Chief Reader Report on Student Responses:

2018 AP® Statistics Free-Response Questions

- Number of Students Scored: 222,501
- Number of Readers: 882
- Score Distribution:

Exam Score	N	%At
5	32,417	14.6
4	47,108	21.2
3	55,483	24.9
2	35,407	15.9
1	52,086	23.4

- Global Mean: 2.88

The following comments on the 2018 free-response questions for AP® Statistics were written by the Chief Reader, Ken Koehler from Iowa State University prepared this document. 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 preparation 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 **Max. Points: 4** **Mean Score: 1.63**

What were the responses to this question expected to demonstrate?

The primary goals of this question were to assess a student's ability to (1) identify various values in regression computer output; (2) interpret the intercept of a regression line in context; (3) interpret the coefficient of determination in context; and (4) identify an outlier from a scatterplot.

How well did the response address the course content related to this question? How well did the responses integrate the skills required on this question?

In part (a), many responses correctly communicated the concept of an intercept using context. Identifying the intercept from the given output was also accomplished by many students.

In part (b), most responses identified the correct value of the coefficient of determination from the given output, and almost all provided context in their interpretation.

In part (c), almost all responses correctly identified the outlier on the scatterplot.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
In part (a), communicating the concept of a slope instead of the concept of an intercept.	When there are zero customers in line, check-out time is predicted to take 72.95 seconds.
In part (a), not communicating that the value 72.95 is a prediction; an estimate of the mean check out time when there are no other customers in line.	When there are zero customers in line, checkout time is predicted to take 72.95 seconds.
In part (a), identifying an incorrect value of the intercept.	When there are zero customers in line, check-out time is predicted to take 72.95 seconds.
In part (b), not describing r^2 as a percent variation explained by a linear model.	The 73.33% of variation in checkout times is explained by the linear model relating mean checkout time to the number of customers ahead in line.
In part (b), describing r^2 as a percent variation of the wrong entity, such as variation in numbers of customers.	The 73.33% of variation in checkout time is explained by the linear model relating mean checkout time to the number of customers ahead in line.

In part (c), indicating that the circled point was far from the regression line, without communicating that it was farther than any other point.	The circled point is the farthest from the regression line (or, the circled point has the largest residual).

Based on your experience at the AP® Reading with student responses, what advice would you offer to teachers to help them improve the student performance on the exam?

- Give students practice identifying and interpreting the intercept of least squares regression line. Highlight the contextual difference between the interpretation of the intercept and the interpretation of the slope.
- Emphasize that the least squares regression line yields a prediction for each specific value of *x*, and the intercept is the prediction for $x = 0$.
- The format of linear regression output is similar across most software packages. Familiarize students with the output format from at least one software package, in addition to their calculators.
- Give students practice interpreting r^2 in a way that enhances the understanding that r^2 is the percent reduction in variation as measured by squared deviations of the *y* values from the least squares regression line relative to squared deviations of the *y* values from $\overline{y}$. This will help students t better understand what is meant by the commonly used statement that r^2 is the percent variation in the *y* values that can be explained by the variation in the *x* values.
- Emphasize that outlier identification requires comparison against the remaining data. Stating a circled point is far from the regression line makes no comparison with the distances from the line for the remaining data points. However, the necessary comparison is accomplished by stating that the circled point is *farthest* from the regression, or stating that the circled point deviates most from the pattern of the other data points.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- In general, review of previous exam questions and chief reader reports will give teachers insight into what constitutes strong statistical reasoning, as well as common student errors and how to address them in the classroom.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. To locate resources to give your students practice interpreting computer output, try searching the community for “output” and filtering for “Classroom-Ready Materials.” You may find worksheets, data sets, spreadsheets for using class data to simulate computer output, and guided notes offered by the author of a Statistics textbook, among other resources.

Question #2 Max. Points: 4 Mean Score: 1.08

What were the responses to this question expected to demonstrate?

The primary goals of this question were to assess a student's ability to (1) calculate the sample size when given the endpoints of a confidence interval for a proportion; (2) explain how bias could be present in a particular survey method; and (3) estimate a proportion from sample data collected using a method designed to decrease bias.

How well did the responses address the course content related to this question? How well did the responses integrate the skills required on this question?

- Most responses correctly identified a potential source of bias arising from face-to-face interviews with the environmental science teacher and its effect on the point estimate.
- Most responses incorporated good communication skills by explaining their answers and showing work for calculations.
- Many responses demonstrated understanding of how to work backwards from a confidence interval to the sample size.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
In working with the confidence interval in part (a), using $\hat{p}=0.5$ in the sample size calculation, even when they used $\hat{p}=0.7$ to calculate margin of error.	Solving $1.96\sqrt{\frac{(0.7)(1-0.7)}{n}}=0.116$ yields $n=\frac{(1.96)^2(0.7)(1-0.7)}{(0.116)^2}=59.95$, so the sample size is 60.
Using the formula for the standard error of a sample mean instead of the standard error for a proportion, and/or using a t^* critical value instead of a z critical value.	Solving $1.96\sqrt{\frac{(0.7)(1-0.7)}{n}}=0.116$ yields $n=\frac{(1.96)^2(0.7)(1-0.7)}{(0.116)^2}=59.95$, so the sample size is 60.
Giving an answer of "about 60" or "approximately 60" for the sample size. In this case, 60 is the exact answer.	Solving $1.96\sqrt{\frac{(0.7)(1-0.7)}{n}}=0.116$ yields $n=\frac{(1.96)^2(0.7)(1-0.7)}{(0.116)^2}=59.95$, so the sample size is 60.

