ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ
ФИНАНСОВО-ЭКОНОМИЧЕСКИЙ ИНСТИТУТ

Кафедра статистики и эконометрики

КОНСПЕКТ ЛЕКЦИЙ

«ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ И СТАТИСТИКИ, ИСПОЛЬЗУЕМЫЕ В ЭКОНОМЕТРИКЕ»

ПО КУРСУ «ЭКОНОМЕТРИКА»
ДЛЯ СТУДЕНТОВ III КУРСА
ДНЕВНОГО ОТДЕЛЕНИЯ ВСЕХ СПЕЦИАЛЬНОСТЕЙ

Казань 2004

Методическая разработка печатается по решению кафедры прогнозирования и статистики: **протокол № 3 от 04.11.2003 г.**

Автор: ст. преп. Кундакчян Резеда Мухтаровна.

Данная методическая разработка составлена в соответствии с утвержденной программой курса «Эконометрика» и предназначена для студентов 3 курса дневного отделения всех специальностей. Она является начальным, вводным этапом изучения и практики построения эконометрических моделей. По сути данный конспект лекций является преддверием к последующим конспектам лекций по курсу «Эконометрика».

Предполагается, что студентами, изучающими предмет «Эконометрика», прослушаны курсы «Математика», «Информатика», «Статистика», «Макроэкономика», «Макроэкономика», «Мировая экономика». Данный конспект лекций, с учетом этого, содержит основные положения теории вероятностей и статистики, составляющие основу статистико-математического инструментария эконометрики и без которых невозможно понимание всего курса эконометрики, тем самым прослеживается преемственность изученных ранее курсов и их практическое использование в эконометрических исследованиях. Студенты могут самостоятельно восстановить основные знания по теории вероятностей и статистики, а также изучить в данном разделе новые понятия, широко используемые в эконометрике.

Несомненно, экономическая составляющая эконометрики является первичной. Именно экономическая теория определяет постановку задачи и исходные предпосылки, а результат, формируемый на статистико-математическом языке, представляет интерес лишь в том случае, если удается его экономическая интерпретация.

Таким образом, студенты, владея эконометрическим инструментарием и используя компьютерные эконометрические пакеты, могут при помощи эконометрического моделирования социально-экономических процессов подтверждать и совершенствовать положения экономической теории, что в конечном итоге ведет к более глубокому пониманию сути происходящих процессов и к осуществлению продуманной и целенаправленной экономической политики.

В конце методической разработки указана рекомендуемая литература по данной теме.

Основные понятия теории вероятностей

Вероятностный эксперимент (испытание, опыт) – это эксперимент, результат которого не предсказуем заранее, т.к. он является случайным в силу сложного состояния различных естественных причин.

Любое действие в экономике по своей сути является вероятностным экспериментом.

Событие – это любой исход или совокупность исходов какого – либо вероятностного эксперимента. Теория вероятностей изучает не любые события, а лишь *случайные.*

Событие, которое может произойти или не произойти в результате некоторого эксперимента, называется **случайным.**

Все события можно классифицировать следующим образом:

Достоверное событие – это событие, которое обязательно произойдет в результате испытания.

Событие называется **невозможным,** если оно не происходит никогда в условиях данного эксперимента (испытания).

Достоверные и невозможные события не являются случайными.

Совместные события – несколько событий называют совместными, если в результате эксперимента наступление одного из них не исключает появления других.

Несовместные события – несколько событий называют несовместными в данном эксперименте, если появление одного из них исключает появление других. Два события называются **противоположными,** если одно из них происходит тогда и только тогда, когда не происходит другое.

Событие, которое нельзя разбить на более простые, называется **элементарным.**

Событие, представимое в виде совокупности (суммы) нескольких элементарных событий, называется **составным.**

Вероятность события – это количественная мера, которая вводится для сравнения событий по степени возможности их появления.

Классическое определение вероятности.

Вероятностью события А – *P(A)* – называется отношение числа m элементарных событий (исходов), благоприятствующих появлению события *A,* к числу n всех элементарных событий в условиях данного вероятностного эксперимента.

$$P(A)=\frac{m}{n} \qquad (1)$$

Из определения вытекают следующие свойства вероятности:

1.Вероятность случайного события есть положительное число, заключенное между 0 и 1:

$$0 \le P(A) \le 1. \qquad (2)$$

2. Вероятность достоверного события A равна 1: $\mathbf{P(A)=1}$ (3)

3. Если событие невозможное, то его вероятность равна

$$0: P(A)=0. \qquad (4)$$

4. Если события A и B несовместны, то

$$P(A+B)=P(A)+P(B). \qquad (5)$$

5. Если события А и В совместны, то вероятность их суммы равна сумме вероятностей этих событий без вероятности их совместного наступления:

$$P(A+B) = P(A) + P(B) - P(AB) \qquad (6)$$

6. Если A и $\overline{A}$ - противоположные события, то $P(\overline{A})=1-P(A).$ (7)

7. Сумма вероятностей событий $A_1, A_2, \ldots, A_n,$ образующих полную группу, равна 1:

$$P(A_1) + P(A_2) + \ldots + P(A_n) = 1. \qquad (8)$$

В экономических исследованиях значения m и n в формуле могут интерпретироваться по-другому. При **статистическом определении** вероятности события A под n понимается количество наблюдений результатов эксперимента, в которых событие A встречалось ровно m раз. В этом случае отношение $\frac{m}{n}$ называется **относительной частотой (частостью) события** A**.**

Зависимые и независимые события

События *A, B* называются **независимыми**, если вероятности каждого из них не зависит от того, произошло или нет другое событие. Вероятности независимых событий называются *безусловными*.

События *A, B* называются **зависимыми**, если вероятность каждого из них зависит от того, произошло или нет другое событие. Вероятность события В, вычисленная в предположении, что другое событие А уже осуществилось, называется *условной вероятностью*.

Если два события А и В – независимые, то справедливы равенства:

$$P(B) = P(B/A),\ P(A) = P(A/B) \text{ или } P(B/A) - P(B) = 0 \qquad (9)$$

Вероятность произведения двух зависимых событий А, В равна произведению вероятности одного из них на условную вероятность другого:

$$P(AB) = P(B) \cdot P(A/B) \text{ или } P(AB) = P(A) \cdot P(B/A) \qquad (10)$$

Вероятность события В при условии появления события А:

$$P(B/A) = \frac{P(AB)}{P(A)} \qquad (11)$$

Вероятность произведения двух **независимых** событий А, В равна произведению их вероятностей:

$$P(AB) = P(A) \cdot P(B) \qquad (12)$$

Если несколько событий попарно независимы, то отсюда еще не следует их независимость в совокупности.

События $A_1, A_2, \ldots, A_n$ $(n>2)$ называются независимыми в совокупности, если вероятность каждого из них не зависит от того, произошли или нет любые события из числа остальных.

Вероятность совместного появления нескольких событий, независимых в совокупности, равна произведению вероятностей этих событий:

$$P(A_1 \cdot A_2 \cdot A_3 \cdot \ldots \cdot A_n) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot \ldots \cdot P(A_n). \qquad (13)$$

Случайная величина.

Случайной величиной (СВ) называют величину, которая в результате наблюдения (испытания) принимает то или иное значение, заранее не известное и зависящее от случайных обстоятельств.

Различают:

- дискретные СВ;
- непрерывные СВ.

Дискретной называют такую СВ, которая принимает отдельные, изолированные (конечные или счетные) значения с определенными вероятностями.

Непрерывной называют такую СВ, которая может принимать любое значение из некоторого конечного или бесконечного числового промежутка (т.е. количество возможных значений непрерывной СВ бесконечно и несчетно).

Большинство СВ, рассматриваемых в экономике, имеет настолько большое число возможных значений, что их удобнее представлять в виде непрерывных СВ.

Дискретная случайная величина

Наиболее полным, исчерпывающим описанием **дискретной СВ** является ее закон распределения. **Законом распределения** случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Его можно задать таблично, аналитически (т.е. в виде формулы) и графически.

При табличном задании закона распределения дискретной СВ X первая строка таблицы содержит ее возможные значения $(x_1, x_2, \ldots x_к)$, а вторая – их вероятности $(p_1, p_2, \ldots p_к)$:

X	x_1	x_2	$\cdots$	$x_к$
P_i	p_1	p_2	$\cdots$	$p_к$

Обычно $x_1 < x_2 < \ldots < x_к$.

Такая таблица называется *рядом распределения* дискретной случайной величины.

Для любой дискретной случайной величины

$$\sum_{i=1}^{n} P(X = x_i) = \sum_{i=1}^{n} p_i = 1 \qquad (14)$$

Если по оси абсцисс откладывать значения случайной величины, по оси ординат – соответствующие им вероятности, то можно получить (соединением точек) ломаную, называемую многоугольником или *полигоном распределения вероятностей.*

Аналитически СВ задается либо функцией распределения, либо плотностью вероятностей.

Функцией распределения СВ *X* называют функцию $F(x)$, определяющую вероятность того, что СВ *X* принимает значение меньшее, чем x, т.е. $F(x) = P(X < x)$ (15)

Эту функцию также называют *функцией накопленной вероятности* или *кумулятивной функцией распределения*.

Из определения вытекают **свойства функции распределения:**

1. Функция распределения случайной величины есть неотрицательная функция, заключенная между нулем и единицей: $0 \le F(x) \le 1.$ (16)

2. $F(x)$ – неубывающая функция,т.е.

$$(x_1 < x_2) \Rightarrow F(x_1) \le F(x_2). \tag{17}$$

3. На минус бесконечности функция распределния равна нулю, на плюс бесконечности – равна единице: $\lim\limits_{x\to -\infty} F(x) = 0,\ \lim\limits_{x\to +\infty} F(x) = 1.$ (18)

4. Вероятность попадания случайной величины X в интервал $[a,b]$ равна приращению ее функции распределения на этом интервале:

$$P(a \le X \le b) = F(b) - F(a). \tag{19}$$

5. $P(X \ge x) = 1 - F(x).$ (20)

6. Если возможные значения СВ *X* принадлежат отрезку $[a, b]$, то

$$F(x) = \begin{cases} 0, \text{ если } x \le a \\ 1, \text{ если } x > b. \end{cases} \tag{21}$$

Непрерывная случайная величина

Для непрерывной СВ нельзя определить вероятность того, что она примет некоторое конкретное значение (точечную вероятность). Так как в любом интервале содержится бесконечное число значений, то вероятность выпадения одного из них асимптотически равна нулю. Таким образом в результате непрерывную СВ нельзя задать таблично. Однако для описания непрерывной СВ может быть использована функция распределения. При этом она является непрерывной неубывающей функцией, изменяющейся от 0 до 1.

Плотностью вероятности (плотностью распределения вероятностей) непрерывной СВ X называются производная ее функции распределения:

$$f(x) = \lim_{\Delta x\to 0} \frac{P(x \le X < x + \Delta x)}{\Delta x} = \lim_{\Delta x\to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = F'(x) \tag{22}$$

Плотность вероятности ***f(x)***, как и функция распределения ***F(x)***, является одной из форм закона распределения, но в отличие от функции распределения она существует только для *непрерывных* случайных величин.

График плотности вероятности называется *кривой распределения*.

Свойства плотности вероятности:

1. $f(x) \geq 0.$ (23)

2. $P(a \leq X \leq b) = \int\limits_a^b f(x)dx.$ (24)

3. Если $f(x)-$ плотность вероятности непрерывной СВ, то функция распределения $F(x) = \int\limits_{-\infty}^{x} f(t)dt.$ (25)

4. $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$ (условие нормировки).

Можно отметить, что для непрерывной СВ справедливы равенства:

$$P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b).$$

Вероятность попадания значений СВ в «хвосты» распределения, т.е. в интервалы (-∞; a) и (b; +∞), равна $1 - P(a \leq X \leq b).$ (27)

Таким образом, с помощью плотности вероятности $f(x)$ непрерывной СВ X можно определить вероятность ее попадания в заданный интервал: $P(a \leq x \leq b)$.

Во многих практических случаях информация о СВ, которую дают закон распределения, функция распределения или плотность вероятностей, является избыточной. Чаще всего используют числа, которые описывают СВ суммарно. Такие числа называют **числовыми характеристиками** СВ. Условно их подразделяют на:

- **характеристики положения** (математическое ожидание, мода, медиана, начальные моменты различных порядков);

- **характеристики рассеивания** (дисперсия, среднее квадратическое отклонение, центральные моменты различных порядков).

