2023 Mathematics *Standards of Learning*
Understanding the Standards – Probability & Statistics

The following standards outline the content of a one-year course in Probability and Statistics. If a one-semester course is desired, the standards with a dagger (†) would apply. The purpose of the course is to present basic concepts and techniques for collecting and analyzing data, drawing conclusions, and making predictions.

Technology tools will be used to assist in teaching and learning. Graphing technologies facilitate visualizing, analyzing, and understanding algebraic and statistical behaviors and provide a powerful tool for solving and verifying solutions.

Data in Context

PS.DC.1† The student will use a statistical cycle to formulate questions, describe types of data, data sources, and constraints within the context of a problem.

Students will demonstrate the following Knowledge and Skills:

a) Define the stages of the statistical cycle and how each stage relates to the others.
b) Formulate questions and conclusions based on context.
c) Understand the type of data relevant to the question at hand (e.g., quantitative versus categorical).
d) Compare and contrast population and sample, and parameter and statistic.
e) Identify and explain constraints of the statistical approach.

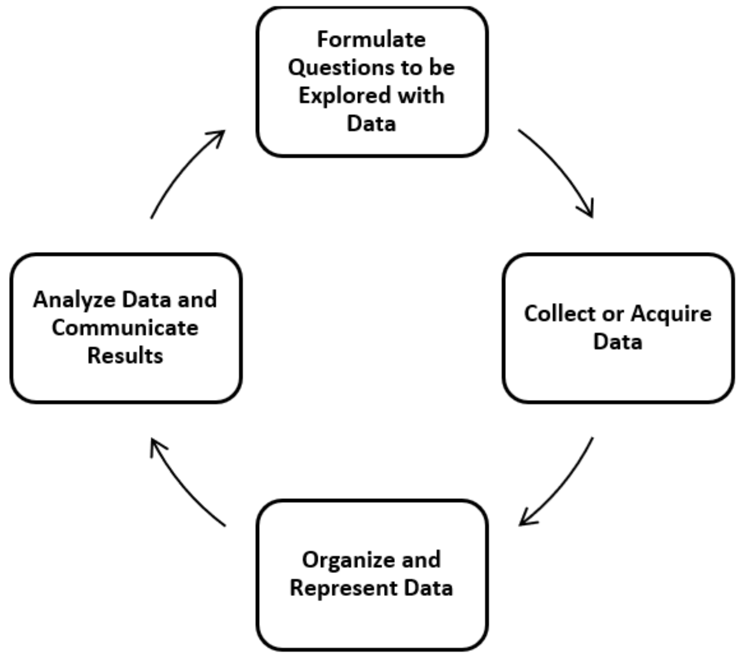
PS.DC.1† The student will use a statistical cycle to formulate questions, describe types of data, data sources, and constraints within the context of a problem. *Additional Content Background and Instructional Guidance:*

• Data are collected for a purpose and have meaning in a context.
• The iterative stages of the statistical cycle (data cycle) include:
  ○ Formulate questions to be explored with data – identify the driving question for the problem being solved.
  ○ Collect or acquire data – collect data with methods appropriate to answer the identified question/problem.
  ○ Organize and represent data – create visual representations to reveal patterns.
  ○ Analyze data and communicate results – apply appropriate statistical methods for description and inference; determine appropriate conclusions and to whom the conclusions apply; and, effectively communicate conclusions based on context and audience.

PS.DC.1† The student will use a statistical cycle to formulate questions, describe types of data, data sources, and constraints within the context of a problem.

Additional Content Background and Instructional Guidance:

Data Cycle
Grade K-Algebra 2



PS.DC.2† The student will compare and contrast data collection methods to plan and conduct an observational study.

Students will demonstrate the following Knowledge and Skills:

a) Investigate and describe sampling techniques (e.g., simple random sampling, stratified sampling, systematic sampling, cluster sampling).

b) Determine which sampling technique is best, given a particular context.

c) Investigate and explain biased influences inherent within sampling methods and various forms of response bias.

d) Use the statistical cycle to plan and conduct an observational study to answer a question or address a problem.

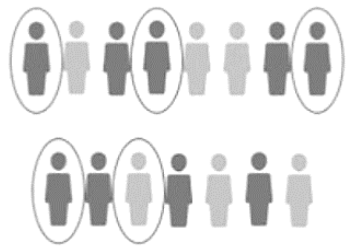
PS.DC.2† The student will compare and contrast data collection methods to plan and conduct an observational study. *Additional Content Background and Instructional Guidance:*

• A population is the entire set of individuals or items from which data is drawn for a statistical study.
• A sample is a data set obtained from a subset that represents an entire population. This data set can be used to make reasonable assumptions about the whole population.
• Sampling is the process of selecting a suitable sample, or a representative part of a population, for the purpose of determining characteristics of the whole population.
• The purpose of sampling is to provide representative information so that population characteristics may be inferred. Probability sampling means that every member of the population has an equal chance of being selected.
• An example of a population would be the entire student body at a school, whereas a sample might be only one grade level in the entire student body at a school. A sample may or may not be representative of the population as a whole. Questions to consider when determining whether an identified sample is representative of the population include:
  ○ What is the target population of the formulated question?
  ○ Who or what is the subject or context of the question?
• Sample size refers to the number of participants or observations included in a study. Statistical data may be more accurate, and outliers may be more easily identified with larger sample sizes.
• Examples of questions to consider in building good samples:
  ○ What is the context of the data to be collected?
  ○ Who is the audience?
  ○ What amount of data should be collected?
• There are four main types of probability sampling techniques:
  ○ In simple random sampling, each member of the group or population has an equal chance of being selected. For example –

PS.DC.2† The student will compare and contrast data collection methods to plan and conduct an observational study.

Additional Content Background and Instructional Guidance:

- A simple random sample of 1,000 employees at a company are selected to participate in a marketing survey. Each employee is assigned a number in the company and a random number generator selects 100 numbers.
- The collection of population items is equally likely to make up the sample, just as in a lottery.

