AP®

Student Performance Q&A:

2014 AP® Statistics Free-Response Questions

The following comments on the 2014 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) calculate conditional proportions from a two-way table; (2) comment on association between two categorical variables as displayed in a graph; and (3) draw an appropriate conclusion from the *p*-value of a chi-square test.

How well did students perform on this question?

The mean score was 1.96, out of a possible 4 points, with a standard deviation of 1.08.

What were common student errors or omissions?

Part (a):

- Some students gave a decimal answer without showing the fraction that produced the answer.
- Some students gave the proportions with *exactly* one activity rather than *at least one* activity.

Part (b):

- Many students described the two distributions separately without using *comparative* language to compare the two distributions.
- Some students used comparative language but were not clear about what was being compared to what, for example saying that the proportion with no activities was larger, without specifying larger *than what*.
- Some students used language suggesting that they were comparing *counts* rather than *proportions*, for example mistakenly saying that more on-campus students than off-campus students participated in two or more activities.
- Some students compared the two groups on only *one* category of the response variable, for example saying only that a higher proportion of off campus than on campus students participated in no activities, without comparing the proportions with one activity or with two or more activities.

Part (c):

- Many students wrote about *accepting* the null hypothesis for their conclusion.
- Some students wrote a conclusion in context that was equivalent to accepting the null hypothesis, for example concluding that there is no association between living situation and participation in extra-curricular activities.
- Some students neglected to state a conclusion in the context of the study about students' living situation and participation in extra-curricular activities.
- Some students did not justify their conclusion by comparing the given *p*-value to a standard significance level or by otherwise referring to the magnitude of the given *p*-value.

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 frequently remind students that they must show work behind numerical calculations.

Teachers should provide many opportunities for students to write paragraphs comparing the distributions of *categorical* variables between two (or more) groups. It is probably more common to give practice with comparing distributions of *quantitative* variables.

Teachers should emphasize to students the importance of proportional reasoning, using proportions rather than counts to make comparisons between groups. Equally important is communicating clearly that proportions should be considered when comparing distributions of categorical variables. Also crucial when writing a comparative paragraph is to be very clear about which group is being compared to which other group.

Drawing an appropriate conclusion from a hypothesis test that results in a large *p*-value is challenging for students. Teachers should make sure that they provide students with such opportunities and provide detailed feedback. In particular, teachers should make clear to students the distinction between *accepting* a null hypothesis and *failing to reject* a null hypothesis, ideally by using many different contextual wordings of each type. In this context the distinction is between concluding that the data provide no evidence of association, as opposed to claiming mistakenly that the data provide evidence of no association. Teachers should also try to help students understand why the logic of significance testing assesses evidence *against* the null hypothesis rather than *for* the null hypothesis.

Teachers should require students to draw conclusions in context from inference procedures. Students should also be encouraged to justify conclusions that they reach from a *p*-value, either by comparing the *p*-value to a common significance level or by commenting on the magnitude of the *p*-value.

Question 2

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; (2) assess whether a claim about randomness is questionable in light of a calculated probability; and (3) judge whether a description of a simulation method achieves a correct simulation of a random process.

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 1.36.

What were common student errors or omissions?

Part (a):

- Many students calculated the probability as if the sampling had been conducted *with* replacement rather than *without* replacement, mistakenly obtaining $\frac{3}{9} \times \frac{3}{9} \times \frac{3}{9}$.
- Some students appeared to recognize that the sampling was conducted without replacement but mistakenly calculated $\frac{1}{9} \times \frac{1}{8} \times \frac{1}{7}$.

Part (b):

- Many students did not justify their conclusion by describing the probability as small or by comparing the probability to a common significance level such as 0.10 or 0.05.
- Some students declared the probability to be small even when they had mistakenly calculated the probability to be a non-small value.
- Some students essentially ignored the probability by arguing that selecting all women is possible with random selection, so there is no reason to doubt the manager's claim about random selection.
- Some students stated their conclusion too strongly by asserting that the manager had *definitely* not selected the people at random.
- Some students mistakenly referred to the probability calculated in part (a) as the probability that the manager had selected people at random.