Важнейшими из них являются:

- математическое ожидание;
- дисперсия;
- среднее квадратическое отклонение.

Математическое ожидание характеризует среднее ожидаемое значение СВ, т.е. приближенно равно ее среднему значению.

Математическое ожидание определяется следующим образом:

Для **дискретной** СВ:

$$M(x) = \sum_{i=1}^{k} x_i p_i, \tag{28}$$

где k – число всех возможных значений СВ x.

Для **непрерывной** СВ:

$$M(x) = \int_{-\infty}^{+\infty} x f(x) dx \tag{29}$$

Таким образом, математическое ожидание рассчитывается в тех случаях, когда желают определить возможное среднее значение исследуемой величины.

Но для детального анализа поведения СВ знания лишь среднего значения явно недостаточно. Существуют отличные друг от друга случайные величины, имеющие одинаковые математические ожидания.

Следовательно, нужна числовая характеристика, которая оценивает разброс возможных значений СВ относительно ее среднего значения (математического ожидания). Такой характеристикой является дисперсия.

Дисперсией $D(X)$ (иногда она обозначается σ_x^2) СВ X называется математическое ожидание квадрата отклонения СВ от ее математического ожидания.

Она рассчитывается по формуле:

$$D(X) = M(X - M(X))^2 = M(X^2) - M^2(X) \qquad (30)$$

При этом для дискретной СВ:

$$D(X) = \sum_{i=1}^{k}(x_i - M(X))^2 p_i = \sum_{i=1}^{k} x_i^2 p_i - M^2(X) \qquad (31)$$

Для непрерывной СВ:

$$D(X) = \int_{-\infty}^{+\infty}(x - M(X))^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - M^2(X) \qquad (32)$$

Свойства дисперсии:

1. $D(C) = 0,$ где $C -$ константа; (33)
2. $D(CX) = C^2 D(X);$ (34)
3. $D(X \pm Y) = D(X) + D(Y),$ где X и $Y -$ независимые СВ; (35)
4. $D(aX + b) = a^2 D(X),$ где a и $b -$ константа. (36)

Средним квадратическим отклонением $\sigma(x)$ СВ X называется квадратичный корень из дисперсии $D(X)$:

$$\sigma = \sqrt{D(x)} \qquad (37)$$

Чтобы оценить разброс значений СВ в процентах относительно ее среднего значения, вводится коэффициент вариации $V(x)$, рассчитываемый по формуле:

$$V(x) = \frac{\sigma(x)}{|M(x)|} \cdot 100\% \qquad (38)$$

Меры разброса (дисперсия, среднее квадратическое отклонение, коэффициент вариации) кроме оценивания рассеивания значений СВ обычно применяются при изучении риска различных действий со случайным исходом, в частности при анализе риска инвестирования в ту или иную отрасль, при оценивании различных активов в портфеле и портфеля активов в целом в финансовом анализе и т.д.

Взаимосвязь случайных величин.

Многие экономические показатели определяются несколькими числами, являясь **многомерными** СВ. Упорядоченный набор $X=(X_1, X_2, ..., X_n)$ случайных величин называется **многомерной (n-мерной) случайной величиной** (или **системой случайных величин, n-мерным вектором**). Например, издержки предприятия включают в себя фиксированную и переменную составляющие; уровень жизни населения подразумевает использование большого числа показателей: ВНП на душу населения, распределение доходов, наличие товаров пи услуг, продолжительность жизни и т.д.

При проведении эконометрического анализа одно из главных мест занимает исследование взаимных связей СВ, при которых реализация одной из СВ влияет на вероятность определенной реализации других СВ.

Для описания n-мерной случайной величины используются следующие понятия:

1. Совместная вероятность

$$P_{X_1 X_2 \ldots X_n}(x_1, x_2, ..., x_n) = P(X_1 = x_1, X_2 = x_2, ... X_n = x_n) \qquad (39)$$

2. Совместная функция распределения

$$F(x_1, x_2, ..., x_n) = P(X_1 < x_1, X_2 < x_2, ... X_n < x_n) \qquad (40)$$

3. Совместная плотность вероятностей

$$f(x_1, x_2, ..., x_n) = \frac{\partial^n F(x_1, x_2, ..., x_n)}{\partial x_1 \partial x_2 ... \partial x_n} \qquad (41)$$

В двумерном случае для случайной величины (X,Y) двумерная вероятность, функция распределения и плотность вероятностей будут определяться:

$$P(X{=}x,\ Y{=}y);\ F(x,y)=P(X<x,\ Y<y);\ f(x,y)=\frac{\partial^2 F(x,y)}{\partial x \partial y} \tag{42}$$

Свойства функции распределения *F(x,y)* и плотности вероятности двумерной случайной величины *f(x,y)* аналогичны свойствам одномерной случайной величины соответственно.

Если необходимо вычислить значения вышеуказанных функций при фиксированных величинах одной или нескольких случайных величин, то эти функции суммируются (усредняются) по лишним переменным. В результате получаются **маргинальные (предельные) вероятности, функции распределения и плотности вероятности** или условия согласованности:

$$P_1(X_1 = x_1) = \sum_{x_2} \dots \sum_{x_n} P_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) \tag{43}$$

$$F_1(x_1) = F(x_1, \dots, \infty) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{+\infty} f_{X_1 X_2 \dots X_n}(t, x_2, \dots, x_n) dt dx_2 \dots dx_n \tag{44}$$

$$f_1(x_1) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{X_1, X_2, \dots X_n}(x_1, x_2, \dots x_n) dx_2 \dots dx_n \tag{45}$$

Для многомерных случайных величин кроме совместной вероятности (плотности вероятностей) определяются условные вероятности (условные плотности вероятностей).

Условным законом распределения одной из одномерных составляющих двумерной случайной величины *(X,Y)* называется ее закон распределения, вычисленный при условии, что другая составляющая приняла определенное значение (или попала в какой-то интервал).

Условная вероятность и условная плотность вероятностей случайной величины X для двумерной случайной величины *(X,Y)* при условии, что случайная величина Y примет значение y (Y=y) определяются по формулам:

$$P(x \mid Y = y) = \frac{P(x,y)}{P(y)};\ \ f(x \mid Y = y) = \frac{f(x,y)}{f(y)}$$

В соответствии с условной вероятностью (условной плотностью вероятности) двух случайных величин можно определить совместную вероятность (совместную плотность вероятности) этих случайных величин:

$$P(x,y)=P(X=x,Y=y)=P(x\,|\,y)P(y)=P(y\,|\,x)P(x) \quad (46)$$

$$f(x,y)=f(x\,|\,y)f(y)=f(y\,|\,x)f(x)) \quad (47)$$

Для **независимых** случайных величин X и Y выполняются следующие соотношения:

$$P(x,y)=P(x)\cdot P(y);\quad f(x,y)=f(x)\cdot f(y);\quad F(x,y)=F(x)\cdot F(y) \quad (48)$$

Построение закона распределения многомерной случайной величины является трудоемким процессом. Поэтому обычно для анализа степени взаимной связи СВ используют следующие числовые характеристики:

- смешанные моменты распределения;
- ковариацию;
- коэффициент корреляции.

Смешанным моментом порядка k,l называется величина:

$$m_{k,l}=M(x^k y^l)=\begin{cases}\sum\limits_x\sum\limits_y x^k y^l P(x,y)-\text{для дискретных СВ};\\ \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} x^k y^l f(x,y)dxdy-\text{для непрерывных СВ}.\end{cases} \quad (49)$$

Например, $M(X)=m_{1,0,}\quad M(Y)=m_{0,1.}$

Центральным моментом порядка k,l называется величина:

$$\mu_{k,l}=M((X-M(X))^k(Y-M(Y))^l) \quad (50)$$

Например, $D(X)=\mu_{2,0};\ D(Y)=\mu_{0,2.}$

Для описания связи между СВ X и Y применяют центральный момент порядка 1,1 $(\mu_{1,1})$, который называется **ковариацией** СВ *X* и *Y* (или корреляционным моментом):

$$\sigma_{x,y} = cov(X,Y) = M((X - M(X))(Y - M(Y)) = \\ = M(XY) - M(X)M(Y) \quad (51)$$

Ковариация является **абсолютной** (зависящей от размерностей) мерой взаимосвязи (co-vary – «совместное изменение») переменных.

$$\sigma_{xy} = \begin{cases} \sum_{x}\sum_{y} x_i y_i P(x_i, y_i) - M(X)M(Y) - \text{для дискретных СВ}; \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy - M(X)M(Y) - \text{для непрерывных СВ}. \end{cases} \quad (52)$$

Свойства ковариации:

1. $\sigma_{xy} = \sigma_{yx}$.

2. $\sigma_{xx} = D(X) = \sigma_x^2$

3. Если X и Y независимые СВ, то $\sigma_{xy} = 0$.

4. $|\sigma_{xy}| \le \sigma_x \sigma_y$.

5. $\text{cov}(a + bX, c + dY) = bd\,\text{cov}(X,Y),\ \text{где } a,b,c,d - \\ - \text{константы}.$

Однако существенным **недостатком** ковариации является ее зависимость от размерностей рассматриваемых СВ. Поэтому при различных единицах измерения СВ одна и та же зависимость может выражаться различными значениями ковариаций. Кроме того, ковариация не позволяет определить силы (строгости) зависимости между рассматриваемыми СВ.

Для устранения данных недостатков вводится относительная мера взаимосвязи (безразмерная величина) – коэффициент корреляции.

Коэффициентом корреляции СВ X и Y называют величину

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sqrt{D(X)D(Y)}} \quad (53)$$

Зависимость между СВ X и Y, характеризуемая коэффициентом корреляции, называется **корреляцией.**

СВ X и Y называется **некоррелированными,** если $\rho_{xy} = 0,$ что равносильно равенству $\sigma_{xy} = 0.$ Если же $\rho_{xy} \neq 0,$ то СВ X и Y называют **коррелированными.**

Свойства коэффициента корреляции:

1. $\rho_{xx} = 1.$

2. $\rho_{xy} = \rho_{yx}.$

3. $-1 \leq \rho_{xy} \leq 1.$

4. Если СВ X и Y независимы, то $\rho_{xy} = 0.$

5. $|\rho_{xy}| = 1,$ тогда, когда $Y = a + bx$ (т.е. между СВ X и Y существует линейная функциональная зависимость).

Заметим, что если X и Y независимые СВ, то X и Y – некоррелированные СВ. Обратное утверждение неверно.

Ранее мы привели основные свойства и формулы расчета дисперсии, в частности дисперсии суммы двух независимых СВ (см. формулу 35).

В случае, когда СВ не являются независимыми, а коррелируют друга, формулы расчета дисперсии их суммы, либо разности имеют вид:

$$D(X \pm Y) = D(X) + D(Y) \pm 2\operatorname{cov}(X, Y) \qquad (54)$$

$$D(X \pm Y) = D(X) + D(Y) \pm 2\rho_{xy}\sigma_x\sigma_y \qquad (55)$$

При независимости случайных величин последние слагаемые в этих формулах обращаются в ноль.

Выборочное наблюдение.

Генеральной совокупностью называется множество всех возможных значений или реализаций исследуемой СВ X при данном реальном комплексе условий.

Выборкой (выборочной совокупностью) называют часть генеральной совокупности, отобранную для изучения.

Число элементов рассматриваемой совокупности называется ее объемом.

Изучение всей генеральной совокупности во многих случаях либо невозможно, либо нецелесообразно в силу больших материальных затрат.

Задачей статистического описания выборки является получение такого ее представления, которое позволит наглядно выявить вероятностные характеристики. Для этого применяются различные формы упорядочения данных в выборке – по возрастанию, по совпадающим значениям, по интервалам и т.п.

