CollegeBoard AP

AP® Statistics
2009 Scoring Guidelines
Form B

The College Board

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 5,600 schools, colleges, universities and other educational organizations. Each year, the College Board serves seven million students and their parents, 23,000 high schools and 3,800 colleges through major programs and services in college readiness, college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT® and the Advanced Placement Program® (AP®). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities and concerns.

Visit the College Board on the Web: www.collegeboard.com.
AP Central® is the official online home for AP teachers: apcentral.collegeboard.com.

AP® STATISTICS
## 2009 SCORING GUIDELINES (Form B)

Question 1

Intent of Question

The primary goals of this question were to assess students' ability to (1) approximate the median and the IQR of a distribution from the boxplot and (2) recalibrate the values of the median and the IQR of the distribution if the same constant value is added to each observation in the distribution.

Solution

Part (a):

Median

Q1   Q3

State Gasoline Taxes (in cents per gallon)
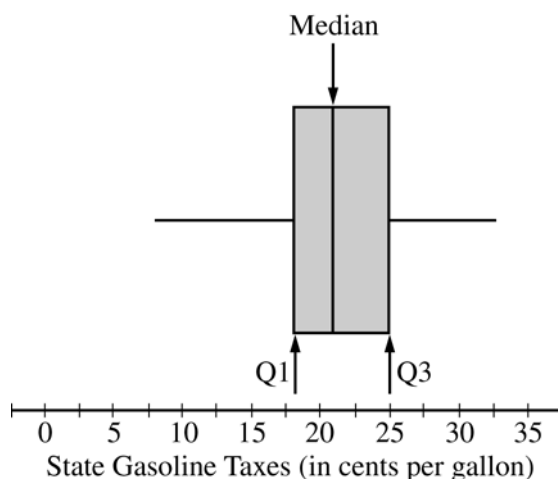
The median and quartiles are marked and labeled on the boxplot above. The median is approximately 21 cents per gallon.

The first and third quartiles are approximately 18 cents per gallon and 25 cents per gallon, respectively. The IQR is Q3 – Q1, which is approximately 25 – 18 = 7 cents per gallon.

Part (b):

After adding 18.4 cents per gallon to each of the state taxes, the median of the combined gasoline taxes would be the median of the state tax plus the federal tax, which is approximately 21 + 18.4 = 39.4 cents per gallon.

Although the quartiles of the combined gasoline taxes will change (Q1 = 18 + 18.4 = 36.4 cents per gallon and Q3 = 25 + 18.4 = 43.4 cents per gallon), the IQR will remain the same as it was for the state taxes at 7 cents per gallon (43.4 – 36.4 = 7).

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the student identifies reasonable values for the median and IQR and justifies them by marking and labeling the boxplot.

Partially correct (P) if the student identifies reasonable values for the median and IQR but does not mark or label the boxplot *OR* identifies, marks, and labels only one value (median or IQR).

Incorrect (I) if the student identifies neither value *OR* identifies only one value but fails to mark and label the boxplot.

Part (b) is scored as follows:

Essentially correct (E) if the student gives a median that is 18.4 cents per gallon larger than the median identified in part (a), gives an IQR that is the same single number found in part (a), *AND* provides a reasonable justification for at least one of these values.

Partially correct (P) if the student provides only one correct value (either the median or the IQR) *AND* provides a justification.

Incorrect (I) if the student gives incorrect values for the median and IQR *OR* provides only one correct value with no justification.

4   Complete Response

> Both parts essentially correct

3   Substantial Response

> One part essentially correct and one part partially correct

2   Developing Response

> One part essentially correct and one part incorrect
>
> *OR*
>
> Both parts partially correct

1   Minimal Response

> One part partially correct and one part incorrect

Question 2

Intent of Question

The primary goal of this question was to assess students' ability to evaluate conditional probabilities as they relate to diagnostic testing.

