Pharmaceutical statistics

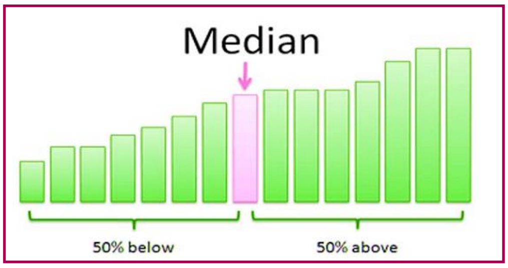
2025-2024

Dr. Rola Aqeel

ARKAN ACADEMY

Median

- The point that divides the distribution into ***two equal parts***, or the point between the upper and lower halves of the distribution.
  - Accordingly, if we have a finite number of values, then the median is the value that divides those values into two 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.



  - It is called the 50^{th} percentile
  - ***How to determine the Median in ordered array?***
    - ✓ When ***n*** is an *odd number*, the median is the value number ***(n+1)/2*** in an ordered array.

★ Example:

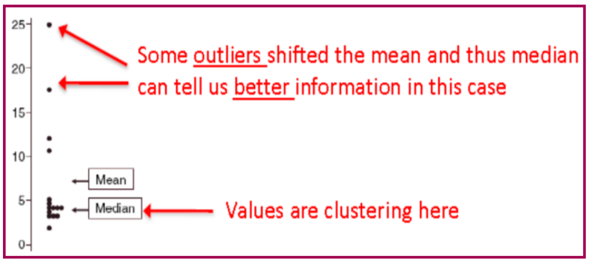
- ✓ What is the median for the following data set: 10, 15, 12, 25, 20.
  - The median is the *value number of* ***(n+1)/2***
    1. Rank of the data: 10, 12, 15, 20, 25.
    2. (5+1)/2=3^{rd} (The third value in the ordered array) is the median so it is ***15***.
- ✓ *When* ***n*** *is an even number*, ***the median is:***
  - The mean of the two middle values ***(n/2)***th and ***((n/2) + 1)***th in an ordered array.

★ Example:

- ✓ What is the median for the following data set: 10, 15, 20, 25, 30, 5
  1. Rank of the data: 5, 10, 15, 20, 25, 30
  2. (6/2) =3^{rd} value which is 15 and ((6/2) + 1) =4^{th} which is 20
  3. Calculate the mean of the 3^{rd} and 4^{th} values (15+20)/2= ***17.5***

★ Example 1:

- ✓ Fifteen patients were provided with their drugs in a child-proof container of a design that they had not previously experienced and the time it took each patient to open the container was measured.
- ✓ The mean = 7.09 seconds
- ✓ Is this the most representative descriptive figure? No
- ✓ Most patients have got the idea straight away and have taken only 2-5 s to open the container.
- ✓ However, four seem to have got the wrong end of the stick and have ended up taking anything up to 25 seconds.
- ✓ These four have contributed a disproportionate amount of time (65.6 s) to the overall total.
- ✓ This then increased the mean to 7.09 s
- ✓ We would not consider a patient who took 7.09 s to be remotely typical. In fact, they would be distinctly slow.



- This problem of ***mean*** values being disproportionately *affected by a minority of outliers* arises quite frequently in biological and medical research.
- *A useful approach in such a case is to use the* ***median***.

★ Example 2:

- ✓ Blood Glucose Level (mg/dl): 80, 81, 82, 83, 84, 84, 86, 86, **180**
- ✓ Mean: 93
- ✓ Median: 84
- ✓ The outlier 180 shifted the mean to higher value, which is not descriptive for the data set in this case.

➢ *Median for Intervals (50th percentile, Q2):*

- ✓ The median is the value that divides the dataset into two equal parts, with half of the observations below it and half above it.
- ✓ This means that the median is the value of the *84.5th observation.*
- ✓ The median falls within a specific *class interval.*
- ✓ To find this interval, you look for the cumulative frequency where the 84.5th observation occurs.

