Central York School District

AP Statistics

Overview

Grade Level: 9 – 12

Course Description:
The AP Statistics course introduces students to the major concepts and tools for collecting, analyzing, and drawing conclusions from data. There are four Statistical Practices along with their concomitant skills: formulate questions, collect data, analyze data, and interpret results. The course content includes analyzing one-variable data, collecting data, probability, random variables and probability distributions, statistical inference for proportions and differences in proportions, statistical inference for two-way tables, and statistical inference for proportions and differences in means. Students use technology, investigations, problem solving, and writing as they build conceptual understanding.

Course Big Ideas:

- BIG IDEA 1: VARIATION AND DISTRIBUTION
  The distribution of measures for individuals within a sample or population describes variation. The value of a statistic varies from sample to sample. How can we determine whether differences between measures represent random variation or meaningful distinctions? Statistical methods based on probabilistic reasoning provide the basis for shared understandings about variation and about the likelihood that variation between and among measures, samples, and populations is random or meaningful.
- BIG IDEA 2: PATTERNS AND UNCERTAINTY
  Statistical tools allow us to represent and describe patterns in data and to classify departures from patterns. Simulation and probabilistic reasoning allow us to anticipate patterns in data and to determine the likelihood of errors in inference.
- BIG IDEA 3: DATA-BASED PREDICTIONS, DECISIONS, AND CONCLUSIONS
  Data-based regression models describe relationships between variables and are a tool for making predictions for values of a response variable. Collecting data using random sampling or randomized experimental design means that findings may be generalized to the part of the population from which the selection was made. Statistical inference allows us to make data-based decisions.

Course Essential Questions:

- How can we reason in the face of uncertainty and chance?
- What story does data tell?
- What procedures generate meaningful data?
- How can we anticipate patterns in randomness?
- How can we make inferences about large populations from samples drawn out of those populations?
- When can we conclude that one variable causes changes in another variable?

Course Competencies:

- Write an appropriate investigative question.
- Describe one-variable data numerically and graphically.
- Describe two-variable data numerically and graphically.
- Identify proper and improper methods for selecting a sample.
- Design a proper experiment.
- Compute probabilities of random events using simulation and/or mathematics.
- Describe distributions of random variables.
- Describe how sample statistics vary in random sampling.
- Construct and interpret confidence intervals to estimate population parameters.
- Perform significance tests to test claims about population parameters.

Course Assessments:

- Unit quizzes
- Unit tests
- Informal and formal writing assignments
- Presentations
- Informal check of daily assignments

Central York School District

AP Statistics

Grade Level: 9 – 12

Unit 1 Part 1: Exploring One-Variable Data

Time: 11 days

Standards:

<u>College Board 2026 CED for AP Statistics</u>

- 1.1.A Identify components within a statistical study.
- 1.1.B Determine an investigative question within a statistical study.
- 1.2.A Identify observational units, variables, parameters, and statistics from a statistical study or data set.
- 1.2.B Identify types of variables.
- 1.2.C Identify types of quantitative variables.
- 1.3.A Construct categorical one-variable tabular representations.
- 1.3.B Describe categorical one-variable tabular representations with summary statistics.
- 1.4.A Construct categorical one-variable graphical representations.
- 1.4.B Justify a claim using categorical one-variable graphical representations.
- 1.4.C Compare multiple categorical one-variable tabular and graphical representations.
- 1.5.A Construct quantitative one-variable graphical representations.
- 1.6.A Describe distributions of quantitative one-variable graphical representations.
- 1.6.B Justify a claim using distributions of quantitative one-variable graphical representations.
- 1.7.A Calculate measures of center and position for quantitative data.
- 1.7.B Calculate measures of variability for quantitative data.
- 1.7.C Calculate different units of measurement for summary statistics.
- 1.7.D Calculate outliers for quantitative data.
- 1.7.E Compare multiple quantitative one-variable summary statistics.
- 1.7.F Justify the selection of a summary statistic for describing quantitative data.
- 1.8.A Construct quantitative one-variable graphical representations of summary statistics.
- 1.8.B Describe quantitative one-variable graphical representations of summary statistics based on the relationship of the mean and the median.
- 1.9.A Compare multiple quantitative one-variable graphical representations.
- 1.9.B Compare multiple quantitative one-variable graphical representations of summary statistics.
- 1.9.C Justify a claim using multiple quantitative one-variable graphical representations.
- 1.9.D Calculate *z*-scores with population parameters.
- 1.9.E Compare *z*-scores as measures of relative position for distributions.

Know:

- A statistical study is a study in which data are collected from a sample to answer an investigative question about a larger population.
- A population consists of all items or individuals of interest.
- A sample is a subset of the population from which data are obtained.
- An observational unit is an item or individual from which information is collected.
- A variable is a characteristic that can change from one individual to another.
- Variables can be classified as categorical or quantitative.
- A parameter is a numerical attribute or summary of the variable of interest for a population.
- A statistic is a numerical attribute or summary of the variable of interest for a sample.
- A discrete variable can take on a countable number of values.

- A continuous variable can take on infinitely many values.
- Percentages, relative frequencies, and rates all provide the same information as proportions.
- Outliers for one-variable data are data points that are unusually small or large relative to the rest of the data.
- A distribution is skewed to the right (positive skew) if the right tail is longer than the left.
- A distribution is skewed to the left (negative skew) if the left tail is longer than the right.
- A distribution is symmetric if the left half is the mirror image of the right half.
- One-variable graphs with one main peak are known as unimodal.
- A gap is a region of a distribution between two data values where there are no observed data.
- Clusters are concentrations of data usually separated by gaps.
- The mean is the sum of all the data values divided by the number of values.
- The median of a data set is the middle value when data are ordered.
- The first quartile, Q1, is the median of the half of the ordered data set from the minimum to the position of the median.
- The third quartile, Q3, is the median of the half of the ordered data set from the position of the median to the maximum.
- Range = maximum – minimum.
- The interquartile range (IQR) = Q3 – Q1.
- Standard deviation s is a way to measure variability of the distribution of a quantitative variable.
- The sample variance is the square of the sample standard deviation, s^2.
- If a distribution is relatively symmetric, then the mean and median are relatively close to one another.
- If a distribution is skewed right, then the mean is usually greater than the median.
- If a distribution is skewed left, then the mean is usually less than the median.
- The mean, standard deviation, and range are not resistant.
- The median and IQR are considered resistant.
- The five-number summary is the minimum data value, the first quartile (Q1), the median, the third quartile (Q3), and the maximum data value.
- A boxplot is a graphical representation of the five-number summary.

Understanding/Key Learning:

- Given that variation may be random or not, conclusions are uncertain.
- Graphical representations and numerical summaries allow us to identify and represent key features of data.
- Claims may be justified (or not) by examining data, including graphical representations and summary statistics.

Do:

- Summarize the distribution of data with a frequency table and/or a relative frequency table.
- Display the distribution of categorical data with a bar chart (or bar graph) of frequencies (counts) or relative frequencies.
- Justify claims about data in context using counts and relative frequencies of categorical data.
- Compare two or more data sets in terms of the same categorical variable using frequency tables, bar graphs, or other representations.
- Display the distribution of a quantitative data with a dotplot, a stem and leaf plot, and/or a histogram.
- Describe the distribution of quantitative data: shape, center, and variability (spread), as well as any unusual features.
- Compute the mean and the median of quantitative data to measure the center of the values.
- Compute the range, interquartile range (IQR), and standard deviation of quantitative data to measure the variability of the values.
- Calculate and interpret the percentile for a value.
- Determine how changing units of measurement affects the values of calculated statistics.
- Determine outliers in quantitative data using the $1.5 \times$ IQR rule.
- Determine outliers in quantitative data using the 2 standard deviations rule.
- Display the distribution of quantitative data with a boxplot.

- Summary statistics of quantitative data, or of sets of quantitative data, can be used to justify claims about the data in context.
- Compare two or more independent samples on center, variability, clusters, gaps, outliers, and other features using graphs.
- Compare two or more independent samples using numerical summaries.
- Calculate and interpret a *z*-score (one type of standardized score) for a value in a set of quantitative data.
- Find the proportion of data values located on a given interval of a normally distributed random variable using technology, such as a calculator, a standard normal table, or computer-generated output.
- Use technology to estimate parameters for some populations given the area of a region under the graph of the normal distribution curve.
- Compare relative positions of points within a data set or between data sets using percentiles and/or *z*-scores.

Unit Essential Questions:
- How can I represent data visually and numerically?
- How do I describe distributions of data?
- How can I compare two groups of data for the same variable?
- Where do values fall, relative to other values?

