Прикладная статистика. Критерии согласия и однородности.

Леонид Иосипой

Программа «Математика для анализа данных»
Центр непрерывного образования, ВШЭ

8 декабря 2020

- Повторение
- Критерии согласия
- Критерии однородности

Повторение

В проверке гипотез делается предположение о процессе, генерирующем данные, и задача состоит в том, чтобы определить, содержат ли данные достаточно информации, чтобы отвергнуть это предположение или нет.

Чтобы иметь возможность отвергнуть предположение, необходимо зафиксировать альтернативу — другое предположение о данных, относительно которого мы будем решать, отвергать основную гипотезу или нет.

Повторение

Правило, позволяющее принять или отвергнуть гипотезу H_0 на основе данных называется статистическим критерием.

Обычно критерий задается при помощи статистики критерия $T(x_1, \ldots, x_n)$ такой, что для нее типично принимать умеренные значения в случае, когда гипотеза H_0 верна, и большие (иногда малые) значения, когда H_0 не выполняется.

Повторение

Статистика критерия T должна обладать важным свойством:

- при верной H_0 статистика T должна иметь известное нам распределение G_0;
- при неверной H_0 должна иметь какое-либо распределение отличное от G_0.

Повторение

Если значение T попало в область, имеющую при выполнении гипотезы H_0 малую вероятность, то можно заключить, что данные противоречат гипотезе H_0.

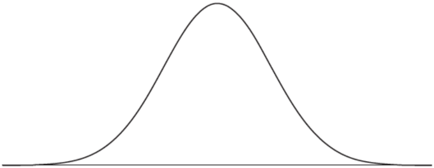
Если произошло обратное, то есть значение T попало в область, имеющую при выполнении гипотезы H_0 большую вероятность, то можно заключить, что данные не противоречат гипотезе H_0.

Вероятности можно посчитать, так как нам известно распределение G_0.

Повторение

Формализация задачи: в случае простых гипотез

выборка:	$\mathbf{X} = (X_1, \ldots, X_n)$
нулевая гипотеза:	$H_0: \ X_i \sim F_0$
альтернатива:	$H_1: \ X_i \sim F_1 \neq F_0$
статистика:	$T(x_1, \ldots, x_n), \ \ T(\mathbf{X}) \sim G_0$ при $\mathbf{X} \sim F_0$
	$T(\mathbf{X}) \nsim G_0$ при $\mathbf{X} \sim F_1$



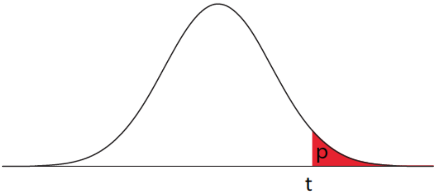
Повторение

реализация выборки: $\mathbf{x} = (x_1, \ldots, x_n)$

реализация статистики: $t = T(\mathbf{x})$

достигаемый уровень значимости или p-value: $p(\mathbf{x}) = \mathbb{P}(T(\mathbf{X}) \geq t \mid H_0)$

(если для T экстремальные значения — большие)



Повторение

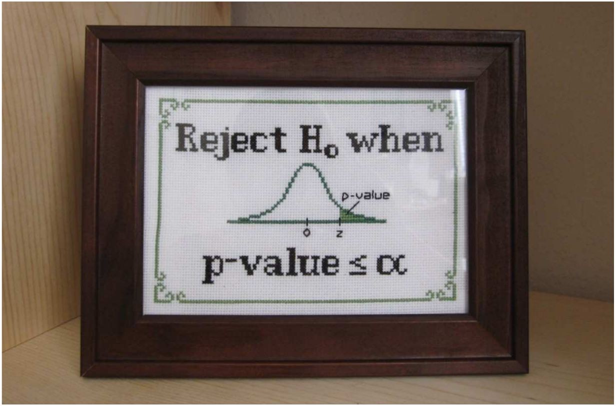
Достигаемый уровень значимости или p-value — это вероятность для статистики T при верной H_0 получить значение t или ещё более экстремальное.

