AP® STATISTICS
2012 SCORING GUIDELINES

Question 5

Intent of Question

The primary goals of this question were to assess students' ability to (1) describe a Type II error and its consequence in a particular study; (2) draw an appropriate conclusion from a *p*-value; (3) describe a flaw in a study and its effect on inference from a sample to a population.

Solution

Part (a):

In the context of the study, a Type II error means failing to reject the null hypothesis that 35 percent of adult residents in the city are able to pass the test when, in reality, less than 35 percent are able to pass the test. The consequence of this error is that the council would not fund the program, and the city would continue to have a smaller proportion of physically fit residents than the council would like.

Part (b):

Because the *p*-value of 0.97 is larger than $\alpha = 0.05$, we fail to reject the null hypothesis. There is not convincing evidence that the proportion of adult residents in the city who are able to pass the physical fitness test is less than 0.35. After all, the sample proportion of $\hat{p} = 0.416$ is actually higher than 0.35, which is in the opposite direction of the alternative hypothesis.

Part (c):

This is not a randomly selected sample because the sample was selected by recruiting volunteers. It seems reasonable to think that volunteers would be more physically fit than the population of city adults as a whole. Therefore, the sample proportion will likely overestimate the population proportion of adult residents in the city who are able to pass the physical fitness test.

Scoring

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

Part (a) is scored as follows:

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

1. Describes the error in context by referring to the proportion of adult residents in the city who are able to pass the physical fitness test.
2. Describes the consequence as not funding the program and/or continuing poor physical fitness of the adult residents in the city.

Question 5 (continued)

Notes

- If a response provides more than one description of a Type II error, score the weakest attempt.
- Referring to the symbolic hypotheses is *not sufficient* for context.
- Referring to funding and/or the city council is *not sufficient* for context.
- If a response describes a Type II error incorrectly, the response can get the consequence component correct if it is consistent with the incorrectly described error.
- If a response provides more than one description of a Type II error, the response can get the consequence component correct if the consequence is clearly linked to one of the error descriptions and is consistent with the error to which it is linked.
- If a response gives an incomplete description of a Type II error (for example, "we fail to reject the null hypothesis that the proportion of adult residents who are able to pass is 0.35"), the response can get the consequence component correct if the consequence is consistent with the partial description of the error.
- If a response provides no description of a Type II error, the response cannot get the consequence component correct.

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

Incorrect (I) if the response correctly completes neither of the two components listed above.

Note: Describing the Type II error only in terms of the consequence (for example, "They don't fund the program when they should") should get credit for the consequence but should not get credit for the error, because there is no reference to the proportion of adult residents in the city who are able to pass the test.

Part (b) is scored as follows:

Essentially correct (E) if the response correctly completes the following three components:

1. Links the p-value to the conclusion by stating that the p-value is greater than $\alpha = 0.05$,
   OR
   by stating that the p-value is large,
   OR
   by correctly interpreting the p-value.
2. Uses context by referring to the proportion of adult residents who are able to pass the test,
   OR
   by referring to the funding of the program.
3. Makes a correct conclusion that describes the lack of evidence for the alternative hypothesis ($H_a : p < 0.35$).

Notes

- If a response includes an incorrect interpretation of the p-value, then the response cannot earn credit for the linkage component, even if the response explicitly compares the p-value to α or describes the p-value as large.
- Referring to the symbolic hypotheses is not sufficient for context.
- Accepting the null hypothesis or some equivalent statement such as "the population proportion is (or is likely to be, or is about) 0.35" cannot receive credit for the conclusion component, even if the student makes additional correct statements about the alternative hypothesis.

Question 5 (continued)

- Stating that the null hypothesis should not be rejected is not sufficient for the conclusion, because it does not address the direction of H_a.
- Correctly addressing the consequence ("They don't fund the program") is sufficient if the response also indicates that the null hypothesis is not being rejected.
- Drawing a conclusion about the sample proportion (for example, "proportion who passed the test") is not sufficient for the conclusion, because it does not properly address the parameter in H_a.

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

Incorrect (I) if the response correctly completes one or none of the three components listed above.

Notes

- A response that says the *p*-value is very large, recognizes that the sample proportion $\left(\hat{p} = 0.416\right)$ is greater than $p = 0.35$, and consequently concludes there is no evidence to support H_a (in context) is scored as essentially correct (E).
- A response that *rejects* H_0 is scored as incorrect (I).

Part (c) is scored as follows:

Essentially correct (E) if the response correctly completes the following three components:

1. States that the sample is not random and/or says that volunteers were used.
2. Describes *how* the sample is "different" with regard to physical fitness or another variable related to the ability to pass the physical fitness test.
3. Addresses the idea of making an inference from the sample to the population by stating that the sample statistic will overestimate the population parameter or that the sample will not be representative of the population.

Notes

- If for the first component a student provides additional proposed flaws (for example, "the sample size is too small"), score the weakest attempt.
- Saying only that the sample is different or not representative does not address *how* the sample is different.
- Saying "physically fit people will be overrepresented" or "the results cannot be generalized" or "the results will be inaccurate" lack a specific reference to the population and is not sufficient for the third component.
- Referring to "bias" is not sufficient for the first component unless the concept of bias is clearly explained (for example, saying that the sample proportion will tend to overestimate the population proportion).
- Incorrect application of statistical concepts (for example, saying that the statistic is "skewed," discussing cause and effect) results in a loss of credit for the third component.