Lesson Essential Questions:
- What are the two types of variables?
- How can we summarize the distribution of a categorical variable?
- How can we visualize the distribution of a categorical variable?
- How can we visualize the distribution of a quantitative variable?
- How can we describe the distribution of a quantitative variable?
- What is the center of the distribution of a quantitative variable?
- How much do the values of a quantitative variable vary?
- How are the distributions of these variables different and the same?
- Where is the relative location of a value in a distribution?

Materials/Resources:
Textbook
Graphing calculator
Online statistical applets (e.g. https://www.stapplet.com/)

Vocabulary:

Data	Distribution	Statistic	Parameter
Observational	Bar graph	Mean	Standardized score
UnitVariability	Dotplot	Median	*Z*-score
Categorical variable	Histogram	Standard deviation	
Quantitative variable	Stemplot	Range	
Discrete variable	Boxplot	Interquartile range (IQR)	
Continuous variable	Skewed left	Outliers	
	Skewed right		
	Symmetric		
	Unimodal		

Assessments:
Unit quizzes
Unit tests
Informal and formal writing assignments

Presentations
Informal check of daily assignments
Progress check (College Board)

Central York School District

AP Statistics

Grade Level: 9 – 12

Unit 1 Part 2: Collecting Data

Time: 8 days

Standards:
<u>College Board 2026 CED for AP Statistics</u> • 1.10.A Determine the components of an investigative question within a statistical study. • 1.10.B Identify a census. • 1.10.C Identify an experiment. • 1.10.D Identify an observational study. • 1.10.E Justify the appropriateness of generalizations for a statistical study. • 1.11.A Identify a sampling method given a description of a study. • 1.11.B Justify the appropriateness of a sampling method. • 1.12.A Identify potential sources of bias in sampling methods. • 1.13.A Identify elements of a well-designed experiment. • 1.13.B Identify experimental designs. • 1.13.C Justify the appropriateness of a particular experimental design. • 1.13.D Justify the appropriateness of the conclusions based on a well-designed experiment.
Know:
• An investigative question should be posed so that the required data can be collected and analyzed. • Methods for data collection that do not rely on chance result in untrustworthy conclusions. • A population consists of all items or subjects of interest. • A sample selected for study is a subset of the population. • In an observational study, treatments are not imposed. • In a retrospective study, investigators examine data for a sample of individuals. • In a prospective study, investigators follow a sample of individuals into the future, collecting data. • A sample survey collects data from a sample in an attempt to learn about the population from which the sample was taken. • In an experiment, different conditions (treatments) are assigned to experimental units (participants or subjects). • It is only appropriate to make generalizations about a population based on samples that are randomly selected or otherwise representative of that population. • A sample is only generalizable to the population from which the sample was selected. • It is not possible to determine causal relationships between variables using data collected in an observational study. • When an item from a population can be selected only once, this is called sampling without replacement. • When an item from the population can be selected more than once, this is called sampling with replacement. • A simple random sample (SRS) is a sample in which every group of a given size has an equal chance of being chosen. • A stratified random sample involves the division of a population into separate groups, called strata, based on shared attributes or characteristics (homogeneous grouping). • A cluster sample involves the division of a population into smaller groups, called clusters. • A systematic random sample is a method in which sample members from a population are selected according to a random starting point and a fixed, periodic interval.

- A census selects all items/subjects in a population.
- There are advantages and disadvantages for each sampling method depending upon the question that is to be answered and the population from which the sample will be drawn.
- Bias occurs when certain responses are systematically favored over others.
- Voluntary response bias may occur when a sample consists of volunteers or people who choose to participate.
- Undercoverage bias may occur when part of the population has a reduced chance of being included in the sample.
- Nonresponse bias may occur when individuals chosen for the sample for whom data cannot be obtained differ from those for whom data can be obtained.
- Problems in the data gathering instrument or process result in response bias.
- Non-random sampling methods introduce potential for bias because they do not use chance to select the individuals.
- The experimental units are the individuals that are assigned treatments.
- An explanatory variable (or factor) in an experiment is a variable whose levels are
- manipulated intentionally.
- The levels or combination of levels of the explanatory variable(s) are called treatments.
- A response variable in an experiment is an outcome from the experimental units that is measured after the treatments have been administered.
- A confounding variable in an experiment is a variable that is related to the explanatory variable and influences the response variable.
- A confounding variable may create a false perception of association between explanatory and response variables.
- A well-designed experiment should include the following: comparison, random assignment/allocation of treatments, replication, and control.
- There are advantages and disadvantages for each experimental design.
- In a completely randomized design, treatments are assigned to experimental units completely at random.
- Random assignment tends to balance the effects of uncontrolled (confounding) variables so that differences in responses can be attributed to the treatments.
- In a single-blind experiment, subjects do not know which treatment they are receiving, but members of the research team do, or vice versa.
- In a double-blind experiment neither the subjects nor the members of the research team who interact with them know which treatment a subject is receiving.
- A control group is a collection of experimental units either not given a treatment of interest or given a treatment with an inactive substance (placebo).
- The placebo effect occurs when experimental units have a response to a placebo.
- For randomized block designs, treatments are assigned completely at random within each block.
- Blocking ensures that at the beginning of the experiment the units within each block are similar to each other with respect to at least one blocking variable.
- A randomized block design helps to separate natural variability from differences due to the blocking variable.
- A matched pairs design is a special case of a randomized block design.
- There are advantages and disadvantages for each experimental design.
- Statistical inference attributes conclusions based on data to the distribution from which the data were collected.
- Statistically significant differences between or among experimental treatment groups are evidence that the treatments caused the effect.
- Random assignment of treatments to experimental units allows researchers to conclude that some observed changes are so large as to be unlikely to have occurred by chance.
- Statistically significant differences between or among experimental treatment groups are evidence that the treatments caused the effect.
- If the experimental units used in an experiment are representative of some larger group of units, the results of an experiment can be generalized to the larger group.

Understanding/Key Learning:
• Given that variation may be random or not, conclusions are uncertain. • Statistical investigative questions establish the purpose of research and direct the way in which data are collected. • The way we collect data influences what we can and cannot say about a population. • Well-designed experiments can establish evidence of causal relationships.
Do:
• Identify the three components of a statistical investigative question. • Determine if a statistical study is an observational study or an experiment. • Classify observational studies. • Describe how to select a simple random sample. • Describe how to select a stratified random sample. • Describe how to select a cluster random sample. • Describe how to select a systematic random sample. • Determine the largest population to which results in a sample can be safely generalized. • Explain the problems in poor data collection protocols. • Identify the experimental units, the explanatory variable(s), the response variable, and the treatments in an experiment. • Determine the composition of the treatments in a multi-factor experiment. • Describe a method to randomly allocate/assign treatments to experimental units. • Design a completely randomized experiment. • Design a randomized block experiment. • Design a matched pairs experiment. • Identify blinding in an experiment. • Describe the placebo effect in an experiment. • Explain the purpose of blocking in an experiment. • Explain the purpose of pairing in an experiment. • Determine if differences in groups are statistically significant using a simulation. • Determine if cause and effect inferences are justified based on the results and methodology of an experiment.

Unit Essential Questions:
- How do I state investigative questions that direct research methods?
- What are poor data collection practices?
- Why is randomness important in data collection?
- How does the design of a statistical study relate to the inferences that can be drawn from the study?

Lesson Essential Questions:
- What are the components of an investigative question?
- Which type of statistical study is being used?
- How is a representative sample obtained?
- What problems can occur when sampling?
- What are the principles of good experiments?
- How are experiments designed?
- What inferences can be drawn from statistical studies?

Materials/Resources:
Textbook
Graphing calculator
Online statistical applets (e.g. https://www.stapplet.com/)

Vocabulary:

Population
Sample
Parameter
Statistic
Statistical inference
Observational study
Prospective
Retrospective
Experiment

Census
Sampling variability
Bias
Simple random sample
Stratified random sample
Cluster random sample
Systematic sample
Unbiased estimators
Convenience sample
Voluntary response sample
Undercoverage bias
Nonresponse
Response bias

Experimental units
Factor
Level
Treatment
Confounding
Extraneous variable
Random assignment
Comparison
Replication
Control
Control group
Placebo
Placebo effect

Blinding
Completely randomized experiment
Blocking
Randomized block experiment
Matched pairs experiment
Statistically significant

Assessments:

Unit quizzes
Unit tests
Informal and formal writing assignments
Presentations
Informal check of daily assignments
Progress check (College Board)

Central York School District

AP Statistics

Grade Level: 9 – 12

Unit 2 Part 1: Probability

Time: 8 days

Standards:

College Board 2026 CED for AP Statistics

- 2.1.A Compare tabular and graphical representations for the relationship between two categorical variables.
- 2.1.B Justify a claim using tabular and graphical representations for the distributions of two categorical variables.
- 2.2.A Calculate summary statistics from two-way tables.
- 2.2.B Compare summary statistics for two categorical variables.
- 2.2.C Justify a claim using summary statistics for two categorical variables.
- 2.3.A Estimate probabilities using simulations.
- 2.4.A Calculate probabilities for events and their complements.
- 2.5.A Justify why two events are mutually exclusive (or disjoint) using joint probability.
- 2.6.A Calculate conditional probabilities.
- 2.7.A Calculate probabilities for independent events and for the union of two events.

