Sanni Ali, Sileshi Lulseged, Girmay Medhin. *Ethiop Med J, 2018, Vol. 56, No. 4*

TEACHING ARTICLE

EMJ Series on Methods and Statistics: Presenting and Summarizing Data – Part I

M Sanni Ali, DVM, MSc, PhD[1,2], Sileshi Lulseged, MD, MMed[3], Girmay Medhin, MSc, PhD[4]

ABSTRACT

Research data is collected on individual units of observation, which can only be interpreted meaningfully if analyzed and summarized using descriptive statistics. This, often done as an initial step in the analysis of a data set, provides simple summaries about the sample and about the observations that have been made. It aims to summarize a sample, rather than use the data to learn about the population that the sample data is believed to represent. The summaries may be in the form of summary statistics or in the form of tables or graphs. These summaries may either form the basis of the initial description of the data as part of a more extensive statistical analysis, or they may be sufficient in themselves for a particular investigation. Some measures that are commonly used to summarize or describe a data set are measures of central tendency or location and measures of dispersion or variability. Measures of central tendency or location include the mean, median and mode, while measures of dispersion or variability include the standard deviation (or variance), the range and interquartile range. The mean is the most informative measure that is used with both discrete and continuous variables (interval and ratio scale) and is strongly influenced by outliers (numerically distant from the rest of the data points). The median is the "mid-way" or central value of data organized in ascending or descending order. The median is a preferred measure of central tendency over the mean (or mode) when data is skewed (non-symmetrical) and is less influenced by outliers in a data set. Mode is the most frequent value(s) in a data set or the highest point of a peak on a frequency distribution and seldom used as a summary statistic. The measures of dispersion (standard deviation/variance, range, and inter-quartile range) are usually used in conjunction with a measure of central tendency, such as the mean or median, to provide an overall description of a data set. A thorough understanding of these measures used in descriptive statistics is critically important to do informed appraisal of the literature and in the analysis and write-up of scientific articles.

INTRODUCTION

Data analysis is the process we follow to understand what is contained within a data set collected from individual units of interest. Individual observations in a data set can always be examined individually. However, individual observations cannot be interpreted meaningfully until and unless described in a summarized and useful way. Indeed, to make sense of individual units in a data set, it is an accepted practice to begin by summarizing data using descriptive statistics, which essentially constitutes the first step in data analysis. Descriptive statistical method allows to summarize and describe a data set within the scope of the sample (1). The method aims to provide simple summaries about the sample and about the observations that have been made, rather than use the data to learn about the population that the sample data is drawn from.

Descriptive summaries on their own do not help to reach conclusions about hypotheses or make inference about the target population on which the investigation is based. These measures can only help to solve initial problems or identify essential issues within a more complex research project.

[1] Faculty of Epidemiology and Population Health, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK and

[2] Center for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK.

[3] School of Medicine, Addis Ababa University, Addis Ababa University, Ethiopia.

[4] Aklilu Lemma Institute of Pathobiology, Addis Ababa University Addis Ababa, Ethiopia.

Descriptive summaries may be sufficient in and of themselves for a particular investigation (2). For example, agencies that carry out Demographic Health Survey and Census finish their investigation with the report based on descriptive summary of the collected data. Descriptive statistics may also form the basis of the initial description of the data as part of a more extensive statistical analysis. Even when a data analysis draws its main conclusions using inferential statistics, For example, in papers reporting on human subjects, typically a table is included giving the overall sample size, sample sizes in important subgroups (e.g., for each treatment or exposure group), and demographic or clinical characteristics such as the average age, the proportion of subjects of each sex, the proportion of subjects with related comorbidities,

Descriptive statistics typically involves either quantitative (summary statistics), or visual (simple-to-understand graphs and/or tables). It is often convenient to summarize a numerical variable in terms of two quantities or measurements, one indicating the average or the center of the data set (e.g. mean, median, and mode) and the other indicating the spread of the values (e.g. range, interquartile range, standard deviation or variance. Summarizing data in tables and/or graphs provides several advantages: 1) familiarity with the data with respect to the scale and distributions of variables, 2) outlier identification (an observation point that is distant from other observations) and a way for processing data, e.g. transformation, 3) revealing possible error in the data, and 4) checking for assumptions required for statistical testing (3, 4). Tabular and graphic presentation of is presented in part II, while the section below (part I) in this Issue of EMJ describes quantitative summary statistics.

