Student Performance Q&A:

2013 AP® Statistics Free-Response Questions

The following comments on the 2013 free-response questions for AP® Statistics were written by the Chief Reader, Allan Rossman of California Polytechnic State University – San Luis Obispo. They give an overview of each free-response question and of how students performed on the question, including typical student errors. General comments regarding the skills and content that students frequently have the most problems with are included. Some suggestions for improving student performance in these areas are also provided. Teachers are encouraged to attend a College Board workshop to learn strategies for improving student performance in specific areas.

Question 1

What was the intent of this question?

The primary goals of this question were to assess a student's ability to (1) use a stem-and-leaf plot to answer a question about a distribution of data; (2) identify and compute an appropriate confidence interval after checking the necessary conditions; and (3) interpret the interval in the context of the data.

How well did students perform on this question?

The mean score was 2.27 out of a possible 4 points, with a standard deviation of 1.10.

What were common student errors or omissions?

Part (a):

- Very few students made substantive errors on this part.

Part (b), Step 1: Identification of procedure, check of conditions:

- Some students had difficulty with identifying the correct procedure; many of these students thought that a one-proportion z-interval was appropriate.
- Students who successfully generated a list of conditions to be checked were not always able to check them correctly.
- Many students mistakenly thought that a sample size less than 30 determined that the normality condition was not satisfied, without realizing that with a small sample size it's

important to check whether the sample data indicate that the population distribution might reasonably be considered normal.

- A surprising number of students entered the data into a calculator and produced a graphical display in addition to the stem-and-leaf plot provided in the question. This was not an error, but was an unnecessary step that wasted valuable time.

Part (b), Step 2: Mechanics of calculating the interval:

- Many students calculated a *z*-interval rather than a *t*-interval.
- Many students used their calculator to determine the correct *t*-interval, but some of these students wrote incorrect supporting work along with the correct answer.
- Some students used incorrect notation with their supporting work, such as writing σ rather than *s* and writing μ rather than $\bar{x}$.

Part (b), Step 3: Interpreting the interval:

- Some students mistakenly thought that the confidence interval estimated the lead level for an *individual* crow, rather than the mean lead level in the population of crows.
- Some students mistakenly thought that the confidence interval estimated the mean lead level in the *sample* of 23 crows, or for a future *sample* of crows, rather than for the population of crows.
- Some students neglected to put their interpretation in the context of lead levels in crows.
- Some students attempted to interpret the confidence *level* rather than the confidence interval.

Based on your experience of student responses at the AP Reading, what message would you like to send to teachers that might help them to improve the performance of their students on the exam?

Provide many opportunities for students to practice with identifying the appropriate inference procedure to address a particular research question. Also emphasize that *t*-procedures are used with *quantitative* data, where the relevant parameter is a population *mean*. Give students frequent opportunities to practice with identifying such procedures both by name and by formula.

Emphasize to students not only what the validity conditions are for particular inference procedures, but also how to check whether the conditions are satisfied in a given context. Students could also benefit by learning about *why* checking validity conditions is necessary.

Emphasizing and giving feedback on proper use of statistical notation is also important, for example in helping students to distinguish between μ and $\bar{x}$ and also between σ and *s*.

Teachers cannot emphasize enough the importance of identifying a parameter clearly in context. This is a challenging task for many students, requiring frequent practice and feedback.

Help students to recognize the difference between interpreting a confidence *interval* and interpreting a confidence *level*. The former must include a mention of the specific interval (endpoints) calculated and must also be clear about what the relevant parameter is. The latter involves mentioning what would happen in the long run under repeated random sampling.

Question 2

What was the intent of this question?

The primary goals of this question were to assess a student's ability to (1) recognize and explain why a particular sampling method is likely to be biased; (2) describe a method for selecting a simple random sample from a population using a computer random number generator; and (3) demonstrate an understanding of the principle of stratification by describing circumstances in which one stratification variable would be better than another.

How well did students perform on this question?

The mean score was 0.93 out of a possible 4 points, with a standard deviation of 0.97.

What were common student errors or omissions?

Part (a):

- Some students wrote about the 500 students in the convenience sample without comparing them to the larger population of students at the university.
- Some students inappropriately compared the 500 students in the convenience sample to students who did not attend the game.
- Many students did not link the characteristics of the students in the convenience sample to the variable of interest: opinion about appearance of university buildings and grounds.
- Many students neglected to mention bias at all, also failing to refer to the parameter of interest.
- Many students supplied generic responses about sampling bias without referring to the study of interest.

