Министерство сельского хозяйства Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«Саратовский государственный аграрный университет

имени Н. И. Вавилова»

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Краткий курс лекций

для студентов I курса

Направление подготовки

36.05.01 Ветеринария

Саратов 2016

ББК 22.1

Д93

Рецензенты:

Доцент кафедры «Электроника колебаний и волн», кандидат физико-математических наук
ФГОУ ВО «СГУ им. Чернышевского».
А.В. Толстиков
Заведующий кафедрой «Информационные технологии и прикладная математика»,
кандидат физико-математических наук, доцент ФГОУ ВО «Саратовский ГАУ»
Г.Н.Камышова

Д93
Математическая статистика: краткий курс лекций для студентов 1 курса направления подготовки 36.05.01 «Ветеринария» / сост. Н.В. Дьяконова //ФГБОУ ВО «Саратовский ГАУ».- Саратов, 2016.-

ISBN

Краткий курс лекций по дисциплине «Математическая статистика» составлен в соответствии с программой дисциплины и предназначен для студентов направления подготовки 36.05.01 «Ветеринария». Краткий курс лекций позволяет студентам освоить основные методы математической статистики, необходимые для анализа процессов и явлений в ходе поиска оптимальных решений практических задач. Курс направлен на формирование ключевых компетенций, необходимых для эффективного решения производственных задач в профессиональной деятельности.

УДК 51

ББК 22.1

ISBN

Введение

Математическая статистика – раздел математики, изучающий методы сбора, систематизации и обработки результатов наблюдений с целью выявления статистических закономерностей. Она имеет исключительно важное значение как для всего процесса обучения студента, так и для последующей профессиональной деятельности специалиста. Методы математической статистики используются при планировании организации производства, анализе технологических процессов, для контроля качества продукции и многих других целей.

Обоснованием математической статистики является теория вероятностей, изучающая закономерности случайных явлений. Знание закономерностей, которым подчиняются массовые случайные события, позволяет предвидеть, как эти события будут протекать.

Курс лекций состоит из двух частей. Первая часть включает основы теории вероятностей: случайные события и их вероятности; случайные величины, их распределение и числовые характеристики; важнейшие предельные теоремы теории вероятностей. Во второй части изложены основы выборочного метода, теории оценивания вероятностных параметров, проверки статистических гипотез, регрессионного и корреляционного анализа.

Лекция 1

ПРЕДМЕТ ТЕОРИИ ВЕРОЯТНОСТЕЙ. СЛУЧАЙНЫЕ СОБЫТИЯ

1.1 Основные понятия

В различных разделах науки и техники нередко возникают ситуации, когда результат каждого из многих проводимых опытов заранее предугадать невозможно, однако можно исследовать закономерности, возникающие при проведении серии опытов. Нельзя, например, точно сказать, какая сторона монеты окажется сверху при данном броске: герб или цифра – но при большом количестве бросков число выпадений герба приближается к половине количества бросков; нельзя заранее предсказать результат одного выстрела из данного орудия по данной цели, но при большом числе выстрелов частота попадания приближается к некоторому постоянному числу. Исследование вероятностных закономерностей массовых однородных явлений составляет предмет **теории вероятностей**.

Основным интуитивным понятием классической теории вероятностей является **случайное событие**.

Определение. Событием называется всякий факт, который может произойти или не произойти в результате опыта.

При этом тот или иной результат опыта может быть получен с различной степенью возможности. Т.е. в некоторых случаях можно сказать, что одно событие произойдет практически наверняка, другое практически никогда.

В отношении друг друга события также имеют особенности, т.е. в одном случае событие А может произойти совместно с событием В, в другом – нет.

Определение. События называются несовместными, если появление одного из них исключает появление других.

Классическим примером несовместных событий является результат подбрасывания монеты – выпадение лицевой стороны монеты исключает выпадение обратной стороны (в одном и том же опыте).

Определение. Полной группой событий называется совокупность всех возможных результатов опыта.

Определение. Достоверным событием называется событие, которое наверняка произойдет в результате опыта. Событие называется невозможным, если оно никогда не произойдет в результате опыта.

Например, если из коробки, содержащей только красные и зеленые шары, наугад вынимают один шар, то появление среди вынутых шаров белого – невозможное событие. Появление красного и появление зеленого шаров образуют полную группу событий.

Определение. События называются равновозможными, если нет оснований считать, что одно из них появится в результате опыта с большей вероятностью.

В приведенном выше примере появление красного и зеленого шаров – равновозможные события, если в коробке находится одинаковое количество красных и зеленых шаров.

Если же в коробке красных шаров больше, чем зеленых, то появление зеленого шара – событие менее вероятное, чем появление красного.

1.2 Классическое определение вероятности

Исходя из общих понятий можно дать определение вероятности.

Определение. Вероятностью события А называется математическая оценка возможности появления этого события в результате опыта. Вероятность события А равна отношению числа, благоприятствующих событию А исходов опыта к общему числу попарно несовместных исходов опыта, образующих полную группу событий.

$$P(A) = \frac{m}{n}$$

Исход опыта является благоприятствующим событию А, если появление в результате опыта этого исхода влечет за собой появление события А.

Очевидно, что вероятность достоверного события равна единице, а вероятность невозможного – равна нулю. Таким образом, значение вероятности любого события – есть положительное число, заключенное между нулем и единицей.

$$0 \le P(A) \le 1$$

Определение. Относительной частотой события А называется отношение числа опытов, в результате которых произошло событие А к общему числу опытов.

Отличие относительной частоты от вероятности заключается в том, что вероятность вычисляется без непосредственного произведения опытов, а относительная частота – после опыта.

Так в рассмотренном выше примере, если из коробки наугад извлечено 5 шаров и 2 из них оказались красными, то относительная частота появления красного шара равна:

$$W(A) = \frac{2}{5}$$

Как видно, эта величина не совпадает с найденной вероятностью.

При достаточно большом числе произведенных опытов относительная частота изменяется мало, колеблясь около одного числа. Это число может быть принято за вероятность события.

Вообще говоря, классическое определение вероятности – довольно относительное.

Это обусловлено тем, что на практике сложно представить результат опыта в виде совокупности элементарных событий, доказать, что события равновероятные.

К примеру, при произведении опыта с подбрасыванием монеты на результат опыта могут влиять такие факторы как несимметричность монеты, влияние ее формы на аэродинамические характеристики полета, атмосферные условия и т.д.

Классическое определение вероятности неприменимо к испытаниям с бесконечным числом исходов. Чтобы преодолеть этот недостаток вводится понятие геометрической вероятности, т.е. вероятности попадания точки в какой – либо отрезок или часть плоскости (пространства).

Так если на отрезке длиной L выделен отрезок длины l, то вероятность попадания наугад взятой точки в отрезок l равна отношению l/L.

1.3 Основные формулы комбинаторики

При вычислении вероятностей часто приходится использовать некоторые формулы **комбинаторики** – науки, изучающей комбинации, которые можно составить по определенным правилам из элементов некоторого конечного множества. Определим основные такие комбинации.

Определение Перестановки – это комбинации, составленные из всех **n** элементов данного множества и отличающиеся только порядком их расположения. Число всех возможных перестановок

$$P_n = n!$$

Определение Размещения – это комбинации из *m* элементов множества, содержащего **n** различных элементов, отличающиеся либо составом элементов, либо их порядком. Число всех возможных размещений

$$A_n^m = n(n-1)(n-2)\ldots(n-m+1)$$

Определение Сочетания – неупорядоченные наборы из *m* элементов множества, содержащего **n** различных элементов (то есть наборы, отличающиеся только составом элементов). Число сочетаний

$$C_n^m = \frac{n!}{m!(n-m)!}.$$

Вопросы для самоконтроля

1. Элементы комбинаторики: перестановки, размещения, сочетания. Примеры.
2. Случайные события и их классификация.
3. Относительная частота случайного события и статистическое определение вероятности.
4. Классическое определение вероятности. Непосредственный подсчёт вероятностей.

СПИСОК ЛИТЕРАТУРЫ

Основная

1. *Вентцель, Елена Сергеевна.* Теория вероятностей и ее инженерные приложения : учебное пособие / Е. С. Вентцель, Л. А. Овчаров. - 4-е изд., стер. - М. : Высш. шк., 2007. - 491 с. - (Учебники для вузов. Математика). - ISBN 978-5-06-005714-0

2. *Ермакова, В.И.* Теория вероятностей и математическая статистика : учебное пособие / ред. В. И. Ермакова. - М. : Инфра-М, 2004. - 287 с. - (Высшее образование). - ISBN 5-16 001561-2

3. *Гмурман, В. Е.* Теория вероятностей и математическая статистика : учебное пособие / В. Е. Гмурман. - 12-е изд., перераб. - М. : Высш. образование, 2006. - 479 с. : ил. - (Основы наук).

4. *Рау, Валерий Георгиевич.* Практический курс математики и общей теории статистики : учебное пособие / В. Г. Рау. - М. : Высш. шк., 2006. - 126 с. : ил. -ISBN 5-06-005529-9

5. *Кириллова, Т. В., Хучраева Т.С.* Элементы математической статистики / Т. В., Хучраева Т.С. Кириллова. - Саратов : Сарат. гос. агр. ун-т, 2004. - 60 с. - ISBN 5-7011-0394-3

Дополнительная

1. *Хучраева, Тамара Сергеевна.* Высшая математика : учебное пособие / Т. С. Хучраева, А. А. Смоленинов, Т. В. Кириллова. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2008. - 156 с. - ISBN 978-5-7011-0559-9

2. *Шириков, Виктор Филиппович*. Теория вероятностей : учебное пособие / В. Ф. Шириков, С. М. Зарбалиев. - М. : КолосС, 2008. - 389 с. - (Учебники и учеб. пособия для студентов высш. учеб. заведений). - ISBN 978-5-9532-0621-1

3. *Данко, Павел Ефимович*. Высшая математика в упражнениях и задачах : в 2 ч. Ч.2 / П. Е. Данко, А. Г. Попов, Т. Я. Кожевникова. - М. : Оникс 21 век ; М. : Экзамен ; М. : Мир и образование, 2003. - 416 с. : ил. - ISBN 5-329-00327-х. - ISBN 5-94666-009-8

4. *Юрьева, Александра Антоновна*. Руководство к решению задач по теории вероятностей и математической статистике. : учебное пособие / А. А. Юрьева. - 2-е изд., испр. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2009. - 208 с.

5. *Прохоров, Ю.В.* Вероятность и математическая статистика : энциклопедический словарь / ред. Ю. В. Прохоров. - Репр. воспроизведение изд. - М. : Большая Российская Энциклопедия, 2003. - 910 с. - (Золотой фонд). - ISBN 5-7107-7433-2

Лекция 2

ОСНОВНЫЕ ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ. АЛГЕБРА СОБЫТИЙ

2.1 Операции над событиями

Определение. События А и В называются равными, если осуществление события А влечет за собой осуществление события В и наоборот.

Определение. Объединением или суммой событий A_k называется событие А, которое означает появление хотя бы одного из событий A_k.

$$A = \bigcup_k A_k$$

Определение. Пересечением или произведением событий A_k называется событие А, которое заключается в осуществлении всех событий A_k.

$$A = \bigcap_k A_k$$

2.2 Основные теоремы о вероятности

Теорема (сложения вероятностей). Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий.

$$P(A+B) = P(A) + P(B)$$

Следствие 1: Если события $A_1, A_2, ..., A_n$ образуют полную группу несовместных событий, то сумма их вероятностей равна единице.

$$\sum_{i=1}^{n} P(A_i) = 1$$

Определение. Противоположными называются два несовместных события, образующие полную группу.

Следствие 2: Сумма вероятностей противоположных событий равна единице.

$$P(A) + P(\overline{A}) = 1$$

Теорема. Вероятность появления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий без вероятности их совместного появления.

$$P(A+B) = P(A) + P(B) - P(AB)$$

Определение. Событие А называется независимым от события В, вероятность события А не зависит от того, произошло событие В или нет. Событие А называется зависимым от события В, если вероятность события А меняется в зависимости от того, произошло событие В или нет.

Определение. Вероятность события В, вычисленная при условии, что имело место событие А, называется условной вероятностью события В.

$$P_A(B) = P(B/A) = P(AB)/P(A)$$

Теорема. (Умножения вероятностей) Вероятность произведения двух событий (совместного появления этих событий) равна произведению вероятности *одного из них* на условную вероятность другого, вычисленную при условии, что первое событие уже наступило.

$$P(AB) = P(A)P(B/A) = P(A)P_A(B)$$

Если события независимые, то $P(B/A)=P(B)$, и теорема умножения вероятностей принимает вид:

$$P(AB)=P(A)P(B)$$

В случае произведения нескольких зависимых событий, вероятность равна произведению одного из них на условные вероятности всех остальных при условии, что вероятность каждого последующего вычисляется в предположении, что все остальные события уже совершились.

$$P(A_1A_2...A_n)=P(A_1)P(A_2/A_1)P(A_3/A_1A_2)...P(A_n/A_1A_2...A_{n-1})$$

Из теоремы произведения вероятностей можно сделать вывод о вероятности появления хотя бы одного события.

Если в результате испытания может появиться n событий, независимых в совокупности, то вероятность появления хотя бы одного из них равна

$$P(A)=1-q_1q_2...q_n$$

Здесь событие А обозначает наступление хотя бы одного из событий A_i, а q_i – вероятность противоположных событий $\overline{A}_1,\overline{A}_2,...,\overline{A}_n$.

