AP STAT DEBRIEF – Question 5

Billy Esra
Bishop Hall Charter School

2017 QUESTION 5: Chi Square Test of Independence
Please view the questions here: https://apcentral.collegeboard.org/pdf/ap-statistics-frq-2017.pdf
You can find scoring guidelines here: https://secure-media.collegeboard.org/ap/pdf/ap17-sg-statistics.pdf
You can find the chief reader report here: https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap17-chief-reader-report-statistics.pdf

Question 5: INTENT OF THE QUESTION

"The primary goal of this question was to assess a student's ability to identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question. More specific goals were to assess a student's ability to (1) state appropriate hypotheses; (2) identify the appropriate statistical test procedure and check appropriate conditions for inference; (3) calculate the appropriate test statistic and *p*-value; and (4) draw an appropriate conclusion, with justification, in the context of the study." (https://secure-media.collegeboard.org/ap/pdf/ap17-sg-statistics.pdf).

SAMPLE SOLUTION and COMMENTS:

This question was a very straight forward Chi-Square Test of Independence question. Excluding the investigative task (question 6), question 5 had the lowest mean score (1.51). The inclusion of a segmented bar graph distracted some students from performing a chi-square test, but most students who attempted the question chose one of the chi-square tests to perform. Students had varying level of success on carrying out the test.

a. The student had to identify the correct hypothesis test by name or formula, state the correct pair of hypotheses, and check the correct conditions.
    - Students could identify the test as the chi-square test of independence (or association) to name the correct test. Without explicitly naming the correct test, they could also get credit for naming if they had the correct formula somewhere in their response. The formula could be just the symbolic formula or the formula with the correct numbers substituted. Using the formula instead of a name only worked if there were no mistakes.
    - Students who wrote Chi-Square (with no formula or errors in the formula), Chi-Square Test of Homogeneity, or Chi-Square Goodness of Fit were not awarded credit for the correct name.

- At least one of the hypothesis needed to be in context and the hypothesis had to represent the population, not the sample.
- The hypotheses should be about the population, not the sample.
- For conditions of inference:
  - Expected counts were to be reported and checked. Students were expected to list all expected counts and note that they were all at least 5. Alternately, they could show that the smallest expected count was at least 5.
  - Students could not include any incorrect conditions, such as Normality.
  - Random sample condition was stated in the prompt so it did not have to be checked by the student.

b. The student had to correctly calculate the chi-square test statistic and calculate the correct *p*-value (or *p*-value range from the table).
   - Students did not have to show any calculations to get credit for the test statistic and the *p*-value. But if they did the work had to be correct (even if the correct values were given).
   - If students made an error along the way (for example using the wrong degrees of freedom), they would lose credit for one component but not both (as long as their answer followed correctly from their miscalculation).
   - Students completely lost credit for this part of their response if they reported the wrong type of test statistic (like *t* or *z*) even if the *p*-value was correct.

c. The student had to write the correct conclusion in context about the alternative hypothesis and justify the conclusion by linking the p-value with a reasonable alpha level.
   - The conclusion could not be about the null hypothesis.
   - Students who found the wrong *p*-value could still get full credit for this part of their response if they made the correct conclusion based on their incorrect *p*-value.
   - Students should not use absolute language in the conclusion (for example, I conclude that there is an association …). There should be some level of uncertainty. There is evidence, or there is not evidence, of an association.

NOTES/COMMON MISTAKES

1. The segmented bar graph was a big distractor for statistically weak students (or perhaps students who had no experience with chi-square). If a student started writing about the segmented bar graph, they almost never recovered and almost universally earned a score of zero.
2. In years past, students did not always have to explicitly say chi-square for homogeneity or chi-square test for association/independence. In fact, they sometimes did not have to name the test at all. If they just typed $\chi^2 = \#$, that was sometimes enough to get credit

for naming the test. But this year, students were expected to choose the correct chi-square test.

