

## Контрольная работа №1. Вариант 2

Базы данных `height1.sav` и `height2.sav` содержат некоторые данные, полученные в ходе опроса студентов второго курса факультета социологии НИУ ВШЭ. База данных `height1.sav` содержит переменные `id` (идентификационный номер респондента) и `height` (рост респондента). База данных `height2.sav` содержит переменные `id` (идентификационный номер респондента), `fheight` (рост отца), `mheight` (рост матери) и `female` (пол респондента).

1. Объедините базы данных `height1.sav` и `height2.sav` по переменной `id`.
2. Найдите и укажите минимальное, максимальное, среднее, медианное значения и среднеквадратичное отклонение роста респондентов. Постройте и кратко опишите коробчатую диаграмму (box plot) роста респондентов.
3. Постройте корреляционную матрицу переменных рост респондента, рост отца респондента и рост матери респондента. Проинтерпретируйте полученные коэффициенты корреляции.
4. Создайте переменную, измеряющую разницу между ростом отца и ростом матери. Постройте и проинтерпретируйте гистограмму этой переменной.
5. Перекодируйте переменные для роста отца и матери в категориальные переменные с категориями “низкий рост”, “средний рост”, “высокий рост”, в каждой из которых находилось бы около трети наблюдений. Постройте и проинтерпретируйте таблицу сопряженности между этими категориальными переменными с процентами по строкам и по столбцам.
6. Оцените и проинтерпретируйте регрессионную модель, в которой зависимой переменной является рост респондента, а независимыми – пол респондента, рост матери и рост отца. Выпишите регрессионное уравнение, проинтерпретируйте величину и статистическую значимость регрессионных коэффициентов, показатель R-квадрат. Сравните величину эффектов роста отца и матери, используя стандартизованные регрессионные коэффициенты. Кратко содержательно проинтерпретируйте полученные результаты.
7. Повторите этот анализ отдельно для юношей и девушек (независимыми переменными, соответственно, будут только рост отца и

рост матери). Представьте и сравните между собой две модели. Обратите внимание на количество наблюдений в двух группах.