PIERIAN DATA

EXERCISE #1 – Data – ANSWER KEY

1. What is data?

   Data is all of the collected observations we have about something.

2. Why do we use visualizations with data?

   Charts and graphs help us see patterns in data that can't be seen in a table.

3. What is the difference between a population and a sample?

   A population describes all of the members of a group, while a sample is a subset of members.

4. Why do we use sampling?

   Time and resources prevent us from measuring every member of a population.

MEASUREMENTS OF DATA

5. What level of measurement describes an employee's education level?

   Ordinal (education levels can be sorted but lack a formal scale).

6. What level of measurement describes the time needed to complete a project?

   Ratio (values have scale and a true zero point).

MATHEMATICAL SYMBOLS & SYNTAX

7. Set up and solve 5^3

$$5^3 = 5 \times 5 \times 5 = 125$$

8. Set up and solve $5!$

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

9. Set up and solve $\sum_{x=1}^{5} x$

$$\sum_{x=1}^{5} x = 1 + 2 + 3 + 4 + 5 = 15$$

MEASURES OF CENTRAL TENDENCY

10. Find the mean value of the series $\{6, 12, 8, 5, 10\}$

$$\frac{6 + 12 + 8 + 5 + 10}{5} = 8.2$$

11. Find the median value of the series $\{7, 3, 11, 6, 9, 9\}$

1. *Sort the series:* $\{3,6,7,9,9,11\}$
2. *Find the middle value(s):* $[7,9]$
3. *If necessary, take the mean of the two middle values:*
$$\frac{7 + 9}{2} = \frac{16}{2} = 8$$

MEASURES OF DISPERSION

12. Find the standard deviation σ of the series $\{2, 10, 8, 6, 3, 7\}$

1. *Find the mean:* $\bar{x} = \dfrac{2 + 10 + 8 + 6 + 3 + 7}{6} = \dfrac{36}{6} = 6$
2. *Apply the formula for population standard deviation:*

$$\sigma = \sqrt{\frac{(2-6)^2 + (10-6)^2 + (8-6)^2 + (6-6)^2 + (3-6)^2 + (7-6)^2}{6}}$$

$$= \sqrt{\frac{16 + 16 + 4 + 0 + 9 + 1}{6}}$$

$$= \sqrt{\frac{46}{6}} = \sqrt{7.67} = \mathbf{2.77}$$

QUARTILES & INTERQUARTILE RANGE (IQR)

13. Divide the following series into quartiles. $\{5, 1, 6, 4, 2, 6, 7, 3, 1, 8, 4, 8\}$
    What are the boundaries of the IQR?

 1. *Sort the series:* $\{1,1,2,3,4,4,5,6,6,7,8,8\}$

 2. *Work from the outside toward the middle:* $\dfrac{\{\cancel{1},\cancel{1},\cancel{2},\cancel{3},\cancel{4}, 4, 5, \cancel{6},\cancel{6},\cancel{7},\cancel{8},\cancel{8}\}}{6\ 5\ 4\ \ 3\ 2\ 1\ !\ 1\ 2\ 3\ 4\ 5\ 6}$

 3. *Find the median value:* $\dfrac{4 + 5}{2} = \mathbf{4.5}$

 4. *Find the upper and lower median values:* $\dfrac{2 + 3}{2} = \mathbf{2.5}, \dfrac{6 + 7}{2} = \mathbf{6.5}$

 5. *These become the boundaries of the interquartile range:* $\mathbf{2.5\ to\ 6.5}$

14. In the above problem, where would the upper fence fall using the 1.5 IQR method?

 1. *Find the width of the IQR:* $6.5 - 2.5 = 4$
 2. *Multiply this by* 1.5: $4 \times 1.5 = 6$
 3. *Add this to the third quartile:* $6.5 + 6 = \mathbf{12.5}$
 Anything above 12.5 *would be considered an outlier*

BIVARIATE DATA

15. Calculate the Pearson Correlation Coefficient for the following table of values:
    We recommend using a spreadsheet!

Height	Weight	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
5	143	-5	-7	35	25	49
7	145	-3	-5	15	9	25
11	147	1	-3	-3	1	9
12	157	2	7	14	4	49
15	158	5	8	40	25	64
Sum: 50	750		**Sum:**	101	64	196
Mean: $\bar{x}$ = 10	$\bar{y}$ = 150					

$$\rho_{X,Y} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2}\sqrt{\Sigma(y - \bar{y})^2}} = \frac{101}{\sqrt{64}\ \sqrt{196}} = \frac{101}{112}$$

$$= \mathbf{0.90}$$

16. Are these values correlated? Why or why not?

A value of 0.90 indicates a strong positive correlation.