При анализе какого-то конкретного показателя X в фиксированный момент времени (либо без учета фактора времени) наблюдаемые значения $x_{1,} x_{2\ldots} x_n$ обычно упорядочивают по неубыванию: $x_1 \leq x_2 \leq \ldots x_{n.}$

Разность между максимальными и минимальными значениями СВ X называется **размахом выборки.**

Пусть количество различных значений в выборке равно $k(k \leq n)$. Значение $x_{i,} i = 1,2 \ldots k$, называется **вариантами.**

Если значение x_i встретилось в выборке n_i раз, то число n_i называется частотой значения x_i, а величина $\varpi_i = \frac{n_i}{n}$ – **относительной частотой значения** $x_{i.}$. Тогда наблюдаемые значения можно сгруппировать в статистический ряд:

x	$x_1 \quad x_2 \quad \ldots \quad x_k$	$\sum_{i=1}^{k} n_i = n;$ $\sum_{i=1}^{k} \frac{n_i}{n} = 1$
n_i	$n_1 \quad n_2 \quad \ldots \quad n_k$	
$\varpi_i = \frac{n_i}{n}$	$\frac{n_1}{n} \quad \frac{n_2}{n} \quad \ldots \quad \frac{n_k}{n}$	

По статистическому ряду можно построить эмпирическую функцию распределения $F^*(x)$:

$$F^*(x) = \frac{n_x}{n} \qquad (1)$$

где n_x – число значений случайной величины X, меньших, чем x;

n – объем выборки.

По определению $F^*(x)$ обладает следующими свойствами:

1. $0 \le F^*(x) \le 1.$

2. Для любых $x_1 < x_2 \quad F^*(x_1) \le F^*(x_2).$

3. $F^*(x) = 0$ при $x \le x_1$; $F^*(x) = 1$ при $x > x_k$.

Эмпирическая функция распределения $F^*(x)$ является оценкой функции распределения $F(x) = p(X < x)$, которая в этом случае называется **теоретической функцией распределения.**

Вычисление выборочных характеристик.

Для любой СВ X кроме определения ее функции распределения желательно указать числовые характеристики, важнейшими из которых является:

- математическое ожидание;
- дисперсия;

- среднее квадратическое отклонение.

Пусть объем генеральной совокупности равен N. Тогда математическим ожиданием СВ X является **генеральное среднее:**

$$\bar{x}_г = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (2)$$

Дисперсией СВ X является **генеральная дисперсия:**

$$D_г = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x}_г)^2 \qquad (3)$$

Корень квадратный из генеральной дисперсии называется **генеральным средним квадратическим отклонением:**

$$\sigma_г = \sqrt{D_г} \qquad (4)$$

Таким образом, для нахождения генеральных числовых характеристики необходим анализ всей генеральной совокупности.

В силу того, что в реальности чаще всего работают с выборками, приходится находить оценки указанных выше генеральных характеристик – выборочные числовые характеристики:

- выборочное среднее;
- выборочную дисперсию;
- выборочное среднее квадратическое отклонение;
- выборочный коэффициент вариации.

Выборочное среднее – это среднее арифметическое наблюдаемых значений выборки.

$$\bar{x}_B = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (5)$$

При задании выборки в виде статистического ряда рассчитывается по следующей формуле:

$$\bar{x}_B = \frac{1}{n}\sum_{i=1}^{n} n_i x_i \qquad (6)$$

Оценкой генеральной дисперсии является выборочная дисперсия:

$$D_B = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_B)^2 \qquad (7)$$

(для дальнейшем $\bar{x}_B$ будем обозначать через $\bar{x}$).

Часто для вычисления D_B применяется следующая формула:

$$D_B = \frac{1}{n}\sum_{i=1}^{k}(x_i^2 - 2x_i \cdot \bar{x}_B + (\bar{x}_B)^2) = \overline{x^2} - \bar{x}^2 \qquad (8)$$

При задании выборки в виде статистического ряда имеем:

$$D_B = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2 \qquad (9)$$

Корень квадратный из выборочной дисперсии называется **выборочным средним квадратическим (стандартным) отклонением.**

$$\sigma_B = \sqrt{D_B} = \sqrt{\frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2} = \sqrt{\overline{x^2} - \bar{x}^2} \qquad (10)$$

По аналогической схеме определяются статистические оценки других числовых характеристик СВ.

Выборочный коэффициент вариации V определяется отношением выборочного среднего квадратического отклонения к выборочной средней, выраженным в %:

$$V = \frac{\sigma_в}{\bar{x}} \cdot 100\% \qquad (11)$$

Коэффициент вариации – безмерная величина, удобная для сравнения величин рассеивания двух выборок, имеющих различные размерности.

Как нами уже отмечалось ранее, наиболее употребляемыми характеристиками связи двух СВ является меры их линейной связи – ковариация и коэффициент корреляции. Их оценками являются:

- выборочная ковариация S_{xy};

- выборочный коэффициент корреляции r_{xy}.

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \overline{xy} - \bar{x}\cdot\bar{y} \qquad (12)$$

$$r_{xy} = \frac{S_{xy}}{\sigma_B(X)\cdot\sigma_B(Y)} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\cdot\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2}\cdot\sqrt{\overline{y^2} - \bar{y}^2}}$$

Здесь $$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n}x_i y_i \qquad (13)$$

Законы распределений случайных величин.

Большинство СВ подчиняется определенному закону распределения, зная который можно предвидеть вероятности попадания исследуемой СВ в определенные интервалы. Это очень важно при анализе экономических показателей, т.к. в этом случае появляется возможность осуществлять продуманную политику с учетом возможности возникновения той или иной ситуации.

Законов распределений достаточно много. К числу наиболее активно использующихся в эконометрическом анализе относятся:

- нормальное распределение (распределение Гаусса);
- распределение χ^2;
- распределение Стьюдента;
- распределение Фишера.

Для удобства использования данных законов были разработаны таблицы критических точек, которые позволяют быстро и эффективно оценивать соответствующие вероятности.

Нормальное распределение

Нормальное распределение (распределение Гаусса) является предельным случаем почти всех реальных распределений вероятности. Поэтому оно используется в очень большом числе реальных приложений теории вероятностей.

СВ X имеет нормальное распределение, если ее плотность вероятности имеет вид:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{\frac{-(x-m)^2}{2\sigma^2}} \qquad (14)$$

Это равносильно тому, что

$$F(x) = \frac{1}{\sqrt{2\pi\sigma}} \cdot \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt \qquad (15)$$

СВ, имеющая нормальное распределение, называется нормально распределенной или нормальной. Графики плотности вероятности и функции распределения нормальной СВ изображены на рис.1 и 2.

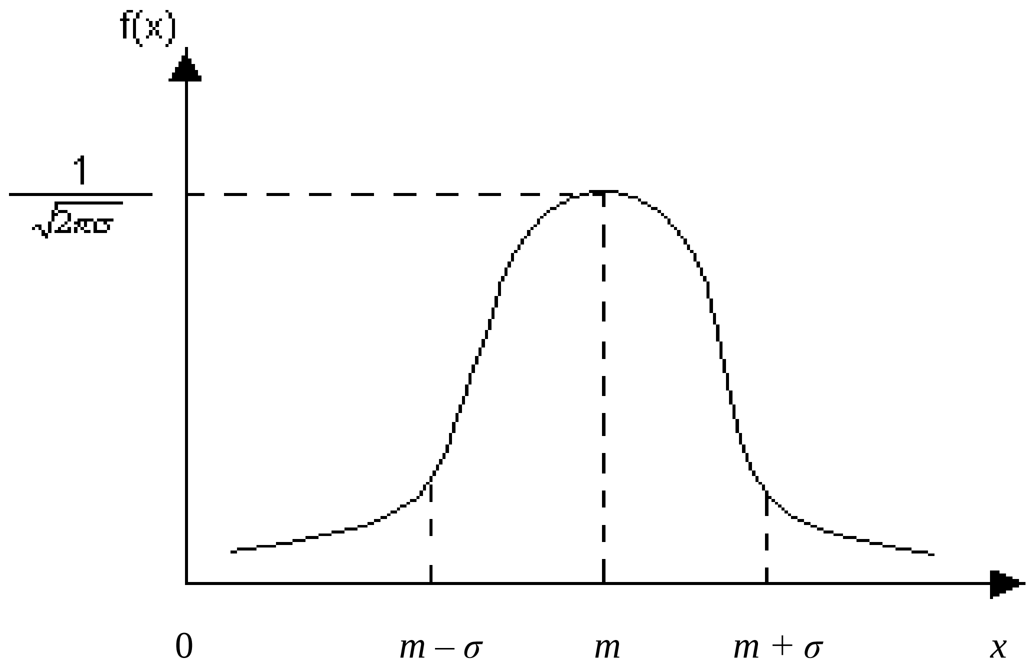


Рис.1. График плотности вероятности нормального распределения СВ X.

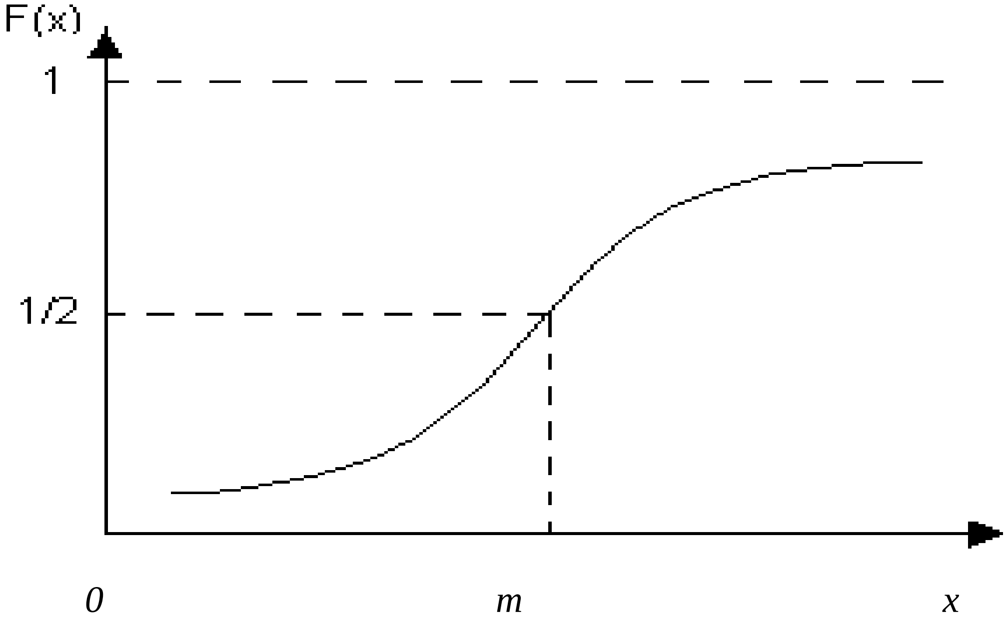


Рис.2. Функция распределения нормальной СВ.

Как видно из формул (1) и (2), нормальное распределение зависит от параметров *m и σ* и полностью определяется ими. При этом $m = M(X)$, $\sigma = \sigma(X)$, т.е. $D(X) = \sigma^2$, $\pi = 3{,}14159\ldots$, $e = 2{,}71828\ldots$.

Если СВ *X* имеет нормальное распределение с параметрами $M(X) = m$ и $\sigma(X) = \sigma$, то символически это можно записать так:

$X \sim N(m, \sigma)$ *или* $X \sim N(m, \sigma^2)$.

Очень важным частным случаем нормального распределения является ситуация, когда $m = 0$ и $\sigma = 1$. В этом случае говорят о **стандартизированном (стандартном) нормальном распределении.**

Стандартизированную нормальную СВ обозначают через U $(U \sim N(0,1))$, учитывая при этом, что

$$f(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}}; \quad F(u) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{u} e^{-\frac{t^2}{2}} dt \quad (16)$$

Для практических расчетов специально разработаны таблицы функций $f(u)$, $F(u)$ стандартизированного нормального распределения, но чаще используется так называемая <u>таблица значений Лапласа $\Phi(u)$.</u> Функция Лапласа имеет вид:

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^u e^{-\frac{t^2}{2}} dt = F(u) - 0{,}5 \qquad (17)$$

Эту таблицу можно использовать для любой нормальной СВ *X* ($X \sim N(m, \sigma)$) при расчете соответствующих вероятностей:

$$P(a \le x \le b) = F\left(\frac{b-m}{\sigma}\right) - F\left(\frac{a-m}{\sigma}\right) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right) \qquad (18)$$

Заметим, что если $X \sim N(m, \sigma)$, то $U = \dfrac{X-m}{\sigma} \sim N(0,1)$.

