AP® CollegeBoard

Chief Reader Report on Student Responses:

2017 AP® Statistics Free-Response Questions

- Number of Students Scored: 215,840
- Number of Readers: 859
- Score Distribution:

Exam Score	N	%At
5	29,350	13.6
4	34,386	15.9
3	53,488	24.8
2	43,576	20.2
1	55,040	25.5

- Global Mean: 2.72

The following comments on the 2017 free-response questions for AP® Statistics were written by the Chief Reader Jessica Utts, University of California, Irvine. 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:** 1.72

What were responses expected to demonstrate in their response to this question?

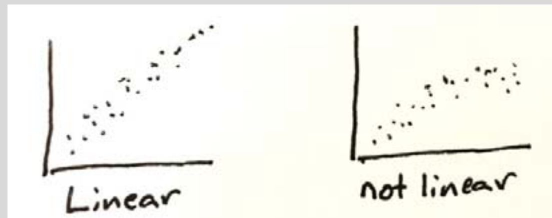
The primary goals of this question were to assess a student's ability to (1) explain statistical terms used when describing the relationship between two variables; (2) interpret the slope of a linear regression equation; and (3) calculate a value of y when given a regression equation, a value of x, and a residual.

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?

- In part (a), responses did an excellent job of explaining what is meant by a positive relationship and including the context. Responses did a moderate job of explaining a strong relationship. Many had trouble explaining what is meant by a linear relationship without using circular reasoning words like "line" and "linear."
- Part (b) required responses to apply a straight-forward definition of slope to a particular context. They were able to do so, but some neglected to explain that the relationship is not exact.
- Part (c) required use of standard formulas for predicted values and residuals in a non-standard way, and some responses only were able to partially complete the exercise by finding the predicted value, but neglected to carry out the final step of using the residual.

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

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Defining a positive relationship by simply saying that there is a positive correlation.	• A <u>positive</u> relationship means that wolves with higher values of length also tend to have higher weights.
• In defining a linear relationship responses failed to link a change in y to a change in x.	• A <u>linear</u> relationship means that when length increases by one meter, weight tends to change by a constant amount, on average.
• Not clearly indicating that a linear relationship has a constant <u>rate</u> of change in the response variable (weight) as the explanatory variable (length) inceases.	• For any change in length the rate of change in weight is the same.
• Using "correlation" to define a linear relationship. In most cases it was not clear if "correlation" was used in statistical manner or if was merely used as substitute for "relationship." A correlation coefficient is a measure of the strength of a linear relationship, but it does not by itself explain the meaning of a linear relationship. It is more appropriate to use correlation to discuss a strong relationship.	• A <u>linear</u> relationship means that when length increases by one meter, weight tends to change by a constant amount, on average. A <u>strong</u> relationship means that there is a high correlation (close to 1) between length and weight.

• Indicating that a relationship is strong when points in the scatterplot are close together or not too scattered. The response should indicate that a relationship is strong when the points in the scatterplot are close to a line, or more generally, a curve.	• A strong relationship means that the points are close to the least squares regression line.
• A graph presented without some explanation may not be provide enough information for the terms in part (a). Including graphs in explanations, when relevant, often helps to strengthen and clarify a response. However, in the explanation of positive, linear, or strong, a single graph was generally not acceptable if it was not accompanied by some written communication or a second graph. The score for a graph-based response with no written communication, or no useful written communication, was scored as acceptable only if it had a pair of graphs with one illustrating the attribute and the other illustrating what the attribute is not.	• The following graphs would illustrate what is meant by "linear" by offering a comparison of linear and not linear. 
• Implying that the slope of a least squares regression line corresponds to an exact relationship between changes in observed values of y as x changes.	Examples of acceptable responses are: • The predicted weight increases by 35.02 kg for each 1-meter increase in length. • Weight increases by 35.02 kg for each 1-meter increase in length, on average.
• Failure to link the increase in the predicted response to an increase of a specific size in the explanatory variable. For instance, an unacceptable response is "For any change in length, the predicted weight increases by 35.02 kg."	• The predicted weight increases by 35.02 kg for each 1-meter increase in length.
• In calculating the actual weight from the regression equation for a specific x when a residual is given, many responses stopped after computing the predicted value.	• Predicted weight $= -16.46 + 35.02(1.4) = 32.568$ kg • Actual weight $= 32.568 + (-9.67) = 22.9$ kg
• Some responses incorrectly replaced the residual with the intercept in the calculation of the actual response.	• Predicted weight $= -16.46 + 35.02(1.4) = 32.568$ kg • Actual weight $= 32.568 + (-9.67) = 22.9$ kg

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?

