

AP[®] STATISTICS

2015 SCORING GUIDELINES

Question 4

Intent of Question

The primary goal of this question was to assess a student's ability to identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question. More specific goals were to assess a student's ability to (1) state appropriate hypotheses; (2) identify the appropriate statistical test procedure and check appropriate conditions for inference; (3) calculate the appropriate test statistic and p -value; and (4) draw an appropriate conclusion, with justification, in the context of the study.

Solution

Step 1: States a correct pair of hypotheses.

Let p_{asp} represent the population proportion of adults similar to those in the study who would have developed colon cancer within the six years of the study if they had taken a low-dose aspirin each day. Similarly, let p_{plac} represent the population proportion of adults similar to those in the study who would have developed colon cancer within the six years of the study if they had taken a placebo each day.

The hypotheses to be tested are $H_0 : p_{\text{asp}} = p_{\text{plac}}$ versus $H_a : p_{\text{asp}} < p_{\text{plac}}$ or equivalently,

$H_0 : p_{\text{asp}} - p_{\text{plac}} = 0$ versus $H_a : p_{\text{asp}} - p_{\text{plac}} < 0$.

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

The appropriate procedure is a two-sample z -test for comparing proportions.

Because this is a randomized experiment, the first condition is that the volunteers were randomly assigned to one treatment group or the other. The condition is satisfied because we are told that the volunteers were randomly assigned to take a low-dose aspirin or a placebo.

The second condition is that the sample sizes are large, relative to the proportions involved. The condition is satisfied because all sample counts are large enough; that is, 15 with colon cancer in aspirin group, 26 with colon cancer in placebo group, $500 - 15 = 485$ cancer-free in aspirin group, and $500 - 26 = 474$ cancer-free in placebo group.

Step 3: Calculates the appropriate test statistic and p -value.

The sample proportions who developed colon cancer are $\hat{p}_{\text{asp}} = \frac{15}{500} = 0.030$ and $\hat{p}_{\text{plac}} = \frac{26}{500} = 0.052$.

The combined sample proportion who developed colon cancer is $\hat{p}_{\text{combined}} = \frac{15 + 26}{500 + 500} = 0.041$.

The test statistic is $z = \frac{0.030 - 0.052}{\sqrt{0.041(1 - 0.041)\left(\frac{1}{500} + \frac{1}{500}\right)}} \approx -1.75$ (-1.7542 from calculator).

The p -value is $P(Z \leq -1.75) = 0.0401$ (0.0397 from calculator), where Z has a standard normal distribution.

AP[®] STATISTICS

2015 SCORING GUIDELINES

Question 4 (continued)

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

Because the p -value is less than the given significance level of $\alpha = 0.05$, we reject the null hypothesis. The data provide convincing statistical evidence that the proportion of all adults similar to the volunteers who would develop colon cancer if they had taken a low-dose aspirin every day is less than the proportion of all adults similar to the volunteers who would develop colon cancer if they had not taken a low-dose aspirin every day.

Scoring

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

Step 1 is scored as follows:

Essentially correct (E) if the response identifies correct parameters *AND* both hypotheses are labeled and state the correct relationship between the parameters.

Partially correct (P) if the response identifies correct parameters *OR* states correct relationships, but not both.

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

Note: Either defining the parameters in context, or simply using common parameter notation with subscripts clearly relevant to the context, such as p_{asp} and p_{plac} , is sufficient.

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. Notes that the use of random assignment satisfies the randomness condition.
3. Checks for approximate normality of the test statistic by citing that all four counts are larger than some standard criterion such as 5 or 10.

Partially correct (P) if the response correctly includes only two of the three components.

Incorrect (I) if the response correctly includes at most one of the three components.

Notes:

- For the randomness component, it is (minimally) acceptable to say “random assignment — check” but not acceptable to say “random — check” or “SRS — check.” The important concept here is that it is random assignment, and not random sampling, that is required. If the response implies that the study used a random sample, the randomness component is not satisfied, regardless of whether random assignment is correctly addressed.
- The normality check may use the expected counts under the null hypothesis in place of observed counts.

