Student Performance Q&A:

2009 AP® Statistics Free-Response Questions

The following comments on the 2009 free-response questions for AP® Statistics were written by the Chief Reader, Christine Franklin of the University of Georgia, Athens. 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) construct an appropriate graphical display for comparing the distributions of two categorical variables; (2) summarize from this graph the relationship of the two categorical variables; and (3) identify the appropriate statistical procedure to test if an association exists between two categorical variables and state appropriate hypotheses for the test.

How well did students perform on this question?

The mean score was 2.02 out of a possible 4 points.

What were common student errors or omissions?

Part (a)

- Using counts (frequencies) instead of percents (relative frequencies) when constructing the graph, which is not appropriate when comparing groups of unequal size
- Providing no label or an incorrect label on the vertical axis of the graph
- Indicating conditioning on one variable (e.g., gender) but drawing the graph as if conditioning on the other variable (e.g., job experience)
- Constructing nonstandard graphs of many varieties

Part (b)

- Failing to fully discuss how males and females compared in all three job experience categories, writing comments only on which gender had more (or fewer) part-time jobs and ignoring the two different categories of part-time job experience
- Struggling to communicate statistical thinking clearly when writing a few sentences about the association between gender and job experience
- Describing the graph as if discussing quantitative data and using terms like *shape*, *center*, *spread*, or *correlation* when none of these is appropriate for describing distributions of categorical data

Part (c)

- Failing to correctly name the appropriate significance test by giving it an incomplete name like "chi-square test," which on its own was not sufficient to earn credit because there are three distinct chi-square tests
- Stating hypotheses in a way that suggested causation, which was not appropriate for this observational study; for example, H_o: "Gender has no effect on job experience"; H_a: "Gender has an effect on job experience."
- Attempting to use symbols instead of, or in addition to, words when stating the hypotheses, which does not work well for a chi-square test of association/independence

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?

In many cases it appeared that students found knowing what type of graph to construct to represent categorical data to be a challenge. This suggests that students need more experience with scenarios involving the construction of graphical displays of categorical data. Even when a student could construct an appropriate graph, it then was often a challenge to know how to interpret the graph; many times the student described the graph as if it represented a quantitative variable. When covering the "exploring data" portion of the AP Statistics curriculum, it is often more typical to focus on quantitative data; however, in the real world, the type of data most often encountered may be categorical. Additionally, the concept of independence with categorical data is a subtle one that needs to be emphasized.

Question 2

What was the intent of this question?

The primary goals of this question were to assess a student's ability to (1) calculate a percentile value from a normal probability distribution; (2) recognize a binomial scenario and calculate an appropriate probability; and (3) use the sampling distribution of the sample mean to find a probability for the mean of five observations.

How well did students perform on this question?

The mean score was 0.84 out of a possible 4 points.

What were common student errors or omissions?

Part (a)

- Demonstrating difficulty with determining the percentile number of students; for example, setting $z = 0.7$ and attempting to calculate the stopping distance
- Interpreting the 70th percentile to be 70 percent centered about the mean and then using the 68–95–99 percent rule to answer the question
- Using calculator syntax to calculate stopping distance without refining the parameters of the normal distribution
- Providing unclear and unlabeled sketches of the approximately normal distribution

Part (b)

- Using the binomial distribution incorrectly to calculate $1\text{-P}(Y \leq 2)$ instead of $1\text{-P}(Y \leq 1)$
- Calculating only one term, $\text{P}(Y = 2)$
- Using 0.7, instead of 0.3, as the probability of a success
- Using calculator syntax instead of setting up the problem and clearly defining the binomial parameters
- Constructing another normal probability without recognizing that the scenario had become binomial

Part (c)

- Failing to recognize that the question was asking about a sampling distribution and not knowing how to define the distribution or its parameters correctly
- Giving a value of $z = 1.72$, and a p-value of 0.0427, without correctly indicating that this probability related to $\text{P}(Z \geq 1.72) = 1\text{-P}(Z < 1.72) = 0.0427$
- Confusing the question with a test of hypothesis and giving the probability $\text{P}(Z \geq 1.72) = 1\text{-P}(Z < 1.72) = 0.0427$ as a p-value

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 vital to help students understand the importance of *always showing complete work* in arriving at a numerical answer. In all parts of the question the responses in which students gave a numerical answer with no justification, or showed work that was simply calculator syntax, did not earn full credit. Teachers should stress the importance of providing written statements or sketches of the distribution. This question asked students to recognize three distinct distributions and scenarios: (1) a population scenario for a continuous variable approximated by a normal distribution, (2) a scenario for a discrete variable using the binomial distribution, and (3) a scenario using the sampling distribution for sample means approximated by a normal distribution. More practice with these scenarios in a similar exam-question format would benefit students by helping them learn how to identify scenarios for using these different types of probability distributions.