Describing a bias based on something other than the situation described, such as problems with the wording of the question or voluntary response bias.	Because the students know that the environmental sciences teacher cares about the environment and they must respond directly to the teacher, some students may say "yes" when they actually do not recycle. This would result in an estimate of the proportion of all students who recycle that is larger than the true proportion.
Failing to contrast how students would respond to the question (what they say) with the truth about whether they recycle (what they do).	Because students know that the environmental sciences teacher cares about the environment and they must respond directly to the teacher, <u>some students may say "yes" when they actually do not recycle.</u> This would result in an estimate of the proportion of all students who recycle that is larger than the true proportion.
Providing an incorrect answer of 0.29 in part (c-ii) because of failure to account for the method used by the statistics teacher. Good responses accounted for the 150 students who were forced to answer "no" when they calculated the numerator and denominator of the point estimate.	The estimate is based on the expectation that 150 students will be required to say "no" and 150 students will answer truthfully. Of the 213 "no" responses, we expect $213-150-63$ were from students who answered truthfully. Consequently, we expect that $150-63=87$ out of 150 students who answered truthfully said "yes," and our estimate of the proportion of all students who answer truthfully is $\frac{87}{150}=0.58$.
In both part (b) and part (c), describing a point estimate as a population parameter.	As indicated in the box above, good responses made it clear that the value 0.58 is an <u>estimate</u> of the population proportion of students who would answer yes to the question.

Based on your experience at the AP® Reading with student responses, what advice would you offer to teachers to help them improve the student performance on the exam?

- Make sure students understand that using $\hat{p}=0.5$ is appropriate when calculating a sample size *before* a sample is collected. However, once a sample has been collected, the value of $\hat{p}$ is based on that sample and the value of $\hat{p}$ from the sample should be used consistently throughout the formula.
- Make sure students know when to use methods for proportions and when to use methods for means.
- Although many calculations in statistics result in approximate answers, make sure students realize that some answers are exact.
- Students need to be able to distinguish the different ways bias can be introduced in a sample survey. Give them practice within many different contexts.
- Remind students that for bias to exist, there must be a systematic difference between the responses from the sample and the truth about the population.

- Make sure students read the problem carefully and follow the lead of the question instead of trying to create a new situation.
- Many AP Statistics items have a "flow" in which preceding parts of a question help answer subsequent parts. Remind students to consider how the parts of an item are connected.
- When describing an estimate, students should use words such as "expect," "predict," and "estimate," to make it clear that students understand the estimate they provide is not the exact value of a population parameter.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- In general, review of previous exam questions and chief reader reports will give teachers excellent insight into what constitutes strong statistical reasoning, as well as common student errors and how to address them in the classroom.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. To locate resources for teaching about bias, for example, try searching the community for "bias" and filtering for "Resource Library." You may find a link to an article about an experiment about bias in the NFL, topical indexes of released Multiple Choice and Free Response Questions as of 2017, and vocabulary quizzes, among other resources.

Question #3 Max. Points: 4 Mean Score: 1.55

What were the responses to this question expected to demonstrate?

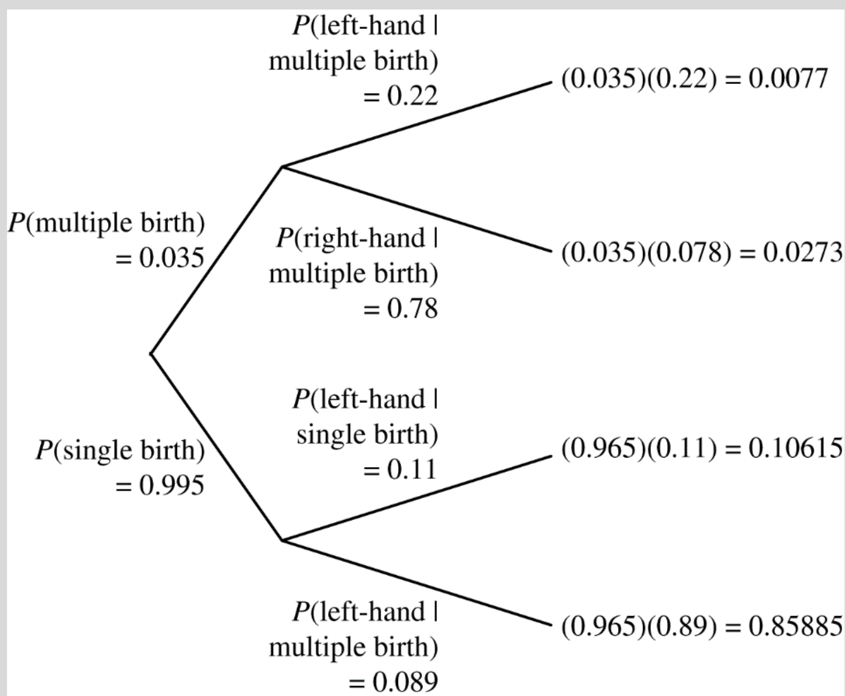
The primary goals of this question were to assess a student's ability to (1) compute a probability based on a weighted mixture of two populations; (2) compute a conditional probability; and (3) recognize a binomial random variable and compute the probability associated with it.

