



Student Performance Q&A: 2005 AP[®] Statistics Free-Response Questions

The following comments on the 2005 free-response questions for AP[®] Statistics were written by the Chief Reader, Brad Hartlaub of Kenyon College in Gambier, Ohio. 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?

Three primary concepts were being assessed. First, the student needed not only to describe center, shape, and spread for a distribution, but to use these three characteristics as a basis for comparison. Second, because the question described only one urban school and one rural school, the student needed to realize that the results of the study could not be generalized to all rural and urban ninth-grade students in the United States. To generalize the results, a random sample of more than one urban school and one rural school would be needed. Third, the student was asked to identify which of two plans would better meet the goals of a similar study. In designing another study, it was important to think about day-to-day variability. For example, eating patterns on weekdays are often different from those on weekends. The plans presented were suggested as a means of assessing whether a student could explain the benefit of using a 7-day period so that both weekdays and weekends were included.

How well did students perform on this question?

The mean score was 1.41 out of a possible 4 points. As was the case in 2004, this exploratory data analysis question was not among the highest scoring questions in 2005. Students did not make comparative statements as expected in part (a), and they had trouble expressing the major concepts in parts (b) and (c).

What were common student errors or omissions?

Overall, students did not communicate their answers clearly.

Part (a)

- Often, students did not address the three common characteristics of a distribution (the center or position, the spread or variability, and the shape).
- When students did address some or all of the characteristics, they simply gave a listing or description of what these characteristics were, without making a comparison of the distributions (for example, the average of the rural distribution has a higher number of calories than the urban distribution).
- For each of the individual characteristics, students tended to have problems:
 - o They were apt to refer to the urban distribution as being skewed to the left or skewed to lower values, not noting that the stems went from smaller values at the top to larger values at the bottom.
 - o They often referred to the rural distribution as being normally distributed, or they sometimes said it was skewed right (because of the single value 51).
 - o Students tended to use the location of blocks of data—sometimes all of the data from each of the distributions—to describe general location of the distributions.
 - o Students often interpreted the term “distribution” as referring only to the overall shape, so they only commented on the difference in shape.
 - o Students were more likely to omit a reference to spread, which wasn’t very different in the two distributions.
 - o Students frequently took the rural distribution as the “norm” and only made statements about the shape of the urban distribution.
 - o Students communicated the measures of spread poorly, often referring to the range or the spread of the data as being the (min, max) pair for each distribution instead of reporting the range as the maximum minus the minimum.
 - o Students frequently only provided output from the calculators, with some general statements about the distributions.

Part (b)

- Some students failed to realize that the sampling unit was the school, and because there was only one school from each area, they could not generalize to all rural and urban ninth graders in the United States. Instead, students focused on the fact that there were only 20 students in each sample, rather than that there were only two schools in the study.
- Students often communicated part of the correct idea that the schools needed to be selected from a wider geographical mix, but they didn’t make explicit either what was wrong with the study or how it could be improved.

- Students thought that the major issue was that the schools were not randomly selected, even though the stem of the problem did not provide information on how the schools were selected.
- Those students who answered “yes” to the question thought that random samples of 20 were sufficient to make the comparisons based on normality, or they appealed to the central limit theorem.

Part (c)

- Students who responded with Plan II often thought that the main reason it was better was because there were more days in Plan II than in Plan I, thus appealing to the “more is better” concept. Students missed the connection to the day-to-day variation in eating habits (weekday eating habits are often different from those on weekends).
- Students often referred to the problem of lurking or confounding variables without addressing any reasonable specific concerns or variables that could cause concern in implementing the plan.
- Some students referred to the eating habits of an individual student instead of making more general statements about the eating habits of groups of students.

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 to read the question carefully and respond to the specific question asked. In responding to part (a), giving a good description of each distribution is not the same as *comparing* the distributions. For questions such as part (b), students must realize that a random sample from the population of interest is needed to generalize results to that population. Some students who may have understood this basic concept were unable to explain clearly that more than one school was needed, even though a random sample of students had been taken from each school in the study. Students often stated that “more days are better” in part (c), but they were unable to clearly articulate the benefit of the 7-day period. Generally, students failed to realize that the 7-day period constituted a week, and that both weekdays and weekends would be included.

Question 2

What was the intent of this question?

Students were asked to compute two different measures of center, the expected value (mean) and the median, for a discrete distribution. Students’ understanding of the properties of the sampling distribution of sample means and, in particular, the variability associated with those sampling distributions was also assessed. The sample mean of 1.25 telephone lines based on a random sample of 20 days was 0.35 telephone lines from the population mean. Because the sample mean from a random sample of 1,000 days is much less variable than one based on 20 days, the probability that the sample mean would be as far from the population mean for a sample of size 1,000 as it was for a sample of size 20 is extremely small. Finally, students were asked to discuss the relationship between the mean and the median relative to the shape of the given distribution.

