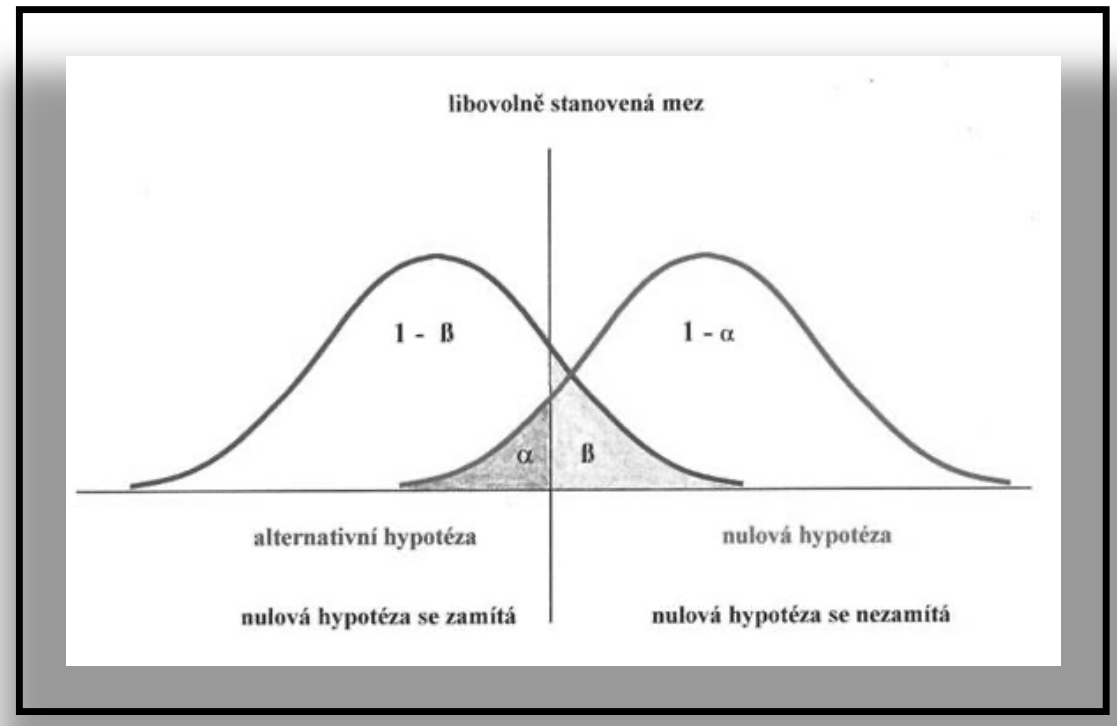
b) nemal rád záporné hypotézy, k práci vždy pristupoval pozitívne a záporné hypotézy od tohto okamihu podľa všetkých známych dokumentov z jeho života už nikdy nepoužíval,

c) podľa miestoprísažných výpovedí grónskych domorodcov Frederick Cook bol v čase jeho výpravy postihnutý silným záchvatom morskej choroby a pevninu nikdy neopustil; chronický skorbut, prejavujúci sa pri dlhodobom polárnom pobyte mu zatienil bádateľské myslenie, ktoré si potom fabulovalo ako chcelo.

Sústredil sa na dve hypotézy: Nultú, v ktorej dal džentlmensky prednosť svojmu rivalovi, a na alternatívnu hypotézu komentovanú slovami:“ Na póle nie je miesto pre dvoch!“ Obaja bádatelia merali v deň, keď boli presvedčení, že sa nachádzajú na póle, niekoľkonásobne svoju vzdialenosť od pólu pomocou presného sextantu vychýrenej viedenskej firmy Schwetz und Sohn, ktorý umožňoval merania so smerodajnou odchýlkou σ=150 m. Výsledky takýchto meraní podľa Cimrmana podliehali normálnemu rozdeleniu. Dosť práce mu dalo rozlúštiť všetky záznamy, niekoľkokrát rozmočené snehom a vysušené vlastným telom polárnikov, ktoré si zaznamenal do tabuľky čiastočne na okraj N.Y.Times a čiastočne na obrus stola (s.z.č. je severná zemepisná šírka po započítaní všetkých korekcií):

č.	Jára Cimrman 5.4.1909		Robert Peary 6.4.1909	
	vzdial. od pólu [m]	s.z.š.	vzdial. od pólu [m]	s.z.š.
1	0	90°00´00´´	0	88°33´30´´
2	-150	90°00´05´´	-1200	88°34´10´´
3	450	89°59´45´´	1200	88°32´50´´
4	300	89°59´50´´	-450	88°33´45´´
5	-150	90°00´05´´	-450	88°33´45´´
6	-150	90°00´05´´	900	88°33´00´´
7	0	90°00´00´´	-600	88°34´00´´
8	-150	90°00´05´´	1800	88°32´30´´
9	-300	90°00´10´´		
10	150	89°59´55´´		
μ	**0**	**90°00´00´´**	**150**	**88°32´50´´**

Dostal 2 nezávislé štatistické súbory nezávislých meraní odchýlok od pólu s rovnakou smerodajnou odchýlkou σ=150 m. Nakreslil si obrázok v materinskom jazyku [3]:



Obr.VII.1. Testovanie štatistických hypotéz podľa Járy Cimrmana

Potom si v duchu povedal: Pravý kopček bude predstavovať normálne rozdelenie meraní Pearyho, ľavý moje. Ak si zvolím nejaké spoločné štatistické kritérium pre oba kopčeky, ktoré bude mať normované normálne rozdelenie N(0;1), vypočítam ho a ak jeho hodnota bude väčšia resp. rovná tabuľkovej hodnote N(0;1) pre nejakú kritickú hodnotu α napr. ako býva zvykom α=0,05, tak rozdiel medzi strednými hodnotami súborov je štatistický nevýznamný, t.j. nesmiem nulovú hypotézu zamietnuť a musím uznať, že Peary na póle bol. Ak však bude vypočítaná hodnota pod kritickou, musím nulovú hypotézu zamietnuť a prijať hypotézu alternatívnu. Chvíľu uprene hľadel na prekrývajúcu sa oblasť a melancholicky prehodil:

„Ach, budem sa na to musieť ešte v príhodnú chvíľu pozrieť...“

Cimrmanov ďalší testovací postup bol nasledovný:

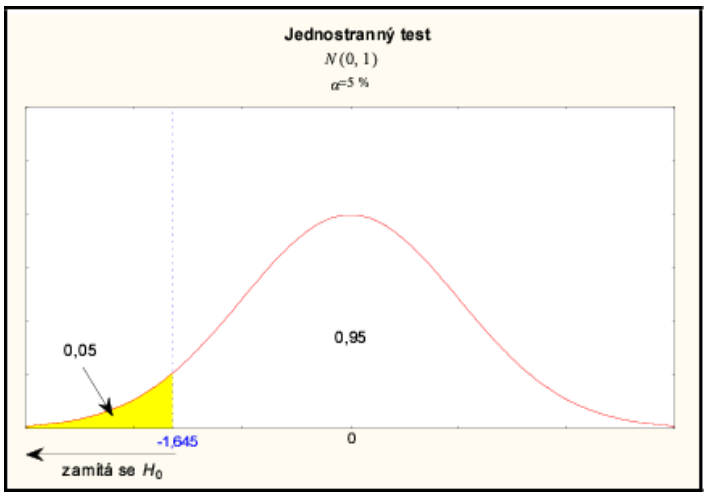
1. Formulácia nultej hypotézy po pretavení do štatistického jazyka znamenala, že stredná hodnota nameraných odchýlok Pearyho - μ_0 bude rovnaká ako jeho - μ_1. Keďže Cimrman vedel, kde stál, to znamenalo, že môže objektívne prijať hypotézu o tom, že Peary dobyl severný pól len vtedy, keď stál na tom istom mieste ako on, matematicky vyjadrené, na hladine významnosti $\alpha = 0,05$ platí

$$\mu_0 = \mu_1 \qquad \textbf{[VII.1]}$$

2. 95%-nú hladinu významnosti, t.j. pre kritickú hodnotu $\alpha = 0,05$ vybral Cimrman ako štandardný postup pre podobné testy. Alternatívna hypotéza v tomto prípade, teda že Cimrman stál na póle a Peary nie, znamenala za rovnakých podmienok, že stredná hodnota Cimrmanových odchýlok vzdialenosti od pólu bude menšia ako Pearyho:

$$\mu_1 < \mu_0 \qquad \textbf{[VII.2]}$$

K nej si tiež nakreslil obrázok [4] :



Obr.VII.2. Ľavostranný test štatistickej hypotézy H_0 proti H_1 pre $\mu_1 < \mu_0$ podľa Járy Cimrmana, ktorému na ľavej ruke po polárnej výprave zostal kompletný počet prstov.

Hodil očkom do štatistických tabuliek, ktoré mal vždy poruke a zistil, že tabuľková hodnota pre ľavostranný test pre $\mu_1<\mu_0$ bude $u_\alpha = u_{0,05} = -u_{1-0,05} = -u_{0,95} = \mathbf{-1,645}$.

3. Aby mohol využiť tabuľky normovaného normálneho rozdelenia N(0;1), objednal si ďalší grog, prižmúril oči nad skutočnosťou, že ide o vcelku málo rozsiahle štatistické súbory a ako testovacie kritérium si zvolil

$$u_p = \frac{\mu_1 - \mu_0}{\sigma.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \textbf{[VII.3]}$$

Vo vzťahu **[VII.3]** už všetko poznal, tak vypočítal

$$u_p = \frac{\mu_1 - \mu_0}{\sigma.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0 - 150}{150.\sqrt{\frac{1}{10} + \frac{1}{8}}} \cong -2,1082$$

Kritickú oblasť pre zamietnutie H_0, teda pre platnosť ľavostrannej alternatívnej hypotézy označil W (podľa nemeckého podstatného mena die Wahr, pretože bol presvedčený, že pravda zvíťazí):

$$W_\alpha = (-\infty; -1,645)$$

4. Po tomto mu zostala už len radostná interpretácia výsledku testu štatistickej hypotézy o štatistickej významnosti resp. bezvýznamnosti rozdielu dvoch stredných hodnôt.

Keďže $u_p < u_\alpha$, vypočítaná hodnota štatistického kritéria jednoznačne spadla do kritickej oblasti pre zamietnutie H_0 a pre prijatie H_1. Hypotézu, že Peary bol na severnom póle je nutné zamietnuť a prijať hypotézu, že na póle bol s 95% pravdepodobnosťou Jára Cimrman.

Ako vždy mal Cimrman pre svoje tvrdenia v zálohe ešte niekoľko poistiek:

- ako si možno všimnúť, pripojil k tabuľke aj jednotlivými polárnymi bádateľmi sextantom namerané hodnoty severnej zemepisnej šírky (s.z.š.), z ktorých vyplývalo, že Peary sa k pólu priblížil len na vzdialenosť asi 160 km, alebo že vôbec nevie narábať s náročným meracím prístrojom a vlastne ani vtedy a možno ani nikdy doteraz Peary nevie, kde je. A vôbec mu nepomohla poznámka v polárnom denníku:
- *Je taký mráz, že sa mi lepia prsty na ten sprostý sextant. Zaokrúhlim to na 89°57′, aby to vyzeralo dôveryhodne. Nedám si po toľkej námahe predsa víťazstvo odriftovať ľadom!*
- posledným tromfom v rukáve však zostal dátum dobytia pólu, ktorý jasne nasvedčoval na Cimrmanovo prvenstvo.

Kongres USA bol Cimrmanovou štatistickou analýzou veľmi zarazený a vzápätí sa strhla prudká hádka medzi republikánmi a demokratmi a vážna vnútropolitická kríza. Kongresmani hodili Pearyho cez palubu. Zlomový okamih nastal, keď Cimrman popisoval ako dlhé hodiny stál na póle a svet sa točil okolo neho ako uprostred kolotoča; ani nie tak kvôli slávnostnosti a neopakovateľnosti chvíle ale preto, lebo odrazu (bol aj trochu zmätený a točila sa mu hlava) nevedel kadiaľ domov. Nech by sa pustil akýmkoľvek smerom, všetko bolo na juh, ale správne tušil, že nie každá cesta je správna. Zachránila ho spásonosná myšlienka, svojou jednoduchosťou hodná génia. Zohol sa tak, pokiaľ mu to tulenie kožušiny dovoľovali, že mal hlavu medzi kolenami a dolu hlavou odrazu pozeral vlastne na sever. Stačilo sa otočiť čelom vzad, presne podľa príručiek c. a k. administratívneho velenia rakúsko-uhorskej armády, a cesta na juh bola otvorená.

Definitívnemu celosvetovému potvrdeniu Cimrmanovho víťazstva zabránili až finanční magnáti z Wall Streetu, ktorí sa zľakli jeho možnej slávy a toho, že Cimrmanová štatistická metóda *CUŠATEHY* by mohla po jej spopularizovaní priviesť krajinu k veľkému hospodárskemu krachu a celú históriu zahmlili. Cimrman však v tom čase už dávno *oral na inom poli*, stopy po jeho polárnych dobrodružstvách a objavoch nachádzame v známych i menej známych umeleckých výbojoch velikána (divadelné hry, živé obrazy a i.).

Skôr než si zovšeobecníme testovanie štatistických hypotéz uvedieme niekoľko poznámok, aj keď patria skôr do širšej problematiky metodológie výskumu [5], [6]:

Hypotéza je určitým predpokladom vysloveným o javoch v našom okolí. **Štatistická hypotéza** je jednoduchý výrok o nejakom správaní sa pravdepodobnostných (náhodných) premenných, ktoré sa dajú kvantitatívne stanoviť – merať, porovnávať. Prvú premennú väčšinou vyberáme z dvoch alebo viacerých úrovní a dá sa zistiť, druhá premenná býva merateľná. Štatistickou hypotézou sú napr. výroky:

Seniori ľahšie podliehajú závislosti na hazardných hrách ako ľudia v strednom veku.

Manželstvá uzavreté pred nadobudnutím plnoletosti sú menej stabilné ako manželstvá uzavreté zrelými partnermi nad 18 rokov.

Vo veľkých mestách je nižšia hranica veku prvej skúsenosti s nepovolenou drogou ako v malých.

Štatistickú hypotézu možno vyjadriť aj implikáciou „ak – tak“, napr.:

Ak sa zvýši počet profesionálnych napr. poradenských psychológov v poradenskej činnosti úradov práce, tak sa zníži dlhodobá nezamestnanosť.

Ak sa zlepší starostlivosť o imigrantov, tak sa zlepší ich uplatnenie na trhu práce.

Ak sa predĺži povinná školská dochádzka, zlepšia sa naši študenti v medzinárodnom porovnávaní úrovne vzdelávania.

Úrovne prvej premennej bývajú vyjadrené väčšinou stupňovaným porovnaním typu „lepší, horší, menší, väčší, mladší, starší, ľahšie, ťažšie," a podobne. V prvom výroku sú úrovne premennej *seniorský vek* a *stredný vek*, porovnanie javu medzi úrovňami (sklon k nekontrolovateľným hazardným hrám) je výrazom „ľahšie", ktoré umožňuje položiť otázky a výskum na potvrdenie tejto hypotézy kvantifikovať (merať čas strávený pri hazardných hrách, vložené resp. prehraté peniaze, relatívna početnosť výskytu gamblerstva v populácii, atď.). Podobne sa dajú kvantifikovať vzťahy medzi premennými vo formulácii hypotéz pomocou implikácie. Na hypotézu sa musí dať jednoznačne odpovedať, napr. formou jej potvrdenia alebo zamietnutia. Toto sa nazýva **testovanie štatistickej hypotézy**.
Hypotézou, obzvlášť štatistickou, aj keď sa to nezdá, nie sú výroky typu:

Ráno vstanem a pravdepodobne sa obarím čajom.
Čo budú robiť študenti VŠ cez prázdniny?
Úrady práce znížia nezamestnanosť o 25% štatisticky na 95% hladine významnosti.

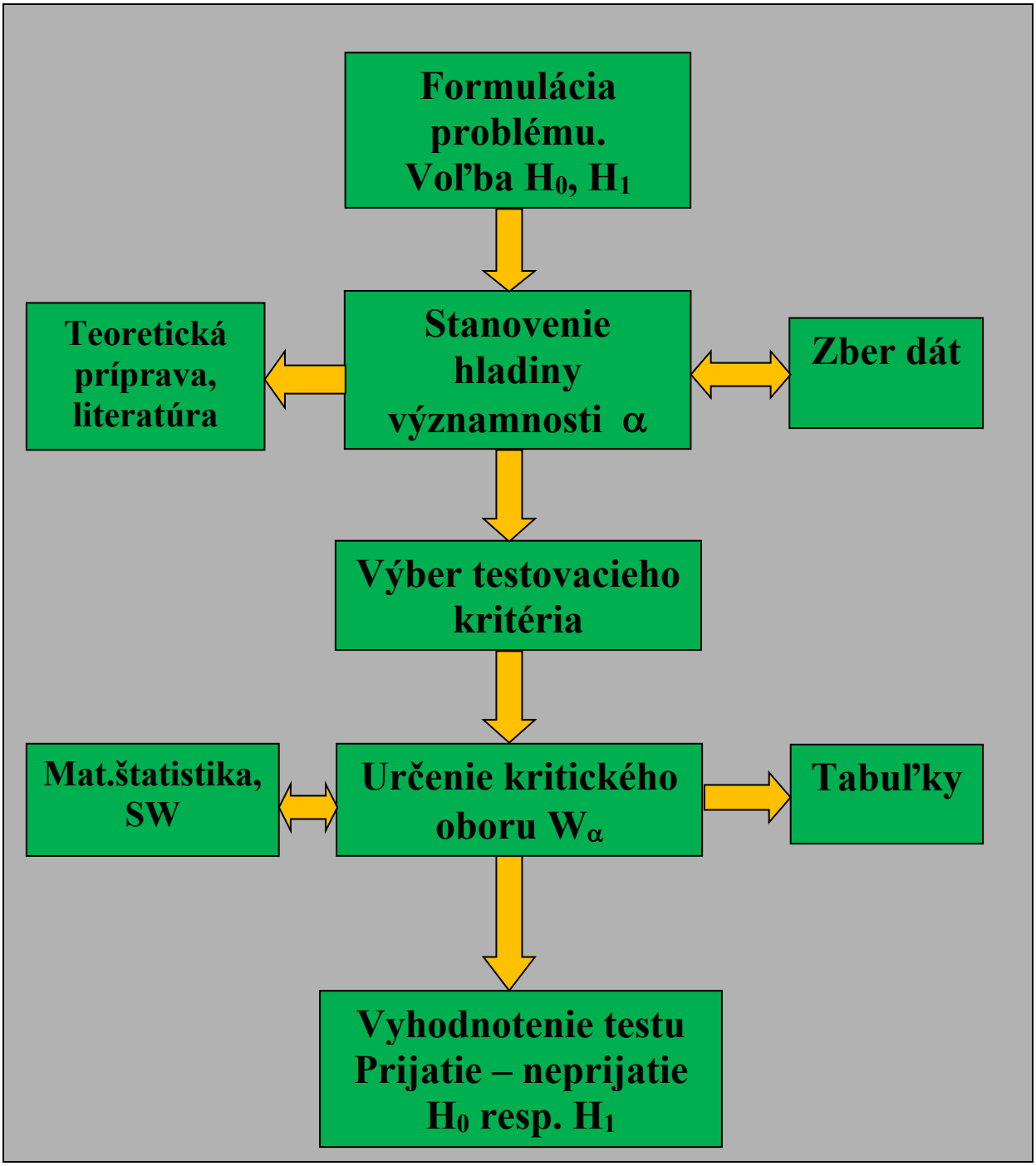
Či sa nám to páči alebo nie, formulovanie hypotéz je v niektorých oblastiach dosť zásadný problém, od ktorého sa odvíja kvalita celého ďalšieho výskumu. V predchádzajúcej kapitole pri hľadaní intervalov spoľahlivosti sme sa už niečo naučili z oblasti používania normovaných výberových štatistík, o práci so štatistickými tabuľkami, o hľadaní intervalov spoľahlivosti pokrývajúcich výsledok štatistickej analýzy a podobne. Keď si čitateľ k tomu priberie zopár užitočných informácií z metodológie testovania hypotéz (odporúčame aspoň [7]), tak nielen, že by nemal mať potom problémy s pochopením obsahu tejto kapitoly, ale testovanie hypotéz sa mu môže stať mimoriadne užitočným štatistickým nástrojom. Na obr.VII.3 je všeobecná schéma testovania štatistických hypotéz.

Takže máme nejaký výrok o rozdelení pravdepodobnosti nejakého znaku, resp. jeho parametrov – štatistickú hypotézu, a overujeme jeho správnosť, čiže testujeme hypotézu. Pri testovaní kladieme oproti sebe dve navzájom si odporujúce hypotézy. Hypotézu, ktorej platnosť overujeme, nazývame testovanou alebo nulovou hypotézou $\mathbf{H_0}$. Je to najčastejšie opak toho, čo chceme dokázať. Oproti nej kladieme alternatívnu hypotézu $\mathbf{H_1}$. Vyššie sme mali náznak testovania štatistickej hypotézy o rovnosti stredných hodnôt 2 súborov s normálnym rozdelením a s rovnakým známym rozptylom σ^2. Zápis je :

$$\boldsymbol{H_0: \mu_0 = \mu_1} \quad \text{oproti} \quad \boldsymbol{H_1: \mu_0 \neq \mu_1} \qquad \textbf{[VII.4.]}$$

Takáto alternatívna hypotéza sa nazýva dvojstranná, jej platnosť môže byť v nejakom intervale naľavo aj napravo od intervalu pre prijatie nulovej hypotézy. Ľavostranná alebo pravostranná alternatívna hypotéza sa zapíše v závislosti od konkrétneho riešeného problému:

$$H_1: \mu_0 < \mu_1 \quad \text{resp.} \quad H_1: \mu_0 > \mu_1 \qquad \textbf{[VII.5.]}$$



Obr.VII.3. Obecná schéma testovania štatistických hypotéz

Na testovanie nulovej hypotézy oproti alternatívnej vyberieme a použijeme nejaké vhodné testovacie kritérium, ktoré sa nazýva aj testovacia štatistika. Aj keď sa už čitateľovi začína zdať, že vie toho zo štatistiky strašne veľa, o čom sa mu ešte prednedávnom ani

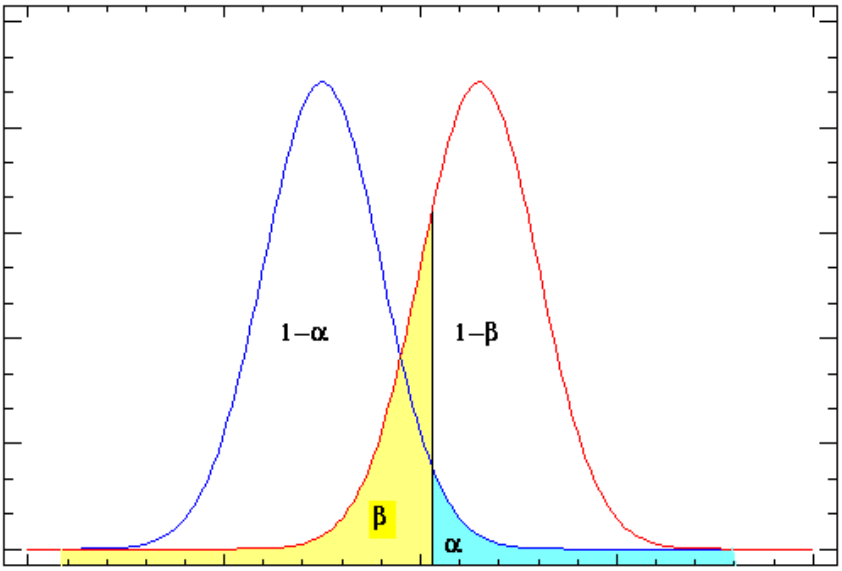
nesnívalo, jeho výber nie je otázkou jednorázového naučenia sa, ako sa to má robiť. Nepomôže ani len jedna jediná publikácia alebo webovská adresa, na ktorú je ochotný sa mrknúť. Ale netreba sa zľaknúť, jednoducho je potrebné s tým pracovať, pozrieť si ako na to išli pri podobných problémoch iní, vyhľadať rôzne príklady, prípadne prijať dobrú radu napr.:

$$testovacia\ štatistika = \frac{pozorovaná\ hodnota - predpokladaná\ hodnota\ pri\ H_0}{smer.odchýlka\ pozor.hodnoty/\sqrt{n}}$$

Keď sa vám to nejakým spôsobom podarí a viac-menej racionálne ste si to aj odôvodnili, je potrebné sa zaoberať oborom hodnôt (intervalom), ktoré môže testovacia štatistika nadobúdať. Celý interval si rozdelíme na dve oblasti:

- **kritický obor W_α** je interval zamietnutia testovanej hypotézy,
- **doplnkový obor** je interval prijatia testovanej hypotézy.

Ak hodnota testovacej štatistiky (kritéria) padne do kritického oboru, H_0 musíme zamietnuť; ak padne do doplnkového oboru, nemožno ju zamietnuť, hypotézu H_0 potom na hladine významnosti testu α prijímame. Pre názornosť si pozrime obrázky VII.4 až VII.6, aj keď proti cimrmanovským obrázkom tiež nie sú výhrady [7, 8].

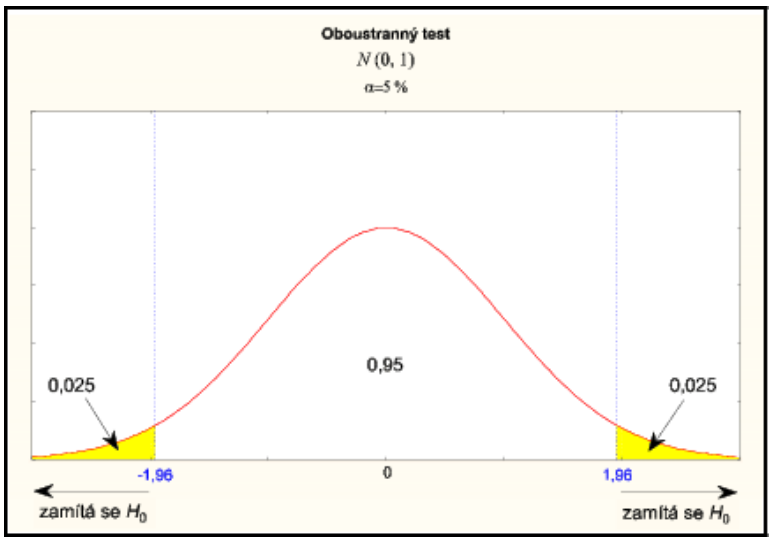


Obr.VII.4: Doplnkový a kritický obor pre prijatie resp. zamietnutie testovanej hypotézy

Obr. VII.4 je názorné zobrazenie oblastí prijatia resp. zamietnutia testovanej hypotézy. Modrý gaussián predstavuje rozdelenie pravdepodobnosti testovacej štatistiky pre nulovú hypotézu $\mathbf{H_0}$. Na zvolenej hladine významnosti α je interval **1-α** (teda všetky hodnoty menšie

ako kritická hodnota **α)** intervalom hodnôt testovacieho kritéria, kde **H_0** nemožno zamietnuť. Interval **α** (všetky hodnoty **>α**) je kritickým oborom **W_α**, v ktorom **H_0** zamietame.

Červený kopček predstavuje rozdelenie pravdepodobnosti testovacej štatistiky pre alternatívnu hypotézu **H_1**. Na zvolenej hladine významnosti **α** je interval **1-β** (teda všetky hodnoty **>** ako kritická hodnota **α)** intervalom hodnôt testovacieho kritéria, kde **H_1** nemožno zamietnuť a **H_0** zamietame. Interval **β** (všetky hodnoty **< α**) je doplnkovým oborom, v ktorom **H_0** nezamietame. Kritickú hodnotu **α** pri testovaní hypotéz nazývame **hladina** alebo **stupeň významnosti testu**. Volí sa podľa potreby **α** = 0,01; 0,05; 0,1 a i. **α** je teda pravdepodobnosť zamietnutia správnej hypotézy, teda miera rizika omylu. Na obr.VII.2. bol znázornený ľavostranný test prijatia resp. neprijatia štatistickej hypotézy H_0 pri hladine významnosti testu α = 0,05 s rozdelením pravdepodobnosti N(0;1). Pre porovnanie je na obr. VII.5 dvojstranný test prijatia resp. neprijatia štatistickej hypotézy H_0 pri hladine významnosti testu α = 0,05 s rozdelením pravdepodobnosti N(0;1). Z obrázkov je jasné, čo predstavuje kritická hodnota α a ako sa s ňou pri jednostranných resp. dvojstrannom teste narába.



