

Семинар 4

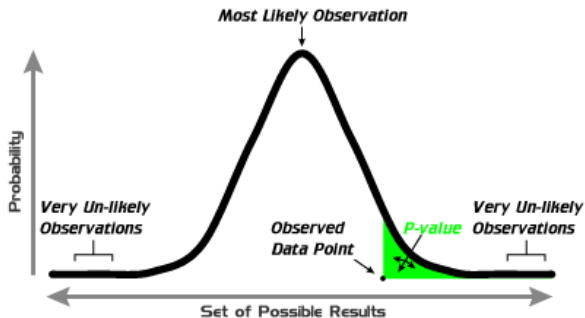
Параметрические критерии

Грауэр Л.В., Архипова О.А.

Санкт-Петербург, 2015

Про p-value

pvalue - вероятность получить более 'экстремальное' значение статистики, чем наблюдаемое (observed) при условии истинности нулевой гипотезы (H_0)



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result arising by chance

Задание 1.

Распределение Т-статистики

Выбрать параметры для двух нормальных распределений с разными мат. ож. ($\mu_0 = a_0$ и $\mu_1 = a_1$) и одинаковыми дисперсиями. 1000 раз сгенерировать по две выборки размера $N=30$. Для каждой выборки посчитать Т-статистику при следующей $H_0 : \mu = a_0$.

На одном графике построить графики распределения Т-статистик и распределение Стьюдента с $N-1$ df.

$$t = \frac{\bar{x} - a_0}{\hat{s}} \sqrt{n}$$

Задание 1.

функции в R

- `rnorm(60, mean = 0, sd = 1)` - генерация случайных нормальных величин
- `var` - исправленная дисперсия, `sd` - стандартное отклонение
- `x<-seq(-10,10,0.01)` - вектор от -10 до 10 с шагом 0.01
- `h1<-hist(sample, add=T, xlim=c(a,b),ylim=c(a,b))`
- `lines(x,dt(x,df=n-1),col="green")` , `dt(x,n-1)` -плотность распределения Стьюдента
`lines(x,dnorm(x,0,1),col="red")`

Задание 1.

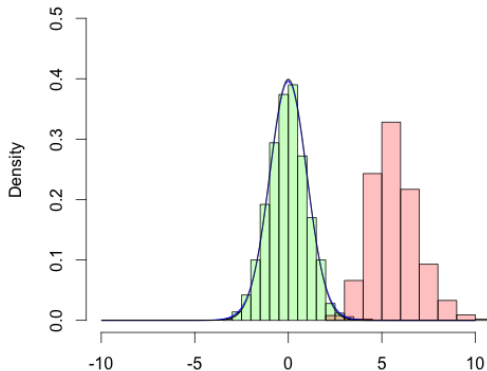
Пример результата

Распределение Т-статистики для выборки, где H_0 верна

Ст. Нормальное распределение

Распределение Стьюдента

Распределение Т-статистики для выборки, где H_0 не верна



Задание 2.

Ошибка первого и второго рода

Выбрать гипотезы H_0 и H_1 . выбрать уровень значимости α и мощность. 1000 раз смоделировать выборку объема $N=100$, для которой верна H_0 , посчитать α . 1000 раз смоделировать выборку объема $N=100$, для которой верна H_1 . Посчитайте мощность. С помощью моделирования определите объем выборки, необходимый для достижения выбранной мощности. Построить графики зависимости мощности от N и от SD .

Функции в R:

- `rnorm(n,mu1,sd)`
- `t.test(sample,alternative=c("t"),mu=mu1)`
- `t.test(sample,alternative=c("t"),mu=mu1)$p.value`

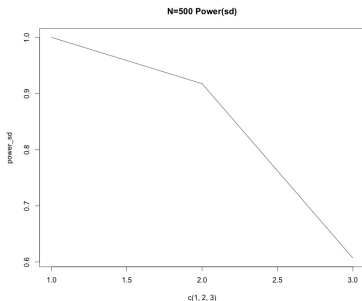
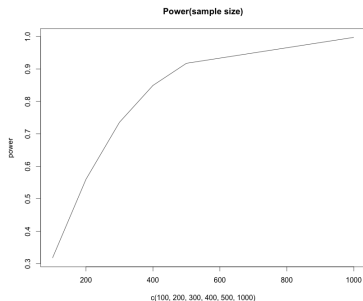
Задание 2.

Пример

$\mu_1=5$, $\mu_2= 5.3$, $sd = 2$, $\alpha=0.05$, $y=0.9$

Для $N=500$:

Если H_0 верна: Из 1000 запусков 954 раза приняли верное решение ($\alpha = 0.046$). Если верна H_1 : Из 1000 запусков 921 раз приняли верное решение ($y=0.92$).



Для проверки моделирования:

```
power.t.test(N,diff,sd,sig.level=0.05,type=c("one.sample"))$power
```

Задание 3.

Критерий Вальда

Проверить гипотезу о ДИСПЕРСИИ нормального распределения.
Выбрать параметры для нормальных распределений - $(\mu, \sigma_1), (\mu, \sigma_2)$.
Выбрать α и β . С помощью критерия Вальда найти N,
удовлетворяющее этим условиям. Проверить с помощью
моделирования ошибку первого и второго рода.

$$c_0 = \frac{\beta}{(1-\alpha)}$$

$$c_1 = \frac{(1-\beta)}{\alpha}$$

Функции в R:

prod(x) - произведение элементов x.

while (condition) {...}

Задание 3.

Пример

$\alpha = 0.05$, $\beta = 0.1$, $\mu = 5$,

$\sigma_0 = 2$, $\sigma_1 = 2.3$,

H_0 верна: 955 раз из 1000 приняли правильное решение. ($\alpha = 0.045$)

H_1 верна: 899 раз из 1000 приняли правильное решение. ($\gamma = 0.89$)

Среднее N : 112.89