

ЭКОНОМЕТРИКА

Введение. Модель и моделирование.

Эконометрические модели их свойства и классификация.

Переменные в моделях и их типы. Этапы моделирования.

Эконометрика – наука, изучающая количественные закономерности и взаимозависимости в экономике методами математической статистики. Можно считать, что эконометрика – «экономика» + «метрика». Это наука об измерении и анализе экономических явлений, о количественных выражениях тех связей и соотношений, которые раскрыты и обоснованы экономической теорией. Это сплав четырех компонент: экономической теории, статистических и математических методов, компьютерных вычислений.

Основная задача курса «Эконометрика» – обучение будущих экономистов методам исследования экономических процессов, основанных на использовании реальных статистических данных и на положениях экономической теории. При изучении взаимосвязей между экономическими показателями применяются положения экономической теории, затем на основе выборочных (статистических) данных строятся модели, проводится оценка их составляющих. Итоговые результаты используются для прогнозирования, принятия решений и уточнения первоначальных положений. Основной особенностью эконометрических моделей является наличие в них случайных составляющих, обусловленное неучтенными факторами, ошибками измерений и др. При этом особую важность приобретает использование основ теории вероятностей и статистики.

Модель – объект любой природы, который создается исследователем с целью получения новых знаний об объекте-оригинале и отражает только существенные (с точки зрения разработчика) свойства оригинала.

Эконометрическая модель – вероятностно – статистическая модель, описывающая механизм функционирования экономической или социально – экономической системы. Можно выделить три основных класса эконометрических моделей: регрессионные модели с одним уравнением, модели временных рядов и системы одновременных уравнений.

В **регрессионных моделях с одним уравнением** объясняемая переменная представляется в виде функции от объясняющих переменных. По виду функции регрессионные модели делятся на линейные и нелинейные.

К моделям **временных рядов** относятся **модели тренда** и **модели сезонности**. Тренд представляет собой устойчивое изменение уровня показателя в течение длительного периода времени. Сезонность характеризует устойчивые колебания уровня показателя в течение небольшого промежутка времени (месяца, квартала, года).

Системы одновременных уравнений описываются системами уравнений, состоящими из тождеств и регрессионных уравнений, в каждом из которых аргументы содержат не только объясняющие переменные, но и объясняемые из других уравнений системы.

Моделирование – процесс построения, изучения и применения моделей.

Экзогенные переменные в модели – переменные, задаваемые «извне», автономно от модели, управляемые и планируемые.

Эндогенные переменные модели – переменные, значения которых формируются в процессе и внутри функционирования анализируемой социально – экономиче-

ской системы в существенной мере под воздействием экзогенных переменных и во взаимодействии друг с другом. В эконометрическом моделировании являются предметом объяснения.

Лаговые переменные – это те переменные (экзогенные или эндогенные), значения которых рассматриваются в различные моменты времени, разделенные некоторым промежутком (лагом). Значения эндогенных переменных в предыдущие моменты времени могут заранее задаваться или вычисляться по уравнениям модели. Тогда эти эндогенные переменные играют роль экзогенных (объясняющих).

Предопределенные переменные модели – все экзогенные переменные модели и лаговые значения эндогенных переменных, которые служат для нахождения значений эндогенных переменных в данный момент времени.

Укажем основные этапы эконометрического исследования.

1. Постановочный. Формулируется цель исследования, определяется набор участвующих в модели экономических переменных. Целью исследования могут быть анализ процесса, прогноз экономических показателей, анализ возможного развития явления при различных значениях экзогенных переменных, выработка управленческих решений. При выборе переменных необходимо теоретическое обоснование каждой экономической переменной.

2. Априорный. Проводится анализ сущности изучаемого объекта, формирование и формализация известной до начала исследования информации.

3. Информационный. Осуществляется сбор необходимой статистической информации, значений экономических переменных.

4. Спецификация модели. В математической форме выражаются обнаруженные связи и соотношения, устанавливается состав экзогенных и эндогенных переменных; формируются исходные предпосылки и ограничения модели. Успех эконометрического моделирования зависит от точности выполненной задачи спецификации.

5. Параметризация. Оцениваются параметры (коэффициенты) выбранной зависимости на основе имеющихся статистических данных.

6. Идентификация. Осуществляется статистический анализ модели и оценка ее параметров.

7. Верификация. Проводится проверка адекватности модели, выясняется, насколько удачно решены проблемы спецификации, идентификации, какова точность расчетов по данной модели, насколько соответствует построенная модель реальному экономическому процессу.

Тема 1. Выборки и оценки.

1.1. Генеральная совокупность и выборка. Основные соглашения

Под генеральной совокупностью (ГС) понимают множество элементов произвольной природы, подлежащих обследованию по одному или нескольким признакам. Каждый из признаков является СВ, распределенной по некоторому закону. Примером обследования ГС может служить республиканская перепись населения. В этом примере каждый объект (человек) обследовался по многим признакам (возраст, образование, социальное положение и др.).

Современная МС оформилась и развивалась под давлением практических потребностей. При решении многочисленных практических задач статистики столкнулись с необходимостью изучения полных ГС с большим числом составляющих ее элементов, не имея реальной возможности изучать каждый элемент в отдельности. Например, современное промышленное предприятие, производящее в массовом масштабе различные типы промышленных изделий, не располагает возможностью установить качество производимой продукции, изучая каждое изделие в отдельности при его изготовлении или при передаче потребителю. В подобных случаях установили, что выходом для заинтересованного лица является изучение ограниченного числа элементов – части всей ГС – в надежде, что полученная таким путем информация будет полезной и достаточно точной для того, чтобы судить о качестве всей совокупности изделий.

Во многих случаях, например, при статистическом контроле качества выпускаемой продукции, когда обследование каждого элемента приводит к его полному уничтожению, обследование всей ГС не имеет смысла.

Часть элементов ГС, отобранных для обследования, называется *выборкой*. Количество отобранных элементов называют *объемом выборки*. Обычно n – объем выборки (количество наблюдений). В результате обследования выборки по некоторому признаку X получаем n значений исследуемого признака: x_1, x_2, \dots, x_n .

В МС рассматривают конкретную и математическую выборки. Конкретная выборка представляет собой совокупность из n чисел, с которыми можно производить различные вычислительные операции. Для объяснения понятия математической выборки мысленно представим себе, что из одной и той же ГС произведены другие выборки объема n . В результате обследования выборок получаем таблицу значений (верхний индекс означает номер выборки):

$$\begin{array}{c} x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)} \\ x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)} \\ \dots \\ x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)} \\ \dots \end{array}$$

Из таблицы видно, что i -е число конкретной выборки является значением некоторой СВ X_i , так как в разных выборках на i -м месте могут стоять, вообще говоря, различные числа. В связи с этим под математической выборкой будем понимать систему СВ (X_1, X_2, \dots, X_n) , компоненты которой X_i , $i = 1, 2, \dots, n$ удовлетворяют нижеследующим условиям А) и Б).

А) Компонента X_i имеет такое же распределение, как и признак в ГС.

Б) X_i – взаимно независимые СВ.

Условие А) означает следующее: каждое выборочное значение является опытной реализацией одной и той же СВ X – изучаемого признака ГС. Поэтому нет оснований считать, что условие А) не выполняется. Принятие условия Б) обусловлено удобством различных теоретических выводов. На практике оно не всегда выполнено.

1.2. Способы оценивания и их свойства.

Как правило, на практике мы не знаем точного вероятностного распределения (дискретного признака) или плотности распределения вероятностей (непрерывного признака). Это означает, что неизвестны также и теоретическое среднее и дисперсия. Мы, тем не менее, можем нуждаться в оценках этих или других характеристик ГС.

Обычно способ оценивания представляет собой формулу для оценивания параметров ГС: $f(X_1, X_2, \dots, X_n)$. Более конкретно, *оценкой* параметра θ по выборке X_1, X_2, \dots, X_n называется произвольная функция $\theta^* = \theta^*(X_1, X_2, \dots, X_n)$ от выборочных значений. Так как предполагаемые выборочные значения X_1, X_2, \dots, X_n являются СВ, то и θ^* является СВ. Как только произведена выборка, у нас будет набор чисел x_1, x_2, \dots, x_n , которые будут использованы для получения численной оценки параметра θ . Таким образом, метод оценивания – формула, оценка – число. Если будут извлекаться другие выборки, то метод оценивания будет один и тот же, а оценки могут меняться от выборки к выборке.

Естественно потребовать, чтобы значения оценки по каждой выборке были близки к истинному значению оцениваемого параметра, которое будем обозначать θ и которого мы не знаем. Обычно к оценкам параметров ГС, которые являются случайными величинами, выдвигаются следующие требования.

а1) *Несмещенность*. θ^* называется несмещенной оценкой θ , если

$$M(\theta^*) = \theta. \quad (1.1)$$

а2) *Эффективность*. Можно получить несколько несмещенных оценок одного и того же параметра. Среди них следует выбрать оценку с наименьшей дисперсией, которую называют *эффективной*.

а3) *Состоятельность*. θ^* называется состоятельной оценкой θ , если с увеличением объема выборки оценка θ^* приближается к истинному значению θ . Формально это записывается с помощью *предела по вероятности*: для любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\theta^* - \theta| < \varepsilon) \rightarrow 1 \quad \text{или} \quad \left(\lim_{n \rightarrow \infty} P(|\theta^* - \theta| > \varepsilon) \rightarrow 0 \right). \quad (1.2)$$

Соотношения (1.2) можно записать так же в виде равенства

$$p \lim_{n \rightarrow \infty} \theta^* = \theta,$$

левая часть которого и означает *предел по вероятности*. Отметим, что предел по вероятности обладает свойствами обычного предела функции: постоянную можно выносить за знак предела $p \lim_{n \rightarrow \infty} C \theta^* = C p \lim_{n \rightarrow \infty} \theta^*$; пределы арифметических операций (суммы, произведения, частного нескольких СВ) находятся по обычным формулам.

Очень часто, для проверки состоятельности оценок используются теорема Чебышева (закон больших чисел) и центральная предельная теорема.

Теорема Чебышева (частный случай). Пусть X_1, X_2, \dots – независимые и одинаково распределенные СВ с математическими ожиданиями $M_{X_k} = m$ и дисперсиями $D_{X_k} = \sigma^2$. Тогда

$$p \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = m. \quad (1.3)$$

Центральная предельная теорема. Пусть X_1, X_2, \dots, X_n – независимые и одинаково распределенные СВ с математическими ожиданиями $M_{X_k} = m$ и дисперсиями $D_{X_k} = \sigma^2$. Введем СВ Y_n и Z_n :

$$Y_n = X_1 + \dots + X_n, M_{Y_n} = nm, D_{Y_n} = n\sigma^2, \quad Z_n = \frac{Y_n - nm}{\sigma\sqrt{n}}, M_{Z_n} = 0, D_{Z_n} = 1.$$

Тогда закон распределения СВ Z_n , которую называют нормированной суммой СВ X_1, X_2, \dots, X_n , стремится при $n \rightarrow \infty$ к нормальному закону с параметрами 0 и 1.

1.3. Примеры построения оценок.

Пример 1.1. Оценка математического ожидания m_X признака X в ГС. Пусть X_1, X_2, \dots, X_n – предполагаемая выборка, по которой мы собираемся оценить m_X . По соглашению А) математическое ожидание каждой компоненты X_i равно $M(X_i) = m_X$ ($i = 1, 2, \dots, n$), т. е. каждая выборочная компонента может служить несмещенной оценкой неизвестной величины m_X . Средняя выборочная так же будет несмещенной оценкой m_X . Действительно, по свойствам математического ожидания

$$M(\bar{X}) = M\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n}(M(X_1) + \dots + M(X_n)) = \frac{m_X + \dots + m_X}{n} = m_X.$$

Рассмотрим далее свойство эффективности оценок. Для этого предположим, что найдено несколько несмещенных независимых оценок X_1, X_2, \dots, X_k параметра m_X , дисперсия каждой из которых $D(X_i) = D_X$, т. е. равна дисперсии признака X . Будем искать линейную комбинацию оценок \hat{X} , удовлетворяющую условиям несмещенности и минимальности дисперсии параметра m_X : $\hat{X} = \lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k$. Из требования несмещенности

$$M(\hat{X}) = M(\lambda_1 X_1 + \dots + \lambda_k X_k) = \lambda_1 M(X_1) + \dots + \lambda_k M(X_k) = (\lambda_1 + \lambda_2 + \dots + \lambda_k) m_X \Rightarrow \lambda_1 + \lambda_2 + \dots + \lambda_k = 1.$$

Из условия минимальности дисперсии получаем

$$D(\hat{X}) = D(\lambda_1 X_1 + \dots + \lambda_k X_k) = \lambda_1^2 D(X_1) + \dots + \lambda_k^2 D(X_k) = (\lambda_1^2 + \lambda_2^2 + \dots + \lambda_k^2) D_X \Rightarrow \lambda_1^2 + \lambda_2^2 + \dots + \lambda_k^2 \rightarrow \min$$

В результате получаем задачу на условный экстремум для функции k переменных:

$$f(\lambda) = \lambda_1^2 + \lambda_2^2 + \dots + \lambda_k^2 \rightarrow \min \text{ при условии } g(\lambda) = \lambda_1 + \lambda_2 + \dots + \lambda_k - 1 = 0.$$

Для решения задачи по методу Лагранжа, составляем функцию Лагранжа

$$F(\lambda, \mu) = f(\lambda) + \mu g(\lambda).$$

Необходимые условия экстремума

$$\frac{\partial F}{\partial \lambda_i} = 2\lambda_i + \mu = 0 \quad (i=1,2,\dots,k); \quad \lambda_1 + \lambda_2 + \dots + \lambda_k = 1 \Rightarrow$$

$$\lambda_i = -\frac{\mu}{2}, \quad \mu = -\frac{2}{k} \Rightarrow \lambda_i = \frac{1}{k}.$$

Так как $d^2F = 3\sum_{i=1}^k (\Delta\lambda_i)^2 \geq 0$, то точка $\lambda_i = \frac{1}{k}$ ($i=1,2,\dots,k$) является точкой

условного минимума функции $f(\lambda)$, т.е. оценка $\hat{X} = \lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k$

имеет минимальную дисперсию при условии $\lambda_1 = \lambda_2 = \dots = \lambda_k = \frac{1}{k}$, равную

$$D(\hat{X}) = \frac{D_X}{k}.$$

Проведенные рассуждения позволяют сделать следующее заключение. Дисперсия оценки \bar{X} , равная $D(\bar{X}) = \frac{D_X}{n}$, обладает свойством минимальности, т.е. \bar{X} является эффективной оценкой математического ожидания среди линейных оценок.

Применим теорему Чебышева к выборке X_1, X_2, \dots, X_n :

$$p \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = m_X \Rightarrow p \lim_{n \rightarrow \infty} \bar{X} = m_X. \quad (1.4)$$

Отсюда следует, что оценка \bar{X} является и состоятельной оценкой математического ожидания ГС.

Пример 1.2. Оценка дисперсии D_X признака X в ГС. Как и в примере 1.1, рассмотрим оценку дисперсии D_X с помощью выборочной дисперсии, которая определяется формулой

$$\begin{aligned} D^* &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n \left((X_i - m_X) - (\bar{X} - m_X) \right)^2 = \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - m_X)^2 - \frac{2}{n} (\bar{X} - m_X) \sum_{i=1}^n (X_i - m_X) + \frac{1}{n} (\bar{X} - m_X)^2 \sum_{i=1}^n 1 = \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - m_X)^2 - 2(\bar{X} - m_X) \left(\frac{1}{n} \sum_{i=1}^n X_i - \frac{m_X}{n} \sum_{i=1}^n 1 \right) + (\bar{X} - m_X)^2 \Rightarrow \end{aligned}$$

$$M(D^*) = \frac{1}{n} \sum_{i=1}^n (X_i - m_X)^2 - (\bar{X} - m_X)^2. \quad (1.5)$$

Для проверки несмещенности найдем математическое ожидание оценки:

$$M(D^*) = M\left(\frac{1}{n} \sum_{i=1}^n (X_i - m_X)^2\right) - M\left((\bar{X} - m_X)^2\right) = D_X - \frac{D_X}{n} = \frac{n-1}{n} D_X.$$

Таким образом, оценка генеральной дисперсии с помощью выборочной дисперсии является смещенной. Чтобы получить несмещенную оценку дисперсии, выборочную дисперсию следует умножить на поправочный коэффициент $\frac{n}{n-1}$, т. е. вместо выборочной дисперсии D^* для оценки следует взять исправленную выборочную дисперсию $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

Для проверки состоятельности оценки D^* найдем предел по вероятности от правой части (1.5). Первое слагаемое $\frac{1}{n} \sum_{i=1}^n (X_i - m_X)^2$ является средним арифметическим n независимых, одинаково распределенных СВ $(X_i - m_X)^2$. По закону больших чисел $p \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n (X_i - m_X)^2\right) = M\left(\frac{1}{n} \sum_{i=1}^n (X_i - m_X)^2\right) = D_X$. Для второго слагаемого $p \lim_{n \rightarrow \infty} (\bar{X} - m_X)^2 = \left(p \lim_{n \rightarrow \infty} \bar{X} - p \lim_{n \rightarrow \infty} m_X\right)^2 = (m_X - m_X)^2 = 0$. Отсюда следует, что $p \lim_{n \rightarrow \infty} D^* = D_X$, т. е. D^* является состоятельной оценкой генеральной дисперсии D_X .

Так как $p \lim_{n \rightarrow \infty} \frac{n}{n-1} = p \lim_{n \rightarrow \infty} \frac{1}{1 - 1/n} = 1$, то

$p \lim_{n \rightarrow \infty} S^2 = p \lim_{n \rightarrow \infty} \frac{n}{n-1} D^* = p \lim_{n \rightarrow \infty} D^* = D_X$, т. е. исправленная дисперсия так же является состоятельной оценкой генеральной дисперсии.

Пример 1.3. Оценка ковариации $\sigma_{XY} = \text{cov}(X, Y)$ признаков X, Y в ГС. Ковариация (корреляционный момент) между СВ X и Y определяется как математическое ожидание произведения центрированных СВ:

$$\sigma_{XY} = M((X - m_X)(Y - m_Y)) = \frac{1}{N} \sum_k (x_k - m_X)(y_k - m_Y),$$

где сумма берется по всем элементам ГС, N – объем ГС. Рассмотрим выборку объема n из двумерной ГС: $(X_1; Y_1), (X_2; Y_2), \dots, (X_n; Y_n)$. По основным соглашениям

$$M((X_i - m_X)(Y_j - m_Y)) = 0 \quad (i \neq j), \quad M((X_i - m_X)(Y_i - m_Y)) = \sigma_{XY}. \quad (1.6)$$

Покажем, что несмещенной оценкой ковариации является выборочный корреляционный момент

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}).$$

Предварительно преобразуем это выражение

$$\begin{aligned} S_{XY} &= \frac{1}{n-1} \sum_{i=1}^n \left((X_i - m_X) - (\bar{X} - m_X) \right) \left((Y_i - m_Y) - (\bar{Y} - m_Y) \right) = \\ &= \frac{1}{n-1} \sum_{i=1}^n (X_i - m_X)(Y_i - m_Y) - \frac{\bar{Y} - m_Y}{n-1} \sum_{i=1}^n (X_i - m_X) - \frac{\bar{X} - m_X}{n-1} \sum_{i=1}^n (Y_i - m_Y) + \\ &\quad + \frac{n}{n-1} (\bar{X} - m_X)(\bar{Y} - m_Y). \end{aligned} \quad (1.7)$$

Заменяя $\bar{X} - m_X = \frac{1}{n} \sum_{j=1}^n X_j - n \frac{m_X}{n} = \frac{1}{n} \sum_{j=1}^n (X_j - m_X)$, $\bar{Y} - m_Y = \frac{1}{n} \sum_{j=1}^n (Y_j - m_Y)$,

получаем

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - m_X)(Y_i - m_Y) - \frac{1}{n-1} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (X_i - m_X)(Y_j - m_Y).$$

Из (1.6) следует, что математическое ожидание каждого слагаемого 1-й суммы равно σ_{XY} . Так как в 1-й сумме n слагаемых, то математическое ожидание 1-й суммы равно $\frac{n}{n-1} \sigma_{XY}$. Во второй сумме n^2 слагаемых, причем для n из них (когда $i = j$) математическое

ожидание равно σ_{XY} , а для остальных равно нулю. Тогда математическое

ожидание 2-й суммы равно $\frac{1}{n-1} \frac{1}{n} n \sigma_{XY} = \frac{1}{n-1} \sigma_{XY}$ и

$$M(S_{XY}) = \frac{n}{n-1} n \sigma_{XY} - \frac{1}{n-1} \sigma_{XY} = \sigma_{XY},$$

т. е. S_{XY} является несмещенной оценкой для σ_{XY} .

Состоятельность оценки S_{XY} доказывается точно таким же образом, как и состоятельность S^2 при оценке дисперсии. Выражение $\frac{1}{n} \sum_{i=1}^n (X_i - m_X)(Y_i - m_Y)$ является средним арифметическим n независимых, одинаково распределенных СВ $(X_i - m_X)(Y_i - m_Y)$. По закону больших чисел

$$p \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n (X_i - m_X)(Y_i - m_Y) \right) = M \left(\frac{1}{n} \sum_{i=1}^n (X_i - m_X)(Y_i - m_Y) \right) = \sigma_{XY}.$$

Предел по вероятности 1-й суммы в (1.7) равен $p \lim_{n \rightarrow \infty} \frac{n}{n-1} \sigma_{XY} = \sigma_{XY}$. Пределы по вероятности остальных слагаемых в (1.7) равны нулю.

Тема 2. Модели парной регрессии.

2.1. Модель парной линейной регрессии.

Между двумя СВ X и Y существует два вида зависимости: функциональная и стохастическая (статистическая). Функциональная зависимость встречается, в основном, в области точных наук и здесь не рассматривается. В случае стохастической зависимости каждому значению одной СВ, например, значению x СВ X соответствует СВ Y_x , называемая условным распределением СВ Y при условии $X=x$. СВ X будем считать независимой (*объясняющей*) переменной или *фактором*, а СВ Y – зависимой (*объясняемой*) переменной или *признаком* (*откликом*).

В модели парной линейной регрессии зависимость между СВ X и Y в ГС представляется в виде

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (2.1)$$

где X – неслучайная величина (*объясняющая переменная*), Y и ε – случайные величины Y – зависимая (*объясняемая*) переменная. Наличие в уравнении случайной составляющей (случайного члена) ε связано с воздействием на зависимую переменную неучтенных в уравнении факторов. Постоянные (β_0, β_1) – параметры уравнения. По выборочным данным оценивается выборочное уравнение регрессии

$$\overline{y_x} = b_0 + b_1 x. \quad (2.2)$$

Здесь (b_0, b_1) – оценки параметров (β_0, β_1) . Метод нахождения оценок – метод наименьших квадратов (МНК).

Для того чтобы регрессионный анализ, основанный на МНК давал наилучшие из всех возможных результаты, должны выполняться **условия Гаусса – Маркова**.