- Some students found it easier to explain the concepts of positive, linear, and strong relationships by formulating responses in the framework of points on the scatterplot described in part (a) instead of using abstract explanations. Responses tended to be less precise when students tried to give more abstract explanations that were based on potential patterns of points in scatter plots. Responses to part (a) also tended to be more off-target when students based them on regression models. There was no mention of any model in part (a), only a description of a scatterplot. Perhaps students were motivated to think about regression models when they read parts (b) and (c) of the question which were based on a regression equation. While it is good practice for students to read the entire question before they begin to any answer parts of the question, they should focus on answering each individual part without pulling in ideas from later parts of the question.
- An explanation of a linear relationship should indicate that is the <u>rate</u> of change in y is the same for any change in x. Examples of inappropriate statements are:
    - Change in y is the same for any change in x. (This implies that a 1-unit increase in x is associated with the same change in y as a 2-unit increase in x.)
    - For any x, the change in y is the same. (This does not relate a change in y to a change in x.)
    - x and y change at the same rate. (This implies a line with slope equal to 1, which is to restricted and does not define all lines.)
    - y is directly proportional to x. This is too restrictive because it implies that y must be 0 when x is 0. (A square is a rectangle but the definition of a square is not a good definition of a rectangle.) Students should avoid the use of "proportional," "directly proportional," and "constant ratio" unless they are comparing rates of change for two variables.
- Use of the word *correlation*. It is often difficult to determine if the response used correlation as a substitute for *relationship* in the English (instead of statistical) meaning of the word. It is better to use *correlation coefficient* to make it clear that correlation is being used in a statistical sense.
- "There is a positive correlation" does not provide an <u>explanation</u> of a positive relationship, nor does it provide an explanation of a linear relationship. A correlation coefficient is a measure of the strength of a linear relationship.
- In using the concept of a correlation coefficient to describe that a linear relationship is strong, it is good practice to provide a range of numerical values to quantify what is strong; for example, a value of a correlation coefficient between 0.7 and 1. Using numerical values also clarifies that correlation is being used in a statistical context.
- An interpretation of a slope of a regression line should relate a specific change in the predicted response to a specific change in the explanatory variable. A correct interpretation is "for each 1-meter increase in length, the weight of wolves is predicted to increase by 35.02 kg." An incorrect interpretation is "for any increase in length, the weight of wolves is predicted to increase by 35.02 kg."
- If asked for a value of a response in a regression problem, use the formula for the least squares regression line to compute the predicted response. If the question gives a value of a residual, use it to compute the actual value of the response from the predicted value.

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 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. Previously released exam questions and participation in the Online Teacher Community are of high value for practice and for advice.

Question #2 **Max. Points:** 4 **Mean Score:** 2.22

What were responses expected to demonstrate in their response to this question?

The primary goals of this question were to assess a student's ability to (1) construct and interpret a confidence interval for a population proportion and (2) use a confidence interval for a proportion to find a confidence interval for a dollar amount that can be calculated using that proportion.

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?

- Part (a) is a standard inference question; to construct and interpret a confidence interval. It addresses specific course content and skills, and responses generally demonstrate good use of those skills, with some common errors.
- Part (b) required an extension of standard content to demonstrate statistical reasoning skills. Responses generally did a nice job of demonstrating this skill, with the most common errors related to not reading the question carefully.

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

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Identification of procedure missing or incorrect.	• The appropriate procedure is a one-sample z-interval for a population proportion.
One or more of the following errors related to the conditions: • Not checking conditions at all. • Omitting the large sample condition ($n\hat{p} \geq 10$ and $n\hat{q} \geq 10$) or verifying only one of the two inequalities. • Mislabeling conditions, such as "Independence" for $n\hat{p} \geq 10$ and $n\hat{q} \geq 10$. • Stating as a condition that the sample or population has a normal distribution (for a categorical variable), or vague reference to a normal distribution. • Inappropriate large sample condition: $n \geq 30$.	• Correct conditions: 1. Random sample 2. Large sample (number of successes $n\hat{p} \geq 10$ and number of failures $n(1-\hat{p}) \geq 10$) • For condition 1, the stem of the problem states that a random sample of customers who asked for a water cup was used. • For condition 2, the number of "successes" (filled cup with soft drink) is 23 and the number of "failures" is 57, both of which are greater than 10.

