



## **AP<sup>®</sup> Statistics 2004 Scoring Guidelines**

**The materials included in these files are intended for noncommercial use by AP teachers for course and exam preparation; permission for any other use must be sought from the Advanced Placement Program<sup>®</sup>. Teachers may reproduce them, in whole or in part, in limited quantities, for face-to-face teaching purposes but may not mass distribute the materials, electronically or otherwise. This permission does not apply to any third-party copyrights contained herein. These materials and any copies made of them may not be resold, and the copyright notices must be retained as they appear here.**

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 4,500 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three 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.

For further information, visit [www.collegeboard.com](http://www.collegeboard.com)

Copyright © 2004 College Entrance Examination Board. All rights reserved. College Board, Advanced Placement Program, AP, AP Central, AP Vertical Teams, APCD, Pacesetter, Pre-AP, SAT, Student Search Service, and the acorn logo are registered trademarks of the College Entrance Examination Board. PSAT/NMSQT is a registered trademark of the College Entrance Examination Board and National Merit Scholarship Corporation. Educational Testing Service and ETS are registered trademarks of Educational Testing Service. Other products and services may be trademarks of their respective owners.

For the College Board's online home for AP professionals, visit AP Central at [apcentral.collegeboard.com](http://apcentral.collegeboard.com).

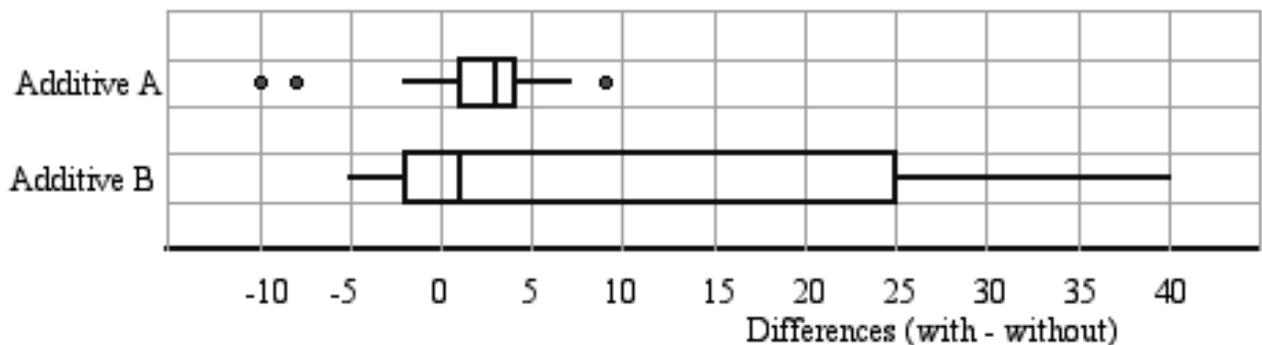
**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 1**

**Solution**

**Part (a):**

	Additive A	Additive B
IQR	$4 - 1 = 3$	$25 - (-2) = 27$
$1.5 \times \text{IQR}$	4.5	40.5
$Q_1 - 1.5 \times \text{IQR}$	$1 - 4.5 = -3.5$	$-2 - 40.5 = -42.5$
$Q_3 + 1.5 \times \text{IQR}$	$4 + 4.5 = 8.5$	$25 + 40.5 = 65.5$
	3 outliers	no outliers



**Part (b i):**

Additive A is better at increasing the mileage in the greatest number of cars. The mileage increased for at least seventy-five percent of the cars with additive A, whereas the mileage decreased for more than twenty-five percent of the cars with additive B.

**Part (b ii):**

Additive B appears to produce a higher mean mileage gain than additive A. The boxplot for additive B clearly shows that the distribution of differences is skewed to the right, which will pull the average towards the larger values. The mean difference for additive B will be substantially greater than the median of 1. On the other hand, the distribution of differences for additive A has much less variability, as seen by comparing the lengths of the two boxes, and appears to be skewed to the left. The mean difference for additive A will be less than the median of 3.

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 1 (cont'd.)**

**Scoring**

Each part is scored as essentially correct, partially correct, or incorrect.

**Part (a)** is essentially correct if parallel boxplots are correctly drawn and the display includes labels. The labels for additives A and B are required, but if the label for either axis (or both axes) is the only item missing then part (a) is still scored as essentially correct.

