**Лабораторная работа №3**

**Задание 1**

Воспользуемся данными npk, иллюстрирующими влияние применения различных удобрений на урожайность гороха (yield). Проведите трехфакторный дисперсионный анализ, где зависимая переменная - это урожайность (yield), а три фактора - типы удобрений (N, P, K). После проведения данного анализа вы получите три значения p - уровня значимости, сделайте вывод о значимости каждого из факторов.

**Задание 2**

Проведите однофакторный дисперсионный анализ на встроенных данных iris. Зависимая переменная - ширина чашелистика (Sepal.Width), независимая переменная - вид (Species). Затем проведите попарные сравнения видов. Какие виды статистически значимо различаются по ширине чашелистика (p < 0.05)?

**Задание 3**

В этой задаче вам дан набор данных («Pilluikin.csv»), в котором представлена информация о температуре нескольких пациентов, которые лечатся разными таблетками и у разных врачей.

Проведите однофакторный дисперсионный анализ с повторными измерениями: влияние типа таблетки (pill) на температуру (temperature) с учётом испытуемого (patient). Каково p-value для влияния типа таблеток на температуру?

**Задание 4**

Теперь Вашей задачей будет провести двухфакторный дисперсионный анализ с повторными измерениями: влияние факторов doctor, влияние фактора pill и их взаимодействие на temperature. Учтите обе внутригрупповые переменные: и тот факт, что один и тот же больной принимает разные таблетки, и тот факт, что один и тот же больной лечится у разных врачей! Каково F-значение для взаимодействия факторов доктора (doctor) и типа таблеток (pill)?

**Задание 5**

Напишите функцию outliers.rm, которая находит и удаляет выбросы. Для обнаружения выбросов воспользуемся самым простым способом, с которым вы не раз встречались, используя график Box plot.

Выбросами будем считать те наблюдения, которые отклоняются от 1 или 3 квартиля больше чем на 1,5 \*  IQR, где  IQR  - межквартильный размах.

На вход функция получает числовой вектор x. Функция должна возвращать модифицированный вектор x с удаленными выбросами.

**Задание 6**

Напишите функцию smart\_cor, которая получает на вход dataframe с двумя количественными переменными. Проверьте с помощью теста Шапиро-Уилка, что данные в обеих переменных принадлежат нормальному распределению.

Если хотя бы в одном векторе распределение переменной отличается от нормального (p - value меньше 0.05), то функция должна возвращать коэффициент корреляции Спирмена. (Числовой вектор из одного элемента).

Если в обоих векторах распределение переменных от нормального значимо не отличается, то функция должна возвращать коэффициент корреляции Пирсона.