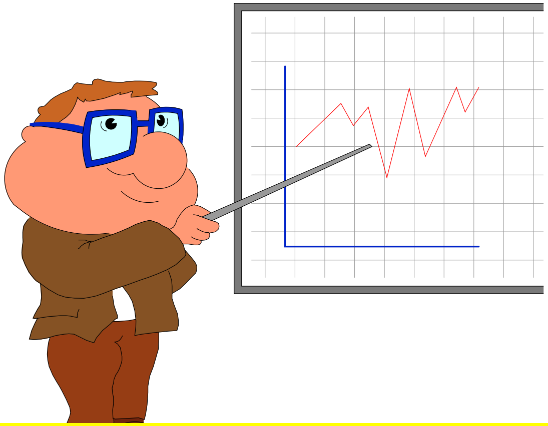
TAIBAH UNIVERSITY
Faculty of Science
Department of Math.

جامعة طيبة
كلية العلوم
قسم الرياضيات

STAT 301



Teacher :

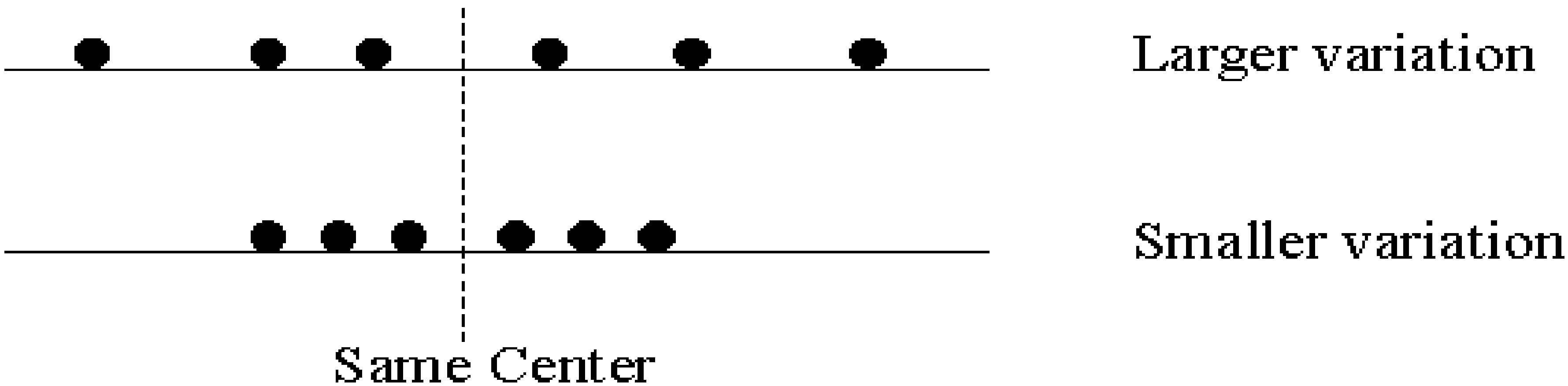
Lesson

6

Measures of Variability (Dispersion or Variation)

Measures of Dispersion

The variation or dispersion in a set of data refers to how spread out the observations are from each other.



The variation is small when the observations are close together. There is no variation if the observations are the same.

Measures of Dispersion

- **Measures of dispersion are important for describing the spread of the data, or its variation around a central value . or express quantitatively the degree of variation or dispersion of values.**

- **There are various methods that can be used to measure the dispersion of a dataset, each with its own set of advantages and disadvantages.**

The Range:

- **The difference between the largest and smallest sample values**
- **If $X_1, X_2, \ldots\ldots\ldots, X_n$ are the values of observations in a sample then range is given by:**

$$Range(X_1, X_2, \ldots\ldots, X_n) =$$
$$\max(X_1, X_2, \ldots\ldots, X_n) - \min(X_1, X_2, \ldots\ldots, X_n)$$

The Range (Example):

find The range of (12, 24, 19, 20, 7) .

