

AP Statistics

Grade(s):	☐ K ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5	☐ 6 ☐ 7 ☐ 8	☐ 9 ☐ 10 ☒ 11 ☒ 12 ☐ Other _____
Discipline/Course:	Discipline: Mathematics Course: AP Statistics		
Course Title:	AP Statistics		
Prerequisite(s):	Algebra II Honors (B or higher) or Probability and Statistics (A or higher), and Teacher Recommendation		
Course Description: <i>Program of Studies</i>	AP Statistics introduces students to the major concepts and tools for formulating questions, collecting and analyzing data, and interpreting results from data. The practices and skills for the course help students understand the statistical problem-solving process. The content, skills, and assessments in AP Statistics focus on exploring data, sampling and experimentation, probability and simulation, and statistical inference. Students use technology, investigations, problem solving, and writing as they build conceptual understanding. Statistics is an excellent option for any student meeting the prerequisites, preparing them for a wide variety of future areas of study including economics and finance, STEM, and social sciences. Students should come prepared with strong literacy, mathematical reasoning, and critical thinking skills.		
Course Essential Questions:	• How do we summarize and interpret data to distinguish meaningful patterns from random variation? • How do data collection methods dictate the validity and scope of conclusions we can draw about a population or a causal relationship? • How can probability rules and distributions be used to model uncertainty and quantify chance outcomes? • How do we draw reliable, evidence-based conclusions about population parameters from sample data while accounting for sampling variability?		

Course Enduring Understandings:	• Measurement and variability are present in all real-world data. Graphical displays and summary statistics allow us to describe patterns, shapes, centers, and spreads, as well as identify unusual features or departures from expected distributions. • The scope of inference depends directly on how data are collected. Random sampling allows generalization to a population, whereas random assignment in well-designed experiments allows for cause-and-effect conclusions. • Probability models describe long-run chance behavior. Simulation and probability rules allow us to anticipate outcomes, model random variables, and quantify uncertainty in decision-making. • Statistical inference uses sample data and probability to estimate unknown population parameters (with margins of error) and assess the strength of evidence for or against claims about populations.	
Duration: Credit:	☐ Semester ☒ Full-Year	☐ 0.5 Credit (s) ☒ 1.0 Credit(s) ☐ 1.5 Credit(s) ☐ N/A
Course Materials/Resources:	The Practice of Statistics 7th Edition, BFW College Board AP Classroom: AP Statistics <u>AP® Statistics Course and Exam Description, Effective Fall 2026</u>	
FPS Course Academic Expectation(s):	☒ Exploring and Understanding (EU) ☐ Synthesizing and Evaluating (SE) ☐ Creating and Constructing (CC) ☒ Conveying Ideas (CI) ☐ Collaborating Strategically (CS) ☐ Using Communication Tools (UCT)	
Year at a Glance (Units):	Unit 1: Exploring One-Variable Data and Collecting Data (~11 classes) Unit 2: Probability, Random Variables, and Probability Distributions (~18 classes) Unit 3: Inference for Categorical Data: Proportions (~11 classes)	

Unit 4: Inference for Quantitative Data: Means (~11 classes)

Unit 5: Regression Analysis (~5 classes)

DRAFT

Unit Number and Title:	Unit 1: Exploring One-Variable Data and Collecting Data
Duration:	~ 11 classes
Resource(s):	The Practice of Statistics 7th Edition: Unit 1 & Unit 3 AP Classroom: Unit 1
Unit Overview:	Starting in Unit 1, students identify components and formulate investigative questions of a statistical study, define and represent categorical and quantitative variables, describe and compare distributions of one-variable data, interpret statistical calculations to assess claims about individual data points or samples, and justify the use of important principles of sampling and experimental design. Students also learn to talk about data in real-world contexts. Variability in data may suggest certain conclusions about the data distribution, but not all variation is meaningful. Statistics allow us to develop shared understandings of uncertainty and variation. Additionally, students learn that depending on how data are collected, they may or may not be able to generalize findings or establish evidence of causal relationships. For example, if random selection is not used to obtain a sample from a population, bias may result, and statistics from the sample could not be assumed to generalize to the population. For data collected using well-designed experiments, statistically significant differences between or among experimental treatment groups are evidence that the treatments caused the effect.
Learning Goals	
Standard(s):	Summarize, represent, and interpret data on a single count or measurement variable. ● S.ID-1: Represent data with plots on the real number line (dot plots, histograms, and box plots). ● S.ID-2: Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets. ● S-ID.3: Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers). ● S-ID.4: Use the mean and standard deviation of a data set to fit it to a normal distribution and to

	estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Understand and evaluate random processes underlying statistical experiments. ● S-IC.1: Understand statistics as a process for making inferences about population parameters based on a random sample from that population. ● S-IC.3: Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each. Make inferences and justify conclusions from sample surveys, experiments, and observational studies. ● S-ID.6b: Evaluate reports based on data.
Essential Question(s):	● How do graphical displays and summary statistics help us identify patterns and variability in data? ● How do sampling methods and experimental design dictate the validity and scope of conclusions?
Enduring Understanding(s):	● Depending on how data is collected, you may or may not be able to generalize findings or establish evidence of causal relations. ● Statistics allow us to develop a shared understanding of uncertainty. ● All data has context. ● In a well-designed experiment, statistically significant differences between or among experimental treatment groups are evidence that the treatments caused the effect.
Learning Goal(s): <i>Students will know and will be able to use their learning to:</i> (Content/ Skills)	Content: (Students will know...) ● a statistical study is a study in which data is collected from a sample to answer an investigative question about a larger population. ● statistical studies are necessary when the population is too large, or it is too difficult to collect data from every item or individual in the population.

- a datum (singular form of data) is a piece of information about an item or individual. A collection of data is called a data set.
- a population consists of all items or individuals of interest. The population size is represented by the symbol n .
- a sample selected for study is a subset of the population from which data are obtained. The number of items in the sample, called the sample size, is represented by the symbol n .
- each component of a statistical study and the resulting calculations can be related to an aspect of the corresponding real-world context from which the components were derived. This identification of a statistical result with the corresponding contextual component is what is meant by “in context.”
- an investigative question for a specific study should have a defined purpose and should not be changed based on the data analysis or results.
- an investigative question should be posed so that the required data can be collected and analyzed.
- a frequency table shows the number of observational units in each category of a categorical variable.
- a relative frequency table shows the proportion of observational units in each category of a categorical variable.
- counts and relative frequencies of categorical variables reveal information that can be used to justify claims about the variables in context.
- bar charts, also called bar graphs, display frequencies (counts) or relative frequencies (proportions) for the categories of a single categorical variable.
- pie charts are used to display frequencies (counts) or relative frequencies (proportions) for categorical data.
- frequency and relative frequency tables, bar charts, and pie charts can be used to compare two or more data sets in terms of the same categorical variable.
- histograms, stem-and-leaf plots, and dotplots provide a visual representation of the distribution of the values of a quantitative variable.
- descriptions of the distribution of one quantitative variable include shape, center, and variability (spread), as well as any unusual features such as outliers, gaps, or clusters in context.
- the shape of the distribution of one quantitative variable is skewed to the right (positively

skewed) if the right tail (toward larger values) is longer than the left. The shape of the distribution is skewed to the left (negatively skewed) if the left tail (toward smaller values) is longer than the right. The shape of the distribution is approximately symmetric if the left half is approximately the mirror image of the right half.

- distributions of one quantitative variable with one main peak are called unimodal. Distributions with two prominent peaks are called bimodal. A distribution in which each frequency or each relative frequency is approximately the same with no prominent peaks is approximately uniform.
- outliers for one quantitative variable are data points that are unusually small or large relative to the rest of the data.
- two commonly used measures of center in the distribution of a quantitative variable are the mean and median.
- the p th percentile is the value that has $p\%$ of the data less than or equal to it when the data set is ordered from smallest to largest. The first and third quartiles are the 25th and 75th percentiles, respectively.
- three commonly used measures of variability (or spread) in the distribution of a quantitative variable are the range, interquartile range, and standard deviation.
- a boxplot is a graphical representation of the five-number summary (minimum, first quartile, median, third quartile, maximum).
- a standardized score measures the number of standard deviations a data value falls above or below the mean.
- an experiment is a study in which a researcher assigns conditions, or treatments, to experimental units to explore an investigative question of interest about the population.
- an observational study is a study where treatments are not imposed.
- a sample is considered random when all observational units in the sample are selected from the population using some type of random mechanism, such as a random number generator.
- each random sampling method has different characteristics that make it more appropriate for sampling populations depending on the question being investigated.
- bias in a sampling method is a systematic error in the sampling procedure that results in a statistic being consistently larger or consistently smaller than the parameter the statistic is used to estimate.

- a well-designed experiment should include the following:
 - comparisons of at least two treatment groups, one of which could be a control group.
 - random assignment of treatments to experimental units.
 - replication.
 - direct control of potential extraneous sources of variation in the response.

Skills: (Students will be able to...)

- identify components within a statistical study.
- determine an investigative question within a statistical study.
- identify observational units, variables, parameters, and statistics from a statistical study or data set.
- identify types of variables (categorical v. quantitative).
- identify types of quantitative variables (discrete v. continuous).
- construct categorical one-variable tabular representations.
- describe categorical one-variable tabular representations with summary statistics.
- construct categorical one-variable graphical representations.
- justify a claim using categorical one-variable graphical representations.
- compare multiple categorical one-variable tabular and graphical representations.
- construct quantitative one-variable graphical representations.
 - histograms, stem-and-leaf plots, and dotplots
- describe distributions of quantitative one-variable graphical representations.
- justify a claim using distributions of quantitative one-variable graphical representations.
- calculate measures of center and position for quantitative data.
 - mean and median.
 - Q1, Q3.
 - pth percentile is the value that has p% of the data less than or equal to it when the data set is ordered from smallest to largest.
- calculate measures of variability for quantitative data.
 - range, interquartile range, and standard deviation.
- calculate different units of measurement for summary statistics.

