



## Student Performance Q&A: 2015 AP<sup>®</sup> Statistics Free-Response Questions

The following comments on the 2015 free-response questions for AP<sup>®</sup> 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) compare features of two distributions of data displayed in boxplots and (2) identify statistical measures that are important in making decisions based on data sets.

#### ***How well did students perform on this question?***

The mean score was 2.24, out of a possible 4 points, with a standard deviation of 1.14.

#### ***What were common student errors or omissions?***

Part (a)

- Many students did not communicate clearly.
- Some students tried to determine the shape or type of distribution from the boxplots.
- Some students described the two boxplots but did not use comparative language.
- Some students omitted the context; students need to clearly identify the variable of interest in their responses.
- Some students did not understand which descriptive statistics can be determined (and compared) from a boxplot and which cannot. For instance, medians can be compared but means cannot. (Only five-number summaries and outliers can be determined from boxplots.)

Part (b)

- Many students did not communicate clearly or gave incomplete answers.
- Some students thought that giving a statistical measure was sufficient justification for choosing one company over the other.

- Some students used statistical language when trying to convey a non-statistical meaning.
- Some students did not understand the five-number summary, for instance thinking that a wider interval means more people were in it.
- Some students referred to the end of the whisker as the maximum when there were outliers.

**Based on your experience of student responses at the AP<sup>®</sup> 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 revisit material from earlier parts of the course on a regular basis to reinforce the ideas and methods.

Teachers should provide students with a lot of practice comparing distributions based on various graphical summaries, including the use of comparative language. For instance, the word “while” may sound like a comparison word, but many students used it without doing a comparison. An example is “The minimum for Corporation A is about \$36,000 while the minimum for Corporation B is about \$41,000.” That statement provides the two minimums but does not compare them. An appropriate statement would be “The minimum for Corporation A is about \$36,000 while the minimum for Corporation B is much higher at about \$41,000.”

Students should be reminded that if you are asked to make a choice, an explicit comparison of the two options needs to be made. It is not sufficient to discuss why one choice is good without explaining why the other choice is not as good.

Teachers should explain that proper justification for a response includes both statistical support and a contextual justification. For instance, saying that the minimum salary is higher for one corporation than the other is not a sufficient contextual justification for choosing one corporation over the other. The response needs to justify why that matters in the context of the problem.

Teachers should discuss what can be learned from boxplots and what cannot. Means and standard deviations cannot be determined from boxplots, and in many cases it is not possible to determine which of two boxplots represents data with a higher mean or standard deviation. Similarly, complete shape information cannot be determined from boxplots. The only definitive measures that can be learned from boxplots are the elements of a five-number summary and outliers that are extreme enough to be illustrated in the boxplot.

Teachers should provide a lot of practice with interpreting data from a variety of contexts, and with writing clear explanations that involve the context. In particular, the context should be incorporated along with statistical information when making conclusions.

## **Question 2**

### ***What was the intent of this question?***

The primary goals of this question were to assess a student’s ability to (1) use confidence intervals to test a question about a proportion and (2) understand the relationship between sample size and margin of error in a confidence interval for a proportion.

### ***How well did students perform on this question?***

The mean score was 1.23, out of a possible 4 points, with a standard deviation of 1.27.

### **What were common student errors or omissions?**

Part (a-i):

- Some students stated that because 0.20 is in the interval, there is evidence that the program IS working (which is analogous to accepting the null hypothesis that  $p = 0.2$ ).
- Some students did not understand how to compare 0.20 to the interval to come to a decision.
- Some students stated that there could be evidence that the program is not working as intended because 0.20 lies so close to the end of the confidence interval.
- Some students mistakenly checked whether 0 was in the interval and thought that there was evidence that the program is not working as intended because 0 is not in the interval.
- Some students attempted to set up a hypothesis test, presumably because they did not recognize how to use a confidence interval to make a decision.
- Some students calculated incorrect endpoints for the interval (for example,  $\pm 0.006$ ).

