Student Performance Q&A:
2001 AP® Statistics Free-Response Questions

The following comments are provided by the Chief Faculty Consultant regarding the 2001 free-response questions for AP Statistics. *They are intended to assist AP workshop consultants as they develop training sessions to help teachers better prepare their students for the AP Exams*. They give an overview of each question and its performance, 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 included. Consultants are encouraged to use their expertise to create strategies for teachers to improve student performance in specific areas.

The number of students taking the AP Statistics Exam increased again this year -- to approximately 42,000. Overall student performance improved compared to the past two years, with higher scores this year on the multiple-choice section of the exam and more consistent performance across the six questions of the free-response section. We hope to see scores continue to improve in the coming year, and this document is intended to help teachers better understand expectations and to assist them in preparing students for the AP Statistics Exam.

General Comments on Exam Performance

Exam performance this year (and in past years) 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.

Areas identified in previous exam reports that continue to be problematic are listed below.

- Many students failed to read questions carefully and, as a result, answered a question different from the one that was asked.

- Many students did not answer questions **in context**. Explanations and conclusions in context are always required for a complete answer.
- More students than in past years stated assumptions when carrying out a hypothesis test, but few 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 their conclusions.

Question 1

The purpose of question 1 was to assess the student's understanding of numerical measures of center and spread and the concept of outliers. To receive full credit for this problem, the student needed to describe a plausible procedure that used the summary statistics provided to determine whether outliers are present in a data set. After describing a general procedure, the student was to use the procedure to determine if there were any outliers in the data set for which summary statistics were given. In part (c), the student was expected to comment on whether 10 inches of rain was an unusual value for this data set.

Common errors in answering question 1 were:

- Many students had a general idea of the meaning of the term outlier but were unable to describe an algorithm for identifying outliers.
- Many students gave implausible descriptions of outliers, such as describing outliers as those values that are within the middle 50% of the data or as any values that were larger than the maximum. Others defined procedures based on using the standard deviation with the median or the IQR with the mean.
- Some students confused the idea of skewness with that of outliers, assuming that if a distribution was skewed there must be outliers. Others assumed normality (in spite of rather strong evidence against this in the given summary statistics) and based decisions on the Empirical Rule or did normal probability calculations in part (c).
- Some students did not recognize that a general procedure for identifying outliers would need to address both ends of the distribution.
- Some students did not understand that the quartiles and the IQR are numbers and not a range of values, making statements such as "10 is in the IQR".
- In part (c), many students described the value of 10 as usual or unusual without supporting this statement.

Question 2

The purpose of question 2 was to evaluate the student's ability to reason about information provided in the form of a probability distribution. Several different approaches to this problem were reasonable—an approach based on comparing the expected cost of machine B to the fixed cost of machine A, an approach based on computing the probability that the cost of machine B would exceed the fixed cost of machine A, and an approach based on simulation. While the majority of students chose the expected value approach, there were also many good responses that took the probability approach, as well as occasional simulation-based responses.

Common errors in answering question 2 were:

- Many students neglected to include the initial costs in the comparison.
- Many students compared a one-year expected cost for machine B to a three-year fixed cost for machine A.
- Some students did not recognize the difference between an expected value and a fixed cost. Others unnecessarily rounded the expected number of repairs because they thought that the expected value had to be an integer.
- Many students who chose the probability approach correctly identified 6.5 repairs as the break-even point, but then incorrectly computed the probability of 7 or more repairs for machine B. Some students stated that this probability was small without providing any support for this statement.
- A surprising number of students ignored the information in the given probability distribution altogether.

Question 3

The purpose of question 3 was to evaluate the student's ability to design and carry out a simulation to estimate the distribution of the number of prize winners in a weekly radio contest. To receive full credit on this question, students needed to describe how they would carry out the simulation and then conduct three trials using the given table of random numbers. Many students had difficulty describing the simulation process, and many responses received low scores because of poor communication.

Common errors in answering question 3 were:

- Many students did not explain how random numbers would be used to represent the coupons in the drawing.
- Many students incorrectly used digits to represent the possible prize value amounts rather than the coupons in the drawing.
- Most students failed to adequately address the non-replacement requirement.
- Some students did not describe or implement a correct stopping rule (i.e., indicate when the simulation should be stopped).
- In part (b), some students failed to report the number of winners for each trial, even though this was underlined in the question.

- Some students seemed to be completely unfamiliar with the use of random number tables. Some even crossed out the random number table provided and replaced it with digits from the random number generator on their calculator.

Question 4

The purpose of question 4 was to assess the student's understanding of some basic principles of experimental design, including blocking and randomization.

