2018

AP Statistics

Sample Student Responses and Scoring Commentary

Inside:

Free Response Question 4

☑ Scoring Guideline

☑ Student Samples

☑ Scoring Commentary

© 2018 The College Board. College Board, Advanced Placement Program, AP, AP Central, and the acorn logo are registered trademarks of the College Board. Visit the College Board on the Web: www.collegeboard.org.

AP Central is the official online home for the AP Program: apcentral.collegeboard.org

AP® STATISTICS
2018 SCORING GUIDELINES

Question 4

Intent of Question

The primary goals of this question were to assess a student's ability to (1) determine whether a cause-and-effect conclusion can be made based on how a study was conducted and (2) set up, perform, and interpret the results of a hypothesis test, in the context of the problem.

Solution

Part (a):

Yes, it would be reasonable to conclude that the new procedure causes a reduction in recovery time, for patients similar to those in the study. The patients in the study were randomly assigned to the two procedures, which reduces the chance that confounding variables will affect the results. Therefore the statistically significant reduction in mean recovery time can be attributed to the new procedure being superior to the standard procedure.

Part (b):

Step 1: State a correct pair of hypotheses.

Let μ_S represent the mean recovery time among all patients similar to those in the study if they were to receive the standard treatment.

Let μ_N represent the mean recovery time among all patients similar to those in the study if they were to receive the new treatment.

The hypotheses to be tested are $\text{H}_0 : \mu_S = \mu_N$ versus $\text{H}_\text{a} : \mu_S > \mu_N$.

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

The appropriate procedure is a two-sample *t*-test for a difference between means.

Because this is an experiment, the first condition is that subjects were randomly assigned to one treatment group or the other. In this case the condition is satisfied because we were told that the subjects were randomly assigned to either the standard or new procedure.

The second condition is that the recovery times of the two populations are normally distributed or the sample sizes are sufficiently large to presume that the distribution of the difference in the sample means is approximately normal. In this case the condition is met because the sample sizes of 110 and 100 are both sufficiently large.

© 2018 The College Board.

Question 4 (continued)

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

The test statistic is $t = \dfrac{\bar{x}_S - \bar{x}_N}{\sqrt{\dfrac{s_S^2}{n_S} + \dfrac{s_N^2}{n_N}}} = \dfrac{217 - 186}{\sqrt{\dfrac{34^2}{110} + \dfrac{29^2}{100}}} \approx 7.13$.

The p-value is the area greater than 7.13 for a t-distribution with $df = 207.18$, which is essentially 0 $\left(8.36 \times 10^{-12}\right)$.

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

Because the p-value is very small, we have sufficient evidence to conclude that for patients similar to the ones in the study, those receiving the new procedure would have less recovery time, on average, than those receiving the standard procedure.

Scoring

This question is scored in three sections. Section 1 consists of part (a); section 2 consists of step 1, step 2, and the test statistic in step 3 in part (b); and section 3 consists of the p-value in step 3 and step 4 in part (b). Sections 1, 2, and 3 are each scored essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

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

1. Correctly states that it is reasonable to make a causal conclusion.
2. Justifies the causal conclusion based on random assignment of patients to procedures (or procedures to patients);

   OR

   justifies the causal conclusion by stating that a randomized experiment was conducted.
3. Includes the context of the situation.

Partially correct (P) if the response satisfies component 1 *AND* provides WEAK justification of the causal conclusion by stating that there was random assignment or a randomized experiment was conducted, but with no context;

OR

by stating that an experiment was conducted or there was assignment (without the word "randomized") *AND* the response includes context of the situation;

OR

by stating that the study design reduces the chance of confounding variables or balances the effects of uncontrolled variables across both groups in context without explicitly referring to the random assignment.

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

Question 4 (continued)

Notes:

- If the response states that it is *not* reasonable to make a causal conclusion because the result could have been due to random chance *AND* explains that there is evidence for a causal conclusion based on random assignment of patients to procedures or by stating that a randomized experiment was conducted, then the response is scored E.
- If the response discusses aspects of an experiment other than random assignment (such as, control, replication, or large samples), then those aspects are considered extraneous and the response can be scored E unless those aspects are incorrect for this study (such as, blocking is a requirement, or the study used blocking, or the study used a placebo) in which case the score should be lowered one level (that is, from E to P, or from P to I).
- If the response correctly states in context that it is reasonable to make a causal conclusion but includes incorrect or contradictory justification (such as, random selection of patients), then the response is scored I.

