

Комиссия 11 октября 2012 г.

Базы данных `ege1.sav` и `ege2.sav` содержат некоторые данные, полученные в ходе опроса студентов второго курса факультета социологии НИУ ВШЭ. База данных `ege1.sav` содержит переменные `id` (идентификационный номер респондента), `female` (пол респондента), `rus` (результат ЕГЭ по русскому языку). База данных `ege2.sav` содержит переменные `id` (идентификационный номер респондента), `math` (результат ЕГЭ по математике), `feduc` (образование отца), `meduc` (образование матери) и `moscow` (место, в котором респондент закончил школу: Москва и Московская область vs. какой-нибудь другой регион).

1. Объедините базы данных `ege1.sav` и `ege2.sav` по переменной `id`.
2. Найдите и укажите минимальное, максимальное, среднее, медианное значение и среднеквадратичное отклонение результатов ЕГЭ по математике. Постройте и кратко опишите гистограмму результатов ЕГЭ по математике.
3. Создайте таблицу сопряженности переменных пол и место окончания средней школы. Приведите и проинтерпретируйте проценты по строкам и столбцам. Кратко опишите полученные результаты.
4. Перекодируйте переменные для образования отца и матери в две дихотомические переменные (0 – высшее образование отсутствует, 1 – есть высшее образование или ученая степень). Создайте индекс образования родителей, просуммировав эти две переменные. Таким образом, индекс может принимать три значения: 0 – высшее образование отсутствует у обоих родителей, 1 – высшее образование есть у одного из родителей, 2 – высшее образование есть у обоих родителей. Убедитесь, что для респондентов, которые не указали образование отца или матери, индекс принимает пропущенное значение. Приведите таблицу распределения созданного индекса.
5. Оцените и проинтерпретируйте регрессионную модель, в которой зависимой переменной является результат ЕГЭ по математике, а независимыми – пол, место окончания средней школы и индекс образования родителей (введенный в модель как категориальная переменная). Выпишите регрессионное уравнение, проинтерпретируйте величину и статистическую значимость регрессионных коэффициентов, показатель R-квадрат. Кратко содержательно проинтерпретируйте полученные результаты.