- calculate outliers for quantitative data.
 - an outlier is a value located more than $1.5 \times \text{IQR}$ above the third quartile or more than $1.5 \times \text{IQR}$ below the first quartile.
- compare multiple quantitative one-variable summary statistics.
 - summary statistics can be used to compare features of two or more independent samples, including center, variability, shape, and outliers.
- justify the selection of a summary statistic for describing quantitative data.
- construct quantitative one-variable graphical representations of summary statistics.
- describe quantitative one-variable graphical representations of summary statistics based on the relationship of the mean and the median.
- compare multiple quantitative one-variable graphical representations.
- compare multiple quantitative one-variable graphical representations of summary statistics.
- justify a claim using multiple quantitative one-variable graphical representations.
- calculate z-scores with population parameters.
- compare z-scores as measures of relative position for distributions.
- determine the components of an investigative question within a statistical study.
- identify a census, observational study & experiment.
- justify the appropriateness of generalizations for a statistical study.
- identify a sampling method given a description of a study;
 - sampling with and without replacement,
 - SRS, stratified, cluster, systematic.
- justify the appropriateness of a sampling method.
- identify potential sources of bias in sampling methods;
 - voluntary response, undercoverage, nonresponse, response.
- identify elements of a well-designed experiment;
 - comparative, random assignment of treatments, replication, control,
 - placebo effect,
 - single- and double-blind.
- identify experimental designs;
 - completely randomized design,
 - blocking,

- | | |
|--|---|
| | <ul style="list-style-type: none">○ matched pairs.● justify the appropriateness of a particular experimental design.● justify the appropriateness of the conclusions based on a well-designed experiment. |
|--|---|

DRAFT

Unit Number and Title:	Unit 2: Probability, Random Variables, and Probability Distributions
Duration:	~18 classes
Resource(s):	The Practice of Statistics 7th Edition: Unit 4 & Unit 5 AP Classroom: Unit 2
Unit Overview:	In this unit, students explore relationships in two-variable categorical data sets that will help them understand some of the rules of probability. Probabilistic reasoning allows statisticians to quantify the likelihood of random events over the long run and make statistical inferences. Simulations and concrete examples can help students understand the abstract definitions and calculations of probability. This unit builds on understandings of simulated or empirical data distributions and fundamental principles of probability to represent, interpret, and calculate parameters for theoretical probability distributions for discrete random variables. When students interpret probabilities and parameters associated with a probability distribution, they are expected to use appropriate units and relate to the context of the situation. This unit also introduces the standard normal distribution, including the empirical rule, sampling distribution of a statistic, and the central limit theorem (CLT).
Learning Goals	
Standard(s):	Summarize, represent, and interpret data on two categorical and quantitative variables. ● S-ID.5: Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data. Understand independence and conditional probability and use them to interpret data. ● S-CP.1: Describe events as subsets of a sample space (the set of outcomes) using characteristics (or categories) of the outcomes, or as unions, intersections, or complements of other events ("or," "and," "not"). ● S-CP.2: Understand that two events A and B are independent if the probability of A and B occurring together is the product of their probabilities, and use this characterization to determine if they are independent.

- **S-CP.3:** Understand the conditional probability of A given B as $P(A \text{ and } B)/P(B)$, and interpret independence of A and B as saying that the conditional probability of A given B is the same as the probability of A, and the conditional probability of B given A is the same as the probability of B.
- **S-CP.4:** Construct and interpret two-way frequency tables of data when two categories are associated with each object being classified. Use the two-way table as a sample space to decide if events are independent and to approximate conditional probabilities.
- **S-CP.5:** Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations.

Use the rules of probability to compute probabilities of compound events.

- **S-CP.6:** Find the conditional probability of A given B as the fraction of B's outcomes that also belong to A, and interpret the answer in terms of the model.
- **S-CP.7:** Apply the Addition Rule, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$, and interpret the answer in terms of the model.
- **S-CP.8 (+):** Apply the general Multiplication Rule in a uniform probability model, $P(A \text{ and } B) = P(A)P(B|A) = P(B)P(A|B)$, and interpret the answer in terms of the model.
- **S-CP.9 (+):** Use permutations and combinations to compute probabilities of compound events and solve problems.

Calculate expected values and use them to solve problems

- **S-MD.1 (+):** Define a random variable for a quantity of interest by assigning a numerical value to each event in a sample space; graph the corresponding probability distribution using the same graphical displays as for data distributions.
- **S-MD.2 (+):** Calculate the expected value of a random variable; interpret it as the mean of the probability distribution.
- **S-MD.3 (+):** Develop a probability distribution for a random variable defined for a sample space in which theoretical probabilities can be calculated; find the expected value.
- **S-MD.4 (+):** Develop a probability distribution for a random variable defined for a sample space in which probabilities are assigned empirically; find the expected value.

