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 μ_d is the mean difference (method A − method B) in the level of *E. coli* bacteria contamination in beef detected by the two methods

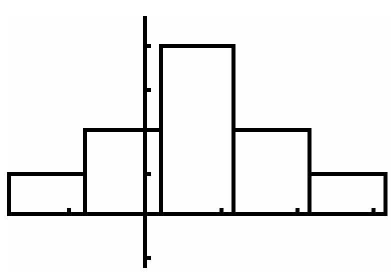
Part 2: Identifies a correct test (by name or by formula) and checks appropriate conditions:

Paired *t*-test $t = \dfrac{\overline{x}_d - 0}{s_d / \sqrt{n_d}}$

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.

Question 4 (continued)

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

$\bar{x}_d = 0.29 \qquad s_d = 0.629727$

$t = \frac{0.29 - 0}{0.629727/\sqrt{10}} = \frac{0.29}{0.199137} = 1.46 \qquad$ d.f. = 9 $\qquad$ *P*-value = 0.179

OR

Calculator: t = 1.4563, *P*-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).

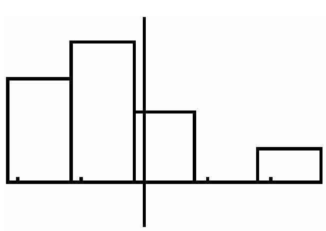
Part 1 is scored as essentially correct (E) if the student states a correct pair of hypotheses. The hypotheses may be stated in terms of μ_A and μ_B. With any nonstandard notation used, the parameters must be identified in context clearly indicating the *population* mean.

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.

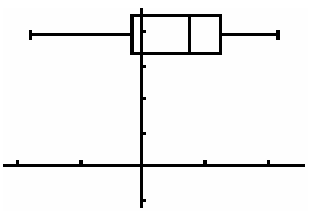
Histogram of the differences (B-A):



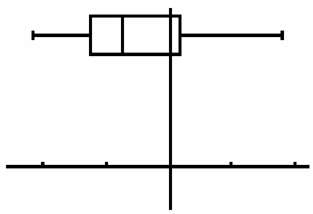
This histogram of differences is roughly symmetric with no apparent outliers. Even though it is hard to judge the overall shape of a distribution with only 10 observations, it appears that the normal distribution is a reasonable option in this case.

Question 4 (continued)

Boxplot of differences (A-B):

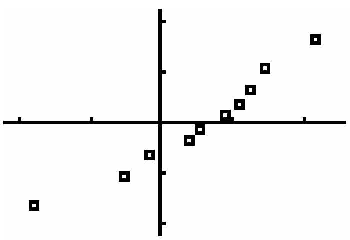


Boxplot of differences (B-A):

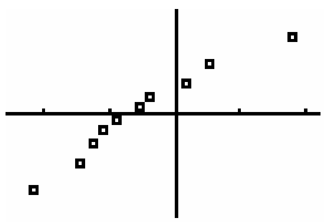


The boxplot of the differences shows that the distribution is approximately symmetric with no outliers, so it is reasonable to proceed with the paired *t*-test.

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



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



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 α and a *p*-value are given together, the linkage between the *p*-value and the conclusion is implied.

If no α 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 μ_d is (-0.16, 0.74).

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.

$t = 0.079$ $p = 0.9377$ df = 18 (pooled) or 17.97 (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

4. Investigators at the U.S. Department of Agriculture wished to compare methods of determining the level of *E. coli* bacteria contamination in beef. Two different methods (A and B) of determining the level of contamination were used on each of ten randomly selected specimens of a certain type of beef. The data obtained, in millimicrobes/liter of ground beef, for each of the methods are shown in the table below.

		Specimen									
		1	2	3	4	5	6	7	8	9	10
Method	A	22.7	23.6	24.0	27.1	27.4	27.8	34.4	35.2	40.4	46.8
	B	23.0	23.1	23.7	26.5	26.6	27.1	33.2	35.0	40.5	47.8

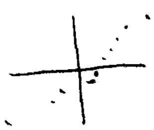
Is there a significant difference in the mean amount of *E. coli* bacteria detected by the two methods for this type of beef? Provide a statistical justification to support your answer.

I. $H_0: \mu_{diff} = 0$ there is no difference in the mean amount of E. coli detected by the two methods

$H_a: \mu_{diff} \neq 0$ " " a difference " " " " " " " " " " " "

$\mu_{diff} = \mu_{method A} - \mu_{method B}$ $\alpha = .05$ = significance level

II. SRS given, population σ unknown, normal quantile plot ≈ linear, ∴ aprox normal



∴ use matched pairs t-test @ 5% significance level

III. $t^* = \frac{.29}{.6297/\sqrt{10}} = 1.456$ p-value = 2[tcdf(1.456, E99, 9)] = .1793

df = 10 - 1 = 9

IV. Since the p-value of .1793 is greater than the significance level of .05, we fail to reject the null hypothesis. In other words, we do not have strong enough evidence to say that there is a significant difference in the mean amount of E coli bacteria detected by the two methods.

GO ON TO THE NEXT PAGE.

4. Investigators at the U.S. Department of Agriculture wished to compare methods of determining the level of *E. coli* bacteria contamination in beef. Two different methods (A and B) of determining the level of contamination were used on each of ten randomly selected specimens of a certain type of beef. The data obtained, in millimicrobes/liter of ground beef, for each of the methods are shown in the table below.

