

# AP<sup>®</sup> STATISTICS 2010 SCORING GUIDELINES

## Question 5

### Intent of Question

The primary goal of this question was to assess students' ability to set up, perform and interpret the results of a significance test. More specific goals were to assess students' ability to (1) state appropriate hypotheses; (2) identify the name of an appropriate statistical test and check appropriate assumptions/conditions; (3) compute the appropriate test statistic and  $p$ -value; (4) draw an appropriate conclusion, with justification, in the context of the study.

### Solution

Step 1: States a correct pair of hypotheses

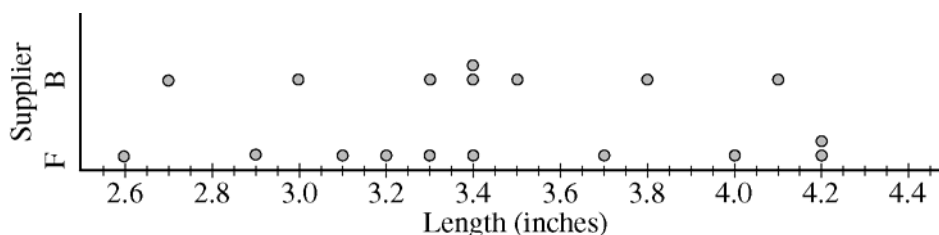
Let  $\mu_B$  represent the population mean length of all adult fish of this species from Buy-Rite Pets, and

let  $\mu_F$  represent the population mean length of all adult fish of this species from Fish Friends.

The hypotheses to be tested are  $H_0: \mu_B = \mu_F$  versus  $H_a: \mu_B < \mu_F$ .

Step 2: Identifies a correct test procedure (by name or by formula) and checks appropriate conditions

The appropriate test is a two-sample  $t$ -test. The first condition is that the samples are independent random samples from the two populations. This was stated in the question. The second condition is that the population distributions of fish lengths are normal. The following dotplots reveal no obvious departures from normality, so it appears reasonable to proceed with the two-sample  $t$ -test.



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

$$\text{The test statistic is: } t = \frac{\bar{x}_B - \bar{x}_F}{\sqrt{\frac{s_B^2}{n_B} + \frac{s_F^2}{n_F}}} = \frac{3.40 - 3.46}{\sqrt{\frac{0.434^2}{8} + \frac{0.550^2}{10}}} \approx -0.259$$

With  $df = 15.99999$ ,  $p$ -value = 0.3996.

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

Because this  $p$ -value is larger than any conventional significance level (such as  $\alpha = 0.10$  or  $\alpha = 0.05$ ), we fail to reject  $H_0$ . The sample data do not provide convincing evidence to conclude that the mean length of the adult fish of the species from Fish Friends is greater than the mean length of the adult fish of the same species from Buy-Rite Pets.

# AP<sup>®</sup> STATISTICS

## 2010 SCORING GUIDELINES

### Question 5 (continued)

#### **Scoring**

Each of steps 1, 2, 3 and 4 is scored as essentially correct (E), partially correct (P) or incorrect (I).

*General Note:* If a two-sample  $t$ -interval approach is taken without addressing the one-sided versus two-sided discrepancy, the student will lose credit for step 1 but still may earn full credit for steps 2, 3 and 4. The correct 95 percent confidence interval with  $df = 16$  is from  $-0.55$  to  $0.43$  inches.

**Step 1** is scored as follows:

Essentially correct (E) if the student uses correct parameters *AND* states correct hypotheses.

Partially correct (P) if the student uses correct parameters *OR* states correct hypotheses but not both.

Incorrect (I) otherwise.

#### *Notes*

- If the null hypothesis is wrong, reduce the score in this step by one level (i.e., E to P, or P to I).
- If the alternative hypothesis is two-sided or in the wrong direction, the student does not get credit for the hypotheses.
- If standard symbols are used for the parameters with appropriate group labels (e.g.,  $\mu_B$ ,  $\mu_F$ ), the parameter component is considered correct.
  - If *generic* standard symbols are used for the parameters (e.g.,  $\mu_1$ ,  $\mu_2$ ), students must clearly identify the parameters with the suppliers.
  - If standard symbols (either with context or generic) are used for the parameters *and* the student attempts to define them, the definitions must be correct and in context, including the concept of mean.
  - If *nonstandard* symbols are used for the parameters, they must be explicitly defined in context and include the concepts of mean and population.
  - If a student does not use symbols in the hypotheses, the response can still receive an E as long as the alternative hypothesis is in the correct direction and it clearly refers to population means in context.