	Use probability to evaluate outcomes of decisions. ● S-MD.5 (+): Weigh the possible outcomes of a decision by assigning probabilities to payoff values and finding expected values. ● S-MD.6 (+): Use probabilities to make fair decisions (e.g., drawing by lots, using a random number generator). ● S-MD.7 (+): Analyze decisions and strategies using probability concepts (e.g., product testing, medical testing, pulling a hockey goalie at the end of a game).
Essential Question(s):	● How do probability rules and theoretical distributions allow us to model uncertainty and anticipate long-run chance behavior? ● How do joint, marginal, and conditional probabilities help us analyze relationships between two categorical variables? ● How do random variables and sampling distributions (including the Central Limit Theorem) bridge probability theory and statistical inference?
Enduring Understanding(s):	● Probability models describe long-run chance behavior, where clear patterns emerge over many repetitions despite short-term randomness. ● Two-way tables, conditional probability, and independence rules allow us to quantify likelihoods and evaluate relationships between two categorical variables. ● Random variables and theoretical probability distributions (e.g., Binomial, Geometric, Normal) provide the parameters (μ and σ) necessary to model sampling variability and establish the foundation for statistical inference.
Learning Goal(s): <i>Students will know and will be able to use their learning to:</i> (Content/ Skills)	Content: (Students will know...) ● a two-way table, also called a contingency table, can be used to summarize and compare data for two categorical variables. ● side-by-side bar charts, segmented bar charts, and mosaic plots are examples of graphs used to display the relationship between two categorical variables. ● summary statistics for two categorical variables can be used to compare distributions for evidence of association between the two variables:

- joint relative frequency.
- marginal relative frequency.
- conditional relative frequency.
- simulation is a way to model random events such that simulated outcomes closely match real-world outcomes.
- the probability of an outcome or event is its long-run relative frequency—that is, its relative frequency over a large number of trials.
- the relative frequency of an outcome or event determined from empirical data can be used to estimate the actual, or true, probability of that outcome or event.
- the law of large numbers states that for independent trials, as the number of trials increases, the long-run relative frequency of the outcome or event gets closer and closer to a single value
- conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$.
- general multiplication rule: $P(A \cap B) = P(A) \cdot P(B|A)$.
- events A and B are independent if and only if knowing whether event A has occurred (or will occur) does not change the probability that event B will occur.
- a random variable is a variable whose values have numerical outcomes that result from a random phenomenon.
- a discrete probability distribution can be determined using the rules of probability or estimated with a simulation.
- a cumulative probability distribution can be represented as a table or function and shows the probability of being less than or equal to each value of the discrete random variable.
- a numerical value measuring a characteristic of a probability distribution of a random variable, or a population, is a parameter. the value of a parameter is a single, fixed value.
- the mean and standard deviation for the probability distribution of a discrete random variable should be interpreted in the context of a specific population.
- a binomial random variable, X, is a discrete random variable that counts the number of successes in repeated independent trials, n, that have only two possible outcomes (success or failure), with the probability of success p and the probability of failure 1– p.
- a normal distribution can be described as a continuous, unimodal, bell-shaped, and symmetric curve.

- a normal curve can be used to model a distribution of data and a continuous random variable.
- the normal distribution, or the normal curve, is identified by two parameters, the mean, μ , and the standard deviation, σ .
- the empirical rule (68-95-99.7 rule) can be used to estimate the area of a region under the graph of the normal distribution curve.
- if the distribution of a random variable is approximately normal, the probability that the random variable takes on values within a particular interval of the random variable is determined by the area under the normal curve within that interval. The total probability or area under the normal curve is 1.
- the boundaries of an interval associated with a given area in a normal distribution can be determined using technology.
- percentiles and proportions may be used to compare relative positions of individual values within a normal distribution or between normal distributions.
- a sampling distribution of a statistic is the distribution of values of the statistic for all possible samples of a given size from a given population.
- the central limit theorem (CLT) states that the sampling distribution of a mean of a random sample has a shape that can be approximated by a normal distribution.

Skills: (Students will be able to...)

- compare tabular and graphical representations for the relationship between two categorical variables.
- justify a claim using tabular and graphical representations for the distributions of two categorical variables.
- calculate summary statistics from two-way tables.
- compare summary statistics for two categorical variables.
- justify a claim using summary statistics for two categorical variables.
- estimate probabilities using simulations.
- calculate probabilities for events and their complements.
 - sample space.
- justify why two events are mutually exclusive (or disjoint) using joint probability.

