

Статистическая проверка статистических гипотез

Статистические гипотезы. Виды ошибок при выдвижении статистических гипотез

Часто необходимо знать закон распределения генеральной совокупности.

Если имеются основания предположить, что он имеет определенный вид, то выдвигают гипотезу: генеральная совокупность распределена по данному закону. Таким образом, в этой гипотезе идет речь о виде предполагаемого распределения.

Другой случай, – когда закон распределения известен, но неизвестны его параметры (среднее, дисперсия). Если есть основания предполагать, что неизвестный параметр Θ равен определенному значению Θ_0 , выдвигают гипотезу $\Theta = \Theta_0$. Таким образом, в этой гипотезе идет речь о предполагаемой величине параметра известного распределения.

Приведенные примеры представляют собой одни из многочисленных вариантов статистических гипотез.

Таким образом, статистической гипотезой называют гипотезу о виде неизвестного распределения или о параметрах известных распределений.

Наряду с первоначально выдвинутой гипотезой рассматривают и противоречащую ей. Если выдвинутая гипотеза будет отвергнута, ее место занимает противоречащая.

Нулевой (основной) гипотезой называют первоначально выдвинутую гипотезу. Гипотезу, противоречащую нулевой, называют конкурирующей (альтернативной) гипотезой.

Условно нулевую гипотезу обозначают H_0 , а альтернативную – H_1 .

Приведу примеры обозначений статистических гипотез и варианты их прочтения:

H_0 : $x = 15$ - основная гипотеза состоит в том, что среднее значение случайной величины X статистически неразлично с 15;

H_1 : $x > 15$ - альтернативная гипотеза состоит в том, что среднее значение случайной величины X статистически различимо и больше 15.

ОБРАТИТЕ ВНИМАНИЕ: о равенстве показателей речи не идет. Корректно говорить «статистически неразлично» или «статистически различимо»

Когда выдвигается гипотеза, всегда существует вероятность, что она может быть правильной или неправильной, поэтому возникает необходимость ее проверки.

Поскольку проверку производят статистическими методами, ее называют статистической проверкой.

При выдвижении гипотезы, независимо от того, статистическая она или нет, автор гипотезы берет на себя определенную ответственность.

Ведь выдвинутая гипотеза, равно как и результаты ее проверки, могут быть ошибочными. Риск, который возникает при выдвижении статистической гипотезы, так и называют ошибкой, причем существуют ошибки I и II рода.

Ошибка I рода состоит в том, что будет отвергнута гипотеза, в то время как она верна. Ошибка I рода оценивается уровнем значимости α .

Ошибка II рода состоит в том, что будет принята гипотеза, в то время как она неверна. Ошибка II рода оценивается мощностью критерия π

При этом последствия таких ошибок могут оказаться весьма различными.

Можно привести примеры, когда ошибка I рода влечет за собой более весомые последствия, чем ошибка II рода, и наоборот.

Пример 1. Идет прием у врача. Исследуя симптомы болезни, врач назначает лечение. Помимо лекарств, назначаемых при данных симптомах, врач выписывает некоторые анализы для подтверждения своего диагноза. При этом возможны следующие варианты:

Ошибка I рода. Назначение данных лекарств было правомерно, т.к. первоначальный диагноз оказался верным, что и подтвердили дополнительные анализы, но врач подверг первоначальный диагноз сомнению, т.е. фактически отверг его.

Ошибка II рода. Назначение данных лекарств недопустимо, т.к. первоначальный диагноз оказался неверным, что и показали дополнительные анализы, но врач назначил их в соответствии с первоначальным диагнозом, который он фактически принял.

Понятно, что в данном примере ошибка II рода приведет к более тяжким последствиям, чем ошибка I рода.

Пример 2. Стоит вопрос о замене строительных материалов, предусмотренных проектом, на другие, поскольку они более доступны и дешевы.

Для этого проводится соответствующая экспертиза. При этом возможны следующие ошибки:

Ошибка I рода. Применение предлагаемых в качестве альтернативы строительных материалов невозможно, но эксперт разрешает их использование, т.к. считает их технические характеристики соответствующими нормам.

