Probability & Statistics Curriculum Maps

<table>
<tr><td>Subject: Probability and Statistics</td><td>Unit 1: Exploring and Understanding Data</td></tr>
<tr><td>Big Idea/Rationale</td><td>Big Idea: Describing and Comparing Distributions of Data

Rationale: Statistics is a way of reasoning designed to help us understand the world in which we live. This unit introduces students to categorical and quantitative variables; the “what” we want to learn about the world. In the world of statistics, students need to learn how to make and interpret graphical displays of the data, as well as, how to calculate meaningful statistics. Once graphs are made and statistics are calculated, students must be able to communicate the findings and results. This unit focuses on the basics which allow students to explore data and communicate the findings to others.</td></tr>
<tr><td>Enduring Understanding (Mastery Objective)</td><td>Data must have context to truly understand what it is telling you.
When analyzing data, it is important to make pictures that are not misleading.
The shape of a distribution provides the basis for most of your key analytic decisions.
Sometimes it is the parts of a distribution that are unusual that are most interesting and important.
Think about the data being analyzed, don’t just perform computations.
Models give us insight but are usually not correct in the real world.</td></tr>
<tr><td>Essential Questions (Instructional Objective)</td><td>What information is important to know when working with data?
What does the type of data you are working with indicate about the statistical analyses you should perform?
Why is independence considered a mathematical issue compared to a real world issue?
What are the key features to discuss when describing a distribution or comparing distributions of data?
Why do we need to standardize scores?
What does it mean when we say that there are millions of normal distributions?
Why do we use models if we know the answers they give us are most likely wrong?</td></tr>
<tr><td>Content (Subject Matter)</td><td>Students will know…
Key Terms – statistics, data, variation, variable, categorical, quantitative, frequency, relative frequency, distribution, contingency table, marginal distribution, conditional distribution, independent, Simpson’s paradox, unimodal, bimodal, multimodal, uniform, symmetric, skewed, outliers, center, spread, mean, median, IQR, standard deviation, percentile, z-score, standardized value, parameter, statistic, normal model, standard normal model
Students will be able to…
Categorize a variable as categorical or quantitative</td></tr>
</table>

	• Create pie charts, bar charts, and segmented bar charts • Find conditional and marginal distributions • Determine if two variables are independent or associated • Create histograms and stem and leaf plots • Determine the shape, center, and spread of a distribution • Create box plots and time series plots • Use technology to calculate summary statistics • Compare two or more distributions • Use a normal model to solve problem • Find normal probabilities and values • Use the 68-95-99.7 rule • Calculate and use z-scores to compare variables
Skills/ Benchmarks (Standards)	S.ID.A.1. Represent data with plots on the real number line (dot plots, histograms, and box plots). S.ID.A.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.A.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.A.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. Use calculators, spreadsheets, and tables to estimate areas under the normal curve. S.ID.B.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. S.IC.B.6. Evaluate reports based on data. S.CP.A.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. *For example, collect data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*
Materials and Resources	• Intro Stats Textbook and Resource Materials • Overhead, transparencies, markers, document camera • Pennies, scales • Graphing calculators
Notes	

Subject: Probability and Statistics	Unit 2: Exploring Relationships Between Variables
Big Idea/Rationale	Big Idea: Describing and modeling relationships between 2 variables Rationale: In the real world, it is important to be able to see relationships between variables. Many times in business, education, medicine, etc., it is important to be able use a model to predict a variable based on another variable, for example using SAT scores to predict college GPA.
Enduring Understanding (Mastery Objective)	• Correlation is a specific type of relationship (linear). Association is the term that applies in all other cases. • Scatterplots can be used to determine the form, strength, and direction of the relationship between two variables. • Correlation does not imply causation. • All models are wrong, but some are useful. • Beware of using a model to extrapolate into the future. • Make sure that the model created is both strong and appropriate.
Essential Questions (Instructional Objective)	• What does the correlation coefficient tell us about a relationship? • What are some of the effects outliers can have on correlation and slope of a regression line? • What are some of the things to look for in a scatterplot that could cause potential issues when creating a linear model? • How do we decide which mathematical function to apply when straightening a curved scatterplot? • How do we determine if a linear model is strong and appropriate? • Why do we re-express curved data? • Why is extrapolation dangerous?
Content (Subject Matter)	*Students will know…* • Key Terms – scatterplot, association, explanatory variable, response variable, correlation coefficient, parameter, linear model, predicted value, residual, line of best fit, coefficient of determination, mean-mean point, extrapolation, interpolation, leverage, influential point, lurking variable, homogenous, re-expression *Students will be able to…* • Differentiate between explanatory and response variables. • Calculate and interpret the correlation coefficient. • Distinguish between correlation and causation. • Describe relationships in scatterplots. • Find the equation of the line of best. • Predict response variables using the regression equation. • Interpret the R-squared, slope, and y-intercept in context. • Read computer generated statistical output.