Know:

- Determine if two categorical variables are associated with graphical representations of two categorical variables.
- In a two-way table, a joint relative frequency is a cell frequency divided by the total for the entire table.
- Summary statistics for two categorical variables can be used to compare distributions and/or determine if variables are associated.
- Patterns in data do not necessarily mean that variation is not random.
- A random process generates results that are determined by chance.
- An outcome is the result of a trial of a random process.
- An event is a collection of outcomes.
- Simulation is a way to model random events, such that simulated outcomes closely match real-world outcomes.
- The relative frequency of an outcome or event in simulated or empirical data can be used to estimate the probability of that outcome or event.
- The law of large numbers states that simulated (empirical) probabilities tend to get closer to the true probability as the number of trials increases.
- The sample space of a random process is the set of all possible non-overlapping outcomes.
- The probability of an event is a number between 0 and 1, inclusive.
- Probabilities of events in repeatable situations can be interpreted as the relative frequency with which the event will occur in the long run.
- The probability that events A and B both will occur, sometimes called the joint probability, is the probability of the intersection of A and B, denoted $P(A \cap B)$.
- Two events are mutually exclusive or disjoint if they cannot occur at the same time. In other words, $P(A \cap B) = 0$.
- The probability that event A will occur given that event B has occurred is called a conditional probability and denoted $P(A \mid B) = P(A \cap B)/P(B)$.
- The multiplication rule states that the probability that events A and B both will occur is $P(A \cap B) = P(A) \cdot P(B \mid A)$.

- 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.
- For independent events, $P(A \cap B) = P(A) \cdot P(B)$.
- The probability that event A or event B (or both) will occur is the probability of the union of A and B, denoted $P(A \cup B)$.
- The addition rule states that the probability that event A or event B or both will occur is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Understanding/Key Learning:

- Information about two categorical variables can be summarized with a two-way table.
- Claims can be justified (or not) by analyzing data.
- Given that variation may be random or not, conclusions are uncertain.
- Simulation allows us to anticipate patterns in data.
- The likelihood of a random event can be quantified.
- Probability distributions may be used to model variation in populations.
- Probabilistic reasoning allows us to anticipate patterns in data.

Do:

- Display the relationship between two categorical variables with side-by-side bar graphs.
- Display the relationship between two categorical variables with segmented bar graphs.
- Display the relationship between two categorical variables with mosaic plots.
- Organize data from two categorical variables with a two-way table.
- Compute marginal relative frequencies in two-way tables.
- Compute conditional relative frequencies in two-way tables.
- Calculate the probability of an event given equally likely outcomes.
- Find the probability of the complement of an event E.
- Calculate the joint probability of the event (A and B) for independent events.
- Calculate the conditional probability of the event $(A \mid B)$.
- Calculate the joint probability of the event (A and B) for dependent events.
- Justify whether two events are independent.
- Calculate the probability of the event (A or B).

Unit Essential Questions:

- Is there an association between two categorical variables?
- How certain can we be about our conclusions?
- How can simulation be used to model chance in our world?
- How can mathematics be used to model chance in our world?

Lesson Essential Questions:

- How are data from two categorical variables organized?
- Is there an association between two categorical variables?
- What does "random" mean?
- How is chance quantified?
- What are mutually exclusive events?
- What is the probability of an event occurring when it depends on the outcome of another event?
- What is the probability of independent events and unions of events?

Materials/Resources:

Textbook
Graphing calculator
Online statistical applets (e.g. https://www.stapplet.com/)

Vocabulary:

Explanatory variable
Response variable
Two-way table (contingency table)
Marginal distribution
Conditional distribution
Association
Statistical tendency

Random process
Outcome
Event
Relative frequency
Law of large numbers
Probability
Simulation
Complement
Sample space

Mutually exclusive
Disjoint
Joint probability
Intersection
Conditional probability
The multiplication rule
The addition rule
Union
Independent events

Assessments:

Unit quizzes
Unit tests
Informal and formal writing assignments
Presentations
Informal check of daily assignments
Progress check (College Board)

Central York School District

AP Statistics

Grade Level: 9 – 12

Unit 2 Part 2: Random Variables and Probability Distributions

Time: 10 days

Standards:

College Board 2026 CED for AP Statistics

- 2.8.A Construct a probability distribution for a discrete random variable.
- 2.9.A Calculate the mean and standard deviation for a discrete random variable.
- 2.9.B Interpret the mean and standard deviation for a discrete random variable.
- 2.10.A Justify why a random variable is or is not a binomial random variable.
- 2.10.B Calculate the mean and standard deviation for a binomial distribution.
- 2.10.C Interpret the mean, standard deviation, and probabilities for a binomial distribution.
- 2.10.D Estimate probabilities of binomial random variables using data from a simulation.
- 2.10.E Calculate probabilities for a binomial distribution.
- 2.11.A Describe a normal distribution.
- 2.11.B Calculate the mean and standard deviation for a normal distribution.
- 2.11.C Calculate percentages from a normal distribution using the empirical rule.
- 2.11.D Calculate the probability that a particular value lies within a given interval of a normal distribution.
- 2.11.E Calculate the associated intervals and areas of a normal distribution.
- 2.11.F Compare measures of relative position for distributions.
- 2.12.A Describe sampling distributions with simulations.

Know:

- The values of a random variable are the numerical outcomes of random behavior.
- A discrete random variable is a variable that can only take a countable number of values.
- The sum of the probabilities over all of the possible values of a random variable must be 1.
- A probability distribution can be represented as a graph, table, or function showing the probabilities associated with values of a random variable.
- A cumulative probability distribution can be represented as a table or function showing the probability of being less than or equal to each value of the random variable.
- An interpretation of a probability distribution provides information about the shape, center, and spread of a population.
- A numerical value measuring a characteristic of a population or the distribution of a random variable is known as a parameter, which is a single, fixed value.
- Parameters for a discrete random variable should be interpreted using appropriate units and within the context of a specific population.
- A probability distribution can be constructed using the rules of probability or estimated with a simulation using random number generators.
- Two random variables are independent if knowing information about one of them does not change the probability distribution of the other.
- A binomial random variable, X, counts the number of successes in n repeated independent trials, each trial having two possible outcomes (success or failure), with the probability of success p and the probability of failure 1 – p.

- The binomial probability function is $P(X = x) = {}_nC_x(p)^x(1 - p)^{n-x}$.
- If a random variable is binomial, its mean $\mu_x = np$ and its standard deviation $\sigma_x = \sqrt{np(1 - p)}$.
- Probabilities and parameters for a binomial distribution should be interpreted using appropriate units and within the context of a specific population or situation.
- A continuous random variable is a variable that can take on any value within a specified domain. Every interval within the domain has a probability associated with it.
- 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 parameters of a normal distribution are the population mean, μ, and the population standard deviation, σ.
- A standard normal distribution is a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.
- The empirical rule for a normal distribution states that in an approximately normal distribution, about 68% of values are within 1 standard deviation of the mean, about 95% of values are within 2 standard deviations of the mean, and about 99.7% of values are within 2 standard deviations of the mean.
- The area under a normal curve over a given interval represents the probability that a particular value lies in that interval.
- The boundaries of an interval associated with a given area in a normal distribution can be determined using z-scores or technology.
- Intervals associated with a given area in a normal distribution can be determined by assigning appropriate inequalities to the boundaries of the intervals.
- Normal distributions can be used to approximate distributions with similar characteristics.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 sampling distribution of a statistic can be simulated by repeatedly generating a large number of random samples from the population assuming known value(s) for the parameter(s).
- A randomization distribution is the distribution of a statistic generated by simulation from repeatedly randomly reallocating, or reassigning, the response values to treatment groups.
- 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.

Understanding/Key Learning:

- Given that variation may be random or not, conclusions are uncertain.
- Simulation allows us to anticipate patterns in data.
- The likelihood of a random event can be quantified.
- Probability distributions may be used to model variation in populations.
- Probabilistic reasoning allows us to anticipate patterns in data.

