Практика №2 ОСНОВНЫЕ ПОНЯТИЯМАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Задачи математической статистики

Математическая статистика возникла и создавалась параллельно с теорией вероятностей в XVII в. Дальнейшее развитие математической статистики (вторая половина XIX и начало XX в.) обязано, в первую очередь П. Л. Чебышеву, А. А. Маркову, А. М. Ляпунову и др.

Математическая статистика – раздел математики, который посвящен способам получения выводов из *опытных данных*(результатов наблюдений). Поскольку математическая статистика изучает общие закономерности массовых явлений в абстрактной форме, безразличной к природе рассматриваемых объектов, ее выводы могут быть применены к самым разнообразным явлениям при условии, что в них реально осуществляются основные положения обшей теоретической схемы.

Итак, задача математической статистики состоит в создании методов сбора и обработки статистических данных для получения научных и практических выводов.

Отсюда вытекают основные разделы математической статистики.

1. Теория оценок.

Эта теория позволяет приближенно вычислить и оценить параметры случайных величин (математическое ожидание, дисперсию и т.д.) *по данным опыта.*

2. Статистическая проверка гипотез.

Эта теория позволяет проверить справедливость интересующих нас гипотез *по данным опыта*.

3. Дисперсионный анализ.

Эта теория позволяет найти слабые (статистические) зависимости между величинами, т. е. направлена на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях. В данном пособии будут рассмотрены только первые два раздела.

Выборка из генеральной совокупности. Вариационный ряд. Гистограмма относительных частот

Определим основные понятия математической статистики. Предположим, что исследуется случайная величина X. Раз за разом воспроизводится опыт, в результате которого случайная величина X принимает какое-то значение. Но закон распределения X нам неизвестен. Как его определить? Очевидно, нет другой возможности, чем раз за разом проводить опыты, фиксировать значения, которые принимает случайная величина, и пытаться по ряду этих значений определить ее. закон распределения, хотя бы приближенный. Совокупность всех значений, которые может принять случайная величина в результате проводимых опытов, называется *генеральной совокупностью* G.

Генеральная совокупность - достаточно большая совокупность однородных объектов, некоторая характеристика которых меняется случайным образом, от объекта к объекту. Наша цель научиться определять, как распределена эта характеристика по всей генеральной совокупности.

Исследовать все элементы генеральной совокупности невозможно и нецелесообразно, т.к. генеральная совокупность может быть очень велика. В связи с этим из генеральной совокупности отбирают определенное число элементов:

$$x_1, \quad x_2, \ldots, \; x_n. \tag{2.1}$$

и их изучают. Результаты исследований показали, что выводы, сделанные по этим элементам, точно характеризуют данное явление. Такой метод называется выборочным. Часть объектов, которая попала на проверку, называется *выборочной совокупностью* X или просто *выборкой*.

Число элементов в генеральной совокупности и в выборке будем называть их объемами. Объем генеральной совокупности равен *N*, а выборки – *n*. Набор чисел (2.1) называют *элементами выборки* или *вариантами*. Расположенные в порядке возрастания значения $x_1, \quad x_2, \ldots, \; x_n$ образуют ряд, называемый *вариационным рядом:*

$x_{\min}, \quad \ldots \quad , x_{\max}$ – вариационный ряд.

Число $R = x_{\max} - x_{\min}$ называется *размахом выборки*.

Для построения вариационного ряда выполним действия:

1. Разделим отрезок $[x_{\min}, \quad x_{\max}]$ на некоторое число *l* интервалов одинаковой длины $\Delta = \left(\dfrac{R}{k}\right)$. Величину *к* приближенно вычисляют по формуле $l = [1 + 3{,}2 \lg n]$, где *n* – объём выборки.

$x_i = x_{i-1} + i\Delta$.

2. Подсчитаем число элементов выборки, попадающих в каждый интервал:

$$n_1, n_2, \quad \ldots \quad , n_m. \tag{2.2}$$

Очевидно, $n_1 + n_2 + \ldots + n_m = n$.

Числа (2.2) называются *частотами попадания в интервал.* Составим табл. 22.

Таблица 22

$x_0 - x_1$	$x_1 - x_2$		$x_{n-1} - x_n$
n_i	n_1		n_m
$\omega_i = \dfrac{n_i}{n}$	$\dfrac{n_1}{n}$		$\dfrac{n_m}{n}$

Элементы третей строки называются *относительными частотами попадания в интервал.* Очевидно, $\frac{n_1}{n} + \frac{n_2}{n} + \ldots + \frac{n_m}{n} = 1$.

Полученная таблица называется *выборочным распределением случайной величины* X .

3) Изобразим выборочное распределение на графике (рис. 2.1). Для этого на оси OX отложим интервалы $x_i - x_{i+1}$ и на каждом из них как на основании построим прямоугольник, площадь которого ω_i, т.е. высота i-го прямоугольника $\frac{\omega_i}{\Delta}$.

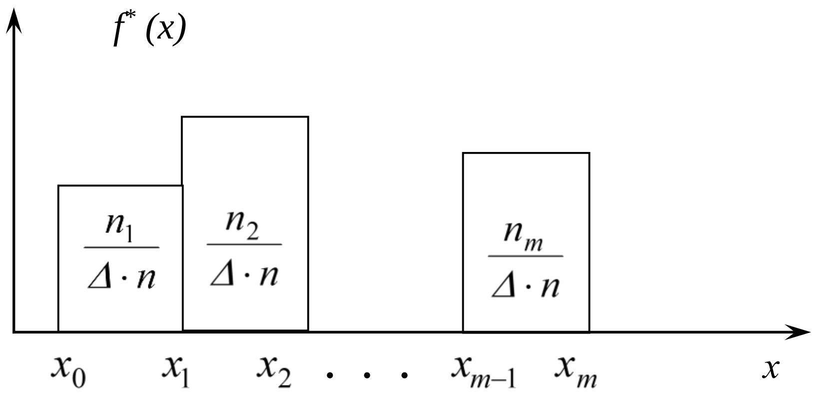


Рис. 2.1. Гистограмма относительных частот

Построенный график называется *гистограммой относительных частот* и представляет собой *выборочный аналог* плотности вероятности случайной величины. Площадь гистограммы равна единице.

