

СТАТИСТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ВЫБОРКИ

Построим математическую модель эксперимента. Для этого продублируем эксперимент в одинаковых условиях независимым образом n раз. В результате опыта множество исходов эксперимента, которое описывается случайной величиной ξ , будет образовывать некоторое конечное множество $\{x_1, x_2, \dots, x_n\}$ из множества возможных значений Ω_ξ случайной величины ξ . Наблюдаемые значения x_i случайной величины ξ называются *вариантами*. Множество Ω_ξ возможных значений случайной величины ξ называется *генеральной совокупностью*, а число элементов множества Ω_ξ – *объёмом генеральной совокупности*. Множество $\{x_1, x_2, \dots, x_n\}$ называется *выборкой* или *выборочной совокупностью*, а число элементов, входящих в выборку, называется *объёмом выборки*. Выборка должна быть *репрезентативной* или *представительной*, то есть должна хорошо отображать свойства генеральной совокупности Ω_ξ .

Исходы эксперимента дискретной случайной величины можно представить в виде таблицы (или *простого статистического ряда*), которая состоит из двух строк, в первой из которой указываются номера эксперимента, а во второй приводятся значения исходов эксперимента. Простой статистический ряд является первоначальной формой статистического ряда.

i – номер эксперимента	1	2	...	n
x_i – исход эксперимента	x_1	x_2	...	x_n

Для оценки распределения случайной величины ξ производят группировку исходных данных. Если исходные данные случайной величины ξ , которые принимают значения $x_1, x_2, \dots, x_n$, располагаются в порядке возрастания, то статистический ряд называется *вариационным*. Определим частоты m_i или частости $\omega_i = m_i/n$ (относительные частоты) появления одинаковых значений случайной величины ξ . В результате получаем сгруппированные статистические ряды следующего вида:

x_i – исход эксперимента	x_1	x_2	...	x_k
m_i – частоты	m_1	m_2	...	m_k
x_i – исход эксперимента	x_1	x_2	...	x_k
m_i/n – частости	m_1/n	m_2/n	...	m_k/n

При записи вариационных рядов необходимо учитывать следующие равенства: $\sum_{i=1}^k m_i = n$, $\sum_{i=1}^k m_i/n = 1$.

Предположим, что случайная величина ξ является непрерывной. Группировка данных заключается в разбиении интервала наблюдаемых значений случайной величины ξ на k частичных интервалов равной длины h : $[x_0; x_1)$, $[x_1; x_2)$, ..., $[x_{k-1}; x_k]$ и нахождении частоты m_i или частоты m_i/n попадания наблюдаемых значений в частичные интервалы. Количество интервалов выбирается произвольно, обычно не меньше трёх, но не более двадцати. Для определения числа интервалов разбиения может быть использована формула $k = 1 + 3,3222 \cdot \lg N$, где N – количество элементов совокупности. В случае равных интервалов величина интервала может быть определена по формуле

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{x_{\max} - x_{\min}}{1 + 3,3222 \cdot \lg N}.$$

В результате составляются интервальные статистические ряды следующего вида $\left(\sum_{i=1}^k m_i = n, \sum_{i=1}^k m_i/n = 1 \right)$:

Интервалы наблюдаемых значений СВ ξ	$[x_0; x_1)$	$[x_1; x_2)$	...	$[x_{k-1}; x_k]$
m_i – частоты	m_1	m_2	...	m_k

Интервалы наблюдаемых значений СВ ξ	$[x_0; x_1)$	$[x_1; x_2)$	...	$[x_{k-1}; x_k]$
m_i/n – частоты	m_1/n	m_2/n	...	m_k/n

Определение 1.1.1 Перечень наблюдаемых значений случайной величины ξ (или интервалов наблюдаемых значений) и соответствующие им частоты $\omega_i = m_i/n$ называется статистическим законом распределения случайной величины ξ .

