# НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ

**СОДЕРЖАНИЕ**

1. ВВЕДЕНИЕ 6

1. ОСНОВНЫЕ ПАРАМЕТРЫ И ОПРЕДЕЛЕНИЯ НОРМАЛЬНОГО ЗАКОНА РАСПРЕДЕЛЕНИЯ 8

1.1. Нормальное распределение 8

1.2. Статистическая гипотеза 8

1.3. Ошибки первого и второго рода. Уровень значимости 9

1.4. Степень свободы параметра 10

1.5. Критическая область. Область принятия гипотезы. 10

1.6. Критерий Стьюдента 11

1.7. Критерий Фишера 13

1.8. Критерий Кохрэна 15

1.9. Критерий Пирсона 15

2. ХАРАКТЕРИСТИКА ПАКЕТА EXCELL 19

3. АЛГОРИТМ РЕШЕНИЯ ЗАДАЧИ 21

4. 4. ПРОВЕРКА ГИПОТЕЗЫ О НОРМАЛЬНОМ ЗАКОНЕ РАСПРЕДЕЛЕНИЯ ДАННЫХ В ВЫБОРКЕ 24

4. РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ О НОРМАЛЬНОМ ЗАКОНЕ РАСПРЕДЕЛЕНИЯ ДАННЫХ В ВЫБОРКЕ 26

ЗАКЛЮЧЕНИЕ 28

ЛИТЕРАТУРА 29

# ВВЕДЕНИЕ

Нормальное (гауссовское) распределение занимает центральное место в теории и практике вероятностно-статистических исследований. В качестве непрерывной аппроксимации к биномиальному распределению его впервые рассматривал А.Муавр в 1733 г. Через некоторое время нор­мальное распределение снова открыли и изучили К.Гаусс (1809 г.) и -П.Лаплас, которые пришли к нормальной функции в связи с ра­ботой по теории ошибок наблюдений.

Цельих объяснения механизма формирования нормально распределенных случайных величин заключается в следующем. Постулируется, что зна­чения исследуемой непрерывной случайной величины формируются под воздействием очень большого числа независимых случайных факторов, при­чем сила воздействия каждого отдельного фактора мала и не может прева­лировать среди остальных, а характер воздействия - аддитивный (т.е. при воздействии случайного фактора *F* на величину *а* получается вели­чина *\_\_\_\_\_\_\_\_\_\_\_,* где случайная "добавка" *\_\_\_\_\_\_* мала и равновероятна по знаку).

Во многих случайных величинах, изучаемых в технике и других областях, естественно видеть суммарный аддитивный эффект большого числа независимых причин. Но центральное место нормального закона не следует объяснять его универсальной приложимостью.

В этом смысле нормальный закон - один из многих типов распределения, имеющихся в природе, однако с относительно большим удельным весом практической приложимости.

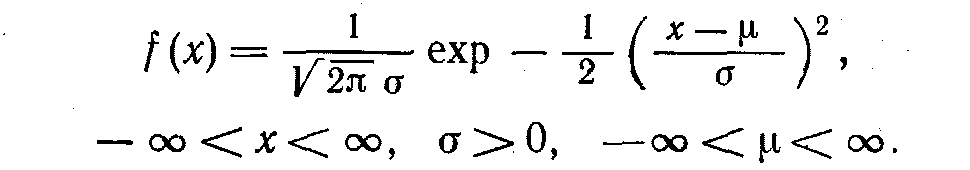
Однако полнота теоретических исследований, относящихся к нормаль­ному закону, а также сравнительно простые математические свойства де­лают его наиболее привлекательным и удобным в применении. Даже в слу­чае отклонения исследуемых экспериментальных данных от нормального закона существует, по крайней мере, два пути его целесообразной эксплуатации: во-первых, использовать нормальный закон в качестве пер­вого приближения (при атом нередко оказывается, что подобное допуще­ние дает достаточно точные с точки зрения конкретных целей исследова­ния результаты); во-вторых. подобрать такое преобразование исследуемой случайной величины, которое видоизменяет исходный "не нормальные" закон распределения, превращая его в нормальный.