Ошибка II рода. Применение альтернативных материалов возможно, но эксперт запрещает замену.

В данном случае ошибка II рода менее тяжела, чем ошибка I рода.

Когда экспериментатор выдвигает ту или иную статистическую гипотезу, он предполагает, что может совершить ошибку.

Решение, принимаемое экспериментатором должно иметь альтернативу, т.е. экспериментатор помимо выдвижения гипотезы должен держать наготове ответ на вопрос: «А что, если Вы ошиблись?»

Про такую ситуацию говорят, что экспериментатор закладывает в гипотезу ошибку того или иного рода. Ошибку какого рода заложить в свою гипотезу экспериментатор решает в зависимости от тяжести последствий при совершении ошибки.

Чтобы было более-менее понятно, о чем идет речь, проиллюстрируем данные соображения на примере приема у врача, описанного выше.

Фактически, назначая дополнительные анализы для подтверждения диагноза, врач закладывает в свою гипотезу ошибку первого рода, т.е. первоначальный диагноз может оказаться верным, но врач не верит без дополнительной проверки.

Кстати, из этих же соображений врач первоначально, пока не выяснит окончательно диагноз, назначает лекарства лишь облегчающие симптоматику, но не решающие все кардинально. Согласитесь, в этом примере ошибка первого рода несет наименьшие последствия, и врач поступает правильно.

Статистические критерии

Когда любой из нас проводит проверку чего-либо, принимает какое-либо решение, всем бывает необходим критерий соответствия полученного результата ожиданиям, тем или иным требованиям и т.д.

Например, при покупке дивана человек оценивает его на соответствие многим критериям: габариты, цвет, форма...

Точно также обстоит дело и в статистике. Только в данном случае необходимы критерии для проверки соответствия выдвинутой статистической гипотезы реальному положению дел. И критерии, соответственно, должны быть статистические.

Статистическим критерием (или просто критерием, критерием согласия) называют критерий проверки гипотезы о предполагаемом законе распределения случайной величины или значениях параметров распределений случайной величины.

При этом значение критерия, вычисленное по экспериментальным данным, называют наблюдаемым значением критерия $K_{\text{набл}}$.

Статистические критерии работают на всем множестве значений числовой прямой в пределах $(-\infty; +\infty)$. При этом вся эта числовая прямая делится на два типа подобластей: критическую и область принятия гипотезы (решения).

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

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

Критическими точками (границами) $k_{\text{кр}}$ называют точки, отделяющие критическую область от области принятия решений.

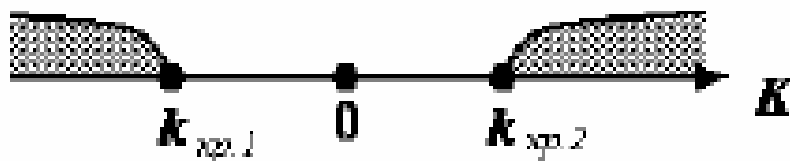
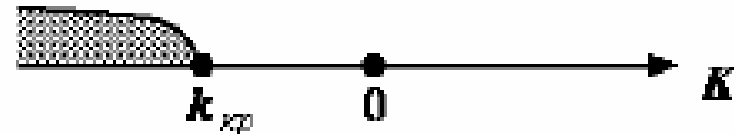
различают одностороннюю и двустороннюю критические области.

Первая, в свою очередь, делится на правостороннюю и левостороннюю.

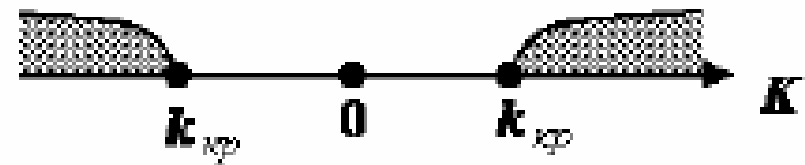
Правосторонней критической областью называют критическую область, определяемую неравенством $K_{\text{набл}} > k_{\text{кр}}$, где $k_{\text{кр}}$ – положительное число (см. рисунок а).