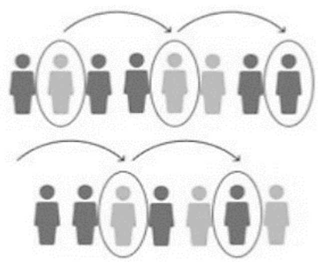


- Stratified sampling involves dividing the population into subpopulations, called strata, where the members are similar in some way. This type of sampling allows more precise conclusions by ensuring that each subgroup is properly represented within the sample. This sampling method involves dividing the population into subgroups (strata) based on the relevant characteristic (e.g., age range, income, career). The overall proportions of the population will determine how many people should be sampled from each group. For example, a company has 1000 employees as described below. They want to survey 100 of the employees about their opinions regarding benefits. Thoughts and opinions about benefits vary among males and females.
  - A company has 600 male employees and 400 female employees. To ensure that the sample reflects the gender balance of the company, the population is sorted into two strata based on gender. Then, a random sampling on each group is used with 60 male employees and 40 female employees, which gives a representative sample of 100 people. Why is it a good idea to survey both males and females?
  - A visual representation of stratified sampling is as follows –

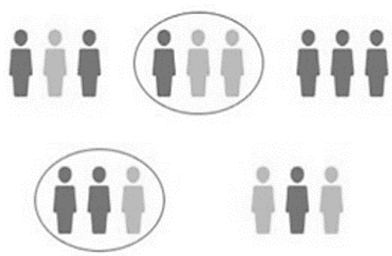
- In systematic sampling, each member of the population is listed with a number. However, instead of randomly generating numbers, individuals are chosen at regular intervals. For example –

PS.DC.2† The student will compare and contrast data collection methods to plan and conduct an observational study. *Additional Content Background and Instructional Guidance:*

- Employees of a company are listed in alphabetical order. From the first 10 numbers, a starting point is randomly selected at number 8. From number 8 onward, every tenth person on the list is selected (8, 18, 28, 38, 48, …) until a sample of 100 people is created.
- Systematic sampling is sometimes used to perform quality control checks with products off an assembly line.
- A visual representation of systematic sampling is as follows –



- Cluster sampling involves dividing the population into subgroups, or clusters. This sampling method is effective when dealing with large and dispersed population, but there is more risk of error in the sample as significant differences may exist between the clusters. An example of cluster sampling is as follows –
    - To determine the unemployment rate in a county, an agency samples households in the county and asks adults in the household how many of them are unemployed.
    - A visual representation of cluster sampling is as follows –



- Observational studies allow researchers to examine the relationships and patterns observed within a population of interest. Surveys are one type of observational study.
- A statistic is a number that describes a sample. A parameter is a number that describes a population.
- The value of a sample statistic varies from sample to sample if the samples are taken repeatedly from the population of interest.
- Bias may limit the degree to which accurate conclusions can be drawn. Errors in collecting and organizing the data may also contribute to bias. If the study overestimates or underestimates a population value, it is considered to be biased.

PS.DC.2† The student will compare and contrast data collection methods to plan and conduct an observational study.

Additional Content Background and Instructional Guidance:

- Bias may influence data analysis and may result in misleading generalizations and only support one opinion or view.
- A study is unbiased if the procedure used produces the same result as the average of the population.
- Sampling bias may occur when members of a population are systematically more likely to be chosen over other members of the same population. In other words, sampling bias does not ensure proper randomization. It is important to note that a large sample size does not make up for bias.
- Response bias is a general term that refers to conditions or factors that influence survey responses. The different types of response bias are demand bias, social desirability bias, dissent bias, acquiescence bias, extreme responses, neutral responding, and question order bias.
    - Demand bias reflects a respondent’s pressure of engaging in a study. Because of this, their behavior and opinions change based on advanced assumptions about the questions.
    - Social desirability bias reflects a respondent’s desire to answer a question in a way they believe is morally or socially acceptable.
    - Dissent bias is when a respondent answers questions negatively. They may not understand what the survey is for, or they might have difficulty understanding the questions.
    - Acquiescence bias means that all survey answers are positive.
    - Extreme response biases are seen in answer selections like “Strongly agree” or “Strongly disagree.” A respondent chooses extreme answers even if it is not their actual viewpoint.
    - Neutral responding is the opposite of extreme response bias. Respondents consistently choose an unbiased answer.
    - Question order bias reflects the order in which the questions are asked. This arrangement can negatively or positively affect the respondent's answer. The order of survey questions must be structured wisely.
- There are multiple uses of probability sampling to include reduction of sample bias, diverse populations, and the creation of an accurate sample. Inherent bias diminishes as sample sizes increases.
- Poor data collection can lead to misleading, meaningless, or biased conclusions.

PS.DC.3† The student will utilize the principles of experimental design to plan and conduct a well-designed experiment.

Students will demonstrate the following Knowledge and Skills:

a) Describe the principles of experimental design, including:
   i) treatment/control groups;
   ii) blinding/placebo effects;
   iii) experimental units/subjects; and
   iv) blocking/matched pairs and completely randomized designs.

b) Evaluate the principles of experimental design to address comparison, randomization, replication, and control within the context of the problem.

c) Compare and contrast controlled experiments and observational studies and the conclusions that may be drawn from each.

d) Use the statistical cycle to plan and conduct a well-designed experiment to answer a question or address a problem.

e) Select a data collection method appropriate for a given context.

PS.DC.3† The student will utilize the principles of experimental design to plan and conduct a well-designed experiment.

Additional Content Background and Instructional Guidance:

- A statistical experimental design includes:
  - comparison of at least two treatment groups, one of which can be a control group;
  - randomization of treatments;
  - replication of more than one experimental unit in each treatment group; and,
  - control of potential confounding variables.
    - A confounder is a variable related to both the treatment and the outcome. When a confounder exists, it is difficult to determine if differences in the outcome are due to the treatment or to the confounder.
- Experiments must be well-designed to generalize findings and detect a cause-and-effect relationship between variables.
- Principles of experimental design include comparison with a control group, randomization, and blindness.
- Almost all experiments involve comparison. randomization, replication, and control. These principles complement each other in trying to increase the accuracy of an experiment and to provide a valid test of significance while retaining the distinctive features of their roles in any experiment.
  - The group that receives the treatment in an experiment is called the experimental group, while the group that does not receive the treatment is called the control group. The control group provides a baseline to see if the treatment has an effect.
  - Blinding is a commonly used method of not telling participants which treatment a subject is receiving. Blinding is a critical part of a randomized control trial. It

