Student Performance Q&A:

2003 AP® Statistics Free-Response Questions

The following comments on the 2003 free-response questions for AP® Statistics were written by the Chief Reader, Roxy Peck of California Polytechnic State University in San Luis Obispo, California. 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 intent of this question was to assess students' ability to construct boxplots and to reason about a distribution using information in a boxplot. Good communication was particularly important on this question. In Part (c) students needed to recognize that testing a hypothesis about means would not answer the question posed because of potential variability in individual times.

How well did students perform on this question?

The mean score for this question was 1.97 out of a possible four points. There were many good responses to this question, and it was the highest scoring question on this year's exam.

What were common student errors or omissions?

Common errors when answering this question included the following:

- Many students did not label the two graphs or did not include a scale, making comparisons difficult.
- Some students did not use a common scale for the two boxplots, again making comparisons difficult.
- Some students created graphical displays other than boxplots, such as stem-and-leaf displays or histograms.
- Some students applied outlier rules of their own invention in order to be able to show outliers in the boxplots.
- In Part (b) many students did not answer the question asked and did not indicate a choice between teachers and students.
- In Part (b) many students incorrectly interpreted the boxplots (often because they were not labeled correctly or because a common scale was not used).

- In Part (b) some students did not base their response on the boxplots, choosing to comment on the behavior of students and teachers based on their own personal experience.
- In Part (c) many students focused on checking assumptions to see if a particular test could be used to test the given hypotheses, rather than focusing on addressing the question that had been asked, which dealt with the appropriateness of the hypotheses.

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 need to be more careful in constructing graphs and should always include labels and scales. When a comparison is to be made, a common scale should be used. Also, encourage students to read the question carefully and then answer the question that has been asked

Question 2

What was the intent of this question?

The intent of this question was to assess students' understanding of the concept of errors in hypothesis testing. A complete response defined the parameter of interest and established hypotheses that were appropriate to the question of interest to the law firm. It also included a definition of type I and type II errors in context and a discussion of the monetary consequences of each type of error from the law firm's perspective.

How well did students perform on this question?

The mean score for this question was 1.62 out of a possible four points. Students who knew the definitions of type I and type II errors and could describe them in context tended to do well, but a very large number of responses earned a score of 0 on this question. This was surprising because it was one of the more straightforward questions on the exam and suggests that the topic of errors in testing is something that may not be receiving adequate coverage in some classes.

What were common student errors or omissions?

Common errors when answering this question included the following:

- Many students did not know the term "parameter," often confusing the "parameter of interest" with the "variable of interest."
- Some students defined the parameter of interest in terms of the sample, clearly confusing the distinction between sample and population. Some students wrote hypotheses in terms of sample values rather than population characteristics.
- Students often confused "consequences" with the decision as to whether to take the case or not, or they thought of consequences in terms of winning or losing the case.
- Many students reversed type I and type II errors. Some who reversed the errors then even went on to reverse a second time and described consequences that were not consistent with the errors as they had defined them! To receive credit for the discussion of consequences, the consequences needed to be consistent with the errors as defined by the students.
- Many students did not describe type I and type II errors in context, instead giving a textbook definition (i.e., reject the null hypothesis when it is true) with no context.

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?

Written communication is important on the AP Statistics Exam. It is important that students understand that conclusions and interpretations need to be in the context of the problem and not in generic terms. It is a good idea to have students practice this by always requiring context in student writing.

Question 3

What was the intent of this question?

The intent of this question was to evaluate students' ability to compute probabilities based on the normal distribution and the binomial distribution. To receive full credit for this question, students were required to correctly compute the requested probabilities, showing the work necessary to support the answer.

How well did students perform on this question?

The mean score for this question was 1.84 out of a possible four points. Students generally did well on Parts (a) and (b), but many were unable to recognize that Part (c) required the use of the binomial probability distribution.

What were common student errors or omissions?

Common errors when answering this question included the following:

- Many students failed to show supporting work, even though the question specifically stated that work should be shown.
- Many students tried to use the Empirical Rule to answer Parts (a) and (b), even though the relevant boundaries were not 1, 2, or 3 standard deviations from the mean.
- In Part (a) some students failed to recognize that the problem required looking at two tail regions and instead focused on only one tail.
- Many students did not recognize that the probability computed in Part (b) was the probability of success for the binomial calculation in Part (c).

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

Students should be encouraged to show supporting work, especially when the problem asks them to do so! Also, make sure students understand that listing calculator commands is not the same thing as showing work.

Question 4

What was the intent of this question?

The intent of this question was to assess students' understanding of some basic principles of experimental design, including the importance of random assignment, the benefit of including a control group, and how using volunteers limits the scope of the inferences that can be drawn.

How well did students perform on this question?

The mean score for this question was 1.69 out of a possible four points. Students did not perform as well on this question as they have on design questions in the past. Poor communication of ideas contributed to the lower scores.

What were common student errors or omissions?

