Általános és gazdasági statisztika

Csugány Julianna

SZÉCHENYI 2020

MAGYARORSZÁG KORMÁNYA

Európai Unió
Európai Szociális Alap

BEFEKTETÉS A JÖVŐBE

MÉDIAINFORMATIKAI KIADVÁNYOK

Általános és gazdasági statisztika

Csugány Julianna

Líceum Kiadó
Eger, 2015

Hungarian Online University – Ágazati informatikai együttműködés létrehozása az új típusú e-learning alapú képzések hazai és nemzetközi elterjesztésére

TÁMOP-4.1.1.C-12/KONYV-2012-0003

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Lektorálta:
dr. Tánczos Tamás

Szerkesztette: Csugány Julianna

ISBN 978-615-5509-64-3

Felelős kiadó: dr. Kis-Tóth Lajos
Készült: az Eszterházy Károly Főiskola nyomdájában, Egerben
Vezető: Kérészy László
Műszaki szerkesztő: Nagy Sándorné

Tartalom

1. Bevezetés ____ *5*

1.1 Célkitűzések, kompetenciák és a tantárgy teljesítésének feltételei ____ **5**

1.1.1 Célkitűzés ____ 5
1.1.2 Kompetenciák ____ 6
1.1.3 A tantárgy teljesítésének feltételei ____ 7

1.2 A kurzus tartalma ____ **8**

1.3 Tanulási tanácsok, tudnivalók ____ **9**

1.4 Statisztikai alapfogalmak ____ **9**

1.4.1 Statisztikai adatforrások ____ 12

2. A statisztikai adatgyűjtés és adatrendszerezés formái ____ *14*

2.1 Célkitűzések és kompetenciák ____ **14**

2.2 Tananyag ____ **14**

2.2.1 Mintavételi technikák ____ 15
Véletlenen alapuló mintavétel ____ 16
Nem véletlenen alapuló mintavétel ____ 17
2.2.2 Statisztikai adatrendszerezés ____ 18
A statisztikai sor ____ 19
A statisztikai tábla ____ 23

2.3 Összefoglalás, kérdések ____ **24**

2.3.1 Összefoglalás ____ 24
2.3.2 Önellenőrző kérdések ____ 25
2.3.3 Gyakorló tesztek ____ 26

3. Viszonyszámok ____ *27*

3.1 Célkitűzések és kompetenciák ____ **27**

3.2 Tananyag ____ **27**

3.2.1 Megoszlási és koordinációs viszonyszámok ____ 28
3.2.2 Dinamikus viszonyszámok ____ 30
3.2.3 Területi/térbeli összehasonlító viszonyszámok ____ 34
3.2.4 Tervfeladat és tervteljesítési viszonyszámok ____ 34
3.2.5 Intenzitási viszonyszámok ____ 35

3.3 Összefoglalás, kérdések ____ **37**

3.3.1 Összefoglalás ____ 37
3.3.2 Önellenőrző kérdések ____ 38
3.3.3 Gyakorló tesztek ____ 38

4. A viszonyszámok gyakorlati alkalmazása: munkaerő-piaci és pénzügyi statisztikák _____ 40

4.1 Célkitűzések és kompetenciák _____ 40

4.2 Tananyag _____ 40

4.2.1 Munkaerő-piaci statisztikák _____ 41

Aktivitási ráta _____ 42

A foglalkoztatási és a munkanélküliségi ráta _____ 43

4.2.2 Pénzügyi statisztikák _____ 44

Viszonyszámok alkalmazása a részvények értékelésénél _____ 45

A vállalati teljesítmény értékelése pénzügyi mutatókkal _____ 45

4.2.3 Demográfiai mutatók _____ 47

4.3 Összefoglalás, kérdések _____ 48

4.3.1 Összefoglalás _____ 48

4.3.2 Önellenőrző kérdések _____ 48

4.3.3 Gyakorló tesztek _____ 49

5. Leíró statisztika _____ 50

5.1 Célkitűzések és kompetenciák _____ 50

5.2 Tananyag _____ 50

5.2.1 Középértékek _____ 51

Számított középértékek: az átlagok _____ 51

A számtani átlag tulajdonságai _____ 54

Helyzeti középértékek _____ 57

Medián _____ 59

Kvartilisek _____ 59

5.2.2 Szóródási mérőszámok _____ 60

Terjedelem _____ 60

Szórás és relatív szórás _____ 61

Átlagos eltérés és átlagos különbség _____ 62

5.2.3 Alakmutatók _____ 63

Az aszimmetria (ferdeség) mérőszámai _____ 64

Csúcsosság _____ 65

5.3 Összefoglalás, kérdések _____ 66

5.3.1 Összefoglalás _____ 66

5.3.2 Önellenőrző kérdések _____ 66

5.3.3 Gyakorló tesztek _____ 66

6. Összetett viszonyszámok (főátlagok) összehasonlítása standardizálás segítségével _____ 69

6.1 Célkitűzések és kompetenciák _____ 69

6.2 Tananyag _____ 69

6.2.1 A heterogenitás kezelése: csoportképzés 70
6.2.2 A standardizáláson alapuló különbségfelbontás módszere 72
6.2.3 A standardizáláson alapuló indexszámítás módszere 75

6.3 Összefoglalás, kérdések 78
6.3.1 Összefoglalás 78
6.3.2 Önellenőrző kérdések 78
6.3.3 Gyakorló tesztek 79

7. Értéken alapuló indexszámítás 82

7.1 Célkitűzések és kompetenciák 82

7.2 Tananyag 82
7.2.1 Időbeli összehasonlítás 83
Egyedi indexek 84
Együttes indexek 86
7.2.2 Területi összehasonlítás 89

7.3 Összefoglalás, kérdések 91
7.3.1 Összefoglalás 91
7.3.2 Önellenőrző kérdések 92
7.3.3 Gyakorló tesztek 92

8. Az indexszámítás gyakorlati alkalmazása: a vállalati teljesítmény mérése, nemzetgazdasági indikátorok, külkereskedelmi statisztikák 94

8.1 Célkitűzések és kompetenciák 94

8.2 Tananyag 94
8.2.1 A vállalati teljesítmény számszerűsítése indexek segítségével 95
8.2.2 A nemzetgazdasági teljesítmény számszerűsítése indexek segítségével 96
8.2.3 Az infláció 98
8.2.4 Külkereskedelmi statisztikák 100

8.3 Összefoglalás, kérdések 100
8.3.1 Összefoglalás 100
8.3.2 Önellenőrző kérdések 101
8.3.3 Gyakorló tesztek 101

9. Ismérvek közötti kapcsolat vizsgálata I.: asszociáció és vegyes kapcsolat 103

9.1 Célkitűzések és kompetenciák 103

9.2 Tananyag 103

9.2.1 Asszociáció 105
Yule mutató 105
Csuprov és Cramer mutató 106
9.2.2 Vegyes kapcsolat 108
Szórásnégyzet-hányados – H^2 mutató 109
Szóráshányados – H mutató 109
9.3 Összefoglalás, kérdések 112
9.3.1 Összefoglalás 112
9.3.2 Önellenőrző kérdések 112
9.3.3 Gyakorló tesztek 113
10. Ismérvek közötti kapcsolat vizsgálata II.: korreláció- és regressziószámítás 115
10.1 Célkitűzések és kompetenciák 115
10.2 Tananyag 115
10.2.1 Korrelációs mérőszámok 116
Kovariancia (C) 116
Pearson-féle lineáris korrelációs együttható (r) 117
Lineáris determinációs együttható (r^2) 117
Korrelációs hányados (η) 118
Rangkorreláció 118
10.2.2 Kétváltozós lineáris regresszió 123
A tapasztalati (mintabeli) regresszió függvény paramétereinek meghatározása és értelmezése 123
Rugalmassági együttható (elaszticitás) 124
KOVARIANCIA 127
PEARSON-FÉLE LINEÁRIS KORRELÁCIÓS EGYÜTTHATÓ 127
LINEÁRIS DETERMINÁCIÓS EGYÜTTHATÓ 127
RUGALMASSÁGI EGYÜTTHATÓ 128
Regressziós becslések és a modell tesztelése 129
10.2.3 Nemlineáris regresszió 130
10.2.4 Illeszkedésvizsgálat 132
10.2.5 Többváltozós regresszió 133
10.3 Összefoglalás, kérdések 134
10.3.1 Összefoglalás 134
10.3.2 Önellenőrző kérdések 134
10.3.3 Gyakorló tesztek 135
11. Idősorelemzési technikák 136
11.1 Célkitűzések és kompetenciák 136
11.2 Tananyag 136
11.2.1 Átlagos abszolút és relatív változás 138
11.2.2 A determinisztikus idősorelemzés 140

Az alapirányzat meghatározása 142
A szezonalitás 153
Additív szezonalitás 158
Multiplikatív szezonalitás 159
Előrejelzés 159
11.2.3 A trendilleszkedés vizsgálata 162

11.3 Összefoglalás, kérdések 163
11.3.1 Összefoglalás 163
11.3.2 Önellenőrző kérdések 164
11.3.3 Gyakorló tesztek 164

12. A statisztikai elemzés szemléltető eszköze: a grafikus ábrázolás 167

12.1 Célkitűzések és kompetenciák 167

12.2 Tananyag 167
12.2.1 A grafikus ábrázolás tartalmi és formai követelményei 168
12.2.2 A grafikus ábrák típusai 169
Oszlop és sáv/szalagdiagram 169
Hisztogram 173
Vonaldiagram 174
Pontdiagram 175
Sugárdiagram vagy poláris görbe 176
Kör vagy tortadiagram 177
A térkép, mint a statisztikai ábrázolás segédeszköze 178

12.3 Összefoglalás, kérdések 180
12.3.1 Összefoglalás 180
12.3.2 Önellenőrző kérdések 181
12.3.3 Gyakorló tesztek 181

13. Összefoglalás 183

13.1 Tartalmi összefoglalás 183

13.2 A statisztika, mint tudományos módszertan hasznosíthatósága a gyakorlatban 185

14. Kiegészítések 186

14.1 Irodalomjegyzék 186
Könyv 186
Elektronikus dokumentumok / források 186

14.2 Médiaelemek összesítése 187
14.2.1 Táblázatjegyzék 187
14.2.2 Ábrajegyzék 189
14.2.3 Külső URL hivatkozások 191

15. Tesztek ___ ***194***

Gyakorlótesztek ___ **194**

Próbavizsga ___ **200**

Záróvizsga ___ **204**

1. BEVEZETÉS

1.1 CÉLKITŰZÉSEK, KOMPETENCIÁK ÉS A TANTÁRGY TELJESÍTÉSÉNEK FELTÉTELEI

1.1.1 Célkitűzés

A gazdasági és társadalmi folyamatok megismeréséhez, megértéséhez számos információra van szükség, melyek feldolgozásához jelentős segítséget nyújt a statisztika eszköztára. A számszaki vizsgálatok lehetőséget teremtenek az adatsorok tömör jellemzésére, megalapozott következtetések megfogalmazására, valamint tendenciák és összefüggések matematikai alapon történő alátámasztására is. A statisztikai munka az adatok összegyűjtésével és rendszerezésével kezdődik, melyet az adatelemzés követ. Az eredmények értelmezése, esetleges illusztrálása és a következtetések megfogalmazása zárja a vizsgálatot.

Az általános és gazdasági statisztika tárgy keretében a hallgató megismeri és elsajátítja

- a statisztikai alapfogalmakat és a speciális szakmai szókincset
- a statisztikai adatgyűjtés és adatrendszerezés formáit
- a viszonyszámok típusait és a képzésük logikáját
- a viszonyszámok gyakorlati alkalmazási területeit, kiemelten a munkaerő-piaci, pénzügyi és demográfiai mutatókat
- a leíró statisztika eszköztárát, az alapeloszlás jellemzők tulajdonságait és értelmezését
- az összetett viszonyszámok (főátlagok) standardizáláson alapuló összehasonlításának lehetőségeit
- az értéken alapuló indexszámítás módszerét
- az indexszámítás gyakorlati alkalmazási területeit, kiemelten a vállalati és nemzetgazdasági teljesítmény mérésének lehetőségeit, valamint a külkereskedelmi statisztikákat
- az ismérvek közötti kapcsolat elemzésének eszközeit, az asszociáció, a vegyes kapcsolat, valamint a korreláció és regressziószámítás módszerét
- azokat az idősorelemzési technikákat, melyek segítségével előrejelzés is készíthető
- az elemzési eredmények grafikus ábrázolásának alapelveit és formáit.

1.1.2 Kompetenciák

Az általános és gazdasági statisztika tárgy képzési céljainak eléréshez tematikusan különböző kompetenciák elsajátítása szükséges.

A hallgató

- ismerje és pontosan használni is tudja a statisztikai alapfogalmakat
- ismerje a releváns statisztikai adatforrásokat, melyekből szekunder információkat gyűjthet a társadalmi és gazdasági jelenségek vizsgálatához
- ismerje a statisztikai adatgyűjtés és adatrendszerezés formáit és legyen képes a tanultakat alkalmazni
- sajátítsa el a viszonyszámok képzési formáját, s képes legyen egyszerű elemzésre a viszonyszámok különböző formáival
- legyen képes olyan gazdasági és társadalmi problémák felismerésére és megoldására, melyeknél a viszonyszámok alkalmazhatók
- alkalmazni tudja a leíró statisztika eszköztárát és képes legyen értelmezni az alapeloszlás jellemzőket
- legyen képes összetett viszonyszámok (főátlagok) standardizáláson alapuló összehasonlítására, a különbséget és változást okozó tényezők hatásának számszerűsítésére
- sajátítsa el az értéken alapuló indexszámítás módszerét, legyen képes egy vagy több termék esetén is az ár, mennyiség és érték változását relatív formában is meghatározni
- ismerje fel azokat a területeket, melyeken az indexszámítás módszerei alkalmazhatók
- sajátítsa el az ismérvek közötti kapcsolatok vizsgálatának módszereit, meg tudja különböztetni a sztochasztikus kapcsolatokat
- alkalmazni és értelmezni tudja az asszociáció, a vegyes kapcsolat és a korreláció mérőszámait
- legyen képes mennyiségi ismérvek közötti kapcsolat kétváltozós lineáris regresszió függvénnyel való jellemzésére
- alkalmazni tudja azokat az idősorelemzési technikákat, melyekkel előrejelzéseket is tud készíteni
- sajátítsa el a grafikus ábrázolás alapelveit, legyen képes alkalmazni a különböző típusú ábrákat elemzési eredményeinek illusztrálására.

1.1.3 A tantárgy teljesítésének feltételei

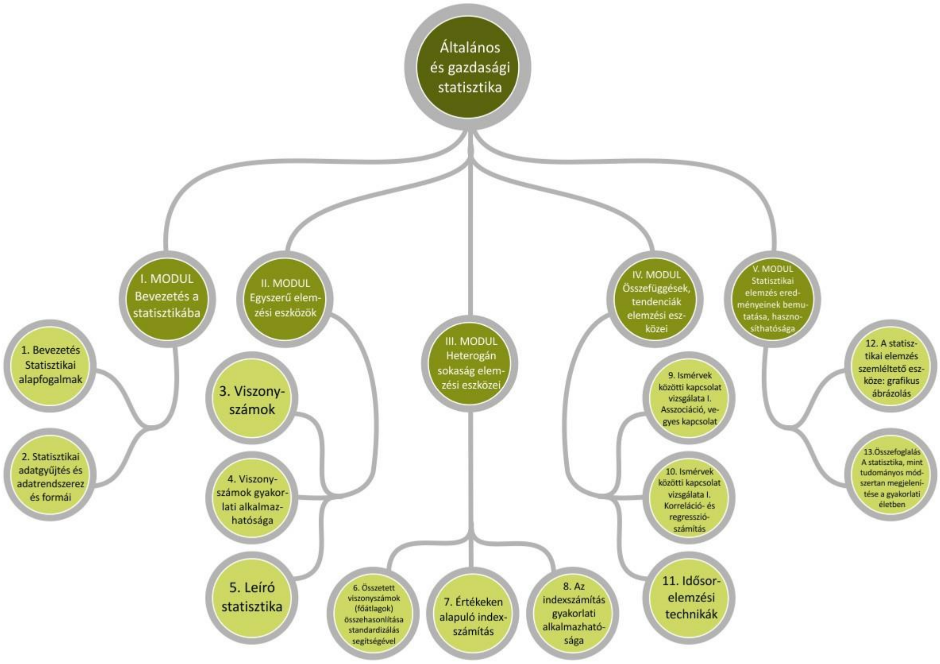
Az általános és gazdasági statisztika tárgy teljesítésének feltétele, hogy a hallgató

- ismeri a statisztikai alapfogalmakat, valamint az adatrendszerezés formáit, így a statisztikai sorok és táblák típusait, sajátosságait
- ki tudja választani adott jelenség vizsgálatánál a megfelelő mintavételi technikát
- átlátja a viszonyszámok rendszerét, meg tudja nevezni a gyakorlati életben előforduló viszonyszámok típusait
- ismeri a viszonyszámként képezhető főbb munkaerő-piaci, pénzügyi és demográfiai mutatókat, s ezek meghatározásának módját
- átlátja a leíró statisztika eszköztárát, ismeri a mutatók tulajdonságait, s tudja értelmezni is az eredményeket
- ismeri az összetett viszonyszámok standardizáláson alapuló összehasonlításának módszereit, átlátja a különbséget és változást okozó tényezők közötti összefüggéseket, s képes értelmezni a komponenseket
- tudja az érték statisztikai meghatározásának módját és felismeri a megjelenési formáit, továbbá ismeri az egyedi és együttes ár-, érték- és volumenindexek számítási módjait, értelmezését és összefüggéseit
- ismeri az indexszámítás gyakorlati alkalmazási területeit, a vállalati és nemzetgazdasági teljesítmény mérésére használt mutatókat, azok tartalmát és számítási formáit
- ismeri az asszociáció és a vegyes kapcsolat mérőszámait, tudja mikor melyik mutató alkalmazható, s értelmezni tudja a kapott eredményeket
- ismerni és értelmezni tudja a korrelációs kapcsolat kimutatására alkalmas mérőszámokat
- rendelkezik a regresszió számításhoz szükséges alapismeretekkel, említeni tud regresszió típusokat, értelmezni tudja egy kétváltozós lineáris regresszió paramétereit, valamint dönteni tud arról, hogy melyik típusú függvény illeszkedik legjobban a tapasztalati értékekre
- elsajátítja a determinisztikus idősorelemzés alapjait, jellemezni tudja az idősor tényezőit, valamint ismeri az alapirányzat és szezonalitás meghatározásának módjait, s tudja a közöttük lévő összefüggéseket
- ismeri a grafikus ábrázolás alapelveit, valamint a grafikus ábrák típusait és alkalmazási lehetőségeit.

1.2 A KURZUS TARTALMA

Az általános és gazdasági statisztika tananyag öt, egymásra jól építkező modulból áll. Az első modul a statisztika alapjaival ismerteti meg a hallgatókat, az alapfogalmakkal, valamint az adatgyűjtés és adatrendszerezés formáival. A második, harmadik és negyedik modul keretein belül tematikusan kerülnek bemutatásra az elemzési módszerek, előbb az egyszerű elemzési eszközök, majd a heterogén sokaság vizsgálati lehetőségei, valamint az összefüggések és tendenciák kimutatásának statisztikai módszerei. Az utolsó modul keretbe zárja a tananyagot az elemzési eredmények bemutatási lehetőségeinek, valamint a tananyag gyakorlati hasznosíthatóságának ismertetésével.

A tananyag végén található kiegészítés részben a forrásmunkák, valamint az ábrák, táblázatok és médiaelemek összesítése van feltüntetve. Az utolsó, Tesztek c. fejezetben a tananyag elsajátítását segítendő feladatok kaptak helyet, melyekkel felmérhető, hogy mennyire sikerült elsajátítani a statisztikai ismereteket.



1. ábra: Az általános és gazdasági statisztika tananyag logikai struktúrája

1.3 TANULÁSI TANÁCSOK, TUDNIVALÓK

A leckék elején gondolattérkép segíti a tananyagban való eligazodást és megkönnyíti átlátni a strukturált módszertani eszköztárakat. A fejezetek végén önellenőrző kérdések és gyakorló feladatok vannak, melyek hozzájárulnak a tananyag hatékony elsajátításához és a tárgy sikeres teljesítéséhez.

1.4 STATISZTIKAI ALAPFOGALMAK

A statisztikai módszerek megismerése előtt tisztázni kell azokat a fogalmakat, melyeket a későbbiek során rendre használni fogunk. Az alapfogalmak megértése és megfelelő használata elengedhetetlen az elemzési eszköztár alkalmazásához.

A statisztikai sokaság a vizsgálni kívánt jelenség valamennyi elemének összessége.

A statisztikai sokaság elemeinek számát tekintve lehet *véges vagy végtelen számú*. A gyakorlatban végtelen számú sokasággal ritkán találkozunk. A statisztikai sokaságot megkülönböztethetjük aszerint is, hogy az adatok adott időpontra vagy időszakra vonatkoznak. Az *álló sokaság* (stock jellegű) adatai befejezett állapotot tükröznek, melyek egy adott időpontban értelmezhetők vagy egy múltbeli időszakra vonatkozó megfigyelés eredményei, így például egy ország népességi adata. A *mozgó sokaság* (flow jellegű) folyamatosságot tükröz, az adatok időszakra értelmezhetők, így például egy utca forgalma egy kiválasztott napon. A statisztikai sokaság tipizálható az elemek egyértelmű meghatározhatósága alapján is. Ha a statisztikai sokaság elemei egyértelműen elkülöníthetők, akkor *diszkrét a sokaság*. Diszkrét sokaság például a népesség, melynek elemei egyértelműen meghatározhatók és jól körülhatárolhatók, az emberek. Ha a sokaság elemeit önkényesen határozzuk meg, vagyis nem egyértelmű az elhatárolás, akkor *folytonos sokaság*ról beszélünk. Folytonos sokaság esetén az alapegység megválasztása önkényes, így például az üdítőital fogyasztás elemzésénél az elfogyasztott mennyiség kifejezhető literben vagy deciliterben, vagyis az űrmérték megválasztása tetszőleges.

A statisztikai sokaság elemei különböző tulajdonságokkal rendelkeznek, melyek alapján összehasonlíthatók és csoportosíthatók.

A statisztikai sokaság elemeinek tulajdonságait ismérvnek nevezzük.

A statisztikai ismérv lehet *közös vagy megkülönböztető* attól függően, hogy a sokaság minden elemére, vagy csak az elemek egy szűkebb körére vonatkozó tulajdonság. Az ismérv attól függően, hogy az elemek milyen típusú jellemzőjét írja le, lehet időbeli, területi, mennyiségi és minőségi. Az *időbeli ismérv* az elemek valamilyen időponthoz vagy időszakhoz kötődő jellemzője. A *területi ismérv* a sokaság elemeit földrajzilag beazonosítható területi egységhez köti. A *mennyiségi ismérv* az elemek számszerűsíthető, mértékegységgel rendelkező jellemzője, míg a minőségi ismérv minden olyan tulajdonság, mely nem tartozik bele az előbbi három kategóriába. A mennyiségi ismérv lehet *diszkrét*, ha az ismérvváltozatok egész számmal fejezhetők ki és nincs köztük átmenet, vagy *folytonos*, ha nem csak egész számokkal fejezhetők ki, van közöttük átmenet.

☸ Jellemezze különböző típusú ismérvek segítségével egy statisztika kurzus hallgatóit!

1. *Példa az ismérvek típusaira*

Az ismérv típusa	Az ismérv megnevezése
Időbeli	A hallgatók születési ideje
Területi	A hallgatók lakhelye
Mennyiségi	A hallgatók testmagassága
Minőségi	A hallgatók neme

Adott ismérv alapján a sokaságban előforduló kimeneteleket ismérvváltozatoknak nevezzük.

Az adott sokaságban, ha egy adott ismérv esetén csak két ismérvváltozat lehetséges, akkor *alternatív ismérv*ről beszélünk. Ilyen például a nem, mint minőségi ismérv, melynek lehetséges ismérvváltozatai a férfi és a nő. Gyakoribb azonban, ha egy sokaságon belül több ismérvváltozattal rendelkező ismérvek alapján különböztetjük meg az elemeket, mely bővebb elemzésre ad lehetőséget. A mennyiségi ismérvek ismérvváltozatait *ismérvérték*eknek nevezzük, hiszen ezek az elemek számértékekkel kifejezhető tulajdonságai.

A matematikai alapokon nyugvó statisztikai módszereket nem csupán számszerűsíthető tulajdonságok esetén alkalmazhatjuk. Lényeges azonban, hogy tisztában legyünk az egyes ismérvek mérési lehetőségeivel, melyek alapjaiban határozzák meg az alkalmazható módszereket. A könnyebb kezelhetőség érdekében a nem számszerű ismérvváltozatokhoz is rendelünk számokat.

Adott ismérv esetén az ismérvváltozatokhoz rendelhető számértékeket mérési skálákon helyezzük el.

A mérési skálák jelentősége a komplexebb elemzéseknél értékelődik fel, ha több, különböző mérési szintű változó közötti kapcsolatot kívánjuk elemezni. Négyféle mérési szintet különböztetünk meg. A mérési skálák hierarchikus sorrendben épülnek egymásra, a felsőbb mérési szint rendelkezik az alatta lévő valamennyi tulajdonságával. Az egyes ismérvtípusok mérési szintekhez kapcsolódnak.

Nominális (névleges) skálán az ismérvváltozatok megkülönböztetésére van lehetőség, jellemzően minőségi és területi ismérveknél használjuk. A számokkal való jellemzés megkönnyíti az elemzéseket, mert átláthatóbbá teszi az adatok sokaságát, de ez esetben az értékek matematikailag nem használhatók. Nominális skálán mérhető minőségi ismérv például a rendszám, ahol a számoknak nincs matematikai tartalma, csupán különbségtételre alkalmas. *Sorrendi (ordinális) skálán* az ismérvértékek sorrendje is megállapítható. Sorrendi skálán mérhető például a vevőelégedettség vagy az osztályzatok. Az ismérvértékek között tehát megállapítható, hogy melyik a jobb vagy a kedvezőbb, tehát az összehasonlítás mellett bővebb elemzésre is ad lehetőséget. Érdekes példa a lottó nyerőszámok mérési szintje, hiszen könnyen beleshetünk abba a tévedésbe, hogy tekintettel arra, hogy sorba állítják a számokat, vagy sorba húzzák őket, rossz skálára helyezzük az értékeket. Lényeges, hogy a sorrendiség nem csupán matematikai, hanem tartalmi kérdés is. Könnyen belátható, hogy statisztikailag nem a számok sorrendjének van jelentősége, hanem annak, hogy a számok egymástól különböznek, vagyis nincs mögöttes tartalma annak, ha nagyobb számot húznak, így a lottó nyerőszámokat nominális skálán mérjük. *Intervallumskálán* az ismérvértékek sorrendje mellett a különbségük is értelmezhető, két érték távolsága számszerűsíthető, vagyis nemcsak azt tudjuk megállapítani, hogy valamely érték jobb, hanem azt is, hogy mennyivel, ugyanakkor az értékek nem összegezhetők. A hőmérséklet az intervallumskálán mérhető ismérvek tipikus példája. Vegyük például, hogy az egyik napon $+20°\mathrm{C}$, a másik napon $+15°\mathrm{C}$ volt a hőmérséklet. A hőmérsékleti értékeket intervallumskálán mérve megállapíthatjuk, hogy $5°\mathrm{C}-\mathrm{kal}$ csökkent a hőmérséklet, vagyis a különbségük értelmezhető, de az értékek nem összegezhetők, illetve egyéb matematikai művelet sem végezhető el. Az *arányskála* a legmagasabb mérési szint, melyen csak mennyiségi ismérvek változatai mérhetők, ugyanis ezen a skálán az ismérvértékek között valamennyi matematikai művelet elvégezhető. Arányskálán mérhető például a forgalom, az árbevétel, melyek összegezhetők, képezhető a különbségük, szorzatuk, s az ismérvértékek arányosíthatók is.

A statisztikai munka a vizsgálni kívánt jelenséghez kötődő adatok összegyűjtésével kezdődik, melyek származhatnak teljes körű vagy részleges adatfelvételből, illetve primer vagy szekunder adatforrásból. A statisztikai elemzések során statisztikai adatokkal, illetve statisztikai mutatószámokkal dolgozunk.

A statisztikai adat a sokaság számszerű jellemzője vagy a sokaság elemeinek száma.

A statisztikai adatok konkrét jelenséghez kötődnek, térben, időben vagy egyéb formában, tehát nem csupán számértékek. Statisztikai adat például az, hogy egy adott vállalat éves árbevétele 12 000 EUR, lényeges, hogy ahhoz a konkrét vállalathoz kapcsolódik. Szintén statisztikai adat az, hogy Magyarországon 2010-ben 6520 gyerek született. Statisztikai adatnak tehát az a számérték minősül, mely kapcsolódik egy konkrét sokasághoz. A statisztikai adatok lehetnek abszolút adatok, melyek közvetlen mérés, illetve megszámolás útján jönnek létre. Matematikai művelet során is létrehozhatóak statisztikai adatok, melyek származtatott értékek, ilyen lehet például, hogy egy országban az átlagos gyümölcsfogyasztás 62 kg vagy a népesség és terület viszonyítása során létrejövő népsűrűségi érték.

Az ismétlődő jelenségek statisztikai jellemzésére szolgáló statisztikai adatokat statisztikai mutatószámoknak nevezzük.

1.4.1 Statisztikai adatforrások

A gazdasági és társadalmi jelenségek jellemzésére számos statisztikai mutatószámot használunk. Hivatalos statisztikai adatforrásnak minősülnek a nemzetközi szervezetek adatbázisai, melyekből eltérő országokra, illetve országcsoportokra, valamint különböző tématerületekre vonatkozóan gyűjthetünk adatokat. Ugyanarra a jelenségre vonatkozó adat – legjellemzőbb példája az országok GDP-je – eltérő lehet attól függően, hogy mely adatbázisban találjuk meg. A számbavételi és módszertani sajátosságokat, illetve egy-egy statisztikai mutatószám tartalmát az adott szervezet oldalán célszerű megnézni, mely az adatok felhasználását is befolyásolhatja.

Az európai országokra vonatkozóan változatos mutatóstruktúrával rendelkezik az Európai Unió statisztikai hivatala, az Eurostat.

1. Eurostat: http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/

A világ országainak legszélesebb körű adatbázisával az OECD (Organization for Economic Co-operation and Development) rendelkezik, valamint az Egyesült Nemzetek Szervezete (United Nations) és a Világbank (Worldbank):

2. OECD adatbázisa: http://stats.oecd.org/
3. ENSZ adatbázisa: http://www.un.org/en/databases/
4. Világbank adatbázisa: http://data.worldbank.org/

Vannak olyan adatbázisok, melyek egy-egy tématerületen végeznek részletesebb adatgyűjtést. Elsősorban pénzügyi, de ezen túl számos gazdasági indikátort tartalmaz a Nemzetközi Valutaalap (International Monetary Fund, IMF) adatbázisa.

5. Nemzetközi Valutaalap adatbázisa: https://www.imf.org/external/data.htm

A Nemzetközi Valutaalapnak speciális pénzügyi statisztikákat tartalmazó adatbázisa is van.

6. International Financial Statistics: http://www.econdata.com/databases/imf-and-other-international/ifs/

A nemzetközi munkaerő-piaci statisztikák fontos forrása a Nemzetközi Munkaügyi Szervezet (International Labour Organization, ILO):

7. Nemzetközi Munkaügyi Szervezet adatbázisa: http://www.ilo.org/global/statistics-and-databases/lang--en/index.htm

A nemzetközi kereskedelem területén gyakran használt adatforrás az ENSZ másik szakosodott adatbázisa:

8. United Nations Conference on Trade and Development: http://unctad.org/en/Pages/Publications/Handbook-of-Statistics.aspx

2. A STATISZTIKAI ADATGYŰJTÉS ÉS ADATRENDSZEREZÉS FORMÁI

2.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A statisztikai munka első lépése az adatgyűjtés, melyet az adatok rendszerezése követ. Az adatok a vizsgálni kívánt jelenségtől függően primer és/vagy szekunder adatforrásból származhatnak. A statisztikai módszerek kiválasztásához az összegyűjtött adatok megfelelő rendszerezésére van szükség. A statisztikai adatrendszerezés alapegysége a statisztikai sor, amelyekből statisztikai tábla állítható össze.

A lecke célja, hogy megismertesse a hallgatókat a statisztikai adatgyűjtés és adatrendszerezés különböző formáival. A hallgató ismerje a statisztikai adatszerzés primer és szekunder formáit, a különböző mintavételi technikákat és azok jellemzőit. Gyakorlati példák segítségével kerülnek bemutatásra az adatgyűjtési módok, melyek hozzásegítik a hallgatót, hogy adott jelenség vizsgálata kapcsán képes legyen kiválasztani a megfelelő mintavételi technikát. Kiemelten fontos, hogy a hallgató tudja az adatrendszerezés alapvető formáit és kritériumait, mely alapot szolgáltat a megfelelő elemzési technika kiválasztásához.

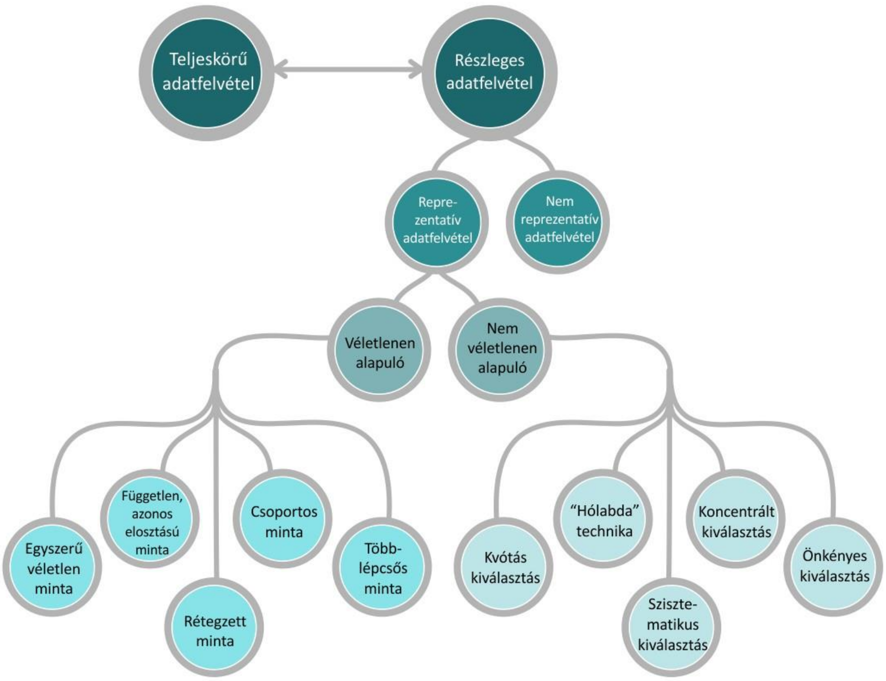
2.2 TANANYAG

A statisztikai adatokkal szemben hármas kritériumot támasztunk: „pontosan, gyorsan, olcsón”. Ezen követelmények alatt azt értjük, hogy az adat legyen elfogadható pontosságú, kellően gyorsan beszerezhető, s ily módon időben álljon rendelkezésre, valamint a lehető legkisebb ráfordítással lehessen hozzájutni. Ezen követelményeknek egyszerre megfelelni nehezen lehet, de törekedni kell rá. Az adatszerzés történhet primer vagy szekunder adatforrásból. Szekunder (másodlagos) forrásnak minősülnek a különféle vállalati kimutatások, nyilvántartások, hivatalok, így például a statisztikai hivatal, az önkormányzatok, vagy a jegybank kiadványai. Primer (elsődleges) adatforrások esetén az elemző maga gyűjti össze a szükséges adatokat, melyeket később vizsgálataihoz felhasznál. Ezen adatforrásoknál mindig szem előtt kell tartani, hogy az adatfelvételnél lehetnek hibák, melyek a későbbi statisztikai vizsgálatok eredményeit torzíthatják. A gyakorlatban ritkán van módunk arra, hogy egy jelenség kapcsán valamennyi elemről rendelkezzünk információval, így a sokaságnak legtöbbször csak egy részét figyeljük meg, s különböző módszerekkel igyekszünk általánosítható, a sokaság egészére vonatkozó következtetéseket levonni.

A minta a vizsgálni kívánt jelenség által érintett sokaság elemeinek egy része.

2.2.1 Mintavételi technikák

Ha a vizsgálni kívánt jelenség kapcsán valamennyi elemre vonatkozóan rendelkezünk információval, akkor teljes körű adatfelvételről beszélünk. Ez csak véges számú sokaság esetén lehetséges és nagyon ritkán fordul elő, ilyen például a népszámlálás[1]. A legtöbb esetben a sokaság egészének jellemzése időigényes és költséges, ezért részleges adatfelvétel történik, melynek során a sokaságnak csak egy részére vonatkozóan vannak adataink. A részleges adatfelvétel során tudatosan kell megtervezni a mintavétel folyamatát. A reprezentativitás kritériuma a statisztikai elemzések elengedhetetlen feltétele, így a mintavételi formáknál a reprezentatív adatfelvétel módozataira helyezzük a hangsúlyt. Ha nem törekszünk általános érvényű következtetések megállapítására, csupán azt helyezzük előtérbe, hogy az adott mintában milyen összefüggések mutathatók ki, nem reprezentatív megfigyelést végzünk.



2. ábra: A statisztikai adatszerzés formái

[1] A népszámlálás esetén teljes körű mintavételre törekszünk, a sokaság minden elemére vonatkozóan gyűjtünk információt, de a válaszadási hajlandóság és az elemek elérhetősége miatt ez a gyakorlatban nem feltétlenül sikerül.

Reprezentatív az a minta, ahol a mintaelemeket meghatározott elvek mentén választjuk ki, s ily módon jól jellemzi az alapsokaságot, melynek köszönhetően a sokaság egészére vonatkozóan vonhatunk le következtetéseket. A sokaság nem teljes körű megfigyeléséből fakadó mintavételi hiba egyértelműen meghatározható.

A mintavétel esetén lényeges a kiválasztási arány, mely a mintába kerülő elemek és a sokaság elemszámának hányadosaként határozható meg. A mintanagyság növelésével a mintavételi hiba csökkenthető. A mintavételi technika kiválasztásánál lényeges, hogy a sokaság homogén vagy heterogén. A sokaság *homogén*, ha a vizsgálni kívánt jelenség szempontjából minden eleme egyforma, így például, ha a kávéfogyasztási szokásokról gyűjtünk információt, a válaszadó neme, életkora nem feltétlenül kiemelt szempont, a vizsgálat szempontjából tehát homogén. A sokaság *heterogén*, ha a vizsgálni kívánt jelenség szempontjából elemei különböznek, így például ha egy vállalatnál a fizetéssel való elégedettséget vizsgálnánk, akkor nem mindegy, hogy milyen beosztású munkavállalókat kérdezünk meg. Szemléletesen példázza a különbséget az az eset, mikor a főiskolai hallgatók körében vizsgálódunk. Ha a szórakozási szokásaikra vagyunk kíváncsiak, akkor a sokaság homogénnek tekinthető, hiszen nemtől, szaktól függetlenül elsősorban az a lényeg, hogy mindannyian főiskolások. Ha a vizsgálatot azonban a tanulási szokásokra vonatkoztatva végezzük el, akkor van jelentősége annak, hogy például milyen szakos a hallgató vagy éppen hányadik évfolyamra jár. A homogén és heterogén megkülönböztetés tehát a vizsgálni kívánt jelenség alapján történik.

Véletlenen alapuló mintavétel

A véletlenen alapuló kiválasztási formák közös jellemzője, hogy a mintaelemek véletlenszerűen, előre meghatározható valószínűséggel kerülnek be a mintába. Homogén sokaság esetén a *független, azonos eloszlású* (FAE) és az *egyszerű mintavétel* (EV) alkalmazható függően a sokaság elemszámától. Véges számú sokaság esetén az egyszerű véletlen mintavétel a leggyakoribb technika, mely esetben minden elem ugyanakkora valószínűséggel kerül a mintába. A FAE módszer végtelen vagy nagy elemszámú minta esetében hasonló feltételekkel alkalmazható. Heterogén sokaság esetén *rétegzett mintavétel*t célszerű alkalmazni, melynek során a mintavételt megelőzően jól elkülöníthető, homogén rétegekre bontjuk a sokaságot, majd az egyes rétegekből külön választunk elemeket. Ha nem rendelkezünk listával a sokaság elemeiről, hanem csak a sokaság elemiből alkotott csoportokat ismerjük, akkor közvetett módon, *csoportos (egylépcsős) mintavétel*t alkalmazhatunk. Ebben az esetben a sokaság számunkra fontos elemeihez indirekt módon jutunk hozzá, egyszerű mintavételi technika segítségével. Ez a közelítés történhet *többlépcsős* formában is, amikor a csoportok távolabb állnak a megfigyelni kívánt elemektől.

Az oktatási miniszter szeretné felmérni a felsőoktatás helyzetét, ezért kíváncsi a felsőoktatásban tanulók véleményére

Mintavételi mód	**Mintába kerülő elemek**
Egyszerű véletlen minta	véletlenszerűen kiválaszt főiskolai/egyetemi hallgatókat
Független, azonos eloszlású minta	véletlenszerűen kiválaszt főiskolai/egyetemi hallgatókat
Rétegzett minta	a felsőoktatásban tanulókat rétegezi pl. képzési szintenként és minden egyes rétegből választ hallgatókat
Csoportos mintavétel	a minisztériumnak könnyebb úgy informálódni, hogy a felsőoktatási intézményeken keresztül kérdezi meg a hallgatókat, hiszen intézményenként van információja a sokaságról, de külön lista, mely csak a hallgatókat tartalmazza nem áll rendelkezésére, így kiválaszt egy intézményt egyszerű mintavétellel és megkérdezi a hallgatóit
Többlépcsős mintavétel	az összes főiskolai/egyetemi hallgatót tartalmazó lista nem áll rendelkezésre, a felsőoktatási intézményekben pedig kari kimutatások vannak, ekkor több lépcsőben jut el a vizsgálni kívánt egyedekhez, a csoportokat egyszerű mintavételi technikák alkalmazásával választja, s így jut el a hallgatókhoz, melyeket teljes körűen vagy véletlenszerűen kérdez meg

Nem véletlenen alapuló mintavétel

A nem véletlenen alapuló mintavételi technikákat akkor alkalmazzuk, ha a mintába kerülő elemeknek nagy a jelentősége a későbbi elemzés szempontjából, ezért előzetes kritériumokat fogalmazunk meg a mintavétellel szemben. A *kvótás kiválasztás* esetén előre meghatározzuk a minta összetételét, vagyis azt, hogy az egyes ismérvváltozatokból hány elemnek kell szerepelnie. Például egy felmérés során hány nőt és hány férfit kell megkérdezni, vagy az egyes korosztályokból hány válaszadónak kell lennie.

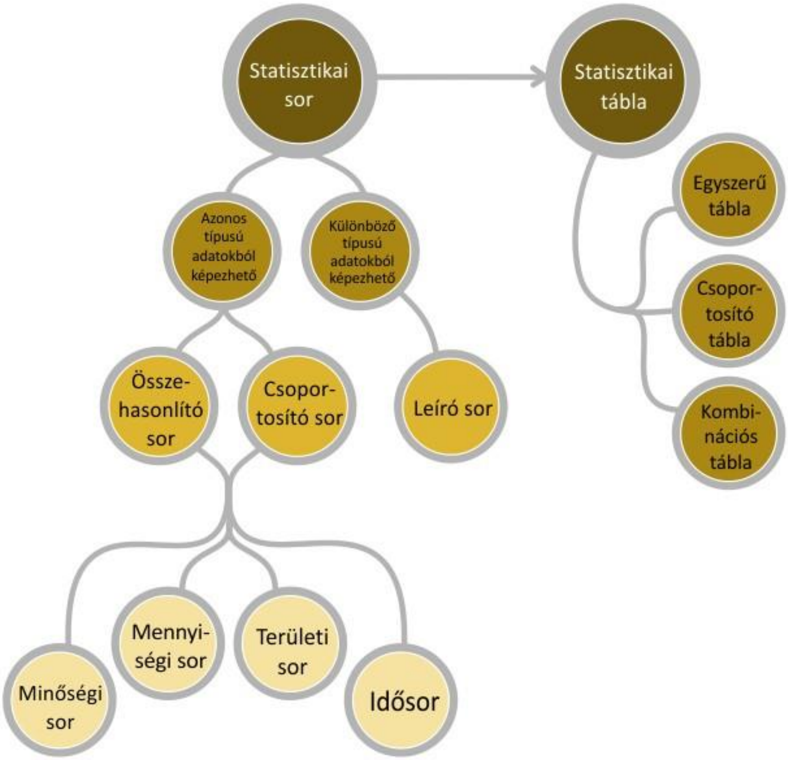
A *hólabda technika* napjainkban elterjedt mintavételi módszer, az internetes oldalakon továbbküldött kérdőívek tipikusan példázzák a módszer lényegét. A közzétett kérdőív az adott személy ismeretségi köréből kiindulva gyűjt a mintába elemeket, mely azonban figyelmen kívül hagyhat az elemzés szempontjából meghatározó egyéneket. A *koncentrált kiválasztás* lényege, hogy a domináns elemek nagyobb mértékben kerüljenek a mintába, ezáltal könnyítve az elemzést. Például a kifejezetten nőknek vagy férfiaknak vagy gyerekeknek készült termékek esetében azt az adott célcsoportot kell megkérdezni, aki érdemben véleményt tud alkotni a termékről. A *szisztematikus mintavétel* adott pontból indulva meghatározott lépésközönként gyűjt elemeket, így például abban az esetben, ha minden ötödik embert kérdezünk meg, aki szembejön velünk az utcán. Kutatásmódszertani oldalról véletlenen alapuló technikának is tekinthető, hiszen véletlen időpontban kezdődik a kutatás, s így véletlen a mintába kerülés is. Statisztikailag a tudatos lépésközök miatt nem véletlenen alapulónak tekintjük. Az *önkényes kiválasztás* a reprezentativitást nagymértékben sértheti, hiszen az adatfelvételt végző szubjektív módon válogatja a mintába az elemeket, mely ily módon a kutatás során komoly problémákhoz vezethet.

Az oktatási miniszter szeretné felmérni a felsőoktatás helyzetét, ezért kíváncsi a felsőoktatásban tanulók véleményére.

Mintavételi mód	Mintába kerülő elemek
Kvótás kiválasztás	Minden felsőoktatási intézményből meghatározott számú hallgatót kérdez meg
„Hólabda” technika	Elkészít egy kérdőívet, majd kiválaszt egy felsőoktatási intézményt, ahová elküldi, ott kitöltik a hallgatók, majd az érintett intézmény és a hallgatóik is továbbküldik más intézményekbe
Szisztematikus kiválasztás	A felsőoktatási nyilvántartó rendszerből, ahol valamennyi hallgató szerepel pl. minden 10. hallgató kerül be a vizsgálatba
Koncentrált kiválasztás	Meghatározott intézményekből választ hallgatókat
Önkényes kiválasztás	Valamely szempont pl. tanulmányi eredmény alapján választja ki a hallgatókat

2.2.2 Statisztikai adatrendszerezés

A statisztikai elemzések megkezdése előtt fontos, hogy az összegyűjtött adatok megfelelő formában álljanak rendelkezésünkre, mely megkönnyíti a módszerek kiválasztását. Az azonos és a különböző típusú adatainkat előbb statisztikai sorokba rendezhetjük, amelyekből statisztikai tábla állítható össze.



3. ábra: A statisztikai sorok és táblák típusai

A statisztikai sor

A statisztikai adatrendszerezés alapegysége a statisztikai sor, amely azonos és különböző típusú adatokból is összeállítható.

A rendelkezésre álló adatok adott ismérv szerinti rendezése a statisztikai sor.

Különböző típusú adatainkat ***leíró sor***okba rendezhetjük. Ebben az esetben az adatok ugyanarra a sokaságra vonatkoznak, de eltérő szempontból jellemzik azt. Például, ha egy adott ország turizmusát kívánjuk jellemezni, az adatokat az alábbi leíró sorba rendezhetjük.

2. *Magyarország turizmusának néhány adata 2013-ban*

Év	Külföldi látogatók száma (ezer fő)	Összesen eltöltött idő (ezer nap)	Átlagosan eltöltött idő (nap/fő)
2013	43 665	101 730	2,3

Forrás: *KSH* (2014)

Ebben az esetben a magyar turizmus helyzete képezi a megfigyelés tárgyát, a sokaság pedig a Magyarországra érkező külföldi turisták, akiknek az egyik számszerű jellemzője, hogy hányan vannak, a másik, hogy öszszesen hány napot töltöttek ott, és egy származtatott adat a kettő hányadosaként, az átlagosan eltöltött idő.

Azonos típusú statisztikai adatokat össze tudunk hasonlítani, illetve csoportosítani lehet őket. Ennek megfelelően megkülönböztetünk ***összehasonlító*** és ***csoportosító sor**ok*at. Attól függően, hogy milyen ismérv szerint történik az összehasonlítás vagy a csoportosítás lehet területi, mennyiségi, minőségi vagy idősor. Az azonos típusú adatokból összeállítható statisztikai sorok között a különbséget az adja, hogy az értékek öszszegezhetők-e. Csoportosító soroknál képezhető összesen rovat abban az esetben, ha valamennyi ismérvváltozatot feltüntetünk. Lényeges, hogy az összegnek kell, hogy legyen gyakorlati jelentősége, mert matematikailag bármely számok összegezhetők, de csoportosító sorba csak olyan adatok rendezhetők, ahol az összeadásnak gyakorlati értelme is van. Ha nincs minden ismérvváltozat szerepeltetve, akkor összehasonlítható sor képezhető. Az összehasonlító sorok legtipikusabb megjelenési formája az idősor, melynek példája lehet, ha az elmúlt években a külföldről érkező vendégek számának alakulását illusztráljuk.

3. *A Magyarországra érkező külföldiek számának alakulása 2011 és 2013 között*

Év	Külföldi látogatók száma (ezer fő)
2011	41 304
2012	43 565
2013	43 665

Forrás: KSH (2014)

Ebben az esetben az évek összegzésének nincs gyakorlati értelme, csupán azt tudjuk mondani, hogy a 3 évben együttesen mennyi vendég érkezett. Más a helyzet abban az esetben, ha egy adott éven belül vizsgálódunk, és negyedéves bontásban közöljük az adatokat, hiszen ez esetben az összeg egy teljes évet ad ki, ami gyakorlatilag is értelmezhető.

4. *A Magyarországra érkező külföldiek számának alakulása 2013 egyes negyedéveiben*

Időszak	Külföldi látogatók száma (ezer fő)
2013. I. negyedév	8 363
2013. II. negyedév	10 640
2013. III. negyedév	14 813
2013. IV. negyedév	9 849
ÖSSZESEN	***43 665***

Forrás: KSH (2014)

A csoportosító soroknál az ismérvváltozatokat osztályoknak, az egyes ismérvváltozatokhoz tartozó elemek számát pedig gyakoriságnak nevezzük.

A gyakoriságok összege megegyezik a megfigyelt elemek számával. A külföldi vendégek számának negyedéves bontású kimutatásánál összesen 43 665 elem alkotja a sokaságot, melyeket attól függően különböztetünk meg, hogy az év melyik negyedévében érkeztek. Az adott negyedévekhez tartozó elemek összessége a gyakoriság, mely megmutatja hány elemű az adott osztály, vagyis az egyes negyedévekben mennyi külföldi látogató érkezett.

A csoportosításnál kevés ismérvváltozat esetén könnyen képezhető osztály. A rangsor megkönnyíti az adatok áttekintését, melynek példája lehet az osztályzatok rendezése.

Ismerjük 20 hallgató félév végi eredményeit statisztikából:

4, 3, 2, 5, 5, 4, 1, 3, 4, 4, 5, 2, 3, 5, 5, 3, 4, 3, 4, 1

Az adatokat monoton növekvő sorrendben feltüntetve, azaz rangsorba rendezve könnyebb alkalmazni a statisztikai módszereket.

1, 1, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5

A rangsorban szereplő ismérvváltozatok alapján gyakorisági sor képezhető.

5. *A hallgatók osztályzatok szerinti megoszlása*

Osztályzat	Hallgatók száma (fő)
Jeles (5)	5
Jó (4)	6
Közepes (3)	5
Elégséges (2)	2
Elégtelen (1)	2
ÖSSZESEN	*20*

A mennyiségi jellemzők csoportosítása sok ismérvváltozat esetén nem egyszerű feladat.

Osztályközös gyakorisági sornak azt a mennyiségi ismérv szerinti csoportosító sort nevezzük, melynél az osztályokat intervallumokkal, más néven osztályközökkel adjuk meg.

Az osztályközös csoportosításnál az osztályközöket úgy kell meghatározni, hogy az ismérvértékek egyértelműen besorolhatók legyenek valamely, de csak egy osztályközbe[2]. Az osztályközök hosszát és számát úgy kell meghatározni, hogy megfelelően tükrözze a sokaság mennyiségi ismérv szerinti összetételét. Az osztályközök optimális száma (k) az alábbi képlettel becsülhető:

$$\mathbf{2^k > N}$$

A szabály alapján a ddig szükséges növelni az osztályközök számát (k), amíg a kettő k-adik hatványa meg nem haladja a sokaság elemszámát (N). Egyenlő osztályközök esetén az osztályköz hossza (h) az alábbi képlettel határozható meg:

$$\text{osztályköz hossza (h)} = \frac{\text{az adatsor legnagyobb értéke} - \text{az adatsor legkisebb értéke}}{\text{osztályközök száma (k)}}$$

Egy statisztikai zárthelyi dolgozatban 100 hallgató pontszámait ismerjük. *Hány darab és milyen hosszúságú osztályköz képezhető, ha tudjuk, hogy a leggyengébb hallgató 16 pontot, míg a legjobb 98 pontot ért el?*

100 elemű sokaság esetén: $2^7 = 128 > 100 \rightarrow k=7$ vagyis az osztályközök optimális száma 7 darab.

A hallgatók elért pontszámai közül a legkisebb érték 16 pont, a legnagyobb pedig 98, így ha tudjuk, hogy 7 osztályköz képezhető optimálisan, akkor egyenlő osztályközös csoportosítás esetén

$$h = \frac{x_{max} - x_{min}}{k} \rightarrow \frac{98 - 16}{7} = 11{,}7 \sim 12$$

Optimálisan tehát 7 db, egyenként 12 egység hosszúságú osztályköz képezhető.

[2] Ezen szabály be nem tartására jó példát szolgáltatnak azok a kérdőívek, melyek az életkorra rákérdezve az intervallumokat úgy adják meg például, hogy 18-25 és 25-35. Ez esetben a 25 éves kitöltőre rá van bízva, hogy melyikbe sorolja magát, a kiértékelésnél azonban ez komoly pontatlanságot eredményez.

A gyakorlatban azonban nem feltétlenül szükséges egyenlő hosszúságú intervallumokat megadni, a nem egyenlő osztályközös felosztás ugyanis kiemelheti a sokaság aránytalanságait.

A statisztikai tábla

A statisztikai adatok elemzésénél leggyakrabban nem csupán egy ismérvre van szükség, hanem igyekszünk a rendelkezésre álló adatainkat minél több ismérv mentén jellemezni.

A statisztikai táblák egymással összefüggő statisztikai sorokat tartalmaznak.

A statisztikai tábla funkciója szerint lehet munkatábla, melyet a számításokhoz használunk és az eredmények közlésére alkalmas közlési tábla. A statisztikai táblának mindig van *címe*, mely utal arra a sokaságra, melyre a táblában lévő adatok vonatkoznak, valamint térben és időben is lehatárolja a vizsgálat tárgyát. Az adatok *forrás*ának feltüntetése eredményközlő táblák esetében kötelező, munkatáblák esetén nem feltétlenül szükséges, mert ott a matematikai számításokon van a hangsúly. A statisztikai tábla szerkesztésénél mindig azt kell szem előtt tartani, hogy *egyszerű és átlátható formában* tüntessük fel az adatokat, melyek mértékegységét a tábla címében vagy a fej, illetve oszloprovatban szerepeltetni kell. Statisztikai táblában üresen hagyott rovat nem lehet, így hiányzó adat esetén a rovatot kihúzzuk, nem ismert adat esetén (…) teszünk, ha kicsi a mértékegység a táblában szereplő többi adathoz képest, akkor 0,0 értéket tüntetünk fel. Csoportosító sort tartalmazó táblánál összesen rovatot kell képezni.

A statisztikai táblákat a benne szereplő statisztikai sorok alapján az alábbiak szerint különböztetjük meg:

– a tábla *egyszerű*, ha nincs benne csoportosító sor, azaz nem képezhető összesen rovat, tehát leíró és/vagy összehasonlító sorok alkotják

6. *A Magyarországra érkező francia és német turisták száma és az általuk átlagosan eltöltött idő 2013-ban*

Származási ország	Látogatók száma (ezer fő)	Átlagosan eltöltött idő (nap/fő)
Németország	3 059	5,1
Franciaország	341	5,4

Forrás: KSH (2014)

– *csoportosító* tábla, melyben van egy csoportosító sor, valamint egy leíró és/vagy összehasonlító sor

7. *A Magyarországra érkező külföldiek számának alakulása az utazás turisztikai célja szerint 2012-ben és 2013-ban*

Utazás célja	Látogatók száma	
	2012	2013
Szabadidős turizmus	13 489	13 546
Üzleti turizmus	1390	1341
ÖSSZESEN	*14 879*	*14 887*

Forrás: KSH (2014)

– *kombinációs* tábla, melyet csak csoportosító sorok alkotnak, így az adatok vízszintesen és függőlegesen is összegezhetők

8. *Magyarországra érkező turisták száma nem és származási helye szerint*

Lakóhely	Balaton	Budapest Közép-Duna-vidék	Alföld	Észak-Magyarország	Dunántúl	Tisza-tó	*Összesen*
Közép-Magyarország	6 445	4 004	3 903	2 363	4 203	270	*21 188*
Dunántúl	5 388	2 547	1 838	749	8 602	194	*19 318*
Észak-Magyarország	1 159	1 354	1 328	3 203	528	280	*7 852*
Alföld	1 989	2 070	4 829	1 335	1 004	315	*11 539*
Összesen	*14 981*	*9 975*	*11898*	*7 650*	*14 337*	*1 059*	*59 897*

Forrás: KSH (2014)

A statisztikai tábla dimenziószáma azt mutatja meg, hogy az adott érték hány ismérvhez tartozik.

A statisztikai adatok rendszerezésénél az egyértelműségre és áttekinthetőségre kell törekedni, a megfelelően strukturált adatokból az elemzéshez szükséges módszerek könnyen kiválaszthatók.

2.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

2.3.1 Összefoglalás

A statisztikai munka első lépése a vizsgálni kívánt jelenség elemzéséhez szükséges adatok összegyűjtése és rendszerezése. Primer és szekunder adatforrásból származhatnak információk, melyek eltérő elemzési módszerek alkalmazását teszik lehetővé. A primer eljárások során lényeges a megfelelő mintavételi technika megválasztása. A mintába az elemek véletlenen és nem véletlenen alapuló kiválasztás útján kerülhetnek. A mintaelemek megválasztásánál a reprezentativitás kritériumait szem előtt kell tartani, ha olyan elemzést kívánunk végezni, mellyel általános érvényű következtetések levonására törekszünk.

A statisztikai adatrendszerezés az adatelemzés közvetlenül megelőző lépése. A statisztikai sor és a statisztikai tábla alapvető jellemzője, hogy a vizsgálni kívánt jelenséget ismérvek segítségével ragadja meg, s rendszerezi. A sorokat képezhetjük azonos és különböző típusú adatokból, ez alapján megkülönböztetünk leíró, összehasonlító és csoportosító sorokat. A statisztikai tábla lényeges jellemzője, hogy címmel és eredményközlés esetén forrással van ellátva, valamint üresen nem hagyhatunk egyetlen rovatot sem. Lényeges továbbá, hogy a különböző típusú sorok és táblák eltérő elemzésekre adnak lehetőséget.

2.3.2 Önellenőrző kérdések

1. Mi a különbség a sokaság és a minta között?
2. Melyek a primer és a szekunder adatforrások jellemzői?
3. Mit tekinthetünk statisztikailag reprezentatív mintának?
4. Milyen véletlenen alapuló mintavételi technikákat ismer?
5. Milyen nem véletlenen alapuló mintavételi technikákat ismer?
6. Mi a statisztikai sor és a statisztikai tábla?
7. Mi jellemzi a leíró, összehasonlító és csoportosító statisztikai sorokat?
8. Mi jellemzi az egyszerű statisztikai táblát?
9. Mi jellemzi a csoportosító statisztikai táblát?
10. Mi jellemzi a kombinációs statisztikai táblát?

2.3.3 Gyakorló tesztek

☸ Állapítsa meg, hogy a vizsgálni kívánt jelenség szempontjából melyik mintavételi technika alkalmazása ajánlott!
Több válasz is megjelölhető!

	Véletlenen alapuló					Nem véletlenen alapuló				
	Független, azonos eloszlású	Egyszerű véletlen	Rétegzett	Csoportos	Többlépcsős	Kvótás	Hólabda	Koncentrált	Szisztematikus	Önkényes
Kávéfogyasztási szokások	X	X					X			
Tévénézési szokások		X					X		X	
Egy vállalatnál dolgozók fizetéssel való elégedettsége			X			X				
Autóvásárlással kapcsolatos tapasztalatok								X		
Középiskolások továbbtanulási tervei				X	X					

☸ Döntse el az alábbi állításokról, hogy igazak (I) vagy hamisak (H)!

