

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 $H_0 : \mu_S = \mu_N$ versus $H_a : \mu_S > \mu_N$.

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.

AP[®] STATISTICS
2018 SCORING GUIDELINES

Question 4 (continued)

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

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

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

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.

**AP® STATISTICS
2018 SCORING GUIDELINES**

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.

**AP® STATISTICS
2018 SCORING GUIDELINES**

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:
 - o States or implies that the null hypothesis is *accepted*
 - o States or implies that the alternative hypothesis has been *proven*
 - o States the conclusion in past tense (unless the response did not satisfy a component of section 2 for the use of past tense)

AP[®] STATISTICS
2018 SCORING GUIDELINES

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.

AP[®] STATISTICS
2018 SCORING GUIDELINES

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