AP® STATISTICS
2011 SCORING GUIDELINES

Question 4

Intent of Question

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

Solution

Step 1: States a correct pair of hypotheses.

Let μ_A represent the mean cholesterol reduction if all such male patients at this hospital are advised on appropriate exercise and diet and also receive a placebo.

Let μ_B represent the mean cholesterol reduction if all such male patients at this hospital are advised on appropriate exercise and diet but receive the drug instead of a placebo.

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

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

The appropriate procedure is a two-sample *t*-test.

When comparing two experimental treatments using a two-sample *t*-test, the subjects must be randomly assigned to the treatments. This condition is stated in the question (10 men were randomly assigned to group A and the remaining 10 men to group B).

The second condition is that the two populations are approximately normally distributed or the sample sizes are sufficiently large. Because of the small sample sizes (10 in each treatment group), we need to check whether it is reasonable to assume that the samples came from populations that are normally distributed. The following dotplots reveal slight skewness and a possible outlier for group B, but it appears reasonable to proceed with the two-sample *t*-test.
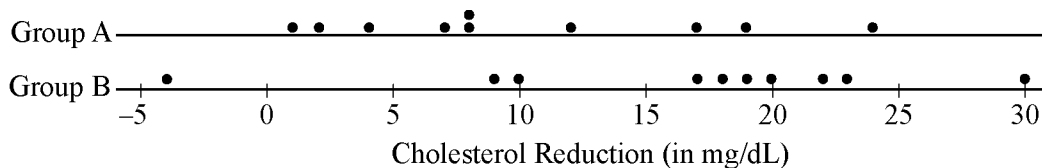
Step 3: Demonstrates correct mechanics, including the value of the test statistic and *p*-value (or the rejection region).

The test statistic is: $t = \dfrac{\overline{x}_A - \overline{x}_B}{\sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}} = \dfrac{10.20 - 16.40}{\sqrt{\dfrac{7.66^2}{10} + \dfrac{9.40^2}{10}}} \approx -1.62$

With df = 17.3, *p*-value ≈ 0.062.

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

Because the p-value is greater than the significance level of $\alpha = 0.01,$ we fail to reject $H_0.$ The data do not provide enough evidence at the 0.01 level of significance to conclude that the drug is effective in producing a mean cholesterol reduction beyond that provided by exercise and dietary advice.

Scoring

Steps 1, 2, 3 and 4 are each scored as essentially correct (E), partially correct (P) or incorrect (I).

Step 1 is scored as follows:

Essentially correct (E) if the response states hypotheses with correct comparisons between the means and defines the population means as the parameters.

Partially correct (P) if the response states hypotheses with correct comparisons between the means *OR* correctly defines the population means as the parameters, but not both.

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

Note: Defining the parameter symbols in context or simply using μ_A and $\mu_B,$ with subscripts clearly relevant to the context is sufficient for defining parameters.

Step 2 is scored as follows:

Essentially correct (E) if the response correctly includes the following three components:
   1. Identifies the correct test procedure (by name or by formula).
   2. Checks for random assignment to treatments.
   3. Checks for normality.

Partially correct (P) if the response correctly includes exactly two of the three components listed above.

Incorrect (I) if the response fails to meet the criteria for E or P.

Notes
- Graphs of both distributions must be produced and described to check the normality condition.
- If the response calls for a *pooled* two-sample *t*-test, step 2 can be scored as E as long as the condition of equal variances is mentioned and checked by comparing the variability in the graphs or the sample standard deviations.
- If the response calls for applying a paired *t*-test, then step 2 is scored as I, but steps 3 and 4 can be scored as E if the test mechanics are correct in step 3 and the conclusion is correct in step 4.

Step 3 is scored as follows:

Essentially correct (E) if both the test statistic and p-value are correctly calculated.

Partially correct (P) if the test statistic is correctly calculated but not the p-value
    OR
if the test statistic is calculated incorrectly, but the correct p-value for the computed test statistic is given.

Incorrect (I) if the response fails to meet the criteria for E or P.

Step 4 is scored as follows:

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

Partially correct (P) if the response provides a correct conclusion, including justification based on the size of the p-value, but not in context
    OR
if the response provides a correct conclusion, written in context, but without justification based on linkage to the p-value.