Частость события $B_i = [x_{i-1}; x_i)$ является случайной величиной с математическим ожиданием, равным вероятности этого события, то есть $M(m_i/n) = P(x_{i-1} < \xi < x_i) = p_i$. Закон больших чисел в схеме Бернулли утверждает, что если эксперимент повторяется n раз при одинаковых условиях, то частость события сходится к вероятности этого же события: $\frac{m_i}{n} \xrightarrow{P} p_i$.

Поэтому во второй строке интервального статистического ряда стоят оценки (приближённые значения) вероятностей p_i .

Определение 1.1.2 Эмпирической функцией распределения случайной величины ξ называется функция $F^*(x)$, определяющая для каждого значения x частность события ($\xi < x$):

$$F^*(x) = \frac{n_x}{n}, \quad (1.1.1)$$

где n_x – число значений x_i , меньших значения x ; n – объём выборки.

Из теоремы Бернулли следует, что при большом объёме выборки эмпирическая функция распределения $F^*(x)$ и функция распределения случайной величины $F(x) = P(\xi < x)$ мало отличаются друг от друга. Отличие эмпирической функции распределения от теоретической состоит в том, что теоретическая функция распределения определяет вероятность события ($\xi < x$), а эмпирическая функция определяет относительную частоту этого же события. На рисунке 1.1.1 изображён график эмпирической функции распределения некоторой случайной величины ξ .

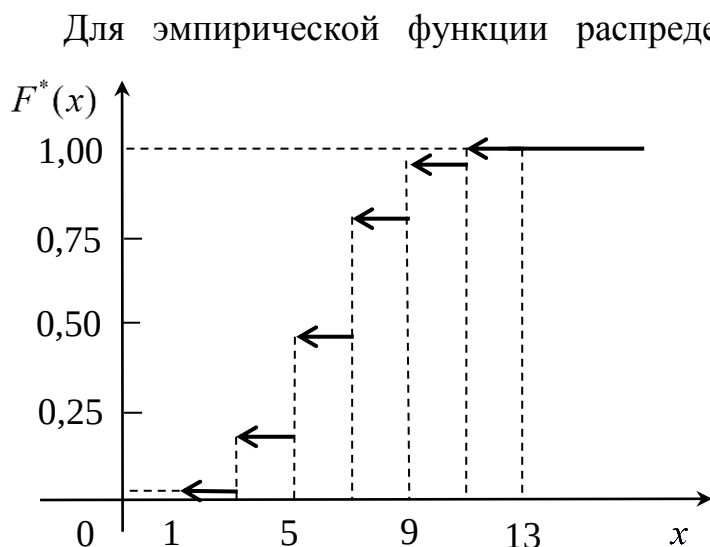


Рисунок 1.1.1 – Эмпирическая функция распределения случайной величины ξ

интегральной функции распределения. Действительно, из определения эмпирической функции распределения следует, что:

- 1) значения эмпирической функции распределения принадлежат отрезку $[0;1]$;
- 2) если x_1 – наименьшее, а x_n – наибольшее наблюдаемое значение, то $F^*(x) = 0$ при значении $x \leq x_1$ и $F^*(x) = 1$ при значении $x > x_n$.

В отличие от интегральной функции распределения $F(x)$ эмпирическая функция распределения $F^*(x)$ является неслучайной – для разных выборок она получается различной.

Основное значение эмпирической функции распределения состоит в том, что она используется в качестве оценки интегральной функции распределения.

Для наглядности вариационные ряды представляют графиками и диаграммами. Наиболее распространёнными графиками являются полигон, гистограмма и кумулята. Полигон применяется как для дискретных, так и для интервальных спастических рядов, а гистограмма – только для изображения интервальных рядов.

Определение 1.1.3 *Полигоном частот* называют ломаную линию, отрезки которой соединяют точки $(x_1; m_1)$, $(x_2; m_2)$, ..., $(x_k; m_k)$, где x_i – варианты выборки, а m_i – соответствующие им частоты.