Как видно из предыдущих рисунков, нормально распределенная СВ *X* ведет себя достаточно предсказуемо. График ее плотности вероятности (рис.1) симметричен относительно прямой $X = m$. Площадь фигуры под графиком плотности вероятности должна оставаться равной единице при любых значениях *m и σ*. Следовательно, чем меньше значение *σ*, тем более крутым является график.

Кроме того, справедливы следующие соотношения:

$P(\; X - M(X) \; < \sigma) = 0{,}68;$ $\qquad P(\; X - M(X) \; < 2\sigma) = 0{,}95;$

$P(\; X - M(X) \; < 3\sigma) = 0{,}9973.$

Другими словами, значения нормально распределенной СВ *X* ($X \sim N(m, \sigma)$) на 99,73 % сосредоточены в области $[\, m - 3\sigma,\ m + 3\sigma \,]$.

Важным является тот факт, что **линейная комбинация произвольного количества нормальных СВ имеет нормальное распределение**. При этом, если $X \sim N(m_x, \sigma_x)$ *и* $Y \sim N(m_y, \sigma_y)$ – независимые СВ, то

$$Z = aX + bY \sim N(m_z, \sigma), \qquad (19)$$

где $m_z = a\, m_x + b\, m_y$; $\sigma_z^{\,2} = a^2 \sigma_x^{\,2} + b^2 \sigma_y^{\,2}$.

Многие экономические показатели имеют нормальный или близкий к нормальному закон распределения. Например, доход населения, прибыль фирм в отрасли, объем потребления и т.д. имеют близкое к нормальному распределение.

Нормальное распределение используется при проверке различных гипотез в статистике (о величине математического ожидания при известной дисперсии, о равенстве математических ожиданий и т.д.).

Часто при моделировании экономических процессов приходится рассматривать СВ, которые представляют собой алгебраическую комбинацию нескольких СВ. Существенную роль в этом играет ряд специально разработанных теоретических законов распределений. К ним относятся χ^2 – распределение, распределения Стьюдента и Фишера.

Распределение χ^2 (хи – квадрат)

Пусть X_i, $i = 1, 2, \ldots, n$ – независимые нормально распределенные СВ с математическими ожиданиями m_i и средними квадратическими отклонениями σ_i соответственно, т.е. $X_i \sim N(m_i, \sigma_i)$.

Тогда СВ $U_i = \frac{(x_i - m_i)}{\sigma_i}$, $i = 1, 2, \ldots, n$, являются независимыми СВ, имеющими стандартизированное нормальное распределение, $U_i \sim N(0,1)$.

СВ χ^2 имеет хи – квадрат распределение с **n** степенями свободы ($\chi^2 \sim \chi_n^2$),

если $\chi^2 = \sum_{i=1}^{n} U_i^2 = U_1^2 + U_2^2 + \ldots + U_n^2$ (20).

Отметим, что **число степеней свободы** (это число обозначается ***ν***) исследуемой СВ определяется числом СВ, ее составляющих, уменьшенным на число линейных связей между ними.

Например, число степеней свободы СВ, являющейся композицией **n** случайных величин, которые в свою очередь связаны **m** линейными уравнениями, определяется числом $\nu = m - n$. Таким образом, $U^2 \sim \chi_1^2$.

Из определения (20) следует, что распределение χ^2 определяется одним параметром – числом степеней свободы ***ν***.

График плотности вероятности СВ, имеющий χ^2 – распределение, лежит только в первой четверти декартовой системы координат и имеет асимметричный вид с вытянутым правым «хвостом» (рис.3). Но с увеличением числа степеней свободы распределение χ^2 постепенно приближается к нормальному:

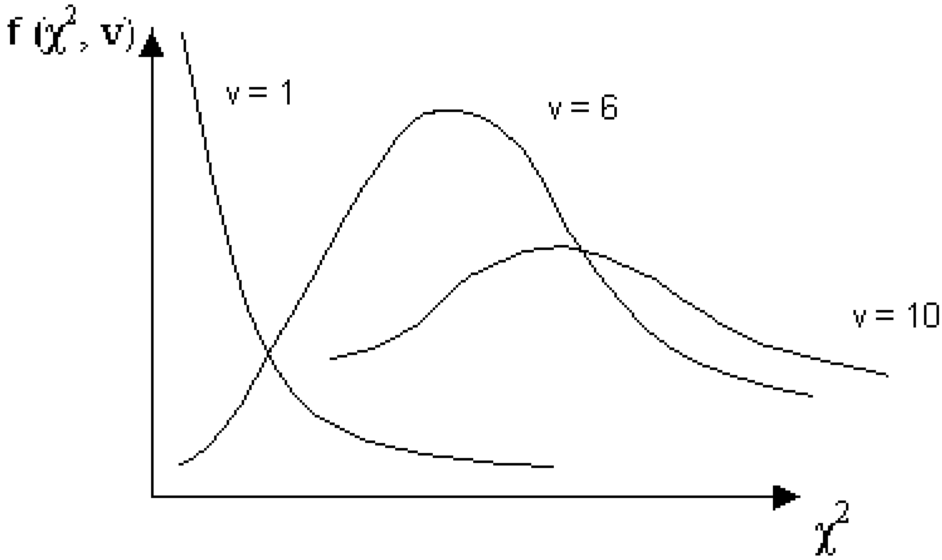


Рис. 3. График плотности вероятности СВ X, имеюший χ^2 – распределение.

$M(\chi^2) = v = n - m,$

$D(\chi^2) = 2v = 2(n - m).$

Если X и Y – две независимые χ^2 – распределенные СВ с числами степеней свободы **n** и **k** соответственно $(X \sim \chi_n^2,\ Y \sim \chi_k^2)$, то их сумма $(X + Y)$ также является χ^2 – распределенной СВ с числом степеней свободы $v = n + k$.

Распределение χ^2 применяется для нахождения интервальных оценок и проверки статистических гипотез. При этом используется таблица критических точек χ^2 – распределения.

Распределение Стьюдента

Пусть СВ $U \sim N(0,1)$, СВ V – независимая от U величина, распределенная по закону χ^2 *с* ***n*** степенями свободы. Тогда величина

$$T = \frac{U}{\sqrt{\frac{V}{n}}} \qquad (21)$$

имеет *распределение Стьюдента (t – распределение) с* **n** *степенями свободы* ($T \sim T_n$).

Из формулы (21) видно, что распределение Стьюдента определяется только одним параметром **n** – числом степеней свободы. График функции плотности вероятности СВ, имеющей распределение Стьюдента, является симметричной кривой (линия симметрии – ось ординат) (рис.4)

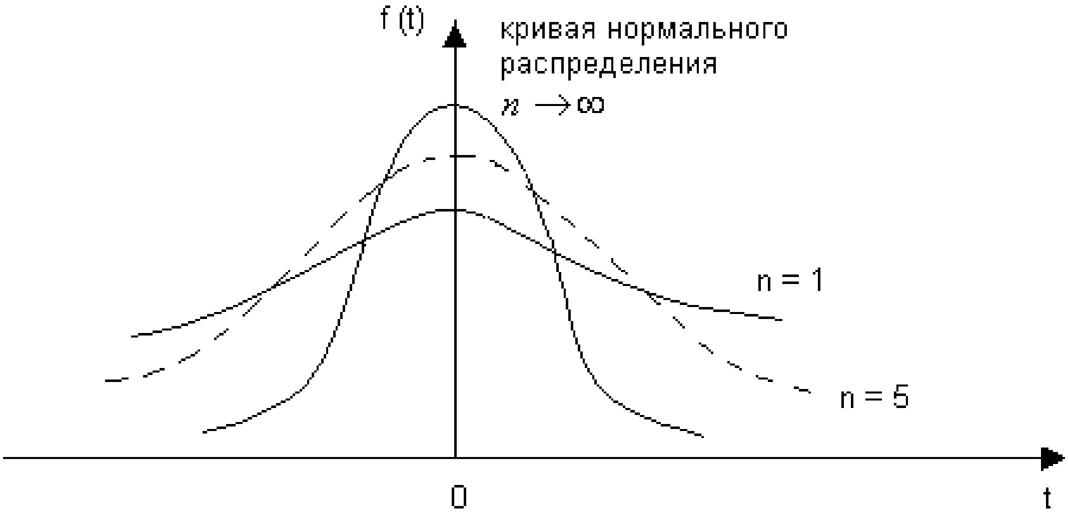


Рис. 4. График функции плотности вероятности СВ X, имеющий распределение Стьюдента.

$M(T) = 0,$

$D(T) = n / (n - 2).$

При этом с увеличением числа степеней свободы распределение Стьюдента приближается к стандартизированному нормальному, причем при $n > 30$ распределение Стьюдента практически можно заменить нормальным распределением.

Распределение Стьюдента применяется для нахождения интервальных оценок, а также при проверке статистических гипотез. При этом активно используется таблица критических точек распределения Стьюдента.

Распределение Фишера

Пусть V и W – независимые СВ, распределенные по закону χ^2 со степенями свободы $v_1 = m$ и $v_2 = n$ соответственно. Тогда величина

$$F = \frac{V/m}{W/n} \qquad (22)$$

имеет распределение Фишера со степенями свободы $v_1 = m$ *и* $v_2 = n$ $(F \sim F_{m,n})$. Таким образом, *распределение Фишера F* определяется двумя параметрами – числами степеней свободы **m** и **n**.

При больших m и n это распределение приближается к нормальному (рис.5). Нетрудно заметить, что $T_n^{\,2} = F_{1,n}$, где T_n – СВ, имеющая распределение Стьюдента с числом степеней свободы $v = n$, $F_{1,n}$ – СВ, имеющая распределение Фишера с числами степеней свободы $v_1 = 1$ *и* $v_2 = n$.

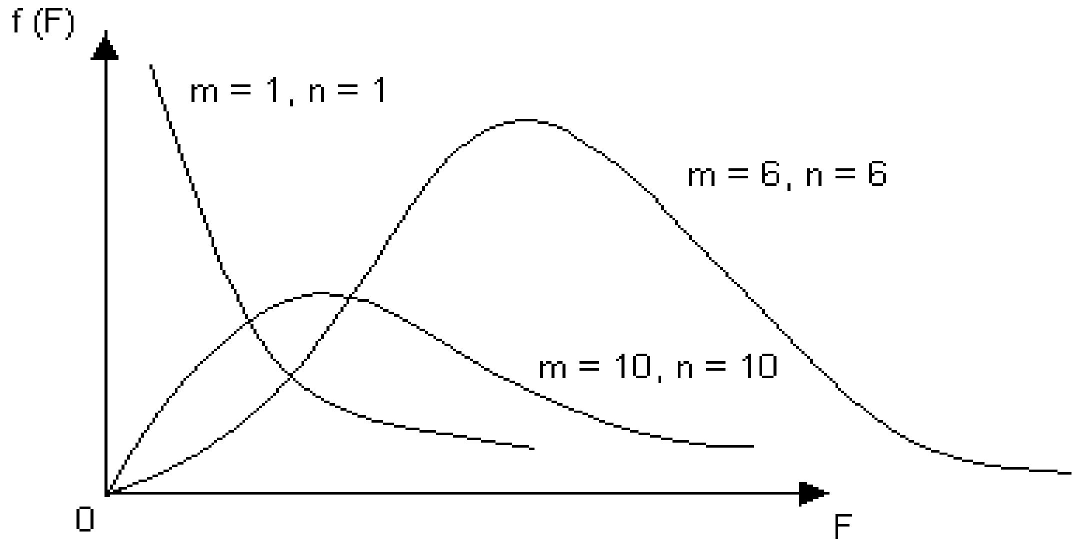


Рис.5. График функции плотности вероятности СВ X, имеющий распределение Фишера.

$$\left.\begin{aligned} M(F) &= \frac{n}{n-2}\cdot(n-2), \\ D(F) &= \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, (n>4) \end{aligned}\right\}$$

Распределение Фишера используется при проверке статистических гипотез в дисперсионном и регрессионном анализах. При этом активно используется таблица критических точек распределения Фишера.

Статистические выводы: оценки и проверка гипотез.

Статистические выводы – это заключения о генеральной совокупности (т.е. законе распределения исследуемой СВ и его параметрах либо о наличии и силе связи между исследуемыми переменными) на основе выборки, случайно отобранной из генеральной совокупности. Или обобщение результатов, полученных по выборке, на генеральную совокупность и есть суть статистических выводов.