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

Notes
- If the conclusion is consistent with the p-value from step 3, and also in context with justification based the size of the p-value, then step 4 is scored as E (even if the p-value in step 3 is incorrect).
- A conclusion in step 4 that is equivalent to "accept H_0" (such as "we conclude that the drug is not effective") is not acceptable for an E. Such a response should be scored as P, provided that the conclusion is in context with justification based on the size of the p-value. Such a response should be scored as I if it lacks either context or linkage to the p-value.

Each essentially correct (E) step counts as 1 point. Each partially correct (P) step counts as ½ point.

4	**Complete Response**
3	**Substantial Response**
2	**Developing Response**
1	**Minimal Response**

If a response is between two scores (for example, 2½ points), use a holistic approach to determine whether to score up or down, depending on the overall strength of the response and communication.

4. High cholesterol levels in people can be reduced by exercise, diet, and medication. Twenty middle-aged males with cholesterol readings between 220 and 240 milligrams per deciliter (mg/dL) of blood were <u>randomly</u> <u>selected</u> from the population of such male patients at a large local hospital. Ten of the 20 males were <u>randomly</u> <u>assigned</u> to group A, advised on appropriate exercise and diet, and also received a placebo. The other 10 males were assigned to group B, received the same advice on appropriate exercise and diet, but received a drug intended to reduce cholesterol instead of a placebo. After three months, posttreatment cholesterol readings were taken for all 20 males and compared to pretreatment cholesterol readings. The tables below give the reduction in cholesterol level (pretreatment reading minus posttreatment reading) for each male in the study.

Group A (placebo)

Reduction (in mg/dL)	2	19	8	4	12	8	17	7	24	1

Mean Reduction: 10.20   Standard Deviation of Reductions: 7.66

Group B (cholesterol drug)

Reduction (in mg/dL)	30	19	18	17	20	−4	23	10	9	22

Mean Reduction: 16.40   Standard Deviation of Reductions: 9.40

1=drug   2=placebo

Do the data provide convincing evidence, at the $\alpha = 0.01$ level, that the cholesterol drug is effective in producing a reduction in mean cholesterol level beyond that produced by exercise and diet?

$H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 > \mu_2$

let $\mu_{1/2}$ = the population mean of the true reduction in cholesterol levels of cholesterol drug / placebo of middle aged males with cholesterol readings between 220–240 mg/dL at the large local hospital (with advice on diet & exercise)

$\alpha = .01$

$df = 17.295$

p-value = .0620

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{16.40 - 10.20}{\sqrt{\frac{9.40^2}{10} + \frac{7.66^2}{10}}} = 1.617$$

Decision: Do not reject H_0 because p > α

Conclusion: There is insufficient evidence, with a p-value of .0620 at an α level of .01, to conclude that the cholesterol drug is effective in producing a reduction in mean cholesterol level beyond that produced by exercise and diet.

Technical Conditions

1) Simple random sample and random assignment of observational units?
—yes, stated in prompt

2) independent samples?
—not stated in prompt, but independence is probable

3) n ≥ 30 or population is normal?
—Sample size is < 30, for both samples but population normalcy is plausible, since the boxplot shows no outliers or skewness,

Boxplots of Reduction in Cholesterol

placebo group

drug group

-5   0   5   10   15   20   -11-30
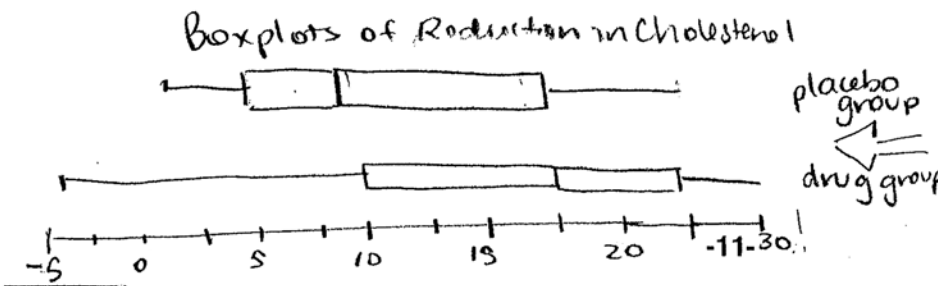
GO ON TO THE NEXT PAGE.