How well did the response address the course content related to this question? How well did the responses integrate the skills required on this question?

- Generally, responses created and used a tree to compute weighted averages of conditional probabilities.
- Many responses reported a correct formula for the conditional probability in part (b).
- Generally, responses presented a probability distribution that was discrete for part (c).

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

<table>
<tr><th>Common Misconceptions/Knowledge Gaps</th><th>Responses that Demonstrate Understanding</th></tr>
<tr><td>Not labeling the tree drawn and implicitly adding two of the branches without communicating which two branches were added. Generally, the responses with labeled trees scored higher if the trees were labeled with words, as opposed to notation. We saw substantial confusion between the following notations:

$P(L)$, $P(M \mid L)$, $P(L \mid M)$,
and $P(L \cap M)$</td><td>

$P(\text{left-hand}) = 0.007 + 0.10615 = 0.11385$</td></tr>
<tr><td>Mixing percentages and probabilities, e.g. $(3.5)(0.22) + (96.5)(0.11)$.</td><td>$(0.035)(0.22) + (0.965)(0.11) + 0.1138$</td></tr>
</table>

Not weighting the percent of left-handed children based on whether a child is from a single birth or multiple birth. That is, simply adding the two conditional probabilities; $0.22+0.11=0.33$.	$(0.035)(0.22)+(0.965)(0.11)+0.1138$
Weighting only multiple births and not weighting single births. The probability of a single birth was not explicitly stated in the stem of the problem, but can be computed as $1-0.035=0.965$.	$(0.035)(0.22)+(0.965)(0.11)+0.1138$
Failing to calculate a cumulative probability and calculating the probability that $X=3$.	$P(X \geq 3)=1-P(X \leq 2)$ $=1-\binom{20}{0}(.11385)^0(.88615)^{20}$ $-\binom{20}{1}(.11385)^1(.88615)^{19}$ $-\binom{20}{2}(.11385)^2(.88615)^{18}$ $=0.402$
Subtracting $P(X \leq 3)$ from 1 instead of subtracting $P(X \leq 2)$ from 1.	$P(X \geq 3)=1-P(X \leq 2)$ $=1-\binom{20}{0}(.11385)^0(.88615)^{20}$ $-\binom{20}{1}(.11385)^1(.88615)^{19}$ $-\binom{20}{2}(.11385)^2(.88615)^{18}$ $=0.402$
Using a normal distribution to approximate the binomial probability. Because $np<5$, the approximation is not sufficiently accurate.	$P(X \geq 3)=1-P(X \leq 2)$ $=1-\binom{20}{0}(.11385)^0(.88615)^{20}$ $-\binom{20}{1}(.11385)^1(.88615)^{19}$ $-\binom{20}{2}(.11385)^2(.88615)^{18}$ $=0.402$

Based on your experience at the AP® Reading with student responses, what advice would you offer to teachers to help them improve the student performance on the exam?

- Remind students to convert percentages to probabilities before doing probability computations.
- Emphasize that it is incorrect to add conditional probabilities.
- Teach the concept of a cumulative distribution and apply that concept to examples of discrete probability distributions before moving to calculators to numerically evaluate results.
- Emphasize that all parameters for calculator functions are expected to be labeled with the meaning of the values being passed to the function. Here are two examples.

  binomcdf(n=20, p=0.11385, lower value =3, upper value=20) for the TI-89 or TI-Nspire

  or

  binomcdf(n=20, p=0.11385, x=2) for other TI calculators
- A cumulative probability is found by summing all the probabilities up to a specific value. Here, the question requested $P(X \geq 3)$, which is the complement of the cumulative distribution function $1 - P(X \leq 2)$.
- Emphasize the meaning of a cumulative probability distribution function. That is, $P(X \leq 2) = \sum_{x=0}^{2} p(x)$.
- Spend time on discrete probability distributions. When finding a cumulative probability, the endpoint for the calculations is very important, e.g. $1 - P(X \leq 2) = P(X \geq 3)$.
- When $np < 5$, the normal approximation to the binomial distribution may not be very accurate. Have students work through examples for which normal approximations are not sufficiently accurate.
- Encourage students to reread the question after responding to determine if the given answer makes sense. When the two segments of the population have 11% and 22% left-handed children, a conclusion that 33% of the children in the population are left-handed is not realistic.
- The formulas needed in this question are given in the test booklet. Students who copied the formulas into their responses generally did better than students who did not provide a formula. It is good practice to give the formula that is being used in the calculations.
- Insist that students carefully label branches of tree diagrams used to compute probabilities, and explicitly indicate which branches are used to calculate a desired probability.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- In general, review of previous exam questions and chief reader reports will give teachers excellent insight into what constitutes strong statistical reasoning, as well as common student errors and how to address them in the classroom.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. To locate resources for teaching about probability, for example, try searching the community for “probability” and filtering for “Resource Library.” You may find a link to a TED Talk about probability and numerous activities applying probability and statistics, among other resources.