Удобно для статистических приложений и свойство "самовоспроизводимости" нормального закона, заключающееся в том, что сумма любого числа нормально распределенных случайных величин тоже подчиняется нормальному закону распределения. Кроме того, с помощью закона нор­мального распределения выведен целый ряд других важных распределений, построены различные статистические критерии

# ОСНОВНЫЕ ПАРАМЕТРЫ И ОПРЕДЕЛЕНИЯ НОРМАЛЬНОГО ЗАКОНА РАСПРЕДЕЛЕНИЯ

## Нормальное распределение

В приложениях статистики чаще всего используется *нормальное (гауссовское) распределение.* Непрерывная случайная величина *Х* называется *распределенной по нормальному закону с параметрами* \_\_\_\_\_\_, если ее плотность распределения есть



.

## Статистическая гипотеза

Часто необходимо знать закон распределения генеральная совокуп­ности. Если он неизвестен, но есть основания предположить, чтоон имеет определенный вид (назовем его А), выдвигают гипотезу: генераль­ная совокупность распределена по закону А. Таким образом, в этой ги­потезе речь вдет о виде предполагаемого распределения.

Возможен случай, когда закон распределения известен, а его параметры неизвестны. Если есть основания предположить, то неизвестный параметр Q равен определенному значению *Q0* , выдвигают гипотезу: *Q* = *Q0.* Таким образом, в этой гипотезе речь идет о предполагаемой величине параметра одного известного распределения.

Возможны и другие гипотезы: о равенстве параметров двух или нескольких распределений, о независимости выборок и многие другие.

**Статистической** называют гипотезу о виде неизвестного распределения или о параметрах известных распределений.

Например статистическими будут гипотезы; генеральная распределена по закону Пуассона, дисперсии двух нормальных совокупностей равны между собой.

В первой гипотезе сделано предположение о виде неизвестного распределения, во второй - о параметрах двух известных распределений.

Наряду с выдвинутой гипотезой рассматривают и противоречивую ей гипотезу. Если выдвинутая гипотеза будет отвергнута, имеет место противоречащая гипотеза. По этой причине эти гипотезы необходимо различать.

Нулевой (основной) называют выдвинутую гипотезу *Н0.*

Конкурирующей (альтернативной) называют гипотезу *Н1*, противоречащую нулевой.

## Ошибки первого и второго рода. Уровень значимости

Выдвинутая гипотеза может быть правильной или неправильной, поэ­тому возникает необходимость проверить ее. Поскольку проверку произво­дят статистическими методами, ее называют статистической. В итоге статистической проверки гипотезы в двух случаях может быть принято неправильное решение, т.е. могут быть допущены ошибки двух родов.

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

Ошибка второго рода состоит в том» что будет принята неправильная гипотеза.

Правильное решение может быть принято также в двух случаях: гипотеза принимается; причем и в действительности она правильная; гипотеза отвергается, причем и в действительности она неверна.

Вероятность совершить ошибку первого рода принято обозначать *q.* Ее называют уровнем значимости. Наиболее часто уровень значимости принимают равным 0,05 или 0,01. Если, например, принят уровень значи­мости, равный 0,05, то это означает, что в пяти случаях из ста мы рис­куем допустить ошибку первого рода (отвергнуть правильную гипотезу).

## Степень свободы параметра

. Степень свободы у какого-либо параметра определяют числом опы­тов, по которым рассчитывают данный параметр, за вычетом количества констант, найденных по этим опытам независимо друг от друга.

## Критическая область. Область принятия гипотезы.

Для проверки нулевой гипотеза используют специально подобранную случайную величину, точное или приближенное распределение которой известно. Ее обозначают t если она распределена по закону Стюдента, *X2* - по закону "хи квадрат", F*-* по закону Фишера, G *-* по закону Кохрэна. Обозначим эту величину *К*

Статистическим критерием (или просто критерием) называется случайная величина К, служащая для проверки нулевой гипотезы.

Для проверки гипотезы по данным выборок вычисляют частные значения входящих в критерий величин и таким образом получают частное (наблюдаемое) значение критерия.

Наблюдаемым значением (К*набл*) называют значение критерия, вычисленное по выборкам. .

