

Прикладная статистика и анализ данных.

Задание 7.

- Дедлайн **30 марта 16:30**. После дедлайна работы не принимаются кроме случаев наличия уважительной причины.
- Выполненную работу нужно отправить на почту `mipt.stats@yandex.ru`, указав тему письма "[asda] Фамилия Имя - задание 7". Квадратные скобки обязательны. Если письмо дошло, придет ответ от автоответчика.
- По задачам 4-5 прислать нужно ноутбук и его pdf-версию (без архивов). Названия файлов должны быть такими: `7.N.ipynb` и `7.N.pdf`, где N — ваш номер из таблицы с оценками.
- Задача 1-3 необходимо оформить в `tex` и прислать pdf или же прислать фотку в правильной ориентации рукописного решения, где все четко видно.
- Решения, размещенные на каких-либо интернет-ресурсах не принимаются. Кроме того, публикация решения в открытом доступе может быть приравнена к предоставлению возможности списать.
- Не забывайте делать пояснения и выводы.

1. (3 балл) Рассмотрим критерий Уилкоксона-Манна-Уитни. Докажите, что при отсутствии совпадений среди X_i и Y_j для статистики Манна-Уитни справедливо $U = V - \frac{m(m+1)}{2}$. Найдите EU при справедливости гипотезы H_0 об однородности выборок.
2. (4 балл) Докажите, что отсутствии нулевых значений и совпадений среди величин $|Z_i|$ для статистики критерия ранговых сумм Уилкоксона выполнено

$$T = \sum_{i \leq j} I \left\{ \frac{Z_i + Z_j}{2} > 0 \right\}$$

3. (3 балл) Докажите, что при выполнении предположений критерия ранговых сумм Уилкоксона случайные векторы $U = (U_1, \dots, U_n)$ и $R = (R_1, \dots, R_n)$ независимы.
4. (3 балл) Пусть (X_1, \dots, X_n) и (Y_1, \dots, Y_n) — связанные выборки, причем разности $\varepsilon_i = Y_i - X_i$ образуют выборку с симметричным распределением. Какой из критериев предпочтительнее использовать для проверки гипотезы об отсутствии сдвига — критерий знаков или критерий ранговых сумм Уилкоксона? Исследуйте случаи, когда ε_i имеют распределения нормальное, Лапласа, Коши.

Подсказка. Исследуйте мощности критериев.

5. (6 балла) Рассмотрим данные о банковских учреждениях из домашнего задания 5. С помощью методов дисперсионного анализа исследуйте, по каким параметрам наблюдается различие между двумя рассматриваемыми классами. При возможности оцените также степень различия (в т.ч. доверительный интервал).