

Обработка результатов наблюдений

В деятельности современного инженера, руководителя, часто возникает необходимость принятия решения в условиях неопределённости. При решении этой задачи наиболее целесообразно использовать методы теории вероятностей и математической статистики. *Математическая статистика – это раздел математики, основная цель которого – получение выводов о массовых явлениях и процессах по данным наблюдений над ними экспериментов.* Эти статистические выводы относятся не к отдельным испытаниям, а представляют собой утверждения об общих характеристиках этого явления (вероятностях, законах распределения и их параметрах, математических ожиданиях, дисперсиях и других моментах) в предположении постоянства условий, порождающих исследуемое явление.

Методы математической статистики применяются в разных отраслях знания: медицине, технике, экономике и могут преследовать различные цели. Но можно указать основные, наиболее *общие задачи математической статистики*: разработка и применение различных методов оценки общих характеристик совокупности, проверка статистических гипотез, выявление взаимосвязей между несколькими характеристиками объектов, согласия наблюдаемых данных с теорией, принятие решений, планирование эксперимента и т.п. Все задачи можно разбить на две большие группы: параметрические и непараметрические. В теории вероятностей описаны вероятностные модели, содержащие параметры (это биномиальное, нормальное, показательное и др. распределения). Если известны вид распределения и область изменения параметров, то имеют место параметрические задачи. Если же нет представления даже о виде плотности, не говоря уже о возможных значениях параметра, то речь идёт о непараметрических задачах.

Математическая статистика опирается на теорию вероятностей и при достижении своей главной цели – оценки характеристик генеральной совокупности по выборочным данным, использует основные понятия теории вероятностей. В таблице 1 приведено соответствие терминов, обозначений, формул вариационного ряда и случайной величины.

*Например, под генеральной совокупностью следует понимать вероятностное пространство $\{\Omega, S, P\}$ (пространство элементарных событий Ω с заданным на нём полем событий S и вероятностями P) и определённая на этом пространстве случайная величина X . Единицей генеральной совокупности называется элементарное событие и отвечающее ему значение случайной величины. Вариационный ряд рассматривается как одна из реализаций распределения признака (случайной величины X). Случайная величина X имеет определённую функцию распределения и вытекающие из неё числовые характеристики. Функцию распределения случайной величины X , её параметры и числовые характеристики называют теоретическими, в отличие от выборочных, эмпирических, которые определяются по выборочным данным.

Материалы, помеченные значком *, предназначены для студентов, углублённо изучающих курс теории вероятностей и математической статистики.

Таблица 1.

Вариационный ряд		Случайная величина	
Обозначения, формулы	Термин	Обозначения, формулы	Термин
—	Дискретный ряд	—	Дискретная случайная величина
—	Интервальный ряд	—	Непрерывная случайная величина
x_i	Варианта	x_i, x	Значение случайной величины
f_i, f	Относительная частота	p_i, p, P	Вероятность
—	Полигон, гистограмма	—	Полигон (многоугольник) распределения вероятностей, кривая распределения
$F_x^*(x) = f(X < x)$	Эмпирическая функция распределения	$F(x) = P(X < x)$	Функция распределения
$\bar{x} = \sum_{i=1}^m x_i f_i$	Средняя арифметическая	$a = M(X) = \sum_{i=1}^n x_i p_i$	Математическое ожидание*
$s^2 = \overline{(x - \bar{x})^2} = \frac{1}{n-1} \sum_{i=1}^m (x_i - \bar{x})^2 n_i$	«Исправленная» дисперсия	$\sigma^2 = M[X - M(X)]^2 = \sum_{i=1}^n (x_i - a)^2 p_i$	Дисперсия*
$s = \sqrt{s^2}$	Среднее квадратическое ожидание	$\sigma = \sqrt{D(X)} = \sqrt{\sigma^2}$	Среднее квадратическое отклонение
Mo	Мода	$Mo(X)$	Мода
Me	Медиана	$Me(X)$	Медиана
$\alpha_k^* = \sum_{i=1}^m x_i^k f_i$	Эмпирический начальный момент k-го порядка	$\alpha_k = \sum_{i=1}^n x_i^k p_i$	Начальный момент k-го порядка*
$\mu_k^* = \sum_{i=1}^m (x_i - \bar{x})^k f_i$	Эмпирический центральный момент k-го порядка	$\mu_k = \sum_{i=1}^n [x_i - M(X)]^k p_i$	Центральный момент k-го порядка*
$As^* = \frac{\tilde{\mu}_3}{s^3}$	Коэффициент асимметрии	$As = \frac{\mu_3}{\sigma^3}$	Коэффициент асимметрии
$Es^* = \frac{\tilde{\mu}_4}{s^4} - 3$	Эксцесс	$Es = \frac{\mu_4}{\sigma^4} - 3$	Эксцесс

* Формула приведена для дискретной случайной величины.

Генеральная совокупность и выборка

Генеральной совокупностью называется вся совокупность, подлежащих изучению объектов или возможных результатов наблюдений над одним объектом. Число изучаемых объектов, составляющих генеральную совокупность, или число возможных результатов наблюдений над одним объектом называется *объёмом генеральной совокупности*. Объём генеральной совокупности может быть как конечным, но с большим объёмом, так и бесконечным.

Наилучшим способом изучения генеральной совокупности является отбор из генеральной совокупности части её элементов (измерений) – называемой *выборочной совокупностью* или *выборкой* и исследование её свойств, а затем обобщение полученных результатов на всю генеральную совокупность. Случайной выборкой (выборкой) объёма n называется последовательность X_1, X_2, \dots, X_n n независимых одинаково распределённых случайных величин, распределение каждой из которых совпадает с распределением исследуемой случайной величины X . Иными словами, случайная выборка – это результат n последовательных и независимых наблюдений над случайной величиной X , представляющей генеральную совокупность.

Основное требование к выборке – хорошо представлять генеральную совокупность, быть представительной, *репрезентативной*. Для обеспечения этого условия выборка должна быть случайной, когда каждому объекту генеральной совокупности обеспечивается одинаковая возможность быть выбранным, содержать независимые результаты наблюдений, иметь правильно определённый объём, с учётом всех конкретных условий.

На практике используются различные способы получения выборки (отбора). Наиболее распространены:

1. отбор, не требующий разбиения генеральной совокупности на части (например, простой случайный бесповторный отбор или простой случайный отбор с повторениями);
2. отбор, при котором генеральная совокупность разбивается на части: типический, механический и серийный.

Простым случайным называется отбор, при котором объекты извлекаются по одному из всей генеральной совокупности. Если при этом выбранный объект перед отбором следующего объекта возвращается в генеральную совокупность, то отбор называется *повторным*. Если объект в генеральную совокупность не возвращается, то отбор *бесповторный*.

Типическим называется отбор, при котором объекты извлекают не из всей генеральной совокупности, а только из некоторой её части. Типический отбор целесообразен, если значения изучаемого признака заметно колеблются в различных частях выборки. *Механическим* называется отбор, при котором генеральная совокупность произвольным образом делится на столько групп, сколько объектов должно войти в

выборку, и из каждой группы извлекается один объект. *Серийным* называется отбор, при котором из генеральной совокупности объекты выбирают не по одному, а группами (сериями). Серийный отбор применяют, когда значение исследуемого признака незначительно колеблется в различных сериях. При *комбинированном отборе* используют все вышеперечисленные методы.

Группировка эмпирических данных

Чаще всего первичные экспериментальные данные (выборочные данные) представляют собой неупорядоченный набор чисел. Процесс изучения выборки начинают с упорядочивания, группировки результатов. *Группировка* заключается в распределении вариантов выборки по группам (интервалам) группировки, каждый из которых содержит некоторый диапазон значений признака.

Цель группировки - получение вариационного или статистического ряда и построение эмпирического распределения (по форме которого делается предположение о форме распределения изучаемого признака в генеральной совокупности, из которой взята выборка).

Вариационный ряд - элементы выборки, расположенные в порядке возрастания элементов. *Статистический ряд* - расположенные в порядке возрастания, различные элементы выборки и число элементов выборки, принимающие различные значения.

Для того, чтобы сгруппировать исходные данные, надо:

1. разбить весь диапазон варьирования признака в выборке между минимальной и максимальной вариантами выборки на интервалы группировки;
2. распределить по этим интервалам выборочные варианты. Если значение варианты совпадает с границей частичного интервала, то ее следует отнести к частичному интервалу, у которого верхняя граница совпадает с этим значением.

Члены дискретного статистического ряда принимают дискретные, изолированные значения (где n - объем выборки).

Таблица 2.

x_i	x_1	x_2	...	x_m	сумма частот
n_i	n_1	n_2	...	n_m	$\sum_{i=1}^m n_i = n.$

Члены непрерывного статистического ряда заполняют некоторый интервал.

Таблица 3.

x_i	$x_1 - x_2$	$x_2 - x_3$...	$x_{m-1} - x_m$
n_i	n_1	n_2	...	n_m

При разбиении всего диапазона варьирования на интервалы группировки следует учитывать следующие соображения:

1. если увеличивать число интервалов группировки и сужать каждый из них, то уменьшится число данных, попадающих в каждый интервал. Так

как выборочные значения случайны, то они случайным образом распределяются по интервалам группировки и, следовательно, картина экспериментального распределения будет содержать много случайных деталей, что мешает установить общие закономерности варьирования признака.

2. при слишком широких интервалах группировки нельзя получить детальной картины распределения, можно упустить важные закономерные подробности формы распределения.

Приблизительно число интервалов можно оценить исходя из объема выборки n можно по формуле Стерджеса: $k = \log_2 n + 1 \approx 3.322 \cdot \lg n + 1$ или по формулам: $k \approx 5 \lg n$, $k \approx \sqrt{n}$. Но эти формулы следует рассматривать только как оценку снизу для k , особенно для больших n . Для приближённого определения k можно воспользоваться таблицей 3. Например, для $n = 50$ можно выбрать $k = 6,6 \approx 7$.

Таблица 4.

Рекомендуемый объем выборки	25 - 40	40 - 60	60 - 100	100 - 200	≥ 200
Число интервалов группировки	5 - 6	6 - 8	8 - 10	8 - 12	10-15

Ширина каждого интервала определяется по формуле:

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{x_{\max} - x_{\min}}{1 + 3.322 \cdot \lg n}, \text{ где } h - \text{ ширина интервала, } x_{\max}, x_{\min} -$$

максимальная и минимальная варианты выборки, найденные непосредственно по таблице исходных данных. Величина h вычисляется с той же точностью, что и исходные данные. Найденные значения определяют с учетом требуемой точности, обычно в сторону увеличения, чтобы не уменьшить общий диапазон варьирования признака.

Нижняя граница первого интервала группировки выбирается так, чтобы x_{\min} попадала приблизительно в середину этого интервала. $x_{H1} = x_{\min} - \frac{h}{2}$. Прибавив к этой величине ширину интервала, получим нижнюю границу второго интервала (или верхнюю границу первого): $x_{B1} = x_{H2} = x_{H1} + h$, $x_{B2} = x_{H3} = x_{H2} + h$, и т. д. Для удобства сгруппированных данных вычисляют срединные значения интервалов группировки x_i : $x_i = x_{Hi} + \frac{h}{2}$, x_{Hi} - нижняя граница i -го интервала. Для удобства, результаты вычислений представляются в виде статистической таблицы.

Таблица 5.

номер интервала, i	границы интервала, $x_{Hi} - x_{Bi}$	срединные значения, x_i	частоты, n_i	накопленные частоты, n_{x_i}	относительные частоты, f_i	накопленные относительные частоты, F_i^*
1	$x_1 - x_2$	x_1	2	2	$2/n$	$2/n$
2	$x_2 - x_3$	x_2	7	9	$7/n$	$9/n$
...
контроль			$\sum n_i = n$		$\sum f_i = 1$	

В четвертом столбце помещают числа, показывающие, сколько раз варианты (относящиеся к каждому интервалу группировки) встречаются в выборке - частоты интервалов n_i .

Накопленная частота интервала - число, полученное последовательным суммированием частот в направлении от первого интервала (до того интервала включительно, для которого определяется накопленная частота) - n_{x_i} .

Относительная частота - отношение частоты к объему выборки: $f_i = \frac{n_i}{n}$. *Накопленная относительная частота* - отношение накопленной частоты к объему выборки: $F_i^* = \frac{n_{x_i}}{n}$. Фактически в восьмом столбце вычислены значения эмпирической функции распределения $F^*(x)$.

Графическое представление данных

Графическое представление повышает наглядность эмпирических распределений. Существует несколько основных способов графического представления данных.

Гистограмма – это ступенчатая фигура из прямоугольников с основаниями, равными ширине интервала группировки h_i и высотами $\frac{n_i}{n \cdot h_i}$. Если на построенной гистограмме соединить отрезками середины верхних сторон прямоугольников, образующих ее, - получим полигон относительных частот этого распределения. Гистограмма используется для представления распределений только непрерывных случайных величин (она является статистическим аналогом функции плотности вероятности случайной величины).

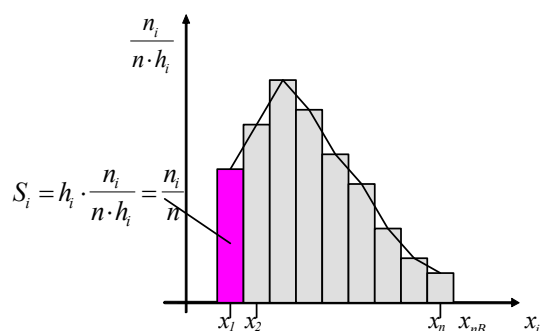


Рис.1. Гистограмма и полигон относительных частот.

$$S = \sum_{i=1}^n S_i = 1.$$

Полигон частот – это ломаная линия, соединяющая точки, соответствующие средним значениям интервалов группировки и частотам этих интервалов. Используется для представления непрерывных и дискретных случайных величин. Если полигон построен по дискретному вариационному ряду случайных величин, то его называют многоугольником распределения частот.

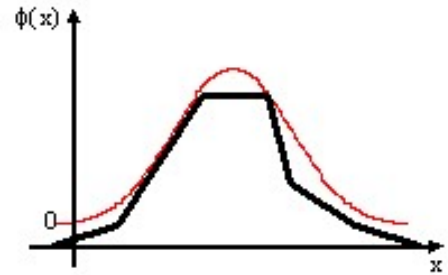
Многоугольники (полигоны) частот должны иметь форму, напоминающую форму соответствующей функции плотности распределения.

Например:

1. Для нормальной выборки полигон имеет одновершинный симметричный

вид:
$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

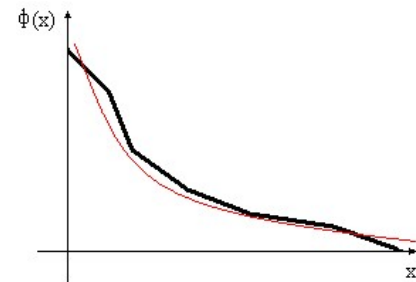
Рис.2 Полигон частот и график функции плотности вероятности



2. экспоненциальное распределение:

$$\varphi(x) = \lambda \cdot e^{-\lambda \cdot x}.$$

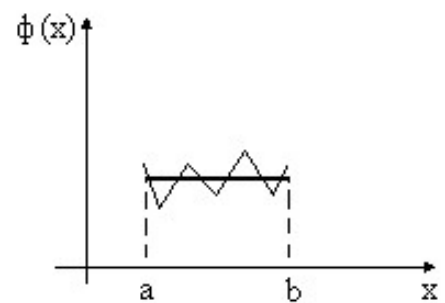
Рис.3 Полигон частот и график функции плотности вероятности



3. равномерное распределение на [a; b]:

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

Рис.4 Полигон частот и график функции плотности вероятности



Эмпирическая функция распределения задаётся формулой:

$$F^*(x) = \frac{\sum_{x_i \leq x} n_i}{n}.$$

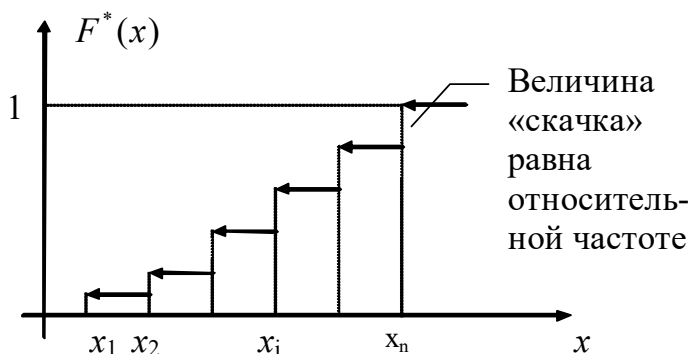
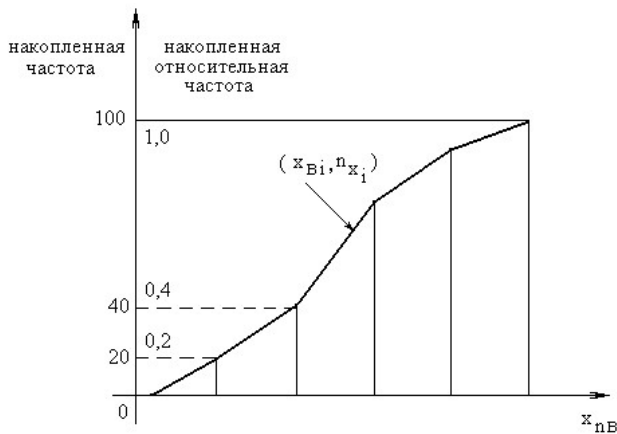


Рис. 5. График эмпирической функции распределения

$$F^*(x) = \begin{cases} 0, & x \leq x_1 \\ \sum_{i=1}^{k-1} \frac{n_{x_i}}{n}, & x_i < x \leq x_{i+1}, \quad i = \overline{1, k} \\ 1, & x > x_k \end{cases} \quad \text{где } \sum_{i=1}^{k-1} \frac{n_{x_i}}{n} - \text{накопленные частоты.}$$

Рис. 6. Кумулята



Кумулята (кривая накопленных частот) — это ломаная линия, полученная соединением отрезками прямых точек, координаты которых соответствуют верхним границам интервалов группировки и накопленным частотам. (чаще используется для представления дискретных данных, так как она имеет более плавную форму, чем гистограмма или полигон).

* Если предполагают, что генеральная совокупность имеет *экспоненциальное распределение*, то можно применить следующий способ проверки: пусть $X = x_1, x_2, \dots, x_n$ — наблюдаемые значения. Составляется $F^*(x) = P(x_i \leq x)$. Если верна гипотеза об экспоненциальном законе распределения выборки X , то эмпирическая функция распределения должна быть «похожа» на реальную функцию распределения: $F(x) = 1 - \exp(-\lambda \cdot x)$.