При исследовании различных параметров генеральной совокупности на основе выборки возможно лишь получение оценок этих параметров. Эти оценки строятся на основе ограниченного набора данных, что влечет за собой вероятность погрешности. Значения оценок могут изменяться от выборки к выборке.

Процесс нахождения оценок по определенному правилу (формуле) называется **оцениванием.**

Цель любого оценивания – получение наиболее точного значения оцениваемой характеристики.

Можно выделить 2 типа оценивания:

1. Оценивание вида распределения.
2. Оценивание параметров распределения.

Процедура оценивания всегда однотипна.

На основе выборки с помощью соответствующей формулы рассчитывается оценка исследуемой характеристики. В качестве оценок параметров распределения генеральной совокупности берутся их **выборочные оценки.** При этом различают 2 вида оценок:

- точечные
- интервальные

После определения оценок обычно встает вопрос об их качестве и статистической значимости.

Точечные оценки и их свойства.

Пусть оценивается некоторый параметр θ наблюдаемой СВ X генеральной совокупности. Пусть из генеральной совокупности извлечена выборка объема $n: x_1, x_2, \ldots x_n,$ по которой может быть найдена оценка θ^* параметра θ.

Например, для нормального закона распределения с плотностью вероятности

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$ параметрами являются математическое ожидание m и среднее квадратическое отклонение σ.

Точечной оценкой θ^* параметра θ называется числовое значение этого параметра, полученное по выборке объема n.

Например, оценками m и $\sigma(x)$ могут быть:

$$m^* = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{и} \quad \sigma^* = \sigma_B = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$ соответственно.

Нетрудно заметить, что оценка θ^* являются функцией от выборки, т.е. $\theta^* = \theta^*(x_1, x_2, \ldots x_n)$.

Так как выборка носит случайный характер, то оценка θ^* является СВ, принимающей различные значения для различных выборок. Любую оценку $\theta^* = \theta^*(x_{1,} x_{2\ldots} x_n)$ называют **статистикой** или **статистической оценкой параметра** θ.

Точностью оценки называют такое число ε, что $\left|\theta - \theta^*\right| \le \varepsilon$. Естественно стремление получить по возможности наиболее точную оценку при данном объеме выборки.

Приведем свойства, выполнимость которых желательна для того, чтобы оценка была признана удовлетворительной.

В силу случайности точечной оценки θ^* она может рассматриваться как СВ со своими числовыми характеристиками – математическим ожиданием $M(\theta^*)$ и дисперсией $D(\theta^*)$. Чем ближе $M(\theta^*)$ к истинному значению θ и чем меньше $D(\theta^*)$, тем лучше будет оценка (при прочих равных условиях). Т.о., качество оценок характеризуется следующими основными свойствами:

- несмещенность;

- эффективность;

- состоятельность.

Оценка θ^* называется **несмещенной оценкой** параметра θ, если ее математическое ожидание равно оцениваемому параметру: $M(\theta^*) = \theta$. В противном случае – оценка называется **смещенной.**

Разность $M(\theta^*) - \theta$ - называется **смещением** или **систематической ошибкой оценивания.** Для несмещенных оценок систематическая ошибка равна нулю. Если $M(\theta^*) > \theta$, то θ^* завышает среднее значение θ.

Свойство несмещенности оценки является важнейшим, но не единственным. Существует несколько возможных несмещенных оценок одного и того же параметра. Выбор будет сделан в пользу той из них, вероятность совпадения которой с истинным значением оцениваемого параметра выше. Оценка должна иметь такую плотность вероятности, которая наиболее «сжата» вокруг истинного значения оцениваемого параметра. Нетрудно заметить, что в этом случае она будет иметь **наименьшую** среди других оценок дисперсию.

Оценка θ^* называется **эффективной оценкой** параметра θ, если ее дисперсия $D(\theta^*)$ меньше дисперсии любой другой альтернативной несмещенной оценки при фиксированном объеме выборки n, т.е. $D(\theta^*) = D_{\min}$.

Оценка называется **асимптотически эффективной,** если с увеличением объема выборки ее дисперсия стремится к нулю, т.е. $D(\theta_n^*) \to 0$ при $n \to \infty$ (индекс n в оценке θ_n^* применяется для подчеркивания объема выборки).

Оценка θ_n^* называется **состоятельной оценкой** параметра θ, если θ_n^* сходится по вероятности к оцениваемому параметру θ при $n \to \infty$. Другими словами, состоятельной называется такая оценка, которая дает истинное значение при достаточно большом объеме выборки вне зависимости от значений входящих в нее конкретных наблюдений.

Справедливо следующее утверждение: если $M(\theta_n^*) \to \theta$ и $D(\theta_n^*) \to 0$ при $n \to \infty$, то θ_n^* – состоятельная оценка параметра θ.

Оценки, являющиеся линейными функциями от выборочных наблюдений, называется **линейными.**

Наиболее употребляемыми методами нахождения точечных оценок является метод моментов, метод максимального правдоподобия, метод наименьших квадратов.

Свойства выборочных оценок.

На начальном этапе в качестве оценки той или иной числовой характеристики (математического ожидания, дисперсии и т.п.) берется выборочная числовая характеристика. Затем, исследуя эту оценку, ее уточняют таким образом, чтобы она удовлетворяла описанным выше свойствам.

Доказано, что выборочное среднее $\bar{x}_в = \frac{1}{n}\sum_{i=1}^{n} x_i$ является несмещенной и состоятельной оценкой математического ожидания *M(X)* генеральной совокупности.

Выборочная дисперсия $D_в = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2$ является смещенной, но состоятельной оценкой дисперсии $D(X) = \sigma^2$ СВ X генеральной совокупности, т.к. доказано, что $D_в = \sigma^2 \frac{n-1}{n}$

Иными словами, выборочная дисперсия оценивает генеральную дисперсию с недостатком**.**

Хотя при $n \to \infty$ $\frac{n-1}{n} \to 1$, и оценка $D_в$ является асимптотически несмещенной, в качестве оценки дисперсии $D(X)$ удобнее брать **исправленную дисперсию:**

$$S^2 = \frac{n}{n-1} D_в = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i (x_i - \bar{x}_в)^2 \qquad (24)$$

Исправленная дисперсия S^2 является несмещенной и состоятельной оценкой дисперсии $D(X)$ СВ X.

Аналогично вводится **исправленное среднее квадратическое отклонение** или так называемый **эмпирический стандарт** S**:**

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k} n_i (x_i - \bar{x}_в)^2} \qquad (25)$$

Отметим, что при n>30 различие между $D_в$ и S^2 $(\sigma_в \ и \ S)$ практически незначимо. Поэтому при большом объеме выборки и ту, и другую оценки можно считать несмещенными.

Относительная частота $\frac{n_i}{n}$ является несмещенной и состоятельной оценкой вероятности $P(X = x_i)$. Аналогично эмпирическая функция распределения $F^*(x) = \frac{n_x}{n}$ (накопленная относительная частота) является несмещенной и состоятельной оценкой (теоретической) функции распределения $F(x) = P(X < x)$.

Интервальные оценки

После получения точечной оценки Θ^*, желательно иметь данные о надежности такой оценки. Особенно важно иметь сведения о точности оценок для небольших выборок (поскольку с возрастанием объема n выборки несмещенность и состоятельность основных оценок гарантируется утверждениями математической статистики).

Точечная оценка может быть дополнена **интервальной оценкой** – интервалом (θ_1, θ_2), внутри которого с наперед заданной вероятностью γ находится точное значение оцениваемого параметра Θ. Определение такого интервала называют **интервальным оцениванием**, а сам интервал – **доверительным интервалом.** При этом γ называют **доверительной вероятностью** – или **надежностью**, с которой оцениваемый параметр Θ попадает в интервал (θ_1, θ_2).

Для определения доверительного интервала заранее выбирают число $\alpha = 1 - \gamma, \quad 0 < \alpha < 1$, называемое **уровнем значимости,** и находят два числа θ_1 и θ_2, зависящих от точечной оценки Θ^*, такие, что

$$P(\theta_1 < \theta < \theta_2) = 1 - \alpha = \gamma \tag{26}$$

В этом случае говорят, что интервал (θ_1,θ_2) накрывает неизвестный параметр θ с вероятностью $(1-\alpha)$. Границы интервала θ_1 и θ_2 называются доверительными, и они обычно находятся из условия $P(\theta>\theta_2)=\alpha/2$

Длина доверительного интервала, характеризующая точность интервальной оценки, зависит от объема выборки n и надежности γ (уровня значимости $\alpha=1-\gamma$). При увеличении величины n длина доверительного интервала уменьшается, а с приближением надежности γ к единице – увеличивается. Выбор α (или $\gamma=1-\alpha$) определяется конкретными условиями. Обычно используется $\alpha=0{,}1;0{,}05;0{,}01$, что соответствует 90, 95, 99%-м доверительным интервалам.

Общая схема построения доверительного интервала:

1. Из генеральной совокупности с известным распределением $f(x,\theta)$ CB X извлекается выборка объема n, по которой находится точечная оценка θ^* параметра θ.

2. Строится CB $Y(\theta)$, связанная с параметром θ и имеющая известную плотность вероятности $f(y,\theta)$.

3. Задается уровень значимости α.

4. Используя плотность вероятности CB Y, определяют два числа l_1 и l_2 такие, что $P(l_1<Y(\theta)<l_2)=\int\limits_{l_1}^{l_2} f(y,\theta)\,dy=1-\alpha$ (27)

5. Выбираются значения l_1 и l_2 из условий

$$P(Y(\theta)<l_1)=\alpha/2; \qquad P(Y(\theta)>l_2)=\alpha/2.$$

Неравенство $l_1 < Y(\theta) < l_2$ преобразуется в равносильное $\theta^* - \delta < \theta < \theta^* + \delta$ такое, что $P(\theta^* - \delta < \theta < \theta^* + \delta) = 1 - \alpha$ (28)

Полученный интервал $(\theta^* - \delta, \theta^* + \delta)$, накрывающий неизвестный параметр θ с вероятностью $1 - \alpha$, и является интервальной оценкой параметра θ.

Интервальная оценка также носит случайный характер, т.к. она напрямую связана с результатами выборки. Она позволяет сделать следующий вывод.

Если построен доверительный интервал, который с надежностью $\gamma = 1 - \alpha$ накрывает неизвестный параметр, и его границы рассчитываются по k выборкам одинакового объема n, то в $(1 - \alpha)k$ случаях построенные интервалы накрывают истинное значение исследуемого параметра.

Поскольку в эконометрических задачах часто приходится находить доверительные интервалы параметров случайных величин, имеющих нормальное распределение, приведем схемы их определения.

I. Доверительный интервал для математического ожидания нормальной CB *при известной дисперсии*

Пусть количественный признак X генеральной совокупности имеет нормальное распределение с заданной дисперсией σ^2 и неизвестным математическим ожиданием *m* *(X ~ N (m,* σ^2*)).*

Построим доверительный интервал для m.

1. Пусть для оценки m извлечена выборка $x_1, x_2, ..., x_n$ объема n. Тогда $m^* = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$

2. Составим СВ $U = \dfrac{\bar{x} - m}{\sigma / \sqrt{n}.}$. Нетрудно показать, что СВ U имеет стандартизированное нормальное распределение, т.е.

$$U - N(0{,}1) \; (f(u) = \frac{1}{\sqrt{2\Pi u}} e^{-\frac{u^2}{2}}).$$

3. Зададим уровень значимости α.

4. Применяя формулу нахождения вероятности отклонения нормальной величины от математического ожидания, имеем:

$$P(|U| \langle u_{a/2} = P\left(\left|\frac{\bar{x} - m}{\delta / \sqrt{n}}\right|\right) \langle u_{a/2} =$$
$$= P(\bar{x} - u_{a/2} \frac{\delta}{\sqrt{n}} \langle \, m \langle \bar{x} + u_{a/2} \frac{\delta}{\sqrt{n}} = 1 - \alpha. \tag{29}$$

Это означает, что доверительный интервал $(\bar{x} - u_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}; \quad \bar{x} + u_{\alpha/2} \dfrac{\sigma}{\sqrt{n}})$ накрывает неизвестный параметр m с надежностью $1 - \alpha.$ Точность оценки определяется величиной $\delta = u_{\alpha/2} \dfrac{\sigma}{\sqrt{n}.}$

Отметим, что по таблице Лапласа число $u_{\alpha/2}$ определяется по таблице значений функции Лапласа из равенства $\Phi(u_{\alpha/2}) = \dfrac{1-\alpha}{2} = \dfrac{\gamma}{2}$.