Solution

Part (a):

The estimated probability of a positive ELISA if the blood sample does not have HIV present is

$$\frac{37}{500} \quad OR \quad \frac{37}{500} = 0.074$$

Part (b):

A total of 489 + 37 = 526 blood samples resulted in a positive ELISA. Of these, 489 samples actually contained HIV. Therefore the proportion of samples that resulted in a positive ELISA that actually contained HIV is

$$\frac{489}{526} \quad OR \quad \frac{489}{526} \approx 0.9297$$

Part (c):

From part (a), the probability that the ELISA will be positive, given that the blood sample does not actually have HIV present, is 0.074. Thus, the probability of a negative ELISA, given that the blood sample does not actually have HIV present, is $1 - 0.074 = 0.926$.

P(new blood sample that does not contain HIV will be subjected to the more expensive test)

> = *P*(*1st ELISA positive* and *2nd ELISA positive* <u>OR</u> *1st ELISA positive* and *2nd ELISA negative* and *3rd ELISA positive* | *HIV not present in blood*)
>
> = *P*(*1st ELISA positive* and *2nd ELISA positive* | *HIV not present in blood*)
> + *P*(*1st ELISA positive* and *2nd ELISA negative* and *3rd ELISA positive* | *HIV not present in blood*)
>
> = (0.074)(0.074) + (0.074)(0.926)(0.074)
>
> = 0.005476 + 0.005070776
>
> = 0.010546776
>
> ≈ 0.0105

> *OR*

P(new blood sample that does not contain HIV will be subjected to the more expensive test)

> = *P*(*1st ELISA positive and not both the 2nd and 3rd are negative*)
> $= (0.074)(1 - 0.926^2)$
> $= (0.074)(0.142524)$
> $= 0.010546776$
> ≈ 0.0105

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the student writes the correct fraction for the estimated probability.

Partially correct (P) if the decimal answer of 0.074 is given with no justification.

Incorrect (I) otherwise.

Part (b) is scored as follows:

Essentially correct (E) if the student writes the correct fraction for the proportion or gives a decimal approximation with justification.

Partially correct (P) if the student writes the wrong fraction but either correctly selects 489 as the numerator or correctly computes $489 + 37 = 526$ as the denominator.

Incorrect (I) otherwise.

Part (c) is scored as follows:

Essentially correct (E) if the student computes the correct probability, showing work.

Partially correct (P) if method is equivalent to *P*(*1st ELISA is positive*) · *P*(*at least one of two subsequent ELISAs is positive*) $= (0.074)(0.1425)$, except that one of the factors is incorrect.
>   OR

The student correctly computes *P*(*at least one of two subsequent ELISAs is positive*) = 0.1425, failing to include the first factor (0.074).
>   OR

The student correctly computes the probability of getting at least two positive ELISAs by testing negative blood three times:

$$\binom{3}{2}(0.074)^2(0.926) + \binom{3}{3}(0.074)^3 = 3(0.074)^2(0.926) + (0.074)^3 \approx 0.015617552$$

Incorrect (I) otherwise.

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 no parts partially correct

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 no parts partially correct

OR

   No parts essentially correct and two parts partially correct

Question 3

Intent of Question

The primary goals of this question were to assess students' ability to (1) state and check appropriate conditions for inference in a study, (2) state the appropriate hypotheses for a statistical test, (3) identify and compute the appropriate test statistic, and (4) make a conclusion in the context of the problem.

Solution

Let A represent the cardiopump treatment, and let B represent the CPR treatment.

Let p_A = proportion of patients who will survive at least one year if treated with the cardiopump.

Let p_B = proportion of patients who will survive at least one year if treated with CPR.

Part (a):

Step 1: State the conditions for inference.

The conditions required for a two-sample z test of equal proportions for an experiment are:
1. Random assignment of treatments to subjects
2. Sufficiently large sample sizes

Step 2: Check the conditions.

