Статистическая проверка статистических гипотез.

В этом разделе мы будем обсуждать сравнительные оценки генеральных параметров по разности, наблюдаемой между сравнительными выборками. Сравнивать приходится данные опыта с контролем, урожайность одной культуры с урожайностью другой, продуктивность одной группы животных с продуктивностью другой и т.д.

О преимуществе той или иной из сравниваемых групп судят обычно по разности между средними долями и другими выборочными показателями – величинами случайными, сопровождаемыми ошибками репрезентативности.

- Вопрос о достоверности выборочной разности с её ошибкой приходится решать исходя из той или иной гипотезы, т.е. предположения или допущения относительно параметров сравниваемых групп, которое выражено в терминах вероятности и может быть проверено по выборочным характеристикам.

Часто необходимо знать закон распределения генеральной совокупности. Если закон распределения неизвестен, но имеются основания предположить ,что он имеет определённый вид (назовём его А), выдвигают гипотезу: генеральная совокупность распределена по закону А. Таким образом, в этой гипотезе речь идёт о виде предполагаемого распределения. Есть гипотезы о предполагаемой величине параметра. Есть и другие гипотезы: о равенстве параметров двух или нескольких распределений, о независимости выборок и.т.д.

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

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

Нулевой (основной) называют выдвинутую гипотезу Но

Конкурирующей (альтернативной) называют гипотезу H_1,которая противоречит Но

Различают гипотезы, которые содержат только одно и более одного предложений.

Простой называют гипотезу, содержащую только одно предпо-ложение.

Сложной называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез.

Ошибки первого и второго рода

- Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость её проверки. Поскольку проверку проводят статистическими методами “её” 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.

Замечание1: Правильное решение может быть принято также в двух случаях:

1) гипотеза принимается, причем и в действительности она правильная;
2) гипотеза отвергается, причем и в действительности она неверна.

Замечание2: Вероятность совершить ошибку первого рода принято обозначать через **α**;

Её 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 имеется риск допустить ошибку первого рода (отвергнуть правильную гипотезу)

Статистический критерий проверки нулевой гипотезы. Наблюдаемое значение критерия.

Для проверки нулевой гипотезы используют специально подобранную случайную величину, точное или приближённое распределение которой известно. Эту величину обозначают через **T** или **Z,** если она распределена нормально, F или V- по закону Фишера-Снедекора,**Т** – по закону Стьюдента, **χ2** – по закону хи - квадрат и.т.д.

Статистическим критерием (или просто критерием) называют случайную величину **К**, которая служит 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** $K_{\text{набл}}$ называют значение критерия, вычисленное по выборкам.

Критическая область. Область принятия гипотезы. Критические точки

После выбора определённого критерия множество всех его возможных значений разбивают на два непересекающихся подмножества: одно из них содержит значения критерия, при которых нулевая гипотеза отвергается, а другая - при которых она принимается.

Критической областью называют совокупность значений критерия, при которых нулевая гипотеза отвергается.

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

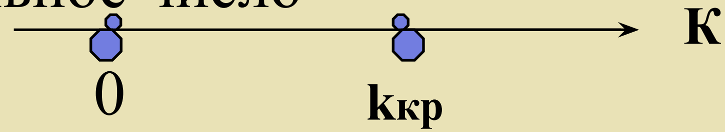
Основной принцип проверки статистических гипотез можно сформулировать так: если наблюдаемое значение критерия принадлежит критической области – гипотезу отвергают, если наблюдаемое значение критерия принадлежит области принятия гипотезы – гипотезу принимают.

Поскольку критерий **К** – одномерная случайная величина, все её значения (возможные) принадлежат некоторому интервалу. Поэтому критическая область и область принятия гипотезы также является интервалами и, следовательно, существуют точки, которые их разделяют.

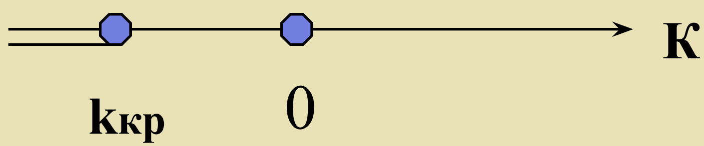
Критическими точками (границами) $К_{кр}$ называют точки, отделяющие критическую область от области принятия гипотезы.

Различают одностороннюю и двустороннюю критические области.

Правосторонней называют критическую область, определяемую неравенством $K > k_{кр}$, где $k_{кр}$ – положительное число



Левосторонней называют критическую область, определяемую
неравенством $K < k_{кр}$, где $k_{кр}$ – отрицательное число.