A minta reprezentativitása az általános érvényű következtetések levonása miatt lényeges.	I
A csoportosító táblában csak csoportosító sorok vannak.	H
A statisztikai tábla dimenziószáma megmutatja, hogy a táblában egy adott érték hány ismérvhez tartozik.	I
Különböző típusú adatokból összehasonlító és csoportosító sorok képezhetők.	H
A kombinációs táblában különböző típusú statisztikai sorok vannak.	H

3. VISZONYSZÁMOK

3.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A statisztikai elemzések egyik legegyszerűbb és leggyakrabban használatos eszközei a viszonyszámok. A viszonyítás során két összefüggő adat hányadosát képezzük azonos vagy különböző típusú adatokból. A változatos eszköztár alkalmazási lehetőségeinek, módszereinek ismerete hozzásegíti a hallgatót egyszerűbb statisztikai eszközökkel megoldható problémák felismeréséhez és megoldásához, melyekkel a mindennapokban is találkozhat.

A lecke célja, hogy megismertesse a hallgatókat a viszonyszámok rendszerével és az egyes típusok kiszámítási formáival. A hallgató ne csupán felületesen tudja értelmezni, hanem értse meg a viszonyszámok logikáját, melyhez elgondolkodtató és a túlzottan matematikai közelítést nélkülöző feladatokkal kívánunk hozzájárulni. Képes legyen felismerni a hallgató olyan problémákat, melyeket a viszonyszámok különböző típusainak alkalmazásával meg is tud oldani.

3.2 TANANYAG

A statisztikai elemzések során a rendelkezésre álló adatokat gyakran nem abszolút formában, hanem viszonyszámként használjuk, mely lehetővé teszi az adatok összehasonlítását és nagyságrendbeli különbségek érzékeltetését is.

Két egymással összefüggő statisztikai adat hányadosát viszonyszámnak nevezzük.

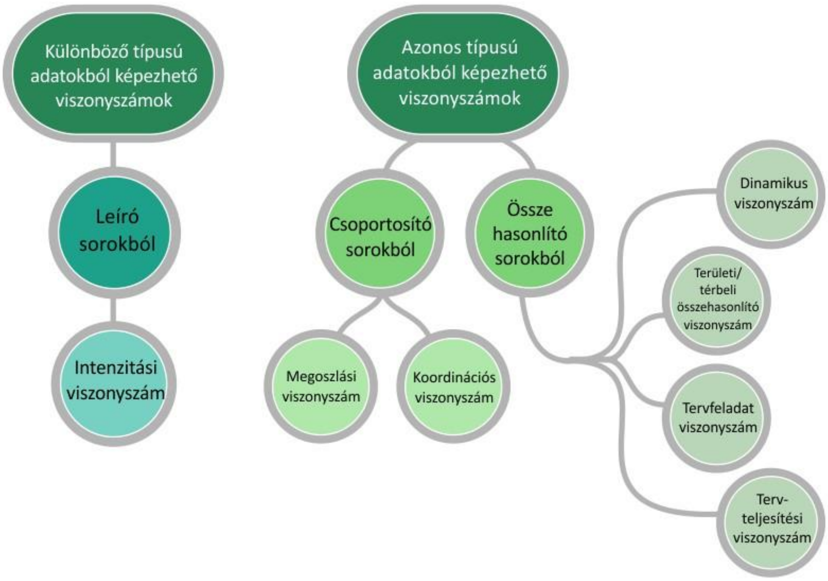
$$\text{Viszonyszám (V)} = \frac{\text{viszonyítandó adat (A)}}{\text{viszonyítási alap (B)}}$$

Az összefüggés alapján bármely két rendelkezésre álló tényező alapján meghatározható a harmadik, ezekben az esetekben a viszonyszámokkal együtthatós formában számolunk:

$$\text{Viszonyítandó adat (A)} = \text{viszonyítási alap (B)} \cdot \text{viszonyszám (V)}$$

$$\text{Viszonyítási alap (B)} = \frac{\text{viszonyítandó adat (A)}}{\text{Viszonyszám (V)}}$$

A viszonyszámoknak különböző típusait különböztetjük meg aszerint, hogy milyen jellegű statisztikai adatokat viszonyítunk egymáshoz. Azonos típusú statisztikai adatok egymáshoz való viszonyát megoszlási, koordinációs, dinamikus, területi/térbeli összehasonlító, tervfeladat vagy tervteljesítési viszonyszámokkal, míg különböző típusú adatok összefüggését intenzitási viszonyszámokkal tudjuk számszerűsíteni. Az azonos típusú adatokból képzett viszonyszámokat tovább csoportosíthatjuk attól függően, hogy csoportosító vagy összehasonlító sorokból képezzük őket.



4. ábra: A viszonyszámok rendszere

A viszonyszámok használatát gyakran az indokolja, hogy eltérő mértékegységű vagy nagyságrendű adatokat nem tudunk összehasonlítani. A rendelkezésre álló statisztikai adatok viszonyszámként gyakran több információt szolgáltathatnak, mint eredeti formájukban. A viszonyszámok így a statisztikai elemzések egyik legfontosabb eszközei.

3.2.1 Megoszlási és koordinációs viszonyszámok

A megoszlási és koordinációs viszonyszámok alkalmazásának elsődleges célja a sokaságon belüli arányok szemléletes érzékeltetése.

A csoportosított sokaság egyes részeinek a sokaság egészéhez viszonyított arányát megoszlási viszonyszámnak nevezzük.

$$\text{Megoszlási viszonyszám } (V_m) = \frac{\text{rész}}{\text{egész}} \rightarrow V_m = \frac{f_i}{\sum f_i}$$

A számítás során a részsokaság adatát osztjuk az összesen adattal, a kapott viszonyszámot pedig 100-zal való szorzás után %-os formában értelmezzük. A megoszlási viszonyszámok összege együtthatós formában

1, százalékos formában pedig 100%. A megoszlási viszonyszámok a nagyságrendek érzékeltetésének hatékony eszközei, melyek segítségével jól jellemezhető a sokaság vagy minta szerkezete.

A koordinációs viszonyszám a csoportosított sokaság egyes részeit viszonyítja egymáshoz.

$$\text{Koordinációs viszonyszám } (V_k) = \frac{\text{rész}}{\text{rész}} \rightarrow V_k = \frac{f_i}{f_j}$$

A koordinációs viszonyszámot ritkábban használjuk, a sokaság részeinek egymáshoz való viszonyát fejezhetjük ki vele, az egyik részsokaság egy egységére mennyi jut a másik részsokaságból. A viszonyítási alapot a könnyebb értelmezhetőség érdekében megnövelhetjük úgy, hogy a két adat hányadosaként kapott értéket megszorozzuk egy tetszőleges értékkel. Ekkor azt fejezi ki, hogy a viszonyítási alap 10 vagy 100 vagy 1000 egységére mennyi jut a másik részsokaságból. A koordinációs viszonyszámokat együtthatós formában értelmezzük.

Összefüggés a megoszlási és koordinációs viszonyszámok között

A megoszlási és a koordinációs viszonyszámok csak csoportosított sokaságból számíthatók. A kiinduló abszolút adatok nélkül kölcsönösen számítható koordinációs viszonyszámból megoszlási, illetve megoszlási viszonyszámból koordinációs.

Egy 80 fős statisztika kurzuson 30 fiú és 50 lány vett részt. *Elemezze viszonyszámok segítségével a kurzus nemek szerinti megoszlását, valamint a fiúk és lányok egymáshoz való viszonyát!*

MEGOSZLÁSI viszonyszámok alkalmazásával a sokaság egyes részeit viszonyíthatjuk a sokaság egészéhez, ebben az esetben a fiúk és lányok számát a kurzus összlétszámához.

9. *A statisztika csoport megoszlásának kiszámítása*

Statisztika csoport	Létszám	
Fiú	30	Fiúk aránya: $\frac{30}{80} = 0{,}375 \xrightarrow{\cdot 100} \mathbf{37{,}5}\ \%$
Lány	50	Lányok aránya: $\frac{50}{80} = 0{,}625 \xrightarrow{\cdot 100} \mathbf{62{,}5}\ \%$
Összesen	*80*	Együtt: $0{,}375 + 0{,}625 = \mathbf{1} \xrightarrow{\cdot 100} \mathbf{100}\ \%$

A megoszlási viszonyszámok alapján a kurzus 37,5 %-a fiú és 62,5 %-a lány.

KOORDINÁCIÓS viszonyszám segítségével a fiúk és lányok egymáshoz való viszonyát mutathatjuk ki.

$$1\ fiúra\ jutó\ lányok\ száma\!: \frac{50}{30} = \mathbf{1,67}\ \ és\ 1\ lányra\ jutó\ fiúk\ száma\!: \frac{30}{50} = \mathbf{0,6}$$

A csoportban egy fiúra 1,67 lány jut, míg egy lányra 0,6 fiú. A könnyebb értelmezhetőség miatt a viszonyítási alapot 10-re emelve azt mondhatjuk, hogy 10 fiúra ~17 lány jut, s 10 lányra 6 fiú, ha 100-ra emeljük, akkor 100 fiúra 167 lány jut, s 100 lányra 60 fiú.

A megoszlási és koordinációs viszonyszámok összefüggései a példán keresztül

Az eredeti adatok hiányában, csupán a koordinációs viszonyszámok ismeretében tudjuk, hogy 1 fiúra 1,67 lány jut. Ebből következtethetünk arra, hogy a teljes sokaság 1 fiúból és 1,67 lányból áll, vagyis 2,67 fő. A rész és teljes sokaság egyszerűsített elemszámát ismerve kiszámítható, hogy

$$a\ fiúk\ aránya\!: \frac{1}{1+1,67} = \frac{1}{2,67} = 0,375 \xrightarrow{\cdot 100} 37,5\ \%$$

$$a\ lányok\ aránya\ \frac{1,67}{2,67} = 0,625 \xrightarrow{\cdot 100} 62,5\ \%$$

Az eredeti adatok hiányában, csupán megoszlási viszonyszámok ismeretében tudjuk, hogy a kurzus 37,5 %-a fiú, 62,5 %-a lány. A két értéket egymáshoz viszonyítva tizedestört formában megkapjuk,

$$1\ fiúra\ jutó\ lányok\ száma\!: \frac{0,625}{0,375} = 1,67$$

$$1\ lányra\ jutó\ fiúk\ száma\!: \frac{0,375}{0,625} = 0,6$$

3.2.2 Dinamikus viszonyszámok

Időbeli összehasonlításra a leggyakrabban használatos elemzési eszköz a dinamikus viszonyszám, melynek segítségével százalékos formában tudjuk kifejezni két időszak között a változás mértékét, illetve megállapítható az iránya is. A viszonyítási alap lehet állandó és változó, melynek megválasztása alapján kétféle viszonyszámot különíthetünk el.

Időbeli összehasonlító soroknál, ha mindig ugyanannak az időszaknak az adatához viszonyítunk, akkor bázisviszonyszámról beszélünk.

$$\text{Bázisviszonyszám} = \frac{\textbf{Tárgy}\text{ időszak adata}}{\textbf{Bázis}\text{ időszak adata}}$$

A gyakorlatban legtöbbször a megfigyelt első értéket tekintjük bázisidőszaknak, ez viszont nem kötelezően van így. A bázisidőszak megválasztásánál azt kell szem előtt tartani, hogy mely év adatához érdemes viszonyítani a megfigyelési értékeket. Fontos megjegyezni, hogy az elsőtől eltérő bázisidőszak esetén a viszonyszámok értelmezésére figyelni kell.

Időbeli összehasonlító soroknál, ha az időszakról időszakra történő változás lényeges, akkor láncviszonyszámokat alkalmazunk, mely mindig az adott időszak adatát a közvetlenül megelőző időszakhoz viszonyítja.

$$\text{Láncviszonyszám} = \frac{\textbf{Tárgy}\text{ időszak adata}}{\text{Tárgyidőszakot }\textbf{megelőző}\text{ időszak adata}}$$

Egy kisboltban vásárlók számáról az elmúlt években ismert, hogy 2010-ben 2000 fő, 2011-ben 2450 fő, 2012-ben 2260 fő, míg 2013-ban 2620 fő fordult meg az üzletben. *Dinamikus viszonyszámok segítségével jellemezze a kisboltban vásárlók számának alakulását az elmúlt években!*

A viszonyszámok kiszámításához célszerű a rendelkezésre álló adatokat statisztikai táblába rendezni, így a kapott eredmények is könnyebben értelmezhetők. A számítás során a hányadosok értékét ajánlott legalább 4 tizedes jegy pontossággal megadni, hogy az eredmények egyértelműen értelmezhetők legyenek.

BÁZISVISZONYSZÁMOK segítségével a kisbolt vásárlóinak számát egy kiválasztott év adatához tudjuk viszonyítani. A legtöbbször az első évet tekintjük bázisidőszaknak, de más is választható.

10. *Munkatábla bázisviszonyszám számításához, ha a 2010-es évet tekintjük bázisidőszaknak*

Év	Vásárlók száma (fő)	A vásárlók számának alakulása 2010-hez viszonyítva (2010 = 100%)	A változás mértéke (%-ban kifejezve) 2010-hez képest
2010	2 000	100	–
2011	2 450	$\frac{\text{tárgy}}{\textbf{bázis}} = \frac{2011}{\textbf{2010}} \rightarrow \frac{2450}{2000} = 1{,}225 \xrightarrow{\cdot 100} \mathbf{122{,}5}$	+ 22,5 %
2012	2 260	$\frac{\text{tárgy}}{\textbf{bázis}} = \frac{2012}{\textbf{2010}} \rightarrow \frac{2260}{2000} = 1{,}13 \xrightarrow{\cdot 100} \mathbf{113{,}0}$	+ 13 %
2013	2 620	$\frac{\text{tárgy}}{\textbf{bázis}} = \frac{2013}{\textbf{2010}} \rightarrow \frac{2620}{2000} = 1{,}31 \xrightarrow{\cdot 100} \mathbf{131{,}0}$	+ 31 %

Útmutató a tábla kitöltéséhez: a tábla fejrovatában jelezni kell a viszonyítási alapot, mely jelen esetben a 2010-ben vásárlók száma, így ez az időszak tekinthető 100 %-nak. A bázisként választott időszakban a bázisviszonyszám értéke 100 %, változás pedig értelemszerűen nincs. A későbbi időszakok adatát mindig a bázisidőszak adatával osztjuk, a hányados értékét 100-zal szorozva kapjuk meg a viszonyszám formát. A változás mértékét a bázisként választott időszakhoz képest úgy határozzuk meg, hogy a bázisviszonyszám értékét 100-hoz viszonyítjuk.

Az eredmények értelmezése: 2010-hez képest a kisboltban vásárlók száma 2011-ben 22,5%-kal, 2012-ben 13%-kal, míg 2013-ban 31%-kal volt magasabb.

Hogyan alakul a kisboltban vásárlók száma akkor, ha a 2012-es év adatát tekintjük bázisidőszaknak?

11. *Munkatábla bázisviszonyszám számításához, ha a 2012-es évet tekintjük bázisidőszaknak*

Év	Vásárlók száma (fő)	A vásárlók számának alakulása a 2012-es évhez viszonyítva (2012 = 100%)	A változás mértéke (%-ban kifejezve) 2012-höz képest
2010	2 000	$\frac{\text{tárgy}}{\textbf{bázis}} = \frac{2010}{\textbf{2012}} \rightarrow \frac{2000}{2260} = 0{,}885 \xrightarrow{\cdot 100} \mathbf{88{,}5}$	– 11,50 %
2011	2 450	$\frac{\text{tárgy}}{\textbf{bázis}} = \frac{2011}{\textbf{2012}} \rightarrow \frac{2450}{2260} = 1{,}0841 \xrightarrow{\cdot 100} \mathbf{108{,}41}$	+ 8,41 %
2012	2 260	100	–

2013	2 620	$\frac{\text{tárgy}}{\textbf{bázis}} = \frac{2013}{\textbf{2012}} \rightarrow \frac{2620}{2260} = 1{,}1593 \xrightarrow{\cdot 100} \mathbf{115{,}93}$	+ 15,93 %

A kisboltban 2010-ben 11,5%-kal kevesebb, míg 2011-ben 8,5%-kal több vásárló fordult meg, mint 2012-ben, 2013-ban a 2012-es évhez képest 15,93%-os emelkedés figyelhető meg.

LÁNCVISZONYSZÁMOK segítségével a kisboltban vásárlók számának alakulását mindig a megelőző időszak adatához viszonyítva tudjuk megadni.

12. *Munkatábla láncviszonyszám számításához*

Év	**Vásárlók száma (fő)**	**A vásárlók számának alakulása az előző évhez viszonyítva (Előző év = 100%)**	**A változás mértéke (%-ban kifejezve) az előző évhez képest**
2010	2 000	—	—
2011	2 450	$\frac{\text{tárgy}}{\textbf{előző}} = \frac{2011}{\textbf{2010}} \rightarrow \frac{2450}{2000} = 1{,}225 \xrightarrow{\cdot 100} \mathbf{122{,}5}$	+ 22,5 %
2012	2 260	$\frac{\text{tárgy}}{\textbf{előző}} = \frac{2012}{\textbf{2011}} \rightarrow \frac{2260}{2450} = 0{,}9224 \xrightarrow{\cdot 100} \mathbf{92{,}24}$	– 7,76 %
2013	2 620	$\frac{\text{tárgy}}{\textbf{előző}} = \frac{2013}{\textbf{2012}} \rightarrow \frac{2620}{2260} = 1{,}1593 \xrightarrow{\cdot 100} \mathbf{115{,}93}$	+ 15,93 %

Útmutató a tábla kitöltéséhez: a tábla fejrovatában jelezni kell a viszonyítási alapot, mely mindig az előző időszak adata, így ez tekinthető 100%-nak. Az első időszakban sem viszonyszám, sem változás nincs, hiszen az előző időszakról nem áll rendelkezésre adat. A későbbi időszakok adatát mindig a megelőző időszak adatával osztjuk, a hányados értékét 100-zal szorozva kapjuk meg a viszonyszám formát. A változás mértékét az előző időszakhoz képest úgy határozzuk meg, hogy a láncviszonyszám értékét 100-hoz viszonyítjuk.

Az eredmények értelmezése: 2010-hez képest 2011-ben 22,5%-kal nőtt a vásárlók száma. 2012-ben 2011-hez képest 7,76%-kal kevesebb vásárló fordult meg az üzletben. 2013-ban 2012-höz képest ismét nőtt a vásárlók száma, 15,93%-kal.

3.2.3 Területi/térbeli összehasonlító viszonyszámok

Területi összehasonlító soroknál használható viszonyszámok, melyek esetében lényeges, hogy földrajzi területhez kötődő statisztikai adatokat hasonlítsunk össze. A viszonyítási alap és a viszonyítandó adat tetszőlegesen felcserélhető attól függően, hogy a viszonyítás célja milyen irányú.

A területi/térbeli viszonyszám két földrajzi egység valamely statisztikai adatát viszonyítja egymáshoz.

$$\text{Területi összehasonlító viszonyszám} = \frac{\text{viszonyítandó terület adata (A)}}{\text{viszonyítási alapként választott terület adata (B)}}$$

A területi összehasonlító viszonyszám együtthatós és százalékos formában is értelmezhető.

Egy vállalat két különböző városban működő telephelyének termeléséről ismert, hogy A településen 10 250 db terméket állítanak elő, míg B településen 8 200 db-ot. *A megfelelő viszonyszám kiszámításával hasonlítsa össze a két telephely termelését!*

$$\frac{A\ városban\ működő\ telephely\ termelése}{B\ városban\ működő\ telephely\ termelése} = \frac{10\ 250}{8\ 200} = \mathbf{1,25}$$

Az A városban működő telephelyen 1,25-ször, vagyis 25%-kal többet termelnek, mint a B városban működő telephelyen.

A viszonyítási alap és a viszonyítandó adat felcserélhető az elemzés céljának megfelelően, attól függően, hogy azt szeretnénk-e hangsúlyozni, hogy az egyik telephely mennyivel többet termel, vagy éppen fordítva, hogy a másik mennyivel kevesebbet.

$$\frac{B\ városban\ működő\ telephely\ termelése}{A\ városban\ működő\ telephely\ termelése} = \frac{8\ 200}{10\ 250} = \mathbf{0,8}$$

A B városban működő telephely termelése 80%-a (4/5-e) az A városban működő telephely termelésének, vagyis 20%-kal marad el attól.

3.2.4 Tervfeladat és tervteljesítési viszonyszámok

A terv és tény adatokat összehasonlító sorokból képezhető viszonyszámokat előirányzott feladat tervezésére és a feladat végrehajtásának számszerűsítésére is használhatjuk.

A tervfeladat viszonyszám a tervezett adatot egy korábbi tényadathoz viszonyítja.

$$\text{Tervfeladat viszonyszám} = \frac{\text{terv}}{\text{tény}}$$

A tervteljesítési viszonyszám a tényleges adatot a tervezett adathoz viszonyítja.

$$\text{Tervteljesítési viszonyszám} = \frac{\text{tény}}{\text{terv}}$$

Egy vállalat a tavalyi évben 18 250 db terméket értékesített. Az idei évre 19 000 db termék értékesítését tűzte ki célul, s végül 18 880 db-ot sikerült eladnia. *A megfelelő viszonyszámok kiszámításával értékelje a vállalat teljesítményét!*

TERVFELADAT VISZONYSZÁM segítségével számszerűsíteni tudjuk, hogy a terv miképpen viszonyul a korábbi időszak tényleges értékesítéséhez.

$$\frac{\text{Idei }\textbf{terv}\text{ezett értékesítés}}{\text{Tavalyi }\textbf{tény}\text{leges értékesítés}} = \frac{19\,000}{18\,250} = \mathbf{1{,}0411} \rightarrow\ +4{,}11\%$$

A vállalat az idei évben a tavalyihoz képest 4,11%-kal kívánta növelni értékesítését.

TERVTELJESÍTÉSI VISZONYSZÁM segítségével számszerűsíteni tudjuk, hogy milyen mértékben sikerült a tervet megvalósítani.

$$\frac{\text{Idei }\textbf{tény}\text{leges értékesítés}}{\text{Idei }\textbf{terv}\text{ezett értékesítés}} = \frac{18880}{19000} = \mathbf{0{,}9937}\ \rightarrow\ -0{,}63\%$$

Az idei tényleges értékesítés 0,63%-kal elmaradt a tervezettől.

Ha ismerjük a korábbi és a tárgyidőszak tényadatát, azokat egymáshoz viszonyítva dinamikus viszonyszámot számolhatunk.

$$\frac{\text{Idei }\textbf{tény}\text{leges értékesítés}}{\text{Tavalyi }\textbf{tény}\text{leges értékesítés}} = \frac{18880}{18250} = \mathbf{1{,}0345} \rightarrow\ +3{,}45\ \%$$

A vállalat tavalyhoz képest 3,45%-kal növelni tudta az értékesítését.

3.2.5 Intenzitási viszonyszámok

A gyakorlatban gyakran viszonyítunk egymáshoz különböző típusú adatokat. Intenzitási viszonyszámokat képezhetünk különböző fajtájú, de azonos mértékegységű adatokból, de gyakoribb és könnyebben felismerhető, ha különböző mértékegységű adatokból származnak.

Az intenzitási viszonyszámokat különböző típusú adatokból képezzük, azt mutatják meg, hogy az egyik sokaság egy egységére mennyi jut a másik, eltérő sokaságból.

A könnyebb értelmezhetőség érdekében sokszor célszerű a viszonyítási alapot megnövelni, s így az adott sokaság 10 vagy 100 vagy 1000 egységére vetített értéket meghatározni. Ebben az esetben a két adat hányadosaként kapott értéket megszorozzuk ezzel az értékkel.

A viszonyítás iránya alapján megkülönböztetünk *egyenes és fordított intenzitási viszonyszámok*at. A fordított intenzitási viszonyszám matematikailag az egyenes intenzitási viszonyszám reciproka, melyet a viszonyítandó adat és a viszonyítási alap felcserélésével határozhatunk meg.

☸ Egy felsőoktatási intézményben összesen 1000 oktató dolgozik és 4200 hallgató tanul. *A megfelelő viszonyszámok segítségével jellemezze az oktatók és hallgatók viszonyát az intézményben!*

INTENZITÁSI viszonyszámok segítségével számszerűsíthető, hogy az intézményben egy oktatóra hány hallgató jut, illetve fordítva, egy hallgatóra hány oktató jut.

$$Egy\ oktatóra\ jutó\ hallgató\text{:}\ \frac{összes\ hallgató}{összes\ oktató} = \frac{4200}{1000} = \mathbf{4{,}2}$$

$$Egy\ hallgatóra\ jutó\ oktató\text{:}\ \frac{összes\ oktató}{összes\ hallgató} = \frac{1000}{4200} = \mathbf{0{,}2381}$$

A felsőoktatási intézményben egy oktatóra 4,2 hallgató jut, míg egy hallgatóra 0,2381 oktató. A viszonyítási alapot megnövelve azt mondhatjuk, hogy 100 oktatóra 420 hallgató jut, míg 100 hallgatóra ~ 24 oktató. Ha csak az intenzitási viszonyszámokat ismernénk, akkor az abszolút adatok ismerete nélkül megállapíthatnánk, hogy az intézményben több, mint négyszer annyi hallgató tanul, mint amennyi oktató dolgozik.

A viszonyítási alaphoz való kötődés alapján *nyers és tisztított intenzitási viszonyszámok*at különböztetünk meg. A tisztított intenzitási viszonyszám a viszonyítási alapnak egy szűkebb részéhez kapcsolódik, mely közelebb áll a viszonyítandó adathoz. A tisztított intenzitási viszonyszám esetén tehát ugyanazt a viszonyítandó adatot a viszonyítási alapnak egy részéhez viszonyítjuk, így értéke biztosan magasabb lesz, mint a nyers intenzitási viszonyszámé. Matematikailag belátható, hogy ugyanazt a számot egy kisebb számmal osztva a hányados értéke biztosan magasabb lesz.

$$\text{Nyers intenzitási viszonyszám} = \frac{\mathbf{A}}{\mathbf{B}} \qquad \text{Tisztított intenzitási viszonyszám} = \frac{\mathbf{A}}{\mathbf{b}}$$

$$\text{ahol } \frac{\mathbf{b}}{\mathbf{B}} = \text{a viszonyítási alapra számítható megoszlási viszonyszám}$$

Ebből következően az alábbi összefüggés írható fel a nyers és tisztított, valamint a viszonyítási alapra számítható megoszlási viszonyszám között:

$$\text{Nyers intenzitási viszonyszám} = \text{Tisztított intenzitási viszonyszám} \cdot \text{Megoszlási viszonyszám}$$

Matematikailag is könnyen belátható az összefüggés: $\frac{\mathrm{A}}{\mathrm{B}} = \frac{\mathrm{A}}{\mathrm{b}} \cdot \frac{\mathrm{b}}{\mathrm{B}}$

A szorzat bármely két tagját ismerve kiszámítható a harmadik viszonyszám.

☸ Egy felsőoktatási intézményben összesen 4000 hallgató tanul, és 1000 oktató dolgozik, akik közül 400 oktat statisztika tárgyat. *A megfelelő viszonyszámok segítségével jellemezze az összes oktató, valamint a statisztikát oktatók és hallgatók viszonyát az intézményben!*

$Egy\ oktatóra\ jutó\ hallgató$: $\frac{összes\ hallgató}{összes\ oktató} = \frac{4000}{1000} = \mathbf{4}$

$Egy\ statisztika\ tárgyat\ oktatóra\ jutó\ hallgató$:

$$\frac{összes\ hallgató}{statisztika\ tárgyat\ oktatók\ összesen} = \frac{4000}{400} = \mathbf{10}$$

A felsőoktatási intézményben egy oktatóra 4 hallgató jut, ha viszont a viszonyítási alapot szűkítjük és csak a statisztika tárgyat oktató kollégákra vetítjük az összes hallgatót, akkor eredményként azt kapjuk, hogy egy statisztikát oktató kollégára 10 hallgató jut.

A viszonyszámok közötti összefüggés e példán keresztül jól illusztrálható.

$Statisztikát\ oktatók\ aránya\ az\ oktatókon\ belül\ (megoszlási\ viszonyszám)$

$$\frac{statisztikát\ oktatók}{összes\ oktató} = \frac{400}{1000} = 0{,}4$$

$$V_{i\,NYERS} = V_{i\,TISZTÍTOTT} \cdot V_M \;\rightarrow\; \mathbf{4 = 10 \cdot 0{,}4}$$

3.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

3.3.1 Összefoglalás

A statisztikai elemzés egyszerű eszközei a viszonyszámok. Azonos és különböző típusú adatok is viszonyíthatók egymáshoz, így csoportosító, összehasonlító és leíró sorokból is képezhetők. A különböző típusok alkalmazási lehetőségeit gyakorlati példák segítségével illusztráltuk. Az időbeli összehasonlítást szolgáló dinamikus, a sokaság vagy minta szerkezetét jellemző megoszlási és koordinációs, a területi összehasonlítást lehetővé tevő területi összehasonlító, valamint a különböző adatok összehasonlítására alkalmas intenzitási viszonyszámokkal a hétköznapokban is gyakran találkozhatunk. Az ellenőrző kérdések és feladatok segítik a hallgatót abban, hogy felmérje, milyen mértékben sikerült ezen egyszerű elemzési módszer alapjait elsajátítania.

3.3.2 Önellenőrző kérdések

1. Mi a viszonyszám?
2. Melyek a megoszlási és koordinációs viszonyszámok közös és eltérő tulajdonságai?
3. Milyen viszonyszám segítségével végezhetünk időbeli összehasonlítást?
4. Mi a különbség a bázis és láncviszonyszámok között?
5. Mely esetben alkalmazható területi összehasonlító viszonyszám?
6. Hogyan határozható meg a tervfeladat viszonyszám?
7. Hogyan határozható meg a tervteljesítési viszonyszám?
8. Milyen viszonyszámot használhatunk különböző típusú adatok összehasonlítására?
9. Mi a különbség egyenes és fordított intenzitási viszonyszám között?
10. Milyen összefüggés van nyers és tisztított, valamint a viszonyítási alapra vetített megoszlási viszonyszám között?

3.3.3 Gyakorló tesztek

☸ Állapítsa meg a viszonyszámok típusát!

	Megoszlási	Koordinációs	Dinamikus	Területi összehasonlító	Tervfeladat	Tervteljesítési	Intenzitási
Az euró árfolyama 1,35 EUR/USD							X
Egy vállalat árbevétele tavalyhoz képest 30%-kal emelkedett			X				
Egy ruházati bolt értékesítésének 30%-a nadrágok eladásából származik	X						
2010-ben az egy főre jutó évi átlagos csokoládéfogyasztás 28 kg volt							X

A tavalyi év kimagasló eredményeinek köszönhetően a vállalat a jövő évre további 10%-os árbevétel növekedéssel számol					X		
Az egy főre jutó GDP az Egyesült Államokban majdnem ötször akkora, mint Magyarországon				X			
Egy múzeum látogatottsági adatai alapján egy férfi látogatóra két nő jutott		X					
Egy vállalat a tervezetthez képest a válság miatt 25%-os visszaesést produkált						X	
2010-hez képest 2013-ban 20%-kal nőtt az importált áruk mennyisége.			X				
Az év eleji leárazásoknál egyes termékekre 50%-os árkedvezmény is igénybe vehető	X						

4. A VISZONYSZÁMOK GYAKORLATI ALKALMAZÁSA: MUNKAERŐ-PIACI ÉS PÉNZÜGYI STATISZTIKÁK

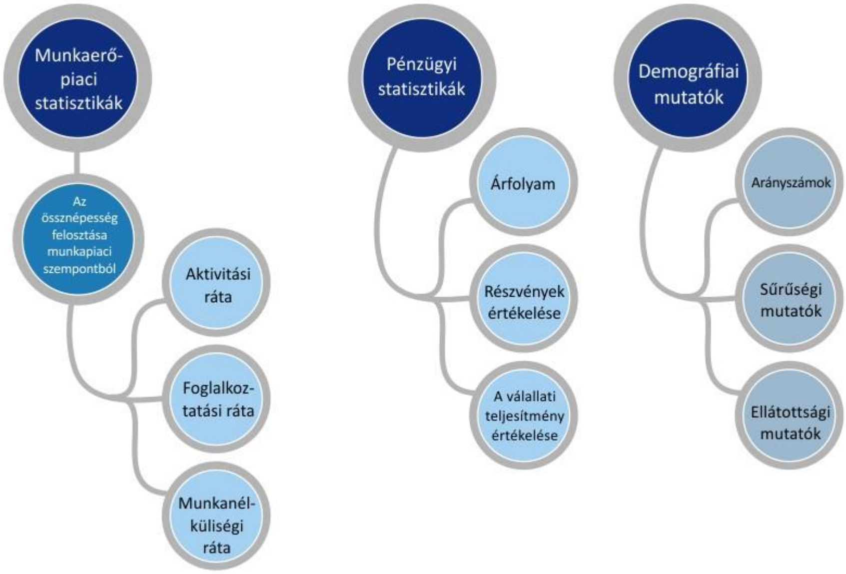
4.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A viszonyszámokat gyakran alkalmazzuk társadalmi és gazdasági jelenségek vizsgálatánál. A leggyakrabban a dinamikus, valamint a megoszlási és az intenzitási viszonyszámokat használjuk, melyek megjelennek a demográfiai és a munkaerő-piaci statisztikákban, valamint a pénzügyi számításokban is.

A lecke célja, hogy megismertesse a hallgatókat a viszonyszámok gyakorlati alkalmazásával, kiemelten a munkaerőpiac elemzésének módszertanával és a pénzügyi területen megjelenő mutatókkal, valamint az intenzitási viszonyszámok különböző megjelenési formáival. A hallgató ismerje a főbb munkaerő-piaci mérőszámokat, azok kiszámítási módját, valamint a vállalatértékelés szempontjából releváns pénzügyi mutatókat, melyek viszonyszámként határozhatók meg. Átfogó célként fogalmazódik meg, hogy a társadalmi és gazdasági élet területén a hallgató ismerje fel a viszonyszámokat.

4.2 TANANYAG

A gazdasági élet valamennyi területén megjelennek a viszonyszámok. A munkaerő-piaci statisztikákban főként megoszlási viszonyszámokat használunk, valamint az egyes munkapiaci kategóriák időbeli változásának számszerűsítésére dinamikus viszonyszámokat. A pénzügyi területen és a demográfiai mutatók számításánál leggyakrabban különböző típusú adatokat viszonyítunk egymáshoz.



5. ábra: A viszonyszámok alkalmazásának főbb területei

4.2.1 Munkaerő-piaci statisztikák

Makroökonómiai szempontból[3] az össznépességet munkaképes és nem munkaképes csoportokra bontjuk. A munkaképes lakosság az, aki életkoránál és egészségügyi állapotánál fogva hajlandó és képes munkát vállalni.

- A munkaképes lakosság statisztikai számbavétele életkor tekintetében változatos, az országszintű, valamint az európai és nemzetközi statisztikák sem egységesek. Az OECD által készített és az európai statisztikákban a 15–64 éves korosztály tartozik ide, míg az ILO által közzétett és például a magyar kimutatásokban is a 15-74 éves korosztályt sorolják ebbe a kategóriába. Vannak kimutatások, ahol az alsó korhatár 16 év, de olyan is, ahol csak 14. A különböző adatbázisokból származó munkaerő-piaci mérőszámok megítélésénél figyelni kell tehát az egyes kategóriák tartalmára.

A munkaképes lakosság további két részre osztható: gazdaságilag aktív és gazdaságilag inaktív csoportokra. Gazdaságilag inaktív az, aki nem vesz részt a társadalmilag szervezett munkavégzésben, ide sorolhatók például háztartásbeliek, de a lottómilliomos is. A gazdaságilag aktívak

[3] A munkaerő-piaci összefüggésekről bővebben olvashat a Közgazdaságtan II. tananyagban.

csoportját a foglalkoztatottak és a munkanélküliek alkotják. A munkanélküliek tovább bonthatók kényszerű és önkéntes formákra.

6. ábra: Az össznépesség felosztása munkapiaci szempontból

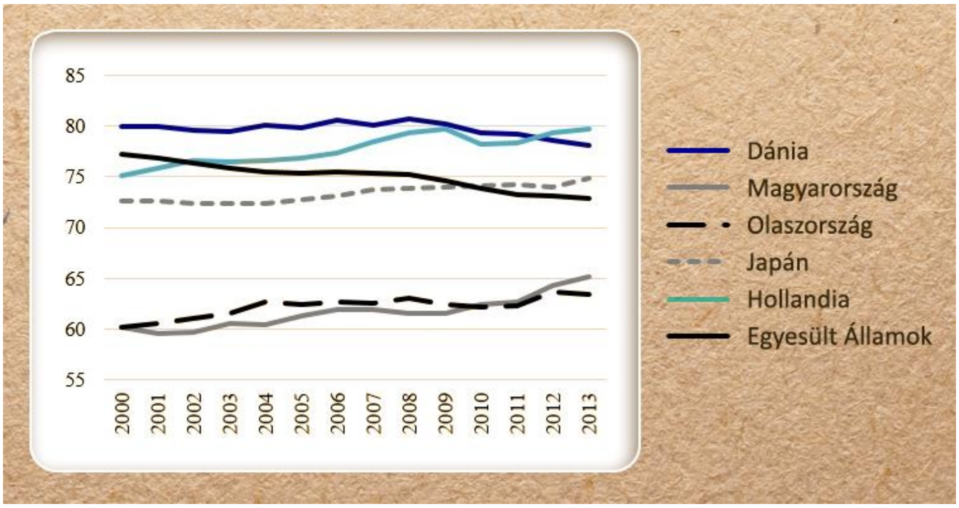
Ezen felosztás alapján számítják a legfontosabb munkaerő-piaci mérőszámokat, így az aktivitási arányt, valamint a foglalkoztatási és a munkanélküliségi rátát.

Aktivitási ráta

Az aktivitási ráta a gazdaságilag aktívak arányát mutatja a munkaképes korú lakosságon belül.

$$\text{Aktivitási ráta} = \frac{\text{Gazdaságilag aktívak}}{\text{Munkaképes korú lakosság}}$$

Az Egyesült Államokban és Japánban az aktivitási arány 70% felett van, míg az európai átlag alig éri el ezt az értéket. Számos országban alig kúszik 60% fölé az arányszám értéke. Az alacsony aktivitási arány komoly munkaerő-piaci problémákról árulkodhat, hiszen a gazdaságilag aktívak képezik a munkapiac kínálati oldalát. Az aktivitási rátának nem csupán az aktuális értéke lehet fontos munkaerő-piaci jelzőszám, hanem az arányszám időbeli változása is szolgáltat információt, hiszen jelzi a munkapiaci változások irányát.



7. ábra: Az aktivitási ráta alakulása néhány országban 2000 és 2013 között

Forrás: OECD (2014)

Az ábrán látszik, hogy a kiválasztott országokban változatos az aktivitási arány és annak elmozdulási iránya is a vizsgált időszakban. Néhány esetben az arányszám emelkedése, máshol a csökkenése figyelhető meg. Az Egyesült Államok aktivitási rátája folyamatosan csökken, míg például Japán vagy Hollandia mutatója emelkedik. Az aktivitási ráta képet ad a munkapiac helyzetéről, a változások hátterében azonban a számlálóban és a nevezőben lévő munkaerő-piaci kategória változása is állhat. A viszonyszámok gyakorlatban történő alkalmazásánál figyelni kell arra, hogy a két adat hányadosaként képzett mutatók változását a viszonyítandó adat és a viszonyítási alap változása egyaránt indokolhatja.

A foglalkoztatási és a munkanélküliségi ráta

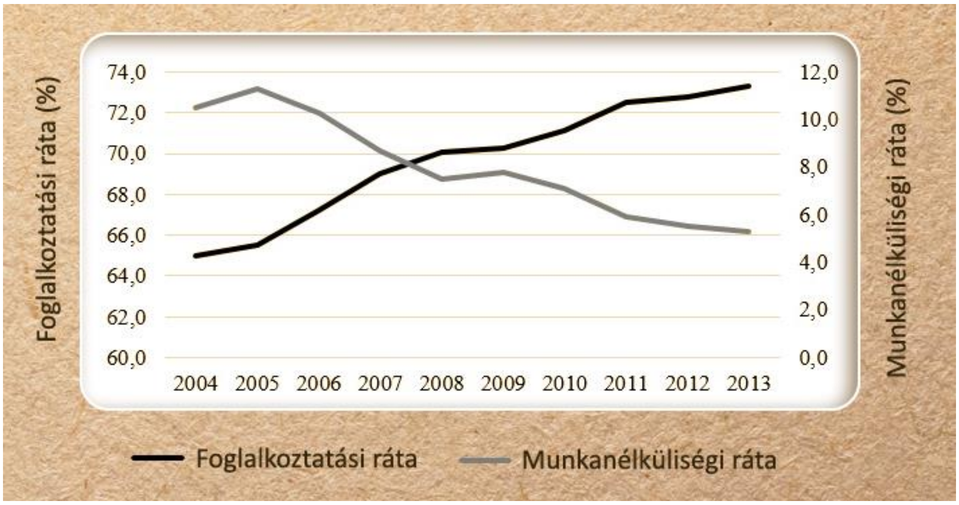
A foglalkoztatási ráta a foglalkoztatottak arányát mutatja a gazdaságilag aktívakon belül.

$$\text{Foglalkoztatási ráta} = \frac{\text{Foglalkoztatottak}}{\text{Gazdaságilag aktívak}}$$

A munkanélküliségi ráta a regisztrált munkanélküliek arányát mutatja a gazdaságilag aktívakon belül.

$$\text{Munkanélküliségi ráta} = \frac{\text{Regisztrált munkanélküliek}}{\text{Gazdaságilag aktívak}}$$

A két mutató megoszlási viszonyszámként határozható meg, ahol az egyik részsokaság a foglalkoztatottak, a másik a munkanélküliek, a munkapiaci felosztás alapján pedig a két csoport együttesen alkotja a gazdaságilag aktív népességet. A foglalkoztatottaknál csak azt tudjuk figyelembe venni, akinek bejelentett munkaviszonya van, míg a munkanélkülieknél a regisztráltakról vannak csak adatok. Az adatszolgáltatás miatt a gyakorlatban nem valósul meg a megoszlási viszonyszámokra vonatkozó azonosság, a két ráta összege nem adja ki a 100%-ot. Nem veszik ugyanis figyelembe azokat, akik nem regisztrálnak a munkaügyi hivatalokban vagy nem legálisan vállalnak munkát. A számbavételi sajátosságok ellenére az időbeli változásoknál kirajzolódik azonban a két ráta azonos viszonyítási alapból fakadó összefüggése, vagyis ha a foglalkoztatási ráta emelkedik, akkor a munkanélküliségi ráta csökken.



8. ábra: Németország foglalkoztatási és munkanélküliségi rátáinak alakulása 2004 és 2013 között

Forrás: EUROSTAT (2014)

4.2.2 Pénzügyi statisztikák

A pénzügyek területén[4] gyakran alkalmaznak viszonyszámokat, számos mutató képezhető két adat hányadosaként. Ebben a részben kizárólag a pénzügyek területén viszonyszámokként képezhető mutatók statisz-

[4] A pénzügyi statisztikákról bővebben olvashat az alábbi szakirodalomban:
Copeland, T. – Koller, T. - Murrin, J. (2000): *Valuation. Measuring and managing the value of companies*. 3rd edition. McKinsey & Company Inc.

tikai jellemzőit mutatjuk be. Nem cél a pénzügyi sajátosságok részletes ismertetése, a mutatók szakmai alkalmazhatóságának megítélése.

A nemzetközi pénzügyek területén a legfontosabb viszonyszámként meghatározható mutató az **árfolyam**, mely két különböző fizetési eszköz értékét viszonyítja egymáshoz. A hivatalos árfolyamok alakulását a nemzeti bankok teszik közzé. Az árfolyam alakulása a nemzetközi kereskedelem volumenét jelentősen befolyásolja. Az euró más valutákhoz viszonyított árfolyamáról az Európai Központi Bank honlapján lehet tájékozódni:

9. Az Európai Központi Bank honlapja: http://www.ecb.europa.eu/stats/exchange/eurofxref/html/index.en.html

A legtöbb viszonyszámot azonban a vállalatértékelés területén alkalmazzák. A részvényekhez, valamint a befektetési döntésekhez számos mutató nyújt segítséget, melyet két egymással összefüggő adat hányadosaként képeznek.

Viszonyszámok alkalmazása a részvények értékelésénél

A részvények kapcsán számos olyan mutató említhető, melyek viszonyszámként határozható meg. Az egyik legfontosabb az **EPS** (Earnings per Share), azaz az egy részvényre jutó nyereség, amely esetében az adózott nyereséget viszonyítják a részvények számához.

$$\mathrm{EPS} = \frac{\text{adózott nyereség}}{\text{részvények száma}}$$

A befektetési döntések meghozatalánál és a részvényekkel kapcsolatban gyakran említett mutató a **P/E** (Price/Earnings), ami szintén viszonyszám. Ez a mutató a részvény árfolyamát viszonyítja az egy részvényre jutó nyereséghez. Befektetési szempontból a mutató minél magasabb értéke kedvező.

$$\mathrm{P/E} = \frac{\text{részvény árfolyama}}{\text{egy részvényre jutó nyereség}} = \frac{\mathrm{P}_0}{\mathrm{EPS}_1}$$

A mutatóról bővebben olvashat az alábbi szakmai portálon:

10. P/E: http://www.investopedia.com/terms/p/price-earningsratio.asp

A vállalati teljesítmény értékelése pénzügyi mutatókkal[5]

A vállalatok teljesítményének megítélése gyakran történik pénzügyi mutatók segítségével, melyek statisztikailag viszonyszámok mintájára képez-

[5] A fejezet Hollóné Kacsó Erzsébet (2011): *Vállalatértékelés*. Mutatók, modellek, üzlet- és vagyonértékelés c. jegyzetére támaszkodik.

hetők. Ezek a mutatók könnyen kiszámíthatók és értelmezhetők, használatuknak a pénzügyek területén azonban korlátaik is vannak.

A vállalat jövedelmezőségének vizsgálatához a ROE, ROA és ROS, valamint a ROI mutatókat használják. A saját tőke jövedelmezőségét mérő **ROE** (Return on Equity) úgy határozható meg, hogy az adózott eredményt a saját tőkéhez viszonyítjuk.

$$\mathrm{ROE} = \frac{\text{Adózott eredmény}}{\text{Saját tőke}}$$

A mutatóról bővebben olvashat az alábbi szakmai portálon:

11. ROE: http://www.investopedia.com/terms/r/returnonequity.asp

A **ROA** (Return on Asset) az eszközarányos megtérülés mutatója, mely lényegében a vállalat eszközeinek jövedelemtermelő képességéről ad képet. A mutató az adózott eredményt az összes eszközhöz viszonyítja.

$$\mathrm{ROA} = \frac{\text{Adózott eredmény}}{\text{Összes eszköz}}$$

A mutatóról bővebben olvashat az alábbi szakmai portálon:

12. ROA: http://www.investopedia.com/terms/r/returnonassets.asp

Az árbevétel arányos jövedelmezőség mutatója a **ROS** (Return on Sales) az adózott eredményt a bevételekhez viszonyítja.

$$\mathrm{ROS} = \frac{\text{Adózott eredmény}}{\text{működéshez tartozó nettó árbevétel és egyéb bevétel}}$$

A mutatóról bővebben olvashat az alábbi szakmai portálon:

13. ROS: http://www.investopedia.com/terms/r/ros.asp

A beruházások értékelésénél használatos még a befektetett tőke megtérülését jelző **ROI** (Return on Investment) mutató. Viszonyszámként számíthatók továbbá a hatékonyságot számszerűsítő forgási sebesség mutatói, valamint a tőkeszerkezeti és eladósodottsági, illetve likviditási mutatók is.

A pénzügyi mutatókról bővebben olvashat az alábbi szakmai oldalon:

14. Pénzügyi mutatók: http://kfknowledgebank.kaplan.co.uk/KFKB/Wiki%20Pages/Financial%20Performance%20Indicators%20%28FPIs%29.aspx

4.2.3 Demográfiai mutatók

A népességhez kapcsolódó mutatók többsége intenzitási viszonyszám. A **sűrűségmutatók** közül a legismertebb a népsűrűség, mely egy adott területegységre vetített lakosok számát mutatja. A mutató mértékegysége általában fő/km^2, melyből látszik, hogy különböző típusú adatok hányadosaként határozható meg.

$$\text{Népsűrűség} = \frac{\text{Az ország teljes népessége}}{\text{Az ország területe}} \rightarrow \frac{\text{fő}}{\text{km}^2}$$

Az **ellátottsági mutatók** is intenzitási viszonyszámok, így például az orvos ellátottság vagy a kulturális ellátottság, ez utóbbi a kulturális intézmények vagy rendezvények számát vetíti a lakosok számára.

$$\text{Orvos ellátottság} = \frac{\text{összes orvos}}{\text{lakosság}} \rightarrow \frac{\text{orvos}}{\text{lakos}}$$

Ezen mutatók közös vonása, hogy a viszonyítási alapot a népesség vagy annak egy meghatározott része képezi. Ezeknél a mutatóknál a viszonyítandó adat és a viszonyítási alap gyakran megcserélődik, így az egyenes és fordított intenzitási viszonyszámokra is példát szolgáltatnak. Ellátottsági mutatószám meghatározható területi egységre vetítve is, ilyen például az egy településre jutó boltok száma.

A demográfiai jellemzésekben fontos szerepet töltenek be az **arányszámok**, melyek például a születések vagy halálozások számát viszonyítják a népességhez, így nyerhető a természetes szaporodás vagy fogyás jelzőszáma. Ezen mutatóknál fordul elő a gyakorlatban nyers és tisztított intenzitási viszonyszám. Tisztított intenzitási viszonyszámra példaként a termékenységi arányszám említhető, mely a szülőképes korú nőkre vetített élveszületéseket mutatja.

A Világbank adatbázisában számos demográfiai mutatószám található:

15. Világbank demográfiai mutatói: http://econ.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/0,,contentMDK:20451597~pagePK:64133150~piPK:64133175~theSitePK:239419,00.html

A nemzetgazdasági indikátorok között is vannak olyanok, melyeket intenzitási viszonyszámként határozhatunk meg. A leggyakrabban alkalmazott mutató az egy főre jutó GDP, vagy a viszonyítási alap szűkítésével megalkotott munkatermelékenységet reprezentáló egy munkásra vetített GDP.

4.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

4.3.1 Összefoglalás

A viszonyszámok a gazdasági és társadalmi jelenségek vizsgálatánál gyakran alkalmazott módszerek. A munkaerő-piaci statisztikákban főként megoszlási és dinamikus viszonyszámokat alkalmazunk, míg a pénzügyek és demográfia területén az intenzitási viszonyszámok előfordulása gyakoribb.

Az össznépesség munkapiaci felosztása alapján számíthatók a főbb munkaerő-piaci indikátorok, így az aktivitási arány, valamint a foglalkoztatási és munkanélküliségi ráta. A pénzügyek területén legismertebb viszonyszám az árfolyam, valamint a részvény értékelésénél fontos P/E, valamint a befektetési döntések meghozatalánál releváns, ROE, ROA, és ROS mutatók. A demográfia területén számos intenzitási viszonyszám képezhető, ide tartozik például a népsűrűség, az ellátottsági mutatók, a születési és halálozási arányszámok. Szintén két összefüggő adat hányadosaként, intenzitási viszonyszámként határozható meg az egy főre, illetve egy munkásra vetített GDP adat is.

4.3.2 Önellenőrző kérdések

1. Melyek a viszonyszámok alkalmazásának fő területei?
2. Melyek a leggyakrabban alkalmazott viszonyszám típusok?
3. Melyek a munkaerőpiac jellemzésére használt viszonyszámok?
4. Hogyan számítható ki az aktivitási ráta?
5. Hogyan számítható ki a foglalkoztatási ráta?
6. Hogyan számítható ki a munkanélküliségi ráta?
7. Mi a közös a foglalkoztatási és munkanélküliségi ráta számításában?
8. Említsen néhány pénzügyek területén alkalmazott viszonyszámot!
9. Milyen viszonyszámokkal jellemezhető a vállalati teljesítmény?
10. Említsen néhány demográfia területén alkalmazható viszonyszámot!

4.3.3 Gyakorló tesztek

☸ Döntse el az alábbi állításokról, hogy igazak (I) vagy hamisak (H)!

Az össznépesség munkapiaci szempontból gazdaságilag aktív és gazdaságilag inaktív csoportokra osztható.

IGAZ **HAMIS**

A foglalkoztatási ráta a foglalkoztatottak népességen belüli arányát mutatja.

IGAZ **HAMIS**

Gazdaságilag inaktív az, aki nem vesz részt a társadalmilag szervezett munkavégzésben.

IGAZ HAMIS

A munkanélküliségi ráta számításánál csak a regisztrált munkanélkülieket tudjuk figyelembe venni.

IGAZ HAMIS

Az aktivitási ráta a gazdaságilag aktívak arányát mutatja a munkaképes lakosságon belül.

IGAZ HAMIS

A népsűrűség kiszámításánál az ország területéhez viszonyítjuk a lakosság számát.

IGAZ HAMIS

A P/E mutató az adózott eredményt viszonyítja a saját tőkéhez.

IGAZ **HAMIS**

Az orvos ellátottság mutatójának meghatározásához a lakosok számát viszonyítjuk az orvosok számához.

IGAZ **HAMIS**

A ROA mutató az összes eszközhöz viszonyítja az adózott eredményt.

IGAZ HAMIS

A termékenységi arányszám nyers intenzitási viszonyszám, amely a szülőképes korú nőkre vetített élveszületéseket mutatja.

IGAZ **HAMIS**

5. LEÍRÓ STATISZTIKA

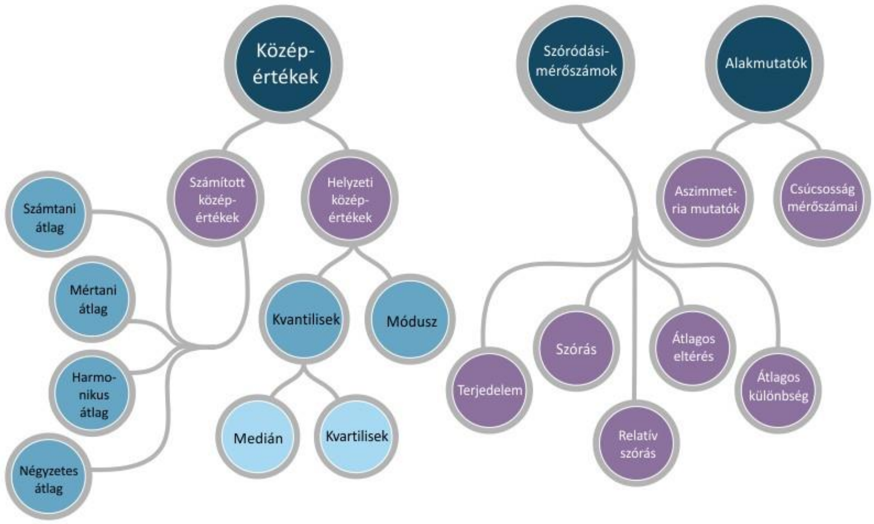
5.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A leíró statisztika eszköztára a sokaság vagy minta mennyiségi ismérv szerinti tömör jellemzésére használható alapeloszlás jellemzőket foglalja magában. A mutatók három fő csoportba sorolhatók, melyek közül a középértékek az adatsor jellemző értékeire, a szóródási mérőszámok az adatok különbözőségére, az alakmutatók pedig a gyakorisági görbe alakjára vonatkozóan szolgáltatnak információt.

A lecke célja, hogy megismertesse a hallgatókat a leíró statisztika eszköztárával. A mutatók sajátosságainak, számítási formáinak és értelmezésének elsajátításával a hallgatók képessé válnak átfogóan jellemezni mennyiségi ismérv szerint egy adatsort statisztikai szempontból. Kiemelten fontos, hogy a hallgatók átlássák a mutatók rendszerét, valamint az egyes indikátorok előnyeit és hátrányait, hogy megfelelő következtetéseket tudjanak levonni a vizsgált jelenség kapcsán.

5.2 TANANYAG

A sokaság vagy minta mennyiségi ismérv szerinti, tömör jellemzésére a leíró statisztika eszköztárát alkalmazzuk. Az alapeloszlás jellemzők szintén a statisztikai módszerek egyszerű, és gyakran használatos eszközei.



9. ábra: A leíró statisztika eszköztára

Az alapeloszlás jellemzők kiszámítási formáinál különbséget kell tennünk az egyedi értékek, valamint az osztályközös gyakorisági sorból számítható értékek között. A mutatók kiszámítása az átlagok kivételével csak egyedi értékekre vonatkoztatva kerül bemutatásra.

5.2.1 Középértékek

A középértékek a sokaság vagy minta olyan jellemző értékei, melyeknek közös tulajdonsága, hogy MINDIG az adatsor legkisebb és legnagyobb értéke közé esnek. Kiszámításuk egyszerű algebrai művelet útján történik, egyértelműen értelmezhetők, ugyanakkor tulajdonságaik alapján figyelni kell az egyes mutatók alapján levonható következtetések általánosíthatóságára.

Számított középértékek: az átlagok

Az átlag valamennyi ember számára közismert fogalom, hiszen a mindennapi életünk során számtalanszor találkozunk vele, így például az átlagkereset, a tanulmányi átlag, az átlagár, vagy az átlagon aluli, átlagon felüli, esetleg átlagos jelző használatánál. Ezek a kifejezések a legtöbb esetben egyszerű számtani átlagot rejtenek, azonban a statisztikában többféle átlagformát különböztetünk meg, s az egyes átlagtípusok alkalmazására is különös hangsúly kerül. Az adatok jellegétől függően megkülönböztetünk számtani, mértani, harmonikus és négyzetes átlagot.

Számtani (aritmetikai) átlag

A legegyszerűbb és leggyakrabban alkalmazott átlagtípus, melynek kiszámítási formája eltér attól függően, hogy egyedi értékek vagy csoportosított, azaz osztályközös gyakorisági sorok állnak rendelkezésünkre.

Egyedi értékek esetén a számtani átlagot úgy kapjuk, hogy az átlagolandó értékeket összeadjuk és elosztjuk az elemek számával.

Egy tanuló a félév során az alábbi érdemjegyeket szerezte: 2, 3, 5, 5, 4

$$Átlageredménye: \bar{x} = \frac{2+3+5+5+4}{5} = \mathbf{3,8}$$

Az elemszám növekedésével egyre hosszadalmasabb művelet a számtani átlag ilyen formán való kiszámítása, így az adatokat célszerű csoportosítani. Ha az adatsorban kevés ismérvérték van, de azok többször is előfordulnak, érdemes *gyakorisági sor*t képezni, mely tömöríti az adatokat úgy, hogy megadja az ismérvértékekhez tartozó elemek számát, vagyis gyakoriságát. Ebben az esetben (gyakorisággal) súlyozott számtani átlagot számítunk.

Gyakorisági sor esetén a súlyozott számtani átlagot úgy kapjuk, hogy az átlagolandó értékeket megszorozzuk az osztályukhoz tartozó gyakoriság értékével és elosztjuk az elemszámmal, mely ebben az esetben a gyakoriságok összegével egyezik meg.

$$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \cdots + f_n \cdot x_n}{f_1 + f_2 + \cdots + f_n} = \frac{\sum f \cdot x}{\sum f}$$

A $\sum$ jel (szumma) összeadást jelent, a jel mögötti kifejezés egészére vonatkozik az összeadás, vagyis súlyozott átlag esetén az osztályonként lévő gyakoriságokat páronként szorozzuk meg a saját értékükkel és végül a szorzatok értékeit összegezzük.

Egy tanuló a félév során az alábbi érdemjegyeket szerezte:
2, 3, 5, 5, 4, 4, 3, 5, 2, 5, 5, 4, 4, 5, 3, 4, 5, 5, 4, 4

Határozza meg az átlageredményét!

A számtani átlag kiszámításához célszerű a rendelkezésre álló adatainkat csoportosítani aszerint, hogy az egyes érdemjegyekből hányat kapott a hallgató.

13. *Munkatábla gyakorisági sor készítéséhez és értelmezéséhez*

Érdem-jegy	A hallgató érdemjegyeinek száma (db)	
		Ebből a gyakorisági sorból azt tudjuk kiolvasni, hogy a hallgató
5	8	← 8 db 5-ös érdemjegyet
4	7	← 7 db 4-es érdemjegyet
3	3	← 3 db 3-as érdemjegyet
2	2	← 2 db 2-es érdemjegyet
1	0	← 0 db 1-es érdemjegyet
Összesen	*20*	← összesen 20 db érdemjegyet szerzett a félév során

Az átlageredmény az érdemjegyekre vonatkozik, és nem a gyakoriságokra.

$$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \cdots f_n \cdot x_n}{f_1 + f_2 + \cdots f_n} \rightarrow \frac{8 \cdot 5 + 7 \cdot 4 + 3 \cdot 3 + 2 \cdot 2 + 0 \cdot 1}{8 + 7 + 3 + 2 + 0} = \mathbf{4{,}05}$$

A számtani átlag kiszámítását tovább bonyolíthatja az adatok számának növekedése. Nagy elemszám esetén ajánlott az adatokat osztályközös gyakorisági sorba rendezni. Az osztályokat úgy kell kialakítani, hogy az elemek egyértelműen besorolhatók legyenek. Lényeges megjegyezni, hogy a valós osztályközök határait a felső határok jelzik, az alsó határokat a megkülönböztetés miatt jelöljük ki, ezért fiktív határként kezeljük[6].