Do:

- Calculate the probability of an event given equally likely outcomes.
- Find the probability of the complement of an event E.
- Calculate the joint probability of the event (A and B) for independent events.
- Calculate the conditional probability of the event (A | B).
- Calculate the joint probability of the event (A and B) for dependent events.
- Justify whether two events are independent.
- Calculate the probability of the event (A or B).
- Represent the probability distribution of a discrete random variable in a table.
- Represent the probability distribution of a discrete random variable with a histogram.
- Find the probability of a value of a random variable.
- Find the probability of an event defined for a random variable.
- Compute and interpret the mean of a discrete random variable.
- Compute and interpret the standard deviation of a discrete random variable.
- Find and interpret the mean of a linear combination of random variables.

- Find and interpret the standard deviation of a linear combination of independent random variables.
- Find and interpret the mean of a linear transformation of a random variable.
- Find and interpret the standard deviation of a linear transformation of a random variable.
- Justify whether a random variable is a binomial random variable.
- Calculate the probability of an event for a binomial random variable.
- Calculate and interpret the mean of a binomial random variable.
- Calculate and interpret the standard deviation of a binomial random variable.
- Calculate the probability of an event for a normal random variable.
- Calculate the mean or standard deviation of a normal random variable, given other information about the distribution.
- Simulate the sampling distribution of a sample statistic.

Unit Essential Questions:
- How certain can we be about our conclusions?
- How can simulation be used to model chance in our world?
- How can mathematics be used to model chance in our world?

Lesson Essential Questions:
- What does "random" mean?
- How is chance quantified?
- What are mutually exclusive events?
- What is the probability of an event occurring when it depends on the outcome of another event?
- What is the probability of independent events and unions of events?
- What is a random variable?
- How much do the values of random variables vary?
- How can random variables be combined?
- How can likelihoods be quantified in binomial settings?
- How can likelihoods be quantified by normal random variables?
- How can we simulate statistics generated by random samples?

Materials/Resources:
Textbook
Graphing calculator
Online statistical applets (e.g. https://www.stapplet.com/)

Vocabulary:

Random variable
Discrete random variable
Probability distribution
Cumulative probability distribution
Parameter
Mean of a random variable
Standard deviation of a random variable
Binomial random variable
Normal random variable
Sampling distribution
Randomization distribution
Central Limit Theorem

Assessments:
Unit quizzes
Unit tests
Informal and formal writing assignments
Presentations

Informal check of daily assignments
Progress check (College Board)

Central York School District

AP Statistics

Grade Level: 9 – 12

Unit 3 Part 1: Inference for One Proportion

Time: 9 days

Standards:

College Board 2026 CED for AP Statistics

- 3.1.A Justify why an estimator is or is not unbiased.
- 3.1.B Calculate estimates for a population parameter.
- 3.2.A Calculate the mean and standard deviation of a sampling distribution for a sample proportion.
- 3.2.B Justify the appropriateness of conditions for the sampling distribution of a sample proportion.
- 3.2.C Interpret the mean, standard deviation, and probabilities for a sampling distribution of a sample proportion.
- 3.3.A Identify an appropriate confidence interval procedure including the parameter for a population proportion.
- 3.3.B Justify the appropriateness of constructing a confidence interval for a population proportion by verifying conditions.
- 3.3.C Calculate an appropriate confidence interval for a population proportion.
- 3.3.D 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.
- 3.4.A Interpret a confidence interval in context for a population proportion.
- 3.4.B Justify a claim based on a confidence interval for a population proportion.
- 3.4.C Identify the relationships among sample size, confidence interval width, confidence level, and margin of error for a population proportion.
- 3.5.A Identify an appropriate testing method for a population proportion including the parameter for the population proportion.
- 3.5.B Identify the null and alternative hypotheses for a population proportion.
- 3.5.C Justify the appropriateness of a hypothesis test for a population proportion by verifying conditions.
- 3.6.A Interpret the p-value of a hypothesis test for a population proportion.
- 3.7.A Calculate an appropriate test statistic and p-value for testing a hypothesis about a population proportion.
- 3.7.B Justify a claim about the population based on the results of a hypothesis test for a population proportion.
- 3.8.A Identify Type I and Type II errors.
- 3.8.B Calculate the probability of Type I and Type II errors.
- 3.8.C Identify the factors that affect the probability of errors in hypothesis testing.
- 3.8.D Interpret Type I and Type II errors.

Know:

- A sampling distribution of a statistic is the distribution of values for the statistic for all possible samples of a given size from a given population.
- A randomization distribution is a collection of statistics generated by simulation assuming known values for the parameters.
- The sampling distribution of a statistic can be simulated by generating repeated random samples from a population.
- When estimating a population parameter, an estimator is unbiased if, on average, the value of the estimator is equal to the population parameter.

- When estimating a population parameter, an estimator exhibits variability that can be modeled using probability.
- A sample statistic is a point estimator of the corresponding population parameter.
- The sampling distribution of the sample proportion, $\hat{p}$, has a mean p and a standard deviation $\sqrt{\frac{p(1-p)}{n}}$.
- If sampling without replacement, the standard deviation of the sample proportion is smaller than what is given by the formula above.
- If the sample size is less than 10% of the population size, the difference between the formula and the actual standard deviation is negligible.
- The sampling distribution of the sample proportion, $\hat{p}$, will have an approximate normal distribution, provided $np \geq 10$ and $n(1-p) \geq 10$.
- Probabilities and parameters for a sampling distribution for a sample proportion should be interpreted using appropriate units and within the context of a specific population.
- The appropriate confidence interval procedure for a one-sample proportion for one categorical variable is a one sample z-interval for a proportion.
- In order to make assumptions necessary for inference on population proportions, means, and slopes, we must check for independence in data collection methods and for selection of the appropriate sampling distribution.
- The standard error of $\hat{p}$ is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
- A margin of error gives how much a value of a sample statistic is likely to vary from the value of the corresponding population parameter.
- The margin of error of $\hat{p}$ is $z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
- The formula for margin of error can be rearranged to solve for n, the minimum sample size needed to achieve a given margin of error.
- In general, an interval estimate can be constructed as a point estimate ± (margin of error).
- For a one sample z-interval for a proportion, the interval estimate is $\hat{p} \pm z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
- Critical values represent the boundaries encompassing the middle C% of the standard normal distribution, where C% is an approximate confidence level for a proportion.
- Confidence intervals for population proportions can be used to calculate interval estimates with specified units.
- Confidence intervals are based on data from random samples, which vary from sample to sample.
- A confidence interval for a population proportion either contains the population proportion or it does not.
- We are C% confident that the confidence interval for a population proportion captures the population proportion.
- In repeated random sampling with the same sample size, approximately C% of confidence intervals created will capture the population proportion.
- Interpreting a confidence interval for a one-sample proportion should include a reference to the sample taken and details about the population it represents.
- When all other things remain the same, the width of the confidence interval for a population proportion tends to decrease as the sample size increases.
- For a population proportion, the width of the confidence interval is proportional to $1/\sqrt{n}$.
- For a given sample, the width of the confidence interval for a population proportion increases as the confidence level increases.
- The width of a confidence interval for a population proportion is exactly twice the margin of error.
- The null hypothesis is the situation that is assumed to be correct unless evidence suggests otherwise, and the alternative hypothesis is the situation for which evidence is being collected.
- For a one-sample z-test for a population proportion, the null hypothesis specifies a value for the population proportion, usually one indicating no difference or effect.
- For a single categorical variable, the appropriate testing method for a population proportion is a one-sample z-test for a population proportion.

- The distribution of the test statistic assuming the null hypothesis is true (null distribution) can be either a randomization distribution or when a probability model is assumed to be true, a theoretical distribution (z).
- When using a z-test, the standardized test statistic can be written as
$$\textit{standardized test statistic} = \frac{\textit{sample statistic}-\textit{null value of the parameter}}{\textit{standard deviation of the statistic}}.$$
- The test statistic for a population proportion is $z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$.
- A p-value is the probability of obtaining a test statistic as extreme or more extreme than the observed test statistic when the null hypothesis and probability model are assumed to be true.
- The p-value is the proportion of values for the null distribution that are as extreme or more extreme than the observed value of the test statistic.
- An interpretation of the p-value of a significance test for a one-sample proportion should recognize that the p-value is computed by assuming that the true population proportion is equal to the particular value stated in the null hypothesis.
- The significance level a is the predetermined probability of rejecting the null hypothesis given that it is true.
- A formal decision explicitly compares the p-value to the significance level a.
- If the p-value ≤ a, reject the null hypothesis. If the p-value > a, fail to reject the null hypothesis.
- Rejecting the null hypothesis means there is sufficient statistical evidence to support the alternative hypothesis.
- Failing to reject the null means there is insufficient statistical evidence to support the alternative hypothesis.
- The conclusion about the alternative hypothesis must be stated in context.
- A significance test can lead to rejecting or not rejecting the null hypothesis, but can never lead to concluding or proving that the null hypothesis is true.
- Small p-values indicate that the observed value of the test statistic would be unusual if the null hypothesis and probability model were true.
- The lower the p-value, the more convincing the statistical evidence for the alternative hypothesis.
- *p*-values that are not small indicate that the observed value of the test statistic would not be unusual if the null hypothesis and probability model were true.
- *p*-values that are not small do not provide evidence that the null hypothesis is true.
- A Type I error occurs when the null hypothesis is true and is rejected (false positive).
- A Type II error occurs when the null hypothesis is false and is not rejected (false negative).
- The significance level a is the probability of making a Type I error, if the null hypothesis is true.
- The power of a test is the probability that a test will correctly reject a false null hypothesis.
- The probability of making a Type II error = 1− power.
- Whether a Type I or a Type II error is more consequential depends upon the situation.
- Since the significance level a is the probability of a Type I error, the consequences of a Type I error influence decisions about a significance level.