После выбора определенного критерия множество всех его возможных значений разбивают на два непересекающихся подмножества; одноиз них содержит значения критерия, при которых нулевая гипотеза отверга­ется, а другое - при которых она принимается.

Критической областью называют совокупность значений критерия, при которых нулевую гипотезу отвергают.

Областью принятия гипотезы (областью допустимых значений)называ­ют совокупность значений критерия, при которых гипотезу принимают.

Основной принцип проверки статистических гипотез можно сформули­ровать так: если наблюдаемое значение критерия принадлежит критичес­кой области - гипотезу отвергают, если наблюдаемое значение критерия принадлежит области принятия гипотезы - гипотезу принимают.

Поскольку критерий *К -* одномерная случайная величина, все ее возможные значения принадлежат некоторому интервалу. Поэтому критическая область и область принятия гипотезы также являются интервалами, и, следовательно, существуют точки, которые их разделяют.

Критическими точками *Ккр* называют точки, отделяющие критичес­кую область от области принятия гипотезы.

Различают, одностороннюю (правостороннюю или левостороннюю) и двустороннюю критические области.

Правосторонней называют критическую область, определяемую нера­венством *К>Ккр* , где К*кр*- положительное число.

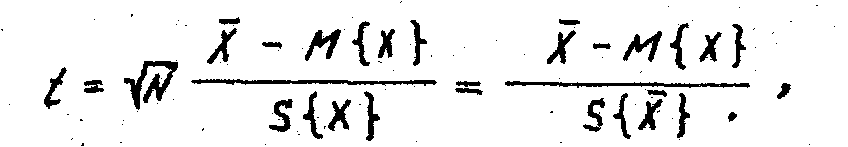
Левосторонней называют критическую область, определяемую нера­венством *К<Ккр* , где К*кр*- отрицательное число.

Односторонней называют правостороннюю или левостороннюю крити­ческую областью.

Двусторонней называют критическую область, определяемую неравенствами K<K1, K>K2, где К2>К1.

## Критерий Стьюдента

t-критерий Стьюдента применяется, когда необходимо сделать статистический вывод, равно ли математическое ожидание M{*Х}* генеральной совокупности некоторому предполагаемому значению *С* или ког­да требуется построить доверительный интервал для M{*Х}*. Обнаруже­но, что случайная величина *t* (при независимых наблюдениях) распреде­лена по закону Стьюдента, если *Х* распределена нормально:



