Statistics and Probability

Course Description and Competencies for Statistics 4740

Course Description and Background

Statistics and Probability is a fourth mathematics course option for students who have completed Algebra II. This course is designed to prepare students for success in postsecondary careers and college courses in a world where all adults need statistical literacy - concepts of data analysis, statistics, and probability - to make informed decisions in life and work. Students build on the conceptual knowledge and skills they mastered in previous mathematics courses in Pre-Kindergarten through Grade 8, Algebra I, Geometry, and Algebra II. This course prepares students for colleges and careers, but is not designed to prepare students for a College Board Advanced Placement exam.

In Statistics and Probability, students will broaden their knowledge of variability and statistical processes. Students will study sampling and experimentation, categorical and quantitative data, probability and random variables, inference, and bivariate data. They will learn what it means to ask a statistical question, collect appropriate and useful data to answer that question, analyze results from that data, and interpret their results to make connections with the initial question. Based on the recommendations found in the Guidelines for Assessment and Instruction in Statistics Education[1] (GAISE), four Guiding Principles serve as a foundation for the course:

Guiding Principle 1.	*Both conceptual understanding and procedural skill should be developed deliberately, but conceptual understanding should not be sacrificed for procedural proficiency.*
Guiding Principle 2.	*Active learning is key to the development of conceptual understanding.*
Guiding Principle 3.	*Real world data must be used wherever possible in statistics education.*
Guiding Principle 4.	*Appropriate technology is essential in order to emphasize concepts over calculations.*

Preceding the GAISE report for PK-12 education, the American Statistical Association released a GAISE College Report[1] which presented five key components of what it means to be statistically educated:

1. ***Students should believe and understand why:***
    a. *Data beat anecdotes.*
    b. *Variability is natural and is also predictable and quantifiable.*
    c. *Random sampling allows results of surveys and experiments to be extended to the population from which the sample was taken.*
    d. *Random assignment in comparative experiments allows cause and effect conclusions to be drawn.*
    e. *Association is not causation.*
    f. *Statistical significance does not necessarily imply practical importance, especially for studies with large sample sizes.*

[1] The "Guidelines for assessment and instruction in statistics education (GAISE) report: A PreK–12 curriculum framework" was authored in 2007 by Christine Franklin, Gary Kader, Denise Mewborn, Jerry Moreno, Roxy Peck, Mike Perry, and Richard Schaffer. The "College report" was authored in 2005 by Martha Aliaga, George Cobb, Carolyn Cuff, Joan Garfield, Rob Gould, Robin Lock, Tom Moore, Allan Rossman, Bob Stephenson, Jessica Utts, Paul Velleman, and Jeff Witmer. Both reports were endorsed by the American Statistical Association and are available for download at http://www.amstat.org/education/gaise/.

g. *Finding no statistically significant difference or relationship does not necessarily mean there is no difference or no relationship in the population, especially for studies with small sample sizes.*

2. ***Students should recognize:***
   a. *Common sources of bias in surveys and experiments.*
   b. *How to determine the population to which the results of statistical inference can be extended, if any, based on how the data were collected.*
   c. *How to determine when a cause and effect inference can be drawn from an association, based on how the data were collected (e.g., the design of the study)*
   d. *That words such as "normal", "random" and "correlation" have specific meanings in statistics that may differ from common usage.*

3. ***Students should understand the parts of the process through which statistics works to answer questions, namely:***
   a. *How to obtain or generate data.*
   b. *How to graph the data as a first step in analyzing data, and how to know when that's enough to answer the question of interest.*
   c. *How to interpret numerical summaries and graphical displays of data - both to answer questions and to check conditions (in order to use statistical procedures correctly).*
   d. *How to make appropriate use of statistical inference.*
   e. *How to communicate the results of a statistical analysis.*

4. ***Students should understand the basic ideas of statistical inference:***
   a. *The concept of a sampling distribution and how it applies to making statistical inferences based on samples of data (including the idea of standard error)*
   b. *The concept of statistical significance including significance levels and p-values.*
   c. *The concept of confidence interval, including the interpretation of confidence level and margin of error.*