Определение 1.1.4 *Полигоном относительных частот* или *полигоном частостей* называют ломаную линию, отрезки которой соединяют точки $(x_1; \omega_1)$, $(x_2; \omega_2)$, ..., $(x_k; \omega_k)$, где x_i – варианты выборки, а ω_i – соответствующие им относительные частоты $\omega_i = m_i/n$ или частости.

При непрерывном распределении случайной величины разбиваем весь интервал, в котором заключены все наблюдаемые значения случайной величины, на интервалы длиной h и определяем m_i – сумму частот вариантов, которые попали в i -й интервал.

Определение 1.1.5 *Гистограммой частот* называется ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h , а высоты равны отношению m_i/h .

Площадь частичного i -го прямоугольника равна $h \cdot m_i/h = m_i$ – сумме частот вариантов, попавших в i -й интервал. Площадь гистограммы частот равна сумме всех частот, то есть объёму выборки n .

Определение 1.1.6 *Гистограммой относительных частот* или *частостей* называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиной h , а высоты равны отношению $\omega_i/h = m_i/nh$.

Площадь частичного i -го прямоугольника равна $h \cdot \omega_i/h = \omega_i$ – относительной частоте вариантов, попавших в i -й интервал. Площадь гистограммы частостей равна сумме всех относительных частот, то есть единице. Если на гистограмме частостей соединить середины верхних сторон элементарных прямоугольников, то полученная замкнутая ломаная линия образует полигон распределения относительных частот. На рисунках 1.1.2 и 1.1.3 изображены полигоны и гистограммы частостей некоторых случайных величин ξ и η , соответственно.

Из принципа построения полигона и гистограммы относительных частот следует, что площадь под полигоном и гистограммой частостей равна единице. В теории вероятностей гистограмме и полигону относительных частот соответствует график плотности распределения вероятностей.

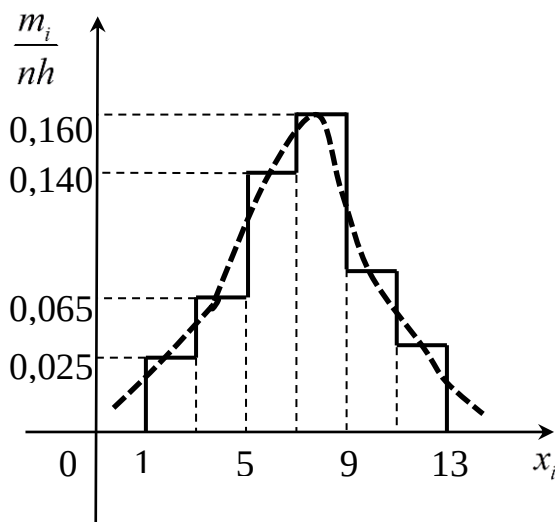


Рисунок 1.1.2 – Полигон и гистограмма частот СВ ξ

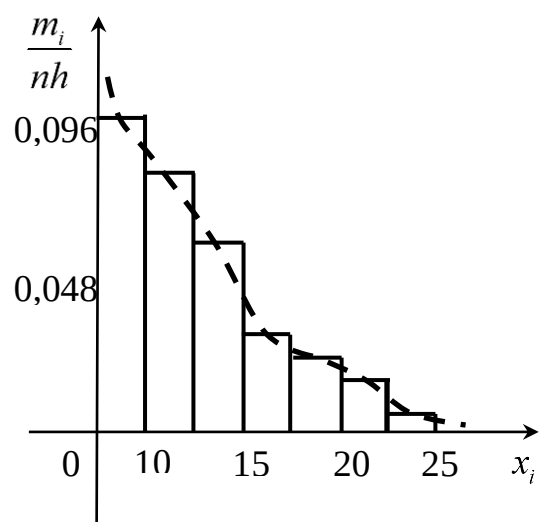


Рисунок 1.1.3 – Полигон и гистограмма частот СВ η

Гистограмму и полигон частот интервального статистического ряда удобно использовать для визуального подбора модели распределения. Так, гистограмма, изображённая на рисунке 1.1.2, напоминает кривую Гаусса, поэтому можно выдвинуть гипотезу о нормальном распределении случайной величины ξ . Вид гистограммы на рисунке 1.1.3 по форме напоминает экспоненциальную кривую. Поэтому можно выдвинуть гипотезу о показательном распределении случайной величины η .