1. If we assume that the relevant characteristics of people who have heart attacks on even-numbered and odd-numbered days are comparable, randomly assigning one treatment to be given on even-numbered days and the other to be given on odd-numbered days is a reasonable approximation to randomly assigning the two treatments to the available subjects.
2. The large sample condition is met because all of the following are at least 5 (or 10):

$$n_A \hat{p}_A = 37 \geq 5 \text{ or } 10, \; n_A(1 - \hat{p}_A) = 717 \geq 5 \text{ or } 10$$
$$n_B \hat{p}_B = 15 \geq 5 \text{ or } 10, \; n_B(1 - \hat{p}_B) = 731 \geq 5 \text{ or } 10$$

Part (b):

Step 1: State a correct pair of hypotheses.

$$H_0 : p_A - p_B = 0 \; (\text{or } p_A = p_B)$$
$$H_a : p_A - p_B > 0 \; (\text{or } p_A > p_B)$$

Step 2: Identify a correct test by name or by formula.

Two-sample z test for proportions

OR

$$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_A} + \frac{\hat{p}(1-\hat{p})}{n_B}}} \quad \text{where} \quad \hat{p} = \frac{n_A\hat{p}_A + n_B\hat{p}_B}{n_A + n_B}.$$

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

$$\hat{p}_A = \frac{37}{754} \approx 0.049 \qquad \hat{p}_B = \frac{15}{746} \approx 0.020 \qquad \hat{p} = \frac{37 + 15}{754 + 746} = \frac{52}{1500} \approx 0.035$$

$$z = \frac{\frac{37}{754} - \frac{15}{746}}{\sqrt{\frac{52}{1500}\left(1 - \frac{52}{1500}\right)\left(\frac{1}{754} + \frac{1}{746}\right)}} \approx 3.066$$

The p-value is 0.0011.

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

Because the p-value of 0.0011 is very small, that is, less than any reasonable significance level such as $\alpha = 0.01$, or $\alpha = 0.05$, we reject the null hypothesis. We have strong evidence to support the conclusion that the proportion of patients who survive when treated with the cardiopump is higher than the proportion of patients who survive when treated with CPR; that is, the survival rate is higher for patients treated with the cardiopump. (*OR*, If all of these patients had been assigned the cardiopump, we have strong evidence that the survival rate would be higher than if all of these patients had been assigned CPR.)

Scoring

This problem is scored in four sections. Section 1 consists of part (a). Section 2 consists of part (b), step 1. Section 3 consists of part (b), steps 2 and 3. Section 4 consists of part (b), step 4. Section 1 is scored as essentially correct (E), partially correct (P), or incorrect (I). Sections 2, 3, and 4 are each scored as essentially correct (E) or incorrect (I).

Section 1 [part (a)] is scored as follows:

Essentially correct (E) if the student correctly states and addresses randomization with reasonable justification *AND* correctly checks that the numbers of successes and failures are at least 5 or 10.

Partially correct (P) if the student correctly states and checks only one condition, *OR* if the student states both conditions correctly but checks neither of them.

Incorrect (I) otherwise.

Section 2 [part (b), step 1] is scored as follows:

Essentially correct (E) if the student states a correct pair of hypotheses.

Note: The hypotheses may be stated using words or using transparent variable notation such as p_{CPR} and p_{CP}, even if the parameters are not defined.

Incorrect (I) otherwise.

Note: Hypotheses that clearly address sample proportions are incorrect. It must be clear that the hypotheses are not about the sample proportions.

Section 3 [part (b), steps 2 and 3] is scored as follows:

Essentially correct (E) if the student identifies a correct test and includes correct mechanics.

Note: The mechanics are considered correct if the student uses a pooled test procedure or an unpooled test procedure, as long as the correct value of z is calculated.

Incorrect (I) otherwise.

Note: If the student writes the formula for the unpooled version but gets the correct values for the test statistic and p-value for the pooled version, section 3 is scored as incorrect.

Section 4 [part (b), step 4] is scored as follows:

Essentially correct (E) if the student states a correct conclusion in the context of the problem.

Note: A correct conclusion in context must explicitly state that the cardiopump survival rate is significantly higher than the CPR survival rate.

Incorrect (I) otherwise.

