# Chapter 2 Statistical Foundations: Descriptive Statistics

Presented in this chapter is a discussion of the types of data and the use of frequency tables; figures (graphs and charts); rates, ratios and proportions, measures of central tendency (mean, median, and mode); measures of variation (range, standard deviation, and variance); and the correlation to summarize data. Measures of position are presented in Chapter 6. Spatz (2011) provides detailed discussions of Chapter 2 topics.

The statistical foundation for assessment is based upon descriptive statistics which includes:

a. Frequency counts (and their presentation devices, i.e., tables and figures);
b. Measures of Central Tendency; and
c. Measures of Variation.

The proper classification of data is critical to selecting appropriate descriptive statistics to quantitatively describe a data set and/or a suitable inferential statistical procedure. For example, correlations are used to summarize a measure's (e.g., test) validity and reliability characteristics.

The two simplest types of data are nominal and ordinal, while the more complex types are interval and ratio. Nominal and ordinal data are typically analyzed with non-parametric statistical procedures. Parametric statistical applications are applied to interval and ratio data. See Spatz (2011, pp. 10-11).

Nominal data consists of names, labels, or categories only, with some underlying connection. The underlying connection for the two nominal categories, boy or girl is gender. Nominal data can't be arranged in an ordering scheme. Categories must be mutually exclusive. Examples are

| Gender | Birth Place | Color |
|---|---|---|
| Boy/Girl | European | Red |
| | African | Yellow |
| | Asian | Blue |

The number of boys or girls is also considered discrete data; because the number associated with each category (boy or girl) indicates the number of children in each category as in 12 boys and 13 girls. Later in this chapter, you will read about measures of central tendency (MCT); the only MCT applicable to nominal data is the mode.

Ordinal data are ordered categories, but distances between categories can't be determined. The ordering of the categories has significance. Later in this chapter, you will read about measures of central tendency (MCT); the only MCT's applicable to ordinal data are the mode and median, never the mean. Each category or grouping has an underlying connection. In the first ordinal data example "High," "Middle," and "Low" income categories, so the underlying connection between the 3 groups is income.

| Income | Grades | Likert Scale |
| --- | --- | --- |
| High | A | Strongly Disagree (SD) |
| Middle | B | Disagree (D) |
| Low | C | No Opinion (NO) |
| | D | Agree (A) |
| | F | Strongly Agree (SD) |

Moving to the Likert scale example, the connection between the 5 categories is strength of agreement or disagreement. The ranking scale (e.g., A to F) is the framework for interpretation. We know an "A" is higher than a "C", but we don't know by exactly by how much. "Strongly Disagree" is an opinion which is very different from "Strongly Agree," but exactly how different is unknown.

Interval data is similar to ordinal data; but, distances between points on a specified measurement scale (e.g., inches, pounds, kilograms, miles, kilometers, or standard scores commonly used in standardized testing) can be identified. There is a uniform "distance" between each position on the measurement scale. Take for example a classroom achievement test with the measurement scale of 1-100 points; a score of 80 is exactly 5 points less than a score of 85, and exactly 5 points more than a score 75. Interval data are considered continuous; because interval data can be added, subtracted, multiplied, or divided which can result in a fractionalized whole number such as 87.33. Even whole numbers, which are associated with a specific measurement scale, are considered continuous. Since interval data has no zero starting point, ratios are meaningless and not computed. Later in this chapter, you will read about measures of central tendency (MCT); all three MCT's are applicable to interval data.

| Temperature | Grades | | Likert Scale | |
| --- | --- | --- | --- | --- |
| 90º F | A | 4.0 | SD | 1 |
| 60º F | B | 3.0 | D | 2 |
| 45º F | C | 2.0 | NO | 3 |
| 32º F | D | 1.0 | A | 4 |
| -32º F | F | 0.0 | SA | 5 |

Forty-five degrees is not twice as hot as 90º F (because there is no true zero starting point), but the distance between 45º F and 90º F is 45 degrees. Now, we see an "A" is two points higher than a "C." And that "SA" is 4 points higher than "SD." In many social science disciplines, it is common practice to "convert" the ordinal Likert or Likert style scale data into interval data by assigning numbers, such as "1" for "Strongly Disagree" or "5" for "Strongly Agree." The "Strongly Disagree," "Disagree," "No Opinion," etc. like the "A," "B," "C" remain ordinal, the addition of the numbers allows the "transformation" of ordinal data into interval data. Among researchers, statisticians, and evaluators, this practice is controversial.

Ratio data is similar to interval data, but with a real zero "0" starting point. Ratios are meaningful. For example, $100 is exactly twice $50 and 12 meters is exactly twice as much as 6 meters. Later in this chapter, you will read about measures of central tendency (MCT); all three MCT's are applicable to interval data.

| Dollars | Height |
|---|---|
| $100 | 12 meters |
| $50 | 6 meters |

There is a less complicated and alternative taxonomy to classify data: Discrete and Continuous. Discrete data represent counts, e.g., # of telephones, usually as whole numbers; nominal and ordinal data are usually considered discrete. Continuous data represent measurements, e.g., life of a light bulb in days, hours or minutes, expressed as whole numbers (e.g., 1, 40, 102, etc.) or whole numbers, with decimal points (e.g., 1.3, 99.5, 131.321, etc.). Interval data, whether factionalized or not, are considered continuous because they are dependent upon established or articulated measurement scale. Discrete data (nominal or ordinal) are used with non-parametric statistics and continuous data (interval, and ratio) with parametric statistics.

We will explore data reduction techniques to make modest, medium, or large data sets (e.g., test scores) more easily understandable. First, we will explore specific tables and figures which will help us to understand numerical distributions (e.g., test scores for a class or training group). Second, we will explore rates, ratios, proportions, and percentages. Third, measures of central tendency (mean, median, and mode) will be examined. Fourth, we will review measures of variation (range, variance, and standard deviation). Fifth, we will consider correlation.

## I. Data Interpretation: Tables and Figures

A. Frequency Tables

1. The frequency table lists categories (also called classes) of scores or other objects of interest along with counts (called frequencies) of the number of scores, etc. that fall into each category. Spatz (2011, pp. 26-30) discusses frequency tables. Frequency tables are used with nominal or ordinal data categories. There are three types:
   a. Simple Frequency Table (Table 2.1)
   b. Relative Frequency Table (Table 2.2)
   c. Cumulative Frequency Table (Table 2.3)

2. Frequency tables are constructed so that a sense of order is imposed upon a number of individual data points. This is usually among the first steps in data analysis. So to ensure reasonably consistent interpretation of tables, follow the guidelines presented below.
   a. Guidelines for Constructing Tables
      (1) Tables should be a simple as possible. Two or three small tables are preferred to a single large table containing many details or variables. Usually, a maximum of three variables can be read with ease.
      (2) Tables should be self-explanatory.
         (a) Codes, abbreviations, or symbols should be labeled and explained in a footnote within the table.
         (b) Each row and column should be clearly and concisely labeled.
         (c) Give the specific units of measurement for the data.
         (d) The title should be clear, concise, and explicit. The title should answer, what?, when?, and where?. The title is routinely separated from the body

of the table by lines, or spaces. In small tables vertical lines separating columns aren't usually necessary.

(3) If not presenting original data, then the source must be fully cited in a footnote in such detail so that an interested reader may obtain his or her own report copy.

(4) When it is necessary to construct large, complicated tables, reserve those to an appendix, as such will break the flow of reading and understanding.

b. Guidelines for Constructing Class Intervals

(1) Each class interval must be mutually exclusive.

(2) Each class interval should be of equal width

(3) Keep the number of class intervals reasonable, no more than 10 or so.

Table 2.1 (Simple Frequency Table)

*XYZ Company Employee Absences by Hour for April-June 2011*

| Employee Absences in Hours | Frequency (Raw count) |
|---|---|
| 0-5 | 153 (59.5%) |
| 6-11 | 37 (14.40%) |
| 12-17 | 24 (9.34%) |
| 18-23 | 18 (7.00%) |
| 24-29 | 12 (4.70%) |
| 30-35 | 13 (5.06%) |
| Total | 257 |

*Note*. Data are from the XYZ Company. Most employees had fewer than 11 hours of absences during April-June 2011; but, 13 were absent over 30 hours.