A felső határnál magasabb értékek már a következő intervallumba esnek, az alsó határok értékét így az adatok nagyságrendjéhez kell igazítani. Valamennyi számítás az egymást követő osztályközök felső határainak figyelembevételével történik.

Osztályközös gyakorisági sor esetén a súlyozott számtani átlagot úgy kapjuk, hogy az intervallumokba rendezett adatoknál az adott és az azt megelőző intervallum felső határainak egyszerű számtani átlagaként kapott osztályközepet megszorozzuk az osztályhoz tartozó gyakoriság értékével és elosztjuk a gyakoriságok összegével.

$$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \cdots f_n \cdot x_n}{f_1 + f_2 + \cdots f_n} = \frac{\sum f \cdot x}{\sum f}$$

Egy csoport statisztikai zárthelyi dolgozatának pontszámait intervallumba rendezett formában ismerjük. *Határozza meg a dolgozatok átlageredményét!*

14. *Munkatábla osztályközös gyakoriság sor adatainak értelmezéséhez*

Elért pontszám	Hallgatói dolgozatok száma (db)
0 – 20	8
21 – 40	12
41 – 60	18
61 – 80	12
81 – 100	10
Összesen	*60*

Ebből az osztályközös gyakorisági sorból azt tudjuk kiolvasni, hogy

← 8 hallgató írt 0 és 20 pont közötti dolgozatot
← 12 hallgató írt 21 és 40 pont közötti dolgozatot
← 18 hallgató írt 41 és 60 pont közötti dolgozatot
← 12 hallgató írt 61 és 80 pont közötti dolgozatot
← 10 hallgató írt 81 és 100 pont közötti dolgozatot
← összesen 60 hallgató írt dolgozatot

A súlyozott számtani átlag kiszámításához meg kell határozni az osztályközepeket, mely az adott és a megelőző osztályköz felső határainak egyszerű számtani átlagaként számítható ki.

[6] Az osztályközök kialakításánál lényegében *elválasztó értékek*et jelölünk ki, melyek később a felső határok lesznek, így például a dolgozatok pontszámánál a 20, 40, 60, 80 érték osztja fel az adatsort. Az egyértelmű besorolás kedvéért az adatok nagyságrendjének figyelembevételével megadjuk az intervallumok alsó határát is. A pontszámok esetében csak egész számokkal dolgozunk, így ha egy dolgozat pontszáma például 20 vagy 21 lehet. A 21 már a következő osztályközbe tartozik, így az alsó határként a 20 után következő egész számot választjuk.

15. *Munkatábla az osztályközepek kiszámításához*

Elért pontszám	OSZTÁLYKÖZÉP	Hallgatói dolgozatok száma (db)
0 - 20	$\frac{0+20}{2} = \mathbf{10}$	8
21 - 40	$\frac{20+40}{2} = \mathbf{30}$	12
41 – 60	$\frac{40+60}{2} = \mathbf{50}$	18
61 – 80	$\frac{60+80}{2} = \mathbf{70}$	12
81 - 100	$\frac{80+100}{2} = \mathbf{90}$	10
Összesen		*60*

$$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \cdots f_n \cdot x_n}{f_1 + f_2 + \cdots f_n} \rightarrow \frac{8 \cdot 10 + 12 \cdot 30 + 18 \cdot 50 + 12 \cdot 70 + 10 \cdot 90}{8 + 12 + 18 + 12 + 10} = 51{,}33$$

A statisztika dolgozatot írt 60 hallgató átlagosan 51,33 pontot ért el.

A számtani átlag tulajdonságai

A számtani átlag kiszámítása akár egyedi értékekből, akár gyakorisági sorokból egyszerű feladat, s minden olyan esetben meghatározható, ahol az adatok összegezhetők. Egy adatsornak csak egy számtani átlaga van, mely megkönnyíti az érték értelmezését. Különböző adatsorok összehasonlíthatók a számtani átlaguk alapján, így például két film esetében a mozilátogatók 1-től 5-ig terjedő skálán mért véleményének átlaga tükrözi azt, hogy átlagosan melyik filmről jobb a kialakított vélemény. Hasonlóképpen két csoport dolgozatainak átlagos pontszáma képet ad arról, hogy melyik csoport teljesített jobban a dolgozat megírásakor. A számtani átlag kiszámításához valamennyi elemre szükség van, így a teljes adatsorra vonatkoztatva vonhatunk le következtetéseket. A számtani átlag legnagyobb hátránya azonban az, hogy érzékeny a kiugróan alacsony vagy kiugróan magas értékekre, melyek torzíthatják az értékét, s ezáltal a levonható következtetések is bizonytalanná válnak.

Egy adatsorban a kiugróan alacsony vagy kiugróan magas értékeket outliernek nevezzük.

Az outlierek kezelése az adatsorban számos kérdést felvet. Statisztikailag indokolt az adatsorból való kizárásuk, melyet azonban gyakorlati oldalról könnyen kritizálhatunk. Az adatsorról átfogó képet ugyanis csak akkor kaphatunk, ha valamennyi rendelkezésre álló adatot figyelembe veszünk.

A számtani átlag torzulásának jellegzetes példája az átlagkeresetek megállapítása, melyen sokan elcsodálkoznak, hiszen valódi keresetük, sőt a környezetükben élőké is elmarad attól. Ez a torzulás a magas keresetűek csekély, de kiugróan magas értékéből fakad. Kérdés persze, hogy mennyire lehetne általánosítható értéket megadni akkor, ha őket kizárnánk az átlagszámításból. Másik tipikus példát a makrogazdasági indikátorok területéről az Európai Unió GDP adatai szolgáltatják. Luxemburg jövedelmi adata a közösségen belül kiugróan magas, így valamennyi GDP értékkel végzett számítás eredménye torzul, s így az Unió átlagos jövedelme is. Vajon Luxemburg kihagyásával lehet átfogó képet alkotni az Európai Unió jövedelmei helyzetéről? A statisztikailag outlier értékek kezelésére így nincs is általánosan alkalmazható eljárás, az adott elemzésben egyedi mérlegelés alapján zárjuk ki vagy hagyjuk benne az elemzésben.

Mértani (geometriai) átlag

A mértani vagy geometriai átlag abban az esetben számítható, ha az átlagolandó értékek nem összegezhetők, viszont szorzatuknak van értelme.

Egyedi értékekből a mértani átlagot úgy kapjuk, hogy az elemeket összeszorozzuk, és annyiadik gyököt vonunk, ahány elemű az adatsor.

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} \rightarrow \bar{x}_g = \sqrt[n]{\pi\, x_n}$$

Ha valamelyik elem többször is előfordul, akkor *súlyozott mértani átlag*formát használunk:

$$\bar{x}_g = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots \cdot x_n^{f_n}} \rightarrow \bar{x}_g = \sqrt[n]{\pi\, x_n^{f_n}}$$

A π jel (pí) szorzást jelent, a jel mögötti kifejezés egészére vonatkozik, vagyis az értékeket összeszorozzuk, ha az elemek többször előfordulnak, akkor előbb az értékek hatványozzuk és a hatványozott értékeket szorozzuk össze.

Ismert, hogy egy szálloda vendégforgalma 2010-ről 2011-re 1%-kal, 2011-ről 2012-re 2%-kal, 2012-ről 2013-ra 5%-kal nőtt. *Állapítsa meg, hogy 2010 és 2013 között átlagosan hogyan változott a vendégforgalom!*

$$\bar{x}_g = \sqrt[3]{1{,}01 \cdot 1{,}02 \cdot 1{,}05} = \mathbf{1{,}0265}$$

A mértani átlag alapján megállapítható, hogy a vizsgált időszak alatt átlagosan 2,65%-kal nőtt a szálloda vendégforgalma.

❁ Az utóbbi években ismert hogyan változott egy szálloda vendégforgalma. *Határozza meg a változás átlagos mértékét!*

16. *A szálloda vendégforgalmának átlagos változása*

Év	Változás az előző évhez képest
2008	-
2009	1,05
2010	1,10
2011	0,95
2012	1,05
2013	1,10

$$\bar{x}_g = \sqrt[5]{1{,}05^2 \cdot 1{,}1^2 \cdot \ 0{,}95} = 1{,}0485$$

2008 és 2013 között a szálloda vendégforgalma átlagosan 4,85%-kal nőtt

A mértani átlag számítását jellegzetesen a dinamikus viszonyszámok átlagolásánál alkalmazzuk. Ebben az esetben nem a százalékos változás mértékeit átlagoljuk, hanem a viszonyszámok tulajdonságai miatt a tizedestört alakokat.

Harmonikus átlag

A harmonikus átlagot akkor használjuk, ha sem az értékek összegének, sem a szorzatuknak nincs értelme, de a reciprokok összegének igen. A harmonikus átlagot általában intenzitási viszonyszámok esetén alkalmazzuk.

Egyedi értékekből a harmonikus átlagot úgy kapjuk, hogy az elemek számát elosztjuk az értékek reciprokainak összegével.

$$\bar{x}_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x_n}}$$

Ha valamelyik elem többször is előfordul, akkor *súlyozott harmonikus átlag*formát használunk:

$$\bar{x}_h = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_n}{x_n}} = \frac{n}{\sum \frac{f_n}{x_n}}$$

❁ Egy irodában három nyomtató működik, melyek közül az egyik 12 oldal/perc, a másik 6 oldal/perc, a harmadik 4 oldal/perc teljesítménnyel nyomtatja ki a színes oldalakat. *Adja meg, hogy mekkora a nyomtatók átlagos teljesítménye, vagyis egy perc alatt átlagosan mennyi színes oldalt nyomtatnak ki az irodában!*

Ebben az esetben sem számtani, sem mértani átlag nem számítható, hiszen a nyomtatók teljesítménye eltérő, ezért azt kell megnéznünk előbb, hogy egy oldalt mennyi idő alatt nyomtatnak ki.

$$\frac{\text{oldal}}{\text{perc}} \rightarrow \frac{\text{perc}}{\text{oldal}} \text{ vagyis az együttes teljesítmény: } \frac{1}{12}+\frac{1}{6}+\frac{1}{4}=\frac{6}{12}=\frac{1}{2}$$

A számítás alapján 1 perc alatt az egyik nyomtató 1/12-ed, a másik 1/6, a harmadik 1/4 papírt nyomtatna ki, együttesen pedig 1/2-edet. Mivel 3 nyomtató van, s a kérdés nem az, hogy egy oldalt menynyi idő alatt nyomtatnak ki, hanem az, hogy egy perc alatt hány oldal ezért használjuk a harmonikus átlagformát:

$$\bar{x}_h = \frac{3}{\frac{1}{12}+\frac{1}{6}+\frac{1}{4}} = \frac{3}{\frac{1}{2}} = 6$$

Harmonikus átlaggal számolva megkaptuk, hogy a három különböző teljesítményű nyomtató együttesen egy perc alatt 6 színes oldalt nyomtat ki.

Négyzetes (kvadratikus) átlag

A négyzetes átlagot gyakorlati esetekben ritkán alkalmazzuk, leginkább más indikátorok, elsősorban a szóródás mérőszámainak meghatározásakor használjuk. A négyzetes átlag alkalmazását eltérő előjelű értékek átlagolása indokolhatja.

Egyedi értékekből a négyzetes átlagot úgy kapjuk, hogy az elemeket négyzetre emeljük, majd ezeket összeadjuk, elosztjuk az elemek számával és végül négyzetgyököt vonunk belőle.

$$\bar{x}_q = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}} = \sqrt{\frac{\sum x_n^2}{n}}$$

$$\text{súlyozott átlagforma: } \bar{x}_q = \sqrt{\frac{f_1 \cdot x_1^2 + f_2 \cdot x_2^2 + \cdots + f_n \cdot x_n^2}{f_1 + f_2 + \cdots f_n}} = \sqrt{\frac{\sum f_n \cdot x_n^2}{n}}$$

Helyzeti középértékek

A helyzeti középértékek az adatsor olyan jellegzetes értékei, melyek helyzetük alapján jellemzik az adatsort.

Módusz

A módusz kiszámítása abban az esetben alkalmas a sokaság jellemzésére, ha az adatsorban egy vagy több elem többször fordul elő. Előnye, hogy nem csak mennyiségi ismérvek esetén használható, hanem minőségi és területi jellemzőknél is.

🕮 **A módusz az adatsor leggyakrabban előforduló eleme, a tipikus érték. Egyedi értékek esetén a móduszt megfigyeléssel határozzuk meg.**

A módusz nem számítható minden adatsorból, ugyanis ha minden elem különböző, akkor nincs módusz. Előfordulhat olyan eset is, amikor több elem is többször, ugyanannyiszor szerepel az adatsorban, ekkor nem egyértelmű a módusz meghatározása. Ezekben az esetekben mindegyik többször előforduló elemet módusznak tekintjük, az adatsort többmóduszú vagy polimodális sornak nevezzük.

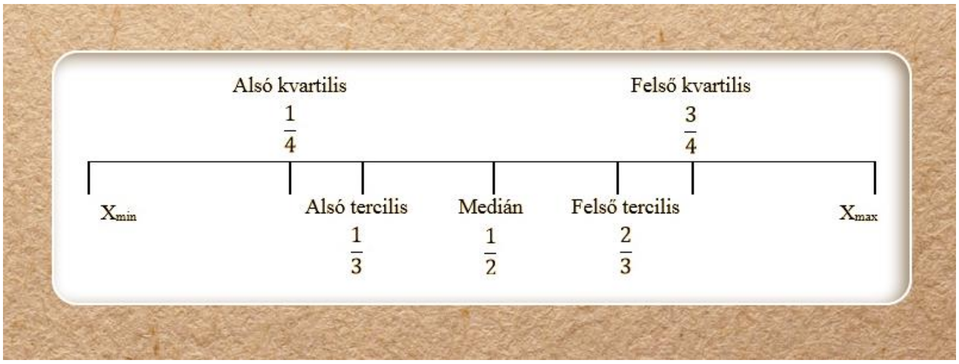
❅ Egy tanuló a félév során az alábbi érdemjegyeket szerezte:
2, 3, 5, 5, 4, 4, 3, 5, 2, 5, 5, 4, 4, 5, 3, 4, 5, 5, 4, 4

Határozza meg az adatsor móduszát!

A leggyakrabban előforduló érdemjegy a jeles (5)

Kvantilisek

A kvantilisek a növekvő sorrendbe rendezett adatsor nevezetes osztópontjai. A kvantilisek közül a legismertebb a *medián*, amely a felezőpont, a harmadoló pontokat *tercilisek*nek, a negyedelő pontokat pedig *kvartilisek*nek nevezzük. A kvantilisek számítási módja ugyanazon logika mentén történik, részletesebben a mediánnal, mint a leggyakrabban alkalmazott osztóponttal és a kvartilisekkel foglalkozunk.



10. ábra: A nevezetes osztópontok elhelyezkedése az adatsorban

Medián

A medián a középső elem, az adatsor felezőpontja. A medián meghatározása egyszerű, mert egyértelmű helye van, s csak egy lehet belőle. Előnye, hogy nem csak mennyiségi, hanem minőségi ismérvek esetén is jól alkalmazható, melyek legalább sorrendi mérési skálán mérhetők. Érzékeny a kiugró értékekre, de alkalmasabb az adatsor jellemzésére, mint az átlag. A mediánt úgy értelmezzük, mint azt az értéket, melynél az adatsor elemeinek fele nagyobb, a fele pedig kisebb.

A medián kiszámításának feltétele, hogy az adatok növekvő sorrendben, rangsorba legyenek rendezve. Egyedi értékek esetén a rangsorba rendezett elemek $\frac{n+1}{2}$ –dik tagja lesz a medián értéke.

Egyedi értékek esetén fontos, hogy az adatsor páros vagy páratlan elemszámú. Páratlan elemszámú adatsor esetén a medián ténylegesen középső eleme az adatsornak, míg páros elemszámnál a két középső érték közé esik, azoknak az egyszerű számtani átlaga.

☸ Egy tanuló a félév során az alábbi érdemjegyeket szerezte:
2, 3, 5, 5, 4, 4, 3, 5, 2, 5, 5, 4, 4, 5, 3, 4, 5, 5, 4, 4

Határozza meg az adatsor mediánját!

Az adatokat előbb rendezzük sorba:
2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5

A mediánt az (n+1)/2 helyen keressük, mely ezen 20 elemű adatsor esetén: (20+1)/2 = 10,5, vagyis páros elemszámú az adatsor, így a medián két érték közé esik, mely ebben az esetben a 10. és 11. elem, vagyis 4 és 4, melyeknek a számtani átlaga is 4 → a medián értéke 4. *A hallgató félév során szerzett érdemjegyeinek fele 4-esnél rosszabb, a fele 4-esnél jobb*.

Kvartilisek

A kvartilisek az adatsor negyedelő pontjai, amelyek közé lényegében a 2/4-nél, azaz ½-nél elhelyezkedő medián is tartozik. Az alsó kvartilis az adatsor ¼-énél helyezkedik el, a felső kvartilis pedig a ¾-énél.

🕮 **Az alsó kvartilis az az érték, melynél az adatsor elemeinek ¼-e kisebb, ¾-e pedig nagyobb. A felső kvartilis az az érték, melynél az adatsor elemeinek ¾-e kisebb, ¼-e pedig nagyobb.**

A kvartilisek kiszámításának is feltétele, hogy az adatok növekvő sorrendbe legyenek rendezve. Egyedi értékek esetén a rangsorba rendezett elemek $\frac{n+1}{4}$ –dik tagja lesz az alsó kvartilis, $\frac{3(n+1)}{4}$ –dik tagja pedig a felső kvartilis értéke.

☸ *Az előző példát folytatva határozza meg az adatsorban az alsó és felső kvartilis értékét!*

Adatok: 2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5

Az alsó kvartilist az (n+1)/4 helyen keressük, mely ezen 20 elemű adatsor esetén: (20+1)/4 = 5,25, vagyis páros elemszámú adatsor esetén a kvartilis is két érték közé esik, mely ebben az esetben az 5. és 6. elem, vagyis 3 és 4. Az alsó kvartilis értékét úgy kapjuk, hogy az alsó értékhez hozzáadjuk a két érték különbségének az ¼-ét, vagyis 3 + (4-3)/4→ az alsó kvartilis értéke **3,25**.

A hallgató félév során szerzett érdemjegyeinek ¼-e 3,25-nél ~3-nál rosszabb, ¾-e 3,25-nél ~ 3-nál jobb.

A felső kvartilist a 3(n+1)/4 helyen keressük, mely ezen 20 elemű adatsor esetén: 3(20+1)/4 = 15,75, vagyis páros elemszámú adatsor esetén a felső kvartilis is két érték közé esik, mely ebben az esetben az 15. és 16. elem, vagyis 5 és 5→ a felső kvartilis[7] értéke **5**.

A hallgató félév során szerzett érdemjegyeinek ¾-e rosszabb, mint jeles (5), ¼-e a példához igazítva nyilván jeles.

5.2.2 Szóródási mérőszámok

A leíró statisztika eszköztárában kiemelt fontosságúak az adatok különbözőségét számszerűsítő mutatók. A terjedelem, a szórás, a relatív szórás, az átlagos eltérés vagy az átlagos különbség kiszámításával átfogóbb képet kaphatunk az adatsorról.

Terjedelem

A terjedelem meghatározása egyszerű, mert azt az intervallumot jelöli ki, melyben az adatsor értékei ingadoznak.

A terjedelem az adatsor legnagyobb és legkisebb értékének a különbsége.

A terjedelem kiszámításához tehát az adatsor két szélső értékére van szükség, a mutató nem jelzi azt, hogy ezen szűkebb vagy tágabb intervallumon belül az adatok miként helyezkednek el. A terjedelmet megbízhatóan meghatározni csak egyedi értékek esetén lehet.

7 A felső kvartilis értékét úgy kapjuk, hogy az alsó értékhez hozzáadjuk a két érték különbségének a ¾-ét.

$$R = X_{max} - X_{min}$$

- Osztályközös gyakorisági sor esetén a terjedelem kiszámítása csak az utolsó osztályköz felső és az első osztályköz alsó határának különbségeként volna lehetséges. Ez azonban nem vezetne valós eredményre, hiszen számos esetben mind az első, mind az utolsó osztályköz nyitott, s csak azért zárjuk le, hogy az alap eloszlásjellemzők kiszámíthatóak legyenek. Éppen ezért ezen terjedelmi értéket nem számítjuk.

A gyakorlatban az elemek elhelyezkedésének eloszlásáról több információt szolgáltat az interkvartilis terjedelem.

Az interkvartilis terjedelem az elemek belső 50%-ának elhelyezkedését mutatja, amely a felső és az alsó kvartilis értékének különbségeként határozható meg.

$$IQR = Q_3 - Q_1$$

Szórás és relatív szórás

A szórás és a relatív szórás a szóródási mérőszámok közül a legfontosabb, melyek kiszámítása és jelölése is eltér sokaság és minta esetén.

Az elemek számtani átlagtól vett átlagos négyzetes eltérését a szórás az alapadatok mértékegységében, a relatív szórás pedig százalékban mutatja.

Sokaság	Egyedi értékek esetén	*Minta*
$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{n}} = \sqrt{\frac{\sum d^2}{n}}$	SZÓRÁS	$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{\sum d^2}{n-1}}$
$V = \frac{\sigma}{\bar{x}}$	RELATÍV SZÓRÁS	$V = \frac{s}{\bar{x}}$

11. ábra: A szórás és relatív szórás kiszámítási módjai sokaság és minta esetén

Az alapadatok mértékegységében kifejezett szórás mellett több, eltérő mértékegységű adatsor összehasonlításra használjuk a relatív szórást. A szórás kiszámításánál előbb vesszük az egyes elemek számtani átlagtól való eltérését, melyet négyzetre emelünk, mert az eltérés mértéke fontos és nem az, hogy az átlagtól pozitív vagy negatív irányban van-e eltérés. Összeadjuk a páronként vett eltérések négyzetre emelt értékét, majd elosztjuk az elemek számával és négyzetgyököt vonunk. A kiszámítási forma alapján látszik, hogy *a szórás értéke mindig POZITÍV*. A relatív szórás esetében az elemek átlagtól való átlagos eltérését viszonyítjuk az átlaghoz. Mivel az átlagtól való eltérés nem lehet nagyobb, mint az átlag maga, ezért a szórás értéke mindig kisebb, mint az átlag értéke. Ebből következően a szórás és az átlag hányadosaként kapott *relatív szórás értéke MINDIG 0 és 1 közé esik*.

☸ Egy tanuló a félév során az alábbi érdemjegyeket szerezte:
2, 3, 5, 5, 4, 4, 3, 5, 2, 5, 5, 4, 4, 5, 3, 4, 5, 5, 4, 4

Határozza meg az adatsorban a terjedelem, az interkvartilis terjedelem, a szórás és a relatív szórás értékét!

$R = X_{max} – X_{min}$ = 5-2 = 3 → az adatok 2 és 5 között egy 3 egység hosszúságú intervallumon belül ingadoznak.

$IQR = Q_3 – Q_1$ = 5 – 3,25 = 1,75 → az adatok középső 50%-a, 3,25 és 5 között egy 1,75 hosszúságú intervallumon belül ingadozik.

$$\sigma = \sqrt{\frac{\sum f \cdot (x - \bar{x})^2}{n}}$$

$$\sigma = \sqrt{\frac{8 \cdot (5 - 4{,}05)^2 + 7 \cdot (4 - 4{,}05)^2 + 3 \cdot (3 - 4{,}05)^2 + 2 \cdot (2 - 4{,}05)^2}{20}} = \mathbf{0{,}9734}$$

$$V = \frac{\sigma}{\bar{x}} = \frac{0{,}9734}{4{,}05} = 0{,}2404 \rightarrow \mathbf{24{,}04}\%$$

Az érdemjegyek az átlagtól átlagosan 0,9734 jeggyel, azaz 24,04%-kal térnek el.

Átlagos eltérés és átlagos különbség

Az adatsor elemeinek különbözőségét többféleképpen számszerűsíthetjük. Az átlagtól való eltérés számszerűsítésének legegyszerűbb formája, ha az átlagtól vett abszolút eltéréseket átlagoljuk.

🕮 **Az átlagos eltérés az elemek számtani átlagtól vett átlagos abszolút eltérését számszerűsíti.**

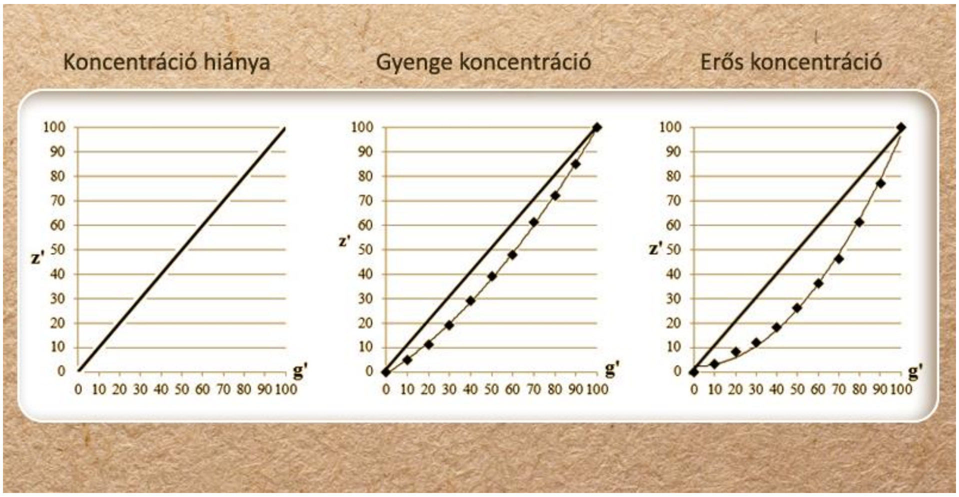
Az értékek nem csak fix értékhez, így a számtani átlaghoz viszonyíthatók, hanem egymáshoz is, az átlagos különbség az adatok egymástól való eltérését határozza meg.

🕮 **Az átlagos különbség az adatok egymástól vett átlagos, abszolút különbségeinek számtani átlaga.**

Ezt a különbséget alkalmazzuk a GINI mutató számításánál, melyet a koncentráció számszerűsítésére használhatunk.

🕮 **A koncentráció az adatok sűrűsödése, tömörülése a sokaság egyik eleme körül.**

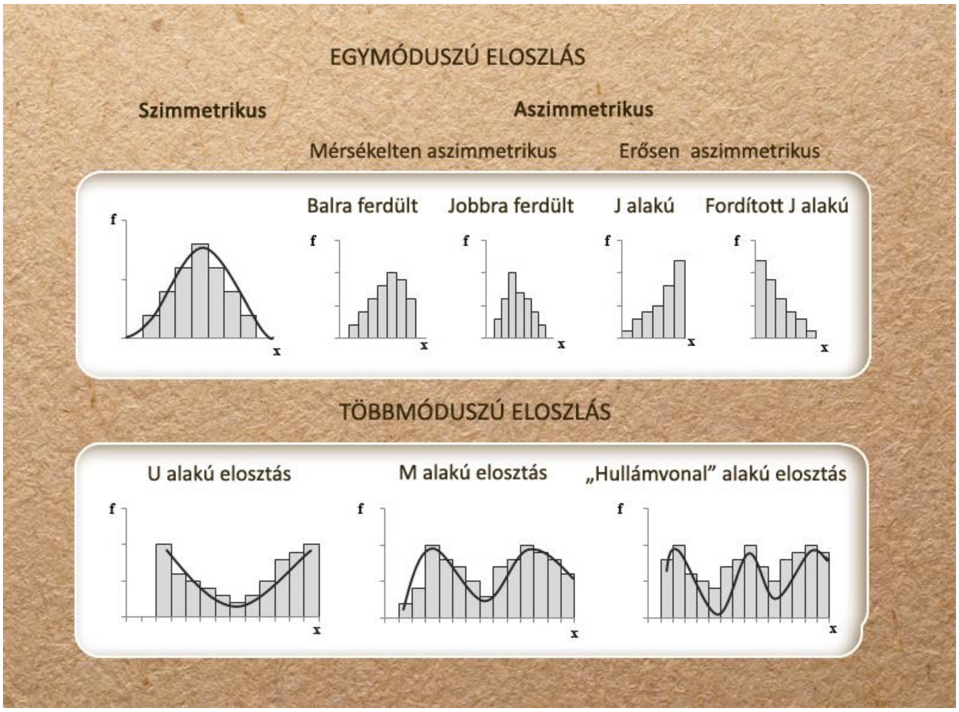
A koncentráció GINI mutatóval való mérése a gyakorlatban a jövedelemegyenlőtlenségek számszerűsítésénél jelenik meg. A GINI koefficiens azt mutatja meg, hogy a világ népessége és jövedelme hogyan oszlik meg. Statisztikai szempontból a koncentrációt célszerűbb a sokkal illusztratívabb Lorenz görbével vizsgálni, mely a relatív gyakoriságok és a relatív értékösszegek alakulását hasonlítja össze. A Lorenz görbét egy egységnyi oldalú négyzetben ábrázoljuk, melynek átlója a koncentráció hiányát jelzi, ezen a vonalon ugyanis a relatív gyakoriság és a relatív értékösszeg megegyezik. Az adatsor alapján számított relatív gyakoriság (g') és a relatív értékösszeg (z') értékeket összekötő görbe minél távolabb esik az átlótól, annál nagyobb fokú a koncentráció.



12. ábra: A koncentráció vizsgálata Lorenz görbével

5.2.3 Alakmutatók

Az alakmutatók segítségével a gyakorisági görbe alakjára vonatkozóan vonhatunk le következtetést. Számszerűsíteni tudjuk, hogy a szimmetrikus eloszláshoz képest ferdeség vagy aszimmetria, illetve csúcsosság tekintetében hogyan tér el az adatsor.



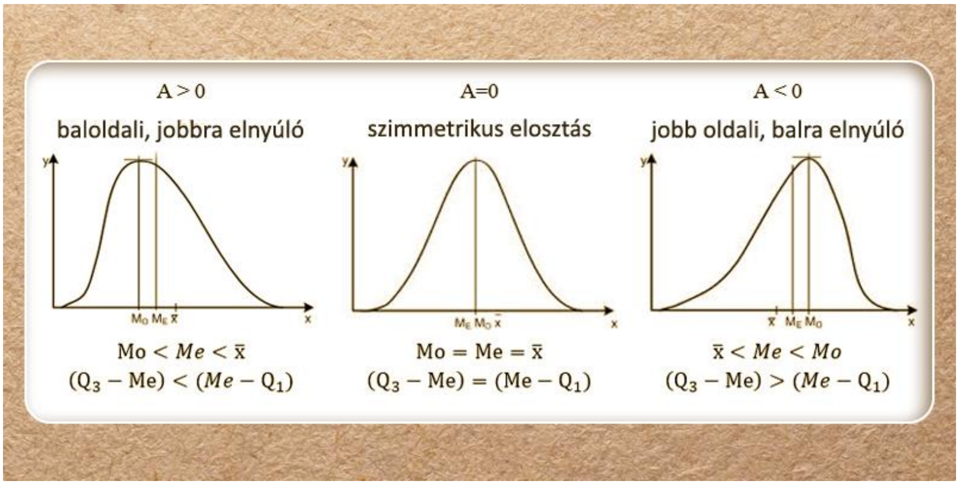
13. ábra: Az empirikus eloszlások típusai

Az aszimmetria (ferdeség) mérőszámai

A leggyakrabban alkalmazott mérőszámok az aszimmetria vagy ferdeség számszerűsítésére a Pearson-féle A mutató és az F mutató. Az *A mutató* a számtani átlag, a módusz és a szórás felhasználásával, az *F mutató* pedig az alsó és felső kvartilis, valamint a medián értékének figyelembevételével határozza meg az aszimmetria mértékét.

$$A = \frac{\bar{x} - Mo}{\sigma} \qquad F = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$$

Az A mutató értékének nincs abszolút korlátja, előjele mutatja meg az aszimmetria irányát.



14. ábra: A szimmetrikus és aszimmetrikus eloszlások tulajdonságai

Szimmetrikus eloszlás esetén a számtani átlag, a módusz és a medián egybeesik, erre a pontra nézve az elemek eloszlása szimmetrikus, melyet egy szabályos haranggörbe ír le. Az, hogy az aszimmetria bal vagy jobb oldali, illetve merre elnyúló a számtani átlag és a módusz helyzetétől függ:

- a móduszhoz képest, ha a számtani átlag nagyobb, akkor *baloldali* az aszimmetria
- a móduszhoz képest, ha a számtani átlag kisebb, akkor *jobboldali* az aszimmetria
- a számtani átlaghoz képest, ha a módusz kisebb, akkor *jobbra elnyúló*
- a számtani átlaghoz képest, ha a módusz nagyobb, akkor *balra elnyúló*

Az F mutató értéke -1 és 1, vagyis abszolút értékben 0 és 1 közé esik. Az F mutató az aszimmetria erősségét is jelzi: minél közelebb van az értéke 0-hoz annál gyengébb, s közeledve a szélső értékekhez egyre erősebb.

Csúcsosság

Az alakmutatók közé tartozik a csúcsosság (angol megfelelőjéből – curtosis – magyarosítva kurtózis), mely a szimmetrikus eloszláshoz képest az adatok nagyságrendi különbségeit szemlélteti, mely alapján lehet csúcsos vagy lapos az eloszlás görbe alakja. A csúcsosság mérőszáma jelzi a normálistól való eltérést: ha az érték pozitív, akkor a görbe csúcsos, míg ha az érték negatív, akkor lapos.

5.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

5.3.1 Összefoglalás

A leíró statisztika eszköztárát gyakran alkalmazzuk a statisztikai elemzések során. A primer adatgyűjtéseknél kiemelt fontosságú, hogy tömören jellemezzük az adatsort, mely számos információt szolgáltat.

A középértékek közös vonása, hogy mindig az adatsor legkisebb és legnagyobb eleme közé esnek. Mind a számított, mind a helyzeti középértékek könnyen meghatározhatók, hátrányuk azonban, hogy csak meghatározott feltételek mentén használhatók megbízhatóan. Kiemelendő, hogy az átlagszámítás formája függ a rendelkezésre álló adataink típusától. Az átlag mellett a gyakorlatban szórást is számítunk, s a két mutató együttesen alkot komplex képet az adatsorról. Az adatokkal való további elemzéshez szükség van a szóródás részletesebb számszerűsítésére, valamint az eloszlás formájára, s ily módon az aszimmetria és a csúcsosság megállapítására, mely a gyakorlatban leginkább grafikus ábrázolás útján történik.

5.3.2 Önellenőrző kérdések

1. Hogyan csoportosítható a leíró statisztika eszköztára?
2. Milyen számított középértékeket ismer?
3. Melyek a számtani átlag számításának előnyei és hátrányai?
4. Mik a kvantilisek?
5. Mi a módusz és melyek a számításának előnyei és hátrányai?
6. Mit számszerűsítenek a szóródási mérőszámok?
7. Mi a különbség a szórás és relatív szórás között?
8. Hogyan mérhető és illusztrálható a koncentráció?
9. Milyen reláció van az átlag és módusz között baloldali aszimmetria esetén?
10. Milyen reláció van az átlag és módusz között jobboldali aszimmetria esetén?

5.3.3 Gyakorló tesztek

❅ Egy adott napon megfigyeltek 100 embert, akik egy hipermarketben vásároltak. *Melyik alapeloszlás jellemző van értelmezve?*

A vásárlók fele 25 €-nál többet, fele ennél kevesebbet költött.

Középértékek

	Átlag
	Módusz
X	Medián
	Alsó kvartilis
	Felső kvartilis

Szóródás és alakmutatók

	Terjedelem
	Szórás
	Relatív szórás
	A mutató
	F mutató

A vásárlók átlagosan 24 €-t költöttek az üzletben.

Középértékek

X	Átlag
	Módusz
	Medián
	Alsó kvartilis
	Felső kvartilis

Szóródás és alakmutatók

	Terjedelem
	Szórás
	Relatív szórás
	A mutató
	F mutató

Az egyes vásárlók által az üzletben elköltött összegek az átlagos vásárlási értéktől átlagosan 4 €-val térnek el.

Középértékek

	Átlag
	Módusz
	Medián
	Alsó kvartilis
	Felső kvartilis

Szóródás és alakmutatók

	Terjedelem
X	Szórás
	Relatív szórás
	A mutató
	F mutató

A legtöbb vásárló 20 € értékben vásárolt.

Középértékek

	Átlag
X	Módusz
	Medián
	Alsó kvartilis
	Felső kvartilis

Szóródás és alakmutatók

	Terjedelem
	Szórás
	Relatív szórás
	A mutató
	F mutató

A vásárlók egynegyede 15 €-nál kevesebbet, háromnegyede ennél többet költött el az üzletben.

Középértékek

	Átlag
	Módusz
	Medián

Szóródás és alakmutatók

	Terjedelem
	Szórás
	Relatív szórás

X	Alsó kvartilis
	Felső kvartilis

	A mutató
	F mutató

☸ Döntse el az alábbi állításokról, hogy igazak (I) vagy hamisak (H)!

A módusz érzékeny az adatsorban lévő kiugróan alacsony vagy kiugróan magas értékekre.	H
A kvantilisek az adatsor nevezetes osztópontjai.	I
A szórás az alapadatok mértékegységében, a relatív szórás pedig százalékban fejezi ki az adatok egymástól vett átlagos eltérését.	H
A Lorenz görbe a koncentráció vizsgálatára alkalmas.	I
Az F mutató a csúcsosság számszerűsítésére alkalmazható.	H

6. ÖSSZETETT VISZONYSZÁMOK (FŐÁTLAGOK) ÖSSZEHASONLÍTÁSA STANDARDIZÁLÁS SEGÍTSÉGÉVEL

6.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

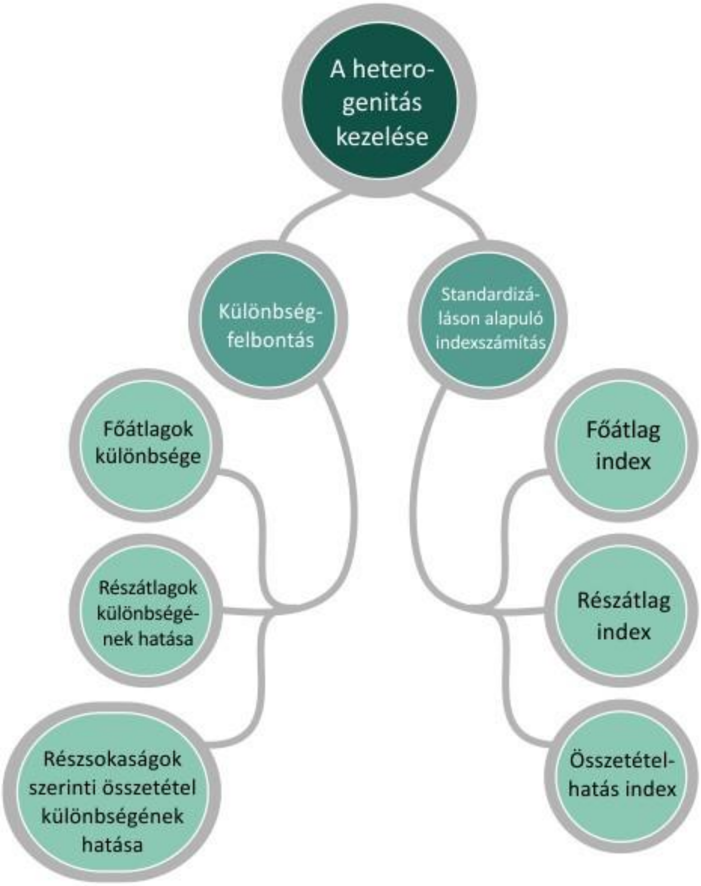
A statisztikai elemzések során gyakran előfordul, hogy a vizsgált sokaság (vagy minta) nem homogén, mely megnehezíti az egyszerű elemzési módszerek használatát. A heterogenitás kezelhető úgy, hogy a vizsgált sokaságot adott ismérv szerint homogén részsokaságokra bontjuk, s vizsgálatainkat nemcsak a teljes sokaságra, hanem a részsokaságokra vonatkoztatva is elvégezzük. Heterogén sokaságok esetén leggyakrabban összehasonlítást végzünk, mely történhet időbeli, illetve területi vagy minőségi jellemzők szerint. A sokaságok közötti különbséget, illetve adott sokaság esetén a változást előidéző tényezők a standardizálás módszerével vizsgálhatók.

A lecke célja, hogy megismertesse a hallgatókat egy olyan statisztikai módszerrel, amelynek segítségével képesek az egyszerű statisztikai módszereket heterogén sokaság esetén is alkalmazni. A hallgató ismerje a különbségfelbontás, valamint a standardizáláson alapuló indexszámítás módszerét és alkalmazási lehetőségeit. A különbséget, illetve változást előidéző komponenseket és indexeket tudja értelmezni, s lássa át a köztük lévő összefüggéseket.

6.2 TANANYAG

A heterogén sokaságok sokféle elemzési lehetőséget kínálnak, azonban az egyszerű statisztikai módszerek alkalmazásához a heterogenitást valahogyan kezelni kell. Ennek egyik formája, hogy csoportképzéssel homogén részsokaságokat hozunk létre, vagyis a sokaságokat csoportosítjuk valamilyen megkülönböztető ismérv szerint (pl. statisztika csoport bontása szakok szerint). Az egyszerű statisztikai módszerek alkalmazásánál, ha egy sokaság valamely vizsgált tényező szempontjából heterogén, akkor a homogén részsokaságokat is vizsgálni kell.

A csoportosított sokaságok minőségi vagy területi, illetve időbeli jellemzők alapján összehasonlíthatóvá és elemezhetővé válnak. A minőségi vagy területi összehasonlítás különbségfelbontással végezhető el, melynek során a két sokaság egészére vonatkoztatott átfogó mérőszám, a főátlag vagy összetett viszonyszám különbségét bontjuk tényezőkre. Adott sokaság időbeli változásának vizsgálatához a standardizáláson alapuló indexszámítás módszerét használjuk, melynek során a későbbi és korábbi időszak főátlagának hányadosát bontjuk elemeire.



15. ábra: Az összetett viszonyszámok (főátlagok) összehasonlításának módszerei

6.2.1 A heterogenitás kezelése: csoportképzés

A heterogén sokaságot az átlag önmagában nem jellemzi megfelelően, szükség van a homogén részsokaságokat jellemző átlagokra is.

A teljes sokaságra vonatkozó átlag a főátlag, a részsokaságra vonatkozó átlag a részátlag. A főátlag a részátlagok részsokaságok abszolút vagy relatív gyakoriságával súlyozott számtani átlaga. A főátlag súlyozott harmonikus átlag formájában is kiszámítható attól függően, hogy mely adatok állnak rendelkezésre.

Az átlagos értékek számítási formájuk alapján intenzitási viszonyszámok, melyek két különböző típusú adat hányadosaként képezhetők, csoportosított sokaság esetén a nem csoportosító sorban lévő adatokból.

Az egyes csoportokra vonatkoztatott viszonyszámokat részviszonyszámnak, a teljes sokaságra vonatkoztatott viszonyszámot pedig összetett viszonyszámnak nevezzük.

Egy vállalatnál a dolgozókat fizikai és szellemi foglalkozásúakra oszthatjuk. Külön-külön meg tudjuk határozni a fizikai és szellemi dolgozók átlagkeresetét, ha ismerjük mennyi a számukra kifizetett összes bér és mennyien vannak. Az adatok ismeretében ugyanezen számítás elvégezhető az összes dolgozóra vonatkoztatva is.

17. *A vállalatnál dolgozó fizikai és szellemi foglalkozásúak átlagbérei*

Állomány csoport	Kifizetett összes bér (Ft)	Létszám (fő)	ÁTLAGBÉR (Ft/fő)	
	A	**B**	$\mathbf{V = \frac{A}{B}}$	
Fizikai	27 500 000	250	**110 000**	← **részviszonyszám / részátlag**
Szellemi	21 750 000	150	**145 000**	← **részviszonyszám / részátlag**
Összesen	*49 250 000*	*400*	***123 125***	← ***összetett viszonyszám / főátlag***

Az összetett viszonyszám ($\overline{V}$) és a főátlag ($\overline{X}$) többféleképpen is meghatározható, a jelöléseikben van eltérés:

$$\overline{V} = \frac{\sum A_i}{\sum B_i} = \frac{\sum(\mathbf{B_i} \cdot \mathbf{V_i})}{\sum \mathbf{B_i}} = \frac{\sum A_i}{\sum \frac{A_i}{V_i}} \qquad \overline{X} = \frac{\sum A_i}{\sum B_i} = \frac{\sum(\mathbf{B_i} \cdot \mathbf{\bar{x}_i})}{\sum \mathbf{B_i}} = \frac{\sum A_i}{\sum \frac{A_i}{\bar{x}_i}}$$

A képletek segítségével az alábbiak szerint számíthatók ki az összes dolgozó átlagkeresete:

$$\frac{\sum A_i}{\sum B_i} = \frac{49\,250\,000}{400} = 123\,125$$

$$\frac{\sum(B_i \cdot V_i)}{\sum B_i} = \frac{250 \cdot 110\,000 + 150 \cdot 145\,000}{400} = 123\,125$$

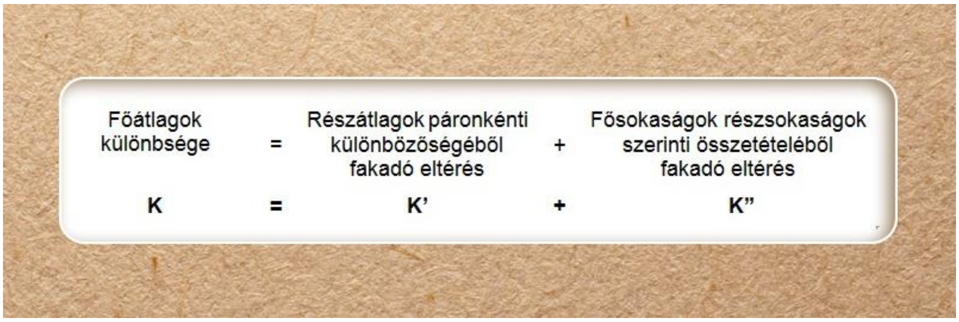
$$\frac{\sum A_i}{\sum \frac{A_i}{V_i}} = \frac{49\,250\,000}{\frac{27\,500\,000}{110\,000} + \frac{21\,750\,000}{145\,000}} = 123\,125$$

A főátlag értékét két tényező befolyásolja. Egyrészt a részátlagok értéke, másrészt a részsokaságok elemszáma (súly). A főátlagnak a legkisebb és legnagyobb részátlag közé kell esnie, hiszen egy átlagos érték, másrészt értéke függ a teljes sokaságon belül a részsokaságok megoszlásától.

Két azonos módon csoportosított sokaság *főátlaga összehasonlítható*, az eltérést egyrészt a sokaságokon belüli azonos részsokasághoz tartozó részátlagok különbözősége, másrészt a teljes sokaság részsokaságok szerinti összetételének eltérése okozza. Az összehasonlítás végezhető minőségi vagy területi és időbeli ismérvek segítségével. A minőségi vagy területi összehasonlításnál alkalmazható különbségfelbontásnál előre rögzíteni kell, hogy mely sokaság adata lesz a viszonyítandó adat, s melyik a viszonyítási alap, ezt az alsó indexben jelölni is szükséges. Időbeli összehasonlításnál azonban kötött a viszonyítás sorrendje, mert az időben későbbi időszak (tárgy) adatát viszonyítjuk a korábbi időszak (bázis) adatához. A bázisidőszakra vonatkozó adatokat 0, míg a tárgyidőszakra vonatkozó adatokat 1 indexekkel jelöljük. Az indexszámításnál tehát egy-egy képlet van a számításokhoz.

6.2.2 A standardizáláson alapuló különbségfelbontás módszere

A különbségfelbontás módszerét heterogén sokaságok minőségi vagy területi összehasonlítására alkalmazhatjuk. A két sokaság főátlagának különbségét egyrészt a részátlagok páronkénti különbözősége, másrészt a fősokaságok részsokaságok szerinti összetételének eltérése okozza.



16. ábra: A különbségfelbontás tényezői és kapcsolódásuk

A standardizálás során a főátlagok különbségét felbontjuk két tényező hatására úgy, hogy egyszerre csak az egyik tényező szerinti különbözőséget vizsgáljuk, a másikat változatlannak tekintjük, vagyis, ha a részátlagok különbözőségének hatását akarjuk kimutatni, akkor az összetétel marad változatlan (standard összetétel), míg ha a részsokaságok szerinti összetétel hatását vizsgáljuk, akkor a részátlagok maradnak változatlanok (standard részátlag). A standard értékek meghatározása úgy történik, hogy K' esetében az első fősokaság (viszonyítandó) összetétele standard ($B_S = B_1$), míg K" esetében a második fősokaság (viszonyítási alap) részátlaga változatlan ($V_S = V_2$).

A különbségfelbontás során elsőként a főátlagok (összetett viszonyszámok) különbségét (K) határozzuk meg, majd külön-külön számszerűsítjük az egyes részviszonyszámok különbözőségének (K') és az összetétel eltérésének (K") a hatását. A módszer lényege, hogy ha az egyes részviszonyszámok különbözőségének hatását akarjuk kimutatni, akkor változatlannak (standard) kell tekinteni a csoportok összetételét, míg ha az összetétel eltérésének hatását vizsgáljuk, akkor a részviszonyszámok maradnak változatlanok. K' meghatározásánál a kisebbítendőhöz tartozó összetétel standard, míg K" kiszámításánál a kivonandóhoz tartozó részviszonyszámok maradnak változatlanok.

$$\mathbf{K}_{1-2} = \overline{V_1} - \overline{V_2} = \frac{\sum B_1 V_1}{\sum B_1} - \frac{\sum B_2 V_2}{\sum B_2}$$

$$\mathbf{K}'_{1-2} = \frac{\sum \mathbf{B_S} V_1}{\sum \mathbf{B_S}} - \frac{\sum \mathbf{B_S} V_2}{\sum \mathbf{B_S}}, ahol\ B_s = B_1$$

$$\mathbf{K}''_{1-2} = \frac{\sum B_1 \mathbf{V_S}}{\sum B_1} - \frac{\sum B_2 \mathbf{V_S}}{\sum B_2}, ahol\ V_S = V_2$$

A kivonás irányát alsó indexben célszerű jelölni.

Előfordulhatnak olyan esetek, amikor csak az egyik tényező magyarázza a különbséget, ebben az esetben az azonos tényezők eltérésének értéke 0.

Ha K' = 0, akkor K = K" VAGY ha K" = 0, akkor K = K'

❆ Egy 76 fős statisztika csoportban műszaki menedzser (MM), emberi erőforrások (EE) és turizmus-vendéglátás (TV) szakos hallgatók írtak zárthelyi dolgozatot. Az oktató kétféle – A és B – feladatsort állított össze. *A feladatsort* 40 hallgató írt, melyből 20 fő MM, 5 fő EE és 15 fő TV szakos. B feladatsort 36 hallgató írt, melyből 18 fő MM, 6 fő EE és 12 fő TV szakos. Az *A feladatsor* kapcsán továbbá ismert, hogy a MM átlagosan elért eredménye 36,2 pont, az EE átlagosan elért eredménye 40,4 pont, a TV átlagosan elért eredménye 35,8 pont. A *B feladatsor* kapcsán továbbá ismert, hogy a MM átlagosan elért eredménye 38,4 pont, az EE átlagosan elért eredménye 32,6 pont, a TV átlagosan elért eredménye 39,2 pont. *Határozza meg az A és B feladatsort írt hallgatók átlagos pontszámát!*

$$\overline{V_A} = \frac{\sum B_A \cdot V_A}{\sum B_A} = \frac{20 \cdot 36{,}2 + 5 \cdot 40{,}4 + 15 \cdot 35{,}8}{40} = 36{,}575 \sim \mathbf{36{,}6}$$

$$\overline{V_B} = \frac{\sum B_B \cdot V_B}{\sum B_B} = \frac{18 \cdot 38{,}4 + 6 \cdot 32{,}6 + 12 \cdot 39{,}2}{36} = \mathbf{37{,}7}$$

18. *A statisztika csoport zárthelyi dolgozaton elért átlagos eredményei*

Szak	A FELADATSOR		B FELADATSOR	
	Létszám (fő)	*Átlagos pontszám*	*Létszám (fő)*	*Átlagos pontszám*
	B_A	V_A	B_B	V_B
MM	20	36,2	18	38,4
EE	5	40,4	6	32,6
TV	15	35,8	12	39,2
Összesen	40	**36,6**	36	**37,7**

A különbségfelbontás módszerével elemezze a feladatsorok átlag-eredménye közötti eltérést! Számszerűsítse, hogy mely tényezők és milyen mértékben járulnak hozzá a különbséghez!

$$K_{B-A} = \overline{V_B} - \overline{V_A} = 37{,}7 - 36{,}6 = \mathbf{1{,}1}$$

A *B feladatsor* átlagos pontszáma 1,1 ponttal magasabb, mint az A feladatsoré. Ezt két tényező magyarázza:

– az egyes feladatsoroknál a különböző szakos hallgatók eltérő átlageredménye (**K'**: részátlagok különbözőségének hatása – standard összetétel)

– a feladatsoroknál a hallgatók szakok szerinti megoszlása (**K"**: összetétel-hatás – standard részátlagok)

K' számításához a standard összetétel kiválasztásánál a kivonás iránya mérvadó, a kisebbítendőhöz (ez esetben B) tartozó összetételt tekintjük változatlannak.

$$K'_{B-A} = \frac{\sum B_B V_B}{\sum B_B} - \frac{\sum B_B V_A}{\sum B_B}$$

$$K'_{B-A} = 37{,}7 - \frac{18 \cdot 36{,}2 + 6 \cdot 40{,}4 + 12 \cdot 35{,}8}{36} = 37{,}7 - 36{,}8 = \mathbf{0{,}9}$$

Értelmezés: az A és B feladatsoroknál az egyes szakokhoz tartozó átlageredmények eltérése miatt 0,9 ponttal lenne magasabb a B feladatsort író hallgatók pontszáma, vagyis ha csak a különböző szakos hallgatók feladatsoronkénti átlageredménye lenne eltérő (azonos szakonkénti megoszlást feltételezve az egyes feladatsoroknál = standard összetétel), akkor a B feladatsor átlageredménye 0,9 ponttal lenne csak magasabb.

K'' számításához a standard részátlag kiválasztásánál is a kivonás iránya mérvadó, a kivonandóhoz (ez esetben A) tartozó részátlagot tekintjük változatlannak. Lényeges, hogy a standard tényezőt K' és K" számításához eltérő sokaságból kell választani!

$$\boldsymbol{K''}_{\boldsymbol{B-A}} = \frac{\sum \boldsymbol{B_B V_A}}{\sum \boldsymbol{B_B}} - \frac{\sum \boldsymbol{B_A V_A}}{\sum \boldsymbol{B_A}} = 36{,}8 - 36{,}6 = \boldsymbol{0{,}2}$$

Értelmezés: az A és B feladatsoroknál az eltérő szakonkénti megoszlás miatt 0,2 ponttal lenne magasabb a B feladatsort író hallgatók pontszáma, vagyis ha csak a különböző szakos hallgatók feladatsorokon belüli megoszlása lenne különböző (azonos szakonkénti részátlagokat feltételezve az egyes feladatsoroknál = standard részátlag), akkor a B feladatsor átlageredménye 0,2 ponttal lenne csak magasabb.

A különbségfelbontás végén, ha jól dolgoztunk a két oldalon azonosságot kell kapnunk, azaz az egyes tényezők összegének meg kell egyezniük a főátlagok különbségével.

Ellenőrzés: K = K' + K'' → ez esetben 1,1 = 0,9 + 0,2 ✓

6.2.3 A standardizáláson alapuló indexszámítás módszere

A standardizáláson alapuló indexszámítás során elsőként a két időszak közötti relatív változást számszerűsítjük. Ezt követően megnézzük a részviszonyszámokban bekövetkező változást (I'), melyet változatlan összetétel mellett számítunk ki, standard a későbbi időszak összetétele marad. Az összetételben bekövetkező változás (I") meghatározásához a részviszonyszámokat tekintjük változatlannak, standardnak a korábbi időszak részviszonyszámait választjuk.

17. ábra: A standardizáláson alapuló indexszámítás tényezői és kapcsolódásuk

❆ Egy elsős évfolyamon műszaki menedzser (MM), emberi erőforrások (EE) és turizmus-vendéglátás (TV) szakos hallgatók tanulnak, akiknek a 2012/2013. tanév két félévének tanulmányi átlageredményeiről az alábbi adatok ismertek. 2013/2014. tanév I. félévére 47 hallgató iratkozott be, melyből 16 fő műszaki menedzser, 5 fő emberi erőforrások és 26 fő turizmus-vendéglátás szakos. 2013/2014. tanév II. félévére a keresztféléves felvételi, illetve a tanulmányok halasztása miatt 52 hallgató iratkozott be, melyből 14 fő műszaki menedzser, 8 fő emberi erőforrások és 30 fő turizmus-vendéglátás szakos. A 2013/2014. tanév I. félév kapcsán továbbá ismert, hogy a MM tanulmányi átlageredménye 3,8; az EE átlagosan átlageredménye 4,4; a TV átlagosan átlageredménye 3,6. A 2013/2014. tanév II. félév kapcsán továbbá ismert, hogy a MM tanulmányi átlageredménye 3,82; az EE átlagosan átlageredménye 4,0; a TV átlagosan átlageredménye 3,5.

Határozza meg az egyes félévek tanulmányi átlageredményeit!

$$\overline{\boldsymbol{V_0}} = \frac{\sum \boldsymbol{B_0} \cdot \boldsymbol{V_0}}{\sum \boldsymbol{B_0}} = \frac{16 \cdot 3{,}8 + 5 \cdot 4{,}4 + 26 \cdot 3{,}6}{47} = \mathbf{3{,}75}$$

$$\overline{\boldsymbol{V_1}} = \frac{\sum \boldsymbol{B_1} \cdot \boldsymbol{V_1}}{\sum \boldsymbol{B_1}} = \frac{14 \cdot 3{,}82 + 8 \cdot 4{,}0 + 30 \cdot 3{,}5}{52} = \mathbf{3{,}66}$$

A 2013/2014. tanév I. félévében az évfolyam egészére vonatkoztatott tanulmányi átlageredmény 3,75; a II. félévben pedig 3,66.

19. *Az évfolyam tanulmányi átlageredményei a 2013/2014. tanévben*

Szak	**2013/2014. tanév I. félév**		**2013/2014. tanév II. félév**	
	Létszám (fő)	***Tanulmányi átlageredmény***	***Létszám (fő)***	***Tanulmányi átlageredmény***
	B_0	V_0	B_1	V_1
MM	16	3,8	14	3,82
EE	5	4,4	8	4,0
TV	26	3,6	30	3,5
Összesen	47	**3,75**	52	**3,66**

Az indexszámításnál mindig a *tárgyidőszakot viszonyítjuk a bázisidőszakhoz*, hiszen az időben előrehaladva történik változás, vagyis a viszonyítás irányát külön nem szükséges jelölni.

A standardizáláson alapuló indexszámítás módszerével elemezze a tanulmányi átlageredmények változását! Számszerűsítse, hogy mely tényezők és milyen mértékben járultak hozzá a változáshoz!

$$\boldsymbol{I} = \frac{\overline{\boldsymbol{V_1}}}{\overline{\boldsymbol{V_0}}} = \frac{\frac{\sum \boldsymbol{B_1 V_1}}{\sum \boldsymbol{B_1}}}{\frac{\sum \boldsymbol{B_0 V_0}}{\sum \boldsymbol{B_0}}} = \frac{3{,}66}{3{,}75} = \boldsymbol{0{,}976} \xrightarrow{\cdot 100} \boldsymbol{97{,}6} \rightarrow \boldsymbol{-2{,}4}\%$$

A II. félévben 2,4 %-kal romlottak a tanulmányi átlageredmények. Ezt két tényező magyarázza:

- az egyes félévekben a különböző szakos hallgatók eltérő tanulmányi átlageredménye (**I'**: részátlagok változásának hatása – standard tárgyidőszaki összetétel)
- az egyes félévekben a hallgatók szakok szerinti összetételének változása (**I''**: összetétel változásának hatása– standard bázisidőszaki részátlagok)

Időbeli változás esetén az indexszámításnál standard tárgyidőszaki összetételt alkalmazunk, vagyis I' kiszámításához a tárgyidőszak csoportok közötti megoszlását tekintjük változatlannak.

$$\boldsymbol{I'} = \frac{\frac{\sum \boldsymbol{B_S} V_1}{\sum \boldsymbol{B_S}}}{\frac{\sum \boldsymbol{B_S} V_0}{\sum \boldsymbol{B_S}}} = \frac{\frac{\sum \boldsymbol{B_1} V_1}{\sum \boldsymbol{B_1}}}{\frac{\sum \boldsymbol{B_1} V_0}{\sum \boldsymbol{B_1}}} = \frac{3{,}66}{\frac{14\cdot 3{,}8+8\cdot 4{,}4+30\cdot 3{,}6}{52}} = \frac{3{,}66}{3{,}78} = \boldsymbol{0{,}968} \xrightarrow{\cdot 100} \boldsymbol{96{,}8} \rightarrow -3{,}2\%$$

Értelmezés: 2013/2014. tanév II. félévében az évfolyam egészét tekintve 3,2%-kal romlottak volna a tanulmányi eredmények az egyes szakok tanulmányi átlageredményeinek változása miatt, vagyis ha

csak a különböző szakos hallgatók tanulmányi átlageredményei változtak volna (azonos szakonkénti megoszlást feltételezve az egyes félévekben = standard összetétel), akkor a II. félévben 3,2 %-kal romlottak volna összességében a tanulmányi eredmények.
Időbeli változás esetén az indexszámításnál standard bázisidőszaki részátlagokat alkalmazunk, vagyis I'' kiszámításához a bázisidőszak csoportonkénti részátlagait tekintjük változatlannak.

$$\boldsymbol{I''} = \frac{\frac{\sum B_1 \boldsymbol{V_S}}{\sum B_1}}{\frac{\sum B_0 \boldsymbol{V_S}}{\sum B_0}} = \frac{\frac{\sum B_1 \boldsymbol{V_0}}{\sum B_1}}{\frac{\sum B_0 \boldsymbol{V_0}}{\sum B_0}} = \frac{3{,}78}{3{,}75} = \mathbf{1{,}008} \rightarrow +\mathbf{0{,}8}\%$$

Értelmezés: 2013/2014. tanév II. félévében az évfolyam egészét tekintve 0,8%-kal javultak volna a tanulmányi eredmények a hallgatók szakonkénti megoszlásának változása miatt, vagyis ha csak a különböző szakos hallgatók létszáma, és ily módon a sokaságok összetétele változott volna (azonos szakonkénti tanulmányi átlageredményeket feltételezve az egyes félévekben = standard részátlag), akkor a II. félévben 0,8%-kal javultak volna összességében a tanulmányi eredmények.