Understanding/Key Learning:

- Given that variation may be random or not, conclusions are uncertain.
- An interval of values should be used to estimate parameters, in order to account for uncertainty.
- The normal distribution may be used to model variation in sample statistics.
- Significance testing allows us to make decisions about hypotheses within a particular context.
- Probabilities of Type I and Type II errors influence inference.

Do:

- Determine whether an estimator of a population parameter is biased or unbiased.
- Verify whether the conditions for the sampling distribution of the sample proportion $\hat{p}$ have been met.
- Compute the mean and standard deviation of the sampling distribution of the sample proportion $\hat{p}$.
- Compute probabilities of events given the sampling distribution of a sample proportion $\hat{p}$.

- Check for independence (random and 10% conditions) and that the sampling distribution is approximately normal (Large Counts condition) before constructing a confidence interval to estimate a population proportion *p*.
- Compute the margin of error for a population proportion *p*.
- Calculate the minimum sample size required to estimate *p* given a required margin of error.
- Check for independence (random and 10% conditions) and that the sampling distribution is approximately normal (Large Counts condition) before constructing a confidence interval for a population proportion *p*.
- Calculate and interpret a one-sample *z* interval for *p*.
- Interpret the confidence level of a confidence interval.
- Evaluate a claim about a parameter using a confidence interval.
- Describe how the margin of error relates to sample size and confidence level.
- State hypotheses for a significance test about a population proportion.
- Check for independence (random and 10% conditions) and that the sampling distribution is approximately normal (Large Counts condition) before performing a significance test about a population proportion *p*.
- Calculate the standardized test statistic and p-value for a one-sample *z* test for *p*.
- State a conclusion for a one-sample *z* test for *p*.
- Interpret the p-value for a one-sample *z* test for *p*.
- Describe a Type I error and give a possible consequence of such an error.
- Describe a Type II error and give a possible consequence of such an error.
- Explain how sample size, significance level, standard error, and the true (alternative) value of the parameter influence the power of a test.

Unit Essential Questions:

- What is an unusual event when sampling from a population?
- What patterns occur when sampling from a population when measuring a categorical variable?
- How is an unknown population parameter estimated?
- How is a claim about an unknown population parameter tested?

Lesson Essential Questions:

- What happens to sampling distributions when sample size increases?
- What is an unbiased estimator?
- What is the sampling distribution of a sample proportion?
- How can a population proportion be estimated?
- What affects margin of error?
- How is a claim about a population proportion tested?
- What does a *P*-value mean?
- What is the conclusion for a test about a population proportion?
- What errors could occur when testing a population proportion?

Materials/Resources:

Textbook
Graphing calculator
Online statistical applets (e.g. https://www.stapplet.com/)

Vocabulary:

Sampling distribution
Randomization distribution
Biased estimator
Unbiased estimator
Point estimate
Point estimator

Confidence level
Critical value
Margin of error
One-sample z interval for *p*

Null hypothesis
Alternative hypothesis
One-sided hypothesis
Two-sided hypothesis
One-sample *z* test for *p*
Standardized test statistic

Type I error
Type II error
Power of a test

Standard error
Sampling distribution of the sample proportion
Random condition
10% condition
Large Counts condition

P-value
Significance level

Assessments:
Unit quizzes
Unit tests
Informal and formal writing assignments
Presentations
Informal check of daily assignments
Progress check (College Board)

Central York School District

AP Statistics

Grade Level: 9 – 12

Unit 3 Part 2: Inference for Two Proportions and Two-way Tables

Time: 8 days

Standards:

<u>**College Board 2026 CED for AP Statistics**</u>

- 3.9.A Calculate the mean and standard deviation of the sampling distribution for the difference between two sample proportions.
- 3.9.B Justify the appropriateness of conditions for the sampling distribution for the difference between two sample proportions.
- 3.9.C Interpret the mean, standard deviation, and probabilities for the sampling distribution for the difference between two sample proportions.
- 3.10.A Identify an appropriate confidence interval procedure including the parameters for the difference between two population proportions.
- 3.10.B Justify the appropriateness of constructing a confidence interval for the difference between two population proportions by verifying conditions.
- 3.10.C Calculate an appropriate confidence interval for the difference between two population proportions.
- 3.10.D Calculate the standard error and margin of error for estimating the difference between two population proportions.
- 3.11.A Interpret a confidence interval in context for the difference between two population proportions.
- 3.11.B Justify a claim based on a confidence interval for the difference between two population proportions.
- 3.12.A Identify an appropriate testing method for the difference between population proportions including the parameters.
- 3.12.B Identify the null and alternative hypotheses for the difference between population proportions.
- 3.12.C Justify the appropriateness of a hypothesis test for the difference between two population proportions by verifying conditions.
- 3.13.A Calculate an appropriate test statistic and p-value for testing a hypothesis for the difference between two population proportions.
- 3.13.B Interpret the p-value of a hypothesis test for the difference between two population proportions.
- 3.13.C Justify a claim about the populations based on the results of a hypothesis test for the difference between two population proportions.
- 3.14.A Describe chi-square distributions.
- 3.14.B Identify an appropriate testing method for comparing distributions in two-way tables of categorical data including the populations and variables.
- 3.14.C Identify the null and alternative hypotheses for a chi-square test for homogeneity or independence.
- 3.14.D Justify the appropriateness of a chi-square test for independence or homogeneity by verifying conditions.
- 3.15.A Calculate expected counts for two-way tables of categorical data.
- 3.15.B Calculate the appropriate test statistic and p-value for a chi-square test for homogeneity or independence.
- 3.15.C Interpret the p-value for the chi-square test for homogeneity or independence.
- 3.15.D Justify a claim about the population(s) based on the results of a chi-square test for homogeneity or independence.

Know:

- The sampling distribution of the difference in sample proportions $\hat{p}_1 - \hat{p}_2$ has mean $p_1 - p_2$ and standard deviation $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$
- If sampling without replacement, the standard deviation of the difference in proportions is smaller than what is given by the formula above.
- If the sample size is less than 10% of the population size, the difference between the formula and the actual standard deviation is negligible.
- The sampling distribution of the difference in proportions will have an approximate normal distribution, provided $n_1p_1 \geq 10$ and $n_1(1-p_1) \geq 10$ and $n_2p_2 \geq 10$ and $n_2(1-p_2) \geq 10$.
- Parameters for a sampling distribution for a difference of proportions should be interpreted using appropriate units and within the context of specific populations.
- The appropriate confidence interval procedure for a two-sample comparison of proportions for one categorical variable is a two-sample z-interval for a difference between population proportions.
- In order to calculate confidence intervals to estimate a difference between proportions, we must check for independence and that the sampling distribution is approximately normal.
- For a comparison of proportions, the interval estimate is $\hat{p}_1 - \hat{p}_2 \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
- Confidence intervals for a difference in proportions can be used to calculate interval estimates with specified units.
- In repeated random sampling with the same sample size, approximately C% of confidence intervals created will capture the difference in population proportions.
- A confidence interval for difference in population proportions provides an interval of values that may provide sufficient evidence to support a particular claim in context.
- For a single categorical variable, the appropriate testing method for the difference of two population proportions is a two-sample z-test for a difference between two population proportions.
- In order to make statistical inferences when testing a difference between population proportions, we must check for independence and that the sampling distribution is approximately normal.
- The test statistic for a difference in proportions is $z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where $\hat{p}_c = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$.
- An interpretation of the p-value of a significance test for a difference of two population proportions should recognize that the p-value is computed by assuming that the null hypothesis is true.
- The results of a significance test for a difference of two population proportions can serve as the statistical reasoning to support the answer to an investigative question about the two populations that were sampled.
- Variation between what we find and what we expect to find may be random or not.
- Expected counts of categorical data are counts consistent with the null hypothesis. In general, an expected count is a sample size times a probability.
- The chi-square statistic measures the distance between observed and expected counts relative to expected counts.
- Chi-square distributions have positive values and are skewed right. Within the family of chi-square density curves, the skewness becomes less pronounced with increasing degrees of freedom.
- The distribution of a test statistic assuming the null hypothesis is true (null distribution) can be either a randomization distribution or, when a probability model is assumed to be true, a theoretical distribution (chi-square).
- The expected count in a particular cell of a two-way table of categorical data can be calculated using the formula $expected\ count = \frac{(row\ total)(column\ total)}{table\ total}$.
- The appropriate hypotheses for a chi-square test for homogeneity are:
  H_0: There is no difference in distributions of a categorical variable across populations or treatments.