Part (a-ii):

- As in part (a-i), some students erroneously concluded that there is evidence that the program is working as intended because 0.20 is in the confidence interval, equivalent to accepting the null hypothesis that  $p = 0.2$ .
- Some students indicated that 0.20 is a plausible value because it is in the confidence interval, but did not recognize that there are other plausible values.
- Some students argued that a particular value within a confidence interval is more or less likely to be the correct estimate of a parameter based on its location within the interval.

Part (b):

- Some students obtained a new margin of error that was greater than 0.06 and did not recognize that this would be impossible with a larger sample size.
- Some students divided the margin of error by 4 (instead of the square root of 4) to get 0.015.
- A few students got the correct numerical answer coincidentally by stating a formula of  $z^* (\text{margin of error}) = 0.06$ , plugging in  $z^* = 2$  based on the empirical rule, then dividing 0.06 by  $z^* = 2$  to arrive answer of 0.03. (This was scored as incorrect.)
- Some students wrote the answer as 0.03 but did not justify how 0.03 was obtained.
- Some students did not provide a numerical answer even though they were asked to determine a value.
- Many students did not recognize that quadrupling the sample size divides the margin of error by two, so they relied on the formula for margin of error and then made calculation errors.

Part (c):

- Some students did not recognize how to use the new margin of error to arrive at a conclusion.
- Some students calculated an incorrect interval in part (b) that contained 0.20 and then came to the conclusion that the program is working as intended (analogous to accepting the null hypothesis that  $p = 0.2$ ).
- Some students had the correct conclusion but did not justify this conclusion by comparing 0.20 to the interval.
- Some students never drew a conclusion.

**Based on your experience of student responses at the AP<sup>®</sup> 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 develop an understanding that a confidence interval provides a range of plausible values for the population parameter. Students should not assign a probability to the likelihood of any one value being the parameter based on the location of that value within the interval.

Teachers should help students develop an understanding that for one sample proportions and means, as the sample size increases, the margin of error will decrease. Specifically, there is an inverse square root relationship between sample size and margin of error.

Teachers should emphasize that it is important to justify conclusions and calculations with a relevant explanation or formula.

Teachers should instruct students to make sure they read the question carefully and provide an answer to the question that is asked. For example, if the question states, “Determine the value...” then a value (a number) should be provided. In addition, once a question is answered there is no need to offer further ideas. Additional incorrect information can reduce a student’s score, even if a correct answer has been provided.

### **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 discrete random variable; (2) calculate the expected value of a discrete random variable; (3) perform a conditional probability calculation from a discrete random variable; and (4) use probabilistic thinking to make a prediction about how an expected value will change given a condition about the random variable.

***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 1.21.

***What were common student errors or omissions?***

Part (a):

- Some students were confused by the meaning of “at least one,” and so answered a question about  $X \leq 1$  or  $X > 1$ .

Part (b):

- Several students rounded the expected value to 2 or said that the expected value was approximately 2, suggesting that they held the misconception that the mean of a discrete random variable has to be a whole number.

Part (c):

- Many students made the mistake of assuming events were independent when they were not. They tried to calculate  $P(X = 3 \text{ and } X \geq 1)$  by multiplying  $P(X = 3)$  and  $P(X \geq 1)$ .

Part (d):

- When asked to “explain,” some students simply parroted the stem of the question and gave an answer without any real explanation.

General:

- Many students did not show their work when calculating probabilities or expected values.
- Some students used incorrect notation. For example,
  - $\bar{x}$  instead of  $E(X)$  or  $\mu_X$
  - $P(0.24)$  instead of  $P(3) = 0.24$
  - $P(3 | 1)$  instead of  $P(3 \geq 1)$
- Some students did not seem to know that appropriate formulas were provided on the formula sheet.
- Some students tried to express probabilistic thinking in words, for instance when explaining how the expected value would change if it is given that  $X \geq 1$ , and did not communicate their ideas well.

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

On computational questions involving probability and random variables, teachers should emphasize the importance of supporting work. Students should show the arithmetic they are performing, even if they use a calculator to do the arithmetic.

Teachers should make it clear to students that writing a generic formula is *not* sufficient for showing work, such as  $P(A | B) = \frac{P(A \cap B)}{P(B)}$  or  $E(X) = \sum x_i p_i$ . Students should show the numerical calculations.