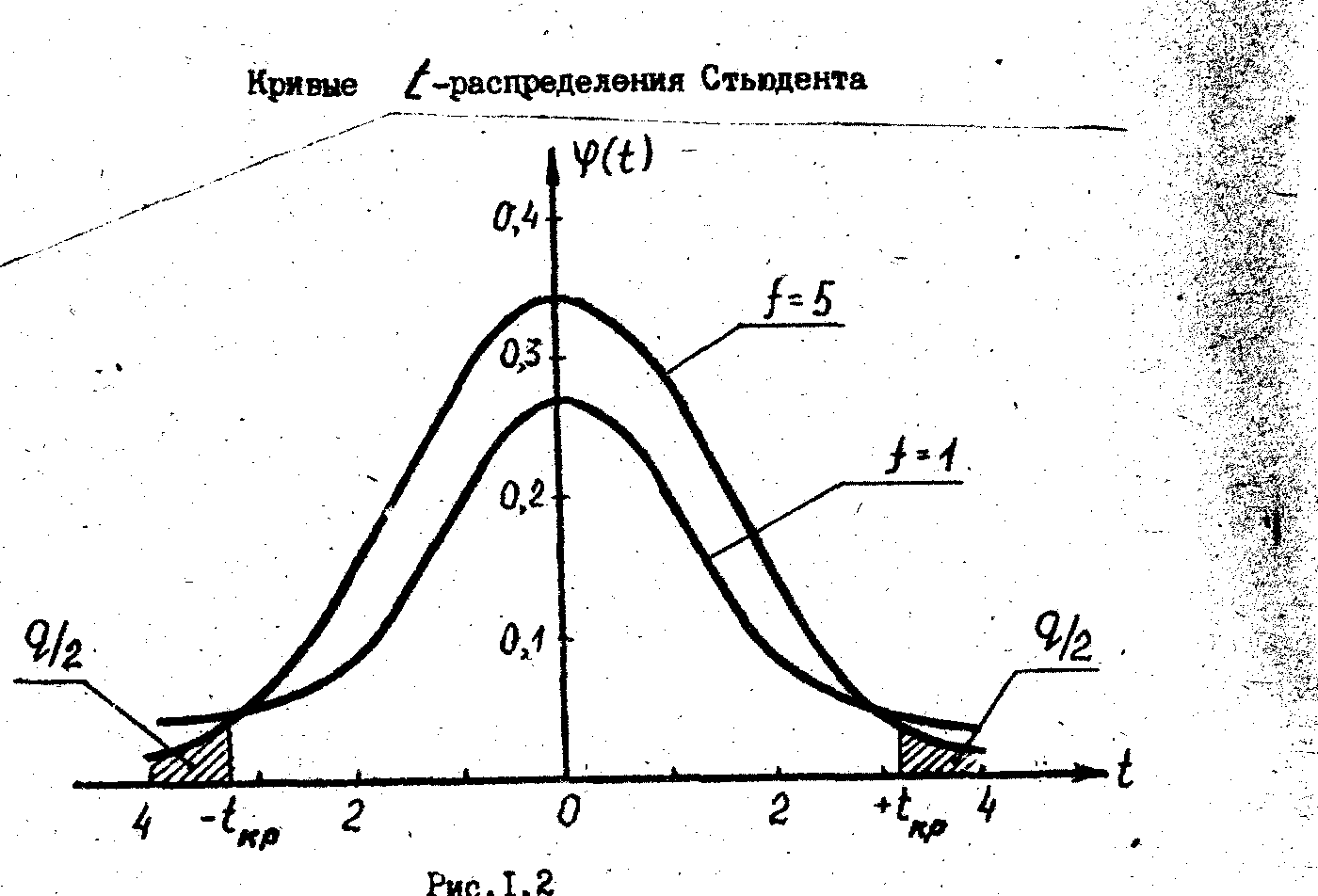
где *N-* общее число наблюдений (объем выборки),

Х - среднее арифметическое случайной переменной Х;

S{Х), S{X}- среднеквадратическое отклонение соответственно единичных значений *Х* и среднего арифметического *Х.*

На рис.1.2 показаны кривые дифференциального закона распределе­ния *Ф(t)* для различных степеней свободы f=N-1 , по которым вычисляют несмещенную оценку дисперсии S2{ Х } . При сравнитель­но небольших N кривая *Ф(t)* более пологая, чем нормальный закон распределения Ф(Х)*.* При N----- кривая *Ф(t)* приближается к кривой нормированного нормального распределения. Из рис.1.2 видно, что t-распределение симметрично относительно t=0, поэтому в таблицах, где даны критические значения t*кр = tq,f* для принятого уровня значимости q и имеющегося чис­ла степеней свободы f , задаются только положительные *tкр .*

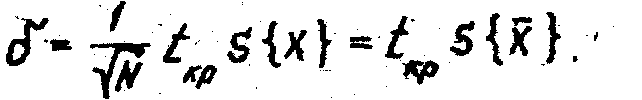
Если при расчете *t* по формуле (1.3) при подстановке в нее вместо М{X} предполагаемого значения С окажется, что t*< tкр* , то можно сделать вывод о том, что гипотеза М{X} = С не проти­воречит результатам наблюдения при принятой уровне значимости *q* .



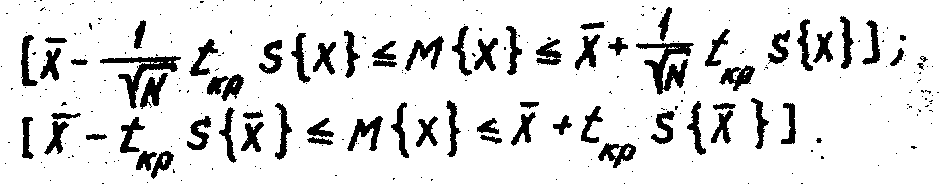
В противном случае эта гипотеза отвергается с тем же уровнем значимости *q.* При этом остается возможность совер­шить ошибку первого рода, т.е. отвергнуть верную гипотезу с вероят­ностью *q* . -

Рассмотрим использование t-критерия Стьюдента для построения доверительного интервала для математического ожидания.

