



Student Performance Q&A: 2016 AP[®] Statistics Free-Response Questions

The following comments on the 2016 free-response questions for AP[®] Statistics were written by the Chief Reader, Jessica Utts of the University of California, Irvine. 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) describe the distribution of a quantitative variable based on a histogram and (2) determine the effect of changing one data value on the mean and the median.

How well did students perform on this question?

The mean score was 1.73 out of a possible 4 points, with a standard deviation of 1.00.

What were common student errors or omissions?

Part (a):

- Many students had difficulty describing variability from a histogram.
- Some students made no mention of the gap or outlier, which is a major feature of the histogram.
- Some students did not include context (tip amounts) or mistook units (dollars) for context.
- Some students provided an acronym they had been taught for describing the distribution (like SOCS or CUSS), but then they did not discuss all of the elements.
- Some students described the frequency of the values in each interval (the heights of the histogram bars), which were not relevant in this context.

Part (b):

- Many students misinterpreted the question, thinking a new value of \$18 was added to the data set.
- Many students did not use the values (\$8 and \$18) in their justifications.
- Some students noted that the mean would increase, but did not clearly justify why it would increase.

- In the justification for the mean increasing, some students noted that one value increased, but did not make the connection between increasing one value in the distribution and increasing the sum of all the values.
- Some students argued that the mean is not resistant to extreme values, but did not indicate that \$18 is an extreme value in this context.
- Many students did not clearly justify why median would not change.
- Some students thought the median would only change if n changed or if \$8 or \$18 was the median.
- Some students included inappropriate modifiers (such as the mean is *likely* to increase or the median *probably* will not change).
- Some students made up a data set that was unrelated to the histogram provided to investigate the effect on the mean and median, but never connected the results from their example back to Robin's tip distribution.

General:

- Some students used statistical vocabulary incorrectly (such as, resistant versus robust).
- Some students contradicted themselves within the response to one part of the question.
- Some students made generally unclear arguments.
- Some students used "it" as a vague reference in their descriptions and explanations when the meaning of "it" was not clear.

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 give more exposure to analyzing distributions of quantitative data based on graphical displays in which some or most of the data values are unknown.

Teachers should provide more opportunities to discuss distributions in context without simply listing features, and stress that context is about the variable(s) of interest and not the units of measurement for those variable(s).

Teachers should give students ample practice writing justifications and receiving critical feedback about them.

Teachers should be sure students can distinguish among the effects of the following changes to the data:

- Changing one value in a data set.
- Adding or removing a value from a data set.
- Transforming all the values in a data set via addition, subtraction, multiplication, division.

Teachers should build in spiral review throughout the course, revisiting topics and emphasizing connections, comparisons and contrasts. Some students seem to have forgotten the basics from the beginning of the course by the time they reach the end.

Question 2

What was the intent of this question?

The primary goals of this question were to assess a student's ability to (1) identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question and (2) assess the effectiveness of treatments in a controlled experiment.

How well did students perform on this question?

The mean score was 1.22 out of a possible 4 points, with a standard deviation of 1.21.

What were common student errors or omissions?

Section 1 (Inference steps 1 & 2: Statement of hypotheses, name of test, and technical conditions)

For the hypotheses:

- Some students reversed the null and alternative hypotheses, stating otherwise-correct hypotheses in context but confusing which one was which.
- Some students gave hypothesis statements in terms of some value of the chi-square statistic, such as $H_0 : \chi^2 = 0$ and $H_a : \chi^2 > 0$.
- Some students correctly stated the null hypothesis in terms of homogeneity of the three proportions (in words or symbols), such as $H_0 : p_A = p_B = p_C$, but then gave as the alternative an (incorrect) statement that all three proportions must differ, such as $H_a : p_A \neq p_B \neq p_C$.

For the technical conditions:

- Many students calculated the expected counts but failed to comment that all are greater than or equal to 5, or stated that all are greater than or equal to 5, but failed to show the values.
- Some students reported incorrect expected counts, such as “all expected counts are 12.5” (based on the incorrect idea that under the null hypothesis the choices of Choco-Zuties and Apple-Zuties are equally likely).
- Many students incorrectly stated that the randomness condition is satisfied because a simple random sample was chosen, for example, “SRS \checkmark ”, or just listing the name of the condition to be checked as “SRS?”.
- Some students gave a large sample as a condition that does not apply to a chi-square test, such as “normality is satisfied because 75 children were included in the study and $75 > 30$.”