Part (c):

- Many students did not notice that the simulation method described is only appropriate for sampling *with* replacement and so mistakenly replied that the simulation method is reasonable because $\frac{2}{6} = \frac{3}{9}$.
- Some students simply responded that a die with six sides cannot be used to represent random selection from nine people.
- Some students correctly stated that the die cannot replicate sampling without replacement, but neglected to be clearer in explaining why this is so.

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 provide students with opportunities to calculate probabilities in situations where sampling is conducted *without* replacement. In other words, students should work with situations where the assumption of independent observations is not appropriate.

Teachers should make sure to expose students to probability questions in realistic situations rather than solely abstract ones involving dice and cards. Even a somewhat contrived but realistic situation can help students to understand how and why probabilities matter in everyday life.

Teachers must emphasize to students the logical reasoning process behind the argument that if an observed result is unlikely to occur under a particular model/assumption/hypothesis, then the occurrence of that observed result provides strong evidence against that model/assumption/hypothesis. This is the reasoning process of statistical significance, which teachers should be sure to present to students in ways that are not purely mechanical.

Students need considerable practice with writing coherent, logical arguments. Teachers should offer these opportunities and provide detailed feedback. Teachers should also resist the temptation to force students to follow a fill-in-the-blanks template for such arguments, so students can develop their own understanding of such logical arguments.

With regard to simulation methods, teachers should ask students to design their own simulation methods to model various random phenomena. Teachers should be sure that students work with situations involving independent events, such as coin tosses or die rolls or sampling with replacement, and also situations involving dependent events, such as sampling without replacement.

Question 3

What was the intent of this question?

The primary goals of this question were to assess a student's ability to (1) perform a probability calculation from a normal distribution; (2) explain an implication of examining the distribution of a sample mean rather than the distribution of a single measurement; and (3) perform a probability calculation involving independent events using the multiplication rule.

How well did students perform on this question?

The mean score was 1.43, out of a possible 4 points, with a standard deviation of 1.10.

What were common student errors or omissions?

Part (a):

- Many students did not clearly state that they were using a normal distribution, either with the word "normal" or with a clearly labeled sketch.
- Many students did not explicitly identify the parameter values (mean and standard deviation) of the normal distribution.
- Some students subtracted in the wrong order in the numerator of the *z*-score calculation, essentially reversing the roles of x and μ.

Part (b):

- Many students did not explicitly answer the question (more likely, less likely, or equally likely).
- Many students did answer the question but provided no justification.
- Many students did not make clear that the distribution of a sample *mean* was relevant.
- Some students mistakenly argued that the distribution of the sample mean could not be determined because the sample size is less than 30.
- Some students calculated $(0.0287)^3$, which is the probability that the number of absences exceeds 140 on all 3 days, rather than the probability that the sample mean would exceed 140.

Part (c):

- Some students mistakenly applied the complement rule, calculating the probability that none of the events would occur as one minus the probability that all of the events would occur.
- Some students mistakenly used a binomial probability calculation with $n = 5$ or $n = 15$ rather than $n = 3$.

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 remind students frequently that when working with probability distributions, they should clearly state the name of the probability distribution and clearly identify its parameter values. Teachers should grade student work closely on this point to encourage clear communication.

When answering questions based on normal distributions, students should be encouraged to draw a well-labeled sketch that identifies the parameter values and shades in the region whose probability is being calculated. Students should refer to the shaded area as a check on the reasonableness of their calculation.

When calculating a z-score, students should be encouraged to express the general formula and also write out the components being subtracted and divided.

Teachers should insist that students answer the question asked, especially when the question asks about which of three options (for example, more likely, less likely, or equally likely) is the case. Teachers should also make sure that student realize that justifying such an answer is also crucial.