При **t=t***кр* разность [X - M{Х}] в (1.3) равна половине шири­ны доверительного интервала *\_\_*  т.е.



Доверительный интервал, в котором с доверительной вероятностью ***P=I-q*** находится математическое ожидание M{X} , определяется следующими выражениями:



Поскольку мате­матическое ожидание М{X} есть истинное, объективно существующее неслучайное значение, а границы интервала - случайные величины (за счет наличия в них случайных величин *X* и S{X})*,* то правильно будет говорить о том, что доверительный интервал (1.5), (1.6) с ве­роятностью **Р = I - q** накрывает *М {X}.*

## Критерий Фишера

Критерий Фишера применяется при проверке гипотезы о равенстве дисперсий двух генеральных совокупностей, распределенных по нормальному закону.

F-критерий Фишера называют дисперсионным отношением, так как он формируется как отношение двух сравниваемых несмещенных оценок дисперсий:

причем в числителе ставится большая из двух дисперсий. Расчетное F сравнивают с \_\_\_\_\_\_\_\_\_\_\_\_\_, которое находятиз таблиц, для степеней свободы \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_где N*1* - число элементов выборки, по который вычислена **\_\_\_\_\_\_\_ .**

N*2* - число элементов выборки, по которым получена оценка дисперсии \_\_\_\_\_\_\_\_.

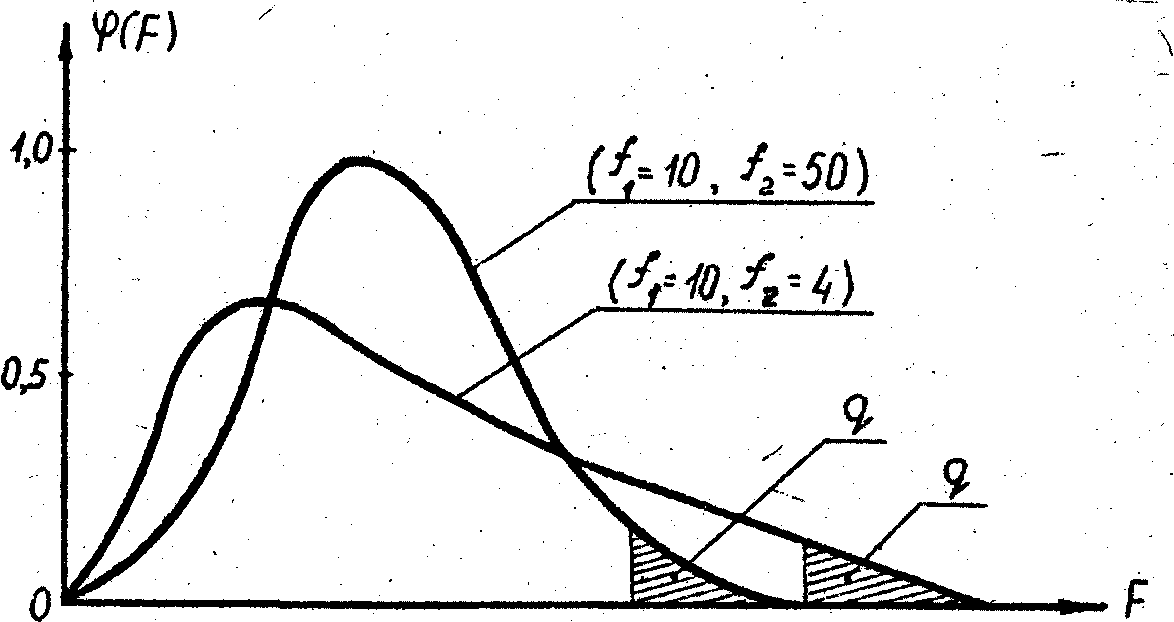
Если *F<Fкр* , то принимается нулевая гипотеза о равенстве генеральных дисперсий \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ при принятом уровне значимости *q.*

На рис. 1.3 показаны кривые распределения *\_\_\_\_\_.* Зачернена об­ласть критических значений F .