• Errors in mechanics: • Using incorrect critical value (wrong z^* or a t^*). • Showing df; using t instead of z. • Using an incorrect formula for the standard error of a sample proportion.	• The correct confidence interval is: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, which is $0.2875 \pm 1.96\sqrt{\frac{0.2875(1 - 0.2875)}{80}}$ The interval is 0.1883 to 0.3867.
• Calculating an unreasonable interval for a proportion—not realizing that a proportion must be between 0 and 1.	• The interval is 0.1883 to 0.3867.
• Not recognizing that the population of interest is all customers who asked for a water cup at this fast-food restaurant.	• When defining the parameter, state that "The population is all customers of this restaurant who ask for a water cup, and p is the proportion of that population that will fill the cup with a soft drink."
• Interpreting confidence level instead of confidence interval.	• We can be 95% confident that in the population of all customers at this fast-food restaurant who ask for a water cup, the proportion that will fill it with a soft drink is between 0.1883 and 0.3867.
Errors in part (b): • Calculating a single value (point estimate) rather than an interval. • Not using the interval from part (a) as directed. • Not showing work.	• Using the confidence interval in Part (a), a 95% interval estimate for the number of customers in June who asked for a water cup but then filled it with a soft drink is 3,000 × 0.1883 to 3,000 × 0.3867, or 565 to 1,160. At a cost of $0.25 per customer, a 95% interval estimate for the cost to the restaurant in June is $141.25 to $290.00.

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?

- In inference questions, ask students to identify the population and parameter of interest. Encourage students to use the language in the stem of the question when defining the parameter.
- Discuss <u>why</u> each condition is being checked for an inference procedure and help students understand how to check the condition. Use applets and hands-on activities to demonstrate what happens when each condition isn't met.
- Insist on proper notation throughout the course and refer students to the formula sheet.
- Emphasize the difference between interpreting a confidence <u>interval</u> and a confidence <u>level</u>. Use hands-on activities and applets involving repeated sampling to illustrate the idea of a confidence level interpretation.

- Give students ample practice in distinguishing categorical variables/data (proportions) from quantitative variables/data (means).
- Have students make a summary chart of inference procedures with the appropriate names, conditions, and formulas for each.

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 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. Previously released exam questions and participation in the Online Teacher Community are of high value for practice and for advice.

Question #3 **Max. Points:** 4 **Mean Score:** 1.72

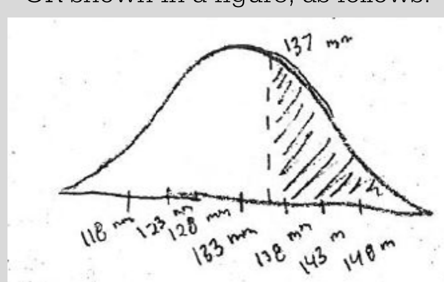
What were responses expected to demonstrate in their response to this question?

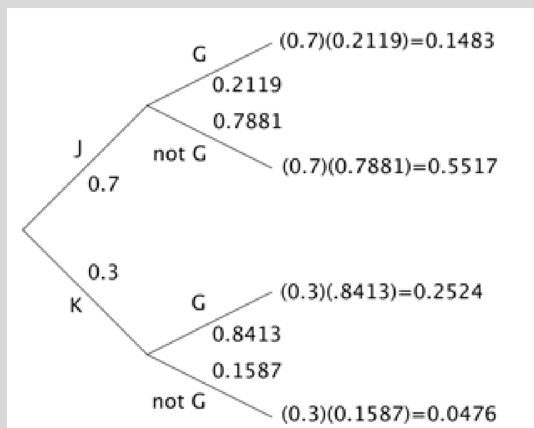
The primary goals of this question were to assess a student's ability to (1) calculate a probability from a normal distribution; (2) calculate a weighted probability from two individual probabilities; and (3) calculate a conditional probability for dependent events when individual and joint probabilities are provided.

