

# AP<sup>®</sup> STATISTICS

## 2016 SCORING GUIDELINES

### Question 5

#### **Intent of Question**

The primary goals of this question were to assess a student's ability to (1) construct and interpret a confidence interval for a population proportion; (2) explain why one of the conditions for inference is necessary; and (3) explain why a suggested procedure for constructing a confidence interval is incorrect.

#### **Solution**

##### **Part (a):**

The appropriate procedure is a one-sample z-interval for a population proportion. The problem stated the conditions for inference have been met, so they do not need to be checked. A 95 percent confidence interval for the population proportion is given as  $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ , which is

$$0.37 \pm 1.96 \sqrt{\frac{(0.37)(0.63)}{1,048}} \approx 0.37 \pm 0.03 = (0.34, 0.40).$$

We are 95 percent confident that the population proportion of all adults in the U.S. who would have chosen the economy statement is between 0.34 and 0.40.

##### **Part (b):**

The condition is necessary because the formula for the confidence interval relies on the fact that the binomial distribution can be approximated by a normal distribution which then results in the sampling distribution of  $\hat{p}$  being approximately normal. The approximation does not work well unless both  $n\hat{p}$  and  $n(1-\hat{p})$  are at least 10.

##### **Part (c):**

The suggested procedure is not appropriate because one of the requirements for using a two-sample z-interval for a difference between proportions is that the two proportions are based on two independent samples. In the situation described the two proportions come from a single sample and thus are not independent.

#### **Scoring**

This question is scored in three sections. Section 1 consists of the mechanics of the confidence interval in part (a), section 2 consists of interpreting the confidence interval in part (a), and section 3 consists of parts (b) and (c).

**Section 1** is scored as follows:

Essentially correct (E) if the response includes the following two components:

1. States the correct procedure by name or formula
2. Calculates the confidence interval

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

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

# AP<sup>®</sup> STATISTICS

## 2016 SCORING GUIDELINES

### Question 5 (continued)

Notes:

- A formula with correct values is sufficient for component 2.
- Component 2 can never be satisfied if the response contains an incorrect formula.

**Section 2** is scored as follows:

Essentially correct (E) if the response includes the following four components for the interval interpretation:

1. Estimates a proportion
2. Infers about the population
3. States 95 percent confidence
4. Includes context

Partially correct (P) if the response satisfies components 1 and 2 *AND* satisfies only one of components 3 or 4;

*OR*

if the response gives a correct interpretation of the confidence *level* in context without interpreting the specific interval.

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

Notes:

- Any indication of an inference to the adults sampled rather than the population does not satisfy component 2 and is scored as I.
- Stating values that are unrealistic as proportions (including blanks) in the interpretation lowers the score by one level (from E to P or from P to I).
- When both the interpretation of the interval and the level are given, only the interpretation of the interval is scored.

**Section 3** is scored as follows:

Essentially correct (E) if the response includes the following two components:

1. Part (b) states that the condition implies the sampling distribution of  $\hat{p}$  is approximately normal *OR* the normal approximation to the binomial distribution is appropriate.
2. Part (c) indicates the procedure is not appropriate because the two proportions come from a single sample (dependent) rather than two (independent) samples.

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

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

Notes:

- Referring to the sampling distribution needing to be normal without explicitly stating  $\hat{p}$  satisfies component 1.
- Referring to normal approximation must specify “to binomial” to satisfy component 1.
- Discussing the two-sample z-interval versus the two-proportion z-interval is extraneous information.

**AP<sup>®</sup> STATISTICS**  
**2016 SCORING GUIDELINES**

**Question 5 (continued)**

- In part (c) component 2 can never be satisfied if a procedure that is not possible for one sample of categorical data is suggested.

**4 Complete Response**

Three parts essentially correct

**3 Substantial Response**

Two parts essentially correct and one part partially correct

**2 Developing Response**

Two parts essentially correct and no parts partially correct

*OR*

One part essentially correct and one or two parts partially correct

*OR*

Three parts partially correct

**1 Minimal Response**

One part essentially correct

*OR*

No parts essentially correct and two parts partially correct

5A1

5A1

5. A polling agency showed the following two statements to a random sample of 1,048 adults in the United States.

Environment statement: Protection of the environment should be given priority over economic growth.

Economy statement: Economic growth should be given priority over protection of the environment.

The order in which the statements were shown was randomly selected for each person in the sample. After reading the statements, each person was asked to choose the statement that was most consistent with his or her opinion. The results are shown in the table.

	Environment Statement	Economy Statement	No Preference
Percent of sample	58%	37%	5%

- (a) Assume the conditions for inference have been met. Construct and interpret a 95 percent confidence interval for the proportion of all adults in the United States who would have chosen the economy statement.

I will construct a one sample z interval for p.

p = the true proportion of US adults who would choose the economy statement

Population: All US adults

$$\hat{p} \pm z^* \sqrt{\frac{p(\pm p)}{n}}$$

$$.37 \pm 1.96 \sqrt{\frac{.37(.63)}{1048}}$$

$$.37 \pm 1.96 (.0149)$$

$$.37 \pm .0292$$

$$(.3408, .3992)$$

I am 95% confident that the true proportion of US adults that would have chosen the economy statement is between 34.08% and 39.92%.