Teachers should make it clear to students that calculator commands like 1-VarStats L1, L2 are *not* sufficient for showing work.

Teachers should help students better understand the meaning of expected value by presenting it as the long-run average if a chance process is repeated many, many times.

Teachers should give students practice with using the formula sheet (provided with the AP Exam) on assessments throughout the year.

Teachers should stress that the multiplication rule for independent events can only be used when two events are independent. When events are not independent, the rule needs to be modified to include conditional probabilities.

Teachers should provide opportunities for students to explain statistical concepts in words.

Teachers should emphasize that statistics has its own very precise language and that careful communication matters.

## Question 4

**What was the intent of this question?**

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

### **How well did students perform on this question?**

The mean score was 1.52, out of a possible 4 points, with a standard deviation of 1.27.

### **What were common student errors or omissions?**

- Many students had trouble defining the parameters appropriately. Some common errors were:
  - Using subscripts that do not clearly convey which group is associated with which parameter (for example,  $p_1$  and  $p_2$ ), and with no explanation of which is which.
  - Defining the parameter symbol as the group rather than as a population proportion associated with the group, such as,  $p_2 =$  placebo group.
  - Defining symbols that refer to (or imply reference to) the sample rather than to a population proportion, such as “ $p_1$  is the proportion of adults who *took* low-dose aspirin daily and then *developed* cancer.”
- Many students had trouble checking the appropriate conditions for the test. For instance:
  - Students incorrectly stated that the randomness condition was satisfied because a simple random sample was chosen, rather than because of random assignment.
  - Students incorrectly stated that the normality condition was satisfied because both groups were larger than 30.
- Some students had problems with the computing or stating the test statistics. Common errors included:
  - Not reporting the value of the test statistic, but reporting only the  $p$ -value.
  - Using the formula for the standard error of the difference in sample proportions as the  $z$ -statistic.
  - Calculating the  $z$ -statistic “by hand” from the formula, plugging in the numbers, simplifying the expression incorrectly, and getting the wrong value.
- Some students had trouble making an appropriate conclusion. Common mistakes included:
  - Not providing an explicit conclusion about the research question, but simply restating a rejection of the null hypothesis in context.
  - Omitting explicit justification for a decision or conclusion by failing to compare the  $p$ -value to the given  $\alpha = 0.05$ , such as, “at  $\alpha = 0.05$  we reject the null hypothesis.”
  - Failure to make any use of the significance level provided in the problem ( $\alpha = 0.05$ ), such as, “since the  $p$ -value is low ...”.
  - Overstating the conclusion to imply that the alternative hypothesis has been proven, such as, “since the  $p$ -value is less than  $\alpha = 0.05$ , we know that taking a low dose aspirin every day will reduce the chance of getting colon cancer.”

### **Based on your experience of student responses at the AP<sup>®</sup> 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 stress the importance of clearly defining parameters used in hypotheses. Some important factors are:

- Making sure subscripts are defined. It is not sufficient to use subscripts of 1 and 2 without describing what they mean.
- Making sure the parameters are explicitly defined to be about the population(s) and not the sample(s). Give students examples of definitions contrasting descriptions of sample quantities (not valid population parameters) to definitions that describe population quantities (parameters). For instance, “the proportion of adult volunteers who took aspirin and then developed colon cancer”

refers to a sample quantity, but “the proportion of all adults similar to the volunteers who would have developed colon cancer if they had taken a daily aspirin” refers to a population parameter.

Teachers should emphasize the distinction between random *samples* (generally used for surveys and some observational studies) and random *assignment* (generally used in experimental studies).

Teachers should 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 those conditions are satisfied.

Teachers should remind students that correct mechanics for hypothesis tests include the reporting of a test statistic, not just a  $p$ -value.

Teachers should let students know that while using technology is fine, enough information from the calculator must be reported to justify a response. For hypothesis tests, at a minimum this should include the value of the test statistic and the  $p$ -value.