Many students were at a loss as to how to work through this question, especially parts (b) and (c). The concept of a sampling distribution, as assessed in part (c), is a notoriously difficult one that requires much attention and practice.

Question 3

What was the intent of this question?

The primary goals of this question were to assess a student's ability to describe (1) a randomization process required for comparing two groups in a randomized experiment and (2) a potential consequence of using self-selection instead of randomization.

How well did students perform on this question?

The mean score was 1.42 out of a possible 4 points.

What were common student errors or omissions?

Part (a)

- Using a stopping rule with a coin toss (or equivalent method) without prior randomization, which did not achieve a randomized design
- Failing to state a device/mechanism for randomization
- Assigning numbers to students without randomization and then using evens/odds or 1–12 and 13–24 to select the members of the two groups
- Neglecting to specify groups in context (e.g., forming "Group 1" and "Group 2" but not indicating which was the dissection group and which was the computer software group)
- Failing to give enough information so that two knowledgeable statistics users would employ the same method to assign the students to the two instructional groups
- Making references to simple random samples when the scenario was an experimental setting asking for random assignment of treatments
- Providing only a design diagram with no explanation
- Picking names or numbers from a hat, bag, or laundry basket but forgetting to *mix* the contents of the container before selecting them
- Forgetting to block on *similar* pretest scores when using a paired design
- Failing to be explicit enough on how to put the students in random order (e.g., simply using the words "randomly assign")
- Overrandomizing where the initial randomization was correct but another randomization was incorrect (e.g., doing proper randomization into two groups and then using a poorly described coin flip to assign treatments to groups)
- Describing inappropriate or poorly designed blocking schemes
- Attempting to form blocks on a characteristic other than pretest, such as gender

Part (b)

- Providing a reasonable characteristic (that was not a mistake) but stating only that students "like it"
- Failing to describe how behaviors associated with the self-selection criterion impacted the changes in the differences (posttest - pretest)
- Referring only to the posttest instead of to the change in score (posttest - pretest), or mentioning only a vague aspect of performance (e.g., "do better," "learn more/less")

- Requiring Exam Readers to infer that the student taking the AP Exam knew that an effect on the posttest score would affect the response variable (change)
- Using terms like *bias*, *observation*, and *voluntary response* unclearly
- Mentioning a characteristic only, without making any connection to performance

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?

Poor communication (i.e., lack of clarity in students' writing) was an issue in the responses to this question, both when students were attempting to clearly describe a randomization scheme and when they were trying to provide *statistical* justification and reasoning for the issues with self-selection. Students often know certain buzzwords to use in a design question; however, without a clear explanation of the statistical concept in context, no credit will be given to a student's answer. The concept of confounding is a challenging one for students to understand and be comfortable explaining, so they need considerable practice with it. In particular, students must learn that a confounding variable needs to be associated with both the explanatory and the response variables.

Question 4

What was the intent of this question?

The primary goals of this question were to evaluate a student's ability to (1) identify and compute an appropriate confidence interval after checking the necessary conditions; (2) interpret the interval in the context of the question; and (3) use the confidence interval to make an inference about whether or not a council member's belief is supported.

How well did students perform on this question?

The mean score was 1.64 out of a possible 4 points.

What were common student errors or omissions?

Part (a), step 1

- Identifying a z confidence interval as the appropriate procedure, rather than a t confidence interval
- Failing to check the sample size condition at all
- Doing an inadequate job of checking the sample size conditions by saying that the samples were large enough but making no reference to a number like 25 or 30, the central limit theorem, or sampling distributions
- Stating that 50 was large enough to assume that the populations or samples or data were approximately normal, rather than that the sampling distribution(s) of the mean(s) was(were) approximately normal

Part (a), step 2

- Using 1.645 as the multiplier in the computation of the interval

- Neglecting to square the standard deviations when computing the standard error and consequently presenting an incorrect final answer
- Believing the interval could not go below 0 and incorrectly truncating the lower bound of the interval at 0

Part (a), step 3

- Omitting the word "mean" and interpreting the confidence interval as applying to the difference in individual response times
- Omitting the word "difference," or any similar wording that indicated that the interval was for a difference in means, and incorrectly stating that the interval was for the "mean response time"
- Omitting context from the interpretation
- Interpreting the confidence level instead of the confidence interval
- Interpreting the confidence interval correctly but then unnecessarily interpreting the confidence level incorrectly
- Writing that the confidence interval was for a "mean proportion" or a "proportion of difference" or employing similar phrasing that used the word "proportion"