Table 2.2 (Relative Frequency Table)

*XYZ Company Employee Absences by Hour for April-June 2011*

| Employee Absences in Hours | Frequency (Proportion) |
|---|---|
| 0-5 | .595 |
| 6-11 | .144 |
| 12-17 | .093 |
| 18-23 | .070 |
| 24-29 | .047 |
| 30-35 | .051 |
| Total | 1.00 |

*Note*. Data are from the XYZ Company. Sixty percent of employees had fewer than 11 hours of absences during April-June 2011; but, 5% were absent over 30 hours.

Table 2.3 (Cumulative Frequency Table)

*XYZ Company Employee Absences by Hour for April-June 2011*

| Employee Absences in Hours | Frequency | Cumulative |
|---|---|---|
| 30-35 | 13 (100%) | 257 |
| 24-29 | 12 (94.9%) | 244 |
| 18-23 | 18 (90.2%) | 232 |
| 12-17 | 24 (83.2%) | 214 |
| 6-11 | 37 (73.9%) | 190 |
| 0-5 | 153 (59.5%) | 153 |

*Note*. Data are from the XYZ Company. Two-hundred fifty-seven (257) employees were absent 17 hours or less during April to June 2011. One hundred percent of employees had less than 35 hours of absence during the quarter; 94.9% of employees were absent 29 or fewer hours; and 90% of employees were absent 23 or less hours.

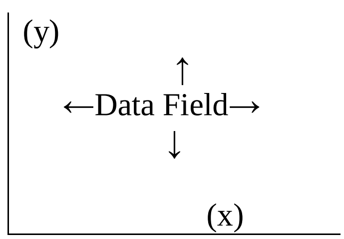
B. Figures (Graphs and Charts)

1. Graphs: Introduction

a. A graph is a method of showing quantitative data using a coordinate system on two axes usually "x" and "y". When constructed correctly, graphs allow a reader to develop an overall grasp of the data. Spatz (2011, pp. 30-33) discusses figures.

(1) Rectangular coordinate graphs are those which consist of two lines which form a right angle. Each line is identified with a scale of measurement.

(2) The independent variable (or classification method) is located on the x-axis and the dependent variable (or frequency) is located on the y-axis. A change in "y" is plotted with respect to a change in "x."




b. General Principals for Constructing Graphs

(1) Keep graphs simple, as they are more effective.

(2) Tables should be self-explanatory.

(a) Use as few coordinate lines and symbols to avoid eye clutter.

(b) Use only the number of coordinate lines necessary to guide the eye.

(c) Outside graph lines should be darker than in-graph coordinate lines.

(d) Place titles at either the top or bottom of the graph.

(3) The measurement scale and its divisions should be clearly identified and marked.

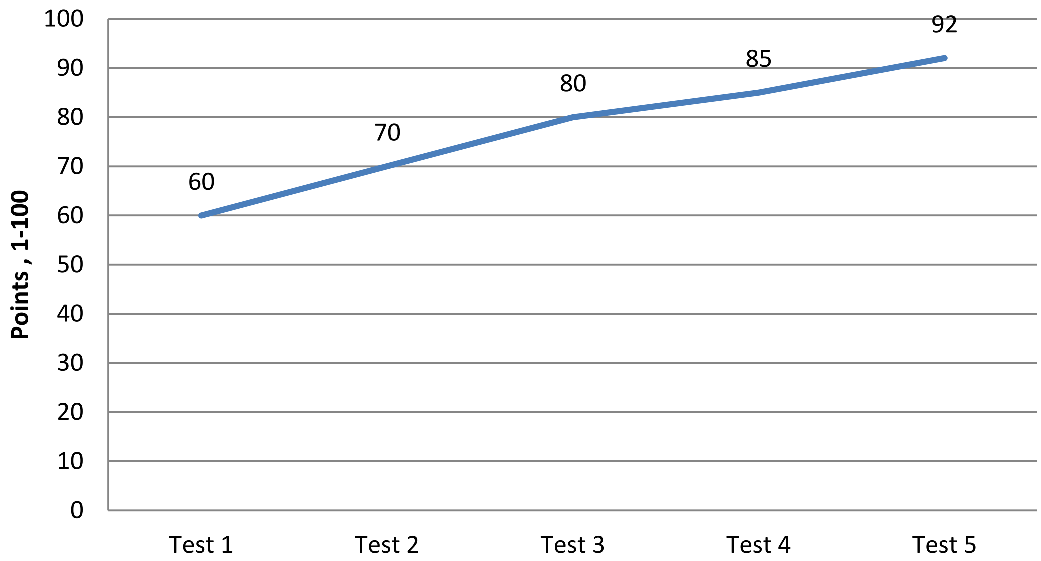
(4) On an arithmetic scale, equal measurement scale increments must represent equal numerical units.

(5) Frequency data are on the vertical or y-axis and classification category is on the horizontal or x-axis.

(6) While it is recommended that one variable be shown on a graph, there are times when two or more a warranted. In such instances, each variable must be separated with the separation devices reported in a legend or key.

2. Arithmetic Scale Line (or Line) Graph
   a. On an arithmetic scale line graph, equal distance is represented as an equal quantity on either axes, but not necessarily between the axes. Intervals on each axis must be constructed to ensure understanding. Remember, we interpret by plotting the x-axis value relative to the y-axis value to note how a change in "x" results in a change in "y".
   b. An example is presented in Figure 2.1 which was constructing using the Chart function in Microsoft Word 2010.

**Figure 2.1 Billy's History Test Results**



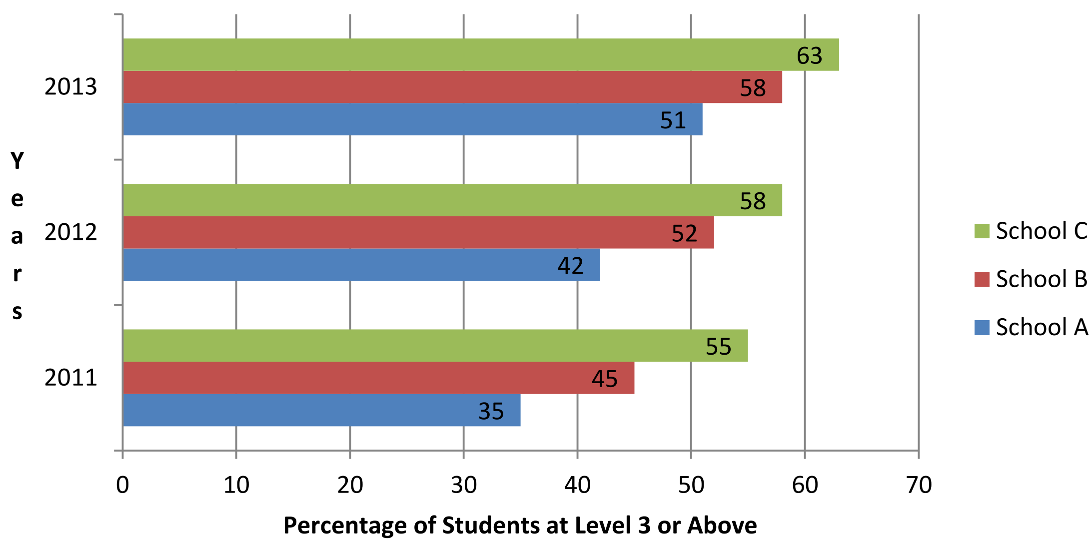

*Note*. Billy took five (5) history tests over the course of the fall semester. His test scores showed steady improvement. The line graph is particularly well suited to show data trends over time.

3. Bar & Column Charts
   a. <u>Bar Charts</u>
      (1) On the horizontal axis (x-axis) of the bar chart are the data values. On the vertical axis (y-axis), are the frequencies or bars representing the frequency for each data value or category.
      (2) The bar chart and its variations are used with nominal or ordinal data. Computer spreadsheet and statistical programs will quickly and effectively produce these images.
      (3) Feeder School FCAT Level 3+ Math Bar Chart (Figure 2.2)
         (a) Figure 2.2 contains coordinate lines with no background fill. Most spreadsheet and statistical programs with graphing capability have several display options.

