Кафедра математической статистики ИВМиИТ КФУ 30 декабря 2020 г.

Научно-популярная статья к выпуску подготовил Казанцев А.В.

Автор: Альбина Галимова, магистрант 2 курса кафедры математической статистики

t-критерий Стьюдента

Альбина Галимова – автор статьи

Альбина Галимова – бакалавр математики (КФУ, 2019). Выпускная квалификационная работа: Галимова А.Р. «Исследование модифицированной модели ARCH(1)». Научный руководитель – доцент С.Г. Халиуллин.

t-критерий Стьюдента – это общее название для класса методов статистической проверки гипотез (статистических критериев), основанных на распределении Стьюдента. Наиболее частые случаи применения t-критерия связаны с проверкой равенства средних значений в двух выборках.

Условия применения статистического критерия Стьюдента:

- Сравниваемые выборки должны происходить из нормально распределенных совокупностей;
- Дисперсии сравниваемых генеральных совокупностей должны быть равны.

Кроме того, в своей исходной форме, t-критерий предполагает независимость сравниваемых выборок.

Проверка указанных требований к данным должна всегда предшествовать формальному статистическому анализу, в котором задействован критерий Стьюдента. Отметим, что условие нормальности распределения данных становится не таким жестким при "больших" объемах выборок, а для выборок с разными дисперсиями существует особая модификация t-критерия (критерий Уэлча).

Предположим нормальность выборочных данных и равенство дисперсий. В этом случае мы можем использовать данный критерий. Правило, по которому принимается или отвергается нулевая гипотеза, называется критерием. Критерий определяется заданием особого подмножества S выборочного пространства $X^{(n)}$, которое называется критической областью: если выборочные данные $x^{(n)}$ попадают в эту область, то нулевая гипотеза отклоняется и принимается альтернатива.

Область $A = S^c = X^{(n)} \setminus S$ называется областью принятия нулевой гипотезы. Обычно критическую область определяют с помощью индикаторной функции $\varphi = \varphi(X^{(n)})$, которая называется критической функцией или просто критерием. Данная функция принимает значение 1, если произошло событие $X^{(n)} \in S$, значение 0 в противном случае. Тогда математическое ожидание $E\varphi(X^{(n)})$ интерпретируется как вероятность отклонения гипотезы, т.е. $E\varphi\left(X^{(n)}\right) = P_\theta\left(X^{(n)} \in S\right) = m(\theta)$ – функция мощности критерия φ. $m(\theta)$ указывает, как часто отклоняется нулевая гипотеза, когда θ – истинное значение параметра.

Существует две компоненты функции риска: $\alpha(\theta) = m(\theta)$ при $\theta \in \Theta_0$ и $\beta(\theta) = 1 - m(\theta)$ при $\theta \in \Theta_1$. Функция $\alpha(\theta)$, $\theta \in \Theta_0$ называется вероятностью ошибки первого рода (вероятность отклонить нулевую гипотезу, когда она верна), $\beta(\theta)$, $\theta \in \Theta_1$ – вероятность ошибки второго рода (вероятность принять нулевую гипотезу, когда она ложна).

Предположим, что отклонение гипотезы H_0, когда она в действительности верна, приводит к более тяжким последствиям, чем ее принятие при справедливости альтернативы. В таком случае мы заинтересованы в первую очередь контролировать вероятность ошибки первого рода. С этой целью заранее фиксируется (выбирается) некоторый уровень α, выше которого вероятность ошибки первого рода не допустима, и критическая область S (критерий φ) определяется таким образом, что $\alpha(\theta) \leq \alpha$, каково бы ни было $\theta \in \Theta_0$.

Это ограничение α на вероятность ошибки первого рода называется уровнем значимости, а сам критерий φ, для которого выполняется это ограничение, – критерием уровня α. Наибольшее значение вероятности ошибки первого рода $\bar{\alpha} = \sup \alpha(\theta)$, $\theta \in \Theta_0$, называется размером критерия φ, и если $\bar{\alpha} = \alpha$, то говорят о критерии φ размера α.

Простейший метод построения критериев значимости состоит в использовании состоятельных оценок тестируемого параметра θ. Рассмотрим простейший случай: θ – скалярный параметр, вероятностная модель не содержит других (мешающих) параметров и проверяется простая гипотеза H_0: $\theta = \theta_0$ при альтернативе H_1: $\theta \neq \theta_0$, где θ_0 – некоторое, априори фиксированное значение параметра θ.

Если $\hat{\theta} = \hat{\theta}\left(X^{(n)}\right)$ – состоятельная оценка θ, дисперсия которой стремится к 0 при $n \to \infty$ как $O(\frac{1}{\sqrt{n}})$, то естественно определить критическую область с помощью неравенства $\sqrt{n}\left|\hat{\theta}_n\left(X^{(n)}\right) - \theta_0\right| > C$. Вероятность ошибки первого рода в этом случае $\alpha(\theta_0, C) = P_{\theta_0}(\sqrt{n}\left|\hat{\theta}_n\left(X^{(n)}\right) - \theta_0\right| > C)$ и приравнивая эту вероятность заданному уровню значимости α, находим критическую константу $C = C(\alpha)$ как квантиль распределения случайной величины $\sqrt{n}\left|\hat{\theta}_n\left(X^{(n)}\right) - \theta_0\right|$ (такой выбор C приводит к критерию уровня α). Если θ ($\neq \theta_0$) – некоторое альтернативное значение параметра, тогда в силу состоятельности оценки $\sqrt{n}\left|\hat{\theta}_n\left(X^{(n)}\right) - \theta_0\right| \to \infty$ (по вероятности) и поэтому вероятность ошибки второго рода стремится к 0 при $n \to \infty$.

