04 Intro to Two-Variable Statistics

Content Area: **Math**
Course(s):
Time Period: **Full Year**
Length: **4 Weeks**
Status: **Published**

General Overview, Course Description or Course Philosophy

This course combines the study of Statistics and Probability with Data Science. The goal is to have students think critically about data in today's data-driven world and understand its role in the 21st Century economy. Furthermore, students will become familiar with the concepts, topics, and techniques used by data scientists and statisticians in their day-to-day work.

Throughout this course, students will engage in project-based observational studies and experiments to develop their understanding of data analysis, sampling, correlation/causation, bias and uncertainty, probability, modeling with data, as well as making and evaluating data-based arguments. Students will also learn about the roles of data scientists, the power of data in society, machine learning, and how data scientists extract knowledge and insights from real-world data.

In this unit, students will understand:

A statistical scenario can include two or more variables. Many scenarios or real world models with two variables can be generalized using regression lines across different types of functions (linear, quadratic, exponential, etc.).

OBJECTIVES, ESSENTIAL QUESTIONS, ENDURING UNDERSTANDINGS

Essential Questions

- What are ways in which two variable data can be displayed in a way which it can be analyzed with ease?
- What are examples in real life where analyzing two variables are useful?

Enduring Understandings

- Real life situations can be modeled and the regression line can be useful for data that lies within the data range
- Regression lines can be used to model data outside of the data range however caution

should be used in this instance due to unpredictable future events.

CONTENT AREA STANDARDS

S.ID

A. Summarize, represent, and interpret data on a single count or measurement variable

B. Summarize, represent, and interpret data on two categorical and quantitative variables

C. Interpret linear models

S.IC

A. Understand and evaluate random processes underlying statistical experiments

B. Make inferences and justify conclusions from sample surveys, experiments, and observational studies

S.CP

A. Understand independence and conditional probability and use them to interpret data

B. Use the rules of probability to compute probabilities of compound events in a uniform probability model

S.MD

A. Calculate expected values and use them to solve problems

B. Use probability to evaluate outcomes of decisions

MA.9-12.I	Exploring Data: Describing patterns and departures from patterns
MA.9-12.I.A	Constructing and interpreting graphical displays of distributions of univariate data (dotplot, stemplot, histogram, cumulative frequency plot)
MA.9-12.I.B	Summarizing distributions of univariate data
MA.9-12.I.B.3	Measuring position: quartiles, percentiles, standardized scores (z-scores)
MA.9-12.I.C	Comparing distributions of univariate data (dotplots, back-to-back stemplots, parallel boxplots)
MA.9-12.I.D	Exploring bivariate data
MA.9-12.I.D.3	Least-squares regression line
MA.9-12.I.D.4	Residual plots, outliers and influential points

MA.9-12.I.E	Exploring categorical data
MA.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.
MA.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).
MA.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.

RELATED STANDARDS (Technology, 21st Century Life & Careers, ELA Companion Standards are Required)

9.1.12.RM.4: Determine when and why it may be appropriate for the government to provide insurance coverage rather than private industry. • 9.1.12.RM.5: Explain what self-insuring is and determine when it is appropriate. • 9.1.12.RM.6: Differentiate the costs benefits and features (e.g., riders, deductibles, umbrella policies) of renter's and homeowner's insurance.