Пример 1.

При проведении 20-ти серий из 10-ти бросков игральной кости число выпадений шести очков оказалось равным 1,1,4,0,1,2,1,2,2,0,5,3,3,1,0,2,2,3,4,1. Составим вариационный ряд: 0,1,2,3,4,5. Статистический ряд для абсолютных и относительных частот имеет вид табл. 23.

Таблица23

x_i	0	1	2	3	4	5

n_i	3	6	5	3	2	1
$\omega_i = \frac{n_i}{n}$	$\frac{3}{20}$	$\frac{6}{20}$	$\frac{5}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$

Числовые характеристики статистического распределения: выборочное среднее, дисперсия

4.1. Параметры распределения

Одна из задач математической статистики: по имеющейся выборке оценить значения числовых характеристик исследуемой случайной величины. Основными числовыми характеристиками выборки являются выборочное среднее $\bar{x}_в$ и выборочная дисперсия $D_в$.

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

$$\bar{x}_в = \frac{x_1 + \ldots + x_n}{n} \text{ или } \bar{x}_в = \frac{\sum_{i=1}^{l} x_i n_i}{n}. \quad (2.3)$$

Выборочное среднее $\bar{x}_в$ служит для точечной оценки математического ожидания $M(X)$ исследуемой случайной величины

$$\bar{x} = \frac{x_1 + \ldots + x_n}{n}$$

Определение. *Выборочной дисперсией* называется

$$D_в(X) = \frac{\sum_{i=1}^{l} \left(x_i - \bar{x}_в\right)^2 \cdot n_i}{n} = \frac{\sum_{i=1}^{l} x_i^{\ 2} \cdot n_i}{n} - \bar{x}_в^{\ 2}. \quad (2.4)$$

Выборочная дисперсия $D_в(X)$ служит для точечной оценки дисперсии $D(X)$.

В отличие от выборочного среднего, выборочная дисперсия является смещенной оценкой дисперсии генеральной совокупности. Пользуясь оценкой $D_в$ вместо D, мы будем совершать некоторую систематическую ошибку, так как её математическое ожидание несколько меньше истинного значения. Чтобы её ликвидировать, достаточно ввести поправку, умножив $D_в$ на $\frac{n}{n-1}$. Таким образом, можно предложить другую оценку дисперсии – *исправленную дисперсию* s^2, вычисляемую по формуле $s^2 = \frac{n}{n-1} D_в = \frac{\sum_{i=1}^{l} (x_i - \bar{x}_в)^2 n_i}{n-1}$. Эта оценка является состоятельной несмещенной оценкой дисперсии D.

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

$$s=\sqrt{s^2}=\sqrt{\frac{\sum_{i=1}^{l}(x_i-\bar{x}_в)^2 n_i}{n-1}}. \qquad (2.5)$$

При больших n поправка $\frac{n}{n-1}$ становится близкой к единице и её применение теряет смысл.

Пример 1. Оптический пирометр установлен на светящуюся нить накала, различными операторами было произведено несколько измерений температуры. Получены следующие результаты (табл.25).

Таблица 25

Температура, °C	925	950	975	1000	1025	500
Число измерений	1	9	6	18	10	2

Требуется найти среднюю квадратическую и вероятную ошибки в предположении, что эта выборка взята из нормально распределенной совокупности.

Решение. Найдем сначала выборочное среднее значение $\bar{x}_в$.

Проще находить среднее для $1000 - x$, а не для x (табл. 26).

Таблица 26

$1000–x_i$	Число измерений n	$n\cdot(1000–x_i)$	$(1000–x_i)^2$	$n\cdot(1000–x_i)^2$
1	2	3	4	5
75	1	75	5625	5625
50	9	450	2500	22500
25	6	150	625	3750
0	18	0	0	0
-25	10	-250	625	6250
-50	2	-100	2500	5000
$\sum_{i=1}^{l}$	46	325	–	43125

$$1000-\bar{x}_в=\frac{\sum_{i=1}^{l}x_i n_i}{n}=\frac{325}{46}=7{,}1 \text{ или } \bar{x}_в=1000-7{,}1=992{,}9°.$$

$$D_в(X)=\frac{\sum_{i=1}^{l}x_i^2\cdot n_i}{n}-\bar{x}_в^2=\frac{43125}{46}-7{,}1^2=887.$$

Найдем исправленную дисперсию s^2,

$$s^2 = \frac{n}{n-1} D_в = \frac{46}{45} \cdot 887 = 906{,}7\ ^{\circ}\text{C}.$$

Статистическая проверка статистических гипотез

Сопоставление высказанной гипотезы относительно генеральной совокупности с имеющимися выборочными данными, сопровождаемое количественной оценкой степени достоверности получаемого вывода и осуществляемое с помощью того или иного статистического критерия, называется проверкой статистических гипотез.

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

Определение. *Нулевой (основной)* называют выдвинутую гипотезу H_0. *Конкурирующей (альтернативной)* называют гипотезу H_1, которая противоречит нулевой.

Если выдвигаемая гипотеза сводится к утверждению о том, что значение некоторого неизвестного параметра генеральной совокупности в точности равно заданной величине, то эта гипотеза называется простой. В других случаях гипотеза называется сложной.

