# AP® Statistics
# 2014 Scoring Guidelines

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 1

**Intent of Question**

The primary goals of this question were to assess a student's ability to (1) calculate conditional proportions from a two-way table; (2) comment on association between two categorical variables as displayed in a graph; and (3) draw an appropriate conclusion from the *p*-value of a chi-square test.

**Solution**

**Part (a):**

The proportion of on campus residents who participate in at least one extracurricular activity is $\frac{17+7}{33} = \frac{24}{33} \approx 0.727$. The proportion of off campus residents who participate in at least one extracurricular activity is $\frac{25+12}{67} = \frac{37}{67} \approx 0.552$.

**Part (b):**

The graph reveals that on campus residents in this sample are more likely to participate in extra-curricular activities than off campus residents. The proportions who participate in two or more extra-curricular activities are similar between the two groups but slightly greater for on campus residents (on campus: 0.212, off campus: 0.179). On campus residents have a greater proportion who participate in one activity (on campus: 0.515, off campus: 0.373) and a smaller proportion who participate in no extracurricular activities (on campus: 0.273, off campus: 0.448) than off campus residents.

**Part (c):**

The *p*-value of 0.23 is greater than conventional significance levels such as $\alpha = 0.10$ or $\alpha = 0.05$ or $\alpha = 0.01$. Therefore, the *p*-value indicates that the sample data do not provide strong enough evidence to conclude that participation in extracurricular activities differs between on and off campus residents in the population of all students at the university.

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 1 (continued)

**Scoring**

Parts (a), (b), and (c) were scored as essentially correct (E), partially correct (P), or incorrect (I).

**Part (a)** is scored as follows:

Essentially correct (E) if the response correctly performs *both* calculations with work shown.

Partially correct (P) if the response correctly performs one of the two calculations with work shown;
*OR*
if the response provides both correct answers with no work shown;
*OR*
if the response calculates the proportion of students involved in exactly one extracurricular activity rather than the proportion of students involved in at least one extracurricular activity for both groups, with work shown.

Incorrect (I) if the response does not meet the criteria for E or P.

*Note*: Answers reported as fractions rather than decimals are acceptable.

**Part (b)** is scored as follows:

Essentially correct (E) if the response correctly compares proportions between the two groups of students for at least two of the three categories.

Partially correct (P) if the response correctly lists proportions for at least two categories for the two groups but does not make an explicit comparison between the two groups;
*OR*
if the response correctly compares the relative values of the proportions between the two groups of students for only one of the three categories.

Incorrect (I) if the response does not meet the criteria for E or P.

*Notes*:

- A response without any reference to percentages or proportions is scored as at most P, (for example, a response that attempts to compare counts).
- A response that treats bar graphs as distributions of a quantitative variable lowers the score one level (that is, from E to P, or from P to I) in part (b).

# AP® STATISTICS
# 2014 SCORING GUIDELINES

### Question 1 (continued)

**Part (c)** is scored as follows:

Essentially correct (E) if the response states a correct conclusion in the context of the study *AND* provides correct justification/decision of that conclusion based on linkage to the *p*-value.

Partially correct (P) if the response provides no conclusion in context but does provide correct justification/decision based on linkage to the *p*-value;
*OR*
if the response provides a correct conclusion in context but with incorrect or missing justification/decision based on linkage to the *p*-value;
*OR*
if the response provides the conclusion in context and correct justification/decision based on linkage to the *p*-value but states a conclusion equivalent to accepting the null hypothesis.

Incorrect (I) if the response does not meet the criteria for E or P.

*Notes*:
- Justification based on the *p*-value can be given by stating a significance level and noting that the *p*-value is larger than the significance level *OR* by simply stating that the *p*-value is large.
- A conclusion that is equivalent to "accept the null hypothesis", either as a stated decision or as a conclusion in context, cannot be scored as E. Such a response is scored as P if it includes both content and correct justification based on linkage to *p*-value. If such a response lacks either context or linkage, it is scored as I.

