*Федеральное государственное бюджетное образовательное учреждение* *высшего образования*

|  |  |
| --- | --- |
| **Изображение выглядит как эмблема, герб, нашивка, символ  Автоматически созданное описание** | ***«Московский государственный технический университет  имени Н.Э. Баумана»***  ***(национальный исследовательский университет)***  ***(МГТУ им. Н.Э. Баумана)*** |

**ФАКУЛЬТЕТ** ФУНДАМЕНТАЛЬНЫЕ НАУКИ\_\_\_\_\_\_\_\_

**КАФЕДРА** ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА И МАТЕМАТИЧЕСКАЯ ФИЗИКА (ФН11)\_

**НАПРАВЛЕНИЕ ПОДГОТОВКИ** МАТЕМАТИКА И КОМПЬЮТЕРНЫЕ НАУКИ (02.03.01)

**Отчет**

**по лабораторной работе № \_1\_**

**Название лабораторной работы: Первоначальная обработка статистических данных**

**Вариант № 2**

**Дисциплина:** Теория вероятности и математическая статистика

Студент группы ФН11-52Б **\_\_\_\_\_\_\_\_\_\_\_\_\_ Кожемякин Г.А.**

(Подпись, дата) (И.О. Фамилия)

Преподаватель  **\_\_\_\_\_\_\_\_\_\_\_\_\_ Облакова Т.В.**

(Подпись, дата) (И.О. Фамилия)

Москва, 2024

**Задание**

По предложенной выборке, заданной для удобства в виде матрицы Х:

1) найдите крайние члены вариационного ряда и размах выборки;

2) осуществите группировку данных (количество интервалов находим по правилу Стерджеса);

3) по сгруппированным данным постройте гистограмму относительных частот;

4) вычислите выборочное среднее и выборочную дисперсию;

5) по виду гистограммы определите возможный закон распределения, оцените параметры этого закона по методу моментов или методу максимального правдоподобия, постройте совмещенные графики гистограммы и плотности предполагаемого закона;

6) найдите эмпирическую функцию распределения и постройте совмещенные графики эмпирической и теоретической функций распределения.

**Исходные данные**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 8.432 | 6.077 | 8.982 | 2.897 | 5.784 | 29.159 | 14.564 | 9.869 | 11.621 | 10.405 |
| 11.589 | 15 | 8.659 | 15.142 | 5.345 | 7.69 | 16.504 | 8.742 | 2.834 | 9.846 |
| 13.697 | 13.167 | 6.846 | 6.727 | 12.614 | 3.655 | 19.627 | 1.572 | 11.4 | 8.384 |
| 4.807 | 7.897 | 13.881 | 6.244 | 13.188 | 12.419 | 8.279 | 9.37 | 7.233 | 15.503 |
| 13.47 | 8.511 | 10.134 | 18.831 | 9.422 | 3.715 | 8.63 | 6.99 | 10.187 | 17.491 |
| 8.262 | 11.019 | 13.02 | 7.619 | 5.1 | 7.757 | 7.772 | 16.063 | 8.736 | 12.214 |
| 7.998 | 5.651 | 5.225 | 8.525 | 12.935 | 9.312 | 5.12 | 6.146 | 19.032 | 15.049 |
| 9.643 | 8.633 | 4.367 | 3.91 | 5.957 | 6.345 | 14.577 | 9.878 | 13.658 | 4.104 |
| 5.752 | 5.693 | 15.207 | 6.776 | 5.055 | 4.96 | 35.182 | 14.399 | 13.383 | 14.144 |
| 3.77 | 6.394 | 11.695 | 8.474 | 13.196 | 7.636 | 11.219 | 10.463 | 6.802 | 10.135 |

**Выполнение задания**

Работа выполнена при помощи языка программирования Python с использованием библиотек для анализа данных: Numpy, Scipy, Matplotlib.

Находим крайние члены вариационного ряда и размах выборки.

Максимум: 35,182

Минимум: 1,572

Размах выборки: Максимум – Минимум = 33,61

Находим число интервалов группировки по правилу Стерджиса:

Ширина интервала: Размах выборки / число интервалов 4,8

Таким образом, была проведена группировка статистических данных. Результатом группировки является интервальный вариационный ряд.

|  |  |  |  |
| --- | --- | --- | --- |
| Интервал | Середина интервала | Частота | Относительная частота |
| [1,57;6,37) | 3,97 | 25 | 0,25 |
| [6,37;11,17) | 8,77 | 40 | 0,4 |
| [11,17;15,98) | 13,58 | 27 | 0,27 |
| [15,98;20,78) | 18,38 | 6 | 0,06 |
| [20,78;25,58) | 23,18 | 0 | 0 |
| [25,58;30,38) | 27,98 | 1 | 0,01 |
| [30,38;35,18) | 32,78 | 1 | 0,01 |

Осуществляем группировку и строим гистограмму относительных частот, используя встроенные функции.

