# CollegeBoard
## Advanced Placement Program

# Student Performance Q&A:
## 2004 AP® Statistics Free-Response Questions

The following comments on the 2004 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?*

The intent of this question was to assess students' ability to use summary information for two distributions (one for each additive) to construct parallel boxplots. Students were asked to check for outliers and to identify any outliers on their graph. Finally, students were asked to interpret the information provided in the boxplots by making recommendations based on two different goals.

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

The mean score for this question was 1.3 out of a possible four points. Although the exploratory data analysis questions have been among the highest scoring questions in the past, this was not the case for 2004. This question tied with Question 6 for having the distinction of being the lowest scoring question this year.

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

Common errors when answering this question included the following:

- In Part (a) many students had trouble with the identification of outliers. Some students failed to show any outliers at all, while others showed too many. Those who showed too many either used an incorrect formula/rule of Median $\pm$ 1.5IQR, or took any point beyond the first and third quartiles as being an outlier.

- In Part (a) many students failed to include labels for the axes.

- In Part (a) some students even failed to label which boxplot went with which data set.

- In Part (a) some students constructed boxplots by treating the information provided in the stem of the problem as a complete data set.

- In Part (a) some students extended the whiskers on their boxplot out to $Q_1 - 1.5\text{IQR}$ and $Q_3 + \text{IQR}$ rather than to actual data points.

- In Part (a) some students created regular boxplots and then identified the outliers on the whiskers.

- In Part (b i) many students focused on comparing the first quartile for additive A with the median for additive B since they had the same value, one. These students did not recognize that the relevant place of comparison is the value zero, not one.

- In Part (b i) many students failed to note that the first quartile for additive A was positive, so at least 75 percent of the values were positive, while the first quartile for additive B was negative, so less than 75 percent of the values were positive.

- In Part (b i) many students referred to the values covered by the box as "the IQR" and did not recognize that IQR is a numerical statistic.

- In Part (b i) some students constructed the boxplots correctly in Part (a) but attempted to use the given values as data to answer Part (b i).

- In Part (b i) some students described the distribution of A but failed to make the comparison to the distribution of B.

- In Part (b i) some students tried to use the mean or median alone to make the comparison, or attempted to answer the question using range and IQR alone.

- In Part (b ii) many students tended to focus only on the extreme values at the upper end of the distribution for additive B.

- In Part (b ii) many students did not understand that, in general, when a distribution is skewed to the right, the mean will be greater than the median.

- In Part (b ii) many students detailed the effect of the skewness in the distribution for additive B but failed to describe the effect of A's shape on its mean.

- In Part (b ii) some students reversed the direction of skewness.

- In Part (b ii) some students who used the mean for comparison in (b i) tended to change their strategy for (b ii).

- In Part (b ii) some students tried to answer the question using the range and IQR.

- In Part (b ii) some students described the shapes of A and B beautifully but never made a decision.

- In Part (b ii) some students described the distribution of B but failed to make the comparison to the distribution of 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?***

Students need to look carefully at the information they are given before making appropriate graphical displays. Identifying outliers and describing the impact of the shape of the distribution on basic

descriptive statistics are important aspects of exploratory data analysis in which students could use more practice. Exposing students to practical problems where different goals or priorities lead to different answers or results would be extremely beneficial.

## Question 2

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

The intent of this question was to evaluate students' understanding of blocking and randomization in the context of practical experiment where different researchers have different opinions. Good communication, especially when describing the criteria used to form the blocks in Parts (a) and (b) and the appropriate method for assigning treatments in Part (c), was particularly important on this problem.

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

The mean score for this question was 1.6 out of a possible four points. This was the highest scoring question for 2004, but poor communication skills cost many students valuable points.

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

Common errors when answering this question included the following:

- Many students failed to clearly indicate the criteria for blocking.

- Many students assumed that the rationale for forming the blocks should also include a discussion of how to assign treatments to subjects. These students lost valuable time that could have been spent on other questions.

- Many students lacked an understanding that blocks, based on the blocking criterion, should be as homogeneous as possible.

- In Part (a) many students said it was important to have one male and one female in each block, since gender was not thought to have an effect on the outcome of the experiment.

- In Part (a) many students gave insufficient criteria (e.g., "block by age," "put the volunteers in order by age," etc.) for establishing their blocks.

- In Part (a) some students formed subgroups of size four of the young (20s), middle-aged (40s), and old (late 50s and 60s) volunteers. Students then randomly selected pairs from each of these subgroups to form the blocks rather than forming blocks with people of the nearest ages.

- In Part (b) many students made the blocks as heterogeneous as possible (e.g., "match the youngest male with the oldest female").