	• Analyze the strength and appropriateness of a linear model. • Identify points of leverage and influential points. • Re-express data to form linear models. • Back solve to find values from re-expressed data.
Skills/ Benchmarks (Standards)	S.ID.B.6. Represent data on two quantitative variables on a scatter plot, and describe how the variables are related. S.ID.B.6A. 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. S.ID.B.6.B. Informally assess the fit of a function by plotting and analyzing residuals. S.ID.B.6.C. Fit a linear function for a scatter plot that suggests a linear association. S.ID.C.7. Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data. S.ID.C.8. Compute (using technology) and interpret the correlation coefficient of a linear fit. S.ID.C.9. Distinguish between correlation and causation.
Materials and Resources	• Intro Stats Textbook and Resource Materials • Overhead, transparencies, markers, document camera • Graphing calculators
Notes	

Subject: Probability and Statistics	**Unit 3: Gathering Data**
Big Idea/Rationale	Big Idea: Generating or collecting data for statistical analysis Rationale: Statistics tries to explain the world based on data. We look beyond randomness for patterns and relationships. Sometimes though, randomness is our salvation. This unit describes different techniques for the proper collection of data from designing good experiments or setting up proper sampling procedures using randomness to avoid biases. In addition to those topics, the design of the survey instrument is important for ensuring the data that is collected is useful. Another method of collecting data is using simulation to examine the results of random phenomena. All of these methods are necessary to collect data that can be analyzed and used to reach meaningful conclusions.
Enduring Understanding (Mastery Objective)	• Randomness is a tool that can be used to reduce bias from data and allow us to make inferences beyond that data themselves. • There are many methods to draw samples from populations, but in order to be useful, there must be a component of randomness. • Bias and sampling error can arise, even when proper statistical techniques are used. • The accuracy of results from a poll depends solely on the size of the sample and not on the size of the population. • Only well-designed experiments can be used to establish cause and effect, observational studies can only show an association. • Experiments are rarely able to allow for generalizations.
Essential Questions (Instructional Objective)	• Why is randomness not the same as equally likely? • What are some of the factors that should be considered when determining which type of sample to draw? • What are some of the considerations to make when designing a questionnaire? • Why do we block in an experiment? • What is the difference between a prospective and retrospective study? • What is the difference between confounding and lurking variables? • How can ethical considerations effect the design of experiments?
Content (Subject Matter)	*Students will know…* • Key Terms – random behavior, random numbers, pseudorandom numbers, simulation, component, outcome, trial, population, sample, sample survey, biased, randomization, census, parameter, statistic, SRS, sampling frame, sampling variability, strata, stratified sample, cluster sample, multistage sample, systematics sample, respondents, voluntary response, convenience sampling, under coverage, nonresponse bias, response bias, observational study, retrospective, prospective, experiment, random assignment, subjects, participants, experimental units, factors, levels, treatments, block,