Выдвинутая нулевая гипотеза может быть правильной или неправильной, поэтому возникает необходимость статистической проверки справедливости основной гипотезы H_0 (табл.27).

Таблица 27

Нулевая гипотеза H_0	Результаты решения относительно H_0	
	Отклонена	Принята
Верна	Ошибка 1-го рода, её вероятность $P(H_1/H_0) = \alpha$	Правильное решение, его вероятность $P(H_0/H_0) = 1 - \alpha$
Неверна	Правильное решение, его вероятность $P(H_1/H_1) = 1 - p$	Ошибка 2-го рода, её вероятность $P(H_0/H_1) = p$

Так как проверка статистических гипотез осуществляется на основании выборочных данных, то решение неизбежно сопровождается вероятностью ошибочного заключения как в ту, так и в другую сторону.

В какой-то небольшой доле случаев нулевая гипотеза может оказаться отвергнутой, в то время как она справедлива. Такую ошибку называют *ошибкой первого рода*, а её вероятность – *уровнем значимости* α. Другими словами это та вероятность, которой можно пренебречь.

Наоборот, в какой-то небольшой доле случаев нулевая гипотеза принимается, в то время как на самом деле она ошибочна. Такую ошибку называют *ошибкой второго рода*. Вероятность ошибки второго рода p. Вероятность $1-p$ называют мощностью критерия.

Для проверки нулевой гипотезы пользуются специально подобранной случайной величиной, распределение которой известно. В общем случае её обозначают K – критерий согласия, устанавливающий, когда полученное в действительности указанное отклонение следует принять несущественным, а когда существенным. Критерий K является функцией от результатов наблюдения.

Критерий Пирсона для проверки гипотезы о виде закона распределения случайной величины

С помощью критерия Пирсона можно проверять гипотезы о различных законах распределения. Для данного критерия в качестве величины, характеризующей степень различия между теоретическим и выборочным законами распределения, выбирается статистика

$$K = \chi^2 = \sum_{i=1}^{l} \frac{\left(n_i - n_i^*\right)^2}{n_i}, \tag{2.8}$$

которая учитывает расхождения между теоретическими n_i и выборочными n_i^* частотами. Эта случайная величина называется χ^2 («хи квадрат») статистикой Пирсона. Смысл ее очевиден: суммируются квадраты отклонений эмпирических частот от теоретических. Целью исследования является сравнение эмпирических и теоретических частот, которые, конечно, отличаются друг от друга, и выяснить, являются ли эти различия несущественными, не опровергающими гипотезу о выбранном законе распределении исследуемой случайной величины, или они настолько велики, что противоречат этой гипотезе.

Можно доказать, что вне зависимости от реального закона распределения генеральной совокупности, закон распределения случайной величины при $n \to \infty$ стремится к закону распределения с числом степеней свободы $k = l - 1 - r$, где r – число параметров предполагаемого распределения, оцененных по данным выборки, l-число интервалов. Для выбранного критерия строится критическая область, определяемая условием

$$P\left(\chi_{\text{выб.}}{}^2 < \chi^2_{\text{крит}}(l;\alpha)\right) = \alpha,$$

где α – уровень значимости. Следовательно, критическая область задается неравенством $\chi_{\text{выб.}}{}^2 < \chi^2_{\text{крит}}(l;\alpha)$ а область принятия гипотезы

$$\chi_{\text{выб.}}{}^2 < \chi^2_{\text{крит}}(l;\alpha).$$

Таким образом, для проверки нулевой гипотезы H_0 *генеральная совокупность распределена по выбранному закону* – нужно вычислить по выборке наблюдаемое значение критерия:

$$\chi^2_{\text{выб.}} = \sum_{i=1}^{l} \frac{\left(n_i - n_i^*\right)^2}{n_i},$$

а по таблице критических точек распределения χ^2 найти критическую точку $\chi^2_{крит}(l;\alpha)$, используя известные значения α и $k = l - 1 - r$.

Если $\chi_{выб.}{}^2 < \chi^2_{крит}(k;\alpha)$ нулевую гипотезу принимают, при $\chi_{выб.}{}^2 > \chi^2_{крит}(k;\alpha)$ ее отвергают.

Замечание. Число параметров распределения *r*, оцененных по данным выборки для нормального и равномерного распределений равно двум, а для показательного – одному.

Из вышесказанного следует, что для поверки данной гипотезы H_0 необходимо найти теоретические и эмпирические частоты.

Пусть получена выборка достаточно большого объема *n* с большим количеством различных значений вариант. Для удобства её обработки разделим интервал от наименьшего до наибольшего из значений вариант на *k* равных частей по методике, предложенной выше. Будем считать, что значения вариант, попавших в каждый интервал, приближенно равны числу, задающему середину интервала. Подсчитав число вариант, попавших в каждый интервал, составим так называемую сгруппированную выборку (табл. 28).

Таблица 28

варианты $\bar{x}_i$	$\tilde{x}_1$	$\tilde{x}_1$	…		$\tilde{x}_k$
частоты n_i	n_1	n_2	…		n_k

Примечание. $\tilde{x}_i = \dfrac{x_i + x_{i+1}}{2}$ – значения середин интервалов, а n_i – число вариант, попавших в *i*-й интервал (эмпирические частоты).

По полученным данным можно вычислить выборочное среднее $\bar{x}_в$ и выборочное среднее квадратическое отклонение $\sigma_в$. Тип закона определим по построенной гистограмме или из общих соображений. Например, ошибки измерений в основном распределены по нормальному закону, а надежность безотказной работы прибора – по показательному. Для проверки предположения, что генеральная совокупность распределена по выбранному закону с параметрами $M(X) = \bar{x}_в$; $D(X) = \sigma_в^2$, необходимо вычислить теоретические частоты по формуле

$$n_i = p_i \cdot n,$$

где $p_i = p_i\left(x_i < X < x_{i+1}\right) = \int\limits_{x_i}^{x_{i+1}} f(x)dx$ – вероятность попадания в i-й интервал , n – объем выборки.