Part (b):

- Many students mentioned using a computer to generate 5-digit numbers, but did not specify to ignore values larger than 70,000.
- Many students neglected to indicate what to do with repeated values in the randomly generated list.
- Some students correctly wrote that 500 unique random integers between 1 and 70,000 should be generated, but then stopped short of saying how to use those numbers to select sample from the population of students.
- Some students ignored the instruction to use a computer, instead describing how to use a random digit table or a very large hat to select the sample of 500 students.

Part (c):

- Many students correctly responded that stratification by campus would be preferable if opinions differed greatly between the two campuses, but neglected to go on to mention that opinions would have to differ less between genders in order for campus to be the preferred stratification variable.
- Many responses did not mention the variable of interest: opinion about appearance of university buildings and grounds.
- Many responses described differences between the campuses with regard to buildings and grounds without referring to students' opinions about appearances or to differences in opinions between genders.

- Some responses only stated that stratification would allow for separate estimates for each campus, without addressing why stratifying by campus would be preferred to stratifying by gender.

Based on your experience of student responses at the AP Reading, what message would you like to send to teachers that might help them to improve the performance of their students on the exam?

Emphasize that students' responses must be expressed in the context of the study. Students generally struggle with providing sufficient detail in questions related to sampling designs and flaws in sampling methods, so teachers should provide many opportunities for students to develop such communication skills and should also hold students to high standards for their descriptions and explanations. Note for students the importance of thinking about and referring to the variable and parameter of interest when answering questions about possible sampling bias.

When asking students to describe a sampling method, help students to realize that their responses need to provide enough detail that someone else could implement the sampling method based *solely* on the description. One strategy for reinforcing this skill might be to have students exchange their descriptions with each other, to see if the other student can follow the description well enough to implement it.

With regard to the topic of stratification, help students understand not only how to select a stratified random sample but *why* such a sampling method might be advantageous. Emphasize that stratification variables are chosen for heterogeneity across strata and homogeneity within strata. Students also need guidance in realizing that when they are asked to select one option over another, their justification should not simply describe why the preferred option is good but must also include a comparison between the two options.

Question 3

What was the intent of this question?

The primary goals of this question were to assess a student's ability to (1) calculate a probability from a normal distribution and (2) apply properties of means and variances of functions of random variables.

How well did students perform on this question?

The mean score was 1.61 out of a possible 4 points, with a standard deviation of 1.05.

What were common student errors or omissions?

Part (a):

- Many students performed the calculation correctly, but some did not justify or communicate their answer as fully as desired, e.g., by not making clear that they used a normal distribution, or by not clearly specifying the parameter (mean and standard deviation) values of the distribution, or by not specifying the boundary value and direction (greater than 850).
- Many responses used incorrect statistical notation, using $\bar{x}$ in place of μ and/or s in place of σ.

- Some students reversed the roles of the boundary value and the mean when calculating a *z*-score, subtracting in the wrong order in the numerator of the *z*-score.
- Some students appeared to consider the question as involving a sample mean, calculating a *z*-score in which the denominator contained the standard deviation σ divided by the square root of 12 (or 13).

Part (b, i):

- Many students did not set up an equation to be solved.
- Some students obtained the correct answer, but did not indicate how they obtained the answer.

Part (b, ii):

- Many students did not set up an equation to be solved, in some cases leading to adding the two variances that should have been subtracted.
- Some students mistakenly considered the total weight of the 12 eggs to be the random variable $12X$ rather than the correct random variable $X_1 + X_2 + \ldots + X_{12}$.
- Some students combined standard deviations rather than variances.

Based on your experience of student responses at the AP Reading, what message would you like to send to teachers that might help them to improve the performance of their students on the exam?

When answering a question involving a probability distribution, students should be sure to name the distribution, identify parameter values, and indicate the region whose probability is being calculated, along with providing the correct answer. When working with a normal probability distribution, these components can be provided in a well-labeled sketch. Provide frequent feedback on how well students communicate these components.

When working with expected values and variances of random variables, urge students to write out equations that illustrate their method of solution and also to apply rules of expected values and variances correctly. Teachers can help to facilitate this by giving many types of questions for students to solve, including some (as in this question) where the desired quantity needs to be solved for algebraically.

Question 4

What was the intent of this question?

The primary goal of this question was to assess students' 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 students' 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.

How well did students perform on this question?