Question # 4 Max. Points: 4 Mean Score: 1.37

What were the responses to this question expected to demonstrate?

The primary goals of this question were to assess a student's ability to (1) determine whether a cause-and-effect conclusion can be made based on how a study was conducted and (2) set up, perform, and interpret the results of a hypothesis test, in the context of the problem.

How well did the responses address the course content related to this question? How well did the responses integrate the skills required on this question?

- Many responses demonstrated a clear understanding of randomized experiments.
- Responses generally demonstrated an understanding that there are steps involved to conduct a significance test.
- Reponses generally recognized that parameters need to be defined.
- Responses generally recognized that hypotheses are required.
- Responses generally recognized that a test statistic must be calculated.
- Responses demonstrated an understanding that conclusions to an inference test are based on the size of the p-value.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
Justifying the causal relationship based on the results of the test, ignoring the design of the study.	We can conclude that the new procedure causes a reduction in recovery time for those patients similar to the patients in this study because the patients were randomly assigned to the two procedures.
Not demonstrating an understanding that the random assignment of treatments allows for the conclusion of a causal relationship as opposed to other aspects of experiment design.	A well-designed experiment has random assignment of treatments, a control group and replication. The random assignment of treatments balances the effects of uncontrolled variables across both groups, enabling researchers to conclude that the new procedure causes a reduction in recovery time for those patients similar to the patients in this study. Other aspects of a well-designed experiment, such as replication and control, are important but do not enable researchers to make a causal conclusion.
Not demonstrating an understanding that random selection allows for generalization to a larger population, while random assignment allows causal conclusions to be made.	In this study, it is not clear whether patients were randomly selected or were volunteers. Regardless of the selection process, because the patients were randomly assigned to either the new procedure or standard procedure, researchers can conclude that the new procedure causes a reduction in recovery time for patients similar to the patients that participated in this study.
Indicating that a causal conclusion cannot be made because the experiment was not replicated.	Even though this is just one study, researchers can conclude that the new procedure causes a reduction in recovery time for those patients similar to the patients in this study because patients were randomly assigned to the new and standard procedures.

Incorrectly stating that the experiment used blocking or a placebo.	Researchers can conclude that the new procedure causes a reduction in recovery time for those patients similar to the patients in this study because patients were randomly assigned to the new and standard procedures.
Stating that there is a causal relationship without providing context for this study.	Researchers can conclude that the new procedure causes a reduction in recovery time for those patients similar to the patients in this study because patients were randomly assigned to the new and standard procedures.
Identifying the correct test as a two-sample *z*-test instead of a two-sample *t*-test.	The correct test for this study is a two-sample *t*-test for a difference in means.
Defining the parameters in terms of the patients in this particular study by using past tense, e.g. $\mu_N =$ the average recovery time for patients who received the new procedure.	μ_N represents the mean recovery time among all patients similar to those in the study if they were to receive the new treatment μ_S represents the mean recovery time among all patients similar to those in the study if they were to receive the standard treatment
Presenting inconsistencies with the direction of the hypotheses, the test statistic, and the *p*-value.	$H_0: \mu_N = \mu_S$ $H_a: \mu_N < \mu_S$ Test statistic = 7.13, and the *p*-value is approximately zero.
Not identifying the test statistic.	Two sample *t*-test, $t = 7.13$
Labeling parameters μ_1, μ_2 without specifying what the subscripts 1 and 2 represent.	μ_1 represents the mean recovery time among all patients similar to those in the study if they were to receive the new treatment μ_2 represents the mean recovery time among all patients similar to those in the study if they were to receive the standard treatment
Using a χ^2 test or a one-sample *t*-test for dependent samples.	The correct test is a two-sample *t*-test.
Identifying a two-sample *t*-test as the correct test but then using σ in the formula instead of s, or using an incorrect formula.	Two-sample *t*-test is $t = \frac{(\bar{x}_N - \bar{x}_S) - 0}{\sqrt{\frac{s_N{}^2}{n_N} + \frac{s_S{}^2}{n_S}}}$