1. Объясняющая переменная X есть величина не случайная.
2. Математическое ожидание случайного члена в любом наблюдении должно быть равно нулю, т. е. $M(\varepsilon_i) = 0 \quad (i = \overline{1, n})$, где i – номер наблюдения.
3. Дисперсия случайного члена должна быть постоянной для всех наблюдений, т. е. $D(\varepsilon_i) = M(\varepsilon_i^2) = \sigma^2 = \text{const} \quad (i = \overline{1, n})$.

Условие *независимости* дисперсии случайного члена от номера наблюдения называется **гомоскедачностью**. Зависимость дисперсии случайного члена от номера наблюдения называется **гетероскедачностью**.

4. Случайные члены не коррелированы между собой: $M(\varepsilon_i \varepsilon_j) = 0 \quad (i \neq j)$.
5. Случайный член распределен нормально, т. е. $\varepsilon_i = N(0, \sigma^2)$.

При выполнении условий Гаусса – Маркова модель называется **классической нормальной регрессионной моделью**.

Рассмотрим теперь основы МНК. По выборочным данным $(x_i, y_i), i = \overline{1, n}$ для каждой пары наблюдений определяем остатки e_i как разности между истинными и расчетными значениями $e_i = y_i - (b_0 + b_1 x_i)$. Суть МНК состоит в минимизации суммы квадратов остатков:

$$\delta(b_0, b_1) = \sum_{i=1}^n (e_i)^2 = \sum_{i=1}^n (b_0 + b_1 x_i - y_i)^2 \rightarrow \min.$$

Необходимые условия экстремума приводят к системе линейных уравнений для нахождения b_0, b_1 :

$$\begin{cases} \frac{\partial \delta}{\partial b_0} = 0 \\ \frac{\partial \delta}{\partial b_1} = 0 \end{cases} \Rightarrow \begin{cases} b_0 n + b_1 \sum_i x_i - \sum_i y_i = 0 \\ b_0 \sum_i x_i + b_1 \sum_i x_i^2 - \sum_i x_i y_i = 0 \end{cases}$$

Разделяя каждое из уравнений на n , приходим к системе

$$\begin{cases} b_0 + b_1 \bar{x} = \bar{y} \\ b_0 \bar{x} + b_1 \overline{x^2} = \overline{xy} \end{cases} \text{ где } \begin{cases} \bar{x} = \frac{1}{n} \sum_i x_i, & \bar{y} = \frac{1}{n} \sum_i y_i, \\ \overline{x^2} = \frac{1}{n} \sum_i x_i^2, & \overline{xy} = \frac{1}{n} \sum_i x_i y_i \end{cases} \quad (2.3)$$

Можно предложить несколько способов решения системы (6.3).

1) Из 2-го уравнения вычтем 1-е, умноженное на \bar{x} :

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}, \quad b_0 = \bar{y} - b_1 \bar{x}. \quad (2.4)$$

2) Матричный способ. Для этого вводим три матрицы:

$$\Phi = \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{pmatrix}, \quad B = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}, \quad A = \begin{pmatrix} \bar{y} \\ \overline{xy} \end{pmatrix}.$$

Матрицу Φ называют информационной матрицей, B – матрица-столбец из неизвестных коэффициентов регрессии; A – матрица столбец свободных членов. Отметим, что определитель матрицы Φ , равный $\det \Phi = \overline{x^2} - (\bar{x})^2 = \sigma_x^2$ всегда отличен от нуля (он равен нулю лишь при условии, что $x_i = C = const (i = \overline{1, n})$). Тогда существует обратная матрица Φ^{-1} , причем

$$\Phi^{-1} = \frac{1}{\sigma_x^2} \cdot \begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}, \quad B = \Phi^{-1} \cdot A. \quad (2.5)$$

Замечание 2.1. При построении регрессии «вручную», рекомендуется применять матричный способ для нахождения коэффициентов регрессии, так как элементы обратной матрицы Φ^{-1} используются для исследования регрессии.

3) Построение регрессии на ЭВМ, например, в среде Excel: «Сервис», «Анализ данных», «Регрессия».

Замечание 2.2. Предположим теперь, что в модели парной линейной регрессии зависимость между СВ X и Y в ГС0 представляется в виде

$$Y = \beta_0 + \beta_1 f(X) + \varepsilon, \quad (2.6)$$

где, как и в (2.1), X – неслучайная величина (*объясняющая переменная*), $f(X)$ – известная скалярная функция, Y и ε – случайные величины Y – зависимая (*объясняемая*) переменная. Путем замены $U = f(X)$ зависимость (2.6) сводится к линейной.

Замечание 2.3. Рассмотрим смысл коэффициентов регрессии (2.2). Предположим, что Y – объем выпускаемой продукции (ден. ед.), X – фактор, влияющий на производство (ден. ед.). Тогда увеличение значения фактора на 1 единицу приводит к увеличению объема выпускаемой продукции на b_1 ($b_1 > 0$) ден. ед.

2.2. Исследование парных регрессионных зависимостей.

а) При исследовании линейных регрессионных зависимостей существенную роль играет сумма квадратов остатков

$$\delta(b_0, b_1) = \sum_{i=1}^n (b_0 + b_1 x_i - y_i)^2. \quad (2.8)$$

Представим выражения (2.4) через выборочный коэффициент корреляции $r_{xy} = r$:

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - (\bar{x})^2} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \frac{\sigma_y}{\sigma_x} = r \frac{\sigma_y}{\sigma_x}, \quad b_0 = \bar{y} - b_1 \bar{x}, \Rightarrow \bar{y}_x = \bar{y} + r \frac{\sigma_y}{\sigma_x} (x - \bar{x}).$$

Тогда

$$\begin{aligned} \delta(b_0, b_1) &= \sum_{i=1}^n (b_0 + b_1 x_i - y_i)^2 = \sum_{i=1}^n \left(r \frac{\sigma_y}{\sigma_x} (x_i - \bar{x}) - (y_i - \bar{y}) \right)^2 = \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 - 2r \frac{\sigma_y}{\sigma_x} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) + r^2 \frac{\sigma_y^2}{\sigma_x^2} \sum_{i=1}^n (x_i - \bar{x})^2 = \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 - 2r \frac{\sigma_y}{\sigma_x} n \sigma_{xy} + r^2 \frac{\sigma_y^2}{\sigma_x^2} n \sigma_x^2 = \sum_{i=1}^n (y_i - \bar{y})^2 - r^2 n \sigma_y^2 = \delta(b_0, b_1). \end{aligned}$$

Так как $\sum_{i=1}^n (y_i - \bar{y})^2 = n \sigma_y^2$, то из последнего равенства

$$\delta(b_0, b_1) = n(1 - r^2) \sigma_y^2. \quad (2.9)$$

Учитывая, что $\delta(b_0, b_1) \geq 0$, из (2.9) получаем, что $1 - r^2 \geq 0 \Rightarrow -1 \leq r \leq 1$. Соотношение (2.9) можно записать также в виде

$$n\sigma_y^2 = r^2 n\sigma_y^2 + \delta(b_0, b_1) = r^2 n\sigma_y^2 + (1 - r^2) n\sigma_y^2. \quad (2.10)$$

Таким образом, дисперсия признака Y представлена в виде суммы двух слагаемых. Первое слагаемое $nr^2\sigma_y^2$ в (2.10) будет характеризовать разброс значений признака Y за счет изменения не случайного фактора X . Разделив его на число степеней свободы, равное 1, получим регрессионную дисперсию

$$S_{\text{regr}}^2 = \frac{nr^2\sigma_y^2}{1} = nr^2\sigma_y^2. \quad (2.11)$$

Величина $\delta(b_0, b_1)$ характеризует разброс значений признака Y за счет случайных (неучтенных) факторов. Число степеней свободы $\delta(b_0, b_1)$ равно $(n - 2)$, так как для ее нахождения использовались два параметра b_0, b_1 , вычисленные по выборке. Разделив $\delta(b_0, b_1)$ на число степеней свободы, получим, так называемую, остаточную дисперсию

$$S_{\text{ost}}^2 = \frac{\delta(b_0, b_1)}{n - 2} = \frac{n(1 - r^2)\sigma_y^2}{n - 2}. \quad (2.12)$$

Если дисперсия (2.11) превышает дисперсию (2.12), то регрессионную зависимость (2.2) можно считать адекватной исходным данным. Сравнение дисперсий прово-

дим по критерию Фишера: если $\frac{S_{\text{regr}}^2}{S_{\text{ost}}^2} = \frac{r^2}{1 - r^2}(n - 2) > F_{1, n-2; 1-\alpha}$, где $F_{1, n-2; 1-\alpha}$ квантиль распределения Фишера, соответствующий уровню значимости α , то с вероятностью $1 - \alpha$ $S_{\text{regr}}^2 > S_{\text{ost}}^2$ и регрессию (2.2) можно считать адекватной исходным данным.

Величину $B = r^2 = r_{xy}^2$ называют коэффициентом детерминации. Он определяет долю дисперсии признака Y , объясненную с помощью уравнения регрессии. Чем ближе B к 1, тем качественнее регрессионная модель, то есть регрессионная модель хорошо аппроксимирует исходные данные.

б) Рассмотрим теперь свойства коэффициентов регрессии b_0, b_1 . Так как они определяются по выборочным данным, то являются случайными величинами. Тот факт, что b_0, b_1 являются несмещенными оценками коэффициентов β_0, β_1 модели (2.1), будет доказан ниже при рассмотрении линейных многофакторных моделей.

Здесь мы покажем, что при выполнении предпосылок РА точечной оценкой дисперсии коэффициента $b_j (j = 0, 1)$ является величина

$$\sigma_{b_j}^2 = S_{\text{ost}}^2 \cdot (\Phi^{-1})_{j,j} \cdot \frac{1}{n} \Rightarrow \sigma_{b_j} = \sqrt{S_{\text{ost}}^2 \cdot (\Phi^{-1})_{j,j} \cdot \frac{1}{n}}, \quad (2.13)$$

где $(\Phi^{-1})_{j,j}$ – диагональные элементы матрицы, обратной к информационной. По 3-му

из условий Гаусса–Маркова $D(y) = D(\varepsilon) = \sigma^2 \Rightarrow D(\bar{y}) = \frac{\sigma^2}{n}$. Из (2.4), по свой-

ствам дисперсии

$$\begin{aligned}
D(b_1) &= D\left(\frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}\right) = \frac{1}{\sigma_x^4} \left(D\left(\frac{1}{n} \sum_{i=1}^n x_i y_i\right) - \bar{x}^2 D(\bar{y}) \right) = \frac{1}{\sigma_x^4} \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right) \frac{\sigma^2}{n} = \\
&= \frac{1}{\sigma_x^4} \left(\overline{x^2} - \bar{x}^2 \right) \frac{\sigma^2}{n} = \frac{1}{\sigma_x^4} \sigma_x^2 \frac{\sigma^2}{n} = \frac{1}{\sigma_x^2} \frac{\sigma^2}{n} = (\Phi^{-1})_{1,1} \cdot \frac{\sigma^2}{n} = (\Phi^{-1})_{1,1} \cdot \frac{S_{\text{ost}}^2}{n},
\end{aligned}$$

так как оценкой σ^2 является остаточная дисперсия S_{ost}^2 . Аналогично

$$D(b_0) = D(\bar{y} - b_1 \bar{x}) = \frac{\sigma^2}{n} + \frac{\bar{x}^2}{\sigma_x^2} \frac{\sigma^2}{n} = \frac{\sigma_x^2 + \bar{x}^2}{\sigma_x^2} \frac{\sigma^2}{n} = \frac{\overline{x^2}}{\sigma_x^2} \frac{\sigma^2}{n} = (\Phi^{-1})_{0,0} \cdot \frac{S_{\text{ost}}^2}{n}.$$

Для проверки значимости коэффициентов $b_j (j = 0, 1)$ введем гипотезу H_0 : коэффициент $b_j = 0 (j = 0, 1)$ при альтернативной гипотезе H_1 : $b_j \neq 0 (j = 0, 1)$. Гипотеза H_0 проверяется по t-критерию Стьюдента. Эмпирическое значение t-критерия для коэффициента $b_j (j = 0, 1)$ находится по формуле

$$T_{\text{emp}} = \frac{|b_j|}{\sigma_{b_j}}. \quad (2.14)$$

Теоретическим значением t-критерия является двусторонний квантиль распределения Стьюдента $t_{n-2, 1-\alpha}^{(2)}$. Если $T_{\text{emp}} > t_{n-2, 1-\alpha}^{(2)}$, то гипотеза H_0 отклоняется, т. е. коэффициент b_j значим ($b_j \neq 0$). В противном случае коэффициент b_j следует выбросить из уравнения регрессии, а саму регрессию пересчитать.

Обозначим через $\beta_j (j = 0, 1)$ коэффициенты истинного уравнения регрессии (которых мы не знаем). Для них можно указать доверительные интервалы, т.е. с вероятностью $1-\alpha$

$$\beta_j \in \left(b_j - \sigma_{b_j} \cdot t_{n-2, 1-\alpha}^{(2)}; \quad b_j + \sigma_{b_j} \cdot t_{n-2, 1-\alpha}^{(2)} \right) \quad j = 0, 1. \quad (2.15)$$

Пример 2.1.

Тема 3. Модели множественной регрессии.

3.1. Модель множественной линейной регрессии.

В модели множественной линейной регрессии зависимость объясняемой переменной Y и факторами (объясняющими переменными) X_1, X_2, \dots, X_k в ГС представляется в виде

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon. \quad (3.1)$$

Данную модель можно представить в матричной форме. Для этого введем матрицу-столбец из коэффициентов $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)^T$ и матрицу-строку из объясняющих переменных $X = (1, X_1, X_2, \dots, X_k)$, где $(M)^T$ – операция транспонирования матрицы M . Тогда модель (3.1) запишется в более короткой форме

$$Y = X \cdot \beta + \varepsilon. \quad (3.2)$$

В (3.1), (3.2) Y и ε – случайные величины Y – зависимая (объясняемая) переменная. Наличие в уравнении случайной составляющей (случайного члена) ε связано с воздействием на зависимую переменную неучтенных в уравнении факторов. Матрица-столбец β представляет собой вектор параметров уравнения. По выборочным данным оценивается выборочное уравнение регрессии

$$\overline{y_x} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k = B \cdot x. \quad (3.3)$$

Здесь $B = (b_0, b_1, b_2, \dots, b_k)^T$ оценки параметров $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)^T$. Как и выше, методом нахождения оценок является метод наименьших квадратов (МНК).

Для того чтобы регрессионный анализ, основанный на МНК давал наилучшие из всех возможных результаты, должны выполняться (как и в случае парной регрессии) **условия Гаусса – Маркова**.

1. Все факторы X_1, X_2, \dots, X_k (выборки их значений) являются детерминированными (неслучайными) величинами.

2. Математическое ожидание случайного члена в любом наблюдении должно быть равно нулю, т. е. $M(\varepsilon_i) = 0 \quad (i = \overline{1, n})$, где i – номер наблюдения.

3. Дисперсия случайного члена должна быть постоянной для всех наблюдений, т. е. $D(\varepsilon_i) = M(\varepsilon_i^2) = \sigma^2 = \text{const} \quad (i = \overline{1, n})$ (условие **гомоскедачности**).

4. Случайные члены не коррелированы между собой: $M(\varepsilon_i \varepsilon_j) = 0 \quad (i \neq j)$.

5. Случайный член распределен нормально, т. е. $\varepsilon_i = N(0, \sigma^2)$.

Дополнительно предполагается, что факторы не коррелированы между собой.

При выполнении условий Гаусса – Маркова модель называется **классической нормальной регрессионной моделью**.

Исходными данными для построения модели (3.3) являются значения факторов $x_{i,j}$ ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, k$; $n > k + 1$) и соответствующие им выборочные значения признака $Y : y_1, y_2, \dots, y_n$. Теоретические значения признака Y определяются по уравнению (3.3). Подставляя в это уравнение выборочные данные, получим систему из n уравнений, каждое из которых выполнено приближенно:

$$y_i \approx b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \dots + b_k x_{i,k} \quad (i = 1, 2, \dots, n).$$

Введем остатки e_i :

$$e_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \dots + b_k x_{i,k} - y_i \quad (i = 1, 2, \dots, n).$$

Как и выше, суть МНК состоит в минимизации суммы квадратов остатков

$$\delta(B) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \dots + b_k x_{i,k} - y_i)^2. \quad (3.4)$$

Необходимые условия экстремума $\frac{\partial \delta(B)}{\partial b_j} = 0 \quad (j = 0, 1, 2, \dots, k)$ дают систему $(k+1)$

линейных уравнений с $(k+1)$ неизвестными $b_0, b_1, b_2, \dots, b_k$. Выпишем эту систему

$$\left\{ \begin{array}{l} b_0 \cdot n + b_1 \sum_{i=1}^n x_{i,1} + b_2 \sum_{i=1}^n x_{i,2} + \dots + b_k \sum_{i=1}^n x_{i,k} = \sum_{i=1}^n y_i \\ b_0 \sum_{i=1}^n x_{i,1} + b_1 \sum_{i=1}^n x_{i,1} x_{i,1} + b_2 \sum_{i=1}^n x_{i,2} x_{i,1} + \dots + b_k \sum_{i=1}^n x_{i,k} x_{i,1} = \sum_{i=1}^n y_i x_{i,1} \\ \vdots \\ b_0 \sum_{i=1}^n x_{i,k} + b_1 \sum_{i=1}^n x_{i,1} x_{i,k} + b_2 \sum_{i=1}^n x_{i,2} x_{i,k} + \dots + b_k \sum_{i=1}^n x_{i,k} x_{i,k} = \sum_{i=1}^n y_i x_{i,k} \end{array} \right. \quad (3.5)$$

Разделим каждое уравнение на n и введем обозначения

$$\overline{x_j} = \frac{1}{n} \sum_{i=1}^n x_{i,j}, \quad \overline{x_j x_m} = \frac{1}{n} \sum_{i=1}^n x_{i,j} x_{i,m}, \quad \overline{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \overline{y x_j} = \frac{1}{n} \sum_{i=1}^n y_i x_{i,j}.$$

Тогда система (3.5) перепишется в виде

$$\left\{ \begin{array}{l} \overline{b_0} + \overline{b_1 x_1} + \overline{b_2 x_2} + \dots + \overline{b_k x_k} = \overline{y} \\ \overline{b_0 x_1} + \overline{b_1 x_1 x_1} + \overline{b_2 x_2 x_1} + \dots + \overline{b_k x_k x_1} = \overline{y x_1} \\ \vdots \\ \overline{b_0 x_k} + \overline{b_1 x_1 x_k} + \overline{b_2 x_2 x_k} + \dots + \overline{b_k x_k x_k} = \overline{y x_k} \end{array} \right. \quad (3.6)$$

Решение системы (3.6) можно записать в матричной форме, аналогичной (2.5). Для этого введем три матрицы

$$\Phi = \begin{pmatrix} 1 & \bar{x}_1 & \cdots & \bar{x}_k \\ x_1 & x_1 x_1 & \cdots & x_k x_1 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_k & x_1 x_k & \cdots & x_k x_k \end{pmatrix}, B = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix}, A = \begin{pmatrix} \bar{y} \\ yx_1 \\ \vdots \\ yx_k \end{pmatrix}$$

и систему (3.6) запишем в виде $\Phi \cdot B = A$. Умножая это уравнение слева на Φ^{-1} , получим решение

$$B = \Phi^{-1} \cdot A. \quad (3.7)$$

Как видно, системы (3.5), (3.6) имеют довольно громоздкий вид. Для его упрощения вводят $(n \times (k + 1))$ -матрицу X и транспонированную к ней матрицу X^T :

$$X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,k} \\ 1 & x_{2,1} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{k,1} & \cdots & x_{k,k} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,k} \end{pmatrix}, X^T = \begin{pmatrix} 1 & 1 & \cdots & 1 & \cdots & 1 \\ x_{1,1} & x_{2,1} & \cdots & x_{k,1} & \cdots & x_{n,1} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ x_{1,k} & x_{2,k} & \cdots & x_{k,k} & \cdots & x_{n,k} \end{pmatrix}.$$

Тогда левую часть системы (3.5) можно записать как $X^T X B$, а правую $X^T Y$. В результате получим матричное уравнение $X^T X B = X^T Y$, решение которого запишется в виде

$$B = (X^T X)^{-1} X^T Y. \quad (3.8)$$

Чтобы формулы (3.7), (3.8) имели смысл, определители матриц Φ и $X^T X$ должны быть отличны от нуля. Легко видеть, что $\Phi = \frac{1}{n} X^T X$, т. е. достаточно чтобы $\det(X^T X) \neq 0$. Иногда это требование закладывается в условия Гаусса–Маркова.

3.2. Некоторые нелинейные многофакторные регрессионные модели.

а) Регрессионные модели вида

$$Y = \beta_0 + \beta_1 f_1(X_1, X_2, \dots, X_k) + \dots + \beta_m f_m(X_1, X_2, \dots, X_k) + \varepsilon, \quad (3.9)$$

где

$$f_1(X_1, X_2, \dots, X_k), f_2(X_1, X_2, \dots, X_k), \dots, f_m(X_1, X_2, \dots, X_k)$$

являются известными функциями от факторов, практически ничем не отличаются от модели (3.1). Некоторые однофакторные модели так же записываются в виде (3.9) с той лишь разницей, что f_1, f_2, \dots, f_m являются функциями одной переменной.

б) Для моделирования и решения задач определения объема выпуска продукции в зависимости от капитальных затрат K и затрат труда L часто используют производственную функцию Кобба–Дугласа

$$Y = A \cdot K^{\beta_1} \cdot L^{\beta_2} \cdot \varepsilon, \quad A > 0, 0 < \beta_1 < 1, 0 < \beta_2 < 1. \quad (3.10)$$

Модель (3.10) сводится к (3.1) путем логарифмирования:

$$\ln Y = \beta_0 + \beta_1 \ln K + \beta_2 \ln L + \ln \varepsilon, \quad \beta_0 = \ln A.$$

Коэффициенты β_1, β_2 имеют смысл эластичности:

$$\frac{\partial Y}{Y} / \frac{\partial K}{K} = \beta_1, \quad \frac{\partial Y}{Y} / \frac{\partial L}{L} = \beta_2,$$

т. е. эластичность выпуска продукции по капиталу и труду равна соответственно β_1 и β_2 . Это означает, что увеличение затрат капитала на 1% приведет к росту выпуска продукции на $\beta_1\%$, а увеличение затрат труда на 1% – к росту выпуска продукции на $\beta_2\%$. Дополнительно, при росте масштаба производства в λ раз, т. е. при росте каждого из факторов K и L в λ раз, выпуск возрастает в $\lambda^{\beta_1 + \beta_2}$ раз. Это означает следующее: если $\beta_1 + \beta_2 > 1$, то функция (3.10) имеет *возрастающую* отдачу от масштаба производства; если $\beta_1 + \beta_2 < 1$, то функция (3.10) имеет *убывающую* отдачу от масштаба производства; если $\beta_1 + \beta_2 = 1$, то функция (3.10) имеет постоянную отдачу от масштаба производства.