Левосторонней критической областью называют критическую область, определяемую неравенством $K_{\text{набл}} < k_{\text{кр}}$, где $k_{\text{кр}}$ – отрицательное число (см. рисунок б).

Двусторонней критической областью называют критическую область, определяемую неравенствами $K_{\text{набл}} < k_{\text{кр.1}}$ и $K_{\text{набл}} > k_{\text{кр.2}}$, где $k_{\text{кр.2}} > k_{\text{кр.1}}$



в)



г)

С логической и понятийной точки зрения все достаточно просто.

Но с практической позиции сразу же возникает вполне естественный вопрос, как отыскать критическую точку?

Для ее отыскания задаются достаточно малой вероятностью – уровнем значимости α .

Уровнем значимости α называют вероятность, при которой событие (в данной определенной задаче) практически невозможно, т.е. это вероятность того, что исследуемое событие при данных условиях не произойдет.

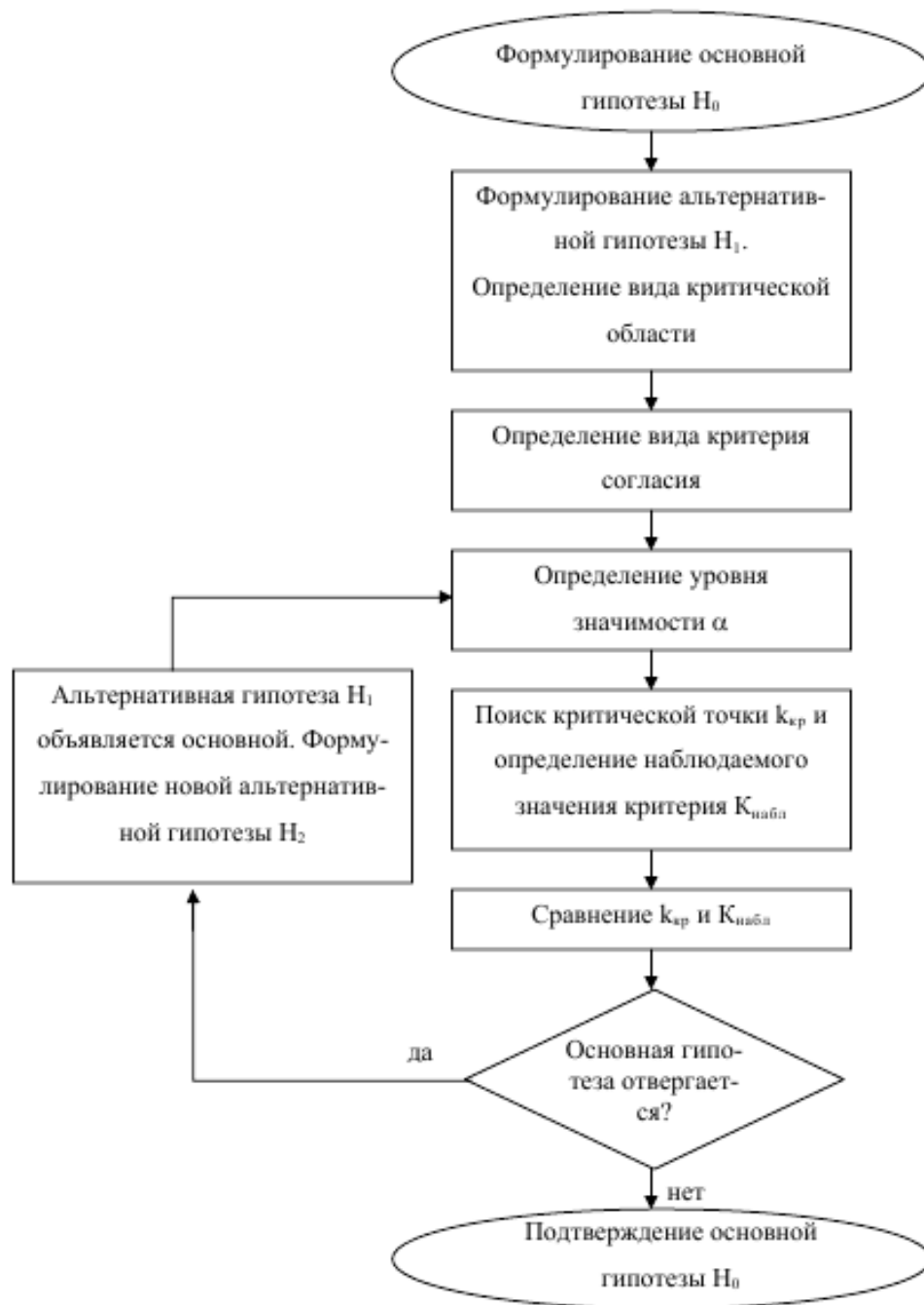
С точки зрения проверки статистических гипотез, уровень значимости – вероятность того, что наблюдаемое значение критерия попадет в критическую область:

$$P(K_{\text{набл}} > k_{\text{кр}}) = \alpha.$$

Вероятность того, что наблюдаемое значение критерия попадет в область допустимых значений называют доверительной вероятностью (надежностью)

$$P = 1 - \alpha.$$

С общих позиций, надежностью называют вероятность того, что имеет место описываемое событие.



Фактически, экспериментатор сам определяет ту степень вероятности, с которой данное событие, а в нашем случае – это выдвинутая гипотеза, не произойдет, т.е., попросту говоря, какова вероятность того, что экспериментатор ошибся, выдвинув свою гипотезу.

Задав уровень значимости, экспериментатор получает возможность найти критическую точку.

Дело в том, что все статистические критерии (или критерии согласия) основываются на различных известных в статистике распределениях: распределении Пирсона, Фишера, Стьюдента и т.д.

Для всех этих распределений уже давно рассчитаны так называемые критические значения, которые представляют собой квантили упомянутых распределений.

Здесь необходимо сделать одно небольшое замечание.

В случае односторонних областей выбор критической точки определяется требованием

$P(K_{\text{набл}} > k_{\text{кр}}) = \alpha$ – при правостороннем критерии
или

$P(K_{\text{набл}} < k_{\text{кр}}) = \alpha$ – при левостороннем критерии.

Однако, в случае двусторонней критической области данное условие примет вид

$$P(K_{\text{набл}} < k_{\text{кр.1}}) + P(K_{\text{набл}} > k_{\text{кр.2}}) = \alpha.$$

Ясно, что критические точки в этой ситуации могут быть выбраны бесчисленным множеством способов. Однако, как правило, критические точки стараются выбрать симметричными относительно нуля. Тогда

$$P(K_{\text{набл}} > k_{\text{кр}}) = P(K_{\text{набл}} < -k_{\text{кр}}),$$

и критерий примет вид

$$P(K_{\text{набл}} > k_{\text{кр}}) = \alpha / 2.$$

Хорошо, – скажете Вы, – с этим понятно. (Хотя на самом деле ничего не понятно). А как определиться с видом критической области: двусторонняя, левосторонняя или правосторонняя?

На самом деле здесь все еще проще. Вид критической области зависит от вида альтернативной гипотезы. Для простоты представим пример выбора критической области в виде стилизованной таблицы

<i>Основная гипотеза:</i>	
$H_0: \bar{x} = a$	
<i>Альтернативная гипотеза:</i>	
<i>Вид гипотезы</i>	<i>Вид критической области</i>
$H_1: \bar{x} \neq a$	Двусторонняя (критические точки ищем при уровне значимости $\alpha/2$)
$H_1: \bar{x} < a$	Левосторонняя (критические точки ищем при уровне значимости α)
$H_1: \bar{x} > a$	Правосторонняя (критические точки ищем при уровне значимости α)

Выше уже говорилось, что при статистической проверке статистических гипотез помимо основной принимается и альтернативная ей гипотеза.

