جامعة الانبار

1987 1408

UNIVERSITY OF ANBAR

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 الثانيه باللغة العربيه : مقياس النزعة المركزيه

اسم المحاضرة الثانيه باللغة الانكليزية : Measures of central tendency

L2

The measure of Central Tendency

Or

The measure of central location

A measure of **central tendency**

is a single value that attempts to describe a set of data by identifying the **central** position within that set of data

In general central tendency is a statistical measure that determines a single value that accurately describes the center of the distribution that represents the entire distribution of scores.

Three measures of central tendency

Mean	• The mathematical average of a distribution
Median	• The value that has the same number of values higher and lower than itself
Mode	• The most common value

Arithmetic Mean: The most familiar measure of central tendency is the arithmetic mean. It is the descriptive measure most people have in mind when they speak of the “average.” we shall refer to the arithmetic mean simply as the mean. The mean is obtained by adding all the values in a population or sample and dividing by the number of values that are added

Properties of the Mean :

1. Uniqueness. For a given data set, there is only one arithmetic mean.

2. Simplicity. The arithmetic mean is easily understood and easy to compute.

3. Since every value in a set of data enters into the computation of the mean, it is affected by each value. Extreme values, therefore, have an influence on the mean and, in some cases, can so distort it that it becomes undesirable as a measure of central tendency. ←Disadvantage

Mean (average)

The **population mean** is donated by µ (pronounced mue)

The sample mean is donated by X¯ (pronounced X – bar)

The Sample Mean When we compute the mean for a sample of values, the procedure just outlined is followed with some modifications in notation.

We use $\bar{x}$ to designate the sample mean and n to indicate the number of values in the sample. The sample mean then is expressed as

The sample mean then is expressed as

1- In raw data

$$\bar{x} = \frac{\sum X}{n}$$

$\sum$ = summation sign (adding all values)

X = no. of observation

n= sample

2- In grouped data

fm = frequency multiplied by mid point

$$\bar{x} = \frac{\sum Fm}{\sum f}$$

$\sum f$ (n) = sample

Example: **The length of stay in two different medical wards in two different district hospitals is 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)?**

Ward A

$$\bar{x} = \frac{\Sigma x}{n} = \frac{4 + 5 + 7 + 3 + 5 + 8 + 10 + 2 + 5 + 8 + 5 + 49}{12} = 9.45$$

d B

2 + 8 + 5 +6 +3 + 4 + 3+ 7 + 3 + 2 +2 + ++++++++++ 9+ 3

=--

23

3.9

$$\text{Grouped data : } \bar{x} = \frac{\Sigma fm}{\sum f} = \frac{1080}{24} = 45$$

Example : **The following data are measurements from 24 subjects of glomerular filtration rate (GFR) measured with renal failure. Use these data to compute the measures of central tendency.**

GFR	f	M or x	Σfx Or Σfm
18 – 29	6	18+29\2 = 23.5	23.5 X 6 = 141
30 - 41	5	30+41\2 = 35.5	177.5
42 – 53	5	47.5	237.5
54 – 65	5	59.5	297.5
66- 77	2	71.5	143
78 – 89	1	83.5	83.5
Total (Σn)(Σf)	24		1080

Median :(50th percentile)

The median of a finite set of values is that value which divides the set into two equal parts such that the number of values equal to or greater than the median is equal to the number of values equal to or less than the median.

If the number of values is odd, the median will be the middle value when all values have been arranged in order of magnitude. When the number of values is even, there is no single middle value. Instead there are two middle values.

Properties of the Median :

1. Uniqueness. there is only one median for a given set of data.
2. Simplicity.
3. It is not as drastically affected by extreme values as is the mean.

Disadvantages of median:

* It provides no information about all values (observations).
* It is less amenable than the mean to tests of statistical significance

In Raw data :

1- If the number of values is odd :

position of median = n+1/2

2- When the number of values is even :

Position of Median = $n/2$

Sol: steps

1st apply formula

2nd ordered array

3rd count (ascending and descending)

Ward A

$$\text{Position of median} = \frac{n}{2} = \frac{12}{2} = 6$$

4 5 7 3 5 8 10 2 5 8 5 49

Ordered array : 2, 3 , 4 , 5, 5, 5, 5, 7, 8, 8, 10, 49

$$= \frac{5+5}{2} = 5$$

Median =5

Ward B

$$\text{Position of median} = \frac{n+1}{2} = \frac{23+1}{2} = 12$$

: 2 8 5 6 3 4 3 7 3 2 2 5 4 4 3 5 4 1 3 2 2 9 3.

Ordered array

2 , 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5,6, 7, 8, 9.

Median = 3

Median in grouped data :

$$\text{Median} = L + \frac{W}{f\ \text{med}} \left(0.5\, n - \sum f\, b \right)$$

Median = L+ R/F X W

Where L: lower limit of class interval containing median.

R= (N/2)-previous cumulative frequency

W= width of C.I

Sol:

Step 1: Construct the cumulative frequency distribution.

Step 2: Decide the class that contain the median.