How well did students perform on this question?

The mean score for this question was 1.25 out of a possible 4 points. Unsupported answers (or answers with no work shown) and poor communication made this question the second lowest scoring question in the exam.

What were common student errors or omissions?

Part (a)

- Many students used incorrect methods for calculating $E(X)$, such as computing the mean of the x values, computing the mean of the $p(x)$ values, or computing the appropriate sum but then dividing by 6. Students who provided the correct answer often did not show their work or used incorrect notation (e.g., x or \bar{x}).

Part (b)

- Many students did not make a connection to part (a) or made a weak one. They simply referred to the “true mean” but not the value. The sample average was often not specifically named but was referred to as “it” or the “value” or the “data.”
- Many students wrote a considerable amount without ever making a comparison, as the question requested. Students must be encouraged to answer the question.
- Some students provided a comparative statement with no justification. Others provided incorrect interpretations of the Central Limit Theorem or the Law of Large Numbers.

Part (c)

- Many students provided an unsupported answer. Others calculated the median of the x values or the median of the $p(x)$ values. Many students ignored the definition given and provided an approximation, interpolation, or interval such as $x < 1$.

Part (d)

- Many students ignored the given distribution. Rather than noticing that the given distribution was skewed right and so the mean should be, and was, greater than the median, students provided generic statements that would apply to any situation (e.g., if the distribution was skewed right, then the mean would be greater than the median). They also confused skewed to the right with skewed to the left. A surprising number of students incorrectly assumed that all distributions are normal.

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?

For parts (a) and (c), which required computation, students often provided the answers only. Teachers must remind students that answers with no support receive no credit: work must be shown. Some students

seemed to have an understanding of what was being asked in part (b) but were unable to communicate this concept effectively. They should be reminded to clearly distinguish between a sample and the sampling distribution. The variance of the sample mean is not the same as the variance of the population, or simply variance, and students need to use the appropriate term. In part (d) students often stated incorrectly that the mean being greater than the median meant that the distribution had to be skewed right. If a population distribution is skewed to the right, the mean is greater than the median. However, it is not necessarily true that the population distribution is skewed right if the mean is greater than the median. Because the population distribution was provided, the student could determine that the population distribution was skewed right, and that led to the mean being greater than the median. Teachers should encourage students to look at the distribution to determine shape and make appropriate connections with measures of center.

There are several topics in the curriculum that many students do not seem to understand and perhaps deserve more attention or examples to help clear up the confusion.

- The difference between a random variable and sample data: While a probability model can be used to generate “sample data” by observing the values of the random variable, the distinction between a random variable and data seems fuzzy, and this causes difficulty for students in exam situations. Emphasizing the difference between a probability model and data, perhaps with dice or cards, is an important activity to complete before students become involved in the details of each separately.
- The difference between discrete and continuous random variables: In particular, many students will take a discrete random variable and treat it as if it were continuous. For example, in part (c) students would get a median of 1.6 or 1.75 because they tried to estimate or interpolate the value that would be the 50th percentile as if there were a continuum of values between 1 and 2.
- Shapes of distributions and descriptive statistics:
 - o Students should know that the shape of a distribution affects the relation of descriptive statistics, not that the statistics affect the shape. So, a distribution skewed to the right should have a mean greater than a median, but the mean being greater than the median does not automatically imply that a distribution is skewed to the right.
 - o Students should recognize that distributions can take on a variety of shapes. While the most common shapes are those that are skewed to the left, skewed to the right, or symmetric, others may be bi-modal, tri-modal, or have very odd shapes, which are difficult to describe.
 - o Students seem to think that skewness in one direction is the same as having outliers. It is possible to have a distribution that is skewed to the left with outliers in the right tail.
 - o Far too often, students give Readers the impression that they think that all symmetric distributions are normal.
- Distinguishing between a sample and a sampling distribution

Question 3

What was the intent of this question?

Students were presented with a data table, scatterplot, residual plot, and computer output from a linear regression analysis. Part (a) asked students to evaluate the appropriateness of a linear model. They were expected to use graphical evidence to make this determination. In part (b) students had to recognize that the estimated slope from the computer output was needed; then they had to use the estimated slope to compute a point estimate of the change in the average cost of fuel per mile for each additional railcar.