Если для для статистики T экстремальными значениями являются большие значения, то это можно записать так:

$$p(\mathbf{x}) = \mathbb{P}(T(\mathbf{X}) \geq t \mid H_0).$$

Гипотеза отвергается при $p(\mathbf{x}) \leq \alpha$, α — уровень значимости, который мы задаем.

Повторение



Повторение

	H_0 верна	H_0 неверна
H_0 принимается	H_0 верно принята	Ошибка второго рода (False negative)
H_0 отвергается	Ошибка первого рода (False positive)	H_0 верно отвергнута

Повторение

Если величина p-value достаточно мала, то данные свидетельствуют против нулевой гипотезы в пользу альтернативы.

Если величина p-value недостаточно мала, то данные не свидетельствуют против нулевой гипотезы в пользу альтернативы.

При помощи инструмента проверки гипотез нельзя доказать справедливость нулевой гипотезы!

Повторение

Вероятность отвергнуть нулевую гипотезу зависит не только от того, насколько она отличается от истины, но и от размера выборки.

По мере увеличения n нулевая гипотеза может сначала приниматься, но потом выявятся более тонкие несоответствия выборки гипотезе H_0, и она будет отвергнута.

Критерии согласия

Пусть у нас есть выборка $X_1, \dots, X_n \sim F$, где F — некоторое неизвестное распределение.

Начнем изучение критериев с критериев согласия, в которых в качестве H_0 будем рассматривать $F \in \mathcal{F}_\theta$, то есть принадлежность F какому-то параметрическому семейству.

Альтернативой H_1 мы будем считать принадлежность F всем остальным распределениям.

Критерии согласия

Критерии согласия так называются, потому что они отвечают на вопрос, согласуется ли наша выборка с каким-то параметрическим семейством или нет.

В англоязычной среде такие тесты называют Goodness of Fit.

Критерии согласия

Для построения критерия согласия достаточно найти некоторое свойство, которые бы выполнялось для всех распределений из нашего класса и на его основе придумать статистику.

При этом сколько-то удовлетворительно мажорировать вероятность ошибки второго рода не удается, поскольку вне нашего параметрического семейства есть сколь угодно похожие на наши распределения.

Но по крайней мере, можно искать критерий, от которого мы ожидаем, что при альтернативе он чаще отвергает нулевую гипотезу.

Критерии согласия

Рассмотрим сперва проверку простой гипотезы.

Мы будем говорить, что произвольная гипотеза H является простой, если $H : F = F_0$, то есть гипотеза состоит из равенства одному распределению.

В противном случае мы будем называть гипотезу сложной.

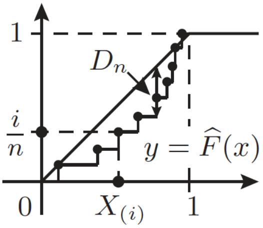
Критерии согласия

1. Критерий Колмогорова.

Критерий Колмогорова базируется на эмпирической функции распределения $\widehat{F}_n$ и ее отклонении от F_0.

Статистика критерия основана на величине

$$D_n = \sup_{u \in \mathbb{R}} \left| \widehat{F}_n(u) - F_0(u) \right|.$$



Критерии согласия

1. Критерий Колмогорова.

Для выборки достаточно большого размера, при верной H_0, значение D_n не должно существенно отклоняться от 0.

Теорема (Гливенко-Кантелли)

Пусть F_0 — функция распределения элементов выборки. Тогда статистика D_n стремится к 0 с вероятностью 1.

Критерии согласия

1. Критерий Колмогорова.

Как количественно охарактеризовать значимость отклонения D_n от нуля на конкретных данных?