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 were able to calculate the normal probability, but made some errors in notation and in showing work.
- Many responses had trouble calculating the weighted probability because they confused independent and dependent events, and/or probability rules for independent events.
- Many responses confused conditional probabilities with joint probabilities.

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

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Use of the t-distribution to solve a normal probability.	• $P(X > 137) = P\left(Z > \frac{137 - 133}{5}\right) = P(Z > 0.8) = 0.2119$
• Use of $\bar{x}$, as in $P(\bar{x} > 137)$, to solve a normal probability.	• $P(X > 137)$
• Use of calculator-speak normalcdf(137, 1000000, 133, 5) withouth identifying parameters and boundary conditions.	• For notation normalcdf(137, 1000000, 133, 5), 137 is the lower bound, 1,000,000 is the upper bound, 133 is μ and 5 is σ
• Generally, not knowing how much work is needed to justify normal probability calculations.	• $P(X > 137) = P\left(Z > \frac{137 - 133}{5}\right) = P(Z > 0.8) = 0.2119$ • OR shown in a figure, as follows: 

• Not knowing the difference between parameters and statistics.	• $P\left(Z > \frac{137 - \mu}{\sigma}\right)$, not $P\left(Z > \frac{137 - \overline{x}}{s}\right)$
• Thinking the Empirical Rule can yield normal probabilities for values other than 1, 2, and 3 standard deviations on either side of the mean.	• $P(X > 137) = P\left(Z > \frac{137 - 133}{5}\right)$ $= P(Z > 0.8)$ $= 0.2119$
• Thinking that 137 can be adjusted as if the normal random variable is discrete and using 137.5 or 138 in place of 137.	• $P(X > 137) = P\left(Z > \frac{137 - 133}{5}\right)$ $= P(Z > 0.8)$ $= 0.2119$
• Not recognizing that two events are mutually exclusive.	• $P(G) = P(G \text{ and } J) + P(G \text{ and } K)$ $= 0.1483 + 0.2524$ $= 0.4007$
• Generally not knowing how much work is needed to justify probability calculations.	• $P(G) = P(G \mid J) \times P(J) + P(G \mid K) \times P(K)$ $= (0.2119)(0.7) + (0.8413)(0.3)$ $= 0.1483 + 0.2524 = 0.4007$
• Not being able to properly create a tree diagram.	• 
• Not being able to find the appropriate probabilities from a tree diagram. For example, thinking the conditional branch on the tree is actually the intersection probability.	• $P(G) = P(G \mid J) \times P(J) + P(G \mid K) \times P(K)$ $= (0.2119)(0.7) + (0.8413)(0.3)$ $= 0.1483 + 0.2524 = 0.4007$
• Not being able to correctly use the probability of a given event in computing a conditional probability, omitting that probability, or switching it with a different probability.	• $P(G) = P(G \mid J) \times P(J) + P(G \mid K) \times P(K)$ $= (0.2119)(0.7) + (0.8413)(0.3)$ $= 0.1483 + 0.2524 = 0.4007$

• Not being able to distinguish between a question that is asking for a conditional probability and one that is asking for an intersection probability; such as, ignoring the "given" in a question.	• $P(J \mid G) = \frac{P(J \text{ and } G)}{P(G)} = \frac{P(G \mid J)P(J)}{P(G)} = \frac{(0.2119)(0.7)}{0.4007} = \frac{0.1483}{0.4007} = 0.3701$
• Assuming two events are independent when they are not.	• $P(J \mid G) = \frac{P(J \text{ and } G)}{P(G)} = \frac{P(G \mid J)P(J)}{P(G)} = \frac{(0.2119)(0.7)}{0.4007} = \frac{0.1483}{0.4007} = 0.3701$
• Not being able to recognize that a calculation from earlier work could be used in subsequent calculations.	• $\frac{(0.7)(\text{answer from part (a)})}{(\text{answer from part (b)})}$

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?