Obr. VII.5.: Dvojstranný test prijatia resp. zamietnutia nulovej hypotézy H_0 [4]

Interval **1-β** je interval zamietnutia nulovej hypotézy **H_0** a nezamietnutia alternatívnej hypotézy **H_1**. Nazýva sa **sila testu**.

Testovaná nulová hypotéza **H_0** sa často formuluje ako predpoklad, že rozdiel medzi porovnávanými charakteristikami je náhodný, napr. rozdiel medzi výberovým priemerom $\bar{x}$ a priemerom základného súboru **μ** je náhodný a potom v skutočnosti nulový: $\bar{x} - \mu = 0$, teda ich rovnosť. V praxi hladinu významnosti testu **α**, na ktorej chceme **H_0** zamietnuť, alebo nezamietnuť, volíme čím nižšiu, aby sme čím viac eliminovali riziko zamietnutia pravdivej

nulovej hypotézy, teda zamietnutia, aj keď testovaná hypotéza platí. V štatistike sa volá **chyba I. druhu**.

Chyba II. druhu β je pravdepodobnosť, že nezamietneme testovanú nulovú hypotézu H_0 aj keď je nesprávna. Správne zamietnutie nepravdivej hypotézy je potom **sila testu 1-β**. Vidíme, že chyby I. a II. druhu sa navzájom vylučujú a dopĺňajú, môžeme si od fyzikov požičať výraz, že sú vzájomne **komplementárne**. Ak zmenšíme pravdepodobnosť chyby I.druhu, teda chybu, že zamietneme pravdivú hypotézu, zväčší sa nám chyba II.druhu, teda, že nezamietneme nepravdivú hypotézu a zároveň sa zmenší sila testu. Podrobne sa touto problematikou zaoberáme, pretože je pre prax dôležitá. Už sme sa s tým stretli pri Bayesovej podmienenej pravdepodobnosti, ako aj v príkladoch použitia niektorých testov a vyšetrovacích metód, ktoré majú vždy istú pravdepodobnosť potvrdenia diagnózy alebo prítomnosti nejakej látky, aj keď pacient chorobu nemá a na druhej strane istú pravdepodobnosť negatívnej odozvy, aj keď ju má. Keďže jedna chyba závisí od druhej nepriamo úmerne, je potom úlohou ich veľkosť vzájomne optimalizovať aj z hľadiska významu pre konkrétny problém. Aj keď väčšinou dávame prednosť znižovaniu chyby I.druhu (často sa aj v štatistike analýzy zaoberajú len chybou I.druhu, alebo napríklad súdna prax v krajinách, nazývaných demokratické, uprednostňuje v prípade nejasností ako omnoho prípustnejšie riziko prepustenie vinného ako potrestanie nevinného), sú prípady, kedy môže byť chyba II.druhu omnoho významnejšia. Príkladom môže byť vypustenie Hubblovho teleskopu na obežnú dráhu zeme bez kontroly optiky. Preverilo sa všetko na potvrdenie očakávaného stavu, teda na elimináciu chyby I.druhu. Ušetrilo sa na kontrole optiky, teda hypotézy, že by teleskop fungoval horšie, asi 20 000 dolárov. Oprava vo vesmírnom priestore stála viac ako 1,5 mld. dolárov. V humanitných vedách, ale hlavne v biomedicínskej štatistike má preto problematika chýb a testovania štatistických hypotéz mimoriadne postavenie.

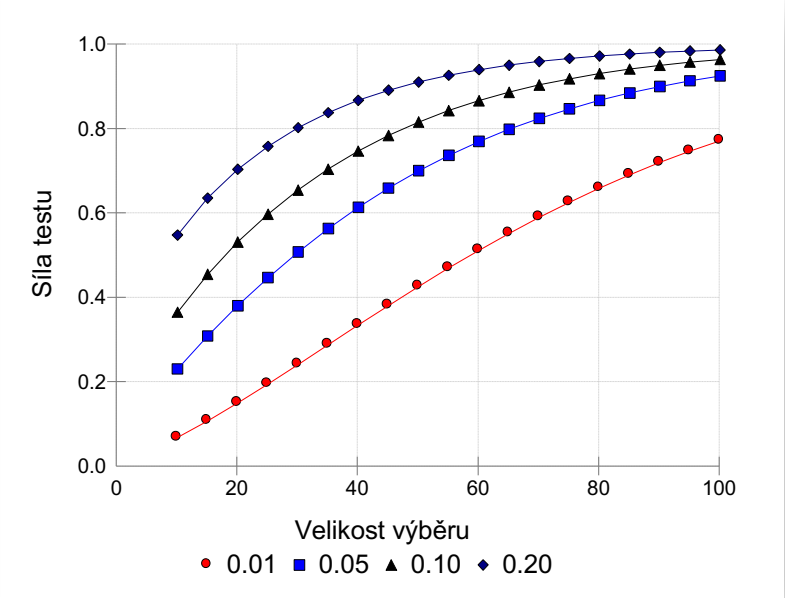
		H_0	
		platí	**neplatí**
Test hypotézy	**pozitívny**	+	**β, chyba II.**
	negatívny	**α, chyba I.**	+

Tab. VII.1.: Vznik chyby I. a II. druhu pri testovaní štatistických hypotéz

Pozrite si ešte raz obrázok VII.4: Sila testu **1 - β** sa zvýši, keď hranicu intervalu **1 - α** posunieme doľava, teda na úkor veľkosti doplnkového intervalu. Sila testu pri danej veľkosti **α** by sa mohla zvýšiť aj oddialením kopčekov, teda je závislá od toho ako je vzdialená hodnota alternatívnej hypotézy napr. priemeru od nulovej, označme si to $\Delta = \bar{x} - \mu$. Vo viacerých kontextoch nazývame **Δ** ako veľkosť účinku, ktorý sledujeme alebo očakávame. Väčšia variabilita roztiahne a „zníži“ kopčeky, zmenší plochu pod nimi a tým aj zníži silu testu. Ich zúženie a zvýšenie sa dosiahne zvýšením rozsahu súboru **n**. Tieto úmernosti možno zhrnúť do vzťahu:

$$S \sim \frac{\Delta . \alpha . \sqrt{n}}{\sigma} \qquad \text{[VII.6.]}$$

Závislosť sily testu **1 - β** od rozsahu súboru sme prevzali zo [7] a je na obr. VII.6. pre rôzne hladiny významnosti testu (**α= 0,01; 0,05; 0,10; 0,20**).



Obr.VII.6.: Závislosť sily testu od rozsahu súboru pri rôznych hodnotách α [7].

Podobné nomogramy na bezvýpočtový odhad potrebného rozsahu výberového súboru je možné nájsť v rôznych publikáciách zaoberajúcich sa medicínskou bioštatistikou napr. [9]. Keďže my radi počítame, uveďme si vzťah pre stanovenie potrebného rozsahu výberu súboru v jednoduchom teste hypotézy o strednej hodnote E(X) normálneho rozdelenia pri známom rozptyle σ^2 (H_0: $\mu=\mu_0$), s testovacím kritériom

$$u_p = \frac{\bar{x} - \mu_0}{\sigma} . \sqrt{n}$$

Je to výber zo základného súboru z normálnym rozdelením N(μ_0, σ^2), z ktorého vyberieme náhodne **n** prvkov a zisťujeme, či je rozdiel medzi $\bar{x}$ a μ_0 na nejakej hladine významnosti testu α štatisticky významný alebo nie. Pre rozsah výberového súboru sa dá použiť pri dvojstrannom teste:

$$n = [\left(u_{1-\frac{\alpha}{2}} + u_{1-\beta}\right) . \frac{\sigma}{\bar{x} - \mu_0}]^2 \qquad \text{resp.}$$

$$n = [\left(u_{1-\alpha} + u_{1-\beta}\right) . \frac{\sigma}{\bar{x} - \mu_0}]^2 \qquad \text{[VII.7.]}$$

pri jednostrannom teste, kde u_p sú hodnoty v tabuľkách normovaného normálneho rozdelenia N(0;1) pre zadané pravdepodobnosti. Často používané hodnoty sú uvedené v tabuľke:

p	0,90	0,95	0,975	0,99
u_p	1,283	1,645	1,960	2,326

Tab. VII.2.: Hodnoty kvantilov N(0;1) používaných v testoch pre najviac používané pravdepodobnosti p = 1-α, resp. p = 1-β.

A čo s tým všetkým? Myslím, že je načase, aby sme spolu začali testovať aspoň niektoré typy štatistických hypotéz. Nebudete na to potrebovať genialitu Járy Cimrmana, stačí trocha pozornosti a usilovnosti.

Pr.VII.1.: Ako jedno z opatrení zlepšenia pracovných podmienok boli na úrade práce vymenené okná za nové. Na jednej strane bola znížená hlučnosť z dopravy zvonku, na druhej strane boli isté pochybnosti či sa novým vnúteným vetraním cez membránové vetracie protipeľové mriežky nezhoršili mikroklimatické podmienky v interiéroch kancelárií a nepribudlo respiračných problémov. Poskytnutím účelového grantu pre zdatných študentov bola škola požiadaná, aby štatisticky zistila a vyhodnotila práceneschopnosť zamestnancov ÚP na respiračné ochorenia pred a po namontovaní nových okien a urobila kvalifikované závery.

Usilovní študenti sa s chuťou pustili do práce:

Krok I: Sformulovali si pracovnú testovanú nultú hypotézu.

H_0: Inštalácia nových okien s novým vetraním a systémom cirkulácie vzduchu má nevýznamný vplyv na priemernú mesačnú práceneschopnosť zamestnancov ÚP v dôsledku respiračných ochorení.

Alternatívna hypotéza H_1: Uvedená zmena má výrazný vplyv na PN pre respiračné choroby.

Krok II: Stanovili si hladinu významnosti testu hypotézy štandardne $\alpha = 0,05$; a pustili sa do zberu dát. V období niekoľkých rokov pred inštaláciou nových okien bola priemerná práceneschopnosť zamestnancov ÚP na respiračné ochorenia 1,95 dní/mesačne s normálnym rozdelením N(1,95; 0,01). Preto mohli položiť: $\mu = 1,95$; $\sigma = 0,1$; $\alpha = 0,05 = \beta$.

$1-\alpha = 1-\beta = 0,95$; $1-\alpha/2 = 0,975$. $u_{1-\alpha} = u_{1-\beta} = 1,645$ a $u_{1-\alpha/2} = 1,96$ (tab.VII.2).

Krok III. Výpočet rozsahu výberového súboru z **[VII.7.]** za podmienky, aby rozdiel neprekročil $\bar{x} - \mu_0 = 0,11$. Pre jednostranný test (ľavostranný alebo pravostranný):

$$n = \left[\left(u_{1-\alpha} + u_{1-\beta}\right).\frac{\sigma}{\bar{x} - \mu_0}\right]^2 = \left[(1,645 + 1,645).\frac{0,1}{0,11}\right]^2 \cong 9$$

Pre obojstranný test:

$$n = \left[\left(u_{1-\frac{\alpha}{2}} + u_{1-\beta}\right).\frac{\sigma}{\bar{x} - \mu_0}\right]^2 = \left[(1,960 + 1,645).\frac{0,1}{0,11}\right]^2 \cong 10$$

Krok IV. Zber údajov: V priebehu niekoľkých rokov získali študenti pre 10 náhodne vybraných zamestnancov priemernú mesačnú práceneschopnosť v dňoch za mesiac s diagnózami *respiračné ochorenia*, ktoré zhrnuli v tabuľke:

i	1	2	3	4	5	6	7	8	9	10	$\bar{x}$	s
d/m	2,2	2,0	1,8	2,3	2,1	1,8	2,4	2,0	1,9	2,1	**2,06**	**0,2**

Tab. VII.3.: Priemerné mesačné PN na respiračné choroby 10 zamestnancov

Krok V. Študenti si zvolili ako testovacie kritérium

$$u_p = \frac{\bar{x} - \mu_0}{\sigma}.\sqrt{n}$$

Krok VI. Obojstranné testovanie hypotézy.

H_0: $\mu_0 = 1,95$

H_1: $\mu \neq 1,95$

$n = 10$; $\alpha = 0,05$; $s \cong 0,2$

$\bar{x} = 2,06 \rightarrow \bar{x} - \mu_0 = 0,11$

$$u_p = \frac{\bar{x} - \mu_0}{\sigma}.\sqrt{n} = \frac{0,11}{0,2}.\sqrt{10} \cong 1,74$$

Test uskutočníme 3 spôsobmi:

a) **Stanovenie kritického oboru**

$W_\alpha = (-\infty; -u_{1-\alpha/2}) \cup (u_{1-\alpha/2}; \infty) = (-\infty; -1,96) \cup (1,96; \infty)$

Keďže $u_p = 1,74 \notin W_\alpha$ tak H_0 na hladine významnosti testu $\alpha = 0,05$ nemôžu zamietnuť.

Nové okna s novým systémom vetrania majú štatisticky zanedbateľný vplyv na práceneschopnosť zamestnancov v dôsledku respiračných ochorení.

b) **Test pomocou intervalu spoľahlivosti.** Hranice 100.(1-α) %-ného empirického intervalu spoľahlivosti strednej hodnoty $\bar{x}$ pri $\alpha = 0,05$ a $s = 0,2$ sú podľa **[VI.5]**:

$\langle x_{min}; x_{max} \rangle = \langle \bar{x} \mp u_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \rangle = \langle 2,06 \mp 1,96 \cdot \frac{0,2}{\sqrt{10}} \rangle = \langle 2,06-0,124; 2,06+0,124 \rangle = \langle 1,936; 2,184 \rangle$.

Pretože $\bar{x} = 1,95 \in \langle 1,936; 2,184 \rangle$, H_0 na hladine významnosti testu $\alpha = 0,05$ nemožno zamietnuť.

c) **Test pomocou p-hodnoty.** Vo IV. kapitole sme sa naučili pracovať s funkciou hustoty pravdepodobnosti a s distribučnou funkciou rozdelenia náhodnej premennej a využívať na to aj EXCEL. **p-hodnotu** pri testovaní hypotéz definujeme ako najmenšiu pravdepodobnosť, pri ktorej možno H_0 zamietnuť. Z takejto definície samozrejme študenti veľa úžitku nemajú, znie to veľmi učene ako celé testovanie hypotéz, ale nie je to veľká veda. Pozrime si spolu obrázok VII.5. pre dvojstranný test prijatia resp. zamietnutia hypotézy. Máme dve oblasti, v ktorých zamietame H_0, žlté chvostíky vľavo a vpravo. Vypočítame pravdepodobnosť p(1,74) pre našu hodnotu testovacieho kritéria $u_p = 1,74$ (z tabuliek alebo v EXCELI funkciou NORMSDIST(u_p)) a hodnotu pravdepodobnosti 1-p(1,74), zoberieme menšiu a vynásobíme 2, pretože máme 2 oblasti. Ak je výsledok menší ako hladina významnosti testu $\alpha = 0,05$, tak H_0 zamietame. Študenti sa toho chopili: p(1,74) = 0,95907, 1-p(1,74) = 0,04093. Menšiu hodnotu vynásobili 2: 2x(1-p(1,74)) = 2x0,04903 = **0,08186 > α** = 0,05, preto H_0 nemožno na tejto hladine zamietnuť.

Krok VII. Ľavostranné testovanie hypotézy.

$H_0: \mu = 1,95$

$H_1: \mu < 1,95$

a) **Test pomocou kritického oboru:**

$W_\alpha = (-\infty; u_\alpha) = (-\infty; u_{0,05}) = (-\infty; -1,645)$

$u_p = 1,74 \notin W_\alpha$ tak H_0 na hladine významnosti testu $\alpha = 0,05$ nemôžu zamietnuť.

b) **Test pomocou intervalu spoľahlivosti.** Hranice 100.(1-α) %-ného empirického intervalu spoľahlivosti strednej hodnoty $\bar{x}$ pri α= 0,05 a s = 0,2 sú

$(-\infty; x_{max}\rangle = \left(-\infty; \bar{x} + u_{1-\alpha}.\frac{s}{\sqrt{n}}\right) = \left(-\infty;\ 2{,}06 + 1{,}645.\frac{0{,}2}{\sqrt{10}}\right) = (-\infty; 2{,}164\rangle$

$\bar{x}$ = **1,95 ∈ (-∞; 2,164⟩**, H_0 na hladine významnosti testu α = 0,05 nemožno zamietnuť.

c) **Test pomocou p-hodnoty.** Pravdepodobnosť p(-∞; 1,74) = p(1,74) = 0,95907.

0,95907 > α = 0,05, preto H_0 nemožno na tejto hladine zamietnuť.

Krok VIII. Pravostranné testovanie hypotézy.

H_0: μ = 1,95

H_1: μ > 1,95

a) **Test pomocou kritického oboru:**

$W_\alpha = (u_{1-\alpha}; \infty) = (u_{0,95}; \infty) = (1{,}645; \infty)$

$u_p = 1{,}74 \in W_\alpha$ tak H_0 na hladine významnosti testu α = 0,05 zamietli v prospech pravostrannej hypotézy.

b) **Test pomocou intervalu spoľahlivosti.** Hranice 100.(1-α) %-ného empirického intervalu spoľahlivosti strednej hodnoty $\bar{x}$ pri α= 0,05 a s = 0,2 sú

$\langle x_{min}; \infty) = \left\langle \bar{x} - u_{1-\alpha}.\frac{s}{\sqrt{n}};\ \infty\right) = \left\langle 2{,}06 - 1{,}645.\frac{0{,}2}{\sqrt{10}};\ \infty\right) = \langle 1{,}956; \infty)$

$\bar{x}$ = **1,95 ∉ ⟨1,956; ∞)**, H_0 na hladine významnosti testu α = 0,05 zamietli v prospech pravostrannej hypotézy.

c) **Test pomocou p-hodnoty.** Pravdepodobnosť p(1,74; ∞) = 1-p(1,74) = 1-0,95907 = 0,04093.

0,04093 < α = 0,05, preto H_0 na hladine významnosti testu α = 0,05 zamietli v prospech pravostrannej hypotézy.

Možno sa vám zdá, že výsledok testu hypotézy, že výmena okien má zanedbateľný vplyv na respiračné ochorenia zamestnancov úradu, nedáva žiaden dôvod k radosti z poznania alebo z estetického zážitku. Potom ale asi nie ste riaditeľ dotyčného úradu ani jeho švagor, ktorý výmenu okien ako zákazku realizoval. (Autori s poľutovaním konštatujú, že dodatočne nedokázali zistiť, odkiaľ čerpali dáta v tomto príklade).

Vďaka usilovným študentom máme dostatočne podrobne prepočítaný dosť jednoduchý a názorný príklad ako návod na testovanie. Uvediem ešte nejaké jednoduché ilustračné príklady, ktorých údaje neboli získané terénnym výskumom:

Pr.VII.2.: V programe zameranom proti domácemu násiliu, ktoré sa vyskytuje asi v 30% rodín s rozptylom 0,0225, sa má zistiť, či zahájenie prevádzok krízových centier pre týrané matky s deťmi sú účinným nástrojom. Predpokladá sa, že možnosť dovolať sa na krízovú telefonickú linku a odísť pri hrozbe násilia do krízového centra zníži jeho výskyt na hodnotu 0,1 (teda asi na 10%).

H_0: Zriadenie krízových centier pre týrané matky s deťmi má zanedbateľný vplyv na výskyt domáceho násilia; $\mu_0 = \mu = 0{,}3$ (t.j. 30% zo zadania). $\sigma = \sqrt{\sigma^2} = 0{,}15$

H_1: $\mu_0 = 0{,}3$; $\mu = 0{,}1$; $\mu < \mu_0$, $\sigma = 0{,}15$

Hladina významnosti testu: $\alpha = 0{,}05 \rightarrow 1\text{-}\alpha = 0{,}95$; $u_{1-\alpha} = 1{,}645$. Kritický obor pre ľavostranný test $\mu < \mu_0$:

$W_\alpha = (-\infty; u_\alpha) = (-\infty; u_{0,05}) = (-\infty; -1{,}645)$

Veľkosť účinku: **$\Delta = \mu - \mu_0 = -0{,}2$**

Výpočet rozsahu výberu na potvrdenie alebo vyvrátenie H_0 podľa **[VII.7.]**:

$$n = \left[\left(u_{1-\alpha} + u_{1-\beta}\right).\frac{\sigma}{\mu - \mu_0}\right]^2 = \left[(1{,}645 + 1{,}645).\frac{0{,}15}{-0{,}2}\right]^2 \cong 6$$

Sila testu 1-β: $\alpha = 0{,}05 \rightarrow u_\alpha = -1{,}645$ (z tabuliek, alebo z EXCELU funkciou **NORMSINV(α)**). Kvantil distribučnej funkcie $\mathbf{u_{1-\beta}}$ dostaneme zo vzťahu:

$$u_{1-\beta} = u_\alpha \pm \frac{\mu - \mu_0}{\sigma}.\sqrt{n} \qquad \textbf{[VII.8]}$$

Znamienko „+“ je pre pravostrannú hypotézu „-„ pre ľavostrannú. Pre našu ľavostrannú hypotézu dostaneme

$$u_{1-\beta} = u_\alpha - \frac{\mu - \mu_0}{\sigma}.\sqrt{n} = -1{,}645 - \frac{-0{,}2}{0{,}15}.\sqrt{6} = 1{,}621133$$

Silu testu **1-β** potom dostaneme opäť z tabuliek N(0;1), alebo v EXCELI funkciou **NORMSDIST($u_{1-\beta}$)**:

$1 - \beta = 0{,}948$ (v bode μ)

Zhrňme si to:

Zo vstupných parametrov sociálneho programu sme dostali 3 údaje dôležité už pri jeho plánovaní:

Minimálny rozsah výberového súboru na zamietnutie alebo nezamietnutie nulovej hypotézy, ktorý by sa sledoval v ďalšom programe

n = 6

Kritický obor pre ľavostranný test $\mu < \mu_0$:

$W_\alpha = (-\infty; u_\alpha) = (-\infty; -1{,}645)$

Testovacie štatistické kritérium, ktoré by sa vypočítalo z terénnych dát, by bolo porovnávané z kritickým oborom na zamietnutie H_0 (ak testovacia štatistika padla do kritického oboru).

Sila testu:

$1 - \beta = 0{,}948$

V prípade zamietnutia H_0 sa musí ešte urobiť sila testu, aká je teda pravdepodobnosť chyby II. druhu (zamietnutie H_0 aj keď je pravdivá). Prvý odhad sily testu hovorí, že pravdepodobnosť zamietnutia H_0 aj keby bola pravdivá je len asi 5% (presnejšie $\beta = 0{,}052$), teda v 1 prípade z 20, čo je veľmi uspokojivé a môžeme začať pracovať na projekte.

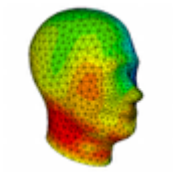
Bolo to ťažké? „Samozrejme, že bolo!“ – počujeme vašu odpoveď – „takmer vôbec nevieme o čo ide!“

Máte pravdu. Je to ťažké. Ale aj to je normálne. Nemáte na to zatiaľ vytvorené v mysli žiadne obvody, a ak vás to trochu upokojí, problematika sily testu býva často obchádzaná a zanedbávaná aj v mnohých učebniciach, prednáškach a štatistikách. Mnohí tomu jednoducho nerozumejú, alebo sa im tým nechcú zaoberať. Maximálne vám povedia, že sa to dá vypočítať nejakým profesionálnym softvérom, odsunú to ďaleko do budúcnosti, možno aj preto, lebo vám nedôverujú a ak by ste to raz v budúcnosti potrebovali, nech vás to niekto iný naučí.

Pritom v celom príklade VII.2. bol oproti predchádzajúcemu textu len jeden jediný nový vzorec, ktorý je potrebné poznať a to **[VII.8]**. (Pre nadanejších by sa to dalo možno pochopiť aj na základe toho, čo sme si povedali v predchádzajúcich kapitolách o rozdelení pravdepodobnosti, o funkciách hustoty pravdepodobnosti, intervaloch a pod., ale to bolo už dávno, že?). Neuvádzame ho pre študentov, aby sa to všetko museli naučiť naspamäť a zbláznili sa, ale aby to mali kde vyhľadať, keby to potrebovali. A kedy by to tak asi mohli potrebovať? Keď budú hrať so štatistikou poctivú hru a hlavne, keď bude v ich analýze H_0 zamietnutá. To je totiž zastavenie sa v polovici cesty, aj keď na takýto prístup natrafíte vo veľkej väčšine prípadov.

Predstavme si, že nulovú hypotézu H_0 z predchádzajúceho príkladu na hladine $\alpha = 0{,}05$ po vykonaní zberu dát a dôkladnom testovaní zamietneme v prospech alternatívnej hypotézy H_1. Zdá sa všetko v poriadku, námaha a financie investované do krízových centier sú naozaj výborný projekt. Ale ak by náhodou z výpočtu sily testu vyšlo, že v ňom je pravdepodobnosť zamietnutia H_0 aj keď je pravdivá napr. 0,8 (t.j. 8 z 10, sila testu $1-\beta=0{,}2$

a $\beta = 0,8$), môžete celý test aj výskum hodiť do koša, pretože prakticky nemá význam, ani žiadnu výpovednú hodnotu. Dobrý deň!



Hypotéz je veľa, štatistické hypotézy majú v nich však predsa len osobitné postavenie, pretože ako bolo naznačené, musia mať isté vlastnosti. Musí to byť nejaký jednoduchý výrok obsahujúci náhodné premenné, ktorých zberom a porovnaním sa dá hypotéza jednoznačne zamietnuť alebo nezamietnuť. Väčšinou stav predstavujúci predchádzajúce status quo pred nejakým zásahom položíme ako nulovú hypotézu, ktorú by sme testovaním najradšej zamietli. Takto na poli bioštatistiky sledujeme napr. účinky liekov, intervencií, liečebných postupov, rôznych diagnostických techník a metodík a pod., preto má v bioštatistike testovanie hypotéz taký veľký význam. Zlé jazyky tvrdia, s prípustnou mierou preháňania, že tým, že sa bioštatistiky zmocnili hlavne zdravotnícki pracovníci, stala sa dosť nezrozumiteľnou ba až „vedeckou“, ale napriek tomu užitočnou interdisciplinárnou činnosťou, ktorej výsledky možno využiť aj v humanitných odboroch. Štatistiky ako prevalencia, úmrtnosť a pod. sa môžu pohybovať aj na styčných hraniciach, napr. štatistika úmrtnosti v súvislosti s drogami. Testovanie hypotéz je jedným z nástrojov, ktoré je užitočné aplikovať aj mimo zdravotníctva a biologických odborov.

