



**AP<sup>®</sup> Statistics  
2006 Scoring Guidelines  
Form B**

**The College Board: Connecting Students to College Success**

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 5,000 schools, colleges, universities, and other educational organizations. Each year, the College Board serves seven million students and their parents, 23,000 high schools, and 3,500 colleges through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT<sup>®</sup>, the PSAT/NMSQT<sup>®</sup>, and the Advanced Placement Program<sup>®</sup> (AP<sup>®</sup>). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities, and concerns.

© 2006 The College Board. All rights reserved. College Board, AP Central, APCD, Advanced Placement Program, AP, AP Vertical Teams, Pre-AP, SAT, and the acorn logo are registered trademarks of the College Board. Admitted Class Evaluation Service, CollegeEd, connect to college success, MyRoad, SAT Professional Development, SAT Readiness Program, and Setting the Cornerstones are trademarks owned by the College Board. PSAT/NMSQT is a registered trademark of the College Board and National Merit Scholarship Corporation. All other products and services may be trademarks of their respective owners. Permission to use copyrighted College Board materials may be requested online at: [www.collegeboard.com/inquiry/cbpermit.html](http://www.collegeboard.com/inquiry/cbpermit.html).

**Visit the College Board on the Web: [www.collegeboard.com](http://www.collegeboard.com).**

**AP Central is the official online home for the AP Program: [apcentral.collegeboard.com](http://apcentral.collegeboard.com).**

AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)

Question 1

**Intent of Question**

The primary purpose of this question is to assess a student's ability to read, interpret, and explain information contained in a cumulative relative frequency plot for a real estate company.

**Solution**

**Part (a):**

This point indicates that 40 percent of the sales agents at this real estate company had sales volume of \$300,000 or less in the month shown.

**Part (b):**

Eighty percent of the sales agents had sales volume of \$800,000 or less and 70 percent of the sales agents had sales volume of \$700,000 or less. Thus,  $0.8 - 0.7 = 0.1$  or 10 percent of the sales agents achieved monthly sales volumes greater than \$700,000 and not exceeding \$800,000.

**Part (c):**

There were no agents whose monthly sales volume was between \$1,000,000 and \$1,100,000.

**Part (d):**

The 80<sup>th</sup> percentile for the distribution of monthly sales volume by these agents during the preceding month is \$800,000. Therefore, an agent making more than \$800,000 will be in the top 20 percent.

**Scoring**

Parts (a), (c), and (d) are scored as essentially correct (E), partially correct (P), or incorrect (I). Part (b) is scored as essentially correct (E) or incorrect (I).

**Part (a)** is essentially correct (E) if both values are correctly identified and interpreted in the appropriate context.

Part (a) is partially correct (P) if incorrect values are read from the plot, but the interpretation is correct OR one of the values is not interpreted correctly in the context of the question.

Part (a) is incorrect (I) if the student fails to recognize the cumulative nature of the graph, for example, the student says 40 percent of the sales agents had a sales volume of \$300,000.

**Part (b)** is essentially correct (E) if 0.1 (or 10 percent) is provided AND work is shown or an appropriate explanation is provided.

Part (b) is incorrect (I) if an answer (0.1, 0.8, or anything else) is provided with no supporting work or explanation.

AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)

Question 1 (continued)

Note: Probability statements are OK and work may be provided on the graph.

Alternative solution to Part (b):

Let  $X$  represent monthly sales volume (in hundreds of thousands of dollars)

$P(7 < X \leq 8) = P(X \leq 8) - P(X \leq 7) = 0.8 - 0.7 = 0.1$ , so 0.1 (or 10 percent) of the sales agents achieved monthly sales volumes greater than \$700,000 and not exceeding \$800,000.

**Part (c)** is essentially correct (E) if the student indicates that none of the sales agents had a monthly sales volume between \$1 million and \$1.1 million.

Part (c) is partially correct (P) if the student recognizes that no events occurred between 10 and 11 but does not provide a correct description in the context of this problem.

