Лекция 6. Статистическая гипотеза. Проверка гипотез

Курбацкий А. Н.

МШЭ МГУ

23 марта 2020

Содержание

Содержание

Идея

Определение

Статистическая гипотеза - это некоторое предположение о свойствах и характеристиках исследуемых генеральных совокупностей.

Это предположение проверяется на основе анализа выборок. В этой теме мы будем иметь дело с параметрическими гипотезами, то есть с гипотезами о параметрах исследуемой ГС (о среднем, доле, дисперсии и т.п.). Непараметрические гипотезы будут рассмотрены позже.

Пример

Если однокурсница сказала вам, средний вес студентки в университете равен 60 кг, то это статистическая гипотеза, которую вы можете проверить, опросив знакомых девушек. Если у опрошенных девушек вес окажется значительно выше 60, то, видимо, это гипотеза неверна.

Но может оказаться, что вы просто общаетесь с крупными девушками! То есть ваша выборка не является показательной. Выборка должна представлять весь университет в уменьшенном масштабе или, как говорят социологи, быть **репрезентативной**. Везде далее мы будем предполагать выполнение этого условия.

Возникает несколько вопросов, с которыми нам предстоит разобраться.

- Во-первых, с какой величиной будем сравнивать гипотетический результат, то есть что вычислять по выборке?
- Во-вторых, когда начинается это "значительно выше" , про которое было сказано в примере?

Про монетку

Если вы захотели проверить предположение о том, что монета является симметричной, то можно взять и подбросить её, скажем, 10 раз и посчитать число гербов.

- Можно ли считать монетку симметричной, если герб выпал 6 раз? А если 10 раз?
- Какие значения числа гербов могут заставить усомниться в симметричности монеты? Ведь все исходы возможны.

Статистическая гипотеза

Определение

Основная или нулевая гипотеза H_0 - это гипотеза, которой мы придерживаемся, пока наблюдения не заставят признать обратное. Ей всегда сопутствует альтернативная гипотеза H_1.

Важно!

- *Статистические методы не позволяют доказать гипотезу. По наблюдениям, которыми мы располагаем, мы можем гипотезу* **опровергнуть***.*
- *И проблема состоит в том, что проверяем мы некоторое следствие, которое верно при выдвинутой гипотезе. Если следствие не соответствуют имеющимся данным, то и гипотеза неверна.*
- *Но если данные согласуются со следствием, то это не означает справедливости гипотезы.*

Пример

- Допустим, вы хотите проверить гипотезу о том, что вы умный. Для проверки гипотезы вы взяли результаты своего ЕГЭ по математике. Если у вас оказалось 25 баллов, то гипотезу, к сожалению, придётся отвергнуть. Если же баллы высокие, то гипотеза данными не опровергается, но при этом нельзя утверждать, что вы умный.
- Если при 10 бросках монеты герб выпал 4-6 раз, то такой результат согласуется с тем, что монета симметрична. Если же число гербов оказалось 10 или 0, то в симметричности возникают сомнения, ведь такое возможно с вероятность $2^{-10} = \frac{1}{1024}$.

Чтобы определиться, когда гипотезу отвергать, а когда не отвергать введем еще два понятия.

Определение

- *Ошибка первого рода - это ситуация, когда H_0 отвергается, хотя она, на самом деле, верна.*
- *Ошибка второго рода - это ситуация, когда H_0 принимается, хотя она неверна.*

Определение

Буквой α обозначается уровень значимости или вероятность ошибки первого рода. Буквой β - вероятность ошибки второго рода.

Естественно хочется сделать как можно меньше сразу обе ошибки, но это, к сожалению, невозможно. При уменьшении ошибки первого рода, увеличивается ошибка второго рода и наоборот. Обычно α берут $0.1, \ 0.05$ или 0.01.

Пример

- Мы совершаем ошибку первого рода, когда не берем съедобный гриб, думая, что он несъедобный.
- Ошибка второго рода выглядит так. Например, суд выдвигает гипотезу H_0: подсудимый невиновен. А он, на самом деле, виновен, но суд признает его невиновным за отсутствием улик (презумпция невиновности). То есть суд принимает гипотезу, хотя она неверна.
- "Ложноположительный результат" при при медицинских анализах - это ошибка какого рода?

