**Гаязов Ильназ Дамирович, группа 09-012, отчет №1**

Раздел 1, задание 1

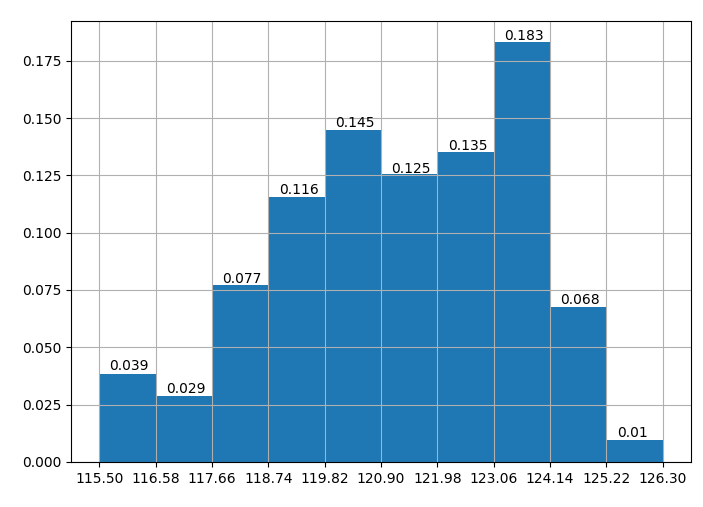
Описательная статистика

Постановка задачи: требуется произвести базовое статистическое исследование предоставленных данных: моментные и базовые характеристики выборки, гистограмма, эмпирическая функция распределения.

* 1. Вычисление характеристик выборки.

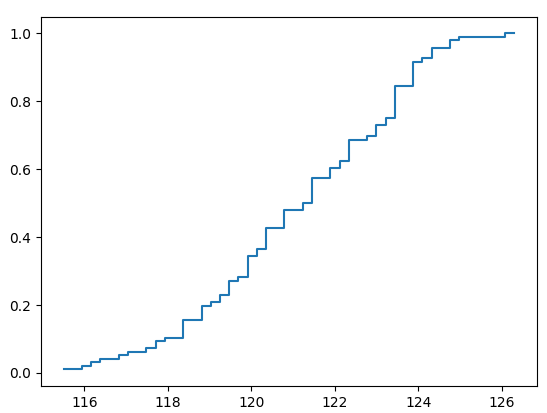
|  |  |
| --- | --- |
| **Характеристика** | **Значение** |
| Объём выборки | 96 |
| Минимум | 115.5 |
| Максимум | 126.3 |
| Размах | 10.8 |
| Среднее | 121.2 |
| Смещённая выборочная дисперсия | 5.5 |
| Несмещенная выборочная дисперсия | 5.6 |
| Стандартное отклонение | 2.3 |
| Асимметрия | -0.3 |
| Медиана | 121.4 |
| Интерквартильная широта | 3.9 |

1.2 Построение гистограммы.



Оценка моды распределения: 123.6

* 1. Построение графика эмпирической функции распределения



* 1. Код программы на Python

*# импорт библиотеки для работы с Excel***import** openpyxl  
**import** math  
*# библиотеки для работы с статистическими данными***import** numpy **as** np  
**import** matplotlib.pylab **as** plt  
**import** statsmodels.api **as** sm  
**from** statsmodels.distributions.empirical\_distribution **import** ECDF  
*# открыть файл для чтения*book = openpyxl.open(**'r1z1.xlsx'**, read\_only=**True**)sheet = book.active  
*# коприуем данные из листа в список*data = []  
**for** row **in** range(2, sheet.max\_row + 1):  
 data.append(sheet[row][0].value)  
*# объем выборки, минимум, максимум и размах*print(**"Объём выборки = "**, len(data))  
print(**"Минимум = "**, min(data))  
print(**"Максимум = "**, max(data))  
print(**"Размах = "**, round(max(data) - min(data), 1))  
*# среднее*sum = 0  
**for** x **in** data:  
 sum += x   
mean = round(sum / len(data), 1)  
print(**"Среднее = "**, mean)  
*# выборочная смещённая дисперсия*sum = 0  
**for** x **in** data:  
 sum += (x - mean)\*\*2  
disp = round(sum / len(data), 1)  
print(**"Смещённая выборочная дисперсия = "**, disp)  
*# выборочная несмещённая дисперсия*s2n = round(len(data) \* disp / (len(data) - 1), 1)  
print(**"Несмещённая выборочная дисперсия = "**, s2n)  
*# стандартное отклонение*s = round(math.sqrt(disp), 1)  
print(**"Стандартное отклонение = "**, s)  
*# асимметрия*sum = 0  
**for** x **in** data:  
 sum += (x - mean)\*\*3  
asymtr = round(sum / (len(data) \* s\*\*3), 1)  
print(**"Ассиметрия = "**, asymtr)  
*# медиана*n = len(data)  
sort\_data = sorted(data)  
**if** n % 2 == 0:  
 print(**"Медиана = "**, (round((sort\_data[n//2-1] + sort\_data[n//2])/2, 1)))  
**else**:  
 print(**"Медиана = "**, sort\_data[n // 2])  
*# интерквартильная широта***if** n % 4 == 0:  
 nkvart = round((sort\_data[n//4-1] + sort\_data[n//4])/2, 1)  
 vkvart = round((sort\_data[3\*n//4-1]+sort\_data[3\*n//4])/2, 1)  
**else**:  
 nkvart = sort\_data[n // 4]  
 vkvart = sort\_data[3 \* n // 4]  
print(**"Интерквартильная широта = "**, vkvart - nkvart)  
*#построение гистограммы*fig = plt.figure(figsize=(8, 6))  
ax = fig.add\_subplot()  
n, bins, pitches = ax.hist(data, bins=(math.ceil(n/10)), density=**True**)  
ax.xaxis.set\_major\_locator(plt.IndexLocator(base=1.08, offset=0))  
**for** i **in** pitches:  
 height = round(i.get\_height(), 3)  
 ax.annotate(height, (i.get\_x() + 0.2, height+0.001))  
ax.grid()  
plt.show()

*# эмпирическая функция распределения*ecdf = sm.distributions.ECDF(data)  
x = np.linspace(min(data), max(data))  
y = ecdf(x)

plt.step(x, y)  
plt.show()