Part (a) is partially correct if only one of the following occurs:

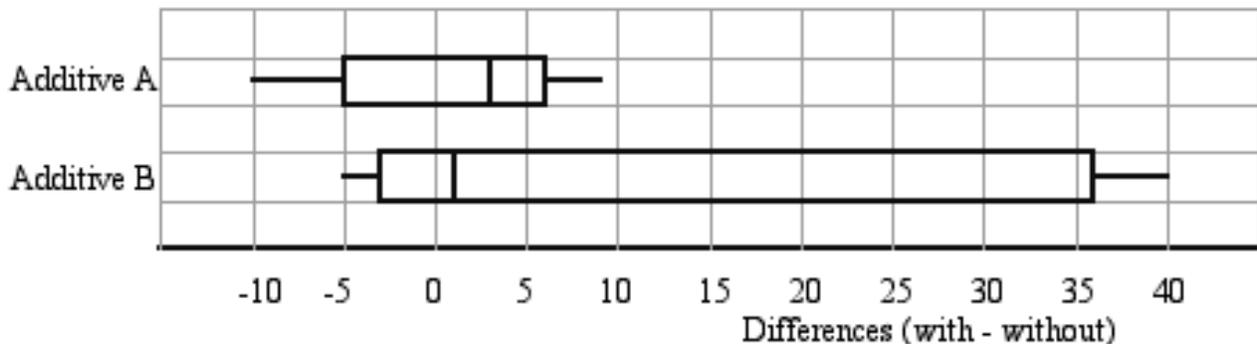
- if neither additive A nor B is labeled, or the labels are reversed, OR
- there is an error in the construction of the boxplots, OR
- outliers are not correct, but otherwise the boxplots are correct.

Note: If the same error is present in the boxplot for Additive A as in the boxplot for Additive B, count this as one error, not two.

Part (a) is incorrect if either of the following occur:

- 2 or 3 of the errors above are committed OR
- the numbers provided in the stem of the problem are used as a data set to form the boxplots below.

**INCORRECT BOXPLOTS USING SUMMARY STATISTICS AS DATA**



Note: If the labeling is reversed for boxplots A and B, score parts (b i) and (b ii) consistently with the reversed labels.

**AP<sup>®</sup> STATISTICS**  
**2004 SCORING GUIDELINES**

**Question 1 (cont'd.)**

**Part (b i)** is essentially correct if additive A is selected as the additive that will increase the mileage in the greatest number of cars and the explanation involves a comparison implicitly or explicitly using the first quartiles from **both** distributions.

Part (b i) is partially correct if any of the following occur:

- additive A is selected, but the only comparison is between the lower quartile of additive A and the median of additive B, OR
- additive A is selected as the additive that will increase the mileage in the greatest number of cars but the comparison is weak or incomplete, such as relying only on variability or only on the median, OR
- additive A is selected and justified by addressing either A or B, but there is no explicit comparison to the other additive.

Part (b i) is scored as incorrect if any of the following occur:

- additive A is selected, but no explanation is provided, OR
- additive A is selected, but the explanation is incorrect, OR
- additive B is selected, OR
- summary statistics are used correctly in part (a), but used as data in part (b).

**Part (b ii)** is essentially correct if additive B is selected as the additive that will produce a higher mean mileage gain and the explanation involves a comparison using the shapes of **both** distributions.

For additive A it is sufficient to refer to the effect of the outliers on the mean.

Part (b ii) is partially correct if any of the following occur:

- additive B is selected as the additive that will produce a higher mean mileage gain but the comparison is weak or incomplete, OR
- additive B is selected and justified by addressing either A or B, but there is no explicit comparison to the other additive, OR
- additive A is selected, because the student argues that medians should be used to measure center instead of means because of the outliers or because the distributions are skewed, OR
- a comparison is made between A and B, but only one side, either left or right, is addressed for both plots.

Part (b ii) is scored as incorrect if any of the following occur:

- additive B is selected, but no explanation is provided,
- additive B is selected, but the explanation is incorrect,
- additive A is selected for any reason other than that stated above for partially correct,
- summary statistics are used correctly in part (a), but used as data in part (b), OR
- the student claims that the means cannot be determined from the information provided.

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 1 (cont'd.)**

**4 Complete Response**

All three parts essentially correct

**3 Substantial Response**

Two parts essentially correct and 1 part partially correct

**2 Developing Response**

2 parts essentially correct and no parts partially correct

OR

One part essentially correct and 2 parts partially correct

OR

3 parts partially correct

**1 Minimal Response**

One part essentially correct and either 0 or 1 part partially correct

OR

No parts essentially correct and 2 parts partially correct

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 2**

**Solution**

**Part (a):**

Block	Volunteers	Ages
1	1, 2	20, 21
2	10, 11	23, 24
3	8, 9	44, 44
4	3, 12	46, 47
5	4, 7	58, 60
6	5, 6	61, 62

Since these researchers believe that the condition of hair changes with age but not gender, the volunteers are sorted from youngest to oldest. The volunteers in the sorted list are paired to form six blocks of size two. More specifically, the youngest two volunteers are placed in the first block. The next two volunteers in the sorted list are placed in the second block. This pairing continues until all six blocks of two are formed, with the oldest two volunteers in the sixth block.