Az indexszámítás végén, ha jól dolgoztunk a két oldalon azonosságot kell kapnunk, azaz az egyes tényezők szorzatának meg kell egyezniük a főátlag index értékével.

Ellenőrzés: I = I' × I" → ez esetben 0,976 = 0,968 × 1,008 ✓

6.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

6.3.1 Összefoglalás

A heterogén sokaságok változatos elemzési lehetőséget rejtenek, az egyszerű statisztikai módszerek alkalmazásához azonban a heterogenitást kezelni kell. Csoportképzéssel homogén részsokaságok képezhetők, melyek segítségével időbeli és minőségi vagy területi összehasonlítás végezhető. A standardizálás módszere lehetővé teszi, hogy különbséget és változást előidéző tényezők hatását számszerűsítsük. A különbségfelbontás során az összehasonlított sokaságok főátlagainak abszolút értékben vett eltérését bontjuk tényezőkre. Az időbeli változás okozta faktorok számszerűsítésére pedig a standardizáláson alapuló indexszámítás módszerét alkalmazhatjuk.

6.3.2 Önellenőrző kérdések

1. Mikor alkalmazható a különbségfelbontás módszere?

2. Mikor alkalmazható a standardizáláson alapuló indexszámítás módszere?
3. Mely tényezőkre bontható a főátlagok eltérését számszerűsítő különbség?
4. Mely tényezőkre bontható a főátlagok változását jelző főátlag index?
5. A részátlagok páronkénti különbözőségéből fakadó eltérés számítása során mi a standard tényező?
6. A fősokaságok részsokaságok szerinti összetételéből fakadó eltérés számítása során mi a standard tényező?
7. Az indexszámítás során hogyan döntjük el a viszonyítás irányát?
8. A részátlag index meghatározásánál mely időszak összetételét tekintjük változatlannak?
9. Az összetételhatás-index számításánál mely időszak részátlagait tekintjük változatlannak?
10. Milyen összefüggés van a különbségfelbontás és indexszámítás tényezői között?

6.3.3 Gyakorló tesztek

☸ Válassza ki a helyes megoldásokat!

Egy vállalatnál dolgozók kereseti viszonyainak ismeretében megállapítható, hogy a férfiak átlagosan 30 000 Ft-tal többet keresnek, mint a nők. Tudjuk, hogy a férfiak és nők iskolai végzettség szerinti megoszlásának eltérése miatt a férfiak átlagosan csak 24 000 Ft-tal keresnének többet. A férfiak és nők között az iskolai végzettségenkénti átlagkeresetek különbözősége miatt

a) a férfiak 6000 Ft-tal kevesebbet keresnének, mint a nők
b) a férfiak 6000 Ft-tal többet keresnének, mint a nők
c) a nők 6000 Ft-tal többet keresnének, mint a férfiak

Egy vállalatnál dolgozók kereseti viszonyainak ismeretében megállapítható, hogy a férfiak átlagosan 30 000 Ft-tal többet keresnek, mint a nők. Tudjuk, hogy a férfiak és nők iskolai végzettségenkénti átlagkeresetek különbözősége miatt a férfiak csak 6 000 Ft-tal keresnének többet. A férfiak és nők iskolai végzettség szerinti megoszlásának eltérése miatt

a) a férfiak 24 000 Ft-tal kevesebbet keresnének, mint a nők
b) a férfiak 24 000 Ft-tal többet keresnének, mint a nők
c) a nők 24 000 Ft-tal többet keresnének, mint a férfiak

Két kiemelt üdülőkörzet adatait megnézve azt láthatjuk, hogy a tavalyi évhez képest 10%-kal esett vissza a vendégek átlagos tartózkodási ideje. Az osztrák, német és lengyel turisták átlagos tartózkodási ideje 2%-kal emelkedett, de a különböző nemzetiségű vendégek megoszlása az egyes években olyannyira megváltozott, hogy ez összességében az átlagos tartózkodási idők mintegy 11,76%-os visszaesését okozta.

a) a főátlag index értéke 110,0
b) a részátlag index értéke 102,0
c) az összetételhatás-index értéke 111,76

Két kiemelt üdülőkörzet adatait megnézve azt láthatjuk, hogy a tavalyi évhez képest 10%-kal esett vissza a vendégek átlagos tartózkodási ideje. Az osztrák, német és lengyel turisták átlagos tartózkodási ideje 2%-kal emelkedett, de a különböző nemzetiségű vendégek megoszlása az egyes években olyannyira megváltozott, hogy ez összességében az átlagos tartózkodási idők mintegy 11,76%-os visszaesését okozta.

a) a főátlag index értéke 90,0
b) a részátlag index értéke 98,0
c) az összetételhatás-index értéke 111,76

Két kiemelt üdülőkörzet adatait megnézve azt láthatjuk, hogy a tavalyi évhez képest 10%-kal esett vissza a vendégek átlagos tartózkodási ideje. Az osztrák, német és lengyel turisták átlagos tartózkodási ideje 2%-kal emelkedett, de a különböző nemzetiségű vendégek megoszlása az egyes években olyannyira megváltozott, hogy ez összességében az átlagos tartózkodási idők mintegy 11,76%-os visszaesését okozta.

a) a főátlag index értéke 110,0
b) a részátlag index értéke 98,0
c) az összetételhatás-index értéke 88,24

☸ Döntse el az alábbi állításokról, hogy igazak (I) vagy hamisak (H)!

Az időbeli összehasonlításra alkalmas indexszámítás során a viszonyítás iránya szabadon megválasztható.	**H**
A főátlag index a részátlag és az összetételhatás index szorzata.	**I**
A részátlagok különbözőségének hatását úgy tudjuk kimutatni, hogy változatlannak tekintjük a fősokaság részsokaságok szerinti összetételét.	**I**
Ha az összehasonlított két sokaság összetétele megegyezik, akkor a részsokaságok szerinti összetétel különbözőségének hatása 1.	**H**
Ha két időszak között a részátlagok értéke nem változik, akkor a főátlag index értéke megegyezik a részátlagindex értékével.	**H**

7. ÉRTÉKEN ALAPULÓ INDEXSZÁMÍTÁS

7.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A gazdasági életben a viszonyszámok változatos formában jelenhetnek meg. Az indexek többnyire dinamikus viszonyszámok, melyeket egy adott sokaságra vonatkozóan a későbbi és korábbi adat hányadosaként képezhetünk. Területi összehasonlító viszonyszám mintájára képezhetők területi indexek is. Az értéken alapuló indexszámítás módszere az ár, a mennyiség és a kettő szorzataként meghatározható érték időbeli és területi összehasonlítására alkalmazható. Az indexek kiszámításával relatív formában számszerűsíthető az árban és a mennyiségben bekövetkező változás, továbbá az érték fogalmának statisztikai bevezetésével lehetővé válik az is, hogy az időbeli és a területi összehasonlítást heterogén termékcsoport esetén is elvégezzük. Az érték kategória a gazdasági számításokban kiemelten fontos, mert segítségével összegezhetővé válnak az eltérő mértékegységben, illetve egységárban kifejezett adatok is.

A lecke célja, hogy megismertesse a hallgatókat az értéken alapuló indexszámítás módszerével, melynek középpontjában az ár, a mennyiség és az érték időbeli és területi összehasonlító vizsgálata áll. A hallgató tudja értelmezni és meghatározni az érték fogalmát, s ismerje fel a megjelenési formáit a gyakorlatban. Az egyedi és együttes indexeket legyen képes értelmezni, ismerje a köztük lévő összefüggéseket és sajátítsa el a számítások logikáját.

7.2 TANANYAG

Az értéken alapuló indexszámítás módszerét ár, mennyiség (volumen) és a kettő szorzataként meghatározható érték adatok időbeli és területi összehasonlítására alkalmazhatjuk.

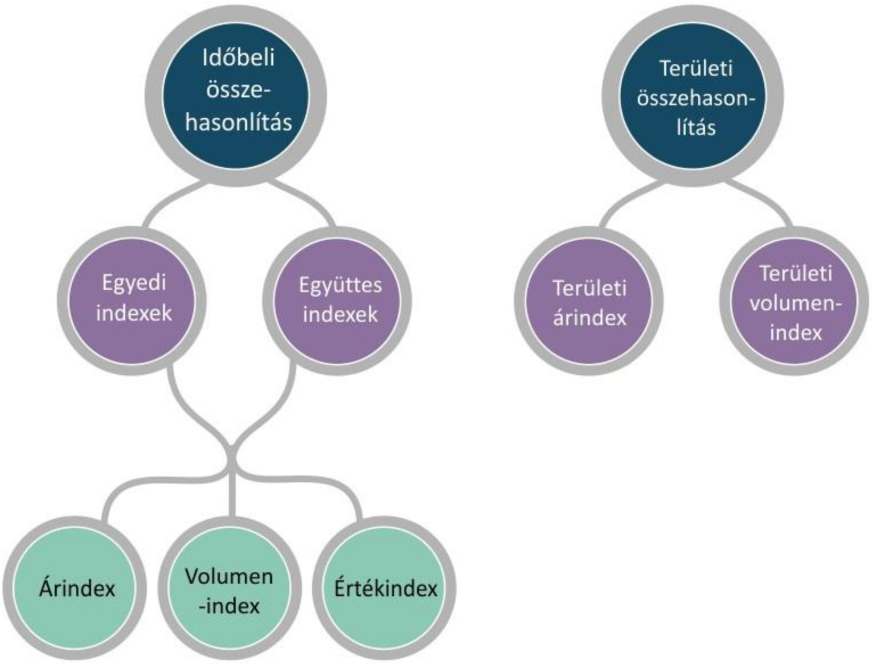
Az ár, mennyiség és érték adatok időbeli vagy területi összehasonlításának eredményeként kapott viszonyszámokat indexszámnak nevezzük.

Az összehasonlítás történhet egy termék esetén, illetve olyan heterogén termékcsoportoknál, melyek különböző fajtájú, leggyakrabban eltérő mértékegységű termékeket tartalmaznak, így közvetlenül nem összegezhetők. Időbeli összehasonlításnál két időszak között az értékben, árakban és mennyiségekben bekövetkező változásokat számszerűsítjük. Területi összevetésnél pedig az árakat és mennyiségeket két földrajzi egység között hasonlíthatjuk össze.

A képletek és számítási formák megértéséhez az alábbi alapösszefüggésre és angol kifejezésekre van szükségünk: az érték (**v**alue) valamely jószág egységárának (**p**rice) és mennyiségének, azaz volumenének (**q**uantity) a szorzataként határozható meg.

$$\text{érték} = \text{ár} \cdot \text{mennyiség} \rightarrow \mathbf{v} = \mathbf{p} \cdot \mathbf{q}$$

Az érték kategóriába tartozik például az árbevétel, a forgalom, a termelési érték vagy az értékesítés.



18. ábra: Az értéken alapuló indexszámítás összefüggésrendszere

7.2.1 Időbeli összehasonlítás

Az értéken alapuló indexszámítás módszerét leggyakrabban időbeli összehasonlításra használjuk. A számítások során a korábbi időszakot (bázis) 0 alsó index-el, a későbbi időszak (tárgy) adatait 1 alsó index-el jelöljük. Egy egyszerű példa segítségével a módszer lényege könnyen megragadható.

Egy bolti vásárlás során különböző termékeket vásárolunk, pl. tejet, kenyeret, gyümölcsöt stb. Feltételezve, hogy ugyanazokat a termékeket vesszük meg, két különböző hónapban össze tudjuk hasonlítani az egyes termékek árát, megvásárolt mennyiségét, illetve azt az összeget, melyet a termékek megvásárlására fordítot-

tunk. Számszerűsíteni tudjuk tehát egy-egy termék esetén a változásokat, illetve azt is, hogy a teljes fogyasztói kosárban miként alakult a két időszak között az elköltött összeg, mennyivel, azaz hány százalékkal emelkedtek vagy csökkentek az árak és a mennyiségek külön-külön és együttesen.

Az ár-, volumen- és értékváltozás számszerűsítésére alkalmas indexeket megkülönböztetjük attól függően, hogy a termékek mely körére vonatkoztatnak, így egyedi és együttes (aggregát) indexeket számíthatunk.

Egy termék (vagy homogén termékcsoport) esetén a mennyiség, ár és érték változását egyedi indexek segítségével számszerűsítjük. Több termék, heterogén termékcsoport esetén a mennyiség, az ár és az érték változását együttes (aggregát) indexek kiszámításával fejezhetjük ki.

Egyedi indexek

Egy adott jószág esetén az ár, a mennyiség és az érték változása egyedi index kiszámításával fejezhető ki.

Az egyedi árindex egy termék árváltozását fejezi ki két időszak között.

$$\textbf{egyedi árindex} = \frac{\textbf{későbbi időszak ár adata}}{\textbf{korábbi időszak ár adata}} \rightarrow \boldsymbol{i_p} = \frac{\boldsymbol{p_1}}{\boldsymbol{p_0}}$$

Az egyedi volumenindex egy termék mennyiségének változását fejezi ki két időszak között.

$$\textbf{egyedi volumenindex} = \frac{\textbf{későbbi időszak mennyiség adata}}{\textbf{korábbi időszak mennyiség adata}} \rightarrow \boldsymbol{i_q} = \frac{\boldsymbol{q_1}}{\boldsymbol{q_0}}$$

Az egyedi értékindex egy termék értékének változását fejezi ki két időszak között.

$$\textbf{egyedi értékindex} = \frac{\textbf{későbbi időszak érték adata}}{\textbf{korábbi időszak érték adata}} \rightarrow \boldsymbol{i_v} = \frac{\boldsymbol{v_1}}{\boldsymbol{v_0}}$$

Összefüggés az egyedi indexek között

Az egyedi értékindex, az egyedi árindex és az egyedi volumenindex szorzata. Ezen összefüggés alapján bármely két egyedi index értékét ismerve kifejezhető a hiányzó harmadik.

$$\mathbf{i_v} = \mathbf{i_p} \cdot \mathbf{i_q} \rightarrow \mathrm{i_p} = \frac{\mathrm{i_v}}{\mathrm{i_q}} \text{ vagy } \mathrm{i_q} = \frac{\mathrm{i_v}}{\mathrm{i_p}}$$

❆ Egy család március hónapban 20 kg kenyeret és 40 l tejet vásárolt, míg áprilisban a kenyérből 22 kg-ot, tejből pedig 38 litert. Márciusban 1 kg kenyér ára 300 Ft, a tej literenkénti ára 200 Ft volt, míg áprilisban a kenyér ára 320 Ft-ra nőtt, míg a tej ára 180 Ft-ra csökkent.

Foglalja táblába az adatokat és határozza meg, hogy mennyit fordított a család az egyes hónapokban kenyér és tej vásárlására együttesen és külön-külön!

	MÁRCIUS			*ÁPRILIS*		
	Elfogyasztott mennyiség	*Egységár*	*A termékek megvásárlására fordított összeg ÉRTÉK*	*Elfogyasztott mennyiség*	*Egységár*	*A termékek megvásárlására fordított összeg ÉRTÉK*
	q_0	p_0	$v_0 = q_0 \cdot p_0$	q_1	p_1	$v_1 = q_1 \cdot p_1$
Kenyér (kg)	20	300	$20 \cdot 300 =$ **6000**	22	320	$22 \cdot 320 =$ **7040**
Tej (l)	40	200	$40 \cdot 200 =$ **8000**	38	180	$38 \cdot 180 =$ **6840**
Együtt			**14 000**			**13 880**

19. ábra: A termékek megvásárlására fordított összeg az egyes hónapokban (09_07_K02)

A kiszámított értékek alapján az alábbiak fogalmazhatók meg:

- márciusban a család 6000 Ft-ot költött kenyér és 8000 Ft-ot tej megvásárlására, így összesen a két termékre együtt 14 000 Ft-ot
- áprilisban a család 7040 Ft-ot költött kenyér és 6840 Ft-ot tej megvásárlására, így összesen a két termékre együtt 13 880 Ft-ot

A tábla értékei alapján látszik, hogy a család áprilisban kenyérből többet fogyasztott, az ára is emelkedett, így a kenyér megvásárlására magasabb összeget kellett fordítania. Ezzel szemben a tejből elfogyasztott mennyiség és az ár is csökkent, ennek köszönhetően a megvásárlásra fordított összeg is alacsonyabb. Összességében a két termék együttes megvásárlására áprilisban a család kevesebbet költött, mint márciusban.

Hogyan változott az egyes termékek elfogyasztott mennyisége, egységára és a termékek megvásárlására fordított összeg áprilisban márciushoz képest?

20. *Az elfogyasztott mennyiség, az egységárak és a termékek megvásárlására fordított összeg változása márciusról áprilisra*

	Elfogyasztott mennyiség változása	Egységár változása	A termékek megvásárlására fordított összeg változása
	$i_q = \frac{q_1}{q_0}$	$i_p = \frac{p_1}{p_0}$	$i_v = \frac{v_1}{v_0}$
Kenyér	$\frac{22}{20} = 1{,}1 \xrightarrow{\cdot 100} \mathbf{110{,}0}$	$\frac{320}{300} = 1{,}0667 \xrightarrow{\cdot 100} \mathbf{106{,}67}$	$\frac{7040}{6000} = 1{,}1733 \xrightarrow{\cdot 100} \mathbf{117{,}33}$
Tej	$\frac{38}{40} = 0{,}95 \xrightarrow{\cdot 100} \mathbf{95{,}0}$	$\frac{180}{200} = 0{,}9 \xrightarrow{\cdot 100} \mathbf{90{,}0}$	$\frac{6840}{8000} = 0{,}855 \xrightarrow{\cdot 100} \mathbf{85{,}5}$

A kiszámított indexeket az alábbiak szerint értelmezhetők:

- a család áprilisban 10%-kal több kenyeret és 5%-kal kevesebb tejet fogyasztott el, mint márciusban
- áprilisban a kenyér ára 6,67%-kal magasabb, míg a tej ára 10%-kal alacsonyabb, mint márciusban
- a család áprilisban 17,33%-kal többet költött a kenyér, és 14,5%-kal kevesebbet a tej vásárlására, mint márciusban

Együttes indexek

Heterogén jószágcsoport, azaz különböző - leggyakrabban eltérő mértékegységű - termékek esetén az árban, mennyiségben és értékben bekövetkező időbeli változást együttes (aggregát) indexek kiszámításával fejezhetjük ki.

Több termék esetén az értékben bekövetkező együttes átlagos változást az együttes (aggregát) értékindex fejezi ki.

$$\mathrm{I_v} = \frac{\sum \mathrm{v_1}}{\sum \mathrm{v_0}} = \frac{\sum \mathrm{p_1 \cdot q_1}}{\sum \mathrm{p_0 \cdot q_0}}$$

Hogyan változott a két termék együttes megvásárlására fordított összeg a két időszak között!

$$I_v = \frac{\sum v_1}{\sum v_0} = \frac{13\,880}{14\,000} = 0{,}9914 \xrightarrow{\cdot 100} \mathbf{99{,}14} \rightarrow -0{,}86\%$$

A család 120 Ft-tal (13880-14000), azaz 0,86%-kal kevesebbet költött (99,14-100) a két termék együttes megvásárlására áprilisban.

Az értékben bekövetkező változást két tényező okozza:

- az árak együttes átlagos változása, melyet az együttes (aggregát) árindex fejez ki
- az elfogyasztott mennyiség együttes átlagos változása, melyet az együttes (aggregát) volumenindex fejez ki

Ha az árakban bekövetkező együttes változást kívánjuk számszerűsíteni, akkor a termékek mennyiségét változatlannak kell tekinteni. A mennyiségben bekövetkező együttes változást pedig a termékek árainak változatlansága mellett tudjuk számszerűsíteni. Ezekben az esetekben a változatlan tényezőt súlynak nevezzük, mely lehet a bázis és a tárgyidőszak adata is. A különböző súlyozás eltéréseinek kiküszöbölésére a leggyakrabban használatos keresztezett formula, a kétféle súlyozású index mértani átlagaként kifejezhető Fisher-féle index.

A bázissúlyozású ár- és volumenindexeket Laspeyres, a tárgyi súlyozású indexeket Paasche indexnek nevezzük. A bázis és tárgyi súlyozású index mérteni átlaga a Fisher index.

***20. ábra:** Az együttes ár- és volumenindexek kiszámítási formái*

A számítás során két valóságban nem létező, fiktív aggregátumot kell kiszámítani: $q_0 \cdot p_1$esetében a korábbi időszak mennyiségét a tárgyidőszak árával szorozzuk, míg $q_1 \cdot p_0$ meghatározásánál a tárgyidőszak mennyiségét szorozzuk a korábbi időszak ár adatával.

	q_0	p_0	$v_0 = q_0 \cdot p_0$	q_1	p_1	$v_1 = q_1 \cdot p_1$	$q_0 \cdot p_1$	$q_1 \cdot p_0$
Kenyér (kg)	20	300	$20 \cdot 300 =$ **6000**	22	320	$22 \cdot 320 =$ **7040**	$20 \cdot 320 =$ **6400**	$22 \cdot 300 =$ **6600**
Tej (l)	40	200	$40 \cdot 200 =$ **8000**	38	180	$38 \cdot 180 =$ **6840**	$40 \cdot 180 =$ **7200**	$38 \cdot 200 =$ **7600**
Együtt			**14 000**			**13 880**	**13 600**	**14 200**

21. ábra: Példa a fiktív aggregátumok kiszámítására

Számszerűsítse, hogy együttesen hogyan változott a kenyér és tej elfogyasztott mennyisége és ára!

Az elfogyasztott mennyiség együttes változása (együttes volumenindexek)

$$I_q^0 = \frac{\sum \boldsymbol{q_1} \cdot p_0}{\sum \boldsymbol{q_0} \cdot p_0} = \frac{14\,200}{14\,000} = 1{,}0143 \xrightarrow{\cdot 100} \mathbf{101{,}43} \rightarrow +1{,}43\%$$

$$I_q^1 = \frac{\sum \boldsymbol{q_1} \cdot p_1}{\sum \boldsymbol{q_0} \cdot p_1} = \frac{13\,880}{13\,600} = 1{,}0206 \xrightarrow{\cdot 100} \mathbf{102{,}06} \rightarrow +2{,}06\%$$

$$I_q^F = \sqrt{I_q^0 \cdot I_q^1} = \sqrt{1{,}0143 \cdot 1{,}0206} = 1{,}0175 \xrightarrow{\cdot 100} \mathbf{101{,}75} \rightarrow +1{,}75\%$$

Márciusról áprilisra a termékekből elfogyasztott mennyiség együttesen átlagosan 1,43%-kal nőtt bázisidőszaki árakkal súlyozva, 2,06%-kal nőtt tárgyidőszaki árakkal súlyozva, 1,75%-kal nőtt mindkét időszak áraival súlyozva.

Az egységárak együttes változása (együttes árindexek)

$$I_p^0 = \frac{\sum q_0 \cdot \boldsymbol{p_1}}{\sum q_0 \cdot \boldsymbol{p_0}} = \frac{13\,600}{14\,000} = 0{,}9714 \xrightarrow{\cdot 100} \mathbf{97{,}14} \rightarrow -2{,}86\%$$

$$I_p^1 = \frac{\sum q_1 \cdot \boldsymbol{p_1}}{\sum q_1 \cdot \boldsymbol{p_0}} = \frac{13\,880}{14\,200} = 0{,}9775 \xrightarrow{\cdot 100} \mathbf{97{,}75} \rightarrow -2{,}25\%$$

$$I_p^F = \sqrt{I_p^0 \cdot I_p^1} = \sqrt{0{,}9714 \cdot 0{,}9775} = 0{,}9745 \xrightarrow{\cdot 100} \mathbf{97{,}45} \rightarrow -2{,}55\%$$

Márciusról áprilisra a termékek egységára együttesen átlagosan 2,86%-kal csökkent bázisidőszaki volumenekkel súlyozva, 2,25%-

kal csökkent tárgyidőszaki volumenekkel súlyozva, 2,55%-kal csökkent mindkét időszak volumeneivel súlyozva.

A kétféle súlyozás közötti választást elsődlegesen a rendelkezésre álló adatok alapján dönthetjük el, amennyiben mindkét index kiszámítható, ajánlatos a keresztezett formula alkalmazása is.

Összefüggések az együttes (aggregát) indexek között

Az együttes indexek esetében, az együttes értékindex értéke megegyezik az ellentétes súlyozású vagy Fisher-féle együttes volumen- és árindex szorzatával.

$$I_v = I_q^0 \cdot I_p^1 = I_q^1 \cdot I_p^0 = I_q^F \cdot I_p^F$$

7.2.2 Területi összehasonlítás

A területi összehasonlításnál a viszonyítás iránya tetszőlegesen megválasztható, alsó indexben jelöljük a viszonyítandó területet és a viszonyítási alapot. Területi összevetésnél az értékindexet nem értelmezzük, az ár és volumenindexeknek viszont nagy a jelentősége.

- Egy városi piacra két faluból (az egyszerűség kedvéért A és B) érkeznek mezőgazdasági termékek. Ismert, hogy milyen mennyiségben érkeznek a termékek a településekről és az egységáruk is. *Hasonlítsa össze az egyes termékek árait és eladásra szánt mennyiségeit a két településen!*

	A falu		B falu		Mennyiségi összehasonlítás A falu B-hez képest	Árak összehasonlítása A falu B-hez képest
	Mennyiség (kg)	Egységár (Ft/kg)	Mennyiség (kg)	Egységár (Ft/kg)		
	q_A	p_A	q_B	p_B	$\frac{q_A}{q_B}$	$\frac{p_A}{p_B}$
Sárgarépa	12 000	140	8 000	150	$\frac{12000}{8000} = 1{,}5 \xrightarrow{\cdot 100} \mathbf{150,0}$	$\frac{140}{150} = 0{,}9333 \xrightarrow{\cdot 100} \mathbf{93,33}$
Hagyma	10 000	100	12 000	80	$\frac{10000}{12000} = 0{,}8333 \xrightarrow{\cdot 100} \mathbf{83,33}$	$\frac{100}{80} = 1{,}25 \xrightarrow{\cdot 100} \mathbf{125,0}$

22. ábra: A két település árainak és eladásra szánt mennyiségeinek összehasonlítása

A táblázat kiszámított értékei alapján a két település kapcsán az alábbi következtetések fogalmazhatók meg:

- *A település*ről 50%-kal több sárgarépát és 16,7%-kal kevesebb (83,3-100) hagymát visznek a piacra
- *A település*en a sárgarépa ára 6,67%-kal alacsonyabb (93,33-100), a hagyma ára pedig 25%-kal magasabb

A viszonyítás az ellentétes irányban is elvégezhető (B-t viszonyítjuk A-hoz), ez esetben viszont a relációt az ellenkező irányból vizsgáljuk, így a kapott eredmény nagyságrendileg eltérő a viszonyítási alap változásának köszönhetően.

Az összehasonlítás nemcsak egy-egy termékféleség esetén lehetséges, hanem többnél is, melyet együttes területi indexekkel végezhetünk el. A területi ár- és volumenindex kiszámításához az időbeli összehasonlításnál megismert módon alkotjuk meg a képleteket. Az index jelölésében a későbbi értelmezés miatt célszerű jelölni a viszonyítás irányát. Súlyként A és B település mennyiségei és árai is szerepelhetnek, a keresztezett formula ez esetben is alkalmazható. A *vizsgálat során elegendő a Fisher indexeket értelmezni*, mert azok információtartalma fontos igazán.

Hasonlítsuk össze a két termék együttes mennyiségét és egységárát a két településen úgy, hogy A-t viszonyítjuk B-hez!

Elsőként ki kell számítani az indexekhez szükséges aggregátumok értékeit.

	A falu		B falu					
	Mennyiség (kg)	*Egységár (Ft/kg)*	*Mennyiség (kg)*	*Egységár (Ft/kg)*				
	q_A	p_A	q_B	p_B	$q_A \cdot p_A$	$q_B \cdot p_B$	$q_A \cdot p_B$	$q_B \cdot p_A$
Sárgarépa	12 000	140	8 000	150	1 680 000	1 200 000	1 800 000	1 120 000
Hagyma	10 000	100	12 000	80	1 000 000	960 000	800 000	1 200 000
					2 680 000	*2 160 000*	*2 600 000*	*2 320 000*

23. ábra: A területi indexek kiszámításához szükséges aggregátumok értékei

Az együttes mennyiség összehasonlítása: területi volumenindex

$$I^{A}_{q\,(\frac{A}{B})} = \frac{\sum q_A \cdot p_A}{\sum q_B \cdot p_A} = \frac{2\,680\,000}{2\,320\,000} = 1{,}1552 \xrightarrow{\cdot 100} \mathbf{115,52}$$

$$I^{B}_{q\,(\frac{A}{B})} = \frac{\sum q_A \cdot p_B}{\sum q_B \cdot p_B} = \frac{2\,600\,000}{2\,160\,000} = 1{,}2037 \xrightarrow{\cdot 100} \mathbf{120,37}$$

$$I^{F}_{q\,(\frac{A}{B})} = \sqrt{I^{A}_{q\,(\frac{A}{B})} \cdot I^{B}_{q\,(\frac{A}{B})}} = \sqrt{1{,}1552 \cdot 1{,}2037} = 1{,}1792 \xrightarrow{\cdot 100} \mathbf{117,92}$$

*A falu*ban (a Fisher index alapján) *átlagosan* 17,92%-kal több terméket lehet megvásárolni.

Az egységárak összehasonlítása: területi árindex

$$I^{A}_{p\,(\frac{A}{B})} = \frac{\sum q_A \cdot p_A}{\sum q_A \cdot p_B} = \frac{2\,680\,000}{2\,600\,000} = 1{,}0308 \xrightarrow{\cdot 100} \mathbf{103,08}$$

$$I^{B}_{p\,(\frac{A}{B})} = \frac{\sum q_B \cdot p_A}{\sum q_B \cdot p_B} = \frac{2\,320\,000}{2\,160\,000} = 1{,}0741 \xrightarrow{\cdot 100} \mathbf{107,41}$$

$$I^{F}_{p(\frac{A}{B})} = \sqrt{I^{A}_{p\,(\frac{A}{B})} \cdot I^{B}_{p\,(\frac{A}{B})}} = \sqrt{1{,}0308 \cdot 1{,}0741} = 1,0522 \xrightarrow{\cdot 100} \mathbf{105,22}$$

A vizsgált termékeket *A falu*ban (a Fisher index alapján) átlagosan 5,22%-kal magasabb áron lehet megvásárolni.

7.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

7.3.1 Összefoglalás

Az értéken alapuló indexszámítás módszerét számos területen alkalmazzuk a gazdasági életben. A mindennapok során, például különböző időpontokban vagy településeken történő bevásárlás alkalmával, ha nem is tudatosan, de ezt a statisztikai metodikát alkalmazzuk.

Az indexszámítás során kulcsfontosságú az ár és mennyiség szorzataként meghatározható érték fogalma, melyet a különböző mértékegységű adatok összegzésére használunk. Egy-egy termék árának, mennyiségének és értékének összehasonlítására egyedi indexeket, heterogén termékcsoport esetén az együttes változás kimutatására együttes indexeket számíthatunk. A viszonyszámok mintájára számítható indexek időbeli és területi összehasonlításra is használhatók.

7.3.2 Önellenőrző kérdések

1. Statisztikailag hogyan határozható meg az érték?
2. Mi a különbség egyedi és együttes index között?
3. Mit mutat meg az egyedi árindex?
4. Mit mutat meg az egyedi volumenindex?
5. Mit mutat meg az egyedi értékindex?
6. Milyen összefüggés van az egyedi indexek között?
7. Mit mutat meg az együttes értékindex?
8. Miért számítható kétféleképpen az együttes ár- és volumenindex?
9. Miért használjuk, és hogyan számítjuk a Fisher-féle keresztezett formulát?
10. Milyen összefüggés van az együttes indexek között?

7.3.3 Gyakorló tesztek

☸ Válassza ki a helyes válaszokat!

Melyik nem tartozik statisztikailag az érték kategóriába?

- forgalom
- **eladási ár**
- árbevétel
- termelési érték

Ismert, hogy a bázissúlyozású árindex értéke 107,82, a láncsúlyozású árindex értéke 109,12. A Fisher index értéke:

- 117,65
- 106, 48
- **108,47**

Az egyedi árindex értéke 100, az egyedi volumenindex értéke 110. Az egyedi értékindex értéke:

- 110
- 90,1
- 100

Egy zöldséges valamennyi termékének együttes értékesítéséből befolyt bevétele tavalyhoz képest 10%-kal csökkent.

- az egyedi értékindex 110,0
- **az együttes értékindex értéke 90,0**
- az együttes volumenindex értéke 90,0

Egy család élelmiszerre fordított kiadásairól ismerjük, hogy a gyümölcsök árindexe 110, a zöldségeké 90. Ez alapján azt tudjuk, hogy

- **a gyümölcsök ára 10%-kal nőtt**
- a zöldségek ára 10%-kal nőtt
- a gyümölcsök és zöldségek árváltozása kiegyenlíti egymást, az együttes árindex 100

☸ Döntse el, hogy az alábbi állítások igazak (I) vagy hamisak (H)!

A Fisher-féle indexeket a bázis és tárgyi súlyozású indexek mértani átlagaként számítjuk ki.	**I**
Az egyedi árindex az egyedi volumenindex és az egyedi értékindex hányadosa.	**H**
Az együttes indexek esetében teljesül az az összefüggés, hogy az együttes értékindex az ellentétes súlyozású ár- és volumenindex szorzata.	**I**
Az értéket ár és mennyiség szorzataként kapjuk.	**I**
Az indexszámítás során a korábbi időszak adatait viszonyítjuk a későbbi időszak adataihoz.	**H**

8. AZ INDEXSZÁMÍTÁS GYAKORLATI ALKALMAZÁSA: A VÁLLALATI TELJESÍTMÉNY MÉRÉSE, NEMZETGAZDASÁGI INDIKÁTOROK, KÜLKERESKEDELMI STATISZTIKÁK

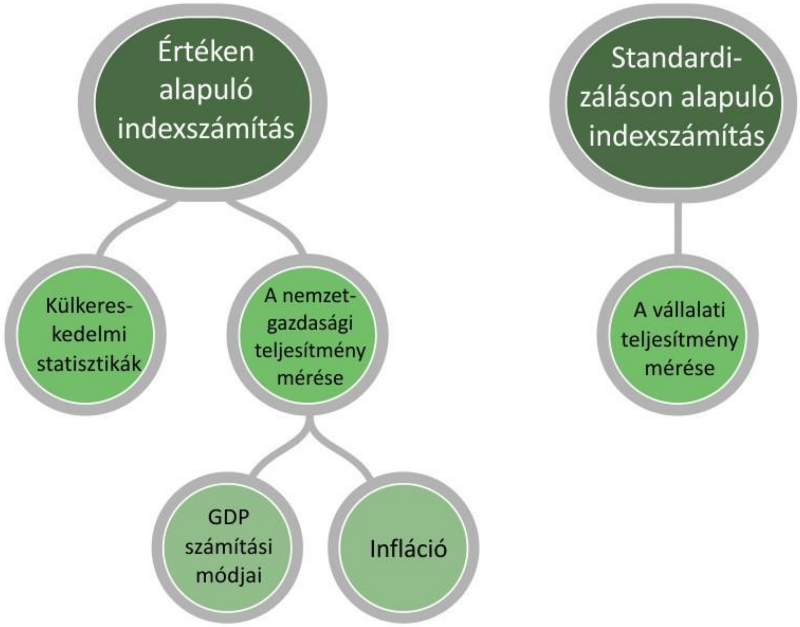
8.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A gazdasági elemzések során gyakran alkalmazunk indexeket. A standardizáláson alapuló indexszámítás főként a vállalati statisztikákban jelenik meg, míg az értéken alapuló indexszámítás módszerét alkalmazzuk a nemzetgazdasági teljesítmény számbavételénél és a külkereskedelmi statisztikákban is. A vállalatok időszakok közötti termelékenységének, fajlagos anyagfelhasználásának, önköltségének, vagy a dolgozók átlagbéreinek összehasonlítására alkalmas a standardizáláson alapuló indexszámítás módszere. A gazdasági teljesítmény, az árszínvonal és a külkereskedelmi forgalom változását az értéken alapuló indexszámítás módszerével számszerűsítjük.

A lecke célja, hogy bemutassa az indexszámítás gyakorlati megjelenési formáit. A hallgató ismerje az indexek fő alkalmazási területeit, sajátítsa el a fő nemzetgazdasági indikátorok számítási módjait. Képes legyen tájékozódni a külkereskedelmi statisztikák világában, valamint rendelkezzen a vállalati teljesítmény méréséhez szükséges módszertani alapokkal.

8.2 TANANYAG

Az indexek mind a mikro-, mind a makroszintű gazdasági elemzésekben gyakran megjelennek. A relatív változások számszerűsítése a vállalati gyakorlatban és nemzetgazdasági szinten is lényeges. A standardizáláson és az értéken alapuló indexszámítás módszerével együttesen és a tényezők szintjén is vizsgálható a vállalati és az országszintű teljesítmény.



24. ábra: Az indexek gyakorlati alkalmazási területei

8.2.1 A vállalati teljesítmény számszerűsítése indexek segítségével

A vállalati teljesítmény lényegében az indexszámítás mindkét formájával számszerűsíthető. A termelékenység, az önköltség és a fajlagos anyagfelhasználás vizsgálatánál a standardizálás, míg a vállalat árbevételének, forgalmának, termelési értékének vagy értékesítésének elemzésénél az érték-, ár- és volumenindex számításának módszerét alkalmazzuk.

A **termelékenység** statisztikailag az összesen megtermelt termékmennyiség és a termelésben részt vevő dolgozók számának a hányadosa. A vállalati szintű termelékenység - a standardizálás szempontjából a főátlag (összetett viszonyszám) - alakulását jelentősen befolyásolja a vállalaton belüli részegységek produktuma – részátlag -, az üzemi szintű termelékenység. A standardizálás módszerével kimutatható, hogy a vállalati szintű termelékenység időben hogyan változik, s adott időszakok közötti változásában mekkora rész tulajdonítható az üzemi szintű termelékenységben bekövetkezett változásoknak és mekkora a létszámbeli összetétel különbségeinek.

Az **önköltség** statisztikai szempontból[8] az összes költség és a megtermelt termékmennyiség hányadosa, mely azt mutatja meg, hogy a termelés egy egységére mekkora költséghányad jut. Különböző technológiával, megváltozott erőforrásokkal termelő vállalatnál a költségek változásának meghatározása különösen fontos, hiszen a vállalat az újításokkal, a modernizációval költségeinek csökkentésére törekszik.

A **fajlagos anyagfelhasználás** az összesen felhasznált anyag és megtermelt termékmennyiség hányadosa, mely azt mutatja meg, hogy egységnyi termék előállításához mennyi anyag felhasználására van szükség. A fajlagos anyagfelhasználás változásának iránya fontos indikátor a vállalatok számára, mely költségstruktúrájukat jelentősen befolyásolja. A standardizálás módszerének alkalmazásához a termékek vagy termékcsoportok szintjén is különbséget kell tenni a fajlagos anyagfelhasználásban – részátlag – melyek változása hatást gyakorol a vállalat egészére. Az összetétel hatás index a termelés szerkezetének változása.

A vállalati gyakorlatban a standardizáláson alapuló indexszámítás módszere alkalmazható az átlagkeresetek változásának számszerűsítésére is vagy az idegenforgalomban az átlagos tartózkodási idők elemzésénél is. A mezőgazdaságban a termelés és a betakarított terület hányadosaként kapott termésátlag változása szintén elemezhető ezzel a módszerrel.

8.2.2 A nemzetgazdasági teljesítmény számszerűsítése indexek segítségével

Az értéken alapuló indexszámítás módszerét alkalmazzuk a nemzetgazdasági teljesítmény méréséhez, amelynek segítségével számszerűsíthetők az árszínvonalban, a volumenben és az értékben bekövetkező változások. Egy ország makrogazdasági teljesítményének két legfontosabb indikátora, a GDP és az infláció.

A nemzetgazdasági teljesítmény mérésére a leggyakrabban használt mutatószám a Bruttó Hazai Termék, angol nevének (Gross Domestic Product) rövidítéséből a **GDP**.

A GDP adott évben, az adott ország területén végső felhasználási céllal megtermelt termékek és szolgáltatások pénzben kifejezett értéke.

A GDP az ország területén megtermelt valamennyi termék és szolgáltatás együttes értéke, amelyet a termékek árainak és mennyiségeinek szorzataként határozhatunk meg. A számítás módjától függően megkülönböztetünk nominál és reál GDP-t. A nominál vagy folyó áras GDP számításánál adott évi mennyiségeket adott évi árakkal szorzunk össze, míg a reál vagy változatlan áras, másképpen vásárlóerő-paritáson

[8] Az önköltség számításának számviteli megközelítése ennél sokrétűbb.

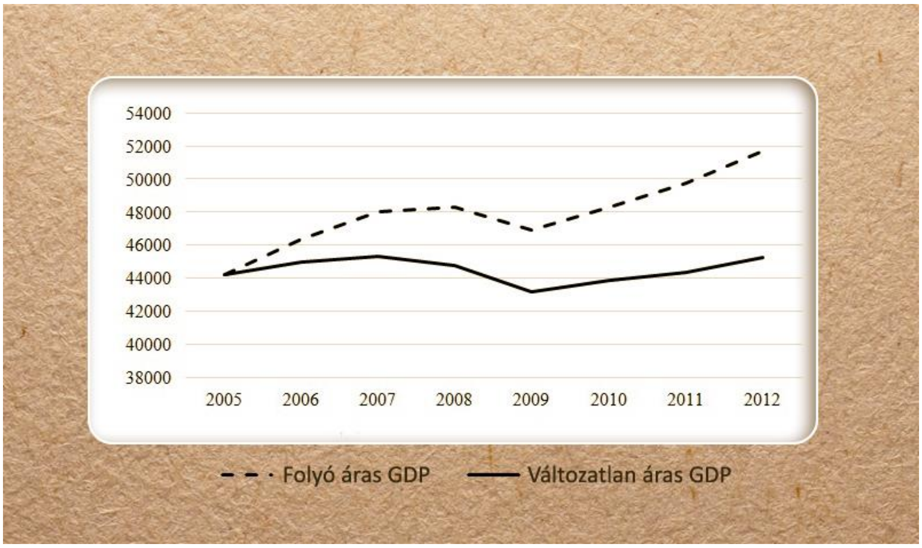
(Purchasing Power Parity, PPP) számított GDP esetén az adott évi mennyiségeket, egy, a korábbi évekre jellemző árral szorozzuk össze, kiküszöbölve így az árváltozásból fakadó hatásokat. Abban az évben, amelyre vonatkozóan az árakat választottuk a reál GDP számításához, a nominál és reál GDP értéke megegyezik.

A makrogazdasági teljesítmény változását indexekkel számszerűsíthetjük. A nominál GDP változása értékindex segítségével fejezhető ki, hiszen összetartozó, vagyis ugyanazon évre vonatkozó ár és mennyiség szorzataként meghatározható értékeket viszonyítunk egymáshoz. Két termék (A, B), két időszak (0,1) esetén az értékindex az alábbi formában számítható:

$$\text{Nominál GDP változása} = \frac{p_A^1 \cdot q_A^1 + p_B^1 \cdot q_B^1}{p_A^0 \cdot q_A^0 + p_B^0 \cdot q_B^0}$$

A nemzetközi összehasonlításokban gyakrabban használják a vásárlóerő-paritáson számított GDP-t, mert ezzel a mutatóval kiszűrhető az infláció hatása, így az index jobban tükrözi a teljesítmény változását. A reál GDP változása esetén a számlálóban és a nevezőben is ugyanaz az ár szerepel, vagyis csak a mennyiségek együttes változását mutatja, ezért a GDP volumenindexének is nevezik.

$$\text{Reál GDP változása} = \frac{p_A^0 \cdot q_A^1 + p_B^0 \cdot q_B^1}{p_A^0 \cdot q_A^0 + p_B^0 \cdot q_B^0}$$



25. ábra: Az USA folyó áras és változatlan (2005. évi) áron számított GDP adatainak alakulása 2005 és 2012 között

Forrás: OECD (2014)

A 25. ábrán az USA folyó áras és változatlan áras GDP adatainak alakulását összehasonlítva jól látszik a számítás módjából fakadó különbség. A változatlan áras GDP számításához a 2005. évi áradatokat vettük alapul, így ebben az évben a kétféle számítású adatsor értéke megegyezik. A folyó áras adatok magasabbak, mint a változatlan áron számítottak, mert ezek az értékek az árváltozás hatását is tartalmazzák. A GDP és az infláció kapcsolatáról bővebben olvashat az alábbi szakmai oldalon.

16. A GDP és az infláció kapcsolata:
http://www.investopedia.com/articles/06/gdpinflation.asp

8.2.3 Az infláció

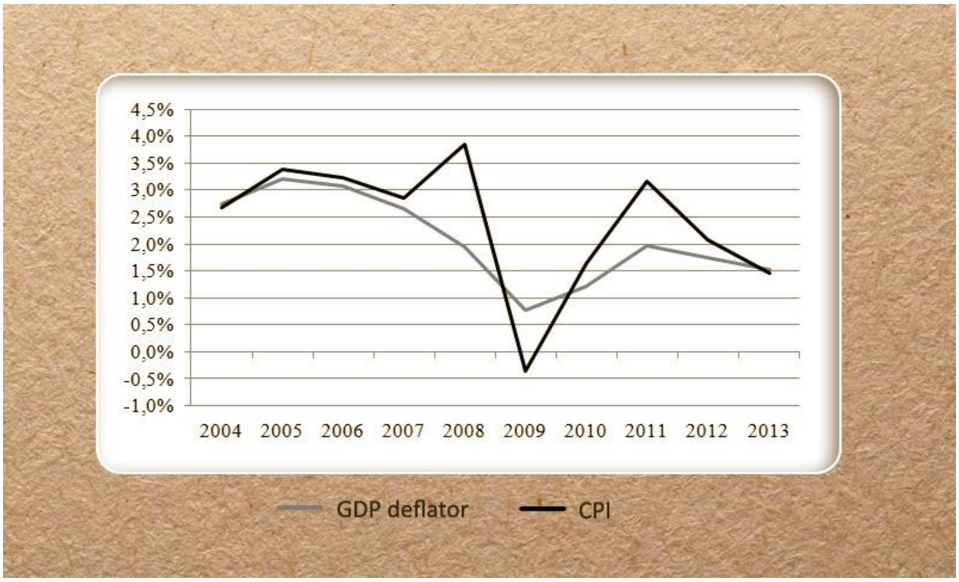
Az infláció a pénz vásárlóerejének csökkenése, az árak együttes változását jelző árindex, mely kétféle formában számítható. A GDP deflátor (GDPD) az együttes árváltozást úgy számszerűsíti, hogy a tárgyévi mennyiségeket tekinti változatlannak. Az infláció számításának elterjedtebb formája a fogyasztói árindex (Consumer Price Index, CPI), amely az előző időszak mennyiségeinek változatlansága mellett méri az árak együttes átlagos változását. A GDP deflátor és a fogyasztói árindex számítási sajátosságairól hallgassa meg az alábbi **hang**anyagot.

Két termék esetén a GDP deflátor az alábbi formában számítható ki:

$$\mathrm{GDPD} = \frac{p_A^1 \cdot q_A^1 + p_B^1 \cdot q_B^1}{p_A^0 \cdot q_A^1 + p_B^0 \cdot q_B^1}$$

Két termék esetén a fogyasztói árindex az alábbi formában fejezhető ki:

$$\mathrm{CPI} = \frac{p_A^1 \cdot q_A^0 + p_B^1 \cdot q_B^0}{p_A^0 \cdot q_A^0 + p_B^0 \cdot q_B^0}$$



26. ábra: Az infláció mérésére alkalmas GDP deflátor és fogyasztói árindex összehasonlítása az USA 2004 és 2013 közötti adatain

Forrás: Világbank (2014)

Az infláció mérési módjainak különbsége jól kivehető a 26. ábrán. A fogyasztói árindex számítása sokkal elterjedtebb, a GDP deflátor kevésbé használt mutató. Ezt az indokolhatja, hogy az árak együttes változása jobban kimutatható akkor, ha a korábbi év mennyiségeit vesszük alapul, hiszen a tárgyévi mennyiségekben közvetett módon már érződik az időszakok közötti árváltozás hatása is. Az értéket képező árak és mennyiségek negatív összefüggésben vannak egymással, emelkedő árak csökkenő mennyiségeket idéznek elő. Ha két időszak között az árak emelkednek, akkor az érintett termékekből vásárolt mennyiségek várhatóan csökkeni fognak. Az árváltozás hatását a fogyasztói árindex úgy méri, hogy az előző, árváltozás előtti időszak mennyiségeit veszi alapul. Statisztikailag a GDP deflátor súlyozása is megfelelő, ám a gyakorlat a fogyasztói árindex logikájára könnyebben építkezik.

Az infláció és a gazdasági teljesítmény szoros összefüggésben van egymással. A számítás módjától függetlenül az árak emelkedése gyakoribb, így a pénzromlás, vagyis az infláció gazdaságokban való előfordulása valószínűbb. Említést kell tenni azonban az árak együttes változásában bekövetkező egyéb folyamatokról is, melyekről az alábbi szakmai oldalon olvashat bővebben:

17. Infláció, defláció, stagfláció: http://www.investopedia.com/exam-guide/cfp/economics-time-value/cfp6.asp

Az együttes indexek számítását a GDP-nél és az inflációnál is látható módon jelentősen befolyásolja a súlyok megválasztása. A súlyozás torzító hatásainak elkerülése érdekében vezették be a Fisher formula használatát, mely a súlyozásbeli különbségeket úgy küszöböli ki, hogy meghatározza az eltérő súlyozású indexek mértani átlagát.

8.2.4 Külkereskedelmi statisztikák

A külkereskedelemben statisztikák készítésénél kiemelten fontos az értéken alapuló indexszámítás. A nemzetközi kapcsolatok intenzitásáról és sajátosságairól kapunk képet a külkereskedelmi termékforgalom, a külkereskedelmi árak és volumen változásának számszerűsítésével. A nemzetközi kereskedelemhez kapcsolódó információkról, statisztikákról a Kereskedelmi Világszervezet (World Trade Organization, WTO) honlapján lehet tájékozódni.

18. World Trade Organization: http://www.wto.org/english/res_e/statis_e/statis_e.htm

A külkereskedelem fontos mérőszáma a **cserearány mutató**, amely az exportált és importált termékek árindexének hányadosa. Az index azt mutatja meg, hogy egységnyi importért a beszámolási időszakban hány százalékkal több vagy kevesebb exportárut kellett értékesíteni, mint a bázisidőszakban[9]. Ha a mutató értéke 1-nél nagyobb akkor a cserearány javul, mert az export volumene nő, míg ha 1-nél kisebb, akkor a cserearány romlik.

A külkereskedelem területén a forgalmi és árstatisztikák készítésének a módszertana az értéken alapuló indexszámítás alapjaira épül, mégis nemzetközi vonatkozásai miatt egyedi. Egy szakértővel a külkereskedelmi árstatisztikák sajátosságairól készített interjút mutatja be az alábbi **videó**.

8.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

8.3.1 Összefoglalás

A vállalati és nemzetgazdasági teljesítmény mérésénél gyakran alkalmazunk indexeket. A termelékenység, az önköltség, a fajlagos anyagfelhasználás és ezek alakulására ható tényezők változásánál a standardizálás módszerét alkalmazzuk. A makrogazdasági kibocsátás, valamint az

[9] Központi Statisztikai Hivatal (2014): Áralakulás. URL: https://www.ksh.hu/thm/1/indi1_2_3.html. Letöltve: 2014. augusztus 19.

árak együttes változásának számszerűsítésére az értéken alapuló indexszámítás alkalmas. A külkereskedelmi statisztikák szintén ezen módszer segítségével készülnek el.

8.3.2 Önellenőrző kérdések

1. Mely kiemelt területeken alkalmazhatók az indexek?
2. Említsen példát a standardizáláson alapuló indexszámítás alkalmazására!
3. Említsen példát az értéken alapuló indexszámítás alkalmazására!
4. Mi a GDP és milyen típusait különböztetjük meg?
5. Mi a vásárlóerő-paritáson számított GDP?
6. Hogyan mérhető a nominál GDP változása?
7. Hogyan mérhető a reál GDP változása?
8. Mi az infláció?
9. Hogyan mérhető az infláció?
10. Mi a cserearány mutató?

8.3.3 Gyakorló tesztek

☸ Válassza ki a helyes megoldást!

A termelékenység megmutatja

- az egy termékegységre jutó dolgozók számát
- **az egy dolgozóra jutó termékek számát**
- az átlagosan megtermelt termékmennyiséget

A GDP értékindexe

- **a nominál GDP változása**
- a reál GDP változása
- a nominál és reál GDP hányadosa

Az infláció mérésére alkalmazott fogyasztói árindex számítása során súlyként használjuk

- a tárgyévi mennyiségeket
- **az előző évi mennyiségeket**
- az előző évi árakat

A cserearány mutató

- **az exportált és importált termékek árindexének hányadosa**
- az importált és exportált termékek árindexének hányadosa

- az export és import hányadosa

A fajlagos anyagfelhasználás

- **az összesen felhasznált anyag és az előállított termékmennyiség hányadosa**
- az összesen felhasznált anyag és az előállított termékmennyiség szorzata
- az előállított termékmennyiség és az összesen felhasznált anyag hányadosa

A nominál GDP

- **adott évi mennyiségek és árak szorzata**
- adott évi árak és egy, a korábbi évekre jellemző mennyiség szorzata
- adott évi mennyiségek és egy, a korábbi évekre jellemző árak szorzata

A reál GDP

- adott évi mennyiségek és árak szorzata
- adott évi árak és egy, a korábbi évekre jellemző mennyiség szorzata
- **adott évi mennyiségek és egy, a korábbi évekre jellemző árak szorzata**

Az infláció mérésére alkalmazott GDP deflátor számítása során súlyként használjuk

- **a tárgyévi mennyiségeket**
- az előző évi mennyiségeket
- a tárgyévi árakat

A GDP volumenindexe

- a nominál GDP változása
- **a reál GDP változása**
- a nominál és reál GDP hányadosa

A külkereskedelmi árstatisztikákban használatos index

- **Fisher-féle**
- Pearson-féle
- Yule-féle

9. ISMÉRVEK KÖZÖTTI KAPCSOLAT VIZSGÁLATA I.: ASSZOCIÁCIÓ ÉS VEGYES KAPCSOLAT

9.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

Az elemző munka során gyakran előfordul, hogy egy-egy jelenség vizsgálatánál nem csupán az adott helyzet leírására törekszünk, hanem arra is kíváncsiak vagyunk, hogy milyen tényezőknek köszönhetően alakult ki ez az állapot. Statisztikailag egy adott jelenség különböző ismérvei, vagyis jellemzői közötti kapcsolatot kívánjuk feltárni. A kapcsolatvizsgálat típusait megkülönböztetjük aszerint, hogy milyen típusú ismérvekről van szó. Elsőként a két nem mennyiségi ismérv kapcsolatát jellemző asszociáció, valamint egy mennyiségi és egy nem mennyiségi ismérv közötti összefüggés számszerűsítésére alkalmas vegyes kapcsolat kerül bemutatásra.

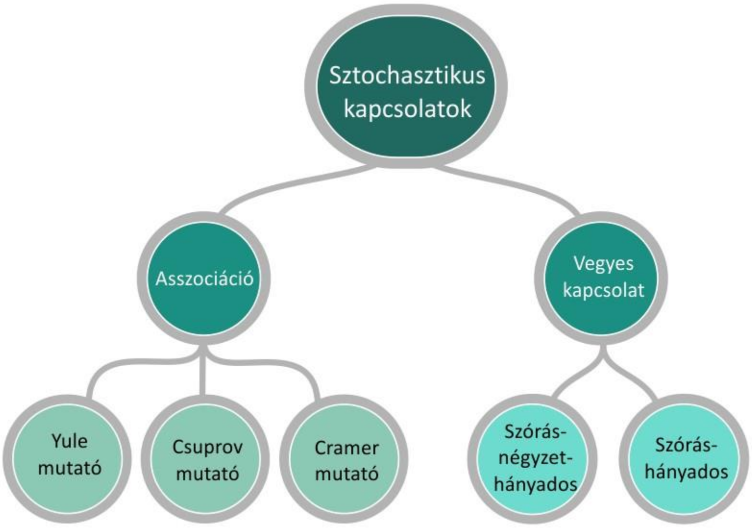
A lecke célja, hogy megismertesse a hallgatókat az ismérvek közötti kapcsolatok statisztikai vizsgálatának alapjaival, valamint részletesen az asszociáció és a vegyes kapcsolat elemzésének módszereivel. A hallgató legyen képes felismerni, hogy adott ismérvek kapcsán, mely kapcsolatvizsgálati technika alkalmazható, s képes legyen e kapcsolat számszerűsítésére is a megismert mutatókkal. A módszerek elsajátításának alapja, hogy a kapott eredményeket a hallgató helyesen tudja értelmezni.

9.2 TANANYAG

A kapcsolatvizsgálat két szélső esete a függetlenség és a függvényszerű kapcsolat. Függetlenség esetén nincs kimutatható összefüggés, míg függvényszerű kapcsolat esetén, az egyik ismérvből egyértelműen lehet következtetni a másikra, vagyis adott ismérvhez tartozó ismérvváltozat minden esetben egy másik ismérv adott ismérvváltozatával együtt fordul elő. A két szélső eset között sztochasztikus kapcsolat van az ismérvek között.

A két vagy több ismérv közötti valószínűségi összefüggést sztochasztikus kapcsolatnak nevezzük. Az egyik ismérv ismérvváltozatából tendenciaszerűen következtethetünk egy másik ismérvhez tartozó ismérvváltozatra.

Az ismérvek közötti kapcsolat feltárásához mutatószámokat alkalmazunk. A mutatószámok értéke minden esetben, abszolút értékben 0 és 1 közötti intervallumban helyezkedik el, ahol a 0 függetlenséget, az 1 pedig függvényszerű kapcsolatot jelöl. A mutatók 0 és 1 közötti értéke sztochasztikus kapcsolat létét jelzi. A mutatószám 0-hoz közeli értéke gyenge, 1-hez közeli értéke pedig erős kapcsolatot fejezi ki.



27. ábra: A kapcsolatvizsgálat alapesetei

Az ismérvek közötti kapcsolat feltárásárának módszere attól függ, hogy milyen típusú ismérveket vizsgálunk:

- ***asszociáció***: két nem mennyiségi ismérv (minőségi vagy területi) közötti kapcsolat – pl. a vizsga sikeressége (sikeres vagy sikertelen) és az előadások látogatása (járt vagy nem járt)
- ***vegyes kapcsolat***: egy nem mennyiségi (minőségi vagy területi) és egy mennyiségi ismérv közötti kapcsolat – pl. a feladatsor típusa (A vagy B) és az elért pontszám

28. ábra: Az asszociáció és a vegyes kapcsolat mérőszámai

9.2.1 Asszociáció

Két nem mennyiségi ismérv közötti kapcsolat vizsgálata esetén a kapcsolat létének kimutatására és erősségének számszerűsítésére van lehetőség. Az asszociáció vizsgálatához megoszlási és koordinációs viszonyszámok is alkalmazhatók, de az összefüggés mérésére leggyakrabban asszociációs együtthatókat – Yule, Csuprov és Cramer – használunk.

Yule mutató

A Yule mutató csak alternatív, azaz két ismérvváltozattal rendelkező ismérveknél számítható. A mutató előjele lehet pozitív és negatív, mely jelzi az ismérvváltozatok összetartozását.

$$\mathrm{Y} = \frac{\mathrm{f}_{11} \cdot \mathrm{f}_{00} - \mathrm{f}_{10} \cdot \mathrm{f}_{01}}{\mathrm{f}_{11} \cdot \mathrm{f}_{00} + \mathrm{f}_{10} \cdot \mathrm{f}_{01}}$$

Tulajdonságai: $0 \leq |Y| \leq 1$

Ha a mutató értéke 0, akkor a két ismérv egymástól független, míg ha az értéke 1, akkor függvényszerű a köztük lévő kapcsolat. A két szélső érték között beszélünk sztochasztikus kapcsolatról.