Part (c) is incorrect (I) otherwise, for example, if the student says the number of agents with \$1 million in sales is the same as the number of agents with \$1.1 million in sales, or if the student says that the frequency of sales of \$1 million is the same as the frequency of sales of \$1.1 million.

**Part (d)** is essentially correct (E) if a minimum monthly sales volume above \$800,000 is identified or the student says anything above \$800,000 will qualify the agent for a bonus AND justification for selecting that value is provided using the complement rule or the graph.

Part (d) is partially correct (P) if the minimum monthly sales volume is specified as \$800,000 with no justification.

Part (d) is incorrect (I) if an incorrect minimum monthly sales volume is specified OR a value is specified without context.

Each essentially correct response is worth 1 point; each partially correct answer is worth  $\frac{1}{2}$  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 determine whether to score up or down depending on the strength of the response and communication.

**AP<sup>®</sup> STATISTICS**  
**2006 SCORING GUIDELINES (Form B)**

**Question 2**

**Intent of Question**

The primary goals of this question are to evaluate a student’s ability to: (1) identify and compute an appropriate confidence interval, after checking the necessary conditions; (2) interpret the interval in the context of the question; and (3) use the confidence interval to conduct an appropriate test of significance.

**Solution**

**Part (a):**

Step 1: Identifies the appropriate confidence interval by name or formula and checks appropriate conditions.

Two sample  $z$  confidence interval for  $p_D - p_N$ , the difference in the proportions of parts meeting specifications for the two shifts OR  $(\hat{p}_D - \hat{p}_N) \pm z^* \sqrt{\frac{\hat{p}_D(1 - \hat{p}_D)}{n_D} + \frac{\hat{p}_N(1 - \hat{p}_N)}{n_N}}$ .

- Conditions:
1. Independent random samples from two separate populations
  2. Large samples, so normal approximation can be used

The problem states that random samples of parts were selected from the two different shifts. We need to assume that these parts were produced independently. That is, each employee works the day shift or night shift, but not both, and the machine quality does not vary over time. Since the sample sizes are both 200 and the number of successes (188 and 180) and the number of failures (12 and 20) for each shift are larger than 10, it is reasonable to use the large sample procedures.

Step 2: Correct mechanics

$$\begin{aligned} \hat{p}_D &= \frac{188}{200} = 0.94 \quad \text{and} \quad \hat{p}_N = \frac{180}{200} = 0.90 \\ (0.94 - 0.9) \pm 2.0537 \sqrt{\frac{0.94 \times 0.06}{200} + \frac{0.9 \times 0.1}{200}} \\ 0.04 \pm 2.0537 \times 0.0271 \\ 0.04 \pm 0.0556 \\ (-0.0156, 0.0956) \end{aligned}$$

Step 3: Interpretation

Based on these samples, we can be 96 percent confident that the difference in the proportions of parts meeting specifications for the two shifts is between  $-0.0156$  and  $0.0956$ .

**AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)**

**Question 2 (continued)**

**Part (b):**

Since zero is in the 96 percent confidence interval, zero is a plausible value for the difference  $p_D - p_N$ , and we do not have evidence to support the manager’s belief. In other words, there does not appear to be a statistically significant difference between the proportions of parts meeting specifications for the two shifts at the  $\alpha = 0.04$  level.

**Scoring**

Part (a) is scored according to the number of correct steps. Each of the first three steps is scored as essentially correct (E) or incorrect (I). Part (b) is scored as essentially correct (E) or incorrect (I).

Notes for Step 1:

The student must identify an appropriate confidence interval and comment on both independence and large sample sizes in order to get this step essentially correct.

Minimum amount of information on independence and large sample sizes needed for an essentially correct response: independence with a check mark AND an indication that the number of successes and the number of failures is larger than 10 (or larger than 5) for both samples.

The student does not need to restate the fact that these are random samples.

Notes for Step 2: An identifiable minor arithmetic error in Step 2 will not necessarily change a score from essentially correct to incorrect.