**Part (b):**

Block	Volunteers	Ages
Female 1	2, 10	20, 24
Female 2	8, 12	44, 46
Female 3	4, 5	60, 62
Male 1	1, 11	21, 23
Male 2	3, 9	47, 44
Male 3	6, 7	61, 58

Since these researchers believe that the condition of hair changes with both age and gender, the women are sorted from youngest to oldest and then the men are sorted from youngest to oldest. The women (men) in the sorted list are paired to form the blocks of size two. More specifically, the youngest two women (men) are placed in a block. The next two youngest women (men) are placed in another block. Finally, the oldest two women (men) are placed in another block.

**Part (c):**

No, the researchers in part (b) should not randomly select three blocks to receive the new formula and then give the current formula to the other three blocks. They blocked on both age and gender to form homogeneous groups because they believe hair condition differs with both age and gender. Giving the youngest or oldest women (men) the same formula defeats the purpose of blocking. In a block design, randomization should be carried out separately within each block. That is, for each block, two random numbers are generated (via a random number generator or a table of random digits) and assigned to the two volunteers. The volunteer with the smallest random number is given the new formula and the other volunteer is given the current formula.

**AP<sup>®</sup> STATISTICS**  
**2004 SCORING GUIDELINES**

**Question 2 (cont'd.)**

**Scoring**

Each part is scored as essentially correct, partially correct, or incorrect.

**Parts (a) and (b)** are each scored as essentially correct if

1. the six blocks are correctly identified by volunteer number or age AND
2. the correct criterion for assigning volunteers to blocks is clearly stated.

**Parts (a) and (b)** are scored as partially correct if either component 1 or 2 above is correct.

- Notes:
1. If at most two volunteers are misplaced, but it is clear that the student is blocking correctly, component 1 can be considered to be correct.
  2. Simply saying “block by age” or “block by age and gender” is not sufficient to get credit for component 2. A student must indicate that the volunteers in each block are similar with respect to age in part (a) or age and gender in part (b). However, if the student has correct blocks but only says to “block by age” in part (a) and “block by age and gender” in part (b), this should only be penalized once.
  3. In part (a), if a student blocks correctly by age, each block happens to contain one male and one female. If the student says or implies that this is an important part of the blocking criterion, then component 2 is incorrect. If they simply comment or notice this fact but do not imply or say that it is part of the criterion, component 2 could be correct.

**Part (c)** is essentially correct if the student recognizes that this is not an appropriate way to assign treatments AND describes an appropriate method for assigning treatments within each block, including a method of randomization that can be implemented by the reader.

Note: Simply saying “use a random number table” or “flip a coin” is not sufficient to get credit for the method of randomization

**Part (c)** is partially correct if the student recognizes that this is not an appropriate way to assign treatments AND either

1. assigns treatments randomly within blocks but lacks a method of randomization, OR
2. assigns treatments within each block with no mention of randomization, OR
3. gives a correct reason why the assignment is not appropriate.

**Part (c)** is incorrect if

1. the student says that this is an appropriate way to assign treatments, OR
2. “NO” is the only response, OR
3. “NO” is the response, but the reason is wrong or absent AND either
  - i) no indication is given about assigning treatments within blocks OR
  - ii) an alternative design is proposed (e.g., crossover, completely randomized).

Note: If in part (b) the student makes heterogeneous blocks (i.e., blocks contain one male and one female of disparate ages), then saying “YES” with an appropriate method of randomization is essentially correct for part (c). In the same situation, if the student says to assign treatments to the blocks with no randomization method, this is scored partially correct for part (c).

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 2 (cont'd.)**

**4 Complete Response**

All three parts essentially correct

**3 Substantial Response**

Two parts essentially correct and 1 part partially correct

**2 Developing Response**

2 parts essentially correct and no parts partially correct

OR

One part essentially correct and 2 parts partially correct

OR

3 parts partially correct

OR

One part essentially correct and 1 part partially correct

**1 Minimal Response**

One part essentially correct and 1 part partially correct

OR

One part essentially correct and 0 parts partially correct

OR

No parts essentially correct and 2 parts partially correct

Note: One part essentially correct and one part partially correct may be given a score of either 1 or 2 holistically.

