

## Домашнее задание 4

Дедлайн: 2025-05-18, 23:59.

Оцениваемые задачи:

1. Величины  $(y_i)$  независимы и одинаково непрерывно распределены. Всего есть 1000 наблюдений. Постройте 95%-й интервал для 90%-го квантиля с помощью выборочных квантилей.  
Если для вычисления необходимых выборочных квантилей использовался код, то приведите его.
2. Есть две выборки:  $x = (2.7, 3.5, 4.2, 6.7)$  и  $y = (1.6, 2.9, 3.9)$ . Все наблюдения независимы. Величины  $(x_i)$  одинаково непрерывно распределены между собой, величины  $(y_i)$  одинаково непрерывно распределены между собой. Проверьте гипотезу  $H_0$  об одинаковом законе распределения в двух выборках, против альтернативной  $\mathbb{P}(x_i > y_j) > 0.5$  на уровне значимости 5%.
  - а) Проведите тест Манна — Уитни, используя точное распределение статистики.
  - б) Проведите тест Манна — Уитни, используя нормальную аппроксимацию. Укажите  $p$ -значение.

Неоцениваемые задачи в удовольствие:

3. Рассмотрим тест знаковых рангов Уилкоксона и связанные пары наблюдений  $(x_i, y_i)$ . При верной  $H_0$  разницы  $D_i = x_i - y_i$  одинаково непрерывно распределены и независимы.  
Рассмотрим сумму знаковых рангов  $WSR = \sum_{i=1}^n \text{sign}(D_i) \text{rank}(|D_i|)$ .  
Найдите ожидание  $\mathbb{E}(WSR)$  и дисперсию  $\text{Var}(WSR)$  при верной  $H_0$ .
4. Величины  $(X_i)$  независимы и одинаково распределены с неизвестными  $\mathbb{E}(X_i) = \mu$  и  $\text{Var}(X_i) = \sigma^2$ . По выборке из 1000 наблюдений оказалось, что  $\bar{X} = 30$ , а несмещённая выборочная дисперсия равна 900.
  - а) Постройте асимптотический 95%-й доверительный интервал для  $\mu$ . Укажите  $p$ -значение для гипотезы  $H_0: \mu = 35$  против альтернативной  $H_a: \mu \neq 35$ .
  - б) Постройте асимптотический 95%-й предсказательный интервал для  $X_{1001}$ .
  - в) Постройте асимптотический 95%-й предсказательный интервал для  $(X_{1001} + X_{1002})/2$ .
5. Бариста Борис заметил, что в последнее время посетители заказывают только капучино и раф. Предположим, что посетители выбирают напиток независимо друг от друга, а вероятность выбора капучино постоянна и равна неизвестному числу  $p$ .  
У Бориса есть только две гипотезы,  $H_0: p = 1/3$  и  $H_a: p = 2/3$ , в которые он до получения данных верит с вероятностями 0.6 и 0.4, соответственно.  
Из первых 100 утренних посетителей  $S = 40$  выбрали капучино. Борис хочет измерить разными способами, насколько этот наблюдаемый результат соотносится с гипотезами.
  - а) Найдите  $\mathbb{P}(H_0 \mid S = 40)$  и  $\mathbb{P}(H_a \mid S = 40)$ .

Борис решил на следующий день повторить эксперимент и снова посчитать  $S_{\text{new}}$ , количество клиентов из первых ста, которые выберут капучино.

  - б) Найдите  $\mathbb{P}(S_{\text{new}} \geq S \mid S = 40, H_0)$  и  $\mathbb{P}(S_{\text{new}} \geq S \mid S = 40, H_a)$ .

- в) Какие из вероятностей можно посчитать без мнения Бориса о  $\mathbb{P}(H_0)$  и  $\mathbb{P}(H_a)$ ?  
 г) Какая из вероятностей называется  $p$ -значением для гипотезы  $H_0$  и статистики  $S$ ?

6. По таблице сопряжённости проверьте гипотезу о независимости двух признаков на уровне значимости 5% против альтернативной гипотезы о зависимости признаков. Укажите  $p$ -значение.

	$X = A$	$X = B$
$Y = C$	50	60
$Y = D$	20	30
$Y = E$	60	50

7. Рассмотрим таблицу сопряжённости

$X = A$	$X = B$	$X = C$	$X = D$
50	70	80	60

- а) На уровне значимости 5% проверьте гипотезу об одинаковых вероятностях  $p_a = p_b = p_c = p_d$  против альтернативной о том, что хотя бы одно из равенств нарушено.  
 б) На уровне значимости 5% проверьте гипотезу об одинаковых вероятностях  $p_a = p_b = p_c = p_d$  против альтернативной о том, что  $p_a \neq p_b = p_c$ .  
 в) На уровне значимости 5% проверьте гипотезу об одинаковых вероятностях  $p_a = p_b = p_c$  против альтернативной о том, что  $p_a \neq p_b = p_c$ .

В каждом случае укажите  $p$ -значение.