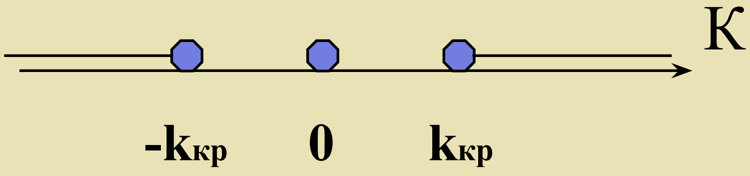


Односторонней называют правостороннюю или левостороннюю критическую область.
Двусторонней называют критическую область, определяемую неравенствами $K < k_1$, $K > k_2$, где $k_2 > k_1$

В частности, если критические точки симметричны относительно
нуля, двусторонняя критическая область определяется неравенствами (в предположении, что **$k_{кр}$>0**):

К< -$k_{кр}$, k>$k_{кр}$, или равносильным неравенством

К > $k_{кр}$.



Для каждого критерия имеются соответствующие таблицы, по которым и находят критическую точку, удовлетворяющую этому требованию.

Замечание 1.

Когда критическая точка уже найдена, вычисляют по данным выборок наблюдаемое значение критерия и, если окажется, что $\mathbf{K_{набл} > k_{кр}}$, то нулевую гипотезу отвергают, если же $\mathbf{K_{набл} < k_{кр}}$, то нет оснований, чтобы отвергнуть нулевую гипотезу.

Замечание 2.

Наблюдаемое значение критерия может оказаться большим $\mathbf{k_{кр}}$ не потому, что нулевая гипотеза ложна, а по другим причинам (малый объём выборки, недостатки методики эксперимента).

В этом случае, отвергнув правильную нулевую гипотезу, совершают ошибку первого рода. Вероятность этой ошибки равна уровню значимости **α**.

Заметим кстати, что в книгах по контролю качества продукции вероятность признать негодной партию годных изделий называют “риском производителя”, а вероятность принять негодную партию – “риском потребителя”.

- <u>**Замечание 3**</u>

Известно, что один пример, подтверждающий справедливость некоторого общего утверждения, ещё не доказывает его. Поэтому более правильно говорить: “данные наблюдений, согласуются с нулевой гипотезой , и, следовательно, не даёт оснований её отвергнуть.”

- На практике для большей уверенности принятия гипотезы её проверяют другими способами или проверяют экспериментом, увеличив объём выборки.

Виды статистических критериев

- В области статистики и биометрии в частности применяют два вида статистических критериев: **параметрические**, построенные на основании параметров данной совокупности (например, ср. арифметического, дисперсии) и представляющие функции этих параметров, и **непараметрические**, представляющие собой функции, зависящие непосредственно от вариант данной совокупности с их частотами.

- При нормальном распределении признака параметрические критерии обладают большей мощностью, чем непараметрические критерии. Они способны более безошибочно отвергать нулевую гипотезу, если она не верна.
- Из параметрических критериев в статистике и биометрии применяют

t-критерий Стьюдента и F-критерий Фишера.

- Первый используют для сравнительной оценки средних величин, второй – для оценки дисперсий.
- ***t-распределение Стьюдента*** зависит только от *числа степеней свободы* k=n-1, причем с увеличением объема выборки t-распределение быстро приближается к нормальному с параметрами a=0 и ср.кв.отклонением равным единице и уже при n >30 не отличается от него.

s_d

Оценка разности средних

- Определение достоверности разности между средними арифметическими или долями двух вариационных рядов можно определить с помощью ошибки разности Sd
- Формула ошибки разности следующая:

$$s_d = \sqrt{s_{\bar{x}_1}^{\,2} + s_{\bar{x}_2}^{\,2}}$$

- где $s_{\bar{x}_1}^{\,2}$- квадрат ошибки средней арифметической 1-го вариационного ряда,
- $s_{\bar{x}_2}^{\,2}$ - квадрат ошибки средней

арифметической 2-го вариационного ряда.

Эта формула используется в тех случаях, когда варианты одной выборки некоррелированы с вариантами другой выборки.

- Критерий достоверности разности между средними арифметическими вычисляют по следующей формуле

$$t_d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2}} \geq t_{st}$$

- Найденный критерий t_d сравнивают с табличным значением критерия Стъюдента при числе степеней свободы $k=n_1+n_2-2$, где n_1 и n_2 — объемы выборок.

- Если $t_d < t_{st}$ - разность считается недостоверной. Это значит, что не получено никакого определённого ответа о разности между соответствующими генеральными параметрами. Если получена благоприятная по смыслу исследования разность между двумя выборочными средними, но эта разность оказалась недостоверной, то это значит, что между соответствующими генеральными средними могут быть любые соотношения, а какие именно неизвестно, но это не может служить доказательством отсутствия разницы между генеральными средними.

