Математическая статистика

(ФБМФ, ФЭФМ)

Проверка гипотез

20.04.2023

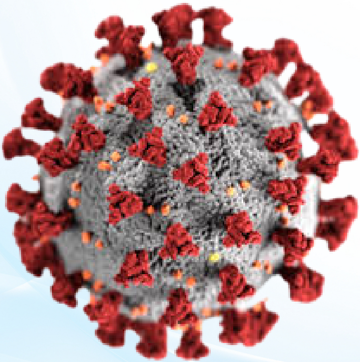
План занятия

- Проверка гипотез
  - Гипотезы
  - Критерии
  - Типы ошибок
  - Мощность

Проверка статистических гипотез

Дизайн эксперимента

Было



Стало

Картинки: https://ru.wikipedia.org/wiki/Коронавирусы, https://static.ngs.ru/news/2021/99/preview/982868ecec8f52e984781677e00b3ae70616830c_764_509.jpg

Хороши ли изменения?

Делим пользователей на две независимые группы:

1. Исследуемая — привитые;
2. Контрольная — плацебо.

Хороши ли изменения?

Делим пользователей на две независимые группы:

1. Исследуемая — привитые;
2. Контрольная — плацебо.

По результатам испытаний получаем, что в случае введения вакцины титр нейтрализующих антител выше в 40 раз. Или, например, среди привитых на 1.5% меньше заболевших.

Хороши ли изменения?

Делим пользователей на две независимые группы:

1. Исследуемая — привитые;
2. Контрольная — плацебо.

По результатам испытаний получаем, что в случае введения вакцины титр нейтрализующих антител выше в 40 раз. Или, например, среди привитых на 1.5% меньше заболевших.

Насколько это сопоставимо с погрешностями?

Хороши ли изменения?

Делим пользователей на две независимые группы:

1. Исследуемая — привитые;
2. Контрольная — плацебо.

По результатам испытаний получаем, что в случае введения вакцины титр нейтрализующих антител выше в 40 раз. Или, например, среди привитых на 1.5% меньше заболевших.

Насколько это сопоставимо с погрешностями?

Ответ дают статистические тесты.

Хороши ли изменения?

Делим пользователей на две независимые группы:

1. Исследуемая — привитые;
2. Контрольная — плацебо.

По результатам испытаний получаем, что в случае введения вакцины титр нейтрализующих антител выше в 40 раз. Или, например, среди привитых на 1.5% меньше заболевших.
Насколько это сопоставимо с погрешностями?

Ответ дают статистические тесты.

Основная гипотеза:

Возможность заболеть одинакова в обеих группах.

Хороши ли изменения?

Делим пользователей на две независимые группы:

1. Исследуемая — привитые;
2. Контрольная — плацебо.

По результатам испытаний получаем, что в случае введения вакцины титр нейтрализующих антител выше в 40 раз. Или, например, среди привитых на 1.5% меньше заболевших.

Насколько это сопоставимо с погрешностями?

Ответ дают статистические тесты.

Основная гипотеза:

Возможность заболеть одинакова в обеих группах.

Альтернативная гипотеза:

Привитые заболевают реже.

Хороши ли изменения?

Делим пользователей на две независимые группы:

1. Исследуемая — привитые;
2. Контрольная — плацебо.

По результатам испытаний получаем, что в случае введения вакцины титр нейтрализующих антител выше в 40 раз. Или, например, среди привитых на 1.5% меньше заболевших.

Насколько это сопоставимо с погрешностями?

Ответ дают статистические тесты.

Основная гипотеза:

Возможность заболеть одинакова в обеих группах.

Альтернативная гипотеза:

Привитые заболевают реже.

Если доля заболевших в первой группе сильно меньше,

то основную гипотезу следует отвергнуть.

Гипотезы

$X = (X_1, \ldots, X_n)$ — выборка.

$\mathscr{X}$ — множество возможных значений эл-тов выборки

$\mathscr{P}$ — множество распределений на $\mathscr{X}$

Гипотезы

$X = (X_1, ..., X_n)$ — выборка.