H_a: There is a difference in distributions of a categorical variable across populations or treatments.

- The appropriate hypotheses for a chi-square test for independence are:
  H_0: There is no association between two categorical variables in a given population or the two categorical variables are independent.
  H_a: Two categorical variables in a population are associated or dependent.
- When comparing distributions to determine whether proportions in each category for categorical data collected from different populations are the same, 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 population from which the data were sampled, the appropriate test is the chi-square test for independence.
- In order to make statistical inferences for a chi-square test for homogeneity or independence we must check for independence (Random and 10% condition) and that the sampling distribution has an approximately Chi-square distribution (Large Counts condition).
- The appropriate test statistic for a chi-square test for homogeneity or independence is the chi-square statistic $\chi^2 = \Sigma \frac{(observed\ count - expected\ count)^2}{expected\ count}$ with degrees of freedom = (number of rows – 1) $\times$ (number of columns – 1).
- The p-value for a chi-square test for independence or homogeneity for a number of degrees of freedom is found using the appropriate table or technology.
- For a test of independence or homogeneity for a two-way table, the p-value is the proportion of values in a chi-square distribution with appropriate degrees of freedom that are equal to or larger than the test statistic.
- An interpretation of the p-value for the chi-square test for homogeneity or independence is the probability, given the null hypothesis and probability model are true, of obtaining a test statistic as, or more, extreme than the observed value.
- A decision to either reject or fail to reject the null hypothesis for a chi-square test for homogeneity or independence is based on comparison of the p-value to the significance level a.
- The results of a chi-square test for homogeneity or independence can serve as the statistical reasoning to support the answer to a research question about the population that was sampled (independence) or the populations that were sampled (homogeneity).

Understanding/Key Learning:

- Given that variation may be random or not, conclusions are uncertain.
- An interval of values should be used to estimate parameters, in order to account for uncertainty.
- The normal distribution may be used to model variation in sample statistics.
- Significance testing allows us to make decisions about hypotheses within a particular context.
- Probabilities of Type I and Type II errors influence inference.
- The chi-square distribution may be used to model variation.

Do:

- Verify whether the conditions for the sampling distribution of a difference in proportions $\hat{p}_1 - \hat{p}_2$ have been met.
- Compute the mean and standard deviation of the sampling distribution of a difference in proportions $\hat{p}_1 - \hat{p}_2$.
- Compute probabilities of events given the sampling distribution of a difference in proportions $\hat{p}_1 - \hat{p}_2$.
- Check for independence (random and 10% conditions) and that the sampling distribution is approximately normal (Large Counts condition) before constructing a confidence interval for a difference in two population proportions $p_1 - p_2$.
- Calculate and interpret a two-sample z interval for $p_1 - p_2$.
- State hypotheses for a significance test about a difference in two population proportions $p_1 - p_2$.

- Check for independence (random and 10% conditions) and that the sampling distribution is approximately normal (Large Counts condition) before performing a significance test about a difference in two population proportions $p_1 - p_2$.
- Calculate the standardized test statistic and p-value for a two-sample z test for $p_1 - p_2$.
- State a conclusion for a two-sample z test for $p_1 - p_2$.
- Interpret the p-value of a significance test.
- Describe a Type I error and give a possible consequence of such an error.
- Describe a Type II error and give a possible consequence of such an error.
- Explain how sample size, significance level, standard error, and the true (alternative) value of the parameter influence the power of a test.
- State hypotheses for a chi-square test for independence.
- Compute expected counts for a two-way table given a null hypothesis.
- Check for independence (random and 10% conditions) and that the sampling distribution is approximately chi-square before performing a chi-square test for independence.
- Calculate the χ^2 test statistic and p-value for a chi-square test for independence.
- State a conclusion for a chi-square test for independence.
- State hypotheses for a chi-square test for homogeneity.
- Check for independence (random and 10% conditions) and that the sampling distribution is approximately chi-square before performing a chi-square test for homogeneity.
- Calculate the χ^2 test statistic and p-value for a chi-square test for homogeneity.
- State a conclusion for a chi-square test for homogeneity.

Unit Essential Questions:

- What is an unusual event when sampling from a population?
- What patterns occur when sampling from two populations when measuring categorical variables?
- How is an unknown difference in population parameters estimated?
- How is a claim about an unknown difference in population parameters tested?
- How can inferences about relationships between categorical variables be made?

Lesson Essential Questions:

- What is the sampling distribution of a difference in sample proportions?
- How can a difference in two population proportions be estimated?
- How can a claim about a difference in two population proportions be tested?
- How is a claim about independence between two categorical variables in one population tested?
- How is a claim about the distribution of a categorical variable in multiple populations tested?

Materials/Resources:

Textbook
Graphing calculator
Online statistical applets (e.g. https://www.stapplet.com/)

Vocabulary:

Difference in proportions Independent samples Sampling distribution of a difference in proportions	Two-sample z test for $p_1 - p_2$ Two-sample z interval for $p_1 - p_2$	Observed counts Expected counts X^2-distribution Two-way table	Association Independence Chi-square test for independence Chi-square test for homogeneity

Assessments:

Unit quizzes
Unit tests
Informal and formal writing assignments
Presentations
Informal check of daily assignments
Progress check (College Board)

Central York School District

AP Statistics

Grade Level: 9 – 12

Unit 4 Part 1: Inference for One Mean

Time: 7 days

Standards:

College Board 2026 CED for AP Statistics

- 4.1.A Calculate the mean and standard deviation of a sampling distribution of a sample mean.
- 4.1.B Justify the appropriateness of conditions for the sampling distribution of a sample mean.
- 4.1.C Interpret the mean, standard deviation, and probabilities for the sampling distribution of a sample mean.
- 4.2.A Describe t-distributions.
- 4.2.B Identify an appropriate confidence interval procedure including the parameter for a population mean or population mean difference.
- 4.2.C Justify the appropriateness of constructing a confidence interval for a population mean or population mean difference by verifying conditions.
- 4.2.D Calculate an appropriate confidence interval for a population mean or population mean difference.
- 4.2.E Calculate the standard error and margin of error for a sample size for a one-sample t-interval.
- 4.3.A Interpret a confidence interval in context for a population mean or population mean difference.
- 4.3.B Justify a claim based on a confidence interval for a population mean or population mean difference.
- 4.3.C Identify the relationships among sample size, confidence interval width, confidence level, and margin of error for a population mean or population mean difference.
- 4.4.A Identify an appropriate testing method and parameter for a population mean or population mean difference with unknown σ.
- 4.4.B Identify the null and alternative hypotheses for a population mean or population mean difference with unknown σ.
- 4.4.C Justify the appropriateness of a hypothesis test for a population mean or population mean difference by verifying conditions.
- 4.5.A Calculate an appropriate test statistic and p-value for testing a hypothesis about a population mean or population mean difference.
- 4.5.B Interpret the p-value of a hypothesis test for a population mean or population mean difference.
- 4.5.C Justify a claim about the population based on the results of a hypothesis test for a population mean or population mean difference.

Know:

- The sampling distribution of the sample mean $\bar{x}$ has mean μ and standard deviation $\sigma/\sqrt{n}$.
- If sampling without replacement, the standard deviation of the sample mean is smaller than what is given by the formula above.
- If the sample size is less than 10% of the population size, the difference between the formula and the actual standard deviation is negligible.
- The central limit theorem (CLT) states that when the sample size is sufficiently large, a sampling distribution of the mean of a random variable will be approximately normally distributed.
- The central limit theorem requires that the sample values are independent of each other and that n is sufficiently large.
- The sampling distribution of the sample mean $\bar{x}$, will have an approximate normal distribution, provided the population is approximately normal or the sample size is large (e.g $n \geq 30$).