2.3 Формула полной вероятности

Пусть некоторое событие А может произойти вместе с одним из несовместных событий $H_1,H_2,...,H_n$, составляющих полную группу событий. Пусть известны вероятности этих событий $P(H_1),P(H_2),...,P(H_n)$ и условные вероятности наступления события А при наступлении события H_i $P(A/H_1),P(A/H_2),...,P(A/H_n)$.

Теорема. Вероятность события А, которое может произойти вместе с одним из событий $H_1,H_2,...,H_n$, равна сумме парных произведений вероятностей каждого из этих событий на соответствующие им условные вероятности наступления события А.

$$P(A)=\sum_{i=1}^{n}P(H_i)P(A/H_i)$$

2.4.Формулы Бейеса (формула гипотез)

Пусть имеется полная группа несовместных гипотез $H_1,H_2,...,H_n$ с известными вероятностями их наступления $P(H_1),P(H_2),...,P(H_n)$. Пусть в результате опыта наступило событие А, условные вероятности которого по каждой из гипотез известны, т.е. известны вероятности $P(A/H_1),P(A/H_2),...,P(A/H_n)$.

Теорема. Вероятность гипотезы после испытания равна произведению вероятности гипотезы до испытания на соответствующую ей условную вероятность события, которое произошло при испытании, деленному на полную вероятность этого события.

$$P(H_i/A)=\frac{P(H_i)P(A/H_i)}{\sum_{i=1}^{n}P(H_i)P(A/H_i)}\quad \text{(i=1,2….n)}$$

Эти формулы называются формулами Бейеса.

Вопросы для самоконтроля

1 Операции над событиями
2 Основные теоремы о вероятности
3 Формула полной вероятности
4.Формулы Бейеса (формула гипотез)

СПИСОК ЛИТЕРАТУРЫ

Основная

1. *Вентцель, Елена Сергеевна.* Теория вероятностей и ее инженерные приложения : учебное пособие / Е. С. Вентцель, Л. А. Овчаров. - 4-е изд., стер. - М. : Высш. шк., 2007. - 491 с. - (Учебники для вузов. Математика). - ISBN 978-5-06-005714-0

2. *Рау, Валерий Георгиевич.* Практический курс математики и общей теории статистики : учебное пособие / В. Г. Рау. - М. : Высш. шк., 2006. - 126 с. : ил. -ISBN 5-06-005529-9

Дополнительная

1. *Хучраева, Тамара Сергеевна.* Высшая математика : учебное пособие / Т. С. Хучраева, А. А. Смоленинов, Т. В. Кириллова. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2008. - 156 с. - ISBN 978-5-7011-0559-9

2. *Шириков, Виктор Филиппович.* Теория вероятностей : учебное пособие / В. Ф. Шириков, С. М. Зарбалиев. - М. : КолосС, 2008. - 389 с. - (Учебники и учеб. пособия для студентов высш. учеб. заведений). - ISBN 978-5-9532-0621-1

3. *Юрьева, Александра Антоновна.* Руководство к решению задач по теории вероятностей и математической статистике. : учебное пособие / А. А. Юрьева. - 2-е изд., испр. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2009. - 208 с.

Лекция 3

НЕЗАВИСИМЫЕ ИСПЫТАНИЯ

3.1 Формула Бернулли

Если производится некоторое количество испытаний, в результате которых может произойти или не произойти событие А, и вероятность появления этого события в каждом из испытаний не зависит от результатов остальных испытаний, то такие испытания называются **независимыми относительно события А**.

Допустим, что событие А наступает в каждом испытании с вероятностью $P(A)=p$. Определим вероятность $P_{m,n}$ того, что в результате n испытаний событие А наступило ровно m раз.

Эту вероятность в принципе можно посчитать, используя теоремы сложения и умножения вероятностей, как это делалось в рассмотренных выше примерах. Однако, при достаточно большом количестве испытаний это приводит к очень большим вычислениям. Таким образом, возникает необходимость разработать общий подход к решению поставленной задачи. Этот подход реализован в формуле Бернулли. (Якоб Бернулли (1654 – 1705) – швейцарский математик)

Пусть в результате n независимых испытаний, проведенных в одинаковых условиях, событие А наступает с вероятностью $P(A) = p$, а противоположное ему событие $\overline{A}$ с вероятностью $P(\overline{A}) = 1 - p$.

Обозначим A_i – наступление события А в испытании с номером i. Т.к. условия проведения опытов одинаковые, то эти вероятности равны.

Если в результате n опытов событие А наступает ровно m раз, то остальные n-m раз это событие не наступает. Событие А может появиться m раз в n испытаниях в различных комбинациях, число которых равно количеству сочетаний из n элементов по m. Это количество сочетаний находится по формуле:

$$C_n^m = \frac{n!}{m!(n-m)!}$$

Вероятность каждой комбинации равна произведению вероятностей:

$$p^m (1-p)^{n-m}$$

Применяя теорему сложения вероятностей несовместных событий, получаем формулу Бернулли:

$$P_{m,n} = \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m}$$

Формула Бернулли важна тем, что справедлива для любого количества независимых испытаний, т.е. того самого случая, в котором наиболее четко проявляются законы теории вероятностей.

3.2. Локальная формула Лапласа

В тех случаях, когда число испытаний n велико, а вероятность p не близка к нулю, для вычисления биномиальных вероятностей используют теоремы Лапласа.

Теорема. Если вероятность p наступления события А в каждом испытании постоянна и отлична от нуля и единицы, а число независимых испытаний достаточно велико, то вероятность $Pn(m)$ может быть вычислена по приближённой формуле.

$$Pn(m) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x), \text{ где } \quad x = \frac{m - np}{\sqrt{npq}}$$

Равенство тем точнее, чем больше *n*

Функция $\varphi(x)$ называется функцией Гаусса.

Свойства функции*:*

1. Функция чётная, т.е. $\varphi(-x) = \varphi(x)$
2. При $x \geq 4$ можно считать, что $\varphi(x) = 0$.

3.3. Интегральная формула Лапласа

Теорема. Если вероятность *p* наступления события А в каждом испытании постоянна и отлична от нуля и единицы, то вероятность $Pn(k1 \leq m \leq k2)$ может быть найдена по приближённой формуле.

$$Pn(k1 \leq m \leq k2) \approx \Phi(x2)\text{-}\Phi(x1),$$

где $x1 = \frac{k_1 - np}{\sqrt{npq}}$, $x2 = \frac{k_2 - np}{\sqrt{npq}}$

Функция $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ называется нормированной функцией Лапласа.

Свойства функции Ф(x):

1. Функция нечётная, т.е $\Phi(-x)=-\Phi(x)$
2. При $x>5$ можно считать, что $\Phi(x)=0.5$

3.4. Формула Пуассона

Теорема. Если число испытаний неограниченно увеличивается $(n \to \infty)$ и вероятность наступления события *A* в каждом испытании неограниченно уменьшается $(p \to 0)$, но так, что их произведение $np=a$ является постоянной величиной, то вероятность $Pn(m)$ приближённо равна $Pn(m)$

$$P_n(m) \approx \frac{a^m e^{-a}}{m!}$$

Вопросы для самоконтроля

1. Независимые испытания. Схема Бернулли.
2. Предельные теоремы в схеме Бернулли.
3. Локальная и интегральная теоремы Муавра-Лапласа.
4. Теорема Пуассона.

СПИСОК ЛИТЕРАТУРЫ

Основная

1. *Вентцель, Елена Сергеевна.* Теория вероятностей и ее инженерные приложения : учебное пособие / Е. С. Вентцель, Л. А. Овчаров. - 4-е изд., стер. - М. : Высш. шк., 2007. - 491 с. - (Учебники для вузов. Математика). - ISBN 978-5-06-005714-0

2. *Гмурман, В. Е.* Теория вероятностей и математическая статистика : учебное пособие / В. Е. Гмурман. - 12-е изд., перераб. - М. : Высш. образование, 2006. - 479 с. : ил. - (Основы наук).

3. *Рау, Валерий Георгиевич.* Практический курс математики и общей теории статистики : учебное пособие / В. Г. Рау. - М. : Высш. шк., 2006. - 126 с. : ил. -ISBN 5-06-005529-9

Дополнительная

1. *Хучраева, Тамара Сергеевна.* Высшая математика : учебное пособие / Т. С. Хучраева, А. А. Смоленинов, Т. В. Кириллова. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2008. - 156 с. - ISBN 978-5-7011-0559-9

2. *Шириков, Виктор Филиппович.* Теория вероятностей : учебное пособие / В. Ф. Шириков, С. М. Зарбалиев. - М. : КолосС, 2008. - 389 с. - (Учебники и учеб. пособия для студентов высш. учеб. заведений). - ISBN 978-5-9532-0621-1

3. *Юрьева, Александра Антоновна.* Руководство к решению задач по теории вероятностей и математической статистике. : учебное пособие / А. А. Юрьева. - 2-е изд., испр. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2009. - 208 с.

Лекция 4

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

4.1 Основные понятия

Выше рассматривались случайные события, являющиеся качественной характеристикой случайного результата опыта. Для получения количественной характеристики вводится понятие случайной величины.

Определение. Случайной величиной называется величина, которая в результате опыта может принимать то или иное значение, причем заранее известно какое именно.

Случайные величины можно разделить на две категории.

Определение. Дискретной случайной величиной называется такая величина, которая в результате опыта может принимать определенные значения с определенной вероятностью, образующие счетное множество (множество, элементы которого могут быть занумерованы).

Это множество может быть как конечным, так и бесконечным.

Например, количество выстрелов до первого попадания в цель является дискретной случайной величиной, т.к. эта величина может принимать и бесконечное, хотя и счетное количество значений.

Определение. Непрерывной случайной величиной называется такая величина, которая может принимать любые значения из некоторого конечного или бесконечного промежутка.

Для задания случайной величины недостаточно просто указать ее значение, необходимо также указать вероятность этого значения.

4.2 Закон распределения дискретной случайной величины

Определение. Соотношение между возможными значениями случайной величины и их вероятностями называется законом распределения дискретной случайной величины.

Таблица соответствия значений случайной величины и их вероятностей называется рядом распределения.

Графическое представление этой таблицы называется многоугольником распределения. При этом сумма все ординат многоугольника распределения представляет собой вероятность всех возможных значений случайной величины, а, следовательно, равна единице.

Определение. Функцией распределения называют функцию F(x), определяющую вероятность того, что случайная величина X в результате испытания примет значение, меньшее *x*.

$$F(x) = P(X < x)$$

Функцию распределения также называют интегральной функцией.

Функция распределения существует как для непрерывных, так и для дискретных случайных величин. Она полностью характеризует случайную величину и является одной из форм закона распределения.

Для дискретной случайной величины функция распределения имеет вид:

$$F(x)=\sum_{x_i<x}P(X=x_i)$$

Знак неравенства под знаком суммы показывает, что суммирование распространяется на те возможные значения случайной величины, которые меньше аргумента *x*.

Функция распределения дискретной случайной величины X разрывна и возрастает скачками при переходе через каждое значение x_i.

Свойства функции распределения:

1) значения функции распределения принадлежат отрезку [0, 1].

$$0\le F(x)\le 1$$

2) F(x) – неубывающая функция.

$$F(x_2)\ge F(x_1)\text{ при } x_2\ge x_1$$

3) Вероятность того, что случайная величина примет значение, заключенное в интервале (*a, b*) , равна приращению функции распределения на этом интервале.

$$P(a\le X<b)=F(b)-F(a)$$

4) На минус бесконечности функция распределения равна нулю, на плюс бесконечности функция распределения равна единице.

5) Вероятность того, что непрерывная случайная величина X примет одно определенное значение, равна нулю.

Таким образом, не имеет смысла говорить о каком – либо конкретном значении случайной величины. Интерес представляет только вероятность попадания случайной величины в какой-либо интервал, что соответствует большинству практических задач.

4.3 Закон распределения непрерывной случайной величины

Непрерывная случайная величина так же как и дискретная может задаваться

Функция распределения полностью характеризует случайную величину, однако, функцией распределения.имеет один недостаток. По функции распределения трудно судить о характере распределения случайной величины в небольшой окрестности той или иной точки числовой оси.

Определение. Плотностью распределения вероятностей непрерывной случайной величины X называется функция *f(x)* – первая производная от функции распределения F(x).

$$f(x)=F'(x).$$

Плотность распределения также называют дифференциальной функцией. Для описания дискретной случайной величины плотность распределения неприемлема.

Теорема. Вероятность того, что непрерывная случайная величина X примет значение, принадлежащее интервалу (a, b), равна определенному интегралу от плотности распределения, взятому в пределах от a до b.

$$P(a<X<b)=\int_a^b f(x)dx$$

Геометрически это означает, что вероятность того, что непрерывная случайная величина примет значение, принадлежащее интервалу (a, b), равна площади криволинейной трапеции, ограниченной осью OX, кривой распределения f(x) и прямыми *x=a* и *x=b*.

Функция распределения может быть легко найдена, если известна плотность распределения, по формуле:

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

Свойства плотности распределения:

1) Плотность распределения – неотрицательная функция.

$$f(x) \geq 0$$

2) Несобственный интеграл от плотности распределения в пределах от - ∞ до ∞ равен единице.

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

4.4 Числовые характеристики дискретных случайных величин

Закон распределения полностью характеризует случайную величину. Однако, когда невозможно найти закон распределения, или этого не требуется, можно ограничиться нахождением значений, называемых числовыми характеристиками случайной величины. Эти величины определяют некоторое среднее значение, вокруг которого группируются значения случайной величины, и степень их разбросанности вокруг этого среднего значения.

