CollegeBoard AP

Student Performance Q&A:

2011 AP® Statistics Free-Response Questions

The following comments on the 2011 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 students' ability to (1) relate summary statistics to the shape of a distribution; (2) calculate and interpret a z-score; and (3) make and justify a decision that involves comparing variables recorded on different scales.

How well did students perform on this question?

The mean score was 1.33 out of a possible 4 points, with a standard deviation of 1.13.

What were common student errors or omissions?

Part (a):

- Students' explanations were generally quite weak.
- Many students realized that there was a problem with the minimum running time being 1.33 standard deviations below the mean, but most found it difficult to clearly articulate why this violated a characteristic of an approximately normal distribution.

Part (b):

- Some students appeared to believe that a z-score should be calculated as [(larger value – smaller value) / standard deviation], rather than as [(value – mean) / standard deviation].
- Many students calculated the z-score but neglected to provide an interpretation or merely restated the name of the calculation.
- Many students' interpretations mentioned only distance from the mean without specifying *direction* from the mean.
- Many students did not recognize the difference between a z-score and a z-statistic, which includes sample size in its calculation.

Part (c):

- Many students performed relevant calculations but did not fully explain the connection between their calculations and their selection of the better player.
- Some students performed normal probability calculations, despite having explained in part (a) why the normal model is not reasonable for approximating the distribution of running times.
- Some students compared the two players on only one variable rather than both variables.
- Some students did not perform any statistical adjustment to the scales before comparing the players.

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?

- Help students to recognize the distinction between a *model* (such as a normal distribution) and observed *data*.
- Provide students with considerable guidance, practice and feedback with justifying their answers statistically, including questions about whether observed data could reasonably be approximated by a particular model.
- Emphasize to students that z-scores (also known as standard scores) can be used as measures of relative location even when the variable is not well modeled by a normal distribution. The interpretation of z-score as number of standard deviations above/below the mean is also reasonable in this case. But caution students against using the z-score to calculate a normal probability if there is reason to doubt that the normal distribution provides a reasonable model for the data.
- Make clear to students the importance of interpretations to accompany calculations. Interpretations can be thought of as definitions applied to the particular context at hand.

Question 2

What was the intent of this question?

The primary goals of this question were to assess students' ability to (1) determine a conditional probability from a table of data; (2) use a table of data to determine whether or not two events are independent; and (3) demonstrate an understanding of the concept of independence by constructing a graph that displays independence between two variables.

How well did students perform on this question?

The mean score was 1.47 out of a possible 4 points, with a standard deviation of 1.21.

What were common student errors or omissions?

Part (a):

- Some students did not show how they calculated their answer.

Part (b):

- Some students revealed mistaken ideas about how to check for independence, for example, by checking whether P(M|Y) = P(Y|M) or whether P(M) = P(Y).
- Many students, even those who knew how to check for independence, exhibited poor communication in their answers, which took many forms:
    - Not explaining how the conclusion about independence followed from probability calculations
    - Not writing complete sentences in the justifications
    - Introducing symbols for events without specifically identifying what they meant
    - Using poor mathematical notation, which made it difficult or impossible to discern what the student intended
    - Writing interpretations of probabilities that were so muddled that it was difficult or impossible to tell what the student meant
- A few students confused independent events with mutually exclusive events.
- A few students calculated relevant probabilities and noted that they were not equal but mistakenly argued that these probabilities were close enough to conclude that the events were independent.
- A few students conducted a chi-square test of independence.

Part (c):

- Many students reproduced the identical graph that was provided for Franklin Township, where gender and party registration were not independent.
- Some students mistakenly assumed that independence required the three party categories to be equally likely.
- Some students drew a reasonable graph displaying independence but incorrectly calculated the marginal distribution of party registration.

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?

- Encourage students to be extremely clear when using notation to define events and when interpreting probabilities. Provide ample opportunities to practice these skills and sufficient feedback on students' performance.
- Emphasize to students the importance of articulating arguments clearly and explaining how conclusions follow from probability calculations. In particular, help students to become comfortable with writing sentences of explanation even when answering questions that concern probability calculations.
- Devote considerable time and attention to helping students understand *concepts* such as independence, in addition to being able to perform calculations of probabilities and conditional probabilities.

Question 3

What was the intent of this question?

The primary goals of this question were to assess students' ability to (1) understand and describe a process for implementing cluster sampling; and (2) describe a statistical advantage of stratified sampling over cluster sampling in a particular situation.

How well did students perform on this question?

The mean score was 1.41 out of a possible 4 points, with a standard deviation of 1.16.

What were common student errors or omissions?

Part (a):

- Many students did not communicate their response clearly.
- Many students did not understand what cluster sampling means.
- Some students did not include a plan for implementing the sampling method.
- Some students did not address the issue of what to do with repeated digits in their implementation plan.
- Some students mentioned the use of a random digit table but did not mention the need to ignore one digit that would not be assigned to any floor.
- Some students adequately described how to select two floors at random but then did not indicate clearly that all eight apartments on those two floors should be selected.