Затем рассмотрим две функции $\Lambda^*(x) = -\ln(1 - F^*(x))$ и $\Lambda(x) = -\ln(1 - F(x)) = -\ln(\exp(-\lambda x)) = \lambda x$. Если гипотеза об экспоненциальном законе выборки верна, то построенная функция $\Lambda^*(x)$ должна быть «похожа» на функцию $\Lambda(x) = \lambda x$. Утверждение проверяется графически (причём данные могут быть сгруппированы): строится график функции $F^*(x)$, а затем $\Lambda^*(x)$ (последний график должен быть близок прямой $\Lambda(x) = \lambda x$).

Числовые характеристики выборки.

Графического представления результатов наблюдений недостаточно для того, чтобы полностью охарактеризовать выборку. Числовые выборочные характеристики дают количественное представление об эмпирических данных и позволяют сравнивать их между собой.

Характеристики положения определяют положение центра эмпирического распределения.

* Наиболее часто применяются *средние степенные k -го порядка*:

$$\bar{x}_k = \frac{\sqrt[k]{\sum_{i=1}^n x_i^k \cdot n_i}}{n} = \frac{\left(\sum_{i=1}^n x_i^k \cdot n_i \right)^{\frac{1}{k}}}{n}, \quad x_i > 0.$$

При $k=1$ получим *среднее арифметическое*: $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$ (для не сгруппированных данных) и $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^k n_i x_i$ (где k - число интервалов группировки, а x_i - середины интервалов) для сгруппированных данных. Среднее арифметическое – это такое значение признака, сумма отклонений а которого выборочных значений признака (с учетом знака отклонения) равна 0.

* При $k=-1$ средняя степенная k -го порядка называется *средней гармонической*:

$$\bar{x}_{\text{гарм}} = \frac{n}{\sum_{i=1}^n \frac{n}{x_i}}, \text{ при } k=0 \text{ средней геометрической: } \bar{x}_{\text{геом}} = \sqrt[n]{\prod_{i=1}^n x_i^{n_i}}, \text{ а при } k=2, \text{ средней}$$

квадратической: $\bar{x}_2 = \frac{\sqrt{\sum_{i=1}^n x_i^2 \cdot n_i}}{n}$. С ростом порядка k степенная средняя возрастает (свойство мажорантности средних): $\bar{x}_{-1} < \bar{x}_0 < \bar{x}_1 < \bar{x}_2 < \dots$.

Средние степенные различных порядков и средняя арифметическая – это аналитические средние. Кроме них в статистическом анализе применяются структурные (порядковые) средние, например, мода и медиана. Эти характеристики не зависят от формы распределения эмпирических данных.

Медианой (Me) называется значение варьирующего признака, приходящееся на середину упорядоченной (ранжированной) совокупности. Часто *Me* называют непараметрической средней, причём $Me \leq \bar{x}$.

Для вычисления медианы данные ранжируют - располагают в порядке возрастания или убывания (если выборка невелика). Для дискретного вариационного ряда с нечётным числом членов n медиана равна срединному варианту $R_{Me} = \frac{n+1}{2}$. Если n четное, то *Me* равна – полусумме (среднему арифметическому) двух срединных вариантов.

$$n = 5: 1, 3, 6, 7, 8 \Rightarrow R_{Me} = \frac{5+1}{2} = 3, \quad Me = 6.$$

Например:

$$n = 6: 1, 2, 3, 6, 7, 9 \Rightarrow Me = \frac{3+6}{2} = 4,5.$$

$$\text{Для сгруппированных данных } Me = X_{MeH} + h \cdot \frac{0,5 \cdot n - n_{X_{Me}-1}}{n_{Me}},$$

где n - объём выборки; X_{Me} - медианный интервал (в нем накопленная частота впервые оказывается больше $\frac{n}{2}$, а накопленная относительная частота - больше 0,5); h - ширина интервала группировки; X_{MeH} - нижняя

граница медианного интервала; n_{Me} - частота медианного интервала;
 $n_{x_{Me}-1}$ - накопленная частота интервала, предшествующего медианному.

Главное свойство медианы состоит в том, что сумма абсолютных величин отклонений вариант от медианы меньше, чем от любой другой величины (в том числе и от \bar{x}): $\sum |x - Me| \Rightarrow \min$

Медиана (Me) не совпадает с \bar{X} , когда форма эмпирического распределения несимметрична. Если распределение сильно несимметрично, то лучшая характеристика центра распределения - медиана, так как \bar{X} теряет практическую ценность (большая часть значений признака оказывается выше или ниже среднего арифметического).

Мода Mo – это значение признака, встречающееся в выборке наиболее часто. Интервал группировки с наибольшей частотой называется модальным. Для вычисления моды применяют формулу:

$$Mo = x_{Mon} + h \cdot \frac{n_{Mo} - n_{Mo-1}}{(n_{Mo} - n_{Mo-1}) + (n_{Mo} - n_{Mo+1})},$$

где x_{Mon} – нижняя граница модального интервала; n_{Mo} – частота модального интервала; n_{Mo-1} – частота интервала, предшествующего модальному; n_{Mo+1} – частота интервала, следующего за модальным.

Если \bar{x} , Mo и Me различны, то это говорит об асимметрии эмпирического распределения. Характеристики положения совпадают только в том случае, когда распределение унимодальное (с одним максимумом) и симметрично.

Пример. Пример вычисления моды и медианы по сгруппированным в интервальный ряд выборочным данным. Сгруппированные результаты наблюдений представлены в таблице 6:

Таблица 6.

Интервалы, (x_{in}, x_{is})	Частоты, n_i	Накопленные частоты, n_{x_i}
0-5	4,6	4,6
5-10	2,4	7,0
10-15	2,5	9,5
15-20	17,7	27,2
20-25	24,0	51,2
25-30	19,5	70,7
30-35	8,1	78,8
35-40	5,7	84,5
40-45	4,6	89,1
45-50	3,4	92,5
50-55	2,3	94,8
55-60	5,2	100,0
Сумма	100,0	-

Модальным (с наибольшей относительной частотой) и медианным (интервалом, где накопленная частота впервые превышает 50) интервалом является интервал $[20,25]$. Подставив числа из таблицы в формулу для вычисления моды и медианы, получим:

$$Mo = 20 + 5 \cdot \frac{24,0 - 17,7}{(24,0 - 17,7) + (24,0 - 19,5)} \approx 22,91 \quad Me = 20 + 5 \cdot \frac{50 - 27,2}{24,0} \approx 24,75$$

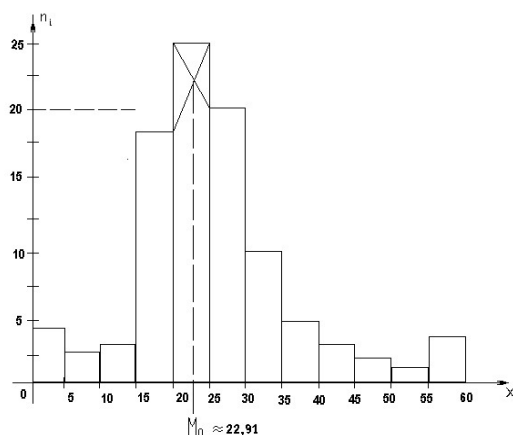


Рис.7. Гистограмма

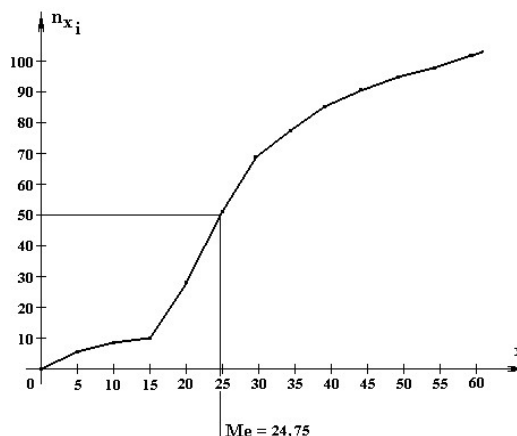


Рис.8. Кумулята

Далее рассмотрим наиболее часто применяемые в статистическом анализе *характеристики рассеяния* выборки.

* *Размах выборки* определяется как $R = x_{\max} - x_{\min}$. *Ко* - коэффициент осцилляции, показывающий относительную варьируемость крайних значений варьируемого признака относительно средней $Ko = \frac{R}{\bar{x}} \cdot 100\%$.

Коэффициент вариации - относительный показатель, который вводится, когда требуется сравнить между собой степень варьирования признаков или сопоставить среднеквадратическое отклонение со средним арифметическим этих признаков: $V = \frac{S}{\bar{x}}$

или $V = \frac{S}{\bar{x}} \cdot 100\%$. V - относительная мера рассеяния признака, показатель однородности выборочных наблюдений. Если V не превышает 10 %, то выборку можно считать однородной, то есть полученной из одной генеральной совокупности (но при $\bar{x} \approx 0$, $V \rightarrow \infty$ каков бы ни был разброс признака, то есть V неприменима в качестве характеристики рассеяния).

Среднее арифметическое абсолютных величин отклонений: $\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}| \cdot n_i}{n}$. учитывает различия всех единиц изучаемой совокупности. *Относительное линейное отклонение*: $Kd = \frac{\bar{d}}{\bar{x}}$ показывает долю усреднённых значений абсолютных отклонений в средней арифметической.

Дисперсией называется средний квадрат отклонений значений признака от средней арифметической: $\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$ для не

сгруппированных данных и $\sigma_s^2 = \frac{1}{n} \cdot \sum_{i=1}^k n_i (x_i - \bar{x})^2$ (где x_i - средние значения интервалов, а $n_i(x_i - \bar{x})^2$ - взвешенные квадраты отклонений) для сгруппированных в интервальный вариационный ряд данных.

На практике чаще вычисляется несмещённая «исправленная» дисперсия $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$ для не сгруппированных данных и $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^k n_i (x_i - \bar{x})^2$ для сгруппированных данных.

Среднеквадратическое отклонение (стандартное отклонение) определяется как $S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \cdot \sum (x_i - \bar{x})^2}$. Размерность стандартного отклонения, в отличие от размерности дисперсии, совпадает с единицами измерения варьируемых признаков.

Вычисление по вышеуказанным формулам не всегда удобно, поскольку при расчетах необходимо каждый раз из x_i вычитать \bar{x} , а затем результат возводить в квадрат - это громоздко; \bar{x} обычно вычислены с погрешностью округления. Это приводит к накоплению ошибки округления результатов - ошибка увеличивается с увеличением объема выборки. Поэтому на практике применяются следующие формулы:

$$S^2 = \frac{1}{n-1} \cdot \left(\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right) = \frac{1}{n-1} \cdot \left(\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n} \right) \text{ для несгруппированных данных}$$

или

$$S^2 = \frac{1}{n-1} \cdot \left(\sum_{i=1}^n n_i x_i^2 - n \cdot \bar{x}^2 \right) = \frac{1}{n-1} \cdot \left(\sum_{i=1}^n n_i x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n} \right) \text{ для сгруппированных данных.}$$

Результаты вычислений представляются в виде таблиц:

Таблица 7.

N	x_i	x_i^2
...
	Σ	Σ

Таблица 8.

N	x_i	n_i	$n_i x_i$	$n_i x_i^2$
...
			Σ	Σ

Определение выборочных моментов высших порядков

В теории вероятностей начальным моментом s -го порядка случайной величины X называется величина $\alpha_s(X) = M(S^s)$, а центральным моментом - $\mu_s(X) = M[(X - MX)^s]$, где MX - математическое ожидание случайной

величины X . Наиболее часто используются моменты $\alpha_1(X) = M(X)$, $\mu_2(X) = D(X)$. Начальный и центральный моменты имеют размерность равную размерности S -й степени случайной величины X .

Очевидно, что центральный момент первого порядка любой случайной величины равен нулю. А в случае симметричного распределения случайной величины X относительно MX равны нулю все центральные моменты нечетного порядка.

Если выборка репрезентативна и велика по объему, то несложно вычислить выборочные аналоги начального и центрального моментов

$$\alpha_s^* \approx \sum_{i=1}^n x_i^s \cdot f_i, \quad \mu_s^* \approx \sum_{i=1}^n (x_i - \bar{x})^s \cdot f_i \quad \text{где} \quad f_i = \frac{n_i}{n}.$$

После вычисления \bar{x} находят центральные моменты $\mu_2^*, \mu_3^*, \mu_4^*$. Результаты удобно представить в виде таблицы:

Таблица 9.

i	x_i	n_i	f_i	$x_i f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
...
-	-	$\sum_{i=1}^k n_i = n$	-	$\sum_{i=1}^k \frac{x_i n_i}{n} = \bar{x}$	-	-	$\sum_{i=1}^k \frac{(x_i - \bar{x})^2}{n} \cdot n_i = \mu_2^*$

(продолжение таблицы 9)

$(x_i - \bar{x})^3$	$(x_i - \bar{x})^3 f_i$	$(x_i - \bar{x})^4$	$(x_i - \bar{x})^4 f_i$
...
-	$\sum_{i=1}^k \frac{(x_i - \bar{x})^3}{n} \cdot n_i = \mu_3^*$	-	$\sum_{i=1}^k \frac{(x_i - \bar{x})^4}{n} \cdot n_i = \mu_4^*$

Несколько проще вычислять центральные моменты высших порядков, используя начальные моменты. Несложно вывести формулы, связывающие эти характеристики: $\mu_1^* = 0$;

$$\mu_2^* = \alpha_2^* - \alpha_1^{*2};$$

$$\mu_3^* = \alpha_3^* - 3 \cdot \alpha_1^* \cdot \alpha_2^* + 2 \cdot \alpha_1^{*3};$$

$$\mu_4^* = \alpha_4^* - 4 \cdot \alpha_1^* \cdot \alpha_3^* + 6 \cdot \alpha_1^{*2} \cdot \alpha_2^* - 3 \cdot \alpha_1^{*4}.$$

При $MX = 0$ начальные и центральные моменты s -го порядка совпадают. Для вычисления начальных моментов по выборочным данным заполняют таблицу:

Таблица 10.

x_i	n_i	f_i	$x_i f_i$	x_i^2	$x_i^2 f_i$	x_i^3	$x_i^3 f_i$	x_i^4	$x_i^4 f_i$
...
Σ	N	1	α_1	—	α_2	—	α_3	—	α_4

Центральные моменты третьего и четвёртого порядка необходимы для вычисления *коэффициента асимметрии* $As = \frac{\mu_3}{\sigma_B^3}$ и *эксцесса* $Es = \frac{\mu_4}{\sigma_B^4} - 3$.

Выборочные коэффициенты асимметрии и эксцесса вычисляются по формулам: $As^* = \frac{\sum_{i=1}^n n_i (x_i - \bar{x})^3}{nS^3}$, $Es^* = \frac{\sum_{i=1}^n n_i (x_i - \bar{x})^4}{nS^4} - 3$.

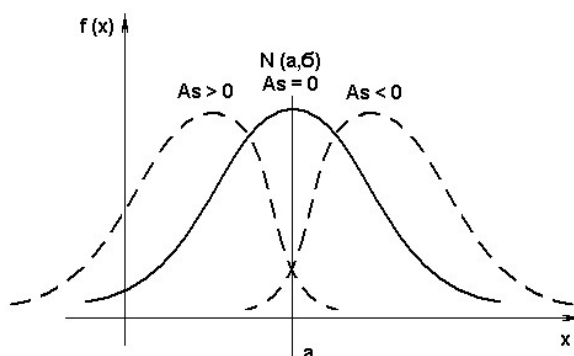


Рис.9. Многоугольники распределения с разными показателями асимметрии.

Асимметрия - характеристика большего или меньшего отклонения многоугольника распределения от симметричного, нормального. Если $As = 0$, график симметричный, если $As > 0$, пологая (длинная) часть графика расположена справа, если $As < 0$ - пологая часть графика слева.

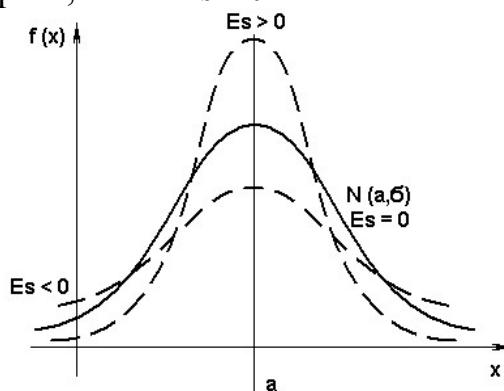


Рис.10. Многоугольники распределения с разными показателями эксцесса.

Эксцесс - характеристика большей или меньшей степени «сглаженности» многоугольника распределения по сравнению с нормальным. Если $Es > 0$ график «круче» нормального, если $Es < 0$ - график более пологий, чем нормальный.

Статистическое оценивание неизвестных параметров распределения.

Часто в приложениях рассматриваются параметрические модели. При этом полагают, что закон распределения случайной величины задан, а вектор параметров $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ неизвестен. Требуется найти оценку $\hat{\theta}$ по случайной выборке (x_1, x_2, \dots, x_n) из генеральной совокупности X .

Если из генеральной совокупности с функцией распределения $F(x, \theta)$, где θ - неизвестный параметр, извлечена выборка, то функция $u(x_1, x_2, \dots, x_n)$, где x_i - варианты выборки, называется векторной функцией или *статистикой*. Значение выборочной функции $\hat{\theta} = u(x_1, x_2, \dots, x_n)$ называется оценкой. Оценки бывают точечными и интервальными.

Точечные оценки параметра θ определяются одним числом. Это значения, максимально близкие к значениям соответствующих параметров генеральной совокупности (иначе называемыми истинными значениями оцениваемых параметров). Наиболее распространены методы получения точечных оценок: метод моментов, метод наибольшего правдоподобия (ММП) и метод наименьших квадратов (МНК).

Метод моментов основывается на том, что эмпирические моменты (или их функции) принимаются за оценки соответствующих теоретических моментов (или их функций) и параметры выражаются через эти моменты.

Идея *метода максимального правдоподобия* такова: для получения оценки неизвестного параметра θ необходимо найти такое значение $\hat{\theta}$, при котором вероятность реализации полученной выборки X_1, X_2, \dots, X_n была бы максимальной. Если предположить, что составляющие случайной величины независимы, то вероятность того, что они примут значения, равные наблюдаемым равна: $L = P(x_1, \theta)P(x_2, \theta) \dots P(x_n, \theta)$.