❆ *Vizsgálja meg a vizsga sikeressége és az előadásra járás közötti kapcsolatot Yule mutató segítségével!*

	Járt előadásra	Nem járt előadásra	Összesen
Sikeres	50	10	60
Sikertelen	15	25	40
Összesen	65	35	100

	Járt előadásra (1)	Nem járt előadásra (0)	Összesen
Sikeres (1)	f_{11}	f_{10}	$f_{1.}$
Sikertelen (0)	f_{01}	f_{00}	$f_{0.}$
Összesen	$f_{.1}$	$f_{.0}$	n

29. ábra: A csoport megoszlása a vizsga sikeressége és az előadásra járás gyakorisága alapján

$$\mathrm{Y} = \frac{\mathrm{f}_{11} \cdot \mathrm{f}_{00} - \mathrm{f}_{10} \cdot \mathrm{f}_{01}}{\mathrm{f}_{11} \cdot \mathrm{f}_{00} + \mathrm{f}_{10} \cdot \mathrm{f}_{01}} = \frac{50 \cdot 25 - 10 \cdot 15}{50 \cdot 25 + 10 \cdot 15} = \frac{1250 - 150}{1250 + 150} = \mathbf{0{,}7857}$$

A vizsga sikeressége és az előadásra járás között erős kapcsolat mutatható ki. A mutató előjele pozitív, vagyis azok a hallgatók, akik jártak előadásra nagy valószínűséggel sikeresen tették le a vizsgát, míg akik nem jártak előadásra, azok sikertelenül.

Csuprov és Cramer mutató

A több ismérvváltozattal rendelkező ismérvek esetén a Csuprov és a Cramer mutatót alkalmazzuk a nem mennyiségi ismérvek közötti kapcsolat kimutatására. Az egyik ismérv változatainak száma s, a másik ismérv változatainak száma t, melyet úgy választunk meg, hogy s < t. A mutatók kiszámításnál abból indulunk ki, hogy az ismérvek függetlensége esetén, a csoportokon belüli megoszlás arányos lenne. Peremgyakoriságnak az egyes ismérvek szerinti csoportok elemszámát nevezzük. Az f* a függetlenség esetén feltételezett gyakoriság értéke, melyet úgy kapunk, hogy a peremgyakoriságok szorzatát osztjuk az elemszámmal.

A Csuprov mutató számításához elsőként χ^2 (khí négyzet) kiszámítására van szükség, melyhez minden részsokaság esetén a tapasztalati és a függetlenség esetén feltételezett gyakoriság értékének különbségét négyzetre emeljük, majd az eredményt elosztjuk a hozzá tartozó függetlenség esetén feltételezett gyakoriság értékével, majd a részsokaságokra kapott értékeket összeadjuk.

$$\chi^2 = \sum \frac{(f - f^*)^2}{f^*}$$

A Csuprov mutató kiszámításához szükség van a χ^2 értékére, valamint az elemszámra és az ismérvekhez tartozó ismérvváltozatok számára is.

$$\boldsymbol{T} = \sqrt{\frac{\chi^2}{n \cdot \sqrt{s-1} \cdot \sqrt{t-1}}}$$

A Cramer mutató számítható a Csuprov mutató segítségével vagy a χ^2, az elemszám és a kevesebb ismérvváltozattal rendelkező ismérv ismérvváltozatainak számával is.

$$\boldsymbol{C} = \frac{T}{T_{max}} = \frac{T}{\sqrt[4]{\frac{s-1}{t-1}}} \qquad \boldsymbol{C} = \sqrt{\frac{\chi^2}{n \cdot (s-1)}}$$

☸ Egy felsőoktatási intézményben megvizsgálták a statisztika vizsga eredményessége és az órai aktivitás közötti összefüggést. Az ezzel kapcsolatos felmérésről a következő adatokat ismerjük:

21. *A statisztika vizsga eredményessége és az órai aktivitás megoszlása*

Statisztika vizsga eredményessége	Órán aktív	Órán kevésbé aktív	Nem járt	Összesen
Sikeres	47	30	7	84
Sikertelen	3	25	8	36
Összesen	50	55	15	120

A megfelelő mérőszám alkalmazásával vizsgálja meg, milyen erősségű sztochasztikus kapcsolat van a statisztika vizsga eredményessége és a hallgatók szakja között!

A kapcsolat erősségének meghatározásához Csuprov és Cramer mutatót kell számítani, mert az órai aktivitás nem alternatív ismérv. Első lépésként meg kell határozni a függetlenség esetén feltételezett gyakoriságok értékeit az összetartozó ismérvváltozatoknál, majd ki kell számítani χ^2 értékét.

Statisztika vizsga eredményessége	Órán aktív		Órán kevésbé aktív		Nem járt		Összesen
	f	f*	f	f*	f	f*	
Sikeres	47	$\frac{84 \cdot 50}{120} = \mathbf{35}$	30	$\frac{84 \cdot 55}{120} = \mathbf{38,5}$	7	$\frac{84 \cdot 15}{120} = \mathbf{10,5}$	84
Sikertelen	3	$\frac{36 \cdot 50}{120} = \mathbf{15}$	25	$\frac{36 \cdot 55}{120} = \mathbf{16,5}$	8	$\frac{84 \cdot 15}{120} = \mathbf{4,5}$	36
Összesen	50	50	55	55	15	15	120

30. ábra: A függetlenség esetén feltételezett gyakoriságok kiszámítása

A vizsga eredményessége és az órai aktivitás	Tapasztalt gyakoriság	Függetlenség esetén feltételezett gyakoriság	$(f-f^*)$	$(f-f^*)^2$	$\frac{(f-f^*)^2}{f^*}$
	f	f*			
Sikeres – órán aktív	47	35	12	144	4,11
Sikeres – órán kevésbé aktív	30	38,5	-8,5	72,25	1,88
Sikeres – nem járt	7	10,5	-3,5	12,25	1,17
Sikertelen – órán aktív	3	15	-12	144	9,60
Sikertelen – órán kevésbé aktív	25	16,5	8,5	72,25	4,38
Sikertelen – nem járt	8	4,5	3,5	12,25	2,72
Összesen	120	120			$\chi^2 = 23,86$

31. ábra: A χ^2 kiszámítása

$$T = \sqrt{\frac{\chi^2}{n \cdot \sqrt{s-1} \cdot \sqrt{t-1}}} = \sqrt{\frac{23,86}{120 \cdot \sqrt{2-1} \cdot \sqrt{3-1}}} = \mathbf{0,375}$$

$$C = \frac{T}{T_{max}} = \frac{0,375}{\sqrt[4]{\frac{2-1}{3-1}}} = \frac{0,375}{\sqrt[4]{\frac{1}{2}}} = \frac{0,375}{0,84} = \mathbf{0,4464}$$

A statisztika vizsga eredményessége és a hallgatók szakja között *közepesnél gyengébb kapcsolat* mutatható ki.

9.2.2 Vegyes kapcsolat

A vegyes kapcsolat egy mennyiségi és egy nem mennyiségi (területi vagy minőségi) ismérv közötti összefüggés. A kapcsolat vizsgálata során a két ismérv közötti kapcsolat szorosságát tudjuk mérni, valamint meg tudjuk határozni, hogy a nem mennyiségi ismérv hány százalékban határozza meg a mennyiségi ismérv alakulását, szóródását.

A kapcsolatvizsgálat első lépése, hogy a nem mennyiségi ismérv szerint csoportosítjuk a sokaságot. A továbblépéshez ismerni kell a teljes sokaság és a csoportok elemszámát is, vagy az elemszám csoportok közötti megoszlásának nagyságát. A mennyiségi ismérv esetén pedig kiszámítjuk a csoportok és a teljes sokaság átlagos értékét.

Az egyes csoportokhoz tartozó átlagot részátlagnak vagy csoportátlagnak, míg a teljes sokaságra számított átlagot főátlagnak nevezzük. A főátlag a csoportátlagok csoportok elemszámával súlyozott számtani átlagaként határozható meg.

A kapcsolat elemzése során az átlagoktól való eltérésekre helyezzük a hangsúlyt. Egy adott érték teljes sokaságra vetített átlagos értéktől való eltérése a teljes eltérés, melyet belső és külső eltérésre bonthatunk fel. A belső eltérés egy adott érték saját csoportján belüli átlagos értéktől való eltérése, míg a külső eltérés a csoportátlagok főátlagtól vett eltérése. Az átlagtól vett átlagos eltérést a szórás fejezi ki, a fentiekhez kapcsolódva megkülönböztetünk teljes, valamint belső és külső szórást.
A vegyes kapcsolat mutatószámainak kiszámításához varianciaanalízisre van szükség, melynek során meg kell határozni a belső, külső és teljes szórásnégyzet értékét.

Eltérés		Szórásnégyzet
$S_K = \sum n \cdot (\bar{x} - \bar{X})^2$	KÜLSŐ	$\sigma_K^2 = \frac{\sum n \cdot (\bar{x} - \bar{X})^2}{N} = \frac{S_K}{N}$
$S_B = \sum n \cdot \sigma^2$	BELSŐ	$\sigma_B^2 = \frac{\sum n \cdot \sigma^2}{N} = \frac{S_B}{N}$
$S = S_B + S_K$	TELJES	$\sigma^2 = \sigma_K^2 + \sigma_B^2 = \frac{S}{N}$

32. ábra: A belső, külső és teljes eltérés, valamint szórásnégyzetek kiszámítási formái (09_09_K06)

Szórásnégyzet-hányados – H^2 mutató

A szórásnégyzet-hányados azt mutatja meg, hogy a nem mennyiségi ismérv hány százalékban határozza meg a mennyiségi ismérv szóródását.

$$H^2 = \frac{\sigma_K^2}{\sigma^2} = \frac{\text{külső szórásnégyzet}}{\text{teljes szórásnégyzet}}$$

Szóráshányados – H mutató

A szóráshányados megmutatja a kapcsolat erősségét egy mennyiségi és egy nem mennyiségi ismérv között. A mutató a szórásnégyzet hányados négyzetgyöke.

$$H = \sqrt{H^2}$$

A vegyes kapcsolat mérőszámainak tulajdonságai

Sztochasztikus kapcsolat esetén a szórásnégyzet-hányados és a szóráshányados értéke is egy 0 és 1 közé eső szám. A két szélső eset közül, ha a mutatók értéke 0, az függetlenségre utal, míg ha az értékük 1, az függvényszerű kapcsolatot jelez.

☸ Egy felsőoktatási intézményben megvizsgálták a statisztika zárthelyi dolgozat eredménye (max. 100 pont) és az írt feladatsor típusa (A vagy B) közötti kapcsolatot. Ismert, hogy hány hallgató írta az A és mennyi a B feladatsort Az oktató kiszámította az egyes feladatsorokhoz tartozó átlageredményeket, valamint azt is, hogy az egyes feladatsorokon belül a hallgatók eredménye az átlagtól átlagosan mennyivel tér el. Az eredmények a következő táblában foglalhatók össze:

22. *A feladatsorokra vonatkozó információk*

Feladatsor típusa	Hallgatók száma	Átlagos eredmény (pont)	Az eredmények szórása (pont)
	n_i	$\bar{x}_i$	σ_i
A	80	57	9
B	60	82	6
Összesen	140		

A táblázat alapján az alábbiak fogalmazhatók meg:

- a 140 fős csoportban 80 hallgató írta az *A feladatsor*t, akik átlagosan 57 pontot értek el, ettől az egyes hallgatók eredménye átlagosan 9 ponttal tér el
- a 140 fős csoportban 60 hallgató írta az *B feladatsor*t, akik átlagosan 82 pontot értek el, ettől az egyes hallgatók eredménye átlagosan 6 ponttal tér el

Jellemezze a feladatsor típusa és a dolgozat eredménye közötti kapcsolatot!

A kapcsolat jellemzéséhez elsőként *a főátlag meghatározására* van szükség, mely a feladatsorok átlageredményeinek hallgatók számával súlyozott számtani átlaga:

$$\overline{X} = \frac{80 \cdot 57 + 60 \cdot 82}{120} = 79$$

A 120 fős csoportban a hallgatók átlagosan 79 pontot értek el.

23. *A csoport átlagos eredményei feladatsoronként és együttesen*

Feladatsor típusa	Hallgatók száma	Átlagos eredmény (pont)	Az eredmények szórása (pont)
	n_i	$\overline{x}_i$	σ_i
A	80	57	9
B	60	82	6
Összesen	140	**79**	

A kapcsolatvizsgálat következő lépése a külső, belső és teljes szórásnégyzet meghatározása.

Külső szórásnégyzet

$$\sigma_K^2 = \frac{\sum n \cdot (\overline{x} - \overline{X})^2}{n} = \frac{80 \cdot (57 - 79)^2 + 60 \cdot (82 - 79)^2}{120} = \mathbf{327,17}$$

$$\rightarrow \sigma_K = \sqrt{327{,}17} = 18{,}09 \sim 18 \text{ pont}$$

A feladatsorokhoz tartozó átlageredmények a teljes csoportra vonatkoztatott átlageredménytől átlagosan 18 ponttal térnek el.

Belső szórásnégyzet

$$\sigma_B^2 = \frac{\sum n \cdot \sigma^2}{n} = \frac{80 \cdot 9^2 + 60 \cdot 6^2}{120} = \mathbf{72} \rightarrow \sigma_B = \sqrt{72} = 8{,}49 \sim 8 \text{ pont}$$

Az egyes feladatsorokon belül a hallgatók eredménye a feladatsorok átlagától átlagosan 8 ponttal tér el.

Teljes szórásnégyzet

$$\sigma^2 = \sigma_K^2 + \sigma_B^2 = 327{,}17 + 72 = \mathbf{399,17}$$

$$\rightarrow \sigma = \sqrt{399{,}17} = 19{,}97 \sim 20 \text{ pont}$$

Az egyes hallgatók eredménye a teljes csoportra vonatkoztatott átlageredménytől átlagosan 20 ponttal tér el.

A vegyes kapcsolat mérőszámai

$$H^2 = \frac{\sigma_K^2}{\sigma^2} = \frac{327{,}17}{399{,}17} = 0{,}8196 \rightarrow \mathbf{81,96}\% \rightarrow H = \sqrt{H^2} = \sqrt{0{,}8196} = \mathbf{0,9053}$$

A feladatsor típusa és a hallgató eredménye között erős kapcsolat mutatható ki (H), a feladatsor típusa 81,96%-ban befolyásolja az elért eredmény alakulását.

9.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

9.3.1 Összefoglalás

A statisztikai elemzések során gyakran kerül sor egy adott jelenség vizsgálatánál különböző tényezők közötti összefüggések vizsgálatára. Az ismérvek típusától függően a sztochasztikus, vagyis valószínűségi kapcsolatok eltérő módon vizsgálhatók. Az asszociáció két nem mennyiségi ismérv közötti összefüggés számszerűsítésére alkalmas módszer. A megfelelő asszociációs mérőszámok – Yule, Csuprov, Cramer –segítségével a kapcsolat létét és erősségét tudjuk megállapítani. A mennyiségi és nem mennyiségi ismérv között lévő vegyes kapcsolat mérése során az elemzési lehetőségek köre bővül. A kapcsolat létén, erősségén túl az is mérhetővé válik, hogy a nem mennyiségi ismérv hány százalékban befolyásolja a mennyiségi ismérv alakulását. A szóráshányados és a szórásnégyzethányados így több információt szolgáltat az ismérvek közötti kapcsolat természetéről. Ezek a kapcsolatvizsgálati módszerek két ismérv esetén alkalmazhatók, lehetővé teszik a nem mennyiségi ismérvekhez kötődő jelenségek statisztikai összefüggéseinek vizsgálatát.

9.3.2 Önellenőrző kérdések

1. Mit értünk sztochasztikus kapcsolat alatt?
2. Mit jelent két ismérv függetlensége és a függvényszerű kapcsolat?
3. Milyen ismérvek közötti kapcsolatot számszerűsít az asszociáció?
4. Melyek az asszociáció mérőszámai?
5. Mit értünk függetlenség estén feltételezett gyakoriság alatt?
6. Milyen ismérvek közötti kapcsolatot számszerűsít a vegyes kapcsolat?
7. Mi a főátlag és hogyan határozható meg?
8. Melyek a vegyes kapcsolat mérőszámai?
9. Mit mutat meg a szórásnégyzet-hányados?
10. Mit számszerűsít a szóráshányados?

9.3.3 Gyakorló tesztek

☸ Válassza ki a megfelelő választ!

A hallgató szakja és zárthelyi dolgozatának eredménye között a kapcsolat milyen mutatóval számszerűsíthető?

Yule Csuprov **Szóráshányados**

Egy vállalatnál a dolgozók iskolai végzettsége és keresete közötti kapcsolat vizsgálata során azt az eredményt kaptuk, hogy a szórásnégyzet-hányados értéke 0,674. Mit jelent ez?

a. a két ismérv között közepesnél erősebb kapcsolat mutatható ki
b. az iskolai végzettség 67,4%-ban határozza meg a kereset alakulását
c. az iskolai végzettségenkénti keresetek a főátlagtól átlagosan 0,674-el térnek el.

A H mutató értéke a vonat típusa és a késés száma között 0,26. Mire következtetünk ebből?

a. a vonat típusa 0,26%-ban befolyásolja a késések számát
b. a vonat típusa és a késések száma között gyenge kapcsolat mutatható ki
c. A vonat típusa 26%-ban befolyásolja a késések számát

A Csuprov mutató értéke a dolgozók beosztása és neme között 0,28. Milyen erősségű a kapcsolat?

Gyenge Közepesnél gyengébb Erős

A vizsga sikeressége (sikeres/sikertelen) és a hallgató lakhelye (főváros/kisváros/falu) közötti kapcsolat milyen mutatóval számszerűsíthető?

Yule **Csuprov** Szóráshányados

Egy vállalatnál a dolgozók iskolai végzettsége és keresete közötti kapcsolat vizsgálata során azt az eredményt kaptuk, hogy a szóráshányados értéke 0,82. Mit jelent ez?

a. a két ismérv között erős kapcsolat mutatható ki
b. az iskolai végzettség 82%-ban határozza meg a kereset alakulását
c. a kereset 82%-ban határozza meg az iskolai végzettséget.

A Cramer mutató értéke két ismérv között 0,64. Melyik lehet a két ismérv?

a. a vendég származási helye és tartózkodási ideje
b. a vendég származási helye és a szállás típusa
c. a vendég tartózkodási ideje és az általa elköltött összeg nagysága

A szóráshányados értéke a vonat típusa és a késés száma között 0,64. Mire következtetünk ebből?

a. **a vonat típusa 40,96%-ban befolyásolja a késések számát**
b. a vonat típusa nem befolyásolja a késések számát
c. A vonat típusa 64%-ban befolyásolja a késések számát

Ha egy csoportban csak annak sikerült teljesíteni a statisztika tárgyat, aki járt előadásra, akkor

a. a tárgy teljesítése és az előadásra járás egymástól független
b. a tárgy teljesítése és az előadásra járás közötti kapcsolat szórás-hányados mutatóval számszerűsíthető
c. **a tárgy teljesítése és az előadásra járás között függvényszerű kapcsolat van**

Egy vállalatnál a dolgozók iskolai végzettsége és keresete közötti kapcsolat vizsgálata során azt az eredményt kaptuk, hogy a H^2 mutató értéke 0,325. Mit jelent ez?

a. a két ismérv között gyenge kapcsolat mutatható ki
b. **az iskolai végzettség 32,5%-ban határozza meg a kereset alakulását**
c. az iskolai végzettség 0,325%-ban befolyásolja a kereset alakulását

10. ISMÉRVEK KÖZÖTTI KAPCSOLAT VIZSGÁLATA II.: KORRELÁCIÓ- ÉS REGRESSZIÓSZÁMÍTÁS

10.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

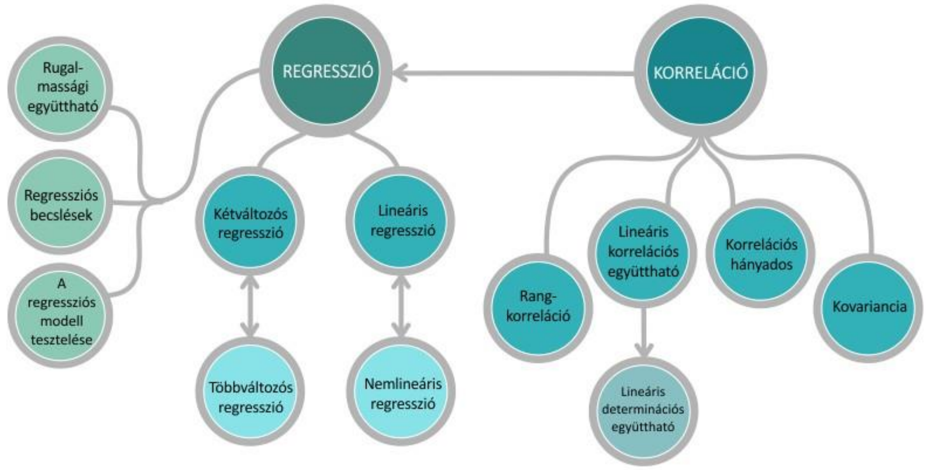
A statisztikai elemzések során gyakran keressük arra a kérdésre a választ, hogy egy adott jelenség alakulását milyen tényezők befolyásolják. A társadalmi és gazdasági folyamatok többnyire számszerűsíthetők, vagyis mennyiségi ismérvek közötti kapcsolatot vizsgálhatunk. Mennyiségi ismérvek esetén nem csupán a kapcsolat létét és erősségét tudjuk számszerűsíteni, hanem lehetővé válik a kapcsolat irányának meghatározása és a tényezők közötti ok-okozati összefüggés feltárása is. A korrelációszámítás segítségével kimutatható két vagy több mennyiségi ismérv közötti sztochasztikus kapcsolat léte, erőssége és iránya. A regresszió az ismérvek közötti kapcsolat természetének matematikai formában való leírására szolgál.

A lecke célja, hogy megismertesse a hallgatókat a mennyiségi ismérvek közötti kapcsolat kimutatására és jellemzésére szolgáló módszerekkel. A hallgató fel tudja ismerni azokat a területeket, amelyeken alkalmazható a korreláció és regressziószámítás, valamint legyen képes átlátni az elemzések struktúráját. A hallgató ismerje a korrelációs mérőszámokat, s tudja azt, hogy az egyes mutatók milyen információt szolgáltatnak a kapcsolatról. Képes legyen egy két ismérv közötti lineáris kapcsolatot leíró regresszió függvény paramétereit értelmezni, s tudja azt, hogyan lehet a különböző formájú regresszió függvények közül kiválasztani azt, amelyik a legjobban illeszkedik a tapasztalati értékekre.

10.2 TANANYAG

A korreláció és regressziószámítás két vagy több mennyiségi ismérv közötti sztochasztikus kapcsolat jellemzésére szolgáló statisztikai módszer. Az ismérvek között pozitív és negatív irányú kapcsolat is lehet, melyet lineáris és nem lineáris regresszió függvény segítségével is lehet leírni. A kapcsolat vizsgálatánál a legtöbb esetben nem teljeskörű a megfigyelés, a vizsgált jelenség elemzéséhez egy adott elemszámú minta áll rendelkezésre, melyből általános érvényű következtetéseket is megfogalmazhatunk.

A lecke részletesen a kétváltozós korreláció és regressziószámítás módszerét mutatja be, vagyis azt az esetet, amikor egy változó magyarázza egy másik változó alakulását. A két változó közötti sztochasztikus kapcsolat létének, erősségének, irányának meghatározására korrelációs mérőszámokat alkalmazunk. A regresszió függvény illesztése akkor célszerű, ha a vizsgált tényezők között szignifikáns kapcsolat mutatható ki. A függvény paramétereinek meghatározása és értelmezése mellett bemutatásra kerül a relatív változások számszerűsítésére alkalmas rugalmassági együttható is. Kitekintő jelleggel a nemlineáris függvényformák és a többváltozós lineáris regresszió is említésre kerül.



33. ábra: A korreláció- és regressziószámítás struktúrája

10.2.1 Korrelációs mérőszámok

A korreláció két vagy több mennyiségi ismérv közötti sztochasztikus kapcsolat létének, szorosságának és irányának kimutatására szolgáló módszer. A korrelációs mérőszámok közül a kovarianciát főként a kapcsolat irányának meghatározására használjuk. A mennyiségi ismérvek mérési szintjétől függően különböző korrelációs mérőszámokat használhatunk a kapcsolat erősségének jellemzésére. Arányskálán mérhető ismérvek esetén a Pearson-féle lineáris korrelációs együtthatót, illetve annak négyzetét a lineáris determinációs együtthatót alkalmazzuk, melyek azonban csak lineáris kapcsolat esetén használhatók megbízhatóan. Ordinális mérési szintű változóknál rangkorrelációs mérőszámokkal mutatható ki a kapcsolat szorossága.

Kovariancia (C)

A kovarianciát elsősorban a mennyiségi ismérvek közötti kapcsolat irányának megállapítására használjuk. A mutatónak nincs abszolút alsó és felső korlátja, így az előjele számít igazán, amely ha pozitív, akkor azonos irányú, míg ha negatív, akkor ellentétes irányú összefüggés mutatható ki az ismérvek között. Az azonos irányú összefüggés azt jelenti, hogy ha az egyik ismérv értékei emelkednek, akkor a másik ismérv értékei is emelkedni fognak (pl. ha nő a hőmérséklet, akkor emelkedni fog a vízfogyasztás is). Ellentétes irányú kapcsolat esetén az egyik ismérv értékeinek emelkedését a másik ismérv értékeinek csökkenése követi (pl. minél több egy autó futott kilométere, annál kevesebbe kerül). A mutató értéke 0, ha nincs kapcsolat, így ha ettől eltérő értéket vesz fel, az utal az ismérvek közötti kapcsolat létére.

$$C = \frac{\sum xy - n \cdot \bar{x} \cdot \bar{y}}{n-1} = \frac{\sum d_x d_y}{n-1} \text{ vagy } C = \frac{\sum xy - n \cdot \bar{x} \cdot \bar{y}}{N} = \frac{\sum d_x d_y}{N}$$

Attól függően, hogy a vizsgált jelenség esetén minden adat vagy csak azok egy része, azaz egy minta áll rendelkezésre, a mutató képletének nevezője eltér. Ha teljes sokaságot tudunk vizsgálni, akkor a nevezőben a teljes elemszám (N) szerepel, míg ha csak a sokaság egy részét vizsgáljuk, vagyis minta alapján vizsgálódunk, akkor a nevezőben n-1 szerepel. A gyakorlatban ritkán fordul elő, hogy a teljes sokaságra vonatkozóan vannak adatok, így a továbbiakban ennek megfelelően egy vizsgált jelenség kapcsán rendelkezésre álló minta segítségével vizsgálódunk.

Pearson-féle lineáris korrelációs együttható (r)

A kapcsolat erősségének és irányának meghatározására arányskálán mérhető ismérveknél a Pearson-féle lineáris korrelációs együtthatót használjuk. A mutató -1 és 1 közötti értéket vehet fel, az előjele a kapcsolat irányára utal, míg abszolút értéke a kapcsolat szorosságát jelzi. Ha a mutató értéke 0, a két változó egymástól független, míg a -1 negatív determinisztikus, a +1 pozitív determinisztikus kapcsolatra utal. A mutató 0-hoz közeli értéke gyenge, a szélső értékekhez közeledve pedig egyre erősebb összefüggést jelez.

$$r = \frac{C}{s_x \cdot s_y} = \frac{C}{\sqrt{\frac{\sum d_x^2}{n-1}} \cdot \sqrt{\frac{\sum d_y^2}{n-1}}} = \frac{C}{\sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} \cdot \sqrt{\frac{\sum (y-\bar{y})^2}{n-1}}}$$

A mutatót a kovarianciából származtatjuk, így a két indikátor előjele megegyezik. A nevezőben x és y változó korrigált tapasztalati (mintabeli) szórásának szorzata szerepel.

Lineáris determinációs együttható (r^2)

A lineáris korrelációs együttható négyzete, értéke 0 és 1 közé esik, melyet 100-al való szorzás után %-os formában értelmezünk. A mutató értelmezésénél szükség van a változók megkülönböztetésére, mely a regresszió által feltárandó ok-okozati összefüggést alapozza meg. Az egyik változó X, melyet független vagy magyarázó változónak, illetve egyszerűen oknak nevezünk, míg Y a függő vagy eredményváltozó, az okozat. A lineáris determináció együttható az a szám, mely megmutatja, hogy a magyarázó változó (X) hány százalékban befolyásolja az eredményváltozó (Y) szóródását, alakulását.

Korrelációs hányados (η)

A korrelációs hányados X változó szerinti csoportosításban mutatja Y változó szóródásának alakulását. A mutató négyzete is értelmezhető (a vegyes kapcsolatnál megismert H és H^2 mutatóval megegyező módon). Mindkét indikátor abszolút értékben 0 és 1 közötti értéket vehet fel: $0 \leq |\eta;\ \eta^2| \leq 1$. Kiszámításukhoz előbb X változó értékeit csoportosítjuk (osztályközös gyakorisági sor), majd a csoportokhoz hozzárendeljük Y átlagos értékeit.

$$\eta = \sqrt{\frac{S_K(y)}{S\,(y)}}$$

Rangkorreláció

A rangkorreláció sorrendi skálán mérhető (ordinális mérési szintű) mennyiségi ismérvek közötti kapcsolat kimutatására alkalmas. A rangkorrelációs együtthatók segítségével megállapítható, hogy az egyik rangsorban elfoglalt hely alakulását, mennyiben befolyásolja a másik rangsorban elfoglalt hely, azaz mennyire van összhang a rangsorok között.

A cégek által végzett piackutatások során gyakran előfordul, hogy a kérdőívekben egy adott termék választása kapcsán különböző szempontokat kell rangsorolni. Tipikus kérdés a „miért választja az adott terméket?”, melyre a különböző válaszok között sorrendet kell felállítani: ár, márka, reklám, ismerősök ajánlása stb.

A rangkorreláció mérésénél azonos elemszámú rangsorokra van szükség, vagyis a rangszámok összegének az összehasonlítandó adatsorokban meg kell egyezni. A legegyszerűbb esetben egyértelmű rangsorok vannak, amelyekben minden rang egyszer fordul elő, vagyis nincs azonosság. A számítást bonyolítja, ha vannak egyforma rangszámok, illetve az is, ha nem csupán kettő, hanem több rangsor közötti összefüggést kívánunk vizsgálni. A rangkorrelációs együtthatók közül a *Spearman-féle ρ (rhó)* alkalmas két rangsor közötti összefüggések vizsgálatára, míg a *Kendall τ* (tau) több rangsor együttes vizsgálatára alkalmazható.

Spearman-féle rangkorrelációs együttható

A Spearman-féle ρ a rangkorreláció mérésére leggyakrabban alkalmazott mérőszám, melynek értéke -1 és 1 közé eshet, ahol a -1 azt jelenti, hogy a rangsor a két esetben teljesen fordított, míg a +1 esetében teljesen megegyezik. Ha az együttható értéke 0, akkor nincs összefüggés.

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot \sum d^2}{n^3 - n}$$

A képletben n a megfigyelt elemek száma, míg a d a két rangsorban elfoglalt helyezés közötti különbség.

☸ Egy diplomamunka pályázatra 10 pályamunka érkezett, melyeket két-két bíráló értékelt. A két bíráló a 10 dolgozatot az alábbi sorrendbe rendezte:

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A bíráló	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
B bíráló	2.	3.	1.	5.	6.	9.	4.	8.	10.	7.

Állapítsa meg, hogy a két bíráló által felállított sorrend mennyire van összhangban!

A rangkorreláció kiszámításához szükség van a két rangsorban elfoglalt helyezés közötti különbségekre[10] és azok négyzetösszegére is.

24. *Munkatábla a rangkorrelációhoz szükséges négyzetösszeg kiszámításához*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
A bíráló	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	
B bíráló	2.	3.	1.	5.	6.	9.	4.	8.	10.	7.	
különbség (d)	-1	-1	2	-1	-1	-3	3	0	-1	3	$\sum$
különbség négyzete (d^2)	1	1	4	1	1	9	9	0	1	9	**36**

A két bíráló értékelésének összefüggése:

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n^3 - n} = 1 - \frac{6 \cdot 36}{10^3 - 10} = 1 - \frac{216}{990} = \mathbf{0,7818}$$

A két bíráló értékelése tehát erős hasonlóságot mutat.

Előfordulnak olyan esetek, amikor a sorszám két értéknél megegyezik, azaz „holtverseny” alakul ki. Ebben az esetben a két egymást követő sorrendi érték egyszerű számtani átlagát rendeljük mindkét értékhez, melyet akkor kapnánk, ha az értékek nem lennének egyformák. Az ilyen rangszámokat kapcsolt rangoknak nevezzük. Például ha egy rangsorban a 4. és 5. elem megegyezik, akkor kétszer szerepelne a 4., ehelyett 4,5 rangot

[10] A kivonás iránya tetszőleges, mert a rangkorreláció számításához a négyzetösszegre van szükség.

kap mindkettő. A kapcsolt rang tehát a két szomszédos rang egyszerű számtani átlaga. Ebben az esetben módosul a képlet:

$$\rho = \frac{\frac{1}{6}(n^3-n)-\left(T_x+T_y\right)-\sum d^2}{\sqrt{\left[\frac{1}{6}(n^3-n)-2T_x\right]\cdot\left[\frac{1}{6}(n^3-n)-2T_y\right]}}$$

$T = \sum \frac{1}{12}(t_j^3 - t_j)$, t a kapcsolt rangok száma, j az azonos rangszámú csoport

A kapcsolt rangok számát úgy határozzuk meg, hogy megnézzük hány esetben van azonos sorszám, így például, ha két 4. helyezett van, akkor mindkettő, 4,5. sorszámot kap, de csak az egyébként 5. sorszámút tekintjük kapcsolt rangúnak. Ha három azonos sorszámú elem van a rangsorban, akkor 2 tekinthető kapcsoltnak. Az azonos rangszámú csoportok száma pedig azon csoportok száma, ahol szükség van korrekcióra.

☸ Egy szakdolgozat pályázatra 10 pályamunka érkezett, melyeket két-két bíráló értékelt. A két bíráló a 10 dolgozatot az alábbi sorrendbe rendezte:

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A bíráló	1.	1.	3.	4.	5.	5.	7.	8.	9.	10.
B bíráló	3.	1.	1.	5.	6.	7.	4.	7.	10.	7.

Állapítsa meg, hogy a két bíráló által felállított sorrend mennyire van összhangban!

Látszik, hogy több esetben is sorrendi azonosság figyelhető meg, ezért kapcsolt rangot kapnak az azonos sorrendi számú értékek és azok alapján határozzuk meg a különbségek négyzetösszegét.

25. *Munkatábla a rangkorrelációhoz szükséges négyzetösszeg kiszámításához*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
A bíráló	1,5.	1,5.	3.	4.	5,5.	5,5.	7.	8.	9.	10.	
B bíráló	3.	1,5.	1,5.	5.	6.	8.	4.	8.	10.	8.	
különbség (d)	-1,5	0	1,5	-1	-0,5	-2,5	3	0	-1	2	Σ
különbség négyzete (d^2)	2,25	0	2,25	1	0,25	6,25	9	0	1	4	**26**

$$T_A = \frac{1}{12}[(2^3-2)+(2^3-2)] = 1 \qquad T_B = \frac{1}{12}[(3^3-3)+(2^3-2)] = 2,5$$

Spearman-féle rangkorrelációs együttható:

$$\rho = \frac{\frac{1}{6}(10^3 - 10) - (1 + 2{,}5) - 26}{\sqrt{\left[\frac{1}{6}(10^3 - 10) - 2 \cdot 1\right] \cdot \left[\frac{1}{6}(10^3 - 10) - 2 \cdot 2{,}5\right]}} = \frac{165 - 3{,}5 - 26}{\sqrt{163 \cdot 160}} = \frac{135{,}5}{161{,}493}$$

$$\rho = \mathbf{0{,}8391}$$

A két bíráló értékelése ez esetben is erős hasonlóságot mutat.

Kendall-féle rangmódszer

A Kendall-féle rangmódszer több változó esetén is alkalmazható. Az előző példához kapcsolódva mérhetővé válik az is, ha nem csupán kettő, hanem *m* számú bíráló, *n* db dolgozatot bírál. Megállapítható, hogy milyen összefüggés mutatható ki a bírálók értékelése között, mennyire hasonlítanak azok egymáshoz. A módszer lényege, hogy a különbségek helyett a rangszámösszegekre helyezi a hangsúlyt, vagyis arra, hogy egy adott elem a rangsorokban összesen hány helyezést kapott. A mutató értéke 0 és 1 között mozog, ahol a 0 teljes egyet nem értést, az 1 pedig teljes egyezést jelent.

☸ Egy szakdolgozat pályázatra 10 pályamunka érkezett, melyeket négy bíráló értékelt, akik az alábbi sorrendet állították fel a dolgozatok között:

Dolgozatok	A bíráló	B bíráló	C bíráló	D bíráló
	rangsora			
(1)	1.	2.	3.	4.
(2)	2.	3.	1.	1.
(3)	3.	1.	2.	3.
(4)	4.	5.	4.	5.
(5)	5.	6.	6.	7.
(6)	6.	4.	7.	2.
(7)	7.	9.	5.	6.
(8)	8.	8.	10.	9.
(9)	9.	7.	8.	10.
(10)	10.	10.	9.	8.

Állapítsa meg, hogy a bírálók által felállított sorrend mennyire van összhangban!

26. *Munkatábla a Kendall-féle rangmódszer alkalmazásához*

Dolgozatok	**A bíráló**	**B bíráló**	**C bíráló**	**D bíráló**	***Rangszám-összeg (R)***	***Átlagtól vett eltérés***	***Átlagtól vett eltérés négyzete***
	rangsora						
(1)	1.	2.	3.	4.	10	-12	144
(2)	2.	3.	1.	1.	7	-15	225
(3)	3.	1.	2.	3.	9	-13	169
(4)	4.	5.	4.	5.	18	-4	16
(5)	5.	6.	6.	7.	24	2	4
(6)	6.	4.	7.	2.	19	-3	9
(7)	7.	9.	5.	6.	27	5	25
(8)	8.	8.	10.	9.	35	13	169
(9)	9.	7.	8.	10.	34	12	144
(10)	10.	10.	9.	8.	37	15	225
					220	*Eltérés négyzet-összeg*	**1130**
				ÁTLAG	**22**		

A számítás során elsőként össze kell adni az egyes dolgozatokhoz tartozó rangszámokat, vagyis azokat a sorrendi értékeket, amelyekkel az egyes bírálók helyezték az adott dolgozatot, ez lesz a rangszámösszeg (R). Ha a rangszámösszegeket összeadjuk és elosztjuk a dolgozatok számával, akkor megkapjuk, hogy az egyes dolgozatokhoz átlagosan mennyi rangszámösszeg tartozik ($\sum R/n = 220/10$). Ezt követően kiszámoljuk, hogy az egyes dolgozatok rangszámösszege menyivel tér el az átlagtól, s a pozitív és negatív eltérés kiküszöbölése érdekében ezeket a különbségeket négyzetre is emeljük. A végén összegzett eltérés négyzetösszeg lesz a Kendall-féle mutató d értéke. A d_{max} a maximális egyetértést jelzi, melyet a bírálók és dolgozatok számát figyelembe véve határozhatunk meg.

$$\boldsymbol{W} = \frac{\boldsymbol{d}}{\boldsymbol{d_{max}}} \leftarrow d_{max} = \frac{m^2(n^3 - n)}{12} = \frac{4^2 \cdot (10^3 - 10)}{12} = 1320$$

$$W = \frac{1130}{1320} = \mathbf{0{,}8561}$$

A kapott eredmény alapján elmondható, hogy a bírálók véleménye nagyon hasonló, vagyis nagy az egyetértés a dolgozatok sorrendjét illetően.

10.2.2 Kétváltozós lineáris regresszió

A regressziós függvényt a korreláció segítségével feltárt kapcsolat matematikai jellemzésére használjuk. Mennyiségi ismérvekről lévén szó nem csupán azt tudjuk statisztikailag meghatározni, hogy van-e, s, ha igen milyen erősségű és irányú kapcsolat az ismérvek között, hanem azt is, hogy a magyarázó változóban bekövetkező változás milyen nagyságrendű változást idéz elő az eredményváltozóban. A vizsgált jelenség kapcsán tehát a kiválasztott változók között előzetesen ok-okozati feltételezünk, melynek alátámasztására alkalmazzuk a regressziós modellt.
A sokaság egészére vonatkozó lineáris regresszió függvény az alábbi formában írható fel:

$$\mathbf{Y} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \cdot \mathbf{X} + \boldsymbol{\varepsilon}$$

A függvény egyenletében Y az eredményváltozó értéke, X a magyarázó változó, ε a véletlen hatás. A változók közötti összefüggés jellemzésére a β_0 és β_1 paraméterekre van szükség. A legtöbb esetben azonban nem rendelkezünk a teljes sokaságra vonatkozóan információval, csak egy minta alapján tudunk következtetéseket levonni. A tapasztalati vagy mintabeli regresszió függvény az alábbi formában írható fel:

$$\hat{\mathbf{y}} = \mathbf{b}_0 + \mathbf{b}_1 \cdot \mathbf{x}$$

A mintabeli regresszió alapján lehetőség van arra, hogy a változók között feltárt összefüggést ne csak a tapasztalati értékeknél, hanem a teljes sokaság vonatkozásában megfogalmazhassuk. Ehhez a modell tesztelésére van szükség (F próba). A mintában feltárt összefüggés segítségével a b_0 és b_1 paraméterekből becsléssel meghatározható az az intervallum is, melybe a sokasági regresszió függvény β_0 és β_1 paraméterei adott valószínűséggel beletartoznak.

A tapasztalati (mintabeli) regresszió függvény paramétereinek meghatározása és értelmezése

A kétváltozós lineáris regresszió függvény paraméterei kétféle formában határozhatók meg. Az ún. normál egyenletek segítségével egy olyan két ismeretlenes egyenletrendszert kell megoldani, melynek eredményeként b_0 és b_1 értékeit kapjuk meg.

$$\sum y \ = b_0 \cdot n + b_1 \cdot \sum x$$
$$\sum xy = b_0 \cdot \sum x + b_1 \cdot \sum x^2$$

A paraméterek közvetlenül, képletek segítségével is meghatározhatók:

$$\mathrm{b_1} = \frac{\sum \mathrm{d_x d_y}}{\sum \mathrm{d_x^2}} \qquad \mathrm{b_0} = \bar{\mathrm{y}} - \mathrm{b_1}\bar{\mathrm{x}}$$

A $\mathbf{b_0}$ paraméter a függvény x=0 helyen felvett értéke, az a szám, mely megmutatja, hogy ha a magyarázó változó (x) értéke 0, akkor mennyi az eredményváltozó (y) értéke. A b_0 paraméter értelmezhetősége függ egyrészt attól, hogy a magyarázó változó értéke lehet-e 0, vagyis van-e a gyakorlati életben értelme x=0 pontnak. A paraméter értelmezhetőségét befolyásolja továbbá az is, hogy a kiszámított érték eleme-e az eredményváltozó értelmezési tartományának, vagyis gyakorlatban felveheti-e a kiszámított értéket a függő változó. Az értelmezhetőség kritériumait minden esetben vizsgálni kell, mert a matematikailag kiszámítható paraméternek nem minden esetben van gyakorlati értelmezése. A $\mathbf{b_1}$ paraméter minden esetben értelmezhető, azt mutatja meg, hogy ha a magyarázó változó értékét egy egységgel növeljük, akkor az eredményváltozóban átlagosan milyen irányú és mekkora változás következik be.

Rugalmassági együttható (elaszticitás)

A rugalmassági együttható megmutatja, hogy Y relatív változása hányszorosa X relatív változásának, azaz egy tetszőlegesen választott x_0 pontban x magyarázó változó 1%-os emelkedése, hány %-os változást okoz az eredményváltozóban, Y-ban.

$$\boldsymbol{E\,(y, x = x_0) = b_1 \cdot \frac{x_0}{b_0 + b_1 \cdot x_0}}$$

☸ Egy főiskolai oktató szeretné feltárni a hallgatók statisztika dolgozatra való felkészülési ideje, valamint a dolgozat eredménye közötti összefüggést.

A szituációról és adatgyűjtésről tekintse meg az alábbi **videó**t:

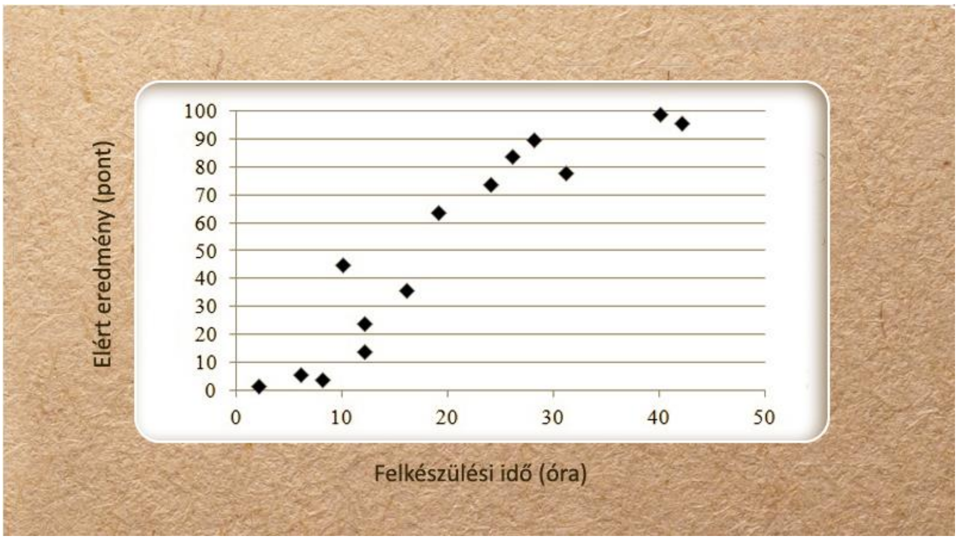
Az elemzéshez az alábbi 14 elemű véletlen minta áll rendelkezésére:

Felkészülési idő (óra)	**Zárthelyi dolgozat pontszáma (max. 100)**
26	84
12	24
42	96
10	45
6	6
31	78
19	64

Felkészülési idő (óra)	**Zárthelyi dolgozat pontszáma (max. 100)**
24	74
28	90
2	2
40	99
12	14
16	36
8	4

A kapcsolatvizsgálat első lépéseként megnézzük az ismérvek típusát, amely meghatározza, hogy mely sztochasztikus kapcsolat mérőszámai alkalmazhatók. Ebben az esetben a felkészülési idő és a pontszám is mennyiségi ismérv, vagyis korrelációszámítással határozhatjuk meg, hogy van-e kapcsolat, s az milyen erősségű és irányú az ismérvek között. Az információk birtokában dönthetünk arról, hogy a kapcsolat természetére vonatkozóan felírunk-e regresszió függvényt. Ha a kimutatott kapcsolat gyenge, akkor a regressziós modell megkérdőjelezhető.

Mennyiségi ismérvek esetén az összetartozó értékpárok pontdiagramon ábrázolhatók, melynek segítségével grafikusan illusztrálható a közöttük lévő összefüggés.



34. ábra: A statisztika vizsgára való felkészülés ideje és a dolgozat eredményessége közötti kapcsolat

- Lineáris regresszió esetén grafikusan azt az egyenest keressük, amely a legjobban illeszkedik a megfigyeléseket reprezentáló pontokra. A legjobban illeszkedő lineáris regresszió függvény felírásának leggyakrabban használatos formája a legkisebb négyzetek módszere (OLS = ordinary least squared method). Ebben az esetben minimalizáljuk a tényleges és regresszió függvényből becsült értékek eltérését, vagyis $\sum(y-\hat{y})^2 \rightarrow min.$

Az ábrán kirajzolódik kapcsolat az ismérvek között. A lineáris determinációs együttható értelmezéséhez és a regressziós modellhez meg kell határozni, hogy melyik lesz a magyarázó változó és melyik az eredményváltozó. Jelen esetben logikusan a felkészülési idő (x) magyarázhatja az elért eredmény (y) alakulását.

A képletekbe való behelyettesítéséhez számítási részeredményekre van szükségünk. Meg kell határozni az ismérvértékek átlagát, valamint az egyes értékek attól való eltérését, s ezek négyzetét, valamint az egyes változóknál számított eltérések páronkénti szorzatát is.

27. *Munkatábla a regresszió függvény paramétereinek meghatározásához*

	Felkészülési idő x	Zárthelyi dolgozat pontszáma y	$d_x = x - \bar{x}$	$d_y = y - \bar{y}$	$d_x d_y$	d_x^2	d_y^2
	26	84	6,286	32,857	206,531	39,510	1079,592
	12	24	-7,714	-27,143	209,388	59,510	736,735
	42	96	22,286	44,857	999,673	496,653	2012,163
	10	45	-9,714	-6,143	59,673	94,367	37,735
	6	6	-13,714	-45,143	619,102	188,082	2037,878
	31	78	11,286	26,857	303,102	127,367	721,306
	19	64	-0,714	12,857	-9,184	0,510	165,306
	24	74	4,286	22,857	97,959	18,367	522,449
	28	90	8,286	38,857	321,959	68,653	1509,878
	2	2	-17,714	-49,143	870,531	313,796	2415,020
	40	99	20,286	47,857	970,816	411,510	2290,306
	12	14	-7,714	-37,143	286,531	59,510	1379,592
	16	36	-3,714	-15,143	56,245	13,796	229,306
	8	4	-11,714	-47,143	552,245	137,224	2222,449
Σ	***276***	***716***	*0,000*	*0,000*	**5544,571**	**2028,857**	**17359,714**
	$\bar{x} = 19,714$	$\bar{y} = 51,143$					

A további számításokhoz szükséges részeredmények:

$$\bar{x} = \frac{\sum x}{n} = \frac{276}{14} = 19{,}714 \qquad \bar{y} = \frac{\sum y}{n} = \frac{716}{14} = 51{,}143$$

$$\sum d_x^2 = 2028{,}857 \qquad \sum d_y^2 = 17359{,}714 \qquad \sum d_x d_y = 5544{,}571$$

Korrelációs mérőszámok segítségével vizsgálja meg a hallgatók statisztika vizsgára való felkészülési ideje, valamint a dolgozatuk eredménye közötti összefüggést!

KOVARIANCIA

$$\mathbf{C} = \frac{\sum \mathrm{d_x d_y}}{\mathrm{n}-1} = \frac{5544{,}571}{14-1} = \mathbf{426{,}51}$$

A hallgatók statisztika vizsgára való felkészülési ideje, valamint a dolgozatuk eredménye között *POZITÍV irányú összefüggés mutatható ki*, vagyis *ha többet készül a vizsgára a hallgató, akkor magasabb pontszámot ér el* (azonos irányú változás).

PEARSON-FÉLE LINEÁRIS KORRELÁCIÓS EGYÜTTHATÓ

$$\boldsymbol{r} = \frac{C}{s_x \cdot s_y} = \frac{C}{\sqrt{\frac{\sum d_x^2}{n-1}} \cdot \sqrt{\frac{\sum d_y^2}{n-1}}}$$

$$\boldsymbol{r} = \frac{426{,}51}{\sqrt{\frac{2028{,}857}{14-1}}\sqrt{\frac{17359{,}714}{14-1}}} = \frac{426{,}51}{12{,}49 \cdot 36{,}54} = \mathbf{0{,}9345}$$

A hallgatók statisztika vizsgára való felkészülési ideje, valamint a dolgozatuk eredménye között *SZOROS, POZITÍV irányú kapcsolat mutatható ki, vagyis minél többet készül valaki, annál jobb eredményt ér el.*

LINEÁRIS DETERMINÁCIÓS EGYÜTTHATÓ

$$\boldsymbol{r^2} = 0{,}9345^2 = \mathbf{0{,}8733} \xrightarrow{\cdot 100} \mathbf{87{,}33}\%$$

A hallgatók statisztika vizsgára való felkészülési ideje (x) *87,33%-ban befolyásolja* a dolgozatuk eredményét (y).

Lineáris összefüggést feltételezve határozza meg a kapcsolatot leíró regresszió függvényt és értelmezze a paramétereit!

A legegyszerűbben képlettel határozhatók meg a paraméterek értékei:

$$\boldsymbol{b_1} = \frac{\sum d_x d_y}{\sum d_x^2} = \frac{5544{,}571}{2028{,}857} = \mathbf{2{,}7329}$$

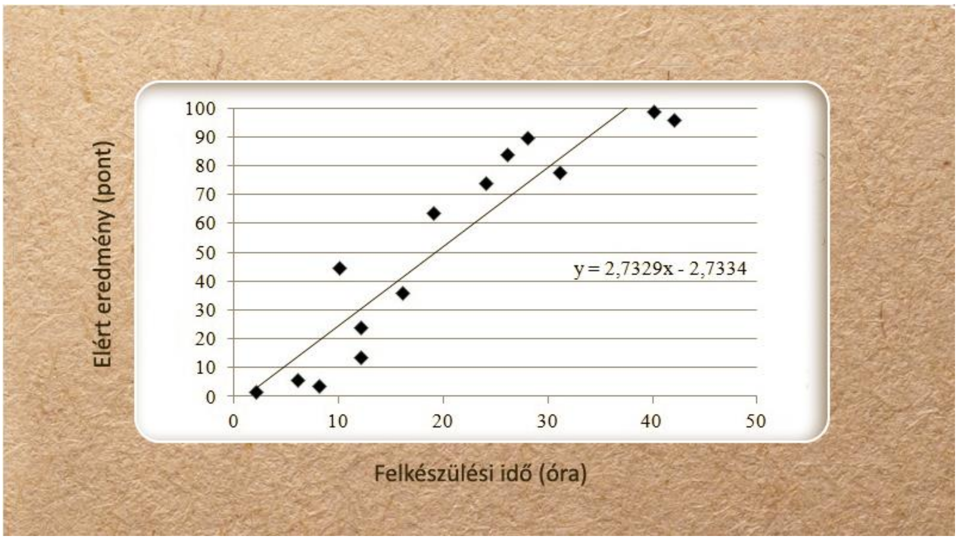
$$\boldsymbol{b_0} = \bar{y} - b_1 \cdot \bar{x} = 51{,}143 - 2{,}7329 \cdot 19{,}714 = -\mathbf{2{,}7334}$$

A kétváltozós lineáris regresszió függvény egyenlete:

$$\boldsymbol{\hat{y} = -2{,}7334 + 2{,}7329 \cdot x}$$

A regresszió függvény b_0 paramétere nem értelmezhető, mert a kapott érték nem eleme *y* értelmezési tartományának, vagyis az eredményváltozó, mely jelen esetben a dolgozat pontszáma nem lehet negatív. A b_1 paraméter alapján megállapítható, hogy ha a hallgató egy órával többet készül a vizsgára, akkor a dolgozatának eredménye átlagosan 2,73 ponttal magasabb.

A megfigyelési értékekre az alábbi lineáris regressziós egyenes illeszthető:



35. ábra: A statisztika vizsgára való felkészülés ideje és a dolgozat eredményessége közötti kapcsolatot leíró lineáris regresszió függvény (09_10_K03)

Határozza meg a dolgozat pontszámának rugalmasságát (elaszticitását) átlagos felkészülési idő esetén, illetve abban az esetben, ha a hallgató 10 órát fordít a felkészülésre!

RUGALMASSÁGI EGYÜTTHATÓ

Átlagos felkészülési idő esetén: $x_0 = 19{,}7$

$$E\,(y, x = x_0) = b_1 \cdot \frac{x_0}{b_0 + b_1 \cdot x_0}$$

$$E\,(y, x = 19{,}7) = 2{,}73 \cdot \frac{19{,}7}{(-2{,}68 + 2{,}73 \cdot 19{,}7)} = \mathbf{1{,}053}$$

Ha a hallgató 19,7 órát fordít a felkészülésre és ezt 1%-kal megnöveli, akkor a dolgozat pontszáma átlagosan 1,053 %-kal fog emelkedni.

10 órás felkészülés esetén: $x_0 = 10$

$$E\ (y, x = x_0) = \ b_1 \cdot \frac{x_0}{b_0 + b_1 \cdot \ x_0}$$

$$E\ (y, x = 10) = \ 2{,}73 \cdot \frac{10}{(-2{,}68 + 2{,}73 \cdot \ 10)} = \mathbf{1{,}109}$$

Ha a hallgató 10 órát fordít a felkészülésre és ezt 1%-kal megnöveli, akkor a dolgozat pontszáma átlagosan 1,109 %-kal fog emelkedni.
Az adatok MS Excel programban történő elemzését mutatja be az alábbi **animáció**:

Regressziós becslések és a modell tesztelése

A mintabeli regresszió függvény lehetőséget biztosít arra, hogy különböző becsléseket végezzünk a sokaságra vonatkozóan. A becslések egy része nem ismert megfigyelési értékekre vonatkozik, míg a másik fő csoportot a sokasági regresszió függvény paramétereinek közelítése képezi.

- A mintából történő következtetés egyik módszere a becslés. A vizsgált jelenség kapcsán rendelkezésre álló minta alapján valamely sokasági paraméter értékét közelítjük egy becslőfüggvény segítségével meghatározott feltételek mellett. Az adott paraméter konkrét értéke is meghatározható (pontbecslés), de leggyakrabban azt az intervallumot (konfidencia intervallum vagy megbízhatósági tartomány) jelöljük ki, amelybe a paraméter értéke adott valószínűséggel (π) beleesik (intervallumbecslés). A módszer jellegéből (következtetés) adódóan a hiba lehetőségével is számolni kell.

A *regresszió függvényből történő egyedi érték becslése* azt jelenti, hogy adott x érték (x_0) esetén a mintabeli regresszió függvény segítségével megbecsülhető az az intervallum, amelybe az eredményváltozó értéke adott valószínűséggel (π) beletartozik. Az *átlagos érték becslése* során pedig adott x érték (x_0) esetén azt az intervallumot becsüljük meg, ahová az eredményváltozó átlagos értéke fog beleesni adott valószínűséggel (π). Meghatározott becslőformulákkal a mintabeli regresszió b_0 és b_1 paraméterei alapján közelíthető a teljes sokaságra vonatkoztatott regressziós modell β_0 és β_1 paramétere. A becslések érzékenyek a minta elemszámára, minél nagyobb a minta, annál jobb a becslés.

Az ismérvek közötti kapcsolat vonatkozásában az általános érvényű következtetések megfogalmazásához szükség van arra, hogy teszteljük valóban megfelelő-e a regressziós modell. A teszt lényege, hogy megbizonyosodjunk arról, hogy nemcsak a mintabeli értékek esetén igazolható a változók közötti összefüggés. A regresszió függvény b_1 és β_1 paramétere fontos e tekintetben, hiszen az mutatja meg, hogy a magyarázó változó hogyan befolyásolja az eredményváltozó alakulását. A hipotézisvizsgálat során arról döntünk, hogy a sokasági regresszió függvény β_1 paramétere eltér-e a nullától. Abban az esetben ugyanis, ha a β_1 paraméter értéke nulla, akkor az egyenletből kiesik a magyarázó változó, hiszen együtthatója nulla, így a függvény konstans. Ez pedig azt jelenti, hogy a választott magyarázó változótól független az eredményváltozó alakulása, vagyis nincs ok-okozati összefüggés a változók között.

A hipotézisvizsgálat a sokaság valamely paraméterére vagy egyéb jellemzőjére vonatkozóan megfogalmazott állítás helyességének vizsgálata, melyet a mintabeli adatok segítségével végezhetünk el. A módszer lényege, hogy az alapállításhoz kapcsolódóan két, egymást kizáró hipotézist fogalmazunk meg. A nullhipotézisben mindig egyenlőség szerepel, az alternatív hipotézisben pedig olyan esemény, mely lefedi a teljes eseményrendszert az alapállítás szempontjából. A paramétertől függő statisztikai próba segítségével döntünk a nullhipotézisben megfogalmazott állításról, elfogadjuk vagy elvetjük, s ez alapján döntünk az alapállításról.

10.2.3 Nemlineáris regresszió

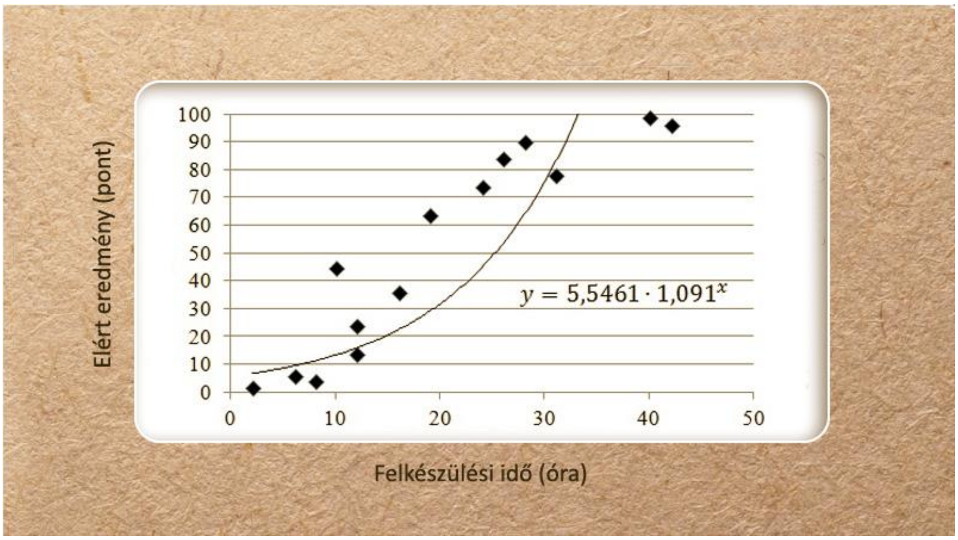
A megfigyelési értékekre nem csak egyenes illeszthető, azonban a leggyakrabban olyan nemlineáris formát alkalmazunk, melyek visszavezethetők egyenesre. Az exponenciális és a hatványkitevős regresszió függvény esetén logaritmikus átalakítással lineáris formát kapunk, ezért nevezik őket lineárisra visszavezethető nemlineáris regresszió függvényeknek.