Part (b)

- Writing a statistically incorrect statement, such as "Because the interval contains 0, the council member's belief is wrong"
- Believing that the interval supported the council member's belief because it included more values on the positive side of 0 than the other
- Believing that the interval supported the council member's belief because it included values as large as 2 minutes
- Basing a conclusion solely on testing hypotheses and making no reference to the confidence interval

Student errors found in every part of the question included:

- Computing two separate confidence intervals instead of a confidence interval for the *difference* in means
- Confusing notation for sample and population means
- Presenting a formula for the confidence interval with incorrect numbers substituted but then using a calculator to compute the correct interval and using the calculator version to answer the remainder of the question
- Referring to the sample standard deviations of 3.7 and 3.2 as σ_1 and σ_2 or calling them population standard deviations

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?

Numerous questions from previously administered AP Statistics Exams are available on AP Central® and can help teachers prepare students to go through the three steps of finding and interpreting a confidence interval. Using questions from past AP Exams may assist in improving students' responses to a confidence-interval problem. For this particular confidence interval, students needed to understand why using the *t*-interval, rather than the *z*-interval, was appropriate. The issue is that the *t*-distribution is a closer approximation for the sampling distribution of the test statistic when the population variances are unknown (which was the focus of one student's response) than using the standard normal distribution. Students are expected to be able to distinguish the situation from one where the population variance is known, where the use of *z* is appropriate. Since most students can use their calculator to find the confidence interval, they need to know the distinction between when to tell the calculator to use *t* and when to tell it to use *z*. If a student tells the calculator to use *z*, then the value for a sigma (the population standard deviation) will be requested, which was not part of this question. When forming a confidence interval for a mean(s) and using a sample standard deviation(s), students should use the *t*-distribution. It is also important for students to focus on what parameter is being estimated with a confidence interval: in this case, the difference between two populations' means.

Question 5

What was the intent of this question?

The primary goals of this statistical inference question were to assess a student's ability to (1) interpret a *p*-value in context; (2) make an appropriate conclusion about the study based on the *p*-value; and (3) based on the conclusion, identify the type of error that could have occurred and a possible consequence of this error in context.

How well did students perform on this question?

The mean score was 0.96 out of a possible 4 points.

What were common student errors or omissions?

Part (a)

- Confusing the *p*-value with the significance level (writing that the *p*-value was the probability of rejecting H_0)
- Interpreting the *p*-value as the probability that H_0 (or H_a) was true (or false)
- Omitting a reference to the difference between proportions obtained in this study; for example, "There is a 7.61 percent chance that the treatment that uses CC alone produces a higher survival rate than CC + MMR, if the true difference between the survival rates is 0"
- Omitting "as large as" in the probability phrase
- Writing "by chance alone" or "as a result of sampling variation" instead of the more complete conditional phrase "if the survival rates for the two treatments (CC alone and CC + MMR) are in fact the same"
- Omitting the conditional phrase "if the survival rates for the two treatments (CC alone and CC + MMR) are in fact the same"
- Omitting the context

Part (b)

- Accepting H_0
- Omitting linkage to the *p*-value in part (a)
- Omitting context

Part (c)

- Confusing Type I and Type II errors
- Providing a "consequence" that was a decision on the part of the statistical analysts rather than an action applied by medical professionals to heart-attack patients
- Lacking specificity with respect to treatments; specifically, failing to distinguish whether both treatments meant CC + MMR and CC, or CC + MMR alone

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?

Poor communication (i.e., lack of clarity in students' writing) was again an issue in responses to this question. This inference question did not ask students to carry out a traditional test of significance with the four-step response template, but instead it required them to clearly articulate the interpretation of a *p*-value and to recognize the type of error potentially committed in the significance test, along with clearly describing a potential consequence of the error in context. The most common mistake students made with the *p*-value interpretation was leaving off the conditional statement. Students need practice with realizing the *p*-value is a probability that is calculated assuming that the null hypothesis is true. Performance on this question emphasizes the importance of giving students practice with writing interpretations of important statistical concepts in a context.

Question 6

What was the intent of this question?

The primary goals of this investigative task were to assess a student's ability to (1) define a parameter and state a correct pair of hypotheses; (2) explain how a particular statistic measures skewness; (3) use the observed value of the statistic and a simulated sampling distribution to make a conclusion about the shape of the population; and (4) create a new statistic and explain how it measures skewness.

How well did students perform on this question?