**Step 2** is scored as follows:

Essentially correct (E) if the student correctly completes all three of the following components:

- Identifies the correct test procedure (by name or by formula)
- Checks for independent random samples
- Checks for normality

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

Incorrect (I) otherwise.

# AP<sup>®</sup> STATISTICS

## 2010 SCORING GUIDELINES

### Question 5 (continued)

#### Notes

- A two-sample  $z$ -test is not a correct test procedure in this case, but if both conditions are checked correctly, this step is scored as partially correct.
- If a student chooses to conduct a pooled  $t$ -test, the equal variance condition must be addressed (e.g., by commenting on similarity of standard deviations or conducting a test for equality of variances) to get credit for choosing the appropriate test procedure.
- To get credit for the check of independent random samples, students must indicate that *more than one* random sample was taken.
- To get credit for the normality condition, students must include correct graphs of both distributions and include an appropriate comment about shape or outliers, such as “neither has outliers,” “both are roughly symmetric,” “no obvious departures from normality,” “approximately normal,” etc.
- Ignore additional conditions listed, as long as they are correct, such as “the sample sizes must be less than 10 percent of the population sizes.” However, if the student includes additional *incorrect* conditions, such as  $np > 10$ , reduce the score in this step by one level (i.e., E to P, or P to I).

**Step 3** is scored as follows:

Essentially correct (E) if the student correctly calculates both the test statistic and  $p$ -value.

Partially correct (P) if the student correctly calculates the test statistic but not the  $p$ -value *OR* omits the test statistic but correctly calculates the  $p$ -value.

Incorrect (I) otherwise.

#### Notes

- It is acceptable for students to use the conservative  $df$  ( $df = 7$ ) or use the  $t$ -table to get a  $p$ -value  $> 0.25$ .
- Students who incorrectly choose a two-sample  $z$ -test lose credit for identifying the correct test procedure in step 2 but can earn full credit in step 3 if they provide the correct  $z$ -statistic ( $z = -0.259$ ) and  $p$ -value ( $p$ -value = 0.3978).
- If the alternative hypothesis is two-sided, the  $p$ -value must be approximately 0.8 to get credit for the  $p$ -value component.
- If a student provides the correct test statistic and/or  $p$ -value but shows additional incorrect work, such as a wrong formula, reduce the score in this step by one level (i.e., E to P, or P to I).

**Step 4** is scored as follows:

Essentially correct (E) if the student provides a correct conclusion in context, also providing justification based on linkage between the  $p$ -value and conclusion.

Partially correct (P) if the student provides a correct conclusion, with linkage to the  $p$ -value, but not in context *OR* provides a correct conclusion in context, but without justification based on linkage to the  $p$ -value.

Incorrect (I) otherwise.

# AP<sup>®</sup> STATISTICS

## 2010 SCORING GUIDELINES

### Question 5 (continued)

#### Notes

- The conclusion must be about the *mean* fish lengths to get credit for context unless the student already lost credit in step 1 for neglecting to include the concept of mean.
- The conclusion must be consistent with the alternative hypothesis to get credit for context unless the student was already penalized for inconsistency with the alternative hypothesis in step 3.
- If the conclusion is consistent with an incorrect  $p$ -value from step 3, and also in context, with justification based on linkage to the  $p$ -value, then this step is scored essentially correct.
- If both a significance level  $\alpha$  and a  $p$ -value are given together, the linkage between the  $p$ -value and the conclusion is implied. If no  $\alpha$  level 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, such as saying: “Because the  $p$ -value is small, we reject the null hypothesis” or “Because the  $p$ -value is large, we do not reject the null hypothesis.”
- If the student chooses to “retain the null hypothesis,” with linkage and/or context, this should be scored partially correct (P). If the student goes on to say something equivalent to “fail to reject” (e.g., “we should not conclude the mean length of fish is greater at Fish Friends”) in context, with linkage, then the response should be scored essentially correct.
- A conclusion in step 4 that is equivalent to “accept  $H_0$ ” (such as “we conclude that the mean fish length is the same from both suppliers”) cannot be scored essentially correct. Such a response should be scored partially correct, provided that the conclusion is in context, with justification based on linkage to the  $p$ -value. Such a response should be scored incorrect if it lacks either context or linkage to the  $p$ -value.
- If a student attempts to interpret the  $p$ -value, but does so incorrectly, then do not give credit for the linkage component.