<b>Alternative Solutions for Step 2</b>	
<b>Procedure</b>	<b>96% Confidence Interval</b>
Calculator	(−0.0155652, 0.0955652)
Wilson Estimator	(−0.0169858, 0.0961937)

**Part (b)** is essentially correct (E) if the student comments on the fact that zero is contained in the confidence interval and the justification links this outcome to a 96 percent confidence level, or a 0.04 significance level, and includes a statement indicating that the data do not support the manager’s belief that there is a difference in the proportion of parts that meet specifications produced by the two shifts.

Part (b) is incorrect (I) if the student says no because zero is in the confidence interval OR simply says no without providing relevant justification.

Note: If a 95 percent confidence interval is used, then the maximum score is 3.

AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)

Question 2 (continued)

**4 Complete Response**

All three steps of the confidence interval in part (a) are essentially correct, and part (b) is essentially correct.

**3 Substantial Response**

All three steps of the confidence interval in part (a) are essentially correct, and part (b) is incorrect.

OR

Two steps of the confidence interval in part (a) are essentially correct and part (b) is essentially correct.

**2 Developing Response**

Two steps of the confidence interval in part (a) are essentially correct, and part (b) is incorrect.

OR

One step of the confidence interval in part (a) is essentially correct, and part (b) is essentially correct.

**1 Minimal Response**

One step of the confidence interval in part (a) is essentially correct, and part (b) is incorrect.

OR

Part (b) is essentially correct.

**AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)**

**Question 3**

**Intent of Question**

The primary goals of this question are to assess a student's ability to: (1) recognize the random variable of interest, identify its probability distribution, and calculate a probability; (2) use basic probability rules to find a different probability; and (3) use the sampling distribution of the sample mean to identify a characteristic of the manufacturing process that will meet a given specification.

**Solution**

**Part (a):**

Let  $D$  represent the distance a randomly selected ball travels. Since  $D$  is normally distributed with a mean of 288 yards and a standard deviation of 2.8 yards, we find

$$P(D > 291.2) = P\left(Z > \frac{291.2 - 288}{2.8}\right) > P(Z > 1.14) = 1 - 0.8729 = 0.1271.$$

**Part (b):**

$$\begin{aligned} P(\text{at least one distance} > 291.2) &= 1 - P(\text{all five distances} \leq 291.2) \\ &= 1 - (1 - 0.1271)^5 \\ &= 1 - (0.8729)^5 \\ &= 1 - 0.5068 \\ &= 0.4932 \end{aligned}$$

**Part (c):**

Since the 99<sup>th</sup> percentile for a standard normal distribution is 2.33, we can set the appropriate z-score equal to 2.33 and solve for the desired mean, say  $M$ . Thus,

$\frac{291.2 - M}{2.8} = 2.33$  or  $M = 291.2 - 2.33 \times 2.8 = 284.676$ . In order to be 99 percent certain that a randomly selected ball does not exceed the maximum distance of 291.2 yards, the mean should be set to 284.676 yards.

**Scoring**

Parts (a), (b), and (c) are scored as essentially correct (E), partially correct (P), or incorrect (I).

**Part (a)** is essentially correct (E) if the student clearly shows ALL three of the following:

- Indicates the distribution is normal.
- Specifies BOTH the mean,  $\mu$ , and the standard deviation,  $\sigma$ .
- Calculates the correct probability.

AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)

Question 3 (continued)

Part (a) is partially correct (P) if the student:

Calculates the correct probability but fails to specify BOTH  $\mu$  and  $\sigma$  in the identification of the normal the distribution.

OR

Completely identifies the distribution as normal with BOTH  $\mu$  and  $\sigma$  correctly specified but fails to calculate the correct probability, e.g., calculates the probability that the ball travels less than 291.2 yards.

OR

Completely identifies the distribution as normal with BOTH  $\mu$  and  $\sigma$  correctly specified but uses the empirical rule to provide an approximate answer.

Part (a) is incorrect (I) if the student:

Reports a correct probability without showing any work.

OR

Calculates and incorrect probability using an inappropriate distribution.

Notes:

Calculator solution is 0.1265. If this is the only information provided, the response is scored as incorrect (I).