(b) Select from among these carefully, using readability and ease of understanding as your guiding criteria.

b. <u>Column Chart</u>
   a. Bars maybe placed on either the vertical or horizontal axes. Microsoft Word calls charts with bars on the horizontal or x-axis column Charts.
      (1) Columns (i.e., the bars) should be colored, shaded, or differentiated in some consistent manner for the reader's understanding. However, all columns belonging to the same data category must be colored, shaded, or differentiated in the same manner from other data associated with another category.
      (2) Space between comparison groups, i.e., nominal or ordinal categories is required.
   b. An example of a column chart is presented in Figure 2.3 Feeder School FCAT Level 3+ Math.

**Figure 2.2 Feeder School FCAT Level Math 3+**



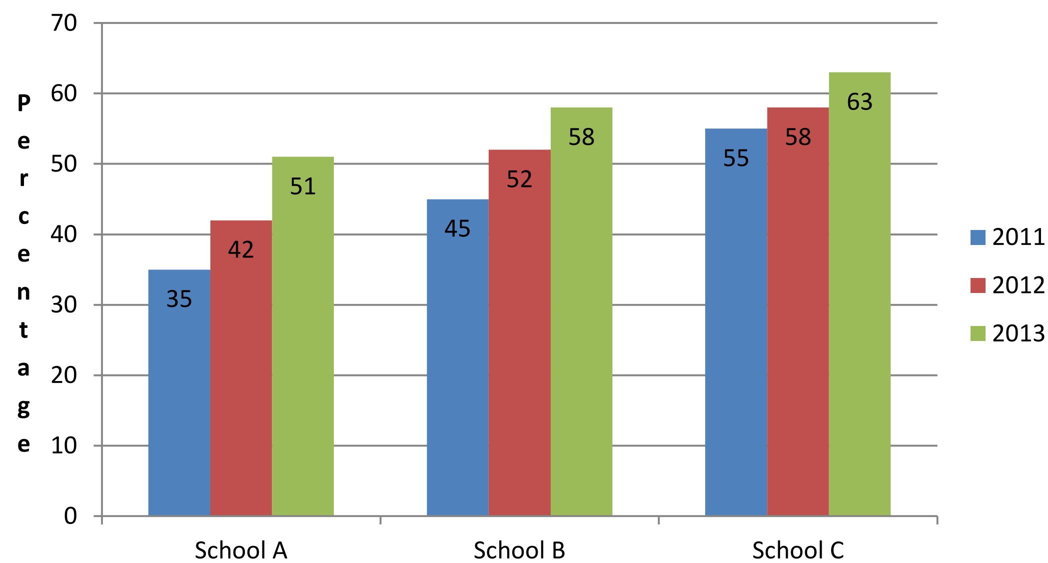

*Note*. The state of Florida administers the Florida Comprehensive Achievement Test (FCAT) which is required for high school graduation. There are five (5) achievement levels, with "1" being lowest and "5" highest. Three "3" is considered minimally proficient. This high school has three "feeder" middle (grades 6-8) schools. Middle school "C" has more students at "proficient" or above in math than either schools "A" or "B."

4. The Pie Chart
   a. The pie chart is used with nominal data. Computer spreadsheets and statistical programs will quickly and effectively produce these images.
   b. An example is presented in Figure 2.4 FCAT Performance Level.

5. Scatter Diagram
   a. Scatter plots are useful for showing the relationship or association between two variables. See also Spatz (2011, p. 101).

b. If the plots (x, y) slope upward and to the right, then a positive relationship is identified. This means that as one variable increases so does the other. See Figure 2.5a.
c. If the plots (x, y) slop downward, then the two variables have an inverse relationship, i.e., as one variable increases, the other decreases. See Figure 2.5b.
d. A graph where the plots are "all over the place" with no discernible order is said to identify no relationship. See Figure 2.5c.
e. If the plots form a curve, then a curvilinear relationship exists.

**Figure 2.3 Feeder School FCAT Math 3+**

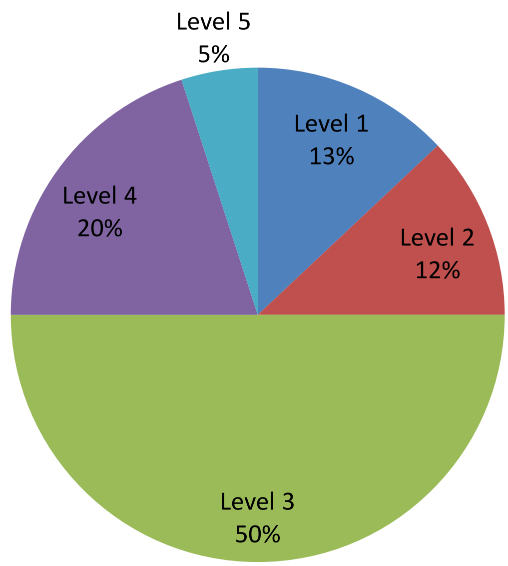



*Note*. The state of Florida administers the Florida Comprehensive Achievement Test (FCAT) which is required for high school graduation. There are five (5) achievement levels, with "1" being lowest and "5" highest. Three "3" is considered minimally proficient. This high school has three "feeder" middle (grades 6-8) schools. Middle school "C" has more students at "proficient" or above in math than either schools "A" or "B."

6. <u>Design, Use, & Interpretation Tips</u>
   a. <u>Figure Design Tips</u>
      (1) If using a black and white printer, don't sort categories by solid color. Cross-hatching or dots are useful for columns; see the pie chart above. For lines, use continuous marks, dashes, dots, symbols, etc. See the line graph above.
      (2) If data are to be compared across time, sort first by the increment of time and then by data category.
      (3) For ease of reading, bars should be arranged in either ascending or descending order.
      (4) If numbers are to be inserted in a column or immediately above it, present the numbers so that the reader may clearly read from left to right.
      (5) Titles should convey the "what", "where", and "when." Other labeling data, such as that presented in the legend or key, should be clear and outside the graph or chart data field.

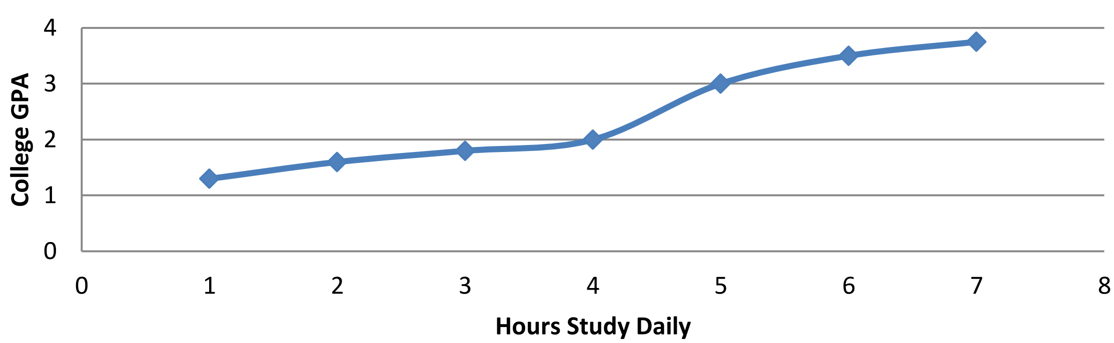
(6) Disclose sources fully. Verification is going to occur and is essential, especially if the report, conclusions, and/or recommendations are controversial.