Measures of central tendency

Central tendency is a statistical measure that identifies a single value as representative or typical of set of data. It aims to provide an accurate description of the entire data (5). As such, measures of central tendency are sometimes called measures of location. The mean, median, and mode are different measures of center in a numerical data set: each try to summarize a data set with a single number to represent a "typical" data point from the data set. All are valid measures of central tendency, but under different conditions.

Mean

The arithmetic mean, customarily just called the mean, is the most popular and well known average value of measures of central tendency. It is calculated as the sum of all the values in a data set divided by the total number of values in the data set. If the data set was based on a series of observations obtained by sampling from a population, the arithmetic mean is the sample mean (denoted $\overline{X}$, spoken as X bar) to distinguish it from the mean of the population from which the sample was obtained, the population mean (denoted as μ or $\mu_{x,}$ spoken as mu or mu x).

The mean, $\overline{X}$, of a set of n values x_1, x_2, x_3, …, x_n is denoted by the mathematical formula:

$$\text{Mean, } \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} \quad (1)$$

where $\sum_{i=1}^{n} x_i$ is read as the sum of all values of xi where I starts from 1 up to n, i.e. the sum of all the values (x_1, $x_{2,}$ x_3, …, x_n). $\sum$ (the Greek capital letter sigma) refers to “the sum of “ and n is the number of data points in the sample.

For example, Table 1 presents a synthetic data set on systolic blood pressure (in mmHg) measurements of 40 individuals. The mean systolic blood pressure measurement in this example is 120.5 mmHg (see Table 1 for the calculation).

Table 1. Systolic blood pressure measurements of 40 Individuals

75	100	100	105	105	105	110	110	110	110
115	115	115	115	115	115	120	120	120	120
120	120	120	120	125	125	125	125	125	125
130	130	130	130	135	135	135	140	140	175

The mean is the most commonly used average because it is readily understood and has several mathematical properties that make it useful, especially as a measure of central tendency:

1) The mean is the only single number, for which the sum of the residuals (deviations) of each value from the estimate (the mean) is always zero. In other words, the values to the left of the mean are balanced by the values to the right of the mean. If values $x_1, \ldots, x_n$ have mean $\overline{X}$, then $(x_1 - \overline{X}) + \ldots + (x_n - \overline{X}) = 0$. Since $x_i - \overline{X}$ is the distance from a given number to (deviations from) the mean.

2) The mean is also the best single predictor in the sense of having the lowest root mean squared error, i.e. it minimizes error in the prediction of any one value in a given data set. In other words, if it is required to use a single value as a "typical" value for a set of known values $x_1, \ldots, x_n$, then the mean of the values does this best, in the sense of minimizing the sum of squared deviations from the typical value: the sum of $(x_i - \overline{X})^2$.

3) The mean of a sample ($\overline{X}$) drawn from the population is an unbiased estimate of the population mean (μ or μ_x). The mean of a sample gets closer to, or converges on, the population mean as the sample size increase. This property is also called the Weak Law of Large Numbers (6).

4) The mean is the most informative measure since it includes every value in the data set as part of the calculation (see Equation 1).

The mean can be used with both discrete and continuous variables (interval and ratio scale), although its use is most often connected with continuous data such as age, blood pressure, body mass index (BMI), CD4 count, and heart rate. The mean values from ordinal scale data (such as Apgar scale, Glasgow coma score, Trauma score, Framingham score, and CHA2DS2VASc score) are generally misleading or invalid due to the lack of consistent level of magnitude between numeric units of the scale (7). The mean is not a useful measure when data is skewed, it will not

accurately represent the data: it loses its ability to provide the best central location for the data because the skewed data is dragging it away from the "typical" value.

The mean has one main disadvantage: it is strongly influenced by outliers. Outliers are values that are unusual or extreme compared to the majority of the data points by being especially small or large in numerical value. In other words, an outlier is an observation that is numerically distant from the rest of the data (See Interquartile Range below). Outlier may be the result of measurement error, coding error, or extreme variability in an observation. It can be identified by visual inspection of the data, graphical displays, or complex modelling. Depending on the number of outliers, they are either statistically transformed (using a complex statistical formula to balance all other values) or excluded from the data set.