	completely randomized design, statistically significant, control, control group, single-blind, double-blind, placebo, placebo effect, matching, confounding *Students will be able to…* • Recognize random outcomes in real-world situations. • Recognize when a simulation might usefully model random behavior. • Know how to perform a simulation using a calculator or random number table. • Describe a simulation so others could repeat it. • Discuss the results of a simulation and draw conclusions. • Understand the value of randomization as a defense against bias. • Understand the size of the sample, not the population determines the precision of estimates. • Draw an SRS from a master list of a population using a calculator or random number table. • Know what to report about a sample as part of a statistical analysis. • Report possible biases in sampling methods. • Recognize when an observational study would be appropriate. • Know the four basic principles of good experimental design. • Recognize the factors, treatments, and levels in an experiment. • Understand the purpose and value of blocking. • Understand when to use a control group, placebo, blinding or double-blinding. • Design a completely randomized or random block experiment. • Know how to report the results of observational studies and experiments.
Skills/ Benchmarks (Standards)	S.IC.A.2. Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation. *For example, a model says a spinning coin falls heads up with probability 0.5. Would a result of 5 tails in a row cause you to question the model*? S.IC.B.3. Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each. S.IC.B.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.B.5. Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.
Materials and Resources	• Intro Stats Textbook and Resource Materials • Overhead, transparencies, markers, document camera • Graphing calculators
Notes	

Subject: Probability and Statistics	**Unit 4: From Randomness to Probability**
Big Idea/Rationale	Big Idea: Use randomness to define statistical significance Rationale: Sometimes randomness and chaos may look the same. Neither of their outcomes can be anticipated with certainty, but random phenomena have a distinguishing feature. In the long run, they settle down in a way that is consistent and predictable. It is this feature of randomness that enables to do Statistics. Are our results significant? To check we need to determine if what we observe arises from chance alone or if something else is going on. Inferential statistics is based on probability and it is the concepts in this unit which give us the ability to use the inferential tools we will learn in future units.
Enduring Understanding (Mastery Objective)	• Probability is the long-run relative frequency of an event • There is a difference between the Law of Large Numbers and the Law of Averages • Whether events are independent or disjoint changes how we calculate probabilities • Probability distributions have centers and spreads • Probability models are nothing more than models. They do not tell us what will happen, just what could happen in the long run. • Bernoulli trials define situations which create specific and common probability distributions.
Essential Questions (Instructional Objective)	• What does it mean to be random? • What is the difference between independent and disjoint events? Why can't an event be both independent and disjoint? • What is meant by reversing the conditioning? • Why don't we add standard deviations when combining random variables? What situation allows us to add variances? • What is the difference between Geometric and Binomial Distributions? • What conditions must be met in order to use the normal model to approximate the binominal?
Content (Subject Matter)	*Students will know…* • Key Terms – random phenomena, probability, trial, outcome, event, independence, Law of Large Numbers, complement rule, disjoint, addition rule, multiplication rule, sample space, general addition rule, conditional probability, general multiplication rule, tree diagram, random variable, discrete random variable, probability model, continuous random variable, expected value, variance, Bernoulli trial, geometric probability, binomial probability, success/failure condition

	Students will be able to… • Understand random phenomena are unpredictable short term, but predictable in the long run. • Be able to recognize random outcomes in real-world situations. • State and use the Law of Large Numbers. • Know basic rules and definitions of probability. • Recognize if event are disjoint or independent. • Determine if a probability assignment is legitimate. • Know how and when to use the addition and multiplication rules. • Know how to use the complement rule. • Understand the concept of conditional probability. • Know how to find the probabilities of compound events. • Know how to make and use tree diagrams and Venn diagrams to solve probability problems. • Be able to recognize random variables. • Find the probability model for discrete random variables. • Know how to compute and use the mean and variance of a random variable. • Know how to determine the new mean and standard deviation after transforming or combining random variables. • Know how to tell if a situation involves Bernoulli trials. • Be able to choose whether to use geometric or binomial model for a random variable. • Calculate the expected value and variance of binomial and geometric distributions. • Determine geometric and binomial probabilities.
Skills/ Benchmarks (Standards)	S.CP.A.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.A.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.A.3. Understand the conditional probability of *A* given *B* as *P*(*A* 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.A.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. *For example, collect data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*