Teachers should insist that students justify their conclusion by using the statistical information from the earlier steps of the process (hypothesis formulation, conditions, and mechanics). This is done by:

- providing a decision to reject or fail to reject the null hypothesis;
- justifying that decision by making an explicit comparison of the  $p$ -value to  $\alpha$ , the significance level (when it is provided); and
- stating a conclusion (a statement in terms of the alternative hypothesis) in the context of the problem.

Teachers should emphasize that a *decision* (reject or fail to reject the null hypothesis) is not enough. Students must also include a *conclusion*, which is an answer to the scientific question asked, in context.

## Question 5

### ***What was the intent of this question?***

The primary goals of this question were to assess a student’s ability to (1) use the information provided by a scatterplot to describe the relationship between two quantitative variables; (2) interpret and use the information given by lines displayed on a scatterplot; and (3) use a regression equation to estimate a predicted value of  $y$  for a given  $x$  value.

### ***How well did students perform on this question?***

The mean score was 2.07, out of a possible 4 points, with a standard deviation of 1.09.

### ***What were common student errors or omissions?***

Part (a):

- Many students failed to mention that the relationship was linear. Others reported the value of the correlation without supporting comment, which is not a sufficient indicator of strength or linearity.
- Some students failed to include context in their description.

Part (b-i):

- Some students identified that the  $y = x$  line is more helpful but without explaining how the  $y = x$  line divides the graph into three regions (above the line, on the line, and below the line) corresponding to the three body shape categories.

- Some students reversed the position of the tall rectangle and short rectangle categories relative to the  $y = x$  line.
- Some students chose Graph 2 for squares and Graph 1 for rectangles.
- Several students referred to the  $y = x$  line as a regression line.

Part (b-ii):

- Many students reported proportions, or relative frequencies, instead of frequencies.
- Some students reversed counts for short rectangle and tall rectangle categories.
- Some students did not use Graph 2 as an aid to count, even when selected in part (b-i).
- Some students used Graph 1 as a counting aid instead of Graph 2.

Part (c):

- Some students did not use the least squares regression line formula given with Graph 1 to predict arm span. Instead, they did one of the following:
  - Predicted arm span from the graph
  - Computed another formula
  - Selected a point on the graph and used that point
- Some students lost credit for failing to show the formula with 61 inserted for height, or for failing to report units of measurement (inches).
- Some students obtained a value that was not reasonable, and did not notice that it could not be correct, for instance by reporting a prediction that was way out of the range of arm spans shown in the graph.

***Based on your experience of student responses at the AP<sup>®</sup> 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 clear handwriting. In a few cases the answers could not be deciphered.

Teachers should give students many types of scatterplots to describe. Use bullets (direction, form, strength) and have students fill in a description of each.

Teachers should not accept answers without context, and always have students report units of measurements.

Teachers should have students calculate as an instruction in assignments and emphasize to students that the work must be shown.

Teachers should give students problems in which a formula they are expected to use is provided. Students need to read carefully to see if a formula is provided; students should realize that when a formula is provided, they do not need to enter the data into their calculator and find the formula.

## **Question 6**

***What was the intent of this question?***

The primary goals of this question were to assess a student's ability to (1) describe how sample data would differ using two different sampling methods; (2) describe the sampling distribution of the sample mean for two different sampling methods; and (3) choose the sampling method that will result in the best estimate of the population mean.



### **How well did students perform on this question?**

The mean score was 1.08, out of a possible 4 points, with a standard deviation of 0.86.

### **What were common student errors or omissions?**

Part (a):

- Most students correctly said that the sample would not be representative of all tortillas made that day and gave an adequate justification (for example, only one line was selected). However, many of these students did not give a *complete justification* for why selecting from only one line wouldn't be representative. It would have been better if students said "because the lines produce tortillas with different mean diameters, selecting from only one line won't produce a representative sample."

Part (b):

- Most students correctly said that the sample came from Method 1 and gave an adequate justification (such as, the histogram is bimodal). However, many of these students did not give a *complete justification* that also referred to the population. It would have been better if students said "because the histogram is bimodal, which is what I would expect when sampling from two production lines that have different means."
- In the stem of part (b), students were told to consider the *shape* of the histogram, but some students focused on center or variability instead or simply restated that Method 1 uses tortillas from both lines.