3.3. Исследование множественных регрессий.

Как и в п. 2.2 под исследованием регрессионных зависимостей понимается проверка построенной регрессии на адекватность исходным данным, установление значимости коэффициентов регрессии и построение доверительных интервалов для них. Приведем (пока без доказательства) основные положения при исследовании.

а) Все исследования проводятся при некотором уровне значимости α . Основную роль при исследовании полученной эмпирической функции регрессии играет сумма квадратов остатков $\delta(B)$, определяемая формулой (3.4) с уже известными коэффициентами регрессии $b_0, b_1, b_2, \dots, b_k$. По сумме $\delta(B)$ строится остаточная дисперсия $S_{\text{ост}}^2$, равная сумме $\delta(B)$ деленной на число степеней свободы. Так как для определения суммы $\delta(B)$ по выборке вычислялись $(k+1)$ коэффициентов регрессии, то число степеней свободы суммы $\delta(B)$ равно $n - (k+1)$, т. е.

$$S_{\text{ост}}^2 = \frac{\delta(B)}{n - (k + 1)}. \quad (3.11)$$

Введем, далее, обозначения

$$\hat{y}_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \dots + b_k x_{i,k} \quad (i = 1, 2, \dots, n) -$$

теоретические значения признака (полученные по уравнению регрессии). Тогда общую сумму квадратов отклонений признака от среднего значения можно представить в виде

$$Q = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 - 2 \sum_{i=1}^n (\hat{y}_i - y_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - y_i)^2.$$

Второе слагаемое равно нулю в силу условий (3.5). Третье слагаемое является суммой квадратов остатков и с его помощью строится остаточная дисперсия (3.11). С помощью первого слагаемого строится регрессионная дисперсия

$$S_{regr}^2 = \frac{1}{k} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2. \quad (3.12)$$

Если дисперсия (3.12) превышает дисперсию (3.11), то регрессионную зависимость (3.2) можно считать адекватной исходным данным. Сравнение дисперсий прово-

дим по критерию Фишера: если $\frac{S_{regr}^2}{S_{ost}^2} > F_{k, n-k-1; 1-\alpha}$, где $F_{k, n-k-1; 1-\alpha}$ квантиль рас-

пределения Фишера, соответствующий уровню значимости α , то с вероятностью $1 - \alpha$ $S_{regr}^2 > S_{ost}^2$ и регрессию (3.2) можно считать адекватной исходным данным.

б) Для исследования адекватности нелинейных регрессионных зависимостей обычно используют *среднюю относительную ошибку аппроксимации*

$$E = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%.$$

Если $E < 10\%$, то регрессионную модель принято считать адекватной исходным данным.

в) Остановимся на свойствах коэффициентов регрессии. Покажем, что вектор B является несмещенной оценкой вектора β . Из (3.8) $B = (X^T X)^{-1} X^T Y$. Заменяя Y по формуле (3.2), будем иметь

$$B = (X^T X)^{-1} X^T (X\beta + \varepsilon) = (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon = \beta + C\varepsilon, \quad (3.13)$$

где $C = (X^T X)^{-1} X^T$. Тогда $M(B) = M(\beta) + CM(\varepsilon) = \beta + 0 = \beta$, так как вектор $\beta = const$, а $M(\varepsilon) = 0$ по условию 2 Гаусса–Маркова. Таким образом, $M(B) = \beta$, что и требовалось показать.

Рассмотрим теперь матрицу дисперсий–ковариаций $M((B - \beta)(B - \beta)^T)$. Диагональными элементами этой матрицы являются дисперсии коэффициентов регрессии $b_0, b_1, b_2, \dots, b_k$. Из (3.13) $B - \beta = (X^T X)^{-1} X^T \varepsilon$.

Несложно показать, что при выполнении предпосылок РА точечной оценкой дисперсии коэффициента $b_j (j = 0, 1, \dots, k)$ является величина

$$\sigma_{b_j}^2 = S_{ost}^2 \cdot ((X^T X)^{-1})_{j,j} = S_{ost}^2 \cdot (\Phi^{-1})_{j,j} \cdot \frac{1}{n} \Rightarrow \sigma_{b_j} = \sqrt{S_{ost}^2 \cdot (\Phi^{-1})_{j,j} \cdot \frac{1}{n}}, \quad (3.14)$$

где $(\Phi^{-1})_{j,j}$ – диагональные элементы матрицы, обратной к информационной.

Для проверки значимости коэффициентов $b_j (j = 0, 1, \dots, k)$ введем гипотезу H_0 : коэффициент $b_j = 0 (j = 0, 1, \dots, k)$ при альтернативной гипотезе H_1 :

$b_j \neq 0$ ($j = 0, 1, \dots, k$). Гипотеза H_0 проверяется по t-критерию Стьюдента. Эмпирическое значение t-критерия для коэффициента b_j ($j = 0, 1, \dots, k$) находится по формуле

$$T_{\text{emp}} = \frac{|b_j|}{\sigma_{b_j}}. \quad (3.14)$$

Теоретическим значением t-критерия является двусторонний квантиль распределения Стьюдента $t_{n-k-1, 1-\alpha}^{(2)}$. Если $T_{\text{emp}} > t_{n-k-1, 1-\alpha}^{(2)}$, то гипотеза H_0 отклоняется, т.е. коэффициент b_j значим ($b_j \neq 0$). В противном случае коэффициент b_j следует выбросить из уравнения регрессии, а саму регрессию пересчитать.

Обозначим через β_j ($j = 0, 1, \dots, k$) коэффициенты истинного уравнения регрессии (которых мы не знаем). Для них можно указать доверительные интервалы, т.е. с вероятностью $1-\alpha$

$$\beta_j \in \left(b_j - \sigma_{b_j} \cdot t_{n-k-1, 1-\alpha}^{(2)}; \quad b_j + \sigma_{b_j} \cdot t_{n-k-1, 1-\alpha}^{(2)} \right) \quad j = 0, 1, \dots, k. \quad (3.15)$$

Тема 4. Моделирование одномерных временных рядов.

Временной ряд – это совокупность значений $Y(t)$ какого либо показателя Y за несколько последовательных моментов времени $t = t_1, t_2, \dots, t_n$. Каждый уровень $Y(t)$ временного ряда формируется под влиянием длительных, кратковременных и случайных факторов. Для удобства временной ряд можно обозначать $Y_t = \{Y_1, Y_2, \dots, Y_n\}$.

Длительные, постоянно действующие факторы оказывают на изучаемое явление определяющее влияние и формируют основную тенденцию – тренд $T(t)$. **Кратковременные**, периодические факторы формируют сезонные колебания ряда $S(t)$. **Случайные** факторы отражаются случайными изменениями уровней ряда $E(t)$.

Модель, в которой временной ряд представлен как сумма перечисленных компонент, т. е. $Y(t) = T(t) + S(t) + E(t)$, называется **аддитивной**.

Модель, в которой временной ряд представлен как произведение перечисленных компонент, т. е. $Y(t) = T(t) \cdot S(t) \cdot E(t)$, называется **мультипликативной**.

Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний. Если амплитуда сезонных колебаний приближенно *постоянная*, используют *аддитивную модель*. Если же амплитуда *возрастает* или *уменьшается*, используют *мультипликативную модель*.

Основная задача эконометрического исследования временного ряда – выявить каждую из перечисленных компонент ряда.

Исследование временного ряда начинают обычно с построения графика. По графику можно определить наличие тренда, сезонных колебаний а так же аддитивность (мультипликативность) модели (см. файл «Исследование Вр ряда.xls», лист 1). Если по графику трудно установить период сезонных колебаний, то это можно сделать вычислением последовательных коэффициентов автокорреляции, т. е. коэффициентов корреляции между рядом Y_t и рядами Y_{t+1}, Y_{t+2}, \dots , которые будем обозначать r_1, r_2, \dots . Если наибольшее значение имеет r_1 , то ряд содержит только тенденцию и не содержит сезонных колебаний. Если же наибольшее значение имеет r_m , то ряд содержит сезонные колебания с периодом m . Например, в файле «Исследование Вр ряда.xls», лист 4

наибольшее значение имеет r_4 , т. е. рассматриваемый временной ряд содержит сезонные колебания с периодом 4 квартала.

Дальнейшее исследование временного ряда проводится по следующим алгоритмам.

АНАЛИЗ АДДИТИВНОЙ МОДЕЛИ

Построение модели $Y = T + S + E$ включает в себя следующие шаги:

- 1) выравнивание исходного ряда методом скользящей средней;
- 2) расчет значений сезонной компоненты;
- 3) устранение сезонной компоненты из исходных уровней ряда ($Y - S$) и получение выровненных данных ($T + E$);
- 4) аналитическое выравнивание уровней ($T + E$), т. е. построение тренда $T(t)$ и расчет значений $T(t)$ с использованием полученного уравнения тренда;
- 5) расчет полученных по модели значений ($T + S$);
- 6) расчет абсолютных ошибок и качества модели.

АНАЛИЗ МУЛЬТИПЛИКАТИВНОЙ МОДЕЛИ

Построение модели $Y = T \cdot S \cdot E$ включает в себя следующие шаги:

- 1) выравнивание исходного ряда методом скользящей средней;
- 2) расчет значений сезонной компоненты;
- 3) устранение сезонной компоненты из исходных уровней ряда (Y / S) и получение выровненных данных ($T \cdot E$);
- 4) аналитическое выравнивание уровней ($T \cdot E$), т. е. построение тренда $T(t)$ и расчет значений $T(t)$ с использованием полученного уравнения тренда;
- 5) расчет полученных по модели значений ($T \cdot S$);
- 6) расчет абсолютных ошибок и качества модели.

Тема 5. Системы одновременных уравнений.

5.1. Основные понятия. Классификация систем одновременных уравнений.

Построение моделей экономических систем очень часто сводится к системам эконометрических соотношений, являющихся уравнениями и тождествами. Уравнение состоит из эндогенных и экзогенных переменных с неопределенными коэффициентами, которые следует определить по выборочным данным. Тождеством называют уравнение, не содержащее случайного члена и в котором все коэффициенты определены.

Наибольшее распространение получили *системы одновременных уравнений*. В таких системах одни и те же эндогенные (зависимые) переменные могут входить и в левую, и в правую части уравнений системы. Ниже буквами y_1, y_2, \dots будут обозначаться эндогенные (зависимые) переменные, x_1, x_2, \dots – экзогенные (независимые) переменные. Общий вид системы одновременных уравнений

$$\begin{cases} y_1 = c_1 + b_{12}y_2 + b_{13}y_3 + \dots + b_{1m}y_m + a_{11}x_1 + a_{12}x_2 + \dots + a_{1k}x_k + \varepsilon_1 \\ y_2 = c_2 + b_{21}y_1 + b_{23}y_3 + \dots + b_{2m}y_m + a_{21}x_1 + a_{22}x_2 + \dots + a_{2k}x_k + \varepsilon_2 \\ \vdots \\ y_m = c_m + b_{m1}y_1 + \dots + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mk}x_k + \varepsilon_k \end{cases} \quad (5.1)$$

Систему (5.1) называют так же *структурной формой модели (структурной моделью)*. Коэффициенты структурной формы модели будем называть *структурными коэффициентами*. Рассмотрим некоторые частные случаи системы (5.1).

1) Система *независимых* уравнений характеризуется тем, что каждая эндогенная переменная выражена только через экзогенные переменные, т. е. коэффициенты b_{ij} в правых частях (5.1) равны нулю. Вид системы независимых уравнений

$$\begin{cases} y_1 = c_1 + a_{11}x_1 + a_{12}x_2 + \dots + a_{1k}x_k + \varepsilon_1 \\ y_2 = c_2 + a_{21}x_1 + a_{22}x_2 + \dots + a_{2k}x_k + \varepsilon_2 \\ \vdots \\ y_m = c_m + a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mk}x_k + \varepsilon_k \end{cases} \quad (5.2)$$

Систему (5.2) называют так же *приведенной формой модели*.

2) Система **рекурсивных** уравнений характеризуется тем, что каждая эндогенная переменная является объясняющей в следующих за ней уравнениях:

$$\left\{ \begin{array}{l} y_1 = c_1 + a_{11}x_1 + a_{12}x_2 + \dots + a_{1k}x_k + \varepsilon_1 \\ y_2 = c_2 + b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \dots + a_{2k}x_k + \varepsilon_2 \\ y_3 = c_3 + b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2 + \dots + a_{3k}x_k + \varepsilon_3 \\ \vdots \\ y_m = c_m + b_{m1}y_1 + \dots + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mk}x_k + \varepsilon_k \end{array} \right. \quad (5.3)$$

Основная задача, связанная с системами одновременных уравнений состоит в определении коэффициентов c_i , b_{ij} , a_{ij} системы (5.1) по выборочным данным. Сложность этой задачи состоит в том, что в общем случае для оценки коэффициентов системы (5.1) неприменим обычный МНК, так как случайные остатки коррелируют с эндогенными переменными. Для приведенной формы модели (5.2) каждое уравнение можно рассматривать самостоятельно и к нему можно применять обычный МНК.

Отметим, что во многих случаях систему (5.1) можно привести к (5.2). При этом коэффициенты приведенной формы модели будут нелинейными функциями структурной формы модели. После оценки коэффициентов приведенной формы по МНК возникает проблема **идентификации**, т. е. задача определения структурных коэффициентов через коэффициенты приведенной формы модели.

5.2. Проблема идентификации.

Возможны следующие ситуации.

1) Структурные коэффициенты однозначно выражаются через коэффициенты приведенной формы модели. В этом случае структурную модель называют **точно идентифицируемой**.

2) Некоторые из структурных коэффициентов не выражаются через коэффициенты приведенной формы модели. Такую структурную модель называют **неидентифицируемой**.

1) Структурные коэффициенты неоднозначно выражаются через коэффициенты приведенной формы модели. Тогда структурную модель называют **сверхидентифицируемой**.

Приведем необходимые и достаточные условия идентифицируемости (сверхидентифицируемости). Эти условия относятся к каждому уравнению структурной модели, т. е. на идентифицируемость проверяется каждое уравнение системы. Если все уравнения системы точно идентифицируемы, то система будет точно идентифицируемой. Если одно или несколько уравнений системы сверхидентифицируемы, а остальные точно идентифицируемы, то и система будет сверхидентифицируемой. Если же хотя бы одно из уравнений неидентифицируемо, то и система неидентифицируема.

Необходимое условие идентификации. Пусть D – число не включенных в уравнение, но присутствующих в системе экзогенных переменных, а G – число включенных в уравнение эндогенных переменных. Если выполнено условие

$$D \geq G - 1, \quad (5.4)$$

то уравнение в структурной модели может быть идентифицировано.

В частности:

1) если $D = G - 1$, то уравнение *точно идентифицируемо* (при выполнении достаточных условий идентификации);

2) если $D > G - 1$, то уравнение *сверхидентифицируемо* (при выполнении достаточных условий идентификации);

3) если же $D < G - 1$, то уравнение неидентифицируемо.

Достаточное условие идентификации. Пусть для рассматриваемого уравнения выполнено необходимое условие идентификации (5.4). Выбросим (мысленно) это уравнение из системы и составим матрицу из коэффициентов при переменных (эндогенных и экзогенных), отсутствующих в исследуемом уравнении. Если ранг полученной матрицы равен $m - 1$, где m – число экзогенных переменных, то рассматриваемое уравнение идентифицируемо.

Пример 5.1. Проверить на идентифицируемость каждое уравнение следующей структурной системы

$$\begin{cases} y_1 = c_1 + b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + \varepsilon_1 \\ y_2 = c_2 + b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + \varepsilon_2 \\ y_3 = c_3 + b_{31}y_1 + b_{32}y_2 + a_{32}x_2 + \varepsilon_3 \end{cases}$$

В данной системе 3 эндогенные переменные y_1, y_2, y_3 и 4 экзогенные переменные x_1, x_2, x_3, x_4 . Предварительно проверим, удовлетворяет ли каждое из уравнений необходимому условию идентификации.

1-е уравнение: $D = 3$, $G = 3$ – отсутствуют три экзогенные переменные x_2, x_3, x_4 , присутствуют три эндогенные переменные y_1, y_2, y_3 . $D > G - 1$, так как $3 > 3 - 1 \Rightarrow$ необходимое условие идентификации выполнено.

2-е уравнение: $D = 1$, $G = 2$ – отсутствует одна экзогенная переменная x_1 , присутствуют две эндогенные переменные y_1, y_2 . $D = G - 1$, так как $1 = 2 - 1 \Rightarrow$ необходимое условие идентификации выполнено.

3-е уравнение: $D = 3$, $G = 3$ – отсутствуют три экзогенные переменные x_1, x_3, x_4 , присутствуют три эндогенные переменные y_1, y_2, y_3 . $D > G - 1$, так как $3 > 3 - 1 \Rightarrow$ необходимое условие идентификации выполнено.

Проверка выполнения достаточных условий идентификации.

1-е уравнение: матрица из коэффициентов при x_2, x_3, x_4 во втором и третьем уравнениях имеет вид

	x_2	x_3	x_4
2 –	a_{22}	a_{23}	a_{24}
3 –	a_{32}	0	0

Минор 2-го порядка $M = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & 0 \end{vmatrix} = -a_{23} \cdot a_{32} \neq 0$, т. е. ранг матрицы равен $2 = 3 - 1$.

Достаточные условия идентификации выполнены для 1-го уравнения. Это уравнение сверхидентифицируемо.

2-е уравнение: матрица из коэффициентов при y_3, x_1 в первом и третьем уравнениях имеет вид

$$\begin{array}{l} 1 - \text{å öðàâíáíèå} \\ 3 - \text{å öðàâíáíèå} \end{array} \begin{array}{cc} y_3 & x_1 \\ b_{13} & a_{11} \\ -1 & 0 \end{array} .$$

Так как определитель этой матрицы $\begin{vmatrix} b_{13} & a_{11} \\ -1 & 0 \end{vmatrix} = a_{11} \neq 0$, то ее ранг равен $2=3-1$. Достаточные условия идентификации выполнены для 2-го уравнения. Это уравнение точно идентифицируемо.

3-е уравнение: матрица из коэффициентов при x_1, x_3, x_4 в первом и втором уравнениях имеет вид

$$\begin{array}{l} 1 - \text{å öðàâíáíèå} \\ 2 - \text{å öðàâíáíèå} \end{array} \begin{array}{ccc} x_1 & x_3 & x_4 \\ a_{11} & 0 & 0 \\ 0 & a_{23} & a_{24} \end{array} .$$

Минор 2-го порядка $M = \begin{vmatrix} a_{11} & 0 \\ 0 & a_{23} \end{vmatrix} = a_{23} \cdot a_{11} \neq 0$, т. е. ранг матрицы равен

$2=3-1$. Достаточные условия идентификации выполнены для 3-го уравнения. Это уравнение сверхидентифицируемо.

Так как каждое уравнение системы идентифицируемо, причем первое и третье уравнения сверхидентифицируемы, то система сверхидентифицируема.

Тема 6. Примеры распределений СВ.

Для приводимых ниже СВ указаны ряды распределения (в случае дискретных СВ), плотность вероятности с функцией распределения (в случае непрерывных СВ), а так же значения числовых характеристик.

6.1. Биномиальное распределение.

Пусть вероятность наступления некоторого события A в каждом из n независимых опытов постоянна и равна p ($0 < p < 1$), $q = 1 - p$ – вероятность не появления события A . Введем СВ X – число наступлений события A в n независимых опытах. Найдем закон распределения и числовые характеристики СВ X .

Очевидно, что множество значений СВ X

$$\Omega_X = \{0, 1, 2, \dots, k, \dots, n\}.$$

Вероятности, с которыми принимаются значения СВ X , определяются по формуле Бернулли:

$$P(X = k) = P_n(k) = C_n^k \cdot p^k \cdot q^{n-k}.$$

Таким образом, СВ X полностью определена. Отметим, что вероятности $P_n(k)$ получаются при разложении выражения $(q+p)^n$ по формуле бинома Ньютона (отсюда и термин «биномиальное распределение»):

$$(q + p)^n = \sum_{k=0}^n C_n^k \cdot q^{n-k} \cdot p^k = \sum_{k=0}^n P_n(k).$$

Числовые характеристики:

$$m_x = np, D(X) = v_2 - v_1^2 = npq, \sigma_x = \sqrt{npq}, v = \sqrt{\frac{q}{np}}.$$

Модой биномиального распределения является число k_0 – *наивероятнейшее число наступлений события A в n независимых испытаниях*. Если $n \cdot p - q$ не является целым числом, то k_0 – целое число, заключенное между $n \cdot p - q$ и $n \cdot p + p$. В частном случае, когда $n \cdot p - q$ является целым числом, биномиальное распределение имеет две моды $n \cdot p - q$ и $n \cdot p + p$. Коэффициенты асимметрии и эксцесса

$$As = \frac{q-p}{\sqrt{npq}}, \varepsilon_x = \frac{1-6pq}{npq}.$$

6.2. Распределение Пуассона.

СВ X имеет распределение Пуассона, если ее возможные значения составляют целые неотрицательные числа $0, 1, 2, \dots, k, \dots$, а соответствующие вероятности выражаются формулой

$$P(X = k) = P_k = \frac{a^k}{k!} e^{-a}, \quad (k = 0, 1, 2, \dots).$$

Числовые характеристики распределения Пуассона:

$$m_x = a, D_x = a, \sigma_x = \sqrt{a}, \nu = \frac{1}{\sqrt{a}}, As = \frac{1}{\sqrt{a}}, \varepsilon_x = \frac{1}{a}.$$

Если параметр a является целым числом, то распределение Пуассона имеет две моды $a-1$ и a . В противном случае Mo есть целое число из промежутка $(a-1, a)$.

6.3. Равномерное распределение.

Будем говорить, что СВ X имеет равномерное распределение на отрезке $[a, b]$ (подчиняется закону $R[a, b]$), если ее плотность вероятности определяется формулой

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b], \end{cases}$$

т. е. равна ненулевой константе на $[a, b]$ и нулю вне $[a, b]$. Числовые характеристики:

$$m_x = \int_{-\infty}^{\infty} xp(x)dx = \int_a^b \frac{x dx}{b-a} = \frac{a+b}{2}, D_x = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{dx}{b-a} = \frac{(b-a)^2}{12},$$

$$\sigma_x = \sqrt{D_x} = \frac{b-a}{2\sqrt{3}}, \mu_3 = 0, \mu_4 = \int_a^b \left(x - \frac{a+b}{2}\right)^4 \frac{dx}{b-a} = \frac{(b-a)^4}{80},$$

$$As = \frac{\mu_3}{\sigma_x^3} = 0, \varepsilon_x = \frac{\mu_4}{\sigma_x^4} - 3 = -1.2, Me = m_x = \frac{a+b}{2}.$$

Равномерное распределение моды не имеет.

6.4. Показательное (экспоненциальное) распределение.