If only the calculator command `Normalcdf (-∞, 291.2, 288, 2.8)` is provided along with 0.1265, then the response should be scored as partially correct (P).

If the calculator command `Normalcdf (-∞, 291.2, 288, 2.8)` is provided along with 0.1265 AND the mean and standard deviation are clearly identified, then the response should be scored as essentially correct (E).

If the calculator command `Normalcdf (-∞, 291.2, 288, 2.8)` AND a shaded/labeled sketch of an appropriate normal distribution are provided along with 0.1265, then the response should be scored as essentially correct (E).

Minor arithmetic or transcription errors will not necessarily lower the score.

**Part (b)** is essentially correct (E) if the student calculates the correct probability and:

Clearly indicates the distribution is binomial AND specifies both  $n$  and  $p$  using the value obtained in part (a)

OR

Correctly applies complement and probability rules using the value obtained in part (a).

Part (b) is partially correct (P) if the student:

Clearly indicates the distribution is binomial AND specifies both  $n$  and  $p$  using the value obtained in part (a), but does not calculate the probability correctly.

OR

Calculates the correct probability using the value obtained in part (a) but fails to completely identify the distribution as binomial with both  $n$  and  $p$  specified.

OR

Indicates a correct procedure for computing the probability but uses a value of  $p$  that is different from the value obtained in part (a).

**AP<sup>®</sup> STATISTICS**  
**2006 SCORING GUIDELINES (Form B)**

**Question 3 (continued)**

Part (b) is incorrect (I) if the student

Provides a probability, but no work is shown.

OR

Obtains a probability with an incorrect solution strategy, e.g.,  $P(\text{at least one distance} > 291.2) = 1 - p^5$   
 or  $P(\text{at least one distance} > 291.2) = 1 - 5p$ , where  $p$  is the solution to part (a).

Notes:

Calculator solution is 0.4916. If the student uses calculator syntax, BOTH  $n$  and  $p$  AND the binomial distribution must be identified to be scored essentially correct.

If only the calculator command  $1 - \text{binomcdf}(5, 0.1265, 0)$  is provided along with the probability 0.4915, then the response should be scored as partially correct.

Alternative solutions using the binomial distribution with  $p = 0.1265$  are:

$$\begin{aligned}
 P(\text{at least one measurement} > 291.2) &= P(B = 1) + P(B = 2) + P(B = 3) + P(B = 4) + P(B = 5) \\
 &= \binom{5}{1} 0.1265^1 (1 - 0.1265)^4 + \binom{5}{2} 0.1265^2 (1 - 0.1265)^3 + \\
 &\quad \binom{5}{3} 0.1265^3 (1 - 0.1265)^2 + \binom{5}{4} 0.1265^4 (1 - 0.1265)^1 + 0.1265^5 \\
 &= 0.368224 + 0.106652 + 0.015445 + 0.001118 + 0.00032 \\
 &= 0.491472
 \end{aligned}$$

$$\begin{aligned}
 P(\text{at least one measurement} > 291.2) &= 1 - P(\text{all five distances} > 291.2) \\
 &= 1 - \binom{5}{0} 0.1265^0 (1 - 0.1265)^5 \\
 &= 1 - (1 - 0.1265)^5 \\
 &= 1 - (0.8735)^5 \\
 &= 1 - 0.5085 \\
 &= 0.4915
 \end{aligned}$$

**Part (c)** is essentially correct (E) if the student clearly shows ALL three of the following:

Identifies the 99<sup>th</sup> percentile for the standard normal distribution

Sets up an appropriate equation

Solves for the desired mean

Part (c) is partially correct (P) if the student:

Recognizes that the 99<sup>th</sup> percentile for the standard normal distribution must be used and sets up the appropriate equation but does not solve the equation for the desired mean.

AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)

Question 3 (continued)

OR

Sets up an appropriate equation using the correct minimum distance but provides an incorrect mean because an incorrect upper tail percentile (say the 99.5<sup>th</sup> percentile) was used.

OR

Sets up an appropriate equation using the correct percentile of the standard normal distribution but provides an incorrect mean because an incorrect minimum distance was used (say 288 yards) was used.

