

## Гипотезы о таксопарках

---

### Опубликовал

sobodv

### Автор или источник

sobopedia

### Предмет

Математическая Статистика (/Subjects/Details?id=5)

### Тема

Теория проверки статистических гипотез (/Topics/Details?id=35)

### Раздел

Проверка гипотез по поводу математического ожидания и дисперсии (/SubTopics/Details?id=125)

### Дата публикации

26.05.2019

### Дата последней правки

26.05.2019

### Последний вносивший правки

sobodv

### Рейтинг

☆☆

## Условие

Остап помышляет о карьере таксиста. В городе работают три таксопарка, в каждом из которых трудоустроены более тысячи таксистов. Остап решил случайным образом опросить таксистов из этих таксопарков об их среднечасовых заработках. Предположим, что заработки случайно взятых таксистов независимы и нормально распределены. В первом таксопарке среднечасовые заработки случайным образом опрошенных таксистов составили 250, 320, 200, 500 и 310 рублей. Для второго таксопарка соответствующие значения составили 600, 180, 150 и 250 рублей. Наконец, в третьем таксопарке он собрал данные о среднечасовых заработках в размерах 500, 100 и 300 рублей.

1. Знакомый таксист сказал Остапу, что **во втором** таксопарке ожидаемый среднечасовой заработок составляет 300 рублей. Помогите Остапу оценить достоверность этой информации, проверив соответствующую гипотезу на уровне значимости 0.05, против альтернативы о том, что ожидаемый среднечасовой заработок **a)** меньше чем 300 рублей, **b)** не равен 300 рублей, **c)** больше 300 рублей. Для каждого из этих случаев рассчитайте p-value - минимальный уровень значимости, при котором нулевая гипотеза не отвергается. Для пункта **b** также рассчитайте мощность критерия.

## Решение

Для начала формализуем условие задачи. Имеются три независимых выборки  $X = (X_1, \dots, X_5)$ ,  $Y = (Y_1, \dots, Y_4)$  и  $V = (V_1, \dots, V_3)$  из распределений  $N(\mu_X, \sigma_X^2)$ ,  $N(\mu_Y, \sigma_Y^2)$  и  $N(\mu_V, \sigma_V^2)$ . соответственно Реализации этих выборок составляют  $x = (250, 320, 200, 500, 310)$ ,  $y = (600, 180, 150, 250)$  и  $v = (500, 100, 300)$  соответственно.

Рассчитаем следующие промежуточные значения:

$$(\hat{\mu}_X | (X = x)) = (\bar{X} | (X = x)) = \bar{x} = \frac{250 + 320 + 200 + 500 + 310}{5} = 316$$

$$(\hat{\mu}_Y | (Y = y)) = (\bar{Y} | (Y = y)) = \bar{y} = \frac{600 + 180 + 150 + 250}{4} = 295$$

$$(\hat{\mu}_V | (V = v)) = (\bar{V} | (V = v)) = \bar{v} = \frac{500 + 100 + 300}{3} = 300$$

$$(\hat{\sigma}_X^2 | (X = x)) = \frac{(250 - 316)^2 + (320 - 316)^2 + (200 - 316)^2 + (500 - 316)^2 + (310 - 316)^2}{5 - 1} = 12930$$

$$(\hat{\sigma}_Y^2 | (Y = y)) = \frac{(600 - 295)^2 + (180 - 295)^2 + (150 - 295)^2 + (250 - 295)^2}{4 - 1} = 43100$$

$$(\hat{\sigma}_V^2 | (V = v)) = \frac{(500 - 300)^2 + (100 - 300)^2 + (300 - 300)^2}{3 - 1} = 40000$$

1. Необходимо проверить, на уровне значимости  $\alpha = 0.05$ , гипотезу о том, что  $H_0 : \mu_X = 300$  против альтернатив а)  $H_1 : \mu_X \leq 300$ , б)  $H_1 : \mu_X \neq 300$  и в)  $H_1 : \mu_X \geq 300$ . В данном случае речь идет о проверке гипотезы о математическом ожидании при неизвестной дисперсии.

Начнем с рассмотрения пункта **б**. Исходя из выбранного уровня значимости критическая область тестовой статистики может быть задана как  $\mathcal{T} = R \setminus [t_{4-1}^{0.025}, t_{4-1}^{0.975}] = R \setminus [-3.18, 3.18]$ , где  $t_{4-1}^k$  - квантиль уровня  $k$  распределения студента с 3 степенями свободы. То есть  $H_0$  отвергается при  $T \notin [-3.18, 3.18]$ , где  $T$  - тестовая статистика.

Найдем реализацию тестовой статистики:

$$(T | X = x) = \frac{295 - 300}{\frac{\sqrt{43100}}{\sqrt{4}}} \approx -0.048$$

Поскольку  $(T | (X = x)) \notin \mathcal{T}$ , то есть  $(-0.048) \in [-3.18, 3.18]$ , то нулевая гипотеза не отвергается на уровне значимости 0.05. А значит информации таксиста из второго таксопарка вполне можно доверять. Для пунктов **а** и **в** дополнения (до  $R$ ) ко множеству  $\mathcal{T}$  будут выглядеть как  $[t_{4-1}^{0.05}, \infty] = [-2.35, \infty]$  и  $[-\infty, t_{4-1}^{0.95}] = [-\infty, 2.35]$  соответственно, в связи с чем нулевая гипотеза в этих пунктах также не отвергается.

Теперь рассчитаем p-value для каждого из пунктов:

$$\text{а) } F_{t_3}(-0.048) = 1 - F_{t_3}(0.048) \approx 0.482$$

$$\text{в) } 1 - \Phi(-0.048) = F_{t_3}(0.048) \approx 0.518$$

$$b) 2 * \min(\Phi(-0.048), 1 - \Phi(-0.048)) \approx 2 * 0.482 = 0.964$$

Чтобы решить этот пункт в R, достаточно воспользоваться следующим кодом:

```
y=c(600,180,150,250) #реализация выборки  
t.test(y, mu=300, alternative = "two.sided") #пункт b  
t.test(y, mu=300, alternative = "less") #пункт a  
t.test(y, mu=300, alternative = "greater") #пункт c
```

Показать решение

Пожалуйста, войдите или зарегистрируйтесь, чтобы оценивать задачи, добавлять их в избранные и совершать некоторые другие, дополнительные действия.