Теорема (Колмогоров)

Пусть F_0 — функция распределения элементов выборки. Если F_0 непрерывна, то для любого $t > 0$, при $n \to \infty$,

$$\mathbb{P}(\sqrt{n}D_n \leq t) \to K(t) := 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2t^2}.$$

$K(t)$ называется функцией Колмогорова, а соответствующее распределение — распределением Колмогорова.

Критерии согласия

1. Критерий Колмогорова.

Быстрая сходимость к предельному закону позволяет пользоваться этим приближением уже при $n \geq 20$.

Условие непрерывности функции распределения необходимо. Например, в схеме Бернулли статистика $\sqrt{n}D_n$ имеет другой предельный закон распределения.

Приведем таблицу некоторых квантилей функции $K(t)$.

α	0,5	0,15	0,1	0,05	0,025	0,01	0,001
$x_{1-\alpha}$	0,83	1,14	1,23	1,36	1,48	1,63	1,95

Критерии согласия

1. Критерий Колмогорова. Резюме

выборка:	$\mathbf{X} = (X_1, \ldots, X_n)$
нулевая гипотеза:	$H_0 : \; X_i \sim F_0, \; F_0$ непрерывна
альтернатива:	$H_1 : \; X_i \not\sim F_0$
статистика:	$\sqrt{n} D_n = \sqrt{n} \cdot \sup_{u \in \mathbb{R}} \left\| \widehat{F}_n(u) - F_0(u) \right\|$
нулевое распределение:	$\sqrt{n} D_n \sim K$ – распределение Колмогорова

Критерии согласия

1. Критерий Колмогорова.

Пример (Гипотеза равномерности)

Критерий Колмогорова можно использовать для проверки гипотезы, что выборка взята из равномерного распределения на отрезке $[0, 1]$. Функция распределения данного закона является непрерывной.

- Зафиксируем уровень значимости, например, $\alpha = 0.05$.
- Вычислим значение статистики $\sqrt{n}D_n$.
- Если оно окажется больше или равно $x_{1-\alpha} = x_{0.95} = 1.36$, то отвергнем гипотезу равномерности.

Критерии согласия

2. Критерий Пирсона (хи-квадрат).

Критерий Пирсона (критерий хи-квадрат) основан уже на другой статистике — частотах.

Этот критерий можно использовать для проверки простой гипотезы о равенстве распределения не только в непрерывном, но и в дискретном случае.

Критерии согласия

2. Критерий Пирсона (хи-квадрат).

Начнем с дискретного случая. Пусть нам дана выборка $X_1, \ldots, X_n$ из дискретного закона

X	a_1	a_2	$\ldots$	a_k
$\mathbb{P}$	p_1	p_2	$\ldots$	p_k

Статистикой критерия является величина

$$T_n = \sum_{i=1}^{k} \frac{(\nu_i - np_i)^2}{np_i},$$

где ν_i — количество значений a_i в реализации $x_1, \ldots, x_n$.

Критерии согласия

2. Критерий Пирсона (хи-квадрат).

Как количественно охарактеризовать значимость отклонения T_n от нуля на конкретных данных?

Теорема (Пирсон)

Пусть реализация $x_1, \ldots, x_n$ *получена из закона* X. *Тогда, при* $n \to \infty$, *распределение статистики* T_n *сходится к закону* χ^2_{k-1}.

Приближение распределения статистики T_n с помощью закона χ^2_{k-1} является достаточно точным при $n \geq 50$ и $np_i \geq 5$ для всех $i = 1, \ldots, k$.

Критерии согласия

2. Критерий Пирсона (хи-квадрат).

Что делать если количество возможных значений X счетно?

В этом случае необходимо «сгруппировать» значения, которые принимаются с малыми вероятностями (причем так, чтобы получилось $np_i \geq 5$ для всех $i = 1, \ldots, k$).

Критерии согласия

2. Критерий Пирсона (хи-квадрат).

