University of Biskra
Department of Mathematics
First Year License
Descriptive Statistics

Detailed Solution – Exercise Sheet No. 3

Exercise 1

Given the distribution:

x_i	2	4	6	8	10
f_i	3	5	4	6	2

1) Total frequency

$$N = 3 + 5 + 4 + 6 + 2 = 20$$

2) Arithmetic mean

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{2 \cdot 3 + 4 \cdot 5 + 6 \cdot 4 + 8 \cdot 6 + 10 \cdot 2}{20}$$

$$\bar{x} = \frac{6 + 20 + 24 + 48 + 20}{20} = \frac{118}{20} = 5.9$$

3) Geometric mean

$$G = \left(\prod x_i^{f_i}\right)^{1/N} = \left(2^3 \cdot 4^5 \cdot 6^4 \cdot 8^6 \cdot 10^2\right)^{1/20}$$

$$G \approx 5.275$$

4) Harmonic mean

$$H = \frac{N}{\sum \frac{f_i}{x_i}} = \frac{20}{\frac{3}{2} + \frac{5}{4} + \frac{4}{6} + \frac{6}{8} + \frac{2}{10}}$$

$$H = \frac{20}{1.5 + 1.25 + 0.6667 + 0.75 + 0.2} = \frac{20}{4.3667} \approx 4.580$$

5) Variance

$$\sigma^2 = \frac{\sum f_i (x_i - \bar{x})^2}{N}$$

Since $\bar{x} = 5.9$,

$$\begin{aligned}
\sigma^2 &= \frac{3(2-5.9)^2 + 5(4-5.9)^2 + 4(6-5.9)^2 + 6(8-5.9)^2 + 2(10-5.9)^2}{20} \\
&= \frac{3(15.21) + 5(3.61) + 4(0.01) + 6(4.41) + 2(16.81)}{20} \\
&= \frac{45.63 + 18.05 + 0.04 + 26.46 + 33.62}{20} \\
&= \frac{123.80}{20} = 6.19
\end{aligned}$$

6) Standard deviation

$$\sigma = \sqrt{6.19} \approx 2.488$$

7) Mode

The largest frequency is 6, corresponding to $x = 8$. Hence,

$$Mo = 8$$

8) Comment

The mean is 5.9 and the standard deviation is about 2.49. Thus, the data are moderately dispersed around the mean.

Exercise 2

Given the grouped distribution:

Class	$[0, 4[$	$[4, 8[$	$[8, 12[$	$[12, 16[$	$[16, 20[$
Frequency	3	7	10	6	4

1) Class midpoints

$$m_1 = 2, \quad m_2 = 6, \quad m_3 = 10, \quad m_4 = 14, \quad m_5 = 18$$

2) Approximate mean

$$N = 3 + 7 + 10 + 6 + 4 = 30$$

$$\bar{x} \approx \frac{\sum f_i m_i}{N} = \frac{2 \cdot 3 + 6 \cdot 7 + 10 \cdot 10 + 14 \cdot 6 + 18 \cdot 4}{30}$$

$$\bar{x} = \frac{6 + 42 + 100 + 84 + 72}{30} = \frac{304}{30} \approx 10.133$$

3) Approximate variance

$$\sigma^2 \approx \frac{\sum f_i (m_i - \bar{x})^2}{N}$$

$$\begin{aligned}
\sigma^2 &\approx \frac{3(2 - 10.133)^2 + 7(6 - 10.133)^2 + 10(10 - 10.133)^2 + 6(14 - 10.133)^2 + 4(18 - 10.133)^2}{30} \\
&\approx \frac{3(66.151) + 7(17.084) + 10(0.018) + 6(14.951) + 4(61.884)}{30} \\
&\approx \frac{198.453 + 119.588 + 0.178 + 89.707 + 247.538}{30} \\
&\approx \frac{655.464}{30} \approx 21.849
\end{aligned}$$

4) Approximate standard deviation

$$\sigma \approx \sqrt{21.849} \approx 4.674$$

5) Coefficient of variation

$$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{4.674}{10.133} \times 100 \approx 46.13\%$$