Intervals	True limits	frequency	Cumulative frequency
10--19	9.5-19.5	4	4
20-29	19.5-29.5	66	70
30-39	29.5-39.5	47	117
40-49	39.5-49.5	36	153
50-59	49.5-59.5	12	165
60-69	59.5-69.5	4	169
Total		169	

- ✓ The 84.5th observation lies <u>between 29.5 and 39.5</u>
- ✓ The formula to calculate the median for grouped data is:

$$\textit{Median} = L + \left(\frac{\frac{n}{2} - F}{f}\right) * h$$

- ✓ ***L*** is the lower boundary of the median class (in this case, 29.5).
- ✓ ***n*** is the total number of observations (n/2 is the rank).
- ✓ ***F*** is the cumulative frequency of the class before the median class.
- ✓ ***f*** is the frequency of the median class.
- ✓ ***h*** is the class width (interval width, or 10 in this case).

$$\textit{Median} = 29.5 + \left(\frac{\frac{169}{2} - 70}{47}\right) * 10 = 29.5 + \left(\frac{14.5}{47}\right) * 10 = 32.6$$

- ▪ To calculate the *arithmetic mean* (average) for grouped data (like in a frequency table), you need the midpoint of each class interval and the corresponding frequencies.

- ✓ Calculate midpoints for each class.
- ✓ Multiply each midpoint by its corresponding frequency.
- ✓ Sum the products.
- ✓ Divide the sum by the total number of observations to get the arithmetic mean.

Mean vs. Median

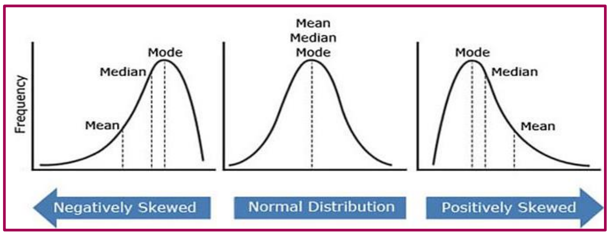
➢ *Uniqueness*

- ➢ Can we have *more than one value* for the indicator? (if we can calculate 1 value)
- ➢ If ***yes***, the indicator is ***not unique.***
- ➢ If ***no***, the indicator is ***unique.***
- ➢ Both *mean* and *median* are ***unique***.

Properties	Mean	Median
Uniqueness	Yes	Yes
Simplicity	Yes	Yes
Robustness to extreme values	**No**	**Yes**

- *Simplicity*
  - ✓ Are these indicators considered to be simple?
  - ✓ Both *mean* and *median* are ***simple***.

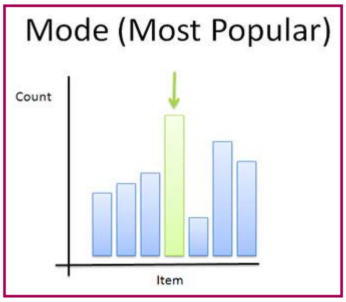
- *Robustness to extreme values*
  - ✓ Is the indicator *highly affected* by extremely high or low values?
  - ✓ If ***yes***, the indicator is ***not robust.***
  - ✓ If ***not***, the indicator is ***robust.***
  - ✓ The *median* is ***robust*** to extreme outliers.
  - ✓ The term 'robust' is used to indicate that a statistic or a procedure will continue to give a reasonable outcome even if some of the data is aberrant.



★ Example:
- For the values: 2, 4, 6, 8, 10 (*median*=6) and for the values: 2, 4, 6, 8, 1000 (*median*=6)
- If the last variable increased to 1000 instead of 10, the *median* would stay the same (6).
- While the *mean* would be hugely inflated!!!!

Mode

- It is the value which occurs *most frequently*
  - ✓ If all values are different there is ***no mode***.
  - ✓ And a set of values may have ***more than one mode***.
- Used for quick estimation and for identifying the <u>most common </u>observation.
- Unlike mean and median, the concept of mode also makes sense for nominal data



★ Example:
- The weights of patients were: 75, 68,74, 68, 68, 75, 84
- The mode is 68

- ***Properties:***
  - ➢ *Not unique* (we can have more than one mode)
  - ➢ *Simple*
  - ➢ *Not Robust,* Less Stable than the Median and Mean.

★ Example: if a dataset {1, 2, 2, 3, 3, 4, 5} has mode 2 and 3, adding a new value of 4 will result in the mode being 4, showing how sensitive the mode is to changes.

- Conversely, the mean and median will not be as dramatically affected by the addition of a single value.