**4 Complete Response**

All three parts essentially correct

**3 Substantial Response**

Two parts essentially correct and one part partially correct

**2 Developing Response**

Two parts essentially correct and one part incorrect
*OR*
One part essentially correct and one or two parts partially correct
*OR*
Three parts partially correct

**1 Minimal Response**

One part essentially correct and two parts incorrect
*OR*
Two parts partially correct and one part incorrect

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 2

**Intent of Question**

The primary goals of this question were to assess a student's ability to (1) calculate a probability; (2) assess whether a claim about randomness is questionable in light of a calculated probability; and (3) determine whether a description of a simulation method achieves a correct simulation of a random process.

**Solution**

**Part (a):**

The probability that all 3 people selected are women can be calculated using the multiplication rule, as follows:

$P$(all three selected are women)

$= P(\text{first is a woman}) \times P(\text{second is a woman} \mid \text{first is a woman}) \times P(\text{third is a woman} \mid \text{first two are women})$

$= \frac{3}{9} \times \frac{2}{8} \times \frac{1}{7} \approx 0.012$

**Part (b):**

The probability calculated in part (a) does provide a reason to doubt the manager's claim that the selections were made at random. The calculation shows that there is only about a 1.2% chance that random selection would have resulted in three women being selected. The probability is small enough that it may cast doubt on the manager's claim that the selections were made at random.

**Part (c):**

No, the process does not correctly simulate the random selection of three women from a group of nine people of whom six are men and three are women. The random selection of three people among nine is done *without* replacement. However, in the simulation with the dice, the three dice rolls in any given trial are independent of one another, indicating a selection process that is done *with* replacement.

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 2 (continued)

**Scoring**

Parts (a), (b), and (c) were scored as essentially correct (E), partially correct (P), or incorrect (I).

**Part (a)** is scored as follows:

Essentially correct (E) if the response correctly computes the probability of selecting the three women, and shows how the probability was computed.

Partially correct (P) if the response shows only one of the following:

Gives the correct probability of $\frac{1}{84}$ (0.012 or 0.011 is acceptable) but does not show how it was computed;
*OR*
Correctly shows how the probability should be computed, but does not carry the computation through correctly;
*OR*
Correctly computes (showing work) only the numerator, or only the denominator of the correct answer. (For example, $\frac{1}{9}\times\frac{1}{8}\times\frac{1}{7}\approx 0.002$, or $\frac{3}{9}\times\frac{2}{9}\times\frac{1}{9}\approx 0.008$, or $\frac{3}{9}\times\frac{3}{8}\times\frac{3}{7}\approx 0.054$);
*OR*
Mistakenly assumes independence and calculates (showing work) the binomial probability $\frac{1}{3}\times\frac{1}{3}\times\frac{1}{3}=\frac{1}{27}\approx 0.037$.

Incorrect (I) if the response does not meet the criteria for E or P.

**Part (b)** is scored as follows:

Essentially correct (E) if the response states that the probability from part (a) is small (or insufficiently small), makes an appropriate decision consistent with the probability being small (or insufficiently small), and does so in the context of this situation.

Partially correct (P) if the response shows only one of the following:

Otherwise satisfies the criteria for an E but does so without any context;
*OR*
States a significance level and makes a decision in context that is appropriate to the given probability in part (a) and the stated significance level, but does not explicitly compare the probability and the significance level;
*OR*
Otherwise satisfies the criteria for an E but does not explicitly make a decision about whether there is reason to doubt the manager's claim. (For example: "The probability of selecting the three women from among the nine employees is very small so it is unlikely to occur by chance.")

Incorrect (I) if the response does not meet the criteria for E or P.