Arithmetic mean is a special case of weighted arithmetic mean where all values have an equal weight of one. In survey data and meta-analysis, it is common for different observations to have different weights (indicating different contributions or representations, some observations or samples have more weights than others). In this case, the mean calculated is called the weighted arithmetic mean (for short, the weighted mean). For example, for a set of n values x_1, x_2, ..., x_n, with weights w_1, w_2, …, w_n, respectively, the weighted mean is defined as:

$$\text{Weighted Mean, } \overline{X} = \frac{\sum_{i=1}^{n} x_i * w_i}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} x_i * w_i = \frac{x_1 * w_1 + x_2 * w_2 + x_3 * w_3 + \ldots + x_n * w_n}{w_1 + w_2 + w_3 + \ldots + w_n} \quad (2)$$

where $\sum_{i=1}^{n} x_i * w_i$ is read as the sum of all the products of the values of x_i and their weights w_i where I starts from 1 up to n, i.e., the sum of all the values multiplied by their respective weights ($x_1 * w_1$, $x_2 * w_2$, $x_3 * w_3$, …, $x_n * w_n$). $\sum_{i=1}^{n} w_i$ is the sum of the weights (w_1, w_2, w_3, … , w_n).

Although not often used as arithmetic mean in descriptive statistics, there is another type of average or mean called the geometric mean. The geometric mean indicates the central tendency or "typical" value of a data set using the product of their values (as opposed to the arithmetic **mean** which uses their sum, as described above). It is defined as the *n*th root of the product of *n* numbers, for example, for a set of n values x_1, x_2, ..., x_n, the geometric mean is defined as:

$$\text{Geometric Mean, } \overline{X} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = \sqrt[n]{x_1 * x_2 * x_3 * \ldots * x_n} \quad (3)$$

Where $\prod_{i=1}^{n} x_i$ is read as the product of all the values ((x_1, x_2, x_3, …, x_n). Π (the Greek capital letter pi) refers to "the product of " and n is the number of data points in the sample.

Median

The median is the "mid-way" or central value of data that has been sorted in (ascending or descending) order of magnitude. Half the values lie below the median and half the values lie above it. It is less influenced by outliers in a data set, and more useful than the mean to describe such data sets. It is also useful for summarizing ordinal data as the magnitude of the difference between values of a data set need not be consistent to determine the median. However, it is not useful for describing nominal data such as gender or blood groups due to the arbitrary selection of numbers to denote this scale (7).

The median is a preferred measure of central tendency over the mean (or mode) when data is skewed (i.e., the frequency distribution for the data is non-symmetrical) or if there are one or two outliers which would make the mean unrepresentative of the majority of the data. In skewed data, unlike the mean, the median best retains its position and is not as strongly influenced by the skewed values (Figures 1A and 1C). Alternatively, geometric mean may be an appropriate summary when the distribution is positively skewed, i.e. when the most of the observations are around the centre but some on the right tail of the distributions.

If there is odd number of observations, the median is simply the middle value; if there is even number of values (n) then there is no middle one, hence the median is calculated as the average of the middle two values.

$$\text{Median} = \frac{(n+1)}{2} \text{ th value of the ordered values} \quad (4)$$

For example, for the data set in Table 1, the 40 systolic blood pressure measurements, the number of observations (n=40) is even. Hence, the median will be the $[(40+1)/2]^{th}$ term, i.e. the 20.5^{th} value which is the average of the 20^{th} and 21^{st} values, 20 (120+120 divided by 2 = 120) (Table 2).

Mode

The mode is the most frequent value(s) in a data set or the highest point of a peak on a frequency distribution. It is most useful to describe bi-modal distributions (when two clusters of data exist) and categorical or nominal data such as blood type (,"A", "B", "AB" and "O"), sex, race, or education level defining the most prevalent characteristics of a sample. The mode can also be calculated for quantitative or continuous data.

The mode is seldom used due to several issues: 1) It is not necessarily unique in a given data set. Hence, there are times one can face difficulty to choose which model value that describe the data set best when there are two or more values that share the highest frequency (multi-modal

distribution). This could be particularly problematic with continuous data. 2) It does not provide a very good measure of central tendency when the most frequent value is far away from the rest of the values in the data set.