5. ***Finally, students should know:***
   a. *How to interpret statistical results in context.*
   b. *How to critique news stories and journal articles that include statistical information, including identifying what's missing in the presentation and the flaws in the studies or methods used to generate the information.*
   c. *When to call for help from a statistician.*

Course Design

The design of this course is inspired by the vision in the GAISE reports. The design has attempted to address each of these five key components of statistical education. To effectively accomplish this vision, the competencies are structured within five strands of statistical and probabilistic thinking that are essential to the development of a statistically literate citizenry. The Probability Strand, while not always addressed in a pure statistics setting, is critical to the advancement of student thinking related to likelihood, predictability, and the relationship in and among events. The remaining four strands are designed in direct response to the GAISE K-12 report, which states:

Statistical problem solving is an investigative process that involves four components:

1. *Formulate Questions (Anticipating Variability—Making the Statistics Question Distinction)*
    a. *Clarify the problem at hand*
    b. *Formulate one (or more) questions that can be answered with data*
2. *Collect Data (Acknowledging Variability—Designing for Differences)*
    a. *Design a plan to collect appropriate data*
    b. *Employ the plan to collect the data*
3. *Analyze Data (Accounting of Variability—Using Distributions)*
    a. *Select appropriate graphical and numerical methods*
    b. *Use these methods to analyze the data*
4. *Interpret Results (Allowing for Variability—Looking beyond the Data)*
    a. *Interpret the analysis*
    b. *Relate the interpretation to the original question*

Further, the incorporation of the new **Mathematical Actions and Processes**[2] were identified to be essential in the overall progression of PK-12 mathematics education. Throughout the implementation of the included competencies, it is essential to connect students to the holistic nature of mathematics that is represented within the Mathematical Actions and Processes.

Mathematical Actions and Processes

Throughout their PK-12 education experience, mathematically literate students will:

Develop a Deep and Flexible Conceptual Understanding. Demonstrate a deep and flexible conceptual understanding of mathematical concepts, operations, and relations while making mathematical and real-world connections.

Develop Accurate and Appropriate Procedural Fluency. Pursue efficient procedures for various computations and repeated processes based on a strong sense of numbers. They will develop a sophisticated understanding of the development and application of algorithms and procedures.

[2] The Mathematical Actions and Processes were included in the 2015 revisions of the Oklahoma Academic Standards for Mathematics. Each of the seven components is based on the Process Standards produced by the National Council of Teachers of Mathematics in 2000 and the interwoven strands of Mathematical Proficiency identified in the 2001 National Research Council report, Adding it Up.

Develop Strategies for Problem Solving. Analyze the parts of complex mathematical tasks and identify entry points to begin the search for a solution. They will select from a variety of problem solving strategies and use corresponding multiple representations (verbal, physical, symbolic, pictorial, graphical, tabular) when appropriate. They will pursue solutions to various tasks from real-world situations and applications that are often interdisciplinary in nature. They will find methods to verify their answers in context and will always question the reasonableness of solutions.

Develop Mathematical Reasoning. Explore and communicate a variety of reasoning strategies to think through problems. They will apply their logic to critique the thinking and strategies of others to develop and evaluate mathematical arguments, including making arguments and counterarguments and making connections to other contexts.

Develop a Productive Mathematical Disposition. Hold the belief that mathematics is sensible, useful and worthwhile. They will develop the habit of looking for and making use of patterns and mathematical structures. They will persevere and become resilient, effective problem solvers.

Develop the Ability to Make Conjectures, Model, and Generalize. Make predictions and conjectures and draw conclusions throughout the problem solving process based on patterns and the repeated structures in mathematics. They will create, identify, and extend patterns as a strategy for solving and making sense of problems.

Develop the Ability to Communicate Mathematically. Discuss, write, read, interpret and translate ideas and concepts mathematically. As they progress, students' ability to communicate mathematically will include their increased use of mathematical language and terms and analysis of mathematical definitions.

Special Thanks

Thank you for the enormous amount of energy and time put into the development of the competencies by Chuck Pack (Tahlequah Public Schools), Jim Lawson (Moore Public Schools), and Dr. Adam Molnar (Oklahoma State University). Your contributions made this work possible.

