PIERIAN DATA

EXERCISE #1 - Data

1. What is data?

the collected obsevation we have about something

2. Why do we use visualizations with data?

helps people undestarnd data

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

population is the entire data, sample is just an extract of the population data

4. Why do we use sampling?

easy to work with

MEASUREMENTS OF DATA

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

Ordinal

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

Ratio

MATHEMATICAL SYMBOLS & SYNTAX

7. Set up and solve 5^3

   5*5*5=125

8. Set up and solve $5!$

   5*(5-1)*(5-2)*(5-3)*(5-4) = 120

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

   1+2+3+4+5 = 15

MEASURES OF CENTRAL TENDENCY

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

    (6+12+8+5+10)/5 = 8.2

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

    3,6,7,9,9,11 = (7 + 9)/2 = 8

MEASURES OF DISPERSION

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

    mean = (2+3+6+7+8+10)/6 = 6

    final = rootsqr((sqr(2-6) + sqr(3-6) + sqr(6-6) + sqr(7-6) + sqr(8-6) + sqr(10-6))/6-1))

    final = sqrroot((16 + 9 + 0 + 1 + 4 + 16)/5)

    final = sqrroot(9.2)

    final = 3.033

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?

(1,1,2,3,4,4,5,6,6,7,8,8)

1 quartile = 2,3 = (2+3)/2 = 2.5

2 quartile = 4,5 = (4+5)/2 = 4.5

3 quartile = 6,7 = (6+7)/2 = 6.5

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

upper fall = 12.5

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	35
	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	206
Mean:	$\bar{x}$ = 10	$\bar{y}$ = 150					

$$\rho_{X,Y} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}} = \frac{101}{\sqrt{64}\ \sqrt{206}} = \frac{101}{8*14.352 = 114.816} \quad 0.879$$

$=$

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

Yes, they are bewteen 1 and -1