Определение. Математическим ожиданием дискретной случайной величины называется сумма произведений всех возможных значений случайной величины на их вероятности.

$$m_x = M(X) = x_1 p_1 + x_2 p_2 + ... + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

С точки зрения вероятности можно сказать, что математическое ожидание приближенно равно среднему арифметическому наблюдаемых значений случайной величины.

Свойства математического ожидани:

1) Математическое ожидание постоянной величины равно самой постоянной.

$M(C) = C$

2) Постоянный множитель можно выносить за знак математического ожидания.

$M(Cx) = CM(x)$

3) Математическое ожидание произведения двух независимых случайных величин равно произведению их математических ожиданий.

$M(XY) = M(X)M(Y)$

Это свойство справедливо для произвольного числа случайных величин.

4) Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий слагаемых.

$M(X+Y) + M(X) + M(Y)$

Это свойство также справедливо для произвольного числа случайных величин.

Однако математическое ожидание не может полностью характеризовать случайный процесс. Кроме математического ожидания надо ввести величину, которая характеризует отклонение значений случайной величины от математического ожидания.

Это отклонение равно разности между случайной величиной и ее математическим ожиданием. При этом математическое ожидание отклонения равно нулю. Это объясняется тем, что одни возможные отклонения положительны, другие отрицательны, и в результате их взаимного погашения получается ноль.

Определение. Дисперсией (рассеиванием) дискретной случайной величины называется математическое ожидание квадрата отклонения случайной величины от ее математического ожидания.

$$D(X) = M[X - M(X)]^2$$

Теорема. Дисперсия равна разности между математическим ожиданием квадрата случайной величины X и квадратом ее математического ожидания.

$$D(X) = M(X^2) - [M(X)]^2$$

Свойства дисперсии.

1) Дисперсия постоянной величины равна нулю.

$D(C) = 0$

2) Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат.

$D(CX) = C^2 D(X)$

3) Дисперсия суммы двух независимых случайных величин равна сумме дисперсий этих величин.

$D(X + Y) = D(X) + D(Y)$

4) Дисперсия разности двух независимых случайных величин равна сумме дисперсий этих величин.

$D(X - Y) = D(X) + D(Y)$

Справедливость этого равенства вытекает из свойства 2.

Определение. Средним квадратическим отклонением случайной величины X называется квадратный корень из дисперсии.

$$\sigma(X) = \sqrt{D(X)}$$

Определение. Модой *Mo* называется наиболее вероятное значение дискретной случайной величины.

4.5 Числовые характеристики непрерывных случайных величин

Пусть непрерывная случайная величина X задана функцией распределения *f(x)*. Допустим, что все возможные значения случайной величины принадлежат отрезку [*a*,*b*].

Определение. Математическим ожиданием непрерывной случайной величины X, возможные значения которой принадлежат отрезку [a,b], называется определенный интеграл

$$M(X) = \int_a^b x f(x) dx$$

Если возможные значения случайной величины рассматриваются на всей числовой оси, то математическое ожидание находится по формуле:

$$M(X) = \int_{-\infty}^{\infty} x f(x) dx$$

При этом, конечно, предполагается, что несобственный интеграл сходится.

Определение. Дисперсией непрерывной случайной величины называется математическое ожидание квадрата ее отклонения.

$$D(X)=\int_{-\infty}^{\infty}[x-M(X)]^2 f(x)dx$$

По аналогии с дисперсией дискретной случайной величины, для практического вычисления дисперсии используется формула:

$$D(X)=\int_{-\infty}^{\infty}x^2 f(x)dx-[M(X)]^2$$

Определение. Средним квадратичным отклонением называется квадратный корень из дисперсии.

$$\sigma(X)=\sqrt{D(X)}$$

Определение. Модой M_0 дискретной случайной величины называется ее наиболее вероятное значение. Для непрерывной случайной величины мода – такое значение случайной величины, при которой плотность распределения имеет максимум.
$f(M_0)=\max$.

Вопросы для самоконтроля

1. Понятие случайной величины, закон её распределения.
2. Дискретная случайная величина. Ряд распределения, многоугольник разделения.
3. Функция распределения F(x) и её свойства.
4. Непрерывная случайная величина. Плотность распределения вероятности f(x) и её свойства. Кривая распределения.
5. Числовые характеристики дискретной и непрерывной случайных величин (математическое ожидание, дисперсия, среднее квадратическое отклонение).
6. Биномиальное распределение дискретной случайной величины, его числовые характеристики.
7. Нормальное распределение непрерывной случайной величины, определение, вид кривой Гаусса, свойства кривой.
8. Вероятность попадания нормально распределённой случайной величины в заданный интервал. Правило 3-х сигм.

СПИСОК ЛИТЕРАТУРЫ

Основная

1. *Вентцель, Елена Сергеевна.* Теория вероятностей и ее инженерные приложения : учебное пособие / Е. С. Вентцель, Л. А. Овчаров. - 4-е изд., стер. - М. : Высш. шк., 2007. - 491 с. - (Учебники для вузов. Математика). - ISBN 978-5-06-005714-0

2. *Гмурман, В. Е.* Теория вероятностей и математическая статистика : учебное пособие / В. Е. Гмурман. - 12-е изд., перераб. - М. : Высш. образование, 2006. - 479 с. : ил. - (Основы наук).

3. *Рау, Валерий Георгиевич.* Практический курс математики и общей теории статистики : учебное пособие / В. Г. Рау. - М. : Высш. шк., 2006. - 126 с. : ил. -ISBN 5-06-005529-9

Дополнительная

1. *Хучраева, Тамара Сергеевна.* Высшая математика : учебное пособие / Т. С. Хучраева, А. А. Смоленинов, Т. В. Кириллова. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2008. - 156 с. - ISBN 978-5-7011-0559-9

2. *Шириков, Виктор Филиппович.* Теория вероятностей : учебное пособие / В. Ф. Шириков, С. М. Зарбалиев. - М. : КолосС, 2008. - 389 с. - (Учебники и учеб. пособия для студентов высш. учеб. заведений). - ISBN 978-5-9532-0621-1

3. *Юрьева, Александра Антоновна.* Руководство к решению задач по теории вероятностей и математической статистике. : учебное пособие / А. А. Юрьева. - 2-е изд., испр. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2009. - 208 с.

Лекция 5

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

5.1 Основные задачи и понятия

Математическая статистика занимается установлением закономерностей, которым подчинены массовые случайные явления, на основе обработки статистических данных, полученных в результате наблюдений. Основными задачами математической статистики являются:

- определение способов сбора и группировки этих статистических данных;

- разработка методов анализа полученных данных в зависимости от целей исследования, к которым относятся:

-оценка неизвестной вероятности события; оценка неизвестной функции распределения; оценка параметров распределения, вид которого известен; оценка зависимости от других случайных величин и т.д.;

-проверка статистических гипотез о виде неизвестного распределения или о значениях параметров известного распределения.

Для решения этих задач необходимо выбрать из большой совокупности однородных объектов ограниченное количество объектов, по результатам изучения которых можно сделать прогноз относительно исследуемого признака этих объектов.

Определим основные понятия математической статистики.

Генеральная совокупность – все множество имеющихся объектов.

Выборка – набор объектов, случайно отобранных из генеральной совокупности.

Объем генеральной совокупности N и объем выборки n – число объектов в рассматриваемой совокупности.

Виды выборки:

Повторная – каждый отобранный объект перед выбором следующего возвращается в генеральную совокупность;

Бесповторная – отобранный объект в генеральную совокупность не возвращается.

5.2 Первичная обработка результатов

Пусть интересующая нас случайная величина **X** принимает в выборке значение $\mathbf{x}_1$ $\mathbf{n}_1$ раз, $\mathbf{x}_2$– $\mathbf{n}_2$ раз, …, $\mathbf{x}_к$ – $\mathbf{n}_к$ раз, причем $\sum_{i=1}^{k} n_k = n,$ где **n** – объем выборки. Тогда наблюдаемые значения случайной величины $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_к$ называют **вариантами**, а $\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_к$ – **частотами**. Если разделить каждую частоту на объем выборки, то получим **относительные частоты** $w_i = \frac{n_i}{n}.$ Последовательность вариант, записанных в порядке возрастания, называют **вариационным** рядом, а перечень вариант и соответствующих им частот или относительных частот – **статистическим рядом**:

$\mathbf{x_i}$	$\mathbf{x}_1$	$\mathbf{x}_2$	…	$\mathbf{x_k}$
$\mathbf{n_i}$	$\mathbf{n}_1$	$\mathbf{n}_2$	…	$\mathbf{n_k}$
$\mathbf{w_i}$	$\mathbf{w}_1$	$\mathbf{w}_2$	…	$\mathbf{w_k}$

5.3 Полигон частот. Выборочная функция распределения и гистограмма.

Для наглядного представления о поведении исследуемой случайной величины в выборке можно строить различные графики. Один из них – **полигон частот**: ломаная, отрезки которой соединяют точки с координатами ($\mathbf{x}_1, \mathbf{n}_1$), ($\mathbf{x}_2, \mathbf{n}_2$),…, ($\mathbf{x_k}, \mathbf{n_k}$), где $\mathbf{x_i}$откладываются на оси абсцисс, а $\mathbf{n_i}$– на оси ординат. Если на оси ординат откладывать не абсолютные ($\mathbf{n_i}$), а относительные ($\mathbf{w_i}$) частоты, то получим **полигон относительных частот** (рис.1)**.**

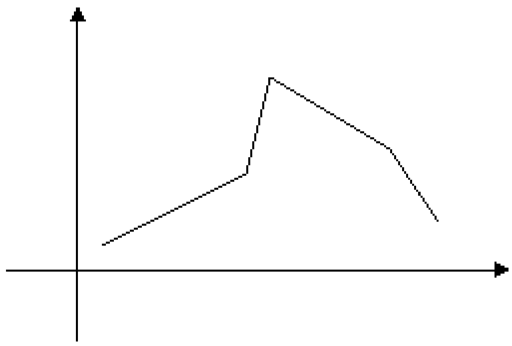


Рис. 1.

По аналогии с функцией распределения случайной величины можно задать некоторую функцию, относительную частоту события **X < x.**

Определение. Выборочной (эмпирической) функцией распределения называют функцию **F*(x)**, определяющую для каждого значения **x** относительную частоту события **X < x.**Таким образом,

$$F^*(x) = \frac{n_x}{n}$$

где $\mathbf{n_x}$ – число вариант, меньших **x**, **n** – объем выборки.

Замечание. В отличие от эмпирической функции распределения, найденной опытным путем, функцию распределения **F(x)** генеральной совокупности называют теоретической функцией распределения. F(x) определяет вероятность события X < x, а F*(x) – его относительную частоту. При достаточно больших **n**, как следует из теоремы Бернулли, **F*(x)** стремится по вероятности к **F(x)**.

Из определения эмпирической функции распределения видно, что ее свойства совпадают со свойствами **F**(**x**), а именно:

1) $0 \le$ **F***(**x**) ≤ 1.

2) **F***(**x**) – неубывающая функция.

3) Если $\mathbf{x}_1$ – наименьшая варианта, то **F***(**x**) = 0 при $\mathbf{x} \le \mathbf{x}_1$; если $\mathbf{x}_\mathbf{к}$ – наибольшая варианта, то **F***(**x**) = 1 при $\mathbf{x} > \mathbf{x}_\mathbf{к}$ **.**

Для непрерывного признака графической иллюстрацией служит **гистограмма**, то есть ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной **h**, а высотами – отрезки длиной $\mathbf{n_i}$ **/h** (гистограмма частот) или $\mathbf{w_i}$ **/h**(гистограмма относительных частот). В первом случае площадь гистограммы равна объему выборки, во втором – единице (рис.2).

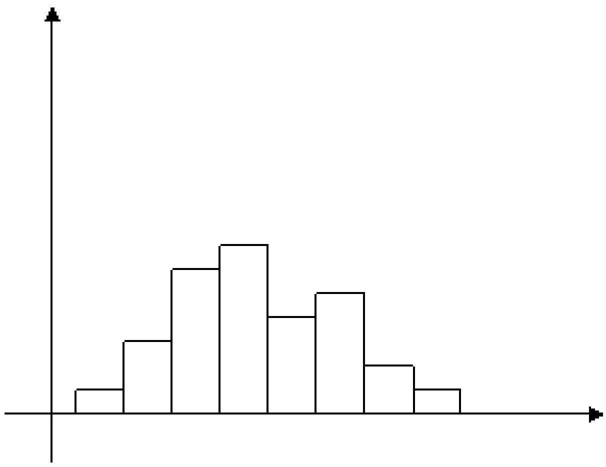


Рис.2

5.4 Числовые характеристики статистического распределения

Одна из задач математической статистики: по имеющейся выборке оценить значения числовых характеристик исследуемой случайной величины.

Определение. Выборочным средним называется среднее арифметическое значений случайной величины, принимаемых в выборке:

$$\bar{x}_B = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n} = \frac{\sum_{i=1}^{k} n_i x_i}{n},$$

где $\mathbf{x_i}$ – варианты, $\mathbf{n_i}$ - частоты.

Замечание. Выборочное среднее служит для оценки математического ожидания исследуемой случайной величины. В дальнейшем будет рассмотрен вопрос, насколько точной является такая оценка.