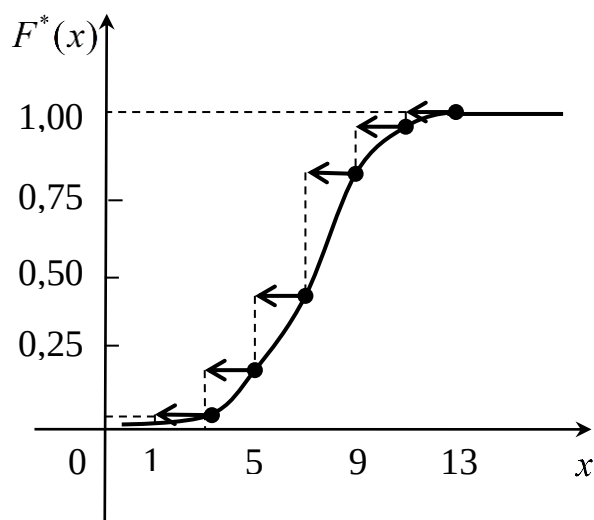


Рисунок 1.1.4 – Кумулята СВ ξ

Для построения кумуляты на оси абсцисс откладывают наблюдаемые значения случайной величины, а на оси ординат – накопленные относительные частоты, то есть накопленные частоты.

Определение 1.1.7 *Накопленной частотью* в точке x называется суммарная частота членов статистического ряда, значения которых меньше x .

Значения накопленных частот являются значениями эмпирической функции распределения $F^*(x)$. В теории

вероятностей кумуляте соответствует график функции распределения $F(x) = P(\xi < x)$. На рисунке 1.1.4 изображена кумулята интервального ряда некоторой случайной величины ξ сплошной линией. При построении используются значения эмпирической функции распределения $F^*(x)$.

СТАТИСТИЧЕСКАЯ ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

Предположим из генеральной совокупности с функцией распределения $F(x, \Theta)$, где Θ – неизвестный параметр, извлечена выборка $x_1, x_2, \dots, x_n$. Задача оценивания неизвестного параметра Θ состоит в построении приближённых формул

$$\Theta \approx u(x_1, x_2, \dots, x_n). \quad (2.1.1)$$

Функцию (2.1.1) называют *выборочной* или *статистической функцией*, а её значение в приближённом равенстве – *оценкой* и обозначают $\hat{\Theta} = u(x_1, x_2, \dots, x_n)$. Любая выборка является конечной и случайной, следовательно, статистические функции $\hat{\Theta}$ также являются случайными. Таким образом, будем рассматривать оценку $\hat{\Theta} = u(x_1, x_2, \dots, x_n)$ неизвестного параметра Θ как случайную величину, а её значение, вычисленное по заданной выборке объёмом n , как одну реализацию случайной величины, то есть как одно из множества её возможных значений.

В математической статистике рассматриваются точечные и интервальные оценки параметров распределения.

Точечная оценка параметра Θ определяется одним числом $\hat{\Theta} = u(x_1, x_2, \dots, x_n)$. *Интервальной оценкой* называют оценку, которая определяется двумя числами $\hat{\Theta}_1$ и $\hat{\Theta}_2$ – концами интервала, накрывающего оцениваемый параметр Θ .

Рассмотрим классификацию точечных оценок. Чтобы точечная оценка неизвестного параметра $\hat{\Theta} = u(x_1, x_2, \dots, x_n)$ была «хорошей» прежде всего с точки зрения точности и надёжности оценок, желательно, чтобы найденные на основании статистических функций $\hat{\Theta} = u(x_1, x_2, \dots, x_n)$ оценки неизвестных параметров были тесно сгруппированы около значений оцениваемых параметров распределения, то есть чтобы рассеивание случайной величины $\hat{\Theta}$ около параметра Θ было по возможности минимальным.

