

Критерии Стьюдента

7/7 баллов получено (100%)

Пересдать

Главная страница курса

Отлично!



Баллов: 1
/ 1

1.

Пусть t — значение статистики критерия Стьюдента для независимых выборок, nu — соответствующее число степеней свободы. Какая из строчек кода вычисляет достигаемый уровень значимости t -критерия при альтернативе $H_1: \mu_1 > \mu_2$?



```
1 stats.t.cdf(t, nu)
```



```
1 2*(1-stats.t.cdf(abs(t), nu))
```



```
1 1-stats.t.cdf(abs(t), nu)
```



```
1 1-stats.t.cdf(t, nu)
```



Правильно



```
1 stats.t.cdf(abs(t), nu)
```



Баллов: 1
/ 1

2.

Выберите ситуации, в которых допустимо проверять нулевую гипотезу $H_0: \mu_1 = \mu_2$ против односторонней альтернативы $H_1: \mu_1 > \mu_2$



В полученных выборках $\bar{X}_1 > \bar{X}_2$



Правильный вариант не выбран



Случай $\mu_1 < \mu_2$ нас не интересует: например, если X_1 — это эффективность изобретённого нами лекарства, а X_2 — эффективность плацебо, не в наших интересах брать двустороннюю альтернативу, потому что мы не хотим ненароком

показать, что лекарство хуже плацебо

Правильный вариант не выбран



Признаки устроены так, что $\mu_1 < \mu_2$ в принципе невозможно

Правильно



Баллов: 1
/ 1

3.

Выберите верные утверждения о проблеме Беренца-Фишера.



Проблема заключается в том, что не существует точного способа сравнить средние двух выборок с неизвестными дисперсиями

Правильно

К счастью, есть хотя бы приближённый способ



Проблема заключается в том, что при сравнении средних двух выборок с неизвестными дисперсиями нельзя даже приближённо посчитать число степеней свободы у нулевого распределения t-критерия

Правильный вариант не выбран



Приближённое решение проблемы Беренца-Фишера достаточно точно при $n_1 = n_2$

Правильно

Если выборки одинакового объёма, можно не волноваться



Приближённое решение проблемы Беренца-Фишера достаточно точно при $[n_1 > n_2] = [\sigma_1 < \sigma_2]$

Правильный вариант не выбран



Баллов: 1
/ 1

4.

Уровень кальция в крови здоровых молодых женщин равен в среднем 9.5 миллиграммам на децилитр и имеет характерное стандартное отклонение 0.4 мг/дл. В сельской больнице Гватемалы для 160 здоровых беременных женщин при первом обращении для ведения беременности был измерен уровень кальция; среднее значение составило 9.57 мг/дл. Можно ли утверждать, что средний уровень кальция в этой популяции отличается от 9.5?

Посчитайте достигаемый уровень значимости. Поскольку известны только среднее и дисперсия, а не сама выборка, нельзя использовать стандартные функции критериев — нужно реализовать формулу достигаемого уровня значимости самостоятельно.

Округлите ответ до четырёх знаков после десятичной точки.

Правильный ответ

```
1 z = (9.57 - 9.5) / (0.4 / np.sqrt(160))
2 p = 2*(1-stats.norm.cdf(abs(z)))
```



Баллов: 1
/ 1

5.

Как вы считаете, это отличие в среднем уровне кальция в крови практически значимо?

- ☐ Нет, раз критерий Стьюдента говорит, что отличие незначимо
- ☐ Да, раз критерий Стьюдента говорит, что отличие значимо
- ☒ Скорее всего нет: стандартное отклонение уровня в популяции намного больше полученного различия между средними

Правильно

Точнее может сказать только специалист в предметной области.



Баллов: 1
/ 1

6.

Имеются данные о стоимости и размерах 53940 бриллиантов:

diamonds.txt

Отделите 25% случайных наблюдений в тестовую выборку с помощью функции `sklearn.cross_validation.train_test_split` (зафиксируйте `random state = 1`). На обучающей выборке настройте две регрессионные модели:

1. линейную регрессию с помощью `LinearRegression` без параметров
2. случайный лес с помощью `RandomForestRegressor` с `random_state=1`.

Какая из моделей лучше предсказывает цену бриллиантов? Сделайте предсказания на тестовой выборке, посчитайте модули отклонений предсказаний от истинных цен. Проверьте гипотезу об одинаковом среднем качестве предсказаний, вычислите достигаемый уровень значимости. Отвергается ли гипотеза об одинаковом качестве моделей против двусторонней альтернативы на уровне значимости $\alpha = 0.05$?



Отвергается, $p < \alpha$



Правильно

```
1 stats.ttest_rel(abs(y_test - lm.predict(X_test)),
2                  abs(y_test - rf.predict(X_test)))
```



Не отвергается, $p > \alpha$



Баллов: 1
/ 1

7.

В предыдущей задаче посчитайте 95% доверительный интервал для разности средних абсолютных ошибок предсказаний регрессии и случайного леса. Чему равна его ближайшая к нулю граница? Округлите до **десятков** (поскольку случайный лес может давать немного разные предсказания в зависимости от версий библиотек, мы просим вас так сильно округлить, чтобы полученное значение наверняка совпало с нашим).

70



Правильный ответ

Получается, случайный лес предсказывает цены бриллиантов в среднем на 70-100 долларов точнее.