Section 2 is scored as follows:

Essentially correct (E) if the response satisfies the following four components:

1. Parameters are defined correctly.
2. Hypotheses imply equality in the null and correct direction in the alternative.
3. Correct test is identified by name or formula.
4. Correct test statistic for a difference in means is calculated.

Partially correct (P) if the response satisfies only two or three of the four components.

Incorrect (I) if the response satisfies at most one of the four components.

Notes:

- If standard symbols are used for the parameters with appropriate group labels (such as, μ_S, μ_N), component 1 is satisfied.
- If the correct test is identified, but the response states an incorrect formula or uses incorrect notation in the formula, component 3 is not satisfied.
- A pooled two-sample *t*-test is acceptable for component 3, but the student must also state and comment on the plausibility of the equal population variances assumption.
- If the response identifies a *z*-test for equal means as the correct test identification, component 3 is not satisfied but component 4 could be satisfied.

Confidence Interval approach:

- If a single two-sample *t*-interval for the difference in means is used, components 3 and 4 can be satisfied. Component 3 is satisfied if the *t*-interval is correctly identified by name or formula. Component 4 is satisfied if the correct interval is calculated. If an alpha level is stated, then an appropriate adjustment to the confidence level must be made because the appropriate test is one-sided.
- If two one-sample *t*-intervals are used, while not a recommended approach, component 3 is not satisfied but component 4 could be satisfied. Component 4 is satisfied if both intervals are calculated correctly.

© 2018 The College Board.

Question 4 (continued)

Section 3 is scored as follows:

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

1. Makes reference to an approximately correct *p*-value that is consistent with the test statistic and alternative hypothesis for a difference in means.
2. Correctly justifies the conclusion based on the size of the *p*-value or the test statistic.
3. Correctly states the conclusion in context.

Partially correct (P) if the response satisfies only two of the three components.

Incorrect (I) if the response does not meet the criteria for E or P or includes a justification not based on the inferential results.

Notes:

Component 1:

- Is satisfied if the response makes reference to a large test statistic without referring to a *p*-value.

Component 2:

- No alpha level is needed to provide justification of the conclusion based on the size of the *p*-value.
- Is satisfied if the response states the *p*-value without reference to size, but it is contiguous to the conclusion and clearly indicates a continuous train of thought.
- A correct interpretation of the *p*-value with a complete explanation that obtaining a test statistic at least this extreme is unlikely due to chance alone is considered justification based on the size of the *p*-value.
- If an incorrect interpretation of the *p*-value is given, the score is lowered one level (that is, from E to P, or from P to I).
- A decision about the null hypothesis (reject H_0 or fail to reject H_0) is not required, but if an incorrect decision is stated based on the given *p*-value then component 2 is not satisfied.
- If a rejection region approach is used, a reasonable critical value replaces the *p*-value.

Component 3:

- A correct conclusion must be related to the alternative hypothesis in order to satisfy component 3.
- The following responses do not satisfy component 3:
  - States or implies that the null hypothesis is *accepted*
  - States or implies that the alternative hypothesis has been *proven*
  - States the conclusion in past tense (unless the response did not satisfy a component of section 2 for the use of past tense)

© 2018 The College Board.

Question 4 (continued)

Confidence Interval approach:

- If a single two-sample *t*-interval for the difference in means is used:
  - Component 1 is satisfied if the response indicates that zero is either included or not included in the calculated interval.
  - Component 2 is satisfied if the response indicates that the bounds are either both above or both below zero (consistent with alternative hypothesis) and uses that as justification for the conclusion.
  - Component 3 is satisfied if the conclusion is stated in context.
- If two one-sample *t*-intervals are used (which is not recommended) the response is scored at most P if all three components are satisfied, otherwise scored I:
  - Component 1 is satisfied if the response states that the intervals do not overlap.
  - Component 2 is satisfied if the conclusion indicates that the confidence interval for the new procedure lies below the confidence interval for the standard procedure.
  - Component 3 is satisfied if the conclusion is stated in context.

Note: If the three sections of the response are scored as E, to earn a score of 4 as a complete response, both conditions in step 2 must be correctly stated and justified. Additional condition(s) inappropriate for a two-sample *t*-test must not be stated. Otherwise, the response earns a score of 3 a substantial response.

© 2018 The College Board.
Visit the College Board on the Web: www.collegeboard.org.

Question 4 (continued)

4 **Complete Response**

Three sections essentially correct with conditions for inference

3 **Substantial Response**

Three sections essentially correct without conditions for inference

OR

Two sections essentially correct and one section partially correct

2 **Developing Response**

Two sections essentially correct and no sections partially correct

OR

One section essentially correct and one or two sections partially correct

OR

Three sections partially correct

1 **Minimal Response**

One section essentially correct

OR

No sections essentially correct and one or two sections partially correct

© 2018 The College Board.