**Figure 2.4. FCAT Performace Level**

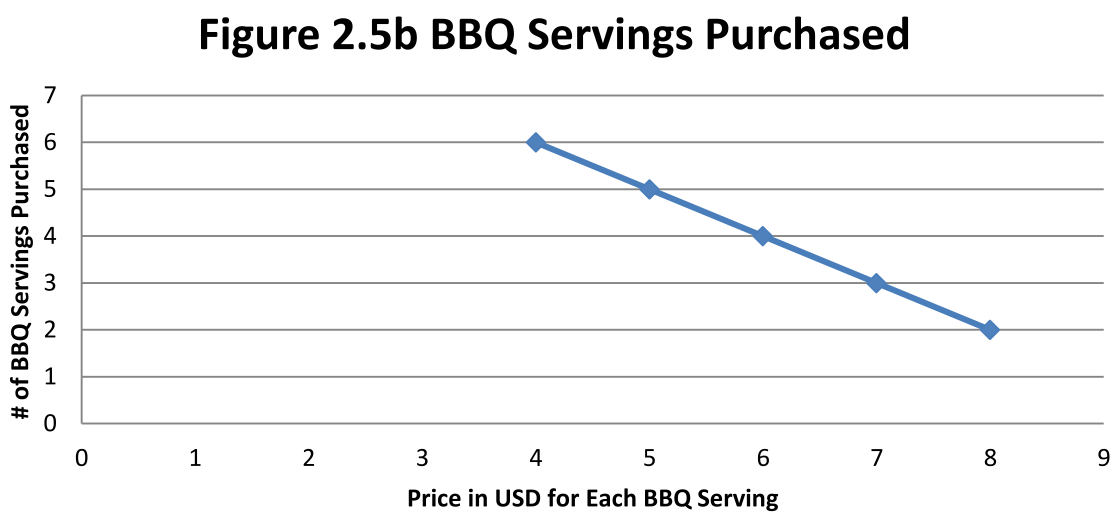



*Note*. The state of Florida administers the Florida Comprehensive Achievement Test (FCAT) which is required for high school graduation. There are five (5) achievement levels, with "1" being lowest and "5" highest. Three "3" is considered minimally proficient. We can see that 50% of examinees scored at Level 3 or minimally proficient, while 25% scored below proficient.

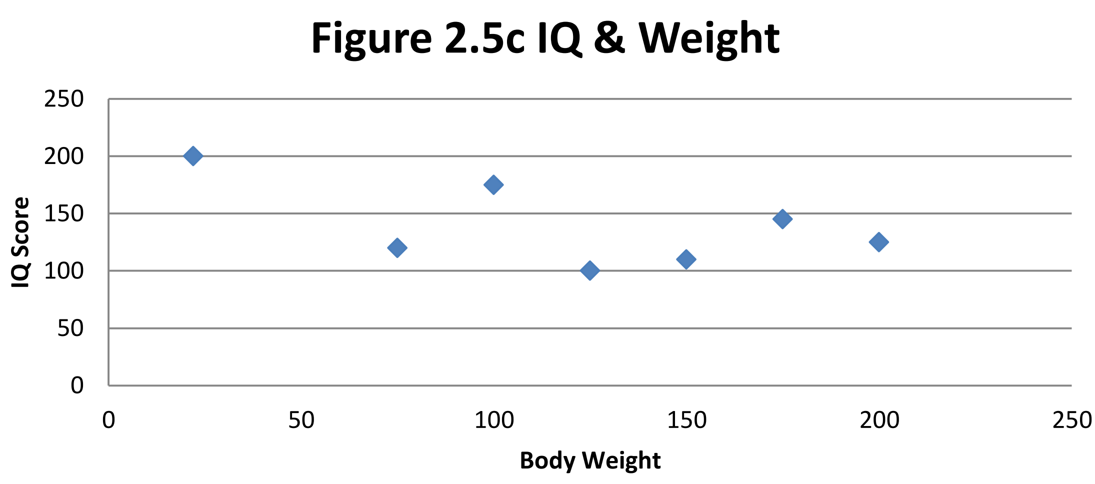
**Figure 2.5a GPA by Hours Study Daily**




*Note*. These are fictitious data; but generally the more one studies, the higher the college GPA.




*Note*. These are fictitious data; but generally the higher an item's price, the fewer in number customers buy.




*Note.* These are fictitious data; but there is no scientific correlation between bodyweight (or any other physical characteristic) and IQ scores.

b. <u>Figure Use Tips</u>
   (1) Determine the precise idea or message to be presented, next select the most appropriate presentation device.
      (a) To present a trend or compare trends, use a line graph.
      (b) To compare quantities, bar charts are most effective.
      (c) To compare a part to a whole, use a pie chart.
   (2) Present only one idea per graph or chart; include only those data which are directly and explicitly related to the idea or "point" you want to convey.

(3) Unless you are comparing data, use a different chart or graph to convey that idea. Smaller charts and graphs are more effective in conveying an idea or making a "point" than large data filled graphics.

c. Figure Interpretation Tips
(1) When drawing or proposing conclusions, consider and ensure that your conclusions reflect the full body of data.
(2) Conduct an extensive review of the available, relevant literature. This will help put your data and interpretation within a context. Your conclusions should make sense, given your data and context.
(3) Remember that tables, graphs, and charts emphasize generalities, at the expense of detail. To compensate, include footnotes as needed and comment on every single table, graph, and chart presented to ensure a full context and data presentation. Correct any possible data distortion(s) which you could reasonably expect via footnoting or commenting. Always refer to the table or figure (graph, or chart) by number and title in text **before** presenting it.

## II. Data Interpretation: Rates, Ratios, Proportions & Percentages

A. Rates

1. A rate measures the frequency of some event with reference to a specified population size. Rates are commonly employed in the health and behavioral sciences.
2. Formula 2.1: $\frac{x}{y} \bullet k$

where x = the number of times an event has occurred during a specific time interval
y = number of persons animals, or objects "exposed" to the event during the same time interval
k = a base or round number (100; 1,000; 10,000; 100,000; 1,000,000; etc.)

3. Example: The Birth Rate is the number of live births to all women of childbearing age (e.g., 15 to 35) in a given time interval. The numerator is the number of live births within the specified time interval. The denominator is the mid-year population estimate of childbearing women.

$$\frac{x}{y} \bullet k = \frac{5,000}{200,000} \bullet 1,000 = 0.025 \bullet 1,000 = 25$$

Thus, for this particular population (usually within a geographical or political jurisdiction) the crude birth rate is 25 live births per 1,000 women of childbearing age.

B. Ratios

1. A ratio expresses the relationship between a numerator and a denominator which may or may not involve a time interval.

2. Formula 2.2: $\frac{x}{y} \bullet k$

   where x = the numerator
   y = the denominator
   k = a base, usually 1 or 100

3. a. Example: You invested $100 into stock. Later you sold the stock for $300.

$$\frac{x}{y} \bullet k = \frac{\$300}{\$100} \bullet 1 = \frac{\$300}{\$100} = \$3$$

   Your return on investment was $3 to $1 for every dollar invested.

   b. Example: You are computing a <u>Cost to Benefit Ratio (BCR)</u>. Program benefit was computed at $2,500,000; program costs were computed at $750,000.

$$\frac{x}{y} \bullet k = \frac{\$2,500,000}{\$750,000} \bullet 1 = \$3.33$$

   The BCR was $3.33 in benefit to every $1.00 the program cost.

   c. Example: You are computing a <u>Return on Investment (ROI)</u> ratio.

$$\frac{x}{y} \bullet k = \frac{\$2000,000}{\$1,500,000} \bullet 1 = 1.33$$

   where: x = Total Benefits - Program Costs; y = Program Costs

   Total calculated program benefit was $3,500,000; total calculated program costs were $1,500,000. Thus, the ROI is $1.33 or the return on investment was $1.33 for every dollar invested.

C. <u>Proportions and Percentages</u>
1. A proportion is an expression where the numerator is always contained in the denominator and when summed, always equals 1.0
2. Formula 2.3: $\frac{x}{y}$

   where x = the number of times an event has occurred during a specific time interval
   y = number of persons animals, or objects “exposed” to the event during the same time interval

3. For an example, see the relative frequency table presented above (Table 2.2).

4. To convert a proportion to a percentage, multiply it by 100. To convert a percentage to a proportion, divide it by 100.