The mean score was 1.32 out of a possible 4 points.

What were common student errors or omissions?

Part (a)

- Failing to understand how to define the parameter of interest; instead Readers saw such attempts to define the parameter as:
    - "The mpg of the cars" (the variable of interest)
    - "All the cars of this model" (the population of interest)
    - "To determine if the manufacturer is misleading customers" (the question of interest)

- Attempting to define the parameter more than once (e.g., writing, “The parameter is . . .” and then later stating, “ μ = . . .”); these were treated as parallel solutions, and the worst attempt was scored
- Using nonstandard notation in the hypotheses, often without explicitly defining the notation
- Using a two-tailed alternative hypothesis

Part (b)

- Reversing the relationship between the mean and the median in a right-skewed distribution (i.e., stating that the mean would be less than the median in a right-skewed distribution)
- Making reasonable statements about the relationship between the mean and the median but not stating that large values of the statistic indicated right skewness
- Stating that large values of the statistic indicated right skewness but only arguing that in a normal (or symmetric) distribution the ratio should be close to 1 and not explaining how the mean and the median were related in a right-skewed distribution
- Stating “large” without any explanation

Part (c)

- Failing to understand that the dotplot approximated the sampling distribution of the statistic (sample mean) / (sample median); in other words, not understanding that the graph showed what values of the statistic would occur when sampling from a normal population
- Believing the dotplot showed sample data (as opposed to simulated values of a sample statistic), describing the shape of the dotplot as approximately normal, and using this to justify that the original sample came from a normal population
- Believing the values in the dotplot came from new samples of size 10 from the original population instead of from a normal population
- Failing to understand how to use the dotplot to make an appropriate conclusion; in other words, not knowing to look for where 1.03, the observed value of the sample statistic, fell in the distribution and then using that relative position to explicitly indicate whether or not a value of 1.03 would be likely to occur by chance when sampling from a normal population
- Stating the relative position of 1.03 without specific numerical evidence from the dotplot (e.g., “1.03 is toward the middle of the distribution”) and then correctly deciding that it was plausible that the original sample came from a normal population
- Believing that 1.03 was unusual enough to conclude that the original population was skewed to the right (e.g., “1.03 is in the tail of the distribution so I conclude that the sample came from a right-skewed population”)
- Stating that the dots were centered around 1, so the sample came from a normal population. Because the sampling distribution was generated by using samples from a normal population, this was not surprising; however, it did not address whether or not 1.03 is unusual.
- Arguing that 1.03 is close to 1 without describing its relative position in the dotplot. It was clear that many students were thinking simply about the absolute difference between 1 and 1.03, without considering the variability in the sampling distribution.

- Stating that the sample size (or number of samples) was large, so the distribution was normal, or claiming that the sample size was too small to make a conclusion
- Stating that the sample came from a normal population and providing no explanation

Part (d)

- Failing to provide a statistic that measured skewness; this typically occurred if the student looked only to the right half of the distribution (e.g., max / Q3) or used a measure of spread (e.g. (max–min) / median)
- Providing a method for identifying skewness but not stating a well-defined statistic; for example, "if (max–med) > (med–min), then the distribution is skewed right"
- Failing to correctly identify the values of the defined statistic, which indicated skewness to the right; for example, looking for values less than 1 when using (med–min) / (max–med)
- Failing to justify the values that indicated right skewness by discussing how right skewness affected the relationship between the components of the statistic
- Trying to use outlier rules to measure skewness; students who used these rules correctly and concluded there was right skewness if there were outliers on the right but not on the left received credit for a reasonable method but not for a reasonable statistic

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 good news with respect to the investigative task this year is that students were given the opportunity to be creative in their responses, and many nicely demonstrated their ability to think beyond the traditional textbook problems by creating a statistic to measure skewness. However, for the standard questions of the investigative task, in particular part (a), many students still struggled with understanding the concept of a parameter and how to define a parameter in context. Furthermore, many students failed to recognize a sampling distribution from the simulation in part (c) and how to work with this sampling distribution. More work with simulations and sampling distributions is encouraged. It is important to help students understand the difference between a population, data (one sample), and sampling distribution of a statistic under repeated sampling.

Also important to emphasize is that most sampling distributions are created assuming that a null hypothesis is true; so, if the observed value of the sample statistic turns out to be in the tail of the sampling distribution, that provides evidence against the null hypothosis. More practice on investigative-task types of questions will help students integrate concepts and statistical tools, thus helping them apply their knowledge in a new setting. Students should also be reminded to allow a sufficient amount of time for responding to the investigative task because it carries more weight than any of the other five free-response questions on the exam.