ГУАП

КАФЕДРА № 41

ОТЧЕТ   
ЗАЩИЩЕН С ОЦЕНКОЙ

ПРЕПОДАВАТЕЛЬ

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| старший преподаватель |  |  |  | Е. К. Григорьев |
| должн., уч. степень, звание |  | подпись, дата |  | инициалы, фамилия |

|  |
| --- |
| ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ № 5. |
| СТАТИСТИЧЕСКИЙ АНАЛИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ |
| по курсу: МОДЕЛИРОВАНИЕ. |
|  |
|  |

РАБОТУ ВЫПОЛНИЛ

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| СТУДЕНТ ГР. № | 4217 |  |  |  | | У. А. Мазориев |
|  |  |  | подпись, дата | |  | инициалы, фамилия |

**Цель работы**

Получить навыки обработки выборочных данных средствами Python.

**Результат выполнения работы**

Программный код, файл с расчетами в формате xlsl можно посмотреть, перейдя по ссылке на GitHub: <https://github.com/Mrx112426/Modelirovanie/tree/main>

**Ход выполнения работы**

Индивидуальное задание: Вариант 10

1. Решение аналитическим способом

Для начала определяться нужное число интервалов для нашей выборки, наша выборка состоит из 50 элементов, в качестве формулы для вычисления оптимального количества интервалов было выбрано соотношение: z = , где N – количество элементов, т.е. z = (рис. 1).

Изображение выглядит как текст, снимок экрана, число, Шрифт

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 1 – Количество интервалов

Далее вычисляется размах выборки, для этого вычисляется разница между максимальным и минимальным значением (рис. 2).

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 2 – Размах выборки

Длина интервала вычисляется по формуле R/z = 15.81/7=2.26 (рис. 3).

Изображение выглядит как текст, снимок экрана, линия, число

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 3 – Длина интервала

Далее формируется таблица с границами интервалов, средними значениями, частотой и относительной частотой (рис. 4):

1. Количество интервалов известно из предыдущих подсчетов
2. Для того чтобы вычислить границы берется минимальное значение – начало для первого интервала, чтобы найти его край нужно прибавить длину интервала, следующий интервал будет начинаться с конца предыдущего, причем если первый интервал [-6,51;4;25), то следующий [-4;25;-1;99), последний интервал включает две границы в себя
3. Для расчета середина интервала берется среднее значение между границами
4. Чтобы определить частоту нужно подсчитать количество значений, который входят в интервал
5. Для подсчета относительной частоты используется формула 1/N \* w(i) где w(i) – частота, а N – кол-во значений

Изображение выглядит как текст, снимок экрана, число, Шрифт

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 4 – Таблица интервалов

После вычисляется значения эмпирической функции распределения (рис. 5)

Изображение выглядит как текст, снимок экрана, число, Шрифт

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 5 – Значения ЭФР

Далее строится график гистограммы выборки (рис. 6).

Изображение выглядит как текст, диаграмма, снимок экрана, число

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 6 – График гистограммы выборки

После строится график полигона частот (рис. 7).

Изображение выглядит как линия, График, диаграмма, скат

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 7 – График полигона частот

Далее строится график эмпирической функции распределения (рис. 8).

Изображение выглядит как текст, рукописный текст, чернила

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 8 – График ЭФР

После считается выборочное среднее M(x), медиана, мода (самое часто-встречающееся значения в выборке. Для подсчета медианы в выборки с четным N используется формула , где – элемент в отсортированный выборке (рис. 9)

Изображение выглядит как текст, снимок экрана, Шрифт, число

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 9 – Подсчитанные значения

Далее считается дисперсии выборки, кф асимметрии и эксцесса для их расчета используется формулы на рисунке 10.

Изображение выглядит как текст, рукописный текст, Шрифт, диаграмма

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 10 – Формулы для расчета

Изображение выглядит как текст, снимок экрана, Шрифт, число

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 11 – Результат расчета

В конечном счете получаются результаты представленные на рисунке 12.