Part (c) is incorrect (I) if the student

Provides the correct mean, 284.676 yards, but no work is shown

OR

A lower tail percentile is used.

OR

An incorrect mean is calculated with an incorrect solution strategy.

Note: Calculator solution is 284.686 yards.

**4 Complete Response**

All 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 two parts partially correct

OR

Three parts partially correct

**1 Minimal Response**

One part essentially correct and either zero or one part partially correct

OR

No parts essentially correct and two parts partially correct

**AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)**

**Question 4**

**Intent of Question**

The primary goal of this question is to assess a student’s ability to conduct a test of significance by stating the hypotheses of interest, checking the necessary conditions, calculating the test statistic and  $p$ -value, and making a conclusion in context.

**Solution**

Step 1: States a correct pair of hypotheses.

Let  $\mu_D$  denote the mean difference (after – before) in dexterity scores for the population of individuals enrolled in the program.

$$H_0 : \mu_D = 0 \text{ versus } H_a : \mu_D > 0$$

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

One sample  $t$ -test or paired  $t$ -test or  $t = \frac{\bar{x}_D}{s_D/\sqrt{n}}$ .

We are told that the 12 people are a random sample. Assume that the differences (after – before) are approximately normal. This check may be done with a histogram, dotplot, stem-and-leaf display, or normal probability plot. The student should note that the normal assumption is not unreasonable because the plot displays no obvious skewness or outliers.

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

$$\bar{x}_D = 0.375, s_D = 0.367$$

$$\text{Degrees of freedom} = 12 - 1 = 11$$

$$t = \frac{0.375}{\frac{0.367}{\sqrt{12}}} = 3.54$$

$$p\text{-value} = 0.002$$

Step 4: States a correct conclusion in the context of the problem.

Since the  $p$ -value is less than 0.05, we can reject the null hypothesis of no difference in favor of the alternative and conclude that, on average, people who completed the program have significantly increased manual dexterity.

**Scoring**

Each of the four steps is scored as essentially correct (E) or incorrect (I).

**AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)**

**Question 4 (continued)**

Notes for Step 2:

Although it is not recommended, a one sample confidence interval for  $\mu_D$  could be used to test the hypotheses in Step 1. An appropriate adjustment to the confidence level must be made since we are conducting a one-sided test. The correct formula is

$$\left( \bar{x}_D - t_{n-1, \alpha}^* \frac{s_D}{\sqrt{n}}, \infty \right) \Rightarrow \left( 0.375 - 1.796 \times \frac{0.367}{\sqrt{12}}, \infty \right) \Rightarrow (0.1847, \infty).$$

The null hypothesis of no change in mean dexterity scores is rejected at the 0.05 level of significance because the right end of this 95 percent one sided confidence interval is above zero. If the  $t$ -value used to constructing the confidence interval does not match the significance level given in the conclusion, then the maximum score for Step 4 is partially correct (P).

If an incorrect two sample procedure is used, then Step 2 is scored as incorrect. The maximum score for a two sample  $t$  procedure is 3.

<b>Incorrect Solutions for Step 2</b>			
<b>Procedure</b>	<b>df</b>	<b>Test Statistic</b>	<b><math>p</math>-value</b>
Two sample $t$ -test	21.98	$t = 1.05$	0.153
Pooled $t$ -test	22	$t = 1.05$	0.152

A response using separate confidence intervals for the two means is also scored as incorrect for Step 2.

Notes for Step 3:

An identifiable minor arithmetic error in Step 3 will not necessarily change a score from essentially correct to incorrect.

If the student argues that the normal distribution is not reasonable, then they may use hypothesis tests for the median.

<b>Other Solutions for Step 3</b>		
<b>Procedure</b>	<b>Test Statistic</b>	<b><math>p</math>-value</b>
Sign Test	$B = 8$	0.0547
Wilcoxon Signed Rank Test	$W = 52$	0.007

If the  $p$ -value is incorrect but the conclusion in Step 4 is consistent with the computed  $p$ -value, Step 4 can be considered essentially correct.

