

Тема 6. Описательные статистики для разных типов данных

Практические задания для самостоятельного выполнения

Задание выполняется по вариантам. Формулировки заданий общие для всех вариантов; конкретный набор данных, подлежащий изучению, выбирается в соответствии с номером своего варианта.

Результаты выполнения задания необходимо представить в виде двух файлов:

- 1) ноутбук в формате *ipynb*, содержащий программный код, результаты его выполнения, а также все необходимые пояснения, выводы и комментарии (в текстовых ячейках);
- 2) файл в формате *pdf* (или *html*), полученный путем экспорта (или вывода на печать) ноутбука из п. 1.

Внимание: в названии файлов должна обязательно присутствовать фамилия автора. Безымянные работы проверяться не будут.

Обратите внимание, что все необходимые для выполнения задания программные конструкции рассмотрены в учебных ноутбуках, размещенных в системе LMS. После изучения этих материалов выполнение задания не потребует больших усилий.

Максимальная оценка за выполнение задания вне аудитории – 1 балл. Дополнительные баллы (от 0 до 3) можно будет получить на следующем практическом занятии по результатам тестирования.

Внимание: самостоятельное и вдумчивое выполнение задания серьезно повышает вероятность успешного прохождения теста (будет проверяться понимание работы принципов работы с инструментарием и, в частности, умение понимать программный код).

Задание 1

Создать выборку из генеральной совокупности, образованной значениями дискретно распределенной случайной величины (дискретное распределение и его параметры выбрать самостоятельно) малого объема ($n \leq 30$). Выполнить исследование полученной выборки. Для этого:

1. Построить полигон частот.
2. Построить полигон относительных частот и теоретический многоугольник распределения на одном графике. Прокомментировать полученные результаты (в текстовой ячейке).
3. Написать функцию для вычисления значений эмпирической функции распределения.
4. На одном графике построить эмпирическую и теоретическую функцию распределения. Прокомментировать полученные результаты (в текстовой ячейке).

5. Вычислить выборочное среднее, выборочную моду и выборочную медиану. Сопоставить полученные значения с математическим ожиданием и медианой случайной величины (вычислить эти значения или использовать известные значения для данного закона распределения). Прокомментировать полученные результаты (в текстовой ячейке).
6. Вычислить выборочную дисперсию и исправленную дисперсию. Сопоставить полученные значения с дисперсией случайной величины (вычислить это значение или использовать известное значение для данного закона распределения). Прокомментировать полученные результаты (в текстовой ячейке).
7. Вычислить выборочное среднее квадратическое отклонение и исправленное среднее квадратическое отклонение. Сопоставить полученные значения со средним квадратическим отклонением случайной величины (вычислить это значение или использовать известное значение для данного закона распределения). Прокомментировать полученные результаты (в текстовой ячейке).

Задание 2

Создать выборку из генеральной совокупности, образованной значениями непрерывно распределенной случайной величины (непрерывное распределение и его параметры выбрать самостоятельно) большого объема ($n \gg 30$). Выполнить исследование полученной выборки. Для этого:

1. Построить гистограмму частот.
2. Построить гистограмму относительных частот и теоретическую плотность распределения на одном графике. Прокомментировать полученные результаты (в текстовой ячейке).
3. На одном графике построить эмпирическую и теоретическую функцию распределения. Прокомментировать полученные результаты (в текстовой ячейке).
4. Вычислить выборочное среднее и выборочную медиану. Сопоставить полученные значения с математическим ожиданием и медианой случайной величины (вычислить эти значения или использовать известные значения для данного закона распределения). Прокомментировать полученные результаты (в текстовой ячейке).
5. Вычислить выборочную дисперсию и исправленную дисперсию. Сопоставить полученные значения с дисперсией случайной величины (вычислить это значение или использовать известное значение для данного закона распределения). Прокомментировать полученные результаты (в текстовой ячейке).
6. Вычислить выборочное среднее квадратическое отклонение и исправленное среднее квадратическое отклонение. Сопоставить полученные значения со

средним квадратическим отклонением случайной величины (вычислить это значение или использовать известное значение для данного закона распределения). Прокомментировать полученные результаты (в текстовой ячейке).

Задание 3

Используя набор данных, выполнить исследование имеющихся в нем признаков. Наборы данных по вариантам представлены в csv-файлах. Имя файла: *Вариант N.3*, где *N* – номер варианта.

Для решения задания необходимо выполнить следующие шаги:

1. Импортировать данные наблюдений из файла. Вывести несколько первых записей для проверки корректности импорта и получения первого представления о значениях признаков.
2. Для каждого из признаков по описанию, данному в варианте, и по наблюдаемым значениям определить тип данных:
 - является ли признак категориальным или количественным,
 - для категориальных признаков – является ли признак номинальным, бинарным или порядковым,
 - для количественных признаков – является ли признак дискретным или непрерывным.

Результаты вместе с объяснениями записать в текстовой ячейке.

3. Выполнить визуализацию статистического распределения признака, соответствующую типу этого признака.
4. Для каждого признака вычислить те из статистических оценок, которое допустимы для вида этого признака:
 - выборочная мода,
 - выборочные первый и третий квартили,
 - выборочная медиана,
 - выборочная средняя,
 - выборочная дисперсия и/или исправленная дисперсия,
 - выборочное среднее квадратическое отклонение и/или исправленное среднее квадратическое отклонение.
5. Для каждого признака объяснить выбор оценок и дать интерпретацию полученным оценкам (записать в текстовой ячейке).