	S.CP.A.5. Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.* S.CP.B.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.B.7. Apply the Addition Rule, P(A or B) = P(A) + P(B) - P(A and B), and interpret the answer in terms of the model. S.CP.B.8. (+) Apply the general Multiplication Rule in a uniform probability model, P(A and B) = P(A)P(B|A) = P(B)P(A|B), and interpret the answer in terms of the model. S.MD.A.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.A.2. (+) Calculate the expected value of a random variable; interpret it as the mean of the probability distribution. S.MD.A.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. *For example, find the theoretical probability distribution for the number of correct answers obtained by guessing on all five questions of a multiple-choice test where each question has four choices, and find the expected grade under various grading schemes.* S.MD.A.4. (+) Develop a probability distribution for a random variable defined for a sample space in which probabilities are assigned empirically; find the expected value. *For example, find a current data distribution on the number of TV sets per household in the United States, and calculate the expected number of sets per household. How many TV sets would you expect to find in 100 randomly selected households?*
Materials and Resources	• Intro Stats Textbook and Resource Materials • Overhead, transparencies, markers, document camera • Graphing calculators
Notes	

Subject: Probability and Statistics	Unit 5: From Data at Hand to the World at Large
Big Idea/Rationale	Big Idea: Use proportions found from samples to make inferences about entire populations. Rationale: We see polls in the news nearly every day. They say some percentage of Americans do this, feel this, or believe this, yet only 1000 or so people respond. This unit introduces students to concepts that allow statisticians to make decisions and inferences about a large group of people based on data collected from a small group of people.
Enduring Understanding (Mastery Objective)	• Samples vary less than individuals do. • When we want to estimate a value from a population, we should use an interval. • The margin of error of a study is determined solely by the sample size, not the size of the population. • Significance tests in statistics use the same logic as trials in the American judicial system. • It is important to think about the conditions under which the data was collected before proceeding with an inference procedure. • The p-value is the key piece in decision-making during a statistical inference procedure.
Essential Questions (Instructional Objective)	• What does the Central Limit Theorem tell us about sampling distributions? • What is the difference between a confidence interval and a confidence level? • How do we determine what the proper null and alternative hypotheses are for a given situation? • What is a p-value? • What are Type I and Type II errors? Why are they important? • How does a hypothesis test relate to a confidence interval? • Why do you pool proportions when conducting two-sample tests?
Content (Subject Matter)	*Students will know…* • Key Terms – sampling distribution model, central limit theorem, standard error, confidence interval, one-proportion z-interval, margin of error, critical value, assumptions, conditions, null hypothesis, alternative hypothesis, p-value, two sided test, one sided test, one-proportion z-test, statistically significant, alpha level, significance level, type I error, type II error, power, effect size, two-proportion z-interval, two-proportion z-test, pooling *Students will be able to…* • Understand the variability of a statistics depends on the size of the sample. • Understand the Central Limit Theorem gives us a model for a sampling distribution sufficiently for large samples regardless of the underlying

	population. • Use a sampling distribution model to make statements about a distribution. • Understand that margin of error changes with sample size and level of confidence. • Know how to examine data for a violation of the necessary conditions for inference. • Construct a one-proportion z-interval. • Write a correct interpretation of a confidence interval. • State the null and alternative hypotheses for a one-proportion z-test. • Know the conditions that make a one-proportion z-test appropriate. • Choose between a one-sided and two-sided test. • Perform a one-proportion z-test. • Write a sentence interpreting the results of a test. • Interpret the meaning of a p-value in non-technical language. • Understand the relationships between confidence intervals and significance tests. • Understand how the critical value is related to the alpha level for a test. • Understand the power of a test is the test's ability to reject a false null. • Understand the power of a test is determined by sample size and effect size. • Know that we do not "accept" a null hypothesis. • Understand that the formula for the standard error in a two-proportion test is based on the fact that variances add. • Know how to find a confidence interval for the difference of two-proportions. • Know how to perform a two-proportion z-test. • Know how to interpret and describe the results of two proportion tests and intervals.
Skills/ Benchmarks (Standards)	S.IC.A.1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population. S.IC.B.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.B.5. Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant. S.IC.B.6. Evaluate reports based on data.
Materials and Resources	• Intro Stats Textbook and Resource Materials • Overhead, transparencies, markers, document camera • Graphing calculators
Notes	