Определение. Выборочной дисперсией называется

$$D_B = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_B)^2}{n} = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2}{n}.$$

Определение. Выборочным средним квадратическим отклонением –

$$\sigma_B = \sqrt{D_B}.$$

Так же, как в теории случайных величин, можно доказать, что справедлива следующая формула для вычисления выборочной дисперсии:

$$D = \overline{x^2} - (\bar{x})^2.$$

Вопросы для самоконтроля

1. Генеральная и выборочная совокупность. Выборочный метод. Повторная, бесповторная, репрезентативная выборки.
2. Статистический и вариационный ряды. Варианты, частоты, относительные частоты.
3. Статистическое распределение выборки. Полигон.
4. Эмпирическая функция распределения $F^*(x)$. Её свойства.
5. Группированное статистическое распределение. Построение гистограммы.
6. Числовые характеристики выборки.

СПИСОК ЛИТЕРАТУРЫ

Основная

1. *Вентцель, Елена Сергеевна.* Теория вероятностей и ее инженерные приложения : учебное пособие / Е. С. Вентцель, Л. А. Овчаров. - 4-е изд., стер. - М. : Высш. шк., 2007. - 491 с. - (Учебники для вузов. Математика). - ISBN 978-5-06-005714-0
2. *Гмурман, В. Е.* Теория вероятностей и математическая статистика : учебное пособие / В. Е. Гмурман. - 12-е изд., перераб. - М. : Высш. образование, 2006. - 479 с. : ил. - (Основы наук).
3. *Рау, Валерий Георгиевич.* Практический курс математики и общей теории статистики : учебное пособие / В. Г. Рау. - М. : Высш. шк., 2006. - 126 с. : ил. -ISBN 5-06-005529-9

Дополнительная

1. *Хучраева, Тамара Сергеевна.* Высшая математика : учебное пособие / Т. С. Хучраева, А. А. Смоленинов, Т. В. Кириллова. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2008. - 156 с. - ISBN 978-5-7011-0559-9
2. *Шириков, Виктор Филиппович.* Теория вероятностей : учебное пособие / В. Ф. Шириков, С. М. Зарбалиев. - М. : КолосС, 2008. - 389 с. - (Учебники и учеб. пособия для студентов высш. учеб. заведений). - ISBN 978-5-9532-0621-1
3. *Юрьева, Александра Антоновна.* Руководство к решению задач по теории вероятностей и математической статистике. : учебное пособие / А. А. Юрьева. - 2-е изд., испр. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2009. - 208 с.

Лекция 6

ЭЛЕМЕНТЫ ТЕОРИИ ОЦЕНОК

6.1. Основные свойства статистических оценок

Пусть Θ* - статистическая оценка неизвестного параметра Θ теоретического распределения. Извлечем из генеральной совокупности несколько выборок одного и того же объема **n** и вычислим для каждой из них оценку параметра Θ: $\Theta_1^*, \Theta_2^*, \ldots, \Theta_k^*$. Тогда оценку Θ* можно рассматривать как случайную величину, принимающую возможные значения $\Theta_1^*, \Theta_2^*, \ldots, \Theta_k^*$. Если математическое ожидание Θ* не равно оцениваемому параметру, мы будем получать при вычислении оценок систематические ошибки одного знака (с избытком, если **M**(Θ*) >Θ, и с недостатком, если **M**(Θ*) < Θ). Следовательно, необходимым условием отсутствия систематических ошибок является требование **M**(Θ*) = Θ.

Определение Статистическая оценка Θ* называется **несмещенной**, если ее математическое ожидание равно оцениваемому параметру Θ при любом объеме выборки:

$$\mathbf{M}(\Theta^*) = \Theta.$$

Смещенной называют оценку, математическое ожидание которой не равно оцениваемому параметру.

Определение Статистическая оценка называется **эффективной**, если она при заданном объеме выборки **n** имеет наименьшую возможную дисперсию.

При рассмотрении выборок большого объема к статистическим оценкам предъявляется еще и требование состоятельности.

Определение **Состоятельной** называется статистическая оценка, которая при **n**→∞ стремится по вероятности к оцениваемому параметру (если эта оценка несмещенная, то она будет состоятельной, если при **n**→∞ ее дисперсия стремится к 0).

В отличие от выборочного среднего, выборочная дисперсия является смещенной оценкой дисперсии генеральной совокупности. Можно доказать, что

$$M(D_В) = \frac{n-1}{n} D_Г,$$

где $\mathbf{D_Г}$ – истинное значение дисперсии генеральной совокупности. Можно предложить другую оценку дисперсии – **исправленную дисперсию s^2**, вычисляемую по формуле

$$s^2 = \frac{n}{n-1} D_B = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2}{n-1}.$$

Такая оценка будет являться несмещенной. Ей соответствует **исправленное среднее квадратическое отклонение**

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2}{n-1}}.$$

Определение. Оценка некоторого признака называется **асимптотически несмещенной**, если для выборки $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$

$$\lim_{n \to \infty} \frac{x_1 + x_2 + \ldots + x_n}{n} = X,$$

где $\mathbf{X}$ – истинное значение исследуемой величины.

6.2. Способы построения оценок

1. Метод наибольшего правдоподобия.

Пусть $\mathbf{X}$ – дискретная случайная величина, которая в результате испытаний приняла значения $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Предположим, что нам известен закон распределения этой величины, определяемый параметром Θ, но неизвестно численное значение этого параметра. Найдем его точечную оценку.

Пусть $\mathbf{p}(\mathbf{x}_i, \Theta)$ – вероятность того, что в результате испытания величина $\mathbf{X}$ примет значение $\mathbf{x}_i$. Назовем **функцией правдоподобия** дискретной случайной величины $\mathbf{X}$ функцию аргумента Θ, определяемую по формуле:

$$\mathbf{L}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n; \Theta) = \mathbf{p}(\mathbf{x}_1,\Theta)\mathbf{p}(\mathbf{x}_2,\Theta)\ldots\mathbf{p}(\mathbf{x}_n,\Theta).$$

Тогда в качестве точечной оценки параметра Θ принимают такое его значение $\Theta^* = \Theta(\mathbf{x}_1,\mathbf{x}_2, \ldots, \mathbf{x}_n)$, при котором функция правдоподобия достигает максимума. Оценку Θ^* называют **оценкой наибольшего правдоподобия**.

Для непрерывной случайной величины с известным видом плотности распределения $\mathbf{f}(\mathbf{x})$ и неизвестным параметром Θ функция правдоподобия имеет вид:

$$\mathbf{L}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n; \Theta) = \mathbf{f}(\mathbf{x}_1,\Theta)\mathbf{f}(\mathbf{x}_2,\Theta)\ldots\mathbf{f}(\mathbf{x}_n,\Theta).$$

2. Метод моментов.

Метод моментов основан на том, что начальные и центральные эмпирические моменты являются состоятельными оценками соответственно начальных и центральных теоретических моментов, поэтому можно приравнять теоретические моменты соответствующим эмпирическим моментам того же порядка.

Если задан вид плотности распределения **f(x**, Θ), определяемой одним неизвестным параметром Θ, то для оценки этого параметра достаточно иметь одно уравнение. Например, можно приравнять начальные моменты первого порядка:

$$\bar{x}_B = M(X) = \int_{-\infty}^{\infty} x f(x;\Theta)dx = \varphi(\Theta),$$

получив тем самым уравнение для определения Θ. Его решение Θ* будет точечной оценкой параметра, которая является функцией от выборочного среднего и, следовательно, и от вариант выборки:

$$\Theta = \psi\,(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n).$$

3. Метод наименьших квадратов.

Если требуется оценить зависимость величин **y** и **x**, причем известен вид связывающей их функции, но неизвестны значения входящих в нее коэффициентов, их величины можно оценить по имеющейся выборке с помощью метода наименьших квадратов. Для этого функция **y** = φ (**x**) выбирается так, чтобы сумма квадратов отклонений наблюдаемых значений $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ от $\varphi(\mathbf{x}_i)$ была минимальной:

$$\sum_{i=1}^{n} (y_i - \varphi(x_i))^2 = \min .$$

При этом требуется найти стационарную точку функции φ(**x**; **a, b, c…**), то есть решить систему:

$$\begin{cases} \sum_{i=1}^{n} (y_i - \varphi(x_i; a, b, c\ldots))\left(\dfrac{\partial \varphi}{\partial a}\right)_i = 0 \\ \sum_{i=1}^{n} (y_i - \varphi(x_i; a, b, c\ldots))\left(\dfrac{\partial \varphi}{\partial b}\right)_i = 0 \\ \sum_{i=1}^{n} (y_i - \varphi(x_i; a, b, c\ldots))\left(\dfrac{\partial \varphi}{\partial c}\right)_i = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{cases}$$

(решение, конечно, возможно только в случае, когда известен конкретный вид функции φ).

6.3 Интервальные оценки

При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, что приводит к грубым ошибкам. Поэтому в таком случае лучше пользоваться **интервальными оценками**, то есть указывать интервал, в который с заданной вероятностью попадает истинное значение оцениваемого параметра. Разумеется, чем меньше длина этого интервала, тем точнее оценка параметра. Поэтому, если для оценки Θ* некоторого параметра Θ справедливо неравенство | Θ* - Θ | < δ, число δ > 0 характеризует **точность оценки** (чем меньше δ, тем точнее оценка). Но статистические методы позволяют говорить только о том, что это неравенство выполняется с некоторой вероятностью.

Определение Надежностью (доверительной вероятностью) оценки Θ* параметра Θ называется вероятность γ того, что выполняется неравенство | Θ* - Θ | < δ. Если заменить это неравенство двойным неравенством – δ < Θ* - Θ < δ, то получим:

$$\mathbf{p}\,(\Theta^* - \delta < \Theta < \Theta^* + \delta) = \gamma.$$

Таким образом, γ есть вероятность того, что Θ попадает в интервал (Θ* - δ, Θ* + δ).

Определение Доверительным называется интервал, в который попадает неизвестный параметр с заданной надежностью γ.

6.4. Построение доверительных интервалов.

1. Доверительный интервал для оценки математического ожидания нормального распределения при известной дисперсии.

Пусть исследуемая случайная величина **X** распределена по нормальному закону с известным средним квадратическим σ, и требуется по значению выборочного среднего $\bar{x}_B$ оценить ее математическое ожидание **a**. Будем рассматривать выборочное среднее $\bar{x}_B$ как случайную величину $\bar{X}$, а значения вариант выборки $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\mathbf{n}$ как одинаково распределенные независимые случайные величины $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_\mathbf{n}$, каждая из которых имеет математическое ожидание **a** и среднее квадратическое отклонение σ. При этом $\mathbf{M}(\bar{X}) = \mathbf{a}$, $\sigma(\bar{X}) = \dfrac{\sigma}{\sqrt{n}}$ (используем свойства математического ожидания и дисперсии суммы независимых случайных величин). Оценим вероятность выполнения неравенства $|\bar{X} - a| < \delta$. Применим формулу для вероятности попадания нормально распределенной случайной величины в заданный интервал:

$\mathbf{p}\,(|X - a| < \delta) = 2\Phi\left(\dfrac{\delta}{\sigma}\right)$. Тогда , с учетом того, что $\sigma(\bar{X}) = \dfrac{\sigma}{\sqrt{n}}$, $\mathbf{p}\,(|\bar{X} - a| < \delta) = 2\Phi\left(\dfrac{\delta\sqrt{n}}{\sigma}\right) = 2\Phi(\mathbf{t})$, где $t = \dfrac{\delta\sqrt{n}}{\sigma}$. Отсюда $\delta = \dfrac{t\sigma}{\sqrt{n}}$, и предыдущее равенство можно переписать так:

$$p\left(\bar{x}_B - \frac{t\sigma}{\sqrt{n}} < a < \bar{x}_B + \frac{t\sigma}{\sqrt{n}}\right) = 2\Phi(t) = \gamma$$

.

Итак, значение математического ожидания **a** с вероятностью (надежностью) γ попадает в интервал $\left(\bar{x}_B - \frac{t\sigma}{\sqrt{n}}; \bar{x}_B + \frac{t\sigma}{\sqrt{n}}\right)$, где значение **t** определяется из таблиц для функции Лапласа так, чтобы выполнялось равенство 2Φ(**t**) = γ.

1. Доверительный интервал для оценки математического ожидания нормального распределения при неизвестной дисперсии.

Если известно, что исследуемая случайная величина **X** распределена по нормальному закону с неизвестным средним квадратическим отклонением, то для поиска доверительного интервала для ее математического ожидания построим новую случайную величину

$$T = \frac{\bar{x}_B - a}{\frac{s}{\sqrt{n}}},$$

где $\bar{x}_B$ - выборочное среднее, **s** – исправленная дисперсия, **n** – объем выборки. Эта случайная величина, возможные значения которой будем обозначать **t**, имеет распределение Стьюдента (см. лекцию 12) с **k = n** – 1 степенями свободы.

Поскольку плотность распределения Стьюдента $s(t,n) = B_n\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}$, где $B_n = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)}$, явным образом не зависит от **a** и σ, можно задать вероятность ее попадания в некоторый интервал (- t_γ , t_γ), учитывая четность плотности распределения, следующим образом: $p\left(\left|\frac{\bar{x}_B - a}{\frac{s}{\sqrt{n}}}\right| < t_\gamma\right) = 2\int_0^{t_\gamma} s(t,n)dt = \gamma$.

Отсюда получаем:

$$p\left(\bar{x}_B - \frac{t_\gamma s}{\sqrt{n}} < a < \bar{x}_B + \frac{t_\gamma s}{\sqrt{n}}\right) = \gamma.$$

Таким образом, получен доверительный интервал для **a**, где t_γ можно найти по соответствующей таблице при заданных **n** и γ.

3. Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения.

Будем искать для среднего квадратического отклонения нормально распределенной случайной величины доверительный интервал вида (**s – δ, s +δ**), где **s** – исправленное выборочное среднее квадратическое отклонение, а для δ выполняется условие: **p** (|σ – **s**| < δ) = γ.

Запишем это неравенство в виде: $s\left(1-\frac{\delta}{s}\right)<\sigma<s\left(1+\frac{\delta}{s}\right)$ или, обозначив $q=\frac{\delta}{s}$,

$$s(1-q)<\sigma<s(1+q). \qquad (1)$$

Рассмотрим случайную величину χ, определяемую по формуле

$$\chi=\frac{s}{\sigma}\sqrt{n-1},$$

которая распределена по закону «хи-квадрат» с **n**-1 степенями свободы Плотность ее распределения

$$R(\chi,n)=\frac{\chi^{n-2}e^{-\frac{\chi^2}{2}}}{2^{\frac{n-3}{2}}\Gamma\left(\frac{n-1}{2}\right)}$$

не зависит от оцениваемого параметра σ, а зависит только от объема выборки **n**. Преобразуем неравенство (1) так, чтобы оно приняло вид $\chi_1<\chi<\chi_2$. Вероятность выполнения этого неравенства равна доверительной вероятности γ, следовательно, $\int_{\chi_1}^{\chi_2}R(\chi,n)d\chi=\gamma.$

Предположим, что **q** < 1, тогда неравенство (1) можно записать так:

$$\frac{1}{s(1+q)}<\frac{1}{\sigma}<\frac{1}{s(1-q)},$$

или, после умножения на $s\sqrt{n-1}$, $\frac{\sqrt{n-1}}{1+q}<\frac{s\sqrt{n-1}}{\sigma}<\frac{\sqrt{n-1}}{1-q}$

Следовательно,

$$\frac{\sqrt{n-1}}{1+q}<\chi<\frac{\sqrt{n-1}}{1-q}.$$

Существуют таблицы для распределения «хи-квадрат», из которых можно найти **q** по заданным **n** и γ, не решая этого уравнения. Таким образом, вычислив по выборке значение **s** и определив по таблице значение **q**, можно найти доверительный интервал (1), в который значение σ попадает с заданной вероятностью γ.

Замечание. Если **q** > 1, то с учетом условия σ > 0 доверительный интервал для σ будет иметь границы

$$0 < \sigma < s(1+q).$$

Вопросы для самоконтроля

1. Основные свойства статистических оценок.
2. Способы построения оценок.
3. Построение доверительных интервалов.

СПИСОК ЛИТЕРАТУРЫ

Основная

1. *Вентцель, Елена Сергеевна.* Теория вероятностей и ее инженерные приложения : учебное пособие / Е. С. Вентцель, Л. А. Овчаров. - 4-е изд., стер. - М. : Высш. шк., 2007. - 491 с. - (Учебники для вузов. Математика). - ISBN 978-5-06-005714-0

2. *Гмурман, В. Е.* Теория вероятностей и математическая статистика : учебное пособие / В. Е. Гмурман. - 12-е изд., перераб. - М. : Высш. образование, 2006. - 479 с. : ил. - (Основы наук).

3. *Рау, Валерий Георгиевич.* Практический курс математики и общей теории статистики : учебное пособие / В. Г. Рау. - М. : Высш. шк., 2006. - 126 с. : ил. -ISBN 5-06-005529-9

Дополнительная

1. *Хучраева, Тамара Сергеевна.* Высшая математика : учебное пособие / Т. С. Хучраева, А. А. Смоленинов, Т. В. Кириллова. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2008. - 156 с. - ISBN 978-5-7011-0559-9

2. *Шириков, Виктор Филиппович.* Теория вероятностей : учебное пособие / В. Ф. Шириков, С. М. Зарбалиев. - М. : КолосС, 2008. - 389 с. - (Учебники и учеб. пособия для студентов высш. учеб. заведений). - ISBN 978-5-9532-0621-1

3. *Юрьева, Александра Антоновна.* Руководство к решению задач по теории вероятностей и математической статистике. : учебное пособие / А. А. Юрьева. - 2-е изд., испр. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2009. - 208 с.

Лекция 7

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

7.1. Понятие статистической гипотезы

Определение. Статистической гипотезой называют гипотезу о виде неизвестного распределения генеральной совокупности или о параметрах известных распределений.

Определение. Нулевой (основной) называют выдвинутую гипотезу $\mathbf{H}_0$. **Конкурирующей (альтернативной)** называют гипотезу $\mathbf{H}_1$, которая противоречит нулевой.

Пример. Пусть $\mathbf{H}_0$ заключается в том, что математическое ожидание генеральной совокупности $\mathbf{a} = 3$. Тогда возможные варианты $\mathbf{H}_1$: а) $\mathbf{a} \neq 3$; б) $\mathbf{a} > 3$; в) $\mathbf{a} < 3$.

Определение. Простой называют гипотезу, содержащую только одно предположение, **сложной** – гипотезу, состоящую из конечного или бесконечного числа простых гипотез.

Определение. Вероятность ошибки первого рода называется **уровнем значимости α.**

Основной прием проверки статистических гипотез заключается в том, что по имеющейся выборке вычисляется значение некоторой случайной величины, имеющей известный закон распределения.

Определение. Статистическим критерием называется случайная величина **К** с известным законом распределения, служащая для проверки нулевой гипотезы.

Определение. Критической областью называют область значений критерия, при которых нулевую гипотезу отвергают, **областью принятия гипотезы** – область значений критерия, при которых гипотезу принимают.

Итак, процесс проверки гипотезы состоит из следующих этапов:

1) выбирается статистический критерий **К**;

2) вычисляется его наблюдаемое значение $\mathbf{K}_{\mathbf{набл}}$ по имеющейся выборке;

3) поскольку закон распределения **К** известен, определяется (по известному уровню значимости α **критическое значение** $\mathbf{k}_{\mathbf{кр}}$, разделяющее критическую область и область принятия гипотезы (например, если $\mathbf{p}(\mathbf{K} > \mathbf{k}_{\mathbf{кр}}) = \alpha$, то справа от $\mathbf{k}_{\mathbf{кр}}$ располагается критическая область, а слева – область принятия гипотезы);

4) если вычисленное значение $\mathbf{K}_{\mathbf{набл}}$ попадает в область принятия гипотезы, то нулевая гипотеза принимается, если в критическую область – нулевая гипотеза отвергается.

Различают разные виды критических областей:

- **правостороннюю** критическую область, определяемую неравенством $\mathbf{K} > \mathbf{k}_{\mathbf{кр}}$ ($\mathbf{k}_{\mathbf{кр}} > 0$);

- **левостороннюю** критическую область, определяемую неравенством $\mathbf{K} < \mathbf{k}_{кр}$ ($\mathbf{k}_{кр} < 0$);

- **двустороннюю** критическую область, определяемую неравенствами $\mathbf{K} < \mathbf{k}_1$, $\mathbf{K} > \mathbf{k}_2$ $(\mathbf{k}_2 > \mathbf{k}_1)$.

7.2 Критерий для проверки гипотезы о вероятности события

Пусть проведено **n** независимых испытаний (**n** – достаточно большое число), в каждом из которых некоторое событие **A** появляется с одной и той же, но неизвестной вероятностью **p**, и найдена относительная частота $\frac{m}{n}$ появлений **A** в этой серии испытаний. Проверим при заданном уровне значимости α нулевую гипотезу $\mathbf{H}_0$, состоящую в том, что вероятность **p** равна некоторому значению $\mathbf{p}_0$.

Критическая область строится в зависимости от вида конкурирующей гипотезы.

1) Если $\mathbf{H}_0$: $\mathbf{p} = \mathbf{p}_0$, а $\mathbf{H}_1$: $\mathbf{p} \neq \mathbf{p}_0$, то критическую область нужно построить так, чтобы вероятность попадания критерия в эту область равнялась заданному уровню значимости α. При этом наибольшая мощность критерия достигается тогда, когда критическая область состоит из двух интервалов, вероятность попадания в каждый из которых равна $\frac{\alpha}{2}$. Поскольку **U** симметрична относительно оси О**y**, вероятность ее попадания в интервалы (-∞; 0) и (0; +∞) равна 0,5, следовательно, критическая область тоже должна быть симметрична относительно О**y**. Поэтому $\mathbf{u}_{кр}$ определяется по таблице значений функции Лапласа из условия $\Phi(u_{кр}) = \frac{1-\alpha}{2}$, а критическая область имеет вид $(-\infty; -u_{кр}) \cup (u_{кр}; +\infty)$.

2) Если конкурирующая гипотеза $\mathbf{H}_1$: $\mathbf{p} > \mathbf{p}_0$, то критическая область определяется неравенством $\mathbf{U} > \mathbf{u}_{кр}$, то есть является правосторонней, причем $\mathbf{p}(\mathbf{U} > \mathbf{u}_{кр}) = \alpha$. Тогда $p(0 < U < u_{кр}) = \frac{1}{2} - \alpha = \frac{1-2\alpha}{2}$. Следовательно, $\mathbf{u}_{кр}$ можно найти по таблице значений функции Лапласа из условия, что $\Phi(u_{кр}) = \frac{1-2\alpha}{2}$. Вычислим наблюдаемое значение критерия по формуле (19.2).

Если $\mathbf{U}_{набл} < \mathbf{u}_{кр}$, то нулевая гипотеза принимается.

Если $\mathbf{U}_{набл} > \mathbf{u}_{кр}$, то нулевая гипотеза отвергается.

3) Для конкурирующей гипотезы $\mathbf{H}_1$: $\mathbf{p} < \mathbf{p}_0$ критическая область является левосторонней и задается неравенством $\mathbf{U} < -\mathbf{u}_{кр}$, где $\mathbf{u}_{кр}$ вычисляется так же, как в предыдущем случае.

Если $\mathbf{U}_{набл} > -\mathbf{u}_{кр}$, то нулевая гипотеза принимается.

Если $\mathbf{U}_{набл} < -\mathbf{u}_{кр}$, то нулевая гипотеза отвергается.

7.3. Критерий для проверки гипотезы о математическом ожидании

Пусть генеральная совокупность **X** имеет нормальное распределение, и требуется проверить предположение о том, что ее математическое ожидание равно некоторому числу $\mathbf{a}_0$. Рассмотрим две возможности.

1) Известна дисперсия σ^2 генеральной совокупности. Тогда по выборке объема **n** найдем выборочное среднее $\overline{x}_B$ и проверим нулевую гипотезу $\mathbf{H}_0$: **M(X)** = $\mathbf{a}_0$.

Учитывая, что выборочное среднее $\overline{X}$ является несмещенной оценкой **M(X)**, то есть $\mathbf{M}(\overline{X})$ = **M(X)**, можно записать нулевую гипотезу так: $\mathbf{M}(\overline{X}) = \mathbf{a}_0$. Для ее проверки выберем критерий

$$U = \frac{\overline{X} - a_0}{\sigma(\overline{X})} = \frac{(\overline{X} - a_0)\sqrt{n}}{\sigma}.$$

Это случайная величина, имеющая нормальное распределение, причем, если нулевая гипотеза справедлива, то **M(U)** = 0, **σ(U)** = 1.

Выберем критическую область в зависимости от вида конкурирующей гипотезы:

- если $\mathbf{H}_1$: $\mathbf{M}(\overline{X}) \neq \mathbf{a}_0$, то $\mathbf{u}_{кр}$: $\Phi(u_{кр}) = \dfrac{1-\alpha}{2}$, критическая область двусторонняя, $U_{набл} = \dfrac{(\overline{x} - a_0)\sqrt{n}}{\sigma}$, и, если $|\mathbf{U}_{набл}| < \mathbf{u}_{кр}$, то нулевая гипотеза принимается; если $|\mathbf{U}_{набл}| > \mathbf{u}_{кр}$, то нулевая гипотеза отвергается.

- если $\mathbf{H}_1$: $\mathbf{M}(\overline{X}) > \mathbf{a}_0$, то $\mathbf{u}_{кр}$: $\Phi(u_{кр}) = \dfrac{1-2\alpha}{2}$, критическая область правосторонняя, и, если $\mathbf{U}_{набл} < \mathbf{u}_{кр}$, то нулевая гипотеза принимается; если $\mathbf{U}_{набл} > \mathbf{u}_{кр}$, то нулевая гипотеза отвергается.

- если $\mathbf{H}_1$: $\mathbf{M}(\overline{X}) < \mathbf{a}_0$, то $\mathbf{u}_{кр}$: $\Phi(u_{кр}) = \dfrac{1-2\alpha}{2}$, критическая область левосторонняя, и, если $\mathbf{U}_{набл} > -\mathbf{u}_{кр}$, то нулевая гипотеза принимается; если $\mathbf{U}_{набл} < -\mathbf{u}_{кр}$, то нулевая гипотеза отвергается.

2) Дисперсия генеральной совокупности неизвестна.

В этом случае выберем в качестве критерия случайную величину

$$T = \frac{(\overline{X} - a_0)\sqrt{n}}{S},$$

где **S** – исправленное среднее квадратическое отклонение. Такая случайная величина имеет распределение Стьюдента с **k** = **n** – 1 степенями свободы. Рассмотрим те же, что и в предыдущем случае, конкурирующие гипотезы и соответствующие им критические области. Предварительно вычислим наблюдаемое значение критерия:

$$T_{набл} = \frac{(\bar{x}_B - a_0)\sqrt{n}}{S}.$$

- если $\mathbf{H_1}$: $\mathbf{M}(\bar{X}) \neq \mathbf{a_0}$, то критическая точка $\mathbf{t_{двуст.кр.}}$ находится по таблице критических точек распределения Стьюдента по известным α и $\mathbf{k} = \mathbf{n} - 1$.