Each essentially correct (E) step counts as 1 point. Each partially correct (P) step counts as  $\frac{1}{2}$  point.

<b>4</b>	<b>Complete Response</b>
<b>3</b>	<b>Substantial Response</b>
<b>2</b>	<b>Developing Response</b>
<b>1</b>	<b>Minimal Response</b>

If a response is between two scores (for example,  $2\frac{1}{2}$  points), use a holistic approach to determine whether to score up or down, depending on the overall strength of the response and communication.

SA

SA

5. A large pet store buys the identical species of adult tropical fish from two different suppliers—Buy-Rite Pets and Fish Friends. Several of the managers at the pet store suspect that the lengths of the fish from Fish Friends are consistently greater than the lengths of the fish from Buy-Rite Pets. Random samples of 8 adult fish of the species from Buy-Rite Pets and 10 adult fish of the same species from Fish Friends were selected and the lengths of the fish, in inches, were recorded, as shown in the table below.

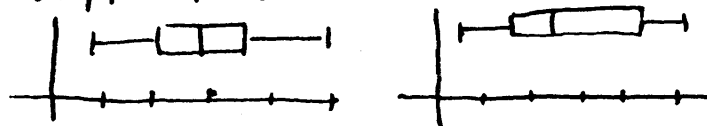
	Length of Fish	Mean	Standard Deviation
Buy-Rite Pets ( $n_B = 8$ )	3.4 2.7 3.3 4.1 3.5 3.4 3.0 3.8	3.40	0.434
Fish Friends ( $n_F = 10$ )	3.3 2.9 4.2 3.1 4.2 4.0 3.4 3.2 3.7 2.6	3.46	0.550

Do the data provide convincing evidence that the mean length of the adult fish of the species from Fish Friends is greater than the mean length of the adult fish of the same species from Buy-Rite Pets?

conduct a two sample t test to estimate the true mean difference in fish lengths  
 $H_0: \mu_F - \mu_B = 0$     $H_a: \mu_F - \mu_B > 0$     $\alpha = 0.05$   
 $\mu_F$  = true mean length of fish at Fish Friends  
 $\mu_B$  = true mean length of fish at Buy-Rite Pets

Conditions:

- Groups of fish are independent and randomly selected
- Assume both samples less than 10% of population at each fish store (At least 80 at BuyRite & 100 fish at Fish Friends)
- Buy Rite Pets      Fish Friends

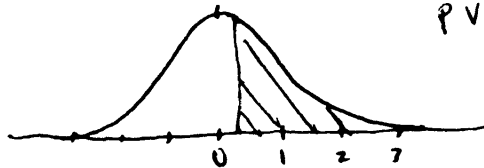


Because the histograms for the data show no apparent outliers or drastic skewedness, assume both distributions normal.

$$t = \frac{3.46 - 3.40}{\sqrt{\frac{0.434^2}{8} + \frac{0.550^2}{10}}} = 0.2585 \quad P(t > 0.2585) = 0.399$$

p value

$$df = 15.9999$$



With a p value  $> \alpha$  ( $0.399 > 0.05$ ) we fail to reject the null & cannot conclude that the mean length of fish from Fish Friends is greater than the mean length of fish from Buy-Rite Pets.

GO ON TO THE NEXT PAGE.

5B

5B

5. A large pet store buys the identical species of adult tropical fish from two different suppliers—Buy-Rite Pets and Fish Friends. Several of the managers at the pet store suspect that the lengths of the fish from Fish Friends are consistently greater than the lengths of the fish from Buy-Rite Pets. Random samples of 8 adult fish of the species from Buy-Rite Pets and 10 adult fish of the same species from Fish Friends were selected and the lengths of the fish, in inches, were recorded, as shown in the table below.