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 2 (continued)

*Notes:*

- Each of the following situations is one in which a response that otherwise would be scored as E should be scored as P, and a response that otherwise would be scored as P should be scored as I:
  - The response includes a statement that the small probability *proves* that the manager did not make the selection at random (or any equivalent wording).
  - The response includes a statement that clearly interprets the probability from part (a) to be the probability that the manager selected the three people at random.
- Each of the following situations is one in which the response is scored as I:
  - The decision is inconsistent with the justification (*e.g.*, "The probability is very small, so there is no reason to doubt the manager's claim").
  - The response states or implies that because the selection of three women was *not impossible,* there is *no reason to doubt* the manager's claim.

**Part (c)** is scored as follows:

Essentially correct (E) if the response answers no *AND* states that the dice outcomes in the proposed simulation are independent *AND* states that the genders of the selected convention attendees are dependent. The table below shows statements that should be considered equivalent to the required statements of independence and dependence.

| Independence of dice outcomes | Dependence of genders |
|---|---|
| • The three dice outcomes are independent.<br>• The probability of rolling a 5 or a 6 is the same on all three dice.<br>• The dice simulation actually simulates sampling with replacement. | • The genders of the three people are dependent (or not independent).<br>• The probability of selecting a woman changes after each selection.<br>• The people are sampled without replacement. |

*OR*

Essentially correct if the response answers no *AND* computes the correct probability that a trial of the simulation will indicate the selection of three women $\left(\frac{1}{3}\times\frac{1}{3}\times\frac{1}{3}\approx 0.037\right)$ *AND* states that the probability is different from the probability found in part (a).

Partially correct (P) if the response correctly answers no and either:

States only that the dice outcomes are independent or states only that the genders of the selected convention attendees are dependent, but not both;

*OR*

Otherwise meets the criteria for E but has poor communication. An example of poor communication is: "No, because it selects with replacement. It isn't possible for the same person to be selected twice." (There is an apparent shift between the two sentences from describing the simulation to describing the actual selection of people, but that is not made clear.)

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 2 (continued)

Incorrect (I) if the response does not meet the criteria for E or P.

*Note:* Pointing out that a sample of three people is more than 10% of the population of nine people should be considered equivalent to stating that the selection of a woman is not independent among the three people selected to attend the convention.

**4 Complete Response**

All three parts essentially correct

**3 Substantial Response**

Two parts essentially correct and one part partially correct

**2 Developing Response**

Two parts essentially correct and one part incorrect

*OR*

One part essentially correct and one or two parts partially correct

*OR*

Three parts partially correct

**1 Minimal Response**

One part essentially correct and two parts incorrect

*OR*

Two parts partially correct and one part incorrect

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 3

**Intent of Question**

The primary goals of this question were to assess a student's ability to (1) perform a probability calculation from a normal distribution; (2) explain an implication of examining the distribution of a sample mean rather than the distribution of a single measurement; and (3) perform a probability calculation involving independent events using the multiplication rule.

**Solution**

**Part (a):**

Because the distribution of the daily number of absences is approximately normal with mean 120 students and standard deviation 10.5 students, the *z*-score for an absence total of 140 students is $z = \frac{140 - 120}{10.5} \approx 1.90$. The table of standard normal probabilities or a calculator reveals that the probability that 140 or fewer students are absent is 0.9713. So the probability that more than 140 students are absent (and that the school will lose some state funding) is $1 - 0.9713 = 0.0287$.

**Part (b):**

High School A would be *less* likely to lose state funding. With a random sample of 3 days, the distribution of the sample mean number of students absent would have less variability than that of a single day. With less variability, the distribution of the sample mean would concentrate more narrowly around the mean of 120 students, resulting in a smaller probability that the mean number of students absent would exceed 140.

In particular, the standard deviation of the sample mean number of absences, $\overline{x}$, is $\frac{\sigma}{\sqrt{n}} = \frac{10.5}{\sqrt{3}} = 6.062$. So the *z*-score for a sample mean of 140 is $\frac{140 - 120}{6.062} \approx 3.30$. The probability that High School A loses funding using the suggested plan would be $1 - 0.9995 = 0.0005$, as determined from the table of standard normal probabilities or from a calculator, which is less than a probability of 0.0287 obtained for the plan described in part (a).