Možno nám v bioštatistike chýba trochu nadhľad nad medicínou. Dokáže dobre spracovať dáta z jednotlivých úzko zameraných problematík, otestovať rôzne podsystémy, zistiť senzitivitu, hladinu významnosti a špecifickosť rôznych medicínskych testov a vyšetrení, teda nájsť rozsah ich použitia, ale aj obmedzenia, a to dosť významné, ktoré sú ich neoddeliteľnou vlastnosťou, či sa nám to páči alebo nie. Spomeňte si na Bayesa v jednej z predchádzajúcich kapitol, kde sme sa mohli dopočítať k pravdepodobnosti pozitívneho výskytu niečoho nedobrého u pacienta pri veľmi slušnom diagnostickom teste len na úrovni tesne nad 30%! Napriek tomu sa mnohí (štatistická väčšina? Nemáme to overené, otestované, bolo by to neandertálske tvrdenie) inak veľmi obetaví nasledovníci Hippokrata v bielom plášti s výsledkom takéhoto testu na 95% hladine významnosti, sa tvária často bohorovne. To je obraz dnešnej medicíny, pracujúcej na úrovni náboženskej viery v jej všemocnosť. Aby nedošlo k omylu, nechceme tu predkladať anarchistické popierania významu medicíny, ako to robia rôzni „alternatívni liečitelia“ najčastejšie s nie úplne čistými úmyslami. Ale je jasné, že ľudský organizmus ako dynamický systém vyvíjajúci sa v čase a v neustálej interakcii so svojim okolím je stvorený tak, že má mnoho spätných väzieb, vracajúcich väčšinu odchýlok samovoľne do rovnovážnej polohy (príroda lieči, lekár uzdravuje), ale aj svoje hranice, kde medicína pomôcť jednoducho nemôže. Bioštatistika ako nadhľad, by mohla viesť k návratu

k prirodzenej úcte a skromnosti medicíny k Božiemu dielu, ale aj ku skromnosti pacientov v ich očakávaní a väčšej starostlivosti a zodpovednosti samých za seba. Mohla by pomôcť očistiť samu seba od množstva neandertálskych účelových štatistík, ktoré sa za ňou skrývajú. Mohla by dať odpoveď na niektoré zaujímavé otázky v mnohých oblastiach, napr.:

Bez falošných nádejí predložiť reálne možnosti medicíny (a zdravotníctva ako dosť postihnutého systému) v riešení rôznych medicínskych problémov.

Inštrumentalizácia vyšetrenia i liečby a odcudzenie pacienta ako negatívny faktor liečebného procesu pri zachovaní priority klinického posúdenia.

Prirodzené možnosti pacienta pri zodpovednom prístupe samého k sebe.

Obrovské preliekovanie populácie, odhalenie systematických odchýlok od normálu v zdravotníckom systéme vplyvom aktivít farmaceutických spoločností. Užitočnosť farmácie a farmakológie vs. ekonomické záujmy farmaceutických firiem. PRIMUM NON NOCERE.

Kvantifikovať pozitívne a negatívne ekonomické, politické a sociálne vplyvy na výkonnosť zdravotníctva.

Bez nejakej túžby po zriaďovaní ďalších úradov by bioštatistika mohla podnietiť vznik komisie s vysokou právomocou, ktorá by mala možnosť postihovať nekalú a klamlivú reklamu orientovanú na zdravie a zdravotníctvo, ktorá už priamo ohrozuje ľudí. Komisia by podľa možnosti mala byť bez zástupcov farmaceutických firiem.

Uvedomiť si mieru niekedy až prekvapujúcej úspešnosti rôznych alternatívnych liečiteľov i „liečiteľov“ ako reálny štatistický jav.

Mnohí asi nájdu ešte ďalšie a dôležitejšie strategické oblasti medicíny a zdravotníctva, ktoré by sa mohla pokúsiť bioštatistika vyriešiť, ale vráťme sa od ódy na radosť z bioštatistiky na zem. Oddýchnime si ďalším príkladom:

Pr.VII.3.: Istá špičková psychiatrická klinika sa rozhodla vykonať štúdiu účinku série vhodných psychoterapeutických sedení s ťažkou depresiou postihnutými študentmi humanitných odborov pred skúškou zo štatistiky v pokojnom prostredí, kde zaručovala, že v okruhu 10 km od kliniky sa nebude žiaden štatistik počas celého programu vyskytovať. Veľkosť účinku (ES = Effect size) kvantifikovala pomocou vhodného dotazníka (napr. HAS, MADRS, alebo vlastným – odborníci vedia o čo ide, ostaným je bližšia informácia teraz zbytočná) pred a po sedeniach, pričom veľkosť skóre z dotazníkov má normálne rozdelenie. Vhodnosť postupu na zníženie depresie, teda účinok, podrobila testu štatistickej hypotézy na vysokej hladine významnosti $\alpha = 0,05$. Pokúsme sa spolu s odborníkmi zahĺbiť do projektu a spolupracovať na tvorbe designu klinickej štúdie:

Aj keď je dostatočná veľkosť účinku **ES** prioritne záležitosťou klinickou a nie len štatistickou, zaviedli niektorí autori [10], [11] na prijateľnú orientáciu nasledujúce orientačné hladiny, ktoré je potrebné spresniť pre konkrétnu úlohu:

ES	Hodnota
Nízka	**0,00 – 0,20**
Stredná	**0,21 – 0,50**
Vysoká	**0,51 – 0,75**

Tab. VII.4.: Hladiny ES podľa [10], [11]

Veľkosť účinku ES si definovali nasledovne:

$$ES = \frac{\mu_1 - \mu_0}{\sigma_{\mu_0}} \qquad [VII.9]$$

μ_0 je stredná hodnota (aritmetický priemer) pred aplikáciou psychoterapeutického programu a μ_1 po aplikácii: $\Delta = \mu_1 - \mu_0$

H_0: Program psychoterapeutických sedení má štatisticky minimálny vplyv na liečbu hlbokej depresie študentov napríklad takej politológie pred skúškou zo štatistiky. $\mu_1 = \mu_0$ na hladine významnosti testu $\alpha = 0,05$; $\beta = 0,20$; $\sigma_{\mu_0} = 50$; $\mu_0 = 45$

H_1: Program psychoterapeutických sedení má štatisticky významný vplyv na liečbu hlbokej depresie študentov politológie pred skúškou zo štatistiky. $\mu_1 < \mu_0$

Rozsah výberového súboru pre klinickú štúdiu podľa **[VII.7.]** bude:

$$n = \left[\left(u_{1-\alpha} + u_{1-\beta}\right).\frac{\sigma}{\bar{x} - \mu_0}\right]^2$$

Pre častejšie používanie kvantilov rozdelenia N(0;1) si ich ešte raz uveď v tabuľke VII.5:

α	u(1-α/2)	u(1-α)	β	u(1-β)
0,001	3,290527	3,090232	0,05	1,644854
0,005	2,807034	2,575829	0,10	1,281552
0,01	2,575829	2,326348	0,15	1,036433
0,05	1,959964	1,644854	0,20	0,841621
0,1	1,644854	1,281552	0,25	0,67449

Tab.: VII.5.: Najčastejšie používané tabuľkové hodnoty kvantilov N(0,1).

Potom rozsah súboru pre testovanie nulovej hypotézy voči ľavostrannej hypotéze $\mu \leq 27$ dostaneme

$$n = [(u_{1-\alpha} + u_{1-\beta}).\frac{\sigma}{\bar{x} - \mu_0}]^2 = [(1{,}645 + 0{,}84).\frac{50}{18}]^2 \cong 48$$

Kritický obor pre ľavostranný test $\mu < \mu_0$:

$W_\alpha = (-\infty; u_\alpha) = (-\infty; -1{,}645)$

Pracovníci kliniky majú pripravené všetko, môžu začať pracovať so študentmi, na konci im dať HAS dotazník a jeho výsledky spracovať. Dokončiť testovanie štatistickej hypotézy a v prípade, že budú môcť zamietnuť H_0, tak pri akceptovateľnej sile testu radostných študentov môžu poslať na opravný termín zo štatistiky a svoju štúdiu publikovať v odbornom časopise.

Ak sa vám zdá, že ste to nejako zvládli a že život je vlastne dosť fádny, teraz to príde! Máme pre čitateľov (nielen pre študentov) úžasnú správu: Aj mnoho iných velikánov vedy okrem Járy Cimrmana chcelo zanechať po sebe stopu a k tomu im najlepšie poslúžila štatistika ako mladá veda, v systematike ešte občas trochu trpiaca detskými chorobami. Dalo by sa takmer zvolať: Čo velikán, to nový test! A máme ich takmer bezbrehé množstvo, prakticky na všetky potrebné a mysliteľné, ba často aj nemysliteľné príležitosti. Dôsledkom toho bývajú mnohé konferencie, pracovné konflikty, oponentské diskusie a tvorivé stretnutia štatistikov neočakávane živé. Je to naozaj otvorené bádateľské pole pre mladých začínajúcich výskumníkov a vedcov na realizáciu svojich intelektuálnych výbojov, ktoré môžu dokonca skončiť, ktohovie, aj ďalším novým, doteraz neznámym testom nazvaným podľa jeho autora. A tak sa stať úspešným, slávnym a nesmrteľným. Nám však prislúcha v tejto chvíli skromnejšia úloha urobiť si aspoň elementárnu systematickú prechádzku zeleným hájom testov štatistických hypotéz.

Asi najschodnejšie je rozdelenie testov podľa obsahu:

- testy dobrej zhody
- testy priemerov
- testy rozptylov
- špeciálne testy (test náhodnosti, nezávislosti, extrémnych odchýlok, a i.

Takmer vo všetkých predchádzajúcich príkladoch testovania sme mlčky alebo aj nahlas predpokladali, že vychádzame z nejakého známeho, najlepšie normálneho rozdelenia. Ak by to nebola pravda, mohli sme sa dopustiť aj závažných chýb vo výpočtoch. V matematike a teda aj v štatistike je niečo iné si len tak povedať a niečo iné si to dokázať alebo aspoň spočítať. Testy dobrej zhody slúžia práve na to, aby v prípade, že si to problém vyžaduje, bolo možné už kvalifikovane povedať, že súbor dát podlieha napr. rovnomernému, exponenciálnemu alebo normálnemu rozdeleniu.

a) Pearsonov parametrický χ^2- test dobrej zhody.

Je veľmi používaný a jednoduchý. Hodnoty pozorovaní zaradíme do **k** tried početnosti: $\mathbf{f_{1p}, f_{2p}, ..., f_{kp}}$. Porovnáme ich s početnosťami, ktoré by zodpovedali teoretickému rozdeleniu (nemusí to byť nevyhnutne len normálne rozdelenie $N(\mu;\sigma^2)$: $\mathbf{f_{1t}, f_{2t}, ..., f_{kt}}$. Testovanou hypotézou H_0 bude hypotéza o zhode medzi pozorovaným a teoretickým rozdelením. Ako testovacie kritérium použijeme štatistiku

$$\chi 2 = \sum_{i=1}^{k} \frac{(f_{i,p} - f_{i,t})^2}{f_{i,t}} \qquad \textbf{[VII.10]}$$

Štatistika χ^2 v prípade platnosti H_0 má χ^2- rozdelenie s (k-1) stupňami voľnosti. H_0 zamietame na hladine významnosti α, ak vypočítaná hodnota $\chi^2 > \chi^2_\alpha(k-1)$, ktorú nájdeme v tabuľkách, pričom $\mathbf{f_{it}}$ v každom intervale musí byť $\mathbf{f_{it} \geq 5}$.

Pr.VII. 4.[12]: Skupinka dopravných policajtov, čerstvých absolventov matfyzu, v rámci boja proti alkoholu za volantom dostala za úlohu zistiť, či víkendové (piatok až nedeľa) dopravné nehody pod vplyvom alkoholu významne prevyšujú rovnaké dopravné nehody v pracovných dňoch týždňa (pondelok až štvrtok). Dostali na to z ministerstva čas 10 mesiacov, nad čím sa pousmiali a ihneď vyhľadali záznamy nehodovosti za posledný rok. Zistili priemerné počty nehôd pod vplyvom alkoholu v rôzne dni týždňa nasledovne:

deň	po	ut	st	št	pi	so	ne	spolu
f_{ip}	15	11	14	9	17	27	26	119

Rozhodli sa pracovať na dosť vysokej hladine významnosti testu $\alpha = 0,05$.

H_0: Rozdelenie nehodovosti pod vplyvom alkoholu v týždni je rovnomerné. Potom teoretické početnosti f_{1t} až $f_{7t} = 119/7 = 17$.

H_1: Rozdelenie nehodovosti pod vplyvom alkoholu v týždni nie je rovnomerné.

Zostrojili tabuľku z pozorovaných, teoretických a vypočítaných hodnôt ($\Delta = f_{i,p} - f_{i,t}$):

deň	$f_{i,p}$	$f_{i,t}$	Δ^2	$\Delta^2 / f_{i,t}$
po	15	17	4	0,235294
ut	11	17	36	2,117647
st	14	17	9	0,529412
št	9	17	64	3,764706
pi	17	17	0	0
so	27	17	100	5,882353
ne	26	17	81	4,764706
spolu	119	119		17,29412
$\bar{x}$	17			
$\chi^2_\alpha(6)$	12,59159			

Keďže $\chi^2 > \chi^2_\alpha(k-1)$ t.j. 17,3>12,6, H_0 o rovnomernom rozdelení nehodovosti pod vplyvom alkoholu v týždni zamietli.

Pouvažovali ešte, že pri rovnomernom rozdelení by musela byť početnosť nehôd v 4 dňoch týždňa (po až št) 4/7 zo 119 → $f_{1,t} = 68$ a v 3 dňoch (pi až ne) 3/7 zo 119 teda $f_{2,t} = 51$. Skutočné hodnoty boli: $f_{1,p} = 49$ a $f_{2,p} = 70$. Keď to vložili do vzťahu **[VII.10]** aby otestovali novú H_0, že nehodovosť v týždni je rovnaká ako cez víkend proti H_1, že cez víkend je väčšia:

$$\chi 2 = \sum_{i=1}^{k} \frac{(f_{i,p} - f_{i,t})^2}{f_{i,t}} = \frac{(49-68)^2}{68} + \frac{(70-51)^2}{51} = 12,39$$

$\chi^2_\alpha(k-1) = \chi^2_{0,05}(1) = 3,84$ podľa tabuliek.

Keďže $\chi^2 > \chi^2_\alpha(k-1)$ t.j. 12,39 >3,84, H_0 o rovnomernom rozdelení nehodovosti pod vplyvom alkoholu v týždni zamietli v prospech H_1, že cez víkend je výrazne vyššia. 9 mesiacov a 29 dní nemuseli už robiť výskum a mohli sa venovať ochrane vodičov a plynulosti dopravy.

Hodnoty kvantilov χ^2- rozdelenia pre rôzne α a rôzne stupne voľnosti možno dostať aj v EXCELI cez funkciu **CHIINV(α; k-1)**. Celý test je možné jednoducho v EXCELI realizovať tak, že si pozorované hodnoty dáme do jedného stĺpca (napr. A1 až A7) a vedľa teoretické hodnoty (B1 až B7), použitím funkcie **CHITEST**(**actual_range**;**expected_range**), kde za **actual_range** vložíme myšou rozsah pozorovaných hodnôt (A1 až A7 v našom prípade) a za **expected_range** zase rozsah očakávaných teoretických hodnôt (B1 až B7). Vráti nám do nového okienka hodnotu pravdepodobnosti χ^2- rozdelenia. Potom použijeme funkciu **CHIINV(p; k-1)** (v našom ilustračnom prípade **CHIINV(vypočítaná hodnota cez CHITEST; 6)**) a dostaneme výslednú hodnotu testovanej štatistiky, ktorú už len porovnáme s tabuľkovou hodnotou získanou cez **CHIINV(α; k-1)** v našom prípade **CHIINV(0,05; 6)**.

b) Kolmogorovov-Smirnovov test dobrej zhody

V častých prípadoch, kedy nemáme pozorované hodnoty rozdelené do tried, alebo ide o malé rozsahy súborov, resp. v triede je početnosť menšia ako 5 používame radšej Kolmogorovov-Smirnovov jednovýberový resp. dvojvýberový test. (K-S test). Pri jednovýberovom K-S teste (napr. zhody s normálnym rozdelením) máme **n** pozorovaní nejakej náhodnej premennej x_i, usporiadané podľa veľkosti do variačného radu. Vypočítame aritmetický priemer $\bar{x}$ a smerodajnú odchýlku σ. Vypočítame ku každej i-tej hodnote podiel **i/n** a hodnotu distribučnej funkcie $F(x_i)$ normálneho rozdelenia, najlepšie pomocou EXCELU a jeho funkcie **NORMDIST(xi, $\bar{x}$; σ;1)** a rozdiel

$$D_i = \left| F(x_i) - \frac{i}{n} \right| \qquad [VII.11]$$

Najväčšiu hodnotu z nich **max(D_i)** porovnáme s tabuľkovými kritickými hodnotami K-S testu $D_{\alpha,n}$. Ak **max(D_i) > $D_{\alpha,n}$** tak H_0 o zhode výberového súboru s normálnym rozdelením zamietame.

Dvojvýberový K-S test posudzuje zhodu rozdelení 2 súborov resp., že 2 výberové súbory pochádzajú z toho istého základného súboru. Ako testovacie kritérium sa v tomto prípade použije štatistika

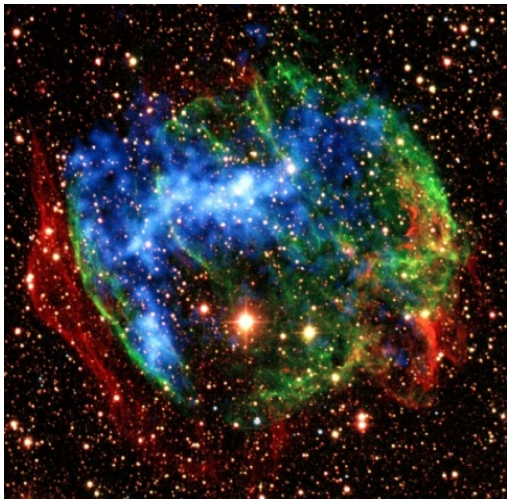
$$D_i = \left| F1(x_{1,i}) - F2(x_{2,i}) \right| \qquad [VII.12]$$

Maximálna hodnota z rozdielov hodnôt distribučnej funkcie 1. a 2.súboru sa porovnáva s tabuľkovou kritickou hodnotou ako v predchádzajúcom prípade. Tabuľky možno nájsť aj na webe napr. [13]. Pre väčšie rozsahy súboru (n>30) kritickú hodnotu možno odhadnúť zo vzťahu:

$$D_{\alpha,n} \cong \sqrt{\frac{1}{2n}.\ln\frac{2}{\alpha}} \qquad [VII.13]$$

Je to dosť jednoduché, ale jednoduchý príklad to najlepšie vysvetlí:

Pr.VII. 5.: *(Pocta Ray Bradburymu a Arthurovi C.Clarkovi, ilustračné foto SITA,AP).*



Umelý ostrov visel nad treťou celkom sympatickou planétou inak nudného hviezdneho systému už niekoľko tisíc jej obehov okolo hviezdy, ale stále sa nevedeli rozhodnúť, či sa na nej nachádza inteligentný život, ba život vôbec. Zakotvili v bode, kde ich nebolo možné zamerať ani nijako spozorovať, aj keby na planéte nejaká vyspelá a technicky zdatná civilizácia bola. Sledovali niektoré úkazy, ako primitívne autodeštrukčné štiepenia atómov a iné premeny energie, odstredivé dráhy niektorých telies alebo intenzívny elektromagnetický smog, ale všetko to mohlo byť len prejavom slepej náhody, nevýznamnými odchýlkami v bežnom usporiadaní subatomárnych častíc pri daných podmienkach. Nakoniec dostali súhlas Najvyššej medzigalaktickej rady na výnimku z etického kódexu a urobili náhodný odber 10 ks zvláštnych dvojnohých a dvojrukých predmetov prevažne na báze kyslíka, vodíka a uhlíka. Všetky techniky a procesy nadviazania kontaktu však úplne zlyhávali, až si niekto spomenul na prastarý test na určovanie stupňa inteligencie nazvaný IQ, používaný kedysi na sledovanie štruktúr omnoho primitívnejších ako nerasty, kamene a kryštály v ich materskom kúte Univerza. Prieskumníci si určili nulovú hypotézu, že ak akokoľvek nízka hladina inteligencie odobraných vzoriek bude podliehať normálnemu rozdeleniu, budú sa planéte ešte nejaký čas venovať, preto zvolili jednovýberový K-S test s výsledkami v tabuľke premietnutej v temnom priestore s mrazivým pozadím horúcich hviezd:

i	x_i [IQ]	i/10	$F(x_i)$	$D= \lvert Fx_i - i/10 \rvert$
1	108	0,1	0,288822	0,188822
2	109	0,2	0,348973	0,148973
3	109	0,3	0,348973	0,048973
4	109	0,4	0,348973	0,051027
5	109	0,5	0,348973	0,151027
6	110	0,6	0,413186	0,186814
7	110	0,7	0,413186	0,286814
8	110	0,8	0,413186	0,386814
9	111	0,9	0,479814	0,420186
10	128	1	0,997583	0,002417
$\bar{x}$	111,3	max D		0,420186
σ	5,9264	D(0,05;10)		0,41

Ako to dali dokopy? V 1.stĺpci je číslo testu IQ po usporiadaní podľa veľkosti. V druhom stĺpci sú zistené hodnoty IQ testu desiatich z planéty náhodne odobratých entít a pod tým ich aritmetický priemer a smerodajná odchýlka. 3.stĺpec je jednoducho vydelené poradové číslo **i** riadku počtom pozorovaní, teda 10, to predstavuje teoretické normálne normované rozdelenie. 4.stlpec dostali pomocou najnovšej medzigalaktickej verzie MS EXCEL a jej funkcie **NORMDIST(x_i, $\bar{x}$; σ;1)**. V poslednom stĺpci urobili absolútnu hodnotu rozdielu **$D=\lvert F(x_i)-i/10 \rvert$**.Maximálna hodnota zo všetkých D je posudzovaná hodnota štatistiky pre test. V poslednom riadku vpravo je tabuľková kritická hodnota testu **D(0,05;10).**

Pretože **max(D) > $D_{\alpha,n}$** tak H_0 o zhode výberového súboru s normálnym rozdelením na hladine významnosti testu α=0,05 zamietli a trochu rozladení, že nenašli ani náznak inteligentného života na planéte ani v celom hviezdnom systéme s nízkou prognózou do budúcnosti, sa odpútali a odleteli do iných zaujímavých kútov Vesmíru. Nebolo tu čo a hlavne komu ponúknuť.

Stretnutie najvyšších predstaviteľov veľmocí a vplyvných krajín Zeme sa dlho nieslo v štandardnom konfrontačnom duchu presadzovania vlastných záujmov. Niekoľkokrát hrozilo jeho konfliktné prerušenie, ale to by nebol dobrý signál svojim voličom. Preto sa nakoniec dohodli na vyhlásení, ktoré malo odpútať pozornosť sveta od reálnych problémov. Bolo konštatované, že po neobyčajnom úsilí a tvorivom riadení najmodernejšej vedy a celej spoločnosti sa pravdepodobne podarilo nadviazať spojenie s mimozemskou civilizáciou, aj keď to vedci ešte zďaleka jednoznačne nepotvrdili. Zato však politici vyjadrili presvedčenie, že pravdepodobní návštevníci s rešpektom a obdivom vnímali naše pokroky v oblasti vedy,

ekonomiky, priemyslu a kultúry, nevynímajúc našu skvelú štatistiku, ktoré budú nepochybne obohatením oboch strán pri blížiacich sa kontaktoch.

Testami priemeru sme sa zaoberali aspoň ilustračne na začiatku kapitoly. Boli to testy parametrov stredných hodnôt jedného súboru (Pr.VII.1.) a dvoch súborov (Jára Cimrman), nazvané v štatistickom žargóne ako ***u-testy***, v ktorých je testovacia štatistika $\mathbf{u_p}$ vzťahovaná na normované normálne rozdelenie N(0;1). Častejšie sa vyskytujú ***t-testy***, kde štatistikou je na prvý pohľad rovnaké kritérium

$$t_p = \frac{\bar{x} - \mu_0}{\sigma} \cdot \sqrt{n}$$ pre jeden súbor, resp.

$$t_p = \frac{\mu_1 - \mu_0}{\sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ resp. pre stredné hodnoty 2 súborov **[VII.14]**

taktiež obojstranné alebo jednostranné (pravo-, ľavostranné) a kritická hodnota $t_p(n-1)$ sa vyhľadá z tabuliek Studentovho rozdelenia. Používa sa predovšetkým v prípade menších rozsahov výberov a pri neznámych rozptyloch.

Pr.VII.6. (Jednovýberový t-test): Veľký boss nebol spokojný hlavne s jeho dlhoročným dílerom *Degešom*. Vedel, bola to zdedená múdrosť otcov, že po čase musí dílerov obmeniť, pretože sa dostávajú do prílišnej rutiny a začínajú špekulovať. Zdalo sa mu, že zisky z *Degešových* obchodov klesajú, ale chcel si to najprv overiť. Dobrí ľudia mu doniesli informácie, ako jazdí v novučičkom drahom aute po pešej zóne, provokuje policajtov a v baroch strieľa po ľuďoch, čo sa mu z akéhokoľvek dôvodu znepáčili; a je k svojim zákazníkom arogantný. To nebolo dobré pre biznis, preto si vzal *Degešove* výsledky posledných dvoch týždňov, aby sa na ne pozrel odborným okom skúseného manažera a štatistika, a preveril, či jeho odovzdané tržby poklesli štatisticky významne oproti požadovaným 10 000.- € za deň. Náročné výpočty pre dôležité rozhodnutia nemohol zveriť ani svojim najbližším a najoddanejším pracovníkom, pretože tí nedokázali spočítať, ani koľko ľudí dnes odkrágľovali. Použil len bežnú hladinu významnosti $\alpha = 0{,}05$; veď išlo o život. Denná tržba *Degeša* za predaj drog v 2 týždňoch v € bola:
8500, 1100, 9400, 7600, 8600,10900, 7400, 7000,7400, 11300, 8800, 8600, 8500 a 8200.
Nalial si slušnú dávku brandy, zapálil voňavú cigaru a schuti sa pustil do práce. Z údajov o tržbe vyrátal aritmetický priemer $\bar{x} = 8800$ s výberovou smerodajnou odchýlkou s = 1436, pretože nepoznal rozptyl σ^2. Nulová hypotéza, ktorú sa rozhodol testovať, bola pre neho

H_0: $\mu = \mu_0$ (*Degeš* ho neokráda) oproti alternatívnej hypotéze H_1: $\mu < \mu_0$ (*Degeš* ho okráda dosť významne). Zvolil si testovaciu štatistiku **[VII.14]**

$$t = \frac{\bar{x} - \mu_0}{s} . \sqrt{n}$$

so **Studentovým t-rozdelením** s (n-1) stupňami voľnosti. Kritická hodnota Studentovho rozdelenia pre $\alpha = 0{,}05$ a počet stupňov voľnosti n - 1 = 13 je $t_{\alpha,n-1} = 1{,}771$ (použil tabuľky, pre istotu si to porovnal s EXCELOM, keď mu druhá karafa brandy prečistila mozog a spomenul si, že EXCEL počíta vždy len obojstrannú hypotézu t.j. s polovinou toho, čo mu vložil $\alpha/2$, tak pre ľavostrannú hypotézu $\mu < \mu_0$ musel do príslušnej funkcie **TINV(α, n-1)** vložiť za α dvojnásobnú hodnotu α=0,1 inak by mu EXCEL počítal len s α=0,025). Kritický obor je pomocou tabuliek a EXCELU

$$\mathbf{W_\alpha = (-\infty, -t_{\alpha,(n-1)}) = (-\infty, -t_{0,05;13}) = (-\infty, -1{,}771)}$$

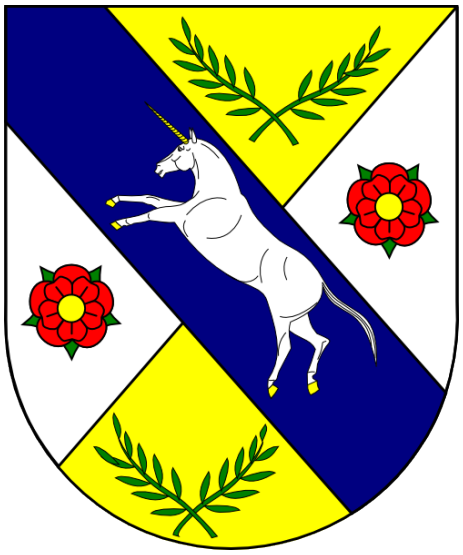
Vypočítal

$$t = \frac{\bar{x} - \mu_0}{s} . \sqrt{n} = \frac{8800 - 10000}{1436} . \sqrt{14} \cong -3{,}130$$

Vedel, že nulovú hypotézu H_0 môže zamietnuť na hladine významnosti testu α **= 0,05** vtedy, keď $|t| > t_{\alpha,n-1} \rightarrow \left|\frac{\bar{x}-\mu_0}{s} . \sqrt{n}\right| > t_{\alpha,n-1}$, resp. keď hodnota testovacieho kritéria padne do kritického oboru, čiže keď $t \in W_\alpha$. A padla.