## III. Data Interpretation: Measures of Central Tendency (MCT), Skewness, and Dispersion

A. Measures of Central tendency (See also Spatz, 2011, pp. 41-50).

1. **Mode ($M_o$):** The mode is the data value in a distribution which occurred the most often. If there are two or more values that occur with the same frequency, then the distribution is said to be bi-modal, tri-modal, etc. The mode is the only MCT used with nominal data.

2. **Median ($\tilde{x}$):** The median of a data set is the mid-point or middle value when the scores are arranged in order of increasing (or decreasing) magnitude so that 50% of the cases fall above and 50% below. The median is a good choice as a representative MCT if there are extreme scores or data values. The median is often denoted by 'x-tilde'. The median can be used with ordinal, interval, and ratio data. To determine the median:
   a. First rank order all of the data values.
   b. If the number of values is odd, then the median is the exact middle value of the list of data values.
   c. If the number of data values is even, the median is the mean between the two middle numbers.
   d. For example
   Odd: 4, 5, 6, 7, 8 (6 is the median)
   Even: 10, 12, 14, 15, 16, 18 (14.5 is the median)

3. **Mean ($\bar{x}$):** The arithmetic mean of a data set is that value obtained by adding the scores and dividing the total by the number of data points.
   a. The mean is the most frequently used MCT; the mean takes into account every score in the distribution.
   b. The mean is affected by extreme scores; in this case, the median may be the most representative score or value for the dataset.
   c. The mean is used only with interval and ratio data.
   d. The formula for computing the mean is presented in Formula 2.4. A training class of 10 trainees completed a posttest where the average score was 91 points out of 100 possible.

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

$$\bar{X} = \frac{910}{10} = 91$$

where: $\bar{X}$ = Mean
$\sum$ = Sum of
x = sum of all the distribution numbers
n= # of numbers in the distribution

4. Measures of Central Tendency: Interpretation.
   a. When the mean, median, and mode are equal or nearly so, the shape of the score plot or distribution is that of the standard normal curve (SNC), Figure 2.8. When the mean, median, and mode are quite different from one another, the shape of the distribution becomes skewed, either positively or negatively.
   b. The mode is used with nominal data; the median is used with interval, or ratio data, where the mean may be distorted due to the influence of extreme outlier data (e.g., test scores); and the mean is the preferred measure of central tendency, but is unstable (i.e. variable), when the data set includes extreme or outlier scores or is small.

5. <u>Skewness</u>
   a. A data distribution is skewed, (i.e., not symmetric) if its "tail" extends to one side (left or right) more than the other.
      (1) The primary indication of a skewed data distribution is that the mean, median, and mode are substantially different from one another.
      (2) When the mean, median, and mode are identical (rarely occurs) or similar, the shape of the distribution approximates the standard normal curve (Figure 2.8). Some refer to this as non-skewed data. Achievement test data (the most common testing strategy in education and training) rarely ever "fall" on the normal curve (i.e., Figure 2.8). Figures 2.6a and 2.7a are more frequent.
      (3) To determine skewness, interval or ratio level data are required. Nominal and ordinal data are never skewed, as a mean cannot be computed from nominal or ordinal data.

   b. Positive Skewness
      (1) In a <u>positively skewed</u> dataset or distribution (Figure 2.6a), the tail goes to the right of the distribution. The mean ( $\bar{X}$ ) and median ( $\tilde{X}$ ) are to the right of the mode ( $\bar{X}$ and $\tilde{X}$ are > Mode) on the "x" axis or the mean is greater than the median ( $\bar{X} > \tilde{X}$ or $\tilde{X} < \bar{X}$ ). Remember, scores or other data values increase on the x-axis going left to right. In Figure 2.6b, the mode is 75, the median is 77.5, and the mean is 81.4. The median and mean are to the right of the mode on the x-axis; also the mean is larger than the median. We compare only the mean and median when there is multi-modal (e.g., bimodal or tri-modal, etc. data). Skewness requires interval or ratio level data; don't try to determine skew for nominal or ordinal data.
      (2) A positive skew can suggest minimal instructional effectiveness as there are more "lower" test scores than "higher" test scores for example on the x-axis. However, a determination of instructional ineffectiveness should be based on all MCT's and relevant additional interpretative criteria (e.g. cut or passing score) or context (e.g., professional judgment).

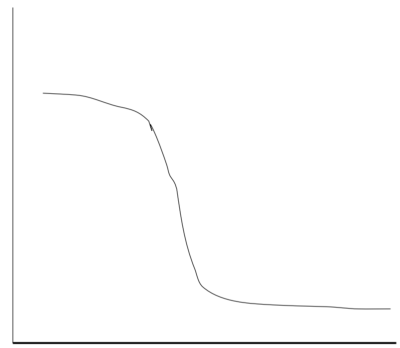
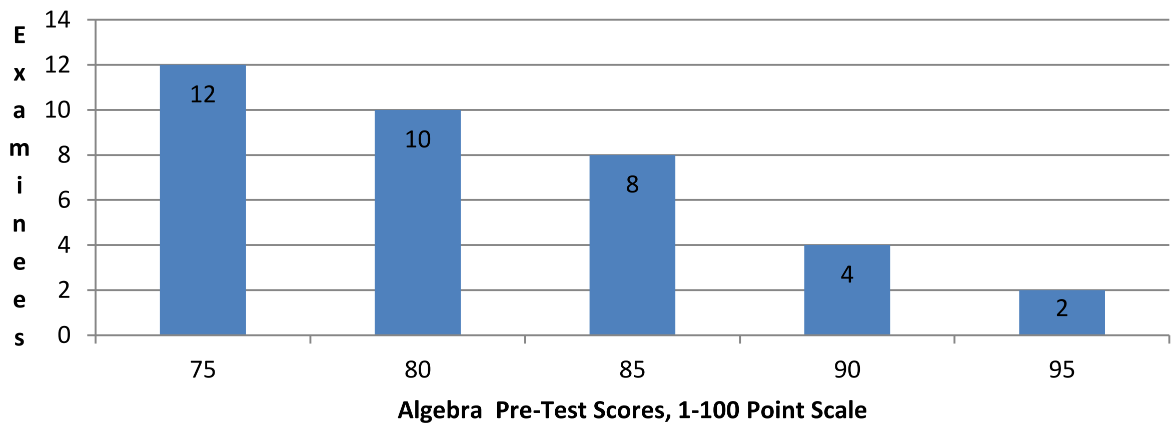


Figure 2.6a Positively Skewed Distributions

Figure 2.6b Positive Skew




*Note*. We see in Figure 2.6b a positive skew resulting from the algebra pre-test. Twenty-two (22) students scored either "75" or "80." Six (6) students scored "90" or higher. A positive skew along with the measures of central tendency, standard deviation and/or a "cut" or passing score may suggest less learning occurred than desired. If a "cut" score or proficiency level was set at "85," then instruction is clearly needed, based on these data.

c. Negative Skewness

(1) In a <u>negatively skewed</u> dataset or distribution (Figure 2.7a), the tail goes to the left of the distribution. The mean ( $\bar{X}$ ) and median ( $\tilde{X}$ ) are to the left of the mode ( $\bar{X}$ and $\tilde{X}$ are < Mode) on the "x" axis or the median is greater than the mean ( $\tilde{X} > \bar{X}$ or $\bar{X} < \tilde{X}$ ). Remember, scores or other data values increase on the x-axis going left to right. Examine Figure 2.7b; the mean (88.6) and median (90) are to the left of the mode (95) on the x-axis. The median (90) is greater than the mean (88.6). We compare only the mean and median when there is multi-modal (e.g., bimodal or tri-modal, etc. data). Skewness requires interval or ratio level data; don't try to determine skew for nominal or ordinal data.

(2) A negatively skewed distribution suggests instructional effectiveness, as there are more “higher” scores than “lower” scores on the x-axis. However, a determination of instructional effectiveness should be based on all MCT’s and relevant additional interpretative criteria (e.g. cut or passing score) or context (e.g., professional judgment).

(3) Instructional effectiveness conclusions are never drawn based only on a negative skew; the MCTs and a “cut score” are also usually included. A “cut or passing score” is a mastery level determination, e.g., 80 out of 100 possible points which shows that examinees have mastered the content and/or skills in the curriculum to an acceptable performance level.

