Прикладная статистика.
Проверка гипотез. Критерии согласия.

Леонид Иосипой

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

10 февраля 2021

- Повторение
- Проверка гипотез
- Критерии согласия

Проверка гипотез

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.

Проверка гипотез

Пример

Предположим, что кто-то подбросил 10 раз монетку, и в 8 случаях она упала гербом вверх. Можно ли считать эту монетку симметричной?

Пусть $X_1, \ldots, X_n \sim \mathbf{B_p}$.

$H_0: \; p = \frac{1}{2}$ (основная гипотеза).
$H_1: \; p \neq \frac{1}{2}$ (альтернативная гипотеза).

Как проверить гипотезу H_0 о том, что $p = 1/2$?

Проверка гипотез

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.

Проверка гипотез

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.

Проверка гипотез

В нашем примере в качестве статистики T можно взять

$$T(x_1, \ldots, x_n) = x_1 + \ldots + x_n.$$

Тогда гипотезе $H_0: \ p = 1/2$ противоречат значения, которые близки к 0 или n.

Более того,

- при верной H_0 имеет биномиальное распределение $\mathbf{B_{n,1/2}}$;
- при верной H_1 имеет биномиальное распределение $\mathbf{B_{n,p}}$, но с $p \neq 1/2$.

Проверка гипотез

Если значение T попало в область, имеющую при выполнении гипотезы H_0 малую вероятность, то можно заключить, что данные противоречат гипотезе H_0 в пользу альтернативы H_1.

Если значение T попало в область, имеющую при выполнении гипотезы H_0 большу́ю вероятность, то можно заключить, что данные не свидетельствуют против гипотезы H_0 в пользу альтернативы H_1.

Проверка гипотез

Формализация задачи:

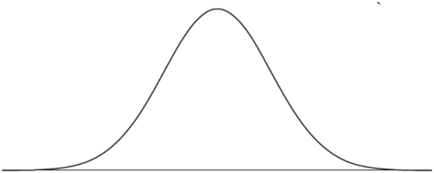
выборка: $\mathbf{X} = (X_1, \ldots, X_n),\ X_i \sim F$

нулевая гипотеза: $H_0:\ F \in \mathcal{F}_0$

альтернатива: $H_1:\ F \in \mathcal{F}_1,\ \mathcal{F}_1 \cap \mathcal{F}_0 = \varnothing$

статистика: $T(x_1, \ldots, x_n),\ \ T(\mathbf{X}) \sim G_0$ при H_0

$T(\mathbf{X}) \nsim G_0$ при H_1



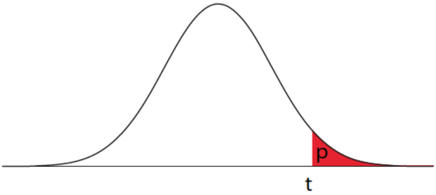
Проверка гипотез

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

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)



Проверка гипотез

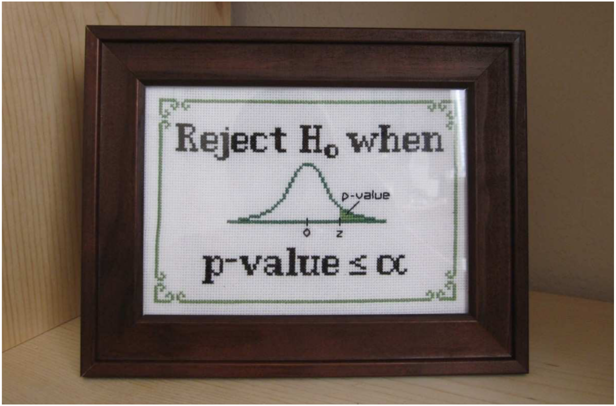
Достигаемый/Фактический уровень значимости (p-value) — это вероятность для статистики T при верной H_0 получить значение t или ещё более экстремальное.

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

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

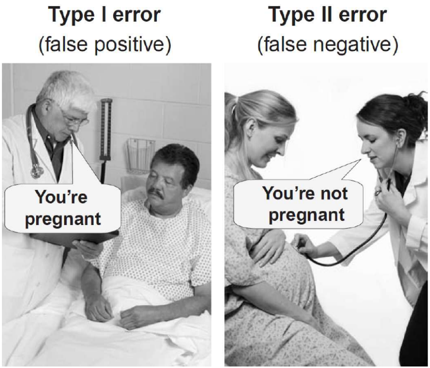
Нулевая гипотеза H_0 отвергается при $p(\mathbf{x}) \leq \alpha$, α — уровень значимости, который мы задаем.