Zabafkal z cigary, uškrnul sa nepríjemným úškľabkom, z ktorého behal mráz po chrbte, zdvihol telefón a stručne prikázal: „Namiešajte kvalitný betón!“

Pr.VII.7. (Dvojvýberový t-test pre stredné hodnoty): Bola to zvláštna, zapadajúcim slnkom do červena podfarbená scéna. Na kopci nad Dunajom v podvečernom napätom tichu boli zoradené jednotky princa Eugena, ktoré len ťažko zakrývali únavu po ťažkom a dlhom prechode kopcov a močiarov popri rieke. Pod kopcom kľačala elitná jednotka tureckých janičiarov, ich odvážny a hrdý veliteľ Turčín Poničan na znak priateľských úmyslov stál pred princovým koňom s vystretými rukami a podával mu svoj zahnutý ozdobený meč. Nebola to pasca, úskok? Ale princa Eugena Savojského upokojil jeho verný pobočník na bojových výpravách, Michael Olahus z významnej vetvy duchovnej šľachty, ktorej predstaviteľom bol člen protiosmanskej ligy a ostrihomský arcibiskup Nicolaus III. Oláh, ktorý vedel, že mladý veliteľ tureckej elitnej gardy náhodou stretol v Hornom Uhorsku svoju matku a spoznal svoje rodné korene i korene celej družiny.



Princ nariadil vojsku krátky tichý odpočinok a zvolal bojovú poradu. Od Turčína Poničana sa dozvedeli aktuálnu situáciu, ale aj neutešený pomer síl: Princ mal necelých 50 000 dosť vyčerpaných mužov, zatiaľ čo sultán Mustafa II. ich mal viac ako 100 000. A začalo byť jasné, prečo sultán napriek výhodnému pomeru síl ustupuje: Chcel po prekročení rieky Tisy zaujať za jej mostom ešte výhodnejšie, prakticky neporaziteľné postavenie. Princ, jeho generáli i janičiari na seba bezradne pozreli, len Michael Olahus v kúte bojového stanu za blikotania fakieľ niečo usilovne počítal. Bola noc 10. septembra 1697.

Mal dva súbory: Vlastnú armádu, v ktorej musel odčítať ranených a úplne vyčerpaných, a po porade s 20 dôstojníkmi získal odhad, teda $\bar{x}_1$ = **40000** mužov s odhadom smerodajnej odchýlky s_1 = **±24500** a osmanské vojsko Mustafu II., ktoré po násilnom zverbovaní obyvateľstva a nasadení otrokov mohlo mať rozsah podľa vyjadrenia 10 janičiarov $\bar{x}_2$ = **120000** mužov so smerodajnou odchýlkou s_2 = **±80000**. Zapísal si n_1= **20** (20 dôstojníkov, ktorí mu poskytli číselný odhad počtu mužov) a n_2 = **10** (janičiarov). Stret na otvorenom poli by bol nezmyslom. Čo by sa však stalo, keby nepriateľa napadli v polovici presunu cez rieku, aké by mali šance? Preložené do štatistického jazyka, dalo by sa prijať H_0, že by sa stredné hodnoty súborov počtu bojovníkov vyrovnali? Teda

$H_0 : \mu_1 = \mu_2$ oproti **$H_1 : \mu_1 \neq \mu_2$; $\alpha = 0{,}05$**

Michael Olahus si opäť prepísal vstupné údaje:

$\bar{x}_1 = 40000$; $s_1 = \pm 24500$; $s_1^2 = 600250000$; $n_1 = 20$

$\bar{x}_2 = 60000$; $s_2 = \pm 40000$; $s_1^2 = 1600000000$; $n_2 = 10$

Ako testovaciu charakteristiku zvolil zo **[VII.14]**

$$t = \frac{\bar{x}_2 - \bar{x}_1}{s}$$

kde **s** je spoločná smerodajná odchýlka, ktorú dostaneme trochu namáhavejšie zo vzťahu

$$s = s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad \textbf{[VII.15]}$$

kde

$$s_p = \sqrt{\frac{(n_1 - 1).s_1^2 + (n_2 - 1).s_2^2}{n_1 + n_2 - 2}} \qquad \textbf{[VII.16]}$$

Po trochu ťažkopádnom výpočte dostal pre **s = 11757,5** a pre testovaciu charakteristiku

t = 1,701036. Tabuľková kritická hodnota $\mathbf{t_\alpha(n_1 + n_2 - 2)}$ **= t$_{0,05}$(28) = 1,701131.** Keďže **t <** $\mathbf{t_\alpha(n_1 + n_2 - 2)}$, H_0 nebolo potrebné zamietnuť a dalo sa predpokladať, že odchýlky stredných hodnôt počtov bojovníkov v oboch zoskupeniach pri polovičnom prechode Turkov cez most na hladine významnosti $\alpha = 0,05$ sú viac-menej vyrovnané.

Princovi Eugenovi došlo, že spánok sa odkladá. K mostu cez Tisu pri málo známom meste Zenta docválali v pravú chvíľu. V ten deň zahynulo alebo sa v Tise utopilo asi 30 000 Turkov, na druhej strane padlo 500 bojovníkov a začal sa postupný úpadok Osmanskej ríše.

V predchádzajúcom príklade nemali vojská veľa času, inak je potrebné ešte pred testovaním urobiť predbežný test na štatistickú blízkosť rozptylov dvoch súborov, nazvaný **Fisherov F-test**, teda **H_0: $\sigma_1^2 = \sigma_2^2$** oproti **H_1: $\sigma_1^2 \neq \sigma_2^2$**:

H_0 sa zamieta na hladine významnosti α, ak

$$F = \frac{s_1^2}{s_2^2} > F(n_1 - 1; n_2 - 1) \qquad \textbf{[VII.17]}$$

kde prvé hodnoty (s_1^2 a $n_1$sú vždy väčšie z dvojice hodnôt) a $F(n_1 - 1; n_2 - 1)$ je tabelovaná hodnota Fisherovho rozdelenia pre 2 stupne voľnosti $(n_1 - 1; n_2 - 1)$.

Michael Olahus sa dopočítal dodatočne, že $\mathbf{F < F(n_1 - 1; n_2 - 1) = 2,947652 < 2,665556}$ a taktiež H_0 nemusel zamietnuť.

Nie bez zaujímavosti je aj testovanie rozptylov ANOVA, alebo jednoduchý párový t-test na rovnosť stredných hodnôt 2 súborov, neparametrický znamienkový test, alebo test odľahlých extrémnych hodnôt, pretože aj keď sa nám nejaká zistená hodnota vo výberovom súbore príliš nepáči, to ešte nie je dôvod, aby sme ju zo súboru vylúčili. To je skôr dôvod, že výber a výberový súbor ovplyvňujeme. Vylúčenie odľahlej hodnoty, zaťaženej veľkou systematickou chybou sa dá urobiť na základe testu (Dixonov, Grubbsov a i.).

Už viackrát som spomenul, že urobiť správny a dobrý výber je dosť veľký problém, spadajúci viac do metodológie ako do štatistiky. Ako urobiť napr. náhodný výber? Možno to skúsiť losovaním zo základného súboru tak, že prvkom základného súboru pridelíme čísla z generátora náhodných čísel a tie potom losujeme. Alebo podobným spôsobom. Problém však môže nastať, a to či generátor náhodných čísel (nejaký softvér, ktorý mám k dispozícii)

naozaj generuje náhodné čísla. Skúsme si otestovať pomocou jedovýberového K-S testu generátor náhodných čísel v EXCELI, ktorý máme k dispozícii:

Pr.VII. 8.: Urobím si v EXCELI tabuľku pre 20 hodnôt F_i rovnomerného rozdelenia (ak je to naozaj náhodný výber), ktoré som dostal funkciou generátora náhodných čísel RAND() v EXCELI. D je max.hodnota d, a $D(\alpha;n) = D(0,05;20) = 0,294$ nájdem v tabuľkách K-S kritických hodnôt.

i	i/20	F_i	$d=\lvert F_i-i/20\rvert$
1	0,05	0,083116	0,033116
2	0,1	0,088404	0,011596
3	0,15	0,112827	0,037173
4	0,2	0,130461	0,069539
5	0,25	0,165427	0,084573
6	0,3	0,172891	0,127109
7	0,35	0,189972	0,160028
8	0,4	0,221692	0,178308
9	0,45	0,260614	0,189386
10	0,5	0,263225	0,236775
11	0,55	0,351411	0,198589
12	0,6	0,485374	0,114626
13	0,65	0,637237	0,012763
14	0,7	0,699771	0,000229
15	0,75	0,702131	0,047869
16	0,8	0,736751	0,063249
17	0,85	0,74413	0,10587
18	0,9	0,793859	0,106141
19	0,95	0,795046	0,154954
20	1	0,904985	0,095015
D	**0,236775**		
D(0,05;20)	**0,294**		

Keďže **$D < D(\alpha;n)$**, nemôžem zamietnuť, že môj generátor náhodných čísel mi dáva naozaj výberový súbor náhodných čísel, nebudem sa musieť poobzerať po inom softvéri. Trochu sa pri teste vášho generátora náhodných čísel pohrajte s tým, ako vložiť náhodné číslo do 3.stĺpca, aby sa vždy opätovne neprepočítavalo a aby ste celý 3.stĺpec mohli usporiadať podľa veľkosti od najmenšieho po najväčšie a potom súbor testovať.

Dotkli sme sa možnosti EXCELU pri testovaní hypotéz. Pokiaľ viete, čo robíte, môže to byť celkom dobrý pomocník. Cez lištu kliknete na ***Údaje*** a potom vpravo hore na ***Analýza dát***, objaví sa vám opäť okno ***Analytické nástroje***. V nich sa dajú nájsť tieto užitočné možnosti:

- *Anova: jeden faktor*
- *Anova: dva faktory s opakovaním*
- *Anova: dva faktory bez opakovania*
- *Dvojvýberový F-test pre rozptyl*
- *Dvojvýberový párový t-test na strednú hodnotu*
- *Dvojvýberový t-test s rovnosťou rozptylov*
- *Dvojvýberový t-test s nerovnosťou rozptylov*
- *Dvojvýberový z-test na strednú hodnotu*

Test ANOVA (analýza rozptylov) predstavuje náročnejšie testovanie viacerých súborov, kde testy zhody stredných hodnôt sú nahradené F-testami zhody rozptylov.

Dvojvýberový F-test pre rozptyl sme si robili už na konci Pr.VII.7 pomocou vzťahu **[VII.17]**.

Musí predchádzať dvojvýberové t-testy.

Pri pozorovaní dvoch závislých kvantitatívnych znakoch X a Y, z ktorých môžeme urobiť výberový súbor **n** dvojíc (x_i, y_i), pričom X má rozdelenie $N(\mu_1; \sigma_1^2)$ a Y zase $N(\mu_2; \sigma_2^2)$, vypočítame rozdiely $d_i = x_i - y_i$, následne aritmetický priemer $\bar{d}$ a výberovú smerodajnú odchýlku **s** podľa **[V.9]**. Nulovú hypotézu H_0: $\mu_1 = \mu_2$ testujeme párovým t-testom s testovacou štatistikou

$$t = \frac{\bar{d}}{s}.\sqrt{n} \qquad [VII.17]$$

H_0 zamietame na hladine významnosti testu α ako vždy ak

$$|t| > t_\alpha(n-1)$$

kde $t_\alpha(n-1)$ je tabelovaná hodnota Studentovho rozdelenia pre (n-1) stupňov voľnosti.

Dvojvýberový t-test s rovnosťou rozptylov sme použili v pr. VII.7, pokiaľ by nám Fisherov F-test nepotvrdil rovnosť rozptylov tak musíme použiť Dvojvýberový t-test s nerovnosťou rozptylov. **z-test** je nám známy **u-test** so štatistikou **u**. Príklad v EXCELI:

Pr.VII. 9.: Výskumná skupina vplyvu hudby na správanie klientov reedukačného centra pre mladistvých s ADHD si urobila 2 náhodné výberové súbory po 10 žiakov. V prvom sa zamerala počas výskumu na upútanie pozornosti na vážnu hudbu, na náročnejšie objasnenie jej krásy a praktické ukážky. Vážnu hudbu počúvali žiaci aj na praxi a pri fyzickej práci. Činnosť druhej skupiny sprevádzala na pozadí intenzívnejšia a z hľadiska variability hudobných výrazových prostriedkov veľmi jednoduchá rocková hudba. V skúmanom období vyhodnocovali v každej skupine prejavy násilia (agresivita, šikana a pod.), ktorej sa jednotliví žiaci dopustili, alebo sa stali jej obeťami. Výsledky sú v tabuľke:

žiak č.	1	2	3	4	5	6	7	8	9	10
klasika	2	0	1	3	0	0	4	2	2	1
rock	0	7	4	2	5	6	6	6	7	2

Párový t-test v EXCELI mal otestovať nulovú hypotézu, že nie je rozdiel v prejavoch násilia žiakov vplyvom druhu počúvanej hudby, teda

H_0: $\mu_1 = \mu_2$ oproti alternatívnej hypotéze H_1: $\mu_1 \neq \mu_2$

Krok 1: Preniesli sme si tabuľku do EXCELU.

Krok 2: Cez *Analýza dát* a *Analytické nástroje* si zvolíme *Dvojvýberový párový t-test na strednú hodnotu*, klikneme na neho. Do riadku *Soubor 1* vložíme myšou rozsah buniek, v ktorých je prvý riadok údajov pre klasickú hudbu (napr. C3:L3), podobne do riadku *Soubor 2* vložíme myšou rozsah buniek, v ktorých je druhý riadok údajov pre rockovú hudbu (napr. C4:L4). *Hypotetický rozdíl středních hodnot* zvolíme **0**, pretože testujeme, že rozdiel medzi strednými hodnotami je nulový. Hladinu významnosti testu *Alfa* dáme 0,05. Do *Výstupní oblast* vložme myšou okienko, kde chceme umiestniť výstupnú tabuľku údajov a stlačme OK.

Dostali sme:

Dvouvýběrový párový t-test na střední hodnotu		
	Soubor 1	*Soubor 2*
Stř. hodnota	1,5	4,5
Rozptyl	1,833333	5,833333
Pozorování	10	10
Pears. korelace	-0,18687	
Hyp. rozdíl stř. hodnot	0	
Rozdíl	9	
t Stat	-3,18198	
P(T<=t) (1)	0,005575	
t krit (1)	1,833113	
P(T<=t) (2)	0,011149	
t krit (2)	2,262157	

Krok 3: Tabuľka je to pekná, ešte si ujasnime, čo v nej máme:

Riadok stredná hodnota je jasná: $\mu_1 = 1,5$; $\mu_2 = 4,5$

Rozptyl: $\sigma_1^2 = 1,833$ a $\sigma_2^2 = 5,833$

Pozorovanie: je počet prvkov každého súboru $n = n_1 = n_2$

Pears.korelace je hodnota Rearsonovho koeficientu korelácie, ktorá určuje stupeň závislosti medzi súborom 1 a 2. Koreláciou sa budem zaoberať čoskoro v niektorej z nasledujúcich kapitol.

Hypotetický rozdiel stredných hodnôt sme si dosadili 0.

Riadok, pre mňa nie úplne pochopiteľne nazvaný *Rozdíl* udáva počet stupňov voľnosti (n-1), pre ktorý z tabuliek vybral pri α=0,05 kritickú hodnotu $t_\alpha(n-1) = t_{0,05}(10-1) = 1,833$.

V nasledujúcom riadku *tStat* je hodnota testovacej štatistiky **t = -3,18198.**

V posledných 4 riadkoch sú dve hodnoty pravdepodobnosti P, ak by sme testovanie robili cez pravdepodobnosť, alebo kritické hodnoty t.

V riadku t krit (1) uvádza tabuľka $t_\alpha(n-1) = t_{0,05}(10-1) = 1,833$. Túto hodnotu môžeme použiť pri jednostrannom testovaní.

V riadku t krit (2) uvádza tabuľka $t_{\frac{\alpha}{2}}(n-1) = t_{0,025}(10-1) = 2,262157$. Túto hodnotu môžeme použiť pri dvojstrannom testovaní. Testujeme už sami:

Krok 4: Dvojstranný test:

H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$

Kritický obor $W_\alpha = (-\infty; -2,262) \cup (2,262; \infty)$; t = -3,182

Výsledok testu: $t \in W_\alpha$, preto H_0 zamietame v prospech H_1, že typ hudby má vplyv na agresivitu správania žiakov reedukačného centra.

Krok 5:

Ľavostranný test:

H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 < \mu_2$

Kritický obor $W_\alpha = (-\infty; -1,833)$

Výsledok testu: $t \in W_\alpha$, preto H_0 zamietame v prospech H_1, že vplyvom klasickej hudby sa prejavuje nižšia agresivita správania žiakov reedukačného centra ako pod vplyvom jednoduchej čisto rytmickej rockovej hudby.

Trvalo nám to dlho a je toho trochu veľa, ale len preto, že sme všetko do podrobnosti vysvetľovali. Ak si to prejdete ešte raz pozorne a uvedomíte si, čo všetko môžete vynechať, zistíte, že máte v rukách skvelý, účinný, stručný a rýchly nástroj na vaše testovanie. Dvojvýberovým t-testom musí predchádzať dvojvýberový F-test rovnosti rozptylov, ale všetko prebieha už viac-menej spôsobom, ako bolo uvedené v predchádzajúcom príklade. Celú krásu testovania hypotéz a radosť z nich si môže študent a čitateľ vychutnať samozrejme len po istom osvojení, teda po prepočítaní istého množstva príkladov a po použití na niektoré vlastné úlohy a riešené problémy. Nevzdať to, neprejsť popri tom bez povšimnutia, neprepásť príležitosť. Neponáhľať sa príliš. V tejto súvislosti si povedzme dávnejší príbeh, s ktorým sa

dalo stretnúť na rôznych web blogoch ako so skutočným príbehom *Huslista v metre*: Bolo chladné januárové ráno roku 2007, na stanici metra vo Washingtone D.C, hral muž na husliach necelú hodinu niekoľko diel J. S. Bacha, medziiným aj jeho husľové koncerty a-mol a E-dur. V priebehu hry prešlo stanicou viac ako 2000 ľudí cestou do práce. Niektorí si ho všimli, poniektorí sa aj na nepatrný okamih pristavili. Zanedlho huslista dostal prvý dolár. Žena bez zastavenia hodila peniaze do klobúka a pokračovala v chôdzi. Po šiestich minútach sa mladý muž oprel o stenu, aby počúval, potom sa pozrel na hodinky a začal znovu kráčať. O desať minút sa pri ňom zastavil trojročný chlapec, ale jeho matka ho náhlivo potiahla. Dieťa sa opäť zastavilo a pozeralo na huslistu, ale matka potiahla silnejšie a dieťa pridalo do kroku, po celý čas otáčajúc hlavu. Toto sa opakovalo ešte aj s niekoľkými ďalšími deťmi, ale každý rodič - bez výnimky – prinútil svoje deti ísť ďalej. Po trištvrte hodine, v ktorej hudobník hral bez prestávky bolo možné bilancovať: Iba šesť ľudí sa zastavilo, aby chvíľu počúvali. Asi dvadsať mu dalo peniaze, ale kráčali ďalej normálnym tempom. Muž vyzbieral celkovo 27 dolárov. Prestal hrať a nastalo ticho. Bez povšimnutia, bez potlesku.V rámci sociálneho experimentu inkognito huslistom bol Joshua Bell, jeden z najväčších súčasných hudobníkov na svete. Hral jedny z najzložitejších a najkrajších diel aké boli kedy napísané, a aj keď to nie je najdôležitejšie, na husliach za 3,5 milióna dolárov. Dva dni predtým Joshua Bell vypredal divadlo v Bostone, kde bola priemerná cena lístka 100 dolárov za sedenie a počúvanie rovnakej hudby, akú hral v metre...

Literatúra k VII. kapitole

[1] Smoljak, L., Svěrák, Z., Cimrman, J.: Dobytí severního pólu, divadelná hra Divadla Járy Cimrmana, Praha 1985
[2] http://magazin.atlas.sk/techmag/prvenstvo-na-severnom-pole-nevyriesena-zahada/668849.html
[3] http://cit.vfu.cz/statpotr/POTR/Teorie/Predn3/hypotezy.htm
[4] http://new.euromise.org/czech/tajne/ucebnice/html/html/node9.html
[5] ONDREJKOVIČ, P. : Úvod do metodológie spoločenskovedného výskumu, Veda, Bratislava 2007
[6] JUSZCZYK, S.: Metodológia empirického výskumu v spoločenských vedách, IRIS, Bratislava 2006.
[7] Gavora, P. a kol. 2010. Elektronická učebnica pedagogického výskumu. Univerzita Komenského, Bratislava 2010.
[8] Meloun, M., Militký, J., 1998: Statistické zpracování experimentálních dat. Praha, East Publishing,
[9] Ptáček,R., Raboch,J.: Určení rozsahu souboru a power analýza v psychiatrickém výskumu, Čes a slov Psychiatr 2010;106(1): 33 -41
[10] Cohen J. Statistical power analysis for the behavioral sciences. San Diego, CA: Academic Press; 1969.
[11] Cohen J. Statistical power analysis for the behavioral science. 2nd ed. Hillsdale, NY: Erlbaum; 1988.
[12] http://www.km.fpv.ukf.sk/upload_publikacie/20120130_90405__1.pdf
[13] http://www.kirp.chtf.stuba.sk/moodle/pluginfile.php/25812/mod_resource/content/0/tabulky1.pdf


Smiešna hypotéza, Adamko.
Kvôli jednému jablčku sa ešte
nikdy nikomu nič nestalo...

VIII. Deň závislosti alebo keď všetko spolu súvisí

V škole sa mladí ľudia učia všetkému, čo je potrebné, aby sa mohli stať profesormi.

Anonymný autor

Provizórna chatrč uprostred neobývanej krajiny, dočasné sídlo sibírskej vedeckej expedície Moskovskej univerzity, vedenej profesorkou M. Bykovovou, ťažko stonala pod nárazmi ľadového severáku a podobné stony vydávala aj duša pani profesorky a s ňou aj celého niekoľkočlenného vedeckého tímu. Mráz im nevadil, na ten boli odmalička zvyknutí, ale akoby sa stratil cieľ ich výskumu a celej expedície, snežný muž Yetti [1]. Už niekoľko dní sa vôbec neukázal. Keď si pani profesorka Bykovová spomenula na jeho širokú chlpatú hruď a ohnivé krvavočervené oči, celá sa roztriasla:

„Naozaj tam nezostalo už ani trochu vodky?!“ – zrevala, až sa na chvíľu odmlčali vonku zúriace živly a výskumníci sa hlbšie zahrabali do svojich pelechov. Destilačný prístroj už dávnejšie rozbil ich meteorológ, ale aj tak boli jeho predpovede nanič, tak ho poslali zohnať nový. Panovala ponurá nálada, vedecký pokrok stagnoval. A nálada by sa tak skoro nezlepšila, keby nemali v expedícií aj malého Aľjošu, jedného z najstarších ale aj najveselších členov výpravy, ktorý sa do nej dostal pre svoju spoločenskú povahu. Okrem kryobiológie sa ako svojmu koníčku venoval štatistike a s vytrvalosťou hraničiacou až s obsesiou si do svojich denníkov zaznamenával každý detail, úplne všetko, od počtu a času ulovených sobov a medveďov, cez výkyvy počasia, pravidelnosť kúpania, až po dĺžku spánku jednotlivých členov výpravy. Preto nebola žiadna náhoda, ako sa to ostatným zdalo, keď veselo prehodil:

„Vydržte chlapci (pani profesorku rodovo neodlišoval), zajtra Yetti určite príde.“

Mysleli si, že si vymýšľa, aby ich nejako rozobral, rozveselil, ba aj pani profesorke prebehlo mysľou, či by ho tiež nemala poslať do tundry niečo zohnať, ale Aľjoša vytiahol svoj denník a začal im ukazovať svoje záznamy. Aj keď pre smäd a zimu nemohli v sporo osvetlenej chatrči dobre vidieť, aby im ubehol nejako čas, predsa len sústredili istú pozornosť na jeho výklad:

„Podľa mojich záznamov by sa tu zajtra mali zastaviť Nenci a Čukoti na svojej pravidelnej obchôdzke so sobími záprahmi a niečo nám priniesť.“

Všetci ožili, takúto pravidelnosť ešte nevypozorovali. Domorodci určite prinesú potrebné potraviny a výstroj, ale hlavne obnovia pre výpravu nevyhnutné zásoby silných domorodých repných destilátov.

„Mám tu zaujímavú tabuľku priemernej dennej spotreby vodky v našej expedícii“ – všetci sa nepokojne zamrvili – „ale istým riadením osudu, alebo vedeckou intuíciou doplnenú aj o počty členmi expedície udávaných kontaktov s Yettim.“

Aľjoša je predsa len chlapák, pomyslela si vedúca výpravy, aj keď rozpráva možno úplne z cesty. Okrem radostnej správy dáva tomu ešte aj vedecký rozmer.

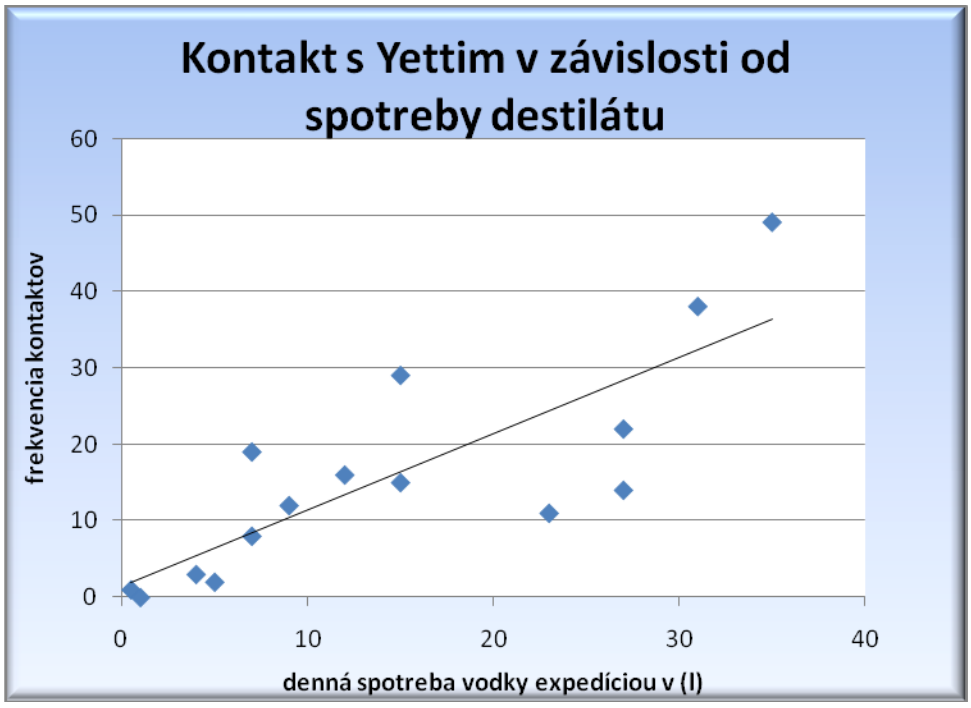
„Vylúčil som v delíriách uvádzané údaje nášho bývalého meteorológa, pravdepodobne výrazne zaťažené systematickou chybou a preto nevedecké,“ – všetci prikyvovali a dávali mu za pravdu, nech je tej pľuhe meteorológovi ľad ľahký!