Az **exponenciális regresszió** függvényt abban az esetben használjuk, ha az adott jelenség növekedése függ a jelenség elért színvonalától. A sokasági regresszió függvény az alábbi formában írható fel:

$\boldsymbol{Y = \alpha \cdot \beta^x \cdot \varepsilon}$ vagy az előző jelölést használva $Y = \beta_0 \cdot \beta_1^x \cdot \varepsilon$

A tapasztalati regresszió függvény ennek mintájára: $\mathbf{\hat{y} = a \cdot b^x}$. Ebből logaritmikus átalakítással az alábbi függvényt kapjuk: $\lg \hat{y} = \lg a + \lg b \cdot x$, ahol legyen $lg\ a = b_0$ és $lg\ b = b_1$, így egy lineáris függvénnyel dolgozhatunk tovább.

Illesszen az előző adatsorra exponenciális regresszió függvényt!



36. ábra: A statisztika vizsgára való felkészülés ideje és a dolgozat eredményessége közötti kapcsolatot leíró exponenciális regresszió függvény (09_10_K04)

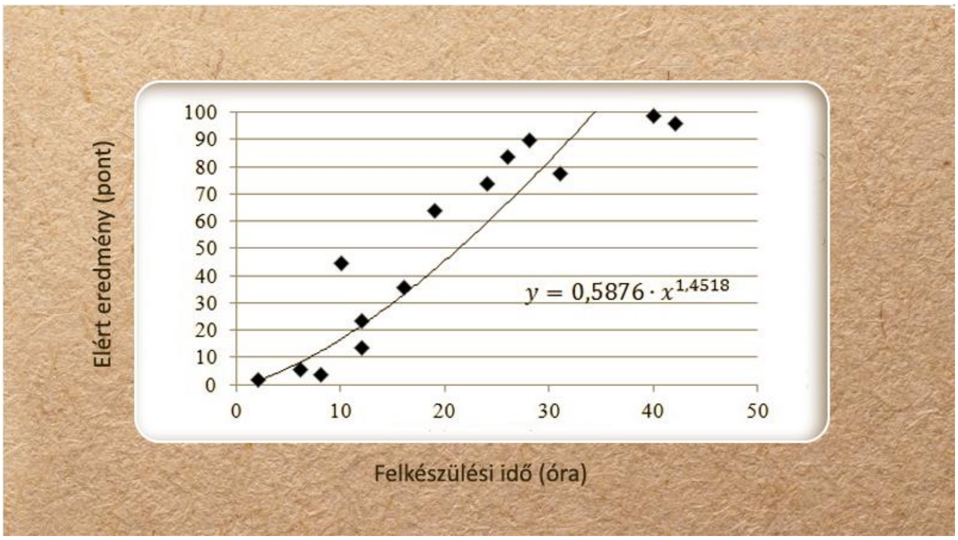
Az exponenciális regresszió függvény paraméterei közül *a* értékét ritkán értelmezzük, a függvény x=0 helyen felvett értéke, ha van értelme. A *b* (β_1) paraméter azt mutatja meg, hogy a magyarázó változó egységnyi növekedése hányszorosára növeli vagy csökkenti az eredményváltozó értékét.

A felkészülési idő és az elért eredmény vonatkozásában a b paraméter értéke 1,091, amely azt jelenti, hogy a*zok a hallgatók, akik 1%-kal több időt fordítanak felkészülésre, átlagosan 9,1 %-kal több pontra számíthatnak.*

A **hatványkitevős regresszió** esetén szintén logaritmikus átalakítás után lineáris függvényt kapunk. A sokasági regresszió függvény az alábbi formában írható fel: $\boldsymbol{Y = \alpha \cdot x^{\beta} \cdot \varepsilon}$

A tapasztalati regresszió függvény alakja: $\hat{y} = a \cdot x^b$ Ebből logaritmikus átalakítással az alábbi függvényt kapjuk: $\lg\hat{y} = \lg a + b \cdot \lg x,$ ahol gyen $lg\ a = b_0$ és $b = b_1$

Illesszen az előző adatsorra hatványkitevős regresszió függvényt!



37. ábra: A statisztika vizsgára való felkészülés ideje és a dolgozat eredményessége közötti kapcsolatot leíró hatványkitevős regresszió függvény (09_10_K05)

A hatványkitevős regresszió függvény minden pontjában azonos a rugalmasság, mely a kitevővel egyezik meg. A *b* paraméter így elaszticitási együtthatóként értelmezhető, mely azt mutatja meg, hogy 1%-kal nagyobb X értékhez, hány %-kal nagyobb vagy kisebb Y érték tartozik. Az *a* paramétert a gyakorlatban ritkán értelmezzük, x=1 helyen felvett érték, ha van értelme.

A felkészülési idő és az elért eredmény vonatkozásában a b paraméter értéke 1,4518, amely azt jelenti, hogy azok a hallgatók, akik 1%-kal több időt fordítanak felkészülésre, 1,4518 %-kal több pontra számíthatnak.

10.2.4 Illeszkedésvizsgálat

A tapasztalati értékekre bármilyen regresszió függvény illeszthető, célszerű azonban azt választani, amelyik a legjobban illeszkedik. A leggyakrabban a lineáris formát választjuk, matematikailag ugyanis ezzel a legkönnyebb dolgozni. A regresszióból történő becslések azonban annál pontosabbak, minél jobban illeszkedő formát választunk.

Az a regresszió függvény illeszkedik legjobban, ahol a tapasztalati és a regresszió függvényből becsült értékek eltérése (reziduum) a legkisebb. A reziduális szórás lényegében a reziduumok (e) átlagtól vett átlagos eltérését mutatja meg, melyet az Y adatsor mértékegységében fejezhetünk ki.

Minél kisebb a reziduális szórás értéke, annál kisebb az eltérés a tapasztalati és a becsült értékek között, vagyis annál jobb a modell.

$$s_e = \sqrt{\frac{\sum e^2}{n-2}}$$

Statisztikailag az illeszkedésvizsgálat során a reziduális szórás értékét határozzuk meg, s amelyik formánál a legkisebb az értéke, az a regresszió illeszkedik legjobban a tapasztalati értékekre. A gyakorlati munka során más indikátorok is lehetnek jelzésértékűek az illeszkedés vizsgálatához, így például az r^2 értéke, amely annál magasabb, minél jobban illeszkedik a függvény a megfigyelési pontokra.

10.2.5 Többváltozós regresszió

A gyakorlati életben ritkán fordul elő, hogy egy változó alakulását csupán egyetlen tényező magyarázza, ezért többváltozós modellekkel dolgozunk. A többváltozós lineáris regresszió függvény az alábbi formában írható fel:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 \dots \beta_n \cdot X_n + \varepsilon$$

A kétváltozós formához hasonlóan itt is van tapasztalati vagy mintabeli regresszió:

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n$$

A standard lineáris regressziós modell paramétereinek meghatározáshoz egy négyes feltételrendszernek kell teljesülnie. A linearitás azt jelenti, hogy az eredményváltozó alakulását a magyarázó változók befolyásolják, vagyis azok lineáris összefüggése adja a várható értékét. A függetlenség kritériuma arra utal, hogy a magyarázó változóknak egymástól függetlennek kell lennie, vagyis magyarázzák az eredményváltozó alakulását, de nem befolyásolják egymást. Teljesül a változókra vonatkoztatott szórásállandóság, azaz a homoszkedaszticitás, valamint a normális eloszlás.

A paraméterek értelmezésére egy többváltozós modellben még nagyobb hangsúly kerül. A b_0 paraméter a függvény $x_1= x_2=\ldots= x_n = 0$ helyen felvett értéke, amely akkor értelmezhető, ha a magyarázó változók értéke lehet 0, illetve a paraméterre kiszámított érték eleme y értelmezési tartományának. A magyarázó változók együtthatóit ceteris paribus[11] értelmezzük, azaz a többi magyarázó változó értékét változatlannak feltételezve. Így például a b_1 paraméter megmutatja, hogy x_1 magyarázó változó értékét egy egységgel növelve, y eredményváltozó értéke a kap-

[11] A közgazdaságtanban gyakran használt kifejezés, melynek jelentése: minden egyéb tényező változatlansága mellett.

csolat irányának megfelelően átlagosan hány egységgel nő vagy csökken miközben a többi magyarázó változó értékét változatlannak tekintjük.

- A többváltozós regressziós modell esetében a magyarázó változók egymástól való függetlensége kritikus. Abban az esetben, ha a magyarázó változók közötti korreláció magas[12], a regressziós modell nem lesz megfelelő. Ez az ún. multikollinearitás, melyet a modellből ki kell szűrni, azaz olyan változókat kell bevonni, melyek befolyásolják az eredményváltozót, de egymás alakulását nem.

10.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

10.3.1 Összefoglalás

A mennyiségi ismérvek közötti kapcsolat vizsgálata kiemelten fontos a gyakorlatban. A korrelációszámítás során az ismérvek közötti kapcsolat létének, szorosságának és irányának meghatározására nyílik lehetőség. A legfontosabb mérőszámai a kovariancia, valamint arányskálán mérhető ismérveknél a lineáris korrelációs és determinációs együttható, ordinális mérési szintű változóknál pedig a Spearman és Kendall-féle rangkorrelációs együttható. Abban az esetben, ha kimutatható kapcsolat van a változók között, akkor a kapcsolat természetének jellemzésére regressziós modell írható fel. Az ismérvek között azonos és ellentétes irányú kapcsolat is lehet, melyre lineáris és nem lineáris függvény is illeszthető. A regresszió a változók között előzetesen meghatározott ok-okozati összefüggés leírására alkalmas matematikai eszköz. A megfigyelési értékekre illesztett mintabeli regresszió függvény segítségével megállapítható, hogy a magyarázó változó egységnyi növekedése milyen irányú és mekkora nagyságrendű változást idéz elő az eredményváltozóban. Gyakran előfordul, hogy egy jelenség alakulását nem csupán egy változó magyarázza, így többváltozós modelleket alkalmazunk.

10.3.2 Önellenőrző kérdések

1. Milyen ismérvek közötti kapcsolatot számszerűsít a korreláció?
2. Mit mutat meg a kovariancia?
3. Milyen mérési szintű ismérvek közötti kapcsolatot számszerűsít a rangkorreláció és melyek a fő mérőszámai?
4. Mit mutat meg a lineáris korrelációs és lineáris determinációs együttható?

[12] Az egymástól függő magyarázó változók kezelésére alkalmas módszer például a főkomponens-analízis.

5. Milyen típusú diagram segítségével ábrázolható két mennyiségi ismérvek közötti összefüggés?
6. Mi a különbség sokasági és tapasztalati regresszió függvény között?
7. Milyen típusú regresszió függvények illeszthetők az adatokra?
8. Hogyan írható fel egy kétváltozós lineáris regresszió függvény és mit jelentenek a paraméterei?
9. Mit mutat meg a rugalmassági együttható?
10. Mi a reziduum és a reziduális szórás? Mire használható?

10.3.3 Gyakorló tesztek

☸ Döntse el az alábbi állításokról, hogy igazak (I) vagy hamisak (H)!

A kovariancia -1 és 1 közötti értéket vehet fel, s előjele a kapcsolat irányát jelzi.	**H**
A Spearman féle mérőszám arányskálán mérhető ismérvek közötti kapcsolat erősségét határozza meg.	**H**
Az r^2 = 0,52 azt jelenti, hogy a magyarázó változó 0,52%-ban befolyásolja az eredményváltozó alakulását.	**H**
A többváltozós regressziós modellek feltétele, hogy a magyarázó és az eredményváltozó egymástól független legyen.	**H**
Az a regresszió függvény illeszkedik legjobban a tapasztalati értékekre, amelynél a reziduális szórás értéke a legnagyobb.	**H**
Hatványkitevős regresszió függvény esetén a rugalmassági együttható megegyezik a függvény b paraméterének értékével.	**I**
A lineáris determinációs együttható azt mutatja meg, hogy a magyarázó változó hány %-ban határozza meg az eredményváltozó alakulását.	**I**
Az y=325-12x regresszió függvény alapján megállapítható, hogy ha a magyarázó változó értékét egy egységgel növeljük, akkor az eredményváltozó 325 egységgel fog emelkedni.	**H**
Az r=0,894 azt jelenti, hogy a mennyiségi ismérvek között pozitív irányú, erős kapcsolat mutatható ki.	**I**
Az y=325-12x regresszió függvény alapján megállapítható, hogy ha a magyarázó változó értékét egy egységgel növeljük, akkor az eredményváltozó 12 egységgel fog csökkeni.	**I**

11. IDŐSORELEMZÉSI TECHNIKÁK

11.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

Az idősorok vizsgálatához változatos statisztikai eszköztár áll rendelkezésünkre. Az eddigiekben olyan módszereket ismertettünk, melyek segítségével egy adott megfigyelési időszak értékei elemezhetők. Az időbeli vizsgálatok célja azonban leggyakrabban az, hogy a rendelkezésre álló adatokból a jövőre vonatkozóan is lehessen információt nyújtani. Ebben a leckében olyan idősorelemzési technikák kerülnek bemutatásra, melyek segítségével a tapasztalati értékekben kimutatható tendencia alapján előrejelzés is készíthető. Az idősorelemzés dekompozíciós módszerének lényege az, hogy az idősor tényezőkre bontható, s ezen tényezők hatása egy-egy tapasztalati érték esetén számszerűsíthető. Determinisztikus idősorelemzés esetén az idősor komponensei egyértelműen különválaszthatók attól függően, hogy milyen időtávon érvényesül a hatásuk az értékek alakulásában. A sztochasztikus idősorelemzés viszont arra a feltételezésre épül, hogy egy adott időbeli érték hatással van a későbbi érték alakulására, vagyis az idősor elemei a korábbi értékek és a véletlen függvényében írhatók le.

A lecke célja, hogy megismertesse a hallgatót a determinisztikus idősorelemzés alapjaival, példákkal gazdagon illusztrálva a módszer lényegét. A hallgató ismerje az idősorok összetevőit, s azok tulajdonságait. A tényezők kapcsolódási formáinak és kiszámítási módjainak ismeretében képes legyen a kimutatott tendencia értelmezésére és a jövőre vonatkozó előrejelzés készítésére is. A hallgató tudja, hogy milyen paraméter alapján lehet dönteni arról, hogy melyik tartósan érvényesülő tendenciát leíró alapirányzat illeszkedik legjobban a tapasztalati értékekre.

11.2 TANANYAG

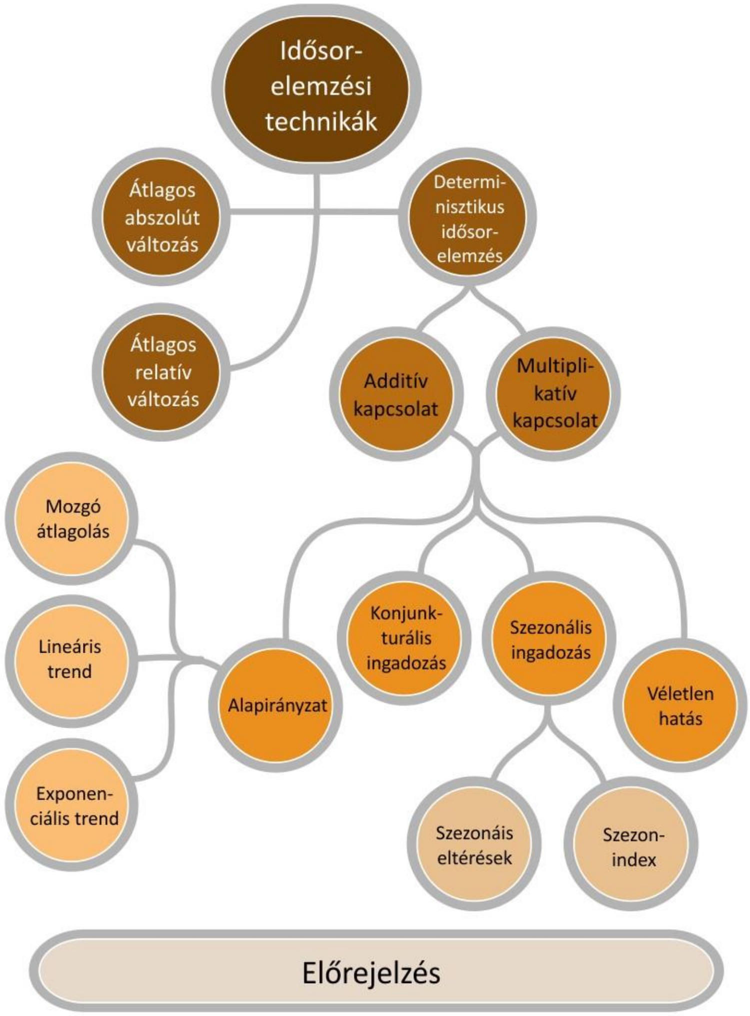
Az idősoros adatok elemzésének legegyszerűbb módszere a dinamikus viszonyszám, mely lehetőséget biztosít a különböző időpontbeli adatok összehasonlítására, a változások irányának és mértékének kimutatására két időbeli adat esetén. Heterogén sokaság esetén indexek számíthatók, melyek szintén két időszak adatainak összehasonlítására használhatók. Ezen módszerek azonban nem alkalmasak az idősorban lévő tendenciák kimutatására. A szabályszerűségek feltárására és a jövőbeni folyamatokra vonatkozó előrejelzés készítésének egyik módszere a determinisztikus idősorelemzés.

Az elemzés során az idősor két típusát különböztetjük meg, amely kiindulópontként szolgál a későbbi vizsgálatokhoz. A *tapasztalati vagy empirikus idősor*t a megfigyelt értékek alkotják, melyekre vonatkozóan az időbeli tendencia leírható. Az *elméleti idősor* az adott jelenség véges időszakra vonatkozó elemeiből áll, melynek nem ismert értékeire vonatkozóan a tapasztalati idősorban feltárt tendenciák alapján előrejelzés készíthető. A tapasztalati idősor tehát mintának, az elméleti idősor pedig sokaságnak tekinthető.

Az idősorelemzési technikák közül elsőként az átlagos abszolút és átlagos relatív változás mérőszámai kerülnek bemutatásra, melyek segítségével az időbeli adatok változásának iránya abszolút értékben és százalékos formában is számszerűsíthető. Ezen indikátorok értékei azonban csak akkor használhatók megbízhatóan előrejelzésre, ha az idősor elemei azonos irányban és megközelítőleg azonos mértékben változnak. A determinisztikus idősorelemzés során az adatokban lévő tendencia kimutatásához komponensekre bontjuk az idősort és ezen tényezők hatásait külön-külön számszerűsítjük, melynek segítségével a megfigyelt időszakon belül és a jövőre vonatkozóan is prognózis készíthető.

A feltárt tendenciák ismeretében a jövőre vonatkozó előrejelzést extrapolációnak, a vizsgált időszakon belüli, nem ismert időpontra történő becslést interpolációnak nevezzük.

Az előrejelzés kapcsán meg kell említeni, hogy egy adott idősorból elfogadható megbízhatósággal legfeljebb a megfigyelési időszak hosszának másfélszereséig lehet. Egyszerű példával illusztrálva egy 10 éves idősor alapján legfeljebb a következő 5 év adata becsülhető. Minél távolabbi időszakra készül az előrejelzés, annál kisebb a megbízhatósága.



38. ábra: Idősorelemzési technikák

11.2.1 Átlagos abszolút és relatív változás

Az idősorban lévő szabályszerűség egyszerűen jellemezhető az átlagos abszolút és átlagos relatív változás[13] mutatóival. Ezek a mérőszámok a megfigyelési értékek időszakról időszakra történő változásait átlagolják, melynek segítségével megállapítható, hogy a megfigyelt adatok milyen

[13] Az átlagos abszolút változást a fejlődés átlagos mértékének, míg az átlagos relatív változást a fejlődés átlagos ütemének is szokták nevezni.

irányban és átlagosan hogyan változtak. A vizsgált jelenség kapcsán egy tapasztalati idősor (minta) áll rendelkezésünkre, melyet a számítások során is figyelembe kell venni.

🕮 **Az átlagos abszolút változás az alapadatok mértékegységében fejezi ki a változás átlagos mértékét.**

Az átlagos abszolút változás kiszámításához az egymást követő időszakok értékeinek különbségét használjuk fel, s ezeknek a különbségeknek a számtani átlagát határozzuk meg annak figyelembevételével, hogy a megfigyelés nem teljeskörű (n-1).

$$\bar{d} = \frac{(y_2 - y_1) + (y_3 - y_2) + \cdots + (y_n - y_{n-1})}{n-1} = \frac{y_n - y_1}{n-1}$$

🕮 **Az átlagos relatív változás százalékos formában fejezi ki a változás átlagos mértékét.**

Az átlagos relatív változás kiszámításához az egymást követő időszakok értékeinek hányadosait használjuk fel, s ezeket átlagoljuk. Dinamikus (lánc) viszonyszámokról lévén szó, mértani átlagot számítunk.

$$\bar{l} = \sqrt[n-1]{\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdot \ldots \cdot \frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

A viszonyítás iránya mindkét mutató esetében kötött, a későbbi időszakot hasonlítjuk a korábbi időszakhoz. A képletek a zárójelek felbontásával, valamint a törtek szorzására vonatkozó ismeretek birtokában leegyszerűsíthetők, s azt láthatjuk, hogy a számításokhoz elegendő ismerni az utolsó és az első időszak adatát. Ebből következően az átlagos abszolút és átlagos relatív változás mutatók értéke érzékeny a két szélső értékre. Ha az első vagy utolsó időszak adata kiugró, akkor a mutatók által jelzett tendencia nem jellemzi megfelelően az adatok változását. Szintén torzítja az idősor tendenciáját a mutatók értéke akkor, ha a vizsgált időszakon belül az adatok jelentős ingadozása figyelhető meg.

☸ Egy vállalat által értékesített termékek számát ismerjük 2009 és 2013 között. *Határozza meg, hogy a vizsgált időszakban átlagosan hogyan változott az értékesítés!*

28. *Egy vállalat által értékesített termékek száma 2009 és 2013 között*

Év	Értékesített termékek száma (ezer db)
2009	240
2010	246
2011	257
2012	289
2013	320

A vizsgált időszakban, azaz az elmúlt 5 évben (n=5) meghatározható, hogy átlagosan mennyivel és milyen mértékben változott az értékesítés.

$$\bar{d} = \frac{y_n - y_1}{n-1} = \frac{320-240}{5-1} = \mathbf{20\ ezer\ db}$$

$$\bar{l} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[5-1]{\frac{320}{240}} = \mathbf{1{,}0745} \rightarrow +7{,}45\%$$

2009 és 2013 között évente átlagosan 20 000 db-bal, azaz 7,45%-kal nőtt a vállalat által értékesített termékek száma.

11.2.2 A determinisztikus idősorelemzés

A determinisztikus idősorelemzés során az idősort elemeire bontjuk, a komponensek attól függően, hogy milyen időtávon érvényesül a hatásuk az alábbiak szerint különíthetők el:

- alapirányzat vagy trend ($\hat{Y}$): hosszú távon érvényesülő tendencia
- konjunkturális ingadozás vagy ciklus (K): középtávú hatás
- szezonális ingadozás vagy szezonalitás (S): rövid távú hatás
- véletlen (V) hatása

Ezen tényezők additív és multiplikatív módon kapcsolódhatnak össze. Additív összefüggés esetén a tényleges érték (Y_t) a tényezők összegeként adódik, a komponensek az alapadatok mértékegységében vannak kifejezve. Multiplikatív kapcsolódás esetén a tényezők közül az alapirányzat az alapadatok mértékegységében van megadva, melyhez a többi komponens relatív módon kapcsolódik, mely matematikailag a trendhez viszonyított arányszámok együtthatós formában lévő szorzataként határozható meg.

Additív	Y_t	=	$\hat{Y}$	+	K	+	S	+	V
Multiplikatív	Y_t	=	$\hat{Y}$	·	K	·	S	·	V
			alapirányzat (trend)		konjunkturális ingadozás		szezonális ingadozás		véletlen hatás

39. ábra: Az idősor elemeinek kapcsolódása

Az **alapirányzat** (trend) az idősorban tartósan érvényesülő tendencia, mely meghatározza a változások fő irányát. A hosszú táv a vizsgált jelenséggel összefüggően értelmezhető, így hossza nem egyértelműen meghatározható.

A **konjunkturális ingadozás** (ciklus) az alapirányzat körüli tartósabb hullámzás, hosszabb időszakra vonatkozó (nem átmeneti) trend fölötti vagy trend alatti mozgás. Az ingadozás periodikus, kevésbé szabályos és nehezen kiszámítható. A középtávon érvényesülő tendenciáknak a gazdasági életben különösen nagy jelentőségük van, a legjelentősebb, közgazdászokról el is nevezett ciklusról, valamint jellemzőikről az alábbi linkeken olvashat:

19. Üzleti ciklusok:
    http://www.britannica.com/EBchecked/topic/86233/business-cycle
20. Kitchin ciklus (3-5 év): http://www.policonomics.com/joseph-kitchin/
21. Juglar ciklus (7-11 év): http://www.policonomics.com/clement-juglar/
22. Kuznets ciklus (15-25 év):
    http://www.dictionaryofeconomics.com/article?id=pde2008_K000045
23. Kondratyev ciklus (45-60):
    http://www.newworldencyclopedia.org/entry/Nikolai_Kondratiev

A **szezonális ingadozás** rövid időszakra (legfeljebb egy év) vonatkozó, szabályszerű hullámzás az alapirányzat körül. Az ingadozás ismétlődő időközönként, azonos hosszúságú periódusokban következik be és kiszámítható módon, ugyanabba az irányba téríti el az értékeket. A szezonalitás tipikus példája az idegenforgalom, ahol az elnevezés is utal a jelenségre. Az előszezonban és utószezonban rendre kevesebb vendég érkezik, míg a főszezonban több mint ami egy alapirányzatból becsülhető lenne. A szezonhatás a gazdaságban számos helyen jelentkezik, például az idényjellegű termékeknél vagy a mezőgazdaságban is.

A **véletlen ingadozás** rövid távon jelentkező, szabálytalan mozgás. Időbeli adatoknál kiemelten fontos, hogy számoljunk azzal, hogy az adatsor értékei egy nem várt esemény hatására eltérnek attól, amire a tendenciák alapján számítani lehetne. Ilyen hatás lehet például a nemzeti valuta leértékelődése, egy nem várt gazdasági esemény vagy válság.

Az idősor komponensei közül részletesen az alapirányzat és a szezonalitás meghatározásával foglalkozunk. A konjunktúra kutatás a gazdaságstatisztika speciális ága, a véletlen hatás pedig maradékelven számszerűsíthető a másik három komponens ismeretében.

Az alapirányzat meghatározása

Az alapirányzat meghatározásának célja az idősorban tartósan érvényesülő tendencia számszerűsítése. A változások fő irányát a mozgó (csúszó) átlagolás módszere az idősor értékeinek átlagolásával határozza meg. Az analitikus trendszámítás során annak a függvénynek az egyenletét adjuk meg, amely a legjobban közelíti az értékeket. A trendfüggvények közül a leggyakrabban alkalmazott lineáris, valamint exponenciális formákkal foglalkozunk.

Mozgóátlagolás

A mozgó (csúszó) átlagolás során a tartósan érvényesülő tendenciát a rendelkezésre álló adatok dinamikus átlaga adja. A számítás első lépése, hogy meghatározzuk a mozgó átlag tagszámát (k), vagyis az átlagolandó értékek számát. A tagszám megválasztásánál az idősor hullámzása és osztottsága mérvadó. Így például, ha negyedéves bontásban vannak különböző évekre vonatkozó adatok, akkor négytagú mozgó átlagot érdemes számítani. Az idegenforgalom esetében szezononkénti adatok állhatnak rendelkezésre, elő-, fő- és utószezon esetében háromtagú mozgó átlag számítása javasolt. A tagszám meghatározásánál célszerű arra figyelni, hogy a későbbiekben a szezonhatás is mérhető legyen. A mozgóátlagolás módszere kevésbé használható éves bontású, illetve osztatlan idősorok esetében, mert a tagszám megválasztása önkényes, így nem biztos, hogy megfelelően simítja ki az adatokat.

A tagszám ismeretében átlagolhatók az adatok úgy, hogy elsőként az első *k* számú adat egyszerű számtani átlagát határozzuk meg, melyet az átlagolt időszak közepére, vagyis (k+1)/2. helyre írunk be. A további értékek meghatározásához az előzőekben átlagolt értékek közül elhagyjuk az elsőt, helyette a (k+1). időszak kerül az átlagba, amely érték helyét is csúsztatjuk. Az átlagba kerülő értékeket így csúsztatjuk végig az adatsoron és írjuk a megfelelő helyekre. Az átlagolás során az idősor elején és végén nem számíthatók értékek, így az adatsor lerövidül. Ha páratlan tagszámú az adatsor, akkor k-1 taggal lesz rövidebb, míg ha páros tagszámú, akkor k taggal.

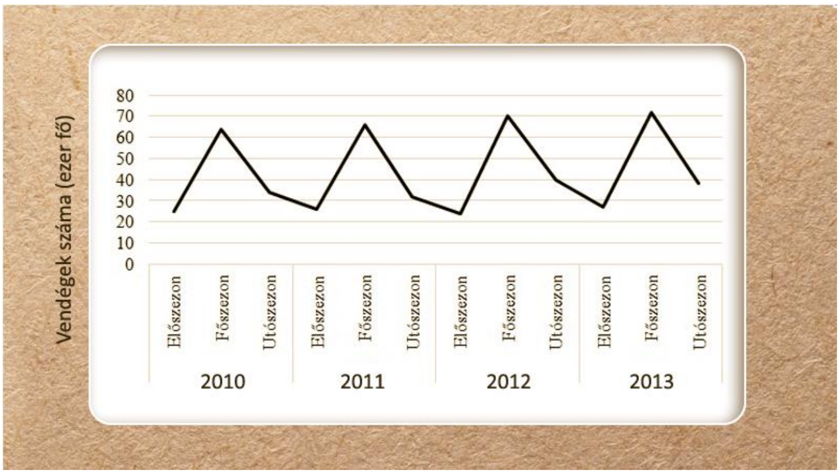
Páros tagszám esetén a mozgó átlagolással kapott átlagos értékek 2-2 érték közé esnek (négytagú mozgó átlag esetén például az első átlag helye a 2. és 3. elem között van – (4+1)/2=2,5), melyeket vissza kell igazítani az adatsorhoz, melyet a centrírozás módszerével végzünk el. Az eljárás lényege, hogy a 2-2 érték közé eső adatokat páronként átlagoljuk, szintén csúszó módszerrel. A számítás menete példán keresztül kerül bemutatásra.

☸ Egy szállodalánc vendégforgalmának alakulásáról az alábbi adatok állnak rendelkezésre. *Ábrázolja az idősort és határozza meg a mozgóátlagolású trend értékeit! Számítsa ki az átlagos abszolút és relatív változás nagyságát!*

29. *A szállodalánc vendégforgalmának szezonális alakulása 2010 és 2013 között*

Év	Idény	Vendégek száma (ezer fő)
2010	*Előszezon*	25
	Főszezon	64
	Utószezon	34
2011	*Előszezon*	26
	Főszezon	66
	Utószezon	32

Év	Idény	Vendégek száma (ezer fő)
2012	*Előszezon*	24
	Főszezon	70
	Utószezon	40
2013	*Előszezon*	27
	Főszezon	72
	Utószezon	38



40. ábra: A szálloda vendégforgalmának szezonális alakulása 2010 és 2013 között

Az ábrán jól látható, hogy az idősorban erőteljesen érvényesül a szezonalitás. A főszezonban voltak a legtöbben, melytől jelentősen lemaradt az utószezon, s még gyengébb az előszezon vendégszám tekintetében. A hosszú távú tendencia feltáráshoz ki kell simítani az idősort, melyre alkalmas a mozgóátlagolás. Elsőként meg kell határozni a mozgó átlag tagszámát, amely ebben az esetben a szezonális bontás miatt három lesz, így páratlan tagszámú adatsorral dolgozunk. A számításhoz 3-3 értéket átlagolunk, az átlagok helyileg a második értékkel kerülnek egy sorba (k+1)/2 = (3+1)/2= 2.

30. *Munkatábla a mozgó átlagok kiszámításához*

Év	*Idény*	*Vendégek száma (ezer fő)* y	*MOZGÓ ÁTLAG (MÁ)* $\hat{y}$
2010	Előszezon	25	-
	Főszezon	64	$\frac{25+64+34}{3}=41$
	Utószezon	34	$\frac{64+34+26}{3}=41{,}33$
2011	Előszezon	26	$\frac{34+26+66}{3}=42$
	Főszezon	66	$\frac{26+66+32}{3}=41{,}33$
	Utószezon	32	$\frac{66+32+24}{3}=40{,}67$
2012	Előszezon	24	$\frac{32+24+70}{3}=42$
	Főszezon	70	$\frac{24+70+40}{3}=44{,}67$
	Utószezon	40	$\frac{70+40+27}{3}=45{,}67$
2013	Előszezon	27	$\frac{40+27+72}{3}=46{,}33$
	Főszezon	72	$\frac{27+72+38}{3}=45{,}67$
	Utószezon	38	-
	Összesen	*518*	

Mozgó átlagolás esetén az átlagos abszolút és átlagos relatív változás meghatározásához a mozgó átlag értékeit és a lerövidült idősor (k-1-el, ez esetben 3-1, azaz kettővel rövidül) elemszámát vesszük alapul.

$$\bar{\boldsymbol{d}} = \frac{\hat{y}_n - \hat{y}_1}{n-1} = \frac{45{,}67 - 41}{10-1} = \mathbf{0{,}519}\text{ ezer fő}$$

Mozgó átlagolás alapján a vizsgált időszakban szezononként átlagosan 519 fővel nőtt a vendégek száma.

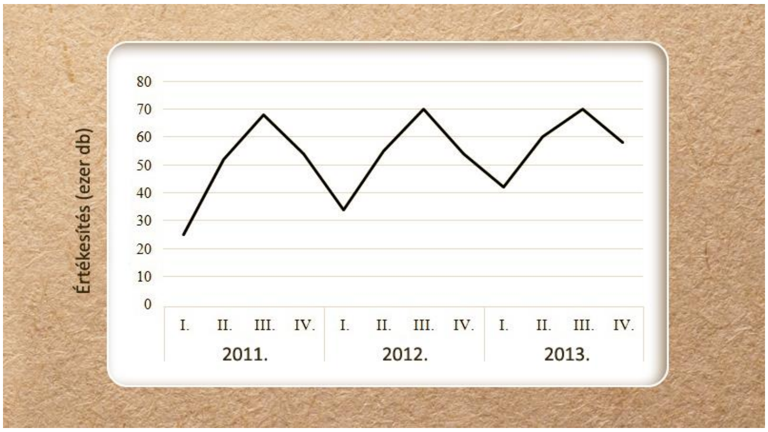
$$\bar{\boldsymbol{l}} = \sqrt[n-1]{\frac{\hat{y}_n}{\hat{y}_1}} = \sqrt[10-1]{\frac{45{,}67}{41}} = \mathbf{1{,}0121} \rightarrow +1{,}21\%$$

Mozgó átlagolás alapján a vizsgált időszakban szezononként átlagosan 1,21 %-kal nőtt a vendégek száma.

☸ Egy vállalat értékesítésére vonatkozóan az elmúlt 3 évben negyedéves bontásban az alábbi adatok állnak rendelkezésre. *Ábrázolja az idősort és határozza meg a mozgóátlagolású trend értékeit! Számítsa ki az átlagos abszolút és relatív változás nagyságát!*

31. *A vállalat értékesítésének alakulása 2011 és 2013 között negyedéves bontásban*

Időszak		*Értékesítés (ezer db)*	*Időszak*		*Értékesítés (ezer db)*	*Időszak*		*Értékesítés (ezer db)*
2011.	I.	25	2012.	I.	34	2013.	I.	42
	II.	52		II.	55		II.	60
	III.	68		III.	70		III.	70
	IV.	54		IV.	54		IV.	58



41. ábra: A vállalat értékesítésének alakulása 2011 és 2013 között negyedéves bontásban (09_11_K04)

Az ábrán jól kirajzolódnak ez egyes negyedévek tendenciái. Az első negyedévek a leggyengébbek, majd az év közepére erősödés figyelhető meg, a III. negyedévben a legmagasabb az értékesítés, ami ismét visszaesik az év vége felé. A negyedéves bontású adatsor esetében a látható hullámzás miatt a négytagú mozgó átlag választása ideális. Ebben az esetben tehát páros tagszámú lesz a mozgó átlag, így centrírozásra is szükség lesz. Előbb kiszámítjuk 4-4 érték átlagolásával a mozgó átlagokat, majd a kiszámított értékek páronkénti átlagolásával centrírozzuk az adatsort.

32. *Munkatábla a centrírozott mozgó átlagok kiszámításához*

		y_i (ezer db)	**Mozgó átlag**	$\hat{y}$ c.m.á centrírozott mozgó átlag
2011.	I.	25		–
			–	
	II.	52		–
			$\frac{25+52+68+54}{4}=49{,}75$	
	III.	68		$\frac{49{,}75+52}{2}=50{,}875$
			$\frac{52+68+54+34}{4}=52$	
	IV.	54		$\frac{52+52{,}75}{2}=52{,}375$
			$\frac{68+54+34+55}{4}=52{,}75$	
2012.	I.	34		$\frac{52{,}75+53{,}25}{2}=53$
			$\frac{54+34+55+70}{4}=53{,}25$	
	II.	55		$\frac{53{,}25+53{,}25}{2}=53{,}25$
			$\frac{34+55+70+54}{4}=53{,}25$	
	III.	70		$\frac{53{,}25+55{,}25}{2}=54{,}25$
			$\frac{55+70+54+42}{4}=55{,}25$	
	IV.	54		$\frac{55{,}25+56{,}5}{2}=55{,}875$
			$\frac{70+54+42+60}{4}=56{,}5$	
2013.	I.	42		$\frac{56{,}5+56{,}5}{2}=56{,}5$
			$\frac{54+42+60+70}{4}=56{,}5$	
	II.	60		$\frac{56{,}5+57{,}5}{2}=57$
			$\frac{42+60+70+58}{4}=57{,}5$	
	III.	70		–
			–	
	IV.	58		–

Az átlagos abszolút és átlagos relatív változás meghatározásához a centrírozott mozgó átlag értékeit és a lerövidült idősor (ez esetben a tagszámmal, azaz néggyel rövidül) elemszámát vesszük alapul.

$$\bar{\boldsymbol{d}} = \frac{\hat{y}_n - \hat{y}_1}{n-1} = \frac{57 - 50{,}875}{8-1} = \mathbf{0{,}875}\ \text{ezer db}$$

Centrírozott mozgó átlagolás alapján a vizsgált időszakban negyedévente átlagosan 875 db-bal nőtt az értékesített termékek száma.

$$\bar{\boldsymbol{l}} = \sqrt[n-1]{\frac{\hat{y}_n}{\hat{y}_1}} = \sqrt[8-1]{\frac{57}{50{,}875}} = \mathbf{1{,}0164} \rightarrow +1{,}64\%$$

Centrírozott mozgó átlagolás alapján a vizsgált időszakban negyedévente átlagosan 1,64 %-kal nőtt az értékesített termékek száma.

Analitikus trendszámítás: lineáris és exponenciális trend

Az analitikus trendszámítás során a megfigyelési értékekben lévő tartósan érvényesülő tendenciát függvénnyel írjuk le. A leggyakrabban lineáris és exponenciális trendet illesztünk az adatokra. A trendfüggvény kizárólag formailag hasonlít a regresszió függvényhez, az idő és a megfigyelt jelenség között nem feltétlenül van ok-okozati kapcsolat[14]. A trend az időbeli adatok szabályszerűségét tehát függvény segítségével tárja fel, a változások irányára és mértékére vonatkozóan ad információt.

A trendfüggvények meghatározásához elsőként az időszakokhoz numerikus értékeket rendelünk, amely lehetővé teszi, hogy a továbbiakban mennyiségi ismérvek elemzésére használatos módszereket idősorok esetében is alkalmazzuk. Ezek az ún. t értékek[15], melyeket kétféle módon rendelhetünk az időszakokhoz.

A **t=1,2..n** vagy monoton t módszer a t értékeket 1-el kezdődően, egyenlő lépésközönként, növekvő sorrendben rendeli az időszakokhoz, vagyis az első időszak adata 1, a második időszak adata 2 és így tovább.

A **∑t=0** módszer az időszakokhoz úgy rendeli a t értékeket, hogy azok összege 0 legyen. Ez a gyakorlatban úgy valósítható meg, hogy az idősor (n+1)/2. eleme lesz 0, az időben visszahaladva egyenlő lépésközönként negatív, előrehaladva pedig pozitív értékeket rendel az időszakokhoz. Például egy 9 éves idősor esetén az 5. év adata lesz 0, a 4. év adata -1, a 6. év adata 1. Páros elemszámú idősor esetén az (n+1)/2. elem két érték közé

[14] A függvények hasonlósága miatt gyakran nevezik a trendfüggvényt speciális regressziónak, amelynek magyarázó változója az idő. A regressziós összefüggés mintájára ez azonban azt jelentené, hogy az adatok változását minden esetben az idő determinálja, vagyis az idő múlása miatt változnak az adatok. A trend számítása során azonban nem az időt, hanem az időszakokhoz rendelt numerikus értékeket vesszük alapul, vagyis az időtényező csak közvetetten jelenik meg,

[15] A t értékek elnevezése az angol time, azaz idő kifejezésből ered, utalva arra, hogy a számértékek időszakokat takarnak.

esik. Ebben az esetben a 0 értéket nem veszi fel egyik időszak sem, a korábbi -1, a későbbi 1 értéket kap, s mivel a kettő közötti lépésköz kettő, ezért az értékek időszakról időszakra kettesével változnak. Például egy 10 éves idősor esetén az 5,5. elem lenne 0, ezt elkerülve az 5. elem -1, a 6. elem 1 értéket kap, míg a 4. elem -3 értéket, a 7. elem 3 értéket.

33. *Példa a t értékek ∑t = 0 módszer alapján történő hozzárendelésére*

Időszak	Hozzárendelt t érték
2010	-1
2011	0
2012	1

Időszak	Hozzárendelt t érték
2009	-3
2010	-1
2011	1
2012	3

A **lineáris trend** az összetartozó értékpárokat közelítő egyenes. Az elméleti idősorra vonatkoztatható sokasági trendfüggvény:

$$\boldsymbol{Y = \beta_0 + \beta_1 t + \varepsilon}$$

A β_0 és β_1 paraméterek értelmezése eltér attól függően, hogy az adatokra t=1,2…n vagy ∑t=0 módszerrel illesztettük a trendet. ε a véletlen hatása, melyet becsült érték lévén figyelembe kell venni.

A megfigyelt értékekre illeszthető tapasztalati trendfüggvény:

$$\boldsymbol{\hat{y} = b_0 + b_1 \cdot t}$$

A trend paraméterei közül β_1 és b_1 előjele kiemelten fontos, mert az mutatja meg a megfigyelési értékek változásának irányát.

A lineáris trendfüggvény paramétereinek meghatározása

t = 1,2 … n (monoton t) módszer esetén a trendfüggvény egyenlete normál egyenletek segítségével fejezhető ki. A számítások során a megfigyelési időszakok száma (n) is fontos. A két ismeretlenes egyenletrendszer megoldása a függvény b_0 és b_1 paraméterének értéke.

$$\boldsymbol{\sum y = b_0 \cdot n + b_1 \cdot \sum t}$$
$$\boldsymbol{\sum ty = b_0 \cdot \sum t + b_1 \cdot \sum t^2}$$

A t = 1,2 … n (monoton t) módszer segítségével felírt lineáris trendfüggvény b_0 paramétere a megfigyelési időszakot közvetlenül megelőző időszak értéke, azaz a függvény t=0 helyen felvett értéke, b_1 paramétere pedig az időszakról időszakra történő átlagos abszolút változást fejezi ki.

∑t = 0 módszer esetén a trendfüggvény egyenlete képletek segítségével adható meg.

$$\boldsymbol{b_0 = \frac{\sum y}{n}} \qquad \boldsymbol{b_1 = \frac{\sum ty}{\sum t^2}}$$

A $\sum t = 0$ módszer segítségével felírt lineáris trendfüggvény b_0 paramétere a megfigyelési időszak (n+1)/2. elemének értéke, a megfigyelési értékek számtani átlaga, b_1 paramétere pedig az időszakról időszakra történő átlagos abszolút változást fejezi ki. Páros számú adatsor esetén, ahol a t értékek kettesével nőnek, az időszakról időszakra történő változás $2 \cdot b_1$ mértékben értelmezhető.

❁ Egy vállalat értékesítésére vonatkozóan az elmúlt 3 évben negyedéves bontásban állnak rendelkezésre adatok (ezer db). *Határozza meg a lineáris trend egyenletét t=1,2… n és ∑t=0 módszerrel is!*

34. *Munkatábla a lineáris trend meghatározásához*

		y_i (ezer db)	*t = 1,2 …n*			**∑t = 0**		
			t	***t · y***	t^2	*t*	***t · y***	t^2
2011.	I.	25	**1**	25	1	**-11**	-275	121
	II.	52	**2**	104	4	**-9**	-468	81
	III.	68	**3**	204	9	**-7**	-476	49
	IV.	54	**4**	216	16	**-5**	-270	25
2012.	I.	34	**5**	170	25	**-3**	-102	9
	II.	55	**6**	330	36	**-1**	-55	1
	III.	70	**7**	490	49	**1**	70	1
	IV.	54	**8**	432	64	**3**	162	9
2013.	I.	42	**9**	378	81	**5**	210	25
	II.	60	**10**	600	100	**7**	420	49
	III.	70	**11**	770	121	**9**	630	81
	IV.	58	**12**	696	144	**11**	638	121
	∑	642	**78**	4415	650	**0**	484	572

A lineáris trendfüggvény egyenletének meghatározása
***t=1,2…n** módszerrel - normál egyenletek*

$$\sum y \;=\; b_0 \cdot n + b_1 \cdot \sum t$$
$$\sum ty = b_0 \cdot \sum t + b_1 \cdot \sum t^2$$

$$642 \;=\; b_0 \cdot 12 + b_1 \cdot 78 \qquad \rightarrow \qquad b_0 = \frac{642 - b_1 \cdot 78}{12}$$
$$4415 = b_0 \cdot 78 + b_1 \cdot 650$$

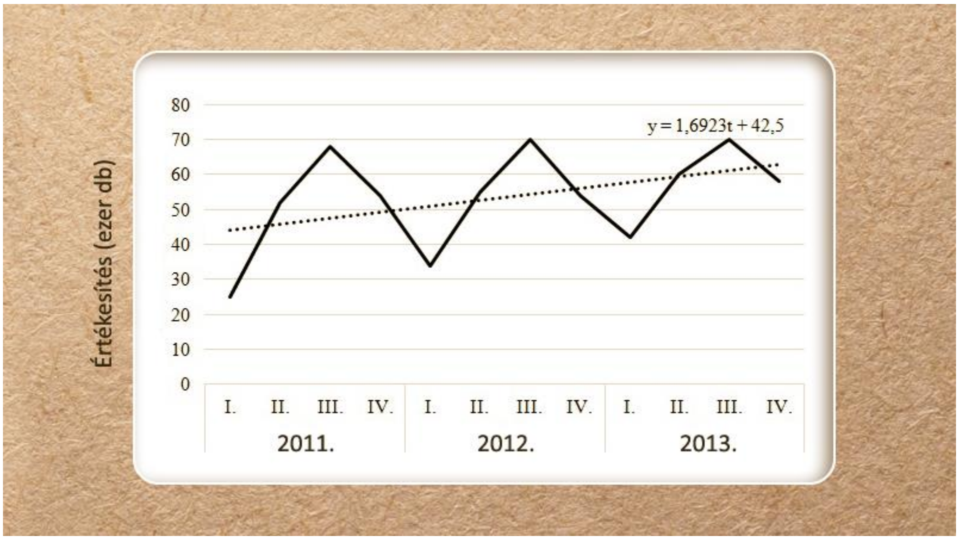
MEGOLDÁS: $4415 = \frac{642 - b_1 \cdot 78}{12} \cdot 78 + b_1 \cdot 650$ egyismeretlenes egyenlet

$$4415 = (642 - b_1 \cdot 78) \cdot 6{,}5 + b_1 \cdot 650 \rightarrow 4415 = 4173 - b_1 \cdot 507 + b_1 \cdot 650$$

$$242 = b_1 \cdot 143 \rightarrow \boldsymbol{b_1 = 1,6923} \rightarrow \boldsymbol{b_0} = \frac{642-1,6923\cdot 78}{12} = \mathbf{42,5}$$

A lineáris trendfüggvény (t=1,2…n) egyenlete: $\hat{\mathbf{y}} = \mathbf{42,5 + 1,6923 \cdot t}$

A lineáris trendfüggvény szerint 2010. IV. negyedévében 42,5 ezer db terméket értékesített a vállalat, s a vizsgált időszakban az értékesítés negyedévente átlagosan 1,6923 ezer db-bal nőtt.



42. ábra: A vállalat értékesítésének alakulása 2011 és 2013 között negyedéves bontásban és az illeszkedő lineáris trendfüggvény (t=1,2 …n)

A lineáris trendfüggvény egyenletének meghatározása
***Σt* = 0** *módszerrel - képletek*

$$\boldsymbol{b_0} = \frac{\sum \boldsymbol{y}}{\boldsymbol{n}} = \frac{642}{12} = 53{,}5 \qquad \boldsymbol{b_1} = \frac{\sum \boldsymbol{ty}}{\sum \boldsymbol{t^2}} = \frac{484}{572} = 0{,}84615$$

A lineáris trendfüggvény (∑t = 0) egyenlete: $\hat{\mathbf{y}} = \mathbf{53,5 + 0,84615 \cdot t}$

A lineáris trendfüggvény szerint 2012. II. és III. negyedéve között 53,5 ezer db terméket értékesítettek, s a vizsgált időszakban az értékesítés negyedévente átlagosan 1,6923 ezer db-bal nő (mivel a t értékek kettesével nőnek, ezért 2· 0,84615 = 1,6923).

Az **exponenciális trend** az összetartozó értékpárokat közelítő gyorsuló függvény. Az elméleti idősorra vonatkoztatható sokasági trendfüggvény:

$$\boldsymbol{Y = \alpha \cdot \beta^t \cdot \varepsilon}$$

Az α és β paraméterek értelmezése ez esetben is eltérő t=1,2…n és $\sum$t=0 módszer esetén, a véletlen hatását (ε) sokaságra vonatkoztatott függvény esetén itt is figyelembe kell venni.

A megfigyelt értékekre illeszthető tapasztalati exponenciális trendfüggvény

$$\boldsymbol{\hat{y} = a \cdot b^t}$$

Az exponenciális trendfüggvény paramétereinek meghatározása

A paraméterek meghatározásához szükség van a függvény logaritmikus átalakításához:

$$\mathrm{lg}\,\hat{\mathrm{y}} = \mathrm{lg}\,\mathrm{a} + \mathrm{lg}\,\mathrm{b} \cdot \mathrm{t}$$

A logaritmizált függvény alapján lineáris összefüggés írható fel, ha *lga* és *lgb* értékek helyett új változókat vezetünk be.

Legyen $lg\,a = b_0$ és $lg\,b = b_1$

Ez esetben az alábbi összefüggést kapjuk:

$$\mathrm{lg}\,\hat{\mathrm{y}} = \mathrm{b}_0 + \mathrm{b}_1 \cdot \mathrm{t}$$

Ezen egyenlethez a lineáris trendfüggvénynél ismertetett módon meghatározhatók a normál egyenletek és a képletek.

t = 1,2 … n (monoton t) módszer esetén az exponenciális trendfüggvény egyenlete is normál egyenletek segítségével fejezhető ki. A két ismeretlenes egyenletrendszer megoldása a függvény b_0 és b_1 paraméterének értéke, melyeket vissza kell alakítani a és b paraméterekké.

$$\boldsymbol{\sum \mathrm{lg}\, y = b_0 \cdot n + b_1 \cdot \sum t}$$
$$\boldsymbol{\sum t \cdot \mathrm{lg}\, y = b_0 \cdot \sum t + b_1 \cdot \sum t^2}$$

A paraméterek visszaalakítása:

$$b_1 = lg\,b \rightarrow \boldsymbol{b} = 10^{b_1}$$

$$b_0 = lg\,a \rightarrow \mathbf{a} = 10^{b_0}$$

A t = 1,2 … n (monoton t) módszer segítségével felírt lineáris trendfüggvény *a* paramétere a megfigyelési időszakot közvetlenül megelőző időszak értéke, azaz a függvény t=0 helyen felvett értéke, *b* paramétere pedig az időszakról időszakra történő átlagos relatív változást fejezi ki.

$\sum$t = 0 módszer esetén a trendfüggvény egyenlete képletek segítségével adható meg.

$$\boldsymbol{b_0 = \frac{\sum lg\, y}{n}} \qquad \boldsymbol{b_1 = \frac{\sum t \cdot lgy}{\sum t^2}}$$

A $\sum t = 0$ módszer segítségével felírt exponenciális trendfüggvény *a* paramétere a megfigyelési időszak (n+1)/2. elemének értéke, a megfigyelési értékek mértani átlaga, *b* paramétere pedig az időszakról időszakra történő átlagos relatív változást fejezi ki. Páros számú adatsor esetén, ahol a t értékek kettesével nőnek az időszakról időszakra történő változás b^2 mértékben értelmezhető.

❅ Egy vállalat értékesítésére vonatkozóan az elmúlt 3 évben negyedéves bontásban állnak rendelkezésre adatok. *Határozza meg az exponenciális trend egyenletét t=1,2… n és ∑t=0 módszerrel is*!

35. *Munkatábla az exponenciális trend meghatározásához*

		y_i (ezer db)	lg y	***t* = 1,2 …*n***			**∑t = 0**		
				t	$t \cdot lgy$	t^2	***t***	$t \cdot lgy$	t^2
2011.	I.	25	1,3979	***1***	1,3979	*1*	***-11***	-15,3773	*121*
	II.	52	1,7160	***2***	3,4320	*4*	***-9***	-15,4440	*81*
	III.	68	1,8325	***3***	5,4975	*9*	***-7***	-12,8276	*49*
	IV.	54	1,7324	***4***	6,9296	*16*	***-5***	-8,6620	*25*
2012.	I.	34	1,5315	***5***	7,6574	*25*	***-3***	-4,5944	*9*
	II.	55	1,7404	***6***	10,4422	*36*	***-1***	-1,7404	*1*
	III.	70	1,8451	***7***	12,9157	*49*	***1***	1,8451	*1*
	IV.	54	1,7324	***8***	13,8592	*64*	***3***	5,1972	*9*
2013.	I.	42	1,6232	***9***	14,6092	*81*	***5***	8,1162	*25*
	II.	60	1,7782	***10***	17,7815	*100*	***7***	12,4471	*49*
	III.	70	1,8451	***11***	20,2961	*121*	***9***	16,6059	*81*
	IV.	58	1,7634	***12***	21,1611	*144*	***11***	19,3977	*121*
	∑	*642*	20,5381	***78***	135,9794	*650*	***0***	4,9635	*572*

Az exponenciális trendfüggvény egyenletének meghatározása **t=1,2…n** módszerrel - normál egyenletek

$$\sum \lg y = b_0 \cdot n + b_1 \cdot \sum t$$
$$\sum t \cdot \lg y = b_0 \cdot \sum t + b_1 \cdot \sum t^2$$

$$20{,}5381 = b_0 \cdot 12 + b_1 \cdot 78 \rightarrow b_0 = \frac{20{,}5381 - b_1 \cdot 78}{12}$$
$$135{,}9794 = b_0 \cdot 78 + b_1 \cdot 650$$

$$135{,}9794 = \frac{20{,}5381 - b_1 \cdot 78}{12} \cdot 78 + b_1 \cdot 650$$

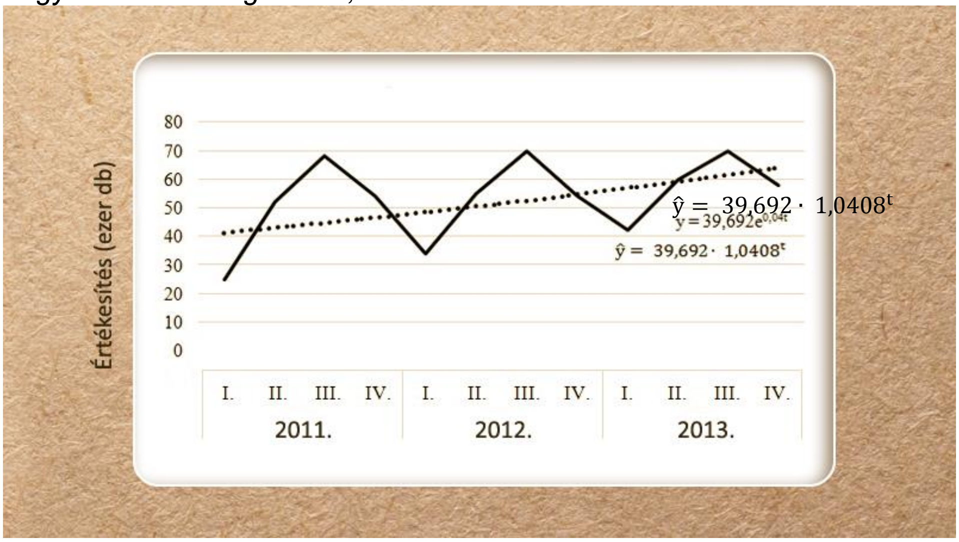
$$b_1 = 0{,}01735 = \lg b \rightarrow \boldsymbol{b} = 10^{0{,}01735} = \mathbf{1{,}0408}$$

$\downarrow$

$$b_0 = \frac{20{,}5381 - 0{,}01735 \cdot 78}{12} = 1{,}5987 = lg\ a \ \rightarrow \mathbf{a} = 10^{1{,}5987} = \mathbf{39{,}692}$$

Az exponenciális trendfüggvény (t=1,2…n) egyenlete: $\boldsymbol{\hat{y} = 39{,}692 \cdot 1{,}0408^t}$

Az exponenciális trendfüggvény szerint 2010. IV. negyedévében 39,962 ezer db terméket értékesítettek, a vizsgált időszakban az értékesítés negyedévente átlagosan 4,08 %-kal nőtt.



*43. ábra: A vállalat értékesítésének alakulása 2011 és 2013 között negyedéves bontásban és az adatokra illeszkedő exponenciális trendfüggvény (t=1,2 …n) (09_*

A lineáris trendfüggvény egyenletének meghatározása $\sum \mathbf{t} = \mathbf{0}$ módszerrel

$$\boldsymbol{b_0} = \frac{\sum \boldsymbol{lg\ y}}{\boldsymbol{n}} = \frac{20{,}5381}{12} = 1{,}7115 \rightarrow \mathbf{a} = 10^{1{,}7115} = \mathbf{51{,}46}$$

$$\boldsymbol{b_1} = \frac{\sum \boldsymbol{t \cdot lgy}}{\sum \boldsymbol{t^2}} = \frac{4{,}9635}{572} = 0{,}00868 \ \rightarrow \boldsymbol{b} = 10^{0{,}00868} = \mathbf{1{,}0202}$$

Az exponenciális trendfüggvény szerint 2012. II. és III. negyedéve között 51,46 ezer db terméket értékesítettek, a vizsgált időszakban az értékesítés negyedévente átlagosan 4, 08%-kal ($1{,}0202^2$) nőtt.

A szezonalitás

A szezonhatás rövid távon érvényesülő periodikus hullámzás. Az idősor elemeinek kapcsolódásától függően szezonális eltéréseket (additív) és szezonindexeket (multiplikatív) számolunk. A szezonális eltéréseket a tényleges és a trendből becsült értékek páronkénti különbségéből (reziduum), a szezonindexeket pedig az értékek páronkénti hányadosaiból határozzuk meg. Az idősor elemeiből kiszűrve a szezonális ingadozást, a szezonhatások ki kell, hogy egyenlítsék egymást.

Szezonális eltérés esetén az ismétlődő periódusokra vonatkozó különbségeknek a számtani átlagát határozzuk meg, melyet nyers szezonális eltérésnek nevezzük. Az átlagokat összeadva 0-t kell kapnunk, így ha a nyers szezonális eltérések összege nem 0, akkor az értékeket korrigálni kell. A korrekciós tényező (k) a nyers szezonális eltérések egyszerű számtani átlaga. A korrigált szezonális eltéréseket úgy kapjuk, hogy a nyers szezonális eltérések értékét csökkentjük a korrekciós tényező nagyságával. Lineáris trend és additív szezonalitás esetén a nyers és korrigált szezonális eltérések megegyeznek.