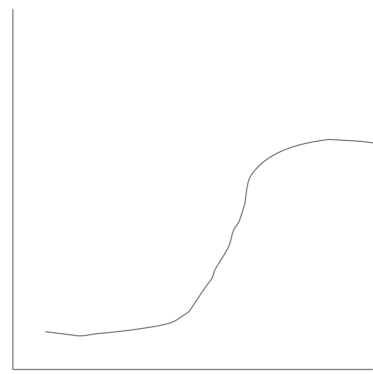


Figure 2.7a Negatively Skewed Distributions

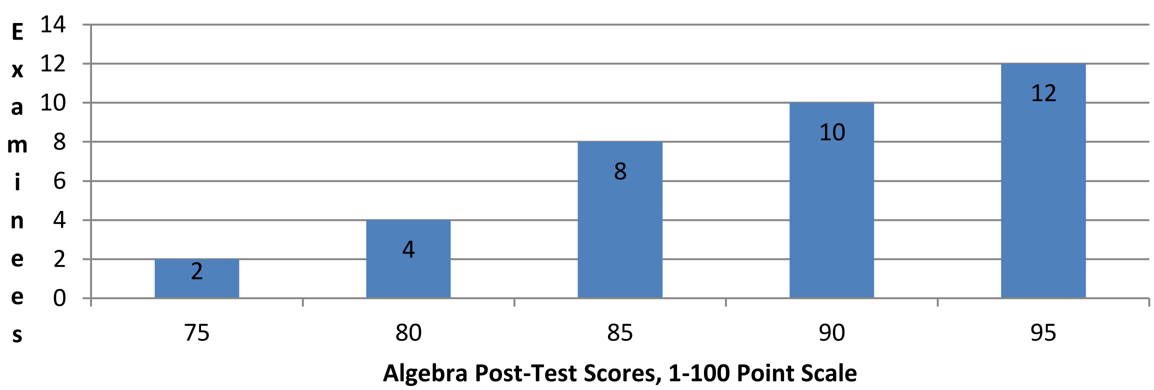



**Figure 2.7b Negative Skew**

*Note.* The algebra post-test indicates that 30 of the 36 students scored above the “cut” or passing score of “85.” A negative skew along with supportive measures of central tendency and with so many meeting or exceeding the “cut” or passing score, we can infer that a desired level of learning occurred.

B. The Standard Normal Curve (See also Spatz, 2011, pp. 34-36.)
   1. The standard normal distribution (SNC) (Figure 2.8) is a normal probability distribution with a mean (average) of zero "0," a standard deviation (σ) of one "1," and is symmetric (i.e., its left half is a mirror image of its right half).
   2. The SNC is a continuous (i.e., the left and right tails go on to infinity or forever) distribution with a bell shape. There are 1000's of bell shaped distributions.
   3. A change in the dataset or distribution's mean (μ) causes the curve to shift to the right or left on whatever measurement scale is being examined.
   4. A change in the dataset or distribution's standard deviation (σ) causes the shape to become more or less peaked, but the basic bell remains.

C. Measures of Variation (or Dispersion) (See also Spatz, 2011, pp. 54-68.)
   1. The **Range** the range is the most primitive of the measures of variation. It is simply the highest data value – the lowest data value. For example, 100 – 57 yields a range of 43. The bigger the range; the more variable (or different from each other) are the individual data values within a data distribution (e.g., a group of test scores).

   2. The **standard deviation** (s or δ) is a summary indices of the degree of variation of individual data values (e.g., test scores) around a distribution's grand mean (N).
      a. Suppose, we have a set of test scores where the mean is 87 and the median is 91; the data distribution is negatively skewed. The grand mean or "N" is 87; we know there are scores above and below the mean.

      b. Since the measurement scale is "points," we know that a score of 77 points is 10 points below the mean (87), whereas a score of 92 is five (5) points above the mean. If there are only a few scores in the distribution, keeping track of the individual score distances from the mean is no problem; however, if there are 20, 30 100, or 1000+ individual scores, a more convenient indicator of individual score distance from the grand or distribution mean is needed; hence, the standard deviation.

      c. There are typically three ways to interpret the standard deviation.
         (1) Method 1. The first method interprets the "δ or s" within the context of the measures of central tendency from the same data set (e.g., a group of classroom achievement test scores). Suppose, for the set of 33 scores: $\bar{X} = 89$, $X = 72$, and $\delta = 11$, and the Mode (Mo) is 71
            (a) We know the distribution is positively skewed (Figure 2.6) as the mean is greater than the Median; there are more scores in the lower end of the distribution than the higher end . It is also likely that there are a few extreme scores (outliers) in the distribution because the mean is much larger than the mode or median.
            (b) We can also "generally," say each individual score is 11 points away from the mean (89). But since the dataset is positively skewed, we know there are many lower scores in the distribution and a few extreme high scores, which pull the mean away from the median.

(c) The teacher, trainer, or instructional designer should investigate why the distribution is positively skewed and whether or not the mean is "inflated" by outlier scores. If the mean is inflated, the outlier scores can be removed and a more "representative" mean computed; we would probably find that the mean is actually closer to the median. If so, a review of the education program's instructional design and teaching strategies may be in order. Students or trainees may need remediation (i.e., re-teaching).

(d) This strategy is used with data from a single group, where there is no comparison or control group.

(2) Method 2. The second interpretative method is to compare "δ or s" from one group with "δ or s" from another group to determine which group had the greater or lesser variation, which may be "good" or "bad" depending on the circumstances. The farther δ is from zero, the more dissimilar or unalike the individual scores are. In a testing situation, larger standard deviations indicate some or many examinees scored lower or higher, compared to the majority of those taking the test.

(a) Suppose, we have two (2) groups with identical means $\bar{X} = 72$, but different standard deviations of $\delta = 3$ & $\delta = 6$.

(b) If we are comparing the two groups, $\delta = 3$ & $\delta = 6$, we can see one standard deviation is larger than the other, indicating that examinee scores in the second group were "more different" than the first group. This might suggest differences in instructional effectiveness. When standard deviations of comparison groups are very different, when they shouldn't be, investigation as to cause is needed.

(c) Method 2 is used when there is a control or comparison group or pretest and posttest. Method 1 is also used.

(3) Method 3. The third standard deviation interpretation strategy is the "Rule of Thumb" or the "Empirical Rule." This interpretation approach requires that the data be normally or approximately distributed (i.e., $\bar{X} = \tilde{X} = M_O$).

(a) This strategy is used mostly for interpreting standardized test results as the "norm" group is very large and when plotted, the scores pattern after the Standard Normal Curve (Figure 2.8). Method 3 is not applied to individual classroom achievement test data.

(b) When using the Empirical Rule, we say that

[1] 68% of cases or data values fall within ±1 σ of the mean.

[2] 95% of cases or data values fall within ±2 σ the mean.

[3] 99.7% of cases or data values fall within ± 3 σ of the mean.

3. **Variance** ($s^2$ or $\sigma^2$) is the squared standard deviation.

a. It is of marginal use to us here but it is very important as variance can be partitioned into its component parts.

b. This is important as it allows for a determination of degree of contribution by a variable or variables to the variance associated with a dependent variable.

## IV. Data Interpretation: Correlation

### A. Nature of the Correlation (See also Spatz, 2011, 89-96, 100-101.)

1. A correlation exists between two variables when one is related to the other. The more variance shared between the variables, the greater the association.
2. As the circles merge (See Figure 2.9) to increasingly overlap or become one, the more variance they share: hence, the higher is the association between the variables represented by the circles.
3. The portion of the two circles which overlap symbolize the degree of commonality or shared variance between the two circles which represent Variable "x" and Variable "y." We say that a specified percent or proportion of Variable "x" is explained, or accounted for, or predicted by Variable "y."
4. The nature and strength of an association or correlation between two variables can be inspected, using a scatter plot (Figure 2.5c) or a statistic, called a correlation coefficient.