Величина L называется функцией правдоподобия, а $\hat{\theta}$, являющаяся точкой максимума этой функции, - оценкой, полученной по ММП. В случае непрерывной случайной величины функция правдоподобия имеет вид: $L = f(x_1, \theta)f(x_2, \theta) \dots f(x_n, \theta)$, где $f(x)$ - функция плотности вероятности случайной величины. В этом случае для нахождения оценок составляется система из m уравнений (где m - число оцениваемых параметров): $\frac{\partial L}{\partial \theta_i} = 0, i = 1, 2, \dots, m$ и выбирается то решение, при котором функция правдоподобия достигает максимума.

Поскольку экстремумы функций L и $\ln L$ достигаются при одних и тех же значениях $\hat{\theta}$, то для упрощения расчётов иногда пользуются логарифмической функцией правдоподобия.

Пример. Случайная величина имеет показательное распределение с плотностью распределения вероятностей $f(x, \theta) = \theta e^{-\theta x}$. Оценить параметр θ по результатам наблюдений x_1, x_2, \dots, x_n .

Функция правдоподобия имеет вид:

$$L = \theta \cdot e^{-\theta x_1} \theta \cdot e^{-\theta x_2} \dots \theta \cdot e^{-\theta x_n} = \theta^n e^{-\theta \sum x_i}.$$

Логарифмируя её, получим: $\ln L = n \ln \theta - \theta \sum x_i$. Дифференцируя полученное равенство по параметру θ , находим: $\frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} - \sum x_i$.

Приравнивая производную к нулю, получим: $\frac{n}{\theta} - \sum x_i = 0$ или $\hat{\theta} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$.

Чтобы оценка неизвестного параметра θ была *качественной*, она должна быть:

1. *состоятельной* - при неограниченном увеличении объема выборки она должна стремиться к истинному значению параметра θ , т.е. $\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|\theta - \hat{\theta}| < \varepsilon) = 1$;
2. *несмещенной* - не должна содержать систематических ошибок, математическое ожидание оценки равно оцениваемому параметру, то есть $M(\hat{\theta}) = \theta$, для асимптотически несмещенных оценок $M(\hat{\theta}) \rightarrow \theta$, при $n \rightarrow \infty$;
3. *эффективной* - несмещенная оценка имеет наименьшую дисперсию по сравнению с другими несмещенными оценками того же параметра генеральной совокупности.

При определении точечной оценки, мы допускаем ошибку, поскольку $\hat{\theta} = u(x_1, x_2, \dots, x_n)$ является случайной функцией. Если X_n - случайная выборка объема n из генеральной совокупности X (с функцией распределения $F(x, \theta)$, зависящей от неизвестного параметра θ), то для параметра θ можно построить интервал $(\underline{\theta}(X_n), \overline{\theta}(X_n))$. Причём $\underline{\theta}(X_n)$ и $\overline{\theta}(X_n)$ являются функциями случайной выборки X_n , такими, что выполняется равенство $P\{\underline{\theta}(X_n) < \theta < \overline{\theta}(X_n)\} = \gamma$.

Интервал $(\underline{\theta}(X_n), \overline{\theta}(X_n))$ называется интервальной оценкой (доверительным интервалом) для параметра θ с доверительной вероятностью γ (где $\underline{\theta}(X_n)$ и $\overline{\theta}(X_n)$ - нижняя и верхняя границы интервальной оценки). Итак, *интервальная оценка* - это интервал со случайными границами, который с заданной вероятностью γ (чаще всего она выбирается равной 0,95; 0,99; 0,999) покрывает неизвестные истинные значения параметра θ .

* Чтобы построить интервальную оценку θ можно использовать так называемую центральную статистику - любую статистику $T(X_n, \theta)$, функция распределения которой $F_T(t) = P(T(X_n, \theta) < t)$ не зависит от параметра θ и верны утверждения:

- функция $F_T(t)$ непрерывна и монотонно возрастает;

- заданы такие числа α и β , что доверительная вероятность может быть представлена в виде $\gamma = 1 - \alpha - \beta$;

- для любой выборки X_n из генеральной совокупности X , функция $T(X_n, \theta)$ является непрерывной и возрастающей (убывающей) функцией параметра θ .

Левосторонней критической границей (квантилью), отвечающей вероятности α , называется такая граница K_α , левее которой вероятность равна α : $\alpha = P(X < K_\alpha) = F(K_\alpha)$, то есть квантиль является решением уравнения: $F(K_\alpha) = \alpha$. *Правосторонней критической границей*, отвечающей вероятности α , называется такая граница B_α , правее которой вероятность равна α . По определению: $\alpha = P(X \geq B_\alpha) = 1 - F(B_\alpha)$. То есть правосторонняя граница является решением уравнения $F(B_\alpha) = 1 - \alpha$. Очевидно, что между правосторонней и левосторонней границами существует соотношение: $K_\alpha = B_{1-\alpha}$.

Двусторонними критическими границами, отвечающими вероятности α , называются такие границы $\underline{B}_\alpha, \overline{B}_\alpha$, внутрь которых случайная величина попадает с вероятностью $1 - \alpha$, а вне – с вероятностью α , причём: $P(X < \underline{B}_\alpha) = P(X \geq \overline{B}_\alpha) = \frac{\alpha}{2}$. Таким образом, двусторонние границы являются решением уравнений $F(\underline{B}_\alpha) = \frac{\alpha}{2}, F(\overline{B}_\alpha) = 1 - \frac{\alpha}{2}$.

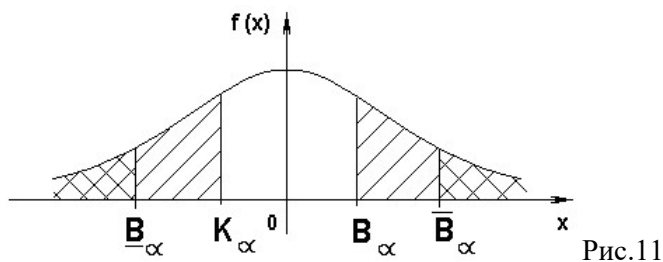


Рис.11

Между односторонними и двусторонними границами существуют соотношения: $\underline{B}_\alpha = K_{\frac{\alpha}{2}} = B_{1-\frac{\alpha}{2}}$, $\overline{B}_\alpha = K_{1-\frac{\alpha}{2}} = B_{\frac{\alpha}{2}}$

* Для любого $q \in (0,1)$ существует единственный корень h_q уравнения $F_T(t) = q$, который называется квантилью уровня q функции распределения $F_T(t)$ случайной величины $T(X_n, \theta)$.

Тогда равенства: $P\{h_\alpha < T(X_n, \theta) < h_{1-\beta}\} = F_T(h_{1-\beta}) - F_T(h_\alpha) = 1 - \beta - \alpha = \gamma$, справедливы для любых возможных значений параметра θ .

Для каждой выборки X_n уравнения $T(X_n, \theta) = h_\alpha$ и $T(X_n, \theta) = h_{1-\beta}$ имеют единственное решение $\underline{\theta}(X_n)$ и $\overline{\theta}(X_n)$. Соответственно, равенства $h_\alpha < T(X_n, \theta) < h_{1-\beta}$ и $\underline{\theta}(X_n) < \theta < \overline{\theta}(X_n)$ равносильны, и для любой выборки выполняются одновременно. Таким образом, $\gamma = P\{h_\alpha < T(X_n, \theta) < h_{1-\beta}\} = P\{\underline{\theta}(X_n) < \theta < \overline{\theta}(X_n)\}$, где $(\underline{\theta}(X_n), \overline{\theta}(X_n))$ - искомые оценки. Таким образом, построение доверительного интервала сводится к выполнению действий:

- построение центральной статистики с известной функцией распределения $F_T(t)$;
- представление доверительной вероятности γ в виде $\gamma = 1 - \alpha - \beta$;

- нахождение квантилей h_α и $h_{1-\beta}$ уровня α и $1-\beta$ функции распределения $F_T(t)$;
 - нахождение значений нижней $\theta(X_n)$ и верхней $\bar{\theta}(X_n)$ границ искомой интервальной оценки путём решения уравнений $T(X_n, \theta) = h_\alpha$ и $T(X_n, \bar{\theta}) = h_{1-\beta}$, соответственно (в том случае, когда $T(X_n, \theta)$ возрастает). Если $T(X_n, \theta)$ убывающая функция, то следует решать уравнения $T(X_n, \theta) = h_{1-\beta}$ и $T(X_n, \bar{\theta}) = h_\alpha$, соответственно.

Аналитические методы проверки гипотез.

Применение критериев согласия требует больших вычислений и поэтому целесообразно перед их использованием проверить гипотезу о характере распределения с помощью более простых методов (обладающих меньшей мощностью).

Для экспоненциального распределения, у которого $M(X) = \frac{1}{\lambda}$; $D(X) = \frac{1}{\lambda^2}$, выборочное среднее \bar{x} и выборочная дисперсия S^2 должны удовлетворять равенству: $S^2 = \bar{x}^2$.

* Пусть дана выборка $\bar{X} = (x_1, x_2, \dots, x_n)$. Известно, что если выборка из экспоненциального распределения, то величины S_1, S_2, \dots, S_{n-1} (где

$$S_i = \left(\frac{x_1 + x_2 + \dots + x_{i-1}}{x_1 + x_2 + \dots + x_i} \right)^{i-1}, \quad i = \overline{1, n-1})$$

взаимно независимы и равномерно распределены на $[0; 1]$. При помощи критерия Колмогорова-Смирнова проверяется гипотеза о равномерном распределении на $[0; 1]$ (при таком способе проверки не требуется использовать неизвестный параметр λ экспоненциального распределения, что повышает точность вычислений).

В распределении Пуассона $M(X) = \lambda$; $D(X) = \lambda^2$. Поэтому выборочное среднее \bar{x} и выборочная дисперсия S^2 должны удовлетворять равенству:

$$S^2 = \bar{x}^2.$$

Если в качестве параметров a и σ нормального распределения $N(a, \sigma)$ принять их выборочные оценки \bar{x} и S , то для проверки гипотезы можно использовать следующие свойства нормального распределения:

- 1) практически все отклонения от среднего значения (99,7%) должны быть меньше $\pm 3s$;
- 2) примерно 2/3 всех отклонений (68,3%) должны быть меньше $\pm s$;
- 3) половина всех отклонений от среднего значения должна быть меньше $\pm 0,637s$;
- 4) выборочные коэффициенты асимметрии γ_3 и эксцесса γ_4 для нормального распределения должны быть близки к нулю.
- 5) Медиана нормально-распределенной случайной величины совпадает с параметром $\alpha = \bar{x}$.

Можно найти среднеквадратические отклонения для A_s и E_s (если они не равны нулю): $\sigma_{A_s} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}$; $\sigma_{E_s} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}$.

Если A_s и E_k отличаются от нуля не более чем на удвоенные σ_{A_s} и σ_{E_s} , то возможно выборка распределена по нормальному закону.

Критерии согласия.

При обработке выборки могут быть получены результаты, которые можно использовать при анализе всей генеральной совокупности, чаще всего при определении функции распределения. Решение таких задач - статистическое, так как оно может оказаться ошибочным, возникают расхождения между эмпирическими и теоретическими функциями распределения.

Статистическими гипотезами называются утверждения относительно распределения генеральной совокупности, соответствующего представлениям об изучаемом явлении, или о значениях параметров известных распределений.

Нулевая гипотеза H_0 - выдвинутая гипотеза; *H_1 - конкурирующая гипотеза*, противоречащая нулевой.

Критерием согласия статистической гипотезы с эмпирическими данными называется правило, позволяющее на основании выборочных данных принять или отвергнуть нулевую гипотезу.

При проверке гипотез возникают ошибки двух типов (их определения представлены в таблице 11).

Таблица 11.

	Решение	
	принять H_0	принять H_1
справедлива H_0	правильное решение с вероятностью $(1-\alpha)$	ошибка первого рода с вероятностью α
справедлива H_1	ошибка второго рода с вероятностью β	правильное решение с вероятностью $(1-\beta)$

Число α называется *уровнем значимости критерия* и задаётся до начала эксперимента. Чаще всего ей приписывают значения: 0,05; 0,01; 0,001 (5%; 1% или 0,1%). Число $(1-\beta)$ называется *мощностью критерия*.

Статистика критерия – это случайная величина, характеризующая меру расхождения между теоретической функцией распределения и статистическим (эмпирическим) распределением.

Критическая область - совокупность значений статистики критерия, при которых нулевую гипотезу отвергают. *Область принятия гипотезы* (область допустимых значений) - совокупность значений статистики, при которых гипотезу принимают. *Критическими* называют точки, отделяющие область принятия гипотезы от критической области.

Критерий согласия χ^2 .

Критерий основан на сравнении эмпирических частот группировки с теоретическими (ожидаемыми) частотами.

Применяется при $n \geq 40$. Если выборочные данные сгруппированы в интервальный вариационный ряд, то число интервалов должно быть не менее 7; теоретические частоты не должны быть менее 5.

Порядок применения критерия.

1. Формулируется гипотеза, выбирается уровень значимости α (обычно 0,05; 0,01; 0,001). H_0 : плотность распределения $f(x)$ генеральной

совокупности, из которой взята выборка, соответствует теоретической модели $f'(x): f'(x) = f(x)$. $H_1: f'(x) \neq f(x)$.

2. Выборка, содержащая результаты независимых наблюдений, представляется в виде интервального вариационного ряда.

3. Рассчитываются выборочные характеристики, используемые в качестве генеральных параметров распределения, с которым предстоит сравнивать эмпирическое распределение.

4. Вычисляются значения теоретических частот n'_i попадания в i -ый интервал группировки. Если вычисленные ожидаемые частоты n'_i некоторых интервалов группировки меньше 5, то соседние интервалы объединяются так, чтобы сумма их ожидаемых частот была больше или равна 5. Соответственно складываются и эмпирические частоты объединенных интервалов.

5. Значение χ^2 - критерия рассчитываются по формуле:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - n'_i)^2}{n'_i},$$

где $\nu = k - l - 1$ - число степеней свободы распределения χ^2 , k - число интервалов группировки после объединения, l - число параметров предполагаемого распределения.

6. По таблице находится критическое значение χ^2_α (χ^2 - критерия для уровня значимости α и числа степеней свободы ν . Например, для нормального распределения $\nu = k - 3$).

7. Формулируется вывод: если $\chi^2 \geq \chi^2_\alpha$, то эмпирическое распределение не соответствует нормальному распределению на уровне значимости α ; в противном случае нет оснований отрицать это соответствие.

Пример применения критерия χ^2 . При помощи критерия согласия Пирсона проверить гипотезу о нормальном распределении генеральной совокупности. Выберем $\alpha = 0,05$. Результаты вычислений целесообразно выполнять в таблице.

Таблица 12.

N	$x_{ni} - x_{vi}$	n_i	$u_{ni} - u_{vi}$	n'_i	$n_i - n'_i$	$\frac{(n_i - n'_i)^2}{n'_i}$
1	12.4-13.2	1	-3.33-(-2.44)	0.315		
2	13.2-14.0	2	-2.44-(-1.56)	2.603		
3	14.0-14.8	9	-1.56-(0.67)	9.603	-0.551	0,024
4	14.8-15.6	15	-0.67-0.22	16.78	1.780	0,189
5	15.6-16.4	17	0.22-1.11	13.973	3.027	0,658
6	16.4-17.2	5	1.11-2.0	5.538	-0.578	0,051
7	17.2-18.0	6	2.0-2.89	6.578		
сумма		50	-	49.882	-	0,922

По таблице рассчитано, что $\bar{x} = 19,4$; $s = 0,9$; $n = 50$. Выберем $\alpha = 0,05$. Число интервалов $k = 4$, следовательно, число степеней свободы $\nu = 4 - 3 = 1$. По таблице из приложения 4 определим $\chi^2_{0,05} = 2,84$. Поскольку наблюдаемое значение $\chi^2 = 0,922 < 2,84$, то гипотеза о нормальном распределении генеральной совокупности принимается на уровне значимости $\alpha = 0,05$.

Критерий согласия λ Колмогорова - Смирнова.

Критерий Колмогорова применяется для проверки гипотез о законах распределения только непрерывных случайных величин (при объеме выборки $n \geq 35$). Эмпирическое распределение должно быть представлено в виде интервального вариационного ряда. Сравниваются эмпирическая функция распределения $F^*(x)$ и теоретическая функция $F(x)$ при известных параметрах распределения. Согласно теореме В.И.Гливенко: если функция распределения $F^*(x)$ случайной величины X непрерывна, то практически ее эмпирическая функция распределения $F(x)$ при $n \rightarrow \infty$ равномерно сходится к $F^*(x)$. В качестве меры расхождения между теоретическим и статистическим распределениями рассматривается максимальное значение модуля разности $D = \max |F^*(x) - F(x)|$.

Колмогоров А.Н. доказал, что какова бы ни была функция распределения $F(x)$ непрерывной случайной величины $D_n = \max |F(x) - F^*(x)| \cdot \sqrt{n}$ при неограниченном увеличении числа независимых наблюдений $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} P(D\sqrt{n} \geq \lambda) = P(\lambda) = 1 - \sum_{n \rightarrow -\infty}^{\infty} (-1)^n e^{-2n^2 \lambda^2}.$$

Функция $K(\lambda) = \sum_{n=-\infty}^{+\infty} (-1)^n \cdot e^{-2n^2 \lambda^2}$ получила название функции Колмогорова.

При $n \rightarrow \infty, \lambda > 0$: $P(D_n) < \lambda) \approx K(\lambda)$ (очень важно, что $K(\lambda)$ не зависит от предполагаемого теоретического распределения).

Порядок применения критерия:

1. Выбирается уровень значимости α . Уровень значимости α принимается чаще всего в пределах 0,01-0,10. Гипотезы H_0 и H_1 формулируется по отношению к функции распределения $F^*(x)$ (функция распределения генеральной совокупности, из которой получена выборка) и $F(x)$ (функция непрерывного теоретического распределения):
 $H_0 : F(x) = F^*(x)$
 $H_1 : F(x) \neq F^*(x);$
2. располагают результаты наблюдений по возрастанию их значений в виде интервального вариационного ряда;
3. вычисляются выборочные характеристики;

4. рассчитываются выборочные значения эмпирических n_{x_i} и теоретических $n_{x_i}^*$ накопленных частот; затем находят эмпирическую функцию

распределения $F^*(x) = \frac{n_{x_i}^*}{n}$ и вычисляют, пользуясь предполагаемой функцией $F(x)$, значения теоретической функции распределения, соответствующие наблюдаемым значениям случайной величины X ;

5. находят для каждого x_i модуль разности между эмпирической и теоретической функциями распределения и вычисляется значение

параметра λ : $\lambda = D\sqrt{n} = \max |F^*(x) - F(x)|\sqrt{n}$ или $\lambda = \frac{\max_i |n_{x_i} - n_{x_i}'|}{\sqrt{n}}$;

6. находят критические значения λ_α в зависимости от уровня значимости α по таблице 13:

Таблица 13.