„Tak som si to spároval a dostal som tabuľku:“

i	Spotreba [l/deň]	Yetti
1	0,5	1
2	1	0
3	4	3
4	5	2
5	7	19
6	7	8
7	9	12
8	12	16
9	15	29
10	15	15
11	23	11
12	27	22
13	27	14
14	31	38
15	35	49

Tabuľka VIII.1. Hodnoty závislosti frekvencie výskytu Yettiho od množstva spotrebovaného alkoholu vedeckou expedíciou.

„V prvom stĺpci je číslo pozorovania. V druhom som si zoradil spotrebu vodky za deň celou expedíciou variačne zoradenú podľa veľkosti, a v treťom je k tomu priradená hodnota frekvencie údajného výskytu Yettiho. V tom sú zahrnuté všetky formy výskytu a kontaktu so snežným mužom od jednoduchého pozorovania, že *sa tu obšmieta*, až po osobnú konverzáciu, družné popíjanie, tancovanie s Yettim či spanie s ním. Keď som si to vyniesol do grafu, dostal som obrázok:“



Obr.VIII.1. Grafické vyjadrenie závislosti frekvencie kontaktov so snežným mužom od množstva spotrebovaného repného destilátu.

Začínalo to byť fakt zaujímavé, závislosť sa javila nespochybniteľne.

„Vzal som 2 body, napr. 3. a 12. a keďže mi ihneď do oka udrela priamka ako obraz lineárnej závislosti, zapísal som si jej rovnicu (my sme si ju uviedli v 1.kapitole na obr.I.9.):“

$$y = a.x + b$$

V nej premenná **x** je tzv. **nezávisle premenná**, v našom prípade je to spotreba alkoholu, pretože ho pijeme nezávisle od čohokoľvek a koľko chceme. Druhá premenná **y** už je závisle premenná, pretože jej hodnota závisí od hodnoty **x**. Je to počet akýchkoľvek kontaktov s Yettim, ktoré uvádzajú členovia expedície po vypití príslušného množstva samohonky. Tak som si do rovnice dosadil najprv hodnoty jedného bodu (x=4; y=3) a potom druhého vybraného bodu (x=27; y=22) a dostal som dve rovnice:“

I. $3 = 4a + b$

II. $22 = 27a + b$

„Z toho som už ľahko dostal smernicu priamky, čiže lineárnej závislosti **a = 0,826** a koeficient **b = -0,304**; ktorý sa príliš nelíši od nuly," – pokračoval zanietene Aľjoša až lapal v mrazivom vzduchu po dychu.

„Viem, že je to len veľmi približný výpočet," uzatváral Aľjoša –„ale nech sa to zoberie z ktorejkoľvek strany, ak zajtra prídu so svojim nákladom domorodci, tak podľa všetkého príde aj Yetti!"

Nastala neopísateľná radosť, vytiahli sa balalajky a harmoniky a ako sa vraví, vo vede šťastie praje pripraveným – našiel sa ešte posledný demižónik repovice. Dlho do noci sa tundrou ozývali spev a hlasné bujaré výkriky na nerozoznanie od Yettiho. Život mal opäť perspektívy.

Dotkli sme sa veľmi zaujímavej oblasti korelácie a regresie.

Korelácia – je obecná závislosť medzi dvomi, a vo všeobecnosti aj medzi viacerými znakmi v štatistickom súbore. Udáva stupeň závislosti.

Regresia – podrobnejší priebeh korelačnej závislosti. Odpovedá na otázku, aká bude hodnota závisle premennej v i-tom bode $\mathbf{y_i}$ ak poznáme hodnotu nezávisle premennej $\mathbf{x_i}$. Hneď na začiatku chceme upozorniť, že je to oblasť, v ktorej sa mimoriadne darí neandertálskej štatistike. Pre nás platí, že tieto otázky možno zodpovedať pre znaky **x** a **y**, ktoré sú závislé. Pre nezávislé to nie je možné. Prečo to hovoríme? To nie je vôbec samozrejmé. Predstavme si dva nezávislé znaky, ktoré majú rastúci trend približne rovnakým tempom napr. rast počtu držiteľov strelných zbraní na jednej strane a rast počtu milovníkov cereálnej stravy na druhej. Aj keď výpočtom príslušných koeficientov a charakteristík dostaneme vysokú závislosť premenných, aj tak tieto dva javy nebudú spolu súvisieť. Ale stačí sa trochu poobzerať po masmediálnom svete a nebudete sa stačiť diviť, čo všetko sa dáva do súvisu, aké „objavné" korelácie boli odhalené a musíme sa im vraj prispôsobiť! Senzačné kauzality novinárov a reklamných pracovníkov neustále udivujú svet. Niekoľko príkladov:

Fajčenie vraj stimuluje rýchlosť myslenia. Inokedy je fajčenie dôsledkom príliš krátkeho kojenia jedinca v dojčenskom veku, tak si to vynahradzuje zástupným dodatočným riešením – cigaretou. Jedenie morských rýb zvyšuje inteligenciu. Časté býva aj zamieňanie príčiny a účinku: Vyššie dane spôsobujú ekonomické problémy firmám a zníženie konkurencieschopnosti. Zníženie konkurencieschopnosti firiem vedie k menšiemu obratu, k menším daňovým odvodom a k nutnosti zvýšiť dane. Toto všetko sa dá podložiť výsledkami výpočtov korelácie a regresnou analýzou. Zdá sa však, že dosť v tom vyniká zdravotnícka štatistika, alebo to, čo sa za ňou často skrýva.

Ak ste vcelku zdravý jedinec v istom veku a náhodou sa dočítate, že vedecké štúdie s výsledkami výpočtov na mnoho desatinných miest potvrdzujú nebývalý nárast neurodegeneratívnych chorôb (Parkinson, Alzheimer a pod.), pritom neuvádzajú, že to viac-menej odpovedá rastu populácie a predlžovaniu ľudského veku; a že už v mladšom veku možno pozorovať prvé vážne príznaky, napríklad zabúdanie, ste v pasci. Lebo zabúdate. Čoskoro zistíte, že vďaka tomuto informačnému šumu sú preplnené čakárne všetkých neurologických ambulancií. Neurológia sa stala veľmi populárnou a výnosnou aktivitou. Ak máte známosti, tak sa po riadnej námahe prepracujete až k Odborníkovi. Ten vám samozrejme predpíše poriadny liek. Ste zachránený od Alzheimera, ale bohužiaľ v tomto svete nič nie je dokonalé, a liek má vedľajšie účinky a trasom začínate vyzerať ako parkinsonik. To nič, pridá sa liek proti trasu, ten však trochu pôsobí na tráviaci systém. Je nutné nasadiť intravenóznu medikamentóznu liečbu, ale aby ste neboli preliekovaný, pridajú sa nejaké vitamíny a výživové doplnky. Nedá sa však nič robiť, nejako vám nefungujú obličky, puchnú vám nohy, aj pankreas si robí čo chce, je potrebné nasadiť lieky, ktoré urobia zásadný obrat vo vašom organizme plus lieky na spanie. Ak máte šťastie, tak vám to len prudko zníži systolický aj diastolický krvný tlak a začínate byť postupne dementný. Vrátite sa na neurológiu, kde Odborník víťazoslávne zvolá: Je to Alzheimer!

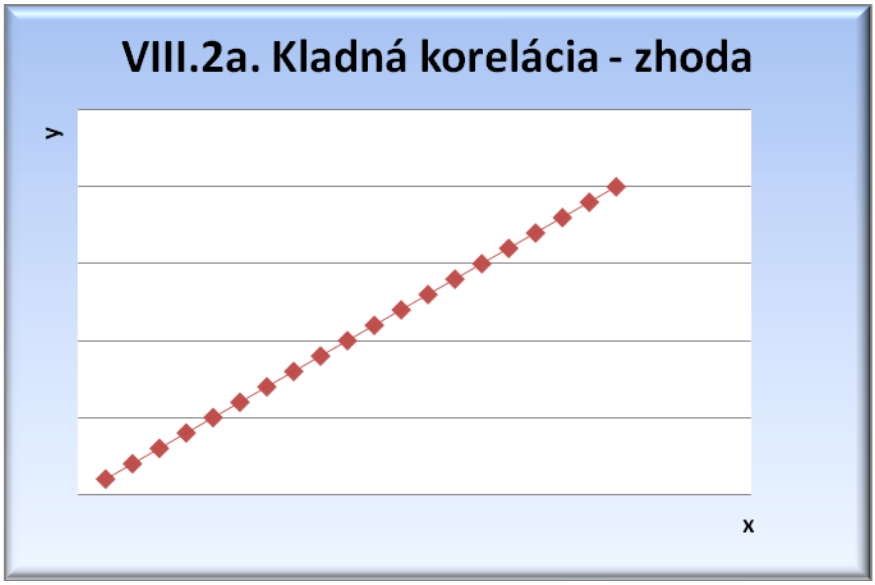
Čo v skutočnosti ponúka regresná analýza?

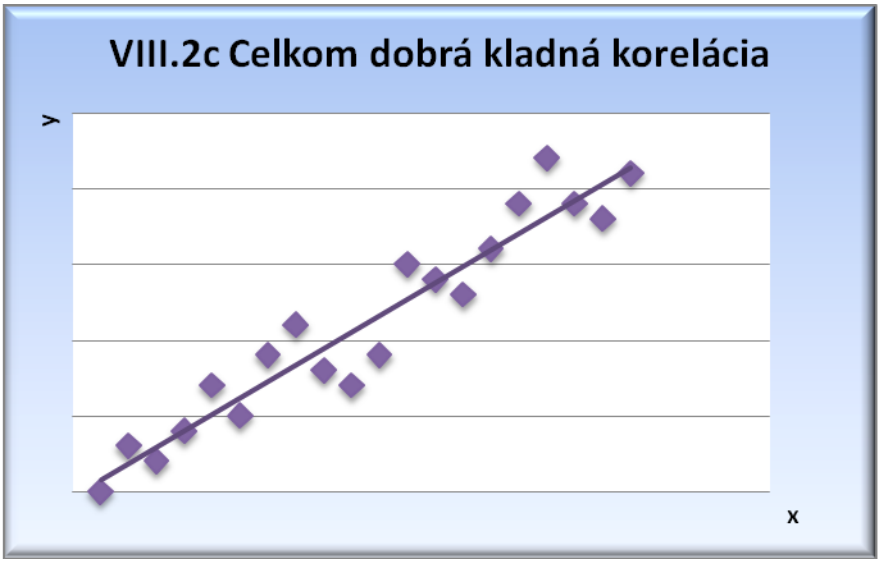
- Môže dať nejaký analytický vzťah, rovnicu, ktorá popisuje vzťah medzi závislými náhodnými premennými.
- Môže potvrdiť teóriu resp. hypotézu, ktorú sme si vytvorili o súvislosti medzi premennými a pritom kvantitatívne odhadnúť veľkosti a znamienka príslušných koeficientov vzťahu.
- Môže predpovedať hodnoty závisle premennej. To je už prognostika resp. predikcia. Ako povedal Niels Bohr, dánsky vedec a nobelista za fyziku, predpovedať niečo je dosť ťažké, hlavne ak ide o budúcnosť. Pozrime sa teda, čo nám korelácia a regresia v skutočnosti umožňuje:

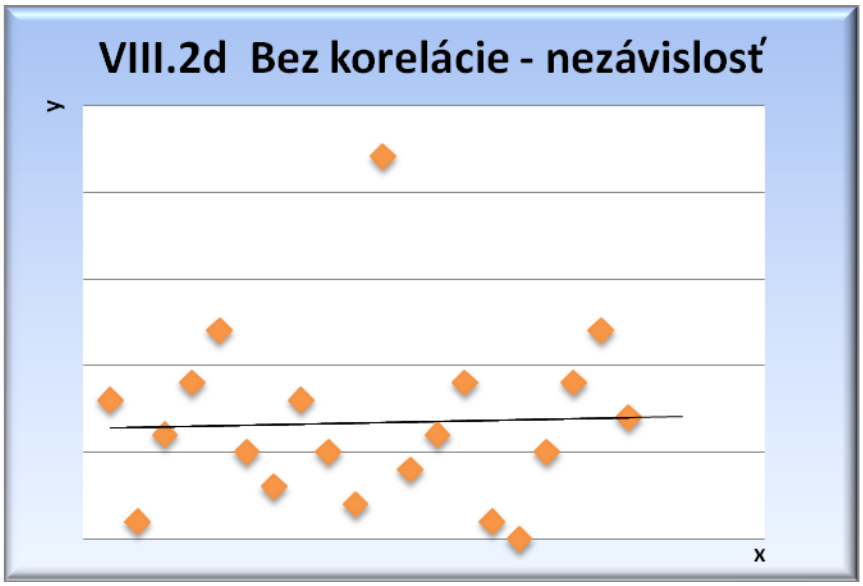
Máme dvojrozmerné pozorovanie. Napr. vzdelanie – nezamestnanosť, vek - miera dlhodobej nezamestnanosti, drogová závislosť – kriminalita, a i. Jeho výsledky, teda dvojice pozorovaných hodnôt, je možné spracovať do tabuliek alebo graficky. Medzi pozorovanými štatistickými znakmi môže nastať jeden z nasledujúcich prípadov:

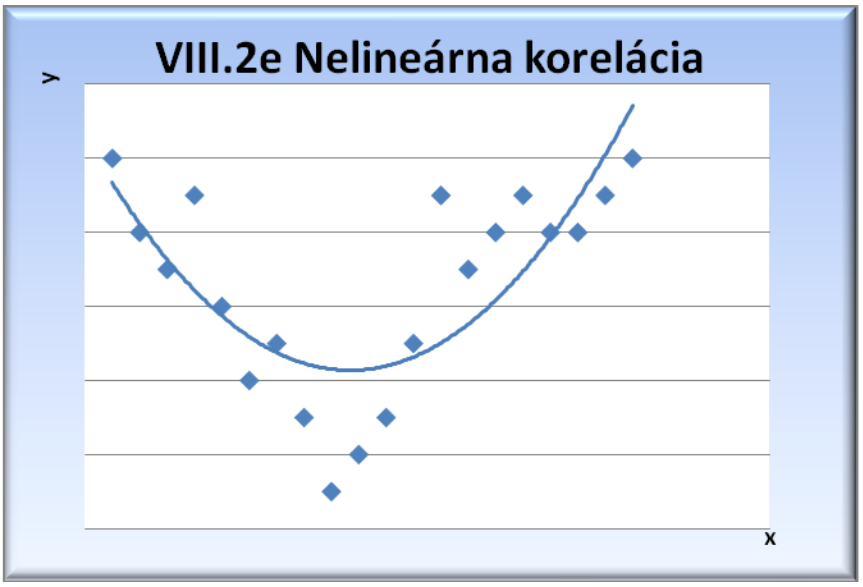
1. Zhoda
2. Protiklad
3. Nezávislosť

Na nasledujúcich obrázkoch **VIII.2a až VIII.2e** sú graficky znázornené možné vzťahy medzi premennými:









Číselná charakteristika pre koreláciu je koeficient korelácie ***r***, ktorý vypočítame ako

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \qquad \textbf{[VIII.1.]}$$

alebo sa používa aj vzťah s **kovarianciou x a y** (**cov(x;y)**):

$$r = \frac{cov(x;y)}{\sigma_x . \sigma_y} = \frac{\overline{x.y} - \overline{x}.\overline{y}}{\sigma_x . \sigma_y} \qquad \textbf{[VIII.2.]}$$

kde $\overline{x.y}$ je aritmetický priemer súčinov $x_i.y_i$ a $\overline{x}.\overline{y}$ je súčin aritmetických priemerov znakov **x** a **y**; σ_x a σ_y sú smerodajné odchýlky znakov **x** a **y**. ***r*** nadobúda hodnoty z intervalu

$$r \in \langle -1;1 \rangle.$$

Nie je to také hrozné ako to na prvý pohľad vyzerá, uveďme si jednu dôležitú tabuľku a radšej jednoduchý príklad:

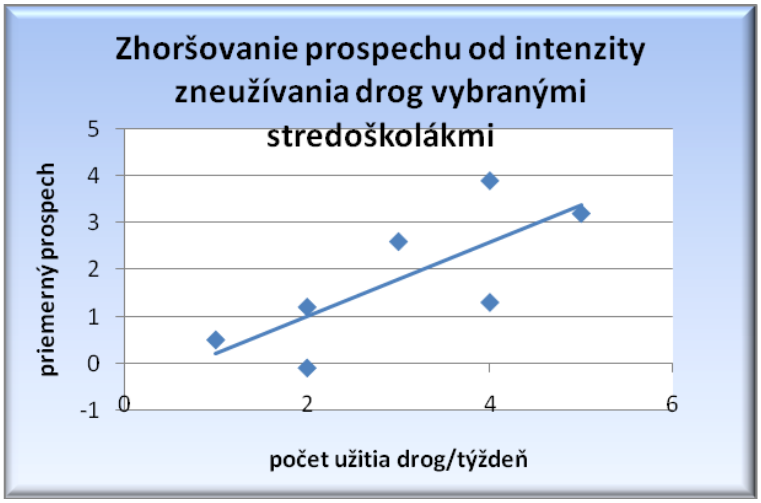
Koeficient korelácie r	Sila závislosti znakov x a y
$\|r\| = 0$	Úplná nezávislosť
$0 \le \|r\| < 0,3$	Prakticky nulová, žiadna závislosť
$0,3 \le \|r\| < 0,5$	Mierna závislosť
$0,5 \le \|r\| < 0,7$	Významná závislosť
$0,7 \le \|r\| < 0,9$	Vysoká závislosť
$0,9 \le \|r\| < 1$	Prakticky úplná závislosť
$\|r\| = 1$	Úplná závislosť

Tabuľka VIII.2: Veľkosť koeficientu korelácie r a sila závislosti znakov x a y

Pr.VIII.1: Sledovalo sa zhoršovanie (nárast) priemeru známok 7 študentov stredných škôl y_i od intenzity (počtu) aplikácie drogy za týždeň x_i. Pre výpočet koeficientu korelácie bola spracovaná pracovná tabuľka **VIII.3**:

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$x_i . y_i$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	1	0,5	-2	-1,3	0,5	4	1,69
2	2	-0,1	-1	-1,9	-0,2	1	3,61
3	2	1,2	-1	-0,6	2,4	1	0,36
4	3	2,6	0	0,8	7,8	0	0,64
5	4	1,3	1	-0,5	5,2	1	0,25
6	4	3,9	1	2,1	15,6	1	4,41
7	5	3,2	2	1,4	16	4	1,96
Súčet	21	12,6	0	0	47,3	12	12,92
priemer	$\bar{x} = 3$	$\bar{y} = 1,8$	$\overline{x.y} = 6,757143$			$\sigma_x = 1,309307$	$\sigma_y = 1,358571$

Grafické znázornenie dáva predstavu o závislosti premenných:



Obr.VIII.3: Závislosť zhoršovania prospechu od intenzity aplikácie drogy vybranými študentmi stredných škôl

Vo vzťahu **[VIII.1.]** resp. **[VIII.2.]** toho poznáme už dosť, tak možno vyhodnotiť silu pozorovanej závislosti, do akej miery jeden štatistický znak vplýva na druhý:

$$r = \frac{cov(x;y)}{\sigma_x.\sigma_y} = \frac{\overline{x.y} - \overline{x}.\overline{y}}{\sigma_x.\sigma_y} = \frac{6,757143 - 3\ .\ 1,8}{1,309307\ .1,35871} = 0,76296 \cong 0,76$$

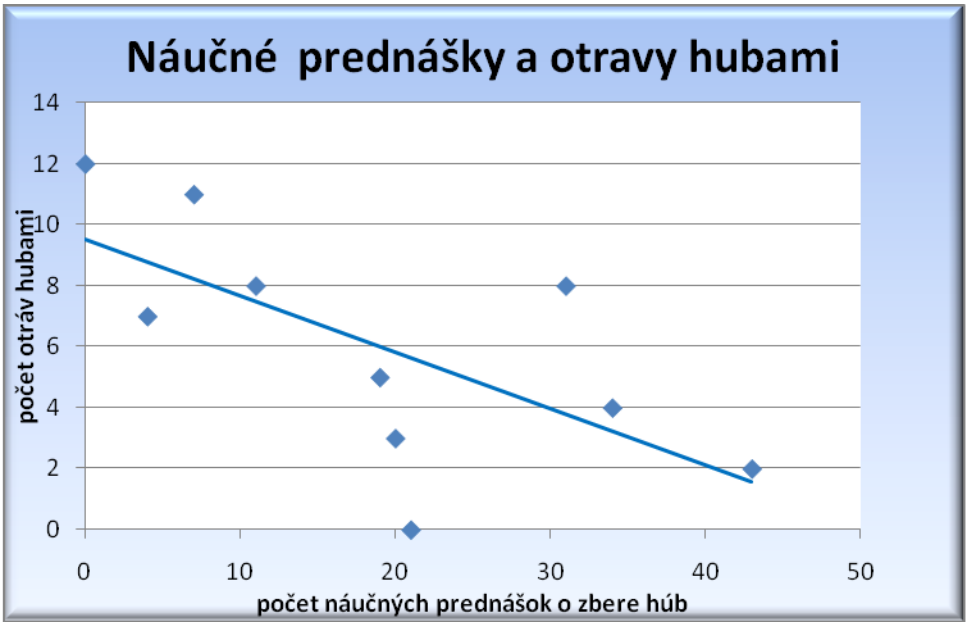
Podľa tabuľky VIII.2 ide o vysokú závislosť medzi užívaním drog a zhoršovaním prospechu v škole pre vybraných stredoškolákov.

Pr.VIII.2: V okresnom meste, správnom centre regiónu s veľmi vysokou nezamestnanosťou obklopenom horami dochádzalo k častým otravám jedovatými hubami. Dobrovoľní humanitárni pracovníci si popri svojej práci u mykológov zistili, že všetky blízke lesy nie sú z hľadiska výskytu jedovatých húb ničím výnimočným. Rozhodli sa počas sezóny usporiadať sériu náučných prednášok o spôsobe správneho zberu húb so zdôraznením jeho nebezpečenstiev s praktickými ukážkami a rozdávaním inštruktážnych plagátikov. Zdalo sa im, že to má význam, ale potrebovali si to štatisticky overiť. Pozbierali výsledky pozorovaní a zhrnuli to do tabuľky:

i	počet	počet otráv	x-xp	y-yp	x.y	(x-xp)**2	(y-yp)**2
1	0	12	-19	6	0	361	36
2	4	7	-15	1	28	225	1
3	7	11	-12	5	77	144	25
4	11	8	-8	2	88	64	4
5	19	5	0	-1	95	0	1
6	20	3	1	-3	60	1	9
7	21	0	2	-6	0	4	36
8	31	8	12	2	248	144	4
9	34	4	15	-2	136	225	4
10	43	2	24	-4	86	576	16
Σ	190	60	0	0	11400	1744	136
priemer	$\overline{x}$ =19	$\overline{y}$ =6	$\overline{x.y}$ =81,8			σ_x=13,21	σ_y=3,69

Tabuľka VIII.4: Počty náučných prednášok a otráv hubami s pomocnými výpočtami

Overili si to najprv graficky a potom výpočtom:



Obr.VIII.4: Grafická korelácia medzi počtom náučných prednášok o zbere húb a počtom otráv hubami v meste vybraného regiónu.

Dosadením do **[VIII.2.]** už ľahko dostaneme hodnotu korelačného koeficientu ***r* = - 0,66** čo predstavuje už významnú závislosť štatistických znakov pri negatívnej korelácii, takže možno predpokladať, že náučné prednášky mali vplyv na pokles počtu otráv jedovatými hubami.

Čitateľ správne tuší, že aj pri týchto výpočtoch vám môže pomôcť veľmi efektívne a veľmi jednoducho napr. EXCEL. Ak máte vstupné dáta v 2.stĺpcoch v EXCELI, v okne ***Analýza dát*** máte funkciu ***Korelace***. Keď si na ňu kliknete, tak len myšou štandardne vložíte ***Vstupnú oblasť***, t.j. bunky, v ktorých máte vložené údaje nezávisle a závisle premennej (x_i a y_i); ***Výstupnú oblasť***, t.j. aspoň jednu voľnú bunku mimo vstupnej oblasti a kliknete OK. Vyhodí vám to minitabuľku s výsledkom výpočtu korelačného koeficientu ***r***, napr. v poslednom príklade, ktorú porovnáte s tab.VIII.2. o sile závislosti. A je to!

	Sloupec 1	*Sloupec 2*
Sloupec 1	1	
Sloupec 2	**-0,66117**	1

Takýmto spôsobom si môžete veľmi rýchlo vypočítať korelačný koeficient z úvodnej sibírskej historky tejto kapitoly: r = 0,81. To už je veľmi slušná závislosť, navodzujúca tušenie, že bola odhalená nejaká neznáma kauzalita. Ale naozaj?

Ako vo viacerých iných, tak aj na tomto pre istotu veľmi vypuklom príklade sa snažíme poukázať na úskalia, ktoré na vás číhajú. Prvé je v nás samotných. Predpokladajme, že sme vynaložili značné úsilie na získanie primárnych alebo vyhrabanie sekundárnych dát a chceme si dokázať nejaký veľmi zaujímavý jav, ktorý si doteraz nikto nevšimol. Teda stať sa slávnym. Môžeme podľahnúť pokušeniu a možno by nám to aj do značnej miery zjednodušilo prácu, povedať, že korelácia, ktorú sme vypočítali je aj dôkazom kauzality pre naše potreby, pritom zovšeobecnenie z vypočítaného na to, čo chceme dokázať, nemusí byť pravdivé. Napríklad, chceme dokázať existenciu Yettiho, potvrdzujú nám našu hypotézu aj výpovede celej sibírskej expedície, dokonca aj nejaká korelácia vychádza dosť presvedčivo. Ťažkosti sú v interpretácii. Aj táto korelácia je pravdivá, keď urobíme syntézu, že je to potvrdenie závislosti počtu tvrdení o kontaktoch s Yettim, zaznamenaných členmi expedície v situácii extrémneho neurofyziologického stavu ovplyvneného vysokou intoxikáciou etanolom, od množstva spotrebovaného destilovaného alkoholu. A nie závislosť počtu reálnych kontaktov od nejakého iného reálneho úkazu, napr. od množstva skonzumovaného destilátu.

Existuje množstvo ľudí a záujmov, ktorí sa vás snažia a budú snažiť zmanipulovať. Bohužiaľ len nepatrná časť z týchto pokusov ide na vrub ľudskej hlúposti. Väčšinou je to vysoko sofistikovaná činnosť, v ktorej nechýbajú kvalitné poznatky zo štatistiky, psychológie, propagandy, tvorby verejnej mienky, gramatiky, vizuálnych vnemov a neviem čoho ešte všetkého.

Krásnu tradíciu práce so slovom a jeho účinkoch na ľudskú myseľ i emócie udáva jeden príbeh z počiatku 20. storočia v Paríži, kde sa v bohémskej štvrti Montmartre stretávali umelci z celého sveta. Pri jednom takomto večernom stretnutí v miestnych rázovitých kaviarničkách sa niekto spýtal Jamesa Joycea, pôvodom írskeho spisovateľa, čo cez deň robil. Samozrejme odpovedal, že písal. Iný predhodil otázku, aby bolo veselo, že čo napísal. Joyce, ako ho hodnotíme dnes, jeden z najväčších spisovateľov 20 .storočia a vôbec postáv svetovej literatúry, odpovedal: *Jednu vetu...*

Nemusíme zachádzať na územie literárnej kritiky, je to len okrajový fakt, že jeho vety sú vybrúsené brilianty, ale pre naše účely je dôležité, že tento spisovateľ zostal vo svojom povolaní pravdivý, autentický, nezapredal svoj talent, neprostituoval slovom. Ako protiklad chceme uviesť, že existuje dosť schopných odborníkov na slovo, pracujúcich v reklame, v žurnalistike, v politike a pod., ktorí využívajú svoje schopnosti na klamanie. Veľmi ľahko zahmlia kritické miesto analýzy, napr. okolo metodiky výskumu sibírskej expedície a zdôraznia jej príslušnosť k renomovanej Moskovskej univerzite a pod. Ak si nedáme pozor, hravo

a bravúrne nás dostanú tam, kde chcú, samozrejme na omnoho „jemnejších“ škálach vyjadrovacích prostriedkov ako v našom príbehu.