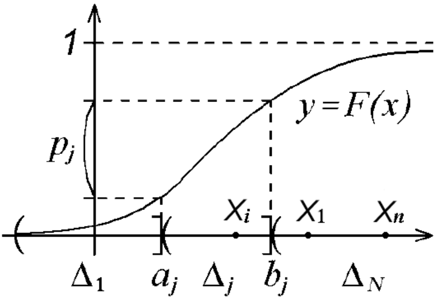
Что делать в непрерывном случае?

В этом случае необходимо дискретизировать распределение.

Критерии согласия

2. Критерий Пирсона (хи-квадрат).

Разобьем множество значений $X \sim F$ на k промежутков (возможно, бесконечных) $\Delta_i = [a_i, b_i]$ (интервалы примыкают друг к другу, то есть $a_j = b_{j-1}$). Положим $p_i = \mathbb{P}(X \in \Delta_i)$.



Критерии согласия

2. Критерий Пирсона (хи-квадрат).

Далее все остается без изменений: статистикой критерия является величина

$$T_n = \sum_{i=1}^{k} \frac{(\nu_i - np_i)^2}{np_i},$$

где теперь ν_i — количество значений $x_1, \ldots, x_n$, которые попали в промежуток Δ_i.

Распределение статистики T_n тоже сходится к закону χ^2_{k-1}.

Критерии согласия

2. Критерий Пирсона (хи-квадрат).

Как на практике выбирать промежутки и их количество?

- Число промежутков k не должно быть слишком большим (все так же необходимо $np_i \geq 5$ для всех $i = 1, \ldots, k$), но и не слишком малым, так как в этом случае дискретное распределение будет плохо аппроксимировать закон распределения F. Обычно на практике берут $k \approx \log n$.
- Сами промежутки чаще всего берут равными по длине. Альтернативным выбором является разбиение возможных значений на равновероятные промежутки.

Критерии согласия

2. Критерий Пирсона (хи-квадрат). Резюме

выборка: $\mathbf{X} = (X_1, \ldots, X_n)$

нулевая гипотеза: $H_0: \ X_i \sim F_0$

альтернатива: $H_1: \ X_i \nsim F_0$

статистика: $T_n = \sum_{i=1}^{k} \frac{(\nu_i - np_i)^2}{np_i}$

нулевое распределение: $T_n \sim \chi^2_{k-1}$ — хи-квадрат с $k-1$ степенью свободы (здесь k — либо количество возможных значений, либо количество промежутков разбиения)

Критерии согласия

2. Критерий Пирсона (хи-квадрат). Резюме

Пример (Генетические законы Менделя)

В экспериментах с селекцией гороха Мендель наблюдал частоты различных видов семян, получаемых при скрещивании растений с круглыми желтыми семенами и растений с морщинистыми зелеными семенами.

Эти данные и значения теоретических вероятностей, определяемые в соответствии с законом Менделя, приведены в следующей таблице.

Критерии согласия

2. Критерий Пирсона (хи-квадрат). Резюме

Пример (Генетические законы Менделя)

Тип семян	Частота	Вероятность
Круглые и желтые	315/556	9/16
Морщинистые и желтые	101/556	3/16
Круглые и зеленые	108/556	3/16
Морщинистые и зеленые	32/556	1/16

Критерии согласия

2. Критерий Пирсона (хи-квадрат). Резюме

Пример (Генетические законы Менделя)

Проверим гипотезу H_0 о согласованности частот с теоретическими вероятностями при помощи критерия хи-квадрат.

- Статистика критерия $T_n \approx 0.47$.
- Это значение приводит к фактическому уровню значимости p-value≈ 0.925.
- Таким образом, согласие наблюдений с гипотезой H_0 очень хорошее.

Критерии согласия

Как думаете, какой критерий чаще лучше работает в непрерывном случае? Критерий Колмогорова или хи-квадрат?

Неоднозначный ответ: критерий Колмогорова. Критерий хи-квадрат слишком универсален и довольно груб из-за потери информации при группировке.

Критерии согласия

Перейдем теперь к сложным гипотезам.