Определение 2.1.1 Оценка $\hat{\Theta} = u(x_1, x_2, \dots, x_n)$ называется *состоятельной*, если при увеличении числа испытаний оценка сходится по вероятности к оцениваемому параметру, то есть если $\lim_{n \rightarrow \infty} P(|\Theta - \hat{\Theta}| < \varepsilon) = 1$.

Определение 2.1.2 Оценка $\hat{\Theta} = u(x_1, x_2, \dots, x_n)$ называется *несмещённой* (оценкой без систематической ошибки), если её математическое ожидание равно оцениваемому параметру, то есть $M(\hat{\Theta}) = \Theta$.

Если последнее равенство не выполняется, то оценка называется *смещённой* (содержащей систематическую ошибку). Наряду с несмещёнными оценками

$\hat{\Theta} = u(x_1, x_2, \dots, x_n)$ применяются асимптотически несмещённые оценки, для которых $M(\hat{\Theta}) \rightarrow \Theta$ при увеличении объёма выборки.

Дисперсия любой несмещённой оценки одного параметра Θ удовлетворяет неравенству Рао-Крамера

$$D(\hat{\Theta}) \geq -\frac{1}{N \cdot M\left(\frac{\partial^2 \ln f(x, \Theta)}{\partial \Theta^2}\right)}, \quad (2.1.2)$$

где $f(x, \Theta)$ – плотность распределения вероятностей случайной величины; N – число произведённых испытаний. В случае дискретной случайной величины плотность распределения вероятностей $f(x, \Theta)$ заменяется функцией распределения вероятностей.

Определение 2.1.3 Оценка $\hat{\Theta} = u(x_1, x_2, \dots, x_n)$, для которой в неравенстве Рао-Крамера (2.1.2) достигается знак равенства, называется *эффективной*.

Определение 2.1.4 Оценка $\hat{\Theta} = u(x_1, x_2, \dots, x_n)$ называется *достаточной*, если она использует всю информацию относительно оцениваемого параметра, содержащуюся в выборке.

Достаточные оценки построены таким образом, что никакие другие оценки не могут дать какой-либо дополнительной информации об оцениваемых параметрах.

Рассмотрим точечные оценки неизвестных параметров распределения.

Несмещённой оценкой генеральной средней (математического ожидания) является *выборочная средняя*

$$\bar{x}_B = \frac{\sum_{i=1}^k m_i x_i}{n}, \quad (2.1.3)$$

где x_i – варианты выборки; m_i – частота варианты x_i ; $\sum_{i=1}^k m_i = n$.

Смещённой оценкой генеральной дисперсии служит *выборочная дисперсия*

$$D_B = \frac{\sum_{i=1}^k m_i (x_i - \bar{x}_B)^2}{n} = \overline{x_B^2} - \bar{x}_B^2. \quad (2.1.4)$$

Выборочная дисперсия является смещённой оценкой генеральной дисперсии D_Γ , так как $M(D_B) = \frac{n-1}{n} \cdot D_\Gamma \neq D_\Gamma$.

Положительное значение квадратного корня из выборочной дисперсии называется *выборочным средним квадратическим отклонением* $\sigma_B = \sqrt{D_B}$.

Несмещённой оценкой генеральной дисперсии служит «исправленная дисперсия», которая определяется по формуле

$$s^2 = \frac{n}{n-1} \cdot D_B = \frac{1}{n-1} \cdot \sum_{i=1}^k m_i (x_i - \bar{x}_B)^2. \quad (2.1.5)$$

Несмещённой оценкой среднего квадратического отклонения генеральной совокупности является «исправленное» среднее квадратическое отклонение

$$s = \sqrt{s^2} = \sqrt{\frac{n}{n-1} \cdot D_B} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^k m_i (x_i - \bar{x}_B)^2}. \quad (2.1.6)$$

Если рассматривается интервальный статистический ряд, то в качестве вариант берутся середины частичных интервалов.