если | $\mathbf{T_{набл}}$ | < $\mathbf{t_{двуст.кр.}}$, то нулевая гипотеза принимается,

если | $\mathbf{T_{набл}}$ | > $\mathbf{t_{двуст.кр.}}$, то нулевая гипотеза отвергается.

- если $\mathbf{H_1}$: $\mathbf{M}(\bar{X}) > \mathbf{a_0}$, то по соответствующей таблице находят $\mathbf{t_{правост.кр.}}(\alpha, \mathbf{k})$ – критическую точку правосторонней критической области. Нулевая гипотеза принимается, если $\mathbf{T_{набл}} < \mathbf{t_{правост.кр.}}$

- при конкурирующей гипотезе $\mathbf{H_1}$: $\mathbf{M}(\bar{X}) < \mathbf{a_0}$ критическая область является левосторонней, и нулевая гипотеза принимается при условии $\mathbf{T_{набл}} > -\mathbf{t_{правост.кр.}}$. Если $\mathbf{T_{набл}} < -\mathbf{t_{правост.кр.}}$, нулевую гипотезу отвергают.

7.4. Критерий для проверки гипотезы о сравнении двух дисперсий

Пусть имеются две нормально распределенные генеральные совокупности **X** и **Y**. Из них извлечены независимые выборки объемов соответственно $\mathbf{n_1}$ и $\mathbf{n_2}$, по которым вычислены исправленные выборочные дисперсии s_X^2 и s_Y^2. Требуется при заданном уровне значимости α проверить нулевую гипотезу $\mathbf{H_0}$: $\mathbf{D(X)} = \mathbf{D(Y)}$ о равенстве дисперсий рассматриваемых генеральных совокупностей. Учитывая несмещенность исправленных выборочных дисперсий, можно записать нулевую гипотезу так:

$$\mathbf{H_0}: \mathbf{M}(s_X^2) = \mathbf{M}(s_Y^2).$$

Пусть $\mathbf{H_1}$: $\mathbf{D(X)} > \mathbf{D(Y)}$. Наблюдаемым значением критерия будет отношение большей из исправленных дисперсий к меньшей: $F_{набл} = \frac{s_б^2}{s_M^2}$. По таблице критических точек распределения Фишера-Снедекора можно найти критическую точку $\mathbf{F_{набл}}(\boldsymbol{\alpha}; \mathbf{k_1}; \mathbf{k_2})$. При $\mathbf{F_{набл}} < \mathbf{F_{кр}}$ нулевая гипотеза принимается, при $\mathbf{F_{набл}} > \mathbf{F_{кр}}$ отвергается.

Если $\mathbf{H_1}$: $\mathbf{D(X)} \neq \mathbf{D(Y)}$, то критическая область является двусторонней и определяется неравенствами $\mathbf{F} < \mathbf{F_1}$, $\mathbf{F} > \mathbf{F_2}$, где $\mathbf{p}(\mathbf{F} < \mathbf{F_1}) = \mathbf{p}(\mathbf{F} > \mathbf{F_2}) = \alpha/2$. При этом достаточно найти правую критическую точку $\mathbf{F_2} = \mathbf{F_{кр}}\left(\frac{\alpha}{2}, \mathbf{k_1}, \mathbf{k_2}\right)$. Тогда при $\mathbf{F_{набл}} < \mathbf{F_{кр}}$ нулевая гипотеза принимается, при $\mathbf{F_{набл}} > \mathbf{F_{кр}}$ отвергается.

Были рассмотрены гипотезы, в которых закон распределения генеральной совокупности предполагался известным. Теперь займемся проверкой гипотез о предполагаемом законе неизвестного распределения, то есть будем проверять нулевую гипотезу о том, что генеральная совокупность распределена по некоторому известному закону. Обычно статистические критерии для проверки таких гипотез называются **критериями согласия**.

7.5. Критерий Пирсона

Достоинством критерия Пирсона является его универсальность: с его помощью можно проверять гипотезы о различных законах распределения.

Пусть получена выборка достаточно большого объема **n** с большим количеством различных значений вариант. Для удобства ее обработки разделим интервал от наименьшего до наибольшего из значений вариант на **s** равных частей и будем считать, что значения вариант, попавших в каждый интервал, приближенно равны числу, задающему середину интервала. Подсчитав число вариант, попавших в каждый интервал, составим так называемую сгруппированную выборку:

варианты………..x_1 x_2 … x_s
частоты…………. n_1 n_2 … n_s ,

где x_i – значения середин интервалов, а n_i – число вариант, попавших в **i**-й интервал (эмпирические частоты).

По полученным данным можно вычислить выборочное среднее $\bar{x}_B$ и выборочное среднее квадратическое отклонение σ_B. Проверим предположение, что генеральная совокупность распределена по нормальному закону с параметрами **M**(**X**) = $\bar{x}_B$,**D**(**X**) = σ_B^2. Тогда можно найти количество чисел из выборки объема **n**, которое должно оказаться в каждом интервале при этом предположении (то есть теоретические частоты). Для этого по таблице значений функции Лапласа найдем вероятность попадания в **i**-й интервал:

$$p_i = \Phi\left(\frac{b_i - \bar{x}_B}{\sigma_B}\right) - \Phi\left(\frac{a_i - \bar{x}_B}{\sigma_B}\right),$$

где a_i и b_i- границы **i**-го интервала. Умножив полученные вероятности на объем выборки п, найдем теоретические частоты: $n_i = n \cdot p_i$. Наша цель – сравнить эмпирические и теоретические частоты, которые, конечно, отличаются друг от друга, и выяснить, являются ли эти различия несущественными, не опровергающими гипотезу о нормальном распределении исследуемой случайной величины, или они настолько велики, что противоречат этой гипотезе. Для этого используется критерий в виде случайной величины

$$\chi^2 = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'} . \quad (1)$$

Смысл ее очевиден: суммируются части, которые квадраты отклонений эмпирических частот от теоретических составляют от соответствующих теоретических частот. Можно доказать, что вне зависимости от реального закона распределения генеральной совокупности закон распределения случайной величины (1) при $n \to \infty$ стремится к закону распределения χ^2 с числом степеней свободы **k = s** – 1 – **r**, где **r**– число параметров предполагаемого распределения, оцененных по данным выборки.

Нормальное распределение характеризуется двумя параметрами, поэтому **k = s** – 3. Для выбранного критерия строится правосторонняя критическая область, определяемая условием

$$p(\chi^2 > \chi^2_{kp}(\alpha, k)) = \alpha,$$

где **α** – уровень значимости. Следовательно, критическая область задается неравенством $\chi^2 > \chi^2_{kp}(\alpha, k),$ а область принятия гипотезы - $\chi^2 < \chi^2_{kp}(\alpha, k)$.

Итак, для проверки нулевой гипотезы **H**$_0$: генеральная совокупность распределена нормально – нужно вычислить по выборке наблюдаемое значение критерия:

$$\chi^2_{набл} = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'},$$

а по таблице критических точек распределения χ^2 найти критическую точку $\chi^2_{kp}(\alpha, k)$, используя известные значения α и **k = s** – 3. Если $\chi^2_{набл} < \chi^2_{kp}$ - нулевую гипотезу принимают, при $\chi^2_{набл} > \chi^2_{kp}$ ее отвергают.

Вопросы для самоконтроля

1. Понятие статистической гипотезы.
2 Критерий для проверки гипотезы о вероятности события.
3. Критерий для проверки гипотезы о математическом ожидании.
4. Критерий для проверки гипотезы о сравнении двух дисперсий.
5. Критерий Пирсона.

СПИСОК ЛИТЕРАТУРЫ

Основная

1. *Вентцель, Елена Сергеевна.* Теория вероятностей и ее инженерные приложения : учебное пособие / Е. С. Вентцель, Л. А. Овчаров. - 4-е изд., стер. - М. : Высш. шк., 2007. - 491 с. - (Учебники для вузов. Математика). - ISBN 978-5-06-005714-0
2. *Гмурман, В. Е.* Теория вероятностей и математическая статистика : учебное пособие / В. Е. Гмурман. - 12-е изд., перераб. - М. : Высш. образование, 2006. - 479 с. : ил. - (Основы наук).
3. *Рау, Валерий Георгиевич.* Практический курс математики и общей теории статистики : учебное пособие / В. Г. Рау. - М. : Высш. шк., 2006. - 126 с. : ил. -ISBN 5-06-005529-9

Дополнительная

1. *Хучраева, Тамара Сергеевна.* Высшая математика : учебное пособие / Т. С. Хучраева, А. А. Смоленинов, Т. В. Кириллова. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2008. - 156 с. - ISBN 978-5-7011-0559-9
2. *Шириков, Виктор Филиппович.* Теория вероятностей : учебное пособие / В. Ф. Шириков, С. М. Зарбалиев. - М. : КолосС, 2008. - 389 с. - (Учебники и учеб. пособия для студентов высш. учеб. заведений). - ISBN 978-5-9532-0621-1
3. *Юрьева, Александра Антоновна.* Руководство к решению задач по теории вероятностей и математической статистике. : учебное пособие / А. А. Юрьева. - 2-е изд., испр. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2009. - 208 с.

Лекция 8

ЭЛЕМЕНТЫ ТЕОРИИ КОРРЕЛЯЦИИ

Понятия корреляции появилось в 19-м веке благодаря работам английских статистов Пирсона и Гильтона. Термин корреляция в переводе с латинского означает соотношение, взаимосвязь. В настоящее время под корреляционным анализом понимают систему математических процедур, направленную на изучение статистических связей между признаками изучаемых объектов.

8.1. Виды зависимостей

Функциональная зависимость это зависимость, при которой каждому значению одной переменной соответствует определенное значение другой. Функциональная зависимость может иметь место как между неслучайными переменными (зависимость скорости падения тела в вакууме от времени) так и между случайными переменными (зависимость стоимости продаж изделий от их числа). **Статистическая или вероятностная зависимость**, это такой вид зависимости, при котором каждому значению одной переменной соответствует множество значений другой переменной. **Корреляционная зависимость** - такая статистическая зависимость между двумя переменными, при которой каждому значению одной переменной ставится в соответствие усредненное значение другой.

Основной задачей корреляционного анализа является выявление связи между признаками и оценкой тесноты этой связи.

Признаки делятся на 2 группы:

- результативный признак (результат) - характеристика эффективности некоторого процесса;

- факторные признаки - показатели, влияющие на результат;

В зависимости от числа факторных признаков выделяют следующие виды корреляционной зависимости:

1. парная корреляция - связь между результативным и одним факторным признаком

2. частная корреляция - связь между результативным и одним факторным признаком при фиксированном значении других факторов

3. множественная корреляция - связь между результативным и несколькими факторными признаками.

Связь называется прямой, если при увеличении факторного признака результативный признак так же увеличивается.

Связь называется обратной если при увеличении факторного признака результативный признак уменьшается.

Если зависимость описывается посредством прямой линии, то она называется линейной, если посредством параболы или иное - нелинейной.

Корреляционный анализ - метод, позволяющий обнаружить зависимость между несколькими случайными величинами. Методами корреляционного анализа решаются следующие задачи:

1) Взаимосвязь. Есть ли взаимосвязь между параметрами?

2) Прогнозирование. Если известно поведение одного параметра, то можно предсказать поведение другого параметра, коррелирующего с первым.

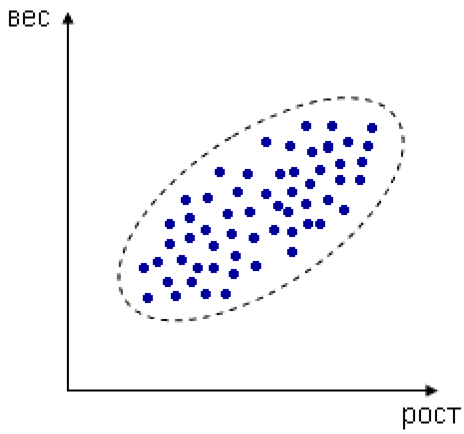
3) Классификация и идентификация объектов. Корреляционный анализ помогает подобрать набор независимых признаков для классификации.

Корреляционный анализ, как и другие статистические методы, основан на использовании вероятностных моделей, описывающих поведение исследуемых признаков в некоторой генеральной совокупности, из которой получены экспериментальные значения x_i и y_i.

Когда исследуется корреляция между количественными признаками, значения которых можно точно измерить в единицах метрических шкал (метры, секунды, килограммы и т.д.), то очень часто принимается модель двумерной нормально распределенной генеральной совокупности. Такая модель отображает зависимость между переменными величинами x_i и y_i графически в виде геометрического места точек в системе прямоугольных координат. Эту графическую зависимость называются также *диаграммой рассеивания* или ***корреляционным полем***.

Допустим, проводится независимое измерение различных параметров у одного типа объектов. Из этих данных можно получить качественно новую информацию о взаимосвязи этих параметров.