6) Quartiles

First compute cumulative frequencies:

$$3, \quad 10, \quad 20, \quad 26, \quad 30$$

First quartile Q_1

$$\frac{N}{4} = \frac{30}{4} = 7.5$$

So Q_1 lies in the class $[4, 8[$.

$$Q_1 = 4 + \frac{7.5 - 3}{7} \cdot 4 = 4 + \frac{4.5}{7} \cdot 4 \approx 6.571$$

Second quartile Q_2

$$\frac{N}{2} = \frac{30}{2} = 15$$

So Q_2 lies in the class $[8, 12[$.

$$Q_2 = 8 + \frac{15 - 10}{10} \cdot 4 = 8 + 2 = 10$$

Third quartile Q_3

$$\frac{3N}{4} = \frac{3 \cdot 30}{4} = 22.5$$

So Q_3 lies in the class $[12, 16[$.

$$Q_3 = 12 + \frac{22.5 - 20}{6} \cdot 4 = 12 + \frac{2.5}{6} \cdot 4 \approx 13.667$$

7) Interquartile range

$$IQR = Q_3 - Q_1 \approx 13.667 - 6.571 = 7.096$$

8) Mean deviation

$$MD \approx \frac{\sum f_i |m_i - \bar{x}|}{N}$$

$$\begin{aligned}
MD &\approx \frac{3|2 - 10.133| + 7|6 - 10.133| + 10|10 - 10.133| + 6|14 - 10.133| + 4|18 - 10.133|}{30} \\
&\approx \frac{3(8.133) + 7(4.133) + 10(0.133) + 6(3.867) + 4(7.867)}{30} \\
&\approx \frac{24.399 + 28.931 + 1.333 + 23.202 + 31.468}{30} \\
&\approx \frac{109.333}{30} \approx 3.644
\end{aligned}$$

9) Modal class

The largest frequency is 10, so the modal class is

$$[8, 12[$$

10) Comment

The coefficient of variation is about 46.13%, which shows a fairly high relative dispersion. Therefore, the distribution is quite spread out.

Exercise 3

Given:

x_i	5	10	15	20	25
f_i	2	4	6	5	3

1) Total frequency

$$N = 2 + 4 + 6 + 5 + 3 = 20$$

2) Mean

$$\bar{x} = \frac{5 \cdot 2 + 10 \cdot 4 + 15 \cdot 6 + 20 \cdot 5 + 25 \cdot 3}{20}$$

$$\bar{x} = \frac{10 + 40 + 90 + 100 + 75}{20} = \frac{315}{20} = 15.75$$

3) Standard deviation

First, compute the variance:

$$\sigma^2 = \frac{\sum f_i (x_i - \bar{x})^2}{N}$$

$$\begin{aligned}
\sigma^2 &= \frac{2(5 - 15.75)^2 + 4(10 - 15.75)^2 + 6(15 - 15.75)^2 + 5(20 - 15.75)^2 + 3(25 - 15.75)^2}{20} \\
&= \frac{2(115.5625) + 4(33.0625) + 6(0.5625) + 5(18.0625) + 3(85.5625)}{20} \\
&= \frac{231.125 + 132.25 + 3.375 + 90.3125 + 256.6875}{20} \\
&= \frac{713.75}{20} = 35.6875
\end{aligned}$$

Thus,

$$\sigma = \sqrt{35.6875} \approx 5.974$$

4) Coefficient of variation

$$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{5.974}{15.75} \times 100 \approx 37.93\%$$

5) Homogeneous or heterogeneous?

Since the coefficient of variation is relatively large, the series is **heterogeneous**.

6) Justification

A coefficient of variation close to 38% indicates an important dispersion relative to the mean.

