

# AP<sup>®</sup> STATISTICS 2007 SCORING GUIDELINES

## Question 4

### Intent of Question

This statistical inference question was developed to assess a student's ability to distinguish paired-data procedures from two-sample procedures and to execute the selected procedure. The ability to provide a complete statistical justification is an important skill that can be evaluated with this standard inference problem.

### Solution

A hypothesis test for the mean difference in the level of *E. coli* bacteria contamination in beef detected by the two methods will be conducted.

**Part 1:** States a correct pair of hypotheses:

$$H_0 : \mu_d = 0$$

$$H_a : \mu_d \neq 0$$

where  $\mu_d$  is the mean difference (method A – method B) in the level of *E. coli* bacteria contamination in beef detected by the two methods

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

$$\text{Paired } t\text{-test } t = \frac{\bar{x}_d - 0}{s_d / \sqrt{n_d}}$$

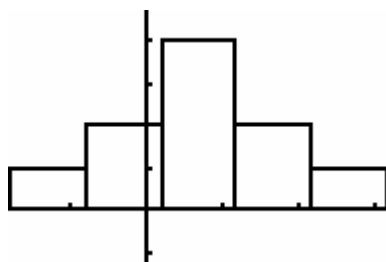
Conditions:

1. Since the observations are obtained on 10 randomly selected specimens, it is reasonable to assume that the 10 data pairs are independent of one another.
2. The population distribution of differences is normal.

The computed differences are:

-0.3 0.5 0.3 0.6 0.8 0.7 1.2 0.2 -0.1 -1.0

Histogram of the differences (A-B):



This histogram of differences is symmetric with no apparent outliers. Even though it is hard to judge the overall shape of a distribution with only 10 observations, it appears that the normal distribution is a reasonable option in this case.

**AP<sup>®</sup> STATISTICS  
2007 SCORING GUIDELINES**

**Question 4 (continued)**

**Part 3:** Correct mechanics, including the value of the test statistic, d.f., and  $P$ -value (or rejection region):

$$\bar{x}_d = 0.29 \quad s_d = 0.629727$$
$$t = \frac{0.29 - 0}{0.629727/\sqrt{10}} = \frac{0.29}{0.199137} = 1.46 \quad \text{d.f.} = 9 \quad P\text{-value} = 0.179$$

OR

Calculator:  $t = 1.4563$ ,  $P\text{-value} = 0.1793$ , d.f. = 9.

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

Since the  $P$ -value is greater than 0.05, we cannot reject  $H_0$ . We do NOT have statistically significant evidence to conclude that there is a difference in the mean amount of *E. coli* bacteria detected by the two methods for this type of beef. In other words, there does not appear to be a significant difference in these two methods for measuring the level of *E. coli* contamination in beef.

**Scoring**

Parts 1, 2, 3, and 4 are scored as essentially correct (E) or incorrect (I).

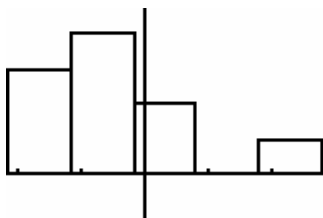
**Part 1** is scored as essentially correct (E) if the student states a correct pair of hypotheses. The hypotheses may be stated in terms of  $\mu_A$  and  $\mu_B$ . With any nonstandard notation used, the parameters must be identified in context clearly indicating the *population* mean.

**Part 2** is scored as essentially correct (E) if the student identifies a correct test (by name or formula) and checks appropriate conditions. The conditions for the paired  $t$ -test are about the differences. If the student says that the 10 differences can be viewed as a SRS of all differences, the answer is acceptable. However, the student does not need to repeat the fact that these specimens can be viewed as a random sample.

It is not acceptable to view all 20 observations as a random sample or two independent samples. If conditions are stated and checked using the two separate samples, part 2 is scored as incorrect (I) for a paired  $t$ -test.

For part 2, a graphical check of normality is required. Graph(s) should be consistent with the data, *AND* students must comment linking the graph to the condition.

Histogram of the differences (B-A):

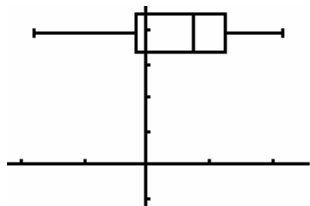


This histogram of differences is roughly symmetric with no apparent outliers. Even though it is hard to judge the overall shape of a distribution with only 10 observations, it appears that the normal distribution is a reasonable option in this case.

# AP<sup>®</sup> STATISTICS 2007 SCORING GUIDELINES

## Question 4 (continued)

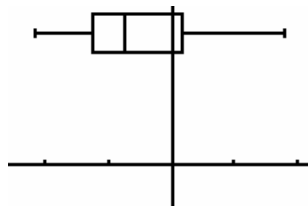
Boxplot of differences (A-B):



Normal Probability Plot of differences (A-B)  
(data on  $x$ -axis):



Boxplot of differences (B-A):



Normal Probability Plot of differences (B-A)  
(data on  $x$ -axis):



The boxplot of the differences shows that the distribution is approximately symmetric with no outliers, so it is reasonable to proceed with the paired  $t$ -test.

The normal probability plot shows linear trend with no obvious departures from linear trend, so the normal model is reasonable.

**Part 3** is scored essentially correct (E) if the student performs correct mechanics when calculating the value of the test statistic and correctly calculates the  $p$ -value for the rejection region.

**Part 4** is scored essentially correct (E) if the student states a correct conclusion in the context of the problem, using the result of the statistical test.

If the  $p$ -value in part 3 is incorrect but the conclusion is consistent with the computed  $p$ -value, part 4 can be considered correct.

In part 4, if both an  $\alpha$  and a  $p$ -value are given together, the linkage between the  $p$ -value and the conclusion is implied.

If no  $\alpha$  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.

Scoring Confidence Interval approach:

A confidence interval may be used to make the inference but must include all four parts to get full credit.

The confidence level must be stated to get credit for part 3.

A 95 percent confidence interval for  $\mu_d$  is  $(-0.16, 0.74)$ .

**AP<sup>®</sup> STATISTICS**  
**2007 SCORING GUIDELINES**

**Question 4 (continued)**

Since zero is included in the 95 percent confidence interval, we cannot reject the null hypothesis at the 0.05 level. We do NOT have statistically significant evidence to conclude that there is a difference in the mean amount of *E.coli* bacteria detected by the two methods for this type of beef. In other words, there does not appear to be a significant difference in these two methods for measuring the level of *E.coli* contamination in beef.

Scoring independent samples *t*-test or confidence interval approach:

If an independent samples *t*-test or confidence interval is done, the maximum score is 3, provided all four parts for independent samples *t*-test are done correctly.

For the independent samples *t*-test or confidence interval, the condition of normality must be checked using two samples separately.

$$t = 0.079 \qquad p = 0.9377 \qquad df = 18 \text{ (pooled) or } 17.97 \text{ (unpooled)}$$

A 95 percent independent-samples (two-sample) confidence interval for  $\mu_A - \mu_B$  is (-7.40, 7.98).

Each part is scored as correct or incorrect.

**4      Complete Response**

Four parts correct

**3      Substantial Response**

Three parts correct

**2      Developing Response**

Two parts correct

**1      Minimal Response**

One part correct