Equality of mean, median and mode in a data set

In a situation where the data is symmetrical and unimodal or does have perfect normal distribution the three measures of central tendency (i.e. the mean, median and mode) are all identical (Figure 1A).

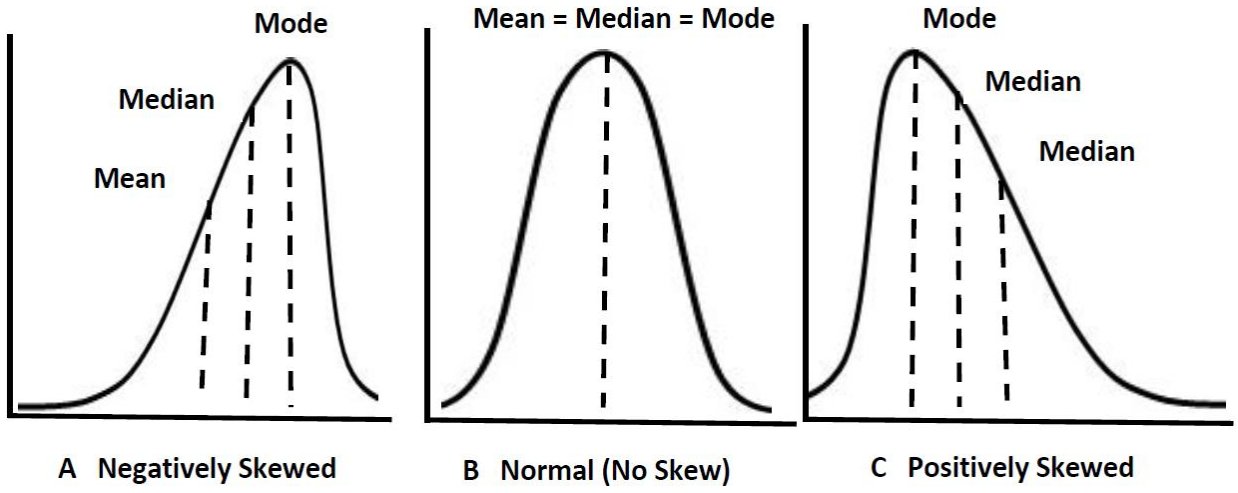

Figure 1: Density plots of different distributions. Negatively skewed (A), perfectly Normal (no skewness, B), and positively skewed (C).

Quantiles

When presenting and analyzing continuous variables, one way of categorizing observations is the use of cut-off points that classify observations into intervals where all intervals take equal proportion of the data points. These cut-off values are called quantiles (8). For example, the systolic blood pressure measurements in Table 1 can be grouped in to (a) ten equal parts using 9 deciles, (b) five classes using 4 quintiles, (c) four classes using 3 quartiles, or (d) three classes using 2 tertiles. Similarly, percentiles or centiles in short, divide a continuous variable in to hundred equal groups. The terms tertiles, quartiles, quintiles, deciles, and percentiles (or centiles) refer to the cut-off points rather than the groups obtained using these cut-off values.

Quartiles are used to form four equal groups each containing 25% of the observations. In this case, only three cut-off values are required and hence, there are three quartiles (often represented as Q_1, Q_2, and Q_3). The second quartile (Q_2) is known as the median since it divides the observations ordered in terms of magnitude in to two equal halves. The lower quartile (Q_1) is located one quarter of the way along the data set when the values have been sorted in order of magnitude and the upper quartile (Q_3) is located three quarters along the data set. Similarly, when using percentiles, the 25th percentile and the 75th percentiles are equal to the first and third quartiles, respectively, whereas the 50th percentile is the same as the second quartile or the median (8). Percentiles are often used to describe cumulative frequency distributions of continuous variables. Generally, the **K**th percentile is the point below which k% of the values of the distribution lie. Hence, for a distribution with n values, the kth percentile is defined as,

$$\text{Kth Percentile} = \frac{K * (n+1)}{100} \text{ th value of the ordered observations} \quad (5)$$

Measures of Spread

A measure of spread, sometimes also called a measure of dispersion or measures of variation, is used to describe the variability that exist in a data set. It is usually used in conjunction with a measure of central tendency, such as the mean or median, to provide an overall description of a data set. Examples of measures of spread include range, interquartile range, variance and standard deviations.