Szezonindexek esetén az ismétlődő periódusokra vonatkozó hányadosoknak a mértani átlagát határozzuk meg, melyet nyers szezonindexnek nevezzük. Az átlagok szorzata 1-et kell, hogy legyen, így ha a nyers szezonindexek szorzata nem 1, akkor az értékeket korrigálni kell. A korrekciós tényező a nyers szezonindexek mértani átlaga. A korrigált szezonindexek értékét úgy kapjuk, hogy a nyers szezonindexek értékét osztjuk a korrekciós tényező nagyságával. Exponenciális trend és multiplikatív szezonalitás esetén a nyers és korrigált szezonindexek megegyeznek.

- ☸ Egy vállalat értékesítésére vonatkozóan ismerjük 2011 és 2013 között negyedéves bontásban az adatokat, valamint a centrírozott mozgó átlag értékeit. *Határozza meg a szezonalitás mértékét!*

36. *A tényleges és a mozgóátlagolású trend értékeinek páronkénti hányadosai és különbségei*

		y_i (ezer db)	Mozgó átlag	$\hat{y}\,c.m.á$ centrírozott mozgó átlag	$\frac{y_i}{\hat{y}\,c.m.á}$	$e = y_i - \hat{y}\,c.m.á$ (reziduum)	$(y_i - \hat{y}\,c.m.á)^2$ e^2
2011.	I.	25		—	—	—	—
			—				
	II.	52		—	—	—	—
			49,75				
	III.	68		50,875	1,3366	17,125	293,2656
			52				
	IV.	54		52,375	1,0310	1,625	2,6406
			52,75				
2012.	I.	34		53	0,6415	-19	361
			53,25				
	II.	55		53,25	1,0329	1,75	3,0625
			53,25				
	III.	70		54,25	1,2903	15,75	248,0625

		$\mathbf{y_i}$ (ezer db)	Mozgó átlag	$\hat{y}\,c.m.$á centrírozott mozgó átlag	$\dfrac{y_i}{\hat{y}\,c.m.á}$	$e = y_i - \hat{y}\,c.m.á$ (reziduum)	$(y_i - \hat{y}\,c.m.á)^2$ e^2
			55,25				
	IV.	54		55,875	0,9664	-1,875	3,5156
			56,5				
2013	I.	42		56,5	0,7434	-14,5	210,25
			56,5				
	II.	60		57	1,0526	3	9
			57,5				
	III.	70		—	—	—	—
			—				
	IV.	58		—	—	—	—
						$\sum e^2$	1130,7968

Additív szezonalitás esetén az értékek páronkénti különbségeit ($y_i - \hat{y}\,c.m.á$) használjuk fel.

37. *Munkatábla a szezonális eltérések meghatározásához mozgóátlagolás esetén*

	I.	II.	III.	IV.	
2011.	—	—	17,125	1,625	
2012.	-19	1,75	15,75	-1,875	
2013.	-14,5	3	—	—	*Összesen*
Nyers szezonális eltérés (az eltérések számtani átlaga)	-16,75	2,375	16,4375	-0,125	1,9375
Korrigált szezonális eltérés: S_i	*-17,234*	*1,891*	*15,9535*	*-0,609*	***0***

Korrekciós tényező: $k = \frac{1,9375}{4} = 0,484 \rightarrow \text{Korrigált} = \text{Nyers} - k$

A tényleges értékesítés az I. negyedévekben átlagosan 17,234 ezer db-bal, a IV. negyedévekben 0,609 db-bal alacsonyabb, a II. negyedévekben átlagosan 1,891 ezer db-bal, a III. negyedévekben 15,9535 ezer db-bal több, mint a centrírozott mozgó átlagból becsült érték.

Multiplikatív szezonalitás esetén az értékek páronkénti hányadosait[16] használjuk fel.

[16] A számítások pontossága érdekében ajánlott legalább négy tizedesjegy pontossággal megadni az értékeket.

38. *Munkatábla a szezonindexek meghatározásához mozgóátlagolás esetén*

		I.	II.	III.	IV.	
$\frac{y_i}{\hat{y}\,c.m.á}$	2010.	—	—	1,3366	1,0310	
	2011.	0,6415	1,0329	1,2903	0,9664	
	2012.	0,7434	1,0526	—	—	*Szorzat*
Nyers szezonindex (változások mértani átlaga)		0,6906	1,0427	1,3132	0,9982	0,9439
Korgált szezonindex: $\boldsymbol{S_i}$		*0,7006*	*1,0578*	*1,3323*	*1,0127*	***1***

Korrekciós tényező: $k = \sqrt[4]{0{,}9439} = 0{,}9857 \rightarrow \text{Korrigált} = \frac{\text{Nyers}}{\text{k}}$

A tényleges értékesítés az I. negyedévekben átlagosan 29,94%-kal alacsonyabb, a II. negyedévekben átlagosan 5,78%-kal, a III. negyedévekben 33,23%-kal, a IV. negyedévekben 1,27%-kal magasabb több, mint a centrírozott mozgó átlagból becsült érték.

❆ Egy vállalat értékesítésére vonatkozóan ismerjük 2011 és 2013 között negyedéves bontásban az adatokat, valamint a lineáris trend egyenletét $\hat{y} = 42{,}5 + 1{,}6923 \cdot t$ (t=1,2,…n). *Határozza meg az egyes negyedévekben a trendből becsült értékeket, valamint a szezonalitás mértékét!*

39. *A tényleges és a lineáris trendből becsült értékek páronkénti különbségei és hányadosai*

		y$\mathbf{y_i}$ (ezer db)	t	$\hat{y} = 42{,}5 + 1{,}6923 \cdot t$	$e = y_i - \hat{y}$	$(y_i - \hat{y})^2 = e^2$	$\frac{y_i}{\hat{y}}$
2011.	I.	25	1	*44,19*	*-19,19*	*368,344*	*0,5657*
	II.	52	2	*45,88*	*6,12*	*37,398*	*1,1333*
	III.	68	3	*47,58*	*20,42*	*417,103*	*1,4293*
	IV.	54	4	*49,27*	*4,73*	*22,380*	*1,0960*
2012.	I.	34	5	*50,96*	*-16,96*	*287,692*	*0,6672*
	II.	55	6	*52,65*	*2,35*	*5,505*	*1,0446*
	III.	70	7	*54,35*	*15,65*	*245,045*	*1,2880*
	IV.	54	8	*56,04*	*-2,04*	*4,155*	*0,9636*
2013.	I.	42	9	*57,73*	*-15,73*	*247,455*	*0,7275*
	II.	60	10	*59,42*	*0,58*	*0,333*	*1,0097*
	III.	70	11	*61,12*	*8,88*	*78,938*	*1,1454*
	IV.	58	12	*62,81*	*-4,81*	*23,113*	*0,9235*
	Σ	*642*		*642*		*1737,462*	

Additív szezonalitás

40. *Munkatábla szezonális eltérések meghatározásához lineáris trend esetén*

	I.	II.	III.	IV.	
2011.	-19,19	6,12	20,42	4,73	
2012.	-16,96	2,35	15,65	-2,04	
2013.	-15,73	0,58	8,88	-4,81	*Összesen*
Nyers = korrigált szezonális eltérés: S_i	*-17,29*	*3,02*	*14,98*	*-0,71*	***0***

Lineáris trend és additív szezonalitás esetén a nyers szezonális eltérések összege 0, vagyis a nyers és korrigált szezonális eltérések megegyeznek.

A tényleges értékesítés az I. negyedévekben átlagosan 17,29 ezer db-bal, a IV. negyedévekben 0,71 ezer db-bal alacsonyabb, a II. negyedévekben átlagosan 3,02 ezer db-bal, a III. negyedévekben 14,98 ezer db-bal több, mint a lineáris trendből becsült érték.

Multiplikatív szezonalitás

41. *Munkatábla a szezonindexek meghatározásához lineáris trend esetén*

	I.	II.	III.	IV.	
2010.	0,5657	1,333	1,4293	1,096	
2011.	0,6672	1,0446	1,288	0,9636	
2012.	0,7275	1,0097	1,1454	0,9235	*Szorzat*
Nyers szezonindex	0,6500	1,1203	1,2823	0,9917	0,9260
Korrigált szezonindex: S_i	*0,6626*	*1,1420*	*1,3072*	*1,0109*	1

Korrekciós tényező: $k = \sqrt[4]{0,926} = 0,981 \rightarrow \text{Korrigált} = \frac{\text{Nyers}}{\text{k}}$

A tényleges értékesítés az I. negyedévekben átlagosan 33,74%-kal alacsonyabb, a II. negyedévekben átlagosan 14,2%-kal, a III. negyedévekben 30,72%-kal, a IV. negyedévekben 1,09%-kal magasabb, mint a lineáris trendből becsült érték.

☸ Egy vállalat értékesítésére vonatkozóan ismerjük 2011 és 2013 között negyedéves bontásban az adatokat, valamint az exponenciális trend egyenletét $\hat{y} = 39,692 \cdot 1,0408^t$ (t=1,2,…n). *Határozza meg az egyes negyedévekben a trendből becsült értékeket, valamint a szezonalitás mértékét!*

42. *A tényleges és az exponenciális trendből becsült értékek páronkénti különbségei és hányadosai*

		y_i (ezer db)	t	$\hat{y} = 39{,}692 \cdot 1{,}0408^t$	$e = y_i - \hat{y}$	$(y_i - \hat{y})^2 = e^2$	$\frac{y_i}{\hat{y}}$
2011.	I.	25	1	41,311	-16,311	266,063	0,6052
	II.	52	2	42,997	9,003	81,055	1,2094
	III.	68	3	44,751	23,249	540,506	1,5195
	IV.	54	4	46,577	7,423	55,100	1,1594
2012.	I.	34	5	48,477	-14,477	209,595	0,7014
	II.	55	6	50,455	4,545	20,654	1,0901
	III.	70	7	52,514	17,486	305,765	1,3330
	IV.	54	8	54,656	-0,656	0,431	0,9880
2013.	I.	42	9	56,886	-14,886	221,605	0,7383
	II.	60	10	59,207	0,793	0,628	1,0134
	III.	70	11	61,623	8,377	70,174	1,1359
	IV.	58	12	64,137	-6,137	37,666	0,9043
	Σ	642				1809,242	

Additív szezonalitás

43. *Munkatábla a szezonális eltérések meghatározásához exponenciális trend esetén*

	I.	II.	III.	IV.	
2011.	-16,311	9,003	23,249	7,423	
2012.	-14,477	4,545	17,486	-0,656	
2013.	-14,886	0,793	8,377	-6,137	*Összesen*
Nyers szezonális eltérés	*-15,225*	*4,780*	*16,371*	*0,210*	*6,136*
Korrigált szezonális eltérés: S_i	*-16,759*	*3,246*	*14,837*	*-1,324*	***0***

Korrekciós tényező: $k = \frac{6{,}136}{4} = 1{,}534 \rightarrow \text{Korrigált} = \text{Nyers} - k$

A tényleges értékesítés az I. negyedévekben átlagosan 16,759 ezer db-bal, a IV. negyedévekben 1,324 ezer db-bal alacsonyabb, a II. negyedévekben átlagosan 3,246 ezer db-bal, a III. negyedévekben 14,837 ezer db-bal több, mint az exponenciális trendből becsült érték.

Multiplikatív szezonalitás

44. *Munkatábla a szezonindexek meghatározásához exponenciális trend esetén*

	I.	II.	III.	IV.	
2011.	0,6052	1,2094	1,5195	1,1594	
2012.	0,7014	1,0901	1,333	0,988	
2013.	0,7383	1,0134	1,1359	0,9043	*Szorzat*
Nyers=Korrigált szezonindex: $\boldsymbol{S_i}$	*0,6793*	*1,1014*	*1,3202*	*1,0118*	***1***

Exponenciális trend és multiplikatív szezonalitás esetén a nyers szezonindexek szorzata 1, vagyis a nyers és a korrigált szezonindexek megegyeznek.

A tényleges értékesítés az I. negyedévekben átlagosan 32,07%-kal alacsonyabb, a II. negyedévekben átlagosan 10,14%-kal, a III. negyedévekben 32,02%-kal, a IV. negyedévekben 1,18%-kal magasabb, mint a lineáris trendből becsült érték.

Előrejelzés

Az idősorelemzés célja, hogy a vizsgált jelenség kapcsán nem ismert időszakokra vonatkozóan is szerezzünk információt. A leggyakrabban jövőre vonatkozó becsléseket készítünk, azaz extrapolációt végzünk. Ezen folyamat során a tartósan érvényesülő tendencia, az alapirányzat, valamint a szezonalitás ismeretében meghatározzuk a megfigyelési időszakot követő értékeket[17]. Az előrejelzés során tehát feltételezzük, hogy a vizsgált időszakon belül feltárt tendencia és periodikus hullámzás folytatódni fog. Az alapirányzat és a szezonalitás additív vagy multiplikatív módon kapcsolódhat egymáshoz, mely a vizsgált adatok tekintetében kétféle előrejelzésre ad lehetőséget.

☸ A korábbi feladatok folytatásaként készítsen előrejelzéseket 2014-re úgy, hogy valamennyi alapirányzatot és szezonalitást kombinálja!

Előrejelzés 2014-re mozgóátlagolás és additív szezonalitás alapján

$$\bar{\mathbf{d}} = \frac{\hat{y}_n - \hat{y}_1}{n-1} = \frac{57 - 50{,}875}{8-1} = \mathbf{0{,}875} \text{ ezer db}$$

Centrírozott mozgó átlagolás alapján a vizsgált időszakban negyedévente átlagosan 875 db-bal nőtt az értékesített termékek száma. Az utolsó centrírozott mozgó átlaggal rendelkező időszak 2013. II. negyedéve volt,

[17] A konjunkturális hatást és a véletlen szerepét most figyelmen kívül hagyjuk.

ezért 2013. III. negyedévében 57+0,875= 57,875 db, 2013. IV. negyedévében 57,875+0,875=58,75 vagy $57 + 0{,}875 \cdot 2 = 58{,}75$ db terméket értékesítettek. ebből következően 2014. I. negyedévében $57 + 0{,}875 \cdot 3 = 59{,}625$ db az értékesített termékmennyiség.

45. *Munkatábla az előrejelzés készítéséhez – mozgóátlagolás és additív szezonalitás*

		$\hat{y}\, c.m.á$	S_i	$Előrejelzés: \hat{y}\, c.m.á + S_i$
2014.	I.	$57 + 0{,}875 \cdot 3 = 59{,}625$	-17,234	$59{,}625 - 17{,}234 = \mathbf{42{,}391}$
	II.	$57 + 0{,}875 \cdot 4 = 60{,}5$	1,891	$60{,}5 + 1{,}891 = \mathbf{62{,}391}$
	III.	$57 + 0{,}875 \cdot 5 = 61{,}375$	15,9535	$61{,}375 + 15{,}9535 = \mathbf{77{,}3285}$
	IV.	$57 + 0{,}875 \cdot 6 = 62{,}25$	-0,609	$62{,}25 - 0{,}609 = \mathbf{61{,}641}$

entrírozott mozgó átlagolás és additív szezonalitás alapján 2014. I. negyedévében várhatóan 42,391 ezer db, a II. negyedévben 62,391 ezer db, a III. negyedévben 77,3285 ezer db, a IV. negyedévben 61,641 ezer db terméket fog értékesíteni a vállalat.

Előrejelzés 2014-re mozgóátlagolás és multiplikatív szezonalitás alapján

A multiplikatív szezonalitással történő előrejelzéshez felhasználható az előbb kiszámított centrírozott mozgó átlag értékei (a számítás menete megegyezik).

46. *Munkatábla az előrejelzés készítéséhez – mozgóátlagolás és multiplikatív szezonalitás*

		$\hat{y}\, c.m.á$	S_i	$Előrejelzés: \hat{y}\, c.m.á \cdot S_i$
2014.	I.	59,625	0,7006	$59{,}625 \cdot 0{,}7006 = \mathbf{41{,}77}$
	II.	60,5	1,0578	$60{,}5 \cdot 1{,}0578 = \mathbf{64}$
	III.	61,375	1,3323	$61{,}375 \cdot 1{,}3323 = \mathbf{81{,}77}$
	IV.	62,25	1,0127	$62{,}25 \cdot 1{,}0127 = \mathbf{63{,}04}$

Centrírozott mozgó átlagolás és multiplikatív szezonalitás alapján 2014. I. negyedévében várhatóan 41,77 ezer db, a II. negyedévben 64 ezer db, a III. negyedévben 81,77 ezer db, a IV. negyedévben 63,04 ezer db terméket fog értékesíteni a vállalat.

Előrejelzés lineáris trend és additív szezonalitás alapján

A lineáris trend esetén t=1,2..n számozás esetén 2013. IV. negyedéve t=12 értéket vett fel. 2014-ben a számozás a 13. értékkel folytatódik, melyeket a trend egyenletébe helyettesítve megkapjuk a trendből becsült értékeket.

47. *Munkatábla az előrejelzés készítéséhez – lineáris trend és additív szezonalitás*

		t	trendből becsült érték $\hat{y}$	S_i	$Előrejelzés: \hat{y} + S_i$
2014.	I.	13	$42{,}5 + 1{,}6923 \cdot 13 = \mathbf{64{,}5}$	- 13,03	64,5-13,03 = **51,47**
	II.	14	$42{,}5 + 1{,}6923 \cdot 14 = \mathbf{66{,}2}$	2,43	66,2+2,43 = **68,63**
	III.	15	$42{,}5 + 1{,}6923 \cdot 15 = \mathbf{67{,}9}$	12,57	67,9+12,57 = **80,47**
	IV.	16	$42{,}5 + 1{,}6923 \cdot 16 = \mathbf{69{,}6}$	–1,97	69,6-1,97 = **67,63**

Lineáris trend és additív szezonalitás alapján 2014. I. negyedévében várhatóan 51,47 ezer db, a II. negyedévben 68,63 ezer db, a III. negyedévben 80,47 ezer db, a IV. negyedévben 67,63 ezer db terméket fog értékesíteni a vállalat.

Előrejelzés lineáris trend és multiplikatív szezonalitás alapján

48. *Munkatábla az előrejelzés készítéséhez – lineáris trend és multiplikatív szezonalitás*

		t	$\hat{y}$	S_i	$Előrejelzés: \hat{y} \cdot S_i$
2014.	I.	13	64,5	0,6626	$64{,}5 \cdot 0{,}6626 = \mathbf{42{,}74}$
	II.	14	66,2	1,142	$66{,}2 \cdot 1{,}142 = \mathbf{75{,}6}$
	III.	15	67,9	1,3072	$67{,}9 \cdot 1{,}3072 = \mathbf{88{,}76}$
	IV.	16	69,6	1,0109	$69{,}6 \cdot 1{,}0109 = \mathbf{70{,}36}$

Lineáris trend és multiplikatív szezonalitás alapján 2014. I. negyedévében várhatóan 42,74 ezer db, a II. negyedévben 75,6 ezer db, a III. negyedévben 88,76 ezer db, a IV. negyedévben 70,36 ezer db terméket fog értékesíteni a vállalat.

Előrejelzés exponenciális trend és additív szezonalitás alapján

Az exponenciális trend esetén t=1,2..n számozás esetén 2013. IV. negyedéve szintén t=12 értéket vett fel. 2014-ben a számozás itt is a 13. értékkel folytatódik, melyeket a trend egyenletébe helyettesítve megkapjuk a trendből becsült értékeket.

49. *Munkatábla az előrejelzés készítéséhez – exponenciális trend és additív szezonalitás*

		t	trendből becsült érték $\hat{y}$	S_i	$Előrejelzés: \hat{y} + S_i$
2014	I.	13	$39{,}692 \cdot 1{,}0408^{13} = 66{,}75$	-16,759	66,75-16,759=**49,99**
	II.	14	$39{,}692 \cdot 1{,}0408^{14} = 69{,}48$	3,246	69,48+3,246=**72,73**
	III.	15	$39{,}692 \cdot 1{,}0408^{15} = 72{,}31$	14,837	72,31+14,837=**87,15**
	IV.	16	$39{,}692 \cdot 1{,}0408^{16} = 75{,}26$	-1,324	75,26-1,324=**73,94**

Exponenciális trend és additív szezonalitás alapján 2014. I. negyedévében várhatóan 49,99 ezer db, a II. negyedévben 72,73 ezer db, a III. negyedévben 87,15 ezer db, a IV. negyedévben 73,94 ezer db terméket fog értékesíteni a vállalat.

Előrejelzés exponenciális trend és multiplikatív szezonalitás alapján

50. *Munkatábla az előrejelzés készítéséhez – exponenciális trend és multiplikatív szezonalitás*

		t	$\hat{y}$	S_i	$Előrejelzés: \hat{y} \cdot S_i$
2014.	I.	13	66,75	0,6793	$66{,}75 \cdot 0{,}6793 = \mathbf{45{,}34}$
	II.	14	69,48	1,1014	$69{,}48 \cdot 1{,}104 = \mathbf{76{,}71}$
	III.	15	72,31	1,3202	$72{,}31 \cdot 1{,}3202 = \mathbf{95{,}46}$
	IV.	16	75,26	1,0118	$75{,}26 \cdot 1{,}0118 = \mathbf{76{,}15}$

xponenciális trend és multiplikatív szezonalitás alapján 2014. I. negyedévében várhatóan 45,34 ezer db, a II. negyedévben 76,71 ezer db, a III. negyedévben 95,46 ezer db, a IV. negyedévben 76,15 ezer db terméket fog értékesíteni a vállalat.

11.2.3 A trendilleszkedés vizsgálata

A megfigyelési értékekre többféle alapirányzat illeszthető, a legpontosabb előrejelzés azzal készíthető, amelyik legjobban közelíti a megfigyelési értékeket. A tényleges és becsült értékek páronkénti eltérésén alapuló reziduális szórás számításával tudjuk megállapítani, hogy melyik trend illeszkedik legjobban az adatokra.

$$\boldsymbol{s_e} = \sqrt{\frac{\sum(y_i - \hat{y})^2}{n-2}} = \sqrt{\frac{\sum e^2}{n-2}}$$

A tapasztalati értékekre az a trend illeszkedik legjobban, amelynél a reziduális szórás a legkisebb.

☸ A korábbi példában egy vállalat 3 éves, negyedéves bontású értékesítési adataira különböző trendeket illesztettünk. *Állapítsa meg, hogy melyik illeszkedik a legjobban!*

Mozgóátlag $\boldsymbol{s_e} = \sqrt{\frac{\sum e^2}{n-2}} = \sqrt{\frac{1130{,}7968}{8-2}} = \mathbf{13{,}728}\ ezer\ db$

A tényleges és a centrírozott mozgó átlagból számított értékek páronkénti eltérése 13 728 db.

Lineáris trend $\boldsymbol{s_e} = \sqrt{\frac{\sum e^2}{n-2}} = \sqrt{\frac{1737{,}462}{12-2}} = \mathbf{13{,}181}\ ezer\ db$

Lineáris trend alapján a tényleges és becsült értékek páronkénti átlagos eltérése 13 181 db.

Exponenciális trend $\boldsymbol{s_e} = \sqrt{\frac{\sum e^2}{n-2}} = \sqrt{\frac{1809{,}242}{12-2}} = \mathbf{13{,}451}\ ezer\ db$

Az exponenciális trend alapján a tényleges és becsült értékek páronkénti átlagos eltérése 13 451 db.

Az alapirányzatok összehasonlítása:

Lineáris trend		**Exponenciális trend**		**Mozgó átlag**
$s_e = 13{,}181$ ezer db	<	$s_e = 13{,}451$ ezer db	<	$s_e = 13{,}728$ ezer db

A reziduális szórás értékei alapján az adatokra legjobban a LINEÁRIS TREND illeszkedik.

11.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

11.3.1 Összefoglalás

Az időbeli ismérv szerinti vizsgálatokhoz változatos eszköztárat mutattunk be. Az idősorelemzési technikák közötti választást befolyásolhatja az, hogy milyen információt szeretnénk kapni az idősoros adatokról. A módszerek egy része csupán az ismert adatok időbeli változását számszerűsíti, ebben a leckében azonban olyan módszerek kerültek bemutatásra, melyek előrejelzésre is alkalmazhatók.

Az ismertetett módszerek közül a mozgóátlagolás alkalmazásával szemben fogalmazódik meg a legtöbb kritika. Az adatok átlagolása könnyen kivitelezhető, de a kiugró adatok ez esetben is komoly problémát jelenthetnek, erős torzító hatásuk lehet. Az idősor lerövidülése szintén pontatlanságot eredményez, mely azonban csökkenthető a tapasztalati értékek növelésével, hiszen minél több adat áll rendelkezésre, annál kisebb az idősor lerövidüléséből fakadó információveszteség. A leggyakrabban lineáris, illetve exponenciális trendet illesztünk az adatokra, mely számítógépes programokkal is könnyen megtehető. A szezonalitás figyelembevétele elsősorban olyan adatsoroknál fontos, ahol a tapasztalati értékekben is jól látszik a periodikus hullámzás.

11.3.2 Önellenőrző kérdések

1. Mi a különbség az átlagos abszolút és átlagos relatív változás között?
2. Determinisztikus idősorelemzés esetén melyek az idősor fő elemei?
3. Hogyan jellemezhető az alapirányzat?
4. Milyen módszerekkel határozható meg az alapirányzat?
5. Mi alapján dönthető el, hogy melyik alapirányzat illeszkedik legjobban?
6. Hogyan írható fel a lineáris trend általános alakban, s mit jelentenek a paraméterek?
7. Hogyan jellemezhető a konjunkturális hatás?
8. Hogyan jellemezhető a szezonalitás?
9. Milyen módszerekkel határozható meg a szezonalitás?
10. Miért lényeges a véletlen hatásával számolni az idősorban?

11.3.3 Gyakorló tesztek

☸ Döntse el az alábbi állításokról, hogy igazak vagy hamisak!

Az a trend illeszkedik legjobban, ahol a reziduális szórás értéke a legkisebb.	I
Lineáris trend és additív szezonalitás esetén a nyers és korrigált szezonindexek megegyeznek.	H
Exponenciális trend és multiplikatív szezonalitás esetén a nyers szezonindexek szorzata 1.	I
Lineáris trend és additív szezonalitás esetén a nyers szezonális eltérések összege 1.	H
Exponenciális trend és multiplikatív szezonalitás esetén a nyers és korrigált szezonális eltérések megegyeznek.	H

☸ Válassza ki a helyes megoldást!

Egy étterem éves forgalmát (ezer Ft) 2000 és 2013 között leíró lineáris trend egyenlete $y=620+12 \cdot t$. A trend alapján az alábbi következtetés vonható le:

- az étterem forgalma a vizsgált időszakban csökkent
- az étterem forgalma 2000-ben 620 ezer Ft volt
- **az étterem forgalma a vizsgált időszakban évente átlagosan 12 ezer Ft-tal nőtt**

Egy étterem 2014-es utószezonjának forgalmára vonatkozóan lineáris trend alapján 1 400 ezer Ft várható, továbbá tudjuk, hogy utószezonban átlagosan 10%-kal alacsonyabb a forgalom, mint a becsült érték. Az előrejelzés szerint 2014. főszezonjában a szezonalitás figyelembevételével számított forgalom

- 1540 ezer Ft
- **1260 ezer Ft**
- 1410 ezer Ft

Egy vállalat termelésének (ezer db) alakulását negyedéves bontásban (ezer Ft) 2009 és 2013 között leíró lineáris trend egyenlete $y=1200-50t$. A trend alapján az alábbi következtetés vonható le:

- a vállalat termelése évente átlagosan 50 ezer db-bal csökkent
- **a vállalat termelése negyedévente átlagosan 50 ezer db-bal csökkent**
- 2013-ban a vállalat termelése 1200 ezer db volt

A lineáris trendből és additív szezonalitás alapján történő előrejelzések szerint 2015. II. negyedévében egy vállalat várhatóan 6500 db terméket fog értékesíteni. A lineáris trend alapján azonban 6200 db lenne. Mekkora a II. negyedévek szezonalitása?

- a II. negyedévekben átlagosan 300 db-bal alacsonyabb az értékesítés
- **a II. negyedévekben átlagosan 300 db-bal magasabb az értékesítés**
- a II. negyedévekben átlagosan 4,82%-kal magasabb az értékesítés

Egy vállalat árbevételére vonatkozóan az alapirányzatok reziduális szórásáról az alábbiakat ismerjük: a tényleges és becsült értékek páronkénti átlagos eltérése lineáris trendnél 625 db, mozgóátlagnál 612 db, exponenciális trendnél 627 db. Melyik alapirányzat illeszkedik legjobban?

- lineáris trend
- exponenciális trend
- **mozgó átlag**

Egy áruház forgalmáról tudjuk, hogy a III. negyedévében átlagosan 12,5%-kal kevesebb terméket értékesítenek. Ez azt jelenti, hogy

- a szezonális eltérés értéke 12,5%
- a szezonindex értéke 112,5
- **a szezonindex értéke 87,5**

12. A STATISZTIKAI ELEMZÉS SZEMLÉLTETŐ ESZKÖZE: A GRAFIKUS ÁBRÁZOLÁS

12.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A statisztikai elemzések eredményeinek szemléletes megjelenítési formája a grafikus ábrázolás. Az ábrák segítségével könnyen átláthatók a tendenciák és az összefüggések, ugyanakkor torzíthatják is a számszaki eredményeket, mely alapján téves következtetések is megfogalmazhatók. Az egyszerűség és áttekinthetőség kritériumait betartva egy statisztikailag precízen szerkesztett diagram a vizsgált probléma kapcsán, tömörítve, számos információt szolgáltathat anélkül, hogy elvesznénk a számok között. A legtöbb esetben az időbeli vizsgálatoknál ábrázoljuk grafikusan az adatokat, de a gyakoriságok, megoszlások és összefüggések kimutatását is megkönnyíti, ezért a vizuális illusztráció módszerét gyakran alkalmazzuk a statisztikai vizsgálatok során.

A lecke célja, hogy megismertesse a hallgatókat a grafikus ábrázolás statisztikában való alkalmazásának lehetőségeivel. A tananyag korábbi részeiben megismert módszerekkel végezhető számítások eredményeit a hallgató képes legyen grafikusan is illusztrálni. Ismerje a főbb ábratípusokat, azok jellemzőit és alkalmazási területeit. Sajátítsa el a statisztikai ábraszerkesztés főbb alapelveit, melyek segítségével statisztikailag precíz formában tudja illusztrálni elemzésének eredményeit.

12.2 TANANYAG

A grafikus ábrázolás során az egyszerűség és az áttekinthetőség kiemelten fontos, a cél ugyanis az, hogy elemzésünk eredményeit tömören, informatív módon illusztráljuk. A vizsgálódás tárgyától függően különböző ábratípusokat választhatunk, az ábrázolás alapelvei azonban minden esetben betartandók. A tartalmi és formai követelmények között a cím és forrásmegjelölés emelendő ki, valamint a célorientált szerkesztés. Az ábrák típusait két csoportra bonthatjuk. A koordináta rendszeren alapuló ábrák esetében egy adott jelenséget két összetartozó jellemző mentén vizsgálunk. Az oszlop, illetve sávdiagram, a vonaldiagram vagy a pontdiagram elsősorban összefüggés és tendenciavizsgálatra, valamint összehasonlításra használható. A nem koordináta rendszeren alapuló ábrák, mint például a kör vagy tortadiagram, illetve a térképen alapuló ábrák egy adott dimenzió mentén való elemzést tesznek lehetővé.

12.2.1 A grafikus ábrázolás tartalmi és formai követelményei

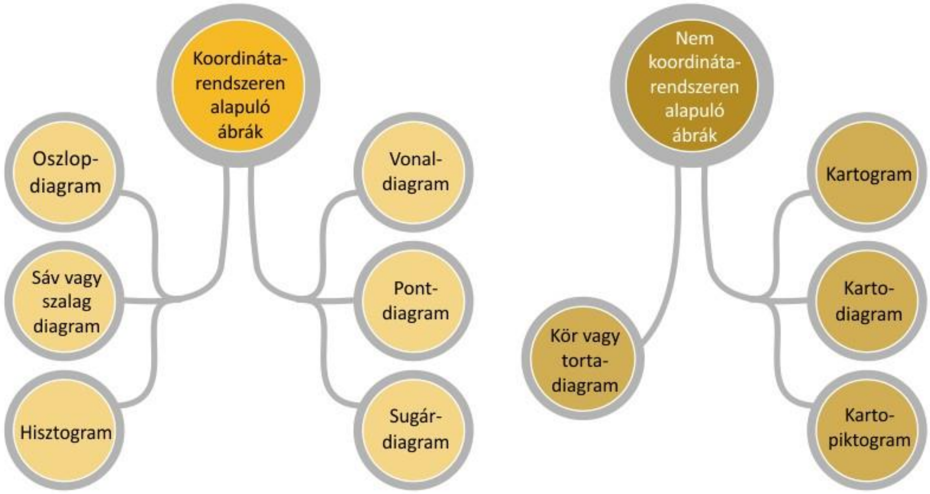
A grafikus ábrázolás első lépéseként át kell gondolni, hogy mi az illusztráció célja, vagyis mit akarunk szemléltetni. Az ábra főként arányokat, tendenciákat érzékeltet, így a vizsgált jelenség kapcsán az ábrázolás célját a diagram típusának megválasztásánál figyelembe kell venni. Az ábrázolni kívánt adatok mennyiségétől, nagyságrendjétől és a céltól függően különböző ábratípusokat választhatunk.

A diagram szerkesztésénél törekedni kell az áttekinthetőségre és az egyszerűségre. Ne akarjunk nagy mennyiségű és sokféle adatot egy diagramba sűríteni, mert az ábra túlságosan összetett és átláthatatlan lesz. Lényeges, hogy már első ránézésre könnyen kivehető legyen, mit ábrázol. A megértést könnyítendő olyan címmel kell ellátni, amely utal a jelenség által érintett sokaságra és annak jellemzőire. Időbeli adatoknál például a címben pontosan le kell határolni a vizsgált időszakot. Összehasonlított adatsoroknál az összehasonlítás alapjául szolgáló ismérvet. Ha egynemű adatokat tüntetünk fel az ábrán, akkor a címben jelölni lehet az adatok mértékegységét is, amely a cím helyett vagy mellett a tengelyeknél is megadható. A cím mellett az ábra kötelező kelléke a forrásmegjelölés, azaz fel kell tüntetni azt, hogy honnan származnak az adatok. Abszolút adatok esetén elegendő a pontos adatforrást megjelölni évszámmal ellátva, így például *OECD (2014).* Primer forrásból származó adatoknál jelölni lehet azt is, hogy saját adatgyűjtés vagy saját kutatás. Származtatott adatok esetén, vagyis ha az elérhető adatokon valamilyen matematikai műveletet végeztünk, az alábbiak szerint tüntetjük fel a forrást: *Saját számítás OECD (2014) alapján*.

A diagram formai megoldásait illetően a tengelyek elnevezése és a jelmagyarázat opcionális, nem szükséges minden esetben feltüntetni. A jelmagyarázat többletinformációt szolgáltat, így kötelező kellék, ha időben többféle adatsor alakulását ábrázoljuk, de felesleges, ha csak egy adott jelenség változását vizsgáljuk. A tengelyek elnevezésére is csak akkor van szükség, ha egyértelművé teszi a diagram mondanivalóját, így például ha a címből nem derül ki, hogy milyen adatokat ábrázoltunk. Időbeli adatoknál nem szükséges az „évek” elnevezést odatenni, mert egyértelműen kivehető a tengelyfeliratokból, hogy miről van szó. A tengelyek beosztását az ábrázolni kívánt adatsornak megfelelően válasszuk meg. Nem feltétlenül szükséges a 0-t kezdőpontként szerepeltetni, de figyelni kell arra, hogy ez ne zavarja az ábra mondanivalóját. A színek és formák megválasztásánál kialakíthatjuk az ábra egyedi stílusát, a statisztikai tartalom figyelembevételével. *Egy ábra akkor jó, ha önmagában, a hozzáfűzött szöveg elolvasása nélkül is világosan értelmezhető az olvasó számára*. A továbbiakban az általános alapelvek alkalmazása különböző diagramtípusokon keresztül kerül bemutatásra.

12.2.2 A grafikus ábrák típusai

A grafikus ábrákat megkülönböztethetjük aszerint, hogy egymáshoz viszonyított jellemzők mentén, koordináta rendszerben ábrázoljuk az adatokat, vagy más, nem viszonyítható módon, speciális formában. Az összetartozó értékpárokra determinisztikus kapcsolat esetén függvény illeszthető, mely főként a pontdiagramhoz kapcsolódóan regresszió és trend esetén fordul elő.

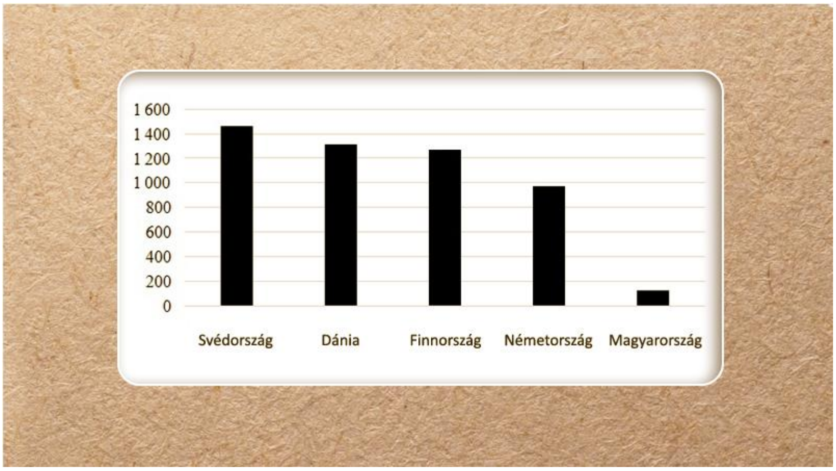


44. ábra: A grafikus ábrák típusai

Oszlop és sáv/szalagdiagram

Az oszlop és a sáv vagy más szóval szalagdiagram[18] abszolút és relatív gyakorisági sorok ábrázolására alkalmazható. Az oszlopok magassága vagy a sávok hosszúsága mutatja a gyakoriságot, mely révén ez a diagramtípus kiválóan alkalmazható bármilyen típusú összehasonlításra.

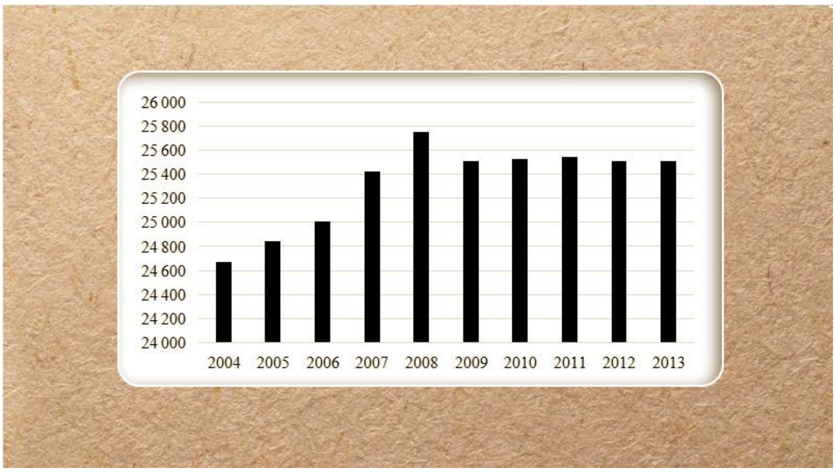
[18] A sáv vagy szalagdiagram lényegében egy 90 fokkal elforgatott oszlopdiagram.



45. ábra: Az egy főre jutó euróban kifejezett K+F ráfordítások adatai néhány európai országban 2012-ben

Forrás: EUROSTAT (2014)

Az ábratípus segítségével jól szemléltethető az ismérvváltozatok gyakoriság szerinti sorrendje, idősoros adatoknál azonban az idő előrehaladását kell előtérbe helyezni, s így a változások iránya lesz jól követhető.

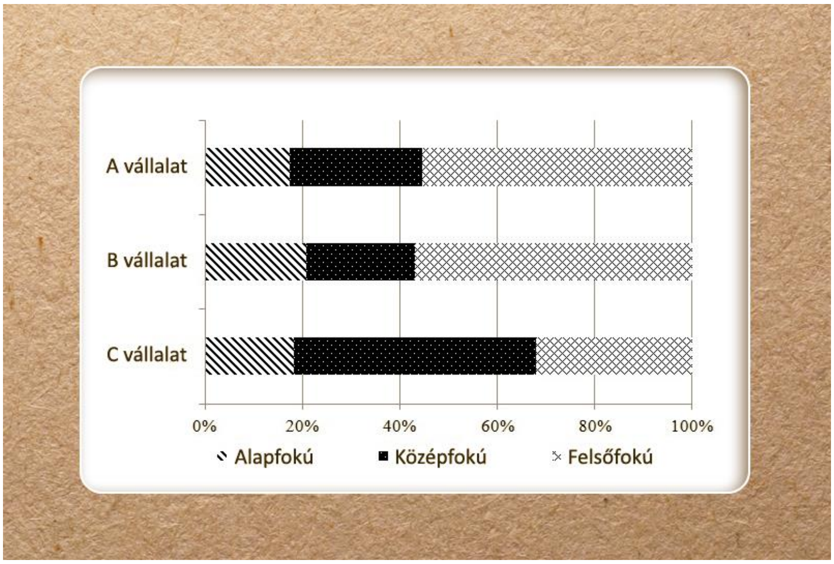


46. ábra: A 15-64 éves korosztályból foglalkoztatottak számának (ezer fő) alakulása Franciaországban 2004 és 2013 között

Forrás: *EUROSTAT* (2014)

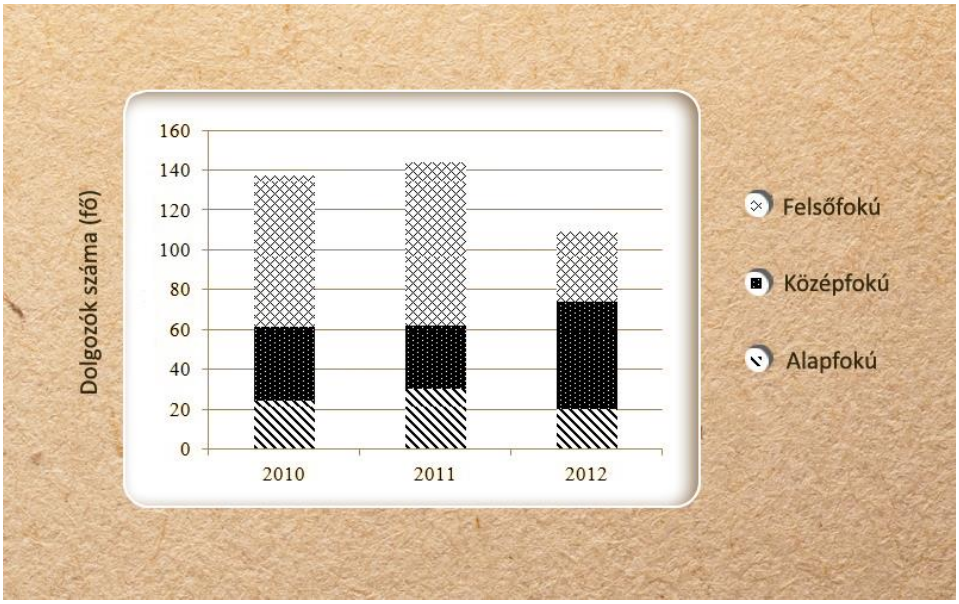
A 46. ábrán a függőleges tengely kezdőpontja nem a 0, melynek köszönhetően a szűk ingadozási sávban mozgó adatok időbeli változásának tendenciája jobban kirajzolódik. A diagram segítségével végzett elemzés során azonban a nagyságrendi különbségtételnél óvatosan kell fogalmazni. Ha kiindulópontként nem 0-t választunk, az ábra megtévesztő is lehet, mert felnagyíthatja a nem túl jelentős eltéréseket is.

Az oszlop és sávdiagramokon több, ugyanazon ismérv szerint jellemzett adatsor is ábrázolható, vagyis többdimenziós vizsgálatra is lehetőséget nyújtanak. Ez esetben a jelmagyarázat kötelező kelléke az ábrának, mert egyértelműen látszódnia kell, hogy milyen adatsorokat hasonlítunk össze, s ezeket a címben is szerepeltetni szükséges. A halmozott oszlop vagy sávdiagram megoszlások összehasonlításának szemléltetésére alkalmas, a sokaságok elemszámától függetlenül érzékeltetni tudjuk az arányokat, de ezen arányok az elemszámbeli eltérés figyelembevételével is illusztrálhatók. Összevethető például két vagy több időszak között egy vállalatnál a dolgozók iskolai végzettség szerinti megoszlása úgy, hogy csupán a megoszlásra helyezzük a hangsúlyt, ekkor 100%-ig halmozott diagramtípust választunk, melynek szerkesztésénél az oszlop magassága vagy a sáv szélessége nem haladhatja meg a 100%-ot. Az ábra megszerkeszthető úgy is, hogy érzékeltetjük a dolgozói létszámban bekövetkező változást is.



47. ábra: Három vállalat dolgozóinak iskolai végzettség szerinti megoszlása

Forrás: fiktív adatok

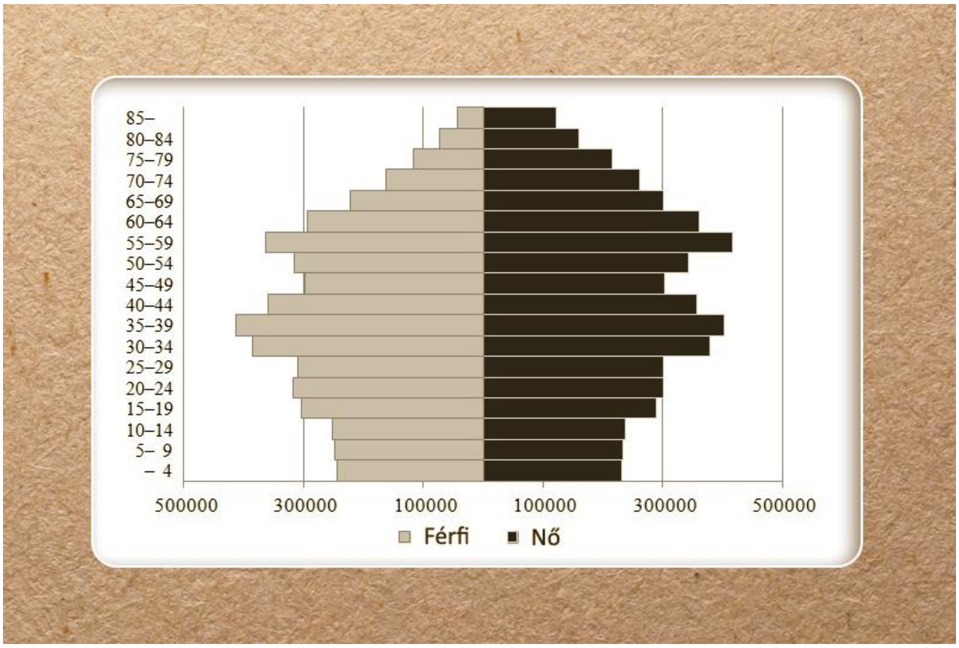


48. ábra: Egy vállalatnál dolgozók iskolai végzettség szerinti összetételének alakulása 2010 és 2012 között

Forrás: fiktív adatok

A sávdiagram speciális, népességstatisztikában alkalmazott formája a korfa (population pyramid). Ezt a diagramtípust a férfiak és nők korcsoportonkénti megoszlásának illusztrálására alkalmazzák. Összetett diagramtípus, középen helyezkedik el az életkort jelölő függőleges tengely, két oldalán pedig a nemek korcsoportonkénti elemszáma, azaz gyakorisága. A korfa a népesség korstruktúráját mutatja a nemek és az életkor figyelembevételével, melynek típusairól az alábbi oldalon olvashat:

24. Korfa: http://www.ehow.com/list_6872808_three-different-types-population-pyramids.html



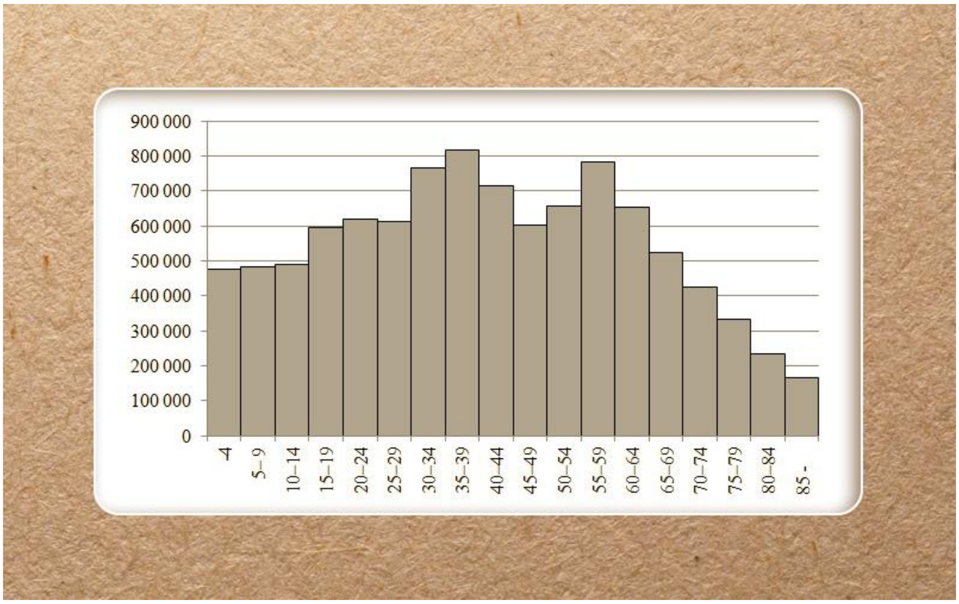
49. ábra: A magyar népesség korfája 2011-ben

Forrás: KSH (2014)

Hisztogram

A hisztogram speciális oszlopdiagram, melyet gyakorisági – főként osztályközös – sorok ábrázolásánál alkalmazunk. Nagy elemszámú adatsor esetén mennyiségi ismérv alapján átfedésmentesen egyenlő és nem egyenlő hosszúságú intervallumok képezhetők, melyekbe az elemek egyértelműen besorolhatók[19]. A hisztogram készítésénél egy olyan oszlopdiagramot szerkesztünk, melynél az oszlopok között nincs térköz.

[19] Az osztályközös gyakorisági sorokról bővebben a 2.2.2 Statisztikai adatrendszerezés fejezetben olvashat.



50. ábra: Magyarország népessége korcsoportonként 2011-ben

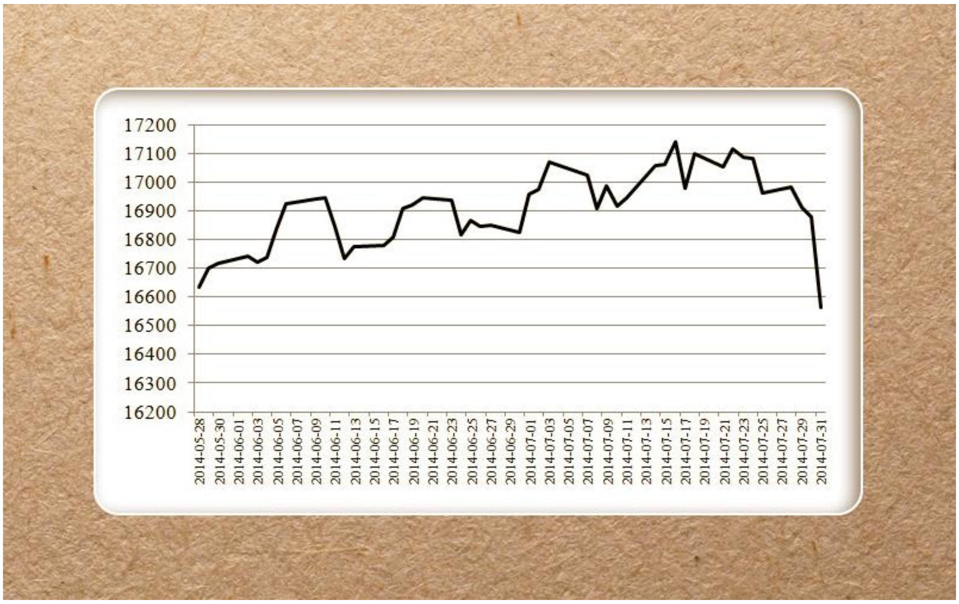
Forrás: KSH (2014)

A hisztogram az osztályok gyakoriságáról és az adatsor eloszlásáról is szolgáltat információt. Az ábra vízszintes tengelyén az osztályokat tüntetjük fel, függőleges tengelyére pedig az abszolút vagy relatív gyakorisági értékek kerülnek. Az ábrán a gyakoriságok alapján kirajzolódik az eloszlásfüggvény alakja, így szemléltethető, hogy az adatsor mennyire szimmetrikus, kivehető az esetleges aszimmetria iránya, valamint a csúcsosság is, melyet a leíró statisztika alakmutatói számszerűsítenek.

Vonaldiagram

A vonaldiagram leggyakrabban idősoroknál alkalmazható, ahol az idősoros adatok folytonos értékek. A diagramtípus jó választás akkor is, ha folytonos mennyiségi ismérv alapján készített gyakorisági sorokat szeretnénk ábrázolni. A vonal a folytonosságot szimbolizálja.

Ez a diagramtípus tendenciák szemléltetésére kiválóan alkalmazható. A változások iránya jól követhető, a vizsgált időszak csúcs- és mélypontjai egyértelműen kivehetők. Az ábraszerkesztés során az elsődleges és másodlagos tengely bevezetésével több, esetlegesen eltérő nagyságrendű adatok is feltüntethetők. Ez utóbbi esetben az ábra összetettsége megnehezítheti a mondanivaló értelmezhetőségét. Az időbeli alakulás vizsgálata több, azonos ismérv mentén jellemzett adatsor esetén is megvalósítható.

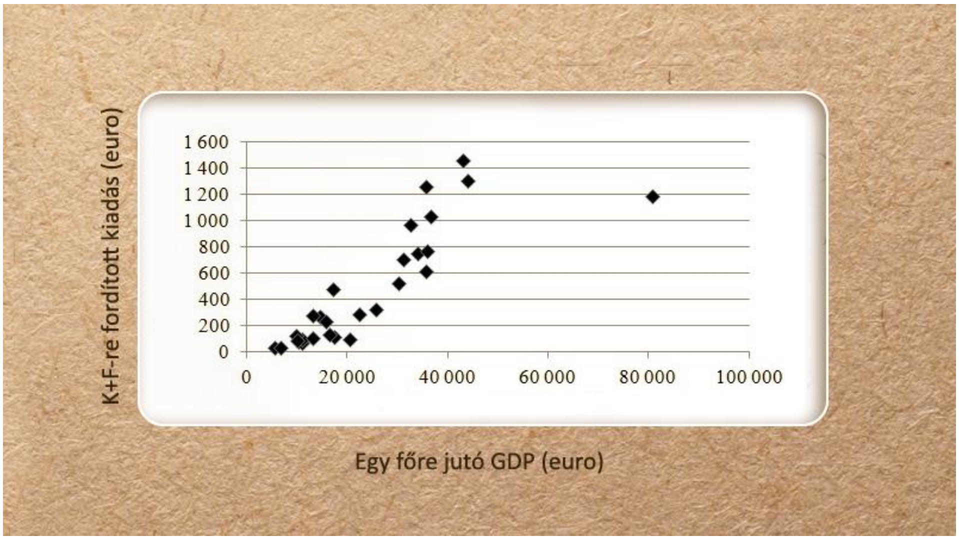


51. ábra: A DOW Jones index alakulása
2014. május 28 és július 31 között

Forrás: *FRED* (2014)

Pontdiagram

A pontdiagramot leggyakrabban mennyiségi ismérvek közötti kapcsolat vizsgálatánál alkalmazzuk, ha diszkrét mennyiségi értékek állnak rendelkezésünkre. Ezen ábratípus esetén fontos kritérium, hogy az értékek közötti átmenet nem értelmezhető, ugyanis ellenkező esetben folytonos ismérvek lévén vonaldiagramot kellene alkalmazni. A két ismérv esetében a tengelyek elnevezése mellett a mértékegységet is fel kell tüntetni.



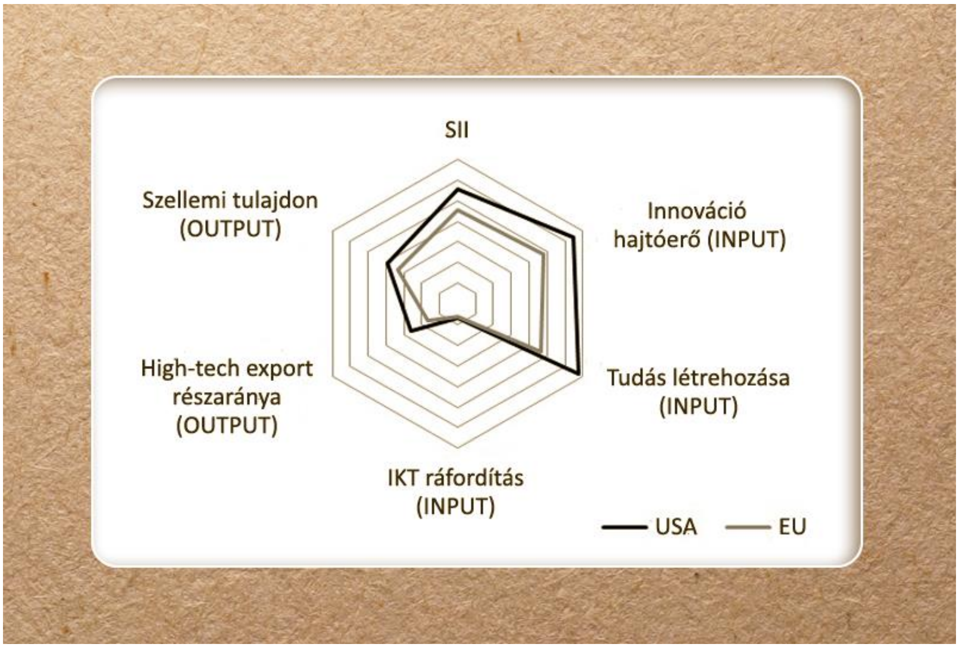
52. ábra: Az egy főre jutó GDP és a K+F-re fordított kiadások közötti összefüggés az Európai Unió tagállamaiban

Forrás: EUROSTAT (2014)

A pontdiagram alapján láthatóvá válik, hogy az összetartozó értékpárokra milyen alakzat illeszkedik leginkább, vagyis milyen típusú regresszió függvényt célszerű az adatokra illeszteni. A fenti ábra kapcsán megfigyelhető, hogy a statisztikai vizsgálatok szempontjából releváns outlier érték is jól kivehető. Ezen ábratípus lehetőséget biztosít arra, hogy az eltérő nagyságrendű adatok közötti kapcsolat is jól illusztrálható legyen.

Sugárdiagram vagy poláris görbe

A sugárdiagram egy polárkoordináta-rendszerben megszerkeszthető ábratípus, melynek segítségével egy adott jelenség több szempontból is vizsgálhatóvá válik. A sokszög formája a vizsgált jellemzők számától függ. Az ábra közepén helyezkedik el a 0, a sokszög csúcsaiban pedig a vizsgált jellemzők, s értékeik. A többdimenziós vizsgálat területi, időbeli és minőségi összehasonlításra is lehetőséget biztosít.



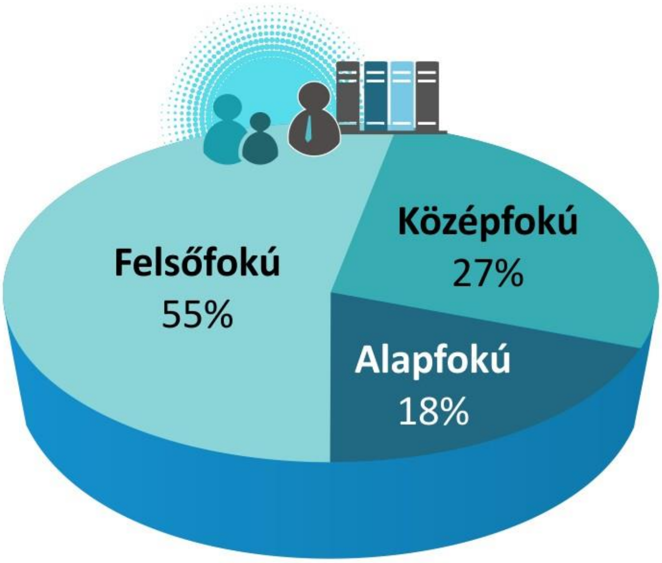
53. ábra: Európa és az USA teljesítménye az összetett innovációs mérőszám egyes részterületein

Forrás: EC – Union Innovation Scoreboard (2012)

Az 53. ábrán az összetett innovációs mérőszám (Summary Innovation Index, SII) és kiemelt részterületeinek európai és amerikai összevetése látszik. A diagramon feltüntethető az adatok mértékegysége is, ám ezt az ábratípust adott jelenség vizsgálata során az összetartozó területeken való produktumok összehasonlításra használjuk elsődlegesen, ahol az alapadatok mértékegysége kevésbé lényeges.