α	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
λ_α	0,828	0,895	0,974	1,073	1,224	1,358	1,510	1,627	1,950

7. формулируется вывод: если наблюдаемое значение $\lambda > \lambda_\alpha$, то гипотеза о согласии теоретического закона распределения с данными выборки опровергается (эмпирическое распределение не соответствует теоретическому на уровне значимости α); в противном случае, если, $\lambda < \lambda_\alpha$ то принимается гипотеза о согласии распределения генеральной совокупности с теоретическим.

Можно поступить иначе: по таблице находят $P(\lambda)$, большая величина вероятности указывает на несущественность расхождений между фактическим и теоретическим распределением.

По опыту известно, что критерий λ является более мощным, чем χ^2 ; он чаще обнаруживает отклонение эмпирического распределения от теоретического.

Пример применения критерия. При помощи критерия согласия Колмогорова проверить гипотезу о нормальном распределении генеральной совокупности. Выберем $\alpha = 0,05$. Результаты вычислений удобно представить в таблице:

Таблица 14.

N	$x_{ni} - x_{ei}$	n_i	x_i	n_{x_i}	$u_i = \frac{x_i - \bar{x}}{S}$	n_{x_i}'	$ n_{x_i} - n_{x_i}' $	$\frac{ n_{x_i} - n_{x_i}' }{\sqrt{n}}$
1	12.4-13.2	12,8	1	1	-2,89	0,098	0,902	0,127
2	13.2-14.0	13,6	2	3	-2,0	1,138	1,862	0,263
3	14.0-14.8	14,4	9	12	-1,11	6,675	5,325	0,753
4	14.8-15.6	15,2	15	27	-0,22	20,648	6,352	0,898
5	15.6-16.4	16,0	17	44	0,67	37,428	6,572	0,929
6	16.4-17.2	16,8	5	49	1,56	47,030	1,97	0,279
7	17.2-18.0	17,6	1	50	2,44	49,633	0,367	0,052
сумма			50					

По таблице рассчитано, что $\bar{x} = 15,4$; $s = 0,9$; $n = 50$. По таблице 15 и таблице 13 определим $\lambda = 0,929$, $\lambda_{0,05} = 1,36$. Поскольку наблюдаемое значение $\lambda < \lambda_{0,05}$ гипотеза о нормальном распределении генеральной совокупности не отвергается на уровне значимости $\alpha = 0,05$.

Критерий W Шапиро-Уилки.

Если объём выборки невелик ($n \leq 40$) более точные выводы даёт критерий W, позволяющий обнаружить отклонения от нормальности распределения уже при $n \geq 10$.

Порядок применения критерия:

1. формулируется H_0 о соответствии распределения генеральной совокупности, из которой получены данные, нормальному распределению. Выберем $\alpha = 0,05$.

2. получаем выборку из $n = 10$ независимых измерений.

3. рассчитывает значение выборочной дисперсии $s^2 = 0,37$

4. ранжируем выборку, располагая выборочные значения в возрастающем порядке.

5. образуем разности Δ_k , для чего из максимального значения x_n вычитаем наименьшее x_1 , затем из x_{n-1} вычитают x_2 и т.д. Если n -чётное, то число разностей $k = \frac{n}{2}$, если n -нечётное, то $k = \frac{n-1}{2}$ (при этом центральная варианта выборки в образовании разностей не участвует).

6. по таблице находим значение коэффициента a_{nk} критерия W, соответствующие $n = 10$ и номерам разностей k .

7. находим произведения $a_{nk} \cdot \Delta_k$, вычисляем величину $b = \sum_{i=1}^k a_{nk} \cdot \Delta_k$.

8. рассчитываем значение W по формуле $W = \frac{b^2}{(n-1) \cdot s^2}$

9. из таблицы находим критическое значение критерия Шапиро-Уилки для уровня значимости $\alpha = 0,05$. Критерий W Шапиро-Уилки строится таким образом, что H_0 принимается при $W > W_{0,05}$, в отличие от других критериев, где гипотеза принимается, если значения критерия меньше критического.

Пример применения критерия. При помощи критерия W Шапиро-Уилки проверить гипотезу о нормальном распределении генеральной совокупности. Выберем $\alpha = 0,05$. Результаты вычислений удобно представить в таблице:

Таблица 15.

N	x_i	k	Δ_k	a_{nk}	$a_{nk} \cdot \Delta_k$
1	11,8	1	2,0	0,5739	1,1478
2	12,0	2	1,2	0,3291	0,3949
3	12,1	3	0,9	0,2141	0,1927
4	12,3	4	0,5	0,1224	0,0612
5	12,6	5	0,0	0,0399	0,0000
6	12,6				
7	12,8				
8	13,0				
9	13,2				
10	13,8				

$$b = \sum_{i=1}^k a_{nk} \cdot \Delta_k = 1,7966, \quad W = \frac{b^2}{(n-1) \cdot s^2} = \frac{1,7966^2}{(10-1) \cdot 0,37} = 0,969$$

Табличное $W_{0,05} = 0,842$ определяется по таблице из приложения 6. Так как наблюдаемое значение меньше табличного: $W = 0,969 > W_{0,05} = 0,842$ можно говорить о соответствии эмпирических данных нормальному распределению на уровне значимости 0,05.

Примеры решения задач

Пример 1. Методика вычисления теоретических частот нормального распределения.

Для вычисления теоретических частот можно воспользоваться интегральной функцией Лапласа $\Phi(z)$:

$$1. \text{ Вычисляют } \bar{X} = \frac{\sum x_i n_i}{n}; \quad \sigma^2 = \frac{\sum x_i^2 n_i}{n} - \left(\frac{\sum x_i n_i}{n} \right)^2. \quad \text{По выборочным}$$

данным вычислим: $\bar{X} = 12,63$, $\sigma = 4,695$.

2. Нормируют случайную величину X , то есть переходят к случайной величине $Z = \frac{X - \bar{X}}{\sigma}$. Причём $z_{in} = \frac{x_{in} - \bar{X}}{\sigma}$, $z_{is} = \frac{x_{is} - \bar{X}}{\sigma}$, наименьшее

$z_{1n} = -\infty$, наибольшее $z_{ns} = +\infty$.

3. Теоретические вероятности p_i попадания X в интервал (x_{in}, x_{is}) вычисляют по формуле: $p_i = \Phi(z_{is}) - \Phi(z_{in})$, где $\Phi(z)$ - функция Лапласа.

4. Теоретические частоты вычисляются по формуле: $n'_i = n \cdot p_i$.

Таблица 16.

i	x_{iH}	x_{iB}	x_i	n_i	z_{iH}	z_{iB}	$\Phi(z_{iH})$	$\Phi(z_{iB})$	p_i	n'_i
1	4	6	5	15	$-\infty$	-1.41	-0.5	-0.4207	0.0793	13.86
2	6	8	7	26	-1.41	-0.99	-0.4207	-0.3389	0.0818	16.36
3	8	10	9	25	-0.99	-0.56	-0.3389	-0.2123	0.1266	25.32
4	10	12	11	30	-0.56	-0.19	-0.2123	-0.0517	0.1606	32.12
5	12	14	13	26	-0.19	0.29	-0.0517	0.1141	0.1658	33.16
6	14	16	15	21	0.29	0.72	0.1141	0.2642	0.1501	30.02
7	16	18	17	24	0.72	1.14	0.2642	0.3729	0.1087	21.74
8	18	20	19	20	1.14	1.57	0.3729	0.4419	0.0689	13.78
9	20	22	21	19	1.57	$+\infty$	0.4419	0.5	0.0382	11.64
-	-	-	-	200	-	-	-	-	≈ 1	≈ 200

Для вычисления теоретических частот можно воспользоваться дифференциальной функцией Лапласа $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

1. $p_i \approx \varphi(x_i) \cdot h = \frac{\varphi(z_i)}{\sigma} h = \frac{x_{ei} - x_{ni}}{\sigma} \varphi(z_i)$, где $z_i = \frac{x_i - \bar{X}}{\sigma}$, h - длина интервала, x_i - середина i -го интервала, $\varphi(z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}}$. Затем вычисляют

$$n'_i = n \frac{h_i}{\sigma} \varphi\left(\frac{x_i - \bar{X}_e}{\sigma}\right).$$

Таблица 17.

$x_{iH} - x_{iB}$	x_i	n_i	z_i	$\varphi(z_i)$	p_i	n'_i
143 - 146	144.5	1	-3.477	0.0009	0.004	0
146 - 149	147.5	2	-2.981	0.0047	0.0023	2
149 - 152	150.5	8	-2.485	0.0182	0.0090	9
152 - 155	153.5	26	-1.989	0.0552	0.0274	27
155 - 158	156.5	65	-1.493	0.1309	0.0649	65
158 - 161	159.5	120	-0.997	0.2427	0.1204	120
161 - 164	162.5	181	-0.501	0.3519	0.1746	175
164 - 167	165.5	201	-0.005	0.3989	0.1979	198
167 - 170	168.5	170	0.491	0.3536	0.1754	175
170 - 173	171.5	120	0.987	0.2451	0.1216	122
173 - 176	174.5	64	1.483	0.1328	0.0659	66
176 - 179	177.5	28	1.979	0.0563	0.0279	28
179 - 182	180.5	10	2.475	0.0186	0.0092	9
182 - 185	183.5	3	2.971	0.0048	0.0024	2
185 - 188	186.5	2	3.467	0.0010	0.0005	0
$h = 3$	-	1000	-	-	-	998

Вычислим $\bar{X} = 165,53$ и $\sigma = 6.048$, тогда $n'_i = 1000 \cdot \frac{3}{6,048} \cdot \varphi(z_i) = \frac{\varphi(z_i)}{2,016}$, где $z_i = \frac{x_i - 165,53}{6,048}$

* Чтобы построить теоретический график плотности вероятности $N(a, \sigma)$, соответствующей выборочным данным, следует найти значение вероятности p_i в характерных точках с координатами $x = \hat{a} = \bar{x}$ (это точка максимума функции), затем в точках $x = \hat{a} \pm \hat{\sigma} = \bar{x} \pm s$ (это точки перегиба), $x = \hat{a} \pm 2\hat{\sigma} = \bar{x} \pm 2s$, $x = \hat{a} \pm 3\hat{\sigma} = \bar{x} \pm 3s$ и через эти точки провести гладкую кривую.

Пример 2. Найти теоретические частоты для равномерно распределенной случайной величины и применить критерий Колмогорова для проверки гипотезы H_0 о равномерном распределении генеральной совокупности.

Таблица 18.

X	0-10	10-20	20-30	30-40	40-50	50-60
n_i	11	14	15	10	14	16
x_i	5	15	25	35	45	55
$f(x_i)$	11/80	14/80	15/80	10/80	14/80	16/80
n_{x_i}	11	25	40	50	64	80
$F(x)$	11/80	25/80	40/80	50/80	64/80	80/80

Объём выборки равен $\sum_{i=1}^6 n_i = 80 = n$. По выборке рассчитаем:

$$\hat{M}(X) = \bar{x} = \frac{1}{80}(5 \cdot 11 + 15 \cdot 14 + 25 \cdot 15 + 35 \cdot 10 + 45 \cdot 14 + 55 \cdot 16) = 31,25 \quad \hat{D}(X) = \sigma_e^2 = \frac{1}{80} \cdot ((5 - 31,25) \cdot 11 + (15 - 31,25) \cdot 14 + (25 - 31,25) \cdot 15 + (35 - 31,25) \cdot 10 + (45 - 31,25) \cdot 14 + (55 - 31,25) \cdot 16) = 295,9375$$

Для оценки параметров распределения можно использовать метод моментов. Известно, что теоретические и выборочные моменты равномерного распределения равны:

$$\bar{X} = \hat{M}(X), \quad M(X) = \frac{a+b}{2},$$

$$\sigma_e^2 = \hat{D}(X), \quad D(X) = \frac{(b-a)^2}{12}, \quad \sigma(X) = \frac{b-a}{2\sqrt{3}}.$$

Приравняв теоретические и выборочные моменты, получим систему

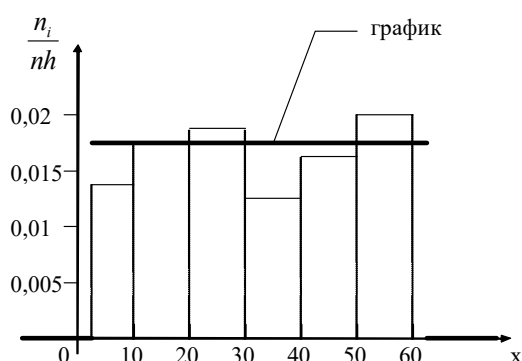
уравнений и решим её:
$$\begin{cases} \frac{a+b}{2} = 31,25 \\ \frac{(b-a)^2}{12} = 295,9375 \end{cases} \Rightarrow \begin{cases} a = 1,46 \\ b = 61,04 \end{cases}.$$
 Теоретическая частота

будет равна $\frac{1}{b-a} = \frac{1}{61,04 - 1,46} = 0,017$.

Таким образом, плотность будет равна
$$f(x) = \begin{cases} 0, & \text{если } x < 1,46 \\ 0,017, & \text{если } 1,46 < x < 61,04 \\ 0, & \text{если } x > 61,04 \end{cases}$$

$As = 0$, поскольку $\mu_3(X) = 0$, $Es = -1,2$

Рис.12. Гистограмма и график функции плотности равномерной случайной величины.



X	f_i/h
0-10	0.0138
10-20	0.0175
20-30	0.0188
30-40	0.0125
40-50	0.0175
50-60	0.02

Таблица 19. Данные для построения гистограммы.

$$h = 10; n'_i = n \cdot p = n \cdot 0,017 \cdot h = 80 \cdot 0,017 \cdot 10 = 13,6.$$

Таблица 20.

N	x_i	n_{x_i}	n'_i	n'_{x_i}	$ n'_{x_i} - n_{x_i} $
1	5	11	13.6	13.6	2.6
2	15	25	13.6	27.2	2.2
3	25	40	13.6	40.8	0.8
4	35	50	13.6	54.4	4.4 - max
5	45	64	13.6	68	4
6	55	80	13.6	81.6	1.6

$$\lambda = \frac{\max |n'_{x_i} - n_{x_i}|}{\sqrt{n}} =$$

$$= \frac{4,4}{\sqrt{80}} = 0,492$$

$$P(\lambda) = P(0,49) = 0,97$$

Так как найденная вероятность не мала (близка к единице), то гипотезу H_0 о равномерном распределении генеральной совокупности не отвергают.

* Используя приложение, можно построить доверительные интервалы для параметров равномерного распределения. Смещёнными оценками параметров будут $\hat{a} = \min_{1 \leq i \leq n} \{x_i\} = 0$,

$\hat{b} = \max_{1 \leq i \leq n} \{x_i\} = 60$. В четвёртом столбце таблицы приложения 10 приведены формулы для верхней и нижней границ доверительного интервала.

$$\hat{a} - \frac{\hat{b} - \hat{a}}{n+1} - u_\alpha \sqrt{\frac{n(\hat{b} - \hat{a})^2}{(n+1)^2(n+2)}} < a < \hat{a} - \frac{\hat{b} - \hat{a}}{n+1} + u_\alpha \sqrt{\frac{n(\hat{b} - \hat{a})^2}{(n+1)^2(n+2)}}$$

$$\hat{b} - \frac{\hat{b} - \hat{a}}{n+1} - u_\alpha \sqrt{\frac{n(\hat{b} - \hat{a})^2}{(n+1)^2(n+2)}} < b < \hat{b} - \frac{\hat{b} - \hat{a}}{n+1} + u_\alpha \sqrt{\frac{n(\hat{b} - \hat{a})^2}{(n+1)^2(n+2)}}.$$

На уровне значимости $\alpha = 0,05$ квантиль стандартного нормального распределения будет равен $u_{0,05} = 1,96$, тогда:

$$0 - \frac{60 - 0}{80 + 1} - 1,96 \sqrt{\frac{80(60 - 0)^2}{(80 + 1)^2(80 + 2)}} < b < 0 - \frac{60 - 0}{80 + 1} + \sqrt{\frac{80(60 - 0)^2}{(80 + 1)^2(80 + 2)}}$$

$$60 - \frac{60 - 0}{80 + 1} - 1,96 \sqrt{\frac{80(60 - 0)^2}{(80 + 1)^2(80 + 2)}} < b < 60 - \frac{60 - 0}{80 + 1} + \sqrt{\frac{80(60 - 0)^2}{(80 + 1)^2(80 + 2)}}$$

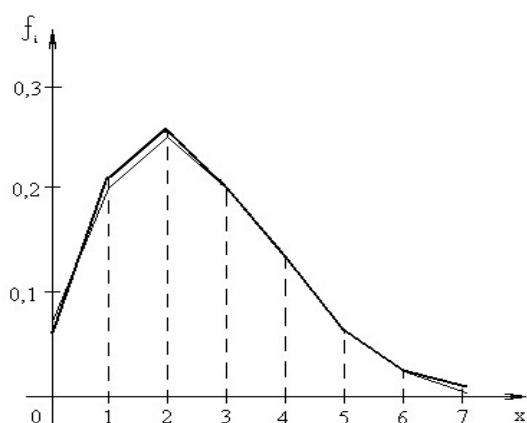
$$\text{Итак, } -2,174 < a < 0,694; \quad 57,826 < b < 60,694$$

Пример 3. Найти теоретические частоты для генеральной совокупности с распределением Пуассона и применить критерий χ^2 для проверки гипотезы H_0 о распределении генеральной совокупности на уровне значимости $\alpha = 0,05$.

По таблице несложно рассчитать выборочные числовые характеристики $\bar{X} = 2,55$; $\sigma_e^2 = 9,03 - 6,5 = 2,53$.

Таблица 21.

x_i	n_i	f_i	$x_i f_i$	x_i^2	$x_i^2 f_i$	$P(x_i)$	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
0	7	0.07	0	0	0	0.08	8	-1	1	1/8
1	21	0.21	0.21	1	0.21	0.20	20	1	1	1/20
2	26	0.26	0.52	4	1.04	0.25	25	1	1	1/25
3	21	0.21	0.63	9	1.89	0.21	21	0	0	0
4	13	0.13	0.52	16	2.08	0.13	13	0	0	0
5	7	0.07	0.35	25	1.75	0.07	7	0	0	0
6	3	0.03	0.18	36	1.08	0.03	3	0	0	0
7	2	0.02	0.14	49	0.98	0.01	1	1	1	1
Σ	100	1	2.55	-	9.03	0.97	98	-	-	1.215

Рис.13. Полигон относительных частот f_i и полигон теоретических относительных частот $P(x_i)$.