- In Part (b) some students thought that blocking by gender meant that each block should contain one male and one female rather than having two individuals with the same gender in each block.

- In Part (c) many students stated that the treatments should be assigned to the volunteers within each block randomly; but they stopped there, failing to provide a description of their randomization method, or they stopped short of completely describing the method. (*Note:* If a student said "flip a coin" or "use a random digit table," this was not considered an adequate description. The student needed to describe how these methods were used to assign the treatments.)

- In Part (c) some students did not understand that randomization should be carried out separately within each block.

- In Part (c) some students simply stated that selecting three blocks out of six was wrong because the selection needed to be done randomly.

***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 read the question carefully and answer the question they have been asked. It would also be helpful if teachers illustrated and discussed the benefits of writing complete, focused responses. Students wasted valuable time on the exam by writing information they had memorized from Content Area II, relevant or not, just to fill up the space provided in the exam booklet.

## Question 3

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

The intent of this question was to assess students' ability to recognize why a common probability distribution should not be used in this problem, to compute an appropriate probability, to make a decision based on the computed probability, and to describe whether or not the sample was representative of a particular population.

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

The mean score for this question was 1.4 out of a possible four points. Many students seemed to recognize that sampling without replacement violated one of the conditions for the binomial distribution, but they were unclear which condition was being violated. After pointing out that the binomial distribution was not appropriate, many students proceeded to use the binomial distribution to calculate the probability in Part (b). Many students had trouble interpreting the probability they had calculated in Part (b) and deciding whether or not the sample was representative of the population of all brontosaurs.

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

Common errors when answering this question included the following:

- Many students did not understand the conditions associated with the proper use of the binomial distribution. This was especially true in the distinction between independent trials and a constant probability of success.

- Many students lacked specificity about the probability that was being discussed in the problem. Students often referred to the probability of selecting "a bone" rather than the probability of selecting a "male femur bone."

- Some students stated that the problem was not binomial because they did not know the number of males and females at the site.

- In Part (b) many students used the binomial distribution, even though they had just argued that the binomial distribution was inappropriate in Part (a).

- In Part (b) many students illustrated a weak understanding of conditional probability.

- In Part (b) many students provided the incorrect solution $\frac{10}{20} \times \frac{9}{20} \times \frac{8}{20} \times \frac{7}{20}$.

- In Part (c) many students did not relate their response to their answer in Part (b).

- In Part (d) many students failed to mention that the bones were not a random sample from the archaeological site.

- In Part (d) many students thought that samples of size 20 could not represent the entire population, regardless of how they were selected.

- In Part (d) many students gave a very good answer to a different question; that is, they discussed why they could not conduct an inference procedure in this situation.

- In Part (d) some students thought that a sample from one area could not represent the entire population.

- In Part (d) some students never indicated if their actual choice was yes or no.

***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?***

Recognizing when common probability distributions should not be used is at least as important as recognizing when they should be used. Asking students to think about the conditions necessary to use a specific probability distribution in a variety of practical settings would help reinforce the fact that these conditions are required for a reason. Experience with both good and bad sampling methods will help students decide whether or not a particular sample is representative of a general population.

## Question 4

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

The intent of this question was to assess students' ability to compute two probabilities and two expected costs, and then to make a recommendation based on the calculated values.

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

The mean score for this question was 1.4 out of a possible four points. Students struggled with the computations required for this problem. Sorting out the information provided and applying the appropriate methods created more trouble than in previous years. Students usually perform very well on parts of problems where calculations are required, but that was not the case for this question.

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

Common errors when answering this question included the following:

- Many students confused the plans (I and II) with the antibiotics (A and B).

- Many students provided numerical answers with no justification or with an incomplete justification.

- Many students had trouble organizing their probabilities and recognizing that independence allowed them to calculate joint probabilities.

- Some students interchanged the probabilities and the costs.

- Some students obtained probabilities larger than one but did not comment that this result is obviously incorrect.

- In Part (b) many students did not understand that the cost for the first drug administered had to be paid whether or not the drug worked.

- In Part (b) many students added $50 to $80 and did not use probabilities to obtain the expected value.

- Many students missed Part (c) because they did not answer the question. Students formulated answers using information from the stem of the question, not from Parts (a) and (b).

- Some students never made a recommendation, while others provided more than one.

***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 read the question carefully and respond to the question that has been asked on the exam. Providing extraneous information reduces the time that is available for other questions, and if this information is incorrect, it must be scored as incorrect. Computing expected costs (or expected values in general) requires the use of both costs (or values of the random variable) and probabilities.