Здесь и далее $f(x)$ - функция плотности распределения вероятностей случайной величины X, выбранного закона распределения. Для простоты вычислений можно воспользоваться приближенной формулой вычисления определенного интеграла

$$p_i = p_i\left(x_i < X < x_{i+1}\right) = \int\limits_{x_i}^{x_{i+1}} f(x)dx = f\left(\tilde{x}_i\right)\cdot\left(x_{i+1} - x_i\right) = f\left(\tilde{x}_i\right)\cdot h, \qquad (2.9)$$

где $\tilde{x}_i = \dfrac{x_i + x_{i+1}}{2}$.

Рассмотрим практические приёмы нахождения теоретических вероятностей и теоретических частот для основных законов распределения.

1. *Нормальный закон распределения.*

Функция плотности распределения задается формулой

$$f(x) = \frac{1}{\sigma_в\sqrt{2\pi}}\cdot e^{\frac{-(x-\bar{x}_в)^2}{2\sigma_в{}^2}} = \frac{1}{\sigma_в}\cdot\varphi\left(\frac{x-\bar{x}_в}{\sigma_в}\right).$$

График функции $f(x)$называют нормальной кривой или кривой Гаусса (рис. 2.3).

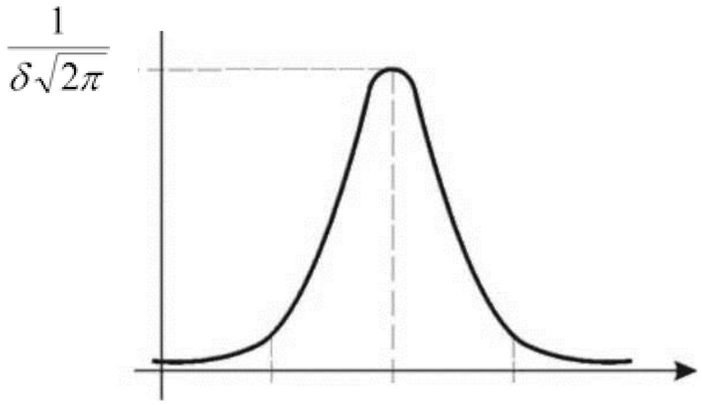


Рис. 2.3. Кривая Гаусса

Тогда вероятности p_i вычисляются по формуле

$$p_i = p_i\left(x_i < X < x_{i+1}\right) = \int\limits_{x_i}^{x_{i+1}} f(x)dx = \Phi\left(\frac{x_{i+1} - \bar{x}_в}{\sigma_в}\right) - \Phi\left(\frac{x_i - \bar{x}_в}{\sigma_в}\right),$$

где $\Phi(x) = \dfrac{1}{2\pi}\int\limits_0^x e^{-\frac{z^2}{2}}dz$– функция Лапласа, её значения находят по прил. 2.

Чаще всего пользуются приближенной формулой

$$p_i = p_i(x_i < X < x_{i+1}) = \int_{x_i}^{x_{i+1}} f(x)dx = \frac{1}{\sigma_в} \cdot \varphi\left(\frac{\tilde{x}_i - \bar{x}_в}{\sigma_в}\right) \cdot \Delta,$$

где $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$ – табулированная функция (см. прил. 1).

Умножив полученные вероятности на объем выборки n, найдем теоретические частоты: $n_i = p_i \cdot n$.

2. *Равномерный закон распределения*.

При использовании критерия Пирсона для проверки гипотезы о равномерном распределении генеральной совокупности с предполагаемой плотностью вероятности

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

необходимо, вычислив по имеющейся выборке значения $\bar{x}_в$ и $\sigma_в^2$, оценить параметры a и b.

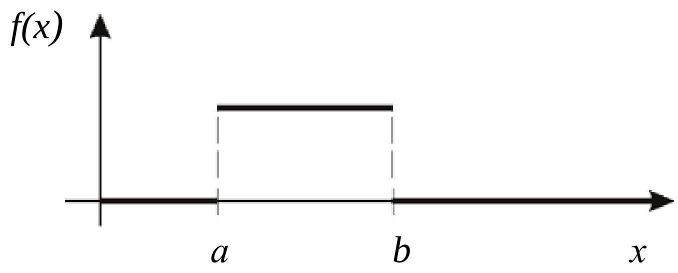


Рис. 2.4. График функции плотности равномерного распределения

Действительно, так как для равномерного распределения $M(X) = \frac{a+b}{2}$ и $\sigma(X) = \sqrt{D(X)} = \sqrt{\frac{(a-b)^2}{12}} = \frac{a-b}{2\sqrt{3}}$, то оценки параметров a и b – концов интервала, в котором наблюдались возможные значения случайной величины X, находим из системы (2.10), где через a^* и b^* обозначены оценки соответствующих параметров равномерного распределения.

$$\begin{cases} \dfrac{b^* + a^*}{2} = \bar{x}_в; \\ \dfrac{b^* - a^*}{2\sqrt{3}} = \sigma_в. \end{cases} \quad (2.10)$$

Решая систему уравнений, получим

$$a^* = \bar{x}_в - \sqrt{3}\sigma_в, \; b^* = \bar{x}_в + \sqrt{3}\sigma_в.$$

Тогда теоретическое распределение будет иметь вид

$$f(x) = \begin{cases} 0, \quad x \notin \left[a^*, b^*\right]; \\ \dfrac{1}{b^* - a^*}, x \in \left[a^*, b^*\right]. \end{cases}$$

Вероятности p_i вычисляются по формуле:

$$p_i = p_i\left(x_i < X < x_{i+1}\right) = \int\limits_{x_i}^{x_{i+1}} f\left(x\right)dx = \int\limits_{x_i}^{x_{i+1}} \frac{1}{b^* - a^*}dx = \frac{x_{i+1} - x_i}{b^* - a^*} = \frac{\Delta}{b^* - a^*}.$$

Умножив полученные вероятности на объем выборки n, найдем теоретические частоты: $n_i = p_i \cdot n$.