$\mathscr{X}$ — множество возможных значений эл-тов выборки

$\mathscr{P}$ — множество распределений на $\mathscr{X}$

$\mathsf{H}_0\colon\ \mathsf{P} \in \mathscr{P}_0$ — основная (нулевая) гипотеза;

$\mathsf{H}_1\colon\ \mathsf{P} \in \mathscr{P}_1$ — альтернативная гипотеза,

где $\mathscr{P}_0, \mathscr{P}_1 \subset \mathscr{P}$ и $\mathscr{P}_0 \cap \mathscr{P}_1 = \emptyset$.

Гипотезы

$X = (X_1, \dots, X_n)$ — выборка.

$\mathscr{X}$ — множество возможных значений эл-тов выборки

$\mathscr{P}$ — множество распределений на $\mathscr{X}$

$\mathsf{H}_0 \colon \ \mathsf{P} \in \mathscr{P}_0$ — основная (нулевая) гипотеза;

$\mathsf{H}_1 \colon \ \mathsf{P} \in \mathscr{P}_1$ — альтернативная гипотеза,

где $\mathscr{P}_0, \mathscr{P}_1 \subset \mathscr{P}$ и $\mathscr{P}_0 \cap \mathscr{P}_1 = \emptyset$.

В параметрическом подходе при $\mathscr{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$:

$$\mathsf{H}_0 \colon \theta \in \Theta_0;$$

$$\mathsf{H}_1 \colon \theta \in \Theta_1,$$

где $\Theta_0, \Theta_1 \subset \Theta$ и $\Theta_0 \cap \Theta_1 = \emptyset$.

Пример (вакцина)

$X = (X_1, \ldots, X_n)$ — результаты у тех, кому ввели вакцину.

$Y = (Y_1, \ldots, Y_m)$ — результаты у тех, кому ввели плацебо.

Пример (вакцина)

$X = (X_1, \dots, X_n)$ — результаты у тех, кому ввели вакцину.

$Y = (Y_1, \dots, Y_m)$ — результаты у тех, кому ввели плацебо.

$\mathsf{H}_0:$ вакцина не отличается от плацебо;

$\mathsf{H}_1:$ вакцина эффективнее плацебо.

Критерии

$X = (X_1, ..., X_n)$ — выборка.

$\mathscr{X}$ — множество возможных значений эл-тов выборки.

Критерии

$X = (X_1, ..., X_n)$ — выборка.

$\mathscr{X}$ — множество возможных значений эл-тов выборки.

Множество $S \subset \mathscr{X}$ называется критерием для проверки H_0 *vs*. H_1, если правило отвержения H_0 выглядит следующим образом

$$\mathsf{H}_0 \text{ отвергается } \Longleftrightarrow X \in S.$$

Критерии

$X = (X_1, \ldots, X_n)$ — выборка.

$\mathscr{X}$ — множество возможных значений эл-тов выборки.

Множество $S \subset \mathscr{X}$ называется **критерием** для проверки H_0 *vs.* H_1, если правило отвержения H_0 выглядит следующим образом

$$\mathsf{H}_0 \text{ отвергается } \Longleftrightarrow X \in S.$$

Пример:

$S = \{x \in \mathscr{X} \mid T(x) \geqslant c\}$ — критерий.

Тогда H_0 отвергается $\Longleftrightarrow T(X) \geqslant c$,

$T(X)$ — **статистика критерия**;

c — **критическое значение**.

Критерии

$X = (X_1, ..., X_n)$ — выборка.

$\mathscr{X}$ — множество возможных значений эл-тов выборки.

Множество $S \subset \mathscr{X}$ называется критерием для проверки H_0 *vs.* H_1, если правило отвержения H_0 выглядит следующим образом

$$\mathsf{H}_0 \text{ отвергается } \Longleftrightarrow X \in S.$$

Пример:

$S = \{x \in \mathscr{X} \mid T(x) \geqslant c\}$ — критерий.

Тогда H_0 отвергается $\Longleftrightarrow T(X) \geqslant c$,

$T(X)$ — статистика критерия;

c — критическое значение.

$T(X) = \overline{X} - \overline{Y}$,

Вакцина эффективнее плацебо

$\Longleftrightarrow \overline{X} - \overline{Y} \geqslant 10$.

Критерии

$X = (X_1, ..., X_n)$ — выборка.

$\mathscr{X}$ — множество возможных значений эл-тов выборки.