Statistics and Probability Competencies

Statistical Questions (Q)

Statistically educated students apply mathematical actions and processes to formulate statistical questions that can be investigated with data.

S.Q.1 Students will understand the distinction between mathematical models and statistical models.

- **S.Q.1.1** Distinguish among different sources of variability, including measurement, natural, induced, and sampling variability.
- **S.Q.1.2** Formulate meaningful statistical questions to clarify the problem at hand.

S.Q.2 Students will distinguish between the distribution of a population, a distribution of sample data, and a sampling distribution.

- **S.Q.2.1** Distinguish between sample statistics and population parameters.
- **S.Q.2.2** Recognize a population distribution has fixed values of its parameters and that these parameter values are typically unknown.
- **S.Q.2.3** Recognize that a sample data distribution is taken from a population distribution and the data distribution is what is seen in practice.
- **S.Q.2.4** Recognize a sampling distribution is the distribution of a sample statistic (such as a sample mean or a sample proportion) obtained from repeated samples.

S.Q.3 Students will identify differences between categorical and quantitative data.

- **S.Q.3.1** Determine whether categorical or quantitative data is appropriate to answer a statistical question.
- **S.Q.3.2** Compare and contrast different potential graphical or visual representations given the same data set.

Collect Data (CD)

Statistically educated students apply mathematical actions and processes to collect, organize, and display relevant data to answer statistical questions.

S.CD.1 Students will distinguish among different types of study designs for collecting data and know the scope of inference for each design type.

- **S.CD.1.1** Distinguish among sample surveys, experiments, and observational studies.
- **S.CD.1.2** Justify which study design will best answer a given research question.
  - **S.CD.1.2.1** Determine the appropriate scope of inference for generalizing results.
  - **S.CD.1.2.2** Explain how sample size impacts the precision with which generalizations can be made.
  - **S.CD.1.2.3** Determine when a cause and effect inference can be drawn from an association, based on how the data were collected.

S.CD.2 Students will identify common sources of bias and the role of randomization in study design.

- **S.CD.2.1** Explain how randomization and sources of bias impact the results of a study.
- **S.CD.2.2** Understand the different roles of random selection and random assignment in study design.

S.CD.3 Students will plan and conduct a study to answer a statistical question.

- **S.CD.3.1** Compare and contrast the benefits of different sampling techniques.
- **S.CD.3.2** Implement a reasonable randomization method for selecting a sample or for assigning treatments in an experiment.

S.CD.4 Students will create sample data distributions and sampling distributions.

- **S.CD.4.1** Create a sample data distribution by taking a sample from a known defined population and summarizing the data in the distribution.
- **S.CD.4.2** Create a sampling distribution of a statistic by taking repeated samples from a population (either hands-on or by simulation with technology).

Analyze Data (AD)

Statistically educated students apply mathematical actions and processes to select and use appropriate statistical methods to analyze data.

S.AD.1 Students will use distributions of quantitative and categorical data to identify the key features of the data collected in context.

- **S.AD.1.1** Summarize and represent the distribution for univariate quantitative data.
  - **S.AD.1.1.1** Describe the shape of the distribution.
  - **S.AD.1.1.2** Describe and analyze the measures of center for the distribution.

S.AD.1.1.3 Describe and analyze the patterns in variability for the distribution.

S.AD.1.1.4 Describe and analyze any outliers, gaps, or other unusual features in the distribution.

S.AD.1.2 Select and create an appropriate display, including dot plots, histograms, and box plots, for univariate data.

S.AD.1.3 Use statistics appropriate to the shape of the data distribution to compare center and variability of two or more different data sets.

S.AD.1.4 Describe and analyze the distribution of univariate categorical data.

S.AD.2 Students will use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages.

S.AD.2.1 Use calculators, computers, or tables to estimate areas under the normal curve.

S.AD.2.2 Recognize that there are data sets for which such a procedure is not appropriate.

S.AD.3 Students will compare two or more groups by analyzing distributions.

S.AD.3.1 Construct appropriate parallel graphical displays of distributions.

S.AD.3.2 Use numerical attributes of distributions to make comparisons between distributions.