Reversing the direction of the hypotheses.	$H_0: \mu_N = \mu_S$ $H_a: \mu_N < \mu_S$
Using sample means in the statements of hypotheses instead of population means.	$H_0: \mu_N = \mu_S$ $H_a: \mu_N < \mu_S$
Using calculator commands instead of probability statements for describing the *p*-value. Failing to label or identify the *p*-value. Providing a *p*-value greater than 1.	*p*-value = $P(t > 7.13) \approx 0$
Comparing the means of the two groups without considering the variability.	$t = \dfrac{(\bar{x}_N - \bar{x}_S) - 0}{\sqrt{\dfrac{s_N^{\ 2}}{n_N} + \dfrac{s_S^{\ 2}}{n_S}}} = -7.13$
Using the *p*-value to draw a conclusion but drawing an incorrect conclusion.	Because the *p*-value is small, there is sufficient evidence to conclude that for patients similar to the ones in the study, those receiving the new procedure would have less recovery time, on average, than those receiving the standard procedure.
Providing in incorrect interpretation of the *p*-value.	Obtaining a test statistic at least this extreme is unlikely due to chance alone therefore there is sufficient evidence to conclude that for patients similar to the ones in the study, those receiving the new procedure would have less recovery time, on average, than those receiving the standard procedure.
Claiming that the conclusion is to "accept the null hypothesis" or "prove the alternative hypothesis."	Because the *p*-value is small, there is sufficient evidence to conclude that for patients similar to the ones in the study, those receiving the new procedure would have less recovery time, on average, than those receiving the standard procedure.
Not using the *p*-value to justify the conclusion.	Because the *p*-value is small, there is sufficient evidence to conclude that for patients similar to the ones in the study, those receiving the new procedure would have less recovery time, on average, than those receiving the standard procedure.

Based on your experience at the AP® Reading with student responses, what advice would you offer to teachers to help them improve the student performance on the exam?

- Help students understand the difference between "random selection" and "random assignment." Random selection is when subjects are randomly selected from the population. It allows the results of the study to be generalized to the population from which the sample was selected. Random assignment is when subjects are randomly assigned to treatments in an experiment. This balances the effects of uncontrolled variables across both groups enabling researchers to conclude that statistically significant differences in the response are caused by the differences in the treatments. This causal relationship can be generalized only to other members of the population who are similar to the subjects in the experiment.

- Help students understand that important elements of a well-designed experiment are the random assignment of treatments to subjects, a comparison group and replication. However, it is the random assignment of treatments to subjects that enables researchers to conclude that statistically significant differences in the response are caused by, or can be attributed to, the differences in the treatments.

- Help students understand that while it is good to replicate experiments, often in practice, this is impractical. A causal relationship can be concluded based on a single experiment.

- Encourage students to always write in the context of the study presented in the problem.

- When students define parameters, encourage them give complete definitions and indicate that parameters are attributes of a population, not a sample.

- Encourage students to identify the test or procedure by name. If they also provide a formula, encourage them to understand the correct formula and to use the correct notation in that formula.

- Help students understand why they cannot "accept" the null hypothesis or "prove" the alternative hypothesis. A statistically significant result provides evidence to support the alternative hypothesis. The conclusion should be written in terms of the evidence it provides, that is, in terms of the alternative hypothesis.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- In general, review of previous exam questions and chief reader reports will give teachers excellent insight into what constitutes strong statistical reasoning, as well as common student errors and how to address them in the classroom.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. To locate resources for teaching about experimental design, for example, try searching the community for the relevant search words "random selection causal" and filtering for "Resource Library." You may find a link to an AP approved activity about an experiment involving beetles, a handout from an AP Annual Conference talk on teaching inference, and an activity to review random sampling and allocation, among other resources.

Question #5 **Max. Points:** 4 **Mean Score:** 0.98

What were the responses to this question expected to demonstrate?

The primary goals of this question were to assess a student's ability to (1) determine which of two histograms represents data with a larger median; (2) calculate the mean of a combined data set when the separate means and sample sizes are known; and (3) calculate the probability that an individual randomly chosen from a finite population will have a value within one standard deviation of the mean, when provided with values for the mean, standard deviation, and all members of the population.

How well did the response address the course content related to this question? How well did the responses integrate the skills required on this question?

- Responses generally addressed the shape of one or both distributions.
- Most responses reported the location of the median for one or both distributions.
- Many responses used correct weights in computing the mean of the combined data set.
- Many responses recognized that the normal distribution may not be used when the distribution is skewed.
- Responses recognized that computing a probability for a randomly selected observation from a finite set of observations is the number of data values that satisfy the event divided by the total number of data values in the set.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
Selecting the incorrect median for each distribution.	The median for High School A is 7 and the median for High School B is 6.
Not demonstrating understanding of how to find the median.	The median for High School A is the between the 100^{th} and 101^{st} data value in the ordered data set and the median for High School B is the 111^{th} data value in the ordered data set.
Describing the shape of the distribution for High School B as skewed left.	The distribution of teaching years for High School B is strongly skewed right and the distribution of teaching years for High School A is less skewed than the distribution for High School B.
Making the decision about the location of a median based on a single bin of the histogram.	The distribution of teaching years for High School B is strongly skewed right, so the median of High School B will be 6 and the median for High School A will be 7.
Not addressing the mean-median gap in a skewed distribution.	Since school B is more strongly skewed than High School A, there is a greater difference between the mean and median in school B.