Вследствие этого целесообразно ввести в рассмотрение вероятность попадания критерия в критическую область при условии, что верна альтернативная гипотеза.

Мощностью критерия π называют вероятность попадания критерия в критическую область при условии, что справедлива конкурирующая гипотеза.

При этом, если вероятность совершения ошибки II рода равна β , то мощность критерия определяется как

$$\pi = 1 - \beta.$$

Если уровень значимости уже выбран, то критическую область следует строить так, чтобы мощность критерия была максимальной.

Фактически, мощность критерия – вероятность того, что ошибка второго рода не будет допущена.

При этом одновременно уменьшить α и β невозможно. При уменьшении одной величины, вторая неизбежно будет возрастать.

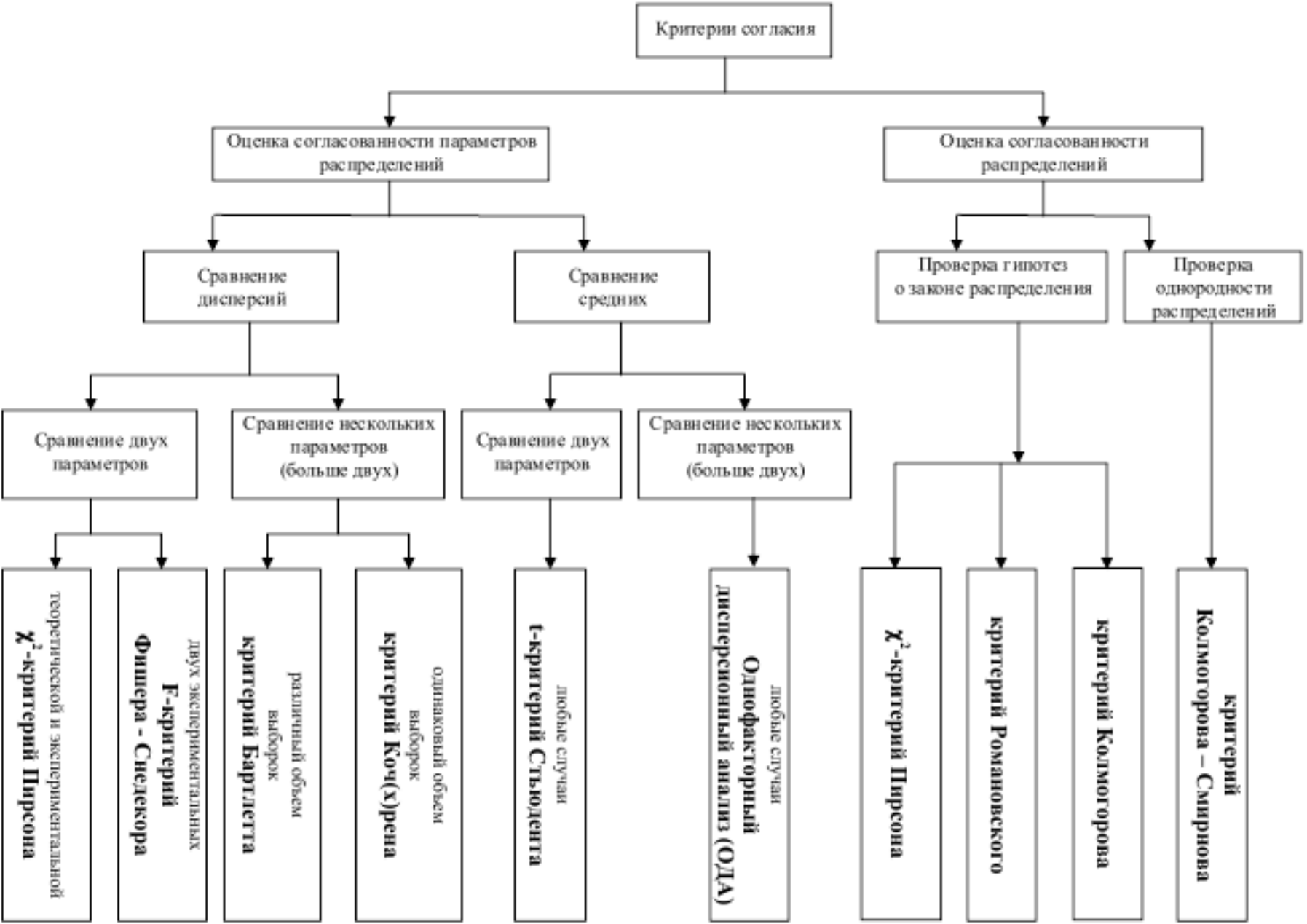
Поскольку при проверке статистических гипотез выбирается уровень значимости α , относительно него и решается вопрос о выборе значения

