

Прикладная статистика и анализ данных.

Задание 9.

- Дедлайн **10 мая 23:59**. 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 9". Квадратные скобки обязательны. Если письмо дошло, придет ответ от автоответчика.
 - Прислать нужно ноутбук и его pdf-версию (без архивов). Названия файлов должны быть такими: `9.N.ipynb` и `9.N.pdf`, где N — ваш номер из таблицы с оценками.
 - Решения, размещенные на каких-либо интернет-ресурсах не принимаются. Кроме того, публикация решения в открытом доступе может быть приравнена к предоставлении возможности списать.
 - В каждой задаче не забывайте делать **пояснения и выводы**.
1. **(4 балла)** Используя теорему о разложении гауссовского вектора получите F -критерий однофакторного дисперсионного анализа для случая независимых выборок.
 2. **(4 балла)** Рассмотрим модель $X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$, $i = 1..n$, $j = 1..k$ однофакторного дисперсионного анализа для случая связанных выборок. Фактор отвечает за степень воздействия на объект, поэтому в качестве альтернативной гипотезы рассматривается $H_1: \beta_1 \leq \dots \leq \beta_k$. Исследуйте поведение мощности критериев Фридмана и Пейджа в зависимости от размера выборки, количества групп, а также вида распределения шума ε_{ij} , если на самом деле $\beta_j = j\beta$, где $\beta > 0$.
 3. **(3 балла)** Выберите несколько человек (не менее 5) с DS-потока по состоянию на осенний семестр и сравните их успеваемость методами дисперсионного анализа, рассмотрев оценки по всем практическим заданиям. Обратите внимание, что в отличие от примера с презентацией в данном случае исследуемым фактором является человек, а задание — мешающий фактор.
 4. **(12 баллов)** На занятии был разобран на Питоне первый этап дисперсионного анализа данных о внебрачных отношениях. Выполните в R следующие задания:
 - (a) Разделите данные на две части в зависимости от того, изменял ли человек супругу(е).
 - (b) Повторите проведенный на занятии анализ. Имеет ли смысл в некоторых случаях рассматривать альтернативную гипотезу вида $H_1: \beta_1 \leq \dots \leq \beta_k$?
 - (c) Проведите post-hoc анализ там, где это необходимо.
 - (d) Исследуйте совместное влияние некоторых факторов методами многофакторного дисперсионного анализа. Поясните, почему вы выбрали именно эти факторы.
 - (e) Исследуйте, какие факторы влияют на факт измены.
 - (f) По результатам исследования сделайте выводы о семейной жизни.