4. High cholesterol levels in people can be reduced by exercise, diet, and medication. Twenty middle-aged males with cholesterol readings between 220 and 240 milligrams per deciliter (mg/dL) of blood were randomly selected from the population of such male patients at a large local hospital. Ten of the 20 males were randomly assigned to group A, advised on appropriate exercise and diet, and also received a placebo. The other 10 males were assigned to group B, received the same advice on appropriate exercise and diet, but received a drug intended to reduce cholesterol instead of a placebo. After three months, posttreatment cholesterol readings were taken for all 20 males and compared to pretreatment cholesterol readings. The tables below give the reduction in cholesterol level (pretreatment reading minus posttreatment reading) for each male in the study.

Group A (placebo)

Reduction (in mg/dL)	2	19	8	4	12	8	17	7	24	1

Mean Reduction: 10.20   Standard Deviation of Reductions: 7.66

Group B (cholesterol drug)

Reduction (in mg/dL)	30	19	18	17	20	−4	23	10	9	22

Mean Reduction: 16.40   Standard Deviation of Reductions: 9.40

Do the data provide convincing evidence, at the $\alpha = 0.01$ level, that the cholesterol drug is effective in producing a reduction in mean cholesterol level beyond that produced by exercise and diet?
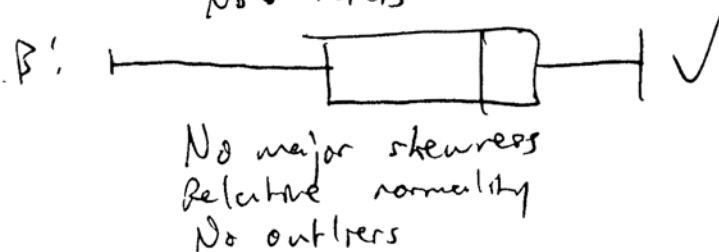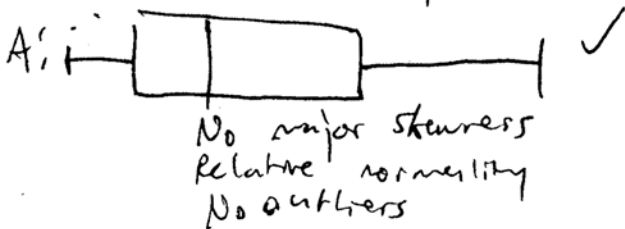
$H_0: M_A = M_B$

$H_a: M_A < M_B$

M_A, M_B = The true mean cholesterol reduction after 3 months of treatment and either a cholesterol drug or placebo

Conditions

2 SRS → Stated

2 Normalities : Graphs

A: ✓

No major skewness
Relative normality
No outliers

B: ✓

No major skewness
Relative normality
No outliers

Independence → Assume that the two groups are independent of each other

Use a 2-Sample t-test to compare M_A to M_B.

$p = \frac{.062}{}$ $t = -1.617$

$\alpha = .01$ $p > \alpha$

$.062 > .01$

We fail to reject H_0 and assume that H_0 is true. The results are not statistically significant at the .01 significance level. There is no significant difference in the reduction of cholesterol levels when a cholesterol drug is given as compared to the placebo.

GO ON TO THE NEXT PAGE.

-11-

High cholesterol levels in people can be reduced by exercise, diet, and medication. Twenty middle-aged males with cholesterol readings between 220 and 240 milligrams per deciliter (mg/dL) of blood were randomly selected from the population of such male patients at a large local hospital. Ten of the 20 males were randomly assigned to group A, advised on appropriate exercise and diet, and also received a placebo. The other 10 males were assigned to group B, received the same advice on appropriate exercise and diet, but received a drug intended to reduce cholesterol instead of a placebo. After three months, posttreatment cholesterol readings were taken for all 20 males and compared to pretreatment cholesterol readings. The tables below give the reduction in cholesterol level (pretreatment reading minus posttreatment reading) for each male in the study.

Group A (placebo)

Reduction (in mg/dL)	2	19	8	4	12	8	17	7	24	1

Mean Reduction: 10.20   Standard Deviation of Reductions: 7.66

Group B (cholesterol drug)

Reduction (in mg/dL)	30	19	18	17	20	−4	23	10	9	22

Mean Reduction: 16.40   Standard Deviation of Reductions: 9.40

Do the data provide convincing evidence, at the $\alpha = 0.01$ level, that the cholesterol drug is effective in producing a reduction in mean cholesterol level beyond that produced by exercise and diet?