Функция распределения и плотность вероятности показательного распределения:

$$F(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases} \Rightarrow p(t) = F'(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}.$$

Числовые характеристики:

$$m_t = \frac{1}{\lambda}, D_t = \frac{1}{\lambda^2}, \sigma_t = \frac{1}{\lambda}, Mo = 0, Me = \frac{\ln(2)}{\lambda}, As = 2, \varepsilon_x = 6.$$

6.5. Нормальное распределение.

СВ X называется нормальным распределением, если множество ее значений совпадает с множеством действительных чисел, а плотность вероятности определяется формулой

$$p(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2 \cdot \sigma^2}}. \quad (6.1)$$

Параметру a можно придавать любое действительное значение, параметр $\sigma > 0$. На рис. 6.1 изображен график функции (3.1) при значениях параметров $a = 4$, $\sigma = 1$.

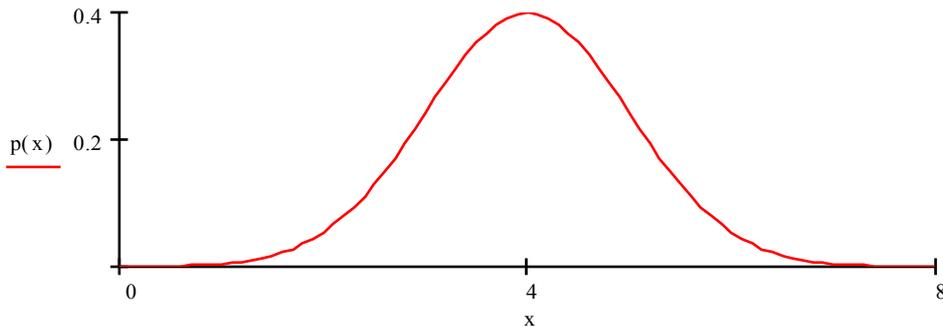


Рис. 6.1

При доказательстве многих предложений, связанных с нормальным распределением, используется интеграл Эйлера – Пуассона

$$J = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 2 \int_0^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}. \quad (6.2)$$

Так как

$$\int_{-\infty}^{\infty} p(x) dx = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2 \cdot \sigma^2}} dx = \left[\begin{array}{l} x-a = t \\ dx = \sigma dt \end{array} \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \frac{\sqrt{2\pi}}{\sqrt{2\pi}} = 1,$$

то основное соотношение для (6.1) выполнено

Найдем числовые характеристики нормального распределения.

$$m_x = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-a)^2}{2 \cdot \sigma^2}} dx = \left[\begin{array}{l} x = a + \sigma t \\ dx = \sigma dt \end{array} \right] = \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}} dt = a.$$

Последний интеграл равен нулю, как интеграл от нечетной функции по симметричному промежутку. Аналогично

$$\mu_2 = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2 \cdot \sigma^2}} dx = \left[\begin{array}{l} x = a + \sigma t \\ dx = \sigma dt \end{array} \right] = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt =$$

$$= \begin{bmatrix} u = t & dv = te^{-\frac{t^2}{2}} dt \\ du & v = -e^{-\frac{t^2}{2}} \end{bmatrix} = \frac{\sigma^2}{\sqrt{2\pi}} \left(-te^{-\frac{t^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \right) = \sigma^2.$$

С помощью той же замены

$$\int_{-\infty}^a p(x) dx = \int_a^{\infty} p(x) dx = \left[\frac{x-a}{\sigma} = t \right] = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{t^2}{2}} dt = \frac{\sqrt{2\pi}}{2\sqrt{2\pi}} = \frac{1}{2}.$$

Таким образом, найдены основные числовые характеристики нормального распределения:

$$M(x) = m_x = a, D(X) = \mu_2 = \sigma^2, \sigma(X) = \sigma, Me(X) = a.$$

Точно так же можно показать, что $\mu_3 = 0, \mu_4 = 3\sigma^4 \Rightarrow As = 0, \varepsilon_x = 0$.

Легко видеть, что функция (6.1) достигает максимума при $x=a$. Отсюда следует, что $Mo(X) = a$.

Функцию распределения нормального закона можно представить в виде

$$F(x) = \int_{-\infty}^x p(t) dt = 0.5 + \Phi\left(\frac{x-a}{\sigma}\right),$$

где $\Phi(x)$ (функция Лапласа, интеграл вероятностей), определяется выражением:

$$\Phi(x) = \frac{1}{\sqrt{2 \cdot \pi}} \int_0^x e^{-t^2/2} dt.$$

Значения функции $\Phi(x)$ можно найти по таблице (приложение 2).

Замечание 6.1. Нормальное распределение с параметрами a и σ обозначают для краткости символом $N(a, \sigma)$.

Вероятность попадания нормально распределенной СВ X в интервал (α, β) определяется формулой

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right). \quad (6.3)$$

Из (6.3)

$$P(|X - a| < \delta) = 2 \cdot \Phi\left(\frac{\delta}{\sigma}\right) \quad (6.4)$$

При $\delta = 3 \cdot \sigma$ из (6.4) получаем, так называемое, *правило трех сигм*:

$$P(|X - a| < 3 \cdot \sigma) = 2 \cdot \Phi(3) = 0.9953 \approx 1,$$

т. е. значения нормально распределенной СВ практически не выходят за пределы интервала $(a - 3 \cdot \sigma, a + 3 \cdot \sigma)$.

Ниже приведены основные (универсальные) распределения, применяемые при проверке статистических гипотез.

6.6. Распределение хи-квадрат (χ^2).

Пусть X_1, X_2, \dots, X_n – независимые в совокупности случайные величины, имеющие нормальное распределение $N(0, 1)$. Сумму квадратов этих СВ называют распределением хи-квадрат с n степенями свободы. Обозначение

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2.$$

Для СВ χ^2 найдено выражение плотности вероятности:

$$p_{\chi^2}(x) = \begin{cases} A_n \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}} & \text{при } x > 0, \\ 0 & \text{при } x \leq 0, \end{cases}$$

где постоянная A_n выбирается из условия нормировки, т. е.

$$\int_0^{\infty} p_{\chi^2}(x) dx = 1.$$

График плотности распределения хи-квадрат при $n=4$ изображен на рис. 6.2.

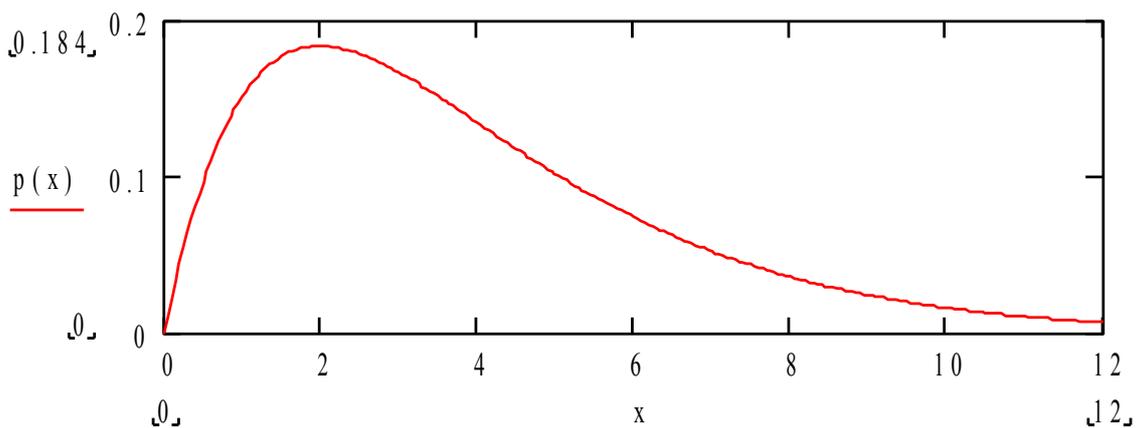


Рис. 6.2

Числовые характеристики распределения хи-квадрат с n степенями свободы:

$$M(\chi^2) = n, D(\chi^2) = 2n, Mo(\chi^2) = n - 2 \quad (n \geq 2).$$

Важнейшим достоинством распределения хи-квадрат является тот факт, что единственный параметр этого распределения – число степеней свободы. На этом факте основаны различные приложения распределения χ^2 .

В приложениях используются, в основном, квантили распределения χ^2 . Отметим, что квантиль K_γ порядка γ любого распределения X определяется как решение одного из уравнений

$$P(X < K_\gamma) = \gamma \quad (\text{или } P(X > K_\gamma) = 1 - \gamma). \quad (6.5)$$

Квантили распределения хи-квадрат с n степенями свободы обозначаются символом $\chi_{n;\gamma}^2$ и находятся по таблице при $\gamma \in \{0,025; 0,05; 0,9; 0,95; 0,975\}$ (приложение 3).

6.7. Распределение Фишера (F-распределение).

Пусть U – распределение хи-квадрат с m степенями свободы, V – распределение хи-квадрат с n степенями свободы. Тогда отношение

$$F = \frac{U \cdot n}{V \cdot m}$$

называется распределением Фишера с числом степеней свободы $(m; n)$ (m – число степеней свободы числителя, n – число степеней свободы знаменателя). Выражение для плотности вероятности F-распределения имеет вид

$$p_F(f) = \begin{cases} C_{m,n} \cdot f^{\frac{m}{2}} \cdot \left(1 + \frac{m}{n} f\right)^{-\frac{m+n}{2}} & \text{и} \ddot{e} f > 0, \\ 0 & \text{и} \ddot{e} f \leq 0, \end{cases}$$

где $C_{m,n}$ выбирается из условия нормировки, т. е. $\int_0^{\infty} p_F(x) dx = 1$.

График плотности распределения Фишера при $m=4$, $n=6$ изображен на рис. 6.3.

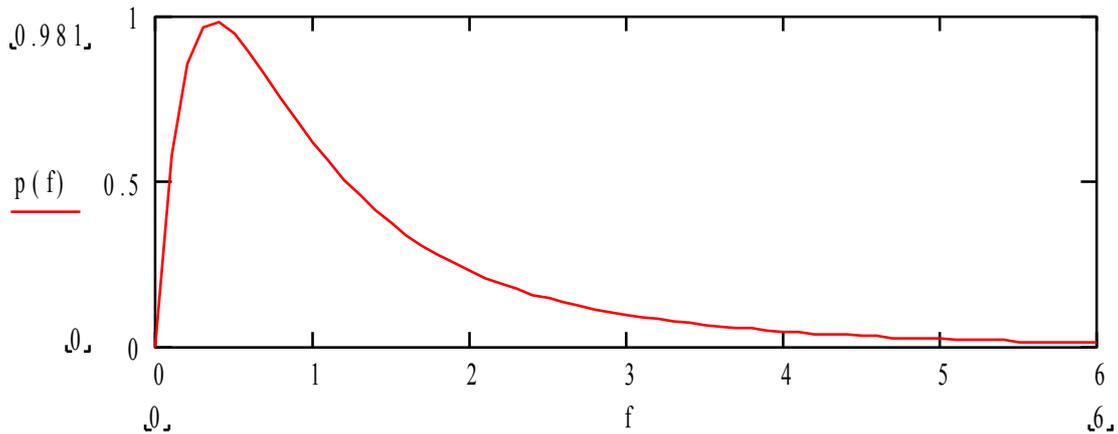


Рис. 6.3

Числовые характеристики F-распределения:

$$M(F) = \frac{n}{n-2}, \quad D(F) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}.$$

Квантили распределения Фишера обозначаются символом $F_{m,n;\gamma}$ и определяются по таблицам (приложение 5).

6.8. Распределение Стьюдента (t-распределение).

Пусть СВ U имеет нормальное распределение с параметрами 0 и 1, т.е. $U = N(0, 1)$, а СВ V имеет распределение хи-квадрат с n степенями свободы. Тогда СВ T , определяемая выражением

$$T = \frac{U}{\sqrt{V}} \cdot \sqrt{n}$$

называется распределением Стьюдента с n степенями свободы. Для распределения Стьюдента найдено выражение плотности вероятности

$$p_T(t) = B_n \cdot \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad t \in (-\infty, +\infty),$$

где постоянная B_n определяется из условия нормировки, т. е

$$\int_{-\infty}^{+\infty} p_T(t) dt = 1.$$

График плотности распределения Стьюдента при $n=6$ изображен на рис. 6. 4.

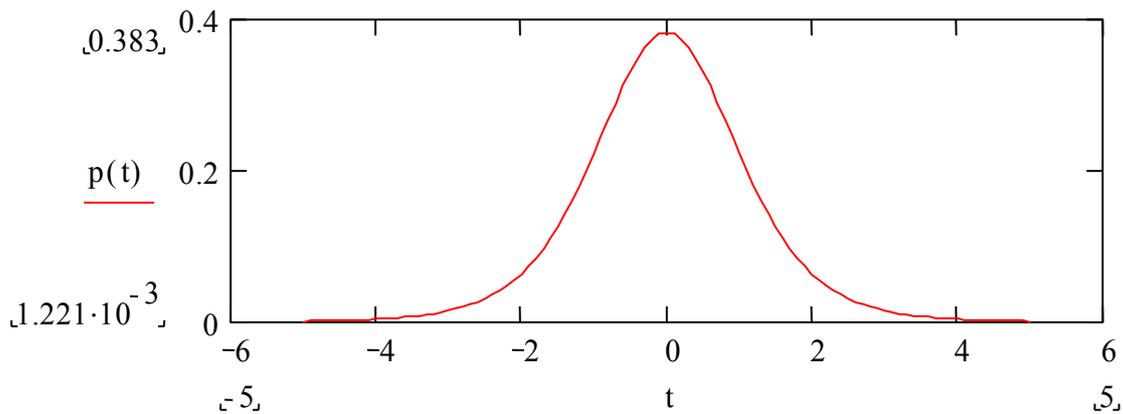


Рис. 6. 4

Числовые характеристики t -распределения с n степенями свободы:

$$M(T) = 0, \quad D(T) = \frac{n}{n-2} \quad (n > 2), \quad Mo(T) = Me(T) = 0.$$

Квантили распределения Стьюдента обозначаются $t_{n; \gamma}$ и находятся по таблице (приложение 4). В связи с четностью функции $p_T(t)$ в приложениях приходится различать односторонние $t_{n; \gamma}$ и двусторонние $t_{n; \gamma}^{(2)}$ квантили. Они определяются как решения следующих уравнений:

$$P\left(T < t_{n; \gamma}\right) = \gamma; \quad P\left(|T| < t_{n; \gamma}^{(2)}\right) = \gamma.$$

Так как $P(T > t_{n; \gamma}) = 1 - \gamma$; $P(T > t_{n; \gamma}^{(2)}) = \frac{1 - \gamma}{2}$, то между $t_{n; \gamma}$ и $t_{n; \gamma}^{(2)}$ легко установить следующую зависимость:

$$t_{n; \gamma}^{(2)} = t_{n; q}, \quad q = 1 - \frac{1 - \gamma}{2} = \frac{1 + \gamma}{2}.$$

В связи с этим для определения квантилей $t_{n; \gamma}^{(2)}$ достаточно иметь подробную таблицу значений квантилей $t_{n; \gamma}$. В приложении 4 собраны как односторонние, так и двусторонние квантили.

Тема 7. Примеры статистических гипотез.

7.1. Понятие о статистических гипотезах и их опытной проверке.

Под статистической гипотезой понимается любое утверждение относительно признака (признаков) в ГС. Это утверждение может касаться вида закона распределения признака в ГС, значений параметров закона распределения, различных соотношений между признаками ГС и др. Выдвигаемую гипотезу называют нулевой и обозначают H_0 . Примеры статистических гипотез: H_0 – признак X в ГС распределен по нормальному закону; H_0 – линейный коэффициент корреляции между признаками X и Y в ГС равен нулю и другие.

Предполагается, что выдвинутую гипотезу H_0 можно проверить статистическими методами. Проверку осуществляют относительно некоторой гипотезы с противоположным содержанием. Ее называют альтернативной гипотезой и обозначают H_1 . Альтернативных гипотез для H_0 может быть несколько. Например, H_0 – значение параметра θ признака в ГС равно 10. К этой гипотезе можно выдвинуть три альтернативных гипотезы: $H_1 - \theta \neq 10$ или $H_1 - \theta > 10$ или $H_1 - \theta < 10$.

Для проверки гипотезы H_0 производят выборку из ГС: x_1, x_2, \dots, x_n . Метод проверки состоит в следующем. На основании интуитивных соображений строится случайная величина W , являющаяся функцией выборочных данных: $W = f(x_1, x_2, \dots, x_n)$. Функцию $f(x_1, x_2, \dots, x_n)$ часто называют *статистикой* или *статистической характеристикой* гипотезы H_0 . Множество Ω_W значений функции W разбивается на две непересекающиеся области Ω_0 и Ω_1 : $\Omega_W = \Omega_0 \cup \Omega_1$ (также на основании интуиции). Если вычисленное по выборке значение функции $W = W_{\text{эмп}}$ попадает в область Ω_0 (*область принятия гипотезы*), то гипотеза H_0 принимается. Другими словами можно сказать, что в этом случае нулевая гипотеза не противоречит выборочным данным. Если же значение $W_{\text{эмп}}$ попадает в область Ω_1 , называемую *критической областью*, то содержание гипотезы H_0 противоречит выборочным данным и ее следует отвергнуть.

Таблица 7.1

Истинные состояния H_0	Принятые решения:	
	H_0 принимается	H_0 отвергается
H_0 верна	Правильное решение	Ошибка первого рода
H_0 неверна	Ошибка второго рода	Правильное решение

При проверке гипотез могут допускаться ошибки. В таблице 4.1 указаны возможные соотношения между истинным состоянием гипотезы и принятым решением.

Ошибка первого рода заключается в том, что отвергается правильная гипотеза H_0 . Вероятность допустить ошибку первого рода называют *уровень значимости* и обозначают α . Эта вероятность совпадает с вероятностью попадания функции W в критическую область, т. е. $\alpha = P(W \in \Omega_1)$. Чтобы уменьшить вероятность ошибки первого рода, область Ω_1 строят таким образом, чтобы вероятность α была не очень большой.

Чаще всего уровень значимости принимается равным 0.1, 0.05, 0.01, 0.001 (или какому-либо значению, заключенному между ними). *Ошибка второго рода* заключается в том, что принимается неверная гипотеза. Отметим, что очень сильное уменьшение уровня значимости ведет, как правило, к увеличению вероятности ошибки второго рода, т. е. одновременная минимизация ошибок первого и второго рода невозможна.

Замечание 7.1. Отметим, что задача построения функции W , которую называют *критерием проверки*, является достаточно сложной. Многие критерии названы именами тех, кто их создал, например, критерий Пирсона, критерий Фишера, критерий Колмогорова и др.

7.2. Гипотезы о соотношениях между дисперсиями.

а) *Проверка гипотезы о равенстве дисперсии нормальной ГС значению σ_0^2 .*

Рассмотрим простейшее приложение распределения хи-квадрат, состоящее в следующем. Пусть из нормально распределенной ГС с признаком $X=N(a, \sigma)$ (a и σ неизвестны) произведена выборка объема n :

$$x_1, x_2, \dots, x_n. \quad (7.1)$$

Находим числовые характеристики выборки (7.1):

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i, \quad S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2, \quad S = \sqrt{S^2}. \quad (7.2)$$

Тогда выражение

$$\frac{(n-1) \cdot S^2}{\sigma^2} \quad (7.3)$$

имеет распределение χ^2 с числом степеней свободы $(n-1)$. На основании этого факта можно создать критерий проверки следующей гипотезы.

Гипотеза $H_0: \sigma^2 = \sigma_0^2$ (дисперсия σ^2 нормального распределения $N(a, \sigma)$ равна σ_0^2) проверяется по выборке (7.1) с характеристиками (7.2) при помощи статистики

$$\chi_{y^{\alpha}}^2 = \frac{(n-1) \cdot S^2}{\sigma_0^2}. \quad (7.4)$$

Если альтернативная гипотеза $H_1: \sigma^2 \neq \sigma_0^2$, то гипотеза H_0 отвергается на уровне значимости α , если $\chi_{y^{\alpha}}^2 < \chi_{n-1; \frac{\alpha}{2}}^2$ или $\chi_{y^{\alpha}}^2 > \chi_{n-1; 1-\frac{\alpha}{2}}^2$ (двусторонний критерий).

Пример 4.1. Для проверки гипотезы $H_0: \sigma^2 = 3$ при уровне значимости $\alpha=0,05$ произведена выборка объема $n=9$ и по выборке вычислена статистическая дисперсия $S^2 = 4,05$. По формуле (7.4)

$$\chi_{y\text{ii}}^2 = \frac{(n-1) \cdot S^2}{\sigma_0^2} = \frac{8 \cdot 4,05}{3} = 10,8.$$

По таблице (приложение 3) находим $\chi_{8;0,025}^2 = 2,18$; $\chi_{8;0,975}^2 = 17,48$. Так как $2,18 < 10,8 < 17,48$, то гипотеза H_0 не противоречит опытным данным.

б) *Сравнение двух дисперсий (F-критерий)*. Простейшее приложение распределения Фишера состоит в следующем: пусть X и Y – нормально распределенные признаки двух генеральных совокупностей. Произведем выборки из этих ГС объемами m и n . По выборкам найдем статистические дисперсии S_x^2 и S_y^2 . Тогда отношение большей дис-

персии к меньшей, например, $\frac{S_x^2}{S_y^2}$ имеет распределение Фишера с числом степеней свободы $(m-1, n-1)$.

Гипотеза $H_0: \sigma_x^2 = \sigma_y^2$ проверяется по результатам двух независимых выборок объемами m и n из нормально распределенных ГС с неизвестными МО, по которым вычисляются статистические дисперсии S_x^2 и S_y^2 . Статистика для проверки гипотезы H_0 строится на основе простейшего приложения F-распределения. Для этого находим отношение большей дисперсии к меньшей, которое имеет F-распределение с числом степеней свободы $(m-1, n-1)$:

$$F_{y\text{ii}} = \frac{S_x^2}{S_y^2} (S_x^2 \geq S_y^2)$$

Гипотеза H_0 отвергается на уровне значимости α при альтернативной гипотезе $H_1: \sigma_x^2 > \sigma_y^2$, если $F_{y\text{ii}} > F_{m-1, n-1; 1-\alpha}$.

в) *Критерий Кочрена*. Введем гипотезу H_0 : дисперсии $\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2$ ($N \geq 2$) N нормально распределенных генеральных совокупностей равны между собой (однородны). Альтернативная гипотеза H_1 : дисперсии неоднородны. Для проверки гипотезы по критерию Кочрена из каждой ГС произведем выборку одного и того же объема m : $x_{i,1}, x_{i,2}, \dots, x_{i,m}$; $i = 1, 2, \dots, N$. Вычислим статистические дисперсии

$$\bar{x}_i = \frac{1}{m} \sum_{k=1}^m x_{i,k}, \quad S_i^2 = \frac{1}{m-1} \sum_{k=1}^m (x_{i,k} - \bar{x}_i)^2.$$

Статистикой для проверки гипотезы является отношение максимальной из статистических дисперсий к сумме всех статистических дисперсий:

$$C_{\hat{a}i\ddot{i}} = \frac{S_{\max}^2}{\sum_{i=1}^N S_i^2}.$$

По таблице критических значений распределения Кочрена при заданном уровне значимости α находим $C_{N,m-1,\alpha}$. Гипотеза H_0 отвергается на уровне значимости α , если $C_{\hat{a}i\ddot{i}} > C_{N,m-1,\alpha}$.