- calculate conditional probabilities.
- calculate probabilities for independent events and for the union of two events.
- construct a probability distribution for a discrete random variable.
- calculate the mean and standard deviation for a discrete random variable.
- interpret the mean and standard deviation for a discrete random variable.
- justify why a random variable is or is not a binomial random variable.
- calculate the mean and standard deviation for a binomial distribution.
- interpret the mean, standard deviation, and probabilities for a binomial distribution.
- estimate probabilities of binomial random variables using data from a simulation.
- calculate probabilities for a binomial distribution.
- describe a normal distribution.
- calculate the mean and standard deviation for a normal distribution.
- calculate percentages from a normal distribution using the empirical rule.
- calculate the probability that a particular value lies within a given interval of a normal distribution.
- calculate the associated intervals and areas of a normal distribution.
- compare measures of relative position for distributions.
- describe sampling distributions with simulations.

Unit Number and Title:	Unit 3: Inference for Categorical Data: Proportions
Duration:	~11 classes
Resource(s):	The Practice of Statistics 7th Edition: Unit 6 & Unit 8 AP Classroom: Unit 3
Unit Overview:	This unit introduces sampling distributions and statistical inference. Students learn that sample statistics can be used to estimate corresponding population parameters and that measures of center (mean) and variability (standard deviation) for these sampling distributions can be determined directly from the population parameters when certain sampling criteria are met. For large enough samples from any population, these sampling distributions can be approximated by a normal distribution. Simulating sampling distributions helps students understand how the values of statistics vary in repeated random sampling from populations with known parameters. This unit also introduces statistical inference. Students analyze categorical data to make inferences about binomial population proportions. Provided that conditions are met, students use statistical inference to construct and interpret confidence intervals to estimate population proportions and perform significance tests to evaluate claims about population proportions. Students begin by learning inference procedures for one proportion and then examine inference methods for a difference between two proportions. They also interpret the two types of errors that can be made in a significance test, the probability of each error occurring, and possible consequences, in context. Students are also introduced to chi-square tests, which can be used when there are two or more categories. Students need to understand how to select between the following two tests: (1) the chi-square test for independence (for associations between categorical variables within a single population), and (2) the chi-square test for homogeneity (for comparing distributions of a categorical variable across populations or treatments). To integrate conceptual understanding, teachers can help students make connections among frequency tables, conditional probability, and calculating expected counts. The chi-square statistic is introduced to measure the distance between observed and expected counts relative to expected counts.

Learning Goals	
Standard(s):	Understand and evaluate random processes underlying statistical experiments. ● S-IC.2: Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation. Make inferences and justify conclusions from sample surveys, experiments, and observational studies. ● S-IC.4: Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling. ● S-IC.5: Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant. <i>Note: Advanced formal z-tests, confidence intervals, and chi-square procedures in Unit 3 extend beyond high school CT standards and standard simulation-based inference.</i>
Essential Question(s):	● How do sample proportions and confidence intervals allow us to estimate unknown population parameters with a quantified margin of error? ● How does a 4-step hypothesis testing framework quantify evidence against a claim about a population proportion? ● How do Chi-Square tests extend inference to assess model fit, homogeneity, and independence across categorical distributions?
Enduring Understanding(s):	● Statistical inference uses sample proportions and normal sampling distributions to estimate unknown population parameters with a specified margin of error. ● Significance tests evaluate claims about population proportions by calculating the probability (p-value) of observing sample results as extreme as those found, assuming the null hypothesis is true. ● Chi-Square procedures allow statisticians to test whether an observed frequency distribution fits

	an expected categorical model (Goodness-of-Fit) or whether two categorical variables are independent across populations.
Learning Goal(s): <i>Students will know and will be able to use their learning to:</i> (Content/ Skills)	Content: (Students will know...) • A sample statistic is a point estimator of the corresponding population parameter and can be thought of as the estimate of the population parameter. • For a population with population proportion p, when the sampled values are independent, the sampling distribution of a sample proportion $\hat{p}$ has a mean $\mu_{\hat{p}}$. • The sampling distribution of the sample proportion $\hat{p}$ is approximately normal provided the sample size is large enough. • A confidence interval is an interval estimate for a population parameter. • a confidence interval can be calculated to estimate the value of a single population proportion. • the interpretation of the confidence level is that, in repeated random sampling with the same sample size, approximately C% of confidence intervals calculated will capture the population proportion, with C representing the numerical value of the confidence level used. • the standard error (SE) of a statistic is an estimate of the standard deviation of the sampling distribution of the statistic. • a hypothesis test is a statistical inference procedure that is used to make a decision about the value of a population parameter. • In the hypothesis testing procedure, the null hypothesis, H_0, is the statement about a parameter that is assumed to be correct unless there is convincing statistical evidence suggesting otherwise. • given that the null hypothesis is true, there is a probability distribution of the test statistic called the null distribution. • the distribution of the test statistic assuming the null hypothesis is true (null distribution) can be approximated by a probability model (e.g., a theoretical distribution such as the standard normal distribution). • the relationship between a p-value and the significance level of a hypothesis test determines whether a result is statistically significant. • the results of a hypothesis test for a population proportion can serve as the statistical reasoning