Notes:
- If the p-value in section 3 is incorrect but the conclusion is consistent with the computed p-value, section 4 can be considered correct.
- In section 4 if both an α and a p-value are given together, the linkage between the p-value and the conclusion is implied. If no α is given, the solution must be explicit about the linkage by giving a correct interpretation of the p-value or explaining how the conclusion follows from the p-value.
- If, instead of a one-sided test, a student correctly performs a two-sided test (chi-square test for homogeneity of proportions or a two-sided z test for comparing two proportions), the final score drops automatically by 1 point.

Each essentially correct (E) response counts as 1 point, and a partially correct (P) response in part (a) 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, 1½ points), use a holistic approach to determine whether to score up or down, depending on the strength of the response and communication.

AP® STATISTICS
2009 SCORING GUIDELINES (Form B)

Question 4

Intent of Question

The primary goals of this question were to assess students' ability to (1) describe assignment of experimental units to treatments in a block design and (2) provide ways to increase the power of an experiment.

Solution

Part (a):

Form three blocks based on the species of bird (blackbirds, starlings, and geese) to accomplish the goal of blocking to create groups of homogeneous experimental units. Within each of the three blocks, carry out a completely randomized design by randomly assigning the birds within each block to one of the two treatments. Within block 1, each bird of a particular species (let's say the blackbirds) will be tagged with a unique random number using a random number generator on a calculator, statistical software, or a random number table. The random numbers will be sorted from lowest to highest. The birds with the lowest 50 numbers in the ordered list will receive treatment 1 (red background with narrow blue stripes). The birds with the highest 50 numbers will receive treatment 2 (blue background with narrow red stripes). This method of randomization should be repeated in the other two blocks.

Part (b):

To increase power (other than by blocking), the researcher could increase the sample size. This reduces the standard error of the sampling distribution. With a smaller standard error, a test is more likely to be able to detect a difference in results from the two treatments, if such a difference exists.

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the student describes a method for randomization that (1) randomly assigns the two treatments to the birds within each of the three blocks and (2) describes the method with sufficient detail so that a knowledgeable statistics user could replicate the method.

Partially correct (P) if the student describes a method for randomization within blocks but does not provide sufficient detail so that a knowledgeable statistics user could replicate the method *OR* the method of randomization does not guarantee sufficient randomization.

Incorrect (I) if there is no randomization or if the randomization is not done within blocks.

Note: The sample sizes within the treatment groups are not required to be equal.

Part (b) is scored as follows:

Essentially correct (E) if the student identifies a correct method for increasing power *AND* links the method to the mechanism for increasing power.

Partially correct (P) if the student identifies a correct method for increasing power but fails to link the method to the mechanism for increasing power.

Incorrect (I) if the student does not identify a correct method for increasing power.

Note: Another correct method for increasing power is to increase the significance level, α. By increasing α, the researcher makes it easier to reject a false null hypothesis, which increases the power of the test.

4   Complete Response

> Both parts essentially correct

3   Substantial Response

> One part essentially correct and one part partially correct

2   Developing Response

> One part essentially correct and one part incorrect
>
> *OR*
>
> Both parts partially correct

1   Minimal Response

> One part partially correct and one part incorrect

Question 5

Intent of Question

The primary goals of this question were to assess students' ability to (1) state the appropriate hypotheses, (2) identify and compute the appropriate test statistic, (3) make a conclusion in the context of the problem for a one-sample t test, and (4) use simulation results to find a simulated p-value to make an inference about the standard deviation.

Solution

Part (a):

Step 1: State a correct pair of hypotheses.

H_0: $\mu = 12.1$

H_a: $\mu \neq 12.1$, where μ is the mean number of fluid ounces dispensed into all juice bottles filled in the past hour

Step 2: Identify a correct test (by name or by formula) and check appropriate conditions. (Stem of the question said to assume that conditions for inference are met.)

One-sample t test for a mean $\;\;OR\;\;$ $t = \dfrac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$

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

test statistic: $t = \dfrac{12.05 - 12.1}{\frac{0.085}{\sqrt{4}}} = \dfrac{-0.05}{0.0425} = -1.176$

p-value: $2 \cdot P(T_{3\,df} < -1.176) = 0.324$

Step 4: State a correct conclusion in the context of the problem, using the result of the computations.