Известно, что для распределения Пуассона $M(X) = D(X) = \lambda$. Значение \bar{x} близко к σ^2 , поэтому можно предположить, что случайная величина имеет Пуассоновское распределение с параметром $\lambda = 2,55$ (поскольку точечная оценка параметра λ вычисляется по формуле:

$$\hat{\lambda} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}.$$

Формула для вычисления вероятностей $P(x_i) = \frac{e^{-2,55}}{x!} \cdot 2,55^x$.

$\chi^2_{набл} = 1,215$, число степеней свободы $\nu = 6 - 3 = 3$,

тогда на уровне значимости $\alpha = 0,05$ табличное $\chi^2 = 7,8$

Так как $1,215 < 7,8$ гипотеза о Пуассоновском распределении генеральной совокупности, из которой взята выборка, не отвергается на уровне значимости $\alpha = 0,05$.

Реализация задачи группировки наблюдений на компьютере с использованием ППП MS EXCEL

1. введите исходные данные
в первый столбец (ячейки A1-A76):

	A
1	Данные
2	2
3	0
4	7
5	3
6	4
7	2

Рис 14.

2. далее в главном меню следует последовательно выбрать пункты СЕРВИС/АНАЛИЗ ДАННЫХ/ОПИСАТЕЛЬНАЯ СТАТИСТИКА и щёлкнуть ОК.

Заполните диалоговое
окно ввода данных и
параметров вывода:

- входной интервал – диапазон, содержащий анализируемые данные;
- группирование – по столбцам или строкам;
- метки – флажок, который указывает, содержит ли первая строка название столбца или нет;
- выходной интервал (достаточно указать левую верхнюю ячейку будущего диапазона), и щёлкните ОК;

Рис.15.

3. результаты вычисления
представлены на рис.3:

Данные	
Среднее	2,835
Стандартная ошибка	0,174
Медиана	3
Мода	3
Стандартное отклонение	1,548
Дисперсия выборки	2,396
Эксцесс	0,284
Асимметричность	0,410
Интервал	7
Минимум	0
Максимум	7
Сумма	224
Счет	79

Рис.16

Группировка данных может быть произведена следующим образом:

1. для определения наибольшего и наименьшего значения выборки можно использовать функции МАКС и МИН: в ячейку A78 поместить формулу =МАКС(A2:A76), в ячейку A79 =МИН(A2:A76) и щёлкнуть ОК.

2. можно упорядочить данные, и тогда первое и последнее значения упорядоченной выборки будут наименьшим и наибольшим значениями выборки. Для этого следует поставить курсор на одну из ячеек, содержащих исходные данные (например A2), в главном меню последовательно выбрать пункты ДАННЫЕ/СОРТИРОВКА:

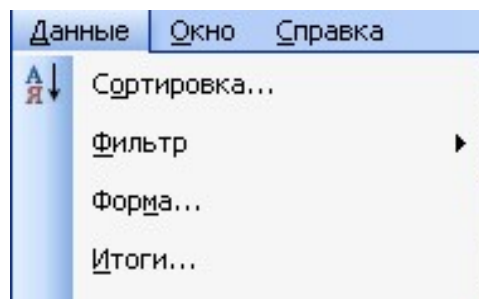


Рис.17

3. в окне сортировать по выбрать поле по которому будет производиться сортировка; выбрать вид сортировки – по возрастанию (пол убыванию), щёлкнуть ОК:

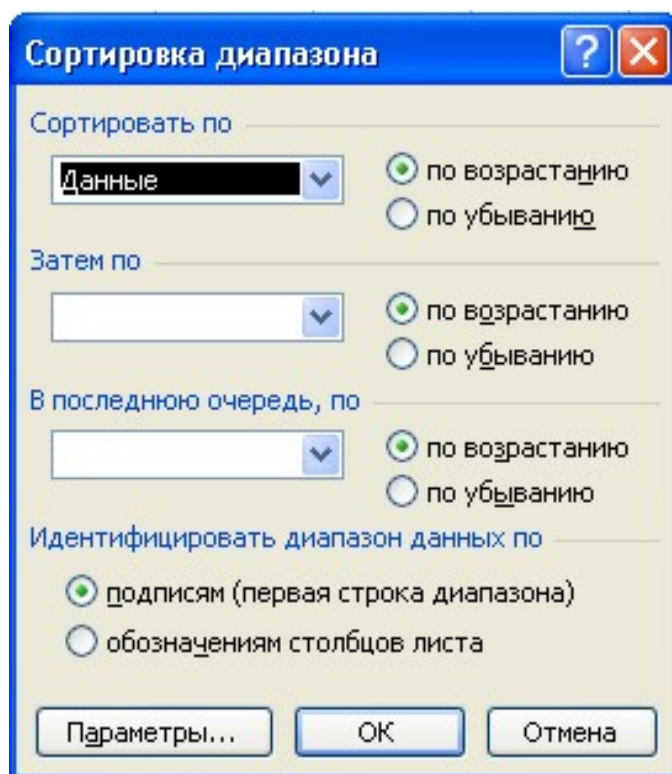


Рис.18

4. если дискретная выборка содержит небольшое количество вариантов, можно подсчитать частоты появления вариантов используя команды **ДАННЫЕ/ИТОГИ/КОЛИЧЕСТВО**:

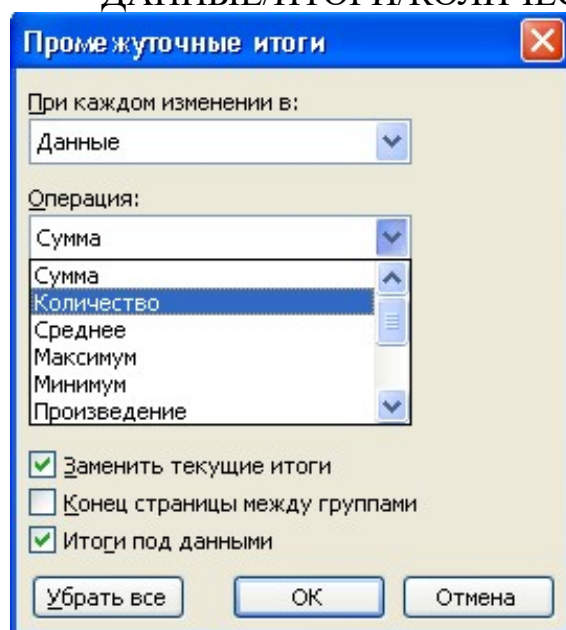


Рис. 19

5. зная количество вариантов, несложно составить таблицу для дальнейших вычислений:

	А	В
1		Данные
2		0
3		0
4		0
5		0
6	0 Количество	4
7		1
8		1
9		1
10		1
11		1
12		1
13		1
14		1
15		1
16		1
17		1
18		1
19		1
20	1 Количество	13
21		2
22		2

Рис.20.

x_i	n_i
0	4
1	13
2	14
3	24
4	16
5	3
6	3
7	2
сумма	79

Рис.21

6. Для группировки дискретных данных удобно использовать сводную таблицу. Выделите ячейки, содержащие исходные данные и в главном меню последовательно выбрать пункты ДАННЫЕ/СВОДНАЯ ТАБЛИЦА:

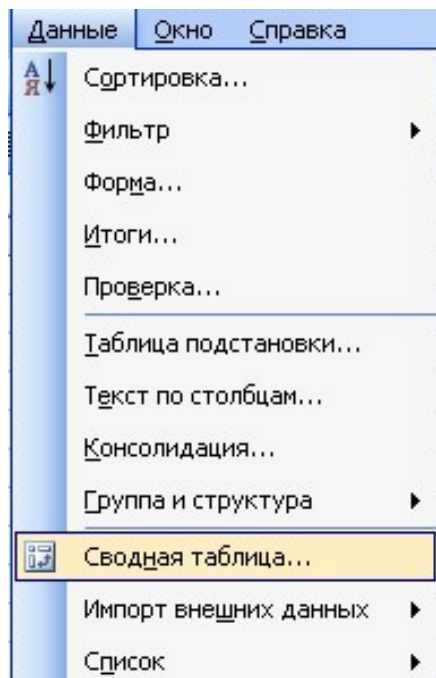


Рис.22.

7. затем необходимо заполнить диалоговые окна ввода данных и параметров вывода:

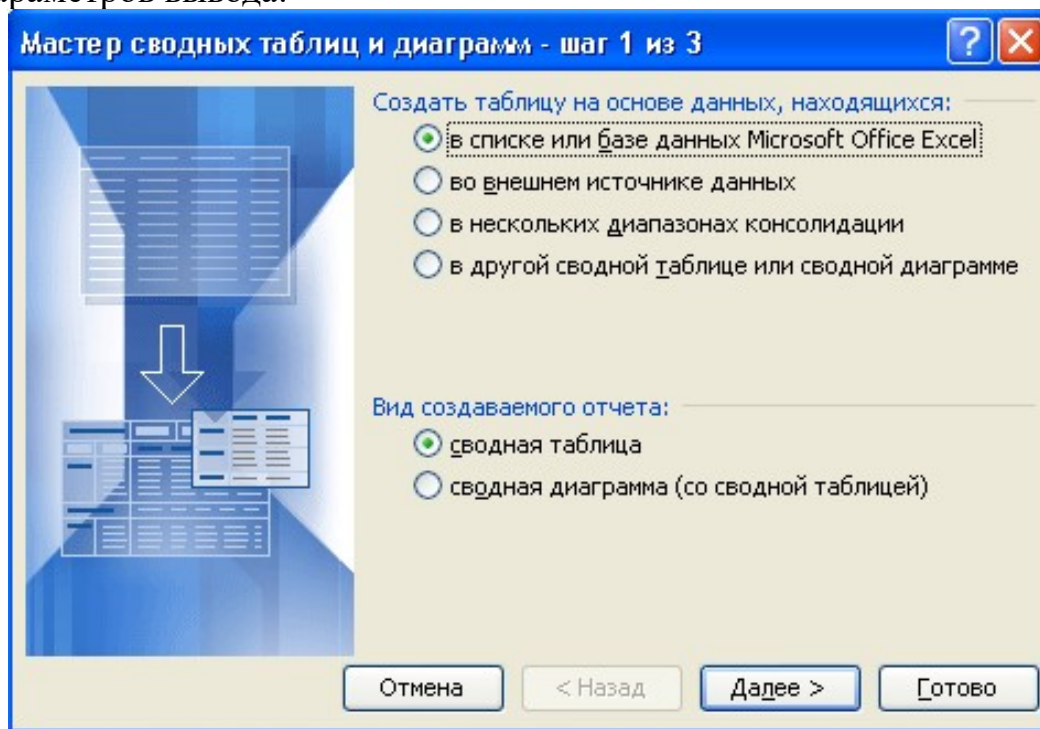


Рис.23.

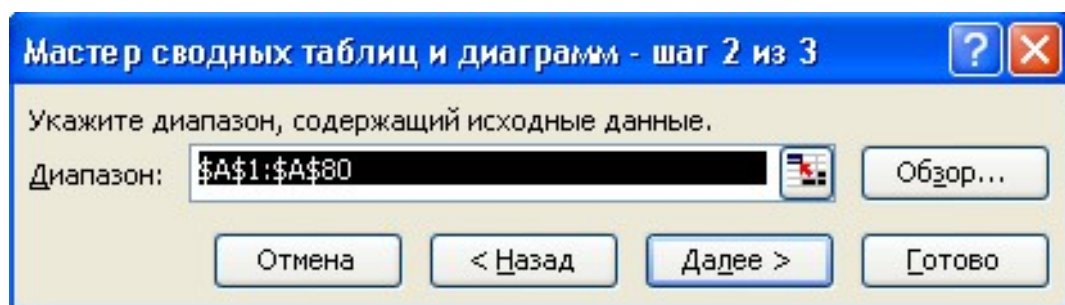


Рис.24.

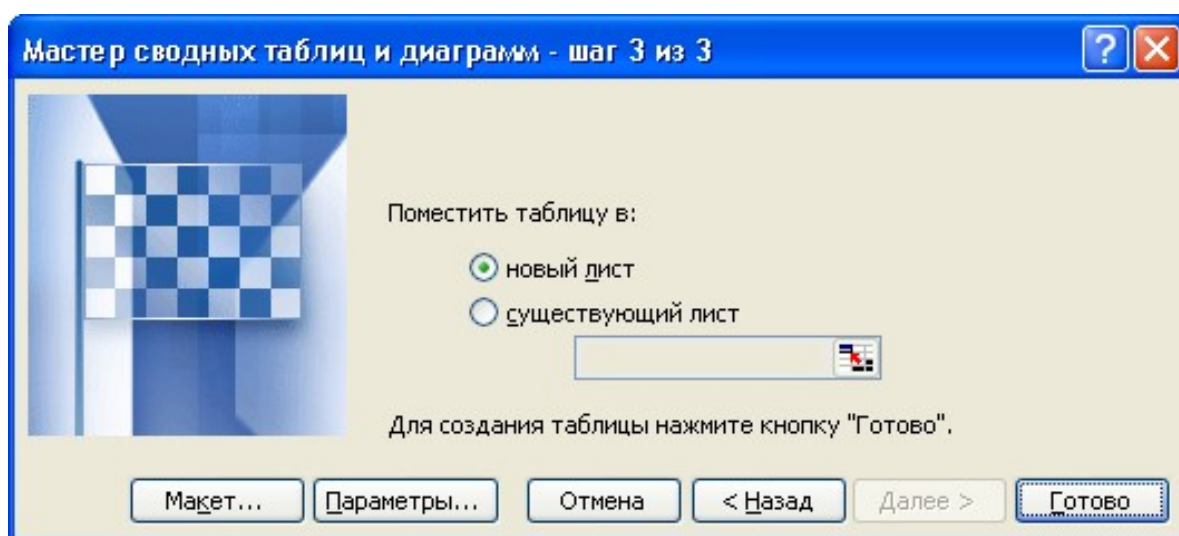


Рис.25.

8. Дважды щёлкнув по кнопке ДАННЫЕ, получим:

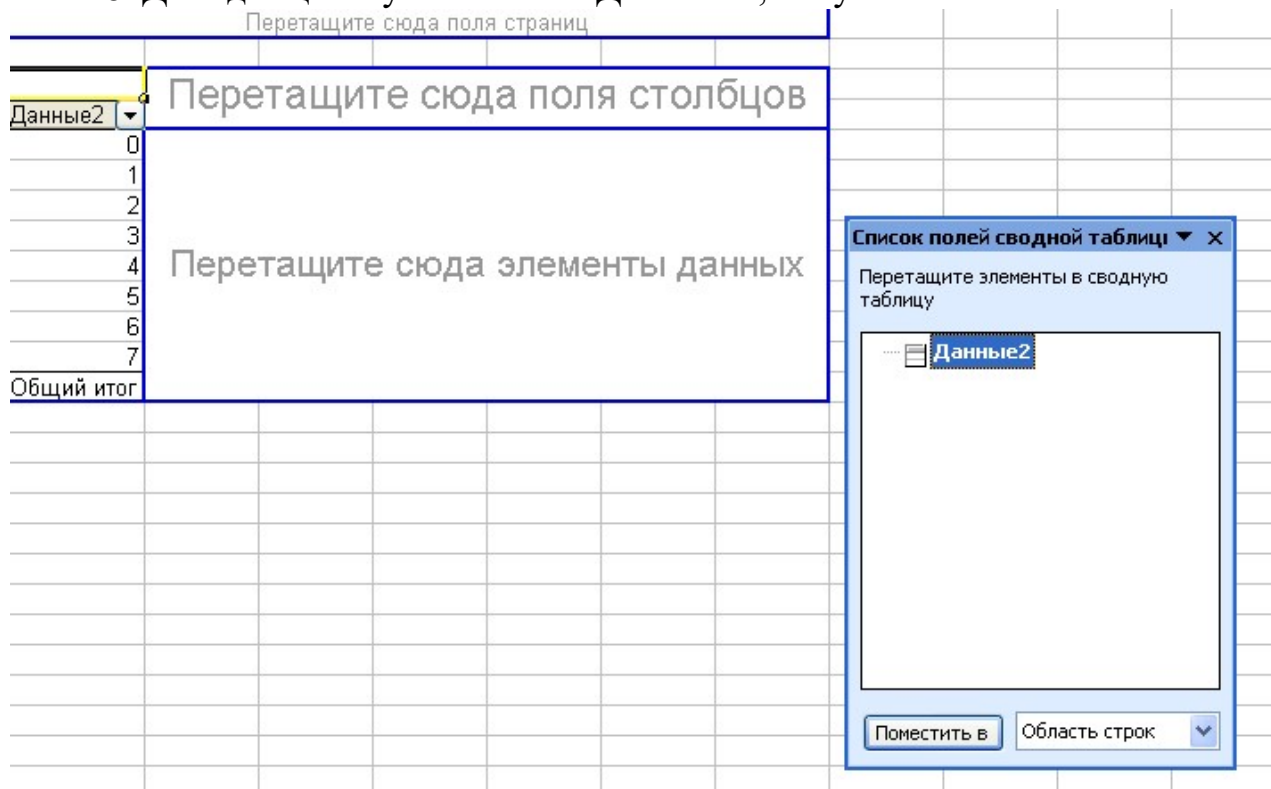


Рис.26.