Part (b):

- Many students did not communicate their response clearly.
- Many students described advantages that were not statistical.
- Some students did not link carpet wear to whether an apartment had children.
- Some students remarked that stratification makes it more likely to obtain apartments with children and apartments without children in the sample, not making clear that stratification *guarantees* that the sample will contain both apartments with children and apartments without children.
- Some students described a reasonable disadvantage of the cluster sampling method without clearly stating how the stratified sampling method would eliminate the disadvantage.
- Some students used statistical words such as "confounding variable" and "bias" without clearly explaining how they related to an advantage of the stratified sampling method.
- Some students repeated information provided in the question, about including six apartments with children and two without children in the sample, without saying more or explaining how this constitutes an advantage of the stratified sampling method.

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?

- Be sure to introduce students to both cluster and stratified sampling methods.
- Emphasize how these two sampling methods differ, both in implementation and in purpose. One challenge associated with this is that both methods involve partitioning the population into subgroups — either clusters or strata — so this similarity can prevent some students from recognizing the differences between the two methods.
- One teaching strategy might be to present a population for which cluster sampling would be reasonable to address one question and for which stratified sampling would be reasonable to address a different question.
- Emphasize the need for randomness with any good sampling method, be it simple random sampling, stratified sampling or cluster sampling. Related to this is the need for students to be thorough and clear in describing how to implement a sampling method in a given situation.
- Above all, make students aware of the importance of communicating their responses clearly. Give students many opportunities to practice good communication, as well as frequent and detailed feedback on how to improve their written communication.

Question 4

What was the intent of this question?

The primary goal of this question was to assess students' ability to set up, perform and interpret the results of a hypothesis test. More specific goals were to assess students' ability to (1) state hypotheses; (2) identify the name of an appropriate statistical test and check appropriate assumptions/conditions; (3) compute the test statistic and p-value; and (4) draw a conclusion, with justification, in the context of the problem.

How well did students perform on this question?

The mean score was 1.53 out of a possible 4 points, with a standard deviation of 1.26.

What were common student errors or omissions?

- Many students did not include all four components of a significance test.
- Even though few students used symbols for sample statistics in their statements of hypotheses, some included written descriptions of hypotheses that pertained to the sample rather than the population.
- Although most students knew to check a randomness condition for inference, they commented on random *sampling*; few correctly referred to the random *assignment* that was used to produce the two treatment groups.
- Many students did not provide both graphs of the sample data and relevant commentary necessary to check the normality condition.
- Some students incorrectly applied a z-test rather than a t-test, mistakenly believing the standard deviations to be from populations rather than samples.
- Some students mistakenly applied a paired t-test.

- Some students provided the correct values for the test statistic and p-value but included additional information that was mistaken, for example, by writing a formula for the test statistic that was based on population parameters rather than sample statistics.
- Many students did not make explicit the connection between the p-value and their conclusion.
- Although few students explicitly wrote "accept the null hypothesis" for their conclusion, many provided a conclusion in context that was equivalent to accepting the null hypothesis.
- Some students did not express their conclusion in the context of this study.

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 throughout the course, when teaching statistical inference as well as when teaching data collection, the important distinction between random *sampling* and random *assignment*. These are not merely different uses of randomness but, more important, allow for different kinds of conclusions. Random sampling allows for generalizing sample results to a larger population, whereas random assignment provides the basis for drawing causal conclusions. Make students aware that this distinction is relevant to checking conditions for inference with a hypothesis test or confidence interval. One way to achieve this is to include many examples of randomized experiments when teaching inference procedures, as well as including many examples of random samples, and focusing students' attention on this distinction.
- Continue to emphasize the big ideas and essential concepts associated with significance tests. For example, frequently remind students that hypotheses are about population parameters rather than sample statistics. Identifying parameters clearly is also a big challenge for many students, so they should receive ample practice with that skill. Continue to make students aware of the importance of always *checking* conditions for inference, 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 data collection procedures.
- Give students considerable practice with and feedback on summarizing conclusions from significance tests. Encourage students to be very clear in stating how their conclusion follows from the p-value. Caution them against writing any conclusion that can be understood as equivalent to accepting the null hypothesis. A p-value that is not small should lead to a conclusion that the sample data do not provide enough evidence to reject the null hypothesis, rather than be interpreted as evidence in favor of the null hypothesis.
- Remind students about the need to express conclusions in the context of the research question presented.

Question 5

What was the intent of this question?

The primary goals of this question were to assess students' ability to (1) determine the equation of the least squares regression line from a computer output; (2) use the slope of the least squares line to compare expected values of the response variable for different values of the explanatory variable; (3) recognize how to determine the proportion of variability in the response variable explained by the least squares line; and (4) use computer output to determine whether the linear relationship between two quantitative variables is statistically significant.