Because the p-value of 0.324 is larger than any reasonable significance level, such as $\alpha = .05$, do not reject the null hypothesis that the mean number of fluid ounces being dispensed is 12.1 fluid ounces. There is not sufficient evidence to conclude that the machine is filling the juice bottles with an average amount different from 12.1 fluid ounces.

Part (b):

In 300 simulated sample standard deviations, the value of the computed standard deviation (0.085) from our sample in part (a) or a value larger than 0.085 occurred only 12 times. This is a simulated p–value of $\frac{12}{300}$ or 0.04. If the actual population standard deviation is 0.05, then we estimate that the chance of observing a sample standard deviation of 0.085 or larger is 4 percent.

Because this simulated p-value is less than a significance level of 5 percent, the sample from part (a) provides strong evidence that the standard deviation of the juice being dispensed exceeds 0.05 ounces.

AP® STATISTICS
2009 SCORING GUIDELINES (Form B)

Question 5 (continued)

Scoring

This problem is scored in four sections. Section 1 consists of part (a), step 1. Section 2 consists of part (a), steps 2 and 3. Section 3 consists of part (a), step 4. Section 4 consists of part (b). Sections 1, 2, and 3 are each scored as essentially correct (E) or incorrect (I). Both step 2 and step 3 in section 2 must be completely correct to earn an E for section 2. Section 4 is scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 [part (a), step 1] is scored as follows:

Essentially correct (E) if the student states a correct pair of hypotheses.

Incorrect (I) otherwise.

Section 2 [part (a), steps 2 and 3] is scored as follows:

Essentially correct (E) if the student identifies a correct test (by name or formula) and includes correct mechanics.

Incorrect (I) otherwise.

Section 3 [part (a), step 4] is scored as follows:

Essentially correct (E) if the student states a correct conclusion in the context of the problem.

Incorrect (I) otherwise.

Section 4 [part (b)] is scored as follows:

Essentially correct (E) if the student finds the correct simulated p-value from the dotplot or states that the actual standard deviation of 0.085 would be unusual if $\sigma = 0.05$ because 0.085 lies in the tail of the distribution *AND* an appropriate conclusion is made.

Partially correct (P) if only the simulated p-value or statement that 0.085 lies in the tail is correct, but the conclusion is weak, wrong, or missing.

Incorrect if the value 0.085 is not linked to the distribution.

Notes:
- If the p-value in section 2 is incorrect but the conclusion is consistent with the computed p-value, section 4 can be considered correct.
- In section 4, if both an α and a p-value are given together, the linkage between the p-value and the conclusion is implied. If no α is given, the solution must be explicit about the linkage by giving a correct interpretation of the p-value or explaining how the conclusion follows from the p-value.

Question 5 (continued)

Each essentially correct (E) response counts as 1 point, and a partially correct (P) response in section 4 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, 1½ points), use a holistic approach to determine whether to score up or down, depending on the strength of the response and communication.

Question 6

Intent of Question

The primary goals of this investigative task were to assess students' ability to (1) evaluate whether a double-blind randomized experiment is possible, (2) construct and interpret a confidence interval for the difference between two proportions, (3) estimate a relative risk and construct and interpret a confidence interval for a relative risk, and (4) compare the confidence interval for the difference between two proportions and the confidence interval for the relative risk.

Solution

Part (a):

Yes, it can. An experiment is said to be double-blind when the control or comparison group and the treatment group are treated exactly alike, except for the treatments of interest, so that neither the patients nor the researchers who care for the patients and take measurements know which treatment has been assigned to any patient. Because treatment A must be administered by pill and treatment B must be administered by injection, then a placebo could be introduced for each treatment group as follows to make the study double-blind:

The patients who get treatment A also get a placebo injection. The patients who get treatment B also get a placebo pill. The physician does not have to know which treatment the patient is receiving if the treatment plus placebo is sent to him or her by a researcher not in contact with the patients (this researcher is the only one who knows which treatment each patient is receiving).

