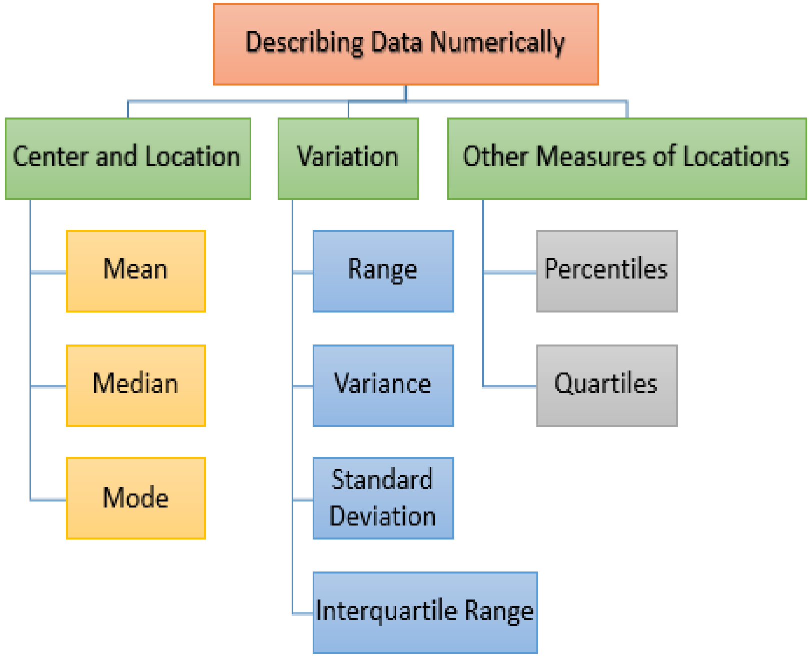
Introduction to Probability and Statistics

Topic (2): “Describing Data with Numerical Measures”



Dr. Heba Ayyoub

Philadelphia University

Solution to the Exercises for the Second Topic

Exercise (1): The mean of 10 numbers is 8. If an eleventh number is now added to the data, the mean becomes 9. What is the value of the new number?

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10}$$

$$8 = \frac{\sum_{i=1}^{10} x_i}{10}$$

$$\sum_{i=1}^{10} x_i = 80$$

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i + x_{11}}{11}$$

$$9 = \frac{80 + x_{11}}{11}$$

$$99 - 80 = x_{11}$$

$$x_{11} = 19$$

Exercise (2): Given that $n = 10$, $\bar{x} = 12$ and $\sum_{i=1}^{9} x_i = 100$, then x_{10} is:

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10}$$

$$12 = \frac{\sum_{i=1}^{9} x_i + x_9}{10}$$

$$12 = \frac{100 + x_9}{10}$$

$$120 - 100 = x_9$$

$$x_9 = 20$$

Exercise (3): A data set $5, c, 8, 2, 3, 7$, where c is unknown observation. Find the value of c if the sample mean is 5.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$5 = \frac{5 + c + 8 + 2 + 3 + 7}{6}$$

$$30 = 25 + c$$

$$c = 5$$

Exercise (4): The mean of n numbers is 5. If the number 12 is now removed from the n numbers, the mean is 4. Find the value of n.

$$\bar{x}_{old} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$5n = \sum_{i=1}^{n} x_i \quad \dots \quad (1)$$

$$\bar{x}_{new} = \frac{\sum_{i=1}^{n} x_i - 12}{n-1}$$

$$4 = \frac{\sum_{i=1}^{n} x_i - 12}{n-1}$$

$$4n - 4 + 12 = \sum_{i=1}^{n} x_i$$

From (1):

$$4n - 4 + 12 = 5n$$

$$8 = 5n - 4n$$

$$n = 8$$

Exercise (5): The mean of 4 numbers is 5, and the mean of 3 other numbers is 12. What is the mean of the 7 numbers together?

$$\bar{x}_1 = \frac{\sum_{i=1}^{4} x_i}{4}$$

$$5 = \frac{\sum_{i=1}^{4} x_i}{4}$$

$$\sum_{i=1}^{4} x_i = 20$$

$$\bar{x}_2 = \frac{\sum_{i=1}^{3} x_i}{3}$$

$$12 = \frac{\sum_{i=1}^{3} x_i}{3}$$

$$\sum_{i=1}^{3} x_i = 36$$

$$\bar{x} = \frac{\sum_{i=1}^{4} x_i + \sum_{i=1}^{3} x_i}{7}$$

$$\bar{x} = \frac{20 + 36}{7} \Rightarrow \bar{x} = 8$$

Exercise (6): For each of the following distributions, find the value of b for the given value of $\bar{x}$.