**Part (c):**

For any one typical school week, the probability is $\frac{2}{5} = 0.4$ that the day selected is not Tuesday, not Wednesday, or not Thursday. Therefore, because the days are selected independently across the three weeks, the probability that none of the three days selected would be a Tuesday or Wednesday or Thursday is $(0.4)^3 = 0.064$.

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 3 (continued)

**Scoring**

Parts (a), (b), and (c) were scored as essentially correct (E), partially correct (P), or incorrect (I).

**Part (a)** is scored as follows:

Essentially correct (E) if the response provides the following three components:
1. Indicates use of a normal distribution and clearly identifies the correct parameter values (showing correct components of a *z*-score calculation is sufficient).
2. Uses the correct boundary value (140, 140.5, or 141 is acceptable).
3. Reports the correct normal probability consistent with components 1 and 2

*OR*

if the response reports a probability of 0.025 with justification based on the empirical rule for an acceptable boundary value (140, 140.5, or 141 is acceptable).

Partially correct (P) if the response correctly provides only two of the three components listed above.

*OR*

if the response provides an incorrect probability of 0.05 with justification based on the empirical rule for an acceptable boundary value (140, 140.5, or 141 is acceptable).

Incorrect (I) if the response does not satisfy the criteria for E or P.

*Note*: An inconsistency in calculations lowers the score for part (a) by one level (that is, from E to P or from P to I).

**Part (b)** is scored as follows:

Essentially correct (E) if the response provides the correct answer of less likely *AND* the following three components:
1. Clearly references the distribution of the sample mean.
2. Indicates that the variability of the distribution is smaller.
3. Indicates that the distribution is centered at 120.

*OR*

if the response provides the correct answer of less likely *AND* the following two components:
1. Correctly calculates the probability that the sample mean would exceed 140 (arithmetic errors are not penalized).
2. Correctly compares this probability to the probability in part (a).

Partially correct (P) if the response provides the correct answer of less likely *AND* only two of the following three components:
1. Clearly references the distribution of the sample mean.
2. Indicates that the variability of the distribution is smaller.
3. Indicates that the distribution is centered at 120.

*OR*

if the response provides the correct answer of less likely *AND* correctly calculates the probability that the sample mean would exceed 140 (arithmetic errors are not penalized) *BUT* does not correctly compare this probability with the probability in part (a).

## Question 3 (continued)

Incorrect (I) if the response does not meet the criteria for E or P, including if the response provides the incorrect answer or provides the correct answer of less likely with no explanation or an incorrect explanation.

*Note*: An equivalent approach is to use the total number of absences for 3 days. The sampling distribution of the total number of absences for the 3 days is approximately normal, with mean $3(120) = 360$ absences and standard deviation $3(6.026) \approx 18.187$ absences. The *z*-score for a total of $3(140) = 420$ absences is: $\frac{420 - 360}{18.187} \approx 3.30$ . Such a response is scored E if the response provides the correct answer of less likely and references the distribution of the sample total, and includes the correct mean and standard deviation.

**Part (c)** is scored as follows:

Essentially correct (E) if the response correctly calculates the probability *AND* shows sufficient work.

Partially correct (P) if the response reports the correct probability but shows no work or does not show sufficient work;
*OR*
if the response uses the multiplication rule involving three events but does so incorrectly and/or with an incorrect probability of not selecting a Tuesday, Wednesday, or Thursday.

Incorrect (I) if the response does not meet the criteria for E or P.