PS.DC.3† The student will utilize the principles of experimental design to plan and conduct a well-designed experiment. *Additional Content Background and Instructional Guidance:*

reduces the bias that affects the results. Blinding is done to address or control for the placebo effect.
    - Before conducting an experiment, an experimental unit (plot) must be defined. A collection of experimental units is known as a block.
    - In a randomized block design, randomization is carried out within each block. In a matched pairs experiment, a block is a pair so the randomization must be carried out within each pair.
- Observations made on experimental units vary considerably. These variations are partly produced by the manipulation of certain variables of interest called treatments, built-in and manipulated deliberately in the experiment to study their influences.
    - Variations (experimental error) can be produced by many unknown sources such as uncontrolled variation in extraneous factors related to the environment, variations in the experimental material other than that due to treatments, etc. These are unavoidable and inherent in the process of experimentation and are produced by a set of unknown factors beyond the control of the researcher.
- In an observational study, we measure or survey members of a sample without trying to affect them. The researcher does not assign the treatment.
- In a controlled experiment, we assign people or things to groups and apply some treatment to one of the groups, while the other group does not receive the treatment.
- Data collection methods are important given the appropriate context as the information collected is used to explain, generate, and analyze solutions. Five key data collection methods (surveys, interviews, focus groups, observations, and content analysis) are described as follows –
    - Surveys are often used to obtain qualitative data in an open-ended, free-text format. Surveys are ideal for documenting perceptions, attitudes, beliefs, or knowledge within a clear, predetermined sample of individuals.
    - Interviews are used to gather information from individuals one-on-one, using a series of predetermined questions or a set of interest areas. Interviews are often recorded and transcribed and can be structured or unstructured. Interview data are often used to create themes, models, or theories.
    - Focus groups are used to gather information in a group setting, either through predetermined interview questions that a facilitator asks of participants or through a script to engage group conversations. Focus groups are used when the sum of a group’s experiences may provide more data than a single person’s experiences in understanding social phenomena.
    - Observations are used to gather information using the senses: vision, hearing, touch, and smell. Observations allow researchers to investigate and document what people do and try to understand why they do it. The number of observations needed will

PS.DC.3†	**The student will utilize the principles of experimental design to plan and conduct a well-designed experiment.**
Additional Content Background and Instructional Guidance:	
	depend on the given research question and the research approach used to conduct the experiment. ○ Content (textual) analysis is used to investigate changes in official views on a specific topic or area to either document the context of certain behaviors or practices or to investigate the experiences and perspectives of a group of individuals.

Descriptive Statistics

PS.DS.1† The student will represent and analyze data visualizations of univariate quantitative data, including dot plots, stemplots, boxplots, cumulative frequency graphs, and histograms, to identify and describe patterns and departures from patterns, using central tendency, spread, clusters, gaps, and outliers, within the context of a problem.

Students will demonstrate the following Knowledge and Skills:

a) Create and interpret graphical displays of data, including dot plots, stemplots, boxplots, cumulative frequency graphs, and histograms, using appropriate technology.

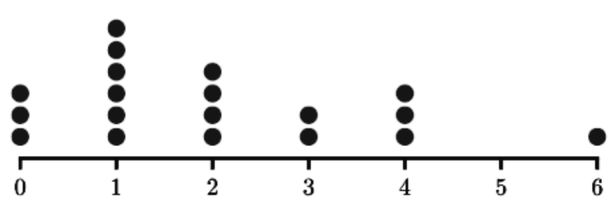
b) Examine the graphs within the context of the problem by analyzing:
   - i) shape;
   - ii) measures of center;
   - iii) spread; and
   - iv) unusual features of the data (e.g., outliers, clusters, gaps).

PS.DS.1† The student will represent and analyze data visualizations of univariate quantitative data, including dot plots, stemplots, boxplots, cumulative frequency graphs, and histograms, to identify and describe patterns and departures from patterns, using central tendency, spread, clusters, gaps, and outliers, within the context of a problem.

Additional Content Background and Instructional Guidance:

- Graphical displays (data visualization) make it easier to communicate and interpret data that may be difficult to understand. Graphical displays reveal patterns, contextualize meaning, identify outliers, and draw conclusions. There are several forms of graphical displays to represent quantitative data to include dot plots, stemplots, boxplots, cumulative frequency graphs, and histograms.
  - Dot plots use dots to represent the number of data points along a number line. The center (median) is the middle data point when all the data points are listed in order from least to greatest. The spread (range) is calculated by subtracting the minimum data value from the maximum data value. The mode (most frequent value) and data distribution can be readily observed.
    - To interpret dot plots, determine the center (median) by finding the middle data point. Then, determine the maximum and minimum values on the graph, using these data to determine the spread (range) of the data set. Determine the overall shape of the graph and note other features of interest.
    - A visual representation of a dot plot is as follows –

Number of children in a family



PS.DS.1† The student will represent and analyze data visualizations of univariate quantitative data, including dot plots, stemplots, boxplots, cumulative frequency graphs, and histograms, to identify and describe patterns and departures from patterns, using central tendency, spread, clusters, gaps, and outliers, within the context of a problem.

Additional Content Background and Instructional Guidance:

- Stemplots are graphical displays where data are split into stems (the largest digit) and leaves (the smallest digit), in order from least to greatest value. Stemplots are also referred to as stem-and-leaf plots.
  - To interpret a stemplot, the numbers are arranged by place value. The largest place-value digits are placed in the stem. Stemplots are useful in that all data points and an overview of the distribution are seen. Stemplots are useful for highlighting the mode and for finding outliers of the data set.
  - A visual representation of a stemplot is as follows –

Age of customers

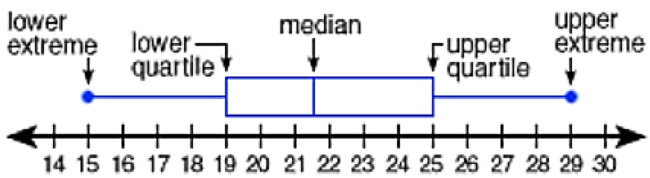
3	3 4 4 5 5 7 8
4	2 2 5
5	0 1 8
6	8
7	2

Key: 1|0 = 10

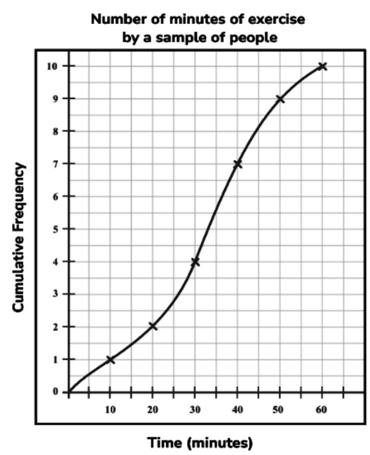
- Boxplots (box-and-whisker plots) visually show the distribution of numerical data and skewness by displaying the data quartiles (percentiles) and averages. Dispersion (also called variability, scatter, or spread) is the extent to which a distribution is stretched or squeezed. A boxplot uses a rectangle to represent the middle half of a set of data and lines (whiskers) at both ends to represent the remainder of the data. The median is marked by a vertical line inside the rectangle.
- The five critical points in a boxplot, commonly referred to as the five-number summary, are lower extreme (minimum), lower quartile, median, upper quartile, and upper extreme (maximum). Each of these points represents the bounds for the four quartiles.
  - To interpret a boxplot, data are divided into sections containing approximately 25% of the data in that set. Boxplots can be presented horizontally or vertically.
  - Using shorthand notation, the lower quartile is represented by Q1, the median is represented by Q2, and the upper quartile is represented by Q3.
  - In the example below, the lower extreme is 15, the lower quartile is 19, the median is 21.5, the upper quartile is 25, and the upper extreme is 29.

