

# AP<sup>®</sup> STATISTICS

## 2012 SCORING GUIDELINES

### Question 4

#### **Intent of Question**

The primary goal of this question was to assess students' 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 students' ability to (1) state appropriate hypotheses; (2) identify the name of an appropriate statistical test and check appropriate assumptions/conditions; (3) calculate 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  $p_{07}$  represent the population proportion of adults in the United States who would have answered "yes" about the effectiveness of television commercials in December 2007. Let  $p_{08}$  represent the analogous population proportion in December 2008.

The hypotheses to be tested are  $H_0: p_{07} = p_{08}$  versus  $H_a: p_{07} \neq p_{08}$ .

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 these are sample surveys, the first condition is that the data were gathered from independent random samples from the two populations. This condition is met because we are told that the subjects were randomly selected in the two different years. Although we are not told whether the samples were selected independently, this is a reasonable assumption given that they are samples of different sizes selected in different years.

The second condition is that the sample sizes are large, relative to the proportions involved. This condition is satisfied because all sample counts (622 "yes" in 2007;  $1,020 - 622 = 398$  "no" in 2007; 676 "yes" in 2008;  $1,009 - 676 = 333$  "no" in 2008) are all at least 10 (or, are all at least 5).

An additional condition may be checked: The population sizes (more than 200 million adults in the United States) are much larger than 10 (or, 20) times the sample sizes.

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

The sample proportions who answered "yes" are:

$$\hat{p}_{07} = \frac{622}{1,020} \approx 0.6098 \quad \text{and} \quad \hat{p}_{08} = \frac{676}{1,009} \approx 0.6700.$$

The combined proportion,  $\hat{p}_c$ , who answered "yes" in these two years is:

$$\hat{p}_c = \frac{622 + 676}{1,020 + 1,009} = \frac{1,298}{2,029} \approx 0.6397.$$

**AP<sup>®</sup> STATISTICS**  
**2012 SCORING GUIDELINES**

**Question 4 (continued)**

The test statistic is:

$$z = \frac{\hat{p}_{07} - \hat{p}_{08}}{\sqrt{\hat{p}_c(1 - \hat{p}_c)\left(\frac{1}{n_{07}} + \frac{1}{n_{08}}\right)}} = \frac{0.6098 - 0.6700}{\sqrt{(0.6397)(1 - 0.6397)\left(\frac{1}{1,020} + \frac{1}{1,009}\right)}} \approx -2.82.$$

The  $p$ -value is  $2P(Z \leq -2.82) \approx 0.0048$ .

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

Because this  $p$ -value is smaller than any common significance level such as  $\alpha = 0.05$  or  $\alpha = 0.01$  (or, because this  $p$ -value is so small), we reject  $H_0$  and conclude that the data provide convincing (or, statistically significant) evidence that the proportion of all adults in the United States who would answer “yes” to the question about the effectiveness of television commercials changed from December 2007 to December 2008.

**Scoring**

Each of 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 parameter symbols in context, or simply using common parameter notation, such as  $p_{07}$  and  $p_{08}$ , with subscripts clearly relevant to the context, 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. Checks for randomness.
3. Checks for normality.

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

Incorrect (I) if the response correctly includes one or none of the three components listed above.

# AP<sup>®</sup> STATISTICS

## 2012 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.

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 decision in context, also providing justification based on linkage between the  $p$ -value and conclusion.

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

*OR*

if the response provides a correct decision 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.

*Note:* If the decision is consistent with an incorrect  $p$ -value from step 3, and also in context with justification based on linkage to the  $p$ -value, then step 4 is scored as E.

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 decide whether to score up or down, depending on the overall strength of the response and communication.

4. A survey organization conducted telephone interviews in December 2008 in which 1,009 randomly selected adults in the United States responded to the following question.

At the present time, do you think television commercials are an effective way to promote a new product?

Of the 1,009 adults surveyed, 676 responded "yes." In December 2007, 622 of 1,020 randomly selected adults in the United States had responded "yes" to the same question. Do the data provide convincing evidence that the proportion of adults in the United States who would respond "yes" to the question changed from December 2007 to December 2008?

$H_0: p_1 = p_2$        $p_1 =$  the true population proportion of U.S. adults who would respond "yes" to the question in December 2007  
 $H_a: p_1 \neq p_2$        $p_2 =$  the true population proportion of U.S. adults who would respond "yes" to the question in December 2008  
 There is a difference in the two proportions.

two-sample z test for a difference in proportions

Conditions

In December 2008, there were 1009 randomly selected adults  
 In Dec. 2007, there were 1020 randomly selected adults.  
 - Both samples randomly selected, condition met.

The two samples are also independent of each other.