**4 Complete Response**

All three parts essentially correct

**3 Substantial Response**

Two parts essentially correct and one part partially correct

**2 Developing Response**

Two parts essentially correct and one part incorrect
*OR*
One part essentially correct and one or two parts partially correct
*OR*
Three parts partially correct

**1 Minimal Response**

One part essentially correct and two parts incorrect
*OR*
Two parts partially correct and one part incorrect

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 4

**Intent of Question**

The primary goals of this question were to assess a student's ability to (1) describe why the median might be preferred to the mean in a particular context; (2) compare the relative merits of two sampling plans; and (3) describe a consequence of nonresponse in a particular study.

**Solution**

**Part (a):**

The median is less affected by skewness and outliers than the mean. With a variable such as income, a small number of very large incomes could dramatically increase the mean but not the median. Therefore, the median would provide a better estimate of a typical income value.

**Part (b):**

Method 2 is better than Method 1. A sample obtained from Method 1 could be biased because of the voluntary nature of the response. It is plausible that class members with larger incomes might be more likely to return the form than class members with smaller incomes. The mean income for such a sample would overestimate the mean income of all class members. With Method 2, despite the smaller sample size, the random selection is likely to result in a sample that is more representative of the entire class and produce an unbiased estimate of mean yearly income of all class members.

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 4 (continued)

**Scoring**

This question is scored in three sections. Part (b) has two components: (1) identifying a relevant characteristic for each sampling method; (2) indicating the effect of the biased method on the estimate of the mean income. Section 1 consists of Part (a); section 2 consists of part (b), component 1; and section 3 consists of part (b), component 2. Sections 1, 2, and 3 are scored as essentially correct (E), partially correct (P), or incorrect (I).

**Section 1** is scored as follows:

Essentially correct (E) if the response includes the following two components:
1. Describes how skewness or outliers affect the mean or do not affect the median.
2. Makes a conjecture about a relevant characteristic of the distribution of incomes, such as skewness or an outlier.

Partially correct (P) if the response includes only one of the two components listed above.

Incorrect (I) if the response does not meet the criteria for E or P.

*Notes*:
- For Component 1, examples of responses that are acceptable include:
  - The mean is affected by skewness (outliers).
  - The median is not affected by skewness (outliers).
  - The mean is greater (less) than the median when there is right (left) skewness or outliers.
- For Component 1, examples of responses that are not acceptable include:
  - Don't use the mean for skewed distributions or distributions with outliers.
  - Use the median for skewed distributions or distributions with outliers.
  - Responses that include an incorrect statement about means and/or medians, such as for right skewed distributions, the median will be higher than the mean.
- It is possible to satisfy *both* components with a single sentence, such as, "If there was a billionaire in the sample, the mean would be higher than the median."
- If a response argues that using the *mean* is a more appropriate way to estimate the typical income, then reduce the score in section 1 by one level (that is, from E to P or from P to I).

**Section 2** is scored as follows:

Essentially correct (E) if the response chooses Method 2 *AND* includes the following two components:
1. Identifies a relevant characteristic of Method 1.
2. Identifies a relevant characteristic of Method 2.

Partially correct (P) if the response chooses Method 2 *AND* includes only one of the two components listed above
*OR*
if the response includes both components but does not choose a method.

Incorrect (I) if the response chooses Method 1 *OR* otherwise does not meet the criteria for E or P.

## Question 4 (continued)

*Notes*:

- Responses that do not explicitly choose Method 2 can still earn an E for section 2 if the choice is clearly implied. The choice of Method 2 is clearly implied if the response only discusses negative characteristics of Method 1 and only discusses positive characteristics of Method 2, such as, Method 1 is biased but Method 2 uses a random sample.
- Responses that *compare* the two methods can satisfy *both* components, such as, saying that Method 1 is more biased or that nonresponse will be less of a problem with Method 2.
- Responses that refer to the nonresponse bias as *voluntary response bias*, *response bias, undercoverage* can still earn an E.
- Discussions of conditions for inference should be considered extraneous and ignored.

**Section 3** is scored as follows:

Essentially correct (E) if the response includes the following two components:

1. Indicates the incomes of responders may be different from the incomes of nonresponders.
2. Indicates the biased sampling method may produce a misleading estimate/conclusion about the mean income, including direction, for example, "The sample mean is likely to be higher than the mean of the population."