Všetky totalitné režimy v posledných obdobiach mali také inštitúcie ako ministerstvá propagandy a školených odborných pracovníkov, ktorí dokázali význam nejakej informácie, ktorá sa nedala úplne poprieť, aspoň jemne posunúť. Koľko západných intelektuálov verilo správam o raji na zemi v Stalinovom Rusku? Ešte donedávna sa všeobecne verilo, že sovietske atómové bomby sú nejakým spôsobom lepšie, obranné, pokusy s nimi menej škodiace atmosfére a planéte, atómový program je viac „mierový“. A keď aj boli nejaké neodškriepiteľné nedostatky, rýchlo sa spustila chvála na Stalina, že dostal krajinu z feudálnej zaostalosti medzi technicky vyspelé jadrové mocnosti. Aby nedošlo k omylu, pre nás sú atómové bomby zlé ako všetky zbrane namierené proti ľuďom a kdekoľvek na svete. Hirošima a Nagasaki sú neprijateľné. Ale kto chce, môže si dnes už vyhľadať informácie, akú cenu museli za to všetko zaplatiť predovšetkým ruskí ľudia. Vývoj americkej atómovej bomby stál pri úrazoch a nehodách život niekoľko ľudí, väčšinou vedcov, ktorých mená sú známe. Štatistické súbory dát úmrtí v tejto oblasti v Rusku spojené so zruinovaním vlastnej ekonomiky sú neporovnateľné a ich rozsah je neuveriteľný, keby som uvádzal len najnižšie odhady, vyzeralo by to ako nejaká moja propaganda [2].

Nebuďme však smutní, že sme takí hlúpi a často naletíme neandertálcom, pretože napr. reklama je úplne profesionálna činnosť s nepredstaviteľným finančným obratom. Tieto finančné prostriedky by do toho nikto nehádzal, keby to nefungovalo. Aj preto nám štatistika môže slúžiť ako veľmi účinný obranný nástroj [3].

Zdá sa vám, že sme príliš rýchlo zanechali testovanie štatistických hypotéz? A cnie sa vám za nimi? Ale to sa dá bez problémov napraviť:

Existuje nejaký veľmi užitočný obecný jav patriaci do studnice poznania ľudstva, napr. že aktivita somnambulikov rastie s narastaním Mesiaca. Základný súbor závislosti javu **Y** na jave **X** s prvkami $(x_k; y_k)$, kde **k** je veľmi veľké kladné celé číslo, sú všetci lunatici. Koeficient korelácie základného súboru označme ρ. (V prípade, ktorý študujete možno jeho veľkosť na základe toho čo sa hovorí, odhadnúť až na $\rho = 0{,}8$). Aby ste si uvedený jav overili, nemôžete testovať celý základný súbor, musíte urobiť výber. Výberový súbor podrobíte skúmaniu, zozbierate prvky výberového súboru $(x_i; y_i)$ a vypočítate si vyššie uvedeným spôsobom výberový koeficient korelácie ***r***. Nakoľko výberová charakteristika ***r*** charakterizuje aj základný súbor a jeho charakteristiku ρ, sa dá zistiť testom významnosti koeficientu korelácie ***r***. Test významnosti ***r*** sa dá urobiť dvojstranne alebo jednostranne:

1. Dvojstranný test významnosti koeficientu korelácie ***r***:

H_0: $\rho = 0$ (žiadna závislosť medzi Y a X), H_1: $\rho \neq 0$ (existuje štatistická závislosť).

Testovacia štatistika pri hladine významnosti testu α má tvar:

$$t = r.\sqrt{\frac{n-2}{1-r^2}} \qquad \textbf{[VIII.3]}$$

ktorá má Studentovo rozdelenie o (n-2) stupňoch voľnosti. Ako vždy H_0 rutinne zamietame na hladine významnosti α ak $|t| > t_{\frac{\alpha}{2}}(n-2)$ kde $t_{\frac{\alpha}{2}}(n-2)$ nájdeme v tabuľkách Studentovho rozdelenia. V opačnom prípade, teda ak $|t| < t_{\frac{\alpha}{2}}(n-2)$, nulovú hypotézu nemožno zamietnuť a na hladine významnosti α predpokladáme, že korelácia nemá štatisticky význam.

2. Ľavostranný test významnosti koeficientu korelácie ***r***:

H_0: $\rho = 0$ (žiadna závislosť medzi Y a X), H_1: $\rho < 0$ (existuje štatistická závislosť, záporná korelácia). H_0 zamietame, ak $t < -t_{\alpha}(n-2)$, v prospech alternatívnej hypotézy, že existuje záporná korelácia medzi pozorovanými znakmi Y a X. Ak $t > -t_{\alpha}(n-2)$, H_0 nezamietame a korelácia nie je štatisticky významná.

3. Pravostranný test významnosti koeficientu korelácie ***r***:

H_0: $\rho = 0$ (žiadna závislosť medzi Y a X), H_1: $\rho > 0$ (existuje štatistická závislosť, kladná korelácia). H_0 zamietame, ak $t > t_{\alpha}(n-2)$, v prospech alternatívnej hypotézy, že existuje kladná korelácia medzi pozorovanými znakmi Y a X. Ak $t < t_{\alpha}(n-2)$, H_0 nezamietame a korelácia nie je štatisticky významná.

Pr.VIII.3: Presvedčte zopár náhodne vybraných svojich priateľov a priateliek – somnambulikov, s ktorými sa radi prechádzate po nočných strechách, aby ich manželský partneri zaznamenávali frekvenciu príhod somnambulizmu v rôznych fázach Mesiaca, napr. cez novomesiac, cez prvú 1/8, cez ¼, ½, ¾ až po spln a zo zozbieraných hodnôt urobte vždy aritmetický priemer. Uložte si to po zaokrúhlení na celé čísla do tabuľky:

Fáza Mesiaca [%]	**0**	**12,5**	**25**	**50**	**75**	**100**
Frekvencia aktivít lunatikov	**2**	**0**	**4**	**5**	**1**	**5**

EXCEL vám pomôže s výpočtom výberového koeficientu korelácie, ale ak chcete počítať postupujte podľa **[VIII.1.]** resp. **[VIII.2.]**:

	Fáza Mesiaca [%]	*Frekvencia aktivít lunatikov*
Fáza Mesiaca [%]	1	
Frekvencia aktivít lunatikov	0,440288	1

Váš výsledok $r = 0{,}44$ nie je veľmi oslňujúci, aj tab.VIII.2 uvádza pre túto hodnotu len miernu závislosť. Urobíte preto test. Vypočítate testovacie kritérium podľa **[VIII.3]**

$$t = r.\sqrt{\frac{n-2}{1-r^2}} = 0{,}440288.\sqrt{\frac{6-2}{1-0{,}1938535}} \cong 0{,}981$$

V tabuľkách Studentovho rozdelenia nájdeme

$$t_{\frac{\alpha}{2}}(n-2) = t_{0{,}025}(4) = 2{,}776$$

$$-t_{\alpha}(n-2) = -t_{0{,}05}(4) = -2{,}132$$

$$t_{\alpha}(n-2) = t_{0{,}05}(4) = 2{,}132$$

Testujeme:

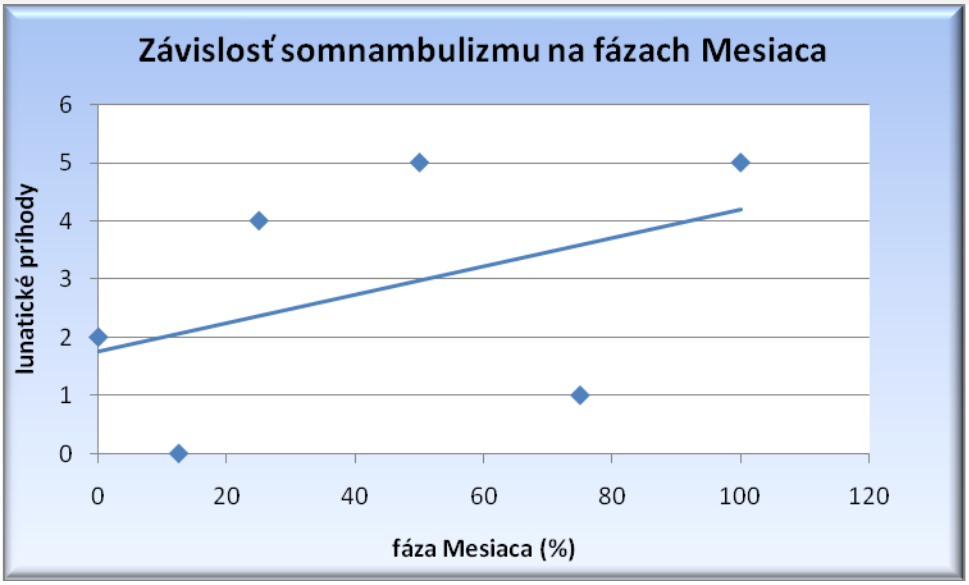
1. Obojstranný test: $|t| < t_{\frac{\alpha}{2}}(n-2)$, nulovú hypotézu nemožno zamietnuť a na hladine významnosti **α=0,05** predpokladáme, že korelácia nemá štatisticky význam.

2. Ľavostranný test: $t > -t_{\alpha}(n-2)$, H_0 nezamietame a korelácia nie je štatisticky významná.

3. Pravostranný test významnosti korelačného koeficientu ***r***:

H_0 nezamietame v prospech alternatívnej hypotézy, pretože $t < t_{\alpha}(n-2)$.

Urobte si vždy aj grafickú výpoveď, obrázok je často dosť názorný.



Obr.VIII.5: Grafická korelácia medzi intenzitou somnambulizmu a fázami Mesiaca

Niekedy sa na posúdenie variability koeficientu korelácie používa **koeficient determinácie** (aj koeficient spoľahlivosti) r^2. V našom príklade r^2= **0,194**. Ten vyjadruje, že len v 19,4% variability lunatických príhod možno pripustiť vplyv fáz Mesiaca. Ostatné?

Koreláciou zisťujeme závislosť dvoch znakov a orientačne jej silu. Môžeme si otestovať koeficient korelácie. Analytické vyjadrenie závislosti dostaneme, keď zistíme matematicky vzťah medzi premennými, čo sme už v začiatku publikácie nazvali funkciou: **y=f(x)**. Z grafického zobrazenia bodov výberového súboru možno často už usúdiť o aký typ závislosti pôjde a vytvoriť si model. Najjednoduchší je model jednoduchej lineárnej závislosti

$$y = a.x + b \qquad \text{[VIII.4]}$$

Z experimentálnych údajov dvojíc $(x_i; y_i)$ lineárna regresia umožňuje určiť koeficienty **a** a **b** a tým vlastne preložiť experimentálne body optimálnou tzv. **regresnou priamkou** metódou najmenších štvorcov tak, aby odchýlky regresnej priamky a experimentálnych hodnôt e_i boli čo najmenšie. Zároveň určí štatistickú spoľahlivosť vypočítaného regresného modelu a dáva v štatisticky „rozumnom“ intervale možnosť predpovedať hodnotu znaku **y**, keď poznáme znak **x**. Koeficient **a** nazývame smernica priamky; **b** je jej kvocient, t.j. hodnota y_0 keď **x=0**. Keď máme **n** experimentálnych bodov pre

$$y_i = a.x_i + b; \text{kde } i = 1,2,3,\dots,n \qquad \text{[VIII.5]}$$

tak pre koeficienty a a b platia vzťahy:

$$a = \frac{\sum_1^n x_i . \sum_1^n y_i - n. \sum_1^n x_i . y_i}{\left(\sum_1^n x_i\right)^2 - n. \sum_1^n x_i^2}$$

$$b = \frac{\sum_1^n x_i^2 . \sum_1^n y_i - \sum_1^n x_i . \sum_1^n x_i . y_i}{n. \sum_1^n x_i^2 - \left(\sum_1^n x_i\right)^2} \qquad \text{[VIII.6]}$$

resp.

$$b = \bar{y} - a.\bar{x} \qquad \text{[VIII.7]}$$

Súčet druhých mocnín (štvorcov) odchýlok regresnej priamky a experimentálnych hodnôt si označme E^2:

$$E^2 = \sum_i e_i = \sum_i [(y_i - (a.x_i + b)]^2 \qquad \text{[VIII.8]}$$

Označme $s = \sqrt{\frac{E^2}{n-2}}$, ktorá sa nazýva štandardná odchýlka rezíduí, alebo regresnej priamky. Ak máme viac modelov, model s minimálnou hodnotou ***s*** je najlepší.

Potom možno odhadnúť smerodajné odchýlky koeficientov priamky vzťahmi:

$$s_a = \frac{s}{\sqrt{\sum_i x_i^2 - \frac{(\sum_i x_i)^2}{n}}}$$

a

$$s_b = s. \sqrt{\frac{\frac{1}{n}.\sum x_i^2}{\sum x_i^2 - \frac{(\sum_i x_i)^2}{n}}} \qquad \textbf{[VIII.9]}$$

Dajú sa na tom robiť testy na rozptyl (ANOVA), testy významnosti koeficientov **a** a **b** a ďalšie kúzla, ale už aj tak to vyzerá dosť náročné, vzorce trochu kostrbaté a až ezoterické. Ale netreba sa ich zľaknúť, naozaj vás privedú k vytúženým hodnotám. Je potrebné vedieť, čo chcete: Priamku preloženú cez experimentálne dáta, ktorá ich bude najlepšie vystihovať. A ako to už v štatistike býva, ešte niekoľko údajov, aby ste sa mohli na ňu na nejakej úrovni spoľahnúť. Napriek zdanlivej náročnosti vzorcov VIII.3 až VIII.9, všetko za vás urobí EXCEL cez funkciu ***Regrese*** v ***Analýza dát***.

Pr.VIII.4.: Rodina, ktorá ponúkala ľuďom zábavu, radosť zo života a splnenie aj skrytých túžob, otvorila novú sieť klimatizovaných kasín v štýle secesných palácov, technologicky vybavených lepšie ako centrum riadenia kozmických letov. Predpokladané zisky sa však nedostavovali, bežná populácia bola už vybavená aspoň pudovým štatistickým myslením, finančná kríza tiež urobila svoje. Najvyššie vedenie rodiny sa rozhodlo zmeniť stratégiu a zistiť, na akú klientelu je nutné sa zamerať. Po predbežnom prieskume sa začínal črtať paradox, že najvýhodnejšie bude osloviť ekonomicky neaktívnu, ale pritom svojprávnu časť populácie – seniorov. Program *Šťastná staroba*, ktorý rodina veľkoryso financovala mal za úlohu zistiť dva jednoduché fakty:

1. Priemernú výšku úspor seniora po 40 až 50 rokoch tvrdej práce.
2. Nájsť vzťah medzi poklesom kognitívnych funkcií a schopnosťou vložiť isté % svojich úspor do hazardných hier.

Keďže rodina mala priateľov na všetkých významných miestach, medziiným sudcov, bankárov, finančníkov a daňových úradníkov, prvý bod ľahko odhadla na približnú hodnotu 250000.- US $ na seniora.

V druhom bode musela skupina odborníkov urobiť výskum. Pokles kognitívnych funkcií seniorov sledovali jednoduchým Folsteinovým testom MMSE s maximálnym skóre 30 bodov pri ich plnom zachovaní. Náhodne vybraných seniorov vzali na platený príjemný a fyzicky nenáročný pobyt v luxusnom zariadení pri mori, ktoré tiež patrilo rodine, kde sa okrem iných zábav v rámci pobytu mohli stretávať v spoločenskej miestnosti, v ktorej si mohli zahrať rôzne hazardné hry so žetónmi, ktoré dostali v rovnakom množstve na začiatku. Pokiaľ by nehrali, alebo ušetrili resp. aj vyhrali, tak za žetóny, ktoré na konci mali, si mohli nakúpiť pri odchode najrôznejšie darčeky v predajniach zariadenia. Úlohou milých obsluhujúcich dievčat bolo obdivne počúvať klientov a bez akéhokoľvek nátlaku navodiť situáciu, aby každý senior, aj keď bol ešte krátko predtým schopný do krvi sa v obchode pohádať pre 10 centov, začal hrať. Technický personál zabezpečil, aby vstup do hry bol sprevádzaný počiatočnou výhrou, ocenenou aj uznaním okolia. Výsledkov experimentu sa potom zmocnili odborníci, ktorí si ich najprv zhrnuli do tabuľky:

i	test MMSE	Zl. úspor	i	test MMSE	Zl. úspor	i	test MMSE	Zl. úspor
1	8	1	11	8	0,6	21	16	0,6
2	20	0,2	12	27	0,2	22	26	0
3	20	0,5	13	12	0,9	23	26	0,1
4	12	0,5	14	12	0,7	24	22	0,4
5	27	0,1	15	16	0,5	25	14	0,4
6	16	0,6	16	26	0,5	26	24	0,4
7	24	0,2	17	14	0,9	27	18	0,4
8	4	0,9	18	14	0,8	28	20	0,6
9	10	0,8	19	24	0,3	29	18	0,3
10	22	0,7	20	18	0,6	30	22	0,3

Tab. VIII.5.: Hodnoty skóre testu kognitívnych schopností MMSE a zlomku úspor, ktoré sú seniori ochotní v určitej situácii utratiť na hazardných hrách.

Potom v EXCELI cez ***Koreláciu*** v ***Analýze dát*** získali štandardným spôsobom minitabuľku:

	test MMSE	*%*
test MMSE	1	
Zlomok úspor	-0,79074	1

Farebne si zvýraznili hodnotu koeficientu korelácie ***r* = -0,79074**, čo predstavuje vysokú závislosť štatistických znakov. V ďalšom kroku v okne ***Analýza dát*** vybrali funkciu ***Regresia***, vložili do ***Vstupnej oblasti Y*** bunky s hodnotami zlomku úspor, ktoré seniori vložili do hazardu; do ***Vstupnej oblasti X*** bunky s hodnotami testu kognitívnych funkcií MMSE seniorov. Klikli si na okienka ***Popisky, Hladina spoľahlivosti 95%*** a ***Výstupná oblasť***, do ktorej vložili nejaké voľné políčko. Ešte klikli na ***Rezídua***, ***Štandardné rezídua*** a ***Graf regresnej priamky***, potvrdili OK. Dostali dosť rozsiahlu výstupnú zostavu, v ktorej si farebne zvýraznili najdôležitejšie potrebné údaje:

VÝSLEDEK

Regresní statistika	
Násobné R	0,790742
Hodnota spolehlivosti R	0,625273
Nastavená hodnota spolehlivosti R	0,61189
Chyba stř. hodnoty	0,165232
Pozorování	30

ANOVA

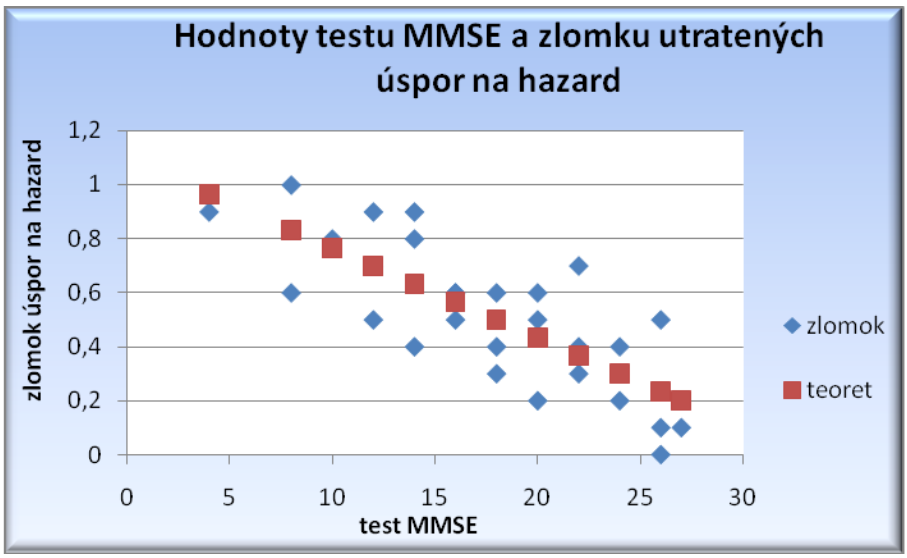
	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	1,275557	1,275557	46,72102	2E-07
Rezidua	28	0,764443	0,027302		
Celkem	29	2,04			

	Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95,0%*	*Horní 95,0%*
Hranice	1,099478	0,092747	11,85462	1,98E-12	0,909495	1,289461	0,909495	1,289461
test MMSE	-0,0333	0,004872	-6,83528	2E-07	-0,04329	-0,02332	-0,04329	-0,02332

REZIDUA

Pozorování	*Očekávané %*	*Rezidua*	*Normovaná rezidua*
1	0,966261	-0,06626	-0,40812
2	0,833043	0,166957	1,028323
3	0,833043	-0,23304	-1,43537
4	0,766435	0,033565	0,206736
5	0,699826	-0,19983	-1,23077
6	0,699826	0,200174	1,232917
7	0,699826	0,000174	0,001071
8	0,633217	-0,23322	-1,43644
9	0,633217	0,266783	1,643175
10	0,633217	0,166783	1,027252
11	0,566609	0,033391	0,205665
12	0,566609	-0,06661	-0,41026
13	0,566609	0,033391	0,205665
14	0,5	-0,2	-1,23185
15	0,5	0,1	0,615923
16	0,5	-0,1	-0,61592
17	0,433391	0,166609	1,026181
18	0,433391	-0,23339	-1,43751
19	0,433391	0,066609	0,410258
20	0,366783	-0,06678	-0,41133
21	0,366783	0,333217	2,052362
22	0,366783	0,033217	0,204593
23	0,300174	-0,00017	-0,00107
24	0,300174	0,099826	0,614852
25	0,300174	-0,10017	-0,61699
26	0,233565	-0,13357	-0,82266
27	0,233565	-0,23357	-1,43858
28	0,233565	0,266435	1,641033
29	0,200261	-0,10026	-0,61753
30	0,200261	-0,00026	-0,00161

a ešte jeden obrázok, ktorý si radšej upravili:



Obr.VIII.6: Grafická regresia (červene teoretická) z údajov skóre testu kognitívnych funkcií a zlomku vynaložených úspor na hazard vybraných seniorov

1.tabuľka, 1.riadok: Násobné R je absolútna hodnota výberového koeficientu korelácie

$|r| = 0{,}790742$

Hodnota predstavuje vysokú závislosť Y na X.

1.tabuľka, 2.riadok: Hodnota spoľahlivosti R je hodnota koeficientu determinácie

$r^2 = 0{,}625273$

Hodnota r^2 znamená, že viac ako 62,5% prípadov za uvedených podmienok sa dá pripísať vplyvu poklesu kognitívnych schopností seniorov na ochotu utratiť veľkú časť svojich úspor na hazardné hry.

Ďalší farebne zvýraznený riadok je Chyba strednej hodnoty: Je to hodnota $s = \sqrt{\frac{E^2}{n-2}}$, ktorú sme si uviedli vyššie a ktorá sa nazýva štandardná odchýlka reziduí, alebo regresnej priamky s

$s = 0{,}165232$

Pod ňou je rozsah výberového súboru: $n = 30$

Druhou tabuľkou je tabuľka analýzy rozptylu ANOVA: V 1.stĺpci sú počty stupňov voľnosti (1 pre regresiu), o niečo ďalej hodnota testovacieho kritéria Fisherovho testu F = 46,72. Ak by chceli odborníci mafie testovať, tak by použili hodnotu **významnosti F = 0.** Tá udáva pravdepodobnosť, ktorej by sa dopustili, keby zamietli H_0 o štatistickej nevýznamnosti lineárneho regresného modelu. Teda ak je rovná nule mohli H_0 zamietnuť na ľubovoľnej hladine významnosti.

Zaujímavejšou je tretia tabuľka hodnôt parametrov lineárneho regresného modelu **y = a.x+b**:

V 1.riadku a 1.stĺpci je hodnota koeficientu **b**, v 2.riadku a 1.stĺpci je hodnota koeficientu **a**.

a = -0,0333

b = 1,099478

Z toho sa dá už zostrojiť rovnicu priamky

$y = -0{,}0333.x + 1{,}099478$

V 2.stĺpci sú smerodajné odchýlky koeficientov

$\sigma_b = 0{,}092747$

$\sigma_a = 0{,}004872$

V ďalšom sú uvedené hodnoty testovacieho kritéria t Studentovho rozdelenia na testovanie H_0: b= 0, resp. H_0: a= 0 a hranice intervalu spoľahlivosti odhadu koeficientov na hladine významnosti $\alpha = 0{,}05$ ako sme si uviedli vo vstupnom excelovskom okienku.

Posledná tabuľka je tabuľka reziduí.

Vedenie rodiny sa nad výsledkami regresnej analýzy spokojne usmievalo, pretože uznávali dobrú poctivú prácu. Rovnica

$$y = -0,0333.x + 1,099478$$

s príslušnými štatistickými ozdôbkami ako sú smerodajné odchýlky a intervaly spoľahlivosti im hovorila, že pokiaľ ich klienti nemajú ohrozené kognitívne funkcie (skóre testu = 30) tak hráči sú ochotní minúť možno len 10% svojich úspor. Zhoršenie na polovicu (skóre = 15) už odhaduje schopnosť utratiť aspoň 60% svojich úspor, čo ďalším znižovaním ešte rastie. Zisky boli zaručené.

Keď si skupinka *Šmejdov* z našich zemepisných šírok išla za *Veľkú mláku* užívať svoje ťažko zarobené peniaze za takmer násilný predražený predaj bezcenného tovaru starým ľuďom, zarazili ich vo vstupných foyeroch kasín všadeprítomné fotografie starších ľudí, žiariacich šťastím a radosťou po veľkej výhre.

V konečnom dôsledku korelácia a regresia nebola až taká ťažká a dostali ste do rúk veľmi zaujímavý a silný nástroj pre vašu prácu. Odhalenie doteraz skrytých ale reálnych súvislostí, je vždy veľkým krokom v poznaní. Trochu náročnejšia je nelineárna regresia, teda spracovanie modelu, v ktorom závislosť medzi závislými znakmi má iný priebeh ako priamku. Nevyhýbajte sa tomu. Nie že by to nebolo riešiteľné, ale spočiatku je vhodné pri istom stupni náročnosti, aby ste sa v tom neutopili, vstúpiť do spolupráce s odborníkom, ktorý to má v pracovnej náplni a musel to vyštudovať, alebo ktorého to jednoducho z nejakého záhadného dôvodu baví.

Závislosti možno skúmať aj medzi kvalitatívnymi znakmi. V prvom kroku je nutnosť overiť, či vôbec existuje štatisticky významná závislosť medzi pozorovanými štatistickými znakmi. Miera závislosti sa potom posudzuje pomocou jej sily resp. intenzity, (viď [3], [4] a rôzna literatúra uvedená v predchádzajúcich kapitolách).

Najpoužívanejšie postupy na overenie závislosti dvoch kvalitatívnych znakov **K** a **L** je:

χ^2- test pre asociačnú tabuľku 2 x 2; (v rozšírení pre kontingenčnú tabuľku)

χ^2- test pre asociačnú tabuľku 2 x 2.

Všeobecný tvar tzv. asociačnej tabuľky 2x2 pri pozorovaní dvoch alternatívnych znakov **K** a **L** možno vo všeobecnosti zapísať v tvare:

K\L	1	0	Súčet
1	*a*	*b*	*a* + *b*
0	*c*	*d*	*c* + *d*
Súčet	*a* + *c*	*b*+ *d*	n

Tab. VIII.6.: Všeobecná asociačná tabuľka 2 x 2

n je počet pozorovaní. ***a,b,c,d*** sú početnosti výskytu všetkých 4 kombinácii úrovne javu **K** a **L**. ***1,0*** sú úrovne alternatívneho znaku (napr. muž - žena, áno - nie, kladný resp. záporný postoj a i.)