$n_1 \hat{p}_1 \geq 10$	$n_2 (1 - \hat{p}_1) \geq 10$	$n_2 \hat{p}_2 \geq 10$	$n_2 (1 - \hat{p}_2) \geq 10$
$1020 \left( \frac{622}{1020} \right) \geq 10$	$1020 \left( 1 - \frac{622}{1020} \right) \geq 10$	$1009 \left( \frac{676}{1009} \right) \geq 10$	$1009 \left( 1 - \frac{676}{1009} \right) \geq 10$
622 $\geq$ 10	398 $\geq$ 10	676 $\geq$ 10	333 $\geq$ 10

All counts of successes and failures at least 10.

It is reasonable to assume that both sample sizes are less than 10% of the population of U.S. adult size in their respective years

$1009(10) = 10090$  adults       $1020(10) = 10200$  adults  
 (more than 10090 adults in 2008, and more than 10200 adults in 2007)

$\hat{p}_1 = .60980$

$\hat{p}_2 = .66997$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{p_c(1-p_c)}{n_1} + \frac{p_c(1-p_c)}{n_2}}} \quad \alpha = .05$$

$$p_c = \frac{676 + 622}{1009 + 1020} = .6397$$

$$Z = \frac{(.60980 - .66997)}{\sqrt{\frac{.6397(1-.6397)}{1020} + \frac{.6397(1-.6397)}{1009}}}$$

$Z = -2.8226$

p-value = .0048

Since our p-value of .0048 is less than  $\alpha = .05$ , reject  $H_0$ . There is evidence to suggest the proportion of adults who would respond "yes" to the question changed from December 2007 to December 2008.

4. A survey organization conducted telephone interviews in December 2008 in which 1,009 randomly selected adults in the United States responded to the following question.

At the present time, do you think television commercials are an effective way to promote a new product?

Of the 1,009 adults surveyed, 676 responded "yes." In December 2007, 622 of 1,020 randomly selected adults in the United States had responded "yes" to the same question. Do the data provide convincing evidence that the proportion of adults in the United States who would respond "yes" to the question changed from December 2007 to December 2008?

2. Proportion Z-Test is appropriate

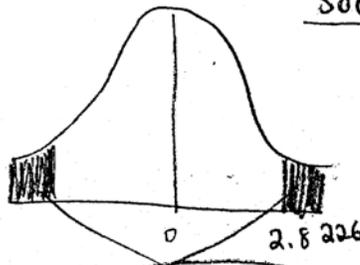
Randomization: 1,009 randomly selected adults in the US in 2008  
1,020 randomly selected adults in the US in 2007

2008	2007
$n_1 = 1009$	$n_2 = 1020$
$x_1 = 676$	$x_2 = 622$
$p_1 = 0.670$	$p_2 = 0.61$

10% condition: 1,009 randomly selected adults in the United States in 2008 and 1,020 randomly selected adults in the United States in 2007 are likely less than 10% of their respective populations.

$H_0: p_1 = p_2$

$H_A: p_1 \neq p_2$



Success/Failure:

- $(1009)(0.670) \geq 10$
- $(1009)(0.33) \geq 10$
- $(1020)(0.61) \geq 10$
- $(1020)(0.39) \geq 10$

$z = 2.8226 = \frac{0.34 - 0}{\sqrt{\frac{(0.670)(0.33)}{1009} + \frac{(0.61)(0.39)}{1020}}}$

Independent response of 1009 randomly selected ~~Groups~~ adults in 2008 is likely independent from 1020 randomly selected adults in 2007.

Since a p-value of 0.00476 is less than any reasonable significance level, we reject the null hypothesis. There is evidence to suggest that the proportion of adults in the United States who would respond "yes" to the question changed from December 2007 to December 2008.

4. A survey organization conducted telephone interviews in December 2008 in which 1,009 randomly selected adults in the United States responded to the following question.

At the present time, do you think television commercials are an effective way to promote a new product?

Of the 1,009 adults surveyed, 676 responded "yes." In December 2007, 622 of 1,020 randomly selected adults in the United States had responded "yes" to the same question. Do the data provide convincing evidence that the proportion of adults in the United States who would respond "yes" to the question changed from December 2007 to December 2008?

$$H_0: \hat{p}_1 = \hat{p}_2$$

$$H_a: \hat{p}_1 \neq \hat{p}_2$$

Random = yes

Independent = yes

< 10% pop = yes

#p<sub>1</sub>, #p<sub>2</sub> > 10 = yes  
#q<sub>1</sub>, #q<sub>2</sub>

$$\hat{p}_1 = \frac{676}{1009} = .6699$$

$$\hat{p}_2 = \frac{622}{1009} = .6164$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SE(\hat{p}_1 - \hat{p}_2)}$$

$$Z = \frac{.0535}{.0213}$$

$$Z = 2.51$$

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

$$SE(\hat{p}_1 - \hat{p}_2) = .0213$$

$$p\text{-value} = .012$$

I reject the null hypothesis and conclude that the data does support that the proportion of adults in the United States who would respond "yes" to the question changed from December 2007 to December 2008.