Partially correct (P) if the response provides only one of the two components listed above.

Incorrect (I) if the response does not meet the criteria for E or P.

*Notes*:

- A single sentence can satisfy the first component of section 2 and the first component of section 3. (For example, "In method 1, rich people are more likely to respond.")
- For component 2, either direction is acceptable but the direction must be consistent with the identified bias. Saying only that Method 2 will be more accurate or more representative does not satisfy component 2.
- If a response addresses possible nonresponse bias in Method 2, the response can still satisfy both components of section 3.
- Responses that focus on the larger sample size in Method 1 can satisfy component 2 if such responses describe the effect as reducing the variability of the estimate. (For example, "I would use Method 1 since the larger sample size would give less variability of the mean.")
- Responses that focus on untruthful survey answers can satisfy component 2 if the effect on the estimate is appropriate. (For example, "People contacted in Method 2 might say they make more money than they actually do. This would make the estimated mean too high.")

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 4 (continued)

**4 Complete Response**

All three sections essentially correct

**3 Substantial Response**

Two sections essentially correct and one section partially correct

**2 Developing Response**

Two sections essentially correct and one section incorrect

*OR*

One section essentially correct and one or two sections partially correct

*OR*

Three sections partially correct

**1 Minimal Response**

One section essentially correct and two sections incorrect

*OR*

Two sections partially correct and one section incorrect

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 5

**Intent of Question**

The primary goal of this question was to assess students' 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 students' 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.

**Solution**

Step 1: States a correct pair of hypotheses.

Let $\mu_{\text{diff}}$ represent the population mean difference in purchase price (woman – man) for identically equipped cars of the same model, sold to both men and women by the same dealer, in the county.

The hypotheses to be tested are $H_0 : \mu_{\text{diff}} = 0$ versus $H_a : \mu_{\text{diff}} > 0$.

Step 2: Identifies a correct test procedure (by name or by formula) and checks appropriate conditions.

The appropriate procedure is a paired *t*-test.

The conditions for the paired *t*-test are:

1. The sample is randomly selected from the population.
2. The population of price differences (woman – man) is normally distributed, or the sample size is large.

The first condition is met because the car models and the individuals were randomly selected. The sample size ($n = 8$) is not large, so we need to investigate whether it is reasonable to assume that the population of price differences is normally distributed. The dotplot of sample price differences reveals a fairly symmetric distribution, so we will consider the second condition to be met.

Step 3: Correct mechanics, including the value of the test statistic and *p*-value (or rejection region).

The test statistic is $t = \dfrac{585 - 0}{\dfrac{530.71}{\sqrt{8}}} \approx 3.12$.

The *p*-value, based on a *t*-distribution with $8 - 1 = 7$ degrees of freedom, is 0.008.

Step 4: States a correct conclusion in the context of the study, using the result of the statistical test.

Because the *p*-value is very small (for instance, smaller than $\alpha = 0.05$), we reject the null hypothesis. The data provide convincing evidence that, on average, women pay more than men in the county for the same car model.

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 5 (continued)

**Scoring**

Each of steps 1, 2, 3, and 4 were scored as essentially correct (E), partially correct (P), or incorrect (I).

**Step 1** is scored as follows:

Essentially correct (E) if the response identifies the correct parameter *AND* states correct hypotheses.

Partially correct (P) if the response identifies the correct parameter *OR* states correct hypotheses, but not both.

Incorrect (I) if the response does not meet the criteria for E or P.

*Note*: Defining the parameter symbol in context or simply using common parameter notation is sufficient.

**Step 2** is scored as follows:

Essentially correct (E) if the response identifies the correct test procedure (by name or by formula) *AND* checks *both* conditions correctly.

Partially correct (P) if the response correctly completes two of the three components (identification of procedure, check of randomness condition, check of normality condition).