## 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, state conclusions in context, and decide whether or not a specific estimate was reasonable for a particular population parameter.

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

The mean score for this question was 1.4 out of a possible four points. Student performance on standard hypothesis-testing questions continues to be disappointing. Many students are still having trouble stating hypotheses, identifying a test, checking appropriate conditions, calculating the test statistic and p-value, and providing a conclusion in the context of the problem. Students also continue to have trouble deciding whether the estimate provided is reasonable based on the sample and the population of interest.

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

Common errors when answering this question included the following:

- Many students had trouble specifying the appropriate hypotheses. Variable names were not used in the statement of hypotheses, null and alternative hypotheses were interchanged, incorrect notation was used (e.g., using p-hat to represent a population proportion), and so on.

- Many students stated hypotheses for one test and then wrote conclusions for a different test. More specifically, students often mixed parts and pieces of a chi-square test of association with those for a two-sample z test for proportions.

- Many students made mistakes in determining if the conditions for the appropriate test were satisfied in this particular problem. Conditions were completely ignored or stated without computing expected counts. Expected counts were computed incorrectly or computed correctly but with no comments made about them.

- Many students did not show their work for computing the test statistic and p-value and often provided only calculator output.

- Some students thought that expected counts had to be whole numbers and frequently made mistakes in writing the correct mechanics for their test.

- Many students did not write their conclusion in the context of the problem.

- In Part (b) many students inappropriately tied their answer to the solution in Part (a).

- In Part (b) many students did not clearly indicate that the sample was drawn from the population of interest.

- In Part (b) some students were concerned with either the size of the sample or the size of the population.

***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?***

Practice on numerous problems dealing with statistical inference is vital for success. Students must be able to recognize the similarities and differences associated with the hypothesis-testing procedures they are learning to master. While the overall idea is similar from one test procedure to another, students need time to assimilate the details and recognize the differences.

## Question 6

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

The intent of this question, known as the investigative task, was to evaluate students' understanding in several content areas of the course and to evaluate their ability to put together major statistical ideas in a new context. Investigating the relationship between confidence intervals (both one and two sided) and hypothesis tests (both one and two sided) provided the framework for this year's task.

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

The mean score for this question was 1.3 out of a possible four points. This question tied with Question 1 for having the distinction of being the lowest scoring question in 2004. The mean score for this question is lower than the mean scores on the investigative tasks from 2003 and 2002. Students continue to struggle with all types of problems dealing with statistical inference.

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

Common errors when answering this question included the following:

- Many students did not realize that conditions needed to be checked for a confidence interval.

- Many students confused the interpretation of a confidence interval with the interpretation of a confidence level.

- In Part (b) many students did not understand the difference between one-sided and two-sided tests.

- Many students never made a connection between the one-sided nature of the test in Part (b) and the two-sided nature of the confidence interval in Part (a).

- Many students did not seem to be familiar with the concept of the duality between two-tailed significance tests and confidence intervals.

- Many students did not know how to find the correct critical value, even though the problem explained how to construct a one-sided interval.

- Many students choose two-sided values for t*, not a one-sided value.

- Many students provided a $z^*$ critical value rather than a $t^*$ critical value.

- Some students used the same critical value in computing the lower bound L in Part (c) as they had when calculating the confidence interval in Part (a).

- A substantial number of students provided satisfactory rationales for Part (d). Many students seemed to understand that when a proposed parameter value falls outside a confidence interval, it is therefore not a plausible value.

- Many students misinterpreted the question to mean "interpret whether the lower bound of the interval changed" rather than "does the interval change the conclusion of the problem?"

- Many students suggested that confidence intervals were less precise than hypotheses tests.

- Some students made a decision with no justification.

- Many students had problems communicating on 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?*

Try to get students to think critically about quantitative information, especially in unfamiliar settings. Expose them to new problems and data sets, and ask them to use the quantitative reasoning skills they have learned to suggest possible analyses and solutions. Classroom investigations of this type can be valuable for all participants.

## General Comments on Exam Performance

Overall performance on the multiple-choice questions was better this year than in the previous two years, but the scores on the free-response questions were lower than the previous two years. The overall mean score was down slightly but very close to those in 2002 and 2003. The best news from 2004 is that all six questions showed good discrimination across the entire range of scores. The most discouraging news is that students continued to perform poorly on standard problems dealing with statistical inference.

Students need to be encouraged to show all of their work and justify their answers. Many students continue to provide solutions with no justification or an incomplete justification. Another issue that continues to be a problem is that students fail to read the question carefully. Thus, they tend to provide information that is not relevant to the question that has been asked. Finally, communication of statistical analyses and concepts continues to be a problem.