Проверка гипотез



Проверка гипотез

	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



Проверка гипотез

Если величина p-value достаточно мала, то данные свидетельствуют против нулевой гипотезы H_0 в пользу альтернативы H_1.

Если величина p-value недостаточно мала, то данные не свидетельствуют против нулевой гипотезы H_0 в пользу альтернативы H_1.

При помощи инструмента проверки гипотез нельзя доказать справедливость нулевой гипотезы!

Проверка гипотез

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.

Проверка гипотез

Задача

Джеймс Бонд говорит, что предпочитает взболтанный мартини, но не смешанный. Проверим, так это или нет.

Проведём слепой тест: n раз предложим ему пару напитков и выясним, какой из двух он предпочитает.

Проверка гипотез

Выборка: $\mathbf{X} = (X_1, \dots, X_n)$, где $X_i \sim \mathbf{B_p}$.

Реализация выборки: $\mathbf{x} = (x_1, \dots, x_n)$ — это бинарный вектор длины n, где

- 1 — Джеймс Бонд выбрал взболтанный мартини
- 0 — Джеймс Бонд выбрал смешанный мартини

H_0: Д.Б. не различает два вида мартини, $p = 1/2$.

H_1: Д.Б. предпочитает взболтанный мартини, $p > 1/2$.

Проверка гипотез

Статистика: $T(x_1, \ldots, x_n) = x_1 + \ldots + x_n$.

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

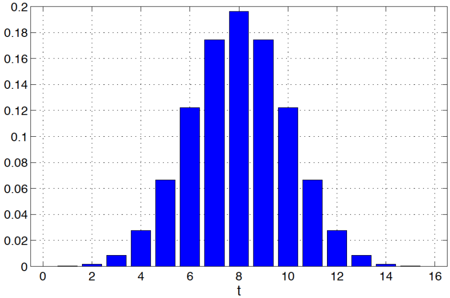
Какие значения T считаются экстремальными?

При альтернативе H_1 экстремальными являются большие значения t (они свидетельствуют против H_0 в пользу H_1).

Проверка гипотез

Если H_0 справедлива и Джеймс Бонд не различает два вида мартини, то T будет иметь распределение $\mathbf{B}_{n,1/2}$.

Пусть $n = 16$, тогда $\mathbf{B}_{n,1/2}$ будет иметь следующий вид:

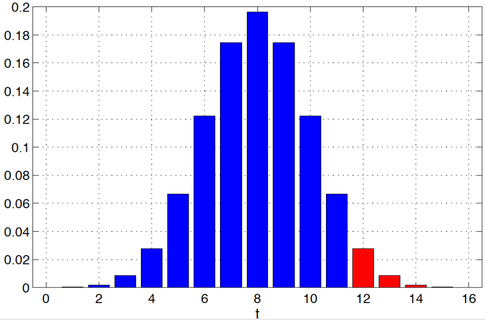


Проверка гипотез

Допустим, что $t = 12$, то есть в 12 случаях из 16 Джеймс Бонд выбрал взболтанный мартини.

Тогда достигаемый уровень значимости p-value равен:

$$\mathbb{P}(T(\mathbf{X}) \geq 12 | H_0) = \frac{2517}{65536} \approx 0.0384.$$



Проверка гипотез

Давайте поменяем альтернативу.

H_1: Джеймс Бонд предпочитает какой-то определённый вид мартини, но неизвестно какой, то есть $p \neq 1/2$.

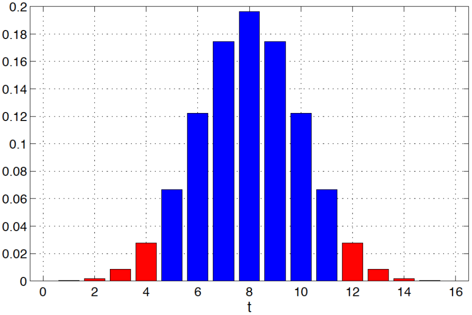
При такой альтернативе и большие, и маленькие значения t свидетельствуют против H_0 в пользу H_1.

Проверка гипотез

Допустим, что $t = 12$, то есть в 12 случаях из 16 Джеймс Бонд выбрал взболтанный мартини.

Тогда достигаемый уровень значимости p-value равен:

$$\mathbb{P}(T(\mathbf{X}) \geq 12 \text{ или } T(\mathbf{X}) \leq 4 | H_0) = \frac{5034}{65536} \approx 0.0768.$$



Проверка гипотез

Чем ниже достигаемый уровень значимости, тем сильнее данные свидетельствуют против нулевой гипотезы в пользу альтернативы.

Достигаемый уровень значимости нельзя интерпретировать как вероятность справедливости нулевой гипотезы!