На практике задача сравнения дисперсий возникает, если требуется сравнить .точность приборов, инструментовили методов измерений. Предпочтительнее тот прибор, инструмент или метод, который обеспечи­вает наименьшее рассеяние результатов измерений, т.е. наименьшую дис­персию.

. .

**Кривые F-распределения Фишера**



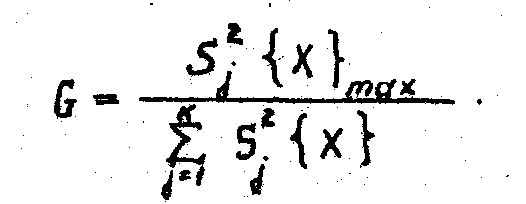
**Рис.1.3**

Если окажется, что нулевая гипотеза справедлива, т.е. генераль­ные дисперсии одинаковы, то различие несмещенных оценок дисперсий незначимо и объясняется случайнымипричинами**,** в частности случайным отбором объектов выборки. Например, если различие несмещенных оценок дисперсий результатов измерений, выполненных двумя приборами, оказа­лось незначимым, то приборы имеют одинаковую точность.

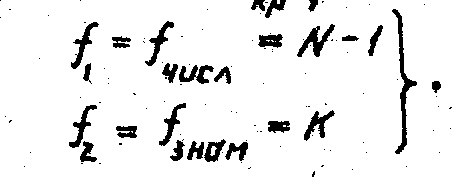
Если нулевая гипотеза будет отвергнута, т.е. генеральные диспер­сии неодинаковы, то различие несмещенных оценок дисперсий значимо и не может быть объяснено случайными причинами, а является следствием того, что сами генеральные дисперсии различны. Например, если разли­чие \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ результатов измерений, произведенных двумя приборами, оказалась значимым, то точность приборов различна.

## Критерий Кохрэна

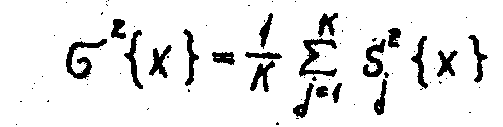
G -критерий Kохрэна применяется для оценки однородности несмещенных оценок дисперсий, вычисленных по одинаковому чис­лу N наблюдений. При этом генеральные совокупности должны быть распределены нормально. Критерий формируется как отношение максимальной из сравниваемых оценок дисперсий к сумме всех K дисперсий;



Если G<G*кр=Gq,f1,f2 ,* то оценки дисперсий признаются однородными или, другими словами, различаются незначимо. В этом слу­чае с уровнем значимости q ммнимается нулевая гипотеза, состоящая в том, что генеральные дисперсии рассматриваемых совокупностей равны между собой: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.Числа степе­ней свободы числителя f1 и знаменателя *f2* определяются условиями

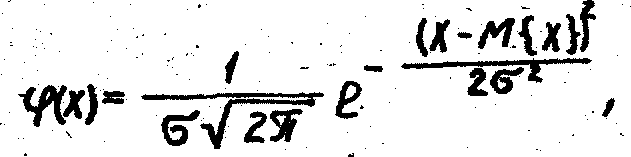


Если требуется оценить генеральную дисперсию, то при условии од­нородности оценок дисперсий целесообразно принять в качестве ее оцен­ки среднее арифметическое несмещенных оценок дисперсий



## Критерий Пирсона

Нормальный закон распределения характеризуется плотностью вероят­ности вида



где **M{X}, \_\_\_\_** — соответственно математическое ожидание и диспер­сия случайной величины. согласованности изучаемого распределения с нормальным

Для проверки гипотезы о соответствии, экспериментального закона распределения случайной величины нормальному применяют критерий Пир­сона или, как его иначе называют, критерий **X2** (хи-квадрат),так как принятие и отклонение гипотезы основаны на **X2** -распределении.

Использование критерия Пирсона основано на сравнении эмпиричес­ких (наблюдаемых) *\_\_\_* и теоретических (вычисленных в предположении нормального распределения) \_\_\_\_\_ частот. Обычно \_\_\_\_ и *\_\_\_\_\_* различны.