AP[®] STATISTICS

2015 SCORING GUIDELINES

Question 4 (continued)

Step 3 is scored as follows:

Essentially correct (E) if the response correctly calculates both the test statistic and a p -value that is consistent with the stated alternative hypothesis.

Partially correct (P) if the response correctly calculates the test statistic but not the p -value;

OR

if the response calculates the test statistic incorrectly but then calculates the correct p -value for the computed test statistic;

OR

if the response reports the correct p -value but no calculations or test statistic are shown.

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

Note: The p -value is considered correct if it is consistent with the alternative stated in the response and the calculated test statistic, even if those are incorrect.

Step 4 is scored as follows:

Essentially correct (E) if the response provides a correct conclusion in context, with justification based on linkage between the p -value and the given $\alpha = 0.05$.

Partially correct (P) if the response provides a correct conclusion, with linkage to the p -value, but not in context;

OR

if the response provides a correct conclusion 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:

- The conclusion must be related to the alternative hypothesis.
- If the p -value is incorrect, then step 4 is scored as E if the response includes proper linkage and a conclusion in context consistent with that p -value.
- If the p -value is less than 0.05, wording that states or implies that the alternative hypothesis is *proven* lowers the score one level (that is, from E to P or P to I) in step 4.
- If the p -value is incorrect and greater than 0.05, wording that states or implies that the null hypothesis is *accepted* lowers the score one level (that is, from E to P or P to I) in step 4.

AP[®] STATISTICS

2015 SCORING GUIDELINES

Question 4 (continued)

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

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

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

4A

4A1

4A

4. A researcher conducted a medical study to investigate whether taking a low-dose aspirin reduces the chance of developing colon cancer. As part of the study, 1,000 adult volunteers were randomly assigned to one of two groups. Half of the volunteers were assigned to the experimental group that took a low-dose aspirin each day, and the other half were assigned to the control group that took a placebo each day. At the end of six years, 15 of the people who took the low-dose aspirin had developed colon cancer and 26 of the people who took the placebo had developed colon cancer. At the significance level $\alpha = 0.05$, do the data provide convincing statistical evidence that taking a low-dose aspirin each day would reduce the chance of developing colon cancer among all people similar to the volunteers?

1) $p_1 \rightarrow$ proportion of people who develop colon cancer after taking low-dose aspirin each day after 6 yrs
 $p_2 \rightarrow$ proportion of people who develop colon cancer after taking a placebo each day after 6 yrs

$$\hat{p}_1 = .03$$

$$\hat{p}_2 = .052$$

$$\hat{p}_c = .041$$

2) $H_0: p_1 = p_2$ 3) $H_A: p_1 < p_2$

4) given random assignment

$$n_1 \hat{p}_c \geq 5 \quad n_1 (1 - \hat{p}_c) \geq 5 \quad n_2 \hat{p}_c \geq 5 \quad n_2 (1 - \hat{p}_c) \geq 5 \quad \left. \begin{array}{l} 20.5 \geq 5 \quad 479.5 \geq 5 \quad 20.5 \geq 5 \quad 479.5 \geq 5 \end{array} \right\} \begin{array}{l} \text{can use} \\ \text{normal} \\ \text{approx} \end{array}$$

assume independent samples

5) 2 proportion z test

$$z = \frac{.03 - .052}{\sqrt{\frac{(.041)(.959)}{500} + \frac{(.041)(.959)}{500}}}$$

$$z = \frac{.03 - .052}{\sqrt{\frac{(.041)(.959)}{500} + \frac{(.041)(.959)}{500}}}$$

6) $z = -1.754$

7)



8) p value = .0397

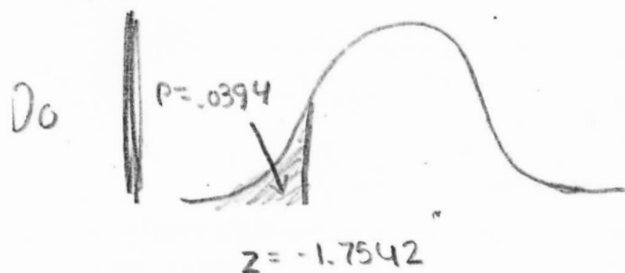
9) since $p < \alpha$, we reject H_0 .