S.AD.4 Students will analyze associations between two variables.

S.AD.4.1 Create two-way tables for bivariate categorical data and analyze for possible associations between the two categories using marginal, joint, and conditional frequencies.

S.AD.4.2 Make predictions and draw conclusions from linear regression models from two-variable quantitative data.

S.AD.4.2.1 Analyze scatter plots for patterns, linearity, outliers, and influential points.

S.AD.4.2.2 Use residual plots to determine the reasonableness of a linear fit.

S.AD.4.2.3 Interpret the slope and intercept of a linear model in the context of the data.

S.AD.4.2.4 Using technology, compute and interpret the correlation coefficient of a linear fit.

S.AD.4.2.5 Understand the implications of extrapolating to make predictions.

S.AD.5 Students will make statistical inferences and evaluate claims from studies.

S.AD.5.1 Construct and interpret confidence intervals for the mean of a normally distributed population and for a population proportion.

S.AD.5.1.1 Explain how a sample statistic and a confidence level are used in the construction of a confidence interval.

S.AD.5.1.2 Explain how changes in the sample size, confidence level, and standard error affect the margin of error of a confidence interval.

S.AD.5.1.3 Construct a confidence interval for the mean of a normally distributed population with a known standard deviation.

S.AD.5.1.4 Construct a confidence interval for a population proportion.

S.AD.5.2 Use confidence intervals to evaluate claims.

S.AD.5.2.1 Evaluate claims for a single population parameter.

S.AD.5.2.2 Evaluate claims for the difference between two population parameters or treatment effects.

Interpret Results (IR)

Statistically educated students apply mathematical actions and processes to develop and evaluate inferences and predictions that are based on data.

S.IR.1 Students will interpret and communicate the results of a statistical analysis in context.

S.IR.1.1 Recognize when the difference between two sample proportions or two sample means is due to random variation or if the difference is statistically significant.

S.IR.1.2 Understand the concept of a confidence interval, including the interpretation of confidence level, margin of error, and statistical significance.

S.IR.1.3 Develop inferences or predictions to construct resulting decisions or recommendations.

S.IR.1.4 Create and evaluate recommendations for areas of future research.

S.IR.2 Students will evaluate practical implications of statistical significance, or lack thereof.

S.IR.2.1 Develop and critique arguments for practical implications based on statistical significance.

S.IR.2.2 Differentiate between correlation and causation when describing the relationship between two variables.

S.IR.2.3 Identify potential lurking variables that may explain an association between two variables.

S.IR.3 Students will critique news stories and journal articles to evaluate claims and conclusions.

S.IR.3.1 Evaluate strengths and weaknesses in the studies or methods used to generate the data.

S.IR.3.2 Evaluate the statistical validity of claims made in the story/article.

Probability (P)

Statistically educated students apply mathematical actions and processes to understand and apply basic concepts of probability.

S.P.1 Students will connect basic probability concepts to statistical analysis.

S.P.1.1 Describe events as subsets of a sample space.

S.P.1.2 Describe the relationship between theoretical and empirical probabilities using the Law of Large Numbers.

S.P.1.3 Determine probabilities, including joint probabilities, conditional probabilities, probabilities of independent events, probabilities of dependent events. Interpret the results.

S.P.1.3.1 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 two events are independent.

S.P.1.3.2 Understand and calculate the conditional probability of A given B as P(A and B)/P(B).

S.P.1.3.3 Interpret independence of A and B as saying that the conditional probability of A given B is the same as the probability of A.

S.P.1.4 Construct and analyze two-way frequency tables of data.

S.P.1.4.1 Use a two-way table as a sample space to decide if events are independent.

S.P.1.4.2 Use a two-way table to evaluate conditional probabilities.

S.P.1.5 Use counting techniques including permutations and combinations to solve mathematical and real-world problems, including determining probabilities of compound events.

S.P.2 Students will use probability to make decisions.

S.P.2.1 Calculate the expected value (weighted average) of a discrete random variable as the mean of its probability distribution.

S.P.2.2 Analyze decisions and strategies using probability concepts and expected values.

S.P.2.3 Analyze decisions about statistical significance based on reported p-values.