ПРИКЛАДНАЯ СТАТИСТИКА

Гипотеза о равенстве средних и доверительные интервалы

Ульянкин Ппилиф [*]

Поговорим про мощность различных процедур. Мы с вами обсудили, что гипотезу о равенстве средних можно проверить с помощью следующей процедуры.

Процедура 1

а. Собираем выборки $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_n$;

б. Находим значение статистики

$$z_{obs} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}};$$

в. Говорим, что по ЦПТ $z_{obs} \overset{asy}{\sim} N(0, 1)$;

г. Находим критическое значение $z_{1-\frac{\alpha}{2}}$;

д. Если мы видим, что $|z_{obs}| < z_{1-\frac{\alpha}{2}}$, мы говорим, что гипотеза не отвергается.

Ту же саму гипотезу можно попробовать проверить с помощью другого алгоритма, основанного на доверительных интервалах.

[*] https://github.com/FUlyankin/matstat_lec

Процедура 2

а. Собираем выборки $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_n$;

б. Находим $\bar{x}$ и $\bar{y}$;

в. Пользуясь ЦПТ и зная, что $\bar{x} \overset{asy}{\sim} N\left(\mu_1, \frac{s_x^2}{n_x}\right)$ и $\bar{y} \overset{asy}{\sim} N\left(\mu_2, \frac{s_y^2}{n_y}\right)$ строим для μ_1 и μ_2 доверительные интервалы;

г. Если доверительные интервалы пересеклись, говорим, что гипотеза не отвергается.

Вроде бы вторая процедура выглядит довольно естественно, однако ей никто не пользуется. Дело в том, что для одинаковых ошибок первого рода, α, ошибка второго рода, β, для процедуры, основанной на доверительных интервалах, окажется выше. Задание состоит в том, чтобы это увидеть.

Для простоты будем дальше предполагать, что $\bar{x} > \bar{y}$. Также будем считать, что обе дисперсии известны и равны единице, $\sigma_x^2 = \sigma_y^2 = 1$. Объёмы выборок одинаковы, $n_x = n_y = n$. Альтернатива односторонняя, то есть наблюдаемое значение статистики нужно искать как $z_{1-\alpha}$.

Решение

Гипотеза о равенстве средних не отвергается, если

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} < z_{1-\alpha}.$$

Упростим выражение

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{2}{n}}} < z_{1-\alpha}.$$

Найдём ошибку второго рода

$$\mathbb{P}\left(H_0 \mid H_a\right) = \mathbb{P}\left(\frac{\bar{x} - \bar{y}}{\sqrt{\frac{2}{n}}} < z_{1-\alpha} \mid \mu_1 \neq \mu_2\right) =$$

$$= \mathbb{P}\left(\frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{2}{n}}} < z_{1-\alpha} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{2}{n}}} \mid \mu_1 \neq \mu_2\right) = \mathbb{P}\left(\mathcal{N}(0,1) < z_{1-\alpha} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{2}{n}}}\right) =$$

$$= \Phi\left(z_{1-\alpha} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{2}{n}}}\right).$$

Построим доверительные интервалы для средних:

$$\bar{x} \pm z_{1-\alpha} \cdot \sqrt{\frac{1}{n}}$$
$$\bar{y} \pm z_{1-\alpha} \cdot \sqrt{\frac{1}{n}}$$

