

Лаба 2

Прочитать описание проекта «Российский мониторинг экономического положения и здоровья населения НИУ ВШЭ (RLMS-HSE)»

<https://www.hse.ru/rlms/description>

Посмотреть фильм ВШЭ (если получится) о данных, которые предстоит анализировать: Основы работы с данными RLMS-HSE

https://www.youtube.com/watch?v=zOlGDW_VovM&list=PLM0wivDx2MreKrzonmYAhIrV_WCfXOpwZ

Скачать подготовленные для анализа данные

<https://docs.google.com/spreadsheets/d/1XGY5ZTxsgT1Zc-vy9e4cVrmE5vOjsIPd/edit?usp=sharing&oid=101164548489542391573&rtpof=true&sd=true>

Задание 2.1

Проверить гипотезу о статистической значимости различия между доходами двух групп работающих и получающих доход граждан Петербурга:

1 группа - имеющие образование среднее и ниже,

2 группа - имеющие среднее специальное или высшее образование

Представить в отчете ход проверки гипотезы (рисунки, таблицы) и выводы.

Для формирования выборки использовать признаки:

- Регион
- ЗАКОНЧЕННОЕ ОБРАЗОВАНИЕ (ГРУППА)
- Ваше основное занятие в настоящее время? = 1 | Вы сейчас работаете
- Сколько денег в течение последних 30 дней Вы получили по основному месту работы после вычета налогов и отчислений? Если все или часть денег Вы получили в иностранной валюте, переведите все в рубли и назовите общую сумму

Для проверки гипотезы использовать jamovi (можно Python), ДА/ДА,
однофакторный дисперсионный анализ, Т-тесты

Инструкция

Help jamovi <https://www.jamovi.org/user-manual.html>

Сохранить только лист с данными в отдельный файл, используя только одну первую строку для заголовка, указав в ней названия переменных, например, регион, пол, возраст, образование и т.д.

Открыть файл в jamovi

Проверить типы данных.

Сохранить файл данных.

Поставить фильтры (добавлять фильтры, нажимая маленький + справа).

Добавить вычисляемую переменную для группировки на два уровня образования:

1 группа - имеющие образование среднее и ниже,

2 группа - имеющие среднее специальное или высшее образование

Для проверки гипотезы использовать однофакторный дисперсионный анализ.

Проверять по критерию Фишера. Вывести значения описательной статистики.

Для проверки гипотезы использовать Т-тест для независимых выборок.
Проверять по критерию Стьюдента. Вывести значения описательной статистики и описательные графики.

Подсказка: ответа должны совпасть.

Задание 2.2

Для выделенных ранее групп проверить гипотезу о равенстве средней продолжительности работы в неделю.

Выполнять только в jamovi (можно Python)

Представить в отчете ход проверки гипотезы (рисунки, таблицы) и выводы.

Инструкция

Для выделенной ранее группы добавить фильтр по переменной «Сколько часов в среднем продолжается Ваша обычная рабочая неделя?»: выбрать только тех, кто указал время.

Для проверки использовать критерий Фишера или Стьюдента.

Полученный вывод подтвердить рисунком.

Задание 2.3

Для работающих, указавших продолжительность работы и получающих доход граждан (**проживающих в любом населенном пункте**) исследовать взаимосвязь двух признаков: курение и употребление алкоголя за последние 30 дней.

Выполнять только в jamovi (можно Python)

Использовать Частотный анализ, Таблицы сопряженности парных выборок.

Представить в отчете ход проверки гипотезы (рисунки, таблицы) и выводы.

Инструкция

Для выделенной ранее группы удалить фильтр по переменной «Регион», добавить фильтры по переменным «Вы курите в настоящее время?» и «В течение последних 30 дней Вы употребляли алкогольные напитки?» и выбрать только тех, кто ответил «да» или «нет».

С помощью Частотный анализ, Таблицы сопряженности парных выборок проверяется гипотеза о независимости признаков, против альтернативы признаки связаны, т.е.

проверяется есть ли связь между тем, что человек курит (фактор риска), и тем, что человек употреблял алкоголь в последние 30 дней (исход).

p – уровень значимости, если его значение меньше 0.05, то гипотезу об отсутствии связи между переменными следует отклонить.