Notes for Step 4:

If both an  $\alpha$  and a  $p$ -value are given, the linkage in Step 4 is implied. If no  $\alpha$  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.

If the hypotheses are reversed in Step 1 (i.e.,  $H_0 : \mu_D > 0$  versus  $H_a : \mu_D = 0$ ), then the conclusion also needs to be reversed. Otherwise, both parts should be scored as incorrect (I).

**Question 4 (continued)**

© 2006 The College Board. All rights reserved.

Visit [apcentral.collegeboard.com](http://apcentral.collegeboard.com) (for AP professionals) and [www.collegeboard.com/apstudents](http://www.collegeboard.com/apstudents) (for students and parents).

**AP<sup>®</sup> STATISTICS**  
**2006 SCORING GUIDELINES (Form B)**

- 4 Complete Response**  
All four steps essentially correct
- 3 Substantial Response**  
Three steps essentially correct
- 2 Developing Response**  
Two steps essentially correct
- 1 Minimal Response**  
One step essentially correct

AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)

Question 5

**Intent of Question**

The primary goals of this question are to assess a student's ability to: (1) identify the response variable, treatments, and experimental units in a study; (2) critique the use of randomization and replication; (3) recognize and explain why a particular variable is a confounding variable.

**Solution**

**Part (a):**

The response variable was the amount of draft. The two treatments were the standard hitch and the new hitch. The experimental units were the two large plots of land.

**Part (b):**

Yes, the two hitches (treatments) were randomly assigned to the two plots (experimental units).

**Part (c):**

No, each treatment (type of hitch) was applied to only one experimental unit (plot of land). Replication is used to repeat the treatments on different experimental units so general patterns can be observed. There is no replication in this study.

**Part (d):**

Although 25 measurements were taken at different locations in the two plots, each hitch was used in one plot (experimental unit) only. Thus, if a difference in the draft is observed we will not know whether the difference is due to the hitch or the plot. In statistical language, the treatments (hitches) are confounded with the plots.

**Scoring**

Parts (a), (b), (c), and (d) are scored as essentially correct (E), partially correct (P), or incorrect (I). Each essentially correct response is worth 1 point; each partially correct answer is worth  $\frac{1}{2}$  point.

**Part (a)** is essentially correct (E) if the response variable, treatments, and experimental units are correctly identified.

Part (a) is partially correct (P) if two of the three components of the experiment are correctly identified.

Part (a) is incorrect (I) if one or less of the three components of the experiment is correctly identified.

Note:

Responses to parts (b), (c) and (d) must be considered with respect to the experimental units identified in part (a).

AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)

Question 5 (continued)

**Part (b)** is essentially correct (E) if the student correctly discusses the use of randomization in this experiment with respect to assignment of the two hitches for use in the two plots OR with respect to the experimental units identified in part (a).

Part (b) is partially correct (P) if the student recognizes the use of randomization in the experiment but provides an incomplete or unclear discussion.

Part (b) is incorrect (I) if the student

Recognizes that randomization was used properly but does not provide a justification. That is, a naked answer of “YES” is scored as incorrect.

OR

Provides a discussion that does not address the issue of randomization, e.g., the student indicates that the plots should be more alike.

**Part (c)** is essentially correct (E) if the student recognizes that replication was not used properly and provides a correct justification.

Part (c) is partially correct (P) if the student recognizes that replication was not used properly but provides an incomplete justification that reveals some understanding of replication.

Part (c) is incorrect (I) if the student

Recognizes that replication was not used properly but does not provide a justification.

OR

Fails to recognize that replication was not properly used, e.g., incorrectly argues that the 25 measurements taken on each experimental unit (plot) provide proper replication.

**Part (d)** is essentially correct (E) if the student provides a valid explanation of confounding in this experiment.

Part (d) is partially correct (P) if the student provides an incomplete explanation that indicates an understanding of confounding in this experiment. For example, the student indicates that differences in plot conditions can affect draft but fails to link this to the inability to distinguish between plots differences and hitch effects.

Part (d) is incorrect (I) if the student

Provides a textbook definition of confounding with no attempt to describe the confounding variable in this experiment.