Изображение выглядит как текст, диаграмма, График, снимок экрана

Автоматически созданное описание

Находим выборочные характеристики (выборочное среднее и выборочную дисперсию):

Имеем, выборочное среднее примерно равно 9,97, выборочная дисперсия примерно равна 26,18.

По виду гистограммы заключаем, что частота максимальна для второго интервала и убывает немонотонно по выборке. Таким образом теоретический закон распределения выборки определенно не является показательным, гауссовским или равномерным. Распределение эмпирических частот похоже на гамма-распределение.

Оценка параметров гамма-распределения.

Приравниваем c и с , решаем системы уравнений и находим параметры , :

Имеем численные значения параметров гамма-распределения: .

Строим совмещенные графики гистограммы и теоретической плотности закона гамма-распределения с полученными параметрами.

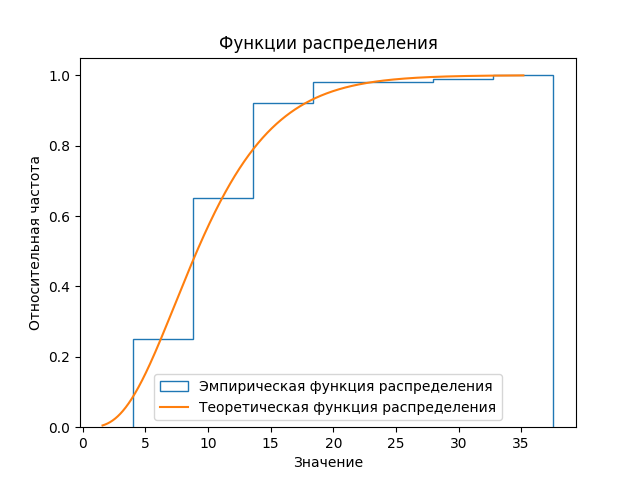
Изображение выглядит как диаграмма, текст, График, снимок экрана

Автоматически созданное описание

Построим эмпирическую функцию распределения по сгруппированным данным из таблицы. Для этого рассчитываем накопленные частоты. Они определяются путем последовательного суммирования относительных частот интервалов. Эмпирическая функция распределения для интервального вариационного ряда принимает вид:

где – количество интервалов, - середина i-го интервала, – относительные частоты.

Построим совмещенные графики эмпирической и теоретической функции распределения.



**Выводы:** в ходе работы были найдены крайние члены вариационного ряда, определен размах выборки, построена гистограмма относительных частот для выборки из ста элементов. Первоначальная обработка результатов позволяют предварительно отнести выборку к гамма-распределению с параметрами .

**Приложение**

Программный код на языке Python, с помощью которого выполнена лабораторная работа.

**import** math  
**import** matplotlib.pyplot **as** plt  
**import** numpy **as** np  
**from** scipy.stats **import** gamma  
  
Data = [[8.432, 6.077, 8.982, 2.897, 5.784, 29.159, 14.564, 9.869, 11.621, 10.405],  
 [11.589, 15, 8.659, 15.142, 5.345, 7.69, 16.504, 8.742, 2.834, 9.846],  
 [13.697, 13.167, 6.846, 6.727, 12.614, 3.655, 19.627, 1.572, 11.4, 8.384],  
 [4.807, 7.897, 13.881, 6.244, 13.188, 12.419, 8.279, 9.37, 7.233, 15.503],  
 [13.47, 8.511, 10.134, 18.831, 9.422, 3.715, 8.63, 6.99, 10.187, 17.491],  
 [8.262, 11.019, 13.02, 7.619, 5.1, 7.757, 7.772, 16.063, 8.736, 12.214],  
 [7.998, 5.651, 5.225, 8.525, 12.935, 9.312, 5.12, 6.146, 19.032, 15.049],  
 [9.643, 8.633, 4.367, 3.91, 5.957, 6.345, 14.577, 9.878, 13.658, 4.104],  
 [5.752, 5.693, 15.207, 6.776, 5.055, 4.96, 35.182, 14.399, 13.383, 14.144],  
 [3.77, 6.394, 11.695, 8.474, 13.196, 7.636, 11.219, 10.463, 6.802, 10.135]]  
  
*# Крайние члены вариационного ряда*max\_Data = np.max(Data)  
min\_Data = np.min(Data)  
  
*# Размах выборки*omega = max\_Data - min\_Data  
  
print(**'Находим крайние члены вариационного ряда.'**)  
print(**f'Максимум: {**max\_Data**} \nМинимум: {**min\_Data**}'**)  
print(**f'Размах выборки: {**omega**}'**)  
  
