02- Exploring Two Variable Data

Content Area: **Mathematics**
Course(s): **AP Statistics**
Time Period: **November**
Length: **5-6 blocks**
Status: **Published**

Transfer Skills

Exam Weighting: 5–7%

Previous Coursework: Students may have been exposed to scatterplots and LSRL ("line of best fit"), but likely do not understand how the line is found or how to determine if it it a good fit.

Developing Understanding: Building on Unit 1 , students will explore relationships in two-variable categorical or quantitative data sets. They will use graphical and numerical methods to investigate an association between two categorical variables. Skills learned while working with two-way tables will transfer to calculating probabilities in Unit 4. Students will describe form, direction, strength, and unusual features for an association between two quantitative variables. They will assess correlation and, if appropriate, use a linear model to predict values of the response variable from values of the explanatory variable. Students will interpret the least-squares regression line in context, analyze prediction errors (residuals), and explore departures from a linear pattern.

Building Course Skills: In Unit 2, students are looking at the relationship between variables. The ability to calculate and describe statistical values, such as a conditional relative frequency or the slope of a regression line, is critical for data analysis because students must be able to analyze patterns before drawing conclusions about the data. Students should be allowed to perform their calculations using technology to help them become more aware of procedural errors. Students will also need practice translating output from technology ("calculator speak') into appropriate statistical language. As any statistician will assert, a numerical calculation is only as good as one's ability to interpret what it means in the real world. Rather than just reporting values from their calculations, students must be able to connect their numerical results to the scenario's context and formulate a verbal response that makes that connection clear. Teachers can model good communication and provide high-quality feedback to help students use accurate statistical language when comparing side-by-side bar graphs, for example, and to avoid common errors in reasoning, such as using the word "line" to explain why a relationship is linear.

Preparing for the AP Exam: Students need ongoing practice with interpretation of vocabulary and calculated values in context. It is typically not sufficient to speak generally about the direction of a relationship, for example. If the question is about a linear model for predicting the weight of a wolf based on its length, students should write that a positive relationship means that longer wolves tend to have higher weights (see 2017 FRQ 1). Students can communicate statistical uncertainty by using words such as "tend to have" and "on average," being careful to be precise with language. For example, when explaining evidence of a linear relationship, the difference between discussing a rate of change, as opposed to a change, is the difference between right and wrong. For the sake of clarity, the word "correlation" should be reserved for discussions

about relationships between two quantitative variables.

Enduring Understandings

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.

Essential Questions

How certain are we that what seems to be a pattern is not just a coincidence?

How does one data point compare to other data points in a distribution?

What kinds of conclusions can we make by using an LSRL?

Student Learning Objectives

TEXT Section 3.1: Scatterplots and Correlation

- Distiguish between explanatory and response variable for quantitative data.
- Make a scatterplot to display the relationship between two quantitative variables.
- Describe the direction, form, and strength of a relationship displayed in a scatterplot and identify any unusual features.
- Interpret the correlation.
- Understand basic properties of correlation, including how the correlation is influenced by unusual points.
- Distiguish correlation from causation.

TEXT Section 3.2: Least-Squares Regression

- Make predictions using regression lines, keeping in mind the dangers of extrapolation.
- Calculate and interpret a residual.
- Interpret the slope and y-intercept of a least squares regression line.
- Determine the equation of a least-squares regression line using technology or computer output.
- Construct and interpret residual plots to assess whether a linear model is appropriate.

- Interpret the standard deviation of the residuals (s) and the coefficient of determination (r^2) and use these values to assess how well the least squares regression line models the relationship between two variables.
- Describe how the least-squares residual line, standard deviation of the residuals, and r^2 are influenced by unusual points.
- Find the slope and y intercept of of the least squares regression line from the means and standard deviations of x and y and their correlation.

TEXT Section 3.3: Transforming to Achieve Linearity

- Use transformations involving powers, roots, or logarithms to create a linear model that describes the relationship between two quantitative variables, and use the model to make predictions.
- Determine which of several models does a better job of describing the relationship between two quantitative variables.