Замечание 7.2. Если объемы выборок различны, то для проверки гипотезы об однородности дисперсий применяется критерий Бартлета. Мы не рассматриваем этот критерий в связи с его громоздкостью.

7.3. Гипотезы о сравнении средних.

а) *Проверка гипотезы о среднем значении нормально распределенной ГС (t – критерий)*. Простейшее приложение распределения Стьюдента состоит в следующем. Пусть из нормально распределенной ГС с признаком $X=N(a, \sigma)$ (a и σ неизвестны) произведена выборка (7.1) с числовыми характеристиками (7.2). Тогда выражение

$$\frac{\bar{X}-a}{S} \cdot \sqrt{n}$$

имеет распределение Стьюдента с $(n-1)$ степенями свободы.

Введем гипотезу $H_0: a = a_0$ (среднее значение a нормального распределения $N(a, \sigma)$ равно a_0 , σ неизвестно). Гипотеза H_0 проверяется по выборке (2.1) с характеристиками (2.2) при помощи статистики $T_{y\ddot{i}i} = \frac{\bar{x}-a_0}{S} \sqrt{n}$, которая имеет распределение Стьюдента с $n-1$ степенями свободы. Правило проверки гипотезы: H_0 отвергается на уровне значимости α при альтернативной гипотезе $H_1: a \neq a_0$ (двусторонний критерий), если $|T_{\text{эмп}}| > t_{n-1; 1-\frac{\alpha}{2}}^{(2)} = t_{n-1; 1-\frac{\alpha}{2}}$.

б) *Критерий сравнения средних в двух нормально распределенных ГС («двухвыборочный t-критерий»)*. Рассмотрим гипотезу $H_0: a_x = a_y$ о равенстве средних значений a_x и a_y двух ГС $X=N(a_x, \sigma_x)$ и $Y=N(a_y, \sigma_y)$ с неизвестными, но равными дисперсиями $\sigma_x^2 = \sigma_y^2$. Гипотеза H_0 проверяется по результатам двух независимых выборок объемами n и m , по которым вычисляются средние значения \bar{x} и \bar{y} , а так же статистические дисперсии S_x^2 и S_y^2 . Статистика для проверки гипотезы H_0 имеет вид

$$T_{y\ddot{i}i} = \frac{\bar{x}-\bar{y}}{\sqrt{(n-1)S_x^2 + (m-1)S_y^2}} \sqrt{\frac{n \cdot m}{n+m} \cdot (n+m-2)}. \quad (7.5)$$

При справедливости гипотезы H_0 статистика (4.5) имеет распределение Стьюдента с $n+m-2$ степенями свободы. Правило применения (двустороннего) критерия заклю-

чается в следующем: гипотеза H_0 отвергается на уровне значимости α при альтернативной гипотезе $H_1: a_x \neq a_y$, если $|T_{y\bar{y}}| > t_{n+m-2; 1-\alpha}^{(2)} = t_{n+m-2; 1-\frac{\alpha}{2}}$.

Замечание 7.3. Предположение о равенстве дисперсий $\sigma_x^2 = \sigma_y^2$ можно проверить с помощью F- критерия. Если $\sigma_x^2 \neq \sigma_y^2$ или если гипотеза « $\sigma_x^2 = \sigma_y^2$ » отклоняется при проверке ее с помощью F- критерия, то вместо (7.5) нужно использовать статистику

$$T_{y\bar{y}} = \frac{\bar{x} - \bar{y}}{\sqrt{n \cdot S_x^2 + m \cdot S_y^2}} \sqrt{n \cdot m}. \quad (7.6)$$

Гипотеза « $a_x = a_y$ » отвергается на уровне значимости α при альтернативной гипотезе $H_1: a_x \neq a_y$, если

$$|T_{y\bar{y}}| > t_{k; 1-\alpha}^{(2)} = t_{k; 1-\frac{\alpha}{2}}.$$

При этом для статистики (7.6) число степеней свободы k – наибольшее целое число, не превосходящее

$$\frac{(n-1)(m-1)(mS_x^2 + nS_y^2)^2}{(m-1)(mS_x^2)^2 + (n-1)(nS_y^2)^2}.$$

Пример 7.2. По результатам двух выборок объемами $n_x=25$ и $n_y=36$ из нормально распределенных ГС вычислены средние значения и статистические дисперсии:

$$\bar{x} = 25,31; \quad S_x^2 = 66,23; \quad \bar{y} = 21,8; \quad S_y^2 = 34,11.$$

Проверить гипотезу H_0 о равенстве средних значений.

Решение. Предварительно проверим гипотезу о равенстве дисперсий. Для этого находим $F_{y\bar{y}} = \frac{S_x^2}{S_y^2} = \frac{66,23}{34,11} = 1,94$. По таблице квантиль $F_{24,35; 0,95} = 1,83$. Так как $1,94 >$

$1,83$, то на уровне значимости $\alpha=0,05$ гипотезу о равенстве дисперсий следует отвергнуть. В связи с этим для проверки гипотезы H_0 выбираем статистику (4.6) ($m=25$, $n=36$):

$$T_{y\bar{y}} = \frac{|\bar{x} - \bar{y}|}{\sqrt{n \cdot S_x^2 + m \cdot S_y^2}} \sqrt{n \cdot m} = \frac{|25,31 - 21,8| \cdot \sqrt{25 \cdot 36}}{\sqrt{36 \cdot 66,23 + 25 \cdot 34,11}} = 1,85.$$

По формуле (4.7) определяем число степеней свободы:

$$\frac{(36-1)(25-1)(25 \cdot 66,23 + 36 \cdot 34,11)^2}{(25-1)(25 \cdot 66,23)^2 + (36-1)(36 \cdot 34,11)^2} = 58,911 \Rightarrow k = 58.$$

По таблице двусторонних квантилей распределения Стьюдента находим $t_{58; 0,95}^{(2)} \approx 2,002$. Так как $1,85 < 2,002$, то гипотезу H_0 о равенстве средних значений можно принять.

Приложения.
 Приложение 1. Таблица значений плотности вероятности
 нормального закона $N(0;1)$.

x	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.3989	.3989	.3989	.3988	.3986	.3984	.3982	.3980	.3977	.3973
0.1	.3970	.3965	.3961	.3956	.3951	.3945	.3939	.3932	.3925	.3918
0.2	.3910	.3902	.3894	.3885	.3876	.3867	.3857	.3847	.3836	.3825
0.3	.3814	.3802	.3790	.3778	.3765	.3752	.3739	.3725	.3712	.3697
0.4	.3683	.3668	.3653	.3637	.3621	.3605	.3589	.3572	.3555	.3538
0.5	.3521	.3503	.3485	.3467	.3448	.3429	.3410	.3391	.3372	.3352
0.6	.3332	.3312	.3292	.3271	.3251	.3230	.3209	.3187	.3166	.3144
0.7	.3123	.3101	.3079	.3056	.3034	.3011	.2989	.2966	.2943	.2920
0.8	.2897	.2874	.2850	.2827	.2803	.2780	.2756	.2732	.2709	.2685
0.9	.2661	.2637	.2613	.2589	.2565	.2541	.2516	.2492	.2468	.2444
1.0	.2420	.2396	.2371	.2347	.2323	.2299	.2275	.2251	.2227	.2203
1.1	.2179	.2155	.2131	.2107	.2083	.2059	.2036	.2012	.1989	.1965
1.2	.1942	.1919	.1895	.1872	.1849	.1826	.1804	.1781	.1758	.1736
1.3	.1714	.1691	.1669	.1647	.1626	.1604	.1582	.1561	.1539	.1518
1.4	.1497	.1476	.1456	.1435	.1415	.1394	.1374	.1354	.1334	.1315
1.5	.1295	.1276	.1257	.1238	.1219	.1200	.1182	.1163	.1145	.1127
1.6	.1109	.1092	.1074	.1057	.1040	.1023	.1006	.0989	.0973	.0957
1.7	.0940	.0925	.0909	.0893	.0878	.0863	.0848	.0833	.0818	.0804
1.8	.0790	.0775	.0761	.0748	.0734	.0721	.0707	.0694	.0681	.0669
1.9	.0656	.0644	.0632	.0620	.0608	.0596	.0584	.0573	.0562	.0551
2.2	.0540	.0529	.0519	.0508	.0498	.0488	.0478	.0468	.0459	.0449
2.1	.0440	.0431	.0422	.0413	.0404	.0396	.0387	.0379	.0371	.0363
2.2	.0355	.0347	.0339	.0332	.0325	.0317	.0310	.0303	.0297	.0290
2.3	.0283	.0277	.0270	.0264	.0258	.0252	.0246	.0241	.0235	.0229
2.4	.0224	.0219	.0213	.0208	.0203	.0198	.0194	.0189	.0184	.0180
2.5	.0175	.0171	.0167	.0163	.0158	.0154	.0151	.0147	.0143	.0139

2.6	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110	.0107
2.7	.0104	.0101	.0099	.0096	.0093	.0091	.0088	.0086	.0084	.0081
2.8	.0079	.0077	.0075	.0073	.0071	.0069	.0067	.0065	.0063	.0061
2.9	.0060	.0058	.0056	.0055	.0053	.0051	.0050	.0048	.0047	.0046
3.0	.0044	.0043	.0042	.0040	.0039	.0038	.0037	.0036	.0035	.0034
3.1	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026	.0025	.0025
3.2	.0024	.0023	.0022	.0022	.0021	.0020	.0020	.0019	.0018	.0018
3.3	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014	.0013	.0013
3.4	.0012	.0012	.0012	.0011	.0011	.0010	.0010	.0010	.0009	.0009
3.5	.0009	.0008	.0008	.0008	.0008	.0007	.0007	.0007	.0007	.0006
3.6	.0006	.0006	.0006	.0005	.0005	.0005	.0005	.0005	.0005	.0004
3.7	.0004	.0004	.0004	.0004	.0004	.0004	.0003	.0003	.0003	.0003
3.8	.0003	.0003	.0003	.0003	.0003	.0002	.0002	.0002	.0002	.0002
3.9	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001
4.0	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.00009	.00009

Приложение 2. Таблица значений функции $\Phi(x)$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0	0.45	0.1736	0.90	0.3159	1.35	0.4115	1.80	0.4641
0.01	0.0040	0.46	0.1772	0.91	0.3186	1.36	0.4131	1.81	0.4649
0.02	0.0080	0.47	0.1808	0.92	0.3212	1.37	0.4147	1.82	0.4656
0.03	0.0120	0.48	0.1844	0.93	0.3238	1.38	0.4162	1.83	0.4664
0.04	0.0160	0.49	0.1879	0.94	0.3264	1.39	0.4177	1.84	0.4671
0.05	0.0199	0.50	0.1915	0.95	0.3289	1.40	0.4192	1.85	0.4678
0.06	0.0239	0.51	0.195	0.96	0.3315	1.41	0.4207	1.86	0.4686
0.07	0.0279	0.52	0.1985	0.97	0.334	1.42	0.4222	1.87	0.4693
0.08	0.0319	0.53	0.2019	0.98	0.3365	1.43	0.4236	1.88	0.4699
0.09	0.0359	0.54	0.2054	0.99	0.3389	1.44	0.4251	1.89	0.4706
0.10	0.0398	0.55	0.2088	1.00	0.3413	1.45	0.4265	1.90	0.4713
0.11	0.0438	0.56	0.2123	1.01	0.3438	1.46	0.4279	1.91	0.4719
0.12	0.0478	0.57	0.2157	1.02	0.3461	1.47	0.4292	1.92	0.4726
0.13	0.0517	0.58	0.219	1.03	0.3485	1.48	0.4306	1.93	0.4732
0.14	0.0557	0.59	0.2224	1.04	0.3508	1.49	0.4319	1.94	0.4738
0.15	0.0596	0.60	0.2257	1.05	0.3531	1.50	0.4332	1.95	0.4744
0.16	0.0636	0.61	0.2291	1.06	0.3554	1.51	0.4345	1.96	0.475
0.17	0.0675	0.62	0.2324	1.07	0.3577	1.52	0.4357	1.97	0.4756
0.18	0.0714	0.63	0.2357	1.08	0.3599	1.53	0.437	1.98	0.4761
0.19	0.0754	0.64	0.2389	1.09	0.3621	1.54	0.4382	1.99	0.4767
0.22	0.0793	0.65	0.2422	1.10	0.3643	1.55	0.4394	2.00	0.4772
0.21	0.0832	0.66	0.2454	1.11	0.3665	1.56	0.4406	2.01	0.4778
0.22	0.0871	0.67	0.2486	1.12	0.3686	1.57	0.4418	2.02	0.4783
0.23	0.0910	0.68	0.2517	1.13	0.3708	1.58	0.4429	2.03	0.4788
0.24	0.0948	0.69	0.2549	1.14	0.3729	1.59	0.4441	2.04	0.4793
0.25	0.0987	0.70	0.258	1.15	0.3749	1.60	0.4452	2.05	0.4798
0.26	0.1026	0.71	0.2611	1.16	0.377	1.61	0.4463	2.06	0.4803
0.27	0.1064	0.72	0.2642	1.17	0.379	1.62	0.4474	2.07	0.4808
0.28	0.1103	0.73	0.2673	1.18	0.381	1.63	0.4484	2.08	0.4812

0.29	0.1141	0.74	0.2704	1.19	0.383	1.64	0.4495	2.09	0.4817
0.30	0.1179	0.75	0.2734	1.20	0.3849	1.65	0.4505	2.10	0.4821
0.31	0.1217	0.76	0.2764	1.21	0.3869	1.66	0.4515	2.11	0.4826
0.32	0.1255	0.77	0.2794	1.22	0.3888	1.67	0.4525	2.12	0.483
0.33	0.1293	0.78	0.2823	1.23	0.3907	1.68	0.4535	2.13	0.4834
0.34	0.1331	0.79	0.2852	1.24	0.3925	1.69	0.4545	2.14	0.4838
0.35	0.1368	0.80	0.2881	1.25	0.3944	1.70	0.4554	2.15	0.4842
0.36	0.1406	0.81	0.291	1.26	0.3962	1.71	0.4564	2.16	0.4846
0.37	0.1443	0.82	0.2939	1.27	0.398	1.72	0.4573	2.17	0.485
0.38	0.148	0.83	0.2967	1.28	0.3997	1.73	0.4582	2.18	0.4854
0.39	0.1517	0.84	0.2995	1.29	0.4015	1.74	0.4591	2.19	0.4857
0.40	0.1554	0.85	0.3023	1.30	0.4032	1.75	0.4599	2.20	0.4861
0.41	0.1591	0.86	0.3051	1.31	0.4049	1.76	0.4608	2.21	0.4864
0.42	0.1628	0.87	0.3078	1.32	0.4066	1.77	0.4616	2.22	0.4868
0.43	0.1664	0.88	0.3106	1.33	0.4082	1.78	0.4625	2.23	0.4871
0.44	0.1700	0.89	0.3133	1.34	0.4099	1.79	0.4633	2.24	0.4875

Таблица значений функции $\Phi(x)$ (продолжение).

x	$\Phi(x)$								
2.25	0.4878	2.70	0.4965	3.15	0.4992	3.60	0.499841	4.05	0.499974
2.26	0.4881	2.71	0.4966	3.16	0.4992	3.61	0.499847	4.06	0.499975
2.27	0.4884	2.72	0.4967	3.17	0.4992	3.62	0.499853	4.07	0.499976
2.28	0.4887	2.73	0.4968	3.18	0.4993	3.63	0.499858	4.08	0.499977
2.29	0.489	2.74	0.4969	3.19	0.4993	3.64	0.499864	4.09	0.499978
2.30	0.4893	2.75	0.4970	3.20	0.4993	3.65	0.499869	4.10	0.499979
2.31	0.4896	2.76	0.4971	3.21	0.4993	3.66	0.499874	4.11	0.499980
2.32	0.4898	2.77	0.4972	3.22	0.4994	3.67	0.499879	4.12	0.499981
2.33	0.4901	2.78	0.4973	3.23	0.4994	3.68	0.499883	4.13	0.499982
2.34	0.4904	2.79	0.4974	3.24	0.4994	3.69	0.499888	4.14	0.499983
2.35	0.4906	2.80	0.4974	3.25	0.4994	3.70	0.499892	4.15	0.499983
2.36	0.4909	2.81	0.4975	3.26	0.4994	3.71	0.499896	4.16	0.499984
2.37	0.4911	2.82	0.4976	3.27	0.4995	3.72	0.499900	4.17	0.499985
2.38	0.4913	2.83	0.4977	3.28	0.4995	3.73	0.499904	4.18	0.499985
2.39	0.4916	2.84	0.4977	3.29	0.4995	3.74	0.499908	4.19	0.499986
2.40	0.4918	2.85	0.4978	3.30	0.4995	3.75	0.499912	4.20	0.499987
2.41	0.492	2.86	0.4979	3.31	0.4995	3.76	0.499915	4.21	0.499987
2.42	0.4922	2.87	0.4979	3.32	0.4995	3.77	0.499918	4.22	0.499988
2.43	0.4925	2.88	0.4980	3.33	0.4996	3.78	0.499922	4.23	0.499988
2.44	0.4927	2.89	0.4981	3.34	0.4996	3.79	0.499925	4.24	0.499989
2.45	0.4929	2.90	0.4981	3.35	0.4996	3.80	0.499928	4.25	0.499989
2.46	0.4931	2.91	0.4982	3.36	0.4996	3.81	0.499931	4.26	0.499990
2.47	0.4932	2.92	0.4982	3.37	0.4996	3.82	0.499933	4.27	0.499990
2.48	0.4934	2.93	0.4983	3.38	0.4996	3.83	0.499936	4.28	0.499991
2.49	0.4936	2.94	0.4984	3.39	0.4997	3.84	0.499938	4.29	0.499991
2.50	0.4938	2.95	0.4984	3.40	0.4997	3.85	0.499941	4.30	0.499991
2.51	0.494	2.96	0.4985	3.41	0.4997	3.86	0.499943	4.31	0.499992
2.52	0.4941	2.97	0.4985	3.42	0.4997	3.87	0.499946	4.32	0.499992
2.53	0.4943	2.98	0.4986	3.43	0.4997	3.88	0.499948	4.33	0.499993

2.54	0.4945	2.99	0.4986	3.44	0.4997	3.89	0.499950	4.34	0.499993
2.55	0.4946	3.00	0.4987	3.45	0.4997	3.90	0.499952	4.35	0.499993
2.56	0.4948	3.01	0.4987	3.46	0.4997	3.91	0.499954	4.36	0.499993
2.57	0.4949	3.02	0.4987	3.47	0.4997	3.92	0.499956	4.37	0.499994
2.58	0.4951	3.03	0.4988	3.48	0.4997	3.93	0.499958	4.38	0.499994
2.59	0.4952	3.04	0.4988	3.49	0.4998	3.94	0.499959	4.39	0.499994
2.60	0.4953	3.05	0.4989	3.50	0.4998	3.95	0.499961	4.40	0.499995
2.61	0.4955	3.06	0.4989	3.51	0.4998	3.96	0.499963	4.41	0.499995
2.62	0.4956	3.07	0.4989	3.52	0.4998	3.97	0.499964	4.42	0.499995
2.63	0.4957	3.08	0.4990	3.53	0.4998	3.98	0.499966	4.43	0.499995
2.64	0.4959	3.09	0.4990	3.54	0.4998	3.99	0.499967	4.44	0.499996
2.65	0.496	3.10	0.4990	3.55	0.4998	4.00	0.499968	4.45	0.499996
2.66	0.4961	3.11	0.4991	3.56	0.4998	4.01	0.499970	4.46	0.499996
2.67	0.4962	3.12	0.4991	3.57	0.4998	4.02	0.499971	4.47	0.499996
2.68	0.4963	3.13	0.4991	3.58	0.4998	4.03	0.499972	4.48	0.499996
2.69	0.4964	3.14	0.4992	3.59	0.4998	4.04	0.499973	4.49	0.499996

Приложение 3. Квантили $\chi^2_{v;P}$ распределения χ^2 .