II. Доверительный интервал для математического ожидания нормальной СВ при неизвестной дисперсии.

В реальной жизни истинное значение дисперсии исследуемой СВ, чаще всего, известно не будет. Это приводит к необходимости использования другой формулы при определении доверительного интервала для математического ожидания СВ, имеющей нормальное распределение.

Для этого из генеральной совокупности СВ X извлекается выборка объема $n: x_1, x_2, \ldots x_n$.

1.В качестве точечной оценки математического ожидания m используется выборочное среднее $\bar{x}$, а в качестве оценки дисперсии σ^2 – исправленная выборочная дисперсия $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, которой соответствует стандартное отклонение $S = \sqrt{S^2}$.

2. Для нахождения доверительного интервала строится статистика $T = \frac{\bar{x} - m}{S/\sqrt{n}}$, имеющая в этом случае распределение Стьюдента с числом степеней свободы $v = n-1$ независимо от значений параметров m и σ^2.

3. Задается требуемый уровень значимости α.

4. Применяется следующая формула расчета вероятности

$$P(|T| < t_{\alpha/2,\,n-1}) = P(-t_{\alpha/2,\,n-1} < T < t_{\alpha/2,\,n-1}) = 1-\alpha, \quad (30)$$

где $t_{\alpha/2,\,n-1}$ - критическая точка распределения Стьюдента, которая находится по соответствующей таблице.

Тогда

$$P(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}) = P(-t_{\alpha/2,n-1} < \frac{\bar{x}-m}{S/\sqrt{n}} < t^2_{\alpha/2,n-1}) =$$

$$= P(\bar{x} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < m < \bar{x} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}) = 1 - \alpha.$$

(31)

Это означает, что интервал $(\bar{x} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}; \bar{x} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}})$ накрывает неизвестный параметр m с надежностью $1-\alpha$.

III. Доверительный интервал для дисперсии нормальной СВ.

Пусть $X \sim N(m, \sigma^2)$ причем m и σ^2 - неизвестны. Пусть для оценки σ^2 извлечена выборка объема $n: x_1, x_2, \ldots x_n$.

1. В качестве точечной оценки дисперсии $D(x)$ используется исправленная выборочная дисперсия: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, которой соответствует стандартное отклонение $S = \sqrt{S^2}$.

2. При нахождении доверительного интервала для дисперсии в этом случае вводится статистика $\chi^2 = \frac{(n-1)S^2}{\sigma^2}$, имеющая χ^2 - распределение с числом степеней свободы $\nu = n-1$ независимо от значения параметра σ^2.

3. Задается требуемый уровень значимости α.

4. Тогда, используя таблицу критических точек $\chi^2 -$ распределения, нетрудно указать критические точки $\chi^2_{1-\alpha/2,\, n-1}$, $\chi^2_{\alpha/2,n-1}$, для которых будет выполняться следующее равенство:

$$P(\chi^2_{1-\alpha/2,\, n-1} < \chi^2 < \chi^2_{\alpha/2,n-1}) = 1-\alpha. \qquad (32)$$

Неравенство $\frac{S^2(n-1)}{\chi^2_{\frac{a}{2},n-1}} < \sigma^2 < \frac{S^2(n-1)}{\chi^1_{1-\frac{a}{2},n-1}}$ (33)

может быть преобразовано в следующее:

$$S\sqrt{(n-1)\Big/\chi^2_{\alpha/2,n-1}} < \sigma < S\sqrt{(n-1)\Big/\chi^2_{1-\alpha/2,n-1}} \quad (34)$$

Таким образом, доверительный интервал $\left(\dfrac{S^2(n-1)}{\chi^2_{\alpha/2,n-1}};\dfrac{S^2(n-1)}{\chi^2_{1-\alpha/2,n-1}}\right)$ накрывает неизвестный параметр σ^2 с надежностью $1-\alpha.$ А доверительный интервал $S\sqrt{(n-1)\Big/\chi^2_{\alpha/2,n-1}};S\sqrt{(n-1)\Big/\chi^2_{1-\alpha/2,n-1}}$ с надежностью $1-\alpha$ накрывает неизвестный параметр $\sigma.$

Статистическая проверка статистических гипотез.

Большинство эконометрических моделей требуют многократного улучшения и уточнения. Для этого необходимо провести расчеты, связанные с установлением выполнимости или невыполнимости тех или иных предпосылок, анализом качества найденных оценок, достоверностью полученных выводов. Обычно эти расчеты проводятся по схеме статистической проверки гипотез.

Во многих случаях необходимо знать закон распределения генеральной совокупности. Если закон распределения неизвестен, но есть основания предположить, что он имеет определенный вид (например, R), выдвигают гипотезу: генеральная совокупность СВ X распределена по закону R

Возможен другой случай, когда закон распределения известен, а его параметры неизвестны. Если есть основания предположить, что неизвестный параметр θ равен ожидаемому числу $\theta_0,$ выдвигают гипотезу: $\theta = \theta_0.$

Статистической гипотезой называется любое предположение о виде закона распределения или о параметрах неизвестного закона распределения. В первом случае гипотеза называется **непараметрической,** а во втором **параметрической.**

Гипотеза $H_{0,}$ подлежащая проверке, называется **нулевой** или нуль – гипотезой (основной). Наряду с нулевой рассматривают гипотезу $H_{1,}$ которая будет приниматься, если отклоняется $H_{0.}$. Такая гипотеза называется **альтернативной (конкурирующей).**

Например, если проверяется гипотеза о равенстве параметра θ некоторому значению $\theta_{0,}$ т.е. $H_0 : \theta = \theta_{0,}$ то в качестве альтернативной могут рассматриваться следующие гипотезы:

$$H_1^{(1)} : \theta \neq \theta_0;$$
$$H_1^{(2)} : \theta > \theta_0;$$
$$H_1^{(3)} : \theta < \theta_0;$$
$$H_1^{(4)} : \theta = \theta_1; \quad (\theta_1 \neq \theta_0).$$

Выбор альтернативной гипотезы определяется конкретной формулировкой задачи, а нулевая гипотеза часто специально подбирается так, чтобы отвергнуть ее и принять тем самым альтернативную гипотезу. Для того, чтобы принять гипотезу о наличии корреляции между двумя экономическими показателями (например, между инфляцией и безработицей), можно опровергнуть гипотезу об отсутствии такой корреляции, взяв ее в качестве нулевой гипотезы.

Гипотезу называют **простой,** если она содержит одно конкретное предложение $(H_0 : \theta = \theta_{0,}\ H_1^{(4)} : \theta = \theta_1)$.

Гипотезу называют **сложной,** если она состоит из конечного или бесконечного числа простых гипотез:

$$H_1^{(1)} : \theta \neq \theta_0; H_1^{(2)} : \theta > \theta_0;\ H_1^{(3)} : \theta < \theta_0.$$

При проверке гипотезы выборочные данные могут противоречить гипотезе H_0. Тогда она **отклоняется.** Если же статистические данные согласуются с выдвинутой гипотезой, то она **не отклоняется**. В последнем случае частот говорят, что нулевая гипотеза принимается. Статистическая проверка гипотез на основании выборочных данных неизбежно связана с риском принятия ложного решения. При этом возможны ошибки двух родов.

1. **Ошибка первого рода** состоит в том, что будет отвергнута правильная нулевая гипотеза.

2. **Ошибка второго рода** состоит в том, что будет принята нулевая гипотеза, в то время как в действительности верна альтернативная гипотеза.

Возможные результаты статистических выводов представлены следующей таблицей:

Результаты проверки гипотезы	Возможные состояния гипотезы	
	верна H_0	верна H_1
Гипотеза H_0 отклоняется	Ошибка первого рода, вероятность α	Правильный вывод, вероятность $(1-\beta)$
Гипотеза H_0 не отклоняется	Правильный вывод, вероятность $(1-\alpha)$	Ошибка второго рода, вероятность β

Последствия указанных ошибок неравнозначны.

Первая приводит к более осторожному, консервативному решению, вторая – к неоправданному риску.

Исключить вообще ошибки 1–ого и 2–ого рода невозможно в силу ограниченности выборки. Поэтому стремятся минимизировать потери от этих ошибок. Отметим, что одновременное уменьшение вероятностей данных ошибок невозможно, т.к. задачи их уменьшения являются конкурирующими, и снижение вероятности допустить одну из них влечет за собой увеличение вероятности допустить другую. В большинстве случаев единственный способ уменьшения вероятности ошибок состоит **в увеличении объема выборки.**

Вероятность совершить ошибку 1–ого рода принято обозначать буквой α и ее называют **уровнем значимости**. Вероятность совершить ошибку 2–ого рода обозначают β. Тогда вероятность не совершить ошибку второго рода $(1-\beta)$ называется **мощностью критерия.**

Обычно значения α задают заранее, «круглыми» числами (например, 0,1; 0,05; 0,01 и т.п.), а затем стремятся построить критерий наименьшей мощности. Т.о., если $\alpha = 0{,}05$, то это означает, что исследователь не хочет совершить ошибку 1–ого рода более чем в 5-ти случаях из 100.

Критерии проверки. Критическая область.

Проверку статистической гипотезы осуществляют на основании данных выборки. Для этого используют специально подобранную СВ (статистику, критерий), точное или приближенное значение которой известно. Эту величину обозначают:

$U\ (или\ Z) -$ если она имеет стандартизированное нормальное распределение;

$T -$ если она распределена по закону Стьюдента;

$\chi 2$ - если она распределена по закону χ^2;

$F -$ если она имеет распределение Фишера.

В целях общности будем обозначать такую СВ через K.

Таким образом, **статистическим критерием** или **статистическим тестом** называют СВ K, которая служит для проверки нулевой гипотезы. После выбора определенного критерия множество всех его возможных значений разбивают на два непересекающихся подмножества:

- одно из них содержит значения критерия, при которых нулевая гипотеза отклоняется;

- другое – при которых она не отклоняется .

Совокупность значений критерия, при которых нулевую гипотезу отклоняют, называют **критической областью**.

Совокупность значений критерия, при которых нулевую гипотезу не отклоняют, называют **областью принятия гипотезы.**

Основной принцип проверки статистических гипотез можно сформулировать так: если наблюдаемое значение критерия K (вычисленное по выборке) принадлежит критической области, то нулевую гипотезу отклоняют. Если же наблюдаемое значение критерия K принадлежит области принятия гипотезы, то нулевую гипотезу не отклоняют (принимают).

Точки, разделяющие критическую область и область принятия гипотезы, называют **критическими.**

В основу определения критических точек и критической области положен принцип практической невозможности маловероятных событий.

Пусть для проверки нулевой гипотезы H_0 служит критерий K. Предположим, что плотность распределения вероятности СВ K в случае справедливости H_0 имеет вид $f(k|H_0)$, а математическое ожидание K равно k_0.