Множество $S \subset \mathscr{X}$ называется критерием для проверки H_0 *vs*. H_1, если правило отвержения H_0 выглядит следующим образом

$$\mathsf{H}_0 \text{ отвергается } \iff X \in S.$$

Пример:

$S = \{x \in \mathscr{X} \mid T(x) \geqslant c\}$ — критерий.

Тогда H_0 отвергается $\iff T(X) \geqslant c$,

$T(X)$ — статистика критерия;

c — критическое значение.

$T(X) = \overline{X} - \overline{Y}$,

Вакцина эффективнее плацебо

$\iff \overline{X} - \overline{Y} \geqslant 10.$

В парам. подходе критерий н-ся двусторонним, если $\mathsf{H}_0 : \theta = \theta_0$ *vs*. $\mathsf{H}_1 : \theta \neq \theta_0$*,*

и односторонним если $\mathsf{H}_0 : \theta = \theta_0$ *vs*. $\mathsf{H}_1 : \theta > \theta_0$ *и* $\mathsf{H}_0 : \theta = \theta_0$ *vs*. $\mathsf{H}_1 : \theta < \theta_0$.

Критерии (продолжение)

Часто критерий имеет вид $S = \{T(x) \geqslant c_\alpha\}$,
где $T(X)$ — статистика критерия.

Критерии (продолжение)

Часто критерий имеет вид $S = \{T(x) \geqslant c_\alpha\}$,
где $T(X)$ — статистика критерия.

α выбирается **ДО** эксперимента,

c_α вычисляется из условия $\mathsf{P}_0(T(X) > c_\alpha) \leqslant \alpha$.

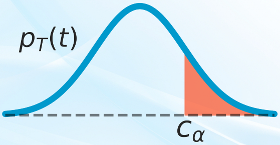
Критерии (продолжение)

Часто критерий имеет вид $S = \{T(x) \geqslant c_\alpha\}$,
где $T(X)$ — статистика критерия.

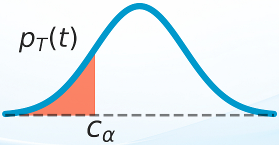
α выбирается **ДО** эксперимента,

c_α вычисляется из условия $\mathrm{P}_0(T(X) > c_\alpha) \leqslant \alpha$.

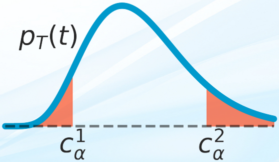
$S = \{T(x) > c_\alpha\}$



$S = \{T(x) < c_\alpha\}$



$S = \{|T(x)| > c_\alpha\}$



Замечание. Выбирать α после эксперимента неправильно.
Так можно подогнать результат под желаемый.
"*Статистика может доказать что угодно, даже истину.*"

Результаты тестирования

При тестировании гипотез может быть два результата:

Результаты тестирования

При тестировании гипотез может быть два результата:

1. $X \in S \Longrightarrow \mathrm{H}_0$ отвергается.

Результаты тестирования

При тестировании гипотез может быть два результата:

1. $X \in S \Longrightarrow \mathrm{H}_0$ отвергается.

   $\overline{X} - \overline{Y} \geqslant 10 \Longrightarrow$ вакцина эффективнее плацебо.

Результаты тестирования

При тестировании гипотез может быть два результата:

1. $X \in S \Longrightarrow \mathrm{H}_0$ отвергается.

   $\overline{X} - \overline{Y} \geqslant 10 \Longrightarrow$ вакцина эффективнее плацебо.

   Результат проверки гипотез является статистически значимым.

Результаты тестирования

При тестировании гипотез может быть два результата:

1. $X \in S \Longrightarrow \mathrm{H}_0$ отвергается.

   $\overline{X} - \overline{Y} \geqslant 10 \Longrightarrow$ вакцина эффективнее плацебо.

   Результат проверки гипотез является статистически значимым.

2. $X \notin S \Longrightarrow \mathrm{H}_0$ не отвергается.

Результаты тестирования

При тестировании гипотез может быть два результата:

1. $X \in S \Longrightarrow \mathrm{H}_0$ отвергается.

   $\overline{X} - \overline{Y} \geqslant 10 \Longrightarrow$ вакцина эффективнее плацебо.

   Результат проверки гипотез является статистически значимым.

2. $X \notin S \Longrightarrow \mathrm{H}_0$ не отвергается.