4A1 4A1

4. The anterior cruciate ligament (ACL) is one of the ligaments that help stabilize the knee. Surgery is often recommended if the ACL is completely torn, and recovery time from the surgery can be lengthy. A medical center developed a new surgical procedure designed to reduce the average recovery time from the surgery. To test the effectiveness of the new procedure, a study was conducted in which 210 patients needing surgery to repair a torn ACL were randomly assigned to receive either the standard procedure or the new procedure.

(a) Based on the design of the study, would a statistically significant result allow the medical center to conclude that the new procedure causes a reduction in recovery time compared to the standard procedure, for patients similar to those in the study? Explain your answer.

Yes. Since the patients were randomly assigned to treatments in this experiment, a cause-and-effect relationship can be drawn and applied to patients similar to those in the study.

(b) Summary statistics on the recovery times from the surgery are shown in the table.

Type of Procedure	Sample Size	Mean Recovery Time (days)	Standard Deviation Recovery Time (days)
Standard	110	217	34
New	100	186	29

Do the data provide convincing statistical evidence that those who receive the new procedure will have less recovery time from the surgery, on average, than those who receive the standard procedure, for patients similar to those in the study?

State: $H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 > 0$

$\alpha = 0.05$

μ_1 is mean recovery time in days for the population of patients recieving the Standard procedure, μ_2 is mean recovery time in days for population of patients recieving the new procedure

Unauthorized copying or reuse of any part of this page is illegal.

GO ON TO THE NEXT PAGE.

© 2018 The College Board.
Visit the College Board on the Web: www.collegeboard.org.

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

Plan: If the conditions are met, we will use a 2-sample t-test for difference between means

Random - the treatments were randomly assigned

Normal/Large Sample - the sample size of both the standard and new procedure is greater than 30, so their sampling distributions are approximately Normal

Since this is a randomized experiment, the 10% condition doesn't need to be checked.

Do: $t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ $\bar{x}_1 = 217$ $s_1 = 34$ $n_1 = 110$
$\bar{x}_2 = 186$ $s_2 = 29$ $n_2 = 100$

$$t = \frac{217 - 186}{\sqrt{\frac{34^2}{110} + \frac{29^2}{100}}} = 7.127$$

$df = 100 - 1 = 99$

p-value = tcdf(lower = 7.127, upper = 1×10^99, df = 99).

p-value = $8.48 \times 10^{-11} \approx 0$

Conclude

Since the p-value of approximately 0 is less than the α of 0.05, we reject H_0 and have convincing evidence that patients who recieve the new procedure will have, on average, a shorter recovery time than those who recieve the standard procedure

© 2018 The College Board.
Visit the College Board on the Web: www.collegeboard.org.

4. The anterior cruciate ligament (ACL) is one of the ligaments that help stabilize the knee. Surgery is often recommended if the ACL is completely torn, and recovery time from the surgery can be lengthy. A medical center developed a new surgical procedure designed to reduce the average recovery time from the surgery. To test the effectiveness of the new procedure, a study was conducted in which 210 patients needing surgery to repair a torn ACL were randomly assigned to receive either the standard procedure or the new procedure.

(a) Based on the design of the study, would a statistically significant result allow the medical center to conclude that the new procedure causes a reduction in recovery time compared to the standard procedure, for patients similar to those in the study? Explain your answer.

Yes, because the patients were randomly assigned to treatment groups. Random assignment is a form of control that allows us to make inferences about cause and effect in a well-designed experiment because we hope it reduces any bias that is a result of confounding variables we didn't directly control, by creating groups that are roughly equivalent in terms of those variables.

(b) Summary statistics on the recovery times from the surgery are shown in the table.

Type of Procedure	Sample Size	Mean Recovery Time (days)	Standard Deviation Recovery Time (days)
Standard	110	217	34
New	100	186	29

N - O

Do the data provide convincing statistical evidence that those who receive the new procedure will have less recovery time from the surgery, on average, than those who receive the standard procedure, for patients similar to those in the study?

we want to test the following hypotheses at the $\alpha = 0.05$ significance level.

$H_0: \mu_{diff} = 0$ days

$H_a: \mu_{diff} < 0$ days

where μ_{diff} = mean difference in recovery times between those who receive the new procedure and those who receive the standard procedure ($\mu_{diff} = \mu_{new} - \mu_{standard}$)

we will conduct a two sample t-test if conditions are met.