Важно!

Надо не забывать, что ошибки первого и второго родов меняются в зависимости от того, какая гипотеза является основной, а какая альтернативной.

Для поверки гипотез используется функция, называемая **статистикой критерия**, которая зависит от выборки.

Определение

Статистикой критерия называется случайная величина, значение которой вычисляется по выборке.

С этого момента, под статистикой будем подразумевать некоторую функцию от выборки.
Для каждой задачи мы будем выбирать уровень значимости и статистику критерия, по значению которой будем делать вывод о справедливости гипотезы. При справедливости основной гипотезы будет известно, с какой вероятностью какое значение принимает статистика критерия. Если эта вероятность очень маленькая, то гипотезу придётся отвергнуть.

Мощность критерия

Определение

Мощностью критерия называется вероятность не совершить ошибку второго рода, то есть $1 - \beta$. А наиболее мощным критерием из всех критериев с уровнем значимости α называется тот, который обладает наибольшей мощностью.

Для сложных альтернатив мощность зависит от неизвестного параметра, поэтому возникает понятие **функции мощности**.

Наиболее мощный критерий

Компромисс между ошибками первого и второго родов можно найти, определив цену каждой ошибки и попытаться минимизировать комбинацию этих цен. Но во многих задачах это будет бессмысленно. Есть ещё один вариант.

Важно!

Зафиксировав уровень значимости, можно уменьшить ошибку второго рода, то есть увеличить мощность, за счёт выбора критической области. При определённых условиях можно получить наиболее мощный критерий! (Оптимальный критерий Неймана-Пирсона).

Замечание

Пирсон, кстати, не тот, который К, а который Э.

Ещё раз о главном

- Гипотеза либо отвергается, либо не отвергается.
- Старайтесь не употреблять слов "принимаем гипотезу" потому что невозможность отвергнуть гипотезу не означает, что она верна и ее стоит придерживаться. Может быть, просто недостаточно оснований или наблюдений, чтобы её отвергнуть.
- Но если очень хочется или удобно работать именно при таком предположение, то, конечно, вместо "не отвергаем гипотезу" можно сказать, мы "принимаем гипотезу" или мы "придерживаемся" этой гипотезы.

Критическая область

Теперь надо определиться, когда наступает момент, с которого мы будем отвергать гипотезу.

Определение

Критической областью *называется область значений статистики критерия, при которых отвергается H_0. А критические значения - это граница критической области.*

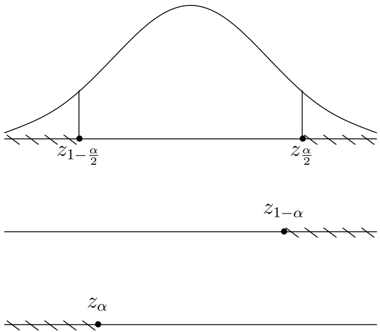
Существует три вида критических областей: левосторонняя, правосторонняя и двусторонняя.

Важно!

Вид критической области определяется видом альтернативной гипотезы.

Критическая область зависит от альтернативы!

Виды критических областей



- Если $H_1 : \theta \neq \theta_0$, то критическая область является двусторонней.
- При $H_1 : \theta > \theta_0$ критическая область является правосторонней,
- а при $H_1 : \theta < \theta_0$ - левосторонней.

Процедура проверки статистических гипотез

Теперь мы готовы узнать, как проверить статистическую гипотезу. Процедура проверки гипотезы состоит из нескольких этапов:

- Сформулировать основную и альтернативную гипотезы и задать уровень значимости α.
- Найти критические значения и построить критическую область.
- Вычислить по выборке значение статистики и посмотреть, попало ли оно в критическую область.
- Сделать вывод. Если значение попало в критическую область, то основная гипотеза отвергается, в противном случае, не отвергается.

