

Solutions

Exercise 1: Javelin Throw Competition Results

A javelin throw competition during an athletics meeting produced the following results:

Throw Length (m) x_i	n_i	C_i	CIn_i	$C_i \times n_i$
[60 – 65[7	62.5	7	437.5
[65 – 70[3	67.5	10	202.5
[70 – 75[6	72.5	16	435.0
[75 – 80[9	77.5	25	697.5
[80 – 85[2	82.5	27	165.0
Total	27			1937.5

such that CIn_i is the cumulative increasing frequency.

1. Calculate the mean \bar{m} of this series.

The mean \bar{m} is calculated using the formula:

$$\bar{m} = \frac{\sum (c_i \times n_i)}{\sum n_i}$$
$$\bar{m} = \frac{1937.5}{27} = 71,75$$

2. Calculate The Mode (Mo) of this series.

The mode is the value of the variable that has the highest partial frequency. For a continuous variable, the modal class is first defined. In our case, the modal class is [75 – 80[because it has the highest frequency 9.

The mode Mo is calculated using the formula:

$$Mo = a_i + \left(\frac{d_1}{d_1 + d_2} \right) \cdot e_i$$

where:

- $a_i = 75$ is the lower boundary of the modal class.
- $e_i = 5$ is the width of the modal class.

- $d_1 = 9 - 6$ is the difference between the frequency of the modal class and the frequency of the previous class.
- $d_2 = 9 - 2$ is the difference between the frequency of the modal class and the frequency of the following class.

Plugging these values into the formula, we find:

$$Mo = 75 + \left(\frac{3}{3 + 7} \right) \cdot 5 = 76.5$$

Therefore, the mode of this series is 76.5 meters.

3. Indicate the median class of this series.

For a continuous statistical series, the median class is determined. The i -th class $[a_i, a_{i+1}[$ is the median class if $CIn_{i-1} \leq \frac{N}{2} < CIn_i$. The median Me is calculated as follows:

$$Me = a_i + \left(\frac{\frac{N}{2} - CIn_{i-1}}{n_i} \right) e_i$$

where:

- a_i : lower boundary of the median class
- e_i : width of the median class
- CIn_{i-1} : cumulative frequency before the median class
- n_i : frequency of the median class

Applying the values from the provided table:

$$Me = 70 + \left(\frac{13.5 - 10}{6} \right) \times 5 \approx 72.92$$

4. percentage of athletes achieved a throw of at least 75 m

This percentage is calculated as:

$$\text{Percentage} = \frac{\text{Frequency of throws} \geq 75m}{\text{Total Frequency}} \times 100$$

Total number of athletes is $N = 27$

Number of athletes who threw at least 75 meters = $n_{\geq 75} = 9 + 2 = 11$.

Percentage of athletes who achieved a throw of at least 75 meters is

$$\left(\frac{n_{\geq 75}}{N}\right) \times 100 = \left(\frac{11}{27}\right) \times 100 \approx 40.74\%.$$

5. percentage of athletes achieved a throw of less than 70 m

This percentage is calculated as:

$$\text{Percentage} = \frac{\text{Frequency of throws } < 70m}{\text{Total Frequency}} \times 100$$

Number of athletes who threw less than 70 meters = $n_{<70} = 7 + 3 = 10$.

Total number of athletes = $N = 27$ (from previous calculation).

Percentage of athletes who achieved a throw of less than 70 meters is

$$\left(\frac{n_{<70}}{N}\right) \times 100 = \left(\frac{10}{27}\right) \times 100 \approx 37.04\%.$$

Exercise 2

The extended table with products of values and frequencies and their total sum:

(x_i)	(n_i)	$(n_i \times x_i)$	$n_i(x_i - \bar{x})^2$	CIn_i
13	3	39	65,4267	3
14	4	56	53,8756	7
15	7	105	49,9023	14
16	14	224	39,0446	28
17	16	272	7.1824	44
18	24	432	2.6136	68
19	15	285	26,5335	83
20	7	140	38,023	90
21	6	126	66.5334	96
22	4	88	74.9956	100
Total	100	1767	424.13	

I (a) **The mean of the data.**

Using the total sum from the third column for calculating the mean:

$$\bar{m} = \frac{\sum(n_i \times x_i)}{\sum n_i} = \frac{1767}{100} = 17.67$$

(b) **The variance of the data is calculated using the formula:**

$$\text{Var} = \frac{\sum n_i \times (x_i - \bar{m})^2}{\sum n_i}$$

Substituting the given totals from the table:

$$\text{Var} = \frac{424.13}{100} = 4.2413$$

Therefore, the variance of the dataset is 4.2413.

(c) **Standard Deviation (σ):** The standard deviation is calculated as the square root of the variance:

$$\sigma = \sqrt{\text{var}} = \sqrt{4.2413} \approx 2.059$$

(d) **Coefficient of Variation (CV):** The coefficient of variation is calculated as the standard deviation divided by the mean, then multiplied by 100 to convert it to a percentage:

$$CV = \left(\frac{\sigma}{\bar{m}}\right) \times 100\% = \left(\frac{2.059}{17.67}\right) \times 100\% \approx 11.655\%$$

Therefore, the standard deviation of the dataset is approximately 2.059, and the coefficient of variation is approximately 11.655%.