PS.DS.1† The student will represent and analyze data visualizations of univariate quantitative data, including dot plots, stemplots, boxplots, cumulative frequency graphs, and histograms, to identify and describe patterns and departures from patterns, using central tendency, spread, clusters, gaps, and outliers, within the context of a problem.

Additional Content Background and Instructional Guidance:



- Cumulative frequency graphs (cumulative frequency diagrams) are useful when representing or analyzing the distribution of large, grouped data sets. These graphs can be used to find estimates for the median value, the lower quartile, and upper quartile for the data set.
  - To interpret a cumulative frequency graph –
    - When the cumulative curve is concave up, data are increasing.
    - When the cumulative curve is linear, data are unchanged.
    - When the cumulative curve is concave down, data are decreasing.
    - When the cumulative curve is horizontal, no data are reported.
  - A visual representation of a cumulative frequency graph is as follows –



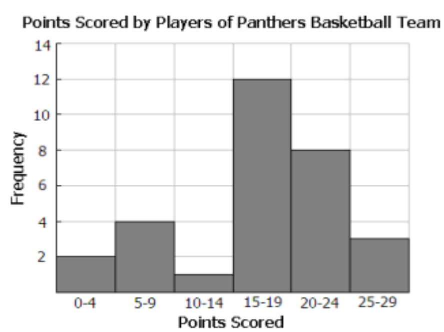
- A histogram is a graph that provides a visual interpretation of numerical data by indicating the number of data points that lie within a range of values, called a class or a bin. The frequency of the data that falls in each class or bin is depicted by the use of a bar.
  - To interpret a histogram, it is to be understood which categories are consecutive and equal intervals. The length or height of each bar is

PS.DS.1† The student will represent and analyze data visualizations of univariate quantitative data, including dot plots, stemplots, boxplots, cumulative frequency graphs, and histograms, to identify and describe patterns and departures from patterns, using central tendency, spread, clusters, gaps, and outliers, within the context of a problem.

Additional Content Background and Instructional Guidance:

determined by the number of data elements (frequency) falling into a particular interval.

- Histograms do not directly display measures of center. Instead, histograms provide an easy-to-read summary for large data sets where displaying individual data points would become cumbersome. The data in a histogram is organized so that ranges of data values can be compared easily; that is, it can be determined which ranges occurred more or less often than others, the range that occurs most often, and where the data is concentrated. Histograms show where the ranges of data are centered or if there are gaps in data.
- A visual representation of a histogram is as follows –



- Measures of central tendency describe how the data cluster or group.
- Measures of dispersion describe how the data spread (disperse) around the center of the data.
- Graphical displays of data are generated using technology to communicate quality visualizations with ease and accuracy.
- Data are collected for a purpose and have meaning in context. The shape, measures of center, spread, and unusual features (e.g., outliers, clusters, and gaps) are essential to data visualization analysis.

PS.DS.2† The student will represent and analyze numerical characteristics of univariate quantitative data sets to describe patterns and departures from patterns within the context of a problem.

Students will demonstrate the following Knowledge and Skills:

a) Interpret measures of central tendency: mean, median, and mode.
b) Interpret measures of spread: range, interquartile range, variance, and standard deviation.
c) Identify possible outliers, using an algorithm.
d) Investigate and explain the influence of outliers on a univariate data set.
e) Investigate and explain ways in which standard deviation addresses variability by examining the formula for standard deviation.

PS.DS.2† The student will represent and analyze numerical characteristics of univariate quantitative data sets to describe patterns and departures from patterns within the context of a problem. *Additional Content Background and Instructional Guidance:*

• Analysis of the descriptive statistical information generated by a univariate data set should include the interplay between central tendency and dispersion as well as among specific measures.
• Description of a univariate data set includes the holistic evaluation of central tendency, variation, shape, and departure from the data pattern.
• The presence of outliers has an impact on the data set and should be considered within the contextual understanding of the problem.
• Data points identified algorithmically as outliers should not be excluded from the data unless sufficient evidence exists to show them to be in error.
    ○ The Interquartile Range (IQR) is the difference between the upper and lower medians (Q1-Q3).
    ○ A commonly used rule says that a data point is an outlier if it is more than $1.5 \cdot IQR$ above the third quartile or below the first quartile. Said differently, low outliers are below $Q_1 - 1.5 \cdot IQR$ and high outliers are above $Q_3 - 1.5 \cdot IQR$.
• To develop an understanding of standard deviation as a measure of dispersion (spread), students should have experience analyzing the formulas for and the relationship between variance and standard deviation.
    ○ Variance (σ^2) and standard deviation (σ) measure the spread of data about the mean in a data set.
    ○ Standard deviation is expressed in the original units of measurement of the data. The greater the value of the standard deviation, the further the data tends to be dispersed from the mean.
    ○ Standard deviation is the square root of the variance and is expressed in the same units as the data set.

PS.DS.2† The student will represent and analyze numerical characteristics of univariate quantitative data sets to describe patterns and departures from patterns within the context of a problem.

Additional Content Background and Instructional Guidance:

Statistics Formulas:

Given:

x represents an element of the data set,
x_i represents the i^{th} element of the data set,
n represents the number of elements in the data set,
μ represents the mean of the data set,
σ represents the standard deviation of the data set, and
σ^2 represents the variance of the data set

standard deviation: $\sigma = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}}$

variance $(\sigma^2) = \dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$

PS.DS.3† The student will represent, compare, and analyze distributions of two or more univariate quantitative data sets, numerically and graphically.

Students will demonstrate the following Knowledge and Skills:

a) Create graphical displays of data, including back-to-back stemplots, parallel dot plots, parallel boxplots, and histograms, using appropriate technology.

b) Compare and contrast two or more univariate data sets, numerically and graphically, within the context of a problem by analyzing:

  i) shape;

  ii) measures of center;

  iii) measures of spread; and

  iv) unusual features of the data (e.g., clusters, gaps, outliers).