9. Чтобы в сводной таблице было подсчитано количество каждой из встречающихся вариантов, щёлкнуть два раза по кнопке ДАННЫЕ и в окне ВЫЧИСЛЕНИЕ ПОЛЯ выбрать операцию КОЛИЧЕСТВО:

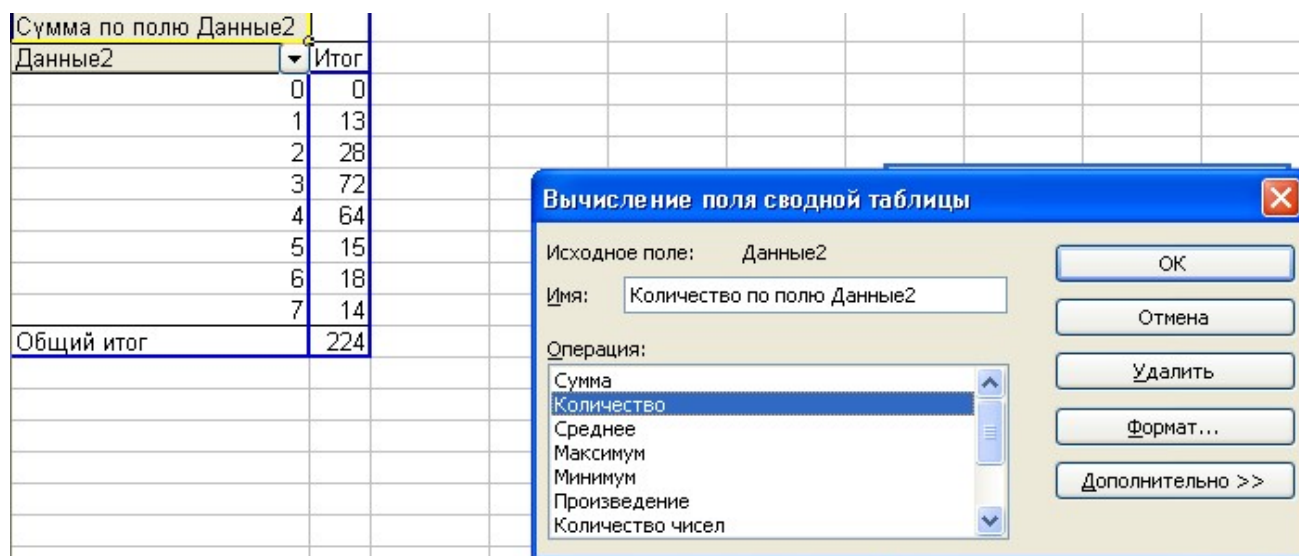


Рис.27.

10. Окончательно, сводная таблица примет вид:

Количество по полю Данные2		
Данные2		Итог
0		4
1		13
2		14
3		24
4		16
5		3
6		3
7		2
Общий итог		79

Рис.28.

Чтобы подсчитать частоты в *интервальном ряду* можно использовать два способа:

1. В первой строке поместим названия столбцов: Данные, N_i – нижняя граница i -го интервала, V_i – верхняя граница i -го интервала, $Ч_i$ – частота появления варианты в i -м интервале. В ячейки A2-A80 введите исходные данные. В ячейках B2-B80 указать нижнюю границу первого интервала, в ячейках C2-C80 – верхнюю границу, а в ячейки столбца D, начиная с D2, введите формулу: $=ЕСЛИ(И(A2>B2;A2<C2);1;0)$. Щёлкните ОК:

Данные	H1	B1	Ч1		
2	-0,5	0,5	=ЕСЛИ(И(A2>B2;A2<C2);1;0)		
0					
7					
3					
4					
2					

Рис.29.

2. Если варианта, попала в выделенный интервал, в ячейке D появится 1, если нет – 0. Подсчитайте количество вариантов, попавших в заданный интервал:

77	3	-0,5	0,5	0
78	1	-0,5	0,5	0
79	1	-0,5	0,5	0
80	5	-0,5	0,5	0
81	сумма			4

Рис.30.

3. Если необходимо скопировать формулу из ячейки D2 в G2, перед именем A2 ставится значок \$: =ЕСЛИ(И(\$A2>E2;\$A2<F2);1;0):

G2	=ЕСЛИ(И(\$A2>E2;\$A2<F2);1;0)						
	A	B	C	D	E	F	G
1	Данные	H1	B1	Ч1	H2	B2	Ч2
2	2	-0,5	0,5	0	0,5	1,5	0
3	0	-0,5	0,5	1	0,5	1,5	0
4	7	-0,5	0,5	0	0,5	1,5	0


Рис.31.

4. Чтобы просуммировать элементы столбца, используют функцию СУММ. В ячейку, где планируется рассчитать сумму, вписывают формулу: =СУММ(D:D), и щёлкают ОК. В результате получим таблицу для дальнейших вычислений:

интервалы	частоты
-0,5-0,5	4
0,5-1,5	13
1,5-2,5	14
2,5-3,5	24
3,5-4,5	16
4,5-5,5	3
5,5-6,5	3
6,5-7,5	2
сумма	79

Рис.32.

Группируя данные можно использовать фильтр. Для этого:

- поставьте курсор на одну из записей (например A2);
- задайте команды меню ДАННЫЕ/ФИЛЬТР/АВТОФИЛЬТР – на поле будет установлен значок  :

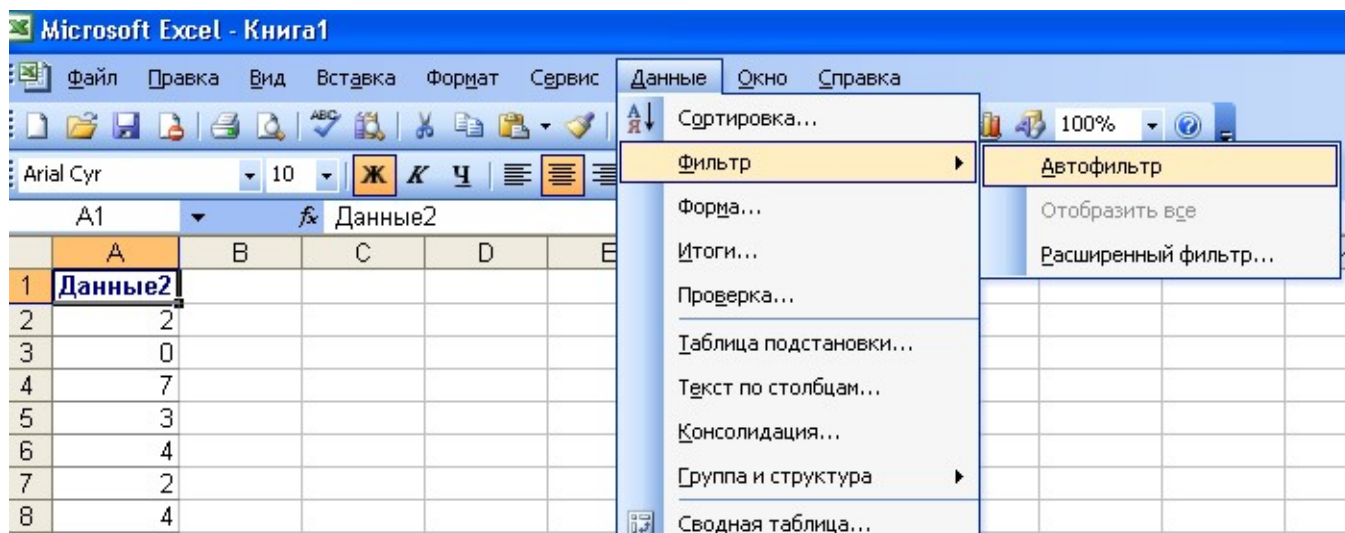


Рис.33.

3. Щёлкните по значку  и задайте условие отбора данных:

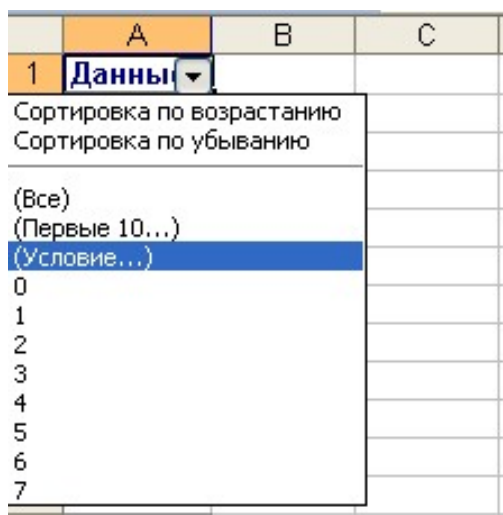


Рис.34.

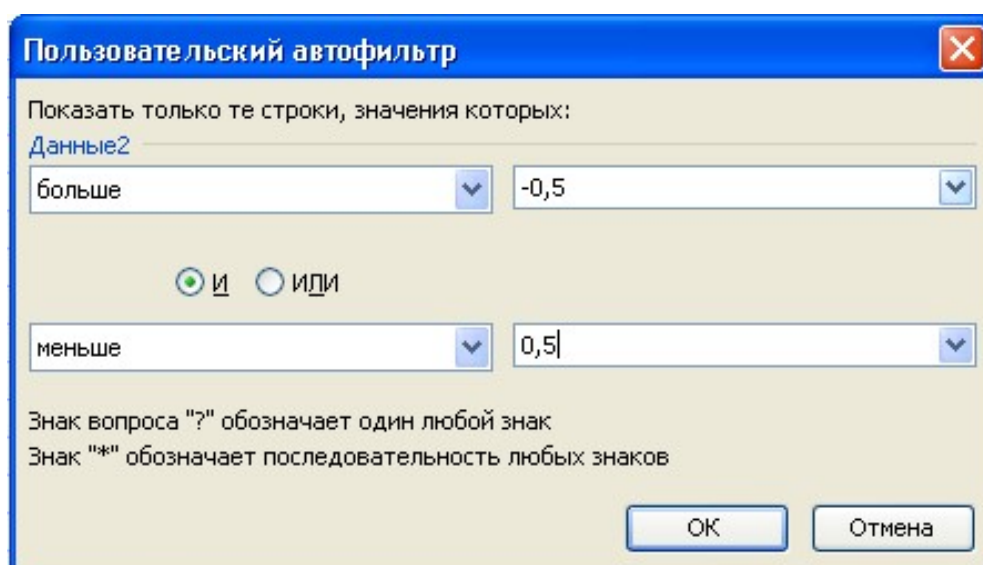



Рис.35.

4. На экран будут выведены все варианты выборки, удовлетворяющие указанному условию:

A screenshot of an Excel spreadsheet. The first cell contains a dropdown menu with the text 'Данный' and a downward arrow. Below it are four empty rows, each with a small square icon in the second column.

Данный	

Рис. 36

С использованием ППП MS EXCEL можно решать и другие задачи, математической статистики (аналитическое решение которых было рассмотрено ранее).

Наиболее часто встречающиеся распределения случайных величин

Биномиальное распределение

Возникает в тех случаях, когда исследователя интересует, сколько раз происходит некоторое событие в серии из определенного числа независимых наблюдений (опытов), выполняемых в одних условиях.

Случайная величина X имеет биномиальное распределение с параметрами n и p , если она принимает значения $0, 1, 2, \dots, n$ с вероятностями:

$$P(X = k) = C_n^k \cdot p^k \cdot (1-p)^{n-k}; \quad k = 0, 1, 2, \dots, n,$$

где p - вероятность “успеха” в испытаниях Бернулли, n - число испытаний, k - число успехов.

$$M(X) = np, \quad D(X) = np(1-p), \quad As = \frac{1-2p}{\sqrt{np(1-p)}}, \quad Es = \frac{1-6p(1-p)}{np(1-p)}$$

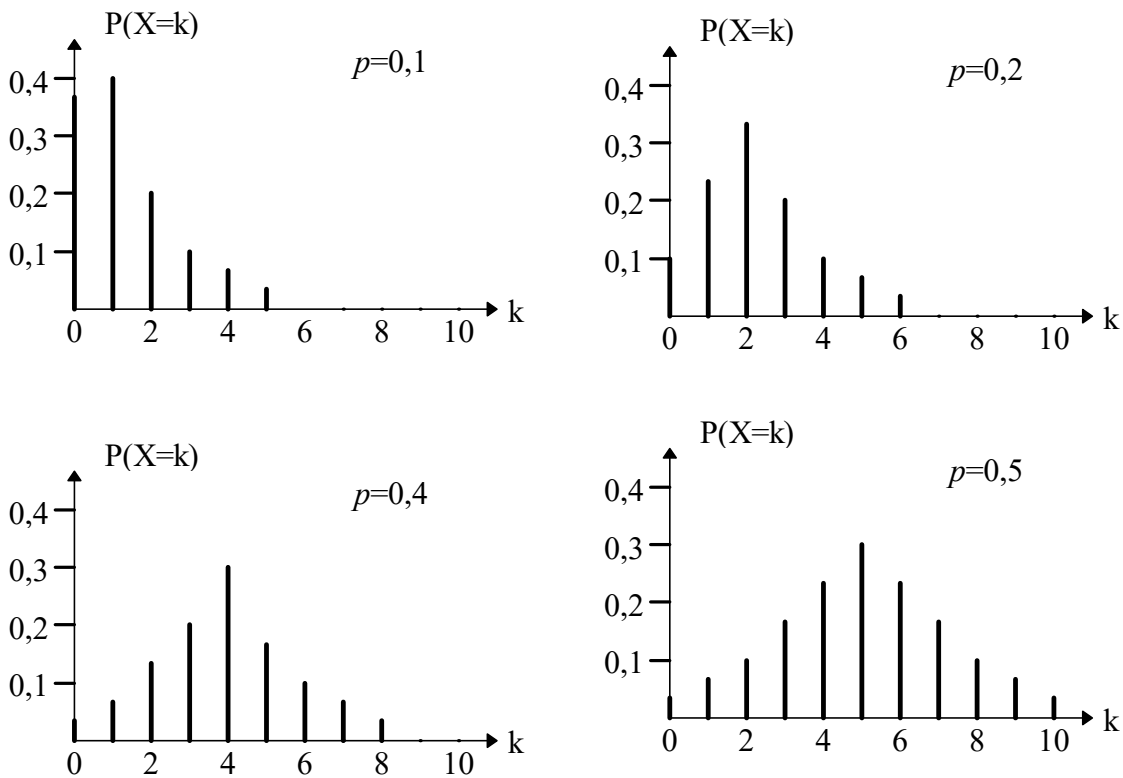


Рис.37. Вид биномиального распределения при различных значениях параметра p при $n=10$.

1). Биномиальное распределение с параметрами n и p может быть аппроксимировано нормальным распределением с параметрами $a = np$ и $\sigma = \sqrt{np(1-p)}$, если выполняются условия: $np(1-p) > 5$, $0.1 \leq p \leq 0.9$; при $np(1-p) > 25$ аппроксимация применяется независимо от p .

2). Биноминальное распределение с параметрами n и p может быть аппроксимировано распределением Пуассона со средним np , при условии, что $p < 0,1$ и n достаточно велико.

Распределение Пуассона.

Возникает в тех случаях, когда исследователя интересуют «редкие» события, которые в серии из большого числа независимых наблюдений (опытов), происходят с очень малой вероятностью.

Случайная величина X , которая принимает только целые неотрицательные значения $0, 1, \dots$ имеет закон распределения Пуассона с параметром λ , если

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

$$M(X) = \lambda, \quad D(X) = \lambda.$$

$$As = \frac{1}{\sqrt{\lambda}}, \quad Es = \frac{1}{\lambda}$$

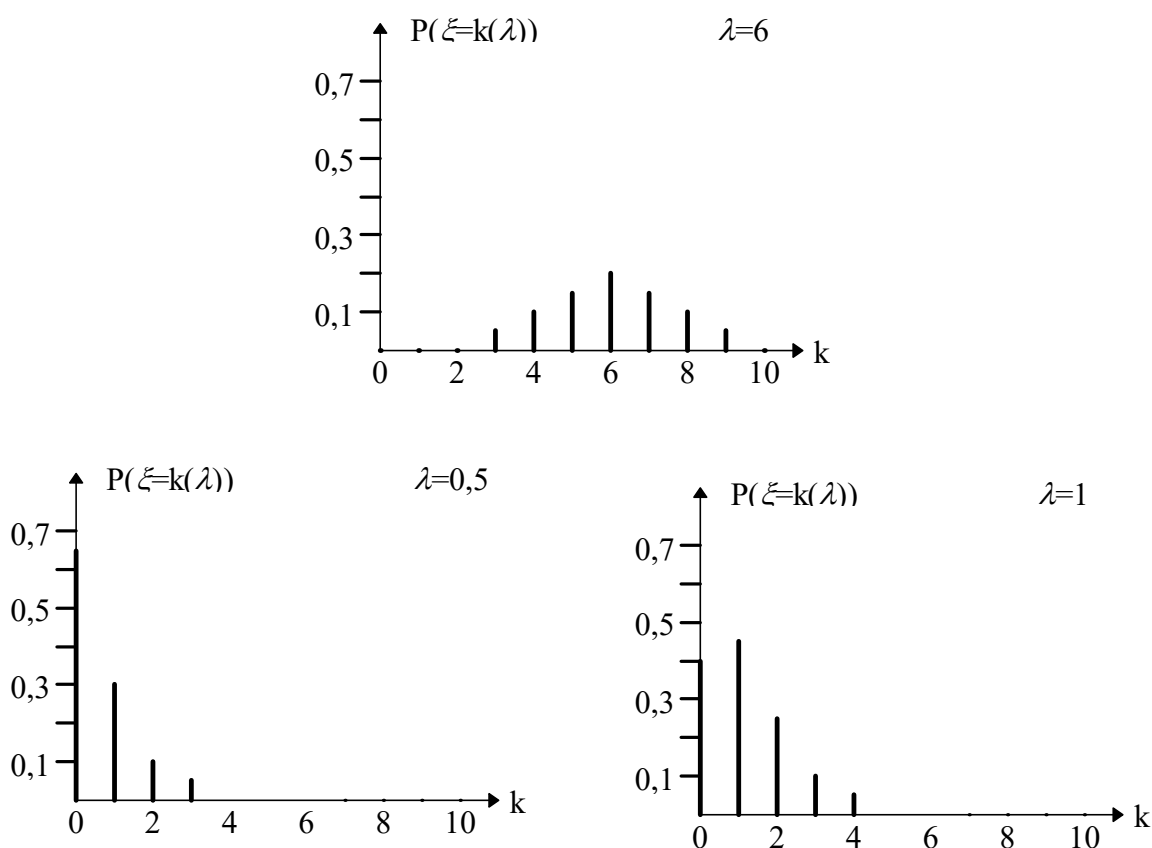


Рис.38. Вид распределения Пуассона для различных значений λ .

1). При больших n и малых p :

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots, \text{ где } \lambda = np.$$

2). При $\lambda > 9$ распределение Пуассона может быть аппроксимировано нормальным распределением с параметрами $a = \lambda$ и дисперсией λ .

$P(X = k(\lambda)) \approx \frac{1}{\sqrt{\lambda}} \varphi\left(\frac{k - \lambda}{\sqrt{\lambda}}\right)$, где φ - плотность нормального распределения $N(0,1)$.

Равномерное распределение.

Равномерное распределение может иметь, например, такая случайная величина как время ожидания общественного транспорта. Случайная величина X равномерно распределена на отрезке $[a; b]$, если её плотность вероятности имеет вид:

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

$$MX^k = \frac{b^{k+1} - a^{k+1}}{(b-a)(k+1)}, \quad k = 1, 2, \dots, \quad D(X) = \frac{(b-a)^2}{12}, \quad As = 0, \quad Es = \frac{b+a}{2}.$$

При помощи линейного преобразования $y = \frac{x-a}{b-a}$ приводится к равномерному распределению на $[0, 1]$.