P	0.025	0.05	0.9	0.95	0.975	P	0.025	0.05	0.9	0.95	0.975
v						v					
1	$0.9 \cdot 10^{-3}$	$0.4 \cdot 10^{-2}$	2.696	3.856	5.035	51	33.22	35.63	64.31	68.56	72.73
2	0.0506	0.103	4.593	5.991	7.367	52	34.08	36.41	65.47	69.81	73.68
3	0.216	0.352	6.249	7.795	9.309	53	34.78	37.27	66.55	71.04	75
4	0.484	0.711	7.759	9.464	11.11	54	35.63	38.07	67.65	72.16	76.09
5	0.831	1.15	9.234	11.06	12.82	55	36.35	38.97	68.79	73.34	77.32
6	1.24	1.64	10.64	12.57	14.42	56	37.2	39.8	69.94	74.45	78.39
7	1.65	2.17	12.02	14.02	15.97	57	37.94	40.61	71.04	75.6	79.72
8	2.18	2.73	13.36	15.45	17.48	58	38.89	41.47	72.18	76.76	80.85
9	2.70	3.33	14.69	16.87	18.98	59	39.63	42.35	73.25	77.85	81.96
10	3.25	4.57	15.98	18.29	20.46	60	40.46	43.18	74.39	79.07	83.27
11	3.823	4.578	17.28	19.72	21.86	61	41.44	44.04	75.47	80.34	84.43
12	4.418	5.239	18.53	21.05	23.35	62	42.18	44.96	76.59	81.39	85.75
13	5.009	5.889	19.83	22.35	24.78	63	43.06	45.7	77.74	82.55	86.85
14	5.6	6.576	21.08	23.63	26	64	43.78	46.59	78.89	83.62	88.06
15	6.212	7.252	22.3	24.98	27.46	65	44.64	47.4	79.98	84.81	89.12
16	6.893	7.938	23.51	26.29	28.77	66	45.4	48.32	81.1	86	90.34
17	7.573	8.696	24.76	27.57	30.16	67	46.25	49.23	82.2	87.24	91.38
18	8.223	9.398	25.96	28.82	31.41	68	47.03	50	83.29	88.28	92.62
19	8.86	10.14	27.2	30.14	32.82	69	47.95	50.87	84.41	89.36	93.67
20	9.573	10.85	28.4	31.4	34.08	70	48.9	51.66	85.55	90.52	94.99
21	10.31	11.63	29.63	32.68	35.46	71	49.68	52.63	86.64	91.68	96.1
22	11	12.35	30.8	33.94	36.73	72	50.46	53.44	87.71	92.86	97.31
23	11.7	13.1	32.03	35.17	38.06	73	51.32	54.32	88.84	93.87	98.37
24	12.43	13.87	33.2	36.39	39.29	74	52.1	55.17	90	95.08	99.66
25	13.13	14.63	34.35	37.65	40.63	75	53.03	55.97	91.07	96.14	100.8
26	13.88	15.4	35.54	38.86	41.87	76	53.72	56.94	92.24	97.32	101.9
27	14.6	16.11	36.72	40.04	43.09	77	54.61	57.78	93.29	98.38	103
28	15.32	16.91	37.91	41.32	44.44	78	55.39	58.62	94.35	99.61	104.3
29	16	17.72	39.07	42.53	45.69	79	56.34	59.51	95.47	100.7	105.4
30	16.76	18.48	40.22	43.72	46.92	80	57.3	60.31	96.57	101.8	106.5
31	17.56	19.29	41.42	44.91	48.3	81	58.05	61.25	97.66	103	107.7
32	18.37	20.07	42.59	46.29	49.45	82	58.96	62.07	98.78	104.1	108.8

33	19.06	20.86	43.75	47.41	50.68	83	59.7	62.99	99.92	105.3	110.1
34	19.81	21.66	44.88	48.63	51.91	84	60.56	63.83	101	106.3	111.2
35	20.62	22.42	46.11	49.8	53.19	85	61.37	64.76	102.2	107.5	112.3
36	21.26	23.29	47.22	50.93	54.39	86	62.13	65.67	103.2	108.5	113.3
37	22.06	24.06	48.39	52.14	55.55	87	63.06	66.49	104.2	109.8	114.7
38	22.91	24.88	49.49	53.37	56.87	88	63.82	67.31	105.4	110.9	115.8
39	23.62	25.7	50.63	54.54	58.07	89	64.84	68.23	106.4	112	116.9
40	24.38	26.5	51.79	55.68	59.24	90	65.63	69.13	107.5	113.1	118
41	25.23	27.37	52.96	56.96	60.54	91	66.51	70	108.7	114.4	119.2
42	26.09	28.15	54.05	58.18	61.74	92	67.45	70.93	109.8	115.4	120.4
43	26.81	28.95	55.23	59.26	62.85	93	68.24	71.73	110.9	116.5	121.4
44	27.6	29.78	56.39	60.47	64.17	94	69.14	72.53	111.9	117.7	122.7
45	28.38	30.58	57.51	61.64	65.33	95	69.93	73.5	113	118.7	123.7
46	29.13	31.45	58.65	62.74	66.45	96	70.83	74.33	114.2	119.9	125
47	29.94	32.34	59.77	63.99	67.79	97	71.61	75.3	115.2	120.9	126.1
48	30.66	33.06	60.86	65.13	68.96	98	72.37	76.22	116.3	122.1	127.2
49	31.61	33.91	62.01	66.26	70.11	99	73.33	77.03	117.4	123.2	128.3
50	32.31	34.78	63.16	67.5	71.24	100	74.1	77.87	118.5	124.2	129.3

Приложение 4. Квантили распределения Стьюдента.

P	Односторонние квантили			Двусторонние квантили		
	0.90	0.95	0.99	0.90	0.95	0.99
n						
1	3.064	6.256	29.81	6.256	12.47	58.12
2	1.882	2.916	6.746	2.916	4.231	9.639
3	1.637	2.347	4.423	2.347	3.176	5.738
4	1.529	2.13	3.717	2.13	2.76	4.492
5	1.474	2.004	3.298	2.004	2.565	3.864
6	1.439	1.938	3.108	1.938	2.444	3.618
7	1.414	1.891	2.978	1.891	2.345	3.448
8	1.396	1.858	2.884	1.858	2.293	3.324
9	1.378	1.821	2.813	1.821	2.253	3.144
10	1.368	1.804	2.717	1.804	2.222	3.088
11	1.36	1.789	2.681	1.789	2.197	3.043
12	1.353	1.778	2.651	1.778	2.176	3.004
13	1.348	1.767	2.625	1.767	2.141	2.972
14	1.343	1.759	2.604	1.759	2.129	2.944
15	1.339	1.751	2.585	1.751	2.118	2.919
16	1.335	1.744	2.568	1.744	2.109	2.838
17	1.332	1.738	2.554	1.738	2.1	2.825
18	1.329	1.733	2.541	1.733	2.093	2.813
19	1.327	1.729	2.529	1.729	2.086	2.803
20	1.324	1.724	2.485	1.724	2.08	2.793
21	1.322	1.72	2.478	1.72	2.074	2.784
22	1.32	1.717	2.471	1.717	2.069	2.776
23	1.319	1.714	2.465	1.714	2.065	2.768
24	1.317	1.711	2.46	1.711	2.06	2.761
25	1.316	1.708	2.455	1.708	2.057	2.755
30	1.31	1.697	2.435	1.697	2.041	2.728

35	1.306	1.689	2.431		1.689	2.029	2.723
40	1.303	1.684	2.417		1.684	2.021	2.704
45	1.301	1.679	2.406		1.679	2.014	2.689
50	1.299	1.675	2.398		1.675	2.009	2.677
55	1.297	1.672	2.391		1.672	2.004	2.667
60	1.296	1.67	2.385		1.67	2	2.658
70	1.294	1.666	2.374		1.666	1.994	2.645
80	1.292	1.663	2.368		1.663	1.99	2.637
90	1.291	1.661	2.364		1.661	1.986	2.63
100	1.29	1.659	2.36		1.659	1.984	2.625
150	1.287	1.653	2.349		1.653	1.975	2.609
200	1.286	1.651	2.343		1.651	1.971	2.601
250	1.285	1.649	2.34		1.649	1.969	2.596
300	1.284	1.648	2.337		1.648	1.967	2.592
500	1.283	1.648	2.334		1.648	1.965	2.586

Приложение 5. Квантили распределения Фишера $F_{m,n;0.95}$

m	1	2	3	4	8	12	16	24	50	100
n										
1	161	199	215	225	239	244	247	249	251	254
2	18.5	19.0	19.2	19.3	19.4	19.4	19.4	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	8.84	8.74	8.69	8.64	8.58	8.53
4	7.71	6.94	6.59	6.39	6.04	5.91	5.84	5.77	5.70	5.63
5	6.61	5.79	5.41	5.19	4.82	4.68	4.60	4.53	4.44	4.36
6	5.99	5.14	4.76	4.53	4.15	4.00	3.92	3.84	3.75	3.67
7	5.59	4.74	4.35	4.12	3.73	3.57	3.49	3.41	3.32	3.23
8	5.32	4.46	4.07	3.84	3.44	3.28	3.20	3.12	3.03	2.93
9	5.12	4.26	3.86	3.63	3.23	3.07	2.98	2.90	2.80	2.71
10	4.96	4.10	3.71	3.48	3.07	2.91	2.82	2.74	2.64	2.54
11	4.84	3.98	3.59	3.36	2.95	2.79	2.70	2.61	2.50	2.40
12	4.75	3.88	3.49	3.26	2.85	2.69	2.60	2.50	2.40	2.30
13	4.67	3.80	3.41	3.18	2.77	2.60	2.51	2.42	2.32	2.21
14	4.60	3.74	3.34	3.11	2.70	2.53	2.44	2.35	2.24	2.13
15	4.54	3.68	3.29	3.06	2.64	2.48	2.39	2.29	2.18	2.07
16	4.49	3.63	3.24	3.01	2.58	2.42	2.33	2.24	2.13	2.01
17	4.45	3.59	3.20	2.96	2.55	2.38	2.29	2.19	2.08	1.96
18	4.41	3.55	3.16	2.93	2.51	2.34	2.25	2.15	2.04	1.92
19	4.38	3.52	3.13	2.90	2.48	2.31	2.21	2.11	2.00	1.88
20	4.35	3.49	3.10	2.87	2.45	2.28	2.18	2.08	1.96	1.84
22	4.30	3.34	3.05	2.82	2.40	2.23	2.13	2.03	1.91	1.78
24	4.26	3.40	3.01	2.78	2.36	2.18	2.09	1.98	1.86	1.73
26	4.22	3.37	2.98	2.74	2.32	2.15	2.05	1.95	1.82	1.69
28	4.20	3.34	2.95	2.71	2.29	2.12	2.02	1.91	1.78	1.65
30	4.17	3.32	2.92	2.69	2.27	2.09	1.99	1.89	1.76	1.62
35	4.12	3.26	2.87	2.64	2.22	2.04	1.94	1.83	1.70	1.57
40	4.08	3.23	2.84	2.61	2.18	2.00	1.90	1.70	1.66	1.51
45	4.06	3.21	2.83	2.58	2.15	1.97	1.87	1.76	1.63	1.48

50	4.03	3.18	2.79	2.56	2.13	1.95	1.85	1.74	1.60	1.44
60	4.00	3.15	2.76	2.52	2.10	1.92	1.81	1.70	1.56	1.39
70	3.98	3.13	2.74	2.50	2.07	1.89	1.79	1.67	1.53	1.35
80	3.96	3.11	2.72	2.49	2.06	1.88	1.77	1.65	1.51	1.32
90	3.95	3.10	2.71	2.47	2.04	1.86	1.76	1.64	1.49	1.30
100	3.94	3.09	2.70	2.46	2.03	1.85	1.75	1.63	1.48	1.28
125	3.92	3.07	2.68	2.44	2.01	1.83	1.72	1.60	1.45	1.25
150	3.90	3.06	2.66	2.43	2.00	1.82	1.71	1.59	1.44	1.22
200	3.89	3.04	2.65	2.42	1.98	1.80	1.69	1.57	1.42	1.19
300	3.87	3.03	2.64	2.41	1.97	1.79	1.68	1.55	1.39	1.15
400	3.86	3.02	2.63	2.40	1.96	1.78	1.67	1.54	1.38	1.13
500	3.86	3.01	2.62	2.39	1.96	1.77	1.66	1.54	1.38	1.11
1000	3.85	3.00	2.61	2.38	1.95	1.76	1.65	1.53	1.36	1.08
∞	3.84	2.99	2.60	2.37	1.94	1.75	1.64	1.52	1.35	1.00

Приложение 6.

Критические точки распределения Кочрена $C(v_1, v_2; \alpha)$, где α – уровень значимости, v_2 – количество выборок, v_1 – число степеней свободы, $v_1 = m-1$ (m – объем каждой выборки).

$\alpha = 0.01$:

v_1	1	2	3	4	5	6	7	10	16	36
2	0.9999	.9950	.9794	.9586	.9373	.9172	.8988	.8539	.7949	.7067
3	.9933	.9423	.8831	.8335	.7933	.7606	.7335	.6743	.6059	.5153
4	.9676	.8643	.7814	.7212	.6761	.6410	.6129	.5536	.4884	.4057
5	.9279	.7885	.6957	.6329	.5875	.5531	.5259	.4697	.4094	.3351
6	.8828	.7218	.6358	.5635	.5195	.4866	.4608	.4084	.3529	.2858
7	.8376	.6644	.5685	.5080	.4659	.4347	.4105	.3616	.3105	.2494
8	.7945	.6152	.5209	.4627	.4226	.3932	.3704	.3248	.2779	.2214
9	.7544	.5727	.4810	.4251	.3870	.3592	.3378	.2950	.2514	.1992
10	.7175	.5358	.4469	.3934	.3572	.3308	.3106	.2704	.2297	.1811
12	.6528	.4751	.3919	.3428	.3099	.2861	.2680	.2320	.1961	.1535
15	.5747	.4069	.3317	.2882	.2593	.2386	.2228	.1918	.1612	.1251
20	.4799	.3297	.2654	.2288	.2048	.1877	.1748	.1501	.1248	.0960
24	.4247	.2871	.2295	.1970	.1759	.1608	.1495	.1283	.1060	.0810
30	.3632	.2412	.1913	.1635	.1454	.1327	.1232	.1054	.0867	.0658
40	.2940	.1915	.1508	.1281	.1135	.1033	.0957	.0816	.0668	.0503
60	.2151	.1371	.1069	.0902	.0796	.0722	.0668	.0567	.0461	.0344
120	.1225	.0759	.0585	.0489	.0429	.0387	.0357	.0302	.0242	.0178

$\alpha = 0.05$:

v_1	1	2	3	4	5	6	7	10	16	36
2	0.9985	.9750	.9392	.9057	.8772	.8534	.8332	.7880	.7341	.6602

3	.9669	.8709	.7977	.7457	.7071	.6771	.6530	.6025	.5466	.4748
4	.9065	.7679	.6841	.6287	.5895	.5598	.5365	.4884	.4366	.3720
5	.8412	.6338	.5981	.5440	.5063	.4783	.4564	.4118	.3645	.3066
6	.7808	.6161	.5321	.4803	.4447	.4184	.3980	.3568	.3135	.2612
7	.7271	.5612	.4800	.4307	.3974	.3726	.3535	.3154	.2756	.2278
8	.6798	.5157	.4377	.3910	.3595	.3362	.3185	.2829	.2462	.2022
9	.6385	.4775	.4027	.3584	.3286	.3067	.2901	.2568	.2226	.1820
10	.6020	.4450	.3733	.3311	.3029	.2823	.2666	.2353	.2032	.1655
12	.5410	.3924	.3624	.2880	.2624	.2439	.2299	.2020	.1737	.1403
15	.4709	.3346	.2758	.2419	.2195	.2034	.1911	.1671	.1429	.1144
20	.3894	.2705	.2205	.1921	.1735	.1602	.1501	.1303	.1108	.0879
24	.3434	.2354	.1907	.1656	.1493	.1374	.1286	.1113	.0942	.0743
30	.2929	.1980	.1593	.1377	.1237	.1137	.1061	.0921	.0771	.0604
40	.2370	.1576	.1259	.1082	.0968	.0887	.0827	.0713	.0595	.0462
60	.1737	.1131	.0895	.0765	.0682	.0623	.0583	.0497	.0411	.0316
120	.0998	.0632	.0495	.0419	.0371	.0337	.0312	.0266	.0218	.0165

НЕКОТОРЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Тема 1. Основные понятия теории вероятностей.

Ниже будут применяться следующие сокращения: СВ – случайная величина; МО – математическое ожидание; СКВ – среднее квадратическое отклонение; ГС – генеральная совокупность; ДА – дисперсионный анализ; РА – регрессионный анализ. Символом \otimes будет обозначаться конец решения или доказательства.

1.1. Случайные события и их вероятности. Повседневный опыт подсказывает, что большинство явлений окружающего нас мира *случайны*, непредсказуемы. С математической точки зрения, *случайное явление* – это такое явление, которое при неоднократном воспроизведении одного и того же опыта (испытания, эксперимента) протекает каждый раз несколько по-иному. Теорией вероятностей называют математическую науку, изучающую закономерности в случайных явлениях.

Исходными для теории вероятностей являются понятия опыта и события. Под *опытом* (экспериментом, испытанием) понимается осуществление конкретного комплекса условий, при которых наблюдается то или другое явление, фиксируется тот или другой результат. В теории вероятностей рассматриваются опыты со случайным исходом. Дополнительно предполагается, что условия опыта можно воспроизводить многократно. Примерами опытов со случайным исходом могут быть: стрельба по мишени; подбрасывание игральной кости; физический опыт по измерению какой-либо величины; наблюдения за явлениями в природе или в обществе; акт выборки одного изделия из партии произведенных изделий, и т.д. Всякий мыслимый результат опыта называется *элементарным событием* и обозначается символом ω . С каждым опытом связано некоторое множество элементарных событий. Введем следующее определение.

а) *Пространством элементарных событий* называется множество элементарных событий, связанных с одним и тем же опытом, таких, что: 1) в результате опыта всегда происходит одно из элементарных событий; 2) все элементарные события взаимоисключающие друг друга, т.е. никакие два из них не могут произойти вместе. Про-

пространство элементарных событий обозначается символом Ω . Если пространство Ω состоит из конечного или счетного числа элементарных событий, то его можно задавать перечислением: $\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$.

Очень часто опыт состоит в измерении некоторой величины X , которая может принять любое значение из интервала $(a; b)$. В этом случае множество Ω состоит из несчетного количества элементарных событий. Его можно задать следующим образом: $\Omega = \{x, x \in (a, b)\}$.

б) Любое подмножество пространства элементарных событий будем называть событием и обозначать символами A, B, C, A_1, B_2 и т.д. Событие A происходит (появляется) в результате опыта, если происходит одно из элементарных событий, принадлежащих A . Будем говорить, что событие A влечет событие B , если A является частью подмножества B . При этом пишут $A \subset B$. События A и B называются эквивалентными, если $A \subset B$ и $B \subset A$.

Пустое множество \emptyset является подмножеством любого множества. Событие, соответствующее пустому множеству, называют невозможным событием и обозначают тем же символом \emptyset . Невозможное событие никогда не происходит в результате опыта.

Аналогично, множество Ω также является подмножеством пространства элементарных событий. Событие, соответствующее Ω называют достоверным событием и обозначают тем же символом Ω . Достоверное событие всегда происходит в результате опыта.

Все другие события в результате опыта либо происходят, либо не происходят. Их называют случайными событиями. Слово случайное в дальнейшем иногда будет опускаться, и случайные события будут называться просто событиями.

Пример 1.1. Множеством всех событий для пространства Ω , состоящего из трех элементарных событий a, b, c , является $F = \{\emptyset, \Omega, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}\}$.

в) *Классическое определение вероятности.* Рассмотрим пространство Ω , связанное с каким-либо опытом, состоящее из конечного числа n элементарных событий. Дополнительно будем предполагать, что все элементарные события *равновозможные*. Это понятие интуитивное и не подлежит определению. Как правило, равновозможность является следствием симметрии предметов, участвующих в эксперименте. В искусственно организованных экспериментах дополнительно поддерживаются условия, обеспечивающие равновозможность появления любого из элементарных событий. Примеры опытов с равновозможными исходами: вытаскивание карты из колоды карт; выбор товара при покупке; заполнение карточки спортлото и др. Равновозможные элементарные события, образующие пространство Ω , называют *исходами (случаями, шансами)* данного опыта.

Пусть A – некоторое событие из пространства Ω . Элементарные события, из которых составлено событие A , будем называть исходами, *благоприятствующими событию A* .

Определение. *Вероятностью события A* называется число $P(A)$, равное отношению числа исходов, благоприятствующих событию A , к числу всех возможных исходов данного опыта, т.е.

$$P(A) = \frac{m_A}{n}. \quad (1.1)$$

В формуле (1.1) знаменатель n – количество всех исходов рассматриваемого опыта, числитель m_A – количество исходов, благоприятствующих событию A . Так как $0 \leq m_A \leq n$, то вероятность события – это число, заключенное в пределах от 0 до 1: $0 \leq P(A) \leq 1$, причем $P(\emptyset) = 0$, $P(\Omega) = 1$. Для равновероятных элементарных событий

$$P(\omega_k) = \frac{1}{n}, k = 1, 2, \dots, n.$$

Пример 1.2. Опыт: подбрасываются две игральные кости. Пусть событие A в этом опыте – сумма очков, выпавших на верхних гранях игральных костей ≥ 10 . Найти вероятность события A .

Решение. Отметим, что исходы данного опыта можно задавать упорядоченными парами чисел $(i; j)$, где i – количество выпавших очков на первой игральной кости, j – на второй. Так как i и j независимо пробегают множество чисел $\{1, 2, \dots, 6\}$, то количество исходов этого опыта $n = 6 \cdot 6 = 36$. Событию A (сумма выпавших очков ≥ 10) благоприятствуют та-

кие исходы: (4;6), (6;4), (5;5), (5;6), (6;5), (6;6), т.е. $m_A = 6$. Тогда $P(A) = \frac{6}{36} = \frac{1}{6}$ ⊗

г) *Геометрические вероятности.* Если опыт сводится к выбору точки из множества, расположенного на прямой или на плоскости или в пространстве, то применяют геометрическое определение вероятности, которое является разновидностью классического определения вероятности. Множество, из которого выбирается точка, будем обозначать Ω и отождествлять его с пространством элементарных событий. Событиями в рассматриваемой ситуации будут подмножества из Ω – интервалы на прямой, геометрические фигуры на плоскости или в пространстве. При дополнительном предположении о равновероятности выбора любых точек из Ω , вероятности событий будут пропорциональны размерам соответствующих подмножеств. Пусть $\mu(M)$ – мера множества M (длина, площадь, объем). Тогда формула (1.1) переписется в виде

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}. \quad (1.2)$$

д) *Статистическое определение вероятности.* Не всегда опыт сводится к системе равновероятных исходов. В таких случаях прибегают к *опытному* или так называемому *статистическому* определению вероятности. Для пояснения рассмотрим опыт, при осуществлении которого появляется или не появляется интересующее нас событие A . Повторим опыт n раз. Пусть в m из них событие A произошло. Тогда отношение $W = m/n$ называется *относительной частотой появления события A* в данной серии из n опытов. Произведем, далее, серию из n_1 опытов, в m_1 из которых появилось событие A . Получим новую относительную частоту $W_1 = m_1/n_1$. Вообще говоря, W_1 отлична от W . Однако практикой установлено, что с увеличением числа опытов относительные частоты стабилизируются и начинают колебаться около вероятности события A . В связи с этим за вероятность события A приближенно можно принять его относительную частоту при большом n :

$$P(A) \cong \frac{m}{n}. \quad (1.3)$$

В этом и состоит суть статистического определения вероятности. В формуле (1.3) n – количество проведенных опытов (достаточно большое число), m – количество опытов, в которых появилось событие A .

Имеется много примеров, когда вероятности событий определяются только статистически: вероятность рождения мальчика (для многих стран эта вероятность равна 0.516); только статистически можно установить вероятность брака при производстве однотипной продукции на предприятии и другие.

1.2. Некоторые теоремы и формулы теории вероятностей.

а) *Операции над событиями (как подмножествами из Ω).*

Событие B состоящее из элементарных событий пространства Ω , не принадлежащих A , называется *противоположным* событию A . Обозначения противоположного события: \bar{A} , "не A ". Событие \bar{A} происходит только в том случае, когда не происходит событие A и наоборот.

Суммой двух событий A и B называется событие, состоящее из элементарных событий, принадлежащих хотя бы одному из событий A или B . Сумму событий обозначают одним из символов $A+B$, " A или B ", $A \cup B$. Событие $A+B$ происходит в результате опыта, если происходит либо A либо B либо оба вместе.

Произведением двух событий A и B называется событие, состоящее из элементарных событий, принадлежащих одновременно и событию A и событию B . Наиболее употребительные обозначения произведения событий: $A \cdot B$, " A и B ", $A \cap B$. Событие $A \cdot B$ происходит в результате опыта, если происходят вместе оба события A и B .

Понятие суммы и произведения легко обобщается на случай любого конечного числа событий. Например, для трех событий $A+B+C=(A+B)+C$, $A \cdot B \cdot C=(A \cdot B) \cdot C$. Надо иметь в виду следующее: сумма $A_1+A_2+\dots+A_n$ происходит в результате опыта, если происходит хотя бы одно из событий A_k ($k=1, 2, \dots, n$); произведение $A_1 \cdot A_2 \cdot \dots \cdot A_n$ происходит, если в результате опыта происходят все события A_1, A_2, \dots, A_n .