Range

The range is the difference between the lowest and highest values in a data set. It the most obvious measure of variability.

$$\text{Range} = \text{Highest value} - \text{Lowest value} \quad (6)$$

It is useful for showing the spread within a data set as well as for comparing the spread between similar data set s. In the systolic blood pressure measurement (Table 1), the highest measurement is 175 and the lowest measurement is 75. Hence, the difference of these two numbers is 100, representing the range. Since range is based solely on the two most extreme values within the data set, there is a risk that it may not be representative of the variability within the data set and can sometimes be misleading. This becomes obvious if one of the two extreme values or both extreme values are outliers. One remedy to reduce the impact of outliers is to use inter-quartile range, which considers the variability of the middle 50% of the data set.

Interquartile range

The inter-quartile range (IQR) is a measure that indicates the extent to which the central 50% of values within the data set are dispersed. It is based upon, and related to, the median or the second quartile (Q_2) hence often reported with the median as a measure of spread. The lower quartile (Q_1) lies halfway between the median (Q_2) and the lowest value in the data set whilst the upper quartile (Q_3) lies half way between the median (Q_2) and the highest value in the data set. The observations between the lower and upper quartiles, Q_1 and Q_3 constitute 50% of the total observations in the data set. The inter-quartile range is calculated by subtracting the lower quartile (Q_1) from the upper quartile (Q_3).

$$\text{IQR} = \text{Upper quartile } (Q_3) - \text{Lower quartile } (Q_1) \quad (7)$$

Quartiles and IQR are used in plotting of box and whisker plots as well as to identify outliers. In a boxplot, an outlier is often defined as a value that is located outside the fences (“whiskers") of the boxplot (e.g. outside 1.5 times the IQR above the upper quartile and below the lower quartile.

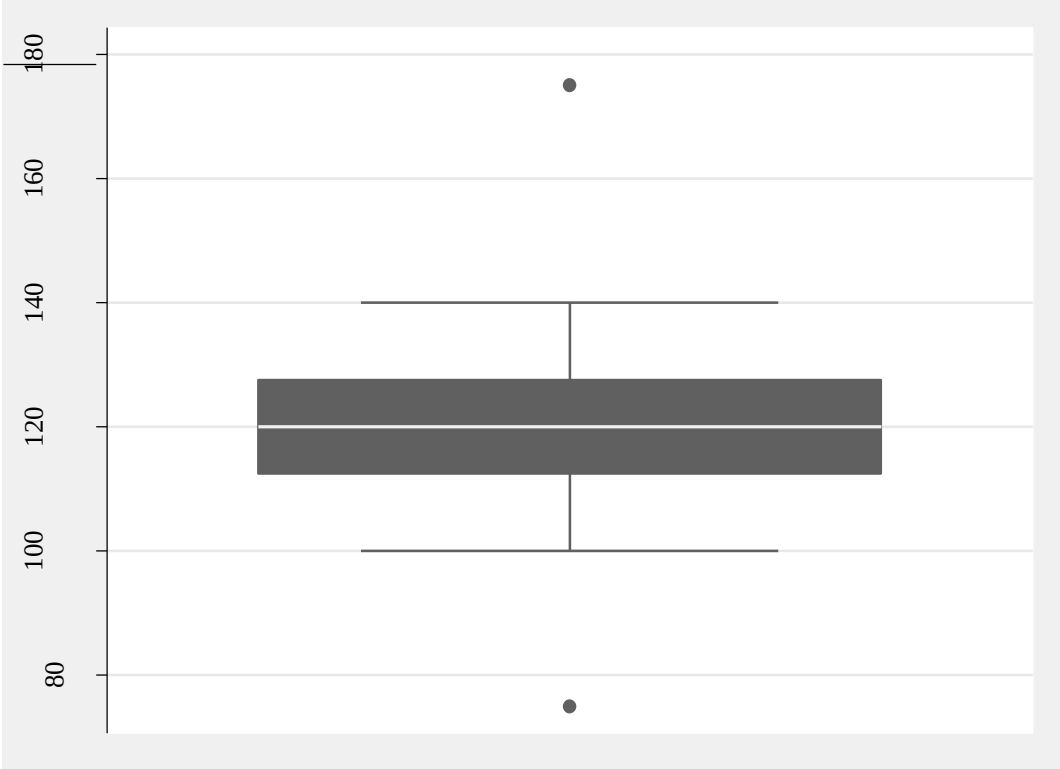

Figure 2: Box plot of systolic blood pressure measurements

For example, in the systolic blood pressure measurements presented in Table 1 and Figure 2, Q_1 and Q_3 are 113.8 and 126.2, respectively. The IQR is 12.4 hence the two values 75 and 175 are

outside 1.5 times the IQR below the lower quartile (Q_1) and above the upper quartile (Q_3), respectively. Hence, both can be considered outliers in the data set.