PS.DS.3† The student will represent, compare, and analyze distributions of two or more univariate quantitative data sets, numerically and graphically.

Additional Content Background and Instructional Guidance:

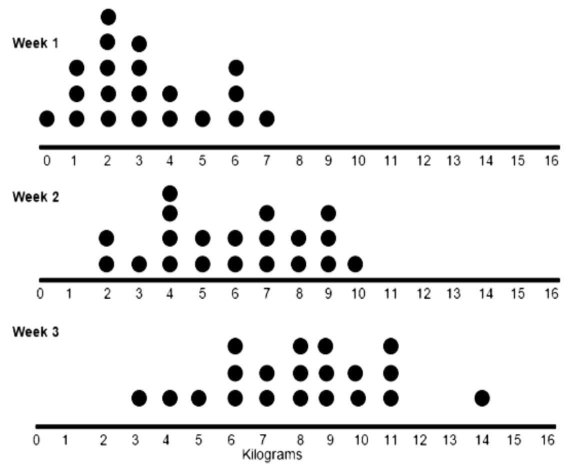
- Descriptions of two or more univariate data sets include the holistic and comparative evaluation of central tendency, variation, shape, and departures from the data patterns.
- Two or more graphical displays should utilize shared numerical scales to allow for accurate comparisons.
- Graphical displays of data are generated using technology to communicate quality visualizations with ease and accuracy.
  - A back-to-back stemplot is a method for comparing two data distributions by attaching two sets of leaves to the same stem in a stemplot. For example, a back-to-back stemplot is constructed below showing the ages of male customers and the ages of female customers:

Male		Female
5, 2, 0	1	5, 8
5, 1	2	1, 6, 9, 9
5, 5, 5, 3, 1	3	
5, 2	4	1, 2, 6, 8
9, 8, 6, 1, 1	5	5
6, 5, 5, 0	6	0, 1
2, 1, 1, 0, 0	7	2

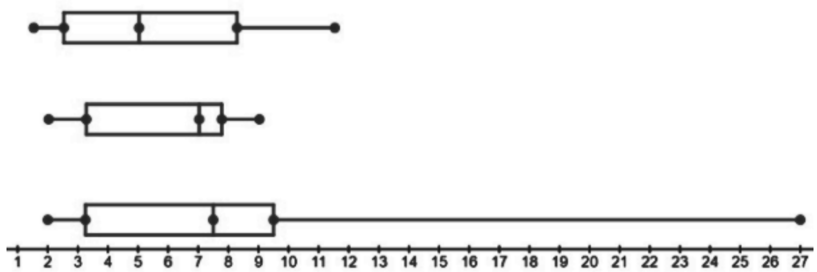
  - Parallel dot plots compare two or more sets of data. These data must be plotted against the same scale. For example, 20 people joined a weight loss class. Weights were taken each week. The overall weight was plotted each week on a parallel dot plot.

PS.DS.3† The student will represent, compare, and analyze distributions of two or more univariate quantitative data sets, numerically and graphically.

Additional Content Background and Instructional Guidance:



- Parallel boxplots are used to compare two (or more) five number summaries and is ideal for comparing multiple samples. For example, the bottom boxplot on this parallel boxplot has more variability than the other two data sets.



- The shape, measures of center, spread, and unusual features (e.g., outliers, clusters, and gaps) are essential to contextual understanding and data visualization analysis.

PS.DS.4 The student will represent and analyze categorical data, using two-way tables and other graphical displays, to describe patterns and relationships.

Students will demonstrate the following Knowledge and Skills:

a) Create and interpret graphical displays of univariate categorical data, including bar graphs within the context of the problem, using appropriate technology.

b) Create and interpret graphical displays comparing distributions of two or more univariate categorical data sets including segmented and side-by-side bar graphs within the context of the problem, using appropriate technology.

c) Generate and interpret a two-way table as a summary of the information obtained from two categorical variables.

d) Calculate and interpret marginal, relative, and conditional frequencies to analyze data in a two-way table within the context of a problem.

PS.DS.4 The student will represent and analyze categorical data, using two-way tables and other graphical displays, to describe patterns and relationships. *Additional Content Background and Instructional Guidance:*
• Graphical and tabular representations allow for the identification and representation of key features of categorical data within the context of the problem. • Summary statistics for two categorical variables allow for comparisons of distributions and may reveal any potential association. • Connections between number lines, tables, and graphs can be reintroduced to convert between one form to another, where number lines can be used to graph one-variable data and the Cartesian Coordinate Plane can be used to graph two-variable data. • Computer software and graphing calculators can be used to generate graphs.

PS.DS.5 The student will represent and analyze quantitative bivariate data with scatterplots to identify and describe the relationship between two variables.

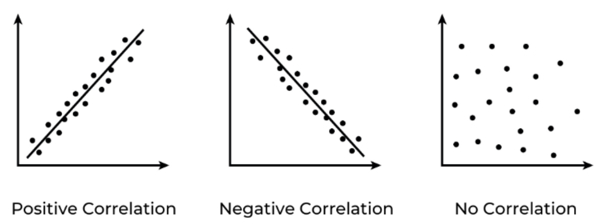
Students will demonstrate the following Knowledge and Skills:

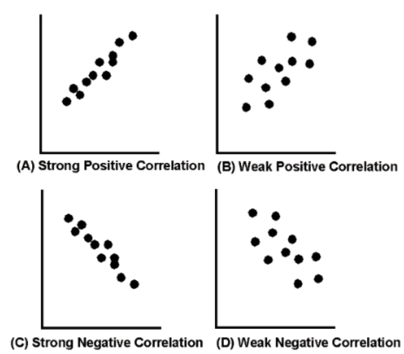
a) Create scatterplots, using appropriate technology.
b) Examine and interpret scatterplots in the context of the problem by analyzing:
   i) the form of relationship for linear and nonlinear trends;
   ii) the direction of the relationship for positive, negative, or no association;
   iii) the strength of the relationship such as strong, moderate, or weak; and
   iv) the presence of unusual features within the data (e.g., clusters, gaps, influential points, outliers).

PS.DS.5 The student will represent and analyze quantitative bivariate data with scatterplots to identify and describe the relationship between two variables.

Additional Content Background and Instructional Guidance:

- The relationship between two variables considers the direction, strength, form, and deviation from the pattern.
- A scatterplot displays a relationship between two variables to determine patterns of association (correlation) and make connections to the linear or nonlinear forms of equations.