Разностью двух событий A и B называется событие $A-B$, состоящее из элементарных событий, принадлежащих A и не принадлежащих B . Возможные обозначения разности событий: $A-B=A \setminus B$. Здесь операция « \setminus » означает теоретико-множественную разность (дополнение множества B в множестве A). В этом плане противоположное событие можно записать в виде $\bar{A} = \Omega \setminus A = \Omega - A$.

Укажем основные свойства операций над событиями:

1. $A+A=A$, 2. $A \cdot A=A$, 3. $A+\Omega=\Omega$, 4. $A+\emptyset=A$, 5. $A \cdot \Omega=A$, 6. $A \cdot \emptyset=\emptyset$,
7. $\bar{\Omega}=\emptyset$, 8. $\bar{\emptyset}=\Omega$, 9. $A+\bar{A}=\Omega$, 10. $A \cdot \bar{A}=\emptyset$, 11. $A+B=B+A$ – перестановочность слагаемых, 12. $A \cdot B=B \cdot A$ – перестановочность сомножителей, 13. $(A+B)+C=A+(B+C)=A+B+C$ – ассоциативность операции сложения, 14. $(A \cdot B) \cdot C = A \cdot (B \cdot C) = A \cdot B \cdot C$ – ассоциативность операции умножения, 15. $\overline{A+B} = \bar{A} \cdot \bar{B}$, 16. $\overline{A \cdot B} = \bar{A} + \bar{B}$ – законы дополнения, 17. $A \cdot (B+C) = (A \cdot B) + (A \cdot C)$, 18. $A+(B \cdot C) = (A+B) \cdot (A+C)$ – законы дистрибутивности.

Ниже приведены простейшие свойства вероятности, вытекающие из определения и разобранных операций над событиями. Некоторые из них очевидны, другие формулируются в виде теорем, которые легко доказываются на основе классического определения вероятности.

б) *Теоремы сложения вероятностей.* События A и B называются *несовместными*, если в их составе нет одних и тех же элементарных событий из Ω . В результате опыта несовместные события не могут произойти вместе. Несовместность означает, что $A \cap B = \emptyset$. Таким же образом вводится понятие несовместности для большего числа событий: события A_1, A_2, \dots, A_k называются *попарно несовместными*, если любые два из них не могут произойти вместе, т. е. $A_i \cap A_j = \emptyset$ ($i \neq j, i=1, 2, \dots, k; j=1, 2, \dots, k$).

Теорема 1.1. Вероятность суммы конечного числа несовместных событий из одного и того же пространства Ω равна сумме вероятностей этих событий:

$$P(A_1 + A_2 + \dots + A_k) = P(A_1) + P(A_2) + \dots + P(A_k). \quad (1.4)$$

На основании теоремы 1.1, доказанной для равновероятных исходов, можно обобщить определение вероятности события для случая, когда элементарные события пространства Ω не обязательно равновероятные. Для этого надо предположить, что вероятности $P(\omega_j)$ элементарных событий пространства Ω установлены (например, с помощью статистического определения). Так как элементарные события по предположению попарно несовместны, то вероятность произвольного события $A \subset \Omega$ можно определить следующей формулой:

$$P(A) = \sum_{\omega_j \in A} P(\omega_j). \quad (1.5)$$

Пусть событие A является частью события B , т. е. $A \subset B$. Можно сказать так же, что A влечет B . Тогда $B = A + (B - A)$, причем A и $(B - A)$ несовместны. Из теоремы 1.1

$$P(B) = P(A) + P(B - A) \Rightarrow P(B - A) = P(B) - P(A).$$

Так как $P(B - A) \geq 0$, то $P(A) \leq P(B)$. Таким образом, если A влечет B , то

$$P(A) \leq P(B). \quad (1.6)$$

Будем говорить, что события A_1, A_2, \dots, A_k составляют *полную группу*, если в результате опыта обязательно происходит одно и только одно из этих событий. Из определения следует, что события, составляющие полную группу, попарно несовместны и их сумма равна достоверному событию Ω . На основании теоремы 1.1 *сумма вероятностей событий, составляющих полную группу, равна 1*, т.е.

$$P(A_1) + P(A_2) + \dots + P(A_k) = 1. \quad (1.7)$$

События A и \bar{A} составляют полную группу, т. е. $P(A) + P(\bar{A}) = 1 \Rightarrow P(\bar{A}) = 1 - P(A)$.

Рассмотрим случай, когда события A и B могут быть совместными. Для вероятности их суммы справедлива следующая

Теорема 1.2. Вероятность суммы двух совместных событий равна сумме вероятностей этих событий минус вероятность их совместного появления.

$$P(A + B) = P(A) + P(B) - P(A \cdot B). \quad (1.8)$$

Теорему 1.2 можно обобщить на случай 3-х, 4-х и т.д. слагаемых. Однако получаемые при этом формулы очень громоздки и неудобны с точки зрения их практического применения. В связи с этим при отыскании вероятности суммы большого числа совместных событий удобнее перейти к противоположному событию. В результате получим формулу

$$P(A_1+A_2+\dots+A_k) = 1 - P(\bar{A}_1 \bar{A}_2 \dots \bar{A}_k). \quad (1.9)$$

в) *Условная вероятность. Теоремы умножения вероятностей.* Очень часто вероятность события зависит от условий проведения эксперимента. *Условной вероятностью события A при условии B* (по-другому, при наличии B) называется вероятность события A вычисленная при условии, что событие B произошло. Условная вероятность обозначается символом $P(A/B)$:

$$P(A/B) = \frac{m_{A \cap B}}{m_B} = \frac{m_{A \cap B} / n}{m_B / n} = \frac{P(A \cap B)}{P(B)} \quad (P(B) \neq 0).$$

Меняя местами A и B, будем иметь

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \quad (P(A) \neq 0).$$

Выражая из этих соотношений $P(A \cdot B)$, приходим к теореме.

Теорема 1.3. Вероятность произведения двух событий равна вероятности одного из них, помноженную на условную вероятность второго при наличии первого, т.е.

$$P(A \cdot B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B). \quad (1.10)$$

Формула (2.7) легко обобщается на случай 3-х сомножителей:

$$P(A \cdot B \cdot C) = P((A \cdot B) \cdot C) = P(A \cdot B) \cdot P(C/(A \cdot B)) = P(A) \cdot P(B/A) \cdot P(C/(A \cdot B)).$$

В общем случае

$$P(A \cdot B \cdot C \dots D \cdot E) = P(A) \cdot P(B/A) \cdot P(C/(A \cdot B)) \dots P(E/(A \cdot B \cdot C \dots D)) \quad (1.11)$$

Событие A называется независимым от события B, если его вероятность не зависит от того, произошло событие B или нет, т. е. $P(A/B) = P(A)$. В противном случае, если $P(A/B) \neq P(A)$, событие A зависит от B.

Зависимость и независимость событий взаимны: если A зависит от B, то и B зависит от A. Действительно, если предположить, что A не зависит от B, то из (1.10) $P(A) \cdot P(B/A) = P(B) \cdot P(A)$. Отсюда видно, что и $P(B/A) = P(B)$. Таким образом, для двух независимых событий вероятность их произведения равна произведению вероятностей:

$$P(A \cdot B) = P(A) \cdot P(B). \quad (1.12)$$

Иногда соотношение (1.10) используется для определения независимости событий, т. е. два события считаются независимыми, если вероятность их совместного появления равна произведению вероятностей.

Отметим, что если A и B независимы, то независимы так же следующие пары событий: (\bar{A}, B) , (A, \bar{B}) , (\bar{A}, \bar{B}) .

Множество событий A, B, C, ..., D, E называется независимым в совокупности, если вероятность каждого из этих событий не зависит от того, происходили или не происходили любые другие события. Всякое множество, образованное из независимых в совокупности событий и противоположных к ним, будет независимо в совокупности. Для независимых в совокупности событий формула (1.11) принимает вид

$$P(A \cdot B \cdot C \dots D \cdot E) = P(A) \cdot P(B) \cdot P(C) \dots P(D) \cdot P(E) \quad (1.13)$$

Пример 1.3. Для сигнализации об аварии установлены три независимо работающих сигнализатора, которые срабатывают при аварии с вероятностями 0.6, 0.7, 0.8. Найти вероятности следующих событий: A – при аварии сработали все три сигнализа-

тора; B – при аварии сработал хотя бы один сигнализатор; C – при аварии сработали два сигнализатора.

Решение. Введем события: A_1 – при аварии сработал 1-й сигнализатор, A_2 – 2-й, A_3 – 3-й. Тогда $A = A_1 \cdot A_2 \cdot A_3$, $B = A_1 + A_2 + A_3$, $C = A_1 \cdot A_2 \cdot \overline{A_3} + A_1 \cdot \overline{A_2} \cdot A_3 + \overline{A_1} \cdot A_2 \cdot A_3$. По условию $P(A_1)=0.6, P(A_2)=0.7, P(A_3)=0.8$, $P(\overline{A_1}) = 1 - 0.6 = 0.4$, $P(\overline{A_2}) = 0.3$, $P(\overline{A_3}) = 0.2$.

Так как события A_1, A_2, A_3 независимы между собой, то $P(A)=P(A_1) \cdot P(A_2) \cdot P(A_3)=0.6 \cdot 0.7 \cdot 0.8=0.336$, $P(B)=1-P(\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}) = 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot P(\overline{A_3})=1-0.4 \cdot 0.3 \cdot 0.2=0.976$. Легко заметить, что событие C является суммой попарно несовместных слагаемых, т.е. $P(C)=0.6 \cdot 0.7 \cdot 0.2+0.6 \cdot 0.3 \cdot 0.8+0.4 \cdot 0.7 \cdot 0.8=0.452$. ⊗

г) *Формула полной вероятности и формула Байеса.* Рассмотрим следующую ситуацию. Некоторое событие A может произойти лишь совместно с одним из событий H_1, H_2, \dots, H_k , составляющих полную группу. Считаются известными вероятности $P(H_1), P(H_2), \dots, P(H_k)$ и условные вероятности $P(A/H_1), P(A/H_2), \dots, P(A/H_k)$. Требуется найти вероятность события A . В этой ситуации событие A можно представить в виде суммы несовместных событий: $A = H_1 \cdot A + H_2 \cdot A + \dots + H_k \cdot A$. По теоремам сложения и умножения вероятность события A определяется по формуле

$$P(A)=P(H_1) \cdot P(A/H_1)+P(H_2) \cdot P(A/H_2)+\dots+P(H_k) \cdot P(A/H_k), \quad (1.14)$$

которую называют *формулой полной вероятности*.

Предположим теперь, что произведен опыт и в результате опыта произошло событие A . Поставим следующую задачу: с какой из гипотез вероятнее всего произошло событие A ? Для ответа на поставленный вопрос следует найти вероятность каждой из гипотез H_1, H_2, \dots, H_k при условии, что событие A произошло, т.е. найти вероятности $P(H_i/A)$ ($i=1, 2, \dots, k$). С учетом формулы (1.10) $P(H_i) \cdot P(A/H_i) = P(A) \cdot P(H_i/A)$. Отсюда получаем формулу Байеса

$$P(H_i / A) = \frac{m(H_i) \cdot \text{Чн}(A / H_i)}{m(A)} \quad (i = 1, 2, \dots, k), \quad (1.15)$$

где $P(A)$ – вероятность события A , вычисляемая по формуле (1.14).

Пример 1.4. В пирамиде 10 винтовок: 3 пристреленные и 7 не пристреленных. Вероятность поражения мишени из пристреленной винтовки равна 0.9, из не пристреленной – 0.5. Стрелок поразил мишень из наудачу выбранной винтовки. Какая, вероятнее всего, была выбрана винтовка: пристреленная или не пристреленная?

Для решения поставленной задачи введем гипотезы: H_1 – выбрана пристреленная винтовка, H_2 – выбрана не пристреленная винтовка и событие A – мишень поражена. По условию задачи $P(H_1) = 3/10 = 0.3$, $P(H_2) = 7/10 = 0.7$, $P(A/H_1) = 0.9$, $P(A/H_2) = 0.5$. Найдем сначала вероятность поражения мишени по формуле полной вероятности: $P(A)$

$$= 0.3 \cdot 0.9 + 0.7 \cdot 0.5 = 0.62. \text{ По формуле Байеса } P(H_1/A) = \frac{0.3 \cdot 0.9}{0.62} = 0.4355, P(H_2/A) =$$

$$\frac{0.7 \cdot 0.5}{0.62} = 0.5645. \text{ Как видно, более вероятно, что была выбрана не пристреленная вин-$$

товка. ⊗

1.3. Повторные испытания. Формула Бернулли.

а) На практике часто приходится встречаться с задачами, в которых один и тот же опыт повторяется многократно. В каждом из этих опытов наблюдается одно и то же событие A . Совокупность опытов (испытаний) называют независимой, если вероятность появления события A в i -м опыте не зависит от исходов предыдущих испытаний. Рассматриваемые ниже совокупности опытов будут предполагаться независимыми. Примером такой ситуации может быть производство однотипных деталей станком-автоматом, который работает в одних и тех же условиях. Однако некоторые из производимых деталей удовлетворяют стандарту (событие A), другие нет.

Введем обозначения: $p = P(A)$ – вероятность появления события A в каждом отдельно взятом опыте; $q = 1-p$ – вероятность не появления. Дополнительно предполагается, что $0 < p < 1$. Поставим следующую задачу: найти вероятность события B_k , состоящего в том, что в n независимых опытах событие A появится ровно k раз. Обозначим эту вероятность через $P_n(k) = P(B_k)$. Событие B_k происходит, если в каких то k опытах из n событие A появляется, а в остальных $n-k$ опытах не появляется. Пусть события A_i ($i=1, 2, \dots, n$), означают тот факт, что A появилось в i -м по счету опыте. Тогда один из вариантов появления события B_k можно задать в виде $A_1 \cdot A_2 \cdot \dots \cdot A_k \cdot \overline{A_{k+1}} \cdot \overline{A_{k+2}} \cdot \dots \cdot \overline{A_n}$. Другие варианты появления события B_k отличаются местами, на которых стоят сомножители A_i , например, $A_1 \cdot \overline{A_2} \cdot A_3 \cdot \dots \cdot A_{k+1} \cdot \overline{A_{k+2}} \cdot \dots \cdot \overline{A_n}$ и т. д. Количество таких вариантов равно числу сочетаний из n элементов по k элементов, т. е. равно $C_n^k = C_n^{n-k}$, причем любые два варианта несовместны между собой. В связи с независимостью опытов, вероятность одного осуществления события B всегда равна $p^k \cdot q^{n-k}$. Приведенные рассуждения позволяют записать выражение вероятности $P_n(k)$ в виде формулы, которую называют *формулой Бернулли*:

$$P_n(k) = C_n^k p^k q^{n-k}. \quad (1.16)$$

б) Рассмотрим, как распределяются в зависимости от k вероятности, вычисляемые по формуле (1.16). С помощью этой формулы находим, что отношение $\frac{P_n(k+1)}{P_n(k)} = \frac{n-k}{k+1} \cdot \frac{p}{q}$. Если $\frac{n-k}{k+1} \cdot \frac{p}{q} > 1 \Leftrightarrow k < n-p-q$, то $P_n(k+1) > P_n(k)$, т. е. вероятности $P_n(k)$ возрастают с увеличением k . Если же $k > n-p-q$, то вероятности $P_n(k)$ уменьшаются с увеличением k . Из этих соотношений видно, что между $n-p-q$ и $n-p-q+1 = n-p+p$ существует целое число k_0 , для которого вероятность $P_n(k_0)$ имеет наибольшее значение. Число k_0 называется *наивероятнейшим числом наступлений события A в независимых испытаниях*. В частном случае, когда $n-p-q$ является целым числом, $P_n(n-p-q) = P_n(n-p+p)$, т. е. существует два наивероятнейших числа.

При больших n формула Бернулли приводит к громоздким вычислениям. Приведем формулировку трех теорем, упрощающих эти вычисления.

в) *Локальная теорема Муавра – Лапласа*. Пусть вероятность наступления некоторого события A в n независимых опытах постоянна и равна p ($0 < p < 1$), $q = 1-p$ – вероятность не появления события A в одном опыте. Обозначим

$$x = \frac{k - np}{\sqrt{npq}}, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Тогда вероятность $P_n(k)$ того, что событие A наступит ровно k раз в этих опытах удовлетворяет при $n \rightarrow \infty$ соотношению

$$\lim_{n \rightarrow \infty} \frac{\sqrt{npq} P_n(k)}{\varphi(x)} = 1.$$

Из локальной теоремы Муавра – Лапласа вытекает следующая приближенная формула. При больших n

$$P_n(k) \cong \frac{1}{\sqrt{npq}} \cdot \varphi(x), \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \quad x = \frac{k - np}{\sqrt{npq}}, \quad k = 0, 1, \dots, n. \quad (1.17)$$

Обычно эта формула применяется при $n > 30$ и $n \cdot p \cdot q \geq 9$. Значения функции $\varphi(x)$ можно найти по таблице для $x \in [0; 4]$ (см. приложение 1). При $x > 4$ $\varphi(x) \cong 0$. При $x < 0$, $\varphi(x) = \varphi(-x)$.

г) *Интегральная теорема Муавра – Лапласа.* Пусть k есть число наступлений события A в n независимых опытах, в каждом из которых эта вероятность постоянна и равна p ($0 < p < 1$). Тогда имеет место соотношение

$$\lim_{n \rightarrow \infty} \left(P \left(a \leq \frac{k - np}{\sqrt{npq}} \leq b \right) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \right) = 0.$$

Если выбрать

$$a = \frac{k_1 - np}{\sqrt{npq}}, \quad b = \frac{k_2 - np}{\sqrt{npq}} \quad (k_1 < k_2), \quad \text{то} \quad P \left(a \leq \frac{k - np}{\sqrt{npq}} \leq b \right) = P_n(k_1 \leq k \leq k_2).$$

Тогда при больших n ($n > 30$, $n \cdot p \cdot q \geq 9$) вероятность $P_n(k_1 \leq k \leq k_2)$ приближенно можно найти по формуле

$$P_n(k_1 \leq k \leq k_2) \cong \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x^2}{2}} dx = \Phi(x_2) - \Phi(x_1), \quad (1.18)$$

где

$$x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

Интеграл, определяющий функцию $\Phi(x)$, «не берущийся». Значения $\Phi(x)$ можно найти по таблице для $x \in [0; 4]$ (см. приложение 2). При $x > 4$ $\Phi(x) \cong 0.5$. Так как функция $\Phi(x)$ нечетная, то для $x < 0$ $\Phi(x) = -\Phi(-x)$.

в) Большой круг задач связан с вычислением вероятностей $P_n(k)$ при малых p (случай редко появляющихся событий). В этих случаях формула (1.17) дает существенные погрешности. Основу для вычисления вероятностей $P_n(k)$ при больших n и малых p дает нижеследующая теорема.

Теорема Пуассона. Пусть вероятность события A при каждом опыте в серии из n независимых опытов равна a/n , где $a > 0$ – постоянная, не зависящая от n . Тогда

$\lim_{n \rightarrow \infty} P_n(k) = \frac{a^k}{k!} \cdot e^{-a}$. Следствие из теоремы Пуассона. При больших n и малых p (обычно $p < 0,1, n \cdot p \cdot q < 9$)

$$P_n(k) \cong \frac{a^k}{k!} \cdot e^{-a}, \text{ где } a = np, k = 0, 1, \dots, n. \quad (1.19)$$

г) Приведем некоторые следствия из интегральной теоремы Муавра – Лапласа. Предварительно рассмотрим вероятность неравенства

$$\left| \frac{k}{n} - p \right| < \varepsilon \Leftrightarrow -\varepsilon n < k - np < \varepsilon n \Leftrightarrow -\varepsilon \cdot \sqrt{\frac{n}{pq}} < \frac{k - np}{\sqrt{npq}} < \varepsilon \cdot \sqrt{\frac{n}{pq}}, \quad (1.20)$$

где ε – положительная постоянная, $\frac{k}{n} = w_n$ – относительная частота появления некоторого события A в n независимых опытах. Остальные величины в (1.20) имеют тот же смысл, что и в п. а). По интегральной теореме Муавра – Лапласа

$$P(|w_n - p| < \varepsilon) = \Phi\left(\varepsilon \sqrt{\frac{n}{pq}}\right) - \Phi\left(-\varepsilon \sqrt{\frac{n}{pq}}\right) = 2 \cdot \Phi\left(\varepsilon \sqrt{\frac{n}{pq}}\right). \quad (1.21)$$

Формула (1.21) устанавливает вероятность того, что в n независимых опытах (при больших n) относительная частота появления события A отклонится от вероятности появления A в одном опыте на величину, меньшую ε . С увеличением n эта вероятность стремится к единице.

1.4. Простейший поток событий

а) Рассмотрим одну ситуацию, где встречается формула (1.19). Введем определение: *потоком событий* называется последовательность однородных событий, появляющихся в случайные моменты времени. Примеры потоков: приход кораблей в порт, поступление вызовов на станцию скорой помощи, приходы посетителей в магазин и др. Одна из задач, связанных с потоками событий, состоит в определении вероятности

$$P_{(t_1, t_2)}(k) \quad (1.22)$$

того, что на интервале времени (t_1, t_2) произойдет ровно k событий потока. Отметим следующие свойства потоков.

а1) *Стационарность*. Это свойство означает, что вероятность (1.22) не зависит от начального момента времени t_1 , а зависит лишь от длины интервала $t = t_2 - t_1$, т. е. формула (1.22) принимает вид $P_t(k)$. Из этого следует, что среднее число событий, появляющихся в единицу времени, постоянно. Обозначим его λ и будем называть *интенсивностью потока*.

а2) *Ординарность*. Это свойство означает, что за бесконечно малый промежуток времени dt произойдет не более одного события потока. Точнее, вероятность попадания на малый промежуток времени dt двух и более событий $P_{dt}(k)$ ($k=2, 3, \dots$) ничтожно мала по сравнению с $P_{dt}(1)$.

а3) *Отсутствие последствия*, т. е. вероятность попадания того или другого числа событий на заданный участок оси $0t$ не зависит от того, сколько событий попало на любой другой не пересекающийся с ним участок. В частности, «будущее потока не зависит от его прошлого».

Поток событий, обладающий указанными тремя свойствами – стационарностью, ординарностью и отсутствием последствия, называется *простейшим* (или стационарным пуассоновским) потоком. Можно доказать, что вероятность $P_t(k)$ того, что за время t произойдет ровно k событий простейшего потока определяется формулой

$$P_t(k) = \frac{(\lambda t)^k}{k!} \cdot e^{-\lambda t}, \quad (1.23)$$

где λ – интенсивность потока (среднее число событий, появляющихся в единицу времени), $k = 0, 1, 2, \dots$.

Пример 1.5. На АТС поступает простейший поток событий с интенсивностью 0.2 (вызов/мин). Найти вероятность того, что за пять минут: а) не придет ни одного вызова; б) придет ровно один вызов; в) придет хотя бы один вызов.