### B. Inspecting the Association

1. Scatter Plots are used to visually inspect a suspected association between two variables.
   a. There are four general scatter plots: no relationship, Figure 2.10a; perfect positive relationship, Figure 2.10b; perfect negative relationship, Figure 2.10c; or curvilinear, Figure 2.10d.
   b. While less precise than a statistical correlation procedure, the scatter plot can be instructive. For example, a visual inspection reveals that in Figure 2.10a, the plots between the two variables are "all over the place," indicating no relationship. In this case, the application of a correlation procedure is most likely not necessary.

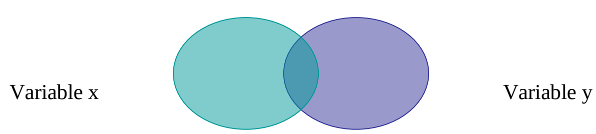



Figure 2.9 Association between Variables

2. The correlation coefficient, *r*, quantifies the strength and nature (positive or negative) of an association.
   a. *r* ranges between –1 and +1 or stated another way $-1 \leq r \leq 1$.
   b. The value of *r* does not change if all of the values of either variable are converted to a different measurement scale.
   c. The value of r is not affected by the choice of x or y for a particular data set.
   d. *r* measures the strength of a linear relationship.
   e. Correlation does not mean causation. Causality is logical not statistical.
   f. Correlation coefficients should be tested for statistical significance. It is possible for an r = .06 to be significant, but that would most likely be due to a very large sample. Ignore such findings as they have no practical significance.

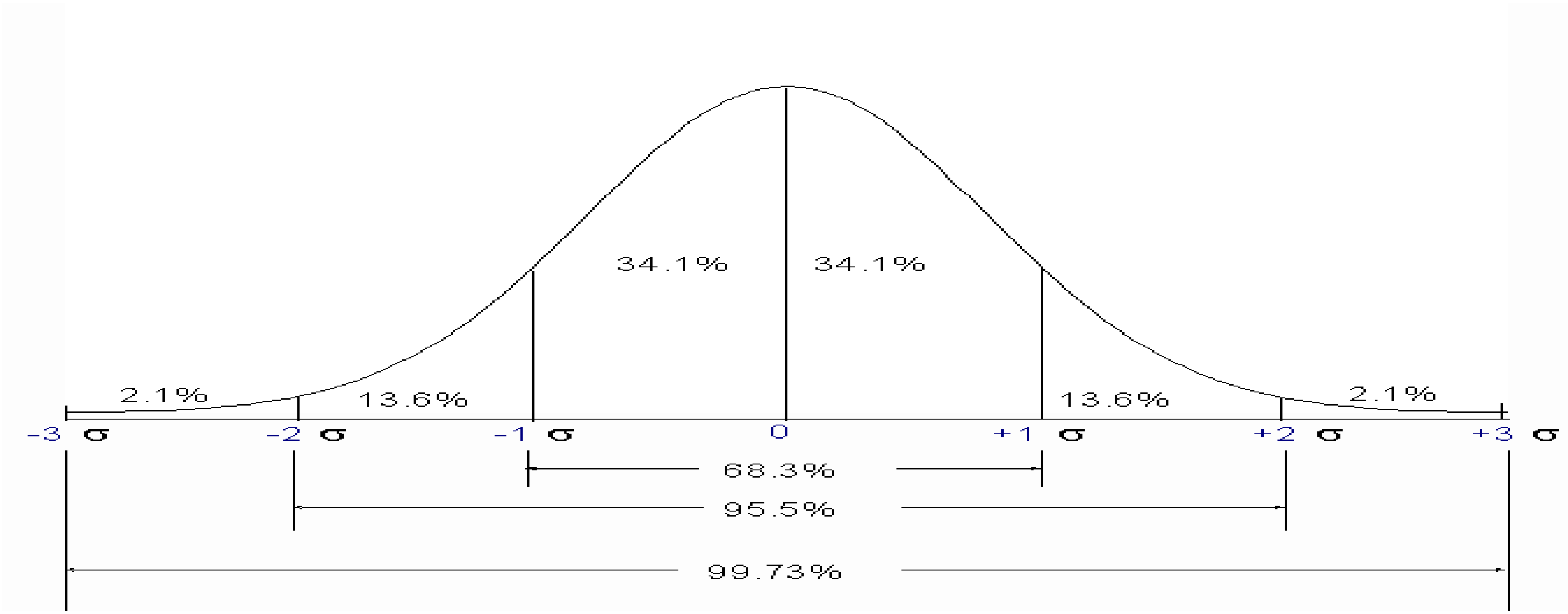



Figure 2.8 Standard Normal Curve (SNC)

3. Coefficient of Determination ($R^2$)
   a. The Coefficient of Determination is the correlation coefficient squared, *r* = .7 thus, .49 or 49% of the variance is shared between Variables "x" and "y."
   b. The coefficient of determination represents the degree of shared variance between the two variables. The higher the squared value, the more the two variables have in common and the more accurate are predictions from one variable to the other.

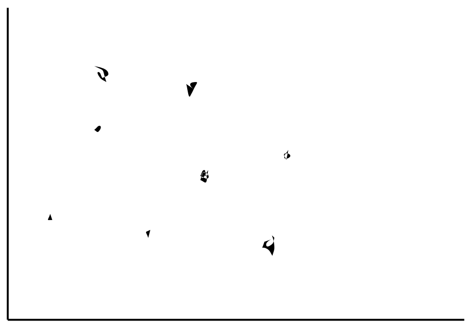


Figure 2.10a No Relationship

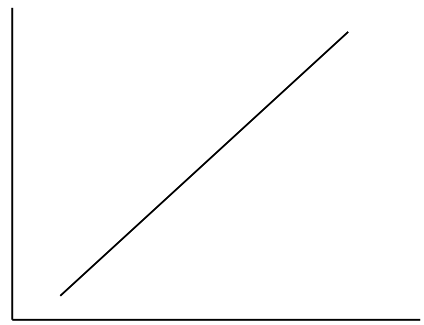


Figure 2.10b Positive Relationship

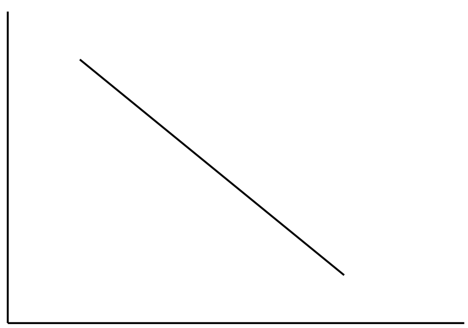


Figure 2.10c Negative Relationship

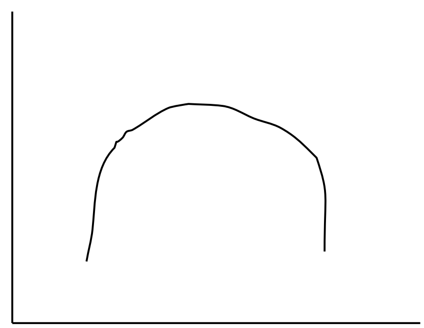


Figure 2.10d Curvilinear Relationship

4. Correlation is important in computing reliability (*r*) coefficients (Chapter 3). A reliability coefficient where *r* = 0, means the test or measuring tool (e.g., attitude survey) is unreliable. When *r* = 1.0, the test or measure is perfectly reliable. For classroom or training session tests, we like to see *r* = 0.70 to 0.80 or higher.

## Review Questions & Application Exercises

Directions. Read each item carefully. There is one correct answer per item.