Common errors in answering question 4 were:

- A large number of students confused blocks with treatment groups, and so incorrectly chose blocking scheme B.
- Many students incorrectly stated that the purpose of blocking was to enable the researcher to test the effects of the forest, rather than to reduce variability in the response by ensuring that each treatment group had equal exposure to the forest.
- In part (b), students often made some generic statement to the effect that "randomization is a good thing" rather than explaining the purpose of randomization in the context of this problem.
- In part (b), many students continued to focus on the effect of the forest, failing to recognize that their blocking scheme had already dealt with the forest effects.
- Some students confused randomization in part (b) with choosing a random sample of trees.
- Some students explained how to randomize rather than why randomization was important.
- Many students who chose blocking scheme B in part (a) because they confused blocks and treatment groups gave answers in part (b) that were inconsistent or contradictory to their answer in part (a).

Question 5

The purpose of question 5 was to evaluate whether the student could carry out a test of hypotheses and state conclusions in context. To receive full credit on this question, the student needed to state hypotheses, identify an appropriate test procedure (by name or by formula), verify (not just state) any necessary assumptions, compute the value of the test statistic and the associated P-value (or rejection region) and then, based on the result of the test, state an appropriate conclusion in context.

Common errors in answering question 5 were:

- Most students did not recognize that the samples were paired and incorrectly chose a two-sample t-test. Some students stated that the samples were paired, but then still proceeded to use the independent samples test.
- Many students did not state or check assumptions, and some stated assumptions but provided no evidence that the assumptions had been checked and were reasonable. (A

✔ by a stated assumption is not sufficient for verifying that the assumption has been met or satisfied.)

- Many students did not identify the test procedure used and/or did not give the computed value of the test statistic or the corresponding degrees of freedom.
- An alarming number of students stated hypotheses in terms of sample means rather than population means.
- Some students did not recognize that the question called for a two-sided test.
- Some students used the calculator to compute the P-value for the test, and then misinterpreted the calculator's scientific notation, reporting an incorrect P-value without recognizing that a P-value must be between 0 and 1.
- Some students misinterpreted small P-values as an indication that the null hypothesis should not be rejected.
- Many students did not give conclusions in context.
- Many students did not link their ultimate conclusion to the outcome of the statistical test.
- A number of students did not recognize the need to employ an inferential procedure, and merely commented on the values of descriptive statistics.

Question 6

Question 6 was the exam's investigative task. As such, its purpose is to evaluate the student's understanding in several course topic areas and to assess ability to integrate statistical ideas and apply them in a new context. This year's investigative task involved using graphical displays to compare two groups, using inference about the slope of a regression line, and using bivariate data given for each of two groups to reason about the group membership of a new data point. It was a very rich problem, with a number of different possible reasonable approaches to the classification problem posed in part (c).

Common errors in answering question 6 were:

In part (a)

- Many students did not recognize that the question asked for a comparison of only the GPA variable, and so constructed bivariate displays (such as scatterplots) to try to answer this question.
- Many students gave graphs that did not have axis labels or scales. Some students used different scales for the two graphs, making comparison difficult.
- Many students did not focus on comparison of the two groups in discussing the important features of the graphs.

In part (b)

- Many students did not realize that a formal inferential procedure was needed, and stated only that there was a high correlation.
- Many students did not know how to carry out a hypothesis test in a regression setting.
- Some students incorrectly used a two-sample t test to compare mean GPA to mean number of credit hours. Others confused the Chi-square test for association in a two-way table with the test for slope in regression.

- <u>Many</u> students did not use the computer output provided, instead preferring to enter the data into their calculators and then trying to duplicate the analysis. If done correctly, this was OK (but was not a good use of time). However, many students produced incorrect values for the test statistic and P-value.
- Some students reversed the null and alternative hypotheses, leading to incorrect conclusions about the relationship.

In part (c)
- Many students failed to recognize the bivariate nature of the problem and unsuccessfully tried to classify the new observation by looking only at one variable or by looking at one variable at a time.
- Some students didn't recognize that they needed to evaluate the new data point with respect to both groups before making a decision.

What Can Teachers Do To Improve Performance in Specific Problem Areas?

- Emphasize conceptual understanding and communication over mechanics.
- Have students practice communicating conclusions and interpreting results *in context* throughout the course.
- Allow sufficient time to cover the entire AP course outline.
- Integrate computer use if possible, and, at a minimum, be sure that students are comfortable with reading computer output.
- Discourage the use of "calculator talk." This is not a good method for communicating what is being done.