Решение. По условию $\lambda=0.2$, $t=5$, $\lambda t=1$. Тогда

$$\begin{aligned} \text{а) } P_5(0) &= \frac{1}{0!} \cdot e^{-1} = 0.36788; \quad \text{б) } P_5(1) = \frac{1}{1!} \cdot e^{-1} = 0.36788; \\ \text{в) } P_5(k \geq 1) &= 1 - P_5(0) = 1 - 0.36788 = 0.73212. \quad \otimes \end{aligned}$$

Тема 2. . Случайные величины (СВ).

2.1. Одномерные дискретные СВ.

а). *Основные определения.* Очень часто в результате эксперимента появляется численное значение, причем заранее (до опыта) нельзя точно предсказать, какое именно значение появится. Математические модели таких экспериментов называются случайными величинами (сокращенно СВ). Примеры СВ: число вызовов, поступивших на станцию скорой помощи в течение суток; величина отклонения точки приземления снаряда от центра цели и др.

В одних экспериментах измеряют одну какую-то величину – такими экспериментами определяются одномерные СВ; в других экспериментах сразу измеряют несколько величин, образующих комплекс – случайные величины в таких экспериментах называют многомерными (векторными) СВ или системами СВ. Обозначения: X, Y, Z, \dots – одномерные СВ; (U, V) – двумерная СВ; $(X^{(1)}, X^{(2)}, \dots, X^{(n)})$ – n -мерная СВ. Одномерные СВ, из которых составлена система, называют ее *компонентами*. При задании многомерной СВ компоненты должны быть упорядочены.

Пример 2.1. Станок изготавливает детали цилиндрической формы. Контролируются следующие размеры детали: L – длина детали; D – ее диаметр; M – масса. СВ (L, D, M) является трехмерной.

По множеству принимаемых значений одномерные СВ делятся на *дискретные* и *непрерывные*. СВ называется дискретной, если множество ее значений конечно или счетное, и непрерывной – если множество значений заполняет некоторый интервал действительных чисел. Обозначения: $\Omega_X = \{x_1, x_2, \dots, x_n, \dots\}$ – множество значений одномерной дискретной СВ X ; $\Omega_Y = \{y, y \in [a, b]\}$ – множество значений одномерной непрерывной СВ Y .

Законом распределения СВ называется любое правило, позволяющее находить вероятности всевозможных событий, связанных со случайной величиной (например, что она примет какое-то значение или значение из некоторого интервала).

Наиболее простую форму можно придать закону распределения дискретной одномерной СВ. Для этого достаточно задать множество ее возможных значений и указать, как часто они появляются в результатах экспериментов, т. е. указать вероятности, с которыми эти значения принимаются. Дополнительное предположение: в результате опыта СВ принимает одно и только одно из своих значений. Из этого следует, что сумма вероятностей всех значений равна единице. Эта единица как-то распределяется между значениями СВ (отсюда и термин «*распределение*»). Дискретную одномерную СВ можно задать таблицей (*рядом распределения*):

Таблица 2.1

Ω_X	x_1	x_2	...	x_n	...
P	p_1	p_2	...	p_n	...

$$\sum_i P(X = x_i) = \sum_i p_i = 1.$$

В верхней строке таблицы перечисляются в порядке возрастания все возможные значения дискретной СВ, в нижней – вероятности этих значений.

Пусть некоторая функция $\varphi(x)$ определена для всех значений дискретной СВ X . Тогда можно определить функцию $Y = \varphi(X)$, которая является СВ и задается следующей таблицей:

Таблица 2.2

$\Omega_{\varphi(X)}$	$\varphi(x_1)$	$\varphi(x_2)$...	$\varphi(x_n)$...
P	p_1	p_2	...	p_n	...

Таблицу 2.2 нельзя считать законом распределения СВ $\varphi(X)$. Для его получения значения $\varphi(x_i)$ в верхней строке следует упорядочить по возрастанию и одинаковые из этих значений записать как одно значение, предварительно сложив соответствующие им вероятности. Обычно случайные величины называют *распределениями*.

б) *Числовые характеристики дискретных одномерных СВ.*

Математическим ожиданием дискретной СВ X называется сумма произведений всех ее значений на соответствующие вероятности. Математическое ожидание будем обозначать одним из символов $M(X)$ или m_x . Предполагая, что дискретная СВ задана таблицей 2.1, запишем математическое ожидание в виде формулы

$$M(X) = m_x = \sum_{i=1}^{n(\infty)} x_i \cdot p_i. \quad (2.1)$$

Так как математическое ожидание СВ и среднее значение измеряемой величины определяются однотипными формулами, то $M(X)$ называют *центром распределения* или *средним значением* СВ. *Центрированной* называют СВ $\overset{o}{X} = X - m_x$, множество зна-

чений которой (см. таблицу 2.2) $\{x_1 - m_x, x_2 - m_x, \dots, x_n - m_x, \dots\}$, а вероятности $P\left(\overset{o}{X} = x_i - m_x\right) = p_i$. Легко видеть, что $M\left(\overset{o}{X}\right) = 0$.

Еще одной из характеристик положения является *мода*. Модой $Mo = Mo(X)$ дискретной СВ X называют значение СВ X , имеющее наибольшую вероятность. Дискретная СВ может иметь несколько мод.

Для характеристики *разбросанности (рассеивания)* значений СВ относительно центра распределения, вводятся понятия дисперсии и среднего квадратического отклонения.

Дисперсией $D(X) = D_x$ СВ X называют математическое ожидание квадрата *центрированной СВ* $\overset{o}{X}$:

$$D(X) = D_x = M\left(\left(\overset{o}{X}\right)^2\right) = \sum_{i=1}^{n(\infty)} (x_i - m_x)^2 \cdot p_i. \quad (2.2)$$

Слово «дисперсия» означает «рассеивание». Отметим, что дисперсия имеет размерность квадрата СВ, что не всегда удобно. Для наглядности в качестве характеристики рассеивания удобнее пользоваться числом, размерность которого совпадает с размерностью СВ. Для этого из дисперсии извлекают квадратный корень и вводят следующее определение.

Средним квадратическим отклонением (сокращенно СКО) или *стандартом СВ* X называется квадратный корень из дисперсии:

$$\sigma_x = \sqrt{D(X)}. \quad (2.3)$$

Безразмерную величину

$$v = \frac{\sigma_x}{m_x} \quad (2.4)$$

называют *коэффициентом вариации*.

Начальные V_j и центральные μ_j моменты j -го порядка определяются формулами:

$$V_j = M\left(X^j\right) = \sum_{k=1}^{n(\infty)} x_k^j \cdot p_k, \quad \mu_j = M\left(\left(\overset{o}{X}\right)^j\right) = \sum_{k=1}^{n(\infty)} (x_k - m_x)^j \cdot p_k. \quad (2.5)$$

На практике эти формулы используются при $j = 1, 2, 3, 4$. Легко видно, что $v_1 = M(X)$, $\mu_1 = 0$, $\mu_2 = D(X)$. С помощью центральных моментов μ_3 и μ_4 определяются безразмерные *коэффициенты асимметрии* As и *эксцесса* ε_x :

$$As = \mu_3 / \sigma_x^3, \quad \varepsilon_x = (\mu_4 / \sigma_x^4) - 3. \quad (2.6)$$

2.2. Одномерные непрерывные СВ.

Так как множество значений непрерывной СВ не является счетным, то задать распределение вероятностей непрерывной СВ в виде таблицы невозможно. Необходи-

мы другие способы задания случайных величин. Введем понятие функции распределения, пригодное для задания как дискретных, так и непрерывных СВ.

а) *Функцией распределения одномерной случайной величины X* называется функция $F(x)$ действительного переменного $x \in R$ (R – множество действительных чисел), равная вероятности выполнения неравенства $X < x$, т. е.:

$$F(x) = P(X < x) = P(X \in (-\infty, x)), \quad (2.7)$$

В формуле (2.7) X – символ случайной величины, x – аргумент функции распределения. Его можно обозначать любой другой буквой. Если в рассмотрении одновременно участвуют СВ X, Y, \dots , то их функции распределения можно обозначать $F_X(x), F_Y(y), \dots$, или $F_1(x), F_2(y), \dots$.

Приведем основные свойства функции $F(x)$, вытекающие из ее определения.

а1) Очевидно $0 \leq F(x) \leq 1, F(-\infty)=0, F(+\infty)=1$.

а2) $F(x)$ неубывающая функция, т. е. если $x_1 < x_2$, то $F(x_1) \leq F(x_2)$

а3) Вероятность попадания СВ в интервал

$$P(a \leq X < b) = F(b) - F(a), \quad (2.8)$$

а4) Если множеством значений СВ X является конечный интервал (a, b) или отрезок $[a, b]$, то функция распределения такой СВ имеет вид

$$F(x) = \begin{cases} 0, & x \leq a, \\ F_0(x), & a < x \leq b, \\ 1, & x > b, \end{cases}$$

где $F_0(x)$ – неубывающая функция, причем $F_0(a) = 0, F_0(b) = 1$.

б) *Плотностью вероятности одномерной непрерывной СВ X в точке x* называется следующий предел (если он существует):

$$p(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X < x + \Delta x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x). \quad (2.9)$$

Таким образом, плотность вероятности является производной от функции распределения. Чтобы формула (2.9) имела смысл, вводится дополнительное предположение, что функция распределения непрерывной СВ непрерывна на всей числовой оси и дифференцируема (за исключением, быть может, конечного числа точек). Выпишем свойства плотности вероятности, вытекающие из формулы (2.9).

б1) Так как функция распределения не убывающая, то плотность вероятности не отрицательна, т. е. $p(x) \geq 0$.

б2) Так как функция распределения $F(x)$ является одной из первообразных для $p(x)$, то

$$P(a \leq X < b) = F(b) - F(a) = F(x)|_a^b = \int_a^b p(x) dx, \quad (2.10)$$

т. е. вероятность попадания непрерывной СВ в интервал, равна интегралу от плотности вероятности по этому интервалу. Полагая в (2.10) $a = -\infty, b = +\infty$, получим соотношение

$$\int_{-\infty}^{+\infty} p(x) dx = 1, \quad (2.11)$$

б3) Если в (2.10) положить $a = -\infty$, $b = x$, получаем выражение функции распределения через плотность вероятности:

$$F(x) = \int_{-\infty}^x p(t) dt. \quad (2.12)$$

б) Если значения непрерывной СВ X лежат в интервале $(a; b)$, то $p(x)=0$, при $x < a$ или $x > b$.

б5) При малых Δx вероятность $P(x \leq X < x + \Delta x) \cong p(\xi) \cdot \Delta x$, $\xi \in [x, x + \Delta x]$.

Таким образом, непрерывную СВ можно задать либо функцией распределения, либо плотностью вероятности. В качестве функции распределения может служить любая непрерывная неубывающая функция $F(x)$, удовлетворяющая условиям $0 \leq F(x) \leq 1$, $F(-\infty) = 0$, $F(+\infty) = 1$. В качестве плотности вероятности можно взять любую неотрицательную кусочно-непрерывную функцию, удовлетворяющую условию (2.11).

в) Числовые характеристики непрерывных одномерных СВ определяются формулами

$$\left. \begin{aligned} M(X) = m_x &= \int_{-\infty}^{+\infty} xp(x)dx, & D(X) &= \int_{-\infty}^{+\infty} (x - m_x)^2 p(x)dx, \\ v_k &= \int_{-\infty}^{+\infty} x^k p(x)dx, & \mu_k &= \int_{-\infty}^{+\infty} (x - m_x)^k p(x)dx \quad (k = 1, 2, 3, 4). \end{aligned} \right\} \quad (2.13)$$

Отметим, что

$$m_x = v_1, \quad \mu_1 = 0, \quad D(X) = \mu_2.$$

Другие числовые характеристики

$$\sigma_x = \sqrt{D(X)}, \quad As = \frac{\mu_3}{\sigma_x^3}, \quad \varepsilon_x = \frac{\mu_4}{\sigma_x^4} - 3, \quad v_x = \frac{\sigma_x}{m_x}.$$

Модой непрерывной СВ называется точка глобального максимума плотности вероятности.

Для непрерывной СВ вводится понятие медианы $Me = Me(X)$, которая определяется из соотношения:

$$P(X \leq Me) = P(X \geq Me) = 0.5,$$

т. е. медиана делит множество значений СВ X на две части, вероятности попадания в которые одинаковы и равны 0.5.

2.3. Двумерные СВ.

а) *Двумерные дискретные СВ.* Многомерными СВ называются векторы, координаты (компоненты) которых являются одномерными СВ. По-другому, многомерные СВ называются векторными СВ или системами СВ. В данном пособии будут рассматриваться, в основном, двумерные СВ.

Двумерные дискретные СВ $(X; Y)$ задаются таблицами с двумя входами. 1-й столбец содержит значения СВ X : $\Omega_X = \{x_1, x_2, \dots, x_m\}$; 1-я строка содержит значения СВ Y : $\Omega_Y = \{y_1, y_2, \dots, y_n\}$. На пересечениях строк и столбцов стоят вероятности, с которыми каждая пара значений принимается: $p_{ij} = P(X=x_i, Y=y_j)$ (таблица 2.3):

Таблица 2.3

Ω_Y	y_1	y_2	...	y_n	$p_{i\bullet}$	
Ω_X	x_1	$p_{1,1}$	$p_{1,2}$...	$p_{1,n}$	$p_{1\bullet}$
	x_2	$p_{2,1}$	$p_{2,2}$...	$p_{2,n}$	$p_{2\bullet}$
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
	x_m	$p_{m,1}$	$p_{m,2}$...	$p_{m,n}$	$p_{m\bullet}$
	$p_{\bullet j}$	$p_{\bullet 1}$	$p_{\bullet 2}$...	$p_{\bullet n}$	1

Таблицу 2.3 называют так же матрицей распределения вероятностей двумерной СВ.

Сумма вероятностей p_{ij} по всем клеткам таблицы 2.3 должна равняться единице.

1-я задача, связанная с двумерными СВ – найти законы распределения компонент X и Y . Для ее решения отметим, что событие $(X=x_i) = (Y=y_1, X=x_i) + (Y=y_2, X=x_i) + \dots + (Y=y_n, X=x_i)$. Отсюда следует, что $P(X=x_i) = p_{i,1} + p_{i,2} + \dots + p_{i,n}$. Аналогичное соотношение можно записать и для вероятности $P(Y=y_j)$. В таблице 2.3 введены дополнительная строка и дополнительный столбец, дающие распределение вероятностей составляющих X и Y двумерной СВ (X, Y) . Таким образом, 1-й и последний столбцы таблицы 2.3 дают закон распределения СВ X , 1-я и последняя строки дают закон распределения СВ Y . Выражения $p_{\bullet j}$ и $p_{i\bullet}$ означают суммирование:

$$p_{\bullet j} = \sum_{i=1}^m p_{i,j} = P(Y=y_j), \quad p_{i\bullet} = \sum_{j=1}^n p_{i,j} = P(X=x_i).$$

2-я задача связана с зависимостью и независимостью компонент. Введем определение: *компоненты X и Y двумерной дискретной СВ (X, Y) называются независимыми между собой, если для всех индексов i, j выполнено условие $P(X=x_i, Y=y_j) = P(X=x_i) \cdot P(Y=y_j)$. Из определения независимости следует, что для независимых СВ $p_{i,j} = p_{i\bullet} \cdot p_{\bullet j}$. Если же хотя-бы для одной пары индексов (i, j) это условие не выполнено, то СВ X и Y зависимы. Свойство зависимости или независимости легко проверяется по таблице 2.3.*

Многомерная СВ $(X^{(1)}, X^{(2)}, \dots, X^{(k)})$ задается в виде таблицы:

i	$X^{(1)}$	$X^{(2)}$...	$X^{(k)}$	p_i
1	$x_{1,1}$	$x_{1,2}$...	$x_{1,k}$	p_1
2	$x_{2,1}$	$x_{2,2}$...	$x_{2,k}$	p_2
\vdots	\vdots	\vdots		\vdots	\vdots
n	$x_{n,1}$	$x_{n,2}$...	$x_{n,k}$	p_n

где

$$\sum_{i=1}^n p_i = 1.$$

б) *Двумерные непрерывные СВ. Функцией распределения двумерной СВ называется функция двух переменных $F(x, y)$, равная вероятности того, что одновременно СВ X принимает значение, меньшее x , а СВ Y – значение, меньшее y :*

$$F(x, y) = P(X < x; Y < y) = P(X \in (-\infty, x); Y \in (-\infty, y)). \quad (2.14)$$

Графически $P(X < x; Y < y)$ означает вероятность попадания случайной точки (X, Y) в бесконечный прямой угол с вершиной в точке (x, y) (рис. 2.1).

y

y

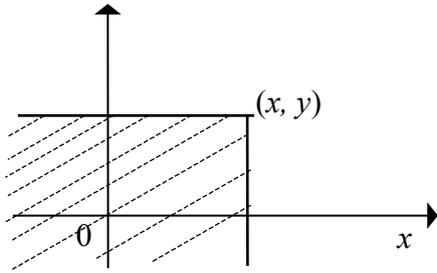


Рис. 2.1

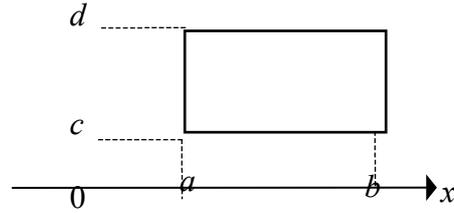


Рис. 2.2

Из определения вытекают следующие свойства $F(x, y)$:

F1) $0 \leq F(x, y) \leq 1$, $F(-\infty, y) = F(x, -\infty) = 0$, $F(+\infty, +\infty) = 1$;

F2) $F(x, y)$ - неубывающая функция по каждой переменной;

F3) вероятность попадания двумерной СВ в прямоугольник (рис. 2.2) определяется формулой

$$P(a \leq X < b, c \leq Y < d) = F(b, d) - F(b, c) - F(a, d) + F(a, c);$$

F4) функции распределения компонент X и Y двумерной СВ (X, Y) (маргинальные функции распределения) находятся по формулам

$$F_X(x) = P(X < x) = F(x, +\infty), \quad F_Y(y) = P(Y < y) = F(+\infty, y)$$

F5) Компоненты X и Y двумерной СВ называются независимыми, если $P(X < x; Y < y) = P(X < x) \cdot P(Y < y)$. Из определения следует, что для независимых СВ функция распределения представляется в виде произведения двух функций, одна из которых зависит только от x а другая только от y :

$$F(x, y) = F_X(x) \cdot F_Y(y).$$

Плотностью вероятности (плотностью распределения) двумерной непрерывной СВ (X, Y) в точке (x, y) называется предел средней плотности распределения вероятностей в прямоугольнике $(x \leq X < x + \Delta x, y \leq Y < y + \Delta y)$ при условии, что $\Delta x \rightarrow 0$ и $\Delta y \rightarrow 0$ (при условии дважды дифференцируемости $F(x, y)$ этот предел существует). Выведем выражение для плотности вероятности через функцию распределения. По определению

$$\begin{aligned} p(x, y) &= \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{P(x \leq X < x + \Delta x, y \leq Y < y + \Delta y)}{\Delta x \cdot \Delta y} = \\ &= \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y) - F(x + \Delta x, y) + F(x, y)}{\Delta x \cdot \Delta y}. \end{aligned}$$

Вычисляя пределы, получаем, что плотность вероятности является смешанной производной второго порядка от функции распределения:

$$p(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = F_{xy}''(x, y) \quad (2.15)$$

Выпишем основные свойства плотности вероятности.

P1) $p(x, y) \geq 0$;

P2) вероятность попадания двумерной СВ (X, Y) в область G определяется формулой

$$P((X, Y) \in G) = \iint_G p(x, y) dx dy;$$

P3) выражение функции распределения через плотность вероятности имеет вид

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(u, v) du dv;$$

P4) при $x = +\infty, y = +\infty$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) dx dy = \iint_{R^2} p(x, y) dx dy = 1.$$

Обычно это свойство называют условием нормировки.

P5) если известна совместная плотность распределения $p(x, y)$ двумерной СВ (X, Y) , то плотности распределения компонент получаются интегрированием $p(x, y)$:

$$p_X(x) = \int_{-\infty}^{+\infty} p(x, y) dy, \quad p_Y(y) = \int_{-\infty}^{+\infty} p(x, y) dx.$$

P6) Если компоненты X и Y двумерной СВ независимы, то плотность распределения представляется в виде произведения двух функций:

$$p(x, y) = p_X(x) \cdot p_Y(y).$$

в). *Числовые характеристики многомерных (двумерных) СВ.* Числовые характеристики двумерных СВ, такие как математическое ожидание, дисперсия, среднее квадратическое отклонение, мода, медиана, коэффициенты асимметрии и эксцесса определяются как векторы, составленные из соответствующих числовых характеристик компонент, т. е. $M(X, Y) = (m_x, m_y)$, $D(X, Y) = (D_x, D_y)$ и т. д.. Поэтому рассмотрим лишь числовые характеристики, устанавливающие связи между компонентами. К ним относятся совместные начальные и центральные моменты. Формулы приводятся для дискретных СВ а в скобках – для непрерывных СВ. Из них получаются выражения для математического ожидания, дисперсии и других характеристик компонент. Предполагается, что дискретная СВ задана таблицей 2.3 ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$), а непрерывная – плотностью распределения $p(x, y)$.

Начальные v_{k+l} и центральные μ_{k+l} моменты порядка $k+l$:

$$v_{k+l} = \sum_{i,j} x_i^k y_j^l p_{i,j} \left(\iint_{R^2} x^k y^l p(x, y) dx dy \right), \quad m_x = v_{1+0}, \quad m_y = v_{0+1};$$

$$\mu_{k+l} = \sum_{i,j} (x_i - m_x)^k (y_j - m_y)^l p_{i,j} \left(\iint_{R^2} (x - m_x)^k (y - m_y)^l p(x, y) dx dy \right).$$

$$D_x = \mu_{2+0}, \quad D_y = \mu_{0+2}, \quad \sigma_x = \sqrt{D_x}, \quad \sigma_y = \sqrt{D_y}, \quad \mu_{1+0} = \mu_{0+1} = 0$$

Особую роль играет центральный момент порядка 1+1. Его называют *корреляционным* моментом или *ковариацией* и обозначают одним из символов $\mu_{1+1} = K_{xy} = \text{cov}(x, y) = \sigma_{xy}$ (ниже, в зависимости от ситуации, будет использоваться любое из этих обозначений):

$$K_{xy} = \sum_{i,j} (x_i - m_x)(y_j - m_y) p_{i,j} = \sum_{i,j} x_i y_j p_{i,j} - m_x m_y$$

$$\left(\iint_{R^2} (x - m_x)(y - m_y) p(x, y) dx dy = \iint_{R^2} xy p(x, y) dx dy - m_x m_y \right)$$

Отношение

$$\rho_{xy} = \frac{K_{xy}}{\sigma_x \cdot \sigma_y},$$

которое является безразмерной величиной, называется *линейным коэффициентом корреляции* между СВ X и Y .