- Probabilities and parameters for a sampling distribution for a sample mean should be interpreted using appropriate units and within the context of a specific population.
- When s is used instead of σ to calculate a test statistic, the corresponding distribution, known as the t-distribution.
- The t-distribution varies from the normal distribution in shape, in that more of the area is allocated to the tails of the density curve than in a normal distribution.
- As the degrees of freedom increase, the area in the tails of a t-distribution decreases.
- Because σ is typically not known for distributions of quantitative variables, the appropriate confidence interval procedure for estimating the population mean of one quantitative variable for one sample is a one-sample t-interval for a mean.
- For one quantitative variable X that is normally distributed, the distribution of $t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ is a t-distribution with $n - 1$ degrees of freedom.
- Matched pairs can be thought of as one sample of pairs. Once differences between pairs of values are found, inference for confidence intervals proceeds as for a population mean.
- In order to calculate confidence intervals to estimate a population mean, we must check for independence and that the sampling distribution is approximately normal.
- The standard error of $\bar{x}$ is $\frac{s}{\sqrt{n}}$.
- The margin of error of $\bar{x}$ is $t^* \frac{s}{\sqrt{n}}$.
- The point estimate for a population mean is the sample mean $\bar{x}$.
- For the population mean for one sample with unknown population standard deviation, the confidence interval is $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$.
- A confidence interval for a population mean either contains the population mean or it does not.
- We are C% confident that the confidence interval for a population mean captures the population mean.
- An interpretation of a confidence interval for a population mean includes a reference to the sample taken and details about the population it represents.
- A confidence interval for a population mean provides an interval of values that may provide sufficient evidence to support a particular claim in context.
- When all other things remain the same, the width of a confidence interval for a population mean tends to decrease as the sample size increases.
- For a population mean, the width of the confidence interval is proportional to $1/\sqrt{n}$.
- For a given sample, the width of the confidence interval for a population mean increases as the confidence level increases.
- The appropriate test for a population mean with unknown is a one-sample t-test for a population mean.
- Matched pairs can be thought of as one sample of pairs. Once differences between pairs of values are found, inference for significance testing proceeds as for a population mean.
- When finding the mean difference μ_d between values in a matched pair, it is important to define the order of subtraction.
- In order to make statistical inferences when testing a population mean, we must check for independence and that the sampling distribution is approximately normal.
- For a single quantitative variable when random sampling with replacement from a population that can be modeled with a normal distribution with mean μ and standard deviation σ, the sampling distribution of $t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$ has a t-distribution with $n - 1$ degrees of freedom.
- An interpretation of the p-value of a significance test for a population mean should recognize that the p-value is computed by assuming that the null hypothesis is true.
- A formal decision explicitly compares the p-value to the significance level a.
- If the p-value $\leq a$, reject the null hypothesis. If the p-value $> a$, fail to reject the null hypothesis.
- The results of a significance test for a population mean can serve as the statistical reasoning to support the

answer to an investigative question about the population that was sampled.

- Random variation may result in errors in statistical inference.

Understanding/Key Learning:

- Random variation may result in errors in statistical inference.
- The *t*-distribution may be used to model variation.
- An interval of values should be used to estimate parameters, in order to account for uncertainty.
- Significance testing allows us to make decisions about hypotheses within a particular context.

Do:

- Verify whether the conditions for the sampling distribution of the sample mean $\bar{x}$ have been met.
- Compute the mean and standard deviation of the sampling distribution of the sample mean $\bar{x}$.
- Explain the Central Limit Theorem in the context of a given population.
- Compute probabilities of events given the sampling distribution of a sample mean $\bar{x}$.
- Check for independence (random and 10% conditions) and that the sampling distribution is approximately normal (Normal/large sample condition) before constructing a confidence interval to estimate a population mean μ.
- Determine the critical value t^* using a table or technology.
- Calculate and interpret a one-sample t interval for μ.
- Check for independence (random and 10% conditions) and that the sampling distribution is approximately normal (Normal/large sample condition) before constructing a confidence interval to estimate a population mean difference μ_{Diff}.
- Calculate and interpret a paired t interval for μ_{Diff}.
- Interpret the confidence level of a confidence interval.
- Evaluate a claim about a parameter using a confidence interval.
- Describe how the margin of error relates to sample size and confidence level.
- State hypotheses for a significance test about a population mean μ.
- Check for independence (random and 10% conditions) and that the sampling distribution is approximately normal (Large Counts condition) before performing a significance test about a population mean μ.
- Calculate the standardized test statistic and p-value for a one-sample t test for μ.
- State a conclusion for a paired t test for μ_{Diff}.
- State hypotheses for a significance test about a population mean difference μ_{Diff}.
- Check for independence (random and 10% conditions) and that the sampling distribution is approximately normal (Large Counts condition) before performing a significance test about a population mean difference μ_{Diff}.
- Calculate the standardized test statistic and p-value for a paired t test for μ_{Diff}
- State a conclusion for a paired t test for μ_{Diff}.
- Explain the relationship between a *C*% confidence interval and two-sided significance test with a significance level of *a*.
- Interpret the p-value of a significance test.
- Describe a Type I error and give a possible consequence of such an error.
- Describe a Type II error and give a possible consequence of such an error.
- Explain how sample size, significance level, standard error, and the true (alternative) value of the parameter influence the power of a test.

Unit Essential Questions:

- What is an unusual event when sampling from a population?
- What patterns occur when sampling from a population when measuring a quantitative variable?
- How is an unknown population parameter estimated?
- How is a claim about an unknown population parameter tested?

Lesson Essential Questions:

- What is the sampling distribution of a sample mean?
- How can a population mean be estimated?
- How can a population mean difference be estimated?
- What is the connection between confidence intervals and significance tests?
- How is a claim about a population mean tested?
- How is a claim about a population mean difference tested?

Materials/Resources:

Textbook
Graphing calculator
Online statistical applets (e.g. https://www.stapplet.com/)

Vocabulary:

Sampling distribution of the sample mean
Random condition
10% condition
Central Limit Theorem
Normal/large sample condition
t-distribution
Degrees of freedom
One-sample *t* interval for μ
One-sample *t* test for μ
Paired data
Mean difference
Difference in means
Paired *t* interval for μ_{Diff}
Paired *t* test for μ_{Diff}

Assessments:

Unit quizzes
Unit tests
Informal and formal writing assignments
Presentations
Informal check of daily assignments
Progress check (College Board)

Central York School District

AP Statistics

Grade Level: 9 – 12

Unit 4 Part 2: Inference for Two Means

Time: 5 days

Standards:

<u>College Board 2026 CED for AP Statistics</u>

- 4.6.A Calculate the mean and standard deviation of a sampling distribution for the difference between two sample means.
- 4.6.B Justify the appropriateness of conditions for the sampling distribution of the difference between two sample means.
- 4.6.C Interpret the mean, standard deviation, and probabilities for a sampling distribution for the difference between sample means.
- 4.7.A Identify an appropriate confidence interval procedure including the parameter for the difference between two population means.
- 4.7.B Justify the appropriateness of constructing a confidence interval for the difference between two population means by verifying conditions.
- 4.7.C Calculate an appropriate confidence interval for the difference between two population means.
- 4.7.D Calculate the standard error and margin of error for estimating the difference between two population means.
- 4.8.A Interpret a confidence interval in context for the difference between two population means.
- 4.8.B Justify a claim based on a confidence interval for the difference between two population means.
- 4.9.A Identify an appropriate testing method for the difference between two population means including the parameters for the difference between the two population means.
- 4.9.B Identify the null and alternative hypotheses for the difference between two population means.
- 4.9.C Justify the appropriateness of a hypothesis test for the difference between two population means by verifying conditions.
- 4.10.A Calculate an appropriate test statistic and p-value for testing a hypothesis for the difference between two population means.
- 4.10.B Interpret the p-value of a hypothesis test for the difference between two population means.
- 4.10.C Justify a claim about the populations based on the results of a hypothesis test for the difference between two population means.

Know:

- The sampling distribution of the difference in sample means $\overline{x}_1 - \overline{x}_2$ has mean $\mu_1 - \mu_2$ and standard deviation $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.
- If sampling without replacement, the standard deviation of the difference in means smaller than what is given by the formula above.
- If the sample size is less than 10% of the population size, the difference between the formula and the actual standard deviation is negligible.
- The sampling distribution of the difference in means will have an approximate normal distribution, provided each sample comes from a normal population or the sample size is large (e.g n ≥ 30).
- Parameters for a sampling distribution for a difference of means should be interpreted using appropriate units and within the context of specific populations.The appropriate confidence interval procedure for one

quantitative variable for two independent samples is a two-sample t- interval for a difference between population means.