Минимальный уровень значимости

- В случае, когда в задаче не дан уровень значимости, возникает естественный вопрос. Какой уровень значимости всё-таки лучше 1%, 2%, 5% или 10%? А может другой? Проверять каждый раз все трудоёмко. И вообще получается, что ответ зависит от того, какой уровень значимости взяли.
- Допустим, мы не отвергли гипотезу при 5% уровне значимости. Но нам хочется знать, с какой вероятностью ошибки первого рода мы её можем отвергнуть. Ошибка в 6% может быть вполне допустимой, а ошибка в 25% это уж слишком много.
- Нам нужна величина, которая позволит указать пороговое значение уровня значимости. И по которому сможем определить, с какой минимальной ошибкой первого рода гипотезу уже можно отвергать.

Определение

Минимальный уровень значимости ($p - value$) - это минимальное значение α, при котором основная гипотеза ещё отвергается.

Содержание

Проверка гипотезы о среднем (диспресия известна)

- В этом параграфе будем учиться проверять гипотезу $H_0 : \mu = \mu_0$ (μ означает среднее ГС, а μ_0 - некоторое предполагаемое нами фиксированное значение). Мы уже знаем, что для проверки гипотезы нужно знать статистику и ее распределение. Оказывается, в зависимости от условий, статистика имеет разный вид.
- Если стандартное отклонение ГС известно, то в этом случае статистика критерия $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ имеет нормальное распределение в предположении справедливости гипотезы H_0. Убедитесь в этом, вспомнив центральную предельную теорему из курса теории вероятностей!
- Для левосторонней области в таблице нормального распределения ищется квантиль уровня α, для правосторонней - квантиль уровня $1-\alpha$, а для двухсторонней - квантили уровня $\frac{\alpha}{2}$ и $1-\frac{\alpha}{2}$. При пользовании таблицами надо не забывать, что $z_\alpha = -z_{1-\alpha}$, в силу чётности функции плотности нормального распределения.

Попробуйте самостоятельно

Пример

Пусть студенты университета в начале учебного года сдают предварительное тестирование, оцениваемое по десятибалльной шкале. Вы предполагаете, что средняя оценка равна 6, и решаете это проверить, опросив несколько человек. Получилась следующая выборка: 9, 5, 7, 7, 4, 10. Из наблюдений прошлых лет известно, что дисперсия $\sigma^2 = 1$*. Проверим гипотезу, что среднее равно 6, на уровне значимости* $\alpha = 0.01$ *против односторонних альтернатив* $\mu > 6$*.*

Подсказка. Так как известна дисперсия генеральной совокупности, то для проверки гипотезы используется статистика $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$.

Решение

- Сформулируем основную и альтернативную гипотезы:

$$H_0 : \mu = 6; \quad H_1 : \mu > 6.$$

- Критическая область является правосторонней. По таблице нормального распределения находим $1 - \alpha = 0.99$ и определяем критическое значение $z_{cr} = 2.33$. Критическая область имеет вид $(2.33; +\infty)$.
- Вычислим значение статистики критерия. Среднее значение $\bar{x} = \frac{9+5+7+7+4+10}{6} = 7$, стандартное отклонение $\sigma = 1$. Значение статистики критерия равно $z = \frac{7-6}{1/\sqrt{6}} \approx 2.45$.
- Вывод. Так как $z \in (2.33; +\infty)$, то основная гипотеза H_0 отвергается.
- Минимальный уровень значимости составляет $1 - z^{-1}(2.45) \approx 0.007$.

Ответ: при данном уровне значимости и такой альтернативе гипотеза отвергается.

Содержание

В случае, когда стандартное отклонение ГС неизвестно, гипотеза проверяется с помощью, так называемой, t-статистики.

Теорема

Для выборки из нормальной ГС с неизвестной дисперсией, статистика критерия дляпроверки гипотезы о среднем имеет вид $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$.

Эта статистика имеет распределение Стьюдента с $n-1$ степенью свободы.

Замечание

Выборочное стандартное отклонение s можно по-прежнему посчитать по формуле $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \sqrt{\frac{1}{n-1}(\sum x_i^2 - n\bar{x}^2)}$.

Пример

Выборка 1, 0, 3, 5, 4, основная гипотеза $\mu = 3$, $\alpha = 0.01$, *односторонние альтернативы* $\mu < 3$.