The mean score was 1.70, out of a possible 4 points, with a standard deviation of 1.57.

What were common student errors or omissions?

- Some students did not realize that this question called for a hypothesis test, as they performed a descriptive analysis by calculating sample proportions and basing their conclusion on them.

Step 1: Stating hypotheses:

- Some students reversed the hypothesis, mistakenly stating the null hypothesis as indicating an association between the variables and the alternative as indicating no association.
- Some students did not refer to a population in their hypotheses.
- Some students did not state hypotheses in the context of this study.
- Some students attempted to state hypotheses in terms of inappropriate parameters such as μ or β or ρ .
- Some students mistakenly described hypotheses in terms of inappropriate or incorrect words such as "correlation" or "effect."

Step 2: Identification of procedure, check of conditions:

- Some students stated an appropriate validity condition in terms of expected counts, but did not clearly demonstrate that the condition had been checked numerically.
- Some students neglected to mention that the condition of having selected a random sample was satisfied.
- Some students listed incorrect or inappropriate conditions, involving normality or Central Limit Theorem or sample size greater than 30.
- Some students made an error in giving the formula for the chi-square test statistic, such as omitting the summation notation or not squaring terms.
- Some students mistakenly referred to the appropriate test procedure as a chi-square test of homogeneity of proportions, rather than a chi-square test of independence or association.

Step 3: Mechanics of calculating test statistic and *p*-value:

- Very few students made substantive errors on this part.
- Some students mistakenly included the row and column totals as cell counts.

Step 4: Summarizing conclusion, in context, based on linkage to *p*-value:

- Some students presented a conclusion that was not consistent with their hypotheses.
- Some students neglected to express the conclusion in the context of this study.
- Some students did not justify the conclusion by linking it to the *p*-value.
- Some students expressed a correct test decision (such as "reject H_0"), but then presented a verbal summary that contradicted that decision.
- Some students presented a conclusion indicating that the alternative hypothesis had been proven to be correct.
- Some students attempted to interpret the *p*-value as the probability of obtaining such an extreme test statistic if there were no association between the variables, but mistakenly omitted part of the interpretation or got it wrong.

Based on your experience of student responses at the AP Reading, what message would you like to send to teachers that might help them to improve the performance of their students on the exam?

By providing many examples and exercises, teachers should help students to recognize questions for which statistical inference is appropriate and necessary, even when the question does not specifically ask for a hypothesis test to be performed. Strive to help students learn not only the steps involved with a hypothesis test, but also help students to understand the overall reasoning process and how those steps relate to each other. Teachers are also encouraged to provide very detailed feedback on student performance with this task.

Students should also receive frequent reminders that hypotheses are always stated in terms of a population (or populations), not in terms of a sample. Teachers should make students aware of the importance of always *checking* conditions for inference, based on the specific details of the study at hand, rather than merely *stating* assumptions for inference, when conducting a significance test or producing a confidence interval. Make students aware that these checks require examination of the sample data and consideration of how the data were collected.

Students should also receive considerable practice and feedback with summarizing conclusions from hypothesis tests. Encourage students to be very clear in stating how their conclusion follows from the p-value. Students should also receive frequent reminders about the need to express conclusions in the context of the study described in the question.

Question 5

What was the intent of this question?

The primary goals of this question were to assess a student's ability to (1) recognize the limited conclusions that can be drawn from an observational study; (2) determine whether a condition for applying a particular inference procedure is satisfied; and (3) draw an inferential conclusion from a simulation analysis.

How well did students perform on this question?

The mean score was 0.57 out of a possible 4 points, with a standard deviation of 0.72.

What were common student errors or omissions?

Part (a):

- Some students replied that drawing a cause-and-effect conclusion is reasonable because the result of the study was statistically significant.
- Many students correctly responded "no, it would not be reasonable to draw a cause-and-effect conclusion", but did not provide a justification.
- Many students appealed to the general idea of a confounding or lurking variable without explaining how that variable prevents drawing a cause-and-effect conclusion.
- Some students argued that significance tests can only establish causation, not cause-and-effect conclusions, without being clear that the design of the study is what determines whether a cause-and-effect conclusion can be drawn from a statistically significant result.
- Some students based their justification on the small sample sizes failing to meet validity conditions for a two-sample z-test, without realizing that the statistically significant result could have been established with an appropriate inference procedure.

- Some students used the word "correlation," which is not strictly correct because the variables in this study were categorical, rather than the more general term "association."