Величина β автоматически будет уменьшаться или возрастать при увеличении или уменьшении α .

Вопрос о выборе величины уровня значимости будет напрямую зависеть от тяжести последствий, вызываемых ошибками I и II рода.

Если ошибка I рода влечет за собой более тяжелые последствия, то величину α выбирают как можно меньше.

Виды критериев согласия и области их применения

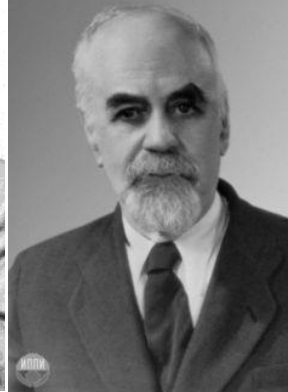
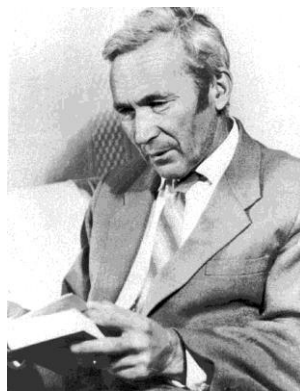


критерий
Колмогорова – Смирнова

критерий Колмогорова

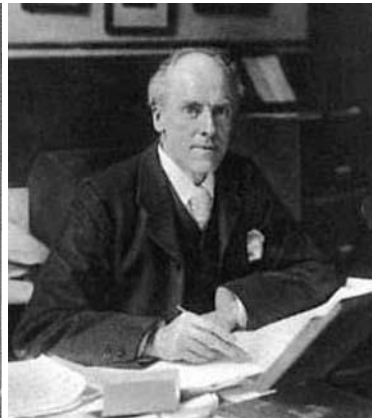
критерий Романовского

χ^2 -критерий Пирсона



любые случаи
Однофакторный
дисперсионный анализ (ОДА)

любые случаи
t-критерий Стьюдента



одинаковый объем
выборки
критерий Коч(х)рена

различный объем
выборки
критерий Бартлетта

двух экспериментальных
F-критерий
Фишера - Снедекора



теоретической и экспериментальной
 χ^2 -критерий Пирсона

Критерии согласия носят названия по имени тех ученых-статистиков, которые их и сформулировали.

Исключение из общей картины на рисунке составляет только один инструмент – однофакторный дисперсионный анализ (ОДА). Данный инструмент НЕ является критерием согласия.

Однако чтобы классификатор инструментов сравнения был полон, ОДА был добавлен к критериям согласия.

Дополнительно замечу, что сам ОДА будет рассмотрен в дальнейшем.

Все критерии согласия рассчитаны на то, что генеральные совокупности рассматриваемых в критериях случайных величин подчиняются нормальному закону.

В противном случае результаты могут быть и неправильными. Кроме того, в критериях согласия рассматриваются так называемые исправленные оценки исследуемых параметров (среднего, дисперсии).

Принцип «работы» всех критериев согласия одинаков: по определенному правилу-алгоритму находим наблюдаемое значение критерия $K_{\text{набл}}$, сравниваем его с критическим значением $k_{\text{кр}}$ распределения, задействованного в данной критерии, и выносим суждение о подтверждении или отвержении основной гипотезы.