Возможно, что расхождение случайно (незначимо) и объясняется малым числом наблюдений, способом их группировки Или другими причина­ми. Возможно, что расхождение частот неслучайно (значимо) и объясня­ется тем, что теоретические частоты вычислены, исходя из неверной ги­потезы о нормальном распределении генеральной совокупности.

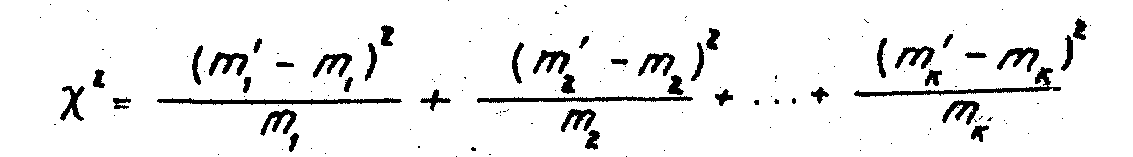
Критерий Пирсона отвечает на поставленный ранее вопрос. Однако, как и любой статистический критерий, он не доказывает справедливость гипотезы, а лишь устанавливает при принятом уровне значимости q ее согласие или несогласие с данными наблюдений.

Пусть по выборке объема \_\_\_ получено эмпирическое распределение.

Допустим, в предположении нормального распределения генеральной совокупности, вычислены теоретические частоты \_\_\_\_\_. При уровне значимости q требуется проверить нулевую гипотезу: генеральная совокупность распределена нормально.

В качестве критерия проверки нулевой гипотезы принимается слу­чайная величина •

или



где *К-* число интервалов (вариант).

Эта величина случайная, так как в различая опытах она принимает различные, заранее неизвестные значения. Чем меньше различаются эмпирические и теоретические частоты, тем меньше значение критерия *(1.9)* и, следовательно, он в известной мере характеризует близость эмпири­ческого и теоретического распределений. Возведением в квадрат разнос­тей частот устраняется возможность взаимного погашения положительных и отрицательных разностей.

При неограниченном возрастании объема выборки ( \_\_\_\_\_\_\_\_\_ ) закон распределения случайной величины (1.9), независимо от того, какому за­кону распределения подчинена генеральная совокупность, стремится к за­кону распределения ***X2***с ***f*** степенями свободы. Поэтому случайная ве­личина (1.9) обозначена ***X2****,* а сам критерий называют критерием сог­ласия "хи квадрат".

Число степеней свободы находят по равенству ***f=K-1-l***где *l-* число параметров предполагаемого распределения, которые оце­нены по данным выборки, а l вызвана тем, что имеется дополнитель­ное ограничение:



т.е.- Теоретическое число элементов совокупности должно быть равно фак­тическому числу элементов.

Поскольку в данном случае, предполагаемое распределение является нормальным, nо оценивают два параметра (математическое ожидание и среднеквадратическое отклонение), поэтому l=2 , и число степеней свободы

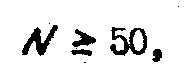


Еслирасчетное (наблюдаемое)значение критерия (1.9).оказалось меньше критического *\_\_\_\_\_* которое находят по таблицам, для соответствующего уровня значимости ***q*** и числа степеней свободы , т.е. если



то нет оснований отвергнуть нулевую гипотезу о нормальности распреде­ления. В противном случае (при \_\_\_\_\_\_\_\_\_\_\_ ) нулевая гипотеза отверга­ется.

При проверке гипотезы о нормальности распределения существует правило, согласно которому общее количество элементов выборки должно быть

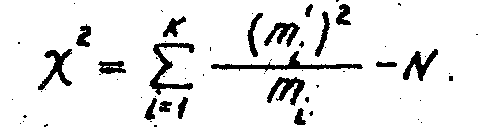


а число элементов, попавших в любой i-и интервал (т.е. значения эмпи­рических частот *\_\_\_\_*),должно быть \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Если в крайние интервалы попадает меньшее число элементов, то они объединяются ссоседними интервалами. Внутренние интервалы объеди­нять запрещается. Общее число интервалов *К ,* оставшихся после объеди­нения, должно удовлетворять условию \_\_\_\_\_\_\_\_\_\_\_\_\_ (1.15)

Иначе число степеней, свободы ***f*** (1.11) окажется равным нулю, и гипо­тезу невозможно будет проверить.