Incorrect (I) if the response does not meet the criteria for E or P.

*Note*: The random sampling condition can be verified by referring to the random selection of car models or to the random selection of male and female car buyers.

**Step 3** is scored as follows:

Essentially correct (E) if the response correctly calculates both the test statistic and the *p*-value.

Partially correct (P) if the response correctly calculates the test statistic but not the *p*-value;
*OR*
if the response calculates the test statistic incorrectly but then calculates the correct *p*-value for the computed test statistic.

Incorrect (I) if the response does not meet the criteria for E or P.

*Note:* If the response identifies a *z*-test for a mean as the correct procedure in step 2, then the response can earn a P in step 3 if both the test statistic and the *p*-value are calculated correctly.

## Question 5 (continued)

**Step 4** is scored as follows:

Essentially correct (E) if the response provides a correct conclusion in context, also providing justification based on linkage between the *p*-value and the conclusion.

Partially correct (P) if the response provides a correct conclusion with linkage to the *p*-value, but not in context;
*OR*
if the response provides a correct conclusion in context, but without justification based on linkage to the *p*-value.

Incorrect (I) if the response does not meet the criteria for E or P.

*Notes*:

- If the conclusion is consistent with an incorrect *p*-value from step 3 and also in context with justification based on linkage to the *p*-value, step 4 is scored as E.
- A response that performs a two-sample *t*-test with correct calculations should fail to reject $H_0$. A conclusion that is equivalent to "accept $H_0$" (such as "we conclude that women pay the same amount as men, on average"), either as a stated decision or as a conclusion in context, cannot be scored as E. Such a response will be scored as P provided that the conclusion is in context with linkage. Such a response will be scored as I if it lacks either context or linkage.

Each essentially correct (E) step counts as 1 point. Each partially correct (P) step counts as ½ point.

**4 Complete Response**

**3 Substantial Response**

**2 Developing Response**

**1 Minimal Response**

If a response is between two scores (for example, 2½ points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 6

**Intent of Question**

The primary goals of this question were to assess a student's ability to (1) calculate and interpret a residual value; (2) answer questions about residual plots; (3) compare associations between two scatterplots; and (4) identify an appropriate explanatory variable to include in a regression model based on residuals from simpler regression models.

**Solution**

**Part (a):**

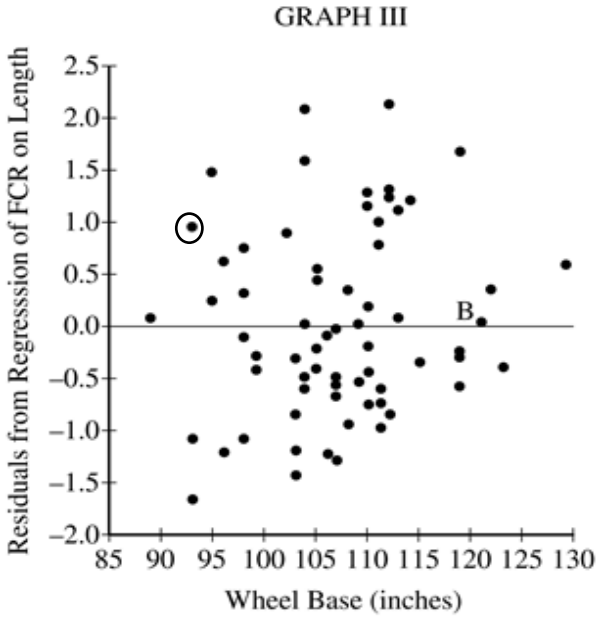
For a car with length 175 inches, the predicted value for the car's FCR, based on the least squares regression line, is

$$\text{predicted FCR} = -1.595789 + 0.0372614(175) \approx 4.92 \text{ gallons per 100 miles.}$$

The actual FCR for the car is 5.88, so the residual is $5.88 - 4.92 = 0.96$. The residual value means that the car's FCR is 0.96 gallons per 100 miles greater than would be predicted for a car of its length.