**AP<sup>®</sup> STATISTICS**  
**2004 SCORING GUIDELINES**

**Question 3**

**Solution**

**Part (a):**

$X$  is not binomial since the trials are not independent and the conditional probabilities of selecting a male change at each trial depending on the previous outcome(s), due to the sampling without replacement.

**Part (b):**

$$P(X = 4) = \left(\frac{10}{20}\right)\left(\frac{9}{19}\right)\left(\frac{8}{18}\right)\left(\frac{7}{17}\right) = \frac{5040}{116280} = 0.043$$

**Part (c):**

No. If males and females were equally represented, the probability of observing four males is small (0.043).

**Part (d):**

No, we can't generalize to the population of all brontosaurus because it is not reasonable to regard this sample as a random sample from the population of all brontosaurus; there is reason to suspect that this sampling method might cause bias.

**Scoring**

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

**Part (a):** Score as:

**Essentially correct** if the response indicates that

- (i) trials are not independent, with an explanation that independence means the outcome on any trial will not impact the probability of success on future trials OR
- (ii) the probability of selecting a male on any given trial depends on the results of previous trials.

**Partially correct** if the response indicates that

- (i) the student is focusing on one of the concepts above, but discussion is weak OR
- (ii) there is sampling without replacement without connection to one of the concepts under Essentially Correct above.

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 3 (cont'd.)**

**Part (b):** Score as:

**Essentially correct** if the probability is correctly computed (with minor arithmetic errors being overlooked), with supporting work or rationale. A statement that this is a hypergeometric distribution (either in Part(a) or Part(b)) will suffice. It is OK if the student leaves the answer as the product of fractions. The probability that all four femurs belong to males can also be

computed by using the formula 
$$P(X = 4) = \frac{\binom{10}{4}\binom{10}{0}}{\binom{20}{4}} = .043.$$

**Partially correct** if there is a correct answer (to 3 decimal places) with incomplete justification.

**Incorrect** if arithmetic errors result in a probability that is negative or greater than one.

**Part (c):** Score as:

**Essentially correct** if the probability provided in Part (b) is interpreted correctly.

**Partially correct** if it is not clear that the student used the probability from Part (b).

**Incorrect** if just a “Yes” or “No” is given without an explanation.

**Part (d):** Score as:

**Essentially correct** if the response indicates that generalization is not possible because this sample

- (i) cannot be viewed as a random sample of all brontosaurus OR
- (ii) there is reason to suspect that this sample might not be representative of the population at large.

**Incorrect** if “No” is given without an explanation.

Note: Discussions about conditions for inference are irrelevant.

Each essentially correct response is worth 1 point; each partially correct answer is worth ½ point.

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 3 (cont'd.)**

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

IF A PAPER IS BETWEEN TWO SCORES (FOR EXAMPLE, 2½ 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  
2004 SCORING GUIDELINES**

**Question 4**

**Solution**

**Part (a):**

Let A be the event “antibiotic A works.”

Let B be the event “antibiotic B works.”

The probability that a child will be cured with Plan I is:

$$\begin{aligned}P(\text{Cure}_I) &= P(A) + P(\text{not } A)P(B) \\ &= 0.6 + (0.4 \times 0.9) \\ &= 0.96\end{aligned}$$

The probability that a child will be cured with Plan II is:

$$\begin{aligned}P(\text{Cure}_{II}) &= P(B) + P(\text{not } B)P(A) \\ &= 0.9 + (0.1 \times 0.6) \\ &= 0.96\end{aligned}$$

**Part (b):**

Treatment with antibiotic A costs \$50, and treatment with antibiotic B costs \$80.

The expected cost per child when Plan I is used for treatment is:

$$\begin{aligned}E(\text{Cost}_I) &= \$50 \times 0.6 + \$130 \times 0.4 \\ &= \$30 + \$52 \\ &= \$82\end{aligned}$$

The expected cost per child when Plan II is used for treatment is:

$$\begin{aligned}E(\text{Cost}_{II}) &= \$80 \times 0.9 + \$130 \times 0.1 \\ &= \$72 + \$13 \\ &= \$85\end{aligned}$$

**Part (c):**

Since the probability that a child will be cured is the same under either plan, some other criterion must be used to make a recommendation. From a financial point of view, Plan I should be recommended because the expected cost per child is less than Plan II.

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

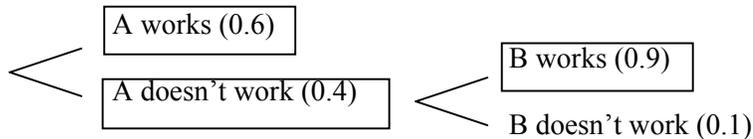
**Question 4 (cont'd.)**

**Scoring**

Each part is scored as essentially correct, partially correct, or incorrect.

