08- Inference for Distributions and Relationships

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

Transfer

NOTE: This is both Unit 8 and Unit 9 from the AP Course Framework.

Exam Weighting: 4-10%

Previous Coursework: NA

Developing Understanding: Unit 6 introduced inference for proportions of categorical data. Unit 8 introduces chi- square tests, which can be used when there are two or more categories. Students need to understand how to select from the following tests: the chi-square test for goodness of fit (for a distribution of proportions of one categorical variable in a population), the chi-square test for independence (for associations between categorical variables within a single population), or the chi-square test for homogeneity (for comparing distributions of a categorical variable across populations or treatments). To integrate conceptual understanding, teachers can make connections between frequency tables, conditional probability, and calculating expected counts. The chi-square statistic is introduced to measure the distance between observed and expected counts relative to expected counts.

Students may be surprised to learn that there is variability in slope. In their experience in previous courses, the slope of the line of best fit does not vary for a particular set of bivariate quantitative data. However, suppose that every student in a university physics course collects data on spring length for 10 different hanging masses and calculates the least-squares regression line for their sample data. The students' slopes would likely vary as part of an approximately normal sampling distribution centered at the (true) slope of the population regression line relating spring length to hanging mass. In this unit, students will learn how to construct confidence intervals for and perform significance tests about the slope of a population regression line when appropriate conditions are met.

Building Course Skills: In Unit 8, students should continue applying the same problem-solving structure to chi-square significance testing: State the hypotheses in words, explicitly identify the correct procedure, verify conditions, calculate the test statistic and the p-value, and then draw a conclusion in context that is directly linked to the p-value. Students should have opportunities to practice the distinctive elements for each type of chi-square test, such as analysis of expected counts, degrees of freedom, verbally stated hypotheses, and two-way tables. When thep-value is large, drawing an appropriate conclusion is challenging for students. Saying there is "no association" between two variables is equivalent to incorrectly "accepting the null hypothesis." Instead, teachers can teach students to use nondeterministic language in their conclusions, that is, "The data do not provide strong enough evidence to conclude that the variables are associated." Students

should have frequent opportunities to practice writing, with detailed feedback to help them improve.

In Unit 9, students should have multiple opportunities to practice interpreting the slope, y-intercept, r^2, standard deviation of the residual s, and standard error of the slope in context from computer output. They should refrain from using deterministic language such as "a I-foot increase in X is associated with a 0.445-point increase in Y," instead framing the association in terms of potential outcomes (i.e., "a predicted 0.445-point increase"). Students should also practice writing "increase" or "additional" for both variables, not just the dependent variable. Students should practice identifying what the question is asking or what needs to be solved. Without careful reading, students often provide answers that are not relevant or required, for example, conducting a significance test when the question does not call for one, or giving the expected number of successes or failures when asked to calculate a probability. Teachers can have them practice identifying the task before they begin, then checking that the response they've provided addresses the task.

Preparing for the AP Exam: When writing hypotheses, students should refer to the population, using language from the question. For example, "The null hypothesis is that the age group at diagnosis and gender are independent (i.e., they are not associated) for the population of people currently being treated for schizophrenia" (see Scoring Guidelines for 2017 FRQ 5). As always, students should name the test and provide evidence verifying appropriate conditions. For chi-square tests, the conditions are (1) random selection or randomized experiment and (2) large counts. Students should be sure to say that all expected counts (rather than actual counts) are at least 5. Students need to clearly present calculations and state the conclusion in context with linkage to p-values. Students should avoid tacitly accepting the null hypothesis. If the p-value is greater than conventional significance levels, the correct conclusion of a chi-square test for independence would be that there is insufficient evidence that there is an association.

Students should pay attention to timing as they work through full-length sections of past exams in order to leave enough time to complete the investigative task, which is weighted more heavily than the other free-response questions. The investigative task includes both familiar course content and questions requiring extended reasoning. As an example of a straightforward application of a topic from this unit, 2007 Form B FRQ 6 part b asks students to find a 95% confidence interval for the slope of a regression line. This familiar task gives students an opportunity to gain confidence and earn some credit, and it serves as an entry to subsequent parts of the question. Although the investigative task will require students to transfer course skills to unfamiliar settings, students who understand course content will have everything they need to complete the task.

Enduring Understandings

Given that variation may be random or not, conclusions are uncertain.

The chi-square distribution may be used to model variation.