   $\overline{X} - \overline{Y} < 10 \Longrightarrow$ нельзя сказать, что вакцина эффект. плацебо

Результаты тестирования

При тестировании гипотез может быть два результата:

1. $X \in S \Longrightarrow \mathrm{H}_0$ отвергается.

   $\overline{X} - \overline{Y} \geqslant 10 \Longrightarrow$ вакцина эффективнее плацебо.

   Результат проверки гипотез является статистически значимым.

2. $X \notin S \Longrightarrow \mathrm{H}_0$ не отвергается.

   $\overline{X} - \overline{Y} < 10 \Longrightarrow$ нельзя сказать, что вакцина эффект. плацебо

   Результат проверки гипотез является стат. не значимым.

Результаты тестирования

При тестировании гипотез может быть два результата:

1. $X \in S \Longrightarrow \mathrm{H}_0$ отвергается.

   $\overline{X} - \overline{Y} \geqslant 10 \Longrightarrow$ вакцина эффективнее плацебо.

   Результат проверки гипотез является статистически значимым.

2. $X \notin S \Longrightarrow \mathrm{H}_0$ не отвергается.

   $\overline{X} - \overline{Y} < 10 \Longrightarrow$ нельзя сказать, что вакцина эффект. плацебо

   Результат проверки гипотез является стат. не значимым.

   Внимание!!! Неправильно говорить "H_0 принимается" — отсутствие доказательств несправедливости H_0 не есть доказательство ее справедливости.

Презумпция невиновности

Реализация выборки — конечный объем информации.

Распределение – бесконечномерный объект.

Мы имеем какой-то набор "фактов" о распределении.

Презумпция невиновности

Реализация выборки — конечный объем информации.

Распределение – бесконечномерный объект.

Мы имеем какой-то набор "фактов" о распределении.

Обвиняемый считается невиновным до тех пор, пока его вина в совершенном преступлении не будет доказана в установленном законом порядке.

Презумпция невиновности

Реализация выборки — конечный объем информации.

Распределение – бесконечномерный объект.

Мы имеем какой-то набор "фактов" о распределении.

Обвиняемый считается невиновным до тех пор, пока его вина в совершенном преступлении не будет доказана в установленном законом порядке.

Обвиняемый	P — неизвестное распределение
Невиновность	$\mathrm{P} \in \mathscr{P}_0$ — основная гипотеза
Виновность	$\mathrm{P} \in \mathscr{P}_1$ — альтернативная гипотеза
Совершенные действия	$X = (X_1, ..., X_n)$ — выборка
Факты	$T(X)$ — статистика критерия
Доказательство	Справедливость утверждения $X \in S$

Типы ошибок

	H_0 верна	H_0 не верна
H_0 не отвергается	OK	ошибка II рода
H_0 отвергается	ошибка I рода	OK

Типы ошибок

	H_0 верна	H_0 не верна
H_0 не отвергается	OK	ошибка II рода
H_0 отвергается	ошибка I рода	OK

Ошибка I рода: признали эффективной плохую вакцину.

Ошибка II рода: признали неэффективной хорошую вакцину.

Типы ошибок

	H_0 верна	H_0 не верна
H_0 не отвергается	OK	ошибка II рода
H_0 отвергается	ошибка I рода	OK

Ошибка I рода: признали эффективной плохую вакцину.

Ошибка II рода: признали неэффективной хорошую вакцину.

$$\mathsf{P}(I_S) = \sup_{\mathsf{P} \in \mathscr{P}_0} \mathsf{P}(X \in S) \text{ — вероятность ошибки I рода}$$

$$\mathsf{P}(II_S) = \sup_{\mathsf{P} \in \mathscr{P}_1} \mathsf{P}(X \notin S) \text{ — вероятность ошибки II рода}$$

Типы ошибок

	H_0 верна	H_0 не верна
H_0 не отвергается	OK	ошибка II рода
H_0 отвергается	ошибка I рода	OK

Ошибка I рода: признали эффективной плохую вакцину.