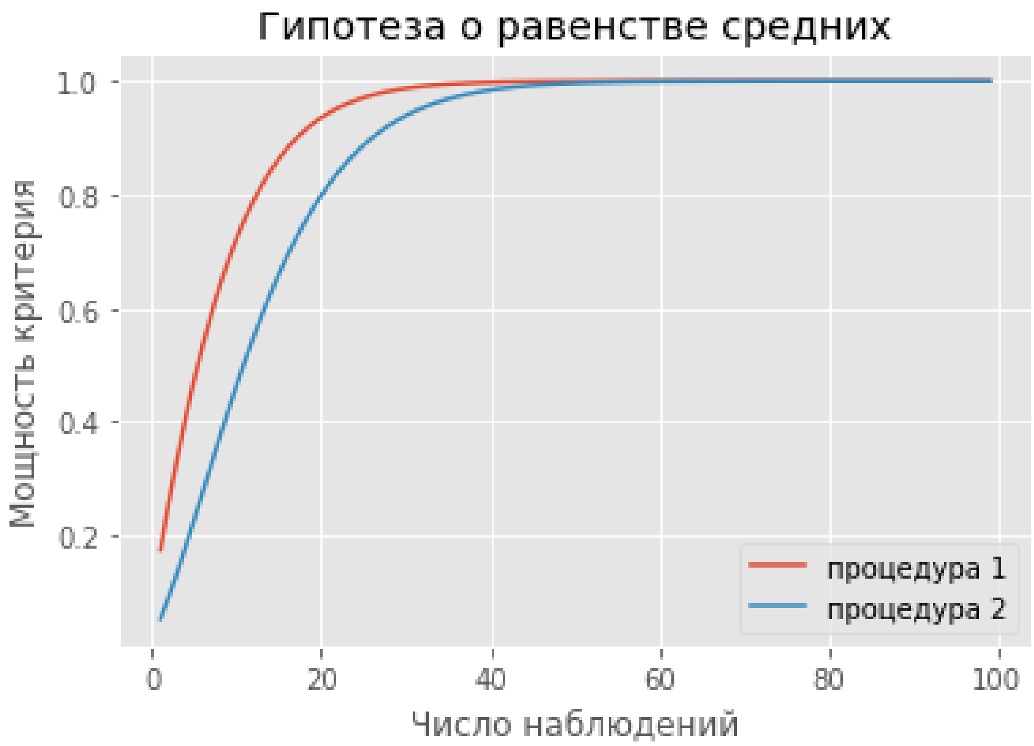
Если доверительные интервалы пересекаются, гипотеза не отвергается, то есть:

$$\begin{aligned}
&\bar{x} - z_{1-\alpha} \cdot \sqrt{\frac{1}{n}} < \bar{y} + z_{1-\alpha} \cdot \sqrt{\frac{1}{n}} \\
&\bar{x} - \bar{y} < 2 \cdot z_{1-\alpha} \cdot \sqrt{\frac{1}{n}} \\
&\frac{\bar{x} - \bar{y}}{\sqrt{\frac{1}{n}}} < 2 \cdot z_{1-\alpha} \\
&\frac{\bar{x} - \bar{y}}{\sqrt{\frac{2}{n}}} < \sqrt{2} \cdot z_{1-\alpha}
\end{aligned}$$

Напомню, что для простоты мы считаем, что $\bar{x} > \bar{y}$. Найдём ошибку второго рода

$$\begin{aligned}
\mathbb{P}\left(H_0 \mid H_a\right) = \mathbb{P}\left(\frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{2}{n}}} < \sqrt{2} \cdot z_{1-\alpha} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{2}{n}}} \mid \mu_1 \neq \mu_2\right) = \\
= \mathbb{P}\left(\mathcal{N}(0,1) < \sqrt{2} \cdot z_{1-\alpha} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{2}{n}}} \mid \mu_1 \neq \mu_2\right) = \Phi\left(\sqrt{2} \cdot z_{1-\alpha} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{2}{n}}}\right).
\end{aligned}$$

Чем больше наблюдений есть в нашем распоряжении, тем больше мощность теста. Нарисуем функции, описывающие мощность в координатах число наблюдений, мощность. На картинке видно, что первая процедура стабильно выигрывает у второй. При бесконечном числе наблюдений разницы не будет, так как мы всегда сможем идеально отделить две альтернативы друг от друга.


Гипотеза о равенстве средних
Мощность критерия
1.0
0.8
0.6
0.4
0.2
0
20
40
60
80
100
Число наблюдений
процедура 1
процедура 2