- In solving problems, model what the students should do by showing all work/steps in a probability problem (and inference).
- When teaching the Empirical Rule, be sure to relate calculated normal probabilities to the rule. Emphasize that the Empirical Rule gives only approximations of normal values and only for 1, 2, and 3 standard deviations from the mean; any sort of interpolation will give incorrect answers.
- When introducing the t-distribution, emphasize that the only time the z-distribution and t-distribution are the same is when the t-distribution is based on an infinite number of degrees of freedom, which is never going to happen. A corollary is that any normal probability problem should be solved using the z-distribution, not the t-distribution.
- In general, it is not a good idea to use "calculator speak" in answering any question.
- Show the students problems where there are multiple parts and the answers for the later parts depend upon the results for the earlier parts.
- If any continuous approximation of a discrete random variable is taught, explain the reason for any adjustment or continuity correction. Hopefully, this will decrease the probability that the student will attempt to use a correction for a continuous distribution.
- After introducing the sampling distribution for the sample mean, go through an example that starts with the assumption of normality, calculates a standard normal probability for a single value of X, then for a value of the sample mean. Explain that technically it is possible (under the assumption of a normal population) to calculate the probability for a single value by using the sample mean based on a sample of size one, but this approach involves more work and could lead to errors later when a sample is not taken from a normal population. Then continue the example where the assumption of a normal population is not valid (best if one sample size is below 30 and the other is above 30).
- Students have trouble with "independence" versus "mutually exclusive." Early on, give an example where there are three events, two are mutually exclusive, two are independent, and two are neither. For example, toss three coins and note whether each is a heads or tails. Let event A represent getting at least two heads, event B represent getting exactly two heads, and event C represent getting all heads or all tails. A and B are neither independent nor mutually exclusive, A and C are independent (not obviously, so it helps to reinforce that independence is based on probability, not appearances), and B and C are mutually exclusive.

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 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. Previously released exam questions and participation in the Online Teacher Community are of high value for practice and for advice.

Question #4 **Max. Points:** 4 **Mean Score:** 1.71

What were responses expected to demonstrate in their response to this question?

The primary goals of this question were to assess a student's ability to use boxplots to (1) compare multiple sets of data; (2) identify which set of data is most likely to have produced a particular summary value; and (3) determine which variable is most useful for classifying a new observation.

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?

- Responses generally indicated an understanding of what is displayed in a boxplot.
- Responses had problems with comparing boxplots to answer specific questions because they tended not to give an actual comparison. Responses understood how to choose which one was best for a given purpose, but did not specify why the others were not as good.

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

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Stating that symmetric boxplots indicate that the distribution is normal.	• Complete shape information cannot be determined from a boxplot.
• Students were asked to describe a similarity and a difference for chemical Z across the three sites. Some students used only a location, such as the maximums or the minimums.	• The median value for the percent of chemical Z in the pottery pieces is similar for all three sites, at about 7%. The ranges for the percent of chemical Z are much different for the three sites, with the smallest range being about 2% (from 6% to 8%) at site II, a much higher range of about 6% (from about 4% to 10%) at site I and the largest range of about 8% (from about 3% to 11%) at site III.
• Using the term range incorrectly – many responses referred to range as an interval instead of a number.	• The ranges for the percent of chemical Z are much different for the three sites, with the smallest range being about 2% (from 6% to 8%) at site II, a much higher range of about 6% (from about 4% to 10%) at site I and the largest range of about 8% (from about 3% to 11%) at site III.

• Some responses described many attributes of the boxplots for chemical Z at each site but never clearly stated what is similar and what is different. This is a “laundry list,” not a comparison.	• The median value for the percent of chemical Z in the pottery pieces is similar for all three sites, at about 7%. The ranges for the percent of chemical Z are much different for the three sites, with the smallest range being about 2% (from 6% to 8%) at site II, a much higher range of about 6% (from about 4% to 10%) at site I and the largest range of about 8% (from about 3% to 11%) at site III.
• For part (b-i), many responses selected site III based on the sums of the medians instead of the sums of the minimums and the sums of the maximums.	• The piece most likely originated at site III. Although values outside of the range of data observed in the samples would be possible, using the available data results in approximate minimum and maximum sums of the percents for the three chemicals, as shown in the table below. The only site that includes 20.5 between the sums of the minimum and maximum values is site III. [Response then includes a table showing the sum of minimum and maximums for each site.]
• Many responses correctly selected site III but did not state why sites I and II were not the best choices.	• The only site that includes 20.5 between the sums of the minimum and maximum values is site III.
• In part (b-ii), many responses had difficulty clearly explaining that the boxplots do not overlap. In order to do this, both a difference in location and small variability needed to be addressed. Some responses described the boxplots as having different means, different variability, or simply said that they have different boxplots. None of these descriptors indicate no overlap.	• Chemical Y would be most useful, because the distribution of the percentages of total weights at the three sites do not overlap. The distributions of chemicals X and Z have substantial overlap.

