

Математическая статистика

Для данных результатов наблюдений случайных величин (СВ) X и Y требуется¹:

1. Представить выборку для СВ X в виде интервального вариационного ряда построив следующую таблицу, где также привести относительные накопленные частоты и ранги.

Интервал		Варианта	Частота	Относительная накопленная частота	ранг
от	до	x_i	n_i	F_i	r_i

2. Построить гистограмму относительных частот СВ X ;
3. Рассчитать для двух выборок:
 - выборочные средние,
 - «исправленные» выборочные стандартные отклонения,
 - максимальное и минимальное значения,
 - моду и медиану;
4. Проверить гипотезу о нормальном распределении генеральной совокупности для СВ X и Y .
5. Подобрать теоретическую кривую и её параметры по данным выборки СВ X .
6. Найти для с.в. X и Y :
 - интервальные оценки математического ожидания (95% доверительный интервал).
 - *интервальные оценки для среднеквадратического отклонения.
7. Построить диаграммы размаха (ящик с усами)
8. Проверить гипотезу о равенстве математических ожиданий генеральных совокупностей из которых сделаны выборки.
9. Рассчитать выборочный коэффициент линейной корреляции между X и Y ;
 - Проверить гипотезу о значимости выборочного коэффициента корреляции;
 - Построить уравнение линейной регрессии Y на X , построить их график уравнения на диаграмме рассеивания.
 - Определить предсказательную способность уравнения линейной регрессии с помощью R^2

Уровень значимости принять равным 0.05.

Звёздочкой отмечены необязательные задания

Для выполнения работы рекомендуется использовать язык программирования R или Python (в среде R Studio или Jupyter), статистические функции программ для работы с электронными таблицами (Microsoft Excel, LibreOffice Calc) или иные математические пакеты.

Готовый Jupyter блокнот для анализа данных: <https://github.com/VetrovSV/ST/blob/master/task1/Python.%20Statistics%202.ipynb>

1 По согласованию с преподавателем можно предложить свой набор данных, скорректировать или изменить задание

Литература

1. Гмурман В.Е. Теория вероятностей и математическая статистика
2. Мастицкий С.Э., Шитиков В.К. (2014) Статистический анализ и визуализация данных с помощью R. – Электронная книга <http://www.ievbras.ru/ecostat/Kiril/R/Mastitsky%20and%20Shitikov%202014.pdf>
- Ю. Е. Воскобойников, Е.И. Тимошенко Математическая статистика (с примерами в Excel) учебное пособие. http://window.edu.ru/resource/305/63305/files/stat_excel.pdf
3. R Studio, сайт разработчика <https://www.rstudio.com/products/rstudio/download/>

Вопросы

1. Каковы исходные данные задачи?
2. Что такое выборочная совокупность? Генеральная совокупность?
3. Какое основное требование к выборочной совокупности? Как его выполнить?
4. Для чего выборка представляется в виде интервального вариационного ряда?
5. Как выбирается число интервалов?
6. Как вычисляется варианта, частота, относительная накопленная частота и ранг?
7. Для чего строится гистограмма частот? Как строится гистограмма?
8. Что такое числовые характеристики случайной величины?
 1. Что они характеризуют?
 2. Расскажите об основных числовых характеристиках
9. Как можно описать случайную величину помимо числовых характеристик.
10. Что такое функция распределения?
11. Что такое диаграмма размаха?
 1. Что характеризует?
 2. Что такое выброс?
12. Что такое статистическая гипотеза?
13. Что такое нулевая гипотеза? Альтернативная? Уровень значимости?
14. Каков алгоритм проверки статистических гипотез?
 1. Способы проверки каких статистических гипотез вы знаете?
 2. Изобразите критическую область и область принятия решения
15. Что такое p-value?
16. Чем отличается теоретическое распределение от эмпирического?
 1. Как подбирались параметры теоретического распределения?
17. Что такое оценки? Какие они бывают?
18. Что такое интервальные оценки? Как вычисляются границы интервалов?
19. Какие существуют статистические критерии сравнения выборок?
20. Как проверялась гипотеза о равенстве мат. Ожиданий? О чем говорят результаты проверки?
21. Что такое линейный коэффициент корреляции?
 1. Зачем проверять его значимость?
 2. Как проверить его значимость?
22. Что такое статистическая зависимость?
 1. Какие ещё зависимости бывают?
 2. Приведите примеры зависимостей
23. Что показывает диаграмма рассеивания?
24. Что такое регрессия?
25. Как строится линия регрессии?
26. Проверить гипотезы используя p-value.