LECTURE 1. Review of Descriptive Statistics

(Measures of Central Tendency , Measures of Variation , and Measures of Position)

Measures of Central Tendency

(Mean, Median, Mode)

Ungrouped Data

Measures of Central Tendency for Ungrouped Data

The number of hours of computer usage in a day of a random sample of 20 students was recorded as follows

2.5	3.2	2.5	1.5	6	1	3.5	3	1	1
4	3	2	2.5	1	1	2.5	4.5	4	3

Determine the value of each measure

Mean ($\bar{x}$) $\dfrac{2.5 + 3.2 + 2.5 + 1.5 + \dots + 3}{20} =$

Median

2.5		1			1		Hours	frequency	
3.2		1			1		1	5	mode (high
2.5		1			1		1.5	1	
1.5		1			1		2	1	
6		1			1		2.5	4	
1		1.5			1.5		3	4	
3		2			2		3.2	1	
3		2.5			2.5		4	2	
1		2.5			2.5		4.5	1	
1		2.5	median=(2.5+2.5)/2		2.5		6	1	
4		2.5			2.5				
3		3			3				
2		3			3				
2.5		3			3				
1		3			3				
1		3.2			3.2				
2.5		4			4				
4.5		4			4				
4		4.5			4.5				
3		6			6				
2.61	mean	2.5	median=(2.5+2.5)/2		1	mode			

Mode

Grouped Data
Measures of Central Tendency for Grouped Data

Ages of randomly selected residents of Tagaytay Highlands.

Age (years)	Frequency f	Midpoint m	$f \cdot m$	Cumulative frequency
0-9	44	4.5		
10-19	42	14.5		
20-29	33	24.5		
30-39	30	34.5		
40-49	27	44.5		
50-59	22	54.5		
60-69	18	64.5		
70-79	9	74.5		
	Total		Total	

Mean ($\bar{x}$) (4pts)
0,1,2,3,4,5,6,7,8,9

Age (years)	f	m	$f \cdot m$	Cumulative frequency
0-9	44	4.5	198	198
Oct-19	42	14.5	609	807
20-29	33	24.5	808.5	1615.5
30-39	30	34.5	1035	2650.5
40-49	27	44.5	1201.5	3852
50-59	22	54.5	1199	5051
60-69	18	64.5	1161	6212
70-79	9	74.5	670.5	6882.5
	225			
	average	30.58889		

Measures of Variability

(Range, Variance, and Standard deviation)

Measures of Variability

Given: The final exam scores of 5 students were 80, 88, 92,90, and 85.

Range

Difference from highest and lowest 92-80 = 12

Sample Variance

Step 1: What is the Sample Mean ($\bar{x}$)

$\bar{x}$= (80+88+92+90+85)/5 = 87

Step 2: Compute for the difference from the mean of each of the given observations. Fill in the column $(x_i - \bar{x})$

Step 3: Fill in the column of square of the difference $(x_i - \bar{x})^2$

Step 4: Sum of the squares $\sum_{i=1}^{n}(x_i - \bar{x})^2$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
80	7	49
88	-1	1
92	5	25
90	-3	9
85	-2	4
$n = 5$		$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 88$

Step 5: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{88}{5-1} = 22$ variance

Sample Standard Deviation

$s = \sqrt{s^2} = \sqrt{22} = 4.69$

Difference of Notations for Population And Sample

	Population	Sample
# of subjects	N	n
Mean	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
Variance	$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$	$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$
Standard deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$	$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$

Symbol	How to read
μ	“Myu” or population mean
σ^2	“sigma squared” or population variance
σ	“sigma” or population standard deviation
$\bar{x}$	“x bar” or sample mean
S^2	“s squared” or sample variance
S	“s” or sample standard deviation

Measures of Variability

(Variance, Standard deviation, and Coefficient of Variation)

The sales department of a popular car company wants to determine the average, variance, standard deviation, and coefficient of variation of the number of cars sold by all of its sales groups in a given month. The following data set is the number of cars sold by all the ten groups of sales department. In this problem the collected data is assumed to be the population.

31	22	18	35	38	30	34	32	20	40

You may use calculator or excel to solve for the following:

Population mean $\mu = \frac{\sum_{i=1}^{N} x_i}{N}$ *(2pts)= 30*

Population Variance $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$. =51. 8

Population Standard Deviation $\sigma = \sqrt{\sigma^2} =$ 7.197

Coefficient of Variation(CV)

$$CV = \frac{\sigma}{\mu} \times 100\%$$

CV= 7.197/30 = 0.2399 x 100% = 23.99

The coefficient of variation is defined only for data sets whose mean is not equal to zero.

Coefficient of Variation

CV is the ratio of standard deviation to the mean, usually expressed in percentage. It is useful when comparing two or more data with different units of measure.

Population

$$CV = \frac{\sigma}{\mu} \times 100\%$$

Sample

$$CV = \frac{s}{\bar{x}} \times 100\%$$

The problem given above solved for the CV of number of cars sold by the ten group sales department. Now, let us look into the revenue or income from the same ten group sales department. Suppose the average revenue from the 10 groups of car dealers in the given problem above is P 210,000 with a standard deviation of P 85,000.

σ =85,000 $\mu = 210{,}000$

What is the Coefficient of Variation of the revenue?

CV = 85,000 / 210,000 = 0.4048 x 100 = 40.48

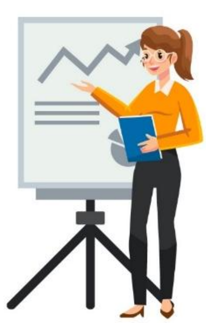
The unit of measure of revenue is Pesos, while the number of cars is just counting. To compare them we need coefficient of variation.

Which has a larger CV, **number of cars sold** or **revenue**?

CV for number of cars is 23.99

CV for revenue is 40.48

Conclusion: There is more variability in the _revenue___than in the__number of cars sold.



The one with larger CV has larger variability.

Question from STEP example for Coefficient of Variation

MODULE 5: EMPLOYMENT **PART C: MAIN JOB IN PAST WEEK**

22	Approximately how many total paid workers are employed in your business/ self-employed enterprise? (DO NOT INCLUDE YOURSELF)	
	1	1
	2-5	2
	6-9	3
	10-15	4
	16-25	5
	26-50	6
	51-200	7
	More than 200	8

26	How much did you personally get from this business in the last work period (excluding any taxes and social security benefits?	Philippine Peso (PHP)

CV number of employees= 0.861273805 CV personal income = 0.734231615

Which has more variability the number of employees or the personal income by the bussiness owner?

The number of employees.