3. *Показательный закон распределения*.

Функция плотности определяется формулой

$$f(x) = \begin{cases} 0, x < 0; \\ \lambda e^{-\lambda x}, x \geq 0. \end{cases}$$

Вычислив по имеющейся выборке значения $\bar{x}_в$ и $\sigma_в^2$, необходимо оценить параметр λ. Для показательного распределения $M(X) = \dfrac{1}{\lambda}$. Следовательно, в качестве оценки параметра λ возьмём величину $\lambda^* = \dfrac{1}{\bar{x}_в}$. Тогда теоретическое распределение будет иметь вид

$$f(x) = \begin{cases} 0, \; x < 0; \\ \dfrac{1}{\bar{x}_в} e^{-\frac{1}{\bar{x}_в}x}, x \geq 0. \end{cases}$$

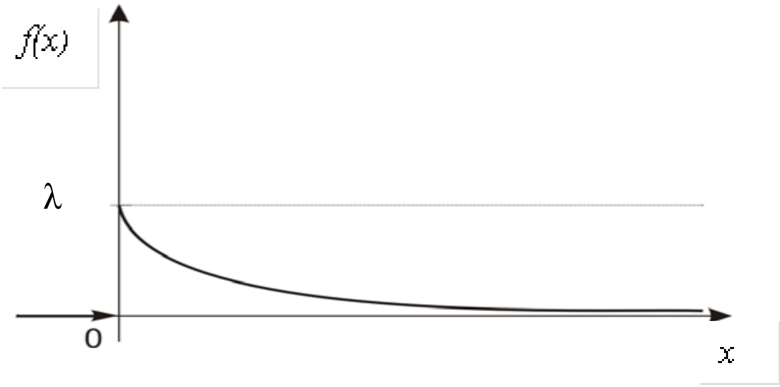


Рис. 2.5. График функции плотности равномерного распределения

Вероятности p_i вычисляются по формуле

$$p_i = p_i\left(x_i < X < x_{i+1}\right) = \int_{x_i}^{x_{i+1}} f\left(x\right)dx = \int_{x_i}^{x_{i+1}} \frac{1}{\bar{x}_в} e^{-\frac{1}{\bar{x}_в}x} dx = e^{-\frac{1}{\bar{x}_в}x_i} - e^{-\frac{1}{\bar{x}_в}x_{i+1}}.$$

Чаще всего пользуются приближенной формулой

$$p_i = p_i\left(x_i < X < x_{i+1}\right) = \int_{x_i}^{x_{i+1}} f\left(x\right)dx = \frac{1}{\bar{x}_в} e^{-\frac{1}{\bar{x}_в}\tilde{x}_i} \cdot \Delta,$$

где $\tilde{x}_i = \dfrac{x_i + x_{i+1}}{2}$.

Тогда теоретические частоты вычисляются по формуле

$$n_i = p_i \cdot n,$$

где n – объем выборки.

Примеры проверки гипотез о законе распределения выборочных данных

Задача 7.

В ходе проведения эксперимента получен следующий набор данных:

55,42	67,49	57,71	64,59	56,01	70,97	71,53	47,66	67,70	82,75
40,89	29,04	59,59	97,18	51,00	67,15	62,16	52,77	53,26	33,04
68,22	96,22	46,60	51,25	58,66	65,12	67,98	61,10	60,44	65,73
53,19	69,11	71,90	71,24	83,94	74,64	73,35	50,80	75,48	59,12
89,03	60,87	60,01	46,90	54,85	27,21	72,91	45,28	49,57	44,11
67,54	78,21	54,19	65,35	26,81	70,84	34,52	60,96	76,75	63,58
93,89	44,32	54,91	48,84	63,08	68,11	71,08	72,17	80,42	59,43
55,41	70,35	62,28	22,61	63,95	100,46	54,59	79,99	41,43	63,39
80,67	62,73	48,82	38,49	77,63	52,98	62,16	43,78	65,55	56,26
42,33	58,28	51,16	83,50	45,74	49,66	53,69	54,96	67,58	79,60

Пусть случайная величина X –наугад взятое значение из набора данных. Требуется для случайной величины X:

1..Составить выборочное распределение.

2. Построить гистограмму и график выборочной функции распределения.

3. Найти состоятельные несмещенные оценки математического ожидания и дисперсии.

4. Построить доверительные интервалы для математического ожидания и дисперсии с уровнем доверия p=0,95.

5. На основании анализа формы построенной гистограммы выдвинуть гипотезу о законе распределения и проверить справедливость гипотезы по критерию Пирсона с уровнем значимости α=0,05.

Решение.

1. *Первым этапом статистического изучения вариации являются построение вариационного ряда* - упорядоченного распределения единиц совокупности по возрастающим (чаще) или по убывающим (реже) значениям признака и подсчет числа единиц с тем или иным значением признака. Для этого сначала построим ранжированный ряд. Ранжированный ряд — это перечень отдельных единиц совокупности в порядке возрастания (убывания) изучаемого признака.