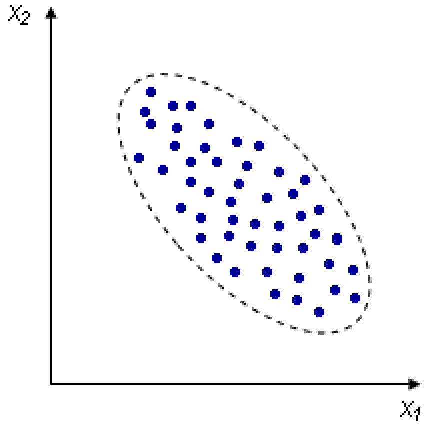
Например, измеряем рост и вес человека и каждое измерение представим точкой в системе прямоугольных координат :

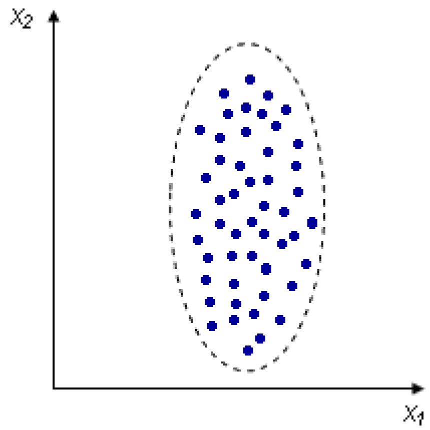


Несмотря на то, что величины носят случайный характер, в общем наблюдается некоторая зависимость - величины коррелируют.

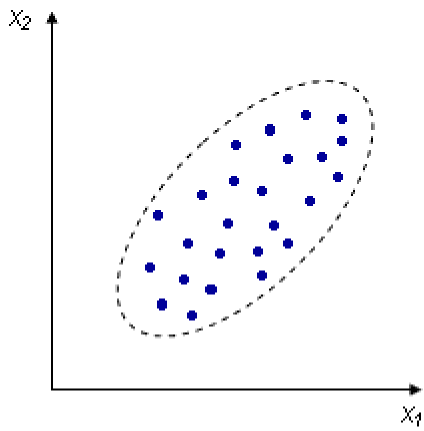
В данном случае это **положительная корреляция** (при увеличении одного параметра второй тоже увеличивается). Возможны также такие случаи:

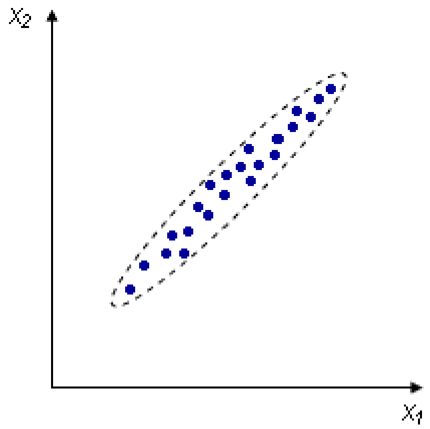
Отрицательная корреляция: Отсутствие корреляции:





Взаимосвязь между переменными необходимо охарактеризовать численно, чтобы, например, различать такие случаи:





Для этого вводится ***коэффициент корреляции***. Он рассчитывается следующим образом:

Есть массив из *n* точек $\{x_{1,i}, x_{2,i}\}$ Рассчитываются средние значения для каждого параметра:

$$\bar{x}_1 = \frac{\sum x_{1,i}}{n}, \quad \bar{x}_2 = \frac{\sum x_{2,i}}{n}$$

И коэффициент корреляции:

$$r = \frac{\sum (x_{1,i} - \bar{x}_1) \cdot (x_{2,i} - \bar{x}_2)}{\sqrt{\sum (x_{1,i} - \bar{x}_1)^2} \cdot \sqrt{\sum (x_{2,i} - \bar{x}_2)^2}}$$

8.2. Коэффициенты корреляции и их свойства

Коэффициент корреляции ***р*** для генеральной совокупности, как правило, неизвестен, поэтому он оценивается по экспериментальным данным, представляющим собой выборку объема ***n*** пар значений (x_i, y_i), полученную при совместном измерении двух признаков *X* и *Y*. Коэффициент корреляции, определяемый по выборочным данным, называется ***выборочным коэффициентом корреляции*** (или просто ***коэффициентом корреляции***). Его принято обозначать символом ***r***.

Коэффициенты корреляции — удобный показатель связи, получивший широкое применение в практике. К их основным свойствам необходимо отнести следующие:

1. Коэффициенты корреляции способны характеризовать только линейные связи, т.е. такие, которые выражаются уравнением линейной функции. При наличии нелинейной зависимости между варьирующими признаками следует использовать другие показатели связи.

2. Значения коэффициентов корреляции – это отвлеченные числа, лежащее в пределах от —1 до +1, т.е. **-1 < r < 1**.

3. При независимом варьировании признаков, когда связь между ними отсутствует, $r = 0$.

4. При положительной, или прямой, связи, когда с увеличением значений одного признака возрастают значения другого, коэффициент корреляции приобретает положительный **(+)** знак и находится в пределах от 0 до +1, т.е. **0 < r 1**.

5. При отрицательной, или обратной, связи, когда с увеличением значений одного признака соответственно уменьшаются значения другого, коэффициент корреляции сопровождается отрицательным **(–)** знаком и находится в пределах от 0 до –1, т.е. **-1 < r <0**.

6. Чем сильнее связь между признаками, тем ближе величина коэффициента корреляции к □1□. Если r = ±1, то корреляционная связь переходит в функциональную, т.е. каждому значению признака *X* будет соответствовать одно или несколько строго определенных значений признака *Y*.

7. Только по величине коэффициентов корреляции нельзя судить о достоверности корреляционной связи между признаками. Этот параметр зависит от числа степеней свободы $k = n - 2$, где: *n* – число коррелируемых пар показателей *X* и *Y*. Чем больше *n*, тем выше достоверность связи при одном и том же значении коэффициента корреляции.

В практической деятельности, когда число коррелируемых пар признаков *X* и *Y* не велико ($\mathbf{n} \leq 30$), то при оценке зависимости между показателями используется следующую градацию:

1) ***высокая степень взаимосвязи*** – значения коэффициента корреляции находится в пределах от 0,7 до 0,99;

2) ***средняя степень взаимосвязи*** – значения коэффициента корреляции находится в пределах от 0,5 до 0,69;

3**)** ***слабая степень взаимосвязи*** – значения коэффициента корреляции находится от 0,2 до 0,49.

Коэффициент ***r*** является случайной величиной, поскольку вычисляется из случайных величин. Для него можно выдвигать и проверять следующие гипотезы:

1. Коэффициент корреляции значимо отличается от нуля (т.е. есть взаимосвязь между величинами):

Тестовая статистика вычисляется по формуле:

$$\xi = \left(0.5 \cdot \ln\left(\frac{1+r}{1-r}\right) - \frac{|r|}{2(n-1)}\right)\sqrt{n-3}$$

и сравнивается с табличным значением коэффициента Стьюдента $t(p = 0.95,\ f = \infty) = 1.96$

Если тестовая статистика больше табличного значения, то коэффициент значимо отличается от нуля. По формуле видно, что чем больше измерений *n*, тем лучше (больше тестовая статистика, вероятнее, что коэффициент значимо отличается от нуля)

2. Отличие между двумя коэффициентами корреляции значимо:

Тестовая статистика:

$$\xi = 0.5 \cdot \ln\left(\frac{(1+r_1)(1-r_2)}{(1-r_1)(1+r_2)}\right) \cdot \frac{1}{\sqrt{\dfrac{1}{n_1-3} + \dfrac{1}{n_2-3}}}$$

Также сравнивается с табличным значением $t(p,\infty)$

Вопросы для самоконтроля

1. Функциональная и корреляционная зависимости между случайными величинами.
2. Основные задачи теории корреляции
3. Коэффициент корреляции и его свойства.
4. Линейная корреляция. Определение параметров прямой регрессии методом квадратов.

СПИСОК ЛИТЕРАТУРЫ

Основная

1. *Вентцель, Елена Сергеевна.* Теория вероятностей и ее инженерные приложения : учебное пособие / Е. С. Вентцель, Л. А. Овчаров. - 4-е изд., стер. - М. : Высш. шк., 2007. - 491 с. - (Учебники для вузов. Математика). - ISBN 978-5-06-005714-0

2. *Гмурман, В. Е.* Теория вероятностей и математическая статистика : учебное пособие / В. Е. Гмурман. - 12-е изд., перераб. - М. : Высш. образование, 2006. - 479 с. : ил. - (Основы наук).

3. *Рау, Валерий Георгиевич.* Практический курс математики и общей теории статистики : учебное пособие / В. Г. Рау. - М. : Высш. шк., 2006. - 126 с. : ил. -ISBN 5-06-005529-9

Дополнительная

1. *Хучраева, Тамара Сергеевна.* Высшая математика : учебное пособие / Т. С. Хучраева, А. А. Смоленинов, Т. В. Кириллова. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2008. - 156 с. - ISBN 978-5-7011-0559-9

2. *Шириков, Виктор Филиппович.* Теория вероятностей : учебное пособие / В. Ф. Шириков, С. М. Зарбалиев. - М. : КолосС, 2008. - 389 с. - (Учебники и учеб. пособия для студентов высш. учеб. заведений). - ISBN 978-5-9532-0621-1

3. *Юрьева, Александра Антоновна.* Руководство к решению задач по теории вероятностей и математической статистике. : учебное пособие / А. А. Юрьева. - 2-е изд., испр. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2009. - 208 с.

Лекция 9

ЭЛЕМЕНТЫ РЕГРЕССИОННОГО АНАЛИЗА

9.1. Понятие регрессии

В практических исследованиях возникает необходимость ***аппроксимировать (описать приблизительно)*** диаграмму рассеяния ***математическим уравнением***. То есть зависимость между переменными величинами *Y* и *X* можно выразить аналитически с помощью формул и уравнений и графически в виде геометрического места точек в системе прямоугольных координат. График корреляционной зависимости строится по уравнениям функции $\bar{y}_x = f(x)$ и $\bar{x}_y = f(y)$, которые называются ***регрессией*** (термин “регрессия” происходит от лат. regressio — движение назад). Здесь $\bar{y}_x$ и $\bar{x}_y$ — средние арифметические из числовых значений зависимых переменных *Y* и *X*.

Для выражения регрессии служат эмпирические и теоретические ряды, их графики — ***линии регрессии***, а также ***корреляционные уравнения (уравнения регрессии) и коэффициент линейной регрессии.*** Показатели регрессии выражают корреляционную связь двусторонне, учитывая изменение средней величины $\bar{y}_x$признака *Y* при изменении значений x_i признака *X*, и, наоборот, показывают изменение средней величины $\bar{x}_y$ признака *X* по измененным значениям y_i признака *Y*. Ряды регрессии, особенно их графики, дают наглядное представление о форме и тесноте корреляционной связи между признаками, в чем и заключается их ценность. И поэтому задача состоит в том, чтобы любую форму корреляционной связи выразить уравнением определенной функции (линейной, параболической и т.д.), что позволяет получать нужную информацию о корреляции между переменными величинами *Y* и *X*, предвидеть возможные изменения признака *Y* на основе известных изменений *X*, связанного с *Y* корреляционно.

9.2. Уравнение линейной регрессии

Обычно признак Y рассматривается как функция многих аргументов — $x_1, x_2, x_3, ...$— и может быть записана в виде:

$$y = a + bx_1 + cx_2 + dx_3 + ... ,$$

где: *a, b, c* и *d* — параметры уравнения, определяющие соотношение между аргументами и функцией. В практике учитываются не все, а лишь некоторые аргументы, в простейшем случае, как при описании линейной регрессии, всего один:

$$y = a + bx \quad (1).$$

В этом уравнении параметр *a* — свободный член; графически он представляет отрезок ординаты (у) в системе прямоугольных координат. Параметр *b* называется коэффициентом регрессии. С точки зрения аналитической геометрии *b*— угловой коэффициент, определяющий наклон линии регрессии по отношению к осям, координат. В области регрессионного анализа этот параметр показывает, насколько в среднем величина одного признака (*Y*) изменяется при изменении на единицу меры другого корреляционно связанного с Y признака X. Наглядное представление об этом параметре и о положении линий регрессии Y по X и X по Y в системе прямоугольных координат дает рисунок 2.1.

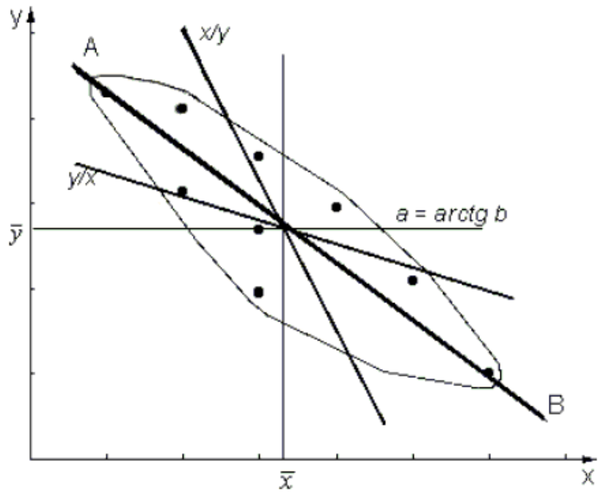


Рис. 2.1.

