Дисциплина: Анализ данных

ФИО студента \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ответьте, пожалуйста, на вопросы:

Задание 1. Проверка отличия распределения от «нормального»

*Откройте файл* ***salary.dta***

Проанализируйте распределение значений переменной **salary**. Проведите тест для проверки того отличается ли распределение значений этой переменной от нормального распределения. Сформулируйте гипотезы и выводы по результатам анализа.

Задание 2. Тест Хи-квадрат

*Откройте файл nlsw88.dta (example datasets)*

Проведите тест Хи-квадрат с парами переменных. Сформулируйте гипотезы. Интерпретируйте результаты анализа. Обоснуйте свои выводы.

a) married и race \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

b) age и married \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Задание 3. Парная корреляция

*Откройте файл auto.dta (example datasets)*

Подчитайте парный коэффициент корреляции между переменными. Обоснуйте выбор коэффициента. Сформулируйте гипотезы. Оцените значимость коэффициента. Сделайте выводы относительно силы и направления взаимосвязи.

a) length и turn \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

b) price и weight \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

c) mpg и price \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

d) foreign и price \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Задание 4. Частная корреляция

*Откройте файл health\_funding.dta*

Посчитайте парный коэффициент корреляции между переменными funding и

disease. Посчитайте частный коэффициент корреляции между этими переменными, исключив влияние переменной visits. Сделайте выводы.

Для каждого задания постройте графики, визуализирующие анализируемые взаимосвязи.

Пришлите с личного почтового ящика на [amelikyan@hse.ru](mailto:amelikyan@hse.ru) следующие файлы:

* Этот файл MS Word с ответами,
* Do-file.