Section 2 (Inference step 3: Mechanics)

- Some students did not report the degrees of freedom.
- Some students tried to calculate the chi-square statistic by hand instead of using the calculator, and then made errors in computation.

Section 3 (Inference step 4: Conclusion)

- Some students over-stated the conclusion to imply that the alternative hypothesis has been proven, such as, “since the p -value is less than $\alpha = 0.05$, we have proven that there is an association between ad type and children’s choice of snack.”
- Some students made an error comparing the p -value to α , such as stating “ p -value = 0.0058 $>$ $\alpha = 0.05$,” thus leading to the wrong decision and conclusion.

Section 4 (part (b): Ad effectiveness)

- Many students summarized information in the table of observed counts (or percentages) for each group separately without making any comparisons.
- Many students compared children’s preference between group A and group B (Choco-Zutie ad to Apple-Zutie ad) — which does not establish ad effectiveness. They did not compare group A to group C, and group B to group C, both of which are needed to establish ad effectiveness or ineffectiveness.
- Some students omitted numerical justification for claims of ad effectiveness.

- Some students made an incorrect argument about ad effectiveness based on the relative numbers choosing a particular snack type, such as stating incorrectly that ad A was effective because children tended to prefer Choco-Zuties.
- Some students provided an incorrect justification for ad effectiveness based on comparing individual chi-square components.

General

- Many students did not realize that the question called for a hypothesis test.
- Many students thought that part (b) of the question (concerning ad effectiveness), was asking for a conclusion to the hypothesis test in 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?

Teachers should caution students to avoid ditto marks (" ") in answering a question as is it not always clear what the student is trying to say.

Teachers should suggest that students review the entire question before beginning to write their response to any parts, and not to assume that each part necessarily requires information from prior parts.

Regarding conditions there are many issues teachers should cover:

- Address the specific conditions needed to justify each inference method and caution students to avoid a “laundry list” (shotgun) approach to listing and checking off conditions, such as:
 Normal \checkmark
 Independent \checkmark
 SRS \checkmark
- Emphasize that a statement regarding conditions, or a comparison that is based on numerical results (such as, expected counts > 5), must be justified with appropriate calculations shown or reference to the numbers given in the problem that support the statement.
- Emphasize the distinction between random assignment and random sampling.
- Avoid the use of abbreviations such as “SRS” as an acceptable way of describing randomness conditions generally, and require that students describe in words (complete sentences) the conditions they are checking and whether or not they are satisfied.

Teachers should inform students that degrees of freedom should always be reported when applicable (for example, a t -test, a chi-square test, etc.).

Teachers should discuss with students that it is good practice to specify what level of α they are using when a level is not given in the problem. Level specification not only conveys that the p -value must be small, but also gives a threshold for “smallness.”

Teachers should emphasize with students the need to clearly state hypotheses in terms of populations, and not use terms that imply that the hypotheses are about the data that were collected in the study or experiment.

Teachers should address with students how to use technology to carry out the mechanics of tests, and in particular what should be reported from the calculator to justify their response. For example, a chi-square test should include the value of the test statistic, the degrees of freedom, and the p -value. Students should avoid “calculator speak” in conveying important information (such as, the degrees of freedom).

Question 3

What was the intent of this question?

The primary goals of this question were to assess a student's ability to (1) identify explanatory and response variables from the description of a research study; (2) indicate and justify whether a study is observational or experimental; and (3) explain what confounding means in the context of a particular study with a specific confounding variable.

How well did students perform on this question?

The mean score was 1.55 out of a possible 4 points, with a standard deviation of 0.85.

What were common student errors or omissions?