II We calculate the mode, median, and quartiles as follows:

(a) Mode Mo

the highest frequency n_i is 24. From the dataset: Mo= 18

(b) Median Me

Since the total frequency is 100 which is even , the median is the average of the 50th and 51th values. In this case:

Median = 18 (since the cumulative frequency reaches 50 at the value 18)

(c) Quartiles

The **first quartile** (Q_1) is the value at the $\left(\frac{N}{4}\right)^{th} = \left(\frac{100}{4}\right)^{th} = 25^{th}$. For this dataset:

$Q_1 = 16$ (since the cumulative frequency reaches 25^{th} at the value 16)

The **third quartile** (Q_3) is the value at the $\left(\frac{3N}{4}\right)^{th} = \left(\frac{3 \times 100}{4}\right)^{th} = 75^{th}$. For this dataset:

$Q_3 = 19$ (since the cumulative frequency reaches 75^{th} at the value 19)

III The range of the data series is the difference between the maximum and minimum values, which is calculated as:

$$\text{Range} = 22 - 13 = 9$$

The interquartile range (IQR) is the difference between the third quartile (Q_3) and the first quartile (Q_1), which is calculated as:

$$\text{IQR} = Q_3 - Q_1 = 19 - 16 = 3$$

Exercise 3

$$\bar{m} = \sum (c_i \cdot f_i)$$

$$\begin{aligned}\bar{m} &= (10 \times 10^3 \times 0.05) + (22.5 \times 10^3 \times 0.14) + (27.5 \times 10^3 \times 0.37) + (35 \times 10^3 \times 0.27) + (45 \times 10^3 \times 0.17) \\ &= 500 + 3,150 + 10,175 + 9,450 + 6,300 + 1,800 \\ &= 31,375 \times 10^3 \text{ DA}\end{aligned}$$

Therefore, the mean salary is $31,375 \times 10^3$ DA.

$$\text{Var} = (0.05 \times (10 \times 10^3 - 31,375 \times 10^3)^2) + \dots + (0.03 \times (60 \times 10^3 - 31,375 \times 10^3)^2)$$

$$\text{Var} = 93.546875 \times 10^6 \text{ DA}$$

Standard Deviation:

$$\sigma = \sqrt{Var} = \sqrt{93.546875 \times 10^6} = 9.67 \times 10^3 DA$$

Coefficient of Variation:

$$CV = \left(\frac{\sigma}{\bar{m}} \right) \times 100\% = \left(\frac{9.67 \times 10^3}{31.375 \times 10^3} \right) \times 100\% = 30.82\%$$

The table below is updated to include the frequency density for each salary range. The frequency density is calculated by dividing the frequency of each range by its width, helping to standardize the comparison across different interval sizes.

x_i (10^3 DA)	Width (10^3 DA)	c_i (10^3 DA)	f_i	Density
<20	20	10	0.05	0.0025
20-25	5	22.5	0.14	0.028
25-30	5	27.5	0.37	0.074
30-40	10	35	0.27	0.027
40-50	10	45	0.14	0.014
50-70	20	60	0.03	0.0015

The mode for a grouped data set can be more accurately determined using frequency densities, especially when the interval widths vary. Below, we outline the calculation for the mode using the corrected density values from the updated salary distribution table.

Assuming the modal class remains 25-30 thousands DA due to its highest frequency density, we use the formula:

$$Mo = a_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \cdot e_i$$

Where:

- (a) a_i is the lower limit of the modal class (25,000 DA).
- $\Delta_1 = f_i - f_{i-1} = (0.074 - 0.028)$.
- f_{i-1} is the frequency density of the class preceding the modal class (0.028).

- $\Delta_2 = f_i - f_{i-1} + 1 = (0.074 - 0.027)$.
- e_i is the width of the modal class (5,000 DA).

Plugging these values into the formula, we calculate the mode as follows:

$$\text{Mode} = 25 \times 10^3 + \left(\frac{(0.074 - 0.028)}{(0.074 - 0.028) + (0.074 - 0.027)} \right) \times 5 \times 10^3$$

$$\text{Mode} = 25 \times 10^3 + \left(\frac{0.046}{0.093} \right) \times 5 \times 10^3 = (25 + 2.473) \times 10^3 = 27.473 \times 10^3 \text{ DA}$$

This calculation confirms that the most typical salary, accounting for varying class widths, is around $27,473 \times 10^3 \text{ DA}$.