22,61	43,78	49,57	53,69	57,71	61,10	65,12	68,11	72,17	80,42
26,81	44,11	49,66	54,19	58,28	62,16	65,35	68,22	72,91	80,67
27,21	44,32	50,80	54,59	58,66	62,16	65,55	69,11	73,35	82,75
29,04	45,28	51,00	54,85	59,12	62,28	65,73	70,35	74,64	83,50
33,04	45,74	51,16	54,91	59,43	62,73	67,15	70,84	75,48	83,94
34,52	46,60	51,25	54,96	59,59	63,08	67,49	70,97	76,75	89,03
38,49	46,90	52,77	55,41	60,01	63,39	67,54	71,08	77,63	93,89
40,89	47,66	52,98	55,42	60,44	63,58	67,58	71,24	78,21	96,22
41,43	48,82	53,19	56,01	60,87	63,95	67,70	71,53	79,60	97,18
42,33	48,84	53,26	56,26	60,96	64,59	67,98	71,90	79,99	100,46

Интервальным вариационным рядом называется упорядоченная совокупность интервалов варьирования значений случайной величины с соответствующими частотами или частностями попаданий в каждый из них значений величины.

Для его построения выполняем следующие действия:

1. Находим размах выборки

$$R = x_{наиб} - x_{наим} = 100.46 - 22.61 = 77.85$$

2. Назначаем число частичных интервалов к:

$$к = 1 + 3{,}322 \ln n.$$

Обычно $k = 9 \div 12$. Выберем к=12

3.Находим длину Δ(шаг разбиения):

$$\varDelta = \frac{R}{k},\ \varDelta = \frac{77.85}{12} = 6.5,\ a_1 = 22.61$$

4. Численность отдельной группы сгруппированного ряда опытных данных называется выборочной частотой. Обозначается:

m_i^* – выборочная частота.

$$\sum_{i=1}^{k} m_i^* = 1$$

Относительная выборочная частота-отношение выборочной частоты данных вариантов к объёму выборки. Обозначается:

p_i^* – относительная выборочная частота.

$p_i^* = \dfrac{m_i^*}{n}$, где i – номер варианты.

Выборочная относительная частота сходится по вероятности к соответствующей вероятности.

$$\sum_{i=1}^{k} p_i^* = 1$$

5. Составляем таблицу.

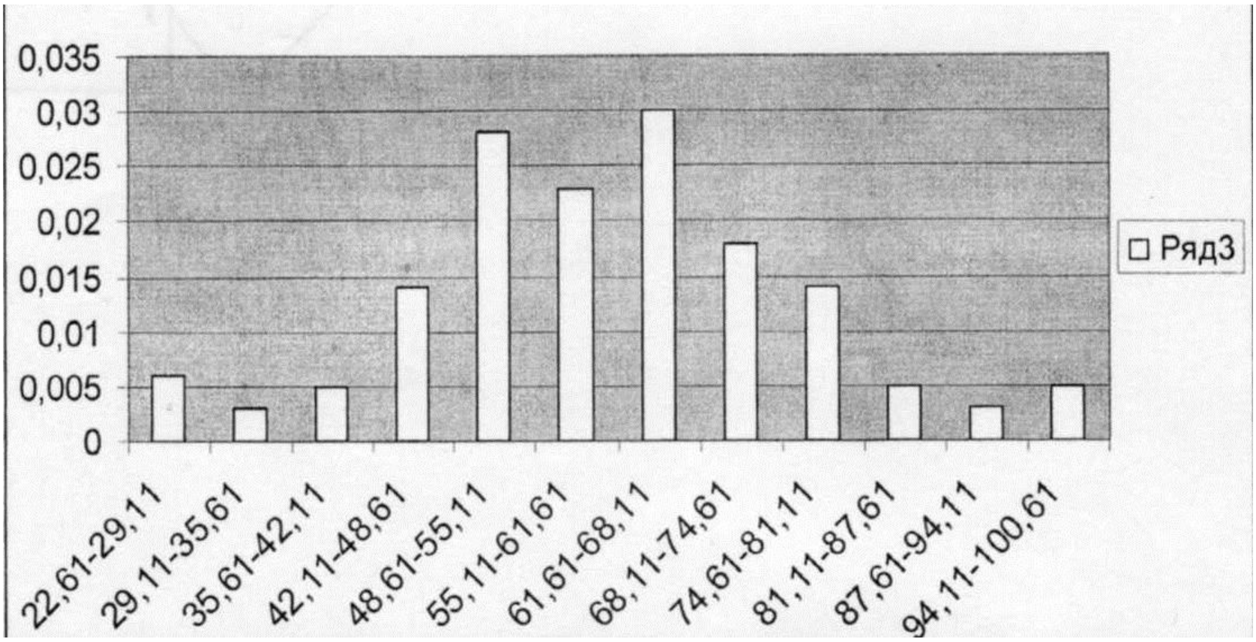
i	$[a_i; a_{i+1}]$	m_i^*	p_i^*	$h_i^* = \frac{p_i^*}{\varDelta}$	$\tilde{x}_i = \frac{a_i + a_{i+1}}{2}$
1	22,61-29,11	4	0,04	0,006	25,86
2	29,11-35,61	2	0,02	0,003	32,36
3	35,61-42,11	3	0,03	0,005	38,86
4	42,11-48,61	9	0,09	0,014	45,36
5	48,61-55,11	18	0,18	0,028	51,86
6	55,11-61,61	15	0,15	0,023	58,36
7	61,61-68,11	20	0,2	0,03	64,86
8	68,11-74,61	12	0,12	0,018	71,36
9	74,61-81,11	9	0,09	0,014	77,86
10	81,11-87,61	3	0,03	0,005	84,36
11	87,61-94,11	2	0,02	0,003	90,86
12	94,11-100,61	3	0,03	0,005	97,36
		$\sum = 100$	$\sum = 1$		

Где h_i^* – плотность относительной частоты;

$\tilde{x}_i$ – середина частичных интервалов.