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?

- Read the question carefully and answer (only) the question asked. If the question asks for a similarity and a difference, clearly organize and state what is similar *and* what is different.
- When asked to identify one choice among three, complete justification includes reasoning for that particular choice, as well as rationales for not choosing the other options.
- Use clear communication within each part of the question. Do not assume that the reader will look back. Instead, incorporate previous work into the answer.

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 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. Previously released exam questions and participation in the Online Teacher Community are of high value for practice and for advice.

Question #5 **Max. Points:** 4 **Mean Score:** 1.51

What were responses expected to demonstrate in their response to this question?

The primary goal of this question was to assess a student's ability to identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question. More specific goals were to assess a student's ability to (1) state appropriate hypotheses; (2) identify the appropriate statistical test procedure and check appropriate conditions for inference; (3) calculate the appropriate test statistic and *p*-value; and (4) draw an appropriate conclusion, with justification, in the context of the study.

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?

- Responses did a good job of identifying hypotheses, stating a decision in terms of the alternative hypothesis, and making a conclusion in context.
- Many responses included mistakes in the details naming the test and carrying out the mechanics.

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>

- Responses incorrectly used the idea of sufficient evidence (given in the stem of the problem) to state their hypotheses:

  H_0 : There is sufficient evidence of no association.
  H_a : There is sufficient evidence of an association.

</td><td>

- H_0 : Age group at diagnosis and gender are independent (that is, they are not associated) for the population of people currently being treated for schizophrenia.
  H_a : Age group at diagnosis and gender are not independent for the population of people currently being treated for schizophrenia.

</td></tr>
<tr><td>

- Some responses had trouble identifying the correct conditions. Problems included:
  - Listing incorrect conditions such as n > 30, or "both samples independent."
  - Stating the condition that expected counts are > 5, but not verifying it by computing them.
  - Stating that the expected count condition is required for normality.

</td><td>

- The expected counts for all 8 cells of the table are at least 5, as seen in the following table, with expected counts shown below observed counts:

<table>
<tr><th></th><th>20 to 29</th><th>30 to 39</th><th>40 to 49</th><th>50 to 59</th><th>Total</th></tr>
<tr><td>Women</td><td>46
56.91</td><td>40
36.22</td><td>21
17.25</td><td>12
8.62</td><td>119</td></tr>
<tr><td>Men</td><td>53
42.09</td><td>23
26.78</td><td>9
12.75</td><td>3
6.38</td><td>88</td></tr>
</table>

</td></tr>
</table>

• Some responses showed incorrect work for the test statistics and/or the p-value.	• $\chi^2 = 2.093 + 0.395 + 0.817 + 1.322 + 2.830 + 0.534 + 1.105 + 1.788 = 10.884$ • The p-value is $P(\chi^2 \geq 10.884) = 0.012$, based on $(4-1)(2-1) = 3$ degrees of freedom.
• Some responses wrote statistical conclusions as definitive statements "we conclude …" or "we prove."	• Because the p-value is very small (for instance much smaller than $\alpha = 0.05$), we would reject the null hypothesis and conclude that the sample data provide strong evidence that there is an association between age group at diagnosis and gender for the population currently being treated for schizophrenia.
• Some responses did not use appropriate linkage between the p-value and a stated level of alpha in making a statistical decision.	• Because the p-value is very small (for instance much smaller than $\alpha = 0.05$), we would reject the null hypothesis…
• Some responses used the bar graph provided and stated conclusions about the sample data, but did not carry out any inference.	• $\chi^2 = 2.093 + 0.395 + 0.817 + 1.322 + 2.830 + 0.534 + 1.105 + 1.788 = 10.884$ • The p-value is $P(\chi^2 \geq 10.884) = 0.012$, based on $(4-1)(2-1) = 3$ degrees of freedom.