Median: The updated table with cumulative frequencies is presented below:

x_i (10^3 DA)	Width (10^3 DA)	c_i (10^3 DA)	f_i	Density	Cumulative Frequency F_i
<20	20	10	0.05	0.0025	0.05
20-25	5	22.5	0.14	0.028	0.19
25-30	5	27.5	0.37	0.074	0.56
30-40	10	35	0.27	0.027	0.83
40-50	10	45	0.14	0.014	0.97
50-70	20	60	0.03	0.0015	1.00

$Me = a_i + \left(\frac{\frac{1}{2} - F_{i-1}}{f_i} \right) e_i$ The total relative frequency F is 1 (as the frequencies are proportions summing to 1). The median class is identified from the cumulative frequency column where it first exceeds 0.5, which is the interval from 25,000 to 30,000 DA.

Applying the formula:

$$\text{Median} = 25 \times 10^3 + \left(\frac{0.5 - 0.19}{0.37} \right) \times 5 \times 10^3 = 25 \times 10^3 + \left(\frac{0.31}{0.37} \right) \times 5 \times 10^3 \approx 29,189 \times 10^3 \text{ DA}$$

This value, $29,189 \times 10^3 \text{ DA}$, represents the median salary, indicating the middle value of the salary distribution.

Calculations of Quartiles and Range

The distribution of salaries can be analyzed through quartiles and the overall range. Here's how we calculate these values for grouped data:

First and Third Quartiles The first quartile (Q_1) and third quartile (Q_3) are calculated using the following formula for each respective quartile:

$$Q_k = a_i + \left(\frac{\frac{k \cdot F}{4} - F_i}{f_i} \right) \times e_i$$

Given the calculated values:

- For Q_1 : Located in the class with boundaries [25,000 DA, 30,000 DA], with $a_i = 25,000$ DA, F_i (cumulative frequency up to the previous class) is 0.19, f_i (frequency of the class) is 0.37, and e_i (class width) is 5,000 DA.
- For Q_3 : Located in the class with boundaries [30,000 DA, 40,000 DA], with $a_i = 30,000$ DA, (F_i is 0.56, f_i is 0.27, and e_i is 10,000 DA.

Applying the formula:

$$Q_1 = 25,000 + \left(\frac{0.25 - 0.19}{0.37} \right) \times 10,000 \approx 26,62 \times 10^3 \text{ DA}$$

$$Q_3 = 30,000 + \left(\frac{0.75 - 0.56}{0.27} \right) \times 10,000 \approx 37,037 \text{ DA}$$

Interquartile Range (IQR) The interquartile range, which measures the spread in the middle 50% of the data, is:

$$IQR = Q_3 - Q_1 = 37,04 \times 10^3 - 26,62 \times 10^3 = 10,42 \times 10^3 \text{ DA}$$

Range The range, indicating the overall spread from the lowest to the highest salary, is calculated as:

$$R = \text{Maximum} - \text{Minimum} = 70,000 - 0 = 70,000 \text{ DA}$$

Exercise 4:

x_i	n_i	<i>Cumulative Frequency</i>
2	1	1
4	2	3
5	2	5
6	1	6
7	3	9
8	2	11
9	1	12
10	1	13

Graphical Determination of the Median

The cumulative frequency curve is plotted, and the median is determined by finding the data point where the cumulative frequency is half the total frequency+1 (7 in this case).

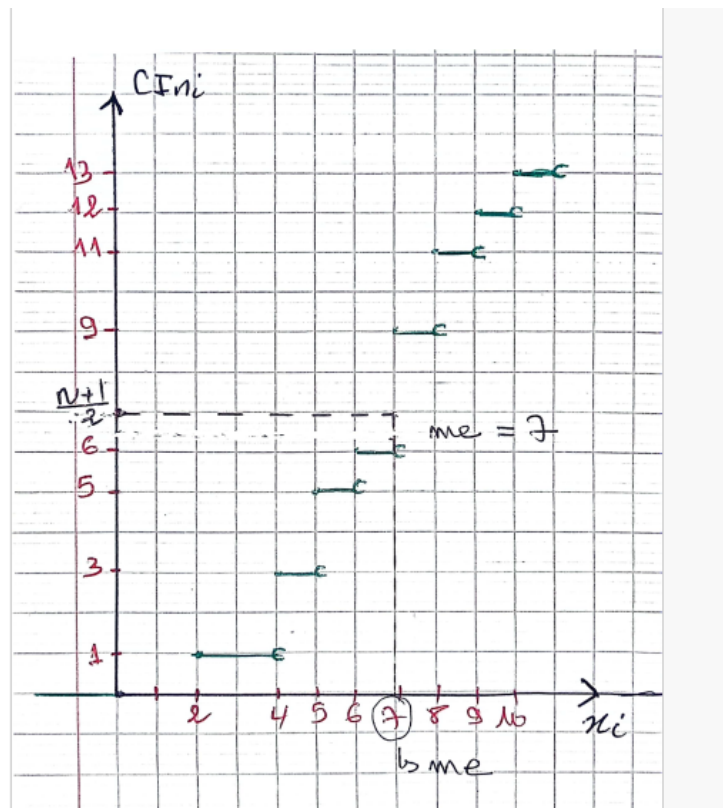


Figure 1: The curve and Median