**Part (a)** is essentially correct if the probabilities of cure are calculated correctly with justification for both plans.

Plan I:



$$P(\text{Cure}_I) = 0.6 + (0.4 \times 0.9) = 0.96$$

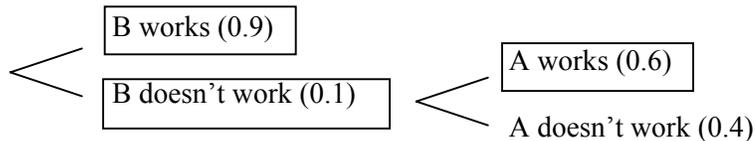
OR

$$P(\text{Cure}_I) = P(A \cup B) = 0.6 + 0.9 - (0.6 \times 0.9) = 0.96$$

OR

$$P(\text{Cure}_I) = 1 - P(\text{not } A)P(\text{not } B) = 1 - (0.4) \times (0.1) = 0.96$$

Plan II:



$$P(\text{Cure}_{II}) = 0.9 + (0.1 \times 0.6) = 0.96$$

OR

$$P(\text{Cure}_{II}) = P(B \cup A) = 0.9 + 0.6 - (0.9 \times 0.6) = 0.96$$

OR

$$P(\text{Cure}_{II}) = 1 - P(\text{not } B)P(\text{not } A) = 1 - (0.1) \times (0.4) = 0.96$$

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 4 (cont'd.)**

**Part (a)** is partially correct if

one of the two probabilities is calculated correctly with justification,

OR

both probabilities are correct with incomplete justifications.

**Part (b)** is essentially correct if the expected costs per child are calculated correctly with justification for both plans.

The expected cost per child when Plan I is used for treatment is:

$$\begin{array}{l} E(\text{Cost}_I) = \$50 \times 0.6 + \$130 \times 0.4 \\ = \$30 + \$52 \\ = \$82 \end{array} \quad \text{OR} \quad \begin{array}{l} E(\text{Cost}_I) = \$50 + 0.4 \times \$80 \\ = \$50 + \$32 \\ = \$82 \end{array}$$

The expected cost per child when Plan II is used for treatment is:

$$\begin{array}{l} E(\text{Cost}_{II}) = \$80 \times 0.9 + \$130 \times 0.1 \\ = \$72 + \$13 \\ = \$85 \end{array} \quad \text{OR} \quad \begin{array}{l} E(\text{Cost}_{II}) = \$80 + 0.1 \times \$50 \\ = \$80 + \$5 \\ = \$85 \end{array}$$

**Part (b)** is partially correct if

the expected cost per child is calculated correctly with justification for one of the two plans,

OR

both expected costs are correct with incomplete justifications,

OR

the expected costs are incorrectly calculated but the probabilities involved add up to 1. For example the following computations would receive a partial.

The expected cost per child when Plan I is used for treatment is:

$$= \$50 \times 0.6 + \$80 \times 0.4 = \$30 + \$32 = \$62$$

The expected cost per child when Plan II is used for treatment is:

$$= \$80 \times 0.9 + \$50 \times 0.1 = \$72 + \$5 = \$77$$

**AP<sup>®</sup> STATISTICS**  
**2004 SCORING GUIDELINES**

**Question 4 (cont'd.)**

In contrast, the following computations would receive an incorrect because the probabilities involved do not add up to 1.

The expected cost per child when Plan I is used for treatment is:

$$= \$50 \times 0.6 + \$130 \times 0.36 = \$30 + \$46.80 = \$76.80$$

The expected cost per child when Plan II is used for treatment is:

$$= \$80 \times 0.9 + \$130 \times 0.06 = \$72 + \$7.80 = \$79.80$$

**Part (c)** is essentially correct if the recommendation contains a statistical argument based on parts (a) and (b). That is, the student must base the recommendation on probabilities from part (a) and expected values from part (b). The following two examples are essentially correct:

Since the probability that a child will be cured is the same under either plan, some other criterion must be used to make a recommendation. From a financial point of view, Plan I should be recommended because the expected cost per child is less than Plan II.

Since the probability that a child will be cured is the same under either plan, some other criterion must be used to make a recommendation. Parents might prefer Plan II, regardless of its higher cost, because their child is more likely to need only the first drug.

**Part (c)** is partially correct if the recommendation contains a statistical argument based only on part (a) or (b) but not both.