Линии регрессии, как показано, пересекаются в точке 0 $(\bar{x}, \bar{y})$, соответствующей средним арифметическим значениям корреляционно связанных друг с другом признаков Y и X. Линия АВ, проходящая через эту точку, изображает полную (функциональную) зависимость между переменными величинами Y и X, когда коэффициент корреляции $r = 1$. Чем сильнее связь между Y и X, тем ближе линии регрессии к АВ, и, наоборот, чем слабее связь между варьирующими признаками, тем более удаленными оказываются линии регрессии от АВ. При отсутствии связи между признаками, когда $r = 0$, линии регрессии оказываются под прямым углом (90°) по отношению друг к другу. Уравнение регрессии тем лучше описывает зависимость, чем меньше рассеяние диаграммы, чем больше теснота взаимосвязи. Уравнение прямой линии пригодно для описания только линейных зависимостей. В случае нелинейных зависимостей математическая запись может отображаться уравнениями параболы, гиперболы и др. Необходимо также сделать одно важное замечание о значении показателей, характеризующих взаимосвязь признаков (коэффициентов корреляции, регрессии и т. п.). Все они дают лишь количественную меру связи, но ничего не говорят о причинах зависимости. Определить эти причины — дело самого исследователя.

9.3. Коэффициенты уравнения парной линейной регрессии

Как уже было определено выше, в случае линейной зависимости уравнение регрессии является уравнением прямой линии. Таких уравнений два:

$$Y = a_1 + b_{y/x}X$$

$$X = a_2 + b_{x/y}Y \quad (2)$$

где: a и b – коэффициенты, или параметры, которые надлежит определить. Значение коэффициентов регрессии вычисляется по формуле:

$$b_{x/y} = r \cdot \frac{\sigma_x}{\sigma_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (y_i - \bar{y})^2}$$
$$b_{y/x} = r \cdot \frac{\sigma_y}{\sigma_x} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} . \quad (3)$$

Коэффициенты регрессии *b* имеют размерность, равную отношению размерностей изучаемых показателей *X* и *Y*, и тот же знак, что и коэффициент корреляции. Коэффициенты *a* определяются по формуле:

$$a_1 = \bar{y} - b_{y/x} \cdot \bar{x}$$

$$a_2 = \bar{x} - b_{x/y} \cdot \bar{y} . \quad (4)$$

Чтобы вычислить этот коэффициенты, надо просто в уравнения регрессии подставить средние значения коррелируемых переменных.

Для оценки качества уравнений регрессии вычисляются остаточные средние квадратические отклонения (или абсолютные погрешности уравнений) по формуле:

$$\sigma_{y/x} = \sigma_y \cdot \sqrt{1 - r^2}$$
$$\sigma_{x/y} = \sigma_x \cdot \sqrt{1 - r^2} . \quad (5)$$

Эти оценки абсолютны и, следовательно, не могут быть сравнимы друг с другом. Поэтому вводят оценки ***относительной погрешности уравнений***, которые выражаются в процентах и с*лужат для точности предсказания (прогнозирования) результатов одного показателя по заранее известным значениям другого*. Относительные погрешности уравнений регрессии определяются по формуле:

$$\sigma'_{y/x} = \frac{\sigma_{y/x}}{\bar{y}} \cdot 100\%$$
$$\sigma'_{x/y} = \frac{\sigma_{x/y}}{\bar{x}} \cdot 100\% . \quad (6)$$

Значение этой оценки, если $r = \pm 1$, равно нулю и, если $r = 0$, максимально. Остаточное среднее квадратическое отклонение характеризует колеблемость *Y* относительно линии регрессии по *X* в прямом уравнении регрессии и, наоборот, в обратном случае. А, следовательно, чем меньше величина относительной погрешности уравнения регрессии, тем точнее будет оно осуществлять прогноз значений одного показателя по заранее известным значениям другого.

9.4. Определение параметров парной линейной регрессии

Определение параметров линейной регрессии – одна из задач регрессионного анализа. Она решается способом наименьших квадратов, основанным на требовании, чтобы сумма квадратов отклонений вариант от линии регрессии была наименьшей. Этому требованию удовлетворяет следующая система нормальных уравнений:

$$an + b\sum x = \sum y,$$
$$a\sum x + b\sum x^2 = \sum xy.$$

Ряды регрессии — это ряды усредненных значений (y_x и x_y) варьирующих признаков *Y* и *X*, соответствующих значениям аргументов x_i и y_i. Поэтому эмпирические уравнения регрессии следует записывать так:

$$y_x = a_{y/x} + b_{y/x}{*}x$$
$$x_y = a_{x/y} + b_{x/y}{*}y \quad (7)$$

Формулы для определения параметров *a* и *b* принимают следующие выражения:

$$a_{y/x} = \bar{y} - b_{y/x} \cdot \bar{x}$$
$$a_{x/y} = \bar{x} - b_{x/y} \cdot \bar{y}. \quad (8)$$

Уравнение линейной регрессии можно выразить в виде отклонений вариант от их средних арифметических:

$$y_x - \bar{y} = b_{y/x} \cdot (x_i - \bar{x})$$
$$x_y - \bar{x} = b_{x/y} \cdot (y_i - \bar{y}). \quad (9)$$

В таком случае система нормальных уравнений для определения параметров *a* и *b* будет следующая:

$$an + b\sum (x_i - \bar{x}) = \sum (y_i - \bar{y});$$
$$a\sum (x_i - \bar{x}) + b\sum (x_i - \bar{x})^2 = \sum (y_i - \bar{y})(x_i - \bar{x}).$$

Если средние $\bar{x}$ и $\bar{y}$ перенести в правую часть уравнения (9), то при $a = \bar{y}$ система нормальных уравнений принимает следующий вид:

$$x_y = \bar{x} + b_{x/y}(y_i - \bar{y})$$
$$y_x = \bar{y} + b_{y/x}(x_i - \bar{x}), \quad (10)$$

Заменив в формуле параметры $b_{y/x}$ и $b_{x/y}$ на их значения из формулы (3), получим систему уравнений парной линейной регрессии:

$$y_x = \bar{y} + \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \cdot (x - \bar{x})$$
$$x_y = \bar{x} + \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (y_i - \bar{y})^2} \cdot (y - \bar{y}).$$

Эти уравнения удобны для определения параметров при отыскивании эмпирических уравнений регрессии в практической работе для точности прогнозирования результатов.

9.5. Графическое представление уравнения парной линейной регрессии

Эмпирические ряды регрессии Y по X и X по Y изображаются в виде линейного графика, при построении которого наиболее точным является использование способа наименьших квадратов, предложенного в 1806 г. К. Гауссом и независимо от него А. Лежандром. В основу этого способа положена теорема, согласно которой сумма квадратов отклонений вариант (x_i) от средней арифметической ($\bar{x}$) есть величина наименьшая, т. е. $\sum (x_i - \bar{x})^2 = \min$. Отсюда и название метода, который нашел широкое применение не только в биологии, но и в технике. Мы уже говорили об этом методе и применяли его, когда находили параметры a и b линейной регрессии, отыскивая эмпирическое уравнение.

При графическом изображении эмпирического уравнения регрессии (например, показатели роста и веса 10 исследуемых), представленного на рисунке 2.2 используется следующая последовательность:

1. Определив форму и направление взаимосвязи между эмпирическими данными на основе данных расчета нормированного коэффициента корреляции, производят расчет уравнений регресии (прямого и обратного).

2. Подставляя в конечный вид уравнений, выражающих зависимость между переменными величинами Y и X, эмпирические данные x_i и y_i находят координаты точек линий регрессии для усредненных значений y_x и x_y.

3. На графике, выполненном в прямоугольной системе координат, на оси x откладывают значения переменных x_i, на оси y – значения y_i и отмечают точками рассчитанные координаты линий регрессии для усредненных значений y_x и x_y (рис.2.2).

4. Две линии регрессии на графике пересекаются в точке М с координатами средних значений показателей x_i и y_i.

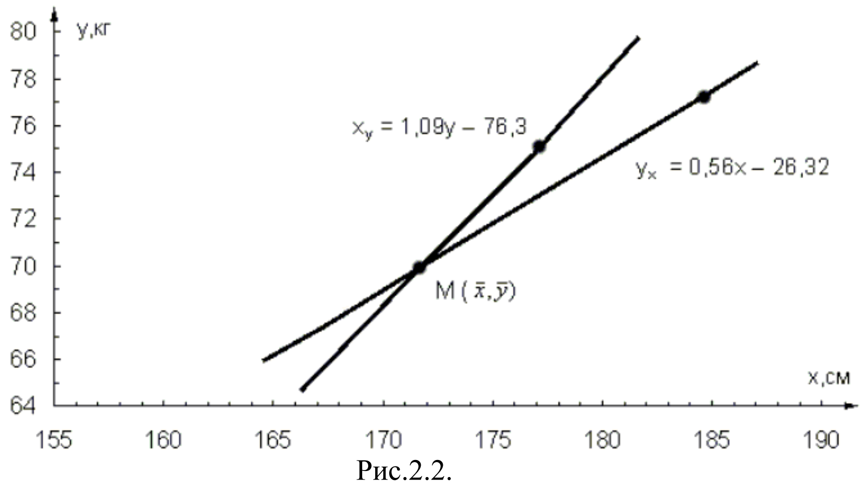


Рис.2.2.

График линий регрессии отражает ряды теоретически ожидаемых значений функции по известным значениям аргумента. При этом, чем сильнее взаимосвязь между величинами x_i и y_i, тем меньше угол между линиями регрессии. При r = ± 1 линии уравнения регресии либо совпадают, либо расположены параллельно, так как корреляционная зависимость между признаками в этом случае переходит в функциональную. И, наоборот, чем слабее зависимость между признаками, тем больше угол между линиями на графике. При r = 0 линии регрессии расположены перпендикулярно.

Вопросы для самоконтроля

1. Понятие регрессии.
2. Уравнение линейной регрессии.
3. Коэффициенты уравнения парной линейной регрессии.
4. Определение параметров парной линейной регрессии.
5. Графическое представление уравнения парной линейной регрессии.

СПИСОК ЛИТЕРАТУРЫ

Основная

1. *Вентцель, Елена Сергеевна.* Теория вероятностей и ее инженерные приложения : учебное пособие / Е. С. Вентцель, Л. А. Овчаров. - 4-е изд., стер. - М. : Высш. шк., 2007. - 491 с. - (Учебники для вузов. Математика). - ISBN 978-5-06-005714-0
2. *Гмурман, В. Е.* Теория вероятностей и математическая статистика : учебное пособие / В. Е. Гмурман. - 12-е изд., перераб. - М. : Высш. образование, 2006. - 479 с. : ил. - (Основы наук).
3. *Рау, Валерий Георгиевич.* Практический курс математики и общей теории статистики : учебное пособие / В. Г. Рау. - М. : Высш. шк., 2006. - 126 с. : ил. -ISBN 5-06-005529-9

Дополнительная

1. *Хучраева, Тамара Сергеевна.* Высшая математика : учебное пособие / Т. С. Хучраева, А. А. Смоленинов, Т. В. Кириллова. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2008. - 156 с. - ISBN 978-5-7011-0559-9
2. *Шириков, Виктор Филиппович.* Теория вероятностей : учебное пособие / В. Ф. Шириков, С. М. Зарбалиев. - М. : КолосС, 2008. - 389 с. - (Учебники и учеб. пособия для студентов высш. учеб. заведений). - ISBN 978-5-9532-0621-1
3. *Юрьева, Александра Антоновна.* Руководство к решению задач по теории вероятностей и математической статистике. : учебное пособие / А. А. Юрьева. - 2-е изд., испр. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2009. - 208 с.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. **Вентцель, Е. С.** Теория вероятностей и ее инженерные приложения : учебное пособие / Е. С. Вентцель, Л. А. Овчаров. - 4-е изд., стер. - М. : Высш. шк., 2007. - 491 с. - (Учебники для вузов. Математика). - ISBN 978-5-06-005714-0

2. **Самарин, Ю. П.** Высшая математика : учебное пособие / Ю. П. Самарин, Г. А. Сахабиева, В. А. Сабахиев. - М. : Машиностроение, 2006. - 432 с. - (В для вузов). - ISBN 5-217-03354-1.

3. **Шириков, В. Ф.** Теория вероятностей : учебное пособие / В. Ф. Шириков, С. М. Зарбалиев. - М. : КолосС, 2008. - 389 с. - (Учебники и учеб. пособия для студентов высш. учеб. заведений). - ISBN 978-5-9532-0621-1

4. **Ермакова, В.И**. Теория вероятностей и математическая статистика : учебное пособие / ред. В. И. Ермакова. - М. : Инфра-М, 2004. - 287 с. - (Высшее образование). - ISBN 5-16 001561-2

5. **Гмурман, В. Е.** Теория вероятностей и математическая статистика : учебное пособие / В. Е. Гмурман. - 12-е изд., перераб. - М. : Высш. образование, 2006. - 479 с. : ил. - (Основы наук).

6. **Белов, А. А.** Теория вероятностей и математическая статистика : учебник / А. А. Белов, Б. А. Баллод, Н. Н. Елизарова. - Ростов н/Д. : Феникс, 2008. - 318 с : ил. - (Высшее образование). - ISBN 978-5-222-14073-4

7. **Шипачев, В. С.** Курс высшей математики : учебник для вузов / Под ред. А. Н. Тихонова. - 3-е изд., испр. - М. : ОНИКС, 2007. - 600 с.

8. **Рау, В. Г.** Практический курс математики и общей теории статистики : учебное пособие / В. Г. Рау. - М. : Высш. шк., 2006. - 126 с. : ил. -ISBN 5-06-005529-9

9. **Хучраева, Т. С.** Высшая математика : учебное пособие / Т. С. Хучраева, А. А. Смоленинов, Т. В. Кириллова. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2008. - 156 с. - ISBN 978-5-7011-0559-9

10. **Юрьева, А. А.** Руководство к решению задач по теории вероятностей и математической статистике. : учебное пособие / А. А. Юрьева. - 2-е изд., испр. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2009. - 208 с.

СОДЕРЖАНИЕ