Решение

- Сформулируем основную и альтернативную гипотезы:
$$H_0 : \mu = 3; \quad H_1 : \mu < 3.$$
- Критическая область является левосторонней. По таблице распределения Стьюдента находим $1 - \alpha = 0.99$, число степеней свободы $n - 1 = 4$ и определяем критическое значение $t_{cr} = -4.6$. Критическая область имеет вид $(-\infty; -4.6)$.
- Вычислим значение статистики. Среднее значение $\bar{x} = \frac{1+0+3+5+4}{5} = 2.6$, выборочная дисперсия $s^2 = \frac{1}{5-1}((1-2.6)^2 + (0-2.6)^2 + (3-2.6)^2 + (5-2.6)^2 + (4-2.6)^2) \approx 4.28$, откуда выборочное стандартное отклонение $s \approx 2.07$. Значение статистики критерия равно $t = \frac{2.6-2}{2.07/\sqrt{5}} \approx 0.64$.
- Вывод. Так как $t \notin (-\infty; -4.6)$, то основная гипотеза H_0 не отвергается.
- Минимальный уровень значимости равен $1 - t^{-1}(0.64) \approx 0.51$.

Ответ: при данном уровне значимости и такой альтернативе гипотеза не отвергается.

Для тренировки

Пример

Преподаватель по информатике заскучал на контрольной. Чтобы немного отвлечься, он стал наблюдать рабочие столы студентов у себя на мониторе и решил оценить, сколько запросов в поисковой системе будет ими введено, если их не останавливать. Не имея возможности следить за всеми, он следил за 12 студентами. Он насчитал общее число обращений $\sum x_i = 216$ *и вычислил* $\sum x_i^2 = 4046$*. На уровне значимости 5% проверьте гипотезу, что среднее число обращений равно 20 против двусторонних альтернатив.*

Решение

Самостоятельно

Содержание

Гипотеза о доле

Теперь мы научимся отвечать на вопрос, какая доля объектов в генеральной совокупности обладает определенным признаком.

Пример

Вы сомневаетесь в том, что доля избирателей некоторого кандидата на предстоящих выборах равна 0.6 *или 60%, и хотите проверить эту информацию. Для этого недостаточно просто опросить большое количество людей, надо ещё уметь определять, насколько сильно полученные результаты должны отличаться от заявленных 60%, чтобы иметь основания опровергать эту информацию.*

Статистика критерия

- Проверка гипотезы $H_0 : p = p_0$ о доле p признака в ГС проводится с помощью z-статистики.
- Статистика критерия для проверки гипотезы о доле равна $z = \frac{m - np_0}{\sqrt{np_0 q_0}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$, которая сходится по распределению к стандартной нормальной величине при $n \to \infty$.
- Здесь n - это объем выборки, m - число объектов в выборке с данным признаком или число "успехов" , p_0 - это предполагаемая доля признака в генеральной совокупности, $\hat{p} = \frac{m}{n}$ - это доля признака в выборке и $q_0 = 1 - p_0$.
- Несмотря на то, что значения статистики мы будем искать в таблице нормального распределения, надо не забывать, что статистика имеет лишь асимптотически нормальное распределение. То есть использовать её стоит только при больших выборках и дополнительно проверять условия $n\hat{p} \geq 5$ и $n\hat{q} \geq 5$.

Пример

Проверим гипотезу, что доля признака в ГС равна 0.1 на уровне значимости $\alpha = 0.05$*, против односторонних альтернатив* $p > 0.1$*. Объем выборки* $n = 100$ *и пусть выборочная доля составила* $\hat{p} = 0.2$*.*

Решение

- Сформулируем основную и альтернативную гипотезы:
$$H_0 : p = 0.1; \quad H_1 : p > 0.1$$
- Условия надёжности использования этих формул $n\hat{p} = 100 \cdot 0.2 = 20 \geq 5$ и $n\hat{q} = 100 \cdot 0.8 = 80 \geq 5$ выполнены.
- Критическая область является правосторонней. По таблице нормального распределения находим $1 - \alpha = 0.95$ и определяем критическое значение $z_{cr} = 1.65$. Критическая область имеет вид $(1.65; +\infty)$.
- Значение статистики критерия равно $z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.2 - 0.1}{\sqrt{0.1 \cdot 0.9/100}} \approx 3.3.$
- Вывод. Так как $z \in (1.65; +\infty)$, то основная гипотеза H_0 отвергается.
- Минимальный уровень значимости составляет $1 - z^{-1}(3.3) \approx 0.0005.$

Ответ: при данном уровне значимости и такой альтернативе гипотеза отвергается.