Тогда вероятность того, что СВ K попадет в произвольный интервал $(k_{1-a/2}, k_{a/2})$, можно найти по формуле:

$$P(k_{1-a/2} < K < k_{a/2}) = \int_{k_{1-a/2}}^{k_{a/2}} f(k|H_0)dk \qquad (35)$$

Зададим эту вероятность равной $1-\alpha$ и вычислим критические точки (квантили) $K-$ распределения $k_{1-a/2}, k_{a/2}$ из условий:

$$P(K \le k_{1-a/2}) = \int_{-\infty}^{k_{1-a/2}} f(k|H_0)dk = \frac{a}{2}, \tag{36}$$

$$P(K \ge k_{a/2}) = \int_{k_{a/2}}^{+\infty} f(k|H_0)dk = \frac{a}{2} \tag{37}$$

Следовательно,

$$P(k_{1-a/2} < K < k_{a/2}) = 1-\alpha,$$

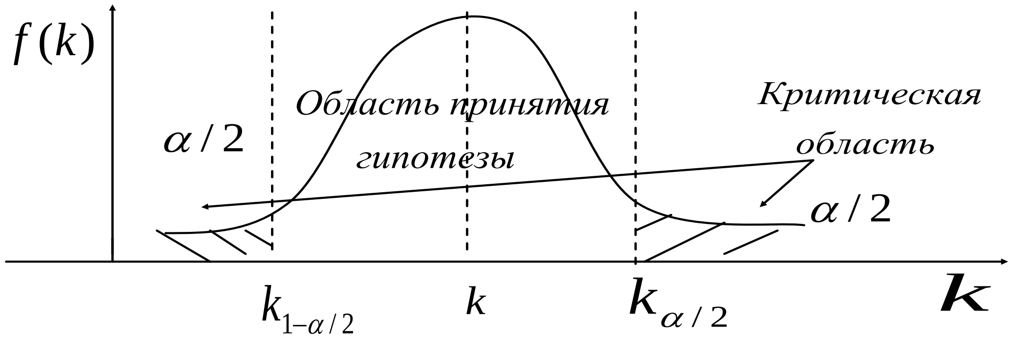
$$P\{(K \le k_{1-a/2}) \cup (K \ge k_{a/2})\} = \alpha \tag{38}$$

Зададим вероятность a настолько малой (0,05; 0,01), что попадание СВ K за пределы интервала $(k_{1-a/2}, k_{a/2})$ можно было бы считать маловероятным событием. Тогда, исходя из принципа практической невозможности маловероятных событий, можно считать, что если H_0 справедлива, то при ее проверке с помощью критерия K по данным одной выборки наблюдаемое значение K должно наверняка попасть в интервал $(k_{1-a/2}, k_{a/2})$.

Если же наблюдаемое значение K попадает за пределы указанного интервала, то произойдет маловероятное, практически невозможное событие. Это дает основание считать, что с вероятностью $1-\alpha$ нулевая гипотеза H_0 несправедлива.

Точки $k_{1-a/2}, k_{a/2}$ называются **критическими.**

Критическая область $(-\infty; k_{1-a/2}) \cup (k_{a/2,} +\infty)$ называется **двусторонней критической областью**. Она определяется в случае, когда альтернативная гипотеза имеет вид $H_1 : \theta \neq \theta_0$.



$H_1 : \theta \neq \theta_0$

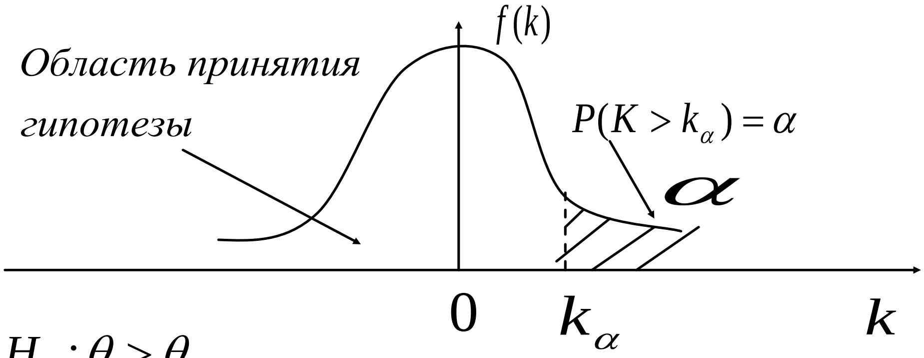
Кроме двусторонней, рассматривают также **односторонние критические области –** правостороннюю и левостороннюю.

Правосторонней называют критическую область $(k_{\alpha,} +\infty)$, определяющуюся из соотношения $P(K > k_\alpha) = \alpha$.

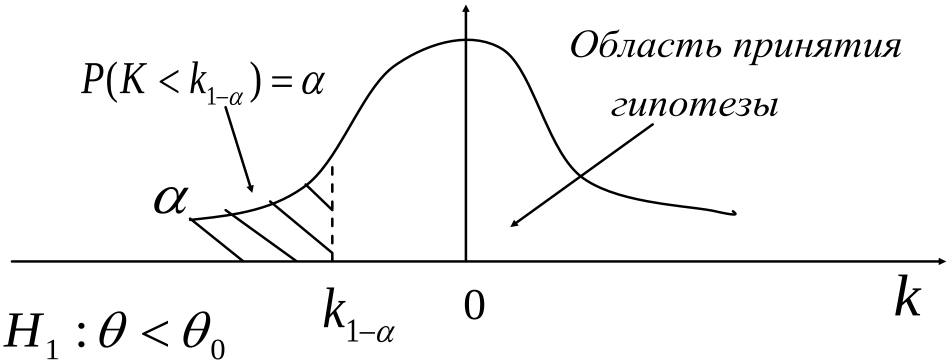
Она используется в случае, когда альтернативная гипотеза имеет вид: $H_1 : \theta > \theta_0$.

Левосторонней называют критическую область $(-\infty, k_{1-\alpha})$, определяющую из соотношения $P(K < k_{1-\alpha}) = \alpha$.

Она используется в случае, когда альтернативная гипотеза имеет вид $H_1: \theta < \theta_0$.



$H_1 : \theta > \theta_0$



$H_1 : \theta < \theta_0$

Общая схема проверки гипотез.

1. Формулировка проверяемой (нулевой - H_0) и альтернативной (H_1) гипотез.

2. Выбор соответствующего уровня значимости α.

3. Определение объема выборки n.

4. Выбор критерия K для проверки H_0.

5. Определение критической области и области принятия гипотезы.

6. Вычисление наблюдаемого значения критерия $K_{набл.}$

7. Принятие статистического решения.

Схемы проверки гипотез и доверительные интервалы.

Проверка гипотез при двусторонней критической области тесно связана с **интервальным** оцениванием. При одном и том же уровне значимости α и объеме выборки n попадание предполагаемого значения исследуемого параметра в доверительный интервал равносильно попаданию соответствующего критерия в область принятия гипотезы. Поэтому для проверки гипотезы в этом случае можно использовать доверительный интервал.

Если прелдполагаемоеое значение исследуемого параметра попадает в этот интервал, то делают вывод, что нет оснований для отклонения выдвигаемой гипотезы.

I. Схема проверки гипотезы о математическом ожидании нормальной СВ при известной дисперсии.

Пусть генеральная совокупность X распределена нормально, причем ее математическое ожидание m неизвестно, а дисперсия σ^2 известна.

Также есть основания предполагать, что $m = m_0$.

Тогда

$$H_0 : m = m_0$$

$$H_1^{(1)} : m \neq m_0 (H_1^{(2)} : m > m_0 ; H_1^{(3)} : m < m_0)$$

Для проверки H_0 извлекается выборка объема $n : x_1, x_2 ..., x_n$ и в качестве критерия строится статистика

$$U = \frac{\bar{x} - m_0}{\sigma / \sqrt{n}} \qquad (1)$$

где $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, $\sigma = \sqrt{\sigma^2}$.

Доказано, что если гипотеза H_0 справедлива, то статистика U имеет стандартизированное нормальное распределение $(U \approx N(0,1))$.

1. Пусть в качестве альтернативной рассматривается гипотеза:

$H_1^{(1)} : m \neq m_0$. Тогда критические точки $u_{\alpha/2}$ и $u_{1-\alpha/2} = -u_{\alpha/2}$ будут определяться по таблице значений функции Лапласа из условия

$$\Phi(u_{\alpha/2}) = \frac{1-\alpha}{2}$$

Если $|U_{набл.}| = \left|\frac{\bar{x} - m_0}{\sigma/\sqrt{n}}\right| < u_{\alpha/2}$ - нет оснований для отклонения H_0.

Если $|U_{набл.}| \geq u_{\alpha/2}$ - гипотеза H_0 отклоняется в пользу альтернативной гипотезы $H_1^{(1)}$.

2. При $H_1^{(2)} : m > m_0$ критическую точку u_α правосторонней критической области находят из равенства

$$\Phi(u_\alpha) = \frac{1-2\alpha}{2}$$

Если $U_{набл.} < u_\alpha$ - нет оснований для отклонения H_0.

Если $U_{набл.} \geq u_\alpha$ - H_0 отклоняют в пользу $H_1^{(2)}$.

2. При $H_1^{(3)} : m < m_0$ критическая точка $u_{1-\alpha} = -u_\alpha$.

Если $U_{набл.} > u_{1-\alpha}$ - нет оснований для отклонения H_0.

Если $U_{набл.} \le u_{1-\alpha}$ - H_0 отклоняют в пользу $H_1^{(3)}$.

II. Схема проверки гипотезы о математическом ожидании нормальной СВ при неизвестной дисперсии.

Пусть генеральная совокупность X имеет нормальное распределение, причем ее математическое ожидание m и дисперсия σ^2 неизвестны.

Пусть есть основания утверждать, что $m = m_0$. Тогда строятся следующие гипотезы:

$$H_0 : m = m_0$$

$$H_1^{(1)} : m \ne m_0 \quad (H_1^{(2)} : m > m_0; H_1^{(3)} : m < m_0).$$

Для проверки H_0 извлекается выборка объема $n : x_1, x_2 ... x_n$; вычисляются выборочное среднее

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i ;$$

исправленная выборочная дисперсия

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 ;$$

стандартное отклонение $S = \sqrt{S^2}$.

Далее строится t - статистика:

$$T = \frac{\bar{x} - m_0}{S / \sqrt{n}}, \qquad (2)$$

имеющая при справедливости H_0 распределение Стьюдента с $\nu = n-1$ степенями свободы. Критическая область строится в зависимости от вида альтернативной гипотезы так же, как и в предыдущем разделе.

1. При $H_1^{(1)}: m \neq m_0$ по таблице критических точек распределения значимости α и числу степеней свободы $\nu = n-1$ находятся критические точки:

$$t_{\alpha/2,n-1} \text{ и } t_{1-\alpha/2,n-1} = -t_{\alpha/2,n-1}.$$

Если $\left|T_{набл.}\right| = \left|\dfrac{\bar{x}-m_0}{S/\sqrt{n}}\right| < t_{\alpha/2,n-1}$ - нет оснований для отклонения H_0.

Если $\left|T_{набл.}\right| \geq t_{\alpha/2,n-1}$- H_0 отклоняют в пользу $H_1^{(1)}$.

2. При $H_1^{(2)}: m > m_0$ определяют критическую точку $t_{\alpha,n-1}$ правосторонней критической области.

Если $T_{набл.} < t_{\alpha,n-1}$ - нет оснований для отклонения H_0.

Если $T_{набл.} \geq t_{\alpha,n-1}$ - H_0 отклоняется в пользу $H_1^{(2)}$.

3. При $H_1^{(3)}: m < m_0$ определяют критическую точку $t_{1-\alpha,n-1} = -t_{\alpha,n-1}$ левосторонней критической области.

Если $T_{набл.} > -t_{\alpha,n-1}$ - нет оснований для отклонения H_0.

Если $T_{набл.} \leq -t_{\alpha,n-1}$ - H_0 отклоняется в пользу $H_1^{(3)}$.

III. Схема проверки гипотезы о величине дисперсии нормальной СВ.

Принятие того или иного решения в экономике часть связано с анализом возможных результатов, точнее разбросе возможных результатов.

Здесь приходится иметь дело с выдвижением и проверкой гипотез о величине дисперсии.

Пусть случайная величина *X ~ N (m, σ²)*; m и σ^2 неизвестны. Проверяется гипотеза о равенстве дисперсии σ^2 нормально распределенной генеральной совокупности X предполагаемому значению σ_0^2. Тогда:

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_1^{(1)} : \sigma^2 \neq \sigma_0^2 ; H_1^{(2)} : \sigma^2 > \sigma_0^2 ; H_1^{(3)} : \sigma^2 < \sigma_0^2 .$$

Для проверки H_0 извлекается выборка объема $n : x_1, x_2 ... x_n$, вычисляются выборочное среднее $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, исправленная выборочная дисперсия $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2$.

Тогда критерий проверки H_0 имеет вид:

$$\chi^2 = \frac{(n-1) \cdot S^2}{\sigma_0^2} \qquad (3)$$

При справедливости H_0 построенная статистика χ^2 имеет χ^2-распределение с $v = n-1$ степенями свободы.

1. При $H_1^{(1)} : \sigma^2 \neq \sigma_0^2$ по таблице критических точек $\chi 2$-распределения по заданному уровню значимости α и числу степеней свободы $v = n-1$ находят критические точки $\chi^2_{1-\alpha/2, n-1}$ и $\chi^2_{\alpha/2, n-1}$ двусторонней критической области.