Part (c) required students to identify the value of r^2 from the computer output and to interpret this value in context. In part (d) students were asked whether it was reasonable to use the linear model to make a prediction for a value of the explanatory variable that is far beyond the range of this data.

How well did students perform on this question?

The mean score was 1.46 out of a possible 4 points. This was a fairly straightforward question that tested four distinct ideas about linear regression. Students who knew these concepts earned high scores.

What were common student errors or omissions?

Part (a)

- Some students argued that a value of r^2 close to 1 or a value of r close to ± 1 indicates that a linear model is appropriate. This is not correct. There are numerous relationships between quantitative variables in which r^2 is close to 1 or r is close to ± 1 , but where the scatterplot shows a nonlinear relationship and the residual plot shows a clear pattern.
- Some students interpreted the question as asking whether the linear model provided the “best fit” for the data, and they proceeded to calculate alternative models—exponential, quadratic, power, and so on.
- Some students generated nonstandard (and awkward) descriptions of the residual plot: “amorphous blob,” “shapeless cloud,” “chaos,” and even “white noise.”
- Some students cast a wide net, hoping to catch *something*—they commented on both plots as well as r , r^2 , or both, and the p -value for the linear regression t -test on the slope. Unfortunately, more is not necessarily better here, since there is a greater potential for making a mistake.

Part (b)

- Many students recognized that this question involved slope and even provided a nice interpretation of slope in context.
- Several students recalculated the slope either by plugging two hypothetical values for the number of railcars into the linear regression equation, or by entering the data set into their calculators and calculating the regression equation for themselves. This took unnecessary time.
- Some students did not seem to be familiar with computer output.

Part (c)

- A number of students referred to R-sq(adj) rather than R-sq on the computer output.
- Many students could not interpret r^2 correctly.
- There were quite a few responses that described r^2 as the “correlation” or that interpreted r^2 as if it were the correlation coefficient (r).

Part (d)

- Some students thought of adding a new point to the data set with $x = 65$ railcars and then discussed how this would affect the linear model (e.g., it would be an influential point).
- Some students would not commit to a yes or no answer, opting instead for “yes, no, maybe so.” For example, “Yes, since r^2 is so high and the residual plot shows random scatter—but no, you might be extrapolating—then maybe so in this case, since 65 isn’t too far away from the rest of the data.”

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 needed to realize that a high r or r^2 value was not sufficient to judge the fit of a linear model in part (a). They should be able to effectively communicate pattern, or the lack of pattern, in plots, using correct statistical terminology. Students should be familiar enough with computer output to use it to answer questions. On past exams, students have sometimes been asked to interpret the estimated slope in the context of the problem. It is also important for students to recognize when the estimated slope is needed to make an estimate of interest, as was requested in part (b). Whether or not it is appropriate to use a regression equation to make a prediction is an important consideration in every regression problem.

Teachers may find the following suggestions helpful for questions of this type.

- Include ample instruction on how to interpret sample computer output in your course.
- Always have students answer questions in context.
- Stress and demand proper use of statistical vocabulary. Use terminology from the *AP Statistics Course Description* and textbooks. Have students practice applying vocabulary in context often.
- Clearly distinguish between the interpretation of r and r^2 .

Students may find the following suggestions helpful for questions of this type.

- When you see data, don’t just start immediately keying it into your calculator.
- Be sure to answer the question that was asked.

- Don't ramble. Say what you have to say and then stop writing. Everything you write will be graded for accuracy.
- Avoid use of terms like “proved,” “caused,” or “normal,” unless you are using them in their statistical sense.
- Remember that a high value of r and r^2 does not necessarily imply a linear relationship.

Question 4

What was the intent of this question?

The intent of this question was to evaluate a student's ability to set up the null and alternative hypotheses appropriate for a given setting, to conduct the appropriate hypothesis test after checking conditions, and to draw conclusions from that test.

How well did students perform on this question?

The mean score was 1.89 out of a possible 4 points. In contrast to previous years, when questions dealing with statistical inference tended to score low, this question was the highest scoring question in 2005.

What were common student errors or omissions?

- Students often used nonstandard notation and then failed to define it.
- Students often forgot to check the appropriate conditions. When they did, they frequently used \hat{p} instead of p . Some students substituted the p -value from their test for the p in the appropriate check of conditions.
- There was a good deal of inconsistent work. Students would substitute values into the appropriate formulas incorrectly and then use a calculator to obtain the correct answer.
- The biggest problem was an inability to clearly communicate the conclusion of the hypothesis test. Many students wrote “accept the null hypothesis” or “the null hypothesis is correct.” Other students used \hat{p} as the p -value for the hypothesis test.

