

AP[®] Statistics 2003 Scoring Guidelines

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

These materials were produced by Educational Testing Service® (ETS®), which develops and administers the examinations of the Advanced Placement Program for the College Board. The College Board and Educational Testing Service (ETS) are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

The College Board is a national nonprofit membership association whose mission is to prepare, inspire, and connect students to college and opportunity. Founded in 1900, the association is composed of more than 4,300 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three million students and their parents, 22