OR

Fails to address the issue of confounding.

Alternative solutions for part (d):

Each treatment was used in only one plot. Therefore, any differences caused by the differences in plots (e.g., soil hardness, moisture level, etc.) cannot be separated from differences in the two treatments.

Because only one plot of land is assigned to each hitch, if no difference is found it could be due to a superior hitch in a poor field (highly compacted) being compared with an inferior hitch in a good field. The effect of the hitch is masked by the differences in the plot

AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)

Question 5 (continued)

- 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 determine whether to score up or down depending on the strength of the response and communication.

**AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)**

**Question 6**

**Intent of Question**

The primary goals of this question are to evaluate a student’s ability to apply the concepts of significance testing to a new situation, in particular to: (1) state hypotheses for a parameter of interest given a research question; (2) recognize that a large sample test should not be used in this setting; (3) identify the possible values of a new test statistic and calculate the probability distribution for this new test statistic, assuming the null hypothesis is true; (4) use the probability distribution of the test statistic under the null hypothesis to identify possible significance levels; (5) conduct a significance test for a small set of data from an initial study; and (6) make a recommendation to improve on the initial study.

**Solution**

**Part (a):**

$$H_0 : p = 0.5 \text{ versus } H_a : p \neq 0.5$$

**Part (b):**

The conditions for the large sample one-proportion  $z$ -test are not satisfied.  $np = n(1 - p) = 8 \times 0.5 = 4 < 5$ .

**Part (c):**

$X$  will follow a binomial distribution with  $n = 8$  and  $p = 0.5$ . The possible values of  $X$  and their corresponding probabilities are given in the table below.

<b>X</b>	<b>Probability</b>
0	0.00391
1	0.03125
2	0.10937
3	0.21875
4	0.27344
5	0.21875
6	0.10937
7	0.03125
8	0.00391

**Part (d):**

No, there is no possible test with a  $p$ -value of exactly 0.05.

The probability that none of the individuals ( $X = 0$ ) or all of the individuals ( $X = 8$ ) prefer Citrus Fresh is  $2 \times 0.003906 = 0.0078$ , which is less than 0.05.

The probability that one or fewer of the individuals ( $X \leq 1$ ) or seven or more of the individuals ( $X \geq 7$ ) prefer Citrus Fresh is  $2 \times (0.003906 + 0.031250) = 0.070312$ , which is greater than 0.05.

**AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)**

**Question 6 (continued)**

**Part (e):**

For the preference data provided,  $X = 2$ . From the table of binomial probabilities computed in part (c), the probability that two or fewer of the individuals ( $X \leq 2$ ) or six or more of the individuals ( $X \geq 6$ ) prefer Citrus Fresh when  $p = 0.5$  is  $2 \times (0.003906 + 0.031250 + 0.109375) = 0.289062$ . Because the  $p$ -value of 0.289062 is greater than any reasonable significance level, say 0.070312, we would not reject the null hypothesis that  $p = 0.5$ . That is, we do not have statistically significant evidence for a consumer preference between Citrus Fresh and Tropical Taste.

**Part (f):**

Increase the number of consumers involved in the preference test. More consumers will give you more data, and you will be better able to detect a difference between the population proportion of consumers who prefer Citrus Fresh and 0.5. The sample proportion in the initial study was only 0.25 (2/8), but we were not able to reject the null hypothesis that  $p = \frac{1}{2}$ . By increasing the number of consumers, a difference of that magnitude would allow the null hypothesis to be rejected. For example, with  $n = 80$  and  $X = 20$  the large sample  $z$ -statistic would be  $z = \frac{0.25 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{80}}} = -4.47$  and the  $p$ -value would be approximately zero.

OR

The large sample test statistic is  $z = \frac{\hat{p} - 0.5}{\sqrt{\frac{(0.5)(0.5)}{n}}} = \sqrt{n} \left( \frac{\hat{p} - 0.5}{0.5} \right)$ , which would get larger as  $n$  is increased,

for the same  $\hat{p}$ . Thus, by using more consumers it would eventually be possible to reject  $H_0$  if there really is a difference and identify the preferred juice.