2. *Построение гистограммы плотностей относительных частот.*

Гистограмма относительных частот — это фигура, состоящая из прямоугольников, опирающихся на интервалы группировки с высотами равными $h_i^* = \dfrac{p_i^*}{\Delta}$. Площадь всей гистограммы должна быть равна 1. Гистограмма является оценкой генеральной функции плотности $f(x)$.



По виду гистограммы мы подбираем подходящий для данного случая теоретический закон распределения:

Сравниваем гистограмму с теоретическими кривыми основных законов (нормальный, показательный, равномерный).

По виду гистограммы можно выдвинуть гипотезу о нормальном законе случайной величины X.

3. Нахождение состоятельных несмещенных оценок математического ожидания и дисперсии. Найдем оценки математического ожидания a^* и дисперсии D^*.

Основными параметрами генеральной совокупности являются математическое ожидание (генеральная средняя) $M(X)$ и среднее квадратическое отклонение s . Это постоянные величины, которые можно оценить по выборочным данным. Оценка генерального параметра, выражаемая одним числом, называется точечной.

Точечной оценкой генеральной средней а является выборочное среднее. Выборочным средним называется среднее арифметическое всех значений величины, встречающихся в выборке.

Если выборочное среднее вычисляется по не сгруппированным данным, то для его определения сумму всех значений делят на количество элементов в выборке. В данном случае определяем по сгруппированным данным(см. табл.):

$$a^* = \bar{x} = \frac{\sum_{i=1}^{k} \tilde{x}_i \cdot m_i^*}{n} = \frac{6083}{100} = 60.83,$$

$$D^* = \bar{s}^2 = \frac{\sum_{i=1}^{k} (\tilde{x}_i - \bar{x})^2 \cdot m_i^*}{n} = \frac{23387.9}{99} = 236.24,$$

$$\bar{s} = \sqrt{\bar{s}^2} = \sqrt{36.24} = 15.37\,.$$

$(\tilde{x}_i - \bar{x})^2 \cdot m_i^*$	$\tilde{x}_i \cdot m_i^*$
4891,6	103,44
1621,08	64,72
1448,04	116,58
2153,89	408,24
1448,3	933,48
91,51	875,4
324,82	1297,2
1330,57	856,32
2610,19	700,74
1660,98	253,08
1803,60	181,72
4003,32	292,08
$\sum = 23387.9$	$\sum = 6083$

$a \approx \bar{x}, \;\; \sigma \approx \bar{s}.$

4. *Построим доверительные интервалы для математического ожидания и дисперсии.*

Найдем доверительный интервал для оценки с надежностью $\gamma = 0{,}95$ неизвестного математического ожидания $M(x)$ нормально распределенного признака X генеральной совокупности, если генеральное стандартное отклонение $\bar{s} = 15{,}37$, выборочная средняя $\bar{x} = 60{,}83$, объем выборки $n = 100$.

Требуется найти доверительный интервал

$$\bar{x} - t \times \frac{\bar{s}}{\sqrt{n}} < M(x) < \bar{x} + t \times \frac{\bar{s}}{\sqrt{n}}$$

Все величины, кроме t, известны. Найдем t из соотношения

$$\Phi(t) = \frac{\gamma}{2} = \frac{0,95}{2} = 0,475$$

По таблице находим $t = 1,96$.

Подставив $t = 1,96$, $\bar{x} = 60,83$, $\bar{s} = 15,37$, $n = 100$, окончательно получим доверительный интервал

Для *M(x)*:

$$J_\gamma = \left(\bar{x} - t \cdot \frac{\bar{s}}{\sqrt{n}} ; \bar{x} + t \cdot \frac{\bar{s}}{\sqrt{n}} \right),$$

$$J_\gamma = (60.1;61.56).$$

Для *D(x)*:

$$J_\gamma = \left(\bar{s}^2 - t \cdot \sigma_{\bar{s}^2} ; \bar{s}^2 + t \cdot \sigma_{\bar{s}^2}\right), \quad J_\gamma = (220.33;252.15).$$

$$\sigma_{\bar{s}^2} = \sqrt{\frac{2}{n-1} \cdot \bar{s}^2} = 33.5.$$

5. *Построенная гистограмма по форме напоминает график плотности вероятности нормального распределения*. Поэтому естественно выдвинуть гипотезу о нормальном распределении случайной величины *X* Проверим справедливость выдвинутой гипотезы по критерию Пирсона с уровнем значимости α=0,05. Тогда гипотетическая функция плотности распределения случайной величины *X* имеет вид:

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

Нормальный закон распределения также называется законом Гаусса.

Нормальный закон распределения занимает центральное место в теории вероятностей. Это обусловлено тем, что этот закон проявляется во всех случаях, когда случайная величина является результатом действия большого числа различных факторов. К нормальному закону приближаются все остальные законы распределения.

Можно легко показать, что параметры a и σ, входящие в плотность распределения являются соответственно математическим ожиданием и средним квадратическим отклонением случайной величины *X*.

График плотности нормального распределения называется нормальной кривой или кривой Гаусса.

Нормальная кривая обладает следующими свойствами:

- Функция определена на всей числовой оси.