Unauthorized copying or reuse of any part of this page is illegal.

GO ON TO THE NEXT PAGE.

© 2018 The College Board.
Visit the College Board on the Web: www.collegeboard.org.

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

-The normal condition is met by the central Limit Theorem because $n=110>30$ and $n=100>30$.

-The random condition is met as given because patients were randomly assigned to treatments.

-The independent condition is met because $10(210) = 2100 <$ total number of patients needing surgery to repair a torn ACL.

Test Statistic:

$$t = \frac{(186-217)-0}{\sqrt{\frac{(29)^2}{100}+\frac{(34)^2}{110}}} = -7.127$$

$t = -7.127$ $p = p(t \leq -7.127) = 8.36 \times 10^{-12}$

since the p-value of 8.36×10^{-12} is less than any reasonable α level, we reject Ho. The data provide convincing evidence that those who receive the new procedure will have less recovery time from the surgery. On average, than those who receive the standard procedure.

GO ON TO THE NEXT PAGE.

© 2018 The College Board.
Visit the College Board on the Web: www.collegeboard.org.

4. The anterior cruciate ligament (ACL) is one of the ligaments that help stabilize the knee. Surgery is often recommended if the ACL is completely torn, and recovery time from the surgery can be lengthy. A medical center developed a new surgical procedure designed to reduce the average recovery time from the surgery. To test the effectiveness of the new procedure, a study was conducted in which 210 patients needing surgery to repair a torn ACL were randomly assigned to receive either the standard procedure or the new procedure.

(a) Based on the design of the study, would a statistically significant result allow the medical center to conclude that the new procedure causes a reduction in recovery time compared to the standard procedure, for patients similar to those in the study? Explain your answer.

No, correlation doesn't prove causation. They could conclude that the new procedure will most likely reduce recovery time compared to the standard procedure, for patients similar to those in the study, but they can't prove that it causes a reduction in recovery time.

(b) Summary statistics on the recovery times from the surgery are shown in the table.

	Type of Procedure	Sample Size	Mean Recovery Time (days)	Standard Deviation Recovery Time (days)
	Standard	110	217	34
less	New	100	186	29

Do the data provide convincing statistical evidence that those who receive the new procedure will have less recovery time from the surgery, on average, than those who receive the standard procedure, for patients similar to those in the study?

Assumptions; SRS, independence, $n_1 p_1 \geq 10$, $n_1 q_1 \geq 10$, $n_2 P_2 \geq 10$, $n_2 q_2 \geq 10$
sample size < 10% population

H_0: $\mu_{new} = \mu_{standard}$

H_a: $\mu_{new} < \mu_{standard}$

$$P\left(z < \frac{186 - 217}{\sqrt{\frac{29^2}{100} + \frac{34^2}{110}}}\right)$$

$$P(z < -7.127)$$

$$p \approx 0$$

Unauthorized copying or reuse of any part of this page is illegal.

© 2018 The College Board.
Visit the College Board on the Web: www.collegeboard.org.

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

Since p is so small, we can reject H_o. This means that there is convincing statistical evidence to conclude that those who receive the new procedure will likely have less recovery time from the surgery, on average, than those who receive the standard procedure, for patients similar to those in the study.

© 2018 The College Board.
Visit the College Board on the Web: www.collegeboard.org.

AP® STATISTICS
2018 SCORING COMMENTARY

Question 4

Overview

The primary goals of this question were to assess a student's ability to (1) determine whether a cause-and-effect conclusion can be made based on how a study was conducted and (2) set up, perform, and interpret the results of a hypothesis test, in the context of the problem.