Before we apply a test, we must check for requirements. According to the passage, this is a simple random sample size of 20. The groups were both randomly assigned, and split up with different treatments (controlled experiment). We are not positive if the groups are normally distributed so we must proceed with caution. Here are the boxplots:

Group A   min: 1 Q1: 4   Med: 8   Q3: 17   max: 24

Group B   min: −4   Q1: 10   Med: 18.5   Q3: 22   max: 30
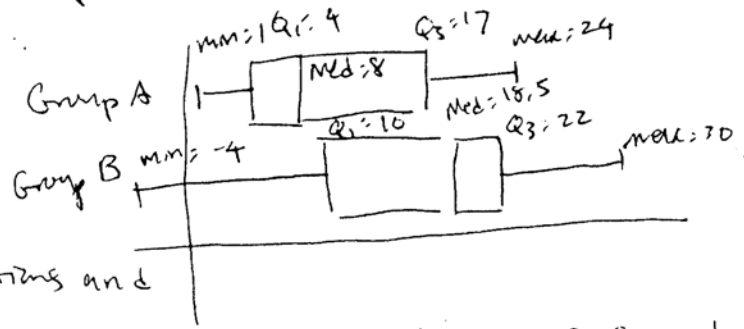
After checking conditions and understanding the data, we can perform a 2-Sample Z-Test

$z = -1.616$
$P = .0529$

$H_0: \bar{x}_A = \bar{x}_B$
$H_a: \bar{x}_A < \bar{x}_B$

According to the p-value, we do not have convincing evidence at the $\alpha = .01$ level that the reduction in cholesterol levels is significantly higher in Group B than Group A, therefore we fail to reject the null.

GO ON TO THE NEXT PAGE.

-11-

AP® STATISTICS
2011 SCORING COMMENTARY

Question 4

Overview

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

Sample: 4A
Score: 4

In step 1 the hypotheses make the correct comparisons, and the parameters are correctly defined. Step 1 was scored as essentially correct. In step 2 the test is not named, but the formula for the test statistic is sufficient. The condition of random assignment of treatments to subjects is stated, and the boxplots, with a summary, establish that it is plausible that the measurements came from normal distributions. Although the response states that no skewness is shown, there is, in fact, some skewness in the boxplots. However, step 2 was scored as essentially correct. In step 3 the test statistic t and the p-value are computed correctly. Step 3 was scored as essentially correct. In step 4 the correct decision is stated and justified: "Do not reject H_0 because $p > a$." The conclusion includes a statement, in context, that "[t]here is insufficient evidence … to conclude that the cholesterol drug is effective." Step 4 was scored as essentially correct. Because all four steps were scored as essentially correct, the response earned a score of 4.

Sample: 4B
Score: 3

In step 1 the hypotheses make the correct comparisons, and the parameters are correctly defined. Step 1 was scored as essentially correct. In step 2 the correct test is named. The boxplots are sketched and summarized well. However, no statement is made that treatments were randomly assigned to subjects. Because one of the three required components is not included, step 2 was scored as partially correct. In step 3 the test statistic t and the p-value are computed correctly. Step 3 was scored as essentially correct. In step 4 the correct decision is made (that is, failing to reject the null hypothesis), and it is justified by either of the inequalities presented, $p > a$ or $0.062 > 0.1.$ The conclusion is stated in context, but because the student also asserts, "[w]e … assume that H_0 is true," step 4 was scored as partially correct. Because two steps were scored as essentially correct and two steps were scored as partially correct, the response earned a score of 3.

Sample: 4C
Score: 2

In step 1 the hypotheses are about the sample means rather than population parameters. Step 1 was scored as partially correct. In step 2 the test is incorrectly named as a two-sample z-test. The response correctly states that the groups were randomly assigned and shows boxplots to check normality, but no description of the boxplots is given. Thus, step 2 was scored as incorrect. In step 3 the test statistic and the p-value are correct for a two-sample z-test. Step 3 was scored as essentially correct. In step 4 the decision is correct and the conclusion is written in context, but the decision is not justified by explicitly comparing the p-value to $\alpha = 0.01.$ Step 4 was scored as partially correct. Because one step was scored as essentially correct, two steps were scored as partially correct, and one step was scored as incorrect, the response earned a score of 2.