- $f(x)>0$
- Ось OX является горизонтальной асимптотой графика плотности вероятности, т.к. при неограниченном возрастании по абсолютной величине аргумента x, значение функции стремится к нулю.
- $f(x)$ достигает max в точке x-a;

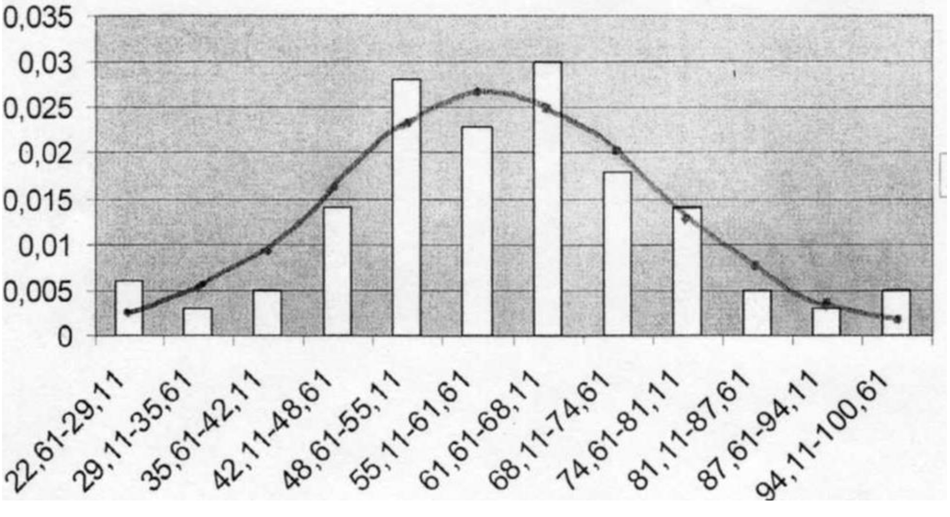
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{\frac{-(x-a)^2}{2\sigma^2}}=\frac{1}{\sigma}\cdot\varphi\left(\frac{x-a}{\sigma}\right)$$

График функции имеет две точки перегиба $x=a\pm\sigma$

Для простоты вычислений составим таблицу:

i	$t_i=\dfrac{\tilde{x}_i-a^*}{\sigma^*}$	$\varphi(t_i)$	$f(\tilde{x}_i)=\dfrac{\varphi(t_i)}{\sigma^*}$	$p_i=f(\tilde{x}_i)\cdot\Delta$
1	-2,28	0,03	0,002	0,013
2	-1,85	0,072	0,005	0,033
3	-1,43	0,144	0,009	0,059
4	-1,01	0,24	0,016	0,104
5	-0,58	0,337	0,022	0,143
6	-0,16	0,394	0,026	0,169
7	0,26	0,386	0,025	0,163
8	0,69	0,314	0,02	0,13
9	1,11	0,216	0,014	0,091
10	1,53	0,124	0,008	0,052
11	1,95	0,06	0,004	0,026
12	2,38	0,024	0,002	0,013
				$\Sigma=$0,0996

Функцию плотности $f(\tilde{x}_i)=\dfrac{\varphi(t_i)}{*}$ изобразим на гистограмме.



Критерий Пирсона - основан на изучении меры расхождения между теоретическим и статистическим распределением, которая в данном случае оценивается по сумме квадратов разности этих расхождений по всем интервалам выборки с учётом «веса» сл. в. $n{\cdot}p_i$..

Пирсон доказал, что значения статистического критерия не зависит от функции распределения *f(x)* и от числа опытов *n*, а зависит от числа частичных интервалов *r* интервального вариационного ряда.

Далее используем правило проверки гипотезы по критерию Пирсона.

1. Вычисляем квантиль $\chi_p^2(r-l-1)$. Для её вычисления нужно воспользоваться табличными распределениями X^2, в которых значения сл. в. находят по заданному уровню значимости α и вычисленному числу свободы υ.

$\upsilon = r - l - 1 = 8 - 2 - 1 = 5$, где

r- число частичных интервалов. Если в некоторых из интервалов значения $m_i^* \prec 5$, то надо объединить расположенные рядом интервалы так, чтобы $m_i^* \succ 5$ или $m_i^* = 5$, тогда число *r*- число из необъединённых интервалов, *l* число неизвестных параметров.

Имеем p=1-α=0,95, m=8, l=2.

По таблице распределения X^2находим

$\chi_{0,95}^2(8-2-1) = \chi_{0,95}^2(5) = 11{,}1.$

2. Находим $X^2_{выбор.}$ по формулам:

$$X^2_{выбор.} = \frac{\sum_{i=1}^{R}(n\cdot p_i^* - n\cdot p)^2}{n\cdot p_i}, \text{или}\ X^2_{выбор.} = \frac{\sum_{i=1}^{R}(m_i^* - m_i)^2}{m_i}.$$

i	p_i^*	$n\cdot p_i^*$	$n\cdot p_i$	$(n\cdot p_i^* - n\cdot p_i)$	$(n\cdot p_i^* - n\cdot p_i)^2$	$\frac{(n\cdot p_i^* - n\cdot p_i)^2}{n\cdot p_i}$
i	*0,09*	*9*	*10,5*	*-1,5*	*2,25*	*0,2*
2	*0,09*	*9*	*10,4*	*-1,4*	*1,96*	*0,2*
3	*0,18*	*18*	*14,3*	*3,7*	*13,69*	*0,96*
4	*0,15*	*15*	*16,9*	*-1,9*	*3,61*	*0,2*
5	*0,2*	*20*	*16,3*	*3,7*	*13,69*	*0,8*
6	*0,12*	*12*	*13*	*-1*	*0,01*	*0,001*

7	*0,09*	*9*	*9,1*	*-0,1*	*0,01*	*0,001*
8	*0,08*	*8*	*9,1*	*-1,1*	*1,21*	*0,133*
						$\sum = X_{выбор.} = 2.495$

Для этого удобно результаты вычислений вносить в следующую таблицу

3. Окончательно имеем

$2.495 = X^2_{выбор.} \prec X^2_{0.95} = 11.1.$что означает: гипотезу о нормальном распределении случайной величины X принимаем, т. к.