**Part (c)** is incorrect if no recommendation is made.

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 4 (cont'd.)**

**4 Complete Response**

All three parts essentially correct

**3 Substantial Response**

Two parts essentially correct and 1 part partially correct

**2 Developing Response**

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and 2 parts partially correct

OR

Three parts partially correct

**1 Minimal Response**

One part essentially correct and either 0 or 1 part partially correct

OR

No parts essentially correct and 2 parts partially correct

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 5**

**Solution**

**Part a:**

$H_0$ : gender and satisfaction with health services offered by the hospital are independent (OR not associated)

$H_a$ : gender and satisfaction with health services offered by the hospital are dependent (OR associated)

Chi-square test for association

$$\text{Test statistic: } \chi^2 = \sum_{\text{all cells}} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Conditions: A random sample has been taken. The expected cell counts are large enough so that the chi-square approximation can be used. (See the table below for the expected cell counts.) That is, all four of the expected cell counts are at least 5 (or the smallest expected cell count is 92.8 which is greater than 5). We can use the chi-square approximation.

Expected counts are printed below observed counts

	Male	Female	Total
1	384 371.20	416 428.80	800
2	80 92.80	120 107.20	200
Total	464	536	1000

$$\begin{aligned} \text{Chi-Sq} &= 0.441 + 0.382 + 1.766 + 1.528 = 4.117 \\ \text{DF} &= 1, \text{ P-Value} = 0.042 \end{aligned}$$

Because the p-value, 0.042, is less than 0.05, we can reject  $H_0$  at significance level 0.05, and conclude that there is evidence of an association between gender and satisfaction with health services offered by the hospital for adult residents of this county.

**Part b:**

Because a random sample has been taken from the population of all county residents, 0.8 is a reasonable estimate for the proportion of all county residents who are satisfied with the services offered by this hospital.

**AP<sup>®</sup> STATISTICS**  
**2004 SCORING GUIDELINES**

**Question 5 (cont'd.)**

**Scoring**

**Part (a):** BECAUSE THE ALTERNATIVE HYPOTHESIS IS GIVEN IN THE PROBLEM, HYPOTHESES AND CONDITIONS WILL BE COMBINED INTO ONE STEP (a1) AND SCORED AS ESSENTIALLY CORRECT OR INCORRECT. Each of the OTHER 2 steps (a2 and a3) of the hypothesis test is scored either as correct or incorrect.

**Step a1:** States a correct pair of hypotheses. Identifies a correct test (by name or by formula) and states and verifies appropriate conditions.

Note: Conditions checked must refer to expected counts. Counts must be verified as sufficiently large. A table of expected counts is sufficient to verify that the student has checked the counts. Stating that 92.8 is the smallest expected count is sufficient to verify that the student has checked counts.

The fact that the observed cell counts are from a random sample does not need to be repeated here.

**Step a2:** Correct mechanics, including the value of the test statistic ( $\chi^2 = 4.117$ ) and p-value (0.042)  
OR  
the test statistic ( $\chi^2 = 4.117$ ) and rejection region ( $\chi^2$  critical value = 3.84 when  $\alpha = 0.05$ ).

**Step a3:** Stating a correct conclusion in the context of the problem linked to both the p-value (or rejection region) and the stated hypotheses.

If an  $\alpha$  and a p-value are given, the linkage 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 p-value in step a2 is incorrect but the conclusion is consistent with the computed p-value, step a3 can be considered as correct.

A conclusion consistent with incorrectly stated hypotheses is scored essentially correct.

NOTE: This chi-squared test can also be done as a two-sample  $z$  test of the difference between two proportions. The major steps of the hypothesis test are still scored as correct or incorrect. That is, the student must state the hypotheses, identify the test by name or formula, check conditions, complete the correct mechanics, and state the correct conclusion in the context of the problem.

$$H_0 : p_1 - p_2 = 0$$

$$H_a : p_1 - p_2 \neq 0$$

where  $p_1$  represents the proportion of males who are satisfied, and  $p_2$  represents the proportion of females who are satisfied.

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 5 (cont'd.)**

Conditions for the two-sample  $z$  hypothesis test of proportions

$$\begin{aligned}n_1\hat{p} &= 371.2 > 5 & n_1(1 - \hat{p}) &= 92.8 > 5 & \text{where } \hat{p} &= \frac{384 + 416}{464 + 536} = 0.8 \\n_2\hat{p} &= 428.8 > 5 & n_2(1 - \hat{p}) &= 107.2 > 5\end{aligned}$$

OR

$$\begin{aligned}n_1\hat{p}_1 &= 384 > 5 & n_1(1 - \hat{p}_1) &= 80 > 5 & \text{where } \hat{p}_1 &= 0.827586 \\n_2\hat{p}_2 &= 416 > 5 & n_2(1 - \hat{p}_2) &= 120 > 5 & \text{where } \hat{p}_2 &= 0.776119\end{aligned}$$

Computer output illustrating the correct mechanics for the two-sample  $z$  test and confidence interval is provided below.