FF = Fish Friends  
BRP = Buy-Rite Pets

	Length of Fish	Mean	Standard Deviation
Buy-Rite Pets ( $n_B = 8$ )	3.4 2.7 3.3 4.1 3.5 3.4 3.0 3.8	3.40	0.434
Fish Friends ( $n_F = 10$ )	3.3 2.9 4.2 3.1 4.2 4.0 3.4 3.2 3.7 2.6	3.46	0.550

Do the data provide convincing evidence that the mean length of the adult fish of the species from Fish Friends is greater than the mean length of the adult fish of the same species from Buy-Rite Pets?

$H_0: \mu_{FF} - \mu_{BRP} = 0$  The null hypothesis is that there is no difference in mean adult fish lengths among the two tropical fish suppliers.

$H_A: \mu_{FF} - \mu_{BRP} > 0$  The alternative hypothesis is that there is a difference in mean adult fish lengths among the two tropical fish suppliers; Fish Friends has greater mean lengths than Buy-Rite pets.

Before we continue we must check the assumptions and conditions. First, we must assume that the lengths of fish are independent of each other. Second, random samples of the fish were taken from each supplier. Third, we must assume that the 8 fish from Buy-Rite Pets and the 10 fish from Fish Friends are less than 10% of the entire population of fish at those stores. We must assume that the two groups of fish are independent of each other. The histograms of each set of data appear to be unimodal and reasonably symmetric.

Because the conditions and assumptions have been satisfied, I can proceed in using a student's t-model to compare the difference of means in my data.

- Assumptions/conditions
- ① Independence ✓
  - ② Random ✓
  - ③ 10% ✓
  - ④ Independence of groups ✓
  - ⑤ nearly normal ✓

$$t = .25869 \quad p = .39958 \quad df = 15,999$$

Because the p-value is so large, I will retain the null hypothesis that there is no difference in mean adult fish lengths among the two tropical fish suppliers. ~~FF~~

GO ON TO THE NEXT PAGE.

5C

5C

5. A large pet store buys the identical species of adult tropical fish from two different suppliers—Buy-Rite Pets and Fish Friends. Several of the managers at the pet store suspect that the lengths of the fish from Fish Friends are consistently greater than the lengths of the fish from Buy-Rite Pets. Random samples of 8 adult fish of the species from Buy-Rite Pets and 10 adult fish of the same species from Fish Friends were selected and the lengths of the fish, in inches, were recorded, as shown in the table below.

	Length of Fish	Mean	Standard Deviation
B-RP = Buy-Rite Pets ( $n_B = 8$ )	3.4 2.7 3.3 4.1 3.5 3.4 3.0 3.8	3.40	0.434
FF = Fish Friends ( $n_F = 10$ )	3.3 2.9 4.2 3.1 4.2 4.0 3.4 3.2 3.7 2.6	3.46	0.550

Do the data provide convincing evidence that the mean length of the adult fish of the species from Fish Friends is greater than the mean length of the adult fish of the same species from Buy-Rite Pets?

Step 1:  $H_0$  - There is no difference (FF-BRP) in length between the fishes

$H_a$  - The difference (FF-BRP) in length between the fishes are greater.

$$H_0: \mu_1 = \mu_2 \quad H_a: \mu_1 > \mu_2$$

Step 2: SRS: We are told the fish of both stores are a random sample from both population.

normality: Since both of samples are below 30 (CLT), we must assume normality.

Independence: We must assume independence for both.

Step 3: 2 Sample T test: Data

List 1: L<sub>1</sub>

List 2: L<sub>2</sub>

Freq 1: 1

Freq 2: 1

$\mu_1 > \mu_2$

pooled: no

$$t = -0.25859$$

$$P = .60037$$

Step 4: Since the p-value, .60037, is significantly above  $\alpha = .05$ , we can accept the null and conclude that there are no differences in length.

GO ON TO THE NEXT PAGE.