Significance testing allows us to make decisions about hypotheses within a particular context.

An interval of values should be used to estimate parameters, in order to account for uncertainty.

The t-distribution may be used to model variation.

Essential Questions

How does increasing the degrees of freedom influence the shape of the chi-square distribution?

Why is it inappropriate to use statistical inference to justify a claim that there is no association between variables?

How can there be variability in slope if the slope statistic is uniquely determined for a line of best fit?

When is it appropriate to perform inference about the slope of a population regression line based on sample data?

Why do we not conclude that there is no correlation between two variables based on the results of a statistical inference for slopes?

Student Learning Objectives

TEXT Section 12.1 Chi-Square Tests for Goodness of Fit

- State appropriate hypotheses and compute the expected counts and chi-square test statistic for a chi-square test for goodness of fit.
- State and Check the Random, 10%, and Large Counts conditions for performing a chi-square test for goodness of fit.
- Calculate the degrees of freedom and P-value for a chi-square test for goodness of fit.
- Perform a chi-square test for goodness of fit.
- Conduct a follow-up analysis when the results of a chi-square test are statistically significant.

TEXT Section 12.2 Inference for Two-Way Tables

- State appropriate hypotheses and compute the expected counts and chi-square test statistic for a chi-square test based on data in a two-way table.
- State and check the Random, 10%, and Large Counts conditions for a chi-square test based on data in a two-way table.
- Calculate the degrees of freedom and P-value for a chi-square test based on data in a two-way table.
- Perform a chi-square test for homogeneity.
- Perform a chi-square test for independence.
- Choose the appropriate chi-square test in a given setting.

TEXT Section 12.3 Inference for Slope

- Check the conditions for performing inference about the slope ß of the population (true) regression line.

- Interpret the values of a, b, s, and SE_b in context, and determine these values from computer output.
- Construct and interpret a confidence interval for the slope ß of the population (true) regression line.
- Perform a significance test about the slope ß of the population (true) regression line.

Vocabulary and Planned Learning Experiences

Vocabulary: chi-square goodness of fit, observed count, expected count, chi-square test statistic, follow-up analysis, chi-square test for homogeneity, chi-square test for independence, degrees of freedom, t-interval for slope, t-test for slope

Planned Learning Experiences:

Simulation: Prepare several bags with an identical mix of at least 250 chips or beads of three colors in different proportions (e.g., red = 0.5, white = 0.3, blue = 0.2). Have each student take a random sample of 25 chips or beads from the bag, calculate the chi-square statistic and plot their value on a class dotplot. Use this graph to introduce the chi-square distribution with 2.

Discussion Groups: Give each group of three to four students an example of a chi-square test involving a two-way table. Have students work together to state appropriate hypotheses, describe a Type 1 and Type 2 error in context, and give a possible consequence of each of those errors.

Graphic Organizer: Have students work in teams of two to three to develop a chart that summarizes the three types of chi-square tests, including when each is appropriate, as well as the hypotheses, conditions, and degrees of freedom.

Note-Taking: Begin by having students use a chart to record the symbols for statistics and parameters that have been used previously to construct confidence intervals: Then, when constructing a confidence interval for the population slope parameter, have
students add a new row for the symbols for the sample slope and population slope. This will reinforce that the slope of the least-squares regression line is a sample statistic and can be used to estimate the population parameter slope.

Error Analysis: Give students some raw data on the distance and cost to fly from their hometown to various major cities. Then introduce some questions justifying a claim and error analysis. For example, how could you refute a claim that the average cost per mile (the population slope) is $0.50 per mile if you believe it to be false?

Notation Read Aloud: Have students read AP Exam questions aloud (e.g., 2011 FRQ 5, 2010 Form B FRQ 6, 2005 Form B FRQ 5, and 2001 FRQ 6), including the given notation. Remind students that the computer output provides the two-sided p-value, and that there are two different p-values in the chart: The top p-value is for the intercept, and the bottom p-value is for the slope. Then have students discuss each of the values in the computer output and carry out a test for the slope of a regression model.

Resources

TEXT: The Practice of Statistics, 6th Edition

AP Classroom and the APCD 2019 Course Description

Rossman-Chance Applets

Stats Medic

Assessments

Reading Quiz: Chi-Square Tests for Goodness of Fit

Reading Quiz: Inference for Two-Way Tables

Reading Quiz: Inference for Slope

TEST: Inference for Distributions and Relationships

Standards

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-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