Гораздо чаще у нас есть гипотеза о принадлежности к параметрическому семейству, например, что выборка нормальная, но с неизвестными параметрами.

Как быть в этом случае?

Критерии согласия

Можно оценить неизвестные параметры состоятельными оценками, но эта процедура может сместить распределение статистик критерия.

Например, в случае с нормальным распределением, оценить среднее и дисперсию с помощью оценок максимального правдоподобия и применить критерий Колмогорова.

Однако в этом случае предельным распределением уже будет распределение Лиллиефорса, а не Колмогорова.

Это замечание крайне важно и зачастую игнорируется малоопытными аналитиками!

Критерии согласия

Рассматривать критерии с подстановкой состоятельных оценок мы не будем, информация о них будет в следующем дополнительном задании.

Вместо этого мы рассмотрим довольно мощные специализированные критерии для некоторых конкретных семейств распределений.

Критерии согласия

1. Проверка показательности

Под гипотезой показательности понимается сложная гипотеза

$$H_0 : F \in \{F_\theta\}_{\theta>0},$$

где класс $\{F_\theta\}_{\theta>0}$ образуют функции распределения вида

$$F_\theta(u) = (1 - e^{-\theta u}) I_{\{u \geq 0\}}.$$

Критерии согласия

1. Проверка показательности

(a) Исключение неизвестного параметра

- Положим $S_k = X_1 + \ldots + X_k$, $k = 1, \ldots, n$.
- Можно доказать, что для показательного закона случайный вектор $(S_1/S_n, \ldots, S_{n-1}/S_n)$, распределен так же, как и вариационный ряд из равномерного распределения на $[0, 1]$ размера $n - 1$.
- Данное преобразование сводит задачу к проверке равномерности. Однако, за исключение «мешающего» параметра θ приходится платить уменьшением размера выборки на 1.

Критерии согласия

1. Проверка показательности

(б) Критерий Гини (Gini)

Этот критерий базируется на статистике

$$G_n = \frac{\sum_{i=1}^{n} X_{(i)}(2i - n - 1)}{n(n-1)\overline{X}},$$

которая при нормировке $12(n-1)(G - 0.5)$ имеет асимптотическое нормальное распределение.

Критерии согласия

1. Проверка показательности

Для проверки показательности существует и ряд других критериев (например, Шапиро-Уилка для экспоненциального случая или Андерсона-Дарлинга).

Другие критерии могут быть основаны на других идеях.

Критерии согласия

2. Проверка нормальности

Под гипотезой нормальности понимается сложная гипотеза

$$H_0 : F \in \{F_{\mu,\sigma}\}_{\mu\in\mathbb{R},\,\sigma>0},$$

где класс $\{F_\theta\}_{\mu\in\mathbb{R},\,\sigma>0}$ образуют функции распределения вида

$$F_{\mu,\sigma}(u) = \Phi\left(\frac{u-\mu}{\sigma}\right),$$

где Φ — функция распределения стандартного нормального закона.

Критерии согласия

2. Проверка нормальности

(a) Критерий Шапиро-Уилка (Shapiro-Wilk)

Этот критерий базируется на статистике

$$SW_n = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n} (X_i - \overline{X})},$$

где a_i — некоторые константы.

Этот тест показывает очень хорошие результаты даже на небольших выборках.

Критерии согласия

2. Проверка нормальности

(б) Критерий Харке—Бера (Jarque-Bera)

Этот критерий использует статистику

$$JB_n = n\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right), \quad S = \frac{\mu_3}{\mu_2^3}, \quad K = \frac{\mu_4}{\mu_2^4},$$

где μ_k — k-ый центрированный выборочный момент:

$$\mu_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k.$$

Этот критерий тоже показывает хорошие результаты на практике.

Критерии согласия

2. Проверка нормальности

Кроме того, нормальность данных можно «проверить» еще визуально при помощи квантильного графика (Q-Q Plot).