**Scoring**

Parts (a) and (b) are combined and scored as essentially correct (E), partially correct (P), or incorrect (I). Parts (c) and (d) are combined and scored as essentially correct (E), partially correct (P), or incorrect (I). Part (e) contains two parts, correct mechanics and conclusion, and is scored essentially correct (E), partially correct (P), or incorrect (I). Part (f) is scored as essentially correct (E), partially correct (P), or incorrect (I).

**Parts (a) and (b)** are essentially correct (E) if both parts are correct.

Parts (a) and (b) are partially correct (P) if one of the two parts is correct.

Notes for parts (a) and (b):

If a one-sided alternative is used in part (a), then the maximum score for the AB component is partially correct (P).

**AP<sup>®</sup> STATISTICS**  
**2006 SCORING GUIDELINES (Form B)**

**Question 6 (continued)**

Other reasonable statements about the conditions for inference not being met are acceptable. Some examples are:

- Since  $n = 8$ , the number of successes and the number of failures will both be less than 10.
- The sample size is too small to use the large-sample inference procedure.

A response in part (b) about a test for comparing proportions from two independent samples (one for Tropical Taste and one for Citrus Fresh) should be scored as incorrect (I).

**Parts (c) and (d)** are essentially correct (E) if both parts are correct.

Parts (c) and (d) are partially correct (P) if one of the two parts is correct.

Notes for parts (c) and (d):

Part (c) is correct if binomial probabilities are correctly calculated for each of the nine possible outcomes using  $n = 8$  and  $p = 0.5$ .

Part (d) is correct if the response includes all three of the following:

- The student states that a 0.05 level test is not possible.
- The justification includes a correct description of the test with significance level 0.00782.
- The justification includes a correct description of the test with significance level 0.07032.

**Part (e)** is essentially correct (E) if both parts are correct.

Part (e) is partially correct (P) if one part is correct.

Notes for part (e):

The mechanics are essentially correct if the student uses the information from the response to part (c) to compute the appropriate  $p$ -value.

- If a large sample  $z$ -test, or a  $t$ -test, is presented, the mechanics are scored as incorrect.
- Using only 0.10937, the probability that  $X$  is 2 from the table in part (c), is incorrect.

To be essentially correct, a correct conclusion must be stated in the context of the problem with linkage between a  $p$ -value and a level of significance  $\alpha$ .

- If the student refers to the rejection region identified in part (d), they do not need to restate the significance level from part (e).
- If the student conducts an  $\alpha = 0.05$  level test by comparing the correct  $p$ -value to 0.05, even though they just said in part (d) that a 0.05 level test is not possible, the response should be scored as essentially correct.

**Part (f)** is essentially correct (E) if the provides a reasonable recommendation in the context of comparing preferences for the two juices with an appropriate statistical justification. For example, the student describes the impact of a larger sample size on the hypothesis testing procedure used in the initial study by saying that increasing the sample size will reduce the chance of making a Type II error, increasing the sample size will increase the power of the test used to detect a difference in juice preferences in part (d), or increasing the sample size will reduce the standard error for the estimated proportion of people who prefer Citrus Fresh.

Part (f) is partially correct (P) if the student provides a reasonable recommendation in the context of comparing preferences for the two juices but does not provides sufficient statistical justification that refers to the initial

AP<sup>®</sup> STATISTICS  
2006 SCORING GUIDELINES (Form B)

Question 6 (continued)

study, e.g., a larger sample is needed to meet the conditions for using a z-test, larger samples are more representative of the population, or increasing the sample size makes the sample or population more normal

OR

Gives a weak recommendation and provides a good statistical justification, for example, providing the subjects with a cracker between juices to reduce possible carryover effects.

Incorrect (I) if the student fails to provide a reasonable recommendation in the context of comparing preferences for the two juices.

Each essentially correct response is worth 1 point; each partially correct response is worth  $\frac{1}{2}$  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 determine whether to score up or down depending on the strength of the response and communication.**