# AP<sup>®</sup> STATISTICS

## 2012 SCORING COMMENTARY

### Question 4

#### Overview

The primary goal of this question was to assess students' 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 students' ability to (1) state appropriate hypotheses; (2) identify the name of an appropriate statistical test and check appropriate assumptions/conditions; (3) calculate the appropriate test statistic and  $p$ -value; (4) draw an appropriate conclusion, with justification, in the context of the study.

#### Sample: 4A

#### Score: 4

In step 1 the hypotheses express the correct relationship between the parameters, and the parameters are correctly defined to be about the populations from which the samples were taken ("true population proportion of U.S. adults who would respond"). Step 1 was scored as essentially correct. In step 2 the test is correctly identified as a two-sample  $z$ -test for proportions, and the corresponding formula is provided, using a combined (pooled) estimate of the hypothesized common population proportion (either would have been sufficient). Both normality and randomness conditions are checked correctly, and the student clearly indicates that *both* samples were randomly selected. The check for approximate normality of the sampling distribution of  $\hat{p}_1 - \hat{p}_2$  is correctly done. The response would have been stronger if these four inequalities had been labeled as the check for normality. The student also addresses both types of independence that are relevant to the situation. Both are optional. The check that both sample sizes are less than or equal to 10 percent of the population sizes is done correctly. This condition verifies that the formula for the standard error gives an estimate that is not much larger than it would be if sampling had been done with replacement. (If a sample size is larger than 10 percent of the population size, another formula should be used.) Furthermore, the student states that the samples are independent of each other. That would be the case if the probability that any given group of 1,009 adults will make up the 2008 sample is equal to the probability that any other given group of 1,009 adults will make up the 2008 sample, *no matter who is selected for the 2007 random sample*. Because no information is given about how the random samples actually were selected, the statement that the two samples are independent of each other cannot be justified from the information given in the stem of the question, but it is a reasonable assumption to make about large national surveys. (Many students did not give a correct definition of what it means for the two samples to be independent, with statements such as "One adult's response doesn't affect another's.") Step 2 was scored as essentially correct. In step 3 the test statistic  $z$  and the corresponding  $p$ -value are correctly computed. Step 3 was scored as essentially correct. In step 4, the correct decision to reject the null hypothesis is justified by stating that the  $p$ -value is less than the (commonly used) significance level of .05. The last sentence presents the conclusion in context. Step 4 was scored essentially correct. Throughout, communication is strong, and the response is well organized and complete. With all four steps essentially correct, the response earned a score of 4.

**AP<sup>®</sup> STATISTICS**  
**2012 SCORING COMMENTARY**

**Question 4 (continued)**

**Sample: 4B**

**Score: 3**

In step 1, although the form of the hypotheses is correct,  $p_1$  and  $p_2$  are defined as the sample proportions with the equations  $p_1 = 0.670$  and  $p_2 = 0.61$ . Step 1 was scored as partially correct. In step 2 the two-proportion z-test is correctly identified, but the formula below does not use the pooled estimate of the hypothesized common population proportion. The two conditions of randomness and normality are checked correctly, with reference to two random samples. The optional independent condition is well done. Step 2 was scored as partially correct. In step 3 there is a minor error in the numerator of the test statistic, but the test statistic and  $p$ -value are correct. Step 3 was scored as essentially correct. In step 4, the correct decision to reject the null hypothesis is justified by saying that the  $p$ -value is less than any reasonable significance level. The final sentence provides correctly worded context. Step 4 was scored as essentially correct. Communication is strong, and the response is well organized. With two steps essentially correct and two steps partially correct, the response earned a score of 3.

**Sample: 4C**

**Score: 2**

In step 1, although the form of the hypotheses is correct, the notation  $\hat{p}_1$  and  $\hat{p}_2$  clearly indicates that the variables are the sample proportions, not the population parameters. Step 1 was scored as partially correct. In step 2 the test is identified by formula, but the formula given for the standard error does not use the pooled estimate of the hypothesized common population proportion. The word “Random” is listed, but there is no indication that it applies to *two* random samples. The normality check is incomplete. Step 2 was scored as incorrect. In step 3 the student makes a transcription mistake of using the 2008 sample size rather than the 2007 sample size in the computation of  $\hat{p}_2$ . Given this error, the values of the test statistic and  $p$ -value are correct. Step 3 was scored as essentially correct. In step 4 the decision is consistent with the  $p$ -value, and the conclusion has good context, but the decision is not justified by appeal to the size of the  $p$ -value (linkage). Step 4 was scored as partially correct. With one step essentially correct, two steps partially correct, and one step incorrect, the response earned a score of 2.