Напомним, что мы проверяем гипотезу принадлежности семейству вида

$$F_{\mu,\sigma}(u) = \Phi\left(\frac{u-\mu}{\sigma}\right),$$

Критерии согласия

2. Проверка нормальности

Идея квантильного графика заключается в следующем:

- Возьмем в качестве приближения F эмпирическую функцию распределения $\widehat{F}_n$.
- Рассмотрим следующий график $y(x) = \Phi^{-1}(\widehat{F}_n(x))$. Если $F \in F_{\mu,\sigma}(u)$, то $y(x) \approx \Phi^{-1}(F(x)) = (x - \mu)/\sigma$.
- Это означает, что данный график не должен сильно отличаться от линейного.

Критерии согласия

2. Проверка нормальности

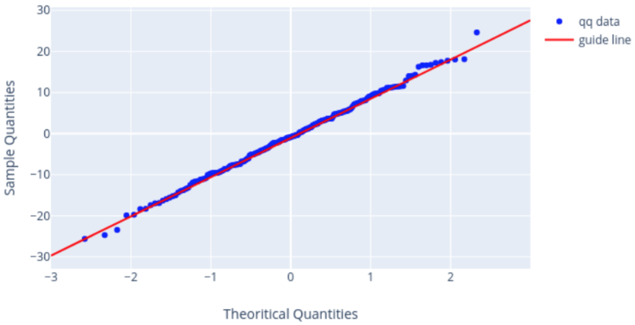
Для реализации этого способа достаточно отметить только точки, которые соответствую «скачкам» $\widehat{F}_n$ и подогнать под это облако точек прямую.

Если точки будут лежать далеко от прямой, то, скорее всего, предположение о том, что выборка взята из нормального распределения, не выполняется.

Обратите внимание, что на графике будут отложены точки $(x_{(i)}, \Phi^{-1}(i/n))$, то есть мы будем сравнивать эмпирические и теоретические квантили. Поэтому график так называется.

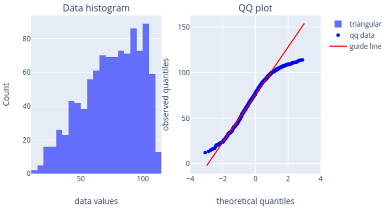
Критерии согласия

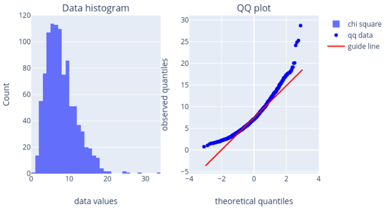
2. Проверка нормальности

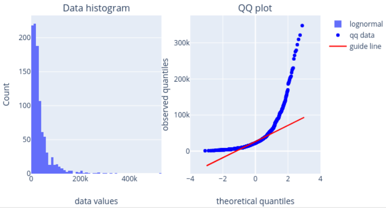


Критерии согласия

2. Проверка нормальности







Критерии согласия

2. Проверка нормальности

Квантильный график можно строить не только для нормального распределения, но и для любого другого семейства сдвига/масштаба (например, для равномерного и экспоненциального распределения).

Кроме того, его можно построить и для двух выборок, чтобы визуально проверить гипотезу о том, что выборки взяты из одного и того же распределения.

Критерии однородности

Теперь перейдем к другого типа критериям — критериям однородности. Они позволяют проверить гипотезу о том, что выборки взяты из одного распределения.

Критерии однородности

Примеры:

1. Лабораторных мышей поместили в двухкомнатные клетки. В одной комнате висело зеркало, а в другой — нет. С целью установить, есть ли у мышей какие-то предпочтения насчет зеркал, измерялась доля времени, которое каждая мышь проводила в каждой из своих двух клеток. По данным долям требуется выяснить, имеется ли значимое предпочтение у мышей или отклонения являются чисто случайными и объясняются «естественной» дисперсией выбранной характеристики.