Exercise 4

Given the data set:

$$2,\ 3,\ 4,\ 4,\ 5,\ 6,\ 12$$

1) Mean

$$\bar{x} = \frac{2+3+4+4+5+6+12}{7} = \frac{36}{7} \approx 5.143$$

2) Median

Since there are 7 observations, the median is the 4^{th} value:

$$Me = 4$$

3) Mode

The most frequent value is 4, so

$$Mo = 4$$

4) Standard deviation

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{7}$$

$$\begin{aligned}
\sigma^2 &= \frac{(2-5.143)^2 + (3-5.143)^2 + (4-5.143)^2 + (4-5.143)^2 + (5-5.143)^2 + (6-5.143)^2 + (12-5.143)^2}{7} \\
&\approx \frac{9.878 + 4.592 + 1.306 + 1.306 + 0.020 + 0.735 + 47.020}{7} \\
&\approx \frac{64.857}{7} \approx 9.265
\end{aligned}$$

$$\sigma = \sqrt{9.265} \approx 3.044$$

5) First Pearson coefficient of skewness

$$Sk_1 = \frac{\bar{x} - Mo}{\sigma} = \frac{5.143 - 4}{3.044} \approx 0.375$$

6) Second Pearson coefficient of skewness

$$Sk_2 = \frac{3(\bar{x} - Me)}{\sigma} = \frac{3(5.143 - 4)}{3.044} \approx 1.126$$

7) Interpretation of the sign

Both coefficients are positive, so the skewness is positive.

8) Nature of the distribution

The distribution is **positively skewed** (right-skewed).

9) Comparison between mean, median, and mode

$$\bar{x} \approx 5.143, \qquad Me = 4, \qquad Mo = 4$$

Hence,

$$Mo = Me < \bar{x}$$

This confirms a right-skewed distribution.

10) Kurtosis parameters

The second central moment is

$$\mu_2 = \sigma^2 \approx 9.265$$

Now compute the fourth central moment:

$$\mu_4 = \frac{\sum (x_i - \bar{x})^4}{7}$$

$$\mu_4 \approx 333.360$$

Thus,

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{333.360}{(9.265)^2} \approx 3.883$$

$$\gamma_2 = \beta_2 - 3 \approx 3.883 - 3 = 0.883$$

11) Type of kurtosis

Since $\gamma_2 > 0$, the distribution is

leptokurtic

Exercise 5

Given:

x_i	2	4	6	8	10
f_i	2	5	8	4	1

1) Total frequency

$$N = 2 + 5 + 8 + 4 + 1 = 20$$

2) Mean

$$\bar{x} = \frac{2 \cdot 2 + 4 \cdot 5 + 6 \cdot 8 + 8 \cdot 4 + 10 \cdot 1}{20}$$

$$\bar{x} = \frac{4+20+48+32+10}{20} = \frac{114}{20} = 5.7$$

3) Necessary central moments

Second central moment

$$\mu_2 = \frac{\sum f_i(x_i-\bar{x})^2}{N}$$

$$\begin{aligned}
\mu_2 &= \frac{2(2-5.7)^2+5(4-5.7)^2+8(6-5.7)^2+4(8-5.7)^2+1(10-5.7)^2}{20} \\
&= \frac{2(13.69)+5(2.89)+8(0.09)+4(5.29)+1(18.49)}{20} \\
&= \frac{27.38+14.45+0.72+21.16+18.49}{20} \\
&= \frac{82.20}{20} = 4.11
\end{aligned}$$

Fourth central moment

$$\mu_4 = \frac{\sum f_i(x_i-\bar{x})^4}{N}$$

$$\begin{aligned}
\mu_4 &= \frac{2(2-5.7)^4+5(4-5.7)^4+8(6-5.7)^4+4(8-5.7)^4+1(10-5.7)^4}{20} \\
&= \frac{2(187.4161)+5(8.3521)+8(0.0081)+4(27.9841)+1(341.8801)}{20} \\
&= \frac{374.8322+41.7605+0.0648+111.9364+341.8801}{20} \\
&= \frac{870.474}{20} \approx 43.524
\end{aligned}$$

4) Kurtosis parameters

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{43.524}{(4.11)^2} \approx 2.577$$

$$\gamma_2 = \beta_2 - 3 \approx 2.577 - 3 = -0.423$$

5) Type of kurtosis

Since $\gamma_2 < 0$, the distribution is

platykurtic

6) Interpretation

The distribution is flatter than the normal distribution.