Common errors when answering this question included the following:

- Many students did not understand the difference between random selection of subjects from a population and random assignment of subjects to treatment groups.
- Many students failed to recognize that the response was the *change* in stress levels.
- Many students incorrectly used statistical terms like bias, confounding, blocking, and so on.
- Some students indicated they believed that randomization eliminates bias and confounding variables.
- In Part (c) some students indicated the problem was that the sample size was too small, thinking that a larger sample would allow generalization even though the subjects were volunteers.

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?

Many students did not appear to be able to use statistical terminology correctly. Failure to understand the terminology of statistics is a big handicap in a course that has a heavy emphasis on communication. Students should be encouraged to be careful in the use of statistical terminology.

Question 5

What was the intent of this question?

The intent of this question was to evaluate whether students could carry out a test of hypotheses and state conclusions in context. To receive full credit for this question, students needed to carry out the test, including identifying appropriate hypotheses, identifying the chi-squared test for independence as the appropriate test procedure, and checking (not just stating) any necessary conditions for the test. Then, based on the result of the test, they needed to give an appropriate conclusion in context.

How well did students perform on this question?

The mean score for this question was 1.35 out of a possible four points. Student performance on hypothesis testing quesitons has been improving over recent years, but scores still remain low. Surprisingly, this was the lowest scoring question on this year's exam. More students (but still not the majority) are addressing necessary conditions and doing a better job of stating a correct conclusion in context, but many still overlook key parts of the hypothesis testing procedure.

What were common student errors or omissions?

Common errors when answering this question included the following:

- A surprising number of students drew conclusions by merely inspecting the data and did not use a formal inference procedure.

- Many students confused the test for independence and the test for homogeneity, causing them to give incorrect hypotheses.
- Many students tried to conduct a two-sample test for means or proportions, often treating the given frequencies as if they were data and using them to compute means and standard deviations.
- Many students failed to adequately address the required conditions (assumptions) for the test.
- Some students stated the required conditions for the test but gave no indication they had checked to determine if these conditions were met.
- Many students did not provide any link between the computations performed as part of the test procedure and the eventual conclusion.
- Many students did not provide a conclusion in context.

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

Emphasize the importance of checking required conditions for statistical procedures. Encourage students to be explicit about what led them to reject or fail to reject the null hypothesis, and to provide a conclusion in the context of the question. Give students practice in determining which of the various test procedures they have learned is the appropriate one to use in a given situation.

Question 6

What was the intent of this question?

Question 6 was the exam's investigative task. As such, its intent was to evaluate students' understanding in several course topic areas and to assess their ability to integrate statistical ideas and apply them in a new context. This year's investigative task involved interpreting a graph, confidence intervals, and formulating and communicating statistical arguments.

How well did students perform on this question?

The mean score for this question was 1.4 out of a possible four points. This question had the same average score as last year, and students still struggled with communicating statistical arguments in an understandable way.

What were common student errors or omissions?

Common errors when answering this question included the following:

- In Part (a) some students did not justify their answer by linking the response to the given information.
- In Part (b) most students failed to recognize that there are required conditions (assumptions) that need to be checked prior to computing a confidence interval.
- In Part (b) many students failed to provide an interpreation of the computed interval, even though one was specifically requested. Some students gave an interpretation of the confidence level rather than an interpretation of the interval itself.
- Many of the students who did provide an interpretation of the confidence interval gave an incorrect interpretation, implying that the population proportion of similar markets with strong demand was a variable quantity and not a fixed, but unknown, value.
- In Part (b) some students used a different confidence level than the one specified in the problem.

- In Part (c) many students used the point estimate of .65 rather than the confidence interval from Part (b) as the basis for making a decision.
- In Part (d) many students did not recognize that, for a particular market, demand will either be strong or weak and so did not recognize that, since all values in the confidence interval were greater than .5, the demand was more likely to be strong than weak.

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 a good idea to give students practice with problems that are investigative in nature. Many students do not seem to have had enough practice with problems that are not "standard."

General Comments on Exam Performance

Exam performance was better this year than in recent years. This was reflected in higher overall scores as well as higher scores on both the free-response and the multiple-choice sections of the exam. As in past years, performance was strongest in the area of describing data and weakest in the area of statistical inference. This was apparent in both the free-response inference questions as well as in the multiple-choice questions dealing with inference. In general, students were much stronger on the mechanical and computational aspects of problems than on parts that required interpretation or conceptual understanding. Communication of results continues to be a weakness, though there has been improvement in this area.

Areas identified in previous exam reports that continue to be problematic include the following:

- Many students failed to read questions carefully and as a result answered a question that was different from the one that had been asked.
- Many students did not answer questions *in context.* Explanations and conclusions in context are always necessary for a complete answer.
- More students stated assumptions when carrying out a hypothesis test but fewer understood that assumptions must also be checked.
- A disappointingly large number of students still seemed to believe that it is okay to draw conclusions by "just looking at the data" and did not seem to understand the need to employ inferential procedures, even when asked to provide statistical evidence to support conclusions.

What can teachers do to improve student performance in specific problem areas? They can:

- emphasize conceptual understanding and communication over mechanics,
- have students practice communicating conclusions and interpreting results throughout the course and not just at the end when inferential techniques are covered,
- allow sufficient time to cover the entire course outline, and
- integrate computer use if possible and, at a minimum, be sure students are comfortable with reading computer output.