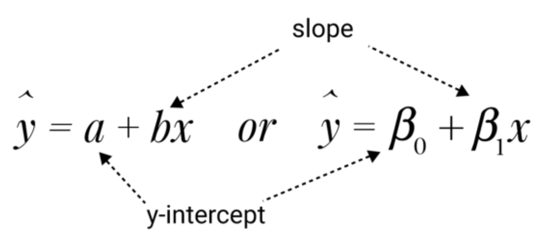
- Graphical displays of data are generated using technology to communicate quality visualizations with ease and accuracy.
- Two variables may be strongly associated without a cause-and-effect relationship existing between them.
- An outlier can be identified by sorting the data in ascending order. A data value that is an abnormal distance relative to the other values in the data set is an outlier. It represents a

PS.DS.5 The student will represent and analyze quantitative bivariate data with scatterplots to identify and describe the relationship between two variables. *Additional Content Background and Instructional Guidance:*
value that "lies outside" (is much smaller or larger than) most of the other values in a set of data. Outliers have a greater effect on the mean and range of a data set but have lesser effect on the median or mode.

PS.DS.6 The student will create and interpret a linear model using the least squares regression method to assess the relationship between two quantitative variables.

Students will demonstrate the following Knowledge and Skills:

a) Create the least squares regression model using technology to interpret the contextual meaning of the slope and *y*-intercept.
b) Using technology, calculate and interpret the correlation coefficient, *r*, within the context of a problem.
c) Using technology, calculate and interpret the coefficient of determination, r^2, within the context of a problem.
d) Use regression lines to make predictions, and identify the limitations of the predictions, such as extrapolation.
e) Calculate and interpret a residual to understand the error of a prediction.
f) Using technology, calculate and interpret the standard deviation of the residuals, *s*.

PS.DS.6 The student will create and interpret a linear model using the least squares regression method to assess the relationship between two quantitative variables. *Additional Content Background and Instructional Guidance:*
• Regression models may allow for predictions based on the existing relationships of the two variables within the context of the problem. • The analysis of the least squares regression examines the appropriateness and reliability of the model.  • Given a collection of pairs (x, y) of numbers, where not all the *x*-values are the same, a line that best fits the data is called the line of least squares regression. • x represents the explanatory variable. • $\hat{y}$ is read *y* "hat". This represents the response variable. • When working with lines of least squares regressions, *b* is often used to reflect the slope (or rate of change in *y* for each increase of one unit in the *x* direction). This is different from what is taught in pre-algebra and algebra courses, when writing the equation of a line in slope-intercept form. • β is read "Beta", which is a Greek letter. β_0 represents the *y*-intercept, and β_1 represents the slope. • a or β_0 represent the *y*-intercept (or the value of *y* when *x* equals 0). • r is the correlation coefficient.

PS.DS.6 The student will create and interpret a linear model using the least squares regression method to assess the relationship between two quantitative variables.

Additional Content Background and Instructional Guidance:

- r^2 is a statistical measure that indicates how much of the variation of a dependent variable is explained by an independent variable in a regression model. r^2 measures the strength of the relationship between a model and the dependent variable on a scale from 0% to 100%.
- Refer to the following linear example:

Marketing Expenses x	Sales y
5000	25000
6700	30000
12000	26000
25000	35000
28000	50000
40000	60000
45000	75000

STATISTICS

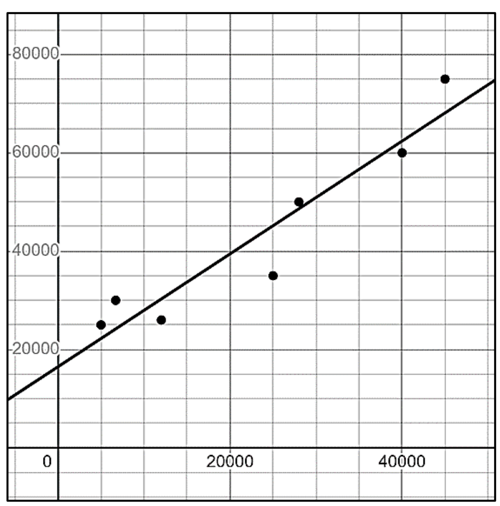
$r^2 = 0.9012$

$r = 0.9493$

PARAMETERS

slope $= b = 1.14695$

y-intercept $= a = 16505.4$



The regression equation is $\hat{y} = 16505.4 + 1.14695x$.

$r^2 = 90.12\%$

Probability

PS.P.1† The student will organize information and apply probability rules to compute probabilities of events within the context of a problem.

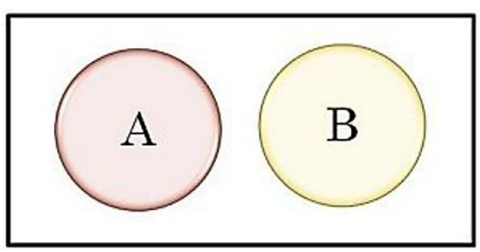
Students will demonstrate the following Knowledge and Skills:

a) Given two or more events, determine whether the events are complementary, dependent, independent, and/or mutually exclusive, and compute the probability of those events.
b) Represent and calculate probabilities using Venn diagrams, tree diagrams, and two-way tables.
c) Apply the addition rule, the multiplication rule, and complementary rule to calculate probabilities.
d) Calculate conditional probabilities to determine the association or independence of two events.

PS.P.1† The student will organize information and apply probability rules to compute probabilities of events within the context of a problem.

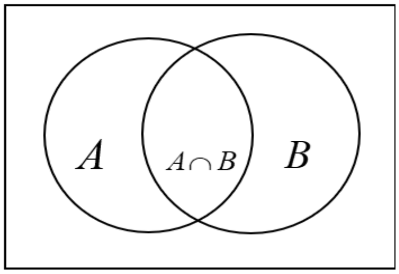
Additional Content Background and Instructional Guidance:

- Probability plays a vital role in everyday life to assess the likelihood or risk of a specific event occurring out of all possible events.
- Venn diagrams, tree diagrams, and two-way tables may be used as organizational tools to determine probabilities.
- Conditional probabilities allow for determination of the association or independence of two events.
- The Fundamental Counting Principle states that if one decision can be made n ways and another can be made m ways, then the two decisions can be made nm ways.
- A sample space is the set of all possible outcomes of a random experiment.
- An event is a subset of the sample space.
- $P(E)$ is a way to represent the probability that the event E occurs.
- Mutually exclusive events are events that cannot both occur simultaneously. A visual representation of a mutually exclusive event is as follows –



- If A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$.
- The complement of event A consists of all outcomes in which event A does not occur.
- P(B|A) is the probability that B will occur given that A has already occurred. P(B|A) is called the conditional probability of B given A.