Variance

Variance is the average of the squared deviations (differences) of individual values from the mean. It measures how far a data set is spread out from their average value or the mean of the data set. A value of zero means that there is no variability; all the values in the data set are the same or identical hence the deviations of individual values from the mean are all zero. Variance is not generally used in data description but it has a central role in statistics. It is used in statistical inference, hypothesis testing such as analysis of variance (ANOVA) and goodness of fit tests which will be discussed in detail in subsequent reviews in this series.

Variance is often represented by σ^2, referring to population variance and s^2, referring to sample variance or Var(X). It is calculated by squaring the deviations of each individual values from the mean, adding them together and dividing by the total number of values (n) for population variance and by n-1 (for sample variance), using the following mathematical formulas, population variance (Equation 6) and sample variance (Equation 7):

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2 = \frac{(x_1-\bar{X})^2+(x_2-\bar{X})^2+(x_3-\bar{X})^2+\ldots+(x_n-\bar{X})^2}{n} \quad (8)$$

$$\text{Var(X)} = = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2 = \frac{(x_1-\bar{X})^2+(x_2-\bar{X})^2+(x_3-\bar{X})^2+\ldots+(x_n-\bar{X})^2}{n\text{ -1}} \quad (9)$$

Where $\sum_{i=1}^{n}(x_i - \bar{X})^2 = (x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + (x_3 - \bar{X})^2 + \ldots + (x_n - \bar{X})^2$ and $\bar{X}$ is the mean described in equation 1

The deviations are squared to overcome any negative deviations from the mean since it is not possible simply to average the deviations. As mentioned before, the deviations from the mean always add up to zero, the positive deviations of values above the mean will balance out the negative deviations of values below the mean. Squaring the deviations, as opposed to ignoring the minus sign, also makes the summary measure mathematically tractable (3, 9).

While working with the sample data, dividing the sum of deviations by (n-1) rather than n, gives a better estimate of the variance of underlying population. The n-1 in the denominator is known as the degree of freedom (df) of the variance. This is the same as the number of independent

deviations (from the mean) available to estimate the variance. The n deviations from the mean should add up to zero. Hence, there are only n-1 rather than n independent deviations available to calculate variance since the last one (i.e. n^{th} deviation) can always be calculated from the rest.

A disadvantage of the variance is that it is measured in square of the units used for individual measurements. For example, in the systolic blood pressure measurements, individual units are mmHg whereas the unit if variance is $mmHg^2$:

Standard deviation

The standard deviation (SD) is a measure that summarizes the amount by which every value within a data set deviates from the mean. It is calculated as the square root of the variance. While variance gives us a rough idea of spread, the standard deviation is more concrete, that provides average distances from the mean across all observations. It is an indication of how closely the values in the data set are bunched around the mean value. The standard deviation is small when the values in a data set are bunched closely around the mean (and it is zero if all the values are equal to the mean) and large when the values are scattered over considerable distance (Figure 3).

Similar to variance, standard deviation is often represented by δ, referring to population standard deviation and s or SD, referring to sample standard deviation.

The standard deviations of n observations, sampled from a population, with means equal to $\bar{X}$ is calculated using the following formula:

$$\delta = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n}} = \sqrt{\sigma^2} \quad (10)$$

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n-1}} = \sqrt{Var\ (X)} \quad (11)$$

