

---

## ЛАБОРАТОРНАЯ РАБОТА №2

### ИМПОРТ ДАННЫХ. РАЗВЕДЫВАТЕЛЬНЫЙ АНАЛИЗ

---

**Задание 1.** Скачайте из учебной группы файл `данные.csv`. Импортируйте из него данные в рабочую среду. Присвойте импортированную таблицу данных в некоторую переменную. Укажите команде импорта, что первая строка содержит заголовки столбцов, а разделителем ячеек в строке является точка с запятой.

**Справка.** Для импорта данных используются команды `read.table(...)`, `read.csv(...)`.

**Задание 2.** Очистите таблицу от NA, присвойте результат **в другую переменную**. Определите, сколько строк в исходной таблице и сколько их в очищенной.

**Справка.** Очистка от NA объекта `x`: `na.omit(x)`. Вычисление количества элементов в объекте `x`: `length(x)`.

**Задание 3.** Отсортируйте очищенную таблицу по двум полям: первичный ключ — пол (по возрастанию), вторичный ключ — балл по математике (по убыванию).

**Справка.** Сортировка объекта `x` по полю `Поле` по возрастанию: `x[order(x$Поле),]`. Сортировка объекта `x` по полю `Поле` по убыванию: `x[order(-x$Поле),]`. Если ключей сортировки несколько, то указанные поля перечисляются через запятые внутри команды `order(...)`.

**Задание 4.** Найдите количество мужчин и женщин в таблице.

**Справка.** Генерирование таблицы распределения значений: `table(x)`.

**Задание 5.** Извлеките в новую переменную данные из столбца `Математика`. Найдите среднее выборочное, выборочную дисперсию, среднееквадратическое отклонение, квартили. Постройте гистограмму.

**Справка.** Среднее выборочное вектора значений: `mean(x)`. Несмещенная дисперсия вектора значений: `var(x)`. Среднееквадратическое отклонение вектора значений: `sd(x)`. Суммарная информация по квартилям: `summary(x)`. Гистограмма: `hist(x)`.

**Задание 6.** Создайте скрипт, который возвращает вектор частот группированной выборки значений результатов по математике. Количество интервалов разбиения задайте по формуле Стёрджесса:  $k = 1 + \lfloor \log_2 n \rfloor$ .

**Справка.** Чтобы создать скрипт, нужно выбрать пункт меню `File | New script`. Запуск скрипта выполняется с помощью пункта меню `Edit | Run all`. Команды в скрипте завершаются точкой с запятой.

Цикл с параметром: `for (<переменная> in <диапазон>) {тело цикла}`.

Условный оператор: `if (<условие>) {да-операторы} else {нет-операторы}`.

Минимум и максимум вектора: `min(x)`, `max(x)`. Логарифм: `log(x, основание)`. Целая часть числа «снизу»: `floor(x)`. Добавление в вектор нового значения: `append(x, значение)`. Логическое И: `x && y`. Логическое ИЛИ: `x || y`. Логическое НЕ: `!x`.

Приведем пример использования конструкций цикла и ветвления. К примеру, пусть  $m$  — числовая матрица размерности  $10 \times 10$ . Вывод всех её элементов, больших 50:

```
for (i in 1:10)
{
  for (j in 1:10)
  {
    if (m[i,j] > 50)
    {
      print(m[i,j])
    }
  }
}
```