

Вариант 1.

1. Рассчитайте показатели описательной статистики (среднюю, ошибку средней, стандартное отклонение, медиану, квартили) *iris\$Petal.Length* используя неявный цикл *tapply()*
2. Рассчитайте показатели описательной статистики (среднюю, ошибку средней, стандартное отклонение, медиану, квартили) по 1-4 столбцам *iris* при помощи цикла *for()*
3. Сгенерируйте последовательность чисел с нормальным распределением при $n = 100$, $\text{mean} = 170$, $\text{sd} = 4$.
4. Постройте график зависимости *stretch(mm)* от *distance(cm)* из фрейма данных *elastic1*.

elastic1

<i>stretch(mm)</i>	46	54	48	50	44	42	52
<i>distance(cm)</i>	183	217	189	208	178	150	249

5. Загрузите таблицу «Коровы_кр», преобразуйте в *tibble* Рассчитать среднюю арифметическую, медиану и стандартное отклонение по удою жиру и белку не удаляя всех строк с пропущенными записями. Очистить таблицу от строк с пропущенными значениями.
6. По данным таблицы «Коровы_кр» определите выбросы по удою, жиру и белку:
 - а) С помощью фильтра Хампеля
 - б) По правилу 6 сигм
 - в) Графическим способом
7. Рассчитайте коэффициент корреляции между *iris\$Sepal.Length* и *iris\$Sepal.Width*. Обоснуйте выбор способа корреляционного анализа.
8. По данным *InsectSprays* рассчитайте влияние группирующего фактора *spray* на зависимую переменную *count*. Обоснуйте выбор способа оценки.
9. Из таблицы *iris* извлечь все строки, содержащие запись "virginica".
10. Из таблицы *state.x77* извлечь все штаты, население (*\$Population*) которых выше среднего
11. Напишите функцию расчёта ошибки средней арифметической

Вариант 2.

1. Изобразить при помощи столбчатой диаграммы средние значения таблицы *mtcars*
2. Рассчитайте показатели описательной статистики (среднюю, ошибку средней, стандартное отклонение, медиану, квартили) по 1-4 столбцам ***iris*** при помощи неявного цикла *apply()*
3. Сгенерируйте вектор из 30 записей с интервалом 0.64 таким образом, что бы последняя запись была равна 85.
4. Используя конструкцию *if...else* произведите кластеризацию возраста деревьев (*\$age*) из фрейма данных ***Orange***: меньше или равно 500 — А, от 501 до 1000 — В, 1001 и более — С.
5. Загрузите таблицу ***heroes_information***(https://raw.githubusercontent.com/Pozdniakov/tidy_stats/master/data/heroes_information.csv). Извлеките все строки, где имеется запись «male». Подсчитать количество и долю пропущенных записей в столбцах *Race*, *Hair color*, *Skin color*
6. Извлеките все строки таблицы ***heroes_information***, где нет выбросов в столбце *Height*. Произведите расчёт описательной статистики столбцов *Height* и *Weight* в зависимости от группирующего фактора *Publisher*.
7. Рассчитайте коэффициент корреляции между ***iris\$Sepal.Length*** и ***iris\$Petal.Width***, оцените достоверность. Обоснуйте выбор способа корреляционного анализа.
8. По данным ***heroes_information*** определите степень влияния группирующего фактора *Race*. Обоснуйте выбор способа оценки.
9. Из таблицы *iris* извлечь все строки, содержащие запись "setosa".
10. Из таблицы *state.x77* извлечь все штаты, население (*\$Population*) которых ниже среднего
11. Напишите функцию расчёта гипотенузы.

Вариант 3.

1. По данным ***InsectSprays*** изобразите графически значения средней арифметической эффективности каждого спрея в зависимости от группирующего фактора *spray*.
2. Рассчитайте показатели описательной статистики (среднюю, ошибку средней, стандартное отклонение, медиану, квартили) по 2-8 столбцам ***state.x77*** при помощи цикла *for()*
3. Сгенерируйте рандомизированный вектор из 10 записей «Blue», 10 «Green», 10 «White», 10 «Red».
4. По данным ***ChickWeight*** рассчитайте показатели описательной статистики переменной *weight* в зависимости от группирующих факторов *Chick* и *Diet* используя соответствующую функцию пакета ***psych***.
5. Загрузите таблицу *semex.csv*. Избавьтесь от всех строк с пропущенными записями. Сравните размерность исходной и обработанной таблиц. Обработанную таблицу преобразуйте в *data.table*.
6. В обработанной таблице определите выбросы при помощи фильтра Хампеля для столбцов *TPI*, *milk*, *fat*, *protein*.
7. В обработанной таблице рассчитайте корреляцию между *TPI* и *milk*, оцените достоверность. Обоснуйте выбор способа корреляционного анализа.
8. Найдите количество записей быков в зависимости от страны происхождения (*semex\$Country*). По данным ***semex*** оцените влияние страны происхождения на прибавку к молоку, жиру и белку. Обоснуйте способ анализа.
9. Из таблицы ***semex*** извлечь быков, родившихся только в США и Канаде.
10. Из таблицы ***semex*** извлечь быков, чей показатель *TPI* находится между первым и третьим квартилями.
11. Напишите функцию расчёта средней взвешанной

Вариант 4.

1. Графически изобразите ранжированный ряд записей *iris\$Sepal.Width*
2. Рассчитайте показатели описательной статистики (среднюю, ошибку средней, стандартное отклонение, медиану, квартили) по 1-4 столбцам *iris* при помощи неявного цикла *apply()*
3. Сгенерируйте вектор из 50 значений, начиная от 100 расположенных по убыванию так что бы интервал между записями был равен 3.
4. Загрузите таблицу *SearchExcel*. Используя конструкцию *if...else* произведите кластеризацию столбца *Udder health*: более 100 -А, 100 и менее — В.
5. Определите количество и долю заполненных ячеек в столбцах *Growth*, *General health* и *Saved feed*. Рассчитайте среднюю арифметическую по этим столбцам в для каждого года рождения.
6. Рассчитайте корреляцию между столбцами *Orange\$circumference* и *Orange\$age*. Определите достоверность. Обоснуйте выбор метода корреляционного анализа.
7. При помощи правила 6 сигм и фильтра Хампеля найдите выбросы для столбцов *Population* и *Area* таблицы *state.x77*.
8. Установите влияние группирующего фактора *diet* на переменную *Weight* по данным *ChickWeight*. Обоснуйте способ анализа.
9. Из таблицы *iris* извлечь записи, не относящиеся к *setosa*
10. Из таблицы *state.x77* извлечь все штаты, население (*\$Population*) которых ниже среднего
11. Напишите функцию нахождения площади круга при условии что *x* — длина окружности.