Describing less than half the data values as a majority.	For High School B, 79 data values fall in the bin 1 to 4.
Not using a weighted sum when calculating the mean of a combined data set with unequal sample sizes.	$\frac{(200)(8.2)+(18)(2.5)}{200+18}=7.73$ years
Using incorrect weights when computing a weighted sum.	$\frac{(200)(8.2)+(18)(2.5)}{200+18}=7.73$
Not using the given mean of 8.2 and estimating it based on frequencies in the histogram.	$\frac{(200)(8.2)+(18)(2.5)}{200+18}=7.73$
Using the median of 7 instead of the correct mean of 8.2 in computing the weighted sum.	$\frac{(200)(8.2)+(18)(2.5)}{200+18}=7.73$
Applying the Empirical Rule to a strongly skewed distribution to estimate a probability.	The sum of the frequencies of the histogram's bins with boundaries included in the interval equals 189. The sum is divided by 221 and the correct probability (0.8552) is reported.
Computing the probabilty based on the normal distribtution when given a strongly skewed distribution.	The sum of the frequencies of the histogram's bins with boundaries included in the interval equals 189. The sum is divided by 221 and the correct probability (0.8552) is reported.
Not adjusting the bin frequency to account for the fact that the data values are integers.	Because the data values are recorded as integers, if a value is within one standard deviation of the mean the value will be in the interval $[1,15]$, or equivalently $[1,16)$.

Based on your experience at the AP® Reading with student responses, what advice would you offer to teachers to help them improve the student performance on the exam?

- To help students avoid the mistake of describing a right skewed distribution as left skewed, provide students with exercises involving both right skewed and left skewed distributions.
- When answering questions involving a statistical measure of center, have students clearly communicate their understanding of the definition of the measure.
- Remind students that the location of the median is the ordered observation at location $\frac{(\text{sample size})+1}{2}$, not $\frac{(\text{sample size})}{2}$.
- Discuss how making a decision based on a small sample from a population differs from making a decision based on information about the values of all members of the population.
- Illustrate that the difference (population mean minus population median) is positive for right shewed distributions and negative for left skewed distributions.
- Develop exercises to help students determine correct weights for computing the mean of a combined data set.
- Develop exercises that encourage students to read carefully and use information given in the stem of the problem.
- Develop exercises to give student more practice with recognizing when to use the empirical distribution to compute probabilities for a population with a discrete set of possible values and when to use a normal distribution.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- In general, review of previous exam questions and chief reader reports will give teachers excellent insight into what constitutes strong statistical reasoning, as well as common student errors and how to address them in the classroom.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. To locate resources for teaching about describing data distributions, for example, try searching the community for "measures of center and skew" and filtering for "Resource Library." You may find a series of data exploration activities and a link to rich source of data at the National Center for Education Statistics, among other resources.

Question 6 **Max. Points: 4** **Mean Score: 0.34**

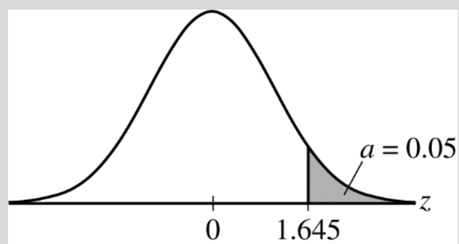
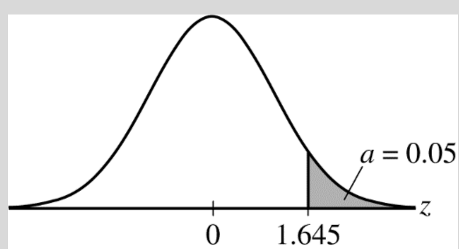
What were the responses to this question expected to demonstrate?

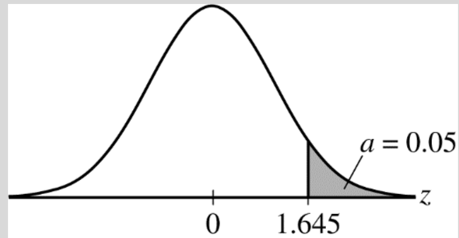
The primary goals of this question were to assess a student's ability to (1) describe what constitutes a Type II error for a specific hypothesis test; (2) specify a rejection region in terms of values of the sample mean; (3) compute the power of a test for a specific value in the alternative hypothesis; (4) recognize the definition of power; and (5) understand the impact of increasing the sample size on the power of a test.

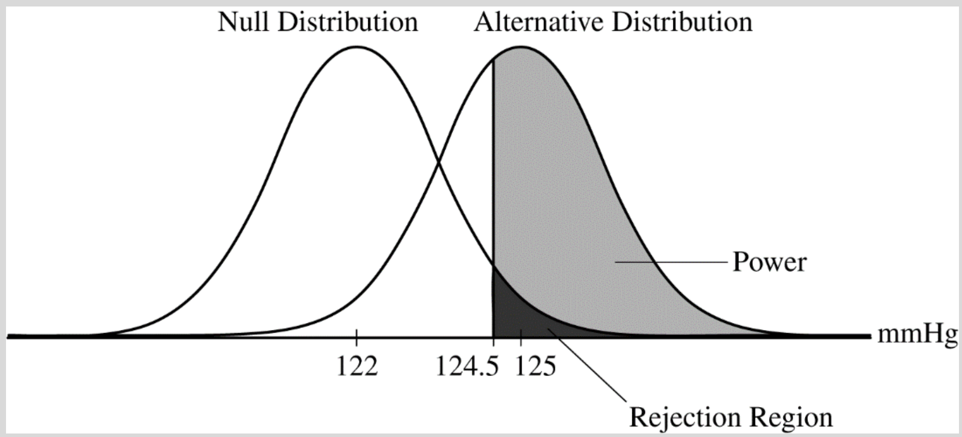
How well did the response address the course content related to this question? How well did the responses integrate the skills required on this question?