Part (a):

- Some students failed to describe the possible levels of the variable; that is, not understanding that variables by definition take on more than one value (smoking and not smoking or Alzheimer's and not Alzheimer's, not just "smoking" or "Alzheimer's").
- Some students failed to describe the variable as a characteristic of the observational unit (that is, the incorrect use of the words "risk" or "chance" or "how many subjects" or "the people who" to describe developing Alzheimer's or not, when those words refer to a statistic taken from a sample).

Part (b):

- Some students provided a circular argument — "it is an observational study because it is not an experiment."
- Some students had vague or imprecise statements regarding why the study was observational, such as "no treatment" or "just manipulating the study."
- There was often confusion among statistical concepts like "treatment," "assignment," "levels," "factors," and/or confusion with different uses of the same word such as "control" vs. "control group."
- Some students believed that a control group is necessary in an experiment.
- Some students thought that assignment must be random or that random selection is necessary for an experiment.
- Some students thought that gathering data on subjects over an extended period of time implies an observational study.
- Some students mentioned ethical considerations only, thus confusing what should happen with what did happen.

Part (c):

- Some students defined a confounding variable as a variable that only affects the response variable without mentioning its relationship to the explanatory variable.
- Some students failed to explain how the confounding variable influences the response variable; for example how lack of exercise affects health without indicating how exercise might influence the development of Alzheimer's. A "how" question suggests that students should provide a plausible mechanism, not just an assertion of a possible influence.

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 emphasize two aspects of how to describe variables. First, make sure the description includes the possible values or categories that could be observed on the observational units, rather than just a name of the variable. Next, make sure that the description is a characteristic an observational unit (such as smoker or not), and not a statistic (such as number of smokers).

Teachers should make sure students are clear about the different vocabulary associated with experimental design vs. sampling (such as, blocking vs. stratifying, randomization or random assignment vs. random selection).

Teachers should develop an understanding that what is necessary for a study to be classified as an experiment is only assignment of treatments. “Good” experiments are randomized comparative experiments. Develop a clear understanding of what makes a study observational, that is, that levels of the explanatory variable (treatments) are not assigned to observational units. Some students described an observational study as a study that just observes an association or effect of smoking on Alzheimer’s. Such observations would be made whether a study is an experiment or observational.

Teachers should remind students to communicate clearly, always write responses in the context of the problem, and use statistical language carefully. For example, the words “association” and “correlation” have specific meanings in statistics.

Teachers should encourage students to avoid generalized statements or definitions. Write responses fully in the context of the problem and be specific. For example, it was not sufficient in this question only to say, “The study is observational because no treatments were imposed.”

Teachers should help students develop a clear understanding of confounding by requiring students to explain in detail how a confounding variable influences the response variable and is associated with the presumed explanatory variable.

In general, teachers should make sure students know to answer the question that is asked and always provide an explanation or justification.

Question 4

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 using basic probability rules or the geometric distribution; (2) recognize that a probability calculation for independent events does not depend on the previous outcomes of those events; and (3) assess whether a claim about the probability of a single event is reasonable based on a calculated probability of a series of those events.

How well did students perform on this question?

The mean score was 0.99 out of a possible 4 points, with a standard deviation of 1.17.

What were common student errors or omissions?

Part (a):

- Many students gave incomplete definitions of parameter values and values of the random variable when using calculator notation.
- Some students calculated the incorrect probability. Two specific cases were that they calculated $P(X = 30)$ using the geometric distribution, or they used the binomial distribution with $n = 32$ trials.
- Many students confused the use of the binomial distribution and the geometric distribution when specifying the required distribution in terms of success and failure.
- Some students gave correct numerical answers but without any justification or work shown.
- Although asked to calculate a probability, many students gave the expected number of successes or failures instead.

Part (b):

- Many students did not realize that the first 30 trials had no influence on the 31st and 32nd trial.
- Many students did not realize that if a conditional probability approach was used, the denominator must be computed as well. (They only computed the numerator.)
- Some students did not recognize that a single probability of the union of two events was required and reported two separate probabilities (for the 31st and 32nd trials).
- Some students failed to realize that the 31st and 32nd trials cannot both be the first failure.
- Some students gave answers without justification (“bald answers”).