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?

It is always good to state the meaning of any parameter used in the null or alternative hypotheses in the context of the problem, but it is essential if nontraditional notation is used. Remind students to actually check the appropriate conditions for a test and not simply list them with a check mark by the side of each one. Being able to correctly state conclusions in the context of a problem when a hypothesis is not rejected is as important as being able to state the conclusions when the hypothesis is rejected.

Question 5

What was the intent of this question?

This question focused on some basic concepts relevant to sample surveys. The focus of part (a) was assessing students' ability to identify a potential source of bias and describe the possible effect on the estimate of interest. Information from a pilot survey, as provided in part (b), could be used to obtain a sample size requirement for a full study. In part (c) students were asked to recognize that a stratified random sample is a valid approach to obtaining estimates at both the state and national levels; they had to clearly describe the structure for taking such a sample.

How well did students perform on this question?

The mean score was 1.11 out of a possible 4 points. This question was the lowest scoring question in 2005 and also ranks as the lowest scoring question in the last three years. Many students confused different types of bias or did not describe how the potential source of bias would affect the estimate of interest. They also did not justify their answer in part (b) and failed to identify the sampling method in part (c).

What were common student errors or omissions?

Part (a)

- The type of bias was often misnamed, sometimes using nonstandard terminology. A common example is the improper use of “voluntary response bias” to describe whether or not a person chooses or “volunteers” to answer the question, which is nonresponse bias.
- Some students did not link the bias described to whether or not the adult head of household had a high school diploma. As an illustration, a student might say, “Not all households had telephones.” Although this is a true statement, it will not result in bias unless one group (those with or those without a high school diploma) is more likely to not have a telephone than the other group.
- The effect of the bias on the survey was often not discussed. For example, a student might clearly indicate that those without a high school diploma were more likely not to have a telephone but fail to indicate that this may lead to the proportion of heads of households without a high school diploma being underestimated.
- Some students realized that the bias affected the estimated proportion of heads of households without a high school diploma, but they suggested that the effect would be the opposite of what would be expected. Using the example above, these students indicated that the proportion of heads of households without a high school diploma would be overestimated instead of correctly specifying underestimated.
- Some students did not clearly indicate they were addressing the estimate of the proportion. As an example, a common response was, “This would make the proportion of heads of households without a diploma lower.”

Part (b)

- Some students had difficulty expressing the margin of error appropriate for this survey. As an example, students often tried to use the formula for the margin of error for a mean, rather than a proportion.
- When determining the required sample size, some students used the variance of a Bernoulli trial, $p(1 - p)$, instead of the standard deviation.
- A few students used 1.645 instead of 1.96 as the appropriate z -value.
- Although some students knew that there is a relationship between the margin of error and a confidence interval, they did not know that the margin of error is half the length of the confidence interval. This led some to inappropriately halve the value for the margin of error, and others to mistakenly double it.
- Calculation and algebra errors were common. These were considered minor errors, but a decimal value was needed so that it could be determined whether or not rounding was done properly. Students often failed to provide this necessary justification.
- Some students provided the correct decimal value but did not round to give the required sample size.
- Students often failed to justify their work by providing the formula they used to find the required sample size n , the decimal value found for n , and the appropriately rounded value for n .

Part (c)

- Students often correctly described the process of taking a stratified random sample but failed to identify it as being a stratified random sample.
- Some students realized that a stratified random sample was needed, but then they stratified by regions, area codes, counties, or some other geographical entity.
- Separate state and national surveys were commonly described, clearly indicating some students do not understand that a stratified random sample can provide both state and national estimates.
- Some students did not describe any type of random process for collecting samples within each state.
- A few students noted that a simple random sample would be taken within each state, but then they proceeded to describe a multistage sampling method or some other sampling method.
- When a multistage sampling method was suggested within each state, some students did not realize that more than one selection was needed at each stage to obtain valid estimates of standard errors for the estimates of the proportions.

- Some students discussed how to take a sample that would not have the bias of part (a) instead of a method that would lead to both state and national estimates.

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 they must answer each part of the question completely. Responses frequently addressed only a portion of part (a). Students should also be taught to use statistical terms properly. As an example, if nonresponse bias is listed as a potential bias, then the description needs to be that of nonresponse and not some other kind of bias. In part (b) the failure to clearly show all work often led to students receiving a lower score than would have otherwise been the case. Answers alone are clearly insufficient. Students need to know the basic purpose of the various sampling methods. They often intuitively described a stratified random sample in part (c) but did not know, or express that they knew, that it was a stratified random sample.