a) $\bar{x} = 1.05$

Value (x_i)	Frequency (f_i)	$x_i f_i$
-2	5	-10
0	2	0
1	b	b
3	9	27
	$16 + b$	$17 + b$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$$

$$1.05 = \frac{16 + b}{17 + b}$$

$$16.8 + 1.05b = 17 + b$$

$$1.05b - b = 17 - 16.8$$

$$0.05b = 0.2$$

$$b = 4$$

b) $\bar{x} = 5.4$

Value (x_i)	Frequency (f_i)	$x_i f_i$
2	5	10
b	2	$2b$
8	2	16
10	1	10
	10	$2b + 36$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$$

$$5.4 = \frac{2b + 36}{10}$$

$$54 - 36 = 2b$$

$$18 = 2b$$

$$b = 9$$

Exercise (7): Find the mean of the following continuous distribution.

Class	Frequency (f_i)	x_i	$x_i f_i$
(1 – 5)	9	$\frac{(1+5)}{2} = 3$	27
(6 – 10)	11	$\frac{(6+10)}{2} = 8$	88
(11 – 15)	16	$\frac{(11+15)}{2} = 13$	208
(16 – 20)	14	$\frac{(16+20)}{2} = 18$	252
	50		575

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{575}{50} = 11.5$$

Exercise (8): If the median of the data 9, 1, x, 12 is 7, what is the value of x?

$$1, x, 9, 12$$

$$\frac{x+9}{2} = 7$$

$$x = 5$$

Exercise (9): You are given n = 10 measurements: 3, 5, 4, 6, 10, 5, 6, 9, 2, 8.

a. Calculate the sample mean .

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{3+5+4+6+10+5+6+9+2+8}{10} = \frac{58}{10} = 5.8$$

b. Find median.

$$2, 3, 4, 5, 5, 6, 6, 8, 9, 10$$

$$\frac{5+6}{2} = 5.5$$

c. Find the mode.

5 and 6

Exercise (10): You are given n = 8 measurements: 4, 1, 3, 1, 3, 1, 2, 2.

a. Find the range.

$$R = 4 - 1 = 3$$

b. Calculate $\bar{x}$.

$$\bar{x} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{17}{8} = 2.125$$

c. Calculate S^2 and S.

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}$$

$$= \frac{45 - 8(2.125)^2}{7}$$

$$= \frac{8.875}{7} \approx 1.27$$

$$S = \sqrt{1.27} \approx 1.13$$

x_i	x_i^2
4	16
1	1
3	9
1	1
3	9
1	1
2	4
2	4
17	**45**

Exercise (11): For a set of 10 numbers, $\sum_{i=1}^{10} x_i$ = 270 and $\sum_{i=1}^{10} x_i^2$ = 9054. Find the mean and the variance.

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{270}{10} = 27$$

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{9054 - 10(27)^2}{9} = 196$$

Exercise (12): For a set of 18 numbers, $\bar{x}$ = 23 and S = 12. Find $\sum_{i=1}^{18} x_i$ and $\sum_{i=1}^{18} x_i^2$.

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10}$$

$$23 = \frac{\sum_{i=1}^{10} x_i}{18}$$

$$\sum_{i=1}^{10} x_i = 414$$

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}$$

$$12^2 = \frac{\sum_{i=1}^{n} x_i^2 - 18(23)^2}{17}$$

$$\sum_{i=1}^{n} x_i^2 = 11970.$$

Exercise (13): The numbers a, b, 6, 4, 7 have mean 5 and variance 4.

Find the values of a, b.

x_i	x_i^2
a	a^2
b	b^2
6	36
4	16
7	49
$\boldsymbol{a+b+17}$	$\boldsymbol{a^2+b^2+101}$

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i}{5}$$

$$5 = \frac{a+b+17}{5}$$

$$a + b = 8 \quad ....(1)$$

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}$$

$$4 = \frac{a^2+b^2+101 - 5(5)^2}{4}$$

$$a^2 + b^2 = 40 \quad ....(2)$$

From (1): $(a+b)^2 = 8^2$

$$a^2 + b^2 + 2ab = 64$$

From (2): $40 + 2ab = 64$

$$ab = 12$$

$$a = \frac{12}{b}$$

From (1): $a + b = 8$

$$\frac{12}{b} + b = 8$$

$$12 + b^2 = 8b$$

$$b^2 - 8b + 12 = 0$$

$$(b-2)(b-6) = 0$$

$$b = 2 \ \ or \ \ b = 6$$

If $b = 2, a = 6$

If $b = 6, a = 2$

Exercise (14): For a set of 20 numbers, $\sum_{i=1}^{20} x_i = 20$ and $\sum_{i=1}^{20} x_i^2 = 96$. For a second set of 30 numbers, $\sum_{i=1}^{30} x_i = 60$ and $\sum_{i=1}^{30} x_i^2 = 236$. Find the mean and standard deviation for the combined set of 50 numbers.