- In order to calculate confidence intervals to estimate a difference of population means, we must check for independence and that the sampling distribution is approximately normal.
- For the difference of two sample means, the margin of error is the critical value (t*) times the standard error (SE) of the difference of two means.
- The standard error of $\bar{x}_1 - \bar{x}_2$ is $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.
- The margin of error of $\bar{x}_1 - \bar{x}_2$ is $t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.
- The point estimate for the difference of two population means is the difference in sample means $\bar{x}_1 - \bar{x}_2$.
- For a difference of two population means where the population standard deviations are not known, the confidence interval is $\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ where ± t* are the critical values for the central C% of a *t*-distribution with appropriate degrees of freedom that can be found using technology.
- In repeated random sampling with the same sample size, approximately C% of confidence intervals created will capture the difference of population means.
- An interpretation for a confidence interval for the difference of two population means should include a reference to the samples taken and details about the populations they represent.
- A confidence interval for a difference of population means provides an interval of values that may provide sufficient evidence to support a particular claim in context.
- When all other things remain the same, the width of the confidence interval for the difference of two means tends to decrease as the sample sizes increase.
- For a quantitative variable, the appropriate test for a difference of two population means is a two-sample t-test for a difference of two population means $\mu_1 - \mu_2$.
- In order to make statistical inferences when testing a difference between population means, we must check for independence and that the sampling distribution is approximately normal.
- For a single quantitative variable, data collected using independent random samples or a randomized experiment from two populations, each of which can be modeled with a normal distribution, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is an approximate *t*-distribution with degrees of freedom that can be found using technology.
- An interpretation of the p-value of a significance test for a two-sample difference of population means should recognize that the p-value is computed by assuming that the null hypothesis is true.
- The results of a significance test for a two-sample test for a difference between two population means can serve as the statistical reasoning to support the answer to an investigative question about the populations that were sampled.

Understanding/Key Learning:

- Random variation may result in errors in statistical inference.
- The *t*-distribution may be used to model variation.
- An interval of values should be used to estimate parameters, in order to account for uncertainty.
- Significance testing allows us to make decisions about hypotheses within a particular context.

Do:

- Verify whether the conditions for the sampling distribution of a difference in means $\bar{x}_1 - \bar{x}_2$ have been met.
- Compute the mean and standard deviation of the sampling distribution of a difference in means $\bar{x}_1 - \bar{x}_2$.
- Compute probabilities of events given the sampling distribution of a difference in means

$\bar{x}_1 - \bar{x}_2$.

- Check for independence (random and 10% conditions) and that the sampling distribution is approximately normal (Large Counts condition) before constructing a confidence interval for a difference in two population means $\mu_1 - \mu_2$.
- Calculate and interpret a two-sample t interval for $\mu_1 - \mu_2$.
- State hypotheses for a significance test about a difference in two population means $\mu_1 - \mu_2$.
- Check for independence (random and 10% conditions) and that the sampling distribution is approximately normal (Large Counts condition) before performing a significance test about a difference in two population means $\mu_1 - \mu_{22}$.
- Calculate the standardized test statistic and p-value for a two-sample t test for $\mu_1 - \mu_2$.
- State a conclusion for a two-sample t test for $\mu_1 - \mu_2$.
- Explain the relationship between a C% confidence interval and two-sided significance test with a significance level of a.
- Interpret the p-value of a significance test.
- Describe a Type I error and give a possible consequence of such an error.
- Describe a Type II error and give a possible consequence of such an error.
- Explain how sample size, significance level, standard error, and the true (alternative) value of the parameter influence the power of a test.

Unit Essential Questions:

- What is an unusual event when sampling from a population?
- What patterns occur when sampling from two populations when measuring quantitative variables?
- How is an unknown difference in population parameters estimated?
- How is a claim about an unknown difference in population parameters tested?

Lesson Essential Questions:

- What is the sampling distribution of a difference in sample means?
- How can a difference in two population means be estimated?
- How can a claim about a difference in two population means be tested?

Materials/Resources:

Textbook
Graphing calculator
Online statistical applets (e.g. https://www.stapplet.com/)

Vocabulary:

Difference in means Sampling distribution of a difference in means	Two-sample t interval for $\mu_1 - \mu_2$	Two-sample t test for $\mu_1 - \mu_2$

Assessments:

Unit quizzes
Unit tests
Informal and formal writing assignments
Presentations
Informal check of daily assignments
Progress check (College Board)

Central York School District

AP Statistics

Grade Level: 9 – 12

Unit 5: Regression Analysis

Time: 5 days

Standards:
<u>**College Board 2026 CED for AP Statistics**</u> • 5.1.A Construct scatterplots depicting the relationship between two quantitative variables. • 5.1.B Describe the characteristics of a scatterplot. • 5.1.C Justify a claim using scatterplots depicting the relationship between two quantitative variables. • 5.2.A Interpret the correlation for a linear relationship. • 5.3.A Calculate a predicted response value using a linear regression model. • 5.4.A Calculate the differences between the observed and predicted values. • 5.4.B Interpret the differences between the observed and predicted values. • 5.4.C Describe the form of association of bivariate data using residual plots. • 5.5.A Calculate the coefficients for the least-squares regression line model. • 5.5.B Interpret coefficients for the least-squares regression line model.
Know:
• Apparent patterns and associations in data may be random or not. • A bivariate quantitative data set consists of observations of two different quantitative variables made on individuals in a sample or population. • An explanatory variable is a variable whose values are used to explain or predict corresponding values for the response variable. • The direction of the association shown in a scatterplot, if any, can be described as positive or negative. • A positive association means that as values of one variable increase, the values of the other variable tend to increase. • A negative association means that as values of one variable increase, values of the other variable tend to decrease. • The form of the association shown in a scatterplot, if any, can be described as linear or non-linear to varying degrees. • The strength of the association is how closely the individual points follow a specific pattern, e.g., linear, and can be shown in a scatterplot. • Strength can be described as strong, moderate, or weak. • Unusual features of a scatter plot include clusters of points or points with relatively large discrepancies between the value of the response variable and a predicted value for the response variable. • The correlation, *r*, gives the direction and quantifies the strength of the linear association between two quantitative variables. • A correlation coefficient close to 1 or −1 does not necessarily mean that a linear model is appropriate. • The correlation, *r*, is unit-free. • The correlation, *r*, is between −1 and 1, inclusive. • A value of r = 0 indicates that there is no linear association. • A value of r = 1 or r = −1 indicates that there is a perfect linear association. • A perceived or real relationship between two variables does not mean that changes in one variable cause changes in the other. That is, association does not necessarily imply causation. • A simple linear regression model is an equation that uses an explanatory variable, *x*, to predict the response

variable, y.

- The predicted response value, denoted by $\hat{y}$, is calculated as $\hat{y} = a + bx$.
- a is the y-intercept of the regression line.
- b is the slope of the regression line.
- Extrapolation is predicting a response value using a value for the explanatory variable that is beyond the interval of x-values used to determine the regression line.
- The residual is the difference between the actual value and the predicted value.
- A residual plot is a plot of residuals versus explanatory variable values or predicted response values.
- Apparent randomness in a residual plot for a linear model is evidence of a linear form to the association between the variables.
- The least-squares regression model minimizes the sum of the squares of the residuals.
- The least-squares regression model contains the point $(\bar{x}, \bar{y})$.
- The slope is the amount that the predicted y-value changes for every unit increase in x.
- The y-intercept value is the predicted value of the response variable when the explanatory variable is equal to 0.

Understanding/Key Learning:

- Given that variation may be random or not, conclusions are uncertain.
- Graphical representations and statistics allow us to identify and represent key features of data.
- Regression models may allow us to predict responses to changes in an explanatory variable.

Do:

- Construct a scatterplot.
- Describe the association between two variables in a scatterplot.
- Interpret a correlation coefficient r.
- Compute a least-squares regression equation.
- Predict a value of the response variable using a least-squares regression equation.
- Compute and interpret a residual.
- Interpret the slope b of a least-squares regression equation.
- Interpret the y-intercept a of a least-squares regression equation.
- Sometimes, the y-intercept of the line does not have a logical interpretation in context.
- Investigate the appropriateness of a selected model using a residual plot.
- Interpret the coefficient of determination r^2.

Unit Essential Questions:

- Is there an association between two quantitative variables?
- How can the association between two quantitative variables be measured?
- How can associations between variables be used to make predictions?

Lesson Essential Questions:

- How are data from two quantitative variables organized?
- Is there an association between two quantitative variables?
- Can the relationship between two quantitative variables be modeled?
- Is the least-squares regression model adequate?
- How is the least-squares regression model interpreted?

Materials/Resources:

Textbook
Graphing calculator
Online statistical applets (e.g. https://www.stapplet.com/)

Vocabulary:

Explanatory variable
Response variable

Scatterplot
Direction
Form
Strength
Correlation coefficient

Least squares regression line
Slope coefficient
Intercept coefficient
Residual
Residual plot

Coefficient of determination

Assessments:

Unit quizzes
Unit tests
Informal and formal writing assignments
Presentations
Informal check of daily assignments
Progress check (College Board)