10) there is sufficient evidence that taking a low-dose aspirin each day would reduce the chance of developing colon cancer among all people similar to the volunteers.

4. A researcher conducted a medical study to investigate whether taking a low-dose aspirin reduces the chance of developing colon cancer. As part of the study, 1,000 adult volunteers were randomly assigned to one of two groups. Half of the volunteers were assigned to the experimental group that took a low-dose aspirin each day, and the other half were assigned to the control group that took a placebo each day. At the end of six years, 15 of the people who took the low-dose aspirin had developed colon cancer and 26 of the people who took the placebo had developed colon cancer. At the significance level $\alpha = 0.05$, do the data provide convincing statistical evidence that taking a low-dose aspirin each day would reduce the chance of developing colon cancer among all people similar to the volunteers?

State $\left\{ \begin{array}{l} p_1 = \text{true proportion of people who took medicine and get colon cancer.} \\ p_2 = \text{true proportion of people who took placebo and get colon cancer.} \\ \alpha = .05 \quad H_0: p_1 = p_2 \quad H_a: p_1 < p_2 \end{array} \right.$

Plan $\left\{ \begin{array}{l} \text{IF conditions are met we will use a 2 sample z test for proportions.} \\ \text{Random: 1000 adult volunteers were randomly assigned to one of two groups.} \checkmark \\ \text{Normal: 26, 474, 15, 485 are all } \geq 10 \checkmark \\ \text{Independent: There are at least } 1000 \times 10 = 10000 \text{ people similar to volunteers.} \checkmark \end{array} \right.$



conclude $\left\{ \begin{array}{l} \text{Because } p\text{-value} < \alpha (.0397 < .05) \text{ it is statistically significant,} \\ \text{we have sufficient evidence to conclude that taking} \\ \text{a low-dose aspirin each day would reduce the} \\ \text{chance of developing colon cancer among all} \\ \text{people similar to the volunteers.} \end{array} \right.$

4. A researcher conducted a medical study to investigate whether taking a low-dose aspirin reduces the chance of developing colon cancer. As part of the study, 1,000 adult volunteers were randomly assigned to one of two groups. Half of the volunteers were assigned to the experimental group that took a low-dose aspirin each day, and the other half were assigned to the control group that took a placebo each day. At the end of six years, 15 of the people who took the low-dose aspirin had developed colon cancer and 26 of the people who took the placebo had developed colon cancer. At the significance level $\alpha = 0.05$, do the data provide convincing statistical evidence that taking a low-dose aspirin each day would reduce the chance of developing colon cancer among all people similar to the volunteers?

State: Determine if ^{the data of} ~~the~~ taking low-dose aspirin each day would reduce the chance of developing colon cancer; is statistically significant. $\alpha = .05$

$$H_0: p_1 = p_2$$

$$H_a: p_1 < p_2$$

Plan: 2 Sample z-test for proportion if conditions are met.

Random: volunteers were randomly assigned to 2 groups.

normal: big enough sample size: $1,000 > 30$.

independent: $1000 \times 10 = 10,000$ - meets the 10% rule.