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 12 – Посчитанные значения аналитическим способом

1. **Центральная тенденция**

Среднее значение (0.41) близко к нулю, что говорит о сбалансированности положительных и отрицательных значений.

Медиана (-0.19) немного меньше среднего, что указывает на небольшую асимметрию в сторону меньших значений.

Мода (-1.38) показывает, что наиболее частое значение находится в отрицательной области.

Размах (15.81) показывает значительное расстояние между минимальным и максимальным значениями.

Дисперсия (9.37) и стандартное отклонение (3.06) указывают на достаточно большой разброс значений.

Коэффициент асимметрии (0.43) положительный, значит, распределение немного скошено вправо (есть больше крупных положительных выбросов).

Коэффициент эксцесса (0.21) положительный, что говорит о том, что выборка немного более "остроконечная", чем нормальное распределение (возможны частые выбросы).

Для наглядности визуализируются график QQ-Plot, который показывает, что распределение имеет небольшое отклонение от нормального, имеются выбросы (рис. 13).

Изображение выглядит как линия, График, текст, диаграмма

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 13 – График QQ-plot

1. Решение через Python

Далее пишется скрипт, который считает метрики посчитанные в 1 части, для проверки правильности подсчетов скрипт проверяется на выборки для 1 части, результаты приведены на рисунках 14–17.

Изображение выглядит как текст, диаграмма, снимок экрана, Шрифт

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 14 – Гистограмма выборки для 1 части

Изображение выглядит как линия, диаграмма, График, текст

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 15 – Гистограмма выборки для 1 части

Изображение выглядит как текст, диаграмма, линия, График

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 16 – ЭФР для 1 части

Изображение выглядит как текст, снимок экрана, Шрифт

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 17 – Результаты подсчетов для 1 части

Значения подсчитанные вручную совпадают со значениями в Python.

Далее подсчитываются метрики и строятся графики для 2 части (рис. 17 – рис. 20)

Изображение выглядит как снимок экрана, График, диаграмма, линия

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 18 – Гистограмма выборки для 2 части

Изображение выглядит как текст, диаграмма, снимок экрана, График

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 19 – Гистограмма выборки для 2 части

Изображение выглядит как текст, диаграмма, График, линия

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 20 – ЭФР для 1 части

Изображение выглядит как текст, снимок экрана, Шрифт

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 21 – Результаты подсчетов для 2 части

Среднее (10.73) и медиана (10.64) очень близки, что говорит об относительной симметричности распределения.

Мода (2.34) сильно отличается от среднего и медианы, что может свидетельствовать о наличии второго пика или скошенности распределения.

Размах (17.59) указывает на значительный диапазон значений.

Дисперсия (13.38) и стандартное отклонение (3.66) показывают, что данные довольно широко разбросаны вокруг среднего.

Коэффициент асимметрии близок к нулю, что означает почти симметричное распределение.

Коэффициент эксцесса (-0.61) меньше 0, что говорит о более "плоском" распределении по сравнению с нормальным (распределение имеет менее выраженный пик и более толстые хвосты).

На рисунке 22 показан график QQ-plot для второй выборки.

Изображение выглядит как линия, График, диаграмма, текст

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 22 – График QQ-Plot для второй части

**Вывод**

В ходе работы были получены навыки анализа выборок, их графического отображения и вычисления метрик для определения типа распределения. Первая выборка имеет положительный коэффициент асимметрии (0.43) и небольшой положительный эксцесс (0.21), что указывает на скошенность вправо и остроконечность распределения. QQ-Plot показывает отклонения от нормального распределения, особенно в хвостах. Первая выборка имеет логнормальное распределение.

Вторая выборка характеризуется почти нулевой асимметрией (0.087) и отрицательным эксцессом (-0.61), что говорит о симметричности и более «плоском» распределении по сравнению с нормальным. График QQ-plot опроверг гипотезу о нормальности, распределение треугольное.