		Specimen									
		1	2	3	4	5	6	7	8	9	10
Method	A	22.7	23.6	24.0	27.1	27.4	27.8	34.4	35.2	40.4	46.8
	B	23.0	23.1	23.7	26.5	26.6	27.1	33.2	35.0	40.5	47.8

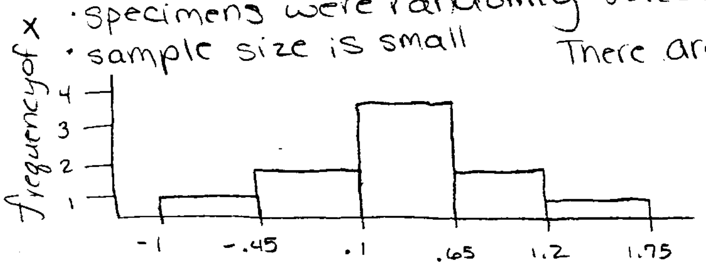
Is there a significant difference in the mean amount of *E. coli* bacteria detected by the two methods for this type of beef? Provide a statistical justification to support your answer.

Matched Pairs T-test

$\bar{x}$ = mean of differences of Method A - Method B

Conditions

- σ is unknown
- we will assume independence in types of beef
- specimens were randomly selected
- sample size is small   There are no outliers or skewness.



Hypothesis

$H_0: \bar{x} = 0$ No difference in amount of E. coli detected.

$H_A: \bar{x} \neq 0$ Significant difference in amount of E. coli detected

Test Statistic

$\alpha = .05$

$t = \frac{\bar{x} - \mu}{\sqrt{\frac{s}{n}}} = \frac{.29 - 0}{\sqrt{\frac{.62971}{10}}} = 1.4563$

p-value = .1793

Conclusion

Since p-value = .1793 > α = .05, we will fail to reject H_0. We can conclude that there is no significant difference in the amount of E. coli detected by Method A and Method B.

GO ON TO THE NEXT PAGE.

4. Investigators at the U.S. Department of Agriculture wished to compare methods of determining the level of *E. coli* bacteria contamination in beef. Two different methods (A and B) of determining the level of contamination were used on each of ten randomly selected specimens of a certain type of beef. The data obtained, in millimicrobes/liter of ground beef, for each of the methods are shown in the table below.

	Specimen									
	1	2	3	4	5	6	7	8	9	10
Method A	22.7	23.6	24.0	27.1	27.4	27.8	34.4	35.2	40.4	46.8
Method B	23.0	23.1	23.7	26.5	26.6	27.1	33.2	35.0	40.5	47.8
A-B	-.3	.5	.3	.6	.8	.7	1.2	.2	-.1	-1

Is there a significant difference in the mean amount of *E. coli* bacteria detected by the two methods for this type of beef? Provide a statistical justification to support your answer.

Hypotheses:

$H_0: \mu_A = \mu_B$, $\mu_{A-B} = 0$ No difference between the two methods

$H_A: \mu_A \neq \mu_B$, $\mu_{A-B} \neq 0$ there is a difference in the dectection of E.Coil levels.

conditions:

Random ✓

indep ✓

at least 10 numbers ✓

>10% population ✓

use T testing

Data:

$n_A = 10$

$\mu_A = 30.94$

$s_A = 8.012$

$n_B = 10$

$\mu_B = 30.65$

$s_B = 8.344$

$n_{A-B} = 10$

$\mu_{A-B} = .29$

$s_{A-B} = .6297$

$t = 1.456$

$df = 9$

p value = .1793 fail to reject null, no significant difference between the mean amounts of E. Coil detected by the two methods

GO ON TO THE NEXT PAGE.

AP® STATISTICS
2007 SCORING COMMENTARY

Question 4

Overview

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 is evaluated with this standard inference problem.

Sample: 4A
Score: 4

This response shows understanding of the inference procedure and connections between its different parts. The matched-pairs t-test is correctly identified. Recognizing the two-sided nature of the inference, the hypotheses are defined by using standard notation in terms of the mean difference in the amount of *E. coli* detected by the two methods and are also described in the context of the problem. The assumption of normality is checked using a normal probability plot for differences and is assessed by commenting on the linearity observed in the plot. Although the normal probability plot is not in the Statistics Course Outline, it was created and interpreted correctly. The mechanics for the paired t-test are provided by showing the correct computations and value of the test statistic, degrees of freedom, and p-value. A correct conclusion statement, at the chosen level of significance, is based on the results of the given computations. The conclusion is stated correctly in the context of the problem. All four parts were scored as essentially correct.

Sample: 4B
Score: 3

This response incorrectly defines the hypotheses in terms of a statistic. The matched-pairs t-test is correctly identified. The hypotheses are incorrectly stated in terms of $\bar{x}$, which is the standard notation for sample means. The assumption of normality is checked using a histogram. Lack of outliers and lack of skewness in the histogram is noted. The correct test statistic (including a formula and computation) and p-value are reported. A correct conclusion statement, at the chosen level of significance, is based on the results of the given computations. The conclusion is stated correctly in the context of the problem.

Sample: 4C
Score: 2

This is a developing response. The test used is not identified by name or formula. However, the calculation of differences indicates the intention of using a matched-pairs t-test. The hypotheses are correctly stated in terms of the mean of differences. The assumption of normality is not checked. Information about mechanics is provided by reporting a correct test statistic, p-value, and degrees of freedom, clearly obtained from calculator output. The conclusion is stated correctly in the context of the problem; however, the link between the reported results and the conclusion is not strong enough.