Vocabulary and Planned Learning Experiences

Vocabulary: scatterplot, explanatory variable, response variable, direction, positive association, negative association, no association, form, strength, unusual features, correlation (r), regression line, predict, slope, y-intercept, extrapolation, least-square regression line, residuals, residual plot, standard deviation of the residuals (s), coefficient of determination (r^2), influential points, high leverage points, outliers, transforming, power model, logarithm, exponential model

Planned Learning Experiences:

Quickwrite: Give students a scatterplot and its associated computer output. Have them identify and describe the meaning of the following values in the context of the problem: slope, y-intercept, coefficient of determination, and standard error of the residuals. Also have them calculate the correlation and explain how they found it. Have students compare their write-ups in groups of three to four.

Reversing Interpretations: Instead of asking students to interpret a residual, give them the residual and the equation of the least-squares regression line and ask them to make a prediction for a particular observation (e.g., “One wolf in the pack had a length of 1.4 m and a residual of −9.87. What does that −9.87 tell us about that particular wolf?”)

Build the Model Solution: Provide students with strips of paper containing portions of the model solution for 2018 FRQ 1 and have them work to assemble the phrases into a solution for the FRQ. Words can be grouped for part a, as follows: [The estimate of the intercept is] [72.95]. [It is] [estimated that] [the average time to] [finish checkout] [if there are no other customers in line] [is 72.95 seconds]. Additional numbers or phrases for part a could include [174.40], [is 174.50 seconds], and [the time to].

Predict and Confirm: Have students toss a handful of M&Ms and record how many land M side up. This is trial 1. Then have them remove the ones that were M side up. For trial 2, have students toss the remaining candies (the ones left over after removing the ones that landed M side up) and record how many land M side up on the second toss. Ask students to think about the trend and make a prediction: Will it be linear? A scatterplot of many trials should show a nonlinear relationship.

Resources

TEXT: The Practice of Statistics, 6th Edition

AP Classroom and the APCD 2019 Course Description

Rossman-Chance Applets

Stats Medic

Assessments

Reading Quiz: Scatterplots and Correlation (Introductory Level)

Reading Quiz: Least Squares Regression (Introductory Level)

Reading Quiz: Transforming to Acheive Linearity (Introductory Level)

Test: Exploring Two-Variable Quantitative Data (AP Level)

Standards

NJSLS Standards in Mathematics Copied and Pasted as well as linked.

NJSLS Standards - Mathematics

MA.S-IC.A.1 Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

MA.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.

MA.S-ID.B.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MA.S-ID.B.6a Fit a function to the data (including with the use of technology); use functions fitted to data to solve problems in the context of the data.

MA.S-ID.B.6b Informally assess the fit of a function by plotting and analyzing residuals, including with the use of technology.

MA.S-ID.B.6c Fit a linear function for a scatter plot that suggests a linear association.

MA.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.

MA.S-ID.C.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.

MA.S-ID.C.9 Distinguish between correlation and causation.

MA.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.
MA.S-ID.B.6	Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
MA.S-ID.B.6a	Fit a function to the data (including with the use of technology); use functions fitted to data to solve problems in the context of the data.
MA.S-ID.B.6b	Informally assess the fit of a function by plotting and analyzing residuals, including with the use of technology.
MA.S-ID.B.6c	Fit a linear function for a scatter plot that suggests a linear association.
MA.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.
MA.S-ID.C.8	Compute (using technology) and interpret the correlation coefficient of a linear fit.
MA.S-ID.C.9	Distinguish between correlation and causation.
MA.S-IC.A.1	Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

Modifications (ELL, Special Education, At-Risk Students, Gifted & Talented, & 504 Plans)

ELL:

- Use visuals
- Introduce key vocabulary before lesson
- Provide peer tutoring

- Guided notes and/or scaffold outline for written assignments

Supports for Students With IEPs:

- Allow extra time to complete assignments or tests
- Guided notes and/or scaffold outline for written assignments
- Work in a small group
- Follow all IEP modifications

At-Risk Students:

- Guided notes and/or scaffold outline for written assignments
- Introduce key vocabulary before lesson
- Work in a small group
- Lesson taught again using a differentiated approach
- Use visuals / Anchor Charts

Gifted and Talented:

- Create an enhanced set of introductory activities (e.g. advance organizers, concept maps, concept puzzles)
- Organize and offer flexible small group learning activities
- Teach cognitive and methodological skills
- Organize integrated problem-solving simulations
- Propose interest-based extension activities

Supports for Students With 504 Plans:

- Follow all the 504 plan modifications
- Text to speech/audio recorded selections