Точечная оценка неизвестного параметра Θ , найденная по выборке объёма n из генеральной совокупности с функцией распределения $F(x, \Theta)$, не позволяет непосредственно ответить на вопрос, какую ошибку мы совершаем, принимая вместо точного значения неизвестного параметра Θ его приближённое значение или оценку $\hat{\Theta} = u(x_1, x_2, \dots, x_n)$.

В связи с этим более выгодно пользоваться интервальной оценкой, основанной на определении некоторого интервала, внутри которого с определённой вероятностью находится неизвестное значение параметра Θ . Предположим, что по результатам выборки объёма n найдена выборочная характеристика $\hat{\Theta} = u(x_1, x_2, \dots, x_n)$, которая является точечной оценкой неизвестного параметра Θ . Чем меньше разность $|\Theta - \hat{\Theta}|$, тем точнее будет оценка. Таким образом, положительное число ε характеризует точность оценки $|\Theta - \hat{\Theta}| < \varepsilon$. При этом точность ε зависит от объёма выборки. Поэтому можно только говорить о вероятности $1 - \alpha$, с которой это неравенство выполняется.

Определение 2.1.5 Доверительной вероятностью оценки называют вероятность $P = 1 - \alpha$ выполнения неравенства $|\Theta - \hat{\Theta}| < \varepsilon$.

Доверительная вероятность точечной оценки показывает, что при извлечении выборки одинакового объёма из одной и той же генеральной совокупности с функцией распределения $F(x, \Theta)$ в $(1 - \alpha)$ 100 % случаях параметр Θ будет покрываться данным интервалом.

Определение 2.1.6 Доверительным интервалом называется интервал $(\hat{\Theta} - \varepsilon; \hat{\Theta} + \varepsilon)$, покрывающий неизвестный параметр Θ с заданной доверительной вероятностью $P = 1 - \alpha$, с уровнем значимости α .

Рассмотрим случайную величину ξ , которая подчинена нормальному закону распределения.

Интервальной оценкой математического ожидания a случайной величины ξ , имеющей нормальное распределение при известном среднем квадратическом отклонении σ и выборочной средней $\bar{x}_B$, является доверительный интервал

$$\bar{x}_B - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < a < \bar{x}_B + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}. \quad (2.1.7)$$

Квантили нормального распределения $u_{\alpha/2}$ находят по таблицам функции Лапласа (приложение 3) из условия $2 \cdot \Phi(u_{\alpha/2}) = 1 - \alpha$. Доверительный интервал покрывает математическое ожидание a с заданной вероятностью или надёжностью $1 - \alpha$. Точность оценки математического ожидания (предельная погрешность) равна $\varepsilon = u_{\alpha/2} \sigma / \sqrt{n}$.

Интервальной оценкой математического ожидания a случайной величины ξ , имеющей нормальное распределение *при неизвестном среднем квадратическом отклонении* σ и известной выборочной средней $\bar{x}_B$, является доверительный интервал

$$\bar{x}_B - t_{\alpha/2; n-1} \cdot \frac{s}{\sqrt{n}} < a < \bar{x}_B + t_{\alpha/2; n-1} \cdot \frac{s}{\sqrt{n}}. \quad (2.1.8)$$

Квантили t -распределения $t_{\alpha/2; n-1}$ находят по таблицам распределения Стьюдента (приложение 5). Доверительный интервал покрывает математическое ожидание a с заданной вероятностью или надёжностью $1 - \alpha$. Точность оценки математического ожидания (предельная погрешность) равна $\varepsilon = t_{\alpha/2; n-1} s / \sqrt{n}$.

