

Measures of dispersion

S. Manikandan

Assistant Editor, JPP

INTRODUCTION

The measures of central tendency are not adequate to describe data. Two data sets can have the same mean but they can be entirely different. Thus to describe data, one needs to know the extent of variability. This is given by the measures of dispersion. Range, interquartile range, and standard deviation are the three commonly used measures of dispersion.

RANGE

The range is the difference between the largest and the smallest observation in the data. The prime advantage of this measure of dispersion is that it is easy to calculate. On the other hand, it has lot of disadvantages. It is very sensitive to outliers and does not use all the observations in a data set.^[1] It is more informative to provide the minimum and the maximum values rather than providing the range.

INTERQUARTILE RANGE

Interquartile range is defined as the difference between the 25th and 75th percentile (also called the first and third quartile). Hence the interquartile range describes the middle 50% of observations. If the interquartile range is large it means that the middle 50% of observations are spaced wide apart. The important advantage of interquartile range is that it can be used

as a measure of variability if the extreme values are not being recorded exactly (as in case of open-ended class intervals in the frequency distribution).^[2] Other advantageous feature is that it is not affected by extreme values. The main disadvantage in using interquartile range as a measure of dispersion is that it is not amenable to mathematical manipulation.

STANDARD DEVIATION

Standard deviation (SD) is the most commonly used measure of dispersion. It is a measure of spread of data about the mean. SD is the square root of sum of squared deviation from the mean divided by the number of observations.

$$SD = \sqrt{\frac{\sum(\chi - \bar{\chi})^2}{n-1}}$$

This formula is a definitional one and for calculations, an easier formula is used. The computational formula also avoids the rounding errors during calculation.

$$SD = \sqrt{\frac{\sum \chi^2 - \frac{(\sum \chi)^2}{n}}{n-1}}$$

In both these formulas $n - 1$ is used instead of n in the denominator, as this produces a more accurate estimate of population SD.

The reason why SD is a very useful measure of dispersion is that, if the observations are from a normal distribution, then^[3] 68% of observations lie between mean ± 1 SD 95% of observations lie between mean ± 2 SD and 99.7% of observations lie between mean ± 3 SD

The other advantage of SD is that along with mean it can be

Access this article online	
Quick Response Code:	Website: www.jpharmacol.com
	DOI: 10.4103/0976-500X.85931

Address for correspondence:

S. Manikandan, Department of Pharmacology, Indira Gandhi Medical College and Research Institute Hospital, Kadirkamam, Pondicherry, India. E-mail: drsmanikandan001@gmail.com

used to detect skewness. The disadvantage of SD is that it is an inappropriate measure of dispersion for skewed data.

APPROPRIATE USE OF MEASURES OF DISPERSION

SD is used as a measure of dispersion when mean is used as measure of central tendency (ie, for symmetric numerical data).

- For ordinal data or skewed numerical data, median and interquartile range are used.^[4]

REFERENCES

1. Swinscow TD, Campbell MJ. Statistics at square one. 10th ed. (Indian). New Delhi: Viva Books Private Limited; 2003.
2. Sundaram KR, Dwivedi SN, Sreenivas V. Medical statistics principles and methods. 1st ed. New Delhi: B.I Publications Pvt Ltd; 2010.
3. Gravetter FJ, Wallnau LB. Statistics for the behavioral sciences. 5th ed. Belmont: Wadsworth – Thomson Learning; 2000.
4. Dawson B, Trapp RG. Basic and Clinical Biostatistics. 4th ed. New York: Mc-Graw Hill; 2004.

How to cite this article: Manikandan S. Measures of dispersion. J Pharmacol Pharmacother 2011;2:315-6.

Source of Support: Nil, **Conflict of Interest:** None declared.