جامعة المستقبل

الاحصاء الحيوي /المحاضرة الثالثة و الرابعة

الفيزياء الطبية /المرحلة الثانية

MEASURES OF DISPERSION AND POSITION

دلائل سعد عبد الزهرة –صفا محمد هادي

MEASURES OF DISPERSION AND POSITION

In statistics, to describe the data set accurately, statisticians must know more than the measures of central tendency. Two data sets with the same mean may have completely different variation or dispersion, so the measures that help us know about the spread data set are called **<u>the measures of</u> <u>dispersion</u>** such as:

1- Range.
2- Variance.
3- Standard deviation.

1- Range:_ The range is the simplest of the three measures and is defined now. The range is the highest value minus the lowest value. The symbol R is used for the range.

R = highest value - lowest value

<u>Disadvantage of range</u>:

a- Based on two values only, largest and smallest.

b- Extremely large or extremely small data can significantly affect the range.

Ex. (1): Calculate the range for the following data set: 5 -7 2 0 -9 16 10 7

Sol: -9 -7 0 2 5 7 10 16

R = highest value - lowest value

R= 16 – (-9) = 25

2- Variance and Standard Deviation.

a- Ungrouped data

$$Population\ variance: \quad \sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

$$Sample\ variance: \quad s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

$$Population\ standard\ deviation: \quad \sigma = \sqrt{\sigma^2}$$

$$Sample\ standard\ deviation: \quad s = \sqrt{s^2}$$

Exp(2): Find the sample variance , standard deviation and the range ,for the amount of European auto sales for a sample of 6 years shown. The data are in millions of dollars.
11.2, 11.9, 12.0, 12.8, 13.4, 14.3

Sol:

$1 - Variance: \qquad s^2 = \frac{\sum(x-x^-)^2}{n-1}$

$x^- = \sum x/n$ = 11.2+ 11.9+ 12.0+ 12.8+ 13.4+ 14.3/6= 75.6/6= 12.6

s^2= $(11.2-12.6)^2 + (11.9-12.6)^2 + (12.0-12.6)^2 + (12.8-12.6)^2 + (13.4-12.6)^2 + (14.3-12.6)^2$/ 5= 1.278

2- Standard deviation:

$$s = \sqrt{1.278} = 1.13$$

3- The range (R) = 14.3 – 11.2 = 3.1

b- Grouped data

$Population\ variance: \qquad \sigma^2 = \frac{\sum f(x_m-\mu)^2}{N}$

$Sample\ variance: \qquad s^2 = \frac{\sum f(x_m-x^-)^2}{n-1}$

$Population\ standard\ deviation: \qquad \sigma = \sqrt{\sigma^2}$

$Sample\ standard\ deviation: \qquad s = \sqrt{s^2}$

Ex(3):
Find the variance and the standard deviation for the data in this frequency distribution table. The data represent the number of miles that 20 runners ran during one week.

Class boundaries	5.5–10.5	10.5–15.5	15.5-20.5	20.5–25.5	25.5–30.5	30.5–35.5	35.5-40.5
Freq. (f_i)	1	2	3	5	4	3	2

Sol:

i	Class boundaries	Freq. (f_i)	xm	$xm.fi$	$(xm - x-)2$	$fi(xm - x-)2$
1	5.5–10.5	1	8	8	272.25	272.25
2	10.5–15.5	2	13	26	132.25	264.5
3	15.5–20.5	3	18	54	42.25	126.75
4	20.5–25.5	5	23	115	2.25	11.25
5	25.5–30.5	4	28	112	12.25	49.00
6	30.5–35.5	3	33	99	72.25	216.75
7	35.5–40.5	2	38	76	182.25	364.5
				∑490		∑1305

$x^{-} = \frac{\sum f.x_m}{n}$, $x^{-} = \frac{490}{20} = 24.5$

$s^2 = \frac{\sum f(x_m - x^{-})^2}{n-1} = \frac{1305}{19} = 68.68$

s= 8.28

<u>Coefficient of Variation</u>

A statistic that allows you to compare standard deviations when the units are different, it denoted by C_{Var}, is the standard deviation divided by the mean. The result is expressed as a percentage.

For samples: $C_{Var} = (s/x^{-}) * 100$

For populations: $C_{Var} = (\sigma/\mu) * 100$

Ex: the mean of the number of sales of cars over a 3-month period is 87, and the standard deviation is 5. The mean of the commissions is 5225 $, and the standard deviation is 773 $. Compare the variations of the two.

Sol:

The coefficients of variation are:

$C_{Var} = (s/x^{-}) * 100 = (5/87) * 100 = 5.75\%$ sales

$C_{Var} = (s/x^{-}) * 100 = (773/5225) * 100 = 14.8\ \%$ commissions

Since the coefficient of variation is larger for commissions, the commissions are more than the sales.

