Journal of Pharmacology & Pharmacotherapeutics
Home
Current issue
Instructions
Submit article

J Pharmacol Pharmacother. 2011 Apr-Jun; 2(2): 140–142.
doi: 10.4103/0976-500X.81920: 10.4103/0976-500X.81920

PMCID: PMC3127352
PMID: 21772786

# Measures of central tendency: The mean

S. Manikandan

*Assistant Editor, Journal of Pharmacology and Pharmacotherapeutics*

**Address for correspondence:** Manikandan S, Department of Pharmacology, Indira Gandhi Medical College and Research Institute, Kadirkamam, Puducherry, India. E-mail: drsmanikandan001@gmail.com



In any research, enormous data is collected and, to describe it meaningfully, one needs to summarise the same. The bulkiness of the data can be reduced by organising it into a frequency table or histogram.[1] Frequency distribution organises the heap of data into a few meaningful categories. Collected data can also be summarised as a single index/value, which represents the entire data. These measures may also help in the comparison of data.

## CENTRAL TENDENCY

Central tendency is defined as "the statistical measure that identifies a single value as representative of an entire distribution."[2] It aims to provide an accurate description of the entire data. It is the single value that is most typical/representative of the collected data. The term "number crunching" is used to illustrate this aspect of data description. The mean, median and mode are the three commonly used measures of central tendency.

## MEAN

Mean is the most commonly used measure of central tendency. There are different types of mean, viz. arithmetic mean, weighted mean, geometric mean (GM) and harmonic mean (HM). If mentioned without an adjective (as mean), it generally refers to the arithmetic mean.

### Arithmetic mean

Arithmetic mean (or, simply, "mean") is nothing but the average. It is computed by adding all the values in the data set divided by the number of observations in it. If we have the raw data, mean is given by the formula

$$\text{Mean} \qquad \bar{X} = \frac{\Sigma X}{n}$$

Where, ∑ (the uppercase Greek letter sigma), *X* refers to summation, refers to the individual value and n is the number of observations in the sample (sample size). The research articles published in journals do not provide raw data and, in such a situation, the readers can compute the mean by calculating it from the frequency distribution (if provided).

$$\text{Mean} \qquad \bar{X} = \frac{\Sigma f X}{n}$$

Where, *f* is the frequency and *X* is the midpoint of the class interval and n is the number of observations.[3] The standard statistical notations (in relation to measures of central tendency) are mentioned in [Table 1]. Readers are cautioned that the mean calculated from the frequency distribution is not exactly the same as that calculated from the raw data. It approaches the mean calculated from the raw data as the number of intervals increase.[4]

## ADVANTAGES

The mean uses every value in the data and hence is a good representative of the data. The irony in this is that most of the times this value never appears in the raw data.

Repeated samples drawn from the same population tend to have similar means. The mean is therefore the measure of central tendency that best resists the fluctuation between different samples.[6]

It is closely related to standard deviation, the most common measure of dispersion.

## DISADVANTAGES

The important disadvantage of mean is that it is sensitive to extreme values/outliers, especially when the sample size is small.[7] Therefore, it is not an appropriate measure of central tendency for skewed distribution.[8]

Mean cannot be calculated for nominal or nonnominal ordinal data. Even though mean can be calculated for numerical ordinal data, many times it does not give a meaningful value, e.g. stage of cancer.

### Weighted mean

Weighted mean is calculated when certain values in a data set are more important than the others.[9] A weight $w_i$ is attached to each of the values $x_i$ to reflect this importance.

$$\text{Weighted mean} = \frac{\Sigma wx}{\Sigma w}$$

For example, When weighted mean is used to represent the average duration of stay by a patient in a hospital, the total number of cases presenting to each ward is taken as the weight.

### Geometric Mean

It is defined as the arithmetic mean of the values taken on a log scale. It is also expressed as the $n^{th}$ root of the product of an observation.

$$\text{Geometric mean (GM)} = \sqrt[n]{(x_1)(x_2)\ldots(x_n)}$$
$$\text{Log (GM)} = \frac{\Sigma(\log x)}{n}$$

GM is an appropriate measure when values change exponentially and in case of skewed distribution that can be made symmetrical by a log transformation. GM is more commonly used in microbiological and serological research. One important disadvantage of GM is that it cannot be used if any of the values are zero or negative.

