

Задание на взаимное оценивание №1

Задание основано на данных опроса населения РМЭЗ НИУ ВШЭ в 2017 году (<https://www.hse.ru/rf/ms/spss>). Для домашнего задания на основе базы РМЭЗ НИУ ВШЭ была составлена выборка из индивидов, находящихся в трудоспособном возрасте. В файле **hw.gdt** (файл Gretl) (или **hw.xls** – файл Excel) содержатся следующие переменные:

Список переменных

- public – 1 если государственное предприятие (в государственной собственности), 0 если частное;
- wage – заработная плата, полученная за последние 30 дней по основному месту работы;
- educ – уровень образования, категориальная переменная (0 для индивидов без среднего образования)
 1. среднее образование
 2. ПТУ, техническое училище
 3. институт, университет, академия
- urban – 1 респондент проживает в городе, 0 в сельской местности;
- male – пол, 1 для мужчин, 0 для женщин;
- age – возраст в годах;
- children – количество детей;
- internet – приходилось ли респонденту в течение последних 12 месяцев пользоваться Интернетом, 1 если да, 0 в противном случае;

Внимание! Для доверительных интервалов и расчетных тестовых статистик при проверке гипотез требуется выписать **формулы**, где будут указаны **степени свободы и значения переменных**, которые соответствуют Вашей выборке. При проверке гипотез требуется указать **расчетное и критическое значение статистики**, а также рассчитать **p-value**. Расчеты можно делать в Excel, Gretl или иной программе. Результаты всех тестов должны быть проинтерпретированы (отвергается или не отвергается гипотеза, на каком уровне значимости при какой альтернативной). На выходе работа должна быть представлена в виде текста (комментариев), формул, результатов, таблиц и графиков **не более 7 страниц**. Задания в тексте обязательно должны быть пронумерованы, согласно пунктам ниже.

Задание

- 1) Опишите вашу **выборку**. Как она была получена? Что является в Вашем случае генеральной совокупностью?
- 2) Рассчитайте **описательные статистики** (минимум, максимум, среднее значение, стандартное отклонение) для всех переменных в Вашей выборке.
- 3) Оцените **перцентили** (5%, 50%, 95%) распределения для заработной платы и возраста респондентов. Определите межквартильный размах.

- 4) Постройте **box-plot** для заработной платы и возраста респондентов. Есть ли выбросы?
- 5) Постройте **гистограммы** распределения для заработной платы и возраста респондентов в выборке. Что можно сказать о скошенности (асимметрии) их распределений?
- 6) Как распределены респонденты по уровню образования? Постройте **гистограмму** распределения и **таблицу** распределения частот. Респонденты с каким уровнем образования имеют наибольшую долю в выборке?
- 7) Постройте 95% и 99% доверительные интервалы для **математического ожидания** и **стандартного отклонения** генеральной совокупности для заработной платы.
- 8) Постройте 90% и 95% доверительный интервал для **доли женщин** в генеральной совокупности.
- 9) Проверьте гипотезу, что **матожидание** заработной платы равно 26000 рублей против двусторонней и правосторонней альтернативной гипотезы.
- 10) Проверьте гипотезу, что средняя заработная плата женщин ниже средней заработной платы мужчин.
- 11) Проверьте гипотезу, что **дисперсия** заработных плат работников, пользующихся Интернетом, выше, чем дисперсия заработной платы работников, не пользующихся Интернетом.
- 12) Проверьте гипотезу, что **доля** работников, имеющих одного ребенка, равна 0,5 при левосторонней альтернативной.
- 13) Постройте **корреляционную таблицу** для переменных в выборке. Проинтерпретируйте результаты (какие переменные имеют наибольшую положительную и отрицательную корреляцию).
- 14) Предположите зависимость заработной платы от возраста респондентов. Постройте **график**, демонстрирующий эту зависимость.