Sample	X	N	Sample p
1	384	464	0.827586
2	416	536	0.776119

Estimate for  $p_1 - p_2$  : 0.0514668

95% CI for  $p_1 - p_2$  : (0.00220611, 0.100727)

Test for  $p_1 - p_2 = 0$  (vs  $p_1 - p_2 \neq 0$ ): Z = 2.03 P-Value = 0.042

Standard deviation for  $p_1 - p_2 = \sqrt{(0.8)(0.2)\left[\frac{1}{464} + \frac{1}{536}\right]} = 0.02536$

This  $p$ -value of 0.042 is less than  $\alpha = 0.05$  so we can reject the null hypothesis. We have sufficient evidence that the proportion of males who are satisfied with the hospital services is different than the proportion of females who are satisfied with the hospital services.

**Part (b)** is scored as essentially correct, partially correct, or incorrect.

Part (b) is scored as essentially correct if 0.8 is viewed as a reasonable estimate because the sample is taken from the population of interest.

NOTE: If a student provides the correct response and then states that the estimate could be improved by conditioning then the response should be scored essentially correct.

**AP<sup>®</sup> STATISTICS**  
**2004 SCORING GUIDELINES**

**Question 5 (cont'd.)**

Part (b) is scored as partially correct

- if the results of the hypothesis test in part (a) are used to make a statement about the applicability of the estimate. That is, 0.8 is viewed as an unreasonable estimate because there is a significant association between gender and satisfaction.
- if the estimate is unreasonable because the number of men and women are not equal in the sample.

Part (b) is scored as incorrect

- if the student simply says yes or no, without any explanation.
- if the student says yes, the sample size is large.

**4 Complete Response**

Part (b) essentially correct and all 3 steps of the hypothesis test are correct

**3 Substantial Response**

Part (b) essentially correct and 2 steps of the hypothesis test correct

OR

Part (b) partially correct or incorrect and 3 steps of the hypothesis test correct

OR

Part (b) partially correct AND either the hypotheses or conditions are correct AND both a2 and a3 are correct

**2 Developing Response**

Part (b) essentially correct and 1 step of the hypothesis test correct

OR

Part (b) partially correct or incorrect and 2 steps of the hypothesis test correct

OR

Part (b) partially correct AND either the hypotheses or conditions are correct AND either a2 or a3 is correct

**1 Minimal Response**

Part (b) essentially correct

OR

Part (b) partially correct or incorrect and 1 step of the hypothesis test is correct

OR

Part (b) partially correct AND either the hypotheses or conditions are correct

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 6**

**Solution**

**Part (a):**

Step 1: States and checks appropriate conditions.

We are told that the sample was randomly selected. Since the sample size is large (e.g.,  $n = 50 > 30$ ), the one sample  $t$  interval should be valid. Alternatively, we could assume that the reduction in cholesterol level after one month is (at least approximately) normally distributed, but we have no way to check this assumption with the information provided.

Step 2: Identifies the appropriate confidence interval by name or formula.

One sample  $t$  interval for  $\mu$ , the mean reduction in cholesterol for the new drug or  $\bar{x} \pm t_{n-1}^* \frac{s}{\sqrt{n}}$ .

Step 3: Correct mechanics.

Degrees of freedom =  $n - 1 = 49$ .

$$\bar{x} \pm t_{n-1}^* \frac{s}{\sqrt{n}} = 24 \pm 2.0096 \frac{15}{\sqrt{50}} = 24 \pm 4.2631 = (19.7369, 28.2631).$$

Step 4: Interprets the confidence interval in context.

We are 95% confident that the mean reduction in cholesterol for the new drug in the population of people with high cholesterol is between 19.74 and 28.26 mg/dl.

**Part (b):**

The decision based on a 95% confidence interval only corresponds to the two-sided test of significance at the 5% level, not necessarily the one-sided test. The confidence interval in (a) is equivalent to testing  $H_0: \mu = 20$  against  $H_a: \mu \neq 20$ . In this test, the tail probability would be doubled, and this two-sided p-value, .066, is larger than .05, failing to reject the null hypothesis. However, in testing  $H_0: \mu = 20$  against  $H_a: \mu > 20$ , the one-sided p-value of .033 is small enough to reject  $H_0$  at the 5% level.

Alternatively, if we had compared the p-value of .033 to an alpha level of .025, the conclusions would match.