3. Many students incorrectly found the expected counts by assuming that expected number would be the same for each cell in the table instead of correctly using the formula for expected counts or expected value matrix in calculator.
4. Some students lost points for conditions by listing the conditions for other inference procedures like Normal, n>30, and np>10. Any incorrect condition checked would cost the students points. The word Normal anywhere in the response would cost them the conditions part of their response. Some students would draw a Normal curve instead of a chi-square curve. The curve alone would not cause them to lose points, but if the curve included a scale with 0 in the middle or was shaded on both sides the student would lose points since the distribution is not Normal.
5. Degrees of Freedom was not explicitly part of the rubric, but if students wrote the wrong degrees of freedom down, they would lose credit for the test statistic (even if it would be the correct answer if they had used the correct degrees of freedom).
6. Linking the p-value and significance level to the decision in context proved difficult for many students. Students could give the correct interpretation of the p-value or explain how the conclusion follows from the p-value and not mention alpha levels (or significance levels) at all.

TEACHER SUGGESTIONS

I have to admit that this is the second year in a row that I was assigned to a chi-square question at the reading. I read and scored the Choco- and Apple-Zuties question last year as well as this chi-square question this year. There were some major differences between the two rubrics which is leading me to rethink how I teach chi-square. When I first realized that I would be reading this question, I was excited because I thought that students would be very successful with this chi-square question. Chi-square is often one of the last topics covered and students generally seem to do well with that topic in my classroom. I was very, very wrong (as I usually am about how students will respond to questions on the test). To me, chi-square questions tend to look a lot like chi-square questions. This question included a segmented bar graph to distract students. Students could use the graph to supplement their chi-square test conclusion, but the question could be answered without referring to the graph at all.

As an AP Statistics teacher, I plan to do a better job at working with students on choosing the appropriate test after finishing up the inference unit. Perhaps I'm too quick to begin review, at large, and could help them with choosing the right test for the right situation. Students should have recognized the categorical variables and immediately thought chi-square when asked "provide convincing statistical evidence." Without performing a chi-square test here, students generally would score a zero on the question.

Perhaps students need a more formulaic method for organizing inference questions. I've started using a "four-corners" approach where students draw two intersecting lines to organize

what goes where. I heard about this at the Best Practices night at the reading a few years ago, and I've started using it with students. So far, anecdotally, it appears to be helping them remember all of the components of an inference procedure answer. It was clear when reading exams that students who were taught some kind of organizational strategy for inference tended to do better on this questions than the students whose responses were more disorganized.

To much consternation of statistics teachers, the rubrics for inference procedure questions have not required students to use formulas. Just reporting the necessary statistics has been enough to earn credit for calculations. But if students do use formulas, they must use the correct formula and fill it in correctly. I would say that many students lost points for trying to write the formula or trying to fill in the formula. Generally speaking, the more that a student writes in their response, the more likely they were to include something that would cause their score to decrease. Reporting only the statistics themselves seems to be a consistent rubric element. Not showing work seemed to work to the advantage of many students. Alternately, showing work correctly (while not required) could get students points for another part of the rubric. For example, if a student wrote the formula correctly and filled in the expected counts into the formula, they could get credit for listing the expected counts even if they had not listed them elsewhere. Also, as mentioned previously writing the correct formula or correctly filling in the counts into the formula, could get students credit for the name of the test.

I think that is important for students to understand why they are checking the conditions for inference. In this case, the expected counts are what we would expect if the two variables were in fact independent. We are therefore testing if the observed counts are far enough from those expected counts, that we have evidence of an association. Students who were just listing conditions often included something that did not make sense for this type of question. Understanding why conditions must be met may help them when deciding which conditions to use for each problem. Understanding why you check for Normality of the sampling distribution for other inference procedures might sway students away from writing it as a condition for chi-square (especially since chi-square is always right skewed).

I loved this question. It was a very straight forward application of the chi-square test for independence, but I really wanted students to perform better than they did, overall. Chi-square questions show up in the AP Statistics FRQ questions almost every year. Though it is always possible that next year's test will not have a Chi-square question, I would make sure that you have built in time in the pacing of your course to cover it well. Chi-square questions could be an easy way for your students to distinguish themselves on the test.