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

Incorrect (I) if the response correctly addresses one or none of the three components listed above.

AP® STATISTICS
2012 SCORING GUIDELINES

Question 5 (continued)

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

SA SA

5. A recent report stated that less than 35 percent of the adult residents in a certain city will be able to pass a physical fitness test. Consequently, the city's Recreation Department is trying to convince the City Council to fund more physical fitness programs. The council is facing budget constraints and is skeptical of the report. The council will fund more physical fitness programs only if the Recreation Department can provide convincing evidence that the report is true.

The Recreation Department plans to collect data from a sample of 185 adult residents in the city. A test of significance will be conducted at a significance level of $\alpha = 0.05$ for the following hypotheses.

$$\begin{cases} H_0 : p = 0.35 \\ H_a : p < 0.35, \end{cases}$$

where p is the proportion of adult residents in the city who are able to pass the physical fitness test.

(a) Describe what a Type II error would be in the context of the study, and also describe a consequence of making this type of error.

A Type II error would be failing to reject H_0, when it is actually false. The Recreation Department would state that 35% of adult residents can pass a physical fitness test, when actually the percentage is less than 35%.

A potential consequence is that no more funding would go to physical fitness programs, when in actuality those programs need the funding.

(b) The Recreation Department recruits 185 adult residents who volunteer to take the physical fitness test. The test is passed by 77 of the 185 volunteers, resulting in a p-value of 0.97 for the hypotheses stated above. If it was reasonable to conduct a test of significance for the hypotheses stated above using the data collected from the 185 volunteers, what would the p-value of 0.97 lead you to conclude?

since the p-value is greater than α, we would fail to reject the null hypothesis, There is insufficient evidence to suggest the percentage of adult residents able to pass a physical fitness test is less than 35%.

(c) Describe the primary flaw in the study described in part (b), and explain why it is a concern.

The primary flaw in the study is bias. The recreation department recruits 185 adults who volunteer to take the physical fitness test. This creates a volunteer-response bias. People who think they will pass the physical fitness test are more likely to volunteer than those who think they would fail. Therefore, the data on adults who passed the fitness test obtained by the Recreation Department is potentially larger than the true proportion of adults in the city that pass.

GO ON TO THE NEXT PAGE.

5B 5B

5. A recent report stated that less than 35 percent of the adult residents in a certain city will be able to pass a physical fitness test. Consequently, the city's Recreation Department is trying to convince the City Council to fund more physical fitness programs. The council is facing budget constraints and is skeptical of the report. The council will fund more physical fitness programs only if the Recreation Department can provide convincing evidence that the report is true.

The Recreation Department plans to collect data from a sample of 185 adult residents in the city. A test of significance will be conducted at a significance level of $\alpha = 0.05$ for the following hypotheses.

$$H_0 : p = 0.35$$
$$H_a : p < 0.35,$$

where p is the proportion of adult residents in the city who are able to pass the physical fitness test.

(a) Describe what a Type II error would be in the context of the study, and also describe a consequence of making this type of error.

A Type II error is when the null failed to be rejected when it was actually false and the alternate hypothesis was true. This means than less than 35% of the city's adults can pass physical tests but the Dept. lacks strong evidence to prove it and the Council would not provide physical fitness program funding and many adult residents in the city would stay unfit.

(b) The Recreation Department recruits 185 adult residents who volunteer to take the physical fitness test. The test is passed by 77 of the 185 volunteers, resulting in a p-value of 0.97 for the hypotheses stated above. If it was reasonable to conduct a test of significance for the hypotheses stated above using the data collected from the 185 volunteers, what would the p-value of 0.97 lead you to conclude?

A p-value of 0.97 is higher than any reasonable α-value, so the null hypothesis that the true proportion of adults in the city unable to pass the test is 35% would fail to be rejected, and the Council would not provide physical fitness program funding.

(c) Describe the primary flaw in the study described in part (b), and explain why it is a concern.

~~No response~~ The sample used only volunteers, and adults that were physically unfit may have been too embarrassed to volunteer while physically fit ones may have been confident in volunteering.

GO ON TO THE NEXT PAGE.

5C 5C

5. A recent report stated that less than 35 percent of the adult residents in a certain city will be able to pass a physical fitness test. Consequently, the city's Recreation Department is trying to convince the City Council to fund more physical fitness programs. The council is facing budget constraints and is skeptical of the report. The council will fund more physical fitness programs only if the Recreation Department can provide convincing evidence that the report is true.

The Recreation Department plans to collect data from a sample of 185 adult residents in the city. A test of significance will be conducted at a significance level of $\alpha = 0.05$ for the following hypotheses.

$$H_0 : p = 0.35$$
$$H_a : p < 0.35,$$

where p is the proportion of adult residents in the city who are able to pass the physical fitness test.

(a) Describe what a Type II error would be in the context of the study, and also describe a consequence of making this type of error.