Интервальной оценкой среднего квадратического отклонения σ случайной величины ξ , имеющей нормальное распределение, является доверительный интервал

$$s \cdot \gamma_1 < \sigma < s \cdot \gamma_2, \quad (2.1.9)$$

$$\text{где } \gamma_1 = \sqrt{\frac{n-1}{\chi_{\alpha/2; n-1}^2}}, \gamma_2 = \sqrt{\frac{n-1}{\chi_{1-\alpha/2; n-1}^2}}, s = \sqrt{\frac{1}{n-1} \sum_{i=1}^k m_i (x_i - \bar{x}_B)^2}.$$

Коэффициенты γ_1 и γ_2 , соответствующие доверительной вероятности $P = 1 - \alpha$ и числу степеней свободы $\nu = n - 1$, находятся по таблице.

Так как $n \rightarrow \infty$, распределение χ^2 приближается к нормальному распределению, то при достаточно большом объёме выборки ($n \geq 50$) доверительный интервал для среднего квадратического отклонения можно найти по формуле

$$P \left(\frac{s}{1 + u_{\alpha/2} / \sqrt{2n}} < \sigma < \frac{s}{1 - u_{\alpha/2} / \sqrt{2n}} \right) = 1 - \alpha, \quad (2.1.10)$$

где $u_{\alpha/2}$ — квантиль стандартизированного нормального распределения, соответствующий доверительной вероятности $1 - \alpha$.

СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ

При решении экономических и инженерных задач приходится делать предположения о виде законов распределения случайных величин или о соотношении между их числовыми характеристиками. Такие предположения в математической статистике называют *гипотезами*.

Определение 3.1.1 Статистическая гипотеза называется *непараметрической*, если в ней сформулированы предположения относительно функции распределения.

Определение 3.1.2 Статистическая гипотеза называется *параметрической*, если в ней сформулированы предположения относительно значений параметров функции распределения исследуемой случайной величины.

Одновременно с выдвигаемой гипотезой законов распределения случайной величины можно выдвинуть и другие альтернативные гипотезы. Если выдвинутая гипотеза отвергается, то принимается другая альтернативная гипотеза. Исходя из этого предположения, гипотезы подразделяются на нулевые и альтернативные.

Определение 3.1.3 Нулевой гипотезой H_0 называют основную или выдвинутую первоначально гипотезу.

При выдвинутой нулевой гипотезе различие между сравниваемыми величинами отсутствуют, а наблюдаемые отклонения объясняются лишь случайными изменениями выборки.

Определение 3.1.4 Альтернативной называется гипотеза H_a , которая составляет конкуренцию нулевой гипотезе H_0 в том смысле, что если нулевая гипотеза отвергается, то за основу принимается альтернативная гипотеза.

Определение 3.1.5 Параметрическая гипотеза называется *простой*, если она содержит только одно предположение относительно параметров распределения.

Например, если a – математическое ожидание случайной величины ξ , то гипотеза $H_0 : a = 0$ является простой гипотезой.

Определение 3.1.6 Параметрическая гипотеза называется *сложной*, если она состоит из конечного или бесконечного числа простых гипотез.

Например, если a – математическое ожидание нормально распределённой случайной величины, то альтернативная гипотеза $H_a : a > 1$ является сложной, так как она состоит из бесконечного множества простых гипотез $H_0 : a = c$, где c – любое число большее единицы.

Определение 3.1.7 *Статистическим критерием* называют случайную величину K , с помощью которой принимаются решения о принятии или отвержении нулевой гипотезы.

Для проверки нулевых гипотез по выборочным данным рассчитываются частные значения входящих в критерий величин и таким образом получают частное или наблюдаемое значение критерия

Определение 3.1.8 *Ошибкой первого рода* называется ошибка отклонения верной нулевой гипотезы H_0 .

Определение 3.1.9 *Уровнем значимости* статистического критерия называется вероятность α совершения ошибки первого рода.

Определение 3.1.10 *Ошибкой второго рода* называется ошибка принятия ложной нулевой гипотезы H_0 .