Testovacie kritérium pre H_0: Štatistické znaky **K** a **L** sú nezávislé, bude v prípade **n>40** (pre **20<n<40** musí byť každá z početnosti **a,b,c,d >5**, v opačnom prípade sa musí použiť **F-test**):

$$\chi^2 = \frac{n.(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} \qquad \textbf{[VIII.10]}$$

Testovacie kritérium sa porovná s tabuľkovou hodnotou $\chi^2_\alpha(1)$. H_0 o nezávislosti znakov **K** a **L** sa zamieta, ak $\chi^2 > \chi^2_\alpha(1)$ v prospech H_1 o závislosti štatistických znakov. Ak sa zistí závislosť, jej sila sa dá odhadnúť pomocou Cramerovho koeficientu kontingencie C:

$$C = \sqrt{\frac{\chi^2}{n+\chi^2}} \qquad \textbf{[VIII.11]}$$

Na pomoc je tu tabuľka:

C	Sila, súvislosť
0,0 – 0,1	žiadna
0,1 – 0,3	malá
0,3 – 0,5	stredná
nad 0,5	veľká

Tab. VIII.7.: Cramerov koeficient kontingencie a sila závislosti

Pr.VIII.5.:Pri sledovaní miery korupcie úradov v krajine bolo náhodným výberom anonymne oslovených 420 bežných občanov a 100 úradníkov s otázkou, či mali skúsenosť s korupciou úradníkov na jednej alebo druhej strane barikády. Výsledky sú v kontingenčnej tabuľke 2 x 2:

status	skúsenosť s korupciou		spolu
	áno	nie	
občan	410	10	420
úradník	5	95	100
spolu	415	105	520

Testovala sa nulová hypotéza H_0, že nie je štatistický rozdiel v skúsenosti s korupciou medzi bežnými občanmi a úradníkmi, oproti ťažšie logicky vysvetliteľnej alternatívnej H_1, že je medzi rôznymi statusmi štatisticky významný rozdiel, na hladine významnosti testu α=0,05.

Testovacie kritérium podľa **[VIII.10]** dáva

$$\chi^2 = \frac{n.(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} = \frac{520.(410.95-10.5)^2}{420.415.105.100} \cong 430$$

$$\chi^2_\alpha(1) = \chi^2_{0,05}(1) = 3{,}841.$$

H_0 o nezávislosti znakov **status** a **skúsenosť s korupciou úradníkov** sa zamieta, pretože $\chi^2 > \chi^2_\alpha(1)$ v prospech H_1 o závislosti štatistických znakov.

Silu tejto závislosti dá kontingenčná analýza pomocou Cramerovho koeficientu kontingencie C (vzťah VIII.11.):

$$C = \sqrt{\frac{\chi^2}{n+\chi^2}} = \sqrt{\frac{430}{520+430}} \cong 0{,}673$$

Na základe tabuľky VIII.7. možno usúdiť, že ide o veľkú závislosť. V niektorých prípadoch, keď máme alternatívne znaky, je špeciálnym prípadom kontingencie **asociácia**, potom predchádzajúcu kontingenčnú tabuľku možno nazvať asociačná tabuľka 2x2 a mierou závislosti je asociačný koeficient $C_{K,L}$:

$$C_{K,L} = \frac{ad-bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}} = \sqrt{\frac{\chi^2}{n}} \qquad [VIII.12]$$

s podobnou interpretáciou ako pri C. Pre závislosť medzi statusom občana resp. úradníka a jeho aktívnou či pasívnou skúsenosťou s korupciou úradníkov vychádza asociácia až $C_{K,L}$ = 0,909.

Je to dosť nepochopiteľné, prečo je taký obrovský rozdiel v tejto skúsenosti medzi bežnými občanmi (častá skúsenosť s korupciou úradníkov) a úradníkmi (takmer nulová skúsenosť), pretože nie je jasné, s kým občania korupčnú skúsenosť vlastne mali. Jedna možnosť je, že všetci občania z nejakých iracionálnych anarchistických a ťažko odhaliteľných dôvodov klamú, pretože krajina si vyberá do štátnej služby len kvalitných uchádzačov v zmysle zákona a cez mimoriadne náročné výberové konania. Navyše 5 úradníkov, ktorí uviedli svoju skúsenosť s korupciou bolo z jedného pracoviska a priznalo, že raz dávno, už je to určite premlčané, vzali spolu od stránky jednu celozrnnú tyčinku máčanú v čokoláde; stále sa rozhodujú, či ju majú zaniesť na Úrad boja proti korupcii.

Pr.VIII.6.: Po porážke fašizmu a skončení II. svetovej vojny nastal čas obnovy. V Nemecku okrem iného bola spoločnosť postavená pred otázky strategického významu a smerovania

v budúcnosti. Celonárodná diskusia bola zameraná na budovanie demokratickej spoločnosti. Bol akútny nedostatok odborníkov na čokoľvek, tak sa diskutovalo, či v niektorých dôležitých oblastiach nezamestnať aj ľudí zo „starých" štruktúr, aby to aspoň nejako fungovalo. Padali otázky, ako je možné budovať demokratickú spoločnosť bez dobrého súdnictva, či je v súdnictve možné zamestnať členov bývalej vládnej NSDAP, ktorých predtým súdili v zmysle požiadaviek totality, ako budú nestranní, ako sa odrazu obrátia a pomôžu demokraciu vybudovať. Do diskusie sa zapojili aj štatistici:

status	**zamestnať v súdnictve členov NSDAP**		**spolu**
	áno	**nie**	
občan	**1025**	**2810**	**3835**
člen NSDAP	**65**	**235**	**300**
spolu	**1090**	**3045**	**4135**

Urobili výber 4135 ľudí, ktorých sa pýtali na ich názor zamestnať v súdnictve bývalých sudcov fašistického režimu pre budovanie demokratickej spoločnosti a testovali H_0, či je na hladine významnosti α=0,05 štatistický rozdiel medzi názormi bežných občanov, ktorí boli väčšinou proti, a členov bývalej NSDAP.

Pomocou **[VIII. 10]** vypočítali testovaciu charakteristiku

$$\chi^2 = \frac{n.(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} \cong 3{,}689$$

Tabuľková hodnota $\chi^2_\alpha(1) = \chi^2_{0,05}(1) = 3{,}841$. H_0 o nezávislosti znakov **občan** a **člen bývalej NSDAP** sa nezamieta, pretože $\chi^2 < \chi^2_\alpha(1)$. Pre istotu urobili podľa **[VIII. 12]** aj výpočet pre asociačný koeficient $C_{K,L}$:

$$C_{K,L} = \sqrt{\frac{\chi^2}{n}} \cong 0{,}023$$

Bolo jasné, že všetky oblasti spoločnosti si uvedomili vážnosť situácie a spoločne sa vyjadrili k budúcnosti. V oblasti súdnictva, ale aj v ďalších oblastiach uskutočnili dôkladnú denacifikáciu spoločnosti. Po niekoľkých desaťročiach sa Nemecko stalo lídrom modernej Európy.

Ponechávame už na čitateľa radosť z ďalšieho testovania kvalitatívnych údajov, určite si príde na svoje. Veď príkladov na to nájde okolo seba neúrekom, stačí sa len poobzerať a mať naozaj otvorené oči.

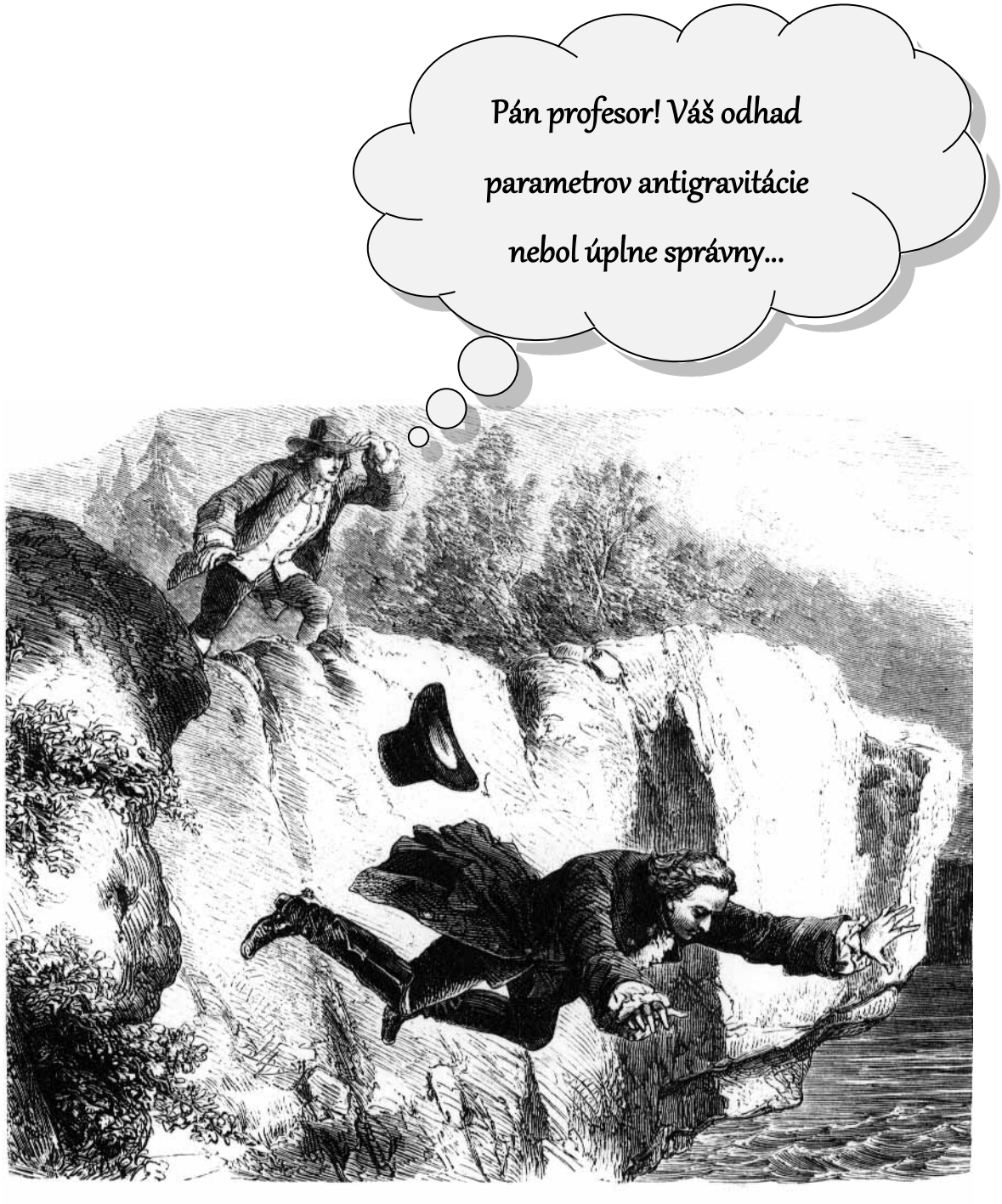
Literatúra k VIII. kapitole:

[1] *Yetti má červené oči*, denník Práca, 16.november 1987 + rôzne blogy

[2] Holloway, D.: Stalin a bomba. Academia, Praha, 2008

[3] Chajdiak, J.: Štatistika v EXCELI, STATIS, Bratislava 2002

[4] Anděl, J.: Matematická statistika. SNTL, Praha, 1985.


Pán profesor! Váš odhad
parametrov antigravitácie
nebol úplne správny...

IX. Zopár myšlienok o (ne)etike výskumu

Morálka je postoj voči ľuďom, ktorých nemáme radi.

Oscar Wilde

V posledných kapitolách sa chceme spolu s vami zamyslieť nad tým, čo je vo verejných aj neverejných diskusiách vo výskume často považované za „nutné zlo“ či „povinnú jazdu“. Budeme hovoriť o etike. Hneď v úvode je nevyhnutné povedať si na rovinu, že „etika výskumu“ je neoddeliteľnou súčasťou morálnej integrity výskumníka, najmä jeho dôveryhodnosti v postupoch výskumného procesu, vďaka ktorému sa dopracuje k platným výsledkom. V praxi to znamená, že ani tie najlepšie etické kódexy nemusia zaistiť bezpečie respondenta, pokiaľ nie je etika súčasťou zvnútornenej morálnej integrity výskumníka.

Samotný pojem etika pochádza z gréckeho „ethos“, čo znamená „charakter“ a jeho vzťah ku mravnosti, čo vo vzťahu k etike výskumu ponúka niekoľko dôležitých otázok. Tie by sme si mali položiť vždy, keď sa chceme pustiť do výskumu, a to buď pre nástojčivý pocit, že som zodpovedný akademik a „mal by som robiť aspoň nejaký výskum“, pre nátlak dekana či vedúceho katedry alebo, v tom najlepšom prípade, chcem zistiť niečo nové a dopátrať sa bližšie k pravde. Ponúkame vám pred realizáciou výskumu niekoľko „etických“ otázok, ku ktorým si neváhajte položiť ďalšie, ktoré vás práve napadnú:

- Bude váš výskum slúžiť v prospech skúmaného problému, prípadne ľudí, ktorí sa výskumu zúčastnia?
- Ktoré etické dilemy najviac ovplyvňujú váš výskum, hlavne výber výskumného problému?
- Aké etické princípy ovplyvňujú dizajn vášho výskumu, výber vzorky respondentov, metodológiu...?
- Máte informovaný súhlas respondentov s účasťou na vašom projekte? Chráni viac vás alebo skúmaného?
- Akými etickými princípmi sa budete riadiť pri rozhodovaní o tom, čo z výskumných zistení zverejníte?
- Aké filozofické, náboženské alebo ideologické východiská ovplyvňujú váš výskum? Majú vplyv na vašu zaujatosť?

Prečo je etika vo výskume dôležitá?

Najtragickejšie príklady potreby etiky výskumu vyplynuli z hrozných zverstiev druhej svetovej vojny, počas ktorej pod zámienkou vedeckého výskumu boli na Židoch, Rómoch či iných rasových/etnických či iných menšinách páchané v nacistických vyhladzovacích koncentračných táboroch zverstvá, aké si uprostred 20. storočia v humanisticky naladenej a kresťanskej Európe plnej „osvietenstva“ nevedel niekto ani len predstaviť. Výsledky odhalení desivých lekárskych experimentov na väzňoch schovávajúcich sa za výskum realizovaný v mene vedeckého pokroku bolo vytvorenie Norimberského kódexu (1949) – etického kódexu, ktorý okrem iného vo svojom úvode uvádza, že účasť na výskume musí byť vždy **dobrovoľná a vedomá**. Čoskoro potom nasledovali ďalšie etické kódexy vrátane Helsinskej deklarácie (1964), ktorá nariaďuje, že všetky výskumné projekty biomedicínskeho výskumu, na ktorých sa zúčastňujú ľudia, musia starostlivo zhodnocovať všetky možné a pravdepodobné riziká aj do budúcna, ktoré z výskumu pre účastníkov vyplývajú. Vznikli rôzne organizácie a rady, ktorých úlohou je kontrolovať etiku výskumu v tzv. rozvojových krajinách, prijímali sa etické kódexy na úrovni krajín, stavovských organizácií, univerzít alebo záujmových združení a etika sa postupne stáva aj v našej „etikou nepobozkanej“ časti strednej Európy neodmysliteľnou súčasťou všetkých serióznych výskumov.

Samozrejme, treba jedným dychom dodať, že extrémne prípady neetického správania sa výskumníkov, akého sme svedkami v totalitných režimoch, sú vo vedeckej komunite skôr výnimkou, ako pravidlom. Ich dôležitým odkazom pre každého výskumníka je poznanie, čo sa môže stať, ak etický rozmer výskumu prestane byť jeho neoddeliteľnou súčasťou. Oponenti môžu namietať, že zverstvá predsa páchali lekári a sociálni pracovníci, zdravotné sestry či psychológovia alebo pedagógovia by sa takéhoto niečoho nikdy nedopustili alebo dopustiť vzhľadom na charakter sociálneho výskumu nemohli. Aj v „našich“ vodách však existuje niekoľko príkladov porušenia etiky hoci s nie tak závažnými následkami ako boli tie v koncentračných táboroch.

Etika výskumu neexistuje vo vákuu

Jedeným z najznámejších sociálnych výskumných projektov, pri ktorých možno hovoriť o etickom zlyhaní, je výskum z roku 1963 o „poslušnosti voči autoritám“ pod vedením známeho amerického sociálneho a experimentálneho psychológa Stanleyho Milgrama. Tento odborník na duševné zdravie sa rozhodol pochopiť podmienky poslušnosti

jednotlivcov voči autoritám. Jeho výskumný protokol vyžadoval oklamať dobrovoľníkov. Experiment sa uskutočnil na Yalovej univerzite. Účastníkov získaval prostredníctvom inzerátu v miestnych novinách, v ktorom ponúkal odmenu 4,50 dolára. V reklame sa uvádzalo, že ide o experiment súvisiaci s učením. Účastníkom bol predstavený iný experimentálny subjekt, ktorý však bol v skutočnosti asistentom experimentátora. Potom boli dobrovoľníci náhodne rozdelení do dvojíc "učiteľ - žiak", pričom všetko bolo nastavené tak, aby skutočným respondentom bol učiteľ. "Učiteľ" videl, ako "žiaka" odviedli do vedľajšej miestnosti, pripútali ho na stoličku, aby z nej nemohol sám vstať, a na zápästie mu pripevnili elektródu napojenú na generátor šokov, ktorý ovládal "učiteľ". Generátor šokov nebol skutočný, ale vyzeral veľmi vierohodne a "učiteľ" nemal mať žiadne pochybnosti o jeho funkčnosti. "Žiak" mal reagovať pridaním príslušného slova do dvojice stlačením 1 zo 4 tlačidiel, čím sa na paneli pred "učiteľom" sediacim vo vedľajšej miestnosti rozsvietilo 1 zo 4 štvorcov.

"Učiteľ" mal k dispozícii mikrofón, pomocou ktorého kládol "žiakovi" otázky, panel, na ktorom sa rozsvietením jeho časti signalizovala odpoveď "žiaka", a generátor elektrických šokov pripojený k "žiakovi". Generátor mal 30 spínačov vedľa seba, označených zľava doprava od 15 V do 450 V s 15 V krokom medzi jednotlivými spínačmi. Okrem hodnoty vo voltoch boli spínače označené v skupinách po 4 slovách, pričom prvá skupina mala názov "ľahký šok" a posledná skupina "nebezpečenstvo: silný šok". Posledné dva spínače za poslednou skupinou boli označené iba XXX. "Učiteľovi" bolo povedané, že experiment bude skúmať vplyv trestu na učenie. V prípade nesprávnych odpovedí "študenta" mu "učiteľ" pomocou spínača na generátore šokov uštedril elektrický šok. Po nesprávnej odpovedi "učiteľ" vždy použil ďalší spínač v poradí, t. j. sila elektrického šoku sa po každej nesprávnej odpovedi zvýšila o 15 V. Experimentátor povedal "učiteľovi", že šoky môžu byť veľmi bolestivé, ale nezanechajú trvalé následky.

Odpovede "žiaka" boli vopred určené a pomer nesprávnych a správnych odpovedí bol približne 3 ku 1. Pri šoku 300 V sa "žiak" kopol do spoločnej steny a od tohto momentu v teste neodpovedal, ale experimentátor povedal "učiteľovi", že mlčanie sa považuje za nesprávnu odpoveď a v teste treba pokračovať. Pri 315 V "študent" opäť kopol do steny a od tej chvíle už o sebe nedával vedieť. Ak "učiteľ" nechcel pokračovať v teste, experimentátor ho vždy povzbudil rovnakými vopred určenými vetami v rovnakom poradí: "Prosím, pokračujte", "Experiment si vyžaduje, aby ste pokračovali", "Je nevyhnutné, aby ste pokračovali" a nakoniec "Nemáte na výber, musíte pokračovať".

Všetci účastníci experimentu pokračovali do 300 V a 65 % z nich dosiahlo koniec stupnice, napriek tomu, že sa často zdráhali a prejavovali obavy o zdravie "žiaka". Experiment sa neskôr opakoval v niekoľkých variantoch (percentá v zátvorkách uvádzajú, koľko respondentov pokračovalo až do konca) - "žiak" nevydával počas celého experimentu žiadny zvuk (100 %), pri 300 V "žiak" búchal do steny (65 %), "žiak" a "učiteľ" sa nachádzali v tej istej miestnosti (40 %), experiment sa uskutočnil v kancelárii, a nie v areáli školy (48 %), "učiteľ" drží ruku "obete" na elektróde (30 %), "učiteľ" dostáva príkazy od experimentátora cez telefón (21 %), "učiteľ" si môže zvoliť silu šoku (2,5 %), experimentátora nahradí bežný občan (20 %), experimentu sa zúčastňujú len ženy (65 %), pri experimente sú prítomní dvaja vedci a jeden z nich sa vzoprie pri 150 V (10 %).

Po experimente sa "učiteľ" a "študent" stretli a respondenti boli informovaní, že boli oklamaní. Napriek tomu bol Milgram kritizovaný za neetickú povahu výskumu. Riešili sa najmä tieto etické otázky:

1. Klamanie účastníkov
2. Silný stres, ktorý účastníci zažili
3. Experiment mohol mať dlhodobý vplyv na sebavedomie
4. Právo kedykoľvek odísť bolo porušené

Experiment Stanleyho Milgrama oklamal dobrovoľné subjekty zapojené do výskumu bez toho, aby od nich získal informovaný súhlas. Protokol tohto experimentu navyše neumožňoval pokusným osobám ukončiť experiment, aj keď niektoré z nich protestovali a žiadali, aby bol zastavený. Okrem toho niektorí aktéri prežívali nadmerný stres, v presvedčení, že môžu inému človeku spôsobiť smrť šokom. Tu je dôležité položiť si otázku: Ospravedlňuje konečný cieľ štúdie jeho prostriedky? Alebo trocha filozoficky: Je eticky správne dosahovať dobrý cieľ zlými prostriedkami? To sú otázky staré ako filozofia či teológia sama – teda asi ako ľudstvo samo. Oscilujú od jednoznačného stanoviska „dobrý cieľ nemožno dosiahnuť zlými prostriedkami“ cez „teórie menšieho zla či väčšieho dobra“ až po jednoduchý cynizmus „účel svätí prostriedky“.

Napriek legitímnym otázkam je požiadavka správať sa eticky nevyhnutnou podmienkou pri všetkých typoch výskumov, obzvlášť na ľuďoch. Povinnosťou výskumníka je preštudovať si etické kódexy svojej profesie, etické kódexy výskumu a prípady etického zlyhania výskumníkov, z ktorých možno čerpať ponaučenie. Pri čítaní etických kódexov netreba zabúdať na „ducha“ kódexu. Veľmi často sa totiž etické pravidlá vyznačujú skôr

dodržiavaním litery než samotného etického princípu – hodnoty, ktorú zastrešuje. Naša výskumnícka etika často začína a končí podpísaním informovaného súhlasu, ktorý neraz chráni viac výskumníka ako skúmaného. Túto skúsenosť vidno na Slovensku často pri výskumoch v marginalizovaných komunitách alebo pri ľuďoch, ktorí si pre vek či sociálne postavenie nevedia účinne vydobyť svoje práva (seniori ubytovaní v zariadeniach sociálnych služieb, ľudia s mentálnym postihnutím, ľudia so skúsenosťou s duševným ochorením a pod.).

Etika výskumu neexistuje vo vákuu. Na jednej strane je univerzálna, ale zároveň veľmi konkrétna vzhľadom na osobitú situáciu či špecifiká respondenta. Autonómia skúmaného, informovaný súhlas alebo dôvernosť či prospech výskumu pre skúmaného je vysoko individuálnou záležitosťou odvíjajúcou sa od skutočných okolností. Hlavnou etickou zásadou výskumu teda nie je ani tak, ČI sa uplatňuje morálny koncept vo výskume, ale AKO sa uplatňuje. To už je na svedomí každého z nás, do ktorého vám vstupovať nemôžeme a ani nechceme.

Záver
alebo nový začiatok

Všetko má svoj čas a svoju chvíľu, každé úsilie pod nebom.
Kaz 3,1

Na chvíľku sa obzrieme späť, len letmým pohľadom na cestu, ktorú sme spolu absolvovali. Dúfajme, že pritom neskamenieme ako Lótová žena, veď nás čaká ešte veľa práce. Niekedy sme sa cítili ako v labyrinte, ale podarilo sa nám z neho dostať. Bola to cesta, občas trochu hrboľatá a do kopca, cesta nesmierne zaujímavou zelenou krajinou elementárneho počtu pravdepodobnosti a štatistiky. Nezablúdili sme do nepreniknuteľných húštin, ani sme sa nešplhali na príliš vysoké končiare, náš výkon zodpovedal našej kondícii. Pre pochopenie často nových a málo zrozumiteľných pojmov boli vždy dopĺňané jednoduchými príkladmi. Dúfam, však, že sme si túto neobyčajnú krajinu obľúbili. Dostali sme základné vedomosti a schopnosti ako pracovať s niektorými objektmi, ako získať výsledok, ako ho interpretovať, hlavne v nasledujúcich oblastiach:

- Výpočet pravdepodobnosti nejakého javu na klasickej predstave pravdepodobnosti.
- Zber štatistických údajov, triedenie, tabuľkové spracovanie dát, početností a relatívnych početností. Štatistický súbor.
- Grafické spracovanie, koláčový, čiarový a stĺpcový graf údajov, histogram, grafy rozdelenia pravdepodobnosti a súčtové grafy.
- Popisné charakteristiky súboru, stredné hodnoty, charakteristiky variability a tvaru.
- Intervaly spoľahlivosti, výber, rozsah výberu.
- Hypotézy a testovanie štatistických hypotéz.
- Štúdium závislostí kvantitatívnych a kvalitatívnych štatistických javov, lineárna regresia, asociačné tabuľky.

Bol to prvý zoznamovací výlet, zostalo toho ešte naozaj dosť na pokročilejší kurz, ale hlavná myšlienka, základné pravidlo, ktoré sa tiahlo celou našou exkurziou ako červená niť a ktoré je potrebné si zvýrazniť znie:

V štatistike nie je matematika to, čoho sa treba obávať

Je zaujímavé, že množstvo rozumných ľudí vie, že pokiaľ by si chceli na klavíri zahrať Čajkovského koncert b-mol, alebo na gitare nejakú Bachovú skladbičku, napr. bourèe zo suity e-mol, BWV 996, musia sa naučiť niečo z teórie hudby, noty, stupnice, akordy, pozrieť sa ako to robili starí majstri, niečo z renesancie, z baroka, klasicizmu, romantizmu i moderny a veľa, veľa cvičiť. Podobne, keď si chcete zahrať hokej alebo futbal, musíte poznať pravidlá, získať nejakú techniku a trénovať. To sa týka vlastne každej činnosti. Pokiaľ sú však postavení pred nejaký pojem, ktorý zaváňa matematikou, okamžite sa stavajú do niektorej z dvoch extrémnych polôh: Na akýkoľvek výkon v matematike musí buď človek padnúť múdry z neba, alebo na to nemá. To samozrejme nie je pravda. Matematika je do značnej miery normálna ľudská činnosť, vyžadujúca poznanie pravidiel hry a isté naberanie kondície. Veríme, že aspoň 1σ - interval čitateľov tejto publikácie v jej závere už tomu rozumie. A nerezignuje. Pretože na rezignáciu sa spoliehajú práve tí, ktorí do štatistiky prenikli, ale vybrali sa neandertálskou cestou jej zneužitia.

