AP® STATISTICS
2009 SCORING GUIDELINES

Question 5

Intent of Question

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

Solution

Part (a):

The *p*-value of 0.0761 measures the chance of observing a difference between the two sample proportions $\left(\hat{p}_{CC} - \hat{p}_{CC+MMR}\right)$ as large as or larger than the one observed, if the survival rates for the two treatments (CC alone and CC + MMR) are in fact the same.

Part (b):

Because the *p*-value of 0.0761 is greater than 0.05, the null hypothesis should not be rejected. That is, there is not sufficient evidence to conclude that the treatment "CC alone" produces a higher survival rate than the standard treatment "CC + MMR."

Part (c):

Because the null hypothesis was not rejected, a Type II error could have occurred. A possible consequence is that CC + MMR would continue as the accepted practice when, in fact, CC alone would result in a higher survival rate.

Scoring

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

Part (a) is scored as follows:

A correct interpretation must include the following three components:

- Correct probability phrase (e.g., "The *p*-value of 0.0761 measures the chance of . . .") that includes the words "as large as" (or something similar).
- Correct conditional phrase (e.g., "if the survival rates for the two treatments are in fact the same").
- Correct context.

Essentially correct (E) if the response includes all three components.

Partially correct (P) if the response includes the first component and one of the other two components
OR
if the probability phrase is complete except for the omission of the words "as large as" (or something similar) and the other two components are included.

Incorrect (I) if the response includes no more than one component.

AP® STATISTICS
2009 SCORING GUIDELINES

Question 5 (continued)

Part (b) is scored as follows:

Essentially correct (E) if a correct conclusion (failure to reject H_0) is provided in context with appropriate linkage to the *p*-value.

Partially correct (P) if a correct conclusion is provided but either the context or linkage is missing
OR
the student "accepts H_0" (or something similar) and provides both context and linkage.

Incorrect (I) if the student rejects H_0
OR
the student provides neither context nor linkage.

Part (c) is scored as follows:

A correct response must include the following two components:
- The type of error named is consistent with the conclusion in part (b).
- A consequence is provided (in context) that is consistent with the conclusion in part (b) and is specific with regard to the treatments. The consequence must address how heart attack patients will be treated as a result of the error.

Essentially correct (E) if the response includes both components.

Partially correct (P) if the response includes only the consequence component
OR
the type of error named is consistent with the conclusion in part (b) *AND* a correct definition (either generic or in context) of that error is given, but the consequence component is either missing or incorrect.

Incorrect (I) if the response does not include the consequence component, apart from the exception given above as the second type of a partially correct response.

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

AP® STATISTICS
2009 SCORING GUIDELINES

Question 5 (continued)

1 **Minimal Response**

One part essentially correct and no parts partially correct

OR

No parts essentially correct and two parts partially correct

5A1

5. For many years, the medically accepted practice of giving aid to a person experiencing a heart attack was to have the person who placed the emergency call administer chest compression (CC) plus standard mouth-to-mouth resuscitation (MMR) to the heart attack patient until the emergency response team arrived. However, some researchers believed that CC alone would be a more effective approach.

In the 1990s a study was conducted in Seattle in which 518 cases were randomly assigned to treatments: 278 to CC plus standard MMR and 240 to CC alone. A total of 64 patients survived the heart attack: 29 in the group receiving CC plus standard MMR, and 35 in the group receiving CC alone. A test of significance was conducted on the following hypotheses.

H_0: The survival rates for the two treatments are equal.

H_a: The treatment that uses CC alone produces a higher survival rate.

This test resulted in a p-value of 0.0761.

(a) Interpret what this p-value measures in the context of this study.

In this study, the probability of observing data as extreme as what was collected, assuming the survival rates for the two treatments are equal, is .0761.

(b) Based on this p-value and study design, what conclusion should be drawn in the context of this study? Use a significance level of $\alpha = 0.05$.

With $\alpha = .05$, p-value $> \alpha$. As a result, the null hypothesis should not be rejected. The data do not suggest the treatment that uses CC alone produces a higher survival rate.

GO ON TO THE NEXT PAGE.

If you need more room for your work for part (b), use the space below.

(c) Based on your conclusion in part (b), which type of error, Type I or Type II, could have been made? What is one potential consequence of this error?

Type II error could have been made, in which the null is not true, however it is not rejected. In such a case, people will ~~only~~ aid by giving MMRs and CCs. when CCs would be more effective. Less patients would survive heart attacks compared to if this error was not made.

GO ON TO THE NEXT PAGE.

5. For many years, the medically accepted practice of giving aid to a person experiencing a heart attack was to have the person who placed the emergency call administer chest compression (CC) plus standard mouth-to-mouth resuscitation (MMR) to the heart attack patient until the emergency response team arrived. However, some researchers believed that CC alone would be a more effective approach.

In the 1990s a study was conducted in Seattle in which 518 cases were randomly assigned to treatments: 278 to CC plus standard MMR and 240 to CC alone. A total of 64 patients survived the heart attack: 29 in the group receiving CC plus standard MMR, and 35 in the group receiving CC alone. A test of significance was conducted on the following hypotheses.

H_0: The survival rates for the two treatments are equal.

H_a: The treatment that uses CC alone produces a higher survival rate.

This test resulted in a p-value of 0.0761.

(a) Interpret what this p-value measures in the context of this study.