Критерии однородности

Примеры:

2. Первая выборка — это пациенты, которых лечили препаратом А. Вторая выборка — пациенты, которых лечили препаратом Б. Значения в выборках — это некоторая характеристика эффективности лечения (уровень метаболита в крови, температура через три дня после начала лечения, срок выздоровления и т.д.) Требуется выяснить, имеется ли значимое различие эффективности препаратов А и Б, или различия являются чисто случайными и объясняются «естественной» дисперсией выбранной характеристики.

Критерии однородности

Примеры:

3. Первая выборка — это значения некоторой характеристики состояния пациентов, записанные до лечения. Вторая выборка — это значения той же характеристики состояния тех же пациентов, записанные после лечения. Требуется выяснить, имеется ли значимое отличие в состоянии пациентов до и после лечения, или различия чисто случайны.

Критерии однородности

Примеры:

4. Первая выборка — это поля, обработанные агротехническим методом А. Вторая выборка — поля, обработанные агротехническим методом Б. Значения в выборках — это урожайность. Требуется выяснить, является ли один из методов эффективнее другого, или различия урожайности обусловлены случайными факторами.

Критерии однородности

Примеры:

5. Первая выборка — это дни, когда в супермаркете проходила промо-акция типа А (красные ценники со скидкой). Вторая выборка — дни промо-акции типа Б (при покупке двух единиц — подарок). Значения в выборках — это показатель эффективности промо-акции (объём продаж, либо выручка в рублях). Требуется выяснить, какой из типов промо-акции более эффективен.

Критерии однородности

Мы будем рассматривать параметрические и непараметрические критерии.

В параметрических критериях мы будем делать предположение о том, что выборки взяты из какого-то параметрического семейства, а в непараметрических — нет.

Как итог: непараметрические критерии менее чувствительные (потому что более общие), но зато они не требуют «идеальных» условий, например, нормальности данных.

Часто бывает так, что при совсем небольших отклонениях от «идеального» распределения непараметрические критерии работают значительно лучше параметрических.

Критерии однородности

Начнем с параметрических критериев. Мы будем различать:

- случаи одной выборки
- случаи двух выборок
  - выборки зависимые
  - выборки независимые

Критерии однородности

1a. Одновыборочный Z-критерий.

выборка: $\mathbf{X} = (X_1, \ldots, X_n)$
$X_i \sim \mathcal{N}(\mu, \sigma^2)$, $\underline{\sigma \text{ известна}}$

нулевая гипотеза: $H_0: \ \mu = \mu_0$ (здесь $\mu_0 \in \mathbb{R}$ задано)

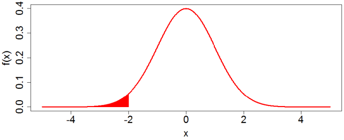
альтернатива: $H_1: \ \mu \neq \mu_0$ или $\mu > \mu_0$ или $\mu < \mu_0$

статистика: $Z_n = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$

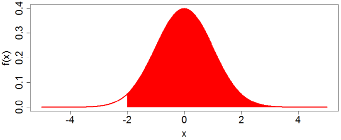
нулевое распределение: $Z_n \sim \mathcal{N}(0, 1)$

Критерии однородности

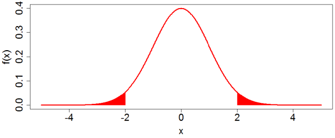
1a. Одновыборочный Z-критерий.



$$H_1 : \mu < \mu_0$$



$$H_1 : \mu > \mu_0$$



$$H_1 : \mu \neq \mu_0$$

Критерии однородности

16. Одновыборочный t-критерий.

выборка: $\mathbf{X} = (X_1, \ldots, X_n)$
$X_i \sim \mathcal{N}(\mu, \sigma^2)$, $\underline{\sigma\text{ неизвестна}}$

нулевая гипотеза: $H_0: \ \mu = \mu_0$ (здесь $\mu_0 \in \mathbb{R}$ задано)

альтернатива: $H_1: \ \mu \neq \mu_0$ или $\mu > \mu_0$ или $\mu < \mu_0$

статистика: $T_n = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$,

где $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$

нулевое распределение: $T_n \sim \mathbf{T}_{n-1}$

Критерии однородности

2а. Двухвыборочный Z-критерий, независимые выборки.

