

Теория вероятностей и математическая статистика

Тестирование статистических гипотез II.

Глеб Карпов

ФКН ВШЭ

Гипотезы о доле бинарного признака

Иначе: о вероятности успеха в схеме Бернулли

Напоминание о выборочной доле

- Введём Y — случайную величину, показывающую количество положительных ответов в выборке, она имеет биномиальное распределение с $E[Y] = np$ и $Var[Y] = np(1-p)$.
- Тогда $\hat{p} = \frac{Y}{n}$ — случайная величина, называемая *выборочной долей* и показывающая долю положительных ответов к размеру выборки.
- Её свойства можно вывести из Y , а именно: $E[\hat{p}] = p$, $Var[\hat{p}] = \frac{p(1-p)}{n}$.
- Как следствие ИТМЛ, если $n > 30$, то:

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right), \quad \text{или} \quad \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0, 1)$$

Гипотезы о доле бинарного признака

Пример 1: правосторонний тест

Компания запустила новую телевизионную рекламную кампанию и хочет проверить, выше ли уровень узнаваемости бренда, чем было исторически ранее (40%).

- Выборка: $n = 200$ опрошенных клиентов
- Реализация выборочной доли: $\tilde{p} = 0.45$ (45% узнали бренд)

Гипотезы:

- $H_0 : p = 0.40$ (новая кампания не более эффективна)
- $H_1 : p > 0.40$ (новая кампания увеличивает узнаваемость бренда)

Гипотезы о доле бинарного признака

Правосторонний тест

Для тестирования $H_0 : p = p_0$ против $H_1 : p > p_0$:

- *Распределение при нулевой гипотезе*, когда мы безраздельно верим в то, что она верна:

$$\hat{p} \sim \mathcal{N}\left(p_0, \frac{p_0(1-p_0)}{n}\right)$$

- Построение критической (отклоняющей) области относительно уровня значимости:

$$P_{H_0}(\hat{p} > K) = \alpha$$

- Преобразование к стандартному нормальному распределению:

$$P_{H_0}(\hat{p} > K) = P_{H_0}\left(\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > \frac{K - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right) = P(Z > z_\alpha)$$

- Восстановление границы отклоняющей области:

$$K = p_0 + z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}}$$

- Правило принятия решения: Отклонить H_0 , если $\tilde{p} > K$

Гипотезы о доле бинарного признака

Пример 1: решение

- **Гипотезы:** $H_0 : p = 0.40$ против $H_1 : p > 0.40$ (правосторонний тест)
- **Данные:** $p_0 = 0.40$, $n = 200$, $\tilde{p} = 0.45$, $\alpha = 0.05$
- **Критическая область:** $z_{0.05} = 1.645$

$$K = 0.40 + 1.645 \cdot \sqrt{\frac{0.40 \cdot 0.60}{200}} \approx 0.457$$

- **Решение:** $\tilde{p} = 0.45 < K = 0.457$, поэтому не отклоняем H_0 .
- **Вывод:** Результаты тестирования не предоставляют достаточных статистически значимых оснований для отклонения нулевой гипотезы. Нет достаточных оснований утверждать, что новая рекламная кампания увеличила узнаваемость бренда выше 40%.

Гипотезы о математическом ожидании при неизвестной дисперсии

Предположения:

1. Случайная выборка $\mathcal{X} = \{X_1, \dots, X_n\}$ из неизвестного распределения с $\mu \equiv E[X_i]$, и $\sigma^2 \equiv \text{Var} [X_i]$,
2. Истинную дисперсию считаем неизвестной,
3. Либо: большая выборка ($n > 30$), либо исследуемый случайный процесс имеет нормальное распределение

Гипотезы о математическом ожидании при неизвестной дисперсии

Пример 2: левосторонний тест

Производственная компания внедрила новый производственный процесс и хочет проверить, уменьшилось ли среднее время производства детали относительно целевого значения 30 минут. Более быстрое производство указывало бы на успешное внедрение новой системы.

- Исследуемая случайная величина: X — время производства. Дисперсия X неизвестна.
- Предполагается, что времена производства следуют нормальному распределению.

Данные:

- Выборка: $n = 16$ различных произведенных деталей
- Выборочное среднее: $\bar{x} = 27.5$ минут
- Выборочное стандартное отклонение: $s = 5$ минут

Гипотезы:

- $H_0 : \mu = 30$
- $H_1 : \mu < 30$

Гипотезы о математическом ожидании при неизвестной дисперсии

Левосторонний тест

Для тестирования $H_0 : E[X] = \mu = \mu_0$ против $H_1 : E[X] = \mu < \mu_0$:

- *Распределение при нулевой гипотезе*, когда мы безраздельно верим в то, что она верна:

$$\bar{X} \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{n}\right), \quad \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t_{n-1} \text{ df}$$

- Построение границы критической области теста относительно установленного уровня значимости:

$$P_{H_0}(\bar{X} < K) = \alpha$$

- Преобразование к t -распределению:

$$P_{H_0}(\bar{X} < K) = P_{H_0}\left(\frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} < \frac{K - \mu_0}{\frac{S}{\sqrt{n}}}\right) = P(t < -t_{(n-1, \alpha)})$$