How well did students perform on this question?

The mean score was 1.59 out of a possible 4 points, with a standard deviation of 1.25.

What were common student errors or omissions?

Part (a):

- Many students did not use a modifier such as "predicted" or "expected" or the ^ ("hat") symbol with the response variable of electricity production.
- Some students incorrectly used modifiers such as "predicted" or "expected" or the ^ ("hat") symbol with the *explanatory* variable of wind velocity.
- Some students interchanged the roles of the two variables.
- Some students did not know how to find the intercept and slope coefficients from the computer output provided.

Part (b):

- Some students neglected to include measurement units (amperes) with their answer or included mistaken measurement units (miles per hour).
- Most students plugged the two values into the least squares equation and calculated the difference in predicted values, rather than the more efficient approach of multiplying the value of the slope by 10.

Part (c):

- Some students did not know how to find the relevant information from the output provided.
- Some students reported the value of adjusted R^2 or the square root of R^2.

Part (d):

- Some students based their answer only on the value of the slope coefficient or the value of R^2, without realizing the need for inference based on the *p*-value.
- Some students did not know how to find the relevant *p*-value from the output provided.
- Many students did not make explicit the connection between the magnitude of the *p*-value and their 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?

- Provide practice with reading and interpreting computer output, ideally from a variety of software packages.
- Emphasize to students what $\hat{y}$ notation indicates, along with the reasons for using modifiers like "expected" or "predicted" on the response variable in expressing and interpreting a least squares equation.
- Help students to get into the habit of always expressing answers in terms of the measurement units involved (e.g., amperes).
- Make students aware of the importance of justifying their conclusions, in particular by commenting on the size of a p-value when addressing the issue of statistical significance.

Question 6

What was the intent of this question?

The primary goals of this question were to assess students' ability to (1) construct and interpret a confidence interval for a population proportion; (2) create a probability tree to represent a particular random process; (3) use a probability tree to calculate a probability; and (4) integrate provided information to create a confidence interval for an atypical parameter.

How well did students perform on this question?

The mean score was 1.31 out of a possible 4 points, with a standard deviation of 1.18.

What were common student errors or omissions?

Part (a):

- Many students did not explicitly state the name or formula for the procedure they used, and some provided an ambiguous name, such as "z-interval."
- Many students did not check conditions for inference.
- Some students incorrectly checked the sample size condition, for example, by merely comparing the sample size to 30, or by simply stating that the Central Limit Theorem holds, or by commenting ambiguously that the distribution is normal.
- Some students reported the interval correctly, presumably because they used a calculator, but then provided more calculations in which they made errors.
- Many students mistakenly referred to the *sample* proportion, rather than the *population* proportion, when they interpreted the interval.
- Some students mistakenly mentioned "mean proportion" in their interpretation.
- Some students did not refer to the context in their interpretation.
- Some students interpreted the confidence *level*, often incorrectly, rather than the confidence *interval*.
- Some students did not recognize that their interval encompassed values that are not reasonable for a proportion, such as negative values and values larger than one.

Parts (b) and (c):

- Some students mistakenly used set notation for a probability, such as "not k" or k^C rather than $1 - k$.
- Some students assumed a particular value for k, such as $k = 0.28$, and then used that value for subsequent calculations.
- Some students mistakenly multiplied, rather than added, the relevant probabilities in part (c).
- Some students used poor notation, such as a probability of a probability, as in $P(0.25 + 0.75k)$.

Part (d):

- Many students did not use their interval from part (a) as the question directed. For example, many set their answer to part (c) equal to the point estimate from part (a), then solved for (k), then used the resulting point estimate to produce a new confidence interval.
- Some students revealed in their interpretations that they did not understand the difference between the parameters p and k.
- Some students made the same interpretation errors as described under part (a).

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?

- Recognize that identifying parameters clearly is a challenging task for students.
- Provide much guidance and practice for determining which inference procedure to use, depending on the type of parameter (e.g., proportion or mean) involved, which in turn depends on the types of variable(s) (categorical or quantitative) in the study.
- Make sure that students know they must check conditions for inference with confidence interval procedures, as well as with significance test procedures.
- Help students to realize that the specific conditions to be checked depend on the type of parameter involved (e.g., proportion or mean, one group or two). Another good habit for students to develop is seeing whether a confidence interval has reasonable values for the context and parameter involved, for example, not giving negative values or values larger than one for estimating a proportion.
- Emphasize the importance of providing an interpretation with every confidence interval calculation.
- Help students to understand the distinction between interpreting a confidence *level* and interpreting a confidence *interval*.
- Encourage students to use good notation with probability calculations.
- Help students to understand the difference between probabilities, which are numbers, and events, which are sets. One implication of this difference is that arithmetic operations such as addition and multiplication apply to probabilities, whereas set operations such as union and complement apply to events.