5A2  
of 2

5A2  
of 2

- (b) One of the conditions for inference that was met is that the number who chose the economy statement and the number who did not choose the economy statement are both greater than 10. Explain why it is necessary to satisfy that condition.

This condition allows the experimenter to assume that the distribution of sample proportions is approximately normal. Because we are estimating the probability of the sample proportion using the normal curve, it is necessary to satisfy this condition in order to proceed with the construction of the interval.

- (c) A suggestion was made to use a two-sample  $z$ -interval for a difference between proportions to investigate whether the difference in proportions between adults in the United States who would have chosen the environment statement and adults in the United States who would have chosen the economy statement is statistically significant. Is the two-sample  $z$ -interval for a difference between proportions an appropriate procedure to investigate the difference? Justify your answer.

No, it is not an appropriate procedure. A condition to run a two sample  $z$  procedure is that there must be two samples that are independent of each other. Because the participants were asked to choose between the two statements, the proportion of participants who chose the environment is dependent on the proportion of participants who chose the economy statement. This survey does not satisfy the conditions for a two-sample  $z$  procedure.

5. A polling agency showed the following two statements to a random sample of 1,048 adults in the United States.

Environment statement: Protection of the environment should be given priority over economic growth.

Economy statement: Economic growth should be given priority over protection of the environment.

The order in which the statements were shown was randomly selected for each person in the sample. After reading the statements, each person was asked to choose the statement that was most consistent with his or her opinion. The results are shown in the table.

	Environment Statement	Economy Statement	No Preference
Percent of sample	58%	37%	5%

- (a) Assume the conditions for inference have been met. Construct and interpret a 95 percent confidence interval for the proportion of all adults in the United States who would have chosen the economy statement.

Parameter: The proportion of all adults in the U.S. who would have chosen the economy statement at 95% confidence.

Assumptions: Random; States it

$$\text{Normal: } (37)/1,048 = 387.76 \geq 10 \quad (.63)/1,048 = 660.24 \geq 10$$

Independent: 10% of all adults in the U.S. > 1,048 adults

Name: One sample  $Z$ -interval

Intervals:  $Z^* = 1.96$

$$1 - .95 = \frac{.05}{2} = .025 \text{ area } \mu \sigma$$

$\downarrow \quad \uparrow \quad \downarrow$   
 $\text{invNorm}(.025, 0, 1)$   
 $-1.96$

$$.37 \pm (1.96) \sqrt{\frac{(.37)(.63)}{1,048}}$$

$$.37 \pm (1.96)(.0149)$$

$$.37 \pm .029$$

$$(.341, .399)$$

Conclusion: We are 95% confident that the true proportion of adults in the U.S. that would have chosen the economy statement is captured by the intervals (.341, .399).

5B2  
of 2

5B2  
of 2

- (b) One of the conditions for inference that was met is that the number who chose the economy statement and the number who did not choose the economy statement are both greater than 10. Explain why it is necessary to satisfy that condition.

It is necessary to satisfy that condition to make sure the distribution is approximately normal. If the condition was not met, then the confidence intervals would have not been correct.

- (c) A suggestion was made to use a two-sample  $z$ -interval for a difference between proportions to investigate whether the difference in proportions between adults in the United States who would have chosen the environment statement and adults in the United States who would have chosen the economy statement is statistically significant. Is the two-sample  $z$ -interval for a difference between proportions an appropriate procedure to investigate the difference? Justify your answer.

The two-sample  $z$ -interval for a difference between proportions would not work because they are coming from the same sample, rather than the two samples needed.

501  
0.42501  
0.42

5. A polling agency showed the following two statements to a random sample of 1,048 adults in the United States.

Environment statement: Protection of the environment should be given priority over economic growth.

Economy statement: Economic growth should be given priority over protection of the environment.

The order in which the statements were shown was randomly selected for each person in the sample. After reading the statements, each person was asked to choose the statement that was most consistent with his or her opinion. The results are shown in the table.

	Environment Statement	Economy Statement	No Preference
Percent of sample	58%	37%	5%

(a) Assume the conditions for inference have been met. Construct and interpret a 95 percent confidence interval for the proportion of all adults in the United States who would have chosen the economy statement.

$P$  = The true proportion of all adults in the United States who chose the economy statement

1. Random sample from population  $\rightarrow$  stated
2.  $n \leq 10\% \cdot N$   $1048 \leq 10$  (all adults in the United States)
3.  $n\hat{p} \geq 10$   $1048(.37) \geq 388$  } Therefore the sampling dist  
 $n(1-\hat{p}) \geq 10$   $1048(.63) \geq 660$  } of  $\hat{p}$  is approx. normal

1 proportion Z interval

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$.37 \pm 1.96 \sqrt{\frac{.37(.63)}{1048}} = .37 \pm .02923$$

$$(.341, .399)$$

We are 95% confident the true proportion of all adults in the United States who chose the economy statement is captured in the interval .341 to .399.