### Harmonic mean

It is the reciprocal of the arithmetic mean of the observations.

$$\text{Harmonic mean (HM)} = \frac{1}{\frac{\Sigma(1/x)}{n}} = \frac{n}{\Sigma(1/x)}$$

Alternatively, the reciprocal of HM is the mean of reciprocals of individual observations.

$$\frac{1}{HM} = \frac{\Sigma(1/x)}{n}$$

HM is appropriate in situations where the reciprocals of values are more useful. HM is used when we want to determine the average sample size of a number of groups, each of which has a different sample size.

## DEGREE OF VARIATION BETWEEN THE MEANS

If all the values in a data set are the same, then all the three means (arithmetic mean, GM and HM) will be identical. As the variability in the data increases, the difference among these means also increases. Arithmetic mean is always greater than the GM, which in turn is always greater than the HM.[5]

The other measures of central tendency (median and mode) and the guidelines for selecting the appropriate measure of central tendency will be dealt with in the subsequent issue.

## Footnotes

**Source of Support:** Nil

**Conflict of Interest:** None declared

## REFERENCES

1. Manikandan S. Frequency distribution. *J Phamacol Pharmacother.* 2011;2:54–6. [PMCID: PMC3117575] [PubMed: 21701652]

2. Gravetter FJ, Wallnau LB. *Statistics for the behavioral sciences.* 5th ed. Belmont: Wadsworth – Thomson Learning; 2000.

3. Rao PS Sundar, Richard J. *Introduction to biostatistics and research methods.* 4th ed. New Delhi, India: Prentice Hall of India Pvt Ltd; 2006.

4. Sundaram KR, Dwivedi SN, Sreenivas V. *Medical statistics principles and methods.* 1st ed. New Delhi, India: BI Publications Pvt Ltd; 2010.

5. Norman GR, Streiner DL. *Biostatistics the bare essentials.* 2nd ed. Hamilton: BC Decker Inc; 2000.

6. Glaser AN. *High Yield Biostatistics.* 1st Ed. New Delhi, India: Lippincott Williams and Wilkins; 2000.

7. Dawson B, Trapp RG. *Basic and Clinical Biostatistics.* 4th ed. New York: Mc-Graw Hill; 2004.

8. Swinscow TD, Campbell MJ. *Statistics at square one.* 10th ed. New Delhi, India: Viva Books Private Limited; 2003.

9. Petrie A, Sabin C. *Medical statistics at a glance.* 3rd ed. Oxford: Wiley-Blackwell; 2009.

## Figures and Tables

## Table 1

Standard statistical notations

| |
|---|
| The population mean is denoted by the Greek letter "μ" (mu) and the sample mean by "$\bar{X}$." (In general, Greek letters are used to identify the characteristics of the population and English alphabets for the sample.) |
| Median and mode do not have a standard notation, but are sometimes mentioned as "Mdn" and "Mo," respectively. |
| The uppercase Greek letter sigma (Σ) denotes adding up a series of numbers whereas the lowercase sigma (σ) denotes standard deviation, which has an entirely different meaning. |
| The statistical measure that describes the sample is called "statistic" and that which describes the population is known as "parameter." |
| Therefore, every population parameter has a corresponding sample statistic. For example, the sample mean ($\bar{X}$) is a statistic and the population mean (μ) is a parameter. Readers can easily remember this by the mnemonic "P for p and S for s," i.e. Population – parameter; Sample – statistic. Beginners might be excited to learn that the statistical tests designed to test the hypotheses about population parameters are called parametric tests and those that do not state hypotheses in terms of population parameters are called "nonparametric tests."[2] |
| The number of subjects in the sample is denoted by N. There is no standard convention on whether to use the uppercase (N) or lowercase (n), but, generally, the lowercase n is used to denote the sample size for a group when there are two or more and the uppercase N for the entire sample.[5] |

Articles from Journal of Pharmacology & Pharmacotherapeutics are provided here courtesy of **Wolters Kluwer -- Medknow Publications**