PS.P.1† The student will organize information and apply probability rules to compute probabilities of events within the context of a problem. *Additional Content Background and Instructional Guidance:*
• Venn diagrams may be used to examine conditional probabilities.



$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$
$$\Rightarrow P(A \cap B) = P(A)P(B|A)$$

- Two events, *A* and *B*, are independent if the occurrence of one does not affect the probability of the occurrence of the other. If *A* and *B* are not independent, then they are said to be dependent.
- If *A* and *B* are independent events, then $P(A \cap B) = P(A)P(B)$.
- The *Law of Large Numbers* states that as a procedure is repeated again and again, the relative frequency probability of an event tends to approach the actual probability.

PS.P.2 The student will represent and interpret situations using discrete random distributions, including binomial distributions.

Students will demonstrate the following Knowledge and Skills:

a) Identify discrete random variables and create a table to represent valid discrete probability distributions within the context of a problem.
b) Calculate and interpret the mean (expected value) and standard deviation for a discrete random variable within the context of a problem.
c) Determine if a discrete random variable satisfies the conditions for a binomial distribution.
d) Design and conduct a simulation of a binomial distribution.
e) Calculate and interpret probabilities from a binomial distribution within the context of a problem.
f) Calculate the mean and standard deviation for binomial distributions.
g) Describe the center, shape, and spread of a discrete random variable within the context of a problem.

PS.P.2 The student will represent and interpret situations using discrete random distributions, including binomial distributions.

Additional Content Background and Instructional Guidance:

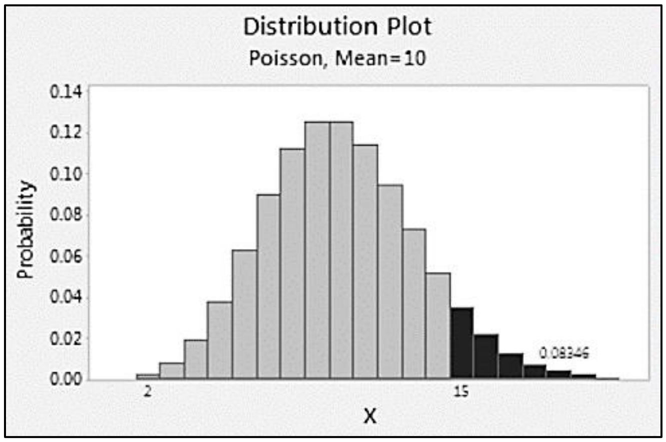
- Random variables are used to model random processes and are relevant to many contextual problems.
- A probability distribution combines descriptive statistical methods and probabilities to form a theoretical model of behavior.
- A discrete distribution describes the probability of occurrence of each value of a discrete random variable. A discrete random variable is a random variable that has countable values, such as a list of non-negative integers.
- Discrete random variables can be represented in tables and bar graphs. Both forms are valuable in creating a picture of the distribution.
- With a discrete probability distribution, each possible value of the discrete random variable can be associated with a non-zero probability.
- With a discrete distribution, unlike with a continuous distribution, the probability that X is exactly equal to some value can be calculated. For example, the discrete Poisson distribution can be used to describe the number of customer complaints within a day. The average number of complaints per day is 10. What is the probability of receiving 5, 10, and 15 customer complaints in one day?

x	$P(X = x)$
5	0.037833
10	0.125110
15	0.034718

PS.P.2 The student will represent and interpret situations using discrete random distributions, including binomial distributions.

Additional Content Background and Instructional Guidance:

A discrete distribution can be viewed on a distribution plot to see the probabilities between ranges. The shaded bars in this example represent the number of occurrences when the daily customer complaints are 15 or more. The height of the bars sums to 0.08346; therefore, the probability that the number of calls per day is 15 or more is 8.35%.



- A binomial random variable is a special class of discrete random variables that can be used to model the likelihood of success over a fixed number of trials.
- Binomial distribution only counts two states, typically represented as 1 (for a success) or 0 (for a failure), given a number of trials in the data. Binomial distribution thus represents the probability for *x* successes in *n* trials, given a success probability *p* for each trial.
- A binomial distribution's expected value, or mean, is calculated by multiplying the number of trials (n) by the probability of successes (p), or $n \times p$. For example, the expected value of the number of heads in 100 trials of heads or tails is 50, or (100 × 0.5).
- The binomial distribution function is calculated as:

$$P_{(x:n,p)} = {}^{n}C_{x}\, p^{x} (1 - p)^{n-x}$$

where:
  - n is the number of trials (occurrences)
  - x is the number of successful trials
  - p is the probability of success in a single trial
  - ${}^{n}C_{x}$ is the combination of n and x. A combination is the number of ways to choose a sample of x elements from a set of n distinct objects where order does not matter, and replacements are not allowed. Note that ${}_{n}C_{x} = \frac{n!}{r!(n-r)!}$, where ! is factorial (so, 4! = 4 × 3 × 2 × 1).

- The mean of the binomial distribution is np, and the variance of the binomial distribution is $np(1-p)$.
  - When $p = 0.5$, the distribution is symmetric around the mean.

PS.P.2 The student will represent and interpret situations using discrete random distributions, including binomial distributions. *Additional Content Background and Instructional Guidance:*
◦ When $p > 0.5$, the distribution curve is skewed to the left. ◦ When $p < 0.5$, the distribution curve is skewed to the right. • Another representation of the binomial distribution is as follows – ◦ The discrete probability distribution $P_p(n \mid N)$ of obtaining exactly n successes out of N Bernoulli trails (where the results of each Bernoulli trial is true with probability p and false with probability $q = 1 - p$). $$P_p(n \mid N) = \binom{N}{n} p^n q^{N-n}$$ $$= \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n},$$ where $\binom{N}{n}$ is the binomial coefficient.

PS.P.3† The student will represent and interpret situations using normal distributions.