**Part (b):**

(i) The point with a wheel base of 93 inches and a residual of 0.96 gallons per 100 miles is circled in graph III below.




(ii) Point B corresponds to a car with an actual FCR that is very close to the FCR that would be predicted for a car with its length by the regression model which predicts FCR using the explanatory variable length.

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 6 (continued)

**Part (c):**

Graph II reveals a moderate association that is positive and linear. In contrast, there is a weak association that is positive and linear in graph III. The association between engine size and residual (from predicting FCR based on length) is stronger than the association between wheel base and residual (from predicting FCR based on length).

**Part (d):**

Engine size is a better choice than wheel base for including with length in a regression model for predicting FCR. The stronger association between engine size and residual (from predicting FCR based on length) indicates that engine size is more useful than wheel base for reducing the variability in FCR values that remains unexplained (as indicated by residuals) after predicting FCR based on length.

**Scoring**

Parts (a), (b), (c), and (d) were scored as essentially correct (E), partially correct (P), or incorrect (I).

**Part (a)** is scored as follows:

Essentially correct (E) if the response provides the following two components:
1. A correct residual value with supporting calculation.
2. A correct interpretation of the residual value, in context.

Partially correct (P) if the response includes only one of the two components listed above.

Incorrect (I) if the response does not meet the criteria for E or P.

*Notes*:
- If the residual value is incorrect, the interpretation should be considered correct if it follows from the incorrect residual value.
- Correct interpretation of the residual must include the correct direction and magnitude of the FCR value away from the predicted FCR value.
- A calculated residual value which is slightly different from 0.96 due to the number of significant digits is acceptable.

**Part (b)** is scored as follows:

Essentially correct (E) if the response provides the following two components:
1. Circles the correct point in graph III.
2. Provides a reasonable interpretation of the car associated with point B having a residual near 0 that refers to predicting FCR based on length.

Partially correct (P) if the response correctly provides only one of the two components listed above.

Incorrect (I) if the response does not meet the criteria for E or P.

## Question 6 (continued)

*Note:* A correct response for the second component must include reference to the observed FCR value of the car represented by point B, not the point B itself.

**Part (c)** is scored as follows:

Essentially correct (E) if the response correctly provides the following three components:
1. A description of form *AND* direction for both graphs.
2. A description of the strength of association for both graphs.
3. A comparison between the two graphs.

Partially correct (P) if the response correctly provides only two of the three components listed above.

Incorrect (I) if the response does not meet the criteria for E or P.

*Notes:*
- Part (c) is focused on the comparison of graph II and graph III. Inferences drawn from patterns in these graphs are considered in part (d).
- Linear is needed for form in graph II.
- Graph III may be described as having no association between wheel base and the residuals of FCR based on length, which is sufficient for describing the form, direction and strength of association of graph III.

**Part (d)** is scored as follows:

Essentially correct (E) if the response indicates the correct choice with a sound justification based on the following two components:
1. The strong(er) association.
2. Reducing the variability that remains unexplained in the model which predicts FCR based on length.

Partially correct (P) if the response indicates the correct choice and provides a justification based on only one of the two components which are listed above.

Incorrect (I) if the response indicates the incorrect choice;
*OR*
if the response indicates the correct choice but does not mention either of the two components which are listed above.

*Note:* Describing the variables in graph II and graph III as residuals is not required but can be used positively in holistic scoring. Incorrect descriptions of graph II or graph III or the variables in graphs are not acceptable.

# AP® STATISTICS
# 2014 SCORING GUIDELINES

## Question 6 (continued)

Each essentially correct (E) part counts as 1 point. Each partially correct (P) part counts as ½ point.

**4 Complete Response**

**3 Substantial Response**

**2 Developing Response**

**1 Minimal Response**

If a response is between two scores (for example, 2½ points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.