Различие состоит лишь в алгоритмах поиска $K_{\text{набл}}$ и привлечении разных распределений для поиска $k_{\text{кр}}$.

Условие подтверждения / отвержения основной гипотезы будем демонстрировать на примере двусторонней критической области, за исключением первого случая.

На примере первого критерия согласия покажем как выглядят условия подтверждения основной гипотезы для всех трех типов критических областей.

Поскольку во всех случаях ситуация будет одна и та же, повторяться, думаем, не имеет смысла.

χ^2 -критерий согласия Пирсона

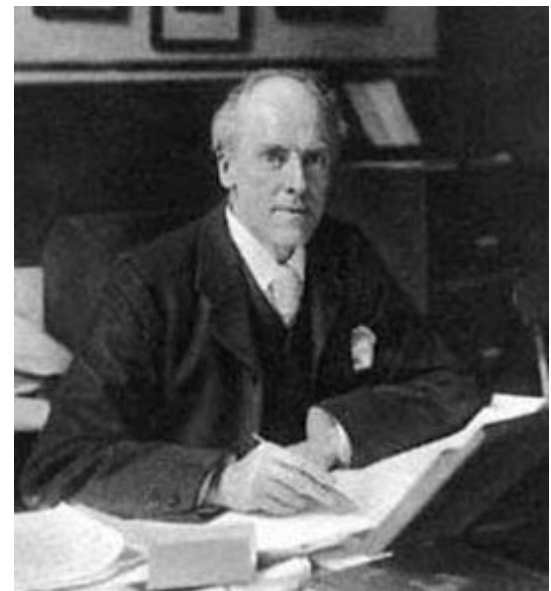
Критерий согласия Пирсона применяется для сравнения теоретического и экспериментального значений дисперсий.

В качестве теоретического значения дисперсии на практике используются значения, регламентированные какими-либо нормативными документами: ГОСТами, ТУ, техническим паспортом и т.п.

Обозначим s^2 – экспериментально полученное значение дисперсии по выборке объема n , σ^2 – теоретическое значение дисперсии.

Основная гипотеза состоит в том, что данные значения дисперсий статистически неразличимы; в краткой записи наше предположение выглядит как

$$H_0: s^2 = \sigma^2.$$



При этом альтернативная гипотеза состоит в том, что

- 1) $H_1: s^2 \neq \sigma^2$ – экспериментальное и теоретическое значения дисперсий статистически различимы – двусторонняя критическая область;
- 2) $H_1: s^2 < \sigma^2$ – теоретическое значения дисперсии превышает экспериментальное – левосторонняя критическая область;
- 3) $H_1: s^2 > \sigma^2$ – теоретическое значения дисперсии меньше экспериментального – правосторонняя критическая область.

Наблюдаемое значение χ^2 -критерия согласия Пирсона определяется по формуле:

$$K_{\chi^2} = \frac{(n - 1) \cdot s^2}{\sigma^2}.$$

Критическая точка определяется как критическое значение χ^2 -распределения Пирсона при заданном уровне значимости α (для двусторонней критической области – $\alpha/2$) с числом степеней свободы $(n - 1)$. Все сказанное укладывается в следующее обозначение:

$$\chi_{\alpha}^2(n - 1).$$

Основная гипотеза подтверждается, если:

- 1) двусторонняя критическая область $K: \chi^2 < \chi^2_{\alpha/2}(n-1)$;
- 2) левосторонняя критическая область $K: \chi^2 > \chi^2_{\alpha}(n-1)$;
- 3) правосторонняя критическая область $K: \chi^2 < \chi^2_{\alpha}(n-1)$.

F-критерий согласия Фишера – Снедекора

Данный критерий согласия применяется для сравнения двух экспериментальных значений дисперсий.