Part (c):

- Some students did not give a direct yes or no answer.
- A number of students did not include context in the response.
- Some students did not give a clear reference to the probability in part (a) or did not relate part (c) to part (a).
- Some students linked the decision to the answer in part (b) instead of to the answer in part (a).
- Many students had overall poor communication.
- Some students thought a hypothesis test was required.
- Some students attempted to use the expected number of 4.5 failures for 30 trials and a 15 percent failure rate as justification, without giving a statistical argument for why 0 failures would therefore provide evidence that the failure rate is less than 15 percent.
- Some students based the decision on incorrect criteria, such as sample size or independence.
- Some students asserted that evidence indicating that the failure rate is likely to be less than 15 percent proves that the failure rate is less than 15 percent.

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 solving probability problems involving a specific distribution, teachers should focus on probability notation rather than calculator notation. When discussing the distribution, require students to specify values for the parameters, and to give clear definitions of those parameters.

Teachers should make sure students have a clear understanding of the characteristics of probability distributions and associated parameters.

Teachers should give students a lot of practice with problems that require the application of basic probability rules and make sure students have a clear understanding of independence and conditional probability. Also make sure students have a clear understanding of the probability of the union and intersection of events.

Teachers should remind students to always answer in context the question asked.

Teachers should give students a lot of practice with communicating the answer to a question in English. Make sure student responses are clear about the meaning of any pronouns used, such as “it” or “that.”

Teachers should give students practice with making predictions and decisions based on probability alone. Do some problems of this sort after inference so students learn that not everything needs a hypothesis test.

Teachers should remind students to give complete justification for answers. Give criteria, comparison, and conclusion to support a decision. Show your work and cross out what you do not want scored.

Question 5

What was the intent of this question?

The primary goals of this question were to assess a student’s ability to (1) construct and interpret a confidence interval for a population proportion; (2) explain why one of the conditions for inference is necessary; and (3) explain why a suggested procedure for constructing a confidence interval is incorrect.

How well did students perform on this question?

The mean score was 1.27 out of a possible 4 points, with a standard deviation of 1.05.

What were common student errors or omissions?

Section 1: part (a): Mechanics of the interval

- Many students did not provide a name for the procedure, or provided the incorrect name or formula for the procedure.
- Some students attempted to use a t -value as the critical value for a proportion procedure.
- Some students used a z critical value for a different percent confidence level.

Section 2: part (a): Interpretation of the interval

- Some students confused sample and population values, by stating that the estimation is for a proportion of the adults “sampled,” “asked,” or “who chose the statement.”
- Some students did not check to see whether the interval they calculated made sense in context; that is, the interval should have contained a proportion.
- Some students combined the interpretation of the confidence interval with an interpretation of confidence level, such as “In repeated samples, we are 95% confident that this interval will contain...”

Section 3: part (b)

- Many students struggled to specify which distribution is approximately normal (for example, sample distribution vs. sampling distribution).
- Some students incorrectly focused on the sample size requirement leading to a representative sample rather than the sampling distribution becoming approximately normal.

Section 3: part (c)

- Some students did not realize that a two-sample z -interval for a difference in proportions and a two-sample z -proportion interval using the calculator are the same thing.
- Some students did not fully communicate what situation would be needed for a two-sample interval, and why that is different from the situation described in this problem.

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 make sure students know how to find the appropriate multiplier for a confidence interval and critical value for a hypothesis test.

Teachers should provide a lot of practice with correctly identifying the population of interest.

Teachers should remind students to always ask if their numerical results make sense in the context of the situation. For instance an interval for a proportion should contain values between 0 and 1.

Teachers should make sure students understand the difference between interpreting a confidence interval and interpreting the level of confidence.

When covering conditions required for statistical inference, teachers should explain why each condition is necessary in addition to checking them. It might be helpful to have students write a sentence for why each condition is necessary. For example, ask students “Why is it important that a random sample is selected for this study?” or “Why is it important that the sample size be greater than 30 before a confidence interval is computed?”

When teaching about different procedures, teachers should use the full names of the procedures rather than what they are called in the calculator. For instance, use “One sample z interval for proportions” rather than “One Samp Z-Int” and “Two sample z interval for a difference in proportions” rather than “2 Prop Z Int.”