LA.RI.11-12.7	Integrate and evaluate multiple sources of information presented in different media or formats (e.g., visually, quantitatively) as well as in words in order to address a question or solve a problem.
LA.11-12.SL.11-12.2	Integrate multiple sources of information presented in diverse formats and media (e.g., visually, quantitatively, orally) in order to make informed decisions and solve problems, evaluating the credibility and accuracy of each source and noting any discrepancies among the data.
CRP.K-12.CRP2	Apply appropriate academic and technical skills.
CRP.K-12.CRP4	Communicate clearly and effectively and with reason.
CRP.K-12.CRP7	Employ valid and reliable research strategies.
CRP.K-12.CRP8	Utilize critical thinking to make sense of problems and persevere in solving them.
CRP.K-12.CRP11	Use technology to enhance productivity.
TECH.8.1.12.A.4	Construct a spreadsheet workbook with multiple worksheets, rename tabs to reflect the data on the worksheet, and use mathematical or logical functions, charts and data from all worksheets to convey the results.
TECH.8.1.12.A.5	Create a report from a relational database consisting of at least two tables and describe the process, and explain the report results.
TECH.8.1.12.C.CS2	Communicate information and ideas to multiple audiences using a variety of media and formats.
TECH.8.1.12.C.CS4	Contribute to project teams to produce original works or solve problems.
TECH.8.1.12.E.CS2	Locate, organize, analyze, evaluate, synthesize, and ethically use information from a variety of sources and media.
TECH.8.1.12.E.CS4	Process data and report results.
TECH.8.1.12.F.CS3	Collect and analyze data to identify solutions and/or make informed decisions.

STUDENT LEARNING TARGETS

Declarative Knowledge

Students will understand that:

- Most statistical scenarios are composed of explanatory and response variables, though there are scenarios where variable classification and identification is unclear.
- A z-score is a numerical measurement describing a value's relationship to the mean of a group of values; the measurement is given in terms of standard deviations from the mean.
- Segmented bar charts are often used to display the relationship between two categorical variables.
- Correlation refers to the connection or mutual relationship between two variables; correlation is measurable.
- Causation indicates that one event is the result of the occurrence of another event.
- A least squares regression line is a line which makes the vertical distance from data points to the related regression line as as small as possible.
- The difference between the actual data point value of the dependent variable and the predicted value based on the regression line is known as the residual value.

Procedural Knowledge

Students will be able to:

- Find and interpret a z-score from data; relate the z-score to the standard deviation from the mean value of the data.
- Distinguish between explanatory and response variables for categorical data.
- Make a segmented bar chart to display the relationship between two categorical variables.
- Determine if there is an association between two categorical variables and describe the association if it exists.
- Determine the explanatory variable in a set of two categorical problems.
- Determine if there is an association between the explanatory and response data.
- Distinguish between explanatory and response variables for quantitative data.
- Create a scatterplot to display the relationship between quantitative variables.
- Describe the direction, form, and strength of a relationship displayed in a scatter plot and identify outliers.
- Estimate and interpret the correlation between 2 quantitative variables from a scatterplot.
- Distinguish correlation from causation.
- Calculate the actual correlation between 2 quantitative variables.
- Apply the properties of correlation.

- Describe how outliers influence/affect correlation and the least squares regression line.
- Use a graphing calculator to determine the least squares regression line from a set of data.
- Describe residual value.
- In the context of the data, interpret the slope and y-intercept of the least squares regression line.
- Explain what is meant by the standard deviation and r^2 in the context of the data.
- Use a residual plot to determine whether a regression model is appropriate.
- Determine the regression equation from non-linear data patterns, including exponential and quadratic.
- Determine if a model is suitable by creating and analyzing the residual statistical plot.
- Find a residual value from the regression equation.

EVIDENCE OF LEARNING

Alternate Assessments

- Portfolios
- Verbal Assessment (instead of written)
- Multiple choice
- Modified Rubrics
- Performance Based Assessments

Formative Assessments

Observations

Task completion

Student journals and notebooks

Cooperative team work

Summative Assessments

Project based learning

Unit assessments

RESOURCES (Instructional, Supplemental, Intervention Materials)

Statistics and Probability with Applications (High School) Third Edition, Starnes & Tabor, 2016

Digital Launchpad book companion

INTERDISCIPLINARY CONNECTIONS

Educational tech applications

ACCOMMODATIONS & MODIFICATIONS FOR SUBGROUPS

See link to Accommodations & Modifications document in course folder.