Kör vagy tortadiagram

A kör vagy a nemzetközi statisztikákban inkább használatos tortadiagram (pie chart) csoportosító sorokból készíthető ábratípusok, melyekkel megoszlást szemléltethetünk. A grafikus ábrák szerkesztésére alkalmas programokban perecdiagram is választható, mely lehetővé teszi az azonos elemekből álló sokaságok megoszlásainak összehasonlítását is. Ezeknél az ábratípusoknál kiemelten fontos az áttekinthetőség és az egyes részek beazonosíthatósága. Célszerű feltüntetni a százalékos megoszlás értékeit is az egyes részeknél, mely az értelmezhetőséget könnyíti, különösen akkor, ha az ábra alapján az arányok nem egyértelműen kivehetők. Sok összetevőből álló sokaságok ábrázolására ezen diagramok kevésbé alkalmasak, mert a jelmagyarázat követhetetlenné válik.



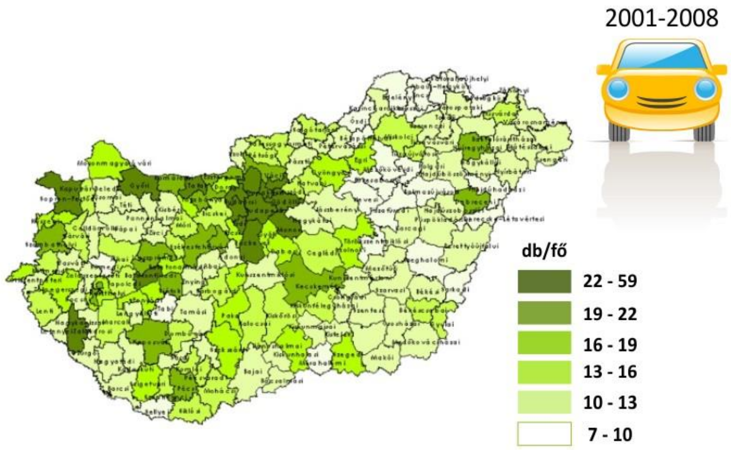
54. ábra: Egy vállalatnál dolgozók iskolai végzettség szerinti megoszlása

Forrás: fiktív adatok

A térkép, mint a statisztikai ábrázolás segédeszköze

A területi összehasonlításoknál a grafikus ábrázolás hasznos segédeszköze lehet a térkép, mely jól áttekinthető formában szolgáltat információt különböző földrajzi egységek adatairól. A térképpel történő ábrázolási formákat főként demográfiai és társadalom-földrajzi elemzések során alkalmazzuk, de speciális területi vizsgálatoknál is látványos megjelenítési eszköz lehet.

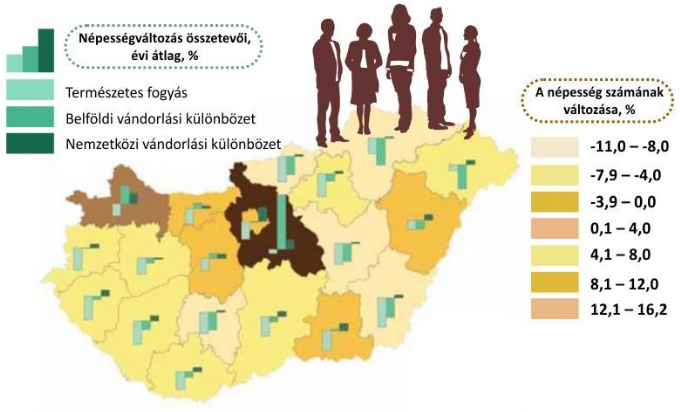
A **kartogram** olyan térképes ábra, melyen eltérő színek és mintázatok jelzik a különböző területi egységeken a vizsgált jelenség intenzitását. A földrajz atlaszokban lévő domborzati, illetve tengerszint feletti magasságokról készített térképek ennek mintájára készülnek, de például a népsűrűség is illusztrálható ilyen formában.



55. ábra: 2001 és 2008 között első alkalommal forgalomba helyezett személygépkocsik száz lakosra jutó száma Magyarországon

Forrás: TeIR KSH T-STAR és Népszámlálás adatbázisok

A **kartodiagram** többletinformációt nyújt a kartogramhoz képest, mert nem csupán egy vizsgált jelenség intenzitását szemlélteti, hanem az adott jelenséghez kapcsolódóan más ismérv alapján diagram segítségével plusz információt is szolgáltat az adott földrajzi egységről.



56. ábra: Magyarország lakónépesség számának változása és annak összetevői a 2001–2011 közötti időszakban

Forrás: KSH - Magyarország társadalmi atlasza (2012)

A kartogramon a területi egységeknél az előzőekben megismert ábratípusok közül leginkább a megoszlás szemléltetésére alkalmas kördiagramot, valamint az időbeli változás vizsgálatához oszlopdiagramot használunk.

A **kartopiktogram** a grafikus ábrázolás nagyon speciális eszköze. Ha a területileg vizsgált jelenség felismerhető módon apró képpel, azaz piktogrammal jól illusztrálható, akkor lehetőség van az intenzitást a piktogram méretezésével is szemléltetni.

57. ábra: Magyarország kullancsfertőzött helyei – fotó (2009. június 15)

Forrás: Alfahír Hírportál (2012)

12.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

12.3.1 Összefoglalás

A statisztikai elemzések eredményei a grafikus ábrázolás segítségével változatos formában jeleníthetők meg. Az egyszerűség, áttekinthetőség és célorientáltság kritériumait betartva olyan ábrák készíthetők, melyekről könnyen felismerhető az adatsorokban lévő tendencia, illetve összefüggések, továbbá összehasonlítás is könnyebben végezhető. A diagramokat címmel és forrással kell ellátni, fel kell tüntetni az adatok mértékegységét, illetve szükség esetén jelmagyarázatot is kell készíteni. Az ábrázolás céljától függően különböző ábratípusok választhatók. Összehasonlító statisztikai sorok oszlop és sávdiagramokon ábrázolhatók, a csoportosító soroknál leginkább kör vagy tortadiagramot használunk, a folytonosságot szemléltető időbeli vizsgálatoknál a vonaldiagram, míg összefüggéseknél a pontdiagram alkalmazása javasolt. A területi összehasonlítások hasznos segédeszköze lehet a térkép.

12.3.2 Önellenőrző kérdések

1. Melyek a grafikus ábrázolás legfőbb alapelvei?
2. Melyek a grafikus ábrák kötelező tartalmi kellékei?
3. Mikor van szükség jelmagyarázatra?
4. Milyen adatok szemléltetésére alkalmas az oszlopdiagram?
5. Mi a korfa?
6. Mi a hisztogram?
7. Milyen adatok szemléltetésére alkalmas a vonaldiagram?
8. Milyen adatok szemléltetésére alkalmas a pontdiagram?
9. Milyen adatok szemléltetésére alkalmas a kördiagram?
10. Mi a különbség kartogram és kartodiagram között?

12.3.3 Gyakorló tesztek

Döntse el, hogy az alábbi állítások igazak (I) vagy hamisak (H)!

Időbeli összehasonlítás grafikus illusztrálására az oszlop- és a vonaldiagram a legalkalmasabb.	I
A pontdiagramot leggyakrabban összefüggés vizsgálatokhoz készítjük, a pontokból kirajzolódik az összetartozó értékpárokra leginkább illeszkedő alakzat.	I
A korfa a népességstatisztikában alkalmazott oszlopdiagram.	H
A kartodiagram a térképen eltérő színekkel jelzi a vizsgált jelenség területi intenzitását.	H
A hisztogram főként osztályközös gyakorisági sorok szemléltetésére alkalmas speciális oszlopdiagram, ahol nincs térköz az oszlopok között.	I
A kartogram a térképen eltérő színű diagramokkal jelzi a vizsgált jelenség alakulását.	H
A sugárdiagram segítségével több dimenzió mentén is végezhető összehasonlítás adatsorok között.	I

A kartogram a térképen eltérő színekkel jelzi a vizsgált jelenség területi intenzitását.	I
A grafikus ábráknál a cím és forrás megjelölése, valamint a jelmagyarázat mindig kötelező.	H
A koordináta-rendszeren alapuló ábráknál a vizsgált jellemzők egymáshoz viszonyított helyzete meghatározó.	I

13. ÖSSZEFOGLALÁS

13.1 TARTALMI ÖSSZEFOGLALÁS

A statisztikai elemzéseknek jelentős szerepe van a gazdasági és társadalmi jelenségek vizsgálatában. A módszerek alkalmazásához elengedhetetlen az alapfogalmak elsajátítása, így a sokaság, az ismérv és a mérési skálák ismerete. A statisztikai munka az adatok összegyűjtésével kezdődik, mely történhet primer és szekunder forrásból. A szekunder adatokat különböző adatbázisokból gyűjthetjük, melyeknél a megbízhatóság kulcsfontosságú, így érdemes a nemzetközi szervezetek adataira támaszkodni. A primer adatszerzés során egy vizsgált jelenség kapcsán önálló kutatás keretében szerzünk információkat, például egy kérdőíves felmérés segítségével. Az adatgyűjtés lehet a teljes sokaságra vonatkozó, ám gyakrabban fordul elő, hogy a soakságnak csak egy részét tudjuk adatokkal jellemezni, így minta áll rendelkezésünkre. A mintába kerülő elemek különböző mintavételi technikák segítségével választhatók ki. A véletlenen alapuló kiválasztás során az elemek mintába kerülési valószínűsége előre meghatározható, míg a nem véletlenen alapuló technikáknál a mintába kerülő elemeknek nagy a jelentősége a későbbi elemzés szempontjából, ezért előzetes kritériumokat fogalmazunk meg a mintavétellel szemben. A reprezentativitás kiemelten fontos, mert ez biztosítja, hogy a részleges adatfelvételből is megfogalmazhatók legyenek általános érvényű következtetések. Az összegyűjtött adatokat statisztikai sorokba, az összefüggő sorokat pedig a továbbiakban statisztikai táblákba rendezhetjük, melyeket címmel és forrásmegjelöléssel is szükséges ellátni.

A viszonyszámok és a leíró statisztika az egyszerű elemzési eszközök közé tartoznak. Az egymással összefüggő adatok hányadosaként képezhető viszonyszámok az élet számtalan területén jelen vannak, típusai könnyen felismerhetők. A sokaság vagy minta szerkezete megoszlási vagy koordinációs viszonyszámmal jellemezhető, összehasonlítás végezhető dinamikus, tervteljesítési, tervfeladat vagy területi összehasonlító viszonyszámokkal. Intenzitási viszonyszámmal különböző típusú adatok is összevethetők. A viszonyszámok megjelennek a munkaerő-piaci és pénzügyi, valamint demográfiai statisztikákban is. A leíró statisztika eszköztárának segítségével mennyiségi ismérv szerint tömören jellemezhetők az adatsorok. A középértékek – átlagok, módusz, kvantilisek - az adatsor tipikus értékeit, a szóródási mérőszámok – terjedelem, szórás, relatív szórás, átlagos különbség és átlagos eltérés - az adatok változékonyságát, az alakmutatók – aszimmetria, csúcsosság mérőszámai - a gyakorisági görbe alakját jellemzik.

A heterogén sokaságok bővebb elemzési lehetőséget rejtenek. A standardizálás módszerének alkalmazásakor a heterogenitást csoportképzéssel kezeljük, mely lehetővé teszi a sokaság egészének minőségi vagy területi és időbeli összehasonlítását is. A különbségfelbontás során az összehasonlított sokaságokat jellemző összetett viszonyszámok különbsége két tényezőre vezethető vissza. A részsokaságokra vonatkoztatott részviszonyszámok eltéréseire, valamint a sokaságok részsokaságok szerinti összetételének különbségeire. A standardizáláson alapuló indexszámítás módszere időbeli összehasonlításra alkalmazható. Az időbeli változást egyrészt a részviszonyszámok változása, másrészt a részsokaságok szerinti összetétel változása okozza. Az érték fogalmának statisztikai bevezetésével lehetővé válik a közvetlenül nem összegezhető heterogén adatsorok elemzése is. Az ár és mennyiség szorzataként meghatározható érték segítségével meghatározható egy vagy több termék esetén az árak, a mennyiségek és az érték változása. Az indexszámítás módszerét alkalmazzák a vállalati és nemzetgazdasági teljesítmény számszerűsítésénél, valamint a külkereskedelmi statisztikákban is.

Az ismérvek közötti kapcsolatok vizsgálata kiemelten fontos, mert gyakran vagyunk kíváncsiak arra, hogy egy adott jelenség kialakulásában mely tényezők és milyen szerepet töltenek be. A sztochasztikus kapcsolatokat aszerint különböztetjük meg, hogy milyen ismérvek között tárjuk fel az összefüggést. Az asszociáció két nem mennyiségi ismérv közötti kapcsolat létének és erősségének kimutatására alkalmas. A vegyes kapcsolat vizsgálata során egy mennyiségi és egy nem mennyiségi ismérv között nézhetjük meg van-e összefüggés, s az milyen erős, továbbá azt is számszerűsíthetjük, hogy a nem mennyiségi ismérv mennyiben befolyásolja a mennyiségi ismérv alakulását. A korrelációszámítás során két mennyiségi ismérv között a kapcsolat léte, erőssége és iránya is mérhető. Mennyiségi ismérvek lévén a kapcsolat természete matematikai formában is leírható, ez a regressziószámítás. A regresszió függvény az ismérvek között előzetesen feltételezett ok-okozati összefüggés feltárására alkalmas módszer.

Az idősorok vizsgálatához változatos eszköztár áll rendelkezésre. A dinamikus viszonyszámok, az átlagolás vagy az indexek csak a megfigyelt adatokban lévő változásokat tudják számszerűsíteni. Az idősoros adatok segítségével azonban lehetőség nyílik megfelelő technikákkal előrejelzések készítésére is. A determinisztikus idősorelemzés lényege, hogy az idősort tényezőkre bontja, s ezen tényezők hatását számszerűsíti. A komponenseket attól függően különböztetjük meg, hogy milyen időtávon érvényesül a hatásuk. Az idősor elemei additív és multiplikatív módon is összekapcsolódhatnak. Az alapirányzat tartósan érvényesülő tendencia, mely mozgóátlagolással, illetve lineáris vagy exponenciális trend segítségével írható fel. A rövid távú periodikus hullámzás, a szezonalitás kimutatására abszolút és relatív formában is van lehetőség. A szezonális eltérések az alapadatok mértékegységében fejezik ki a trendtől való eltéréseket, míg a szezonindexek százalékos formában. A középtávon érvényesülő konjunktúra hatás főként a gazdasági folyamatokban jelentős. Az idősor elengedhetetlen eleme a véletlen hatás, melyet maradékelven tudunk meghatározni. Az előrejelzésekhez minimálisan az alapirányzat és a szezonalitás ismerete szükséges.

A statisztika elemzés eredményei szemléletesen megjeleníthetők grafikus ábrák segítségével. A diagramok szerkesztésénél az egyszerűség és áttekinthetőség kritériuma lényeges, formai és tartalmi követelménye, hogy címe és forrásmegjelölés legyen, valamint a mértékegység is fel legyen tüntetve. A különböző típusú ábrák eltérő elemzési lehetőséget kínálnak, így az adatok megjelenítésének célorientáltnak kell lennie. Az oszlop-, sáv-, vonal- és sugárdiagram leginkább összehasonlításra használható, pontdiagram segítségével az összefüggések ábrázolhatók, míg a szerkezeti jellemzők illusztrálására a kör vagy tortadiagram a legalkalmasabb. A térkép a területi összehasonlítások hasznos segédeszköze lehet.

13.2 A STATISZTIKA, MINT TUDOMÁNYOS MÓDSZERTAN HASZNOSÍTHATÓSÁGA A GYAKORLATBAN

A statisztika olyan tudományos módszertan, amely gyakorlati tevékenység is. A módszerek alkalmazásához ugyanis megfelelő adatokra van szükség, ezért az adatgyűjtés precíz kivitelezése legalább olyan fontos, mint az elemzési eszközök szakszerű használata. Számos olyan jelenség van, melyről nem történik rendszeres adatszolgáltatás, ezért magunknak kell információkat gyűjteni és azokat elemezni. A statisztika nem csupán a tudományos kutatásokban alkalmazható, hanem hétköznapi folyamatok elemzésénél is hasznos lehet. Az élet számos területén találkozunk olyan problémákkal, melyeket a megismert statisztikai módszerekkel számszakilag is alátámasztva tudunk megoldani. Sokszor észre sem vesszük, hogy statisztikai eszközöket használunk, így például, ha az átlagfizetésekről vagy éppen a tanulmányi átlageredményről beszélgetünk, vagy amikor a moziban vagy iskolában a fiúk és lányok arányát taglaljuk, esetleg amikor a boltban elköltött összegeken gondolkodunk. A tananyag a különböző módszereket olyan gyakorlati példákon keresztül igyekezett bemutatni, melyek rávilágítanak arra, hogy a statisztika tudománya nem is feltétlenül tudatosan, de jelen van a mindennapjainkban, ezért az elemzési eszközök elsajátításával a gyakorlati életben is hasznosítható tudásra tehetünk szert.

14. KIEGÉSZÍTÉSEK

14.1 IRODALOMJEGYZÉK

Könyv

BREALEY, R. A. – MYERS, S. C.:*Principles of Corporate Finance*. Fourth Edition. McGraw Hill, 1991.

HOLLÓNÉ KACSÓ ERZSÉBET: *Vállalatértékelés*. Mutatók, modellek, látható- és láthatatlan vagyonértékelési eljárások. Főiskolai jegyzet, Eger, Eszterházy Károly Főiskola, 2011.

ILYÉSNÉ MOLNÁR EMESE – LOVASNÉ AVATÓ JUDIT: *Statisztika feladatgyűjtemény I.* Budapest, Perfekt Kiadó, 2006.

ILYÉSNÉ MOLNÁR EMESE – LOVASNÉ AVATÓ JUDIT: *Statisztika feladatgyűjtemény II.* Budapest, Perfekt Kiadó, 2006.

KERÉKGYÁRTÓ GYÖRGYNÉ – MUNDRUCZÓ GYÖRGY – SUGÁR ANDRÁS: *Statisztikai módszerek és alkalmazásuk a gazdasági, üzleti elemzésekben*. Budapest, Aula Kiadó, 2001.

KORPÁS ATTILÁNÉ DR. (szerk.): *Általános statisztika I.* Budapest, Nemzeti Tankönyvkiadó, 1996.

KORPÁS ATTILÁNÉ DR. (szerk.): *Általános statisztika II.* Budapest, Nemzeti Tankönyvkiadó, 1997.

SÁNDORNÉ DR. KRISZT ÉVA – DR. CSESZNÁK ANITA – ORSZÁG GÁBORNÉ: *Statisztika I.* Budapest, Nemzedékek tudása Tankönyvkiadó, 2013.

SOLT KATALIN: *Makroökonómia.* Tatabánya, TRI-Mester, 2001.

SZŰCS ISTVÁN (szerk): *Alkalmazott statisztika*. Budapest, AGROINFORM Kiadó és Nyomda Kft, 2002.

Elektronikus dokumentumok / források

EURÓPAI BIZOTTSÁG: *Eurostat* [adatbázis] [2014. július] <URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>

EUROPEAN COMMISSION: *Union Innovation Scoreboard* [adatbázis] [2014. június] <URL: http://ec.europa.eu/enterprise/policies/innovation/policy/innovation-scoreboard/index_en.htm>

FEDERAL RESERVE BANK OF ST. LOUIS (FRED) Economic Research: *Dow Jones Industrial Average* (DJIA) [2014. augusztus 2.] <URL: http://research.stlouisfed.org/fred2/series/DJIA/downloaddata>

KÖZPONTI STATISZTIKAI HIVATAL (KSH) *statisztikái* [adatbázis] [2014. július] <URL: http://www.ksh.hu/>

KÖZPONTI STATISZTIKAI HIVATAL: *Magyarország társadalmi atlasza*. Budapest, 2012. [elektronikus dokumentum] 2014. augusztus 3.] <URL: https://www.ksh.hu/docs/hun/xftp/idoszaki/pdf/tarsatlasz.pdf>

KÖZPONTI STATISZTIKAI HIVATAL: *Áralakulás*. [elektronikus dokumentum] 2014. augusztus 19.] < URL: https://www.ksh.hu/thm/1/indi1_2_3.html>

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD) *statisztikái* [adatbázis] [2014. július] <URL: http://stats.oecd.org/>

TERÜLETI ELEKTRONIKUS INFORMÁCIÓS RENDSZER (TEIR) KSH T-STAR és Népszámlálás [adatbázis] [2014. augusztus 3.] <URL: http://www.terport.hu/kartogramok/a-szemelygepkocsi-allomany-1992-es-2008-kozotti-ill-a-magyarorszagon-elso-alkalommal-for>

VILÁGBANK *statisztikái* [adatbázis] [2014. július] <URL: http://data.worldbank.org/>

ALFAHÍR HÍRPORTÁL: Magyarország kullancsfertőzött helyei – fotó (2009. június 15) [2014. augusztus 3.] <URL: http://alfahir.hu/node/31010>

14.2 MÉDIAELEMEK ÖSSZESÍTÉSE

14.2.1 Táblázatjegyzék

1. Példa az ismérvek típusaira 10
2. Magyarország turizmusának néhány adata 2013-ban 20
3. A Magyarországra érkező külföldiek számának alakulása 2011 és 2013 között 20
4. A Magyarországra érkező külföldiek számának alakulása 2013 egyes negyedéveiben 21
5. A hallgatók osztályzatok szerinti megoszlása 21
6. A Magyarországra érkező francia és német turisták száma és az általuk átlagosan eltöltött idő 2013-ban 23
7. A Magyarországra érkező külföldiek számának alakulása az utazás turisztikai célja szerint 2012-ben és 2013-ban 24
8. Magyarországra érkező turisták száma nem és származási helye szerint 24
9. A statisztika csoport megoszlásának kiszámítása 29
10. Munkatábla bázisviszonyszám számításához, ha a 2010-es évet tekintjük bázisidőszaknak 32
11. Munkatábla bázisviszonyszám számításához, ha a 2012-es évet tekintjük bázisidőszaknak 32
12. Munkatábla láncviszonyszám számításához 33
13. Munkatábla gyakorisági sor készítéséhez és értelmezéséhez 52
14. Munkatábla osztályközös gyakoriság sor adatainak értelmezéséhez 53
15. Munkatábla az osztályközepek kiszámításához 54

16. A szálloda vendégforgalmának átlagos változása 56
17. A vállalatnál dolgozó fizikai és szellemi foglalkozásúak átlagbérei 71
18. A statisztika csoport zárthelyi dolgozaton elért átlagos eredményei.. 74
19. Az évfolyam tanulmányi átlageredményei a 2013/2014. tanévben 77
20. Az elfogyasztott mennyiség, az egységárak és a termékek megvásárlására fordított összeg változása márciusról áprilisra.. 86
21. A statisztika vizsga eredményessége és az órai aktivitás megoszlása.. 106
22. A feladatsorokra vonatkozó információk................................... 110
23. A csoport átlagos eredményei feladatsoronként és együttesen.. 111
24. Munkatábla a rangkorrelációhoz szükséges négyzetösszeg kiszámításához 119
25. Munkatábla a rangkorrelációhoz szükséges négyzetösszeg kiszámításához 120
26. Munkatábla a Kendall-féle rangmódszer alkalmazásához.......... 122
27. Munkatábla a regresszió függvény paramétereinek meghatározásához 126
28. Egy vállalat által értékesített termékek száma 2009 és 2013 között.. 140
29. A szállodalánc vendégforgalmának szezonális alakulása 2010 és 2013 között.. 143
30. Munkatábla a mozgó átlagok kiszámításához 144
31. A vállalat értékesítésének alakulása 2011 és 2013 között negyedéves bontásban... 145
32. Munkatábla a centrírozott mozgó átlagok kiszámításához 146
33. Példa a t értékek $\sum t = 0$ módszer alapján történő hozzárendelésére 148
34. Munkatábla a lineáris trend meghatározásához 149
35. Munkatábla az exponenciális trend meghatározásához 152
36. A tényleges és a mozgóátlagolású trend értékeinek páronkénti hányadosai és különbségei... 154
37. Munkatábla a szezonális eltérések meghatározásához mozgóátlagolás esetén 155
38. Munkatábla a szezonindexek meghatározásához mozgóátlagolás esetén 156
39. A tényleges és a lineáris trendből becsült értékek páronkénti különbségei és hányadosai.. 156
40. Munkatábla szezonális eltérések meghatározásához lineáris trend esetén... 157

41. Munkatábla a szezonindexek meghatározásához lineáris trend esetén ... 157
42. A tényleges és az exponenciális trendből becsült értékek páronkénti különbségei és hányadosai ... 158
43. Munkatábla a szezonális eltérések meghatározásához exponenciális trend esetén ... 158
44. Munkatábla a szezonindexek meghatározásához exponenciális trend esetén ... 159
45. Munkatábla az előrejelzés készítéséhez – mozgóátlagolás és additív szezonalitás ... 160
46. Munkatábla az előrejelzés készítéséhez – mozgóátlagolás és multiplikatív szezonalitás ... 160
47. Munkatábla az előrejelzés készítéséhez – lineáris trend és additív szezonalitás ... 161
48. Munkatábla az előrejelzés készítéséhez – lineáris trend és multiplikatív szezonalitás ... 161
49. Munkatábla az előrejelzés készítéséhez – exponenciális trend és additív szezonalitás ... 162
50. Munkatábla az előrejelzés készítéséhez – exponenciális trend és multiplikatív szezonalitás ... 162

14.2.2 Ábrajegyzék

1. ábra: Az általános és gazdasági statisztika tananyag logikai struktúrája ... 8
2. ábra: A statisztikai adatszerzés formái ... 15
3. ábra: A statisztikai sorok és táblák típusai ... 19
4. ábra: A viszonyszámok rendszere ... 28
5. ábra: A viszonyszámok alkalmazásának főbb területei ... 41
6. ábra: Az össznépesség felosztása munkapiaci szempontból ... 42
7. ábra: Az aktivitási ráta alakulása néhány országban 2000 és 2013 között ... 43
8. ábra: Németország foglalkoztatási és munkanélküliségi rátáinak alakulása 2004 és 2013 között ... 44
9. ábra: A leíró statisztika eszköztára ... 50
10. ábra: A nevezetes osztópontok elhelyezkedése az adatsorban ... 58
11. ábra: A szórás és relatív szórás kiszámítási módjai sokaság és minta esetén ... 61
12. ábra: A koncentráció vizsgálata Lorenz görbével ... 63
13. ábra: Az empirikus eloszlások típusai ... 64
14. ábra: A szimmetrikus és aszimmetrikus eloszlások tulajdonságai65

15. ábra: Az összetett viszonyszámok (főátlagok) összehasonlításának módszerei 70
16. ábra: A különbségfelbontás tényezői és kapcsolódásuk 72
17. ábra: A standardizáláson alapuló indexszámítás tényezői és kapcsolódásuk 76
18. ábra: Az értéken alapuló indexszámítás összefüggésrendszere 83
19. ábra: A termékek megvásárlására fordított összeg az egyes hónapokban (09_07_K02)... 85
20. ábra: Az együttes ár- és volumenindexek kiszámítási formái..... 87
21. ábra: Példa a fiktív aggregátumok kiszámítására...................... 88
22. ábra: A két település árainak és eladásra szánt mennyiségeinek összehasonlítása 89
23. ábra: A területi indexek kiszámításához szükséges aggregátumok értékei 90
24. ábra: Az indexek gyakorlati alkalmazási területei...................... 95
25. ábra: Az USA folyó áras és változatlan (2005. évi) áron számított GDP adatainak alakulása 2005 és 2012 között........................... 97
26. ábra: Az infláció mérésére alkalmas GDP deflátor és fogyasztói árindex összehasonlítása az USA 2004 és 2013 közötti adatain.. 99
27. ábra: A kapcsolatvizsgálat alapesetei 104
28. ábra: Az asszociáció és a vegyes kapcsolat mérőszámai........ 104
29. ábra: A csoport megoszlása a vizsga sikeressége és az előadásra járás gyakorisága alapján 105
30. ábra: A függetlenség esetén feltételezett gyakoriságok kiszámítása.. 107
31. ábra: A χ^2 kiszámítása 108
32. ábra: A belső, külső és teljes eltérés, valamint szórásnégyzetek kiszámítási formái (09_09_K06)... 109
33. ábra: A korreláció- és regressziószámítás struktúrája............. 116
34. ábra: A statisztika vizsgára való felkészülés ideje és a dolgozat eredményessége közötti kapcsolat 125
35. ábra: A statisztika vizsgára való felkészülés ideje és a dolgozat eredményessége közötti kapcsolatot leíró lineáris regresszió függvény (09_10_K03).. 128
36. ábra: A statisztika vizsgára való felkészülés ideje és a dolgozat eredményessége közötti kapcsolatot leíró exponenciális regresszió függvény (09_10_K04).. 131
37. ábra: A statisztika vizsgára való felkészülés ideje és a dolgozat eredményessége közötti kapcsolatot leíró hatványkitevős regresszió függvény (09_10_K05) 132
38. ábra: Idősorelemzési technikák 138
39. ábra: Az idősor elemeinek kapcsolódása................................ 141

40. ábra: A szálloda vendégforgalmának szezonális alakulása 2010 és 2013 között 143
41. ábra: A vállalat értékesítésének alakulása 2011 és 2013 között negyedéves bontásban (09_11_K04) 145
42. ábra: A vállalat értékesítésének alakulása 2011 és 2013 között negyedéves bontásban és az illeszkedő lineáris trendfüggvény (t=1,2 …n) 150
43. ábra: A vállalat értékesítésének alakulása 2011 és 2013 között negyedéves bontásban és az adatokra illeszkedő exponenciális trendfüggvény (t=1,2 …n) (09_ 153
44. ábra: A grafikus ábrák típusai 169
45. ábra: Az egy főre jutó euróban kifejezett K+F ráfordítások adatai néhány európai országban 2012-ben 170
46. ábra: A 15-64 éves korosztályból foglalkoztatottak számának (ezer fő) alakulása Franciaországban 2004 és 2013 között 170
47. ábra: Három vállalat dolgozóinak iskolai végzettség szerinti megoszlása 171
48. ábra: Egy vállalatnál dolgozók iskolai végzettség szerinti összetételének alakulása 2010 és 2012 között 172
49. ábra: A magyar népesség korfája 2011-ben 173
50. ábra: Magyarország népessége korcsoportonként 2011-ben ... 174
51. ábra: A DOW Jones index alakulása 2014. május 28 és július 31 között 175
52. ábra: Az egy főre jutó GDP és a K+F-re fordított kiadások közötti összefüggés az Európai Unió tagállamaiban 176
53. ábra: Európa és az USA teljesítménye az összetett innovációs mérőszám egyes részterületein 177
54. ábra: Egy vállalatnál dolgozók iskolai végzettség szerinti megoszlása 178
55. ábra: 2001 és 2008 között első alkalommal forgalomba helyezett személygépkocsik száz lakosra jutó száma Magyarországon 179
56. ábra: Magyarország lakónépesség számának változása és annak összetevői a 2001–2011 közötti időszakban 179
57. ábra: Magyarország kullancsfertőzött helyei – fotó (2009. június 15) 180

14.2.3 Külső URL hivatkozások

1. Eurostat: http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/12
2. OECD adatbázisa: http://stats.oecd.org/ 12
3.ENSZ adatbázisa: http://www.un.org/en/databases/ 12

4.Világbank adatbázisa: http://data.worldbank.org/ 12
5.Nemzetközi Valutaalap adatbázisa:
https://www.imf.org/external/data.htm 13
6.International Financial Statistics: http://www.econdata.com/
databases/imf-and-other-international/ifs/ 13
7.Nemzetközi Munkaügyi Szervezet adatbázisa:
http://www.ilo.org/global/statistics-and-databases/lang--
en/index.htm 13
8.United Nations Conference on Trade and Development:
http://unctad.org/en/Pages/Publications/Handbook-of-
Statistics.aspx 13
9. Az Európai Központi Bank honlapja:
http://www.ecb.europa.eu/stats/exchange/eurofxref/html/index.en.h
tml... 45
10. P/E: http://www.investopedia.com/terms/p/price-earningsratio.asp . 45
11. ROE: http://www.investopedia.com/terms/r/returnonequity.asp....... 46
12. ROA: http://www.investopedia.com/terms/r/returnonassets.asp 46
13. ROS: http://www.investopedia.com/terms/r/ros.asp........................ 46
14. Pénzügyi mutatók:
http://kfknowledgebank.kaplan.co.uk/KFKB/Wiki%20Pages/Financi
al%20Performance%20Indicators%20%28FPIs%29.aspx 46
15. Világbank demográfiai mutatói:
http://econ.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS
/0,,contentMDK:20451597~pagePK:64133150~piPK:64133175~the
SitePK:239419,00.html 47
16. A GDP és az infláció kapcsolata:
http://www.investopedia.com/articles/06/gdpinflation.asp............ 98
17. Infláció, defláció, stagfláció: http://www.investopedia.com/exam-
guide/cfp/economics-time-value/cfp6.asp 100
18. World Trade Organization:
http://www.wto.org/english/res_e/statis_e/statis_e.htm 100
19. Üzleti ciklusok:
http://www.britannica.com/EBchecked/topic/86233/business-
cycle 141
20. Kitchin ciklus (3-5 év): http://www.policonomics.com/joseph-
kitchin/ 141
21. Juglar ciklus (7-11 év): http://www.policonomics.com/clement-
juglar/... 141
22. Kuznets ciklus (15-25 év):
http://www.dictionaryofeconomics.com/article?id=pde2008_
K000045 141

23. Kondratyev ciklus (45-60): http://www.newworldencyclopedia.org/entry/Nikolai_ Kondratiev 141
24. Korfa: http://www.ehow.com/list_6872808_three-different-types-population-pyramids.html 172

15. TESZTEK

GYAKORLÓTESZTEK

1. A kávéfogyasztási szokásokról készít felmérést. Milyen mintavételi technikát alkalmazna az adatgyűjtéshez? Több válasz is megjelölhető!

Véletlenen alapuló		**Nem véletlenen alapuló**	
X	Független, azonos eloszlású		Kvótás
X	Egyszerű véletlen	X	Hólabda
	Rétegzett		Koncentrált
	Csoportos		Szisztematikus
	Többlépcsős		Önkényes

2. A tévénézési szokásokról készít felmérést. Milyen mintavételi technikát alkalmazna az adatgyűjtéshez? *Több válasz is megjelölhető!*

Véletlenen alapuló		**Nem véletlenen alapuló**	
	Független, azonos eloszlású		Kvótás
X	Egyszerű véletlen	X	Hólabda
	Rétegzett		Koncentrált
	Csoportos	X	Szisztematikus
	Többlépcsős		Önkényes

3. A minta reprezentativitása az általános érvényű következtetések levonása miatt lényeges.

   IGAZ HAMIS

4. A csoportosító táblában csak csoportosító sorok vannak.

   IGAZ **HAMIS**

5. Az euró árfolyama 1,35 EUR/USD. Milyen típusú ez a viszonyszám?

Csoportosító sorból képzett	**Összehasonlító sorból képzett**
Megoszlási	Dinamikus
Koordinációs	Területi összehasonlító
Leíró sorból képzett	Tervfeladat

X	Intenzitási		Tervteljesítési

6. Egy vállalat árbevétele tavalyhoz képest 30%-kal emelkedett. Milyen típusú ez a viszonyszám?

Csoportosító sorból képzett		Összehasonlító sorból képzett	
	Megoszlási	X	Dinamikus
	Koordinációs		Területi összehasonlító
Leíró sorból képzett			Tervfeladat
	Intenzitási		Tervteljesítési

7. Egy ruházati bolt értékesítésének 30%-a nadrágok eladásából származik. Milyen típusú ez a viszonyszám?

Csoportosító sorból képzett		Összehasonlító sorból képzett	
X	Megoszlási		Dinamikus
	Koordinációs		Területi összehasonlító
Leíró sorból képzett			Tervfeladat
	Intenzitási		Tervteljesítési

8. 2010-ben az egy főre jutó évi átlagos csokoládéfogyasztás 28 kg volt. Milyen típusú ez a viszonyszám?

Csoportosító sorból képzett		Összehasonlító sorból képzett	
	Megoszlási		Dinamikus
	Koordinációs		Területi összehasonlító
Leíró sorból képzett			Tervfeladat
X	Intenzitási		Tervteljesítési

9. Az össznépesség munkapiaci szempontból gazdaságilag aktív és gazdaságilag inaktív csoportokra osztható.

IGAZ **HAMIS**

10. A foglalkoztatási ráta a foglalkoztatottak népességen belüli arányát mutatja.

IGAZ **HAMIS**

11. A népsűrűség kiszámításánál az ország területéhez viszonyítjuk a lakosság számát.

IGAZ HAMIS

12. A P/E mutató az adózott eredményt viszonyítja a saját tőkéhez.

IGAZ **HAMIS**

13. Egy adott napon megfigyeltek 100 embert, akik egy hipermarketben vásároltak, s azt tapasztalták, hogy a vásárlók fele 25 €-nál többet, fele ennél kevesebbet költött. Melyik alapeloszlás jellemző értéke a 25 €?

Középértékek

	Átlag
	Módusz
X	Medián
	Alsó kvartilis
	Felső kvartilis

Szóródás és alakmutatók

	Terjedelem
	Szórás
	Relatív szórás
	A mutató
	F mutató

14. Egy adott napon megfigyeltek 100 embert, akik egy hipermarketben vásároltak, s azt tapasztalták, hogy a vásárlók átlagosan 24 €-t költöttek az üzletben. Melyik alapeloszlás jellemző értéke a 24 €?

Középértékek

X	Átlag
	Módusz
	Medián
	Alsó kvartilis
	Felső kvartilis

Szóródás és alakmutatók

	Terjedelem
	Szórás
	Relatív szórás
	A mutató
	F mutató

15. A módusz érzékeny az adatsorban lévő kiugróan alacsony vagy kiugróan magas értékekre.

IGAZ **HAMIS**

16. A kvantilisek az adatsor nevezetes osztópontjai.

IGAZ HAMIS

17. Egy vállalatnál dolgozók kereseti viszonyainak ismeretében megállapítható, hogy a férfiak átlagosan 30 000 Ft-tal többet keresnek, mint a nők. Tudjuk, hogy a férfiak és nők iskolai végzettség szerinti megoszlásának eltérése miatt a férfiak átlagosan csak 24 000 Ft-tal keresnének többet. A férfiak és nők között az iskolai végzettségenkénti átlagkeresetek különbözősége miatt

– a férfiak 6000 Ft-tal kevesebbet keresnének, mint a nők
– **a férfiak 6000 Ft-tal többet keresnének, mint a nők**
– a nők 6000 Ft-tal többet keresnének, mint a férfiak

18. Az időbeli összehasonlításra alkalmas indexszámítás során a viszonyítás iránya szabadon megválasztható.

IGAZ **HAMIS**

19. A főátlag index a részátlag és az összetételhatás index szorzata.

IGAZ HAMIS

20. Két kiemelt üdülőkörzet adatait megnézve azt láthatjuk, hogy a tavalyi évhez képest 10%-kal esett vissza a vendégek átlagos tartózkodási ideje. Az osztrák, német és lengyel turisták átlagos tartózkodási ideje 2%-kal emelkedett, de a különböző nemzetiségű vendégek megoszlása az egyes években olyannyira megváltozott, hogy ez összességében az átlagos tartózkodási idők mintegy 11,76%-os viszszaesését okozta.
    - a főátlag index értéke 110,0
    - **a részátlag index értéke 102,0**
    - az összetételhatás-index értéke 111,76
21. Melyik nem tartozik az érték kategóriába?

    forgalom **eladási ár** árbevétel termelési érték
22. Ismert, hogy a bázissúlyozású árindex értéke 107,82, a láncsúlyozású árindex értéke 109,12. A Fisher index értéke:

    117,65 106, 48 **108,47**
23. A Fisher-féle indexeket a bázis és tárgyi súlyozású indexek mértani átlagaként számítjuk ki.

    IGAZ HAMIS
24. Az egyedi árindex az egyedi volumenindex és az egyedi értékindex hányadosa.

    IGAZ **HAMIS**
25. A termelékenység megmutatja
    - az egy termékegységre jutó dolgozók számát
    - **az egy dolgozóra jutó termékek számát**
    - az átlagosan megtermelt termékmennyiséget
26. A GDP értékindexe
    - **a nominál GDP változása**
    - a reál GDP változása
    - a nominál és reál GDP hányadosa
27. Az infláció mérésére alkalmazott fogyasztói árindex számítása során súlyként használjuk
    - a tárgyévi mennyiségeket
    - **az előző évi mennyiségeket**
    - az előző évi árakat

28. A csereárány mutató
   - **az exportált és importált termékek árindexének hányadosa**
   - az importált és exportált termékek árindexének hányadosa
   - az export és import hányadosa

29. A hallgató szakja és zárthelyi dolgozatának eredménye között a kapcsolat milyen mutatóval számszerűsíthető?

   Yule Csuprov **Szóráshányados**

30. Egy vállalatnál a dolgozók iskolai végzettsége és keresete közötti kapcsolat vizsgálata során azt az eredményt kaptuk, hogy a szórásnégyzet-hányados értéke 0,674. Mit jelent ez?
   - a két ismérv között közepesnél erősebb kapcsolat mutatható ki
   - **az iskolai végzettség 67,4%-ban határozza meg a kereset alakulását**
   - az iskolai végzettségenkénti keresetek a főátlagtól átlagosan 0,674-el térnek el.

31. A H mutató értéke a vonat típusa és a késés száma között 0,26. Mire következtetünk ebből?
   - a vonat típusa 0,26%-ban befolyásolja a késések számát
   - **a vonat típusa és a késések száma között gyenge kapcsolat mutatható ki**
   - A vonat típusa 26%-ban befolyásolja a késések számát

32. A Csuprov mutató értéke a dolgozók beosztása és neme között 0,28. Milyen erősségű a kapcsolat?

   Gyenge Közepesnél gyengébb Erős

33. A kovariancia -1 és 1 közötti értéket vehet fel, s előjele a kapcsolat irányát jelzi.

   IGAZ **HAMIS**

34. A Spearman féle mérőszám arányskálán mérhető ismérvek közötti kapcsolat erősségét határozza meg.

   IGAZ **HAMIS**

35. Az r^2 = 0,52 azt jelenti, hogy a magyarázó változó 0,52%-ban befolyásolja az eredményváltozó alakulását.

   IGAZ **HAMIS**

36. A többváltozós regressziós modellek feltétele, hogy a magyarázó és az eredményváltozó egymástól független legyen.

IGAZ **HAMIS**

37. Egy étterem éves forgalmát (ezer Ft) 2000 és 2013 között leíró lineáris trend egyenlete y=620+12t. A trend alapján az alábbi következtetés vonható le:

- az étterem forgalma a vizsgált időszakban csökkent
- az étterem forgalma 2000-ben 620 ezer Ft volt
- **az étterem forgalma a vizsgált időszakban évente átlagosan 12 ezer Ft-tal nőtt**

38. Az a trend illeszkedik legjobban, ahol a reziduális szórás értéke a legkisebb.

IGAZ HAMIS

39. A lineáris trendből és additív szezonalitás alapján történő előrejelzések szerint 2015. II. negyedévében egy vállalat várhatóan 6500 db terméket fog értékesíteni. A lineáris trend alapján azonban 6200 db lenne. Mekkora a II. negyedévek szezonalitása?

- a II. negyedévekben átlagosan 300 db-bal alacsonyabb az értékesítés
- **a II. negyedévekben átlagosan 300 db-bal magasabb az értékesítés**
- a II. negyedévekben átlagosan 4,82%-kal magasabb az értékesítés

40. Lineáris trend és additív szezonalitás esetén a nyers és korrigált szezonindexek megegyeznek.

IGAZ **HAMIS**

41. Időbeli összehasonlítás grafikus illusztrálására az oszlop- és a vonaldiagram a legalkalmasabb.

IGAZ HAMIS

42. A pontdiagramot leggyakrabban összefüggés vizsgálatokhoz készítjük, a pontokból kirajzolódik az összetartozó értékpárokra leginkább illeszkedő alakzat.

IGAZ HAMIS

43. A korfa a népességstatisztikában alkalmazott oszlopdiagram.

IGAZ **HAMIS**

44. A kartodiagram a térképen eltérő színekkel jelzi a vizsgált jelenség területi intenzitását.

IGAZ **HAMIS**

PRÓBAVIZSGA

1. Egy vállalatnál dolgozók fizetéssel való elégedettségéről készít felmérést. Milyen mintavételi technikát alkalmazna az adatgyűjtéshez? Több válasz is megjelölhető!

	Véletlenen alapuló		**Nem véletlenen alapuló**
	Független, azonos eloszlású	X	Kvótás
	Egyszerű véletlen		Hólabda
X	Rétegzett		Koncentrált
	Csoportos		Szisztematikus
	Többlépcsős		Önkényes

2. A statisztikai tábla dimenziószáma megmutatja, hogy a táblában egy adott érték hány ismérvhez tartozik.

IGAZ HAMIS

3. A tavalyi év kimagasló eredményeinek köszönhetően a vállalat a jövő évre további 10%-os árbevétel növekedéssel számol. Milyen típusú ez a viszonyszám?

	Csoportosító sorból képzett		**Összehasonlító sorból képzett**
	Megoszlási		Dinamikus
	Koordinációs		Területi összehasonlító
	Leíró sorból képzett	X	Tervfeladat
	Intenzitási		Tervteljesítési

4. Az egy főre jutó GDP az Egyesült Államokban majdnem ötször akkora, mint Magyarországon. Milyen típusú ez a viszonyszám?

	Csoportosító sorból képzett		**Összehasonlító sorból képzett**
	Megoszlási		Dinamikus
	Koordinációs	X	Területi összehasonlító
	Leíró sorból képzett		Tervfeladat
	Intenzitási		Tervteljesítési

5. Gazdaságilag inaktív az, aki nem vesz részt a társadalmilag szervezett munkavégzésben.

IGAZ HAMIS

6. Az orvos ellátottság mutatójának meghatározásához a lakosok számát viszonyítjuk az orvosok számához.

IGAZ **HAMIS**

7. Egy adott napon megfigyeltek 100 embert, akik egy hipermarketben vásároltak, s azt tapasztalták, hogy az egyes vásárlók által az üzletben elköltött összegek az átlagos vásárlási értéktől átlagosan 4 €-val térnek el. Melyik alapeloszlás jellemző értéke a 4 €?

Középértékek

	Átlag
	Módusz
	Medián
	Alsó kvartilis
	Felső kvartilis

Szóródás és alakmutatók

	Terjedelem
X	Szórás
	Relatív szórás
	A mutató
	F mutató

8. A szórás az alapadatok mértékegységében, a relatív szórás pedig százalékban fejezi ki az adatok egymástól vett átlagos eltérését.

IGAZ **HAMIS**

9. Két kiemelt üdülőkörzet adatait megnézve azt láthatjuk, hogy a tavalyi évhez képest 10%-kal esett vissza a vendégek átlagos tartózkodási ideje. Az osztrák, német és lengyel turisták átlagos tartózkodási ideje 2%-kal emelkedett, de a különböző nemzetiségű vendégek megoszlása az egyes években olyannyira megváltozott, hogy ez összességében az átlagos tartózkodási idők mintegy 11,76%-os viszszaesését okozta.
   - **a főátlag index értéke 90,0**
   - a részátlag index értéke 98,0
   - az összetételhatás-index értéke 111,76

10. Ha az összehasonlított két sokaság összetétele megegyezik, akkor a részsokaságok szerinti összetétel különbözőségének hatása 1.

IGAZ **HAMIS**

11. Az egyedi árindex értéke 100, az egyedi volumenindex értéke 110. Az egyedi értékindex értéke:
   - **110**
   - 90,1
   - 100

12. Az együttes indexek esetében teljesül az az összefüggés, hogy az együttes értékindex az ellentétes súlyozású ár- és volumenindex szorzata.

    IGAZ HAMIS

13. A fajlagos anyagfelhasználás
    - **az összesen felhasznált anyag és az előállított termékmennyiség hányadosa**
    - az összesen felhasznált anyag és az előállított termékmennyiség szorzata
    - az előállított termékmennyiség és az összesen felhasznált anyag hányadosa
14. A nominál GDP
    - **adott évi mennyiségek és árak szorzata**
    - adott évi árak és egy, a korábbi évekre jellemző mennyiség szorzata
    - adott évi mennyiségek és egy, a korábbi évekre jellemző árak szorzata
15. A vizsga sikeressége (sikeres/sikertelen) és a hallgató lakhelye (főváros/kisváros/falu) közötti kapcsolat milyen mutatóval számszerűsíthető?
    - Yule
    - **Csuprov**
    - Szóráshányados
16. Egy vállalatnál a dolgozók iskolai végzettsége és keresete közötti kapcsolat vizsgálata során azt az eredményt kaptuk, hogy a szóráshányados értéke 0,82. Mit jelent ez?
    - **a két ismérv között erős kapcsolat mutatható ki**
    - az iskolai végzettség 82%-ban határozza meg a kereset alakulását
    - a kereset 82%-ban határozza meg az iskolai végzettséget.
17. Az a regresszió függvény illeszkedik legjobban a tapasztalati értékekre, amelynél a reziduális szórás értéke a legnagyobb.

    IGAZ **HAMIS**

18. Hatványkitevős regresszió függvény esetén a rugalmassági együttható megegyezik a függvény *b* paraméterének értékével.

IGAZ HAMIS

19. Egy étterem 2014-es utószezonjának forgalmára vonatkozóan lineáris trend alapján 1 400 ezer Ft várható, továbbá tudjuk, hogy utószezonban átlagosan 10%-kal alacsonyabb a forgalom, mint a becsült érték. Az előrejelzés szerint 2014. főszezonjában a szezonalitás figyelembevételével számított forgalom

– 1540 ezer Ft
– **1260 ezer Ft**
– 1410 ezer Ft

20. Exponenciális trend és multiplikatív szezonalitás esetén a nyers szezonindexek szorzata 1.

IGAZ HAMIS

21. A hisztogram főként osztályközös gyakorisági sorok szemléltetésére alkalmas speciális oszlopdiagram, ahol nincs térköz az oszlopok között.

IGAZ HAMIS

22. A kartogram a térképen eltérő színű diagramokkal jelzi a vizsgált jelenség alakulását.

IGAZ **HAMIS**

ZÁRÓVIZSGA

A FELADATSOR

23. Az autóvásárlással kapcsolatos tapasztalatokról készít felmérést. Milyen mintavételi technikát alkalmazna az adatgyűjtéshez? *Több válasz is megjelölhető!*

	Véletlenen alapuló		**Nem véletlenen alapuló**
	Független, azonos eloszlású		Kvótás
X	Egyszerű véletlen		Hólabda
	Rétegzett	X	Koncentrált
	Csoportos		Szisztematikus
	Többlépcsős		Önkényes

24. Különböző típusú adatokból összehasonlító vagy csoportosító sorok képezhetők.

IGAZ **HAMIS**

25. Egy múzeum látogatottsági adatai alapján egy férfi látogatóra két nő jutott. Milyen típusú ez a viszonyszám?

	Csoportosító sorból képzett		**Összehasonlító sorból képzett**
	Megoszlási		Dinamikus
X	Koordinációs		Területi összehasonlító
	Leíró sorból képzett		Tervfeladat
	Intenzitási		Tervteljesítési

26. Egy vállalat a tervezetthez képest a válság miatt 25%-os visszaesést produkált. Milyen típusú ez a viszonyszám?

	Csoportosító sorból képzett		**Összehasonlító sorból képzett**
	Megoszlási		Dinamikus
	Koordinációs		Területi összehasonlító
	Leíró sorból képzett		Tervfeladat
	Intenzitási	X	Tervteljesítési

27. A munkanélküliségi ráta számításánál csak a regisztrált munkanélkülieket tudjuk figyelembe venni.

IGAZ HAMIS

28. A ROA mutató az összes eszközhöz viszonyítja az adózott eredményt.

IGAZ HAMIS

29. Egy adott napon megfigyeltek 100 embert, akik egy hipermarketben vásároltak, s azt tapasztalták, hogy a legtöbb vásárló 20 € értékben vásárolt. Melyik alapeloszlás jellemző értéke a 20 €?

Középértékek

	Átlag
X	Módusz
	Medián
	Alsó kvartilis
	Felső kvartilis

Szóródás és alakmutatók

	Terjedelem
	Szórás
	Relatív szórás
	A mutató
	F mutató

30. A Lorenz görbe a koncentráció vizsgálatára alkalmas.

IGAZ HAMIS

31. Egy vállalatnál dolgozók kereseti viszonyainak ismeretében megállapítható, hogy a férfiak átlagosan 30 000 Ft-tal többet keresnek, mint a nők. Tudjuk, hogy a férfiak és nők iskolai végzettségenkénti átlagkeresetek különbözősége miatt a férfiak csak 6 000 Ft-tal keresnének többet. A férfiak és nők iskolai végzettség szerinti megoszlásának eltérése miatt

– a férfiak 24 000 Ft-tal kevesebbet keresnének, mint a nők
– **a férfiak 24 000 Ft-tal többet keresnének, mint a nők**
– a nők 24 000 Ft-tal többet keresnének, mint a férfiak

32. A részátlagok különbözőségének hatását úgy tudjuk kimutatni, hogy változatlannak tekintjük a részsokaságok összetételét.

IGAZ HAMIS

33. Egy zöldséges valamennyi termékének együttes értékesítéséből befolyt bevétele tavalyhoz képest 10%-kal csökkent.

– az egyedi értékindex 110,0
– **az együttes értékindex értéke 90,0**
– az együttes volumenindex értéke 90,0

34. Az értéket ár és mennyiség szorzataként kapjuk.

IGAZ HAMIS

35. A reál GDP

– adott évi mennyiségek és árak szorzata
– adott évi árak és egy, a korábbi évekre jellemző mennyiség szorzata
– **adott évi mennyiségek és egy, a korábbi évekre jellemző árak szorzata**

36. Az infláció mérésére alkalmazott GDP deflátor számítása során súlyként használjuk
    - **a tárgyévi mennyiségeket**
    - az előző évi mennyiségeket
    - a tárgyévi árakat
37. A Cramer mutató értéke két ismérv között 0,64. Melyik lehet a két ismérv?
    - a vendég származási helye és tartózkodási ideje
    - **a vendég származási helye és a szállás típusa**
    - a vendég tartózkodási ideje és az általa elköltött összeg nagysága
38. A szóráshányados értéke a vonat típusa és a késés száma között 0,64. Mire következtetünk ebből?
    - **a vonat típusa 40,96%-ban befolyásolja a késések számát**
    - a vonat típusa nem befolyásolja a késések számát
    - a vonat típusa 64%-ban befolyásolja a késések számát
39. A lineáris determinációs együttható azt mutatja meg, hogy a magyarázó változó hány %-ban határozza meg az eredményváltozó alakulását.

    IGAZ HAMIS
40. Az y=325-12x regresszió függvény alapján megállapítható, hogy ha a magyarázó változó értékét egy egységgel növeljük, akkor az eredményváltozó 325 egységgel fog emelkedni.

    IGAZ **HAMIS**
41. Egy vállalat termelésének (ezer db) alakulását negyedéves bontásban (ezer Ft) 2009 és 2013 között leíró lineáris trend egyenlete y=1200-50t. A trend alapján az alábbi következtetés vonható le:
    - a vállalat termelése évente átlagosan 50 ezer db-bal csökkent
    - **a vállalat termelése negyedévente átlagosan 50 ezer db-bal csökkent**
    - 2013-ban a vállalat termelése 1200 ezer db volt
42. Lineáris trend és additív szezonalitás esetén a nyers szezonális eltérések összege 1.

    IGAZ **HAMIS**
43. A sugárdiagram segítségével több dimenzió mentén is végezhető összehasonlítás adatsorok között.

    IGAZ HAMIS
44. A kartogram a térképen eltérő színekkel jelzi a vizsgált jelenség területi intenzitását.

    IGAZ HAMIS

B FELADATSOR

45. A középiskolások továbbtanulási terveiről készít felmérést. Milyen mintavételi technikát alkalmazna az adatgyűjtéshez? *Több válasz is megjelölhető!*

	Véletlenen alapuló		**Nem véletlenen alapuló**
	Független, azonos eloszlású		Kvótás
	Egyszerű véletlen		Hólabda
	Rétegzett	X	Koncentrált
X	Csoportos		Szisztematikus
X	Többlépcsős		Önkényes

46. A kombinációs táblában különböző típusú statisztikai sorok vannak.

IGAZ **HAMIS**

47. 2010-hez képest 2013-ban 20%-kal nőtt az importált áruk mennyisége. Milyen típusú ez a viszonyszám?

	Csoportosító sorból képzett		**Összehasonlító sorból képzett**
	Megoszlási	X	Dinamikus
	Koordinációs		Területi összehasonlító
	Leíró sorból képzett		Tervfeladat
	Intenzitási		Tervteljesítési

48. Az év eleji leárazásoknál egyes termékekre 50%-os árkedvezmény is igénybe vehető. Milyen típusú ez a viszonyszám?

	Csoportosító sorból képzett		**Összehasonlító sorból képzett**
X	Megoszlási		Dinamikus
	Koordinációs		Területi összehasonlító
	Leíró sorból képzett		Tervfeladat
	Intenzitási		Tervteljesítési

49. Az aktivitási ráta a gazdaságilag aktívak arányát mutatja a munkaképes lakosságon belül.

IGAZ HAMIS

50. A termékenységi arányszám nyers intenzitási viszonyszám, amely a szülőképes korú nőkre vetített élveszületéseket mutatja.

IGAZ **HAMIS**

51. Egy adott napon megfigyeltek 100 embert, akik egy hipermarketben vásároltak, s azt tapasztalták, hogy a vásárlók egynegyede 15 €-nál kevesebbet, háromnegyede ennél többet költött el az üzletben. Melyik alapeloszlás jellemző értéke a 15 €?

Középértékek

	Átlag
	Módusz
	Medián
X	Alsó kvartilis
	Felső kvartilis

Szóródás és alakmutatók

	Terjedelem
	Szórás
	Relatív szórás
	A mutató
	F mutató

52. Az F mutató a csúcsosság számszerűsítésére alkalmazható.

IGAZ **HAMIS**

53. Két kiemelt üdülőkörzet adatait megnézve azt láthatjuk, hogy a tavalyi évhez képest 10%-kal esett vissza a vendégek átlagos tartózkodási ideje. Az osztrák, német és lengyel turisták átlagos tartózkodási ideje 2%-kal emelkedett, de a különböző nemzetiségű vendégek megoszlása az egyes években olyannyira megváltozott, hogy ez összességében az átlagos tartózkodási idők mintegy 11,76%-os visszaesését okozta.

- a főátlag index értéke 110,0
- a részátlag index értéke 98,0
- **az összetételhatás-index értéke 88,24**

54. Ha két időszak között a részátlagok értéke nem változik, akkor a főátlag index értéke megegyezik a részátlag index értékével.

IGAZ **HAMIS**

55. Egy család élelmiszerre fordított kiadásairól ismerjük, hogy a gyümölcsök árindexe 110, a zöldségeké 90. Ez alapján tudjuk, hogy

- **a gyümölcsök ára 10%-kal nőtt**
- a zöldségek ára 10%-kal nőtt
- a gyümölcsök és zöldségek árváltozása kiegyenlíti egymást, az együttes árindex 100

56. Az indexszámítás során a korábbi időszak adatait viszonyítjuk a későbbi időszak adataihoz.

IGAZ **HAMIS**

57. A GDP volumenindexe

- a nominál GDP változása
- **a reál GDP változása**
- a nominál és reál GDP hányadosa

58. A külkereskedelmi árstatisztikákban használatos index

Fisher-féle Pearson-féle Yule-féle

59. Ha egy csoportban csak annak sikerült teljesíteni a statisztika tárgyat, aki járt előadásra, akkor
- a tárgy teljesítése és az előadásra járás egymástól független
- a tárgy teljesítése és az előadásra járás közötti kapcsolat szóráshányados mutatóval számszerűsíthető
- **a tárgy teljesítése és az előadásra járás között függvényszerű kapcsolat van**

60. Egy vállalatnál a dolgozók iskolai végzettsége és keresete közötti kapcsolat vizsgálata során azt az eredményt kaptuk, hogy a H^2 mutató értéke 0,325. Mit jelent ez?
- a két ismérv között gyenge kapcsolat mutatható ki
- **az iskolai végzettség 32,5%-ban határozza meg a kereset alakulását**
- az iskolai végzettség 0,325%-ban befolyásolja a kereset alakulását

61. Az r=0,894 azt jelenti, hogy a mennyiségi ismérvek között pozitív irányú, erős kapcsolat mutatható ki.

IGAZ HAMIS

62. Az y=325-12x regresszió függvény alapján megállapítható, hogy ha a magyarázó változó értékét egy egységgel növeljük, akkor az eredményváltozó 12 egységgel fog csökkeni.

IGAZ HAMIS

63. Egy vállalat árbevételére vonatkozóan az alapirányzatok reziduális szórásáról az alábbiakat ismerjük: a tényleges és becsült értékek páronkénti átlagos eltérése lineáris trendnél 625 db, mozgóátlagnál 612 db, exponenciális trendnél 627 db. Melyik alapirányzat illeszkedik legjobban?
- lineáris trend
- exponenciális trend
- **mozgó átlag**

64. Exponenciális trend és multiplikatív szezonalitás esetén a nyers és korrigált szezonális eltérések megegyeznek.

IGAZ **HAMIS**

65. A grafikus ábráknál a cím és forrás megjelölése, valamint a jelmagyarázat mindig kötelező.

IGAZ **HAMIS**

66. A koordináta-rendszeren alapuló ábráknál a vizsgált jellemzők egymáshoz viszonyított helyzete meghatározó.

IGAZ HAMIS