- Восстановление границы отклоняющей области:

$$K = \mu_0 - t_{(n-1, \alpha)} \frac{S}{\sqrt{n}}$$

- Правило принятия решения: Отклонить H_0 , если $\bar{x} < K$

Гипотезы о математическом ожидании при неизвестной дисперсии

Пример 2: решение

- **Гипотезы:** $H_0 : \mu = 30$ против $H_1 : \mu < 30$ (левосторонний тест)
- **Данные:** $\mu_0 = 30$, $s = 5$, $n = 16$, $\bar{x} = 27.5$, $\alpha = 0.05$
- **Критическая область:** $t_{15,0.05} = 1.753$

$$K = 30 - 1.753 \cdot \frac{5}{\sqrt{16}} = 27.81$$

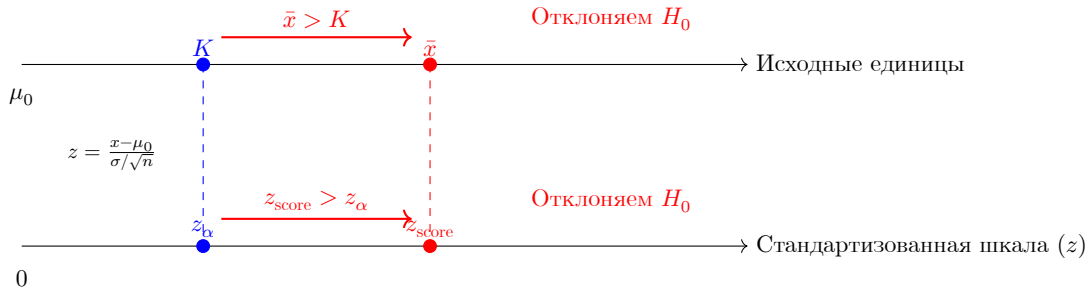
- **Решение:** $\bar{x} = 27.5 < K = 27.81$, поэтому отклоняем H_0 .
- **Вывод:** Имеются достаточно статистически значимые основания для утверждения, что новый производственный процесс уменьшил среднее время производства детали относительно целевых 30 минут. Консервативная гипотеза отклоняется в пользу альтернативной.

Test-score: два эквивалентных подхода

- Ранее мы сравнивали наблюдаемое значение (например, $\bar{x}$ или $\tilde{p}$) с границей принятия решения K в исходных единицах измерения.
- Альтернативный подход: преобразовать и наблюдаемое значение, и границу решения в стандартизованную шкалу.
- **Ключевая идея:** стандартизация — это монотонное преобразование, которое сохраняет отношения ”больше/меньше” между числами.
- Тип score зависит от распределения:
 - z_{score} — когда используем стандартное нормальное распределение (тестирование математических ожиданий при известной дисперсии или тест долей признаков)
 - t_{score} — когда используем t -распределение (неизвестная дисперсия)
- Если наблюдаемое значение $> K$ в исходных единицах, то score $>$ критическое значение в стандартизованной шкале.
- Если наблюдаемое значение $< K$ в исходных единицах, то score $<$ критическое значение в стандартизованной шкале.

Test-score: визуализация преобразования

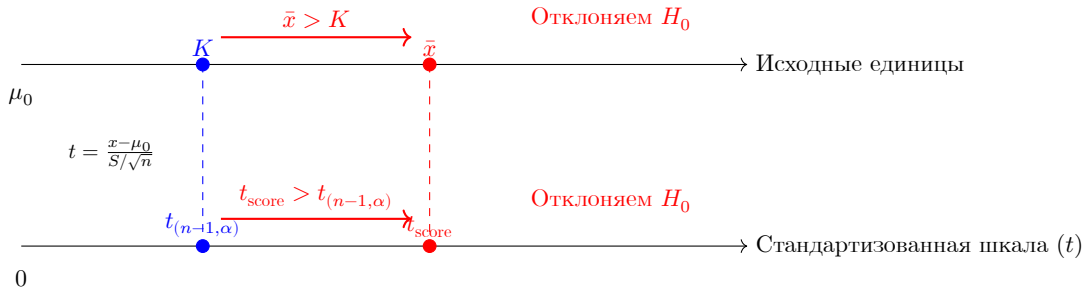
Пример: z -score для тестирования математического ожидания



Преобразование сохраняет отношение порядка. Решение одинаково в обеих шкалах.

Test-score: визуализация преобразования

Пример: t -score для тестирования математического ожидания



Преобразование сохраняет отношение порядка. Решение одинаково в обеих шкалах.