- Responses tended to describe Type II error in the context of the problem.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

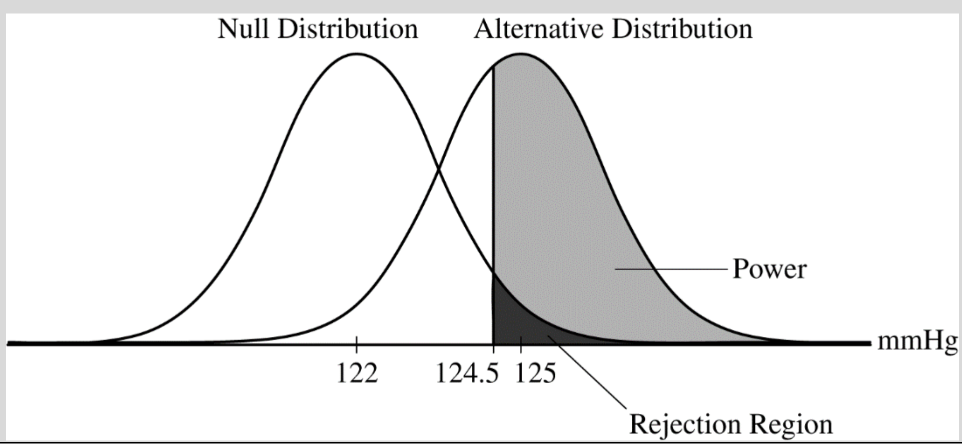
Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
Confusing Type I error with Type II error.	A Type II error in this context would be that the population mean systolic blood pressure of the employees at the corporation is greater than 122 mmHg but the hypothesis test conducted by the investigators fails to reject that the population mean systolic blood pressure is equal to 122 mmHg.
Not recognizing that when population standard deviation is known the standard normal distribution may be used as opposed to a t-distribution.	 $1.645 = \frac{\bar{x} - 122}{1.5}$, so the values of the sample mean where the null hypothesis would be rejected are $\bar{x} \geq 124.4675$.
Not recognizing the standard deviation for the sampling distribution of the sample mean is $\frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1.5$.	 $1.645 = \frac{\bar{x} - 122}{1.5}$, so the values of the sample mean where the null hypothesis would be rejected are $\bar{x} \geq 124.4675$.

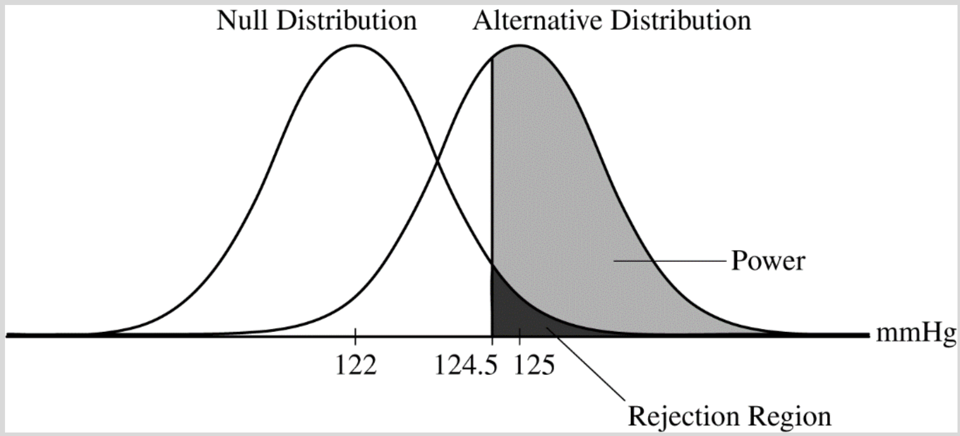
<table>
<tr><td>Using the lower tail region rather than the upper tail region to compute the probability that the null hypothesis will be rejected. This is synonymous with incorrectly using $\mu < 122$ as thealternative hypothesis.</td><td>

