

الكلية : كلية طب العام الفرع : طب الاسرة والمجتمع المرحلة: الثالثة أستاذ المادة : د بديعه ثامر يحيى اسم المادة باللغة العربية : الاحصاء الحياتي اسم المادة باللغة الانكليزيه : Biostatistics اسم المحاضرة الثالثة باللغة العربيه :مقياس التباين

اسم المحاضرة الثالثة باللغة الانكليزية : Measures of variability

## L3

## Measures of variability

# Measures of dispersion

## Measures of variability(measures of dispersion)

- A measure of dispersion conveys information regarding the amount of variability present in a set of data.
- If all values are the same there is no dispersion
- If they are not all the same, dispersion is present in the data.
- The amount of dispersion may be small when the values though different are close together
- Dispersion refers to how variable (spread out)

measure of variability are sometimes called( measure of spread )knowing the dispersion of data can be as important as knowing its central tendency



• Figure 2.5.1 shows the frequency polygons for two populations that have equal means but different amounts of variability. Population B, which is more variable than population A, is more spread out. If the values are widely scattered, the dispersion is greater. Other terms used synonymously with dispersion include variation, approach and conttor.

- These two measures together help us sum up a distribution.
- •Measures of central tendency tell us about typical or central.
- •Measures of variability reveal how far from the typical or central that the distribution tends to vary

### There are five measure of measure of variability :

**1.Range:** One way to measure the variation in a set of values is to compute the range. The range is the difference between the largest and smallest values in a set of observations. If we denote the range by R, the largest value by xL, and the smallest value by xS. Range = Lx - Sx.

2- Variance is the "sum of the squared deviation of the values from the mean

- divided by sample size minus one". ( $S^2$ )
- The Variance When the values of a set of observations lie close to their mean, the dispersion is less than when they are scattered over a wide range. Since this is true, it would be intuitively appealing if we could measure dispersion relative to the scatter of the values about their mean. Such a measure is realized in what is known as the variance. In computing the variance of a sample of values.

- Variance is calculated by the following equations:
- (Ungrouped data , raw data )



• X = sample score  $X^- =$  mean of sample n = number of sample size

#### **3- The standard deviation :**

- The variance represents squared. To obtain a measure of dispersion in original units, we merely take the square root of the variance. The result is called the standard deviation. In general, the standard deviation of a sample is given by
- SD =  $\pm \sqrt{S^2}$



- Importance of standard deviation: The most common measures of dispersion for continuous data are the variance and standard deviation, both describe how much the individual values in a data set vary from the mean or average value.
- \* The variance and standard deviation, usually accompanied by the mean, help to know how a set of data values distributes around its mean.
- \* A small standard deviation means that the values in the data set are close to the middle of the data set, on average while a large standard deviation means that the values in the data set are farther away from the middle, on average.

The coefficient of variation (CV) allows us to compare the variation of wo same variables (or more) different variables

## **Advantage of Coefficient of Variation:**

When two data set distributions have means of different magnitude, a omparison of the C.V. is therefore much more meaningful than a omparison of their respective standard deviation.

#### SD

CV % = ----- X 100

5- Standard Error of the Sample Mean (SE):
\* The sample mean is unlikely to be exactly equal to the populati mean

\* The standard error measures the variability of the mean of th sample

as an estimate of the true value of the mean for the populat SD

SE = \_\_\_\_\_\_\_n

#### Example :

- The length of stay in two different medical wards in two different district hospitals are given below:
- Ward A: 4 5 7 3 5 8 10 2 5 8 5 49
- Ward B: 2 8 5 6 3 4 3 7 3 2 2 5 4 4 3 5 4 1 3 2 2 9 3.
- Using appropriate statistical method(s), compare the duration of stay in the two hospitals. (By measuring mean, median & mode.
- Range , Variance ,Standard deviation , Coefficient variation and standard error

- 1. Range
  - Ward A :
  - Range = Lx Sx = 49 2 = 47
- 2-Variance :
- Ward A

 $\Sigma (X^{-} - X)^{2} = 1780.2$   $S^{2} = ----- = ----- = 161.8$  n-1 = 12-1

(X<sup>-</sup> -X )<sup>2</sup>

(X<sup>-</sup> - X )

12

=9.25

ΣΧ

X<sup>-</sup> = ------



X	X <sup>2</sup>
4	16
5	25
7	49
3	9
5	25
8	64
10	100
2	4
5	25
8	64
5	25
49	2401
Total =111	Total= 2807

 $3-SD = \pm \sqrt{S^2} = \pm \sqrt{161} = \pm 12.7$ 

SD 12.7 4- CV = ----- X100 = ----- x 100 = 137.3%X 9.25 SD 12.7 **5- SE** = ----- = **3.7**  $\sqrt{\mathbf{n}}$   $\sqrt{12}$ 

واجب Ward B واجب

#### • Measure of variability in Grouped data :

Age(years)	Number of death due to accidents	Number of death due to cancer
25 - 34	30	15
35 – 44	35	10
45 –54	40	25
55 – 64	41	30
65 – 74	20	35
75 – 84	16	15
85 – 94	18	10
Total	200	140

• Use these data to compute the measures of central tendency and measures of variability

- Death due to accident
- 1. Range = Lx Sx = 94 25 = 69



#### Death due to accident

Age	f	m	fm	m²	fm²
25-34	30	25+ 34\2= 29.5	885	870.25	26107.5
35 - 44	35	39.5	1382.5	1560.25	54608.7
45 - 45	40	49.5	1980	2450.25	98010
55 - 64	41	59.5	2439.5	3540.25	145150. 5
65 – 74	20	69.5	1390	4830.25	96605
75 - 84	16	79.5	1272	6320.25	101124
85 – 94	18	89.5	1611	8010.25	144184.
Total	200		9960		665790

 $SD = \pm \sqrt{S^2} = \pm \sqrt{853.2} = \pm 24.2$ 



واجب بيتي Death due to cancer

# • Q1\ The Table below pertains to 66 birth weights born to villages in a rural district:

Birth weight (kg)	No. of births
2.0-	8
2.3-	14
2.6-	16
2.9-	22
3.2-	5
3.5-	0
3.8-4.1	1
Total	66

• Choose the most suitable statistical measure(s) to describe the variable in hand. Do present the distribution graphically.

#### Q2\A study about gout & serum calcium of two male samples, found the

following results: (as shown in the Table) :

class interval	Frequency of	Frequency of Serum	
	Uric acid	Calcium	
4-5	5	6	
6-7	10	5	
8-9	3	10	
10-11	2	2	
12-13	14	4	
14-15	8	8	
16-17	11	3	
18-19	7	7	
Total	60	45	

#### **Construct the following:**

- **1** Cumulative frequency distribution and cumulative relative frequency distribution.
- 2- Choose the most appropriate measure to compare between these distributions.
- **3 Compare these two frequency distributions graphically**