Test-score: преимущества подхода

- **Универсальность:** одна и та же критическая точка (z_α или $t_{df,\alpha}$) используется для всех тестов с соответствующим распределением.
- **Удобство:** не нужно вычислять границу решения K для каждого конкретного случая.
- **Эквивалентность:** оба подхода (сравнение в исходных единицах и сравнение score) дают одинаковый результат, так как стандартизация — монотонное преобразование.
- **Гибкость:** подход работает для разных типов тестов (односторонние, двусторонние) и разных распределений.

p-value: концепция

- **p-value** (p-значение) — это вероятность получить наблюдаемое значение статистики или более экстремальное значение *при условии, что нулевая гипотеза верна*.
- p-value показывает, насколько "необычны" полученные нами данные при предполагаемой верной нулевой гипотезе.
- **Правило решения через p-value:** Отклонить H_0 , если $p\text{-value} < \alpha$.
- Чем меньше p-value, тем сильнее свидетельство против нулевой гипотезы.
- p-value — это альтернативный способ принятия решения, эквивалентный сравнению score с критическим значением.

p-value: пример

Пример 3: ah, s, here we go again

Исследовательская команда изучает популяции жирафов. Исследователи верят, что средняя длина шеи жирафов огромная — около 20 метров. Однако в выборке было получено выборочное среднее 2.8 метров. Команда хочет проверить, действительно ли средняя длина шеи меньше того значения, в которое они верят.

- Случайная величина для анализа: X — длина шеи. Дисперсию X предположим известной и правдивой.
- Предполагается, что длины шеи следуют нормальному распределению.

Данные:

- Выборка: $n = 20$ жирафов
- Выборочное среднее: $\bar{x} = 2.8$ метров
- Известная дисперсия X : $\text{Var} [X] = 0.16$.

Гипотезы:

- $H_0 : \mu = 20$
- $H_1 : \mu < 20$

Решение через score

- Гипотезы: $H_0 : \mu = 20$ против $H_1 : \mu < 20$ (левосторонний тест)
- Данные: $\mu_0 = 20$, $\sigma = 0.4$, $n = 20$, $\bar{x} = 2.8$, $\alpha = 0.05$
- z -score:

$$z_{\text{score}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.8 - 20}{0.4/\sqrt{20}} = \frac{-17.2}{0.0894} \approx -192.4$$

- Критическое значение: $z_{0.05} = 1.645$
- Решение через score: $z_{\text{score}} = -192.4 < -z_{0.05} = -1.645$, поэтому отклоняем H_0 .

p-value: вычисление и интерпретация

- **p-value** для левостороннего теста:

$$p\text{-value} = P_{H_0}(Z < z_{\text{score}}) = P(Z < -192.4) \approx 0$$

- p-value практически равна нулю, что означает крайне малую вероятность получить такое экстремальное значение при верной нулевой гипотезе.
- **Решение через p-value:** $p\text{-value} \approx 0 < \alpha = 0.05$, поэтому отклоняем H_0 .
- **Вывод:** Средняя длина шеи жирафов статистически значимо меньше 20 метров. Наблюдаемое значение 2.8 метра крайне противоречит гипотезе о том, что средняя длина равна 20 метрам.
- **Эквивалентность:** Оба подхода (сравнение z_{score} с z_{α} и сравнение p-value с α) дают одинаковый результат.

Пример 4: время подготовки студентов

правосторонний t -тест

Случайная выборка времени (в часах), потраченного на подготовку к экзаменам за неделю:

28, 57, 42, 35, 61, 39, 55, 46, 49, 38.

Предполагаем нормальность распределения и неизвестную дисперсию.

- Найти $\bar{x}$ и s .
- Проверить $H_0 : \mu = 40$ против $H_1 : \mu > 40$.

Пример 4: время подготовки студентов

Решение (слайд 1 из 2)

- По данным: $n = 10$, $\bar{x} = 45$, $s \approx 10.54$.
- Гипотезы: $H_0 : \mu = 40$ против $H_1 : \mu > 40$, уровень значимости α .
- При H_0 :

$$T = \frac{\bar{X} - 40}{S/\sqrt{10}} \sim t_9.$$

- Граница критической области в исходных единицах:

$$\bar{X} > K, \quad K = 40 + t_{9,\alpha} \frac{S}{\sqrt{10}}.$$

- Подставляя числа:

$$K_{0.10} \approx 40 + 1.383 \cdot \frac{10.54}{\sqrt{10}} = 44.61, \quad K_{0.05} \approx 46.11, \quad K_{0.01} \approx 49.40.$$

- Сравнение с $\bar{x} = 45$: $45 > 44.61$, но $45 < 46.11$ и $45 < 49.40$.
- Вывод в исходной шкале: H_0 отклоняется только при $\alpha = 10\%$; при $\alpha = 5\%$ и $\alpha = 1\%$ не отклоняется.

Пример 4: время подготовки студентов

Решение (слайд 2 из 2)

- Наблюдаемое значение статистики:

$$t_{\text{score}} = \frac{45 - 40}{10.54/\sqrt{10}} \approx 1.50.$$

- Критические значения: $t_{9,0.10} \approx 1.383$, $t_{9,0.05} \approx 1.833$, $t_{9,0.01} \approx 2.821$.
- Сравнение: $1.50 > 1.383$, но $1.50 < 1.833$ и $1.50 < 2.821$.
- Вывод: H_0 отклоняется только при $\alpha = 10\%$; при $\alpha = 5\%$ и $\alpha = 1\%$ оснований для отклонения H_0 нет.