Measures of Position

In addition to measures of central tendency and measures of variation, there are measures of position or location. These measures include:

1-standard scores.

2- percentiles.

3- deciles, and quartiles.

z score or **standard score:** it represents the number of standard deviations that a data value falls above or below the mean.

For samples, the formula is: $z = (x - x-)/s$

For populations, the formula is: $z = (x - \mu)/\sigma$

Ex: A student scored 65 on a calculus test that had a mean of 50 and a standard deviation of 10; she scored 30 on a history test with a mean of 25 and a standard deviation of 5. Compare her relative positions on the two tests.

Sol:

$z_1 = (x-x^{-})/s = 65-50/10 = 1.5$

$z_2 = (x-x^{-})/s = 30-25/5 = 1$

Since the z score for calculus is larger, her relative position in the calculus class is higher than her relative position in the history class.

*** Note that if the z score is positive, the score is above the mean. If the z score is 0, the score is the same as the mean. And if the z score is negative, the score is below the mean.

As the name implies, quartiles divide the data set into four equal parts. Therefore, the first quartile, Q_1, is the 25th percentile, the second quartile, Q_2 is the 50th percentile (or the median), and the third quartile, Q_3, is the 75th percentile. The difference between the third and first quartiles is interred quartile range (IQR).

$$IQR = Q3 - Q1$$

25% of data	25% of data	25% of data	25% of data

Min. value Q1 Q2 Q3 Max. value median

Ex: The following are the ages of nine employees of an insurance company

$$47\ 28\ 39\ 51\ 33\ 37\ 59\ 24\ 33$$

a- Find the values of three quartiles

b- When does the age 28 fall in relation to the ages of these employees.

c- Find the inter quartile range (IQR).

Sol:

Arrange the data in increasing order:

$$24, 28, 33, 33, 37, 39, 47, 51, 59$$

Q_2 is the median of all values $\rightarrow$ $Q_2 = 37$

Q_1 is the median of values $(24, 28, 33, 33) \rightarrow Q1 = (28 + 33)/2 = 30.5.$

Q_3 is the median of values $(39, 47, 51, 59) \rightarrow Q3 = (47 + 51)/2 = 49.$

d- The age 28 fall in the first 25% of the ages.

e- The inter quartile range $(IQR) = Q3 - Q1 = 49 - 30.5 = 18.5$ years.

Percentiles:

divide the data set into 100 equal groups. Each data set has 99 percentiles, data must be ranked in increasing order to compute percentiles. The k^{th} percentile is denoted by P_k, where k is an integer range from (1 –99). For example, the 25^{th} percentile which is denoted by P_{25}, is defined to be that numerical value such that at most 25% of the values are smaller than it and at most 75% are larger than it in an ordered data set.

For ungrouped data,

The percentile corresponding to a given value (x) is computed by using the formula:

$$\text{Percentile} = \frac{No.of\ values\ below\ x + 0.5}{Total\ No.of\ values} * 100$$

Ex. A teacher gives a 20-point test to 10 students. Find the percentile rank of a score of 12. Scores: $18, 15, 12, 6, 8, 2, 3, 5, 20, 10$**.**

Sol:

Ordered set: $2, 3, 5, 6, 8, 10, 12, 15, 18, 20.$

$$\text{Percentile} = \frac{No.of\ values\ below\ x + 0.5}{Total\ No.of\ values} * 100$$

……… (H.W)

To Finding the value Corresponding to a Given Percentile:

- Let p be the percentile and n the sample size.
- Arrange the data in order.
- Compute $c = (n \times p)/100$.
- If c is not a whole number, round up to the next whole number. If c is a whole number, use the value halfway between c and $c + 1$.
- The value of c is the position value of the required percentile.

***Ex*:** For the following data set: $2, 3, 5, 6, 8, 10, 12, 15, 18, 20$. Find the values of the 25th and 80th percentile.

Sol:

a. $n = 10, \quad p = 25 \ c = (10 \times 25)/100 = 2.5.$

Hence round up to c = 3. Thus, the value of the 25th percentile is the value x = 5.

b. $n = 10, \quad p = 80 \ c = (10 \times 80)/100 = 8.$

Thus the value of the 80th percentile is the average of the 8th and 9th values.

$$x = (15 + 18)/2 = 16.5.$$

QUESTIONS:

1- Calculate the range for the following data set: 12,5,11,9,7,4,8

2- Mathematical law of sample variance is …………………

3- The standard deviation divided by the mean is called

a- Range *b- standard score* *c*- Coefficient of Variation

4- Mathematical law of Standard deviation is …………………