Самостоятельно

Пример

Допустим, вы думаете, что 25% студентов вашего университета ни разу не пропустили ни одной лекции. В выборочном опросе из 75 случайных студентов таких оказалось 15 человек. Проверьте свою гипотезу на уровне значимости $\alpha = 0.02$ *против двусторонних альтернатив. В ответ запишите минимальный уровень значимости с точностью до четвертого знака.*

Самостоятельно

Пример

Допустим, вы думаете, что 25% студентов вашего университета ни разу не пропустили ни одной лекции. В выборочном опросе из 75 случайных студентов таких оказалось 15 человек. Проверьте свою гипотезу на уровне значимости $\alpha = 0.02$ *против двусторонних альтернатив. В ответ запишите минимальный уровень значимости с точностью до четвертого знака.*

Решение

Самостоятельно означает, что сами делаете.

Содержание

Гипотеза о дисперсии

Наконец, перейдем к проверке гипотезы о равенстве дисперсии σ^2 некоторому значению σ_0^2. Это необходимо делать, когда приходится пользоваться предположениями о дисперсии.

- Чтобы проверить гипотезу $H_0 : \sigma^2 = \sigma_0^2$ надо рассмотреть статистику $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$, имеющую χ^2-распределения с $n-1$ степенью свободы.
- Обратите внимание, что в таблице в случае правосторонней критической области по значению $1-\alpha$ находят квантиль, обозначаемую χ_r^2, в случае левоосторонней по значению α находят χ_l^2, в случае двусторонней критической области в таблице по значениям $1-\frac{\alpha}{2}$ и $\frac{\alpha}{2}$ находят χ_r^2 и χ_l^2 соответственно.

Попробуйте сами

Пример

Допустим мы предполагаем, что стандартное отклонение в стобалльном рейтинге студентов равно 15. И решаем проверить это, оценив рейтинг знакомых. Получилась выборка объема $n = 20$*, у которой выборочная дисперсия равна 196.*
Наша задача эквивалентна проверке гипотезы о равенстве дисперсии 225. Уровень значимости возьмём $\alpha = 0.1$*, а альтернативы рассмотрим двусторонние.*

Решение

- Сформулируем основную и альтернативную гипотезы:
$$H_0 : \sigma^2 = 225; \quad H_1 : \sigma^2 \neq 225.$$
- Критическая область является двусторонней. По таблице χ^2-распределения находим $\alpha/2 = 0.05$, $1 - \alpha/2 = 0.95$, число степеней свободы $n - 1 = 20 - 1 = 19$ и определяем критические точки $\chi^2_l = 10.12$, $\chi^2_r = 30.14$. Критическая область имеет вид $(0; 10.12) \cup (30.14; +\infty)$.
- Значение статистики критерия равно $\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(20-1)\cdot 196}{225} \approx 16.55$.
- Вывод. Так как $\chi^2 \notin (0; 10.12) \cup (30.14; +\infty)$, то основная гипотеза H_0 не отвергается.
- Минимальный уровень значимости составляет $2 \cdot (1 - (\chi^2)^{-1}(16.55)) \approx 0.76$.

Ответ: при данном уровне значимости и такой альтернативе гипотеза не отвергается.

Содержание

Где и что почитать?

Тема. Статистические гипотезы. Уровень значимости, минимальный уровень значимости, критическая область, ошибки первого и второго рода, мощность критерия. Критерий отношения правдоподобия. Проверка гипотезы о среднем одной выборки. Проверка гипотезы о доле признака и дисперсии нормального распределения. ([Ф,Л], глава 15; [Т,М], глава 5).

- Фадеева Л. Н., Лебедев А. В., Теория вероятностей и математическая статистика: учебное пособие. - 2-е изд., перераб. и доп. - М.: Эксмо, 2010. - 496 с. – (Новое экономическое образование).
- Тюрин Ю. Н., Макаров А.А., Анализ данных на компьютере: учебное пособие. - 4-е изд., перераб. - М.: ИД Форум, 2008. - 368 с., ил. - (Высшее образование).