*# Число интервалов группировки по правилу Стерджиса*l = math.trunc(1 + math.log(100, 2))  
print(**f'Число интервалов: {**l**}'**)  
  
*# Построение гистограммы относительных частот*h = (max\_Data - min\_Data) / l  
print(**f'Ширина интервала: {**round(h, 3)**}'**)  
  
print(**'Границы интервалов:'**)  
intervals = [min\_Data **for** i **in** range(l+1)]  
**for** i **in** range(l+1):  
 intervals[i] = intervals[i] + i \* h  
 print(round(intervals[i], 2), end=**' - '**)  
  
print(**'\nСередины интервалов:'**)  
middles = [0 **for** i **in** range(l)]  
**for** i **in** range(l):  
 middles[i] = intervals[i] + h / 2  
 print(round(middles[i], 2), end=**' - '**)  
  
frequencies = [0 **for** i **in** range(l)]  
  
**for** i **in** range(l):  
 **for** j **in** range(10):  
 **for** k **in** range(10):  
 **if** intervals[i] <= Data[j][k] <= intervals[i+1]:  
 frequencies[i] += 1  
  
rel\_frequencies = [0 **for** i **in** range(l)]  
  
print(**'\nКоличество попаданий в интервал:'**)  
**for** i **in** range(l):  
 print(**f'{**i+1**}: {**frequencies[i]**}'**)  
 rel\_frequencies[i] = frequencies[i] / 100  
print(**'Относительные частоты:'**)  
**for** i **in** range(l):  
 print(**f'{**i+1**}: {**rel\_frequencies[i]**}'**)  
  
midpoints = [(intervals[i] + intervals[i+1]) / 2 **for** i **in** range(len(intervals)-1)]  
  
plt.bar(midpoints, rel\_frequencies, width=(intervals[1] - intervals[0]), align=**'center'**, label=**'Гистограмма'**, edgecolor=**'black'**)  
  
plt.plot(midpoints, rel\_frequencies, linestyle=**'-'**, color=**'red'**, label=**'Полигон'**)  
  
plt.xlabel(**"Значения"**)  
plt.ylabel(**"Относительная частота"**)  
plt.title(**'Гистограмма относительных частот'**)  
plt.show()  
  
*# Выборочное среднее и выборочная дисперсия*Data\_mean = np.mean(Data)  
  
Data\_var = 0  
  
**for** i **in** range(10):  
 **for** j **in** range(10):  
 Data\_var += (Data[i][j] - Data\_mean) \*\* 2  
  
Data\_var /= 99  
  
print(**f'Выборочное среднее: {**round(Data\_mean, 3)**}'**)  
print(**f'Выборочная дисперсия: {**round(Data\_var, 3)**}'**)  
  
**''' По виду гистограммы заключаем, что распределение эмпирических частот похоже на гамма-распределение  
Оценка параметра  
E(X) = lambda / alpha, D(X) = lambda / (alpha)^2 '''**lam = (Data\_mean \* Data\_mean) / Data\_var  
alpha = Data\_mean / Data\_var  
  
print(**'Оценка параметров'**)  
print(**f'Лямбда: {**round(lam, 3)**}'**)  
print(**f'Альфа: {**round(alpha, 3)**}'**)  
  
*# Построение гамма-распределения*Gam = gamma.rvs(a=lam, scale=1/alpha, size=10000)  
hist, bin\_edges = np.histogram(Gam, bins=intervals, density=**True**)  
hist = hist / np.sum(hist)  
plt.figure(figsize=(10, 6))  
plt.bar(bin\_edges[:-1], hist, width=np.diff(bin\_edges), align=**'edge'**, edgecolor =**'black'**)  
x = np.linspace(min(intervals), max(intervals), 1000)  
pdf = gamma.pdf(x, a=lam, scale=1/alpha)  
pdf \*= h  
plt.plot(x, pdf, color = **'red'**)  
plt.xlabel(**'Значения'**)  
plt.ylabel(**'Относительная частота'**)  
plt.title(**'Гистограмма и график гамма-распределения'**)  
plt.show()  
  
*# Построение эмпирической функции распределения.  
# Строим ступенчатую функцию распределений*Array = []  
**for** i **in** range(10):  
 **for** j **in** range(10):  
 Array.append(Data[i][j] + round(h / 2, 3))  
Array.sort()  
plt.hist(Array, histtype=**'step'**, cumulative=**True**, bins=7, density=**True**, label=**'Эмпирическая функция распределения'**)  
  
x = np.linspace(min(intervals), max(intervals), 1000)  
cdf = gamma.cdf(x, a=lam, scale=1/alpha)  
plt.plot(x, cdf, label=**'Теоретическая функция распределения'**)  
plt.xlabel(**'Значение'**)  
plt.ylabel(**'Относительная частота'**)  
plt.title(**'Функции распределения'**)  
plt.legend()  
plt.show()