Question 6

What was the intent of this question?

The intent of this question, known as the investigative task, was to evaluate students' understanding of concepts they should certainly know by asking them to use correct statistical thinking to answer questions that go beyond what they have encountered in the classroom. The two-sample confidence interval in part (a) is one of the important methods that students should be able to use. Students needed to identify an appropriate two-sample procedure, check the conditions for the validity of the procedure, carry out the calculations, and interpret the result interval. In part (b) students were given instructions for constructing a plot based on the four means provided. They were not expected to be familiar with this plot but to follow the instructions and properly label and scale it. In part (c) students were required to compare settings and environments, as well as the relationship between the two, using statistical thinking.

How well did students perform on this question?

The mean score was 1.62 out of a possible 4 points. This question was the second highest scoring question in 2005 and ranks as the highest scoring investigative task in the last three years. Students struggled with the two-sample confidence interval in part (a), but they did very well with the plot in part (b) and seemed to understand the major ideas in part (c), although their explanations could have been better.

What were common student errors or omissions?

Part (a)

- Many students seemed to believe that the given structure of the data presentation suggested a paired analysis and did not realize that they had two independent samples.
- Many students either did not address conditions or failed to fully “test” the normality condition (either assuming the sample size was too small or not producing and commenting on appropriate plots of the sample data). Students often listed conditions without seeming to fully understand their application in this particular problem.

- Students were required to identify the procedure by name or by formula. When both name and formula were included, they often were not consistent, or the calculations shown were not consistent with the endpoint values given—students accompanied the calculator answer with some work but then made mistakes in their hand calculations. Sometimes the interval endpoints appeared without any support.
- Students did do better on interpreting the calculation, but this area could still be improved.

Part (b)

- Students did not always include both a vertical axis label and an indication of the “inside/outside” lines.
- Students sometimes struggled with the axis scale, often constructing the first line (inside) and then not having room on the page for the second line with higher values (outside). Sometimes they adjusted the scale, and sometimes they used an improper scale.

Part (c)

- Many students failed to fully address the relationship aspect; they appeared to think that discussion of the maximum point was sufficient.
- Students often did not fully justify their comparisons using the earlier numerical data. It was very unusual for students to discuss the implication of the nonoverlapping confidence intervals. Credit was often given for any justification, rather than requiring students to justify each statement individually. Some students relied on discussion of environmental conditions instead.
- The question asked students to address four main issues but they frequently did not address all four; their responses were often difficult to follow because they did not address each issue in turn.

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?

The most common error in part (a) was forgetting to check the conditions needed for a two-sample t procedure. Although some responses listed the conditions, these conditions often were not checked. Students confused the paired t and the two sample t confidence intervals. Although the two-sample t confidence interval was clearly appropriate here, many students used the paired t confidence interval. Interpreting the confidence interval in the context of the problem is important, and something that many students still find challenging. For the two sample t confidence intervals, the parameter of interest is the “difference in the means” and not the “mean of the differences.” Remind students that once they think they have an answer to the investigative portion of the problem, they should review the response to be sure it has a statistical foundation. Supporting the conclusions in part (c) with confidence intervals to the extent possible and otherwise with means was an important part of the statistical thinking for this investigative task. The best responses were well organized into three parts, one for each of setting, environment, and the relationship between the two, with justification given in each part.

General Comments on Exam Performance

Overall performance on the multiple-choice questions was down slightly from 2004 but is very similar to overall performance on the multiple-choice sections in 2003 and 2002. Scores on the free-response section are up slightly from those in 2004 and are comparable to the overall averages on the free-response sections in 2003 and 2002. The overall average was very close to the average in 2004. The best news from 2005 is that students seemed to understand the major ideas included in the investigative task. The worst news from 2005 is that students did not show their work and did not answer the questions that they were asked on the exam.

General Recommendations for Teachers

- Emphasize to students that they must answer the question as stated in context. Remind them to avoid stock phrases that are vague, ambiguous, or incomplete.
- Explain to students that when they are told to perform a particular task in the stem of a question, they need to do exactly that as clearly as possible. For example, when told to “compare” two items, the students should make statements that clearly make a comparison rather than making independent statements about each.
- Tell students to check to see if their answers make sense. For example, if they think that a distribution is skewed to the right, but they end up computing a mean that is less than the median, they should know that they have made an error somewhere. If they cannot find the error, they should at least comment on the inconsistency to show that they are not blindly accepting unreasonable results.
- Give students more practice with a wide variety of settings, so that they become comfortable making the appropriate applications and explaining the results. Communication, especially in the context of the question, continues to be a major problem for many students.