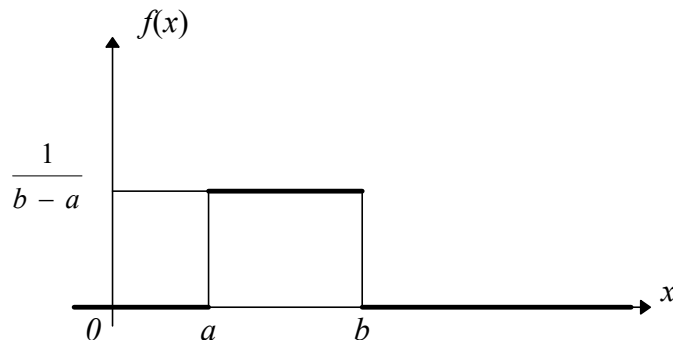


Рис.39. Функция плотности равномерного распределения.

Показательное (экспоненциальное) распределение.

Показательное распределение широко используется в теории надёжности (например, такое распределение может иметь случайная величина - время безотказной работы прибора).

Неотрицательная случайная величина X имеет показательное распределение с параметром $\theta > 0$, если её плотность задана формулой:

$$f(x, \theta) = \theta \cdot e^{-\theta x}, \quad x \geq 0.$$

$$M(x) = \frac{1}{\theta}, \quad D(x) = \frac{1}{\theta^2}, \quad As = 2, \quad Es = 6, \quad Me = \frac{\ln 2}{\theta}, \quad M(X^k) = \frac{k!}{\theta^k}.$$

$$F(x, \theta) = 1 - e^{-\theta x}, \quad x \geq 0.$$

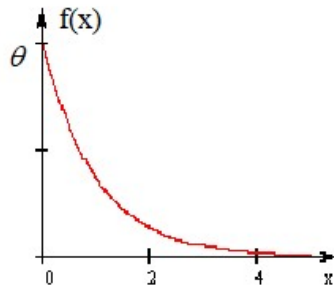


Рис 40. График плотности распределения.

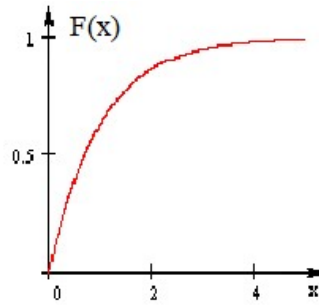


Рис.41.График функции распределения.

Нормальное распределение.

В силу действия закона больших чисел, большинство случайных величин, анализируемых в приложениях – нормальное (например, распределения по росту, весу, доходам и т.п.) Случайная величина X имеет нормальное распределение вероятностей с параметрами a и σ ($X \sim N(a, \sigma)$), если ее плотность распределения задана формулой:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, -\infty < x < \infty.$$

В этом случае ее функция распределения имеет вид:

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt.$$

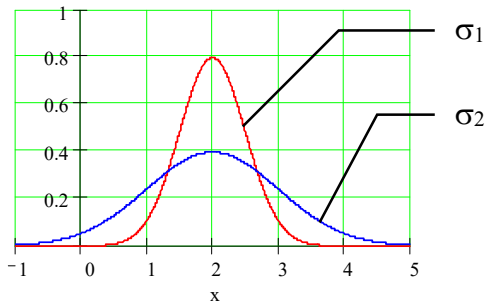


Рис.42.Функция плотности нормального распределения при $\sigma_1 = 0,5$, $\sigma_2 = 1$ и $a = 2$.

$x = a$ - ось симметрии графика плотности. Максимальное значение функции плотности равно $\frac{1}{\sqrt{2\pi}\sigma}$.

Точки $a - \sigma$ и $a + \sigma$ являются точками перегиба. Значение плотности в этих точках равно $\frac{1}{\sqrt{2\pi}\sigma}$.

$$M(X) = a, D(X) = \sigma^2, As = Es = 0, Mo = Me = \bar{x}.$$

$N(0,1)$ - стандартное нормальное распределение с плотностью

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{ где } a = 0, \sigma = 1.$$

Функция распределения нормального стандартного распределения имеет следующий вид $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy$ и носит название функции

Лапласа. Свойства $\varphi(x)$:

1. область определения $(-\infty; +\infty)$;
2. $f(x) > 0$;
3. $\lim_{x \rightarrow \pm\infty} f(x) = 0$;
4. в точках $x = a \pm \sigma$ нормальная кривая имеет перегиб.

Распределение χ^2 .

Пусть случайные величины $\xi_1, \xi_2, \dots, \xi_n$ - независимы и имеют стандартное нормальное распределение $N(0, 1)$, тогда случайная величина χ^2 определяемая как $\chi_n^2 = \xi_1^2 + \xi_2^2 + \dots + \xi_n^2$ имеет распределение χ^2 с n степенями свободы (χ_n^2 (для любых $n \geq 1$) положительна).

Функция плотности χ_n^2 при $x > 0$ задаётся следующим образом:

$$f(x) = \frac{1}{2^{n/2}} \cdot \frac{1}{\Gamma(n/2)} \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}, \text{ где } \Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx - \text{гамма функция.}$$

Показательное распределение с параметром $\theta = \frac{1}{2}$ является χ^2 распределением с двумя степенями свободы.

$$M(\chi_n^2) = n, \quad D(\chi_n^2) = 2n.$$

Функция распределения χ_n^2 при $x < 0$ равна 0, а при $x \geq 0$ задается следующей формулой:

$$F(x) = \frac{1}{2^{n/2} \cdot \Gamma(n/2)} \cdot \int_0^x t^{\frac{n}{2}-1} \cdot e^{-\frac{t}{2}} dt.$$

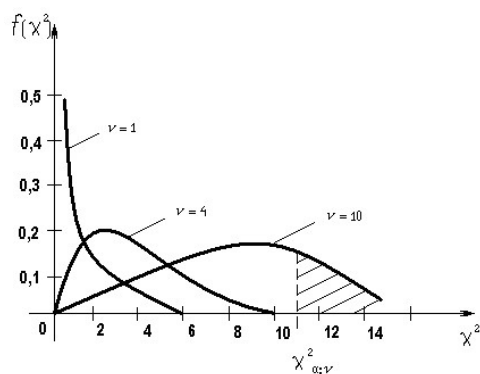


Рис.43. Функция плотности распределения χ^2 с различным числом степеней свободы.

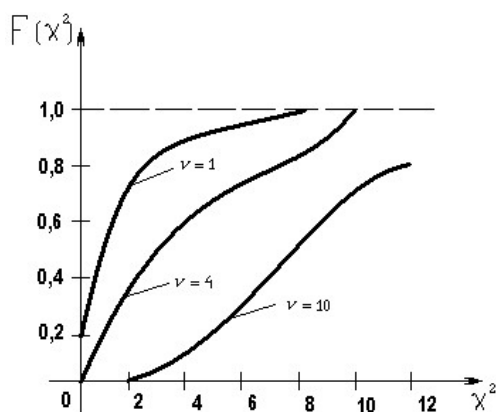


Рис 44. Функция распределения χ^2 с различным числом степеней свободы.

Распределение Стьюдента.

Пусть случайные величины $\xi_1, \xi_2, \dots, \xi_n$ - независимы и имеют стандартное нормальное распределение $N(0, 1)$. Случайная величина t_n , определяемая следующим образом $t_n = \frac{\xi_0}{\sqrt{\frac{1}{n} \cdot \sum_{i=1}^n \xi_i^2}}$, имеет распределение

Стьюдента. n - число степеней свободы.

Плотность распределения Стьюдента равна:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \cdot \sqrt{n\pi}} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

График функции плотности симметричен относительно $x = 0$.

$$M(t_n) = 0, \quad D(t_n) = \frac{n}{n-2}.$$

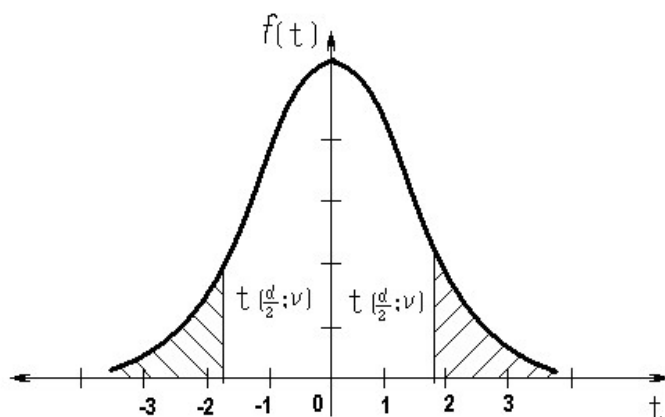


Рис.45. Функция плотности распределения Стьюдента с различным числом степеней свободы n .

Распределение Фишера.

Пусть случайные величины X_1, X_2, \dots, X_m и Y_1, Y_2, \dots, Y_n независимы и имеют нормальное распределение с параметрами $MX_i = MY_j = 0$, $DX_i = DY_j = 1$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$. Тогда безразмерная случайная величина

$$F = \frac{n \sum_{i=1}^m X_i^2}{m \sum_{j=1}^n Y_j^2}$$
 имеет распределение Фишера, с плотностью вероятности:

$$f(x) = \begin{cases} \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right) \nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right) \Gamma\left(\frac{\nu_2}{2}\right)} \frac{x^{\frac{\nu_1}{2}-1}}{(\nu_2 + \nu_1 x)^{\frac{\nu_1 + \nu_2}{2}}}, & x > 0 \\ 0, & x \leq 0 \end{cases}, \text{ где } \nu_1 = m - \text{число степеней}$$

свободы числителя, а $\nu_2 = n$ - число степеней свободы знаменателя.

Распределение Фишера зависит от двух параметров: $\nu_1 = m$ и $\nu_2 = n$ (число степеней свободы числителя и знаменателя).

$$MX = \frac{\nu_2}{\nu_2 - 2}, \nu_2 > 2, \quad DX = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}, \nu_2 > 4.$$

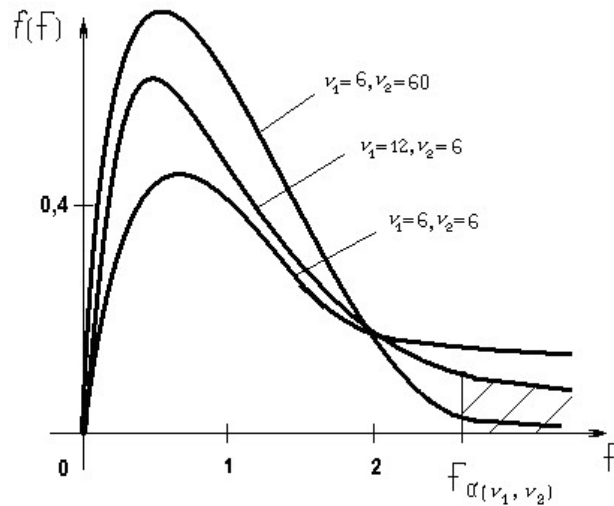


Рис.46. Функция плотности распределения Фишера с различным числом степеней свободы.

Приложение 2.

Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

	0	1	2	3	4	5	6	7	8	9
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	3726	3712	3697
0,4	3683	3668	3652	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	2168	2444
1,0	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
2,0	0,0540	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	0043
3,0	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

Приложение 3.

Таблица значений функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0,00	0,0000	0,45	0,1736	0,90	0,3159	1,35	0,4115	1,80	0,4641	2,50	0,4938
0,01	0,0040	0,46	0,1772	0,91	0,3186	1,36	0,4131	1,81	0,4649	2,52	0,4941
0,02	0,0080	0,47	0,1808	0,92	0,3212	1,37	0,4147	1,82	0,4656	2,54	0,4945
0,03	0,0120	0,48	0,1844	0,93	0,3238	1,38	0,4162	1,83	0,4664	2,56	0,4948
0,04	0,0160	0,49	0,1879	0,94	0,3264	1,39	0,4177	1,84	0,4671	2,58	0,4951
0,05	0,0199	0,50	0,1915	0,95	0,3289	1,40	0,4192	1,85	0,4678	2,60	0,4953
0,06	0,0239	0,51	0,1950	0,96	0,3315	1,41	0,4207	1,86	0,4680	2,62	0,4956
0,07	0,0279	0,52	0,1985	0,97	0,3340	1,42	0,4222	1,87	0,4693	2,64	0,4959
0,08	0,0319	0,53	0,2019	0,98	0,3365	1,43	0,4236	1,88	0,4699	2,66	0,4961
0,09	0,0359	0,54	0,2054	0,99	0,3389	1,44	0,4251	1,89	0,4706	2,68	0,4963
0,10	0,0398	0,55	0,2088	1,00	0,3413	1,45	0,4265	1,90	0,4713	2,70	0,4965
0,11	0,0438	0,56	0,2123	1,01	0,3438	1,46	0,4279	1,91	0,4719	2,72	0,4967
0,12	0,0478	0,57	0,2157	1,02	0,3461	1,47	0,4292	1,92	0,4726	2,74	0,4969
0,13	0,0517	0,58	0,2190	1,03	0,3485	1,48	0,4306	1,93	0,4732	2,76	0,4971
0,14	0,0557	0,59	0,2224	1,04	0,3508	1,49	0,4319	1,94	0,4738	2,78	0,4973
0,15	0,0596	0,60	0,2257	1,05	0,3531	1,50	0,4332	1,95	0,4744	2,80	0,4974
0,16	0,0636	0,61	0,2291	1,06	0,3554	1,51	0,4345	1,96	0,4750	2,82	0,4976
0,17	0,0675	0,62	0,2324	1,07	0,3577	1,52	0,4357	1,97	0,4756	2,84	0,4977
0,18	0,0714	0,63	0,2357	1,08	0,3599	1,53	0,4370	1,98	0,4761	2,86	0,4979
0,19	0,0753	0,64	0,2389	1,09	0,3621	1,54	0,4382	1,99	0,4767	2,88	0,4980
0,20	0,0793	0,65	0,2422	1,10	0,3643	1,55	0,4394	2,00	0,4772	2,90	0,4981
0,21	0,0832	0,66	0,2454	1,11	0,3665	1,56	0,4406	2,02	0,4783	2,92	0,4982
0,22	0,0871	0,67	0,2486	1,12	0,3686	1,57	0,4418	2,04	0,4793	2,94	0,4984
0,23	0,0910	0,68	0,2517	1,13	0,3708	1,58	0,4429	2,06	0,4803	2,96	0,4985
0,24	0,0948	0,69	0,2549	1,14	0,3729	1,59	0,4441	2,08	0,4812	2,98	0,4986
0,25	0,0987	0,70	0,2580	1,15	0,3749	1,60	0,4452	2,10	0,4821	3,00	0,49865
0,26	0,1026	0,71	0,2611	1,16	0,3770	1,61	0,4463	2,12	0,4830	3,10	0,49903
0,27	0,1064	0,72	0,2642	1,17	0,3790	1,62	0,4474	2,14	0,4838	3,20	0,49931
0,28	0,1103	0,73	0,2673	1,18	0,3810	1,63	0,4484	2,16	0,4846	3,30	0,49952
0,29	0,1141	0,74	0,2703	1,19	0,3830	1,64	0,4495	2,18	0,4854	3,40	0,49966
0,30	0,1179	0,75	0,2734	1,20	0,3849	1,65	0,4505	2,20	0,4861	3,50	0,49977
0,31	0,1217	0,76	0,2764	1,21	0,3869	1,66	0,4515	2,22	0,4868	3,60	0,49984
0,32	0,1255	0,77	0,2794	1,22	0,3883	1,67	0,4525	2,24	0,4875	3,70	0,49989
0,33	0,1293	0,78	0,2823	1,23	0,3907	1,68	0,4535	2,26	0,4881	3,80	0,49993
0,34	0,1331	0,79	0,2852	1,24	0,3925	1,69	0,4545	2,28	0,4887	3,90	0,49995
0,35	0,1368	0,80	0,2881	1,25	0,3944	1,70	0,4554	2,30	0,4893	4,00	0,499968
0,36	0,1406	0,81	0,2910	1,26	0,3962	1,71	0,4564	2,32	0,4898	4,50	0,499997
0,37	0,1443	0,82	0,2939	1,27	0,3980	1,72	0,4573	2,34	0,4904	5,00	0,499997
0,38	0,1480	0,83	0,2967	1,28	0,3997	1,73	0,4582	2,36	0,4909		
0,39	0,1517	0,84	2995	1,29	0,4015	1,74	0,4591	2,38	0,4913		
0,40	0,1554	0,85	0,3023	1,30	0,4032	1,75	0,4599	2,40	0,4918		
0,41	0,1591	0,86	0,3051	1,31	0,4049	1,76	0,4608	2,42	0,4922		
0,42	0,1628	0,87	0,3078	1,32	0,4066	1,77	0,4616	2,44	0,4927		
0,43	0,1664	0,88	0,3106	1,33	0,4082	1,78	0,4625	2,46	0,4931		
0,44	0,1700	0,89	0,3133	1,34	0,4099	1,79	0,4633	2,48	0,4934		

Приложение 4.

Квантили стандартизованного нормального распределения для наиболее употребительных значений доверительной вероятности $1 - \alpha$:

Доверительная вероятность, $1 - \alpha$	0.90	0.95	0.99	0.9973	0.999
Квантили $u_{\frac{\alpha}{2}}$	1.64	1.96	2.58	3.00	3.37

Приложение 5.

Таблица значений $\chi^2_{\alpha, \nu}$

ν	α					
	0,20	0,10	0,05	0,02	0,01	0,001
1	1,642	2,706	3,841	5,412	6,635	10,827
2	3,219	4,605	5,991	7,824	9,210	13,815
3	4,642	6,251	7,815	9,837	11,345	16,266
4	5,989	7,779	9,488	11,668	13,277	18,467
5	7,289	9,236	11,070	13,388	15,086	20,515
6	8,558	10,645	12,592	15,033	16,812	22,457
7	9,803	12,017	14,067	16,622	18,475	24,322
8	11,030	13,362	15,507	18,168	20,090	26,125
9	12,242	14,684	16,919	19,679	21,666	27,877
10	13,442	15,987	18,307	21,161	23,209	29,588
11	14,631	17,275	19,675	22,618	24,725	31,264
12	15,812	18,549	21,026	24,054	26,217	32,909
13	16,985	19,812	22,362	25,472	27,688	34,528
14	18,151	21,064	23,685	26,783	29,141	36,123
15	19,311	22,307	24,996	28,259	30,578	37,697
16	20,465	23,542	26,296	29,633	32,000	39,252
17	21,615	24,769	27,587	30,995	33,409	40,790
18	22,760	25,989	28,869	32,346	34,805	42,312
19	23,900	27,204	30,144	33,687	36,191	43,820
20	25,038	28,412	31,410	35,020	37,566	45,315
21	26,171	29,615	32,671	36,343	38,932	46,797
22	27,301	30,813	33,924	37,659	40,289	48,268
23	28,429	32,007	35,172	38,968	41,638	49,728
24	29,553	33,196	36,415	40,270	42,980	51,179
25	30,675	34,382	37,652	41,566	44,314	52,620
26	31,795	35,563	38,885	42,856	45,642	54,052
27	32,912	36,741	40,113	44,140	46,963	55,476
28	34,027	37,916	41,337	45,419	48,278	56,893
29	35,139	39,087	42,557	46,693	49,588	58,302
30	36,250	40,256	43,773	47,962	50,892	59,703

Приложение 6.