Обозначим:

s^2_1 – экспериментальное значение дисперсии, полученное по выборке объема n_1 в первой серии опытов;

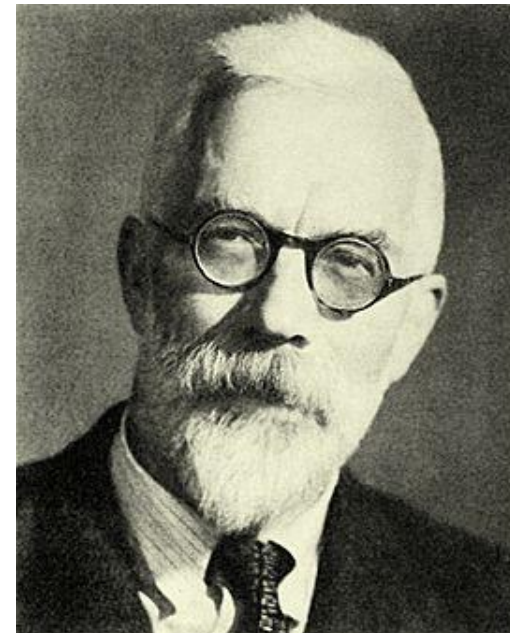
s^2_2 – экспериментальное значение дисперсии, полученное по выборке объема n_2 во второй серии опытов.

Причем, $s^2_1 > s^2_2$.

Основная и альтернативная гипотезы имеют вид:

$$H_0: s^2_1 = s^2_2$$

$$H_1: s^2_1 \neq s^2_2.$$



Наблюдаемое значение F-критерия согласия Фишера определяется по формуле:

$$K_F = \frac{s_1^2}{s_2^2}.$$

Критическая точка определяется как критическое значение F-распределения Фишера при заданном уровне значимости (или $\alpha/2$) с числами степеней свободы $(n_1 - 1; n_2 - 1)$:

$$F_\alpha (n_1 - 1; n_2 - 1).$$

При определении критического значения следует помнить, что первым в скобках стоит значение числа степеней свободы для той дисперсии, которая находится в числителе формулы наблюдаемого значения критерия.

Критерий согласия Бартлетта



Заключается в сравнении нескольких дисперсий (больше двух) по выборкам различного объема.

Главное условие применения критерия согласия Бартлетта – объем выборок должен быть не менее 4 испытаний.

Обозначим:

s_1^2 – экспериментальное значение дисперсии, полученное по выборке объема n_1 в первой серии опытов;

s_2^2 – экспериментальное значение дисперсии, полученное по выборке объема n_2 во второй серии опытов;

... s_i^2 – экспериментальное значение дисперсии, полученное по выборке объема n_i в i -серии опытов.

При этом, некоторые объемы могут быть одинаковыми; если же все выборки имеют одинаковый объем, то предпочтительнее пользоваться критерием Коч(х)рена, описанном ниже.

Основная гипотеза имеет вид:

$$H_0: s_1^2 = s_2^2 = \dots = s_i^2.$$

Следует понимать, что формулировка альтернативной гипотезы в виде математического соотношения достаточно проблематична, т.к. отдельные значения дисперсий могут и совпадать между собой.

Однако основная гипотеза состоит в статистической неразличимости ВСЕХ значений дисперсий, и проверка будет состоять в оценке выполнимости именно этого требования.

Соответственно, альтернативная гипотеза будет состоять в том, что основная гипотеза не выполняется. Если же вдруг встанет вопрос о попарном сравнении, то лучше воспользоваться критерием Фишера – Снедекора.

Обозначим через \bar{s}^2 среднюю арифметическую экспериментальных дисперсий, взвешенную по числам степеней свободы:

$$\bar{s}^2 = \frac{\sum_{m=1}^i k_m s_m^2}{\sum_{m=1}^i k_m}$$

где $k_i = (n_i - 1)$ – число степеней свободы i -серии опытов.

В качестве критерия проверки основной гипотезы о статистической неразличимости (или однородности) дисперсий, т.е. наблюдаемого значения критерия, принимается случайная величина

$$B = V/C,$$