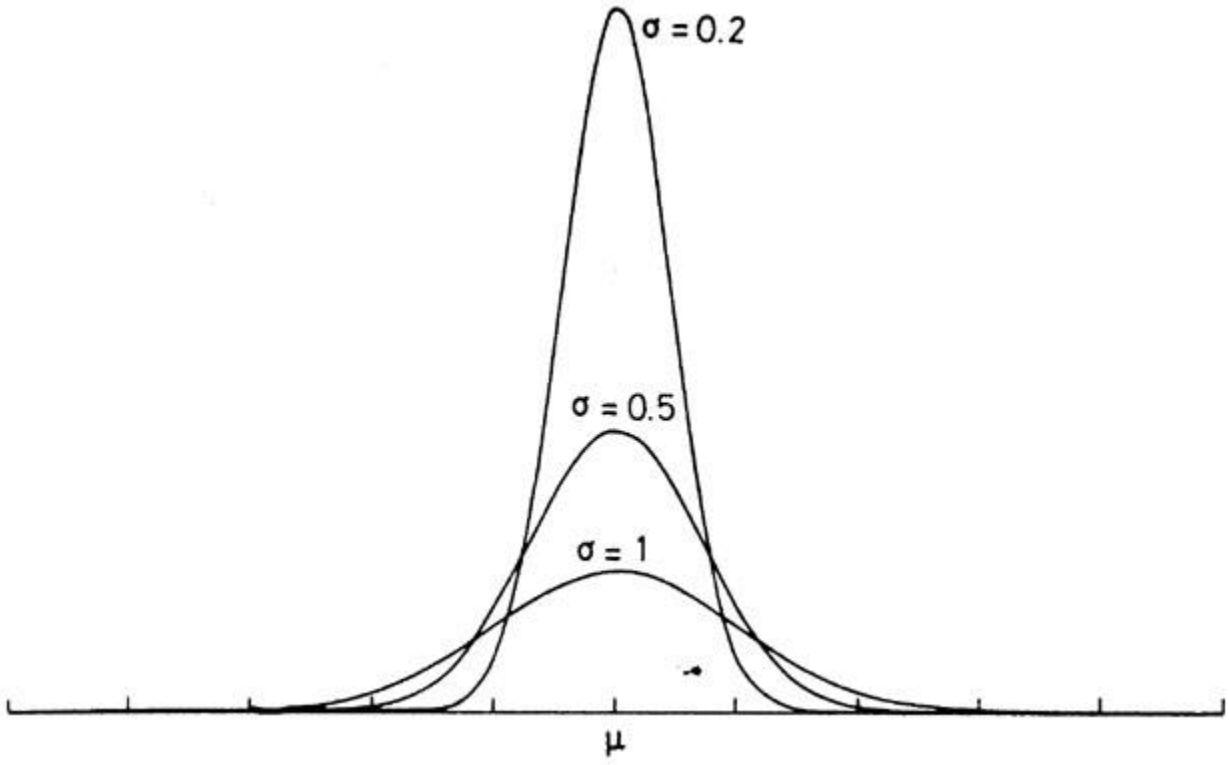

Figure 3: Density plots of Normal distributions with the same mean (μ) but different variations (standard deviations, δ).

Table 2 shows how to calculate standard deviation and other measures of location and spread (range, interquartile range, and variance) using the example data set on systolic blood pressure measurements (in mm Hg) presented in Table 1.

The standard deviation is the most robust and widely used measure of variation usually presented in conjunction with the mean. Unlike the range and inter-quartile range, it takes into account every value in the data set. Unlike variance, standardized deviation is measured in the same units as the mean and, like the mean, it summarizes a great deal of information in one number and has useful mathematical properties. For example, using the standard deviation, we could tell the percentage of values or observations that are within a certain distance (standard deviations, SD) from the mean. For normally distributed variable, 68.26% of the observations lie within 1SD from the mean; 95.44% of the observations lie within 2 SD from the mean; and 99.7% the observations lie within 3 SD from the mean; and 95% of the observations lie between in mean-1.96 SD and mean+1.96 SD (9).