★ Example:

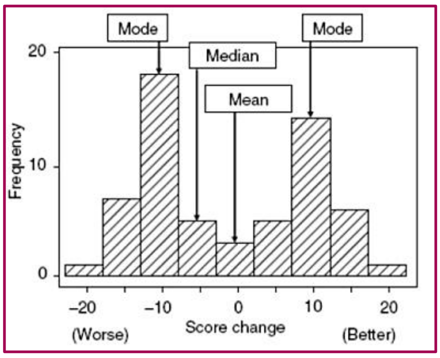
- ✓ The condition of sixty patients with arthritis is recorded using a global assessment variable.
- ✓ A positive score indicates an improvement and a *negative* one a deterioration in the patient's condition after treatment.
- ✓ The mean is (0.77).
- ✓ Do you think the mean is the best descriptive parameter for these data? ***No***
- ✓ A histogram of the data shows that there *are two distinct sub-populations*.

Table 2.3 Individual changes in a global assessment variable following treatment with an anti-inflammatory (61 patients)

Score changes		
11	−9	−8
0	−9	2
−5	−15	−11
11	−13	−12
−13	−13	10
7	−18	−11
7	−13	9
−12	9	14
10	14	−9
−12	10	17
−10	−9	−14
6	11	−6
13	−11	13
−11	14	12
10	10	−6
−9	21	−9
9	6	2
8	−13	5
−12	−6	−7
10	−9	−12
1		

- ➢ Slightly under half the patients have improved quality of life but for the remainder, their lives are made worse.

- Neither the ***mean*** nor the ***median*** indicator remotely describes the situation (Bimodal data). ***only modes achieve this.***
- The mean is particularly *unhelpful* as it indicates a value that is very untypical very few patients show changes close to **zero.**
- We need to describe the fact that in this case, *there are two distinct groups.*
- The data consisted of values clustered around some central points.



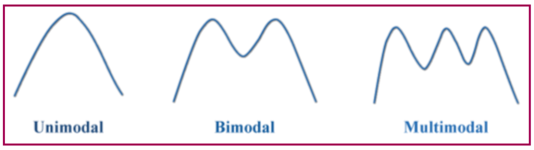
- ***Mode (Most Frequent Observation):***
  - ➢ The mode is the value that occurs most frequently in the dataset.
  - ➢ For grouped data, the mode can be approximated using the class with the highest frequency (*the modal class*).
  - ➢ The modal class is the class interval with the highest frequency.

Intervals	True limits	frequency	Cumulative frequency
10--19	9.5-19.5	4	**4**
20-29	19.5-29.5	66	**70**
30-39	29.5-39.5	47	**117**
40-49	39.5-49.5	36	**153**
50-59	49.5-59.5	12	**165**
60-69	59.5-69.5	4	**169**
Total		169	

- ➢ From the previous calculations, this interval is assumed to be between ***19.5 and 29.5***.
- ➢ *Find the Mode*: For grouped data, a simple way to approximate the mode is to take the midpoint of the modal class.
- ➢ The midpoint is the average of the lower and upper boundaries of the modal class:

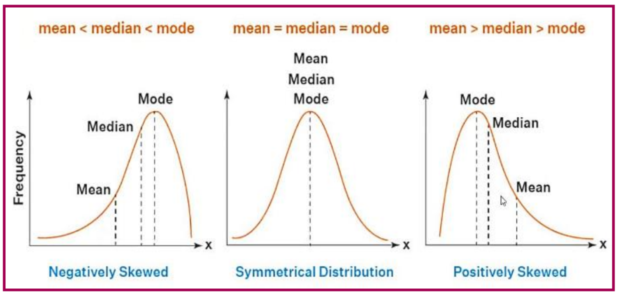
$$Mode = \frac{19.5 + 29.5}{2} = 24.5$$

- ***Data distribution can be:***
  - *Unimodal* or *Polymodal* in the case with several clustering.



  - To be more precise, we use terms such as:
    - ✓ *Bimodal* or *Trimodal* to describe the exact number of clusters.

- How mean, median, and mode are ***related***?
  - For symmetric distributions: the ***mean*** and ***median*** are ***equal***.
  - For skewed distributions with a single mode the ***three measures differ***
    - ✓ ***Mean*** > ***median*** > ***mode*** (positively skewed distributions)
    - ✓ ***Mean*** < ***median*** < ***mode*** (negatively skewed distributions)



- ***How to choose the measure of central tendency to use???***
  - As the ***mean*** is strongly affected by outliers, but the ***median*** isn't.
    - ✓ If the data set contains qualitative data, use the ***mode***
    - ✓ If there is an outlier (or two) in a set of data, use the ***median***
    - ✓ In all other situations, use the ***mean***

ARKAN
ACADEMY

علم في كل مكان

Arkan academy

Arkanacademy

www.arkan-academy.com

+962 790408805