SHUBHAM TRIBEDI / 1811100002037

EXERCISE #1 – Data

1. What is data?

Data is a collection of facts, such as numbers, words, measurements, observations, or just descriptions of things. Qualitative vs Quantitative. These are individual pieces of factual information recorded and used for the purpose of analysis. It is the raw information from which statistics are created.

2. Why do we use visualizations with data?

Data visualization helps to tell stories by curating data into a form easier to understand, highlighting the trends and outliers. A good visualization tells a story, removing the noise from data and highlighting the useful information.
Effective data visualization is a delicate balancing act between form and function.

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

A population is the entire group that you want to draw conclusions about.
A sample is the specific group that you will collect data from.
The size of the sample is always less than the total size of the population. In research, a population doesn't always refer to people.

4. Why do we use sampling?

Sampling means selecting the group that you will actually collect data from in your research. If I am researching the opinions of students in your university, I might be surveying a sample of 100 students.
In statistics, sampling allows us to test a hypothesis about the characteristics of a population.

MEASUREMENTS OF DATA

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

Ordinal and Nominal level of measurement. Because we can rank or order their education level as well as can be categorized distinctly.

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

The time needed to complete a project have Ordinal, Nominal and Interval level of measurement.
Nominal, because the time data can be categorized.
Ordinal, because the time data can be both categorized and ranked or ordered.
Interval, because the tim data can be seperated into different intervals.

MATHEMATICAL SYMBOLS & SYNTAX

SHUBHAM TRIBEDI / 1811100002037

7. Set up and solve 5^3

5^3 = 5x5x5 = (5x5)x5 = 25x5 = 125

8. Set up and solve $5!$

5! = 5x4x3x2x1 = 20x3x2x1 = 60x2x1 = 120x1 = 120

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

Summation of x = 1 to 5, x is 1+2+3+4+5 = 3+3+4+5 = 6+4+5 = 10+5 = 15

MEASURES OF CENTRAL TENDENCY

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

mean({6,12,8,5,10}) = (6+12+8+5+10)/5 = (18+8+5+10)/5 = (26+5+10)/5 = (31+10)/5 = 41/5 = 8.2

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

Median({7,3,11,6,9,9}) = (((n/2)th element) + (((n/2)+1)th element)) /2 considering the sorted array i.e. {3,6,7,9,9,11} = (7+9)/2 = 16/2 = 8. Hence the median is 8

SHUBHAM TRIBEDI / 1811100002037

MEASURES OF DISPERSION

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

Count, N: 6
Sum, Σx: 36
Mean, x̄: 6
Variance, s^2: 9.2

Steps

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2},$$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{N - 1}$$

$$= \frac{(2 - 6)^2 + \ldots + (7 - 6)^2}{6 - 1}$$

$$= \frac{46}{5}$$

$$= 9.2$$

$$s = \sqrt{9.2}$$

$$= 3.0331501776206$$

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?

Arrange data in order: {1,1,2,3,4,4,5,6,6,7,8,8} Median of the Arranged Data: 4+5/2 = 4.5

Lower Half of the Data According to the Median: {1,1,2,3,4,4}
Calculate first quartile (Q1): Mean of the Middle two values of the lower half. i.e. 2+3/2 = 2.5

Upper Half of the Data according to the median: {5,6,6,7,8,8}
Calculate third quartile (Q3): Mean of the Middle Two values of the upper half. i.e. 6+7/2 = 6.5

Calculate interquartile range (IQR)= Q3-Q1 = 6.5 - 2.5 = 4

Calculate lower boundary LB = Q1 - (1.5*IQR) = 2.5 - (1.5*4) = 2.5 - 6 = -4.5Calculate upper boundary UB = Q3 + (1.5*IQR) = 6.5 + (1.5*4) = 6.5 + 6 = 12.5

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

In the above problem the 1.5 IQR i.e. 1.5*4 = 6
So, our fences will be 6 below Q1 an 6 above Q3
Hence, Lower Fence = 2.5 - 6 = -3.5
Upper Fence = 6.5 + 6 = 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	25
	11	**147**	1	-3	-3	1	9
	12	**157**	2	7	14	4	49
	15	**158**	3	8	24	9	64
Sum:	50	750		**Sum:**	85	48	196
Mean:	$\bar{x}=$ 10	$\bar{y}=$ 150					

$$\rho_{X,Y} = \frac{\sum(x-x)(y-\bar{y})}{\sqrt{\Sigma(x-x)^2}\sqrt{\Sigma(y-\bar{y})^2}} = \frac{85}{\sqrt{48}\sqrt{196}} = \frac{85}{96.99}$$

$$= 0.8763$$

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

Yes, these values are correlated as the pearson correlation co-efficient's value is between 0 and 1. So they are correlated as per the rule.