Table 2. Computing different summary measures of location and spread for the systolic blood pressure measurements (in mmHg) in Table 1

Values (x_i)	Deviations, $x_i - \bar{x}$	Squares of Deviations	Quartile
75	75 - 120.25 = -45	$-45^2 = 2025$	1
100	100 - 120.25 = -20.25	$-20.25^2 = 410.06$	1
100	100 - 120.25 = -20.25	$-20.25^2 = 410.06$	1
.	.	.	.
.	.	.	.
.	.	.	.
115	115 - 120.25 = -5.25	$-5.25^2 = 27.56$	2
115	115 - 120.25 = -5.25	$-5.25^2 = 27.56$	2
115	115 - 120.25 = -5.25	$-5.25^2 = 27.56$	2
.	.	.	.
.	.	.	.
.	.	.	.
120	120 - 120.25 = -0.25	$-0.25^2 = 0.0625$	3
120	120 - 120.25 = -0.25	$-0.25^2 = 0.0625$	3
120	120 - 120.25 = -0.25	$-0.25^2 = 0.0625$	3
.	.	.	.
.	.	.	.
.	.	.	.
140	140 - 120.25 = 19.75	$19.75^2 = 390.06$	4
140	140 - 120.25 = 19.75	$19.75^2 = 390.06$	4
175	175 - 120.25 = 54.75	$54.75^2 = 2997.56$	4
Total number (n) = 40		$\bar{x} = 120.25$	
Total sum = $\sum_{i=1}^{40} x_i = 4810$		$\sum_{i=1}^{40}(x_i - \bar{x})^2 = 9097.5$	
Mean = $\frac{1}{n}\sum_{i=1}^{n} x_i = 4810/40 = 120.25$		Var(x) = $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 9097.5/39 = 233.26$	
Median = ((n+1)/2)th value = 20.5th value = Average of the 20th and 21st values = (120+120)/2 = 120		SD = $\sqrt{Var(x)} = \sqrt{233.26} = 15.27$ Range = Largest- Smallest = 175-75 = 150	
Mode = 120		IQR = $Q_3 - Q_1 = 126.2 - 113.8 = 12.4$	

Standard Error of the mean (SE)

Standard Error of the mean is a measure of sampling variability of the mean, i.e., when we have several samples drawn from a population, the mean of the sample means is equal to the population mean as described in the previous series on distributions. However, the variation in the sample means is different from the standard deviation of observations in a sample. The standard deviations of the sample means is referred to as the standard error (SE) of the mean.

Standard error of the mean measures the average deviation of individual sample mean from the population mean. Let us assume that there is a defined population where the true population mean of a given variable, say Y, is μ_y. Let us consider all possible samples of size n (assume that there are only m possible samples of size n). For one of these possible samples, say kth sample, it is easy

to calculate sample mean, $\bar{X}_k$. Hence, there will be one mean for each sample resulting in m different sample means for m possible samples of size n. Each of these sample means is meant to serve as an estimate of the true population mean implying the need for the measure of variability among these sample means taking the population mean as a reference point. This measure of variability is called standard error of the mean and calculated as follows:

$$\text{SE}(\bar{X}) = \sqrt{\frac{\sum_{k=1}^{m}(\bar{X}_k - \mu)^2}{m}} \quad (12)$$

It is important to distinguish standard deviation from standard error and the relationship between them. Standard deviation tells the amount of variation between observations in a given sample of size n. On the other hand, standard error tells us variation between sample means obtained from all possible samples of size n generated from the same underlying population. Without requiring to take all possible samples of size n from a given population, standard error for the mean can also be calculated using the SD obtained from one random sample of size n using the following formula:

$$\text{SE}(\bar{X}) = \frac{SD}{\sqrt{n}} \quad (13)$$

References

1. Prem S Mann, Introductory Statistics (6th ed). John Wiley & Sons, 2006.
2. Trochim, William M. K. (2006). "Descriptive statistics". Research Methods Knowledge Base. Retrieved 15 September 2018
3. Elise Whitley and Jonathan Ball. Statistics review 1: Presenting and summarizing data. Critical Care 2002, 6: 143.
4. J Martin Bland and Douglas G Altman. Transformations, means, and confidence intervals. BMJ 1996, 312: 1079.
5. Gerald Van Belle et al. Biostatistics: a methodology for the health sciences. Vol. 519. John Wiley & Sons, 2004.
6. Robert H Riffenburg. Statistics in Medicine (3rd ed). Elsevier, 2012.
7. Gary M Gaddis and Monica L Gaddis. Introduction to biostatistics: Part 2, descriptive statistics. Annals of Emergency Medicine 1990, 19: 309-315
8. Douglas G Altman and J Martin Bland. Quartiles, quintiles, centiles, and other quantiles. BMJ 1994, 309: 996.
9. Betty R Kirkwood and Jonathan AC Sterne. Essential medical statistics. John Wiley & Sons, 2010.