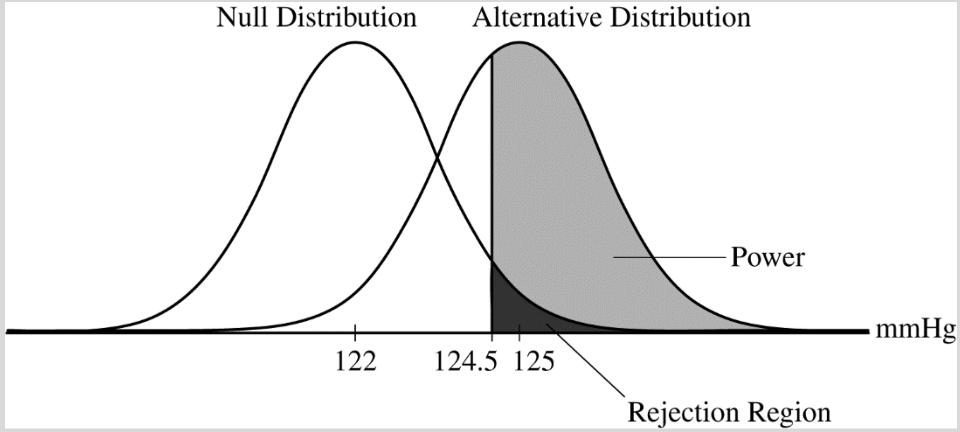
$1.645 = \frac{\bar{x} - 122}{1.5}$, so the values of the sample mean where the null hypothesis would be rejected are $\bar{x} \geq 124.4675$.</td></tr>
<tr><td>Not recognizing that the probability to reject the null hypothesis when the population mean is actually 125 is the area greater than the value found in part (b).</td><td>

$\bar{X} \sim \text{Normal}\left(\mu = 125, \frac{\sigma}{\sqrt{n}} = 1.5\right)$

$P(\bar{X} \geq 124.4675 \mid \mu = 125) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{124.4675 - 125}{1.5}\right)$

$= P(z \geq -0.355) = 0.6387$</td></tr>
<tr><td>Using 125 as a sample mean rather than the new population mean.</td><td></td></tr>
</table>

	$\bar{X} \sim \text{Normal}\left(\mu = 125, \frac{\sigma}{\sqrt{n}} = 1.5\right)$ $P(\bar{X} \geq 124.4675 \mid \mu = 125) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{124.4675 - 125}{1.5}\right)$ $= P(z \geq -0.355) = 0.6387$
Using 15 as the standard deviation rather than 1.5, the standard deviation of the sampling distribution for the sample mean.	 $\bar{X} \sim \text{Normal}\left(\mu = 125, \frac{\sigma}{\sqrt{n}} = 1.5\right)$ $P(\bar{X} \geq 124.4675 \mid \mu = 125) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{124.4675 - 125}{1.5}\right)$ $= P(z \geq -0.355) = 0.6387$

Not stating which distribution was used for calculations.	 $\bar{X} \sim \text{Normal}\left(\mu = 125, \frac{\sigma}{\sqrt{n}} = 1.5\right)$ $P(\bar{X} \geq 124.4675 \mid \mu = 125) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{124.4675 - 125}{1.5}\right)$ $= P(z \geq -0.355) = 0.6387$
Not identifying the statistical term for the probability found in part (c) as power.	The statistical term for the probability that the null hypothesis is rejected when the population mean is 125 mmHg is power.
Not clearly recognizing that power increases as sample size increases because the standard deviation of the sampling distribution of the sample mean decreases.	As sample size increases the power will increase. The standard deviation of the sampling distribution will decrease and therefore the minimum value of the sample mean that will result in rejecting the null hypothesis will decrease. As a result, the probability of rejecting the null hypothesis that the population mean is 122 mmHg when the population mean is actually 125 mmHg will increase.
Responses did not recognize that when the standard deviation of the sampling distribution of the sample mean decreases the value found in part (b) will also decrease.	As sample size increases the power will increase. The standard deviation of the sampling distribution will decrease and therefore the minimum value of the sample mean that will result in rejecting the null hypothesis will decrease. As a result, the probabilbiyt of rejecting the null hypothesis that the population mean is 122 mmHg when the population mean is actually 125 mmHg will increase.

Based on your experience at the AP® Reading with student responses, what advice would you offer to teachers to help them improve the student performance on the exam?

- Develop exercises to help students better understand the difference between Type I error with Type II error.
- Develop exercises to help student realize that the when population standard deviation is known the standard normal distribution may be used as opposed to a *t*-distribution.
- Emphasize the difference between a population distribution and a sampling distribution to help students recognize when the population standard deviation σ should be used and when the standard deviation of the sampling distribution $\frac{\sigma}{\sqrt{n}}$ should be used.
- Responses did not recognize that the probability to reject the null hypothesis when the population mean is actually 125 is the area greater than the value found in part (b). Encourage students to write a probability statement when asked to calculate a probability that designates what probability they are calculating. For example, $P(\bar{X} \geq 124.4675)$.
- Responses used 125 as a sample mean rather than the new population mean. Stress the difference between a population mean and a sample mean.
- Encourage students to clearly define what distribution they are using for probability calculations.
- Responses did not indicate that power increases as sample size increases because the standard deviation of the sampling distribution of the sample mean decreases. Encourage students to show full work when performing calculations to enable them to more clearly recognize the effect of changing a variable.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- In general, review of previous exam questions and chief reader reports will give teachers excellent insight into what constitutes strong statistical reasoning, as well as common student errors and how to address them in the classroom.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. To locate resources for teaching about power, for example, try searching the community for “power” and filtering for “Resource Library.” You may find a simulation to introduce power, a quiz, and a significance test vocabulary review, among other resources.