выборки: $\mathbf{X} = (X_1, \dots, X_{n_1})$
$\mathbf{Y} = (Y_1, \dots, Y_{n_2})$

$X_i \sim \mathcal{N}(\mu_1, \sigma_1^2), \; Y_i \sim \mathcal{N}(\mu_2, \sigma_2^2)$

$\underline{\sigma_1 \text{ и } \sigma_2 \text{ известны}}$

нулевая гипотеза: $H_0: \; \mu_1 = \mu_2$

альтернатива: $H_1: \; \mu_1 \neq \mu_2$ или $\mu_1 > \mu_2$ или $\mu_1 < \mu_2$

статистика: $$Z_n = \frac{\overline{X} - \overline{Y}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

нулевое распределение: $Z_n \sim \mathcal{N}(0, 1)$

Критерии однородности

2б. Двухвыборочный t-критерий, независимые выборки.

выборки: $\mathbf{X} = (X_1, \ldots, X_{n_1})$
$\mathbf{Y} = (Y_1, \ldots, Y_{n_2})$

$X_i \sim \mathcal{N}(\mu_1, \sigma_1^2),\ Y_i \sim \mathcal{N}(\mu_2, \sigma_2^2)$

$\underline{\sigma_1 \text{ и } \sigma_2 \text{ неизвестны}}$

нулевая гипотеза: $H_0: \ \mu_1 = \mu_2$

альтернатива: $H_1: \ \mu_1 \neq \mu_2$ или $\mu_1 > \mu_2$ или $\mu_1 < \mu_2$

статистика: $$T_n = \frac{\overline{X} - \overline{Y}}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

нулевое распределение: $T_n \approx \mathbf{T}_k$ для некоторого $k \in \mathbb{N}$

Критерии однородности

26. Двухвыборочный t-критерий, независимые выборки.

Задача сравнения средних двух нормально распределённых выборок при неизвестных и неравных дисперсиях известна как проблема Беренса-Фишера.

Точного решения этой задачи нет.

Критерии однородности

2б. Двухвыборочный t-критерий, независимые выборки.

Однако рассмотренная аппроксимация достаточно точна в двух ситуациях:

1. Если выборки одинакового объема $n_1 = n_2$.
2. Если знак неравенства между n_1 и n_2 такой же, как между σ_1 и σ_2, то есть выборка с большей дисперсией имеет больший объем.

Критерии однородности

3. Двухвыборочный t-критерий, зависимые выборки.

Рассмотрим теперь случай связанных выборок, то есть элементы X_i и Y_i соответствуют одному и тому же объекту, но измерения сделаны в разные моменты (например, до и после обработки).

Сравнение средних в связанных выборках ничем не отличается от сравнения среднего разности $D_i = X_i - Y_i$ с нулём.

Критерии однородности

3. Двухвыборочный t-критерий, зависимые выборки.

выборки: $\mathbf{X} = (X_1, \ldots, X_n)$

$\mathbf{Y} = (Y_1, \ldots, Y_n)$

$X_i \sim \mathcal{N}(\mu_1, \sigma^2), \; Y_i \sim \mathcal{N}(\mu_2, \sigma^2)$

σ неизвестна

нулевая гипотеза: $H_0: \; \mu_1 = \mu_2$

альтернатива: $H_1: \; \mu_1 \neq \mu_2$ или $\mu_1 > \mu_2$ или $\mu_1 < \mu_2$

статистика: $$T_n = \frac{\overline{X} - \overline{Y}}{S/\sqrt{n}},$$

где $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \overline{D})^2, \; D_i = X_i - Y$

нулевое распределение: $T_n \sim \mathbf{T}_{n-1}$

Спасибо за внимание!