That the first class with the value of cumulative frequency equal at least N/2

(N = the total frequency/2)

fb = (the frequency before class median)

L = (the lower boundary of the class median)

GFR	f	M or x	Σfx Or Σfm	cf
18 – 29	6 fb	18+29\2 = 23.5	23.5 X 6 = 141	6
30 - 41	5 fb	30+41\2 = 35.5	177.5	11
42 – 53	5 fm	47.5	237.5	16
54 – 65	5	59.5	297.5	21
66- 77	2	71.5	143	23
78 – 89	1	83.5	83.5	24
Total (Σn)(Σf)	24		1080	

n\2 = 24\2 =12

$$\text{Median} = L + \frac{W}{f\ med}\,(0.5\,n - \sum f\,b) = 42 + \frac{12}{5}\,(12 - 11) = 44.4$$

Mode:

The mode of a set of values is that value which occurs most frequently. If all the values are different there is no mode; on the other hand, a set of values may have more than one mode

Advantage of Mode:

Sometimes gives a clue about the etiology of the disease.

Disadvantages of Mode:

- With a
- small number of observations, there may be no mode.
- It is less amenable to tests of statistical significance

Sol:

Mode = الرقم صاحي اكثر تكرار

Ward A:

4 5 7 3 5 8 10 2 5 8 5 49

Mode = 5

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.

Mode = 3

Ward A = $\bar{X}$= 9.25, median = 5 , Mode 5→ WHY?????

Ward B = $\bar{X}$= 3.9 . median = 3 , Mode 3 → WHY?????

Mode in grouped data

Mode :الرقم صاحب اكثر تكرار

- Mode = 6 (model class 18 -29)

Skewness :

Data distributions may be classified on the basis of whether they are symmetric or asymmetric. If a distribution is symmetric, the left half of its graph (histogram or frequency polygon) will be a mirror image of its right half. When the left half and right half of the graph of a distribution are not mirror images of each other, the distribution is asymmetric.

If the graph (histogram or frequency polygon) of a distribution is asymmetric, the distribution is said to be skewed. If a distribution is not symmetric because its graph extends further to the right than to the left, that is, if it has a long tail to the right, we say that the distribution is skewed to the right or is positively skewed. If a distribution is not symmetric because its graph extends further to the left than to the right, that is, if it has a long tail to the left, we say that the distribution is skewed to the left or is negatively skewed

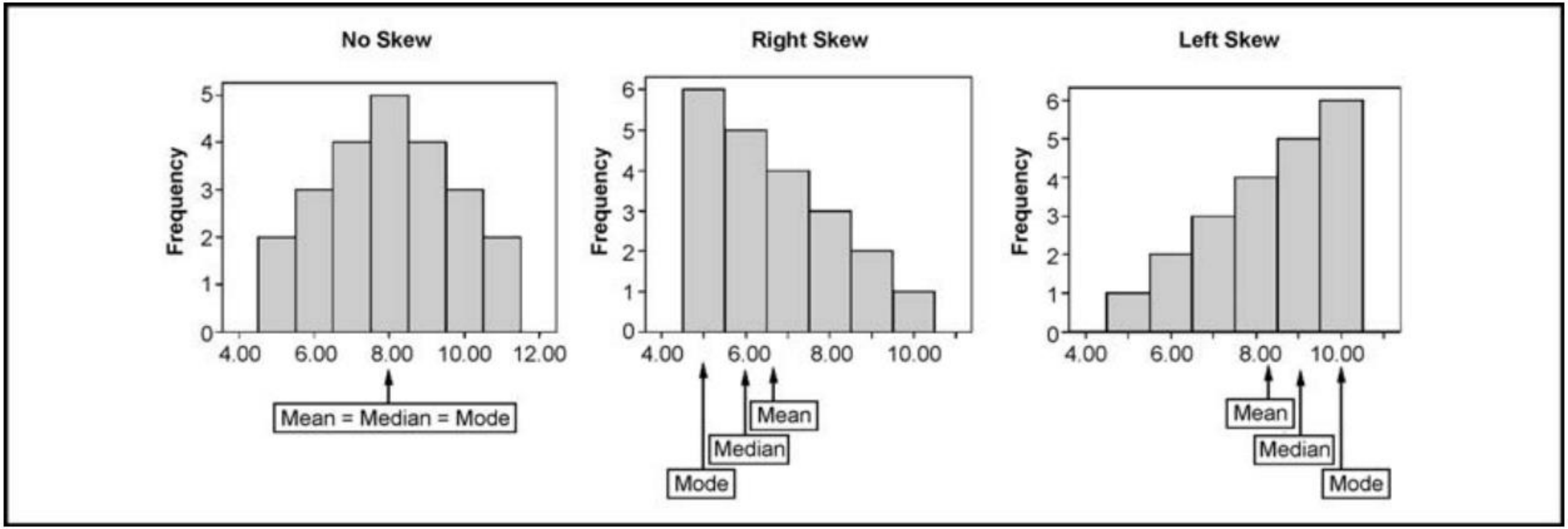

FIGURE 2.4.1 Three histograms illustrating skewness.