- to support the answer to an investigative question about the population that was sampled.
- confidence intervals and tests for difference of proportions.
 - a Type I error occurs when there is convincing statistical evidence that the alternative hypothesis is true (due to the small p-value), but it is not.
 - a Type II error occurs when there is not convincing statistical evidence that the alternative hypothesis is true (due to the large p-value), but it is.
 - in some studies, making a Type I error may have more serious consequences than making a Type II error. In other studies, making a Type II error may have more serious consequences than making a Type I error. The consequences of each error should be considered prior to conducting the study.
 - the power of a hypothesis test is the probability that a hypothesis test will correctly reject the false null hypothesis.
 - based on the sample data, a confidence interval can be calculated to estimate the difference between two population proportions.
 - the appropriate testing method for the difference between two population proportions is a two-sample z-test for the difference between two population proportions.
 - the chi-square statistic measures the distance between observed and expected counts relative to expected counts.
 - to determine whether the distributions of a categorical variable for two or more populations are different, the appropriate test is the chi-square test for homogeneity.
 - to determine whether row and column variables in a two-way table of categorical data might be associated in the single population from which the data were sampled, the appropriate test is the chi-square test for independence.

Skills: (Students will be able to...)

- justify why an estimator is or is not unbiased.
- calculate estimates for a population parameter.
- calculate the mean and standard deviation of a sampling distribution for a sample proportion.
- justify the appropriateness of conditions for the sampling distribution of a sample proportion.

- interpret the mean, standard deviation, and probabilities for a sampling distribution of a sample proportion.
- identify an appropriate confidence interval procedure including the parameter for a population proportion.
- justify the appropriateness of constructing a confidence interval for a population proportion by verifying conditions.
- calculate an appropriate confidence interval for a population proportion.
- calculate the standard error and margin of error of a sample statistic for a confidence interval for a population proportion, and estimate a given sample size from the margin of error.
- interpret a confidence interval in context for a population proportion.
- justify a claim based on a confidence interval for a population proportion.
- identify the relationships among sample size, confidence interval width, confidence level, and margin of error for a population proportion.
- identify an appropriate testing method for a population proportion, including the parameter for the population proportion.
- identify the null and alternative hypotheses for a population proportion.
- justify the appropriateness of a hypothesis test for a population proportion by verifying conditions.
- interpret the p-value of a hypothesis test for a population proportion.
- calculate an appropriate test statistic and p-value for testing a hypothesis about a population proportion.
- justify a claim about the population based on the results of a hypothesis test for a population proportion.
- identify and calculate the probability of Type I and Type II errors.
- calculate the power of a hypothesis test
- interpret Type I and Type II errors.
- identify the factors that affect the probability of errors in hypothesis testing.
- calculate the mean and standard deviation of the sampling distribution for the difference between two sample proportions.
- justify the appropriateness of conditions for the sampling distribution for the difference between two sample proportions.

- interpret the mean, standard deviation, and probabilities for the sampling distribution for the difference between two sample proportions.
- identify an appropriate confidence interval procedure including the parameters for the difference between two population proportions.
- justify the appropriateness of constructing a confidence interval for the difference between two population proportions by verifying conditions.
- calculate an appropriate confidence interval for the difference between two population proportions.
- calculate the standard error and margin of error for estimating the difference between two population proportions.
- interpret a confidence interval in context for the difference between two population proportions.
- justify a claim based on a confidence interval for the difference between two population proportions.
- identify an appropriate testing method for the difference between population proportions, including the parameters.
- identify the null and alternative hypotheses for the difference between population proportions.
- justify the appropriateness of a hypothesis test for the difference between two population proportions by verifying conditions.
- calculate an appropriate test statistic and p-value for testing a hypothesis for the difference between two population proportions.
- interpret the p-value of a hypothesis test for the difference between two population proportions.
- justify a claim about the populations based on the results of a hypothesis test for the difference between two population proportions.
- describe chi-square distributions.
- identify an appropriate testing method for comparing distributions in two-way tables of categorical data, including the populations and variables.
- identify the null and alternative hypotheses for a chi-square test for homogeneity or independence.
- justify the appropriateness of a chi-square test for independence or homogeneity by verifying conditions.
- calculate expected counts for two-way tables of categorical data.