Ошибка II рода: признали неэффективной хорошую вакцину.

$$\mathsf{P}(I_S) = \sup_{\mathsf{P} \in \mathscr{P}_0} \mathsf{P}(X \in S) \text{ — вероятность ошибки I рода}$$

$$\mathsf{P}(II_S) = \sup_{\mathsf{P} \in \mathscr{P}_1} \mathsf{P}(X \notin S) \text{ — вероятность ошибки II рода}$$

Решается задача

$$\begin{cases} \mathsf{P}(I_S) \leqslant \alpha \\ \mathsf{P}(II_S) \to \min\limits_S \end{cases}$$

Типы ошибок

	H_0 верна	H_0 не верна
H_0 не отвергается	OK	ошибка II рода
H_0 отвергается	ошибка I рода	OK

Ошибка I рода: признали эффективной плохую вакцину.

Ошибка II рода: признали неэффективной хорошую вакцину.

$$\mathsf{P}(I_S) = \sup_{\mathsf{P}\in\mathscr{P}_0} \mathsf{P}(X \in S)$$ — вероятность ошибки I рода

$$\mathsf{P}(II_S) = \sup_{\mathsf{P}\in\mathscr{P}_1} \mathsf{P}(X \notin S)$$ — вероятность ошибки II рода

Решается задача

$$\begin{cases} \mathsf{P}(I_S) \leqslant \alpha \\ \mathsf{P}(II_S) \to \min\limits_S \end{cases}$$

В наихудшем случае плохая вакцина будет признана эффективной с вер-тью не более α.

Типы ошибок

	H_0 верна	H_0 не верна
H_0 не отвергается	OK	ошибка II рода
H_0 отвергается	ошибка I рода	OK

Ошибка I рода: признали эффективной плохую вакцину.

Ошибка II рода: признали неэффективной хорошую вакцину.

$$\mathsf{P}(I_S) = \sup_{\mathsf{P}\in\mathscr{P}_0} \mathsf{P}(X \in S)$$ — вероятность ошибки I рода

$$\mathsf{P}(II_S) = \sup_{\mathsf{P}\in\mathscr{P}_1} \mathsf{P}(X \notin S)$$ — вероятность ошибки II рода

Решается задача

$$\begin{cases} \mathsf{P}(I_S) \leqslant \alpha \\ \mathsf{P}(II_S) \to \min\limits_S \end{cases}$$

В наихудшем случае плохая вакцина будет признана эффективной с вер-тью не более α.

Критерий S имеет уровень значимости α, если $\mathsf{P}(I_S) \leqslant \alpha$. Обычно $\alpha = 0.05$

Типы ошибок

	H_0 верна	H_0 не верна
H_0 не отвергается	OK	ошибка II рода
H_0 отвергается	ошибка I рода	OK

Ошибка I рода: признали эффективной плохую вакцину.

Ошибка II рода: признали неэффективной хорошую вакцину.

$$\mathsf{P}(I_S) = \sup_{\mathsf{P} \in \mathscr{P}_0} \mathsf{P}(X \in S) \text{ — вероятность ошибки I рода}$$

$$\mathsf{P}(II_S) = \sup_{\mathsf{P} \in \mathscr{P}_1} \mathsf{P}(X \notin S) \text{ — вероятность ошибки II рода}$$

Решается задача

$$\begin{cases} \mathsf{P}(I_S) \leqslant \alpha \\ \mathsf{P}(II_S) \to \min\limits_S \end{cases}$$

В наихудшем случае плохая вакцина будет признана эффективной с вер-тью не более α.

Критерий S имеет уровень значимости α, если $\mathsf{P}(I_S) \leqslant \alpha$. Обычно $\alpha = 0.05$

Замечание. Выбирать α после эксперимента неправильно.
Так можно подогнать результат под желаемый.

Мощность

Обычно альтернатива сложная:

$H_0: \theta = \theta_0$ *vs*. $H_1: \theta > \theta_0$

H_0: X_i имеет норм. распр. *vs*. H_1: распр. X_i отличается от норм.

Мощность

Обычно альтернатива сложная:

$\mathrm{H}_0 : \theta = \theta_0$ *vs.* $\mathrm{H}_1 : \theta > \theta_0$

$\mathrm{H}_0 : X_i$ имеет норм. распр. *vs.* $\mathrm{H}_1 :$ распр. X_i отличается от норм.

Для сравнения критериев на распределениях, соответствующих альтернативной гипотезе, вводится функция *мощности*:

$\beta_S(\mathsf{P}) = \mathsf{P}(X \in S)$, где $\mathsf{P} \in \mathscr{P}_1$.