Subject: Probability and Statistics	Unit 6: Learning About the World
Big Idea/Rationale	Big Idea: Use the means found from samples to make decisions about populations. Rationale: In the last unit we looked at how polls are used to create estimates for populations. This unit examines averages. For example, do boys have higher SAT scores than girls? Is the manufacturers claim on tread life of tires valid?
Enduring Understanding (Mastery Objective)	• The t-distribution is like the Normal distribution, but is a family of curves instead of one curve. • The shape of the t-distribution curve is determined by degrees of freedom which is one less than the size of the sample. • When we look to compare means it is important to determine if the samples are independent. • We don't pool variances with means because we cannot make the assumption that the variances of two different groups are the same. • When data are paired we find the differences between each observation instead of the difference in means.
Essential Questions (Instructional Objective)	• What does it mean when we say the t-distribution approaches the normal distribution when degrees of freedom increase? • How do we identify if samples are independent or paired? • Why can we pool with proportions, but not with means? • How do we check for the nearly normal condition? • Why must we have an underlying assumption of normality to use the procedures in this unit?
Content (Subject Matter)	*Students will know…* • Key Terms –Student's T, degrees of freedom, pooling, paired data *Students will be able to…* • Know the assumptions required for t-tests and t-intervals. • Compute and interpret the results of a t-test. • Compute and interpret a t-based confidence interval. • Use computer generated output to analyze data. • Recognize situations that require inference on means for independent samples. • Recognize when a pooled test may be appropriate but explain why we may not do it anyway. • Perform a two-sample t-test and interpret the results. • Compute a two-sample t-interval and interpret the results. • Recognize whether a design that compares two groups is paired or not. • Find a paired confidence interval and recognize the procedure is that same

	as a one-sample t-interval. • Conduct a paired t-test and recognize the procedure is that same as a one-sample t-test.
Skills/ Benchmarks (Standards)	S.IC.A.1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population. S.IC.B.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.B.5. Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant. S.IC.B.6. Evaluate reports based on data.
Materials and Resources	• Intro Stats Textbook and Resource Materials • Overhead, transparencies, markers, document camera • Graphing calculators
Notes	

Subject: Probability and Statistics	**Unit 7: Inference When Variables are Related**
Big Idea/Rationale	Big Idea: Perform inference procedures to determine if there is a relationship between two variables. Rationale: One of the most important parts of statistics is to study the relationship between two variables. Might one of the variables provide us information about another? Sometimes we may see what appears to be a relationship, but there might not really be one. In this unit, we learn how to determine if the relationships we see in a subset of data are real, or is it just attributable to the sample we are examining.
Enduring Understanding (Mastery Objective)	• The Chi-Square tests are used to analyze counted data. • The Chi-Square tests are a simpler way to compare proportions across more than two groups as opposed to conducting many two-proportion z-tests. • The appropriate Chi-Square test is determined by analyzing how the data was collected. • We do not construct confidence intervals for Chi-Square analyses. • Even though there may be no association of variables in a population, we may still see evidence of an association in the sample.
Essential Questions (Instructional Objective)	• What are the differences and similarities between a test of independence and a test of homogeneity? • What do we do if the expected counts for a cell are too small? • How do we determine the degrees of freedom for any test? • Why do we need to check the residual plot before proceeding with inference procedures on the slope of a regression line? • What are the key things to look for when analyzing computer output for regression?
Content (Subject Matter)	*Students will know…* • Key Terms – goodness-of-fit, homogeneity, standardized residual, chi-square component, chi-square model, standard deviation of residuals, prediction interval *Students will be able to…* • Recognize when to use goodness-of-fit, independence, or homogeneity tests. • Understand how to calculate degrees of freedom. • Display and interpret counts in a two-way table. • Use chi-square tables to find critical values. • Compute chi-square statistics by hand or using technology. • Interpret chi-square test results. • Examine regression data to find violations of conditions and assumptions for inference.

	• Examine residuals for normality and similar spread. • Test the hypothesis that the true regression slop is 0. • Find the confidence interval for the slope of a regression line. • Summarize regression in words. • Interpret inference results for slopes. • Read computer generated statistical summaries.
Skills/ Benchmarks (Standards)	S.IC.A.1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population. S.IC.B.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.B.5. Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant. S.IC.B.6. Evaluate reports based on data.
Materials and Resources	• Intro Stats Textbook and Resource Materials • Overhead, transparencies, markers, document camera • Graphing calculators
Notes	