- | | |
|--|---|
| | <ul style="list-style-type: none">• calculate the appropriate test statistic and p-value for a chi-square test for homogeneity or independence.• interpret the p-value for the chi-square test for homogeneity or independence.• justify a claim about the population(s) based on the results of a chi-square test for homogeneity or independence. |
|--|---|

Unit Number and Title:	Unit 4: Inference for Quantitative Data: Means
Duration:	~11 classes
Resource(s):	The Practice of Statistics 7th Edition: Unit 7 AP Classroom: Unit 4
Unit Overview:	In this unit, students analyze quantitative data to make inferences about population means. Students learn that t^* and t-tests are used for inference with means when the population standard deviation, σ , is unknown. Using s for σ in the formula for z gives a slightly different value, t . The distribution of t depends on sample size and has more area in the tails than a normal distribution. The boundaries for rejecting a null hypothesis using a t-distribution tend to be further from the mean than for a normal distribution. Students discover how and why conditions for inference with proportions and means are similar and different.
Learning Goals	
Standard(s):	Understand and evaluate random processes underlying statistical experiments. • S-IC.2: Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation. Make inferences and justify conclusions from sample surveys, experiments, and observational studies. • S-IC.4: Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling. • S-IC.5: Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant. <i>Note: Unit 4 extends beyond high school state standards in statistical tests and inferences.</i>

Essential Question(s):	- How does the Central Limit Theorem justify using normal and t-distributions to model the sampling variability of sample means? - Why do t-distributions replace z-distributions when estimating population means, and how do degrees of freedom impact interval width? - How do single-sample, paired, and two-sample t-procedures allow us to test claims and draw conclusions about population means?
Enduring Understanding(s):	- The Central Limit Theorem and sampling distributions establish that sample means vary predictably around the true population mean for sufficiently large samples. t-distributions account for the additional variability introduced by using the sample standard deviation (s) to estimate the unknown population standard deviation (σ). t-confidence intervals and t-tests apply a standard 4-step framework (State, Plan, Do, Conclude) to estimate population means or evaluate whether observed differences between groups are statistically significant.
Learning Goal(s): <i>Students will know and will be able to use their learning to:</i> (Content/ Skills)	Content: (Students will know...) - for a population with population mean μ and population standard deviation σ, when the sampled values are independent, the sampling distribution of the sample mean has mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. - for a quantitative variable, if the population distribution can be modeled by a normal distribution, the sampling distribution of the sample mean, $\bar{x}$, can be modeled with a normal distribution regardless of the sample size. t-distributions, also called Student's t-distributions, form a family of symmetric, bell-shaped, standardized distributions with wider tails than that of the standard normal distribution. - the appropriate confidence interval procedure for estimating the population mean of a quantitative variable for one sample is a one-sample t-interval for a population mean - the interpretation of the confidence level is as follows: In repeated random sampling with the same sample size from the same population, approximately $C\%$ of confidence intervals created

will capture the population mean or population mean difference, where C represents the numerical value of the confidence level used.

- when interpreting a C% confidence interval for a population mean or population mean difference, we say we are C% confident the interval contains the value of the population mean or population mean difference, where a represents the lower limit and b represents the upper limit.
- the appropriate test for a population mean with unknown population standard deviation σ is a one-sample t-test for a population mean.
- the hypothesis testing procedure for the matched pairs design is a one-sample t-test for the population mean difference
- for a population mean or population mean difference, is t, where t has degrees of freedom $n - 1$. The t-statistic has a t-distribution with degrees of freedom $n - 1$ when the null hypothesis is true.
- for two independent populations with population means μ_1 and μ_2 and population standard deviations σ_1 and σ_2 , when the sampled values are independent, the sampling distribution of the difference in sample means $(\bar{x}_1 - \bar{x}_2)$ has a mean $\mu_{(\bar{x}_1 - \bar{x}_2)}$ and standard deviation $\sigma_{(\bar{x}_1 - \bar{x}_2)}$.
- based on the sample data, a confidence interval can be calculated to estimate the difference between two population means.
- the appropriate confidence interval procedure for two independent samples is a two-sample t-interval for the difference between population means.
- the confidence interval for the difference between two population means is calculated based on samples from two populations; the computed interval may or may not contain the value for the difference between the two population means.
- the appropriate test for the difference between two population means is a two-sample t-test for a difference between two population means.
- the results of a hypothesis test for a two-sample t-test for a difference between two population means can serve as the statistical reasoning to support the answer to an investigative question about the two populations that were sampled.

Skills: (Students will be able to...)

- calculate the mean and standard deviation of a sampling distribution of a sample mean.
- justify the appropriateness of conditions for the sampling distribution of a sample mean.
- interpret the mean, standard deviation, and probabilities for the sampling distribution of a sample mean.
- describe t-distributions.
- identify an appropriate confidence interval procedure including the parameter for a population mean or population mean difference.
- justify the appropriateness of constructing a confidence interval for a population mean or population mean difference by verifying conditions.
- calculate an appropriate confidence interval for a population mean or population mean difference.
- calculate the standard error and margin of error for a sample size for a one-sample t-interval.
- interpret a confidence interval in context for a population mean or population mean difference.
- justify a claim based on a confidence interval for a population mean or population mean difference.
- identify the relationships among sample size, confidence interval width, confidence level, and margin of error for a population mean or population mean difference.
- identify an appropriate testing method and parameter for a population mean or population mean difference with unknown σ .
- identify the null and alternative hypotheses for a population mean or population mean difference with unknown σ .
- justify the appropriateness of a hypothesis test for a population mean or population mean difference by verifying conditions.
- calculate an appropriate test statistic and p-value for testing a hypothesis about a population mean or population mean difference.
- interpret the p-value of a hypothesis test for a population mean or population mean difference.
- justify a claim about the population based on the results of a hypothesis test for a population mean or population mean difference.
- calculate the mean and standard deviation of a sampling distribution for the difference between two sample means.