Solution:

$$Range = 24 - 7 = 17$$

- **One of the simplest measures of variability to calculate.**
- **Depends only on extreme values and provides no information about how the remaining data is distributed.**

The Population Variance:

If $X_1, X_2, \ldots\ldots\ldots\ldots, X_N$ are the population values, then the population variance is:

$$\sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \ldots\ldots\ldots + (X_N - \mu)^2}{N}$$

Using summation form:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2$$

Where μ is population mean

The Sample Variance:

If $X_1, X_2, \ldots\ldots\ldots\ldots, X_n$ are the sample values, then the sample variance is:

$$S^2 = \frac{\left(X_1 - \overline{X}\right)^2 + \left(X_2 - \overline{X}\right)^2 + \ldots\ldots\ldots + \left(X_n - \overline{X}\right)^2}{n-1}$$

Using summation form:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

The Sample Variance:

Where:

$$\overline{X} = \sum_{i=1}^{n} X_i / n$$ **is the sample mean.**

Note:

(n –1) : is called the degrees of freedom (df) associated with the sample variance S^2.

The Sample Standard Deviation :

The standard deviation is another measure of variation. It is the square root of the variance, i.e., it is:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

Example 1 :

Compute the sample variance and standard deviation of the following observations (ages in year): 10, 21, 33, 53, 54.

Solution

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\sum_{i=1}^{5} X_i}{5} = \frac{10+21+33+53+54}{5}$$

$$= \frac{171}{5} = 34.2 \qquad \text{(year)}$$

Example 1 :

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^{5}(X_i - 34.2)^2}{5-1}$$

$$= \frac{(10-34.2)^2 + (21-34.2)^2 + (33-34.2)^2 + (53-34.2)^2 + (54-34.2)^2}{4}$$

$$= \frac{1506.8}{4} = 376.7$$

The sample standard deviation is:

$$S = \sqrt{S^2} = \sqrt{376.7} = 19.41$$

The Sample Variance(another formula):

Another Formula for Calculating S^2:

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n-1}$$

(It is simple and more accurate)

The Sample Variance(another formula):

For the previous Example,

X_i	**10**	**21**	**33**	**53**	**54**	$\sum X_i = 171$
X_i^2	**100**	**441**	**1089**	**2809**	**2916**	$\sum X_i^2 = 7355$

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n-1}$$

$$= \frac{7355 - (5)(34.2)^2}{5-1} = \frac{1506.8}{4} = 376.7$$

4. Coefficient of Variation

A coefficient of variation is computed as a ratio of the standard deviation of the distribution to the mean of the same distribution.

$$CV = \frac{SD}{\bar{x}}$$

Example. Comments on Children in a community

Measure	Height	weight
Mean	*40 inch*	*10 kg*
SD	*5 inch*	*2 kg*
CV	***0.125***	***0.20***

Comment: Since the coefficient of variation for weight is greater than that of height, we would tend to conclude that weight has more variability than height in the population.

5. Outliers

Interquartile range rule=IQR

Outliers are decided using *1.5_IQR* rule:

Steps

- ☐ Calculate Q_1 and Q_3
- ☐ $IQR = Q_3 - Q_1$
- ☐ Lower and upper cut points:
  - $L = Q_1 - 1.5 * IQR$
  - $U = Q_3 + 1.5 * IQR$
- ☐ Any number less than L or larger than U are called outliers

Example. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Solution:

$$IQR = 8 - 3 = 5$$

$$L = 3 - 1.5 * 5 = -4.5$$

$$U = 8 + 1.5 * 5 = 15.5$$

No outliers in this data set

Example. 32, 17, 56, 42, 88, 103, 25, 33, 55, 19

Solution:

$$IQR = 56 - 25 = 31$$

$$L = 25 - 1.5 * 31 = -21.5$$

$$U = 56 + 1.5 * 31 = 102.5$$

One outlier: 103