Если $\chi^2_{1-\alpha/2, n-1} < \chi^2_{набл.} < \chi^2_{\alpha/2, n-1}$ - нет оснований для отклонения H_0.

Если $\chi^2_{набл.} < \chi^2_{1-\alpha/2,n-1}$ или $\chi^2{}_{набл.} \geq \chi^2{}_{\alpha/2,n-1}$ - H_0 отклоняется в пользу $H_1^{(1)}$.

2. При $H_1^{(2)}: \sigma^2 > \sigma_0^2$ определяют критическую точку $\chi^2_{\alpha,n-1}$ правосторонней критической области.

Если $\chi^2_{набл.} < \chi^2_{\alpha,n-1}$- нет оснований для отклонения H_0.

Если $\chi^2_{набл.} \geq \chi^2_{\alpha,n-1}$- H_0 отклоняется в пользу $H_1^{(2)}$.

3. При $H_1^{(3)}: \sigma^2 < \sigma_0^2$ находят критическую точку $\chi^2_{1-\alpha,n-1}$ левосторонней критической области.

Если $\chi^2_{набл.} > \chi^2_{1-\alpha,n-1}$- нет оснований для отклонения H_0.

Если $\chi^2_{набл.} \leq \chi^2_{1-\alpha,n-1}$- H_0 отклоняется в пользу $H_1^{(3)}$.

IV. Схема проверки гипотезы о равенстве $M(X)$ двух нормальных СВ при известных дисперсиях.

Пусть $X \sim N\ (m_x,\ \sigma_x^2)$ и $Y \sim N\ (m_y,\ \sigma_y^2)$, причем их дисперсии σ_x^2 и σ_y^2 - известны (из предыдущих наблюдений или определены теоретически).

По двум выборкам $x_1, x_2 ... x_n$ и $y_1, y_2 ... y_k$ объемов n и k соответственно необходимо проверить гипотезу $M(X) = M(Y)$, т.е.

$$H_0 : M(X) = M(Y)$$

$$H_1^{(1)} : M(X) \neq M(Y) \quad \left(H_1^{(2)} : M(X) > M(Y)\right)$$

$$\left(H_1^{(3)} : M(X) < M(Y)\right).$$

В качестве критерия проверки H_0 принимается СВ U:

$$U = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{k}}} \qquad (4)$$

При справедливости H_0 СВ $U \sim N\ (0,\ 1)$.

1. При $H_1^{(1)} : M(X) \neq M(Y)$ по таблице функции Лапласа определяют 2 критические точки $u_{1-\alpha}$ и $u_{\alpha/2}$ из условий:

$$\Phi(u_{\alpha/2}) = \frac{1-\alpha}{2},\ u_{1-\alpha/2} = u_{\alpha/2}.$$

Если $|U_{набл.}| < u_{\alpha/2}$ - нет оснований для отклонения H_0.

Если $|U_{набл.}| \geq u_{\alpha/2}$ - H_0 отклоняется в пользу $H_1^{(1)}$.

2. При $H_1^{(2)} : M(X) > M(Y)$ критическую точку u_α правосторонней критической области находят их равенства: $\Phi(u_\alpha) = \frac{1-2\alpha}{2}$.

Если $U_{набл.} < u_\alpha$ - нет оснований для отклонения H_0.

Если $U_{набл.} \geq u_\alpha$ - H_0 отклоняется в пользу $H_1^{(2)}$.

3. При $H_1^{(3)}: M(X) < M(Y)$ критическая точка $u_{1-\alpha}$ левосторонней критической области определяется из соотношения $u_{1-\alpha} = -u_{\alpha}$.

Если $U_{набл.} > u_{1-\alpha}$ - нет оснований для отклонения H_0.

Если $U_{набл.} \le u_{1-\alpha}$ - H_0 отклоняется в пользу $H_1^{(3)}$.

V. Схема проверки гипотезы о равенстве математических ожиданий двух нормальных СВ при неизвестных дисперсиях.

Пусть $X \approx N(m_x, \sigma_x^2)$ и $Y \approx N(m_y, \sigma_y^2)$,причем их дисперсии σ_x^2 и σ_y^2 неизвестны. Выдвигается гипотеза о равенстве математических ожиданий:

$$H_0: M(X) = M(Y)$$

$$H_1: M(X) \ne M(Y) \quad \left(H_1^{(2)}: M(X) > M(Y);\ H_1^{(3)}: M(X) < M(Y)\right).$$

При этих условиях в качестве критерия проверки H_0 принимают СВ T:

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)S_x^2 + (k-1)S_y^2}} \cdot \sqrt{\frac{nk(n+k-2)}{n+k}} \qquad (5)$$

где n, k - объемы выборок $x_1, x_2 ... x_n$ и $y_1, y_2 ... y_k$ соответственно

$$\bar{x} = \frac{1}{n}\sum x_i; \quad S_x^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2,$$

$$\bar{y} = \frac{1}{k}\sum y_i; \quad S_y^2 = \frac{1}{k-1}\sum (y_i - \bar{y})^2.$$

При справедливости H_0 построенная статистика T имеет t-распределение Стьюдента с $\nu = n + k - 2$ степенями свободы.

1. При $H_1^{(1)}: M(X) \neq M(Y)$ с помощью таблицы критических точек распределения Стьюдента по заданному уровню значимости α и числу степеней свободы $\nu = n + k - 2$ определяются критические точки $t_{1-\alpha/2, n+k-2}$ и $t_{\alpha/2, n+k-2}$ $(t_{1-\alpha/2, n+k-2} = -t_{\alpha/2, n+k-2})$ двусторонней критической области.

Если $|T_{набл.}| < t_{\alpha/2, n+k-2}$ - нет оснований для отклонения H_0.

Если $|T_{набл.}| \geq t_{\alpha/2, n+k-2}$ - H_0 отклоняется в пользу $H_1^{(1)}$.

2. При $H_1^{(2)}: M(X) > M(Y)$ находят критическую точку $t_{\alpha, n+k-2}$ правосторонней критической области.

Если $T_{набл.} < t_{\alpha, n+k-2}$ - нет оснований для отклонения H_0.

Если $T_{набл.} \geq t_{\alpha, n+k-2}$ - H_0 отклоняется в пользу $H_1^{(2)}$.

3. При $H_1^{(3)}: M(X) < M(Y)$ находят критическую точку левосторонней критической области $t_{1-\alpha, n+k-2} = -t_{\alpha, n+k-2}$.

Если $T_{набл.} > -t_{\alpha, n+k-2}$ - нет оснований для отклонения H_0.

Если $T_{набл.} \leq -t_{\alpha, n+k-2}$ - H_0 отклоняется в пользу $H_1^{(3)}$.

VI. Схема проверки гипотезы о равенстве дисперсий двух нормальных СВ.

При сравнении двух экономических показателей иногда, в первую очередь, проводят анализ разброса значений рассматриваемых СВ. Например, при решении инвестирования в одну из отраслей остро стоит проблема риска вложений. При сравнивании уровня жизни двух стран среднедушевые доходы могут быть примерно одинаковы. Необходимо сопоставить разброс в доходах.

Анализ проводится путем сравнения дисперсий исследуемых СВ.

Пусть $X \approx N(m_x, \sigma_x^2)$ и $Y \approx N(m_y, \sigma_y^2)$, причем их дисперсии σ_x^2 и σ_y^2 неизвестны. Выдвигается гипотеза о равенстве дисперсий σ_x^2 и σ_y^2.

$H_0 : \sigma_x^2 = \sigma_y^2$

$H_1^{(1)} : \sigma_x^2 \neq \sigma_y^2 \quad (H_1^{(2)} : \sigma_x^2 > \sigma_y^2)$.

По независимым выборкам $x_1, x_2 ... x_n$ и $y_1, y_2 ... y_k$ объемов n и k соответственно определяется:

$\bar{x}, \bar{y}, S_x^2$ и S_y^2 (для определенности пусть $S_x^2 \geq S_y^2$, в противном случае эти величины можно переобозначить).

В качестве критерия проверки H_0 принимают СВ

$$F = \frac{S_x^2}{S_y^2} , \qquad (6)$$

определяемую отношением большей исправленной выборочной дисперсии к меньшей.

Если H_0 верна, то данная статистика F имеет F - распределение Фишера с $v_1 = n-1$ и $v_2 = k-1$ степенями свободы.

1. При $H_1^{(1)} : \sigma_x^2 \neq \sigma_y^2$ по таблицам критических точек распределения Фишера по уровню значимости α и числам степеней свободы v_1 и v_2 определяется критическая точка $F_{\alpha/2, v_1, v_2}$.

Если $F_{набл.} < F_{\alpha/2, v_1, v_2}$ - нет оснований для отклонения H_0.

Если $F_{набл.} \geq F_{\alpha/2, v_1, v_2}$ - H_0 отклоняется в пользу $H_1^{(1)}$.

2. При $H_1^{(2)}: \sigma_x^2 > \sigma_y^2$ определяется критическая точка F_{α, v_1, v_2}.

Если $F_{набл.} < F_{\alpha, v_1, v_2}$ - нет оснований для отклонения H_0.

Если $F_{набл.} \geq F_{\alpha, v_1, v_2}$ - H_0 отклоняется в пользу $H_1^{(2)}$.

В основном, при проверке гипотезы о равенстве дисперсий в качестве альтернативной гипотезы в большинстве случаев используется гипотеза $H_1^{(2)}$.

VII. Схема проверки гипотезы о значимости коэффициента корреляции.

Одним из важнейших элементов эконометрического анализа является установление наличия связи между различными показателями (между ценой и спросом, доходом и потреблением, инфляцией и безработицей).

Обычно анализ начинают с простой линейной зависимости. Для того чтобы установить наличие значимой линейной связи между двумя СВ X и Y, следует проверить гипотезу о статистической значимости коэффициента корреляции. В этом случае используется следующая гипотеза:

$$H_0: \rho_{xy} = 0$$

$$H_1^{(1)}: \rho_{xy} \neq 0.$$

Для проверки H_0 по выборке $(x_1, y_1), (x_1, y_2) ... (x_n, y_n)$ объема n строится статистика:

$$T = \frac{r_{xy} \cdot \sqrt{n-2}}{\sqrt{1 - r_{xy}^2}} \qquad (7)$$

где r_{xy} - выборочный коэффициент корреляции.

При справедливости H_0 статистика T имеет распределение Стьюдента с $v = n - 2$ степенями свободы.

По таблице критических точек распределения Стьюдента по заданному уровню значимости α и числу степеней свободы $n-2$ определяем критическую точку $t_{\alpha/2,n-2}$.

Если $\left|T_{набл.}\right| < t_{\alpha/2,n-2}$ - то нет оснований для отклонения H_0.

Если $\left|T_{набл.}\right| \geq t_{\alpha/2,n-2}$- то H_0 отклоняется в пользу альтернативной гипотезы $H_1^{(1)}$.

Если H_0 отклоняется, то фактически это означает, что коэффициент корреляции статистически значим (существенно отличен от нуля). Следовательно, X и Y- коррелированны, т.е. между ними существует линейная связь.

Список рекомендуемой литературы:

1. Эконометрика: Учебник / Под редакцией И.И.Елисеевой. – М.: Финансы и статистика, 2002
2. Практикум по эконометрике: Учебное пособие / И.И.Елисеева, С.В.Курышева, Н.М. Гордеенко и др.; Под ред. И.И.Елисеевой. – М.: Финансы и статистика, 2002
3. Магнус Я.Р., Катышев Л.К., Пересецкий А.А. Эконометрика. Начальный курс: Учебник. – 4 изд. – М.: Дело, 2000
4. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики: Учебник для вузов: В 2 т. – М.: ЮНИТИ, 2001
5. Доугерти К. Введение в эконометрику. – М.: Инфра-М, 2001
6. Кулинич Е.И. Эконометрия. – М.: Финансы и статистика, 1999
7. Бородич С.А. Эконометрика: Учеб пособие / С.А. Бородич. – Мн.: Новое знание, 2001
8. Кремер Н.Ш., Путко Б.А. Эконометрика: Учебник для вузов / Под ред. проф. Н.Ш. Кремера. – М.: ЮНИТИ-ДАНА, 2002