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 6 (cont'd.)**

**Part (c):**

The critical value for the lower confidence bound is the 95<sup>th</sup> percentile (instead of the 97.5<sup>th</sup> percentile) of the  $t$  distribution with 49 degrees of freedom.

$$t^* = 1.676$$

and

$$L = 24 - 1.676 \frac{15}{\sqrt{50}} = 24 - 3.5553 = 20.4447$$

**Part (d):**

Yes, the decision would change. Since the lower bound  $L$  is more than 20, the agency would now be convinced that  $\mu$  is greater than 20 and the new drug is statistically significantly better than the current drug.

**Scoring**

**Part (a)** is scored as

**Essentially correct** if all four steps (check conditions, identify procedure, calculate interval, interpret interval) are correct. Each step is scored as correct or incorrect, no partial credit is given for the steps.

- Ok if only state “ $n$  is large” or if assume population distribution is normal with some justification/recognition that this is only an assumption.
- Student can use either  $t$  interval or  $z$  interval but needs to give name of procedure (state  $t$  or  $z$  somewhere) or provide critical value.
- A correct interpretation of the confidence level does not count for step 4. An incorrect interpretation of level prevents credit for correct interpretation of interval.

**Partially correct** if two or three steps are correct.

**Incorrect** if at most one step is correct.

**Part (b)** is scored as

**Essentially correct** if the student discussion includes:

1. the confidence interval is two-sided and the test is one-sided, and
2. a quantitative linkage between the procedures (e.g., doubling the p-value, halving the level of significance, using 90% as the confidence level [with clear connection to the  $\alpha = .05$  significance level]).

**AP<sup>®</sup> STATISTICS**  
**2004 SCORING GUIDELINES**

**Question 6 (cont'd.)**

**Partially correct if**

- Student states only that the confidence interval is two-sided and the test of significance is one-sided.
- Student has the right concept but discussion is very poor.

**Incorrect if**

- Student solution is only a restatement of the conclusions of each procedure.
- Student solution is only a restatement that the results differ.
- Student solution states that a 90% confidence interval is shorter and would not include 20 (but gives no indication of how 90% relates to the level of significance of the test) or that they might not differ if a different level of  $\alpha$  was used (with no value stated).
- Student says confidence intervals and test are different procedures (e.g., vary in specificity) and lead to different conclusions.
- Student says one uses the  $t$  value and the other uses the  $z$  value.

**Note:** Students may refer to the duality between confidence intervals and hypothesis tests in their solutions, but they must describe this duality in order to receive credit.

**Part (c)** is scored as

**Essentially correct if**

1. a reasonable one-sided value of  $t^*$  is given ( $\pm 1.676$  or  $\pm 1.684$ ), and
2.  $L$  is calculated correctly using the value of  $t^*$  provided (20.445 or 20.428).

**Partially correct** if the student identifies an incorrect reasonable critical value ( $z^*$ , two-sided  $t^*$  but see note below, or wrong df) and then uses this value to calculate  $L$ .

**Incorrect** if the student identifies a nonsensical critical value (e.g., uses the test statistic, the p-value,  $\alpha$ ,  $t^*/2$ , 20) and/or obtains a lower bound larger than the sample mean (unless it is from an incorrectly substituted one-sided  $t^*$  value).

**Part (d)** is scored as

**Essentially correct if**

1. the student correctly justifies whether or not the conclusion has changed, and
2. the student makes a correct conclusion from the one-sided confidence interval and supports their conclusion (e.g., compares  $L$  to the value of 20, believes  $\mu \geq L$ , is 20 in the interval).

**Partially correct** if student gives only component 2 or has component 1 with weak justification.

**Incorrect** if an answer (yes or no) is provided with no reasonable explanation.

**NOTE:** If the student's work in part (c) is merely a reworking of the lower endpoint in (a) and they obtain the same lower bound and comment that the calculation is the same, they will receive at most one point among parts (a), (c) and (d). "ACD Rule"

**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 6 (cont'd.)**

Each Essentially Correct response counts as 1 point, each Partially Correct response counts as ½ point.

<b>4</b>	<b>Complete Response</b>	4 points
<b>3</b>	<b>Substantial Response</b>	3 points
<b>2</b>	<b>Developing Response</b>	2 points
<b>1</b>	<b>Minimal Response</b>	1 point

IF A PAPER IS BETWEEN TWO SCORES (FOR EXAMPLE, 2½ POINTS) USE A HOLISTIC APPROACH TO DETERMINE WHETHER TO SCORE UP OR DOWN DEPENDING ON THE STRENGTH OF THE RESPONSE AND COMMUNICATION.