Part (c):

- Most students correctly said that Method 2 will result in less variability in the diameters on a given day and gave an adequate justification (for example, the sample comes from only one production line). However, many of these students did not give a *complete justification* for why selecting from only one line would result in less variability. It would have been better if students said "because the production lines have different means, using a sample from both production lines would likely result in more variable diameters" or "the tortilla diameters would have a range of about 0.6 inches when selecting from both lines but only about 0.4 inches when selecting from one line only."

Part (d):

- Many students did not describe *all three characteristics* of the sampling distribution (shape, center, variability).
- Many students were unable to identify the shape of the sampling distribution of the sample mean as *approximately normal*. Some of these students repeated the population shape (bimodal) but most of these students did not describe the shape at all. Among the students who were able to identify the shape, few were able to give a justification for why the shape is approximately normal.
- Many students did not remember to divide by  $\sqrt{200}$  when calculating the standard deviation of the sampling distribution of the sample mean. Some of these students repeated the population standard deviation (0.11 inch), and many did not include the standard deviation at all.
- Many students used *incorrect notation or sloppy language*, such as  $\bar{x} = 6$  instead of  $\mu_{\bar{x}} = 6$ . Also, some students stated that the shape is "normal" instead of "approximately normal."

Part (e):

- Although most students answered Method 1, few students were able to describe the sampling distribution of the 365 sample means for Method 2 as having roughly half the sample means around 5.9 inches and the other half of the sample means around 6.1 inches. In some cases, it was not clear that the student was describing *more than one* mean. In other cases, students did not imply that the sample means *will vary* from the population means by correctly using phrases such as "the sample means will cluster *around* 5.9 or *around* 6.1."

- Nearly all students, including the ones who earned credit for part (e), did not make the connection between part (d) and part (e). Because the standard deviation in part (d) measures the variability in the sampling distribution of the sample mean for Method 1, this value could have been used to help justify the answer in part (e). For example, “In Method 1, the sample means will typically be about 0.0078 inches from 6, but in Method 2, the sample means will typically be about 0.1 inches from 6.”

Part (f):

- Many students answered Method 1 and gave an adequate justification based on what could happen on a single day. However, some of these students did not imply that the sample mean will vary from the population mean by using phrases such as “the sample mean will be *around* 5.9 or *around* 6.1.”
- Many students did not see the connection between part (f) and part (e), but this was likely due to the poor performance on part (e).

**Based on your experience of student responses at the AP<sup>®</sup> 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 require students to provide complete explanations.

Teachers should remind students to read the question carefully and do what the question asks.

Teachers should give opportunities for students to consider using a numerical justification when appropriate.

Teachers should remind students that they must always address shape, center, and variability when describing any distribution, including sampling distributions.

Teachers should help students understand when the Central Limit Theorem applies; that is, when taking large random samples from the entire population. Remind students that the Central Limit Theorem says that the sampling distribution of the sample mean will only be *approximately* normal with large sample sizes, not exactly normal.

Teachers should remind students to divide by  $\sqrt{n}$  when calculating the standard deviation of the sampling distribution of the sample mean.

Teachers should give students practice thinking about sampling distributions in *unfamiliar contexts*. For example, to estimate the standard deviation of the sampling distribution of the sample median for samples of size 5 from a population of 1000 students, explain how to simulate samples of size 5, record the median for each sample, do this many times, and calculate the standard deviation of the simulated distribution of sample medians.

Teachers should be sure that students are always conscious of sampling variability and that estimates from samples are rarely exact estimates of corresponding population parameters.

The purpose of the investigative task (question 6) is to assess a student’s ability to *integrate statistical ideas* and apply them in a new context or in a non-routine way. Teachers can prepare students by showing them examples of previous investigative tasks or asking similar questions on assessments. Explain that the investigative part of the task generally occurs in the later parts of the question.

Teachers should remind students that the parts of a free-response question are often connected, especially for the investigative task. Teach them that they may need to use answers from one part to justify an answer in another part.