Part (b):

- Many students stated a correct condition to check, but did not verify that the condition had been checked correctly by plugging in appropriate values from the study.
- Some students calculated a reasonable value to check but then neglected to indicate the boundary value (for example, 5 or 10) with which to compare their calculated value.
- Some students mentioned an inappropriate validity condition, such as $n > 30$ or normally distributed population.

Part (c):

- Many students did not take into account the observed data from the study, e.g., by not calculating the observed difference in the success proportions between the two groups to be $\frac{0}{11} - \frac{8}{17} \approx -0.47$ and then using that value to determine the approximate *p*-value from the simulation results.
- Many students did not appear to understand the role of the simulation results in the inference process, e.g., by simply describing the shape, center, and variability of the distribution of simulation results.
- Many students continued to express concerns about the small sample sizes, apparently not realizing that the simulation analysis took the sample sizes into account and eliminated the need for a normally distributed sampling distribution.
- Some students calculated the approximate *p*-value from the simulation analysis correctly, but failed to compare that *p*-value to a common significance level or to say that the *p*-value was small.
- Some students analyzed the data and simulation results correctly and reached the correct decision to reject the null hypothesis, but neglected to state their conclusion in the context of the study.
- Some students analyzed the data and simulation results correctly and reached the correct decision to reject the null hypothesis but then stated a two-sided conclusion rather than the appropriate one-sided conclusion.

Based on your experience of student responses at the AP Reading, what message would you like to send to teachers that might help them to improve the performance of their students on the exam?

Teachers should clarify to students that whenever a question calls for a choice to be made, that choice must be justified with statistical arguments. A very important idea to emphasize to students is that whether a cause-and-effect conclusion can reasonably be drawn from a study depends on how the study was designed, specifically on whether the subjects were randomly assigned to groups. With regard to the issue of confounding, help students realize that explaining how a variable is confounding requires giving a plausible connection between the confounding variable and the explanatory variable, and also between the confounding variable and the response variable.

Teachers should also help students to understand why validity conditions need to be checked before applying an inference procedure. Encourage students to check validity conditions based on the actual data obtained in the study. The reasons behind checking validity conditions should also be emphasized to students.

Teachers cannot emphasize enough the reasoning process behind statistical inference, specifically behind statistical significance and *p*-values. Teachers should present simulation analyses often, and ask students to conduct their own simulation analyses, and also emphasize how inferential conclusions are drawn from such simulation analyses: Assess whether a result as extreme as the observed data occurs rarely in the simulation results that would reveal that such a result would rarely happen by chance alone, which would provide strong evidence against the null hypothesis. Students should also be helped to understand that simulation analyses do not require conditions about approximate normality and that they should always justify decisions by comparing the p-value to a standard significance level or else appeal to the relative size of the *p*-value.

Question 6

What was the intent of this question?

The primary goals of this question were to assess a student's ability to (1) summarize information provided in a time plot that involves trend components; (2) perform calculations related to a summary statistic not previously studied; and (3) compare and contrast information conveyed by the summary statistics with the data.

How well did students perform on this question?

The mean score was 2.14, out of a possible 4 points, with a standard deviation of 1.04.

What were common student errors or omissions?

Part (a):

- Most students provided a good response that compared the centers of the two distributions, the variability of the two distributions, and did so in context.
- Some students confused year-to-year variability revealed in the time series plots with variability in the distributions of frequencies.
- Some students made correct observations about the centers and variability of the two distributions, but neglected to *compare* the centers and variability.

Part (b):

- Many students focused too much on year-to-year fluctuations, often providing a detailed list of such fluctuations, without describing any overall trends.
- Many students used very imprecise language in attempting to describe trends.

Parts (c) and (d):

- Most students completed these parts correctly.
- Some students neglected to plot the point on the graph.
- A few students miscalculated the moving average as the average of other moving averages values, rather than as an average of data values.

Part (e):

- Many students identified reasonable information for both sub-parts, but did not clearly link the information back to aspects of Graph B.
- Some students' descriptions of trends and of reduced variability were too vague and imprecise to be scored as essentially correct.

Based on your experience of student responses at the AP Reading, what message would you like to send to teachers that might help them to improve the performance of their students on the exam?

Teachers should encourage students that when asked to compare two distributions of data, to use comparative language rather than supply a "laundry list" of features. Students should also be given considerable practice, and model solutions, for identifying overall tendencies without dwelling too much on individual fluctuations from those tendencies.

Caution students to read questions carefully to make sure that they address the question asked.