This means that the probability of receiving a test statistic (in this case, ~~Z~~) that great or greater was .0761. It measures the probability that the difference in the sample ~~proportions~~ could have resulted assuming the null hypothesis, that the proportions were equal, is true.

(b) Based on this p-value and study design, what conclusion should be drawn in the context of this study? Use a significance level of $\alpha = 0.05$.

There is not sufficient evidence to reject the null hypothesis H_0, because the p-value is greater than α at .05. Cannot reject the hypothesis that the survival rates for the two treatments are equal.

GO ON TO THE NEXT PAGE.

If you need more room for your work for part (b), use the space below.

(c) Based on your conclusion in part (b), which type of error, Type I or Type II, could have been made? What is one potential consequence of this error?

Since we did not reject, a Type II error could have been made - meaning that, in reality, the alternative Hypothesis was indeed true and the null hypothesis should have been rejected. A potential consequence of this error is assuming that, since there is, according to the conclusion no difference between the treatments, people should use either when in fact it would be better to use CC alone.

5C1

5. For many years, the medically accepted practice of giving aid to a person experiencing a heart attack was to have the person who placed the emergency call administer chest compression (CC) plus standard mouth-to-mouth resuscitation (MMR) to the heart attack patient until the emergency response team arrived. However, some researchers believed that CC alone would be a more effective approach.

In the 1990s a study was conducted in Seattle in which 518 cases were randomly assigned to treatments: 278 to CC plus standard MMR and 240 to CC alone. A total of 64 patients survived the heart attack: 29 in the group receiving CC plus standard MMR, and 35 in the group receiving CC alone. A test of significance was conducted on the following hypotheses.

H_0: The survival rates for the two treatments are equal.

H_a: The treatment that uses CC alone produces a higher survival rate.

This test resulted in a *p*-value of 0.0761.

(a) Interpret what this *p*-value measures in the context of this study.

This p-value measures the probability of the difference of the number of people who survived that used the two different treatments.

(b) Based on this *p*-value and study design, what conclusion should be drawn in the context of this study? Use a significance level of $\alpha = 0.05$.

At $\alpha = 0.05$, do not reject Ho. There is not significant evidence that the treatment that uses CC alone produces a higher survival rate.

GO ON TO THE NEXT PAGE.

If you need more room for your work for part (b), use the space below.

(c) Based on your conclusion in part (b), which type of error, Type I or Type II, could have been made? What is one potential consequence of this error?

Based on the conclusion in part (b) that Ho should not be rejected, a Type II error could have been made - meaning that Ho was not rejected when it should have been. One potential consequence of this error is that the conclusion of the results, that the treatment that uses CC alone does not produce a higher survival rate would be wrong and therefore we would not be treating heart attack patients in the most effective way possible.

AP® STATISTICS
2009 SCORING COMMENTARY

Question 5

Overview

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

Sample: 5A
Score: 4

The response to part (a) includes a correct probability phrase ("the probability of observing data as extreme as what was collected") and a correct conditional phrase ("assuming the survival rates for the two treatments are equal"), stated in the context of the problem. Part (a) was scored as essentially correct. In part (b) the response includes a correct conclusion in context, with appropriate linkage to the *p*-value. Part (b) was scored as essentially correct. The response to part (c) names an error that is consistent with part (b) and provides a correct definition of that error, although the definition is not required. Part (c) provides a consequence, in context, that is consistent with part (b), is specific with regard to the treatments, and addresses how patients would be treated as a result of the error. Part (c) was scored as essentially correct. The entire answer, based on all three parts, was judged a complete response and earned a score of 4 points.

Sample: 5B
Score: 3

The response to part (a) includes a correct probability phrase ("the probability of receiving a test statistic . . . that great or greater") and a correct conditional phrase ("assuming the null hypothesis . . . is true"). However, neither of these two components is stated in the context of the problem. Consequently, part (a) was scored as partially correct. In part (b) the response includes a correct conclusion in context, with appropriate linkage to the *p*-value. Part (b) was scored as essentially correct. The response to part (c) names an error that is consistent with part (b) and provides a correct definition of that error, although the definition is not required. The student also provides a consequence, in context, that is consistent with part (b), is specific with regard to the treatments, and addresses how patients would be treated as a result of the error. Note that the response to part (c) includes a phrase that implies acceptance of the null hypothesis ("there is, according to the conclusion no difference between the treatments"). However, the student was not penalized for this statement because this part of the question focuses on the possible error made, rather than the conclusion of the test. Therefore part (c) was scored as essentially correct. With two parts essentially correct and one part partially correct, the entire answer was judged a substantial response and earned a score of 3 points.

Sample: 5C
Score: 2

The response to part (a) does not include a correct probability phrase; among other problems, there are no phrases such as *as large as* or *as extreme as*. Neither is there a correct conditional phrase. Part (a) was scored as incorrect. In part (b) the response includes a correct conclusion in context; however, there is no

Question 5 (continued)

linkage to the *p*-value. Part (b) was scored as partially correct. The response to part (c) names an error that is consistent with part (b) and provides a correct definition of that error, although the definition is not required. Part (c) provides a consequence, in context, that is consistent with part (b), is specific with regard to the treatments, and addresses how patients would be treated as a result of the error. Part (c) was scored as essentially correct. With one part essentially correct, one part partially correct, and one part incorrect, the entire answer was judged a developing response and earned a score of 2 points.