Определение 3.1.11 Нулевой непараметрической гипотезой называется гипотеза относительного общего вида функции распределения случайной величины ξ , то есть гипотеза вида $H_0 : F(x) = F_0(x)$.

Проверка гипотезы о предполагаемом распределении производится на основании вычисления некоторой выборочной статистики (критерия), распределение которой получено из предположения истинности нулевой гипотезы и сравнения наблюдаемого значения этой выборочной статистики с критическим значением. Приняв ту или иную гипотезу, из неё выводят определённое следствие и рассматривают, насколько оно оправдывается на опыте, то есть проверяется согласие принятой гипотезы с опытом.

Для проверки нулевой гипотезы о виде закона распределения случайной величины рассмотрим один из наиболее распространённых критериев согласия, а именно критерий χ^2 Пирсона. Данный критерий позволяет проверить гипотезу о согласии данных выборки с конкретным распределением для любой случайной величины, как непрерывной, так и дискретной.

Критерий χ^2 – один из наиболее используемых критериев для проверки гипотезы о виде закона распределения случайных величин. Указанный критерий позволяет проводить проверку гипотезы соответствия опытного закона распределения теоретическому закону распределения не только в случаях, когда последний известен полностью, но и тогда, когда параметры теоретического закона распределения определяются на основе опытных данных. Таким образом, χ^2 Пирсона даёт возможность произвести проверку согласия эмпирической функции распределения с гипотетической функцией $F(x)$, принадлежащей к некоторому множеству функций определённого вида (нормальных, показательных, биномиальных и так далее).

Предположим, что проведено n независимых опытов, в каждом из которых случайная величина ξ приняла определённое значение. Результаты опытов оформляются в виде статистического интервального ряда, который содержит k интервалов, с указанием частот m_i , попадания случайной величины ξ в заданный интервал, где $i = \overline{1; k}$, а $\sum_{i=1}^k m_i = n$. При этом в каждом из интервалов желательно иметь 5 – 10 наблюдаемых значений случайной величины ξ .

Рассмотрим схему применения критерия χ^2 Пирсона.

1. Исходя из теоретического или предполагаемого закона распределения, находят вероятности p_i попадания случайной величины ξ в каждый из k полученных интервалов.

2. Вычисляем значение критерия χ^2 (выборочную статистику), которое соответствует опытным данным по формуле

$$\chi^2 = \sum_{i=1}^k \frac{(m_i - np_i)^2}{np_i}. \quad (3.1.1)$$

3. Определяем число ν степеней свободы распределения по формуле

$$\nu = k - r - 1. \quad (3.1.2)$$

где r - число параметров предполагаемого закона распределения, вычисленных опытным путём, а k - число частичных интервалов.

4. Критерий χ^2 построен таким образом, что чем ближе к нулю наблюдаемое значение χ^2 , тем вероятнее, что нулевая гипотеза справедлива. Поэтому для проверки нулевой гипотезы применяется критерий χ^2 с правосторонней критической областью. Для того чтобы проверить нулевую гипотезу, необходимо найти по таблицам квантилей χ^2 -распределения по заданному уровню значимости α и числу степеней свободы ν критическое значение $\chi_{\alpha; \nu}^2$, удовлетворяющее условию $P(\chi^2 \geq \chi_{\alpha; \nu}^2) = \alpha$. Сравнивая наблюдаемое значение выборочной статистики χ^2 , вычисленное по формуле (3.1.1), с критическим значением $\chi_{\alpha; \nu}^2$, принимаем одно из двух решений:

а) если $\chi_{набл}^2 \geq \chi_{\alpha; \nu}^2$, то нулевая гипотеза H_0 отвергается в пользу альтернативной H_a , то есть считается, что гипотетическая функция $F(x)$ не согласуется с результатами эксперимента;

б) если $\chi_{набл}^2 < \chi_{\alpha; \nu}^2$, то считается, что нет оснований для отклонения нулевой гипотезы H_0 , то есть гипотетическая функция $F(x)$ хорошо согласуется с результатами эксперимента.