Part (b):

Step 1: Identify the appropriate confidence interval by name or formula and check appropriate conditions. (The question states that the conditions for inference have been met.)

Two-sample z confidence interval for the difference of two proportions.
   OR
The formula for the confidence interval given in step 2 is provided.

Step 2: Correct mechanics.

Let $\hat{p}_A = \dfrac{38}{154} = 0.2468 \cong 0.25$ and $\hat{p}_B = \dfrac{16}{164} = 0.0976 \cong 0.10$. Then, the 95 percent confidence interval is

$$(\hat{p}_A - \hat{p}_B) \pm z^* \sqrt{\frac{\hat{p}_A(1 - \hat{p}_A)}{n_A} + \frac{\hat{p}_B(1 - \hat{p}_B)}{n_B}} = \left(\frac{38}{154} - \frac{16}{164}\right) \pm 1.96\sqrt{\frac{\frac{38}{154}\left(1 - \frac{38}{154}\right)}{154} + \frac{\frac{16}{164}\left(1 - \frac{16}{164}\right)}{164}}$$

$$= 0.1492 \pm 1.96 \times 0.0418$$

$$= 0.1492 \pm 0.0818$$

$$(0.0674, 0.2310)$$

Step 3: Interpretation.

At the 95 percent confidence level, the range of plausible values for the true difference in the 15-year survival rates for the two treatments is 0.0674 to 0.2310. Because this interval is entirely above zero, this suggests that the 15-year survival rate is higher for treatment A.

Part (c):

The estimate of the relative risk is

$$\frac{\hat{p}_A}{\hat{p}_B} = \frac{\frac{38}{154}}{\frac{16}{164}} \approx 2.5292 \approx 2.53$$

Part (d):

A 95 percent confidence interval for the relative risk is found by evaluating

$e^{0.3868}$ to $e^{1.469}$, which is 1.47 to 4.34.

At the 95 percent confidence level, the range of plausible values for the relative risk is 1.47 to 4.34. People are between 1.47 and 4.34 times more likely to survive 15 or more years with treatment A than with treatment B.

Part (e):

When the proportions of people who survive are low, as is the case with 0.25 and 0.10, it may be more meaningful or vivid to know that a patient's chance of survival with treatment A is 1.47 to 4.34 times what it would be with treatment B rather than to know that the difference in the proportions of people who survive is 0.07 to 0.23, which does not sound like very much.

Scoring

This problem is scored in four sections. Section 1 consists of part (a). Section 2 consists of part (b). Section 3 consists of part (c) and part (d). Section 4 consists of part (e). Sections 1, 2, 3, and 4 are each scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 [part (a)] is scored as follows:

Essentially correct (E) if the student (1) communicates what is meant by a double-blind study and (2) clearly describes how this experiment can be performed as a double-blind experiment.

Partially correct (P) if only one component is correct.

Incorrect (I) otherwise.

Section 2 [part (b)] is scored as follows:

Essentially correct (E) if the correct confidence interval is identified and constructed *AND* interpreted in context.

Partially correct (P) if the correct confidence interval is identified and constructed, but the interpretation is not in context or no interpretation is given.
>   *OR*
There are calculation errors with the appropriate confidence interval, but the interpretation follows correctly from the interval and is in context.

Incorrect (I) if the confidence interval and interpretation are not reasonable.

Section 3 [part (c) and part (d)] is scored as follows:

Essentially correct (E) if the student correctly (1) sets up and calculates the estimated relative risk, (2) sets up and calculates the confidence interval for the relative risk, and (3) interprets the confidence interval in context.

Partially correct (P) if the student provides only two correct components.

Incorrect (I) otherwise.

Section 4 [part (e)] is scored as follows:

Essentially correct (E) if the student (1) provides an advantage of using the confidence interval for the relative risk and (2) addresses both intervals.

Partially correct (P) if the first component is correct.

Incorrect (I) otherwise.

Each essentially correct (E) response counts as 1 point, and each partially correct (P) response 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 determine whether to score up or down, depending on the strength of the response and communication.