- justify the appropriateness of conditions for the sampling distribution of the difference between two sample means.
- interpret the mean, standard deviation, and probabilities for a sampling distribution for the difference between sample means.
- identify an appropriate confidence interval procedure including the parameter for the difference between two population means.
- justify the appropriateness of constructing a confidence interval for the difference between two population means by verifying conditions.
- calculate an appropriate confidence interval for the difference between two population means.
- calculate the standard error and margin of error for estimating the difference between two population means.
- interpret a confidence interval in context for the difference between two population means.
- justify a claim based on a confidence interval for the difference between two population means.
- identify an appropriate testing method for the difference between two population means, including the parameters for the difference between the two population means.
- identify the null and alternative hypotheses for the difference between two population means.
- justify the appropriateness of a hypothesis test for the difference between two population means by verifying conditions.
- calculate an appropriate test statistic and p-value for testing a hypothesis for the difference between two population means.
- interpret the p-value of a hypothesis test for the difference between two population means.
- justify a claim about the populations based on the results of a hypothesis test for the difference between two population means.

Unit Number and Title:	Unit 5: Regression Analysis
Duration:	~5 classes
Resource(s):	The Practice of Statistics 7th Edition: Unit 2 AP Classroom: Unit 5
Unit Overview:	In this unit, students explore relationships in two-variable quantitative data sets. Students describe form, direction, strength, and unusual features for an association between two quantitative variables. They assess correlation and, if appropriate, use a linear model to predict values of the response variable from values of the explanatory variable. Students interpret the least-squares regression line in context, analyze prediction errors (i.e., residuals), and explore departures from a linear pattern.
Standard(s):	Summarize, represent, and interpret data on two categorical and quantitative variables. • S-ID.6: Represent data on two quantitative variables on a scatter plot, and describe how the variables are related. a. Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models. b. Informally assess the fit of a function by plotting and analyzing residuals. c. Fit a linear function for a scatter plot that suggests a linear association. Interpret linear models • S-ID.7: Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data. • S-ID.8: Compute (using technology) and interpret the correlation coefficient of a linear fit. • S-ID.9: Distinguish between correlation and causation. <i>Note: Unit 5 extends beyond CCSS into inference for regression.</i>

Essential Question(s):	• How can the relationship between two variables be modeled, and what are the limitations of using that model for prediction? • How does residual analysis evaluate the appropriateness and fit of a linear regression model?
Enduring Understanding(s):	• Scatterplots and correlation coefficients measure the strength, direction, and form of linear associations between two quantitative variables, though correlation does not imply causation. • Least-squares regression models and residual plots allow us to make predictions and evaluate whether a linear relationship is appropriate for bivariate data.
Learning Goal(s): <i>Students will know and will be able to use their learning to:</i> (Content/ Skills)	Content: (Students will know...) • a bivariate quantitative data set consists of observations of ordered pairs from two quantitative variables, collected from the same individuals in a sample or population, and can be used to construct a scatterplot. • a description of the association shown in a scatterplot includes form, direction, strength, and unusual features. • graphical representations between two quantitative variables are best displayed with scatterplots. • correlation r; or the correlation coefficient, measures the direction and strength of the linear association between two quantitative variables. • correlation does not imply causation. • a regression line is a line that models how a response variable y changes as an explanatory variable x changes, often expressed as $\hat{y} = a + bx$. • linear regression is the process of calculating the regression line either by hand or with technology. • a residual is the difference between an actual value of y and the value of y predicted by the regression line. $\text{residual} = \text{actual } y - \text{predicted } y = y - \hat{y}$. • least-squares regression line is the line that makes the sum of the squared residuals as small as possible. • a residual plot is a scatterplot that plots the residuals on the vertical axis and the explanatory variable on the horizontal axis.

- if the residual plot shows only random scatter, the regression model is appropriate.
- if there is a leftover curved pattern in the residual plot, consider using a regression model with a different form.

Skills: (Students will be able to...)

- construct scatterplots depicting the relationship between two quantitative variables.
- describe the characteristics of a scatterplot.
- justify a claim using scatterplots depicting the relationship between two quantitative variables.
- interpret the correlation for a linear relationship.
- calculate a predicted response value using a linear regression model.
- calculate the differences between the observed and predicted values.
- interpret the differences between the observed and predicted values.
- describe the form of association of bivariate data using residual plots.
- calculate the coefficients for the least-squares regression line model.
- interpret coefficients for the least-squares regression line model.