Таблица критических значений $t_{\alpha, \nu}$, удовлетворяющих условию $P(t \geq t_{\alpha, \nu}) = \int_{t_{\alpha, \nu}}^{\infty} p(t, \nu) dt = \alpha$

ν, α	0,40	0,30	0,20	0,10	0,050	0,02 5	0,001	0,005	0,001	0,0005
1	0,325	0,727	1,376	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,33	31,60
3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,22	12,94
4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	5,032	5,893	6,859
6	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,405
8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,294	4,781
10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,260	0,540	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,259	0,539	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,259	0,538	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,258	0,537	0,868	1,345	1,761	2,145	2,624	3,977	3,787	4,140
15	0,258	0,536	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,258	0,535	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,257	0,534	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,257	0,534	0,862	1,330	1,734	2,101	2,552	2,878	3,611	3,922
19	0,257	0,533	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,257	0,533	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,257	0,532	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,256	0,532	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,256	0,532	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,256	0,531	0,857	1,318	1,711	2,064	2,492	2,797	3,46	3,745
25	0,256	0,531	0,856	1,316	1,708	2,060	2,485	2,787	3,45	3,725
26	0,256	0,531	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,256	0,531	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,256	0,530	0,855	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,256	0,530	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,256	0,530	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,255	0,529	0,851	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,255	0,528	0,849	1,298	1,676	2,009	2,403	2,678	3,262	3,495
60	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
80	0,254	0,527	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,415
100	0,254	0,526	0,845	1,290	1,660	1,984	2,365	2,626	3,174	3,389
200	0,254	0,525	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,339
500	0,253	0,525	0,842	1,283	1,648	1,965	2,334	2,586	3,106	3,310
∞	0,253	0,524	0,842	1,282	1,645	1,96	2,326	2,576	3,090	3,291

Приложение 7.

Вспомогательные коэффициенты a_{nk} для проверки нормальности распределения по критерию W Шапиро-Уилки (n – объём совокупности, k – номер сравниваемой пары).

k	N							
	3	4	5	6	7	8	9	10
1	0,7071	0,6872	0,6646	0,6431	0,6233	0,6052	0,5888	0,5739
2		0,1677	0,2413	0,2806	0,3031	0,3164	0,3244	0,3291
3				0,0875	0,1401	0,1743	0,1976	0,2141
4						0,0561	0,0947	0,1224
5								0,0399

k	n							
	11	12	13	14	15	16	17	18
1	0,5601	0,5475	0,5359	0,5251	0,5150	0,5056	0,4968	0,4886
2	0,3315	0,3325	0,3325	0,3318	0,3306	0,3290	0,3273	0,3253
3	0,2260	0,2347	0,2412	0,2460	0,2495	0,2521	0,2540	0,2553
4	0,1429	0,1585	0,1707	0,1802	0,1878	0,1939	0,1988	0,2027
5	0,0695	0,0922	0,1099	0,1240	0,1353	0,1447	0,1524	0,1587
6		0,0303	0,0539	0,0727	0,0880	0,1005	0,1109	0,1197
7				0,0240	0,0433	0,0593	0,0725	0,0837
8						0,0196	0,0359	0,0496
9								0,0163

k	n							
	19	20	21	22	23	24	25	26
1	0,4808	0,4734	0,4643	0,4590	0,4542	0,4493	0,4450	0,4407
2	0,3232	0,3211	0,3185	0,3156	0,3126	0,3098	0,3069	0,3043
3	0,2561	0,2565	0,2578	0,2571	0,2563	0,2554	0,2543	0,2533
4	0,2059	0,2085	0,2119	0,2131	0,2139	0,2145	0,2148	0,2151
5	0,1641	0,1686	0,1736	0,1764	0,1787	0,1807	0,1822	0,1836
6	0,1271	0,1334	0,1399	0,1443	0,1480	0,1512	0,1539	0,1563
7	0,0932	0,1013	0,1092	0,1150	0,1201	0,1245	0,1283	0,1316
8	0,0612	0,0711	0,0804	0,0878	0,0941	0,0997	0,1046	0,1089
9	0,0303	0,0422	0,0530	0,0618	0,0696	0,0764	0,0823	0,0876
10		0,0140	0,0263	0,0368	0,0459	0,0539	0,0610	0,0672
11				0,0122	0,0228	0,0321	0,0403	0,0476
12						0,0107	0,0200	0,0284
13								0,0094

Приложение 8.

Значения функции $P(\lambda) = 1 - K(\lambda) = P(D \geq \lambda) = 1 - \sum_{k=-\infty}^{+\infty} (-1)^k e^{-k^2 \lambda^2}$

λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$
≤0,29	1,00000	0,76	0,6104	1,23	0,0970	1,70	0,0062	2,17	0,0002
0,30	0,99999	0,77	0,5936	1,24	0,0924	1,71	0,0058	2,18	0,0001
0,31	0,99998	0,78	0,5770	1,25	0,0879	1,72	0,0054	2,19	0,0001
0,32	0,99995	0,79	0,5605	1,26	0,0836	1,73	0,0050	2,20	0,0001
0,33	0,99991	0,80	0,5441	1,27	0,0794	1,74	0,0047	2,21	0,0001
0,34	0,99993	0,81	0,5280	1,28	0,0755	1,75	0,0044	2,22	0,0001
0,35	0,9997	0,82	0,5120	1,29	0,0717	1,76	0,0041	2,23	0,0001
0,36	0,9995	0,83	0,4962	1,30	0,0681	1,77	0,0038	2,24	0,0001
0,37	0,9992	0,84	0,4806	1,31	0,0646	1,78	0,0035	2,25	0,0001
0,38	0,9987	0,85	0,4653	1,32	0,0613	1,79	0,0033	2,26	0,0001
0,39	0,9981	0,86	0,4503	1,33	0,0582	1,80	0,0031	2,27	0,0001
0,40	0,9960	0,87	0,4355	1,34	0,0551	1,81	0,0029	2,28	0,0001
0,41	0,9972	0,88	0,4209	1,35	0,0522	1,82	0,0027	2,29	0,0001
0,42	0,9945	0,89	0,4067	1,36	0,0495	1,83	0,0025	2,30	0,0001
0,43	0,9926	0,90	0,3927	1,37	0,0469	1,84	0,0023	2,31	0,000046
0,44	0,9903	0,91	0,3791	1,38	0,0444	1,85	0,0021	2,32	0,000042
0,45	0,9874	0,92	0,3657	1,39	0,0420	1,86	0,0020	2,33	0,000038
0,46	0,9840	0,93	0,3527	1,40	0,0397	1,87	0,0019	2,34	0,000035
0,47	0,9800	0,94	0,3399	1,41	0,0375	1,88	0,0017	2,35	0,000032
0,48	0,9753	0,95	0,3275	1,42	0,0354	1,89	0,0016	2,36	0,000030
0,49	0,9700	0,96	0,3154	1,43	0,0335	1,90	0,0015	2,37	0,000027
0,50	0,9639	0,97	0,3036	1,44	0,0316	1,91	0,0014	2,38	0,000024
0,51	0,9572	0,98	0,2921	1,45	0,0298	1,92	0,0013	2,39	0,000022
0,52	0,9497	0,99	0,2809	1,46	0,0282	1,93	0,0012	2,40	0,000020
0,53	0,9415	1,00	0,2700	1,47	0,0266	1,94	0,0011	2,41	0,000018
0,54	0,9325	1,01	0,2594	1,48	0,0250	1,95	0,0010	2,42	0,000016
0,55	0,9228	1,02	0,2492	1,49	0,0236	1,96	0,0009	2,43	0,000014
0,56	0,9124	1,03	0,2392	1,50	0,0222	1,97	0,0009	2,44	0,000013
0,57	0,9013	1,04	0,2296	1,51	0,0209	1,98	0,0008	2,45	0,000012
0,58	0,8896	1,05	0,2202	1,52	0,0197	1,99	0,0007	2,46	0,000011
0,59	0,8772	1,06	0,2111	1,53	0,0185	2,00	0,0007	2,47	0,000010
0,60	0,8643	1,07	0,2024	1,54	0,0174	2,01	0,0006	2,48	0,000009
0,61	0,8508	1,08	0,1939	1,55	0,0164	2,02	0,0006	2,49	0,000008
0,62	0,8368	1,09	0,1857	1,56	0,0154	2,03	0,0005	2,50	0,0000075
0,63	0,8222	1,10	0,1777	1,57	0,0145	2,04	0,0005	2,55	0,0000044
0,64	0,8073	1,11	0,1700	1,58	0,0136	2,05	0,0004	2,60	0,0000026
0,65	0,7920	1,12	0,1626	1,59	0,0127	2,06	0,0004	2,65	0,0000016
0,66	0,7764	1,13	0,1555	1,60	0,0120	2,07	0,0004	2,70	0,0000010
0,67	0,7604	1,14	0,1486	1,61	0,0112	2,08	0,0004	2,75	0,0000006
0,68	0,7442	1,15	0,1420	1,62	0,0105	2,09	0,0003	2,80	0,0000003
0,69	0,7278	1,16	0,1356	1,63	0,0098	2,10	0,0003	2,85	0,00000018
0,70	0,7112	1,17	0,1294	1,64	0,0092	2,11	0,0003	2,90	0,00000010
0,71	0,6945	1,18	0,1235	1,65	0,0086	2,12	0,0002	2,95	0,00000006
0,72	0,6777	1,19	0,1177	1,66	0,0081	2,13	0,0002	3,00	0,00000003
0,73	0,6609	1,20	0,1122	1,67	0,0076	2,14	0,0002		
0,74	0,6440	1,21	0,1070	1,65	0,0071	2,15	0,0002		
0,75	0,6272	1,22	0,1019	1,69	0,0066	2,16	0,0002		

Таблица 10. Оценки параметров важнейших распределений.

Название распределения	Оцениваемые параметры	Точечная оценка	Интервальная оценка
<p>Распределение Пуассона: $x = 0, 1, 2, \dots$ $p(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$ $MX = \lambda, DX = \lambda$</p>	λ	<p>X_i - результаты n независимых наблюдений случайной величины X. $\hat{\lambda} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}, M\hat{\lambda} = \lambda, D\hat{\lambda} = \frac{\lambda}{n}$ <i>несмещённые</i></p>	$\frac{\hat{\lambda} - \lambda}{\sqrt{\hat{\lambda}}} \sim N(0,1) \quad \hat{\lambda} - \frac{u_\alpha \sqrt{\hat{\lambda}}}{\sqrt{n}} < \lambda < \hat{\lambda} + \frac{u_\alpha \sqrt{\hat{\lambda}}}{\sqrt{n}}$ <p>где u_α - квантиль $N(0,1)$, $\alpha = 0,05$.</p>
<p>Биноминальное распределение: $p(X = x) = C_N^x p^x (1-p)^{N-x}$ $x = 0, 1, 2, \dots, N$ $MX = pN, DX = p(1-p)N$</p>	p	$\hat{p} = \frac{\sum_{i=1}^n x_i}{nN} = \frac{\bar{x}}{N}$ $M\hat{p} = p, D\hat{p} = \frac{p(1-p)}{nN}$ <p><i>несмещённые</i></p>	$\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{nN}}} \sim N(0,1)$ $\hat{p} - u_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{nN}} < p < \hat{p} + u_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{nN}} \quad \alpha = 0,05$
<p>Равномерное распределение: для X на $[a, b]$ $f(x) = \frac{1}{b-a}$ $a \leq x \leq b$ $MX = \frac{a+b}{2}$ $DX = \frac{(b-a)^2}{12}$</p>	a, b	$\hat{a} = \min_{1 \leq i \leq n} \{x_i\}, \hat{b} = \max_{1 \leq i \leq n} \{x_i\}$ $M\hat{a} = a + \frac{b-a}{n+1}, M\hat{b} = b - \frac{b-a}{n+1}$ $D\hat{a} = \frac{n(b-a)^2}{(n+1)^2(n+2)}$ $D\hat{b} = \frac{n(b-a)^2}{(n+1)^2(n+2)}$ <p><i>смещённые</i></p>	$\frac{\hat{a} - a - \frac{\hat{b} - \hat{a}}{n+1}}{\sqrt{\frac{n(\hat{b} - \hat{a})^2}{(n+1)^2(n+2)}}} \sim \frac{\hat{b} - b - \frac{\hat{b} - \hat{a}}{n+1}}{\sqrt{\frac{n(\hat{b} - \hat{a})}{(n+1)^2(n+2)}}} \xrightarrow{n \rightarrow \infty} N(0;1)$ <p>*</p>

<p>Показательное (экспоненциальное) распределение:</p> $p(x) = \theta e^{-\theta x} \quad x \geq 0$ $MX = \frac{1}{\theta}, \quad DX = \frac{1}{\theta^2}$	θ	$\hat{\theta} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}} \quad M\hat{\theta} = \frac{1}{\theta} \quad D\hat{\theta} = \frac{n}{\theta^2}$	$\frac{2n\theta}{\hat{\theta}} \sim \chi^2_{(V=2n)}$ $\frac{\chi^2_{v=2n, \frac{1+\gamma}{2}} \cdot \hat{\theta}}{2n} < \theta < \frac{\chi^2_{v=2n, \frac{1-\gamma}{2}} \cdot \hat{\theta}}{2n}, \quad \gamma = 0,95$
<p>Нормальное распределение $N(a, \sigma)$:</p> $\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$ $-\infty < x < \infty$ $MX = a, \quad DX = \sigma^2$	a, σ	$a^*(x_1, x_2, K, x_n) = \frac{1}{n} \cdot \sum_{k=1}^n x_k$ $\sigma^{2*}(x_1, x_2, K, x_n) = \frac{1}{n} \cdot \sum_{k=1}^n (x_k - a)^2$ <p>где a - известно; смещённая</p> $s^{2*}(x_1, x_2, K, x_n) = \frac{1}{n-1} \cdot \sum_{k=1}^n (x_k - \bar{X})^2 \quad \text{где}$ <p>$n > 1$, несмещённая</p>	<p>для a: $\left(\bar{X} - \frac{t_\gamma S}{\sqrt{n}}; \bar{X} + \frac{t_\gamma S}{\sqrt{n}} \right)$, где t_γ - квантиль распределения Стьюдента;</p> <p>для σ^2: $\frac{(n-1)S^2}{\chi^2_{q;n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{p;n-1}}$, где $p = \frac{\alpha}{2}$, $q = 1 - \frac{\alpha}{2}$</p>

* Доверительные интервалы для параметров равномерного распределения:

$$\hat{a} - \frac{\hat{b} - \hat{a}}{n+1} - u_\alpha \sqrt{\frac{n(\hat{b} - \hat{a})^2}{(n+1)^2(n+2)}} < a < \hat{a} - \frac{\hat{b} - \hat{a}}{n+1} + u_\alpha \sqrt{\frac{n(\hat{b} - \hat{a})^2}{(n+1)^2(n+2)}}, \quad \alpha = 0,05$$

$$\hat{b} - \frac{\hat{b} - \hat{a}}{n+1} - u_\alpha \sqrt{\frac{n(\hat{b} - \hat{a})^2}{(n+1)^2(n+2)}} < b < \hat{b} - \frac{\hat{b} - \hat{a}}{n+1} + u_\alpha \sqrt{\frac{n(\hat{b} - \hat{a})^2}{(n+1)^2(n+2)}}, \quad \alpha = 0,05$$

** Для равномерно распределённой на $[0, \theta]$, случайной величины X точечной оценкой параметра θ будет число

$$\hat{\theta} = \frac{n+1}{n} x_n^*, \quad x_n^* = \max_{1 \leq k \leq n} x_k, \quad \text{а доверительный интервал примет вид: } (x_n^*; \frac{x_n^*}{\sqrt[n]{\varepsilon}}).$$

Если случайная величина X равномерно распределёна на $[\theta_1, \theta_2]$, то точечные оценки параметров будут равны:

$$\hat{\theta}_1 = \frac{nx_1^*}{n-1} - \frac{x_n^*}{n-1}; \quad \hat{\theta}_2 = \frac{nx_n^*}{n-1} - \frac{x_1^*}{n-1}; \quad x_1^* = \min_{1 \leq k \leq n} x_k; \quad x_n^* = \max_{1 \leq k \leq n} x_k.$$

Литература:

1. С.А.Айвазян, В.С.Мхитарян, Прикладная статистика в задачах и упражнениях, М., ЮНИТИ, 2001, 270 с.
2. В.А.Колемаев, О.В.Староверов, В.Б.Турундаевский, Теория вероятностей и математическая статистика, М., Высшая школа, 1991, 400 с.
3. Справочник по теории вероятностей и математической статистике, В.С.Королук, Н.И.Портенко, А.В.Скороход, А.Ф.Турбин, М., Наука, 1985, 640 с.
4. Н.А.Микулик, Г.Н.Рейзина, Решение технических задач по теории вероятностей и математической статистике (справочное пособие), Минск, Вышэйшая школа, 1991, 164 с.
5. И.Г.Венецкий, В.И.Венецкая, Основные математико-статистические понятия и формулы в экономическом анализе, м., Статистика, 1979, 447 с.
6. Основы математической статистики (учебное пособие для институтов физической культуры), М., Физкультура и спорт, 1990, ред. В.С.Иванова.
7. В.А.Колемаев, В.Н.Калинина, Теория вероятностей и математическая статистика (учебник), М., ИНФРА-М, 1997, 302 с.
8. Прохоров Ю.В. «Характеризация класса распределений распределением некоторой статистики» - «Теория вероятностей и ее применения», 1965, Т10№3, с 479-487.
9. Н.Ш.Кремер, Теория вероятностей и математическая статистика, М., ЮНИТИ, 2001, 543 с.
10. А.И.Карасёв, З.М.Аксютин, Т.И.Савельева, «Курс высшей математики для экономических вузов», М., Высшая школа, 1982, 317 с.