- The mean:

$$\bar{x} = \frac{\sum_{i=1}^{20} x_i + \sum_{i=1}^{30} x_i}{50} = \frac{20 + 60}{50} = 1.6$$

- The standard deviation:

$$S^2 = \frac{\left(\sum_{i=1}^{20} x_i^2 + \sum_{i=1}^{30} x_i^2\right) - n\bar{x}^2}{n-1}$$

$$= \frac{(96 + 236) - 50(1.6)^2}{49}$$

$$= 4.16$$

$$S = \sqrt{4.16} = 2.04$$

Exercise (15): Find the variance of the following distribution.

Value (x_i)	Frequency (f_i)	$x_i f_i$	$x_i^2 . f_i$
3	4	12	36
4	5	20	80
5	8	40	200
6	3	18	108
	20	**90**	**424**

$$S^2 = \frac{\sum_{i=1}^{n}(x_i^2 . f_i) - \frac{(\sum_{i=1}^{n} x_i f_i)^2}{\sum_{i=1}^{n} f_i}}{\sum_{i=1}^{n} f_i - 1} = \frac{424 - \frac{90^2}{20}}{19} = 1$$

Exercise (16): Given the following data: 340, 300, 520, 340, 320, 290, 260, 330.

1) Calculate the first quartile.

$$x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}, x_{(5)}, x_{(6)}, x_{(7)}, x_{(8)}$$

$$260, 290, 300, 320, 330, 340, 340, 520$$

$$g = \frac{p}{100}(n+1) = \frac{25}{100}(8+1) = 2.25$$

$$Q_1 = x_{(2.25)} = x_{(2)} + (2.25 - 2)\left(x_{(3)} - x_{(2)}\right) = 290 + 0.25(300 - 290) = 292.5$$

2) Calculate the median.

$$Q_2 = m = \frac{320 + 330}{2} = 325$$

3) Calculate the third quartile.

$$g = \frac{p}{100}(n+1) = \frac{75}{100}(8+1) = 6.75$$

$$Q_3 = x_{(6.75)} = x_{(6)} + (6.75 - 6)\left(x_{(7)} - x_{(6)}\right) = 340 + 0.75(340 - 340) = 340$$

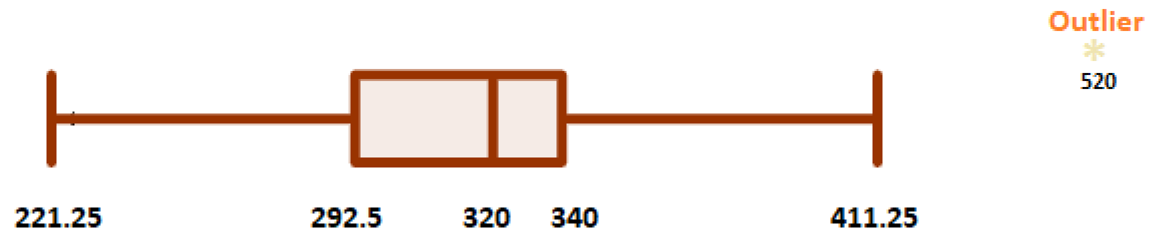
4) Calculate the Interquartile Quartile Range.

IQR = $Q_3 - Q_1 = 340 - 292.5 = 47.5$

5) Draw the box-plot.

Lower fence = $Q_1 - 1.5(IQR) = 292.5 - 1.5(47.5) = 221.25$

Upper fence = $Q_3 + 1.5(IQR) = 340 + 1.5(47.5) = 411.25$



6) Describe the shape of the distribution.

Skewed to the left

7) Determine if there is an outlier, and list the outlier if exist.

520

Exercise (17): True or False

1) The mean is sensitive to extreme values in a dataset. **True**

2) The median is the middle value of a sorted dataset and is unaffected by outliers. **True**

3) The mode is the value that occurs most frequently in a dataset. **True**

4) The range is the difference between the largest and smallest values in a dataset. **True**

5) The interquartile range (IQR) measures the spread of the middle 50% of the data. **True**

6) A data set can have no mode, one mode, or multiple modes. **True**

7) The variance and standard deviation both measure the spread or variability of the data. **True**

8) The standard deviation is equal to the square of the variance. **True**

9) A z-score tells how many standard deviations a data point is from the mean. **True**

10) In a skewed distribution, the mean is always closer to the peak than the median. **False**

11) The five-number summary includes the minimum, maximum, mean, median, and mode. **False**

12) A boxplot is used to show the spread and center of a dataset but does not display individual data points. **True**

13) The empirical rule applies to all datasets, regardless of the shape of the distribution. **False**

14) A negative z-score indicates a value below the mean. **True**

15) The mode of a dataset can only be determined from a bar chart. **False**