# AP<sup>®</sup> STATISTICS

## 2010 SCORING COMMENTARY

### Question 5

#### Overview

The primary goal of this question was to assess students' ability to set up, perform and interpret the results of a significance test. More specific goals were to assess students' ability to (1) state appropriate hypotheses; (2) identify the name of an appropriate statistical test and check appropriate assumptions/conditions; (3) compute the appropriate test statistic and  $p$ -value; (4) draw an appropriate conclusion, with justification, in the context of the study.

#### Sample: 5A

##### Score: 4

For step 1 the student states correct hypotheses using standard symbols, including context-specific subscripts. Although not necessary, the student also defines the symbols used in the hypotheses, including the concepts of population ("true") and mean in context. Step 1 was scored as essentially correct. For step 2 the student correctly identifies the test at the top of the page, states that samples were "independent and randomly selected" and constructs boxplots to assess normality, stating that it is reasonable to assume both (population) distributions are normal because there are "no ... outliers or drastic skewedness." The student refers to the boxplots as "histograms," but this was considered a very minor error. Step 2 was scored as essentially correct. For step 3 the student uses the formula to correctly calculate the test statistic and  $p$ -value and includes a correct sketch. Step 3 was scored as essentially correct. For step 4 the student correctly fails to reject the null hypothesis, provides a justification with linkage to the  $p$ -value by stating that the  $p$ -value is greater than a significance level of 0.05 and answers in context by addressing the mean lengths of the fish. Step 4 was scored as essentially correct. With all four steps essentially correct, the response earned a score of 4.

#### Sample: 5B

##### Score: 3

For step 1 the student states correct hypotheses using standard symbols, including context-specific subscripts. Although not necessary, the student also restates the correct hypotheses in words, including the concept of mean in context. Step 1 was scored as essentially correct. For step 2 the student discusses the relevant conditions for a two-sample  $t$ -test, including the independent random samples condition and normality condition. However, the student does not include the "histograms" that are referred to in the response, which makes the normality condition incorrect. The response goes on to say that "a student's  $t$ -model" will be used to compare the difference in means. Because the response includes a correct identification of the test procedure and a correct check of the independent random samples condition, step 2 was scored as partially correct. For step 3 the student calculates the correct test statistic and  $p$ -value. Step 3 was scored as essentially correct. For step 4 the student uses the phrase "retain the null hypothesis" without a further explanation of what this means. However, because the response justifies the decision with linkage to the  $p$ -value ("[b]ecause the  $p$ -value is so large") and the conclusion is about the difference in means with context, step 4 was scored as partially correct. With two steps essentially correct and two steps partially correct, the response earned a score of 3.



**AP<sup>®</sup> STATISTICS**  
**2010 SCORING COMMENTARY**

**Question 5 (continued)**

**Sample: 5C**

**Score: 2**

For step 1 the student initially states the hypotheses in words and correctly identifies the direction of the alternative hypothesis using the word “greater” in combination with the symbolic representation “ $\mu > 0$ .” However, the verbal hypotheses never address the concept of mean, so no credit was granted for the parameters. Furthermore, based on the hypotheses using symbols, it is unclear if the student is using a one-sample or two-sample test. Because the response failed to get credit for the parameters but did get credit for the direction of the hypotheses, step 1 was scored as partially correct. For step 2 the student indicates that random samples were taken from both populations and that normality should be checked. However, the student does not attempt to address the normality condition by graphing the data. In what the student calls “Step 3,” the response includes a correct identification of the test procedure. Because two of the three required components are present, step 2 was scored as partially correct. For step 3, the student provides a correct test statistic but an incorrect  $p$ -value. Presumably this student reversed the sign of the alternative hypothesis when using a calculator to conduct the test. Because only the test statistic is correct, step 3 was scored as partially correct. For step 4 the student decides to “accept the null,” which can be at best partially correct, provided that the response is in context and links the  $p$ -value to the conclusion. The response correctly links the  $p$ -value to the conclusion by comparing it to “ $\alpha = .05$ ,” but the concept of mean is missing. However, given that the student already lost credit in step 1 for neglecting the concept of mean, no additional deduction was made here for lack of context. Step 4 was scored as partially correct. With all four steps partially correct, the response earned a score of 2.