1. The definition, "a numerical measurement describing some characteristic of a population," defines:
   a. Statistic
   b. Sample
   c. Parameter
   d. Census

2. The definition, "consists of numbers representing counts or measurements," defines
   a. Quantitative data
   b. Continuous data
   c. Discrete data
   d. Nominal data

3. This level of data can be ordered and has discernible distances between points.
   a. Ordinal data
   b. Nominal data
   c. Interval data
   d. Ratio data

4. The definition, "data that may be arranged in some order but differences between data values cannot be determined or are meaningless," defines
   a. Nominal data
   b. Ordinal data
   c. Interval data
   d. Ratio data

5. To make the statement that a score of 30 is twice that of 15 assumes at least which type of data?
   a. Nominal data
   b. Ordinal data
   c. Interval data
   d. Ratio data

6. Which one of the following is the most stable MCT?
   a. Mean
   b. Median
   c. Mode
   d. Midrange

7. Which <u>one</u> of the following statements concerning MCT's is <u>inaccurate</u>?
   a. Data tend to be approximately symmetric when all MCT's are or are approximately equal.
   b. Multi-modal data occur when 2 or more values are the most frequent.
   c. The median is appropriate for nominal data.
   d. If data are asymmetric, it is best to report the mean and median.

8. Which <u>one</u> of the following statements concerning skewness is <u>inaccurate</u>?
   a. Data skewed to the left are negatively skewed.
   b. Data skewed to the right are positively skewed.
   c. Data where the mean, median, or mode is similar is said to be asymmetric.
   d. Distributions skewed to the left are more common in staff training programs.

9. Which one of the following MCT indices is applied only to nominal level data?
   a. Mean
   b. Median
   c. Mode
   d. Midrange

10. The term "middle score" defines which one of the following MCT?
    a. Mean
    b. Median
    c. Mode
    d. Midrange

11. Which one of the following MCT's is most affected by extreme scores?
    a. Mean
    b. Median
    c. Mode
    d. Midrange

12. The definition, "a measure of the variation of scores about the mean," defines
    a. Range
    b. Standard deviation
    c. Z-score
    d. Percentile rank

13. The definition, "...is a normal probability distribution with a mean of zero and a standard deviation of one…" defines:
    a. The standard normal curve
    b. The t-distribution
    c. T-score distribution
    d. Asymmetric distribution

14. If a test with a normal distribution has a mean of 50 and a standard deviation of 10, approximately two-thirds of the group received scores between
    a. 40 and 50
    b. 40 and 60
    c. 50 and 60
    d. 30 and 70

15. Concerning "*r*", which one of the following statements is <u>inaccurate</u>?
    a. Ranges from -1 to +1, exclusive.
    b. The closer the "*r*" value is to zero, the more accurate is the assumption of no association.
    c. The value of "*r*" is unaffected by measurement scale conversion.
    d. Measures the strength of an association.

16. Which one of the following statements concerning "*r*" is <u>incorrect</u>?
    a. When squared, "*r*" is an estimate of shared variance between two variables.
    b. When squared, "*r*" is called the coefficient of determination.
    c. A series of bi-variate plots on a grid sloping downward to the right represents positive relationship.
    d. A series of bi-variate plots on a grid sloping upward to the right represents positive relationship.

17. A correlation between two variables establishes a causal relationship.
    a. True
    b. False

18. If the Coefficient of Determination is .49, which one of the following statements is <u>not</u> accurate?
    a. Represents the degree of shared variance between two variables.
    b. $r = 0.7$
    c. $r = 0.07$
    d. Is the opposite of the Coefficient of Alienation

## Application Exercises
## Exercise A: Customer Service Staff Training Program

<u>Situation</u>: You have been asked to analyze test scores from a customer service staff development program. There were 24 junior level employees who participated in the 8 hours training course. Scores are:

65, 66, 67, 71, 72, 73, 73, 74, 76, 78, 78, 78, 81, 81, 82, 83, 84, 85, 86, 87, 91, 92, 93, & 94.

a. Construct a simple frequency table and a frequency table with class intervals.
b. Construct a histogram.
c. Descriptive Statistics are:
   Group 1: $\bar{x}$ = 79.58, $\tilde{x}$ = 79.5, $M_o$ = 78, s = 8.5, $s^2$ = 72.3, & R = 29
d. Is this distribution skewed; if so, in what direction? How do you know?
e. Describe this distribution (i.e., identify and define each MCT and MV in "c" and interpret ± two standard deviations). Assume an adequate sample size.
f. A competitor training program covers similar materials in four hours and reports the following descriptive statistics:
   Group 2: $\bar{x}$ = 83, $\tilde{x}$ = 82, $M_o$ = 86, s = 5.5, $s^2$ = 30.25, & R = 29
   Assuming costs for the two programs are similar, which one would you select and why?

### Exercise B: Vocational Training

<u>Situation</u>: You have again been asked to analyze test scores, from a high school word-processor skill improvement class. The maximum number of points range from zero to 100. There are 24 word-processors. Scores are:

68, 73, 73, 75, 76, 83, 83, 86, 89, 89, 90, 90, 91, 91, 92, 92, 93, 94, 94, 94, 96, 96, 98, and 98.

a. Construct a simple frequency table and then one with class intervals.
b. Construct a histogram.
c. Descriptive statistics computed by your research assistant are: $\bar{x}$ = 87.6, $\tilde{x}$ = 90.5, Mode = 94, δ = 3.1, $\delta^2$ = 9.61, & R = 30
d. Is this distribution skewed; if so, in what direction? How do you know?
e. Describe this distribution (i.e., identify and define **each** MCT and MV in "c" and interpret the standard deviation to two places, plus and/ minus).

### Answers: Test Items

1. c, 2. b, 3. c, 4. b, 5. d, 6. a, 7. c, 8. c, 9. c, 10. b, 11. a, 12. b, 13. a, 14. b, 15. a, 16. c, 17. b, 18. c,

### Application Exercise A

1a) Draw a frequency table with class intervals and 1b) Draw a histogram.
1c) These data are given to you.
1d) The distribution doesn't appear to be skewed as the mean, median, and mode are similar.
1e) The mean (most stable MCT) of this distribution is 79.58 with a midpoint (median) of 80 points. The most frequently occurring score (mode) is 78. The distribution appears nearly normal, as the mean, median and mode are nearly identical. The range, highest - lowest score is 29 points. The standard deviation is 8.5 points. We expect that approximately 95% of examinees scored between 62.58 and 96.58 points [17 (8.5 * 2) - 79.58 + 17 (8.5 * 2)]. Remember, 95% is two (2) standard deviations above and below the distribution's mean.
1f) We'd probably stay with the current program who's mean, median, and mode are higher than those of the competitor training program even though its four hours shorter. Also, the current training program's lower standard deviation (indicates that examinee scores were more alike) which suggests more a uniform training effect than the competitor program.

| Group 1 | Group 2 |
|---|---|
| (Current Program) | (Competitor Program |
| $\bar{x}$ = 87.6 | $\bar{x}$ = 83 |
| $\tilde{x}$ = 90.5 | $\tilde{x}$ = 82, |
| $M_o$ = 94 | $M_o$ = 86 |
| $\delta$ = 3.1 | s = 5.5 |
| $\delta^2$ = 9.61 | $s^2$ = 30.25 |
| R = 30 | R = 29 |

### Application Exercise B

2a) Draw a simple frequency table and one with class intervals and 2b) Draw a histogram.

2c) These data are given to you.

2d) Distribution appears to be negatively skewed, as the mean and median are less than the mode. A negatively skewed distribution suggests training effectiveness.

2e) The mean (most stable MCT) of this distribution is 87.6 with a midpoint (median) of 91 points. The most frequently occurring score (mode) is 94. The range, (highest - lowest score) is 8.69 points. The standard deviation is 8.5 points. We expect that approximately 95% of examinees scored between 70.22 and 104.98 points. (Of course, it is impossible for there to be 4.98 points more than 100. This happens when sample sizes are small and is an indication that the data are not normally distributed. Here, just ignore the "overage".)

### Reference

Spatz, C. (2011). *Basic statistics: Tales of distributions* (10th ed.). Belmont, CA: Wadsworth.

### Appendix 2.1 Math Symbols

| | |
|---|---|
| $f$ | frequency of |
| $\geq$ | greater than or equal to |
| $\leq$ | less than or equal to |
| $<$ | less than |
| $>$ | greater than |
| $\approx$ | approximately equal to |
| $=$ | equal to |
| $\neq$ | not equal to |
| $+$ | Add |
| $\div$ | Divide |
| $-$ | subtract |
| $\pm$ | plus or minus or add and subtract |
| $\bullet$ | multiply |