- Если в выборочном исследовании оказалась достоверная разница между выборочными показателями, то такая же разница по знаку будет и между соответствующими генеральными параметрами. Основной вывод может быть перенесён на генеральную совокупность.
- Для коррелированных выборок используется следующая формула разности между средними:

$$s_d = \sqrt{s_{\bar{x}_1}^{2} + s_{\bar{x}_2}^{2} - 2r \cdot s_{\bar{x}_1} \cdot s_{\bar{x}_2}}$$

F-критерий Фишера

- Во многих задачах математической статистики, в особенности в дисперсионном анализе, важную роль играет распределение Фишера (F-распределение), названное так по фамилии известного английского математика Р.А. Фишера (1925 г.)
- На практике часто применяется случайная величина

$$F = \frac{S_1^2}{S_2^2}$$

распределенная по закону Стьюдента с числом степеней свободы $k_1 = n_1 - 1$ (для большей дисперсии) и $k_2 = n_2 - 1$ (для меньшей дисперсии).

- Функция F-распределения табулирована для 5%-ного (Р = 0,05) и 1%-ного (Р = 0,01) уровней значимости и чисел степеней свободы $k1$ для большей дисперсии и $k2$ для меньшей. Критические точки для F-критерия содержатся в таблице Приложений. В этой таблице степени свободы для большей дисперсии расположены в верхней строке (по горизонтали), а степени свободы для меньшей дисперсии — в первой графе (по вертикали).
- Если сравниваемые выборки извлечены из одной и той же генеральной совокупности или из разных совокупностей с равными друг другу дисперсиями, то величина F-критерия не превысит критические точки *(Fкр),* указанные в таблице. Если же выборки взяты из разных совокупностей с дисперсиями, не равными друг другу, то $F\text{ф} \geq F\text{кр}$ и нулевая гипотеза должна быть отвергнута.

- *Пример .* Известны данные о влиянии кобальта на массу тела кроликов. Рассчитанные для этих данных дисперсии таковы: в опытной группе s21 = 2596,3, в контроле s22 = 3579,0. Дисперсионное отношение *F*=3579,0/2596,3 = 1,3. В табл. 2 Приложения для 5%-ного уровня значимости (Р=0,05) и чисел степеней свободы *k1 = 9*—1=8 (см. верхнюю строку таблицы) и *k2 = 8*—1 =7 (см. первую графу той же таблицы) находим *Fkp* =3,5. Так как *Fф* < *Fkp,* нулевая гипотеза остается в силе (Р>0,05). Это означает, что генеральные параметры сравниваемых групп σ21 = σ22 и что применение *t*-критерия для проверки Н0-гипотезы в отношении оценки разности между выборочными средними и имеет достаточные основания.
- *F*-критерий можно применить и для оценки разности между долями из неравновеликих выборок, и при оценке разности между коэффициентами вариации.

Критерий согласия Пирсона

- Имеется несколько критериев согласия: («хи-квадрат») К. Пирсона, Колмогорова, Смирнова и др. Ограничимся описанием *применения критерия Пирсона к проверке гипотезы о нормальном распределении генеральной совокупности* (критерий аналогично применяется и для других распределений, в этом состоит его достоинство). С этой целью будем сравнивать эмпирические (наблюдаемые) и теоретические (вычисленные в предположении нормального распределения) частоты.

- Случайно ли расхождение частот? Возможно, что расхождение случайно (незначимо) и объясняется либо малым числом наблюдений, либо способом их группировки, либо 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 выше вопрос. Правда, как и любой критерий, он не доказывает справедливость гипотезы, а лишь устанавливает на принятом уровне значимости её согласие или несогласие с данными наблюдений.

x_i n_i' x_1 x_2 x_s

- Итак, пусть по выборке объёма *n* получено эмпирическое распределение:
- варианты……

эмп. частоты...

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

- В качестве критерия проверки нулевой гипотезы примем случайную величину

$$\chi^2 = \frac{\sum (n_i - n_i')^2}{n_i'}$$

$k = s - 1 - r$

- Число степеней свободы находят по равенству k=s-1-r, где *s* – число групп (частичных интервалов) выборки; r - число параметров предполагаемого распределения, которые оценены по данным выборки.
- Если $\chi^2_{набл} < \chi^2_{кр}$ - нет оснований отвергнуть нулевую гипотезу.
- Если $\chi^2_{набл} > \chi^2_{кр}$ - нулевая гипотеза отвергается.