Рассмотрим некоторые наиболее часто встречаемые виды t-критерия на примерах. Вычисления были выполнены в среде R.

Одновыборочный t-критерий

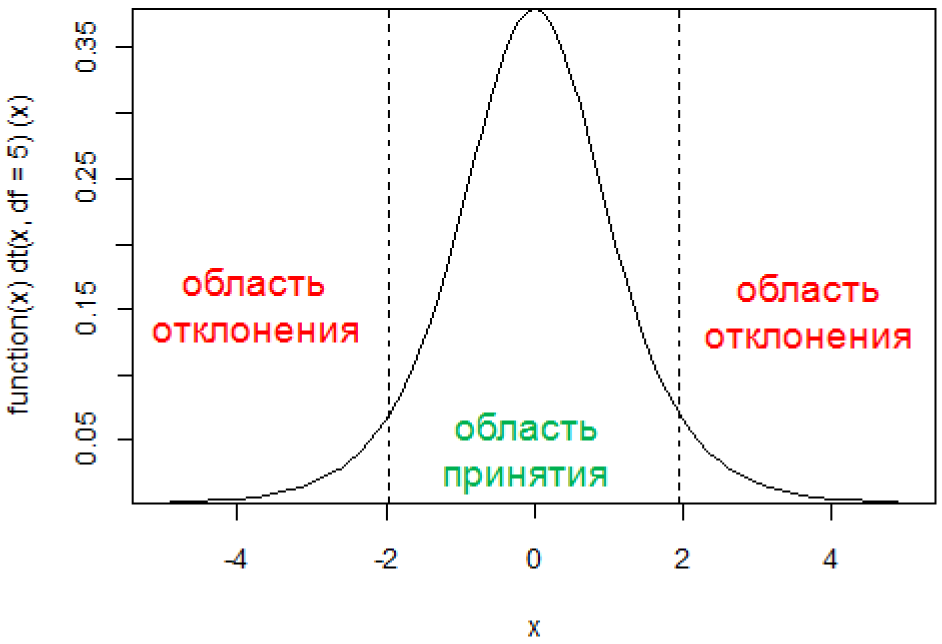
Этот вариант критерия Стьюдента служит для проверки нулевой гипотезы (H_0) о равенстве среднего значения (μ_1) генеральной совокупности, из которой была взята выборка, некоторому известному значению (μ_0):

$$H_0: \mu_1 = \mu_0$$

В общем виде проверка этой гипотезы выполняется при помощи t-критерия, который рассчитывается как отношение разницы между выборочным средним и известным значением к стандартной ошибке выборочного среднего:

$$t = \frac{\bar{x} - \mu_0}{S_{\bar{x}}}$$

Рассчитанное значение критерия мы можем далее интерпретировать следующим образом, исходя из свойств t-распределения: если это значение попадает в так называемую область отклонения нулевой гипотезы, то мы вправе отклонить проверяемую нулевую гипотезу. Область отклонения нулевой гипотезы для критерия Стьюдента определяется заранее принятым уровнем значимости (например, $\alpha = 0{,}05$) и числом степеней свободы.



Эквивалентным подходом к интерпретации результатов теста будет следующий: допустив, что нулевая гипотеза верна, мы можем рассчитать, насколько велика вероятность получить t-критерий, равный или превышающий реальное значение, которое мы рассчитали по имеющимся выборочным данным. Если эта вероятность оказывается меньше, чем заранее принятый уровень значимости (например, $P < 0.05$), то мы вправе отклонить проверяемую гипотезу.

Рассмотрим следующий пример. Предположим, у нас имеются данные по суточному потреблению энергии, поступающей с пищей (кДж/сутки), для 11 женщин (пример заимствован из книги Altman D. G. (1981) Practical Statistics for Medical Research, Chapman & Hall, London).

X = (5260, 5470, 5640, 6180, 6390, 6515, 6805, 7515, 7515, 8230, 8770)

Среднее значение для них = 6753.6

Нас интересует: отличается ли это выборочное среднее от установленной нормы в 7725 кДж/сутки? Их разница = 7725 – 6753.6 = 971.4, достаточно приличная. Узнаем на сколько она велика статистически.

Значение статистики получаем = -2.8208, с 10 степенями свободы, значение p-value = 0.01814 (при $\alpha = 0.05$). Следовательно, мы можем отклонить проверяемую нулевую гипотезу о равенстве средних и принять альтернативу, рискуя ошибиться с вероятностью менее 5 %.