A Type II error would be failing to reject the null hypothesis when there is actually enough evidence to reject it. If this occurred, the [redacted] City Council would not believe the report about less than 35% of adults passing a fitness test was true, so they would NOT fund more physical fitness programs. As a result, the adults of the city would have less access to physical fitness programs and will not be in as good shape.

(b) The Recreation Department recruits 185 adult residents who volunteer to take the physical fitness test. The test is passed by 77 of the 185 volunteers, resulting in a p-value of 0.97 for the hypotheses stated above. If it was reasonable to conduct a test of significance for the hypotheses stated above using the data collected from the 185 volunteers, what would the p-value of 0.97 lead you to conclude?

The p value of .97 would lead us to fail to reject the null hypothesis, and conclude that 35% of the adult residents in the city CAN pass a physical fitness test.

(c) Describe the primary flaw in the study described in part (b), and explain why it is a concern.

The primary flaw in the study above is [redacted] voluntary response. People who are in better physical condition and know they can pass a physical fitness test will be more likely to volunteer for the test. Therefore, the sample in part (b) is not an accurate representation of all adult residents in the city; it is more biased towards the more physically fit residents.

GO ON TO THE NEXT PAGE.

AP® STATISTICS
2012 SCORING COMMENTARY

Question 5

Overview

The primary goals of this question were to assess students' ability to (1) describe a Type II error and its consequence in a particular study; (2) draw an appropriate conclusion from a *p*-value; (3) describe a flaw in a study and its effect on inference from a sample to a population.

Sample: 5A
Score: 4

In part (a) the response begins with a generic definition of a Type II error that lacks context. However, the next sentence clearly describes the Type II error in context by referring to the percentage of adult residents who can pass the physical fitness test. The student also nicely addresses the alternative hypothesis ("when actually the percentage is less than 35%") rather than just saying that the null hypothesis is not true, and then describes a correct consequence of making a Type II error. Because the response includes both required components, part (a) was scored as essentially correct. In part (b) the response provides linkage between the *p*-value and the conclusion by explicitly comparing the *p*-value to α. The student also correctly addresses the lack of evidence for the alternative hypothesis and does so in context. Because the response includes all three required components, part (b) was scored as essentially correct. In part (c) the student states that using volunteers causes a bias because "[p]eople who think they will pass the physical fitness test are more likely to volunteer." The response also clearly explains why this is a concern by stating that the proportion of residents able to pass in the sample "is potentially larger than the true proportion of adults in the city that [are able to] pass." Because the response includes all three required components, part (c) was scored as essentially correct. With all three parts scored as essentially correct, the response earned a score of 4.

Sample: 5B
Score: 3

In part (a) the response begins with a generic definition of a Type II error that lacks context. However, the next sentence describes the Type II error in context, talking about the proportion of the city's adults who can pass physical tests. Finally, the student provides two correct consequences for a Type II error, although only one is necessary. Because the response includes both required components, part (a) was scored as essentially correct. In part (b) the response provides linkage between the *p*-value and the conclusion by stating that a "*p*-value of 0.97 is higher than any reasonable α value." The student also provides context by referring to "the true proportion of adults in the city unable to pass the test." The misstatement of "unable" to pass instead of "able" to pass was viewed as a minor error in the context of the entire response. Finally, the student also states that the null hypothesis would fail to be rejected and that the council would not provide funding. Referring to the lack of funding is sufficient for addressing the alternative hypothesis, because it is clear that the student is not accepting the null hypothesis. Because the response includes all three required components, part (b) was scored as essentially correct. In part (c) the student states that using volunteers is a flaw because "physically fit ones may have been confident in volunteering." However, the response does not address the concern that it would be inappropriate to use this sample to make an inference about the population. Because the response includes two of the three required components, part (c) was scored as partially correct. With two parts scored as essentially correct and one part scored as partially correct, the response earned a score of 3.

AP® STATISTICS
2012 SCORING COMMENTARY

Question 5 (continued)

Sample: 5C
Score: 2

In part (a) the student does not correctly describe a Type II error and instead makes the common mistake of defining it as "failing to reject the null hypothesis when there is actually enough evidence to reject it." However, the response did receive credit for the consequence component, because the consequence is consistent with the decision to fail to reject the null hypothesis. Because the response includes one of the two required components, part (a) was scored as partially correct. In part (b) the response does not provide sufficient linkage by stating that the *p*-value is large or that the *p*-value is larger than α. Although the conclusion is stated in context, the student incorrectly accepts the null hypothesis by concluding that "35% of the adult residents in the city CAN pass a physical fitness test." Because the response includes only one of the three required components, part (b) was scored as incorrect. In part (c) the response indicates that the "flaw in the study … is voluntary response" and that "[p]eople who are in better physical condition … will be more likely to volunteer for the test." Furthermore, the student says that the sample "is not an accurate representation of all adult residents in the city," clearly addressing the idea of making an inference from the sample to the population. Because the response includes all three required components, part (c) was scored as essentially correct. With one part scored as essentially correct, one part scored as partially correct, and one part scored as incorrect, the response earned a score of 2.