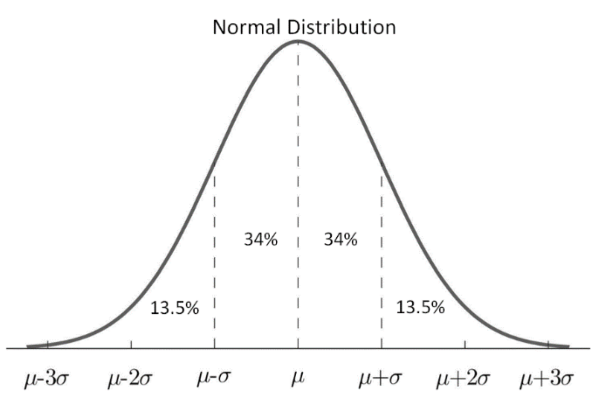
Students will demonstrate the following Knowledge and Skills:

a) Compare and contrast discrete and continuous distributions.
b) Represent probability as the area under the curve of a normal distribution using the Empirical Rule and graphing technology.
c) Describe the center, shape, and spread of normal distributions within the context of a problem.
d) Compare and contrast two or more sets of normally distributed data using *z*-scores, percentiles, or probabilities within the context of a problem.
e) Standardize a data value from a normal distribution and interpret the *z*-score within the context of a problem.
f) Calculate and interpret probabilities of a normal distribution using technology within the context of a problem.

PS.P.3† The student will represent and interpret situations using normal distributions.

Additional Content Background and Instructional Guidance:

- A discrete distribution is one in which the data can only take on certain values, for example integers. A continuous distribution is one in which data can take on any value within a specified range (which may be infinite).
- The normal curve is a probability distribution and the total area under the curve is 1.
- The normal distribution is a type of continuous random variable that is frequently used to represent and compare contextual situations.
- The normal distribution is a symmetrical curve defined by its mean and standard deviation.
- A normal distribution curve is the family of symmetrical, bell-shaped curves defined by the mean and the standard deviation of a data set. The arithmetic mean (μ) is located on the line of symmetry of the curve and is approximately equivalent to the median and mode of the data set.
- For a normal distribution, approximately 68 percent of the data fall within one standard deviation of the mean, approximately 95 percent of the data fall within two standard deviations of the mean, and approximately 99.7 percent of the data fall within three standard deviations of the mean. This is often referred to as the Empirical Rule or the 68 – 95 – 99.7 rule.



NOTE: This chart illustrates percentages that correspond to subdivisions in one standard deviation increments. Percentages for other subdivisions require the table of Standard Normal Probabilities or a graphing utility.

PS.P.3† The student will represent and interpret situations using normal distributions.

Additional Content Background and Instructional Guidance:

- The mean and standard deviation of a normal distribution affect the location and shape of the curve. The vertical line of symmetry of the normal distribution falls at the mean. The greater the standard deviation, the wider (“flatter” or “less peaked”) the distribution of the data.
- A z-score derived from a particular data value tells how many standard deviations that data value falls above or below the mean of the data set. It is positive if the data value lies above the mean and negative if the data value lies below the mean.
- A standard normal distribution is the set of all z-scores. The mean of the data in a standard normal distribution is 0 and the standard deviation is 1. This allows for the comparison of unlike normal data.
- Areas under the curve represent probabilities associated with continuous distributions.
- Familiarity with the center, shape, and spread of the standard normal curve is a foundation for using normal distributions to compute probabilities and make comparisons.

Inferential Statistics

PS.IS.1 The student will apply properties of sampling distributions and inference procedures to make decisions about population proportions.

Students will demonstrate the following Knowledge and Skills:

a) Describe the shape, center, and spread of the sampling distribution of a proportion within the context of a problem.

b) Given a problem, construct a one sample *z* confidence interval:

  i) identify the basic conditions for inference: random sample, independence, and normality;

  ii) calculate a confidence interval using technology; and

  iii) interpret the interval within the context of the problem.

c) Explain how changes in confidence level and sample size affect width of the confidence interval and margin of error.

d) Calculate and interpret a point estimate and margin of error of a confidence interval for a proportion within the context of the problem.

e) Explain how and why the hypothesis testing procedure allows one to reach a statistical decision.

f) Given a problem, apply the one sample *z* hypothesis testing procedures:

  i) construct appropriate null and alternate hypotheses;

  ii) identify the basic conditions for inference: random sample; independence, and normality;

  iii) calculate and interpret the *p*-value using technology;

  iv) determine and justify whether to reject the null hypothesis; and

  v) interpret the results within the context of the problem.

g) Use the statistical cycle to plan and conduct a statistical study about a proportion to answer a question or address a problem with inference.

PS.IS.1 The student will apply properties of sampling distributions and inference procedures to make decisions about population proportions. *Additional Content Background and Instructional Guidance:*
• The sampling distribution is the foundation for inferential statistics for a proportion. • A primary goal of sampling is to estimate the value of a population proportion based on a statistic. • Confidence intervals provide plausible values based on the confidence level for a population proportion. • Sampling distributions have less variability with larger sample sizes. • Confidence intervals and tests of significance are complementary procedures. • Tests of significance assess the extent to which sample data support a hypothesis about a population proportion. • Statistically significant results do not necessarily mean the results are practically significant within the context of the problem.

PS.IS.2 The student will apply properties of sampling distributions and inference procedures to make decisions about populations.

Students will demonstrate the following Knowledge and Skills:

a) Describe the shape, center, and spread of the sampling distribution of a mean within the context of a problem.

b) Calculate and interpret a point estimate and a margin of error for a confidence interval of a mean within the context of a problem.

c) Describe the use of the Central Limit Theorem in satisfying the assumptions and conditions for inference about a mean.

d) Identify the properties of a *t* distribution.

e) Given a problem, construct a one sample *t* confidence interval:

  i) identify the basic conditions for inference: random sample, independence, and approximate normality;

  ii) calculate a confidence interval using technology; and

  iii) interpret the interval within the context of the problem.

f) Given a problem, apply the one sample t hypothesis testing procedures:

  i) construct appropriate null and alternate hypotheses;

  ii) identify the basic conditions for inference: random sample, independence, and approximate normality;

  iii) calculate and interpret the *p* value using technology;

  iv) determine and justify whether to reject the null hypothesis; and

  v) interpret the results within the context of the problem.

PS.IS.2 The student will apply properties of sampling distributions and inference procedures to make decisions about populations. *Additional Content Background and Instructional Guidance:*
• The sampling distribution is the foundation for inferential statistics for a mean. • A primary goal of sampling is to estimate the value of a population mean based on a statistic. • Confidence intervals provide plausible values based on the confidence level for a population mean. • The Central Limit Theorem allows the researcher to safely assume that the sampling distribution of the mean will be normal if sample size is sufficiently large. • Sampling distributions have less variability with larger sample sizes. • Confidence intervals and tests of significance are complementary procedures. • The *t* distribution allows for conservative estimates of the normal distribution because small samples are more variable. • Tests of significance assess the extent to which sample data support a hypothesis about a population mean.