Сравнение двух независимых выборок

При сравнении двух выборок проверяемая нулевая гипотеза заключается в том, что обе выборки происходят из нормального распределенных генеральных совокупностей с одинаковыми средними значениями:

$$H_0: \mu_1 = \mu_2$$

Поскольку эти генеральные средние мы оцениваем при помощи выборочных средних значений, формула t-критерия приобретает вид

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}}$$

В знаменателе приведенной формулы находится стандартная ошибка разницы между выборочными средними, которая в общем виде рассчитывается как

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

где s_1^2 и s_2^2 выборочные оценки дисперсии. При соблюдении условия о равенстве групповых дисперсий приведенная формула приобретает более простой вид. Интерпретация t-критерия, рассчитанного для двух выборок, выполняется точно так же, как и в случае с одной выборкой.

Рассмотрим пример о суточном расходе энергии у худощавых женщин и женщин с избыточным весом, приведенный в книге Питера Дальгаарда (Dalgaard P (2008) Introductory statistics with R. Springer).

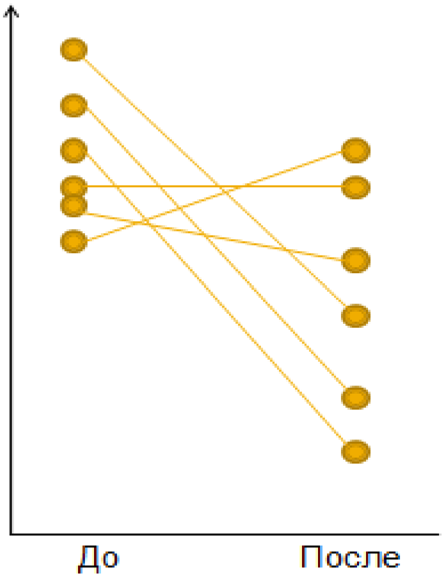
X = (9.21, 7.53, 7.48, 8.08, 8.09, 10.15, 8.40, 10.88, 6.13, 7.90, 11.51, 12.79, 7.05, 11.85, 9.97, 7.48, 8.79, 9.69, 9.68)

У = (изб., худ., худ., худ., худ., худ., худ., худ., худ., худ., изб., изб., худ., изб., изб., худ., изб., изб., изб.)

Соответственно для каждого типа получим средние значения: для худощавых = 8.07, для избыточного веса = 10.30.

Проверим гипотезу об отсутствии разницы между средними при помощи t-теста. Получим значение статистики = -3.8555, степени свободы = 15.919 и значение p-value = 0.001411. Это означает, что средние значения потребления энергии у женщин из рассматриваемых весовых групп статистически значимо различаются. Отвергая эту гипотезу, мы рискуем ошибиться с вероятностью около 0.1%.

Сравнение двух зависимых (парных) выборок



Зависимыми или парными являются две выборки, если они содержат результаты измерений какого-либо количественного признака для одного и того же объекта. Во многих исследованиях какой-то определенный отклик измеряется у одних и тех же объектов до и после экспериментального воздействия. При такой схеме эксперимента исследователь более точно оценивает эффект воздействия, так как прослеживает его у одних и тех же объектов.

В общем виде критерий Стьюдента можно представить так

$$t = \frac{\text{оценка параметра} - \text{истинное значение параметра}}{\text{статистическая ошибка оценки параметра}}$$

Нас интересует «истинное значение параметра» - среднее изменение какого-либо количественного признака как результат экспериментального воздействия, например, δ. Оценкой этого истинного параметра является наблюдаемое (выборочное) среднее изменение признака. Тогда t-критерий примет вид

$$t = \frac{\bar{d} - \delta}{S_{\bar{d}}}$$

Если нулевая гипотеза заключается в равенстве истинного эффекта нулю, то формула для парного критерия Стьюдента примет вид

$$t = \frac{\bar{d}}{S_{\bar{d}}}$$

В книге П. Дальгаарда (Dalgaard 2008) есть пример о суточном потреблении энергии, измеренном у одних и тех же 11 женщин до и после определенного события:

До = (5260, 5470, 5640, 6180, 6390, 6515, 9805, 7515, 7515, 8230, 8770)

После = (3910, 4220, 3885, 5160, 5645, 4680, 5265, 5975, 6790, 6900, 7335)

Сравним каждое значение (После - До) и усредним результат, получим значение = -1320.5. Оценим на сколько статистически значимо это среднее значение отличается от нуля при помощи парного критерия Стьюдента.

Получили значение статистики = 11.9414, со степенями свободы = 10 и значением p-value = $3.059 * 10^{-7}$. Как можно увидеть, значение p-value намного меньше 0.05, что позволяет сделать заключение о наличии существенной разницы в потреблении энергии у исследованных женщин до и после некоторого события.

Список литературы

1. Володин И. Н. Лекции по теории вероятностей и математической статистике. – Казань: (Издательство), 2006. – 271 с.

2. Классические методы статистики: t-критерий Стьюдента - https://r-analytics.blogspot.com/2012/03/t.html

3. t-критерий Стьюдента для независимых совокупностей - https://medstatistic.ru/methods/methods.html