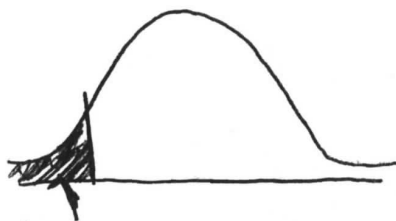
Do:

$$x_1 = 15$$

$$n = 500$$

$$x_2 = 26$$

$$n = 500$$



$$z = -1.75$$

$$p = .0397$$

Conclude: We can conclude that this data is statistically significant because the p-value is less than $\alpha = .05$ ($.04 < .05$). So we can conclude that taking a low dose aspirin ~~does not~~ every day does reduce the chance of getting colon cancer.

AP[®] STATISTICS

2015 SCORING COMMENTARY

Question 4

Overview

The primary goal of this question was to assess a student's ability to identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question. More specific goals were to assess a student's ability to (1) state appropriate hypotheses; (2) identify the appropriate statistical test procedure and check appropriate conditions for inference; (3) calculate the appropriate test statistic and p -value; and (4) draw an appropriate conclusion, with justification, in the context of the study.

Sample: 4A

Score: 4

In step 1 correct hypotheses are given and the parameters are clearly defined as population values for each group. The phrase "proportion of people who develop colon cancer after taking low-dose aspirin each day" clearly defines one of the parameters. Step 1 was scored as essentially correct. In step 2 a correct procedure is identified by name and the formula is shown with the numbers substituted in. The response notes "given random assignment" and provides formulas and expected counts under the null hypothesis for the normality check. The use of expected counts here rather than observed counts is acceptable. Though these formulas are in terms of population parameters rather than sample quantities, this was judged a minor error. As all three components are correct, step 2 was scored as essentially correct. In step 3 correct values of the test statistic and p -value are provided, so step 3 was scored as essentially correct. In step 4 linkage and a correct decision are provided by the statement "since $p < \alpha$, we reject H_0 ," and a correct conclusion in context is given. 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

Correct hypotheses are given in step 1. However, the parameters are incorrectly defined by stating the "true proportion of people who took medicine and get colon cancer." Use of the word "true" appears to reference a population quantity, but use of the past tense (took rather than take) appears to reference sample quantities and not population values. In addition, the phrasing does not imply a conditional proportion (of those who could have taken aspirin, the proportion who get colon cancer). With one correct component and one incorrect component, step 1 was scored as partially correct. In step 2 the response identifies the name of the test, correctly addresses the random assignment condition, and provides all four observed counts, indicating that they are all greater than or equal to 10. As all three components are satisfied, step 2 was scored as essentially correct. In step 3 correct values of the test statistic and p -value are included, and step 3 was scored as essentially correct. In step 4 the comparison of the p -value to 0.05 provides linkage. The phrase "it is statistically significant" is ambiguous. The response has a correct conclusion in context. Although an explicit decision (such as reject H_0) is not provided, it is not necessary when linkage and a correct conclusion in context are given. Step 4 was scored as essentially correct. Because three steps were scored as essentially correct, and one step was scored as partially correct, the response was initially scored as a 3½. Looking holistically at the paper, the error in step 1 was viewed as serious, and the statement about statistically significant was ambiguous. Therefore, the response earned a score of 3.

AP[®] STATISTICS
2015 SCORING COMMENTARY

Question 4 (continued)

Sample: 4C
Score: 2

In step 1 the response states hypotheses in terms of common parameters but uses subscripts that are not clearly relevant to the context of the problem and are not defined in the response. As a result, it is not possible to determine whether or not the relationship between parameters given in the alternative hypothesis is correct. Therefore neither component of step 1 is satisfied, and step 1 was scored as incorrect. In step 2 the response identifies a correct procedure by name, correctly addresses the randomness condition, but provides an incorrect check of the normality condition ($1,000 > 30$). Because two of the three components are satisfied, step 2 was scored as partially correct. In step 3 the test statistic and p -value are correctly stated, and step 3 was scored as essentially correct. In step 4 the response provides linkage between the p -value and $\alpha = .05$, and a conclusion is provided in context. But the conclusion is incorrect as it asserts definitively that taking a low dose of aspirin does reduce the chance of getting colon cancer. 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.