Štatistika má vstupy, samotnú analýzu v užšom slova zmysle a výstupy. Neveľké množstvo aplikovanej matematiky prevažuje len vo fáze analýzy, ale to sa dá osvojiť si naozaj bez väčších či dokonca neriešiteľných problémov. Ťažšie je to v oblasti vstupov, teda už pri výskumnom zámere, pri výbere otázok, ktoré sa budú riešiť, metodík, ktoré sa budú používať, pri získavaní vstupných údajov a pri odhadoch možných chýb a problémov spojených s celým procesom. Tu platí druhé zlaté pravidlo:

Výsledok štatistickej analýzy nie je lepší ako vstupné údaje

Podobne je to s výstupmi. Interpretácia dosiahnutých výsledkov je neoddeliteľná súčasť štatistiky, pritom od matematiky je už dosť ďaleko. Platí tretie pravidlo:

Štatistika je taká, akí sú ľudia, ktorí ju robia

Táto publikácia sa nesústredila len na matematické problémy štatistickej analýzy, aj keď im dáva prioritu, ale v labyrinte metód sa snažila poukázať aj na to, na čo je potrebné si dať pozor. Je to rovnako dôležité, ako samotné výpočty. Neandertálci, ktorí sa nejakým zvláštnym „prirodzeným“ výberom často dostávajú na dôležité a vplyvné miesta v spoločnosti, nemajú strach z teroristov, z anarchistov, z rebelov či revolucionárov, ale určite ho majú zo štatistickej gramotnosti populácie. Takto sa štatistika stáva mocnou zbraňou v rukách aspoň trochu rozmýšľajúcich ľudí. A v tomto kontexte môže byť aj veľmi svieža a zábavná. Keď sa náhodne vyskytne akákoľvek odchýlka, fluktuácia od rovnovážnej hodnoty v málo vzdelanej spoločnosti, môže sa zdať oproti nízkej úrovni, z ktorej sa odchýlila, významná. Taká spoločnosť sa potom nestačí diviť, aké „osobnosti“ jej vo všetkých oblastiach, v politike, v kultúre a umení, vo vzdelávaní či riadení, v mediálnom prostredí i v športe, prudko vyrástli. Pokiaľ sa vyskytne významná fluktuácia na podstatne vyššej úrovni vzdelania a poznania, je v príslušnom pomere pravdepodobnejšie, že je naozaj významná.

Zhrňme si teda v závere aj niektoré poznatky o tom, ako nás štatistika, keď to pripustíme, valcuje, ako s nami manipuluje. A ako sa brániť, veď často ide o majetok, dokonca o život:

- S veľkou rezervou absorbujte do svojho poznania všetky bombastické a prakticky denne servírované štatistiky z mediálneho priestoru. Pokiaľ ide o komerčné a tobôž bulvárne média, tak je potrebné si uvedomiť, že im ide len o štatistiku sledovanosti resp. čítanosti pre zadávateľov reklamy.
- Opatrne narábajte so štatistikami predkladanými z prostredia politických subjektov v demokratickom systéme, ktoré neustále vedú boj o vašu pozornosť a v konečnom dôsledku aj o priazeň. Ich štatistiky, niekedy aj pravdivé, bývajú často zámerne neúplné. Pokiaľ sa nachádzate v systéme, ktorý je vzdialený od štandardov demokratického štátu, neverte vládnym a politickým štatistikám vôbec.
- Neverte príliš presným a ťažko dostupným štatistikám.
- Neverte holým priemerom, uvádzaným bez príslušnej variability, väčšinou niečo dôležité skrývajú.
- Neverte štatistikám, kde sa objektívne nedá urobiť vhodný (nestranný, objektívny, náhodný) výber.

- Neverte štatistikám bez potrebného vysvetlenia, hlavne v reklame. Pozor na grafy, často sa vás snažia vizuálne ovplyvniť.
- Veľmi opatrne prijímajte správy o rôznych zaujímavých až neuveriteľných závislostiach, koreláciách a pod. Niekedy stačí zistiť, či ide o naozaj závislé javy. Taktiež je dobre si uvedomiť, čo je príčina a čo dôsledok nejakého javu.
- Je dobre si vždy položiť niekoľko otázok: Kto to hovorí a čo hovorí? Je to zmysluplné, je to úplné alebo niečo chýba? Aký je výber? Dal sa urobiť nestranný výber? Bol dostatočný? Je interpretácia naozaj odpoveďou na pôvodnú otázku?
- Nepodliehajte štatistikám (t.j. nenakupujte, nemeňte prudko stravovacie zvyky a pod. pod ich vplyvom), pokiaľ im naozaj nerozumiete. Väčšinou sú za nimi komerčné záujmy.
- Všimnite si, že napriek tomu, že rastie vyspelosť medicíny a farmaceutického priemyslu, nie je populácia zdravšia, práve naopak, pocit „chorobnosti“ je neobyčajne vysoký.
- Uvedomte si, ako podliehalo zdravotníctvo v posledných desaťročiach módnym vlnám, uveďme niektoré: najprv to boli problémy výživy (napr. cholesterol, rôzne vitamíny), neskôr sa to priklonilo viac k ortopédii a kĺbom (nutnosť „vyživovať“ nadmerne opotrebované kĺby), následne k neurológii (Alzheimer, Parkinson a pod.), v súčasnosti badať trend smerom k psychiatri. Každá módna vlna priniesla mohutný vzostup príslušných ochorení a záchranu zo strany farmaceutických spoločností.

Určite si tento zoznam doplníte o vlastné skúsenosti s tým, ako na nás neandertálska štatistika priamo alebo nepriamo vplýva, ohlupuje nás, ťahá nám z peňaženky financie, alebo dokonca ohrozuje život. Tým sa štatistika môže stať aj návodom na prežitie, alebo aspoň na znižovanie vplyvu neandertálcov v spoločnosti.

Všetko to, čo nechceme aby robili iní nám, samozrejme nemôžeme robiť ani my iným. Takže normálne narábanie so štatistikou sa týka predovšetkým nás samotných. A v každej situácii, dokonca aj vtedy, keď sa to pre nás nezdá byť najvýhodnejším riešením.

Vybrali sme sa spolu na veľmi zaujímavú neľahkú púť, ktorú teraz niekto viac, niekto menej úspešne končíme. Bola to však predovšetkým cesta za dobrodružstvom poznania. Je to samozrejme len jedna z mnohých možných ciest, ktorými sa dalo ísť, a určite nie je bez omylov a nedostatkov, ako všetko ľudské počínanie. Možno jej vytknúť už samotný výber príkladov, udalostí a príbehov, ktoré mali na štatistiku vplyv; taktiež nedostatky v citáciách autorov, či v presnosti a precíznosti niektorých tvrdení a výpočtov. Viacerí čitatelia dokonca

nebudú s mnohými mojimi tvrdeniami súhlasiť a vtedy budem spokojný, pretože v nich boli vyvolané záchvevy kritického rozumu, a to je omnoho cennejšie ako fakt, že sa to prípadne obráti proti mne.

Spolu sme si mohli overiť, ako to v štatistike funguje, ako sa našim snažením dopracujeme k poznatkom, postupom a vedomostiam, ktoré môžu ľuďom pomáhať, ale aj škodiť. Prešli sme od prvých nesmelých a nezrelých pokusov pochopiť základné metódy pravdepodobnosti a štatistiky až k súčasným trochu sofistikovanejším a po všetkých stránkach náročným projektom a postupom výskumu. Je však toho ešte omnoho viac, čo nevieme.

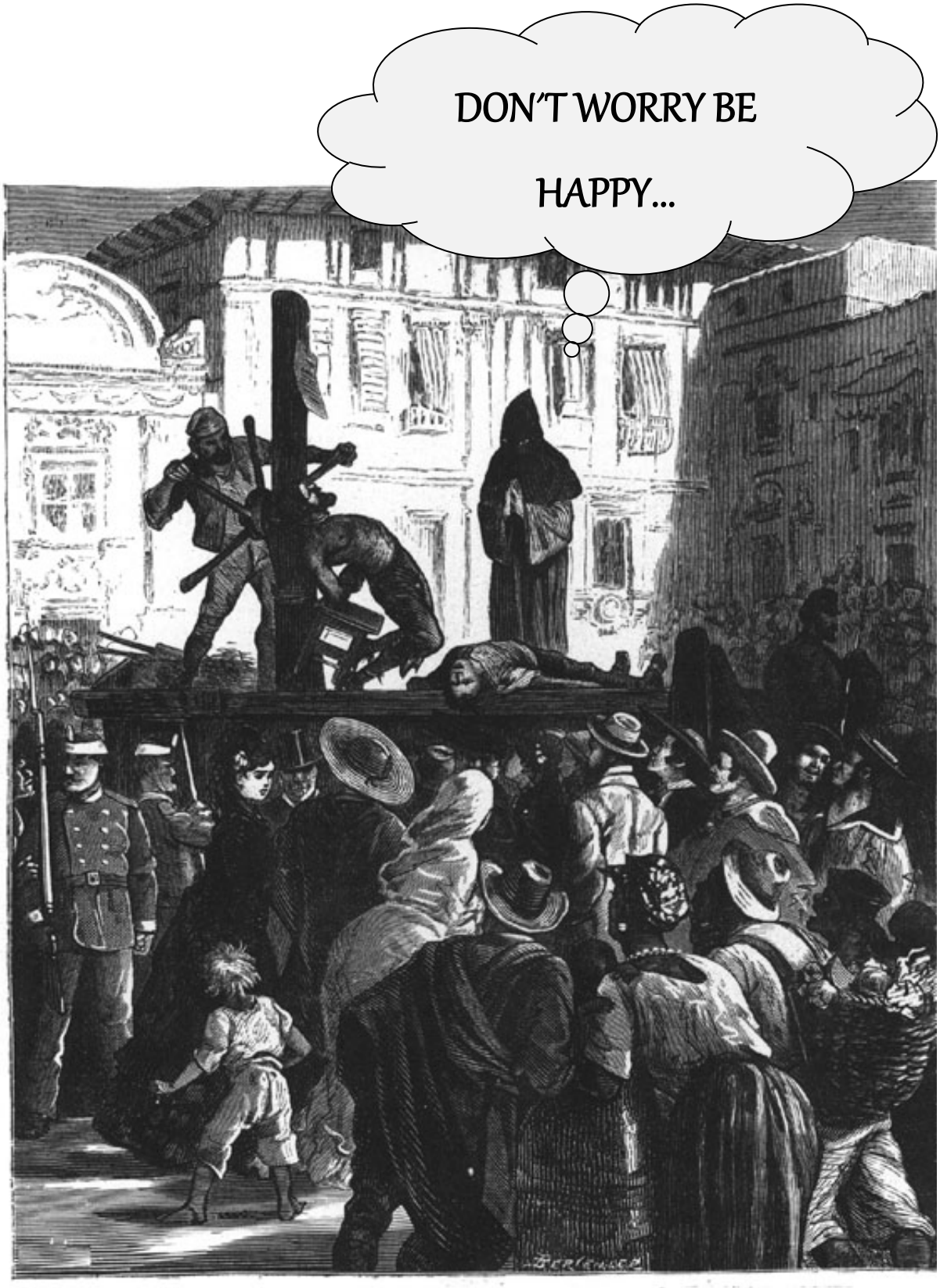
Štatistika, ako sme častejšie zdôrazňovali, nie je výklad sveta. A určite neplatí, že kto rozumie štatistike, rozumie všetkému alebo všetko vie. Pokiaľ naše poznanie do určitej miery vzrástlo, dvojnásobne by mala narásť naša skromnosť. Páči sa nám jedna legenda o otcovi kybernetiky, ktorý bol aj veľkým priekopníkom matematiky náhodných procesov, o americkom vedcovi Norbertovi Wienerovi (1894-1964), známom svojou zábudlivosťou. Keď sa jeho rodina presťahovala do nového domu, manželka mu pre istotu napísala novú adresu na kus papiera.

„Prosím ťa, neblázni," – ohradil sa Wiener – „veď niečo tak dôležitého nemôžem zabudnúť!"

Papier si však strčil do vrecka. Neskôr ho zaujal nejaký matematický alebo štatistický problém a ako ho riešil, popísal všetky papiere, ktoré mal poruke, a aj vo vreckách. Papiere sa povaľovali po celej pracovni, ale keď chcel ísť večer domov, odrazu nevedel kam. Nakoniec sa pobral, ako jediné možné riešenie, na svoju starú adresu, kde na schodišti pred domom sedelo malé dievčatko.

„Prepáč maličká, ale nevieš náhodou, kam sa presťahovali Wienerovci?"

„To je v poriadku, oci, mama ma poslala, aby som ťa doviedla domov..."


DON'T WORRY BE
HAPPY...

T a b u ľ k y

Tabuľka P1: Kvantily normovaného normálneho rozdelenia N(0,1)

P	u_p	P	u_p	P	u_p	P	u_p
0,50	0,000000000	0,75	0,674490	0,950	1,644854	0,975	1,959964
0,51	0,025068908	0,76	0,706303	0,951	1,654628	0,976	1,977368
0,52	0,050153583	0,77	0,738847	0,952	1,664563	0,977	1,995393
0,53	0,075269862	0,78	0,772193	0,953	1,674665	0,978	2,014091
0,54	0,100433721	0,79	0,806421	0,954	1,684941	0,979	2,033520
0,55	0,125661347	0,80	0,841621	0,955	1,695398	0,980	2,053749
0,56	0,150969215	0,81	0,877896	0,956	1,706043	0,981	2,074855
0,57	0,176374165	0,82	0,915365	0,957	1,716886	0,982	2,096927
0,58	0,201893479	0,83	0,954165	0,958	1,727934	0,983	2,120072
0,59	0,227544977	0,84	0,994458	0,959	1,739198	0,984	2,144411
0,60	0,253347103	0,85	1,036433	0,960	1,750686	0,985	2,170090
0,61	0,279319034	0,86	1,080319	0,961	1,762410	0,986	2,197286
0,62	0,305480788	0,87	1,126391	0,962	1,774382	0,987	2,226212
0,63	0,331853346	0,88	1,174987	0,963	1,786613	0,988	2,257129
0,64	0,358458793	0,89	1,226528	0,964	1,799118	0,989	2,290368
0,65	0,385320466	0,900	1,281552	0,965	1,811911	0,990	2,326348
0,66	0,412463129	0,905	1,310579	0,966	1,825007	0,991	2,365618
0,67	0,439913166	0,910	1,340755	0,967	1,838424	0,992	2,408916
0,68	0,467698799	0,915	1,372204	0,968	1,852180	0,993	2,457263
0,69	0,495850347	0,920	1,405072	0,969	1,866296	0,994	2,512144
0,70	0,524400513	0,925	1,439531	0,970	1,880794	0,995	2,575829
0,71	0,553384720	0,930	1,475791	0,971	1,895698	0,996	2,652070
0,72	0,582841507	0,935	1,514102	0,972	1,911036	0,997	2,747781
0,73	0,612812991	0,940	1,554774	0,973	1,926837	0,998	2,878162
0,74	0,643345405	0,945	1,598193	0,974	1,943134	0,999	3,090232

Pre **P<0,5** platí vzťah $u_p = - u_{1-p}$

Tabuľka P2: Distribučná funkcia N(0;1). Pre x<0: $\Phi(-x) = 1 - \Phi(x)$
Pre kvantily N(0;1) platí: $x_p = -x_{1-p}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,500000	0,503989	0,507978	0,511966	0,515953	0,519939	0,523922	0,527903	0,531881	0,535856
0,1	0,539828	0,543795	0,547758	0,551717	0,555670	0,559618	0,563559	0,567495	0,571424	0,575345
0,2	0,579260	0,583166	0,587064	0,590954	0,594835	0,598706	0,602568	0,606420	0,610261	0,614092
0,3	0,617911	0,621720	0,625516	0,629300	0,633072	0,636831	0,640576	0,644309	0,648027	0,651732
0,4	0,655422	0,659097	0,662757	0,666402	0,670031	0,673645	0,677242	0,680822	0,684386	0,687933
0,5	0,691462	0,694974	0,698468	0,701944	0,705401	0,708840	0,712260	0,715661	0,719043	0,722405
0,6	0,725747	0,729069	0,732371	0,735653	0,738914	0,742154	0,745373	0,748571	0,751748	0,754903
0,7	0,758036	0,761148	0,764238	0,767305	0,770350	0,773373	0,776373	0,77935	0,782305	0,785236
0,8	0,788145	0,791030	0,793892	0,796731	0,799546	0,802337	0,805105	0,807850	0,81057	0,813267
0,9	0,815940	0,818589	0,821214	0,823814	0,826391	0,828944	0,831472	0,833977	0,836457	0,838913
1,0	0,841345	0,843752	0,846136	0,848495	0,850830	0,853141	0,855428	0,857690	0,859929	0,862143
1,1	0,864334	0,866500	0,868643	0,870762	0,872857	0,874928	0,876976	0,879000	0,881000	0,882977
1,2	0,884930	0,886861	0,888768	0,890651	0,892512	0,894350	0,896165	0,897958	0,899727	0,901475
1,3	0,903200	0,904902	0,906582	0,908241	0,909877	0,911492	0,913085	0,914657	0,916207	0,917736
1,4	0,919243	0,920730	0,922196	0,923641	0,925066	0,926471	0,927855	0,929219	0,930563	0,931888
1,5	0,933193	0,934478	0,935745	0,936992	0,938220	0,939429	0,940620	0,941792	0,942947	0,944083
1,6	0,945201	0,946301	0,947384	0,948449	0,949497	0,950529	0,951543	0,952540	0,953521	0,954486
1,7	0,955435	0,956367	0,957284	0,958185	0,959070	0,959941	0,960796	0,961636	0,962462	0,963273
1,8	0,964070	0,964852	0,965620	0,966375	0,967116	0,967843	0,968557	0,969258	0,969946	0,970621
1,9	0,971283	0,971933	0,972571	0,973197	0,973810	0,974412	0,975002	0,975581	0,976148	0,976705
2,0	0,977250	0,977784	0,978308	0,978822	0,979325	0,979818	0,980301	0,980774	0,981237	0,981691
2,1	0,982136	0,982571	0,982997	0,983414	0,983823	0,984222	0,984614	0,984997	0,985371	0,985738
2,2	0,986097	0,986447	0,986791	0,987126	0,987455	0,987776	0,988089	0,988396	0,988696	0,988989
2,3	0,989276	0,989556	0,989830	0,990097	0,990358	0,990613	0,990863	0,991106	0,991344	0,991576
2,4	0,991802	0,992024	0,992240	0,992451	0,992656	0,992857	0,993053	0,993244	0,993431	0,993613
2,5	0,993790	0,993963	0,994132	0,994297	0,994457	0,994614	0,994766	0,994915	0,995060	0,995201
2,6	0,995339	0,995473	0,995604	0,995731	0,995855	0,995975	0,996093	0,996207	0,996319	0,996427
2,7	0,996533	0,996636	0,996736	0,996833	0,996928	0,997020	0,997110	0,997197	0,997282	0,997365
2,8	0,997445	0,997523	0,997599	0,997673	0,997744	0,997814	0,997882	0,997948	0,998012	0,998074
2,9	0,998134	0,998193	0,998250	0,998305	0,998359	0,998411	0,998462	0,998511	0,998559	0,998605
3,0	0,998650	0,998694	0,998736	0,998777	0,998817	0,998856	0,998893	0,998930	0,998965	0,998999
3,1	0,999032	0,999065	0,999096	0,999126	0,999155	0,999184	0,999211	0,999238	0,999264	0,999289
3,2	0,999313	0,999336	0,999359	0,999381	0,999402	0,999423	0,999443	0,999462	0,999481	0,999499
3,3	0,999517	0,999534	0,999550	0,999566	0,999581	0,999596	0,999610	0,999624	0,999638	0,999651

Tabuľka P3: Kvantily Studentovho t – rozdelenia pre k–stupňov voľnosti, pre obojstranný t-test (2α)

k\p	0,9	0,95	0,975	0,99	0,995
1	3,077684	6,313752	12,70620	31,82052	63,65674
2	1,885618	2,919986	4,302653	6,964557	9,924843
3	1,637744	2,353363	3,182446	4,540703	5,840909
4	1,533206	2,131847	2,776445	3,746947	4,604095
5	1,475884	2,015048	2,570582	3,364930	4,032143
6	1,439756	1,943180	2,446912	3,142668	3,707428
7	1,414924	1,894579	2,364624	2,997952	3,499483
8	1,396815	1,859548	2,306004	2,896459	3,355387
9	1,383029	1,833113	2,262157	2,821438	3,249836
10	1,372184	1,812461	2,228139	2,763769	3,169273
11	1,363430	1,795885	2,200985	2,718079	3,105807
12	1,356217	1,782288	2,178813	2,680998	3,054540
13	1,350171	1,770933	2,160369	2,650309	3,012276
14	1,345030	1,761310	2,144787	2,624494	2,976843
15	1,340606	1,753050	2,131450	2,602480	2,946713
16	1,336757	1,745884	2,119905	2,583487	2,920782
17	1,333379	1,739607	2,109816	2,566934	2,898231
18	1,330391	1,734064	2,100922	2,552380	2,878440
19	1,327728	1,729133	2,093024	2,539483	2,860935
20	1,325341	1,724718	2,085963	2,527977	2,845340
21	1,323188	1,720743	2,079614	2,517648	2,831360
22	1,321237	1,717144	2,073873	2,508325	2,818756
23	1,319460	1,713872	2,068658	2,499867	2,807336
24	1,317836	1,710882	2,063899	2,492159	2,796939
25	1,316345	1,708141	2,059539	2,485107	2,787436
26	1,314972	1,705618	2,055529	2,478630	2,778715
27	1,313703	1,703288	2,051830	2,472660	2,770683
28	1,312527	1,701131	2,048407	2,467140	2,763262
29	1,311434	1,699127	2,045230	2,462021	2,756386
30	1,310415	1,697261	2,042272	2,457262	2,749996

Tabuľka P4: Kvantily χ^2 – rozdelenia pre k–stupňov voľnosti

k\p	0,001	0,005	0,01	0,025	0,05	0,1
1	1,5708E-06	3,92704E-05	0,000157	0,000982	0,003932	0,015791
2	0,00200100	0,010025084	0,020101	0,050636	0,102587	0,210721
3	0,02429759	0,071721775	0,114832	0,215795	0,351846	0,584374
4	0,09080404	0,206989093	0,297109	0,484419	0,710723	1,063623
5	0,21021260	0,411741904	0,554298	0,831212	1,145476	1,610308
6	0,38106676	0,675726778	0,872090	1,237344	1,635383	2,204131
7	0,59849375	0,989255685	1,239042	1,689869	2,167350	2,833107
8	0,85710483	1,344413088	1,646497	2,179731	2,732637	3,489539
9	1,15194955	1,734932909	2,087901	2,700390	3,325113	4,168159
10	1,47874346	2,155856482	2,558212	3,246973	3,940299	4,865182
11	1,83385267	2,603221895	3,053484	3,815748	4,574813	5,577785
12	2,21420933	3,073823653	3,570569	4,403789	5,226029	6,303796
13	2,61721815	3,565034584	4,106915	5,008751	5,891864	7,041505
14	3,04067253	4,074674969	4,660425	5,628726	6,570631	7,789534
15	3,48268448	4,600915599	5,229349	6,262138	7,260944	8,546756
16	3,94162785	5,142205451	5,812213	6,907664	7,961646	9,312236
17	4,41609273	5,697217119	6,407760	7,564186	8,671760	10,08519
18	4,90484882	6,264804719	7,014911	8,230746	9,390455	10,86494
19	5,40681605	6,843971456	7,632730	8,906517	10,11701	11,65091
20	5,92104075	7,433844283	8,260398	9,590778	10,85081	12,44261
21	6,44667658	8,033653456	8,897198	10,28290	11,59131	13,23960
22	6,98296847	8,642716463	9,542492	10,98232	12,33801	14,04149
23	7,52923981	9,260424795	10,19572	11,68855	13,09051	14,84796
24	8,08488159	9,886233535	10,85636	12,40115	13,84843	15,65868
25	8,64934365	10,51965217	11,52398	13,11972	14,61141	16,47341
26	9,22212686	11,16023749	12,19815	13,84391	15,37916	17,29189
27	9,80277697	11,80758738	12,87850	14,57338	16,15140	18,11390
28	10,3908792	12,46133599	13,56471	15,30786	16,92788	18,93924
29	10,9860535	13,12114895	14,25645	16,04707	17,70837	19,76774
30	11,5879511	13,78671996	14,95346	16,79077	18,49266	20,59923

Tabuľka P4: Kvantily χ^2 – rozdelenia pre k–stupňov voľnosti (pokračovanie)

k\p	0,9	0,95	0,975	0,99	0,995	0,999
1	2,70554397	3,841459149	5,023886	6,634897	7,879439	10,82757
2	4,60517019	5,991464547	7,377759	9,21034	10,59663	13,81551
3	6,25138846	7,814727764	9,348404	11,34487	12,83816	16,26624
4	7,77944034	9,487729037	11,14329	13,2767	14,86026	18,46683
5	9,23635694	11,07049775	12,8325	15,08627	16,7496	20,51501
6	10,6446407	12,59158724	14,44938	16,81189	18,54758	22,45774
7	12,0170366	14,06714043	16,01276	18,47531	20,27774	24,32189
8	13,3615661	15,50731306	17,53455	20,09024	21,95495	26,12448
9	14,6836566	16,91897762	19,02277	21,66599	23,58935	27,87716
10	15,9871792	18,30703805	20,48318	23,20925	25,18818	29,5883
11	17,2750085	19,67513757	21,92005	24,72497	26,75685	31,26413
12	18,5493478	21,02606982	23,33666	26,21697	28,29952	32,90949
13	19,8119293	22,3620325	24,7356	27,68825	29,81947	34,52818
14	21,0641442	23,68479131	26,11895	29,14124	31,31935	36,12327
15	22,3071296	24,99579013	27,48839	30,57791	32,80132	37,6973
16	23,5418289	26,29622761	28,84535	31,99993	34,26719	39,25235
17	24,7690353	27,58711164	30,19101	33,40866	35,71847	40,79022
18	25,9894231	28,86929943	31,52638	34,80531	37,15645	42,3124
19	27,2035711	30,14352721	32,85233	36,19087	38,58226	43,8202
20	28,4119806	31,41043286	34,16961	37,56623	39,99685	45,31475
21	29,6150894	32,67057337	35,47888	38,93217	41,40106	46,79704
22	30,8132823	33,92443852	36,78071	40,28936	42,79565	48,26794
23	32,0068997	35,17246163	38,07563	41,6384	44,18128	49,72823
24	33,1962443	36,4150285	39,36408	42,97982	45,55851	51,1786
25	34,381587	37,65248413	40,64647	44,3141	46,92789	52,61966
26	35,5631712	38,88513865	41,92317	45,64168	48,28988	54,05196
27	36,7412168	40,11327205	43,19451	46,96294	49,64492	55,47602
28	37,9159226	41,33713813	44,46079	48,27824	50,99338	56,89229
29	39,0874698	42,55696777	45,72229	49,58788	52,33562	58,30117
30	40,2560238	43,77297178	46,97924	50,89218	53,67196	59,70306


To je zvláštne?! Pri modelových cvičeniach všetko fungovalo...

„Väčšina ľudí je vždy na strane väčšiny.“
Tomáš Janovic

„Že je publikácia napísaná „s láskou a humorom“, tak ako jej autori uvádzajú v podnázve, je viditeľné v každej jej vete. Láska k matematike a asi aj k antike, veľká rozhľadenosť autorov a netradičný (naozaj humorný) štýl ju robia výnimočnou. Odmyslieť si štatistiku a kniha sa dá čítať aj ako príjemná a zaujímavá beletria. K jej atraktivite prispieva grafika, vtipné „meme“ starých obrázkov a jej celkový dizajn.“

doc. PhDr. Jurina Rusnáková, PhD.
Ústav romologických štúdií, Fakulta sociálnych vied a zdravotníctva
Univerzita Konštantína Filozofa v Nitre

„Dôležitou súčasťou publikácie v každej jej časti sú príklady využitia spracovaných tematických okruhov pri štúdiu, či praxi pomáhajúcich profesii. Študenti, či pomáhajúci profesionáli v paxi môžu pomocou týchto príkladov rozšíriť svoje vedeckovýskumné kompetencie, ako a jrealizovať vedeckovýskumné aktivity s vysokým spoločenským významom, ktoré budú prispievať k identifikácii sociálnych problémov, či efektívnemu nastaveniu možných sociálnych intervencií.“

prof. PaedDr. Peter Jusko, PhD.
Pedagogická fakulta
Univerzita Mateja Bela v Banskej Bystrici

ISBN: 978-80-8132-273-0