502  
of 2

502  
of 2

- (b) One of the conditions for inference that was met is that the number who chose the economy statement and the number who did not choose the economy statement are both greater than 10. Explain why it is necessary to satisfy that condition.

This condition is necessary because it proves the sampling distribution of  $\hat{p}$  is approximately normal. In order to complete a 1 proportion z interval, the sampling distribution must be approximately normal.

- (c) A suggestion was made to use a two-sample z-interval for a difference between proportions to investigate whether the difference in proportions between adults in the United States who would have chosen the environment statement and adults in the United States who would have chosen the economy statement is statistically significant. Is the two-sample z-interval for a difference between proportions an appropriate procedure to investigate the difference? Justify your answer.

No, the two-sample z-interval is not an appropriate procedure to investigate the difference. A z proportion z-interval is not used to measure differences, to measure differences you use a matched pair t-test or interval in which you use means not proportions, and you must know  $\sigma$ .

# AP<sup>®</sup> STATISTICS

## 2016 SCORING COMMENTARY

### Question 5

#### Overview

The primary goals of this question were to assess a student's ability to (1) construct and interpret a confidence interval for a population proportion; (2) explain why one of the conditions for inference is necessary; and (3) explain why a suggested procedure for constructing a confidence interval is incorrect.

#### Sample: 5A

##### Score: 4

In part (a) the response identifies the procedure as a “one sample  $z$  interval for  $p$ ,” which satisfies component 1 of section 1. The response shows calculations and the correct interval, which satisfies component 2. With both components of the mechanics of the confidence interval satisfied, section 1 was scored as essentially correct. In part (a) the response indicates the confidence interval is estimating a proportion, which satisfies component 1 of section 2. The response indicates the correct population as “US adults that would have chosen,” which satisfies component 2. The response correctly states “95% confident” and interprets the interval in context, which satisfies components 3 and 4. With the four components of the interpretation of the confidence interval satisfied, section 2 was scored as essentially correct. In part (b) the response states that the condition “allows the experimenter to assume that the distribution of sample proportions is approximately normal.” The response continues by stating the reason such a condition is necessary is so that the normal curve can be used in the development of the confidence interval. This response satisfies component 1 of section 3. In part (c) the response recognizes that what is given are dependent proportions and what is needed are two proportions from two independent samples, satisfying component 2. With both components satisfied, section 3 was scored as essentially correct. Because all three sections were scored as essentially correct, the response earned a score of 4.

#### Sample: 5B

##### Score: 3

In part (a) the response identifies the procedure as a “one sample  $z$ -interval.” The identification alone does not satisfy component 1 of section 1. However, the response also includes the correct formula with the correct values in the formula, satisfying both component 1 and component 2. With both components satisfied for the mechanics of the confidence interval, section 1 was scored as essentially correct. In part (a) the response indicates that the confidence interval is estimating a proportion, which satisfies component 1 of section 2. The response indicates the correct population as “all adults in the U.S. that would have chosen,” which satisfies component 2. The response correctly states “95% confident” and interprets the interval in context, which satisfies components 3 and 4. With the four components of interpretation satisfied, section 2 was scored as essentially correct. In part (b) the response indicates a need for a distribution to be approximately normal but does not indicate which distribution would fulfill the condition. Thus component 1 of section 3 is not satisfied. In part (c) the response indicates that the suggested procedure would not be appropriate because two samples are needed and only one sample was used in the poll. The requirements for component 2 are satisfied. With only one of the two components satisfied, section 3 was scored as partially correct. Because two sections were scored as essentially correct, and one section was scored as partially correct, the response earned a score of 3.

**AP<sup>®</sup> STATISTICS**  
**2016 SCORING COMMENTARY**

**Question 5 (continued)**

**Sample: 5C**

**Score: 2**

In part (a) the response states the correct procedure and shows a correct interval, satisfying component 1 and component 2 of section 1. With both components of the mechanics of the confidence interval satisfied, section 1 was scored as essentially correct. In part (a) the response indicates the confidence interval is estimating a proportion, which satisfies component 1 in section 2. However, use of the phrase “all adults in the United States who chose” is describing those individuals who participated in the poll and thus is referring to the sample proportion and not the population proportion. Therefore the response does not satisfy component 2. The response correctly states “95% confident” and interprets the interval in context, which satisfies components 3 and 4. However, because the response does not satisfy component 2 in the interpretation of the confidence interval, section 2 was scored as incorrect. In part (b) the response indicates a need for the sampling distribution of the sample proportion to be approximately normal, which satisfies component 1 of section 3. In part (c) the response suggests an incorrect procedure of “matched pair  $t$  test or interval” where no pairing exists; therefore, it does not satisfy component 2. Because the response satisfies only one of the two required components, section 3 was scored as partially correct. Because one section was scored as essentially correct, one section was scored as partially correct, and one section was scored as incorrect, the response earned a score of 2.