Sample: 4A
Score: 4

In part (a) component 1 of section 1 is satisfied because the response states that there is a causal relationship. Component 2 is satisfied because the causal relationship is justified based on the random assignment of the treatments. Component 3 is satisfied because the conclusion is given in context by using the word "patients." All three components of section 1 are satisfied. Section 1 was scored as essentially correct. In part (b) the parameters are correctly defined by using standard notation, μ, for a population mean. The generic subscripts 1 and 2 are used to denote the two different means; however, definitions are provided that clearly indicate that the subscript 1 refers to the standard procedure, and the subscript 2 refers to the new procedure. The definitions neglect to identify the parameters as the "population" mean, but this was overlooked in scoring because commonly accepted notation is used. Component 1 of section 2 is satisfied. Component 2 of section 2 is satisfied because the hypotheses indicate equality in the null and the correct direction in the alternative. Component 3 of section 2 is satisfied because the two-sample *t*-test is correctly identified by stating the correct formula for the test statistic, and the name of the test is given. Only the name or the formula is required to satisfy component 3. Component 4 of section 2 is satisfied because the correct test statistic, 7.127, is stated. All four components of section 2 are satisfied, and section 2 was scored as essentially correct. In part (b) the response makes reference to a correct *p*-value for a difference in means that is consistent with the test statistic and alternative hypothesis, satisfying component 1 of section 3. The response uses the small *p*-value to justify the conclusion that there is evidence to reject the null, satisfying component 2 of section 3. The response provides a correct conclusion by stating that there is evidence to support the alternative hypothesis, and this conclusion is stated in the context of the study. All three components of section 3 are satisfied, and section 3 was scored as essentially correct. Because the response was scored as essentially correct in three sections, the conditions of the test must be checked. Two conditions are required: the treatments must be randomly assigned, and the number of subjects in each treatment group must be large. The response clearly states that the treatments are randomly assigned. The response states that both samples are greater than 30; therefore, the "sampling distributions [of the difference in the sample means] are approximately normal." The response correctly notes that the condition that the sample size can be no larger than 10 percent of the population is unnecessary for this study because the study is a randomized experiment, not a random sample. Because three sections were scored as essentially correct, and conditions were stated and checked, the response earned a score of 4.

Sample: 4B
Score: 3

In part (a) component 1 of section 1 is satisfied because the response correctly states that it is reasonable to make a causal conclusion. Component 2 of section 1 is satisfied because the response justifies the causal relationship based on the random assignment of patients to procedures. Although not necessary, the response provides a very nice explanation of why the random assignment of patients to procedures reduces the possible effects of uncontrolled variables to allow for the conclusion of a causal relationship. Component 3 of section 1 is satisfied because the response is stated in context. All three components in section 1 are satisfied, and section 1 was scored as essentially correct. In part (b) component 1 of section 2 is not satisfied because the parameters are not correctly defined. The response defines the population mean for the difference in means for two dependent samples, which is not correct. Component 2 of section 2 is satisfied because the hypotheses indicate equality in the null and the correct

© 2018 The College Board.

Question 4 (continued)

direction in the alternative. The direction is able to be determined because the response stated that $\mu_{\text{D}} = \mu_{\text{new}} - \mu_{\text{standard}}$ and $\mu_{\text{D}} < 0,$ so it is clear that the alternative is testing that the mean for the new procedure is less than the mean for the standard procedure. Component 3 of section 2 is satisfied because the response states that a two-sample *t*-test will be conducted and uses the correct formula for the test statistic. Only the name or the formula is required to satisfy component 3. Component 4 of section 2 is satisfied because the response states the correct test statistic for a difference in means. All four components of section 2 are satisfied, and section 2 was scored as partially correct. In part (b) component 1 of section 3 is satisfied because the response provides a correct *p*-value that is consistent with the test statistic and the alternative hypothesis. The conclusion is justified based on the small size of the *p*-value, satisfying component 2 of section 3. Component 3 of section 3 is satisfied because the conclusion is stated in the context of the study. All three components are satisfied and section 3 was scored as essentially correct. Because two sections were scored as essentially correct, and one section was scored as partially correct, the response earned a score of 3.

Sample: 4C
Score: 2

In part (a) the response states that it is incorrect to conclude a causal relationship; therefore, component 1 of section 1 is not satisfied. To be scored as partially correct, component 1 must be satisfied with weak justification. Therefore section 1 was scored as incorrect. In part (b) component 1 of section 2 is satisfied because the parameters are correctly identified using standard notation, μ_{new} and μ_{standard}. Component 2 of section 2 is satisfied because the hypotheses indicate equality in the null and the correct direction in the alternative. Component 3 of section 2 is not satisfied because the response identifies the test as a two-sample *z*-test instead of a two-sample *t*-test. Component 4 of section 2 is satisfied because the correct test statistic for a difference in means is provided. Three of the four components of section 2 are satisfied, and section 2 was scored as partially correct. In part (b) component 1 of section 3 is satisfied because the response makes reference to an approximately correct *p*-value that is consistent with the test statistic and alternative hypothesis by stating that $P \approx 0.$ Component 2 of section 3 is satisfied because the response refers to a small *p*-value and uses that as justification for evidence to support the alternative hypothesis. Component 3 is satisfied because the conclusion is correctly stated in context. All three components of section 3 are satisfied, and section 3 was scored as essentially correct. Because one section was scored as essentially correct, one section was scored as partially correct, and one section was scored as incorrect, the response earned a score of 2.

© 2018 The College Board.