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 that the null hypothesis is about the population and not the sample. It may be helpful to point out that no association (independence) is synonymous with status quo/no change similar to a null hypothesis of $\mu_1 = \mu_2$.
- Remind students that it is important for them to name the test they are performing. Help students distinguish between chi-square test for independence and chi-square test for homogeneity. The distinction has to do with the sampling method, whether there was one sample taken or multiple samples taken.
- Help students understand not only the correct conditions needed for hypothesis tests, but also emphasize why those conditions are necessary. Remind students to well label their work and clearly communicate what they are doing.
- Emphasize to students how the values that they obtain from the calculator were calculated. Ideally, students should be familiar with how to calculate values by formula and calculator.
- Emphasize to students that statistical conclusions are not definitive and that there should be some level of uncertainty.
- Highlight the importance of good communication; students should state why they are making the decisions that they are making.

- Emphasize to students the difference between evidence from a sample (which can be obtained from the graph) and convincing statistical evidence (which can be obtained from a hypothesis test).

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 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. Previously released exam questions and participation in the Online Teacher Community are of high value for practice and for advice.

Question #6 **Max. Points:** 4 **Mean Score:** 0.99

What were responses expected to demonstrate in their response to this question?

The primary goals of this question were to assess a student's ability to (1) calculate probabilities associated with treatment and control group memberships for two different methods of random assignment and (2) justify which random assignment method is more appropriate in a given situation.

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?

- Responses generally did a good job of calculating probabilities for the two methods, but did not show sufficient work to know how they were found.
- Responses generally were able to identify the preferred randomization method in the context of the problem, but did not communicate well about the reason why it would be preferred in this situation.

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

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Although many responses listed the correct probabilities in the table in (i) of parts (a) and (b), the justification was often missing or incomplete.	• $P(\text{Arrangement A}) = P(TT) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$
• In (ii) of parts (a) and (b), responses failed to include BOTH Arrangements A and D when asked for the probability that both men end up in the same treatment group.	• Part (a-ii): $P(\text{A}) + P(\text{D}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ • Part (b-ii): $P(\text{A}) + P(\text{D}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$
• In part (ii), when attempting to combine the probabilities of arrangements A and D, some responses incorrectly multiplied the two probabilities rather than adding them.	• Part (a-ii): $P(\text{A}) + P(\text{D}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ • Part (b-ii): $P(\text{A}) + P(\text{D}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$
• In part (ii), when attempting to combine the probabilities of arrangements A and D, some responses incorrectly subtracted $P(\text{A}) \times P(\text{D})$ from $P(\text{A}) + P(\text{D})$.	• Part (a-ii): $P(\text{A}) + P(\text{D}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ • Part (b-ii): $P(\text{A}) + P(\text{D}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

• Some responses included only a benefit of the chip method or only a drawback of the coin method, but not both.	• The chip method gives equal probability to all possible arrangements, but the coin method does not, as shown in the tables from parts (a-i) and (b-i). Furthermore, the coin method is more likely to result in imbalanced treatment groups with regard to students and teachers, based on the probabilities in parts (a-ii) and (b-ii).
• Although many responses said it would be good to have equally likely arrangements or to avoid imbalanced treatment groups, almost no responses explained *why* this is important.	• If food preferences for teachers are different from those of students, an imbalance is a problem. For example, if one treatment group consists entirely of students, it would be impossible to know if a difference in the response variable is due to the treatment (type of meal) or the role of the person at the school (teacher or student).
• Some responses never explicitly made a choice, even though they included correct comments about the chip and coin methods.	• Use the chip method.

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?

- For any probability calculation, students should provide justification that is easy to follow.
- Make sure students read the question carefully, including information in the initial stem of the question.
- Make sure students know the difference between $P(\text{A or D})$ and $P(\text{A and D})$.
- When using the general addition rule, remind students that $P(\text{A and D}) = 0$ when events are mutually exclusive.
- When asked to make a choice between options, make sure students explain why they are choosing what they are choosing, and why they are not choosing what they are not choosing.
- Suggest that students ask "so what?" or "why does this matter?" at the end of an explanation, especially when choosing a data collection method. Correctly addressing why the choice matters is often the difference between a substantial response and a complete response.
- Make sure students answer the question.

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 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. Previously released exam questions and participation in the Online Teacher Community are of high value for practice and for advice.