В целях контроля вычислений формулу (1.9) целесообразно преобра­зовать к виду



В табл.1.4 приведен пример расчета наблюдаемого значения крите­рия ***\_\_\_\_*** по известным эмпирическим и теоретическим частотам.

Если ­­­­\_\_\_\_\_\_\_\_\_ , то нет оснований отвергнуть нулевую гипоте­зу. Т.е., расхождение эмпирических и теоретических частот незначимо. Следовательно, данные наблюдений согласуются с гипотезой о нормальном распределении генеральной совокупности.

# ХАРАКТЕРИСТИКА ПАКЕТА EXCELL

Microsoft Office является единственным пакетом, установленным на большинстве компьютеров. Excel — это организатор любого типа данных, будь они числовыми, текстовыми или какими-нибудь еще. Поскольку в этой программе есть много встроенных вычислительных возможностей, большинство людей обращаются к Excel, когда нужно создать таблицы для финансовых расчетов, работать со статистическими данными. С помощью программы можно сделать свои отчеты (например, созданные в Word) более профессиональными и "пробить" дополнительное финансирование с помощью потрясающих деловых презентаций (вроде тех, что создаются в Microsoft PowerPoint). Excel позволяет создавать диаграммы или таблицы для различных финансовых расчетов, хра­нить какие-либо списки или даже сводить данные из различных таблиц.

Excel — это великий хранитель списков (хотя их принято называть в Excel *базами данных)* и создатель таблиц. Поэтому Excel как нельзя лучше подходит для отслеживания информации о продаваемых товарах, об обслуживаемых клиентах, о служащих, которых вы контролируете, и т.д.

Каждая единица информации (например, имя, адрес, число продаж в ме­сяц и др. информация) занимает свою собственную ячей­ку (клетку) в создаваемой рабочей таблице. В каждой рабочей таблице 256 столбцов (из которых в новой рабочей таблице на экране видны, как правило, только первые 10 или 11 (от А до J или К) и 65 536 строк (из которых обычно видны только первые 15-20). Если умножить 256 на 65 536, то получится, что в каждой рабочей таблице 16 777 216 пустых клеток. Каждая новая рабочая книга содержит три чистых листа рабочих таблиц.

Вся помещаемая в электронную таблицу информация хранится в от­дельных клетках рабочей таблицы. Но ввести информацию можно только в текущую клетку. С помощью адреса в строке формул и табличного курсора Excel ука­зывает, какая из 16 миллионов клеток рабочей таблицы является те­кущей. В основе системы адресации клеток рабочей таблицы — так называемой системы А1 — лежит комбинация буквы (или букв) столбца и номера строки.

Excel являет­ся таким замечательным инструментом для выполнения расчетов по формулам, а также для хранения информации в виде списков и таблиц. Это дает возможность намного упростить работу со статистическими данными, которые рассчитываются по сложным формулам. В программе заложены множество групп формул, в том числе и статистических, или пользователь может сам записать формулу.

# АЛГОРИТМ РЕШЕНИЯ ЗАДАЧИ

Ввод и проверка данных

Оценка кол-ва интервалов по формуле:

Подсчет ширины интервалов:

Находим среднее значение и подсчитываем кол-во полученных отрезков

1

K >K

Построение таблицы с занесением данных границ интервалов, кол-ва точек в заданном интервале и частоты:

2

2

Уменьшаем кол-во интервалов до необходимого значения

1

Построение экспериментальной функции распределения

95% ?

Проверка правильности ввода данных

ВЫЧИСЛЯЕМ теоретические частоты, используя критерий Пирсона

Нахождение середин частичных интервалов

2

Вычисляем среднее значение \_\_\_

Вычисляем среднеквадратичное отклонение \_\_\_

Построение таблицы данных

3

**Нормирование границ интервалов**

**3**

**Вычислений теоретических вероятностей**

**Определение наблюдаемого значения \_\_\_ - статистики**

P**i=1**

**Нахождение числа степеней свободы и уровня значимости**

**Вычисление \_\_\_\_**

**2**

Предположение о нормальном законе распределения подтвердилось