Teachers should stress good communication. Have students practice forming responses that include the answer to the question AND a full explanation of why that answer is correct. For example, don’t just state that a two-sample procedure is inappropriate when only one sample is available, but also state that what would make it an appropriate procedure if results were based on two independent samples.

Question 6

What was the intent of this question?

The primary goals of this question were to assess a student’s ability to (1) use a scatterplot to comment on a report about the relationship between two variables and interpret the slope for the least-squares regression line summarizing this relationship; (2) describe the relationship between two variables in a scatterplot when a categorical variable is introduced and compare a characteristic of the distribution of a variable for different categories of individuals in a scatterplot; and (3) describe how the associations between two variables for each category of individuals in a scatterplot differ from the overall association in the same scatterplot.

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 0.97.

What were common student errors or omissions?

Part (a):

- Most students recognized the positive association between number of semesters and starting salary shown in the scatterplot and correctly stated that the scatterplot supports the newspaper report. However, some students went on to discuss additional characteristics of the scatterplot (strength, form) that were unnecessary to answer this part. Because the association is not very strong, this led some students to incorrectly say that the report was not supported or only weakly supported.

Part (b):

- Some students were not able to identify the numerical value of the slope from the computer output.
- Instead of interpreting the slope, some students simply described the positive association between the two variables.
- Most students did not use non-deterministic language for the change in the value of the response variable. For example, many students said something like “For each additional semester, the starting salary will increase 1159.40 euros.” This implies that an increase of 1159.40 euros is guaranteed for each additional semester. A correct response implies that the actual values vary above or below the value provided by the model. For example, “For each additional semester, the *predicted* starting salary will increase 1159.40 euros” or “For each additional semester, the starting salary will increase 1159.40 euros, *on average*.”
- Some students attempted to use non-deterministic language, but were unsuccessful. For example, students said things like “about 1159.40 euros” or “approximately 1159.40 euros.” These were not considered correct because the words “about” and “approximately” are also used when describing a rounded value or describing that the value 1.1594 is an estimate provided by a sample. Likewise, the expression “according to the model” wasn’t considered correct because it was unclear if the student knew the model is non-deterministic.
- Many students didn’t use proper units for the starting salaries. The slope provided by the output was 1.1594, but the unit is thousands of euros. Many students said things like 1.1594 euros, 1159.40 dollars, or just 1.1594.

Part (c):

- Many students were able to identify one characteristic of the association (direction), but did not address the other two (strength, form).
- Some students did not describe the association in context by referring to the variable names (number of semesters, starting salary).

Part (d):

- Many students did not make any comparison or only made a partial comparison. For example, some students only listed the median salaries for each major but never ranked them. Other students said something like “Chemistry has a higher median starting salary than business or physics” but never compared business to physics.

Part (e):

- Many students were too general in their description of how to modify the newspaper report. Students said things like “the newspaper should account for major” or “the newspaper should include a scatterplot for each major” without providing any additional details.

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?

As general advice, teachers should make sure students answer the question that is asked — and only that question. In question 6, part (a), the question was about the direction of the association only, not the strength.

Teachers should make sure students are familiar with computer output, especially in a regression context.

Teachers should make sure students know that the slope of a least-squares regression line describes the change in the predicted value of the response variable for each 1-unit increase in the explanatory variable, and that they can interpret the slope in context. Emphasize using proper units when discussing the response.

When interpreting the results of regression, teachers should make sure students use unambiguous non-deterministic language. Also, make sure students understand that statistical models provide estimated values and that actual values are likely to differ from the value provided by the model.

Teachers should give students practice describing different types of scatterplots, and make sure to address direction, form, strength, and any unusual features (e.g., outliers, clusters).

In general, teachers should remind students to include the context of the situation for any description, comparison, interpretation, explanation, and justification.

Teachers should remind students that when asked to make a comparison, they should use words like “greater than” or “less than” and make sure to compare all groups. Also, when students are asked to modify, change, or improve something, they should include specific details about what needs modification and how it should be modified.

Teachers should explain to students that the investigative task typically has a flow, where students are asked to put different parts together in the end. When pulling things together, students should explicitly refer to previous parts and not expect a reader to make the connection for them.