Teachers should help students to understand the fundamental distinction between the distribution of a population and the sampling distribution of a statistic such as a sample mean. One strategy for helping students to distinguish between these distributions is to draw well-labeled sketches of the two distributions that specify the variables being represented by the sketches. Related to this is that teachers should emphasize to students that the standard deviation of a sample mean is the population standard deviation divided by the sample size, regardless of how large the sample size is and even regardless of the shape of the population distribution.

Teachers are encouraged to ask students to work through probability questions, like this one, that use a variety of probability rules. Teachers should also urge students to always ask whether their answer to a probability question seems reasonable, especially to question probability calculations that result in an answer larger than one!

Question 4

What was the intent of this question?

The primary goals of this question were to assess a student's ability to (1) describe why the median might be preferred to the mean in a particular context; (2) compare the relative merits of two sampling plans; and (3) describe a consequence of nonresponse in a particular study.

How well did students perform on this question?

The mean score was 1.79, out of a possible 4 points, with a standard deviation of 1.16.

What were common student errors or omissions?

Part (a):

- Many students provided only generic statements about how outliers or skewness affect the mean and do not affect the median, without making the connection to a conjecture about outliers or skewness in the distribution of incomes.
- Some students neglected to mention the effect of skewness or outliers on the mean, for example by saying only that one should not use the mean when outliers or skewness are evident.
- Many students exhibited inaccuracy with their use of statistical terms, for example by writing that outliers skew the mean or that means are biased when there is skewness.

Part (b):

- Some students did not explicitly choose one of the methods.
- Some students chose Method 1 based on the larger sample size, ignoring or dismissing issues of sampling bias.
- Many students gave good arguments for weaknesses of Method 1 but did not specify advantages of Method 2.
- Some students named the type of bias in Method 1 incorrectly, and other students only named the type of bias without describing the bias.
- Some students discussed the possibility of nonresponse with Method 2, even though the question stated that all responses would be obtained with Method 2.
- Many students did not describe the effect of the bias in Method 1 on the sample mean, even though the question specifically asked for this.
- Many students used the ambiguous statement that "results will be more accurate" when commenting on Method 2's effect on the estimate.
- Some students discussed conditions for inference in choosing between the methods, often incorrectly, for example by saying that a large sample size would produce a normally distributed sample.

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 to always relate their responses to the context of the question in a meaningful way. Students should also be reminded frequently that they must always justify their choices, especially when the question calls for a choice between options. Closely related to this is that students should be taught to give reasons in favor of one option as well as reasons against the other option.

Students should also be cautioned to be careful not to misuse statistical terminology such as "bias" and "skew" that are so often used casually in everyday conversations.

Teachers should emphasize the understanding of phenomena such as nonresponse bias, perhaps concentrating less on the names of such phenomena.

Teachers should present students with examples where a larger sample size is not necessarily better. For example, taking a larger sample size with a biased sampling method actually makes it less likely that the sample estimate will be close to the population parameter.

Question 5

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 1.10, out of a possible 4 points, with a standard deviation of 1.20.

What were common student errors or omissions?

- Some students did not realize that this question called for a hypothesis test, as they simply conducted a descriptive analysis.
- Many students did not recognize the paired nature of the data and mistakenly performed a two-sample t-test for independent samples.
- Many students defined the parameter incorrectly.
- Many students did not make clear that they understand why conditions for inference need to be checked.
- Some students used a z-test even though the population standard deviation is unknown.
- Some students wasted time by typing the data into their calculator, even though all relevant graphs and statistics were provided in the question.
- Many students who mistakenly obtained a p-value that was not small went on to provide a conclusion that amounted to accepting the null hypothesis.
- Some students provided an incorrect interpretation of the p-value.

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 help students to recognize that questions about whether data provide convincing evidence of something call for a hypothesis test and not solely a descriptive analysis.

Teachers should provide many, many opportunities for students to identify the appropriate inference technique to apply to a particular research question. Students can be given exercises that ask simply for identification of the correct procedure, without necessarily having to implement the analysis fully. Particular attention should be paid to the question of whether the data are paired.

Teachers should emphasize to students throughout the course that how the data were collected determines how they should be analyzed. This point is especially important when discussing paired designs and analysis of paired data.

Teachers should encourage students to follow the four steps described in the scoring guidelines when applying a significance test procedure.

Teachers should help students to understand the purpose of checking conditions for inference, as well as learning to check those conditions correctly. One approach for encouraging this understanding is to provide some questions where conditions for inference are not satisfied, and examine the consequences of that.

Teachers should teach students not to use z-procedures with quantitative data but rather to use t-procedures.

Teachers should occasionally provide unnecessary information (graphs, statistics) to give students practice with identifying what is relevant from what is not.

Recognizing that identifying parameters is a challenging task for many students, teachers should provide considerable experience and feedback with this aspect of hypothesis testing.

Teachers should emphasize to students that it is never appropriate to state a conclusion that amounts to accepting the null hypothesis. Also inappropriate are conclusions claiming that an alternative hypothesis has been proven. Instead, teachers should emphasize that the only two acceptable conclusions from a hypothesis test are to *reject the null hypothesis* or *fail to* reject the null hypothesis. Teachers should help students to understand the subtle distinctions among these conclusions.

Teachers should insist that students' conclusions from a hypothesis test be expressed in the context of the question and address the research question stated in the question.

Question 6

What was the intent of this question?

The primary goals of this question were to assess a student's ability to (1) calculate and interpret a residual value; (2) answer questions about residual plots; (3) compare associations between two scatterplots; and (4) identify an appropriate explanatory variable to include in a regression model based on residuals from simpler regression models.

How well did students perform on this question?

The mean score was 1.29, out of a possible 4 points, with a standard deviation of 0.99.

What were common student errors or omissions?

Part (a):

- Some students interpreted the residual in terms of distance from the predicted FCR but did not indicate *direction*.
- Some students did not include the *magnitude* of the residual in their interpretation.

Part (b):

- Most students incorrectly stated that the point represented a car with an actual FCR close to the predicted FCR from a regression model with *wheel base*, failing to realize that the residual was based on a regression model with *length*.

Part (c):

- Some students described the association in each scatterplot but neglected to *compare* them with clearly comparative language. An example of a correct comparison would be saying that the association between variables is stronger in graph II than in graph I.
- Some students compared the strengths and directions of the associations in the two scatterplots but did not comment on the form of association, for example by describing the form in graph II as linear.
- Most students incorrectly interpreted the residuals displayed on the vertical axis as if they were based on separate regression models, one with wheel base as the explanatory variable and one with engine size as the explanatory variable. These students failed to realize that the residuals in both graphs were calculated from the regression model using *length* as the explanatory variable.

Part (d):

- Misinterpreting the residuals, as described in the previous bullet point, invariably led students to incorrectly select wheel base as the additional variable to include in a regression model.
- Among student who correctly selected engine size, most did not provide a strong justification based on the principle of explaining additional variability in FCR left unexplained by the simple regression model using length.

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?

Students should learn that interpreting residuals requires commenting on both *magnitude* and *direction*. Teachers might also emphasize that residuals represent the *unexplained* variability in a statistical model, and a common goal is to reduce/explain as much of the variability as possible.

Teachers should frequently remind students that when asked to compare two distributions of data, or even two scatterplots as in this question, they should be sure to use comparative language rather than supply a "laundry list" of features.

Teachers should let students know that the investigate task often begins with questions for which fairly standard tools are appropriate but then often asks a question or two that requires students to extend their knowledge in a novel way. One good way to prepare students for such questions is to give them practice with investigative tasks from previous exams, helping students to recognize where the investigative part begins, and emphasizing that students are not supposed to know the new ideas in advance but are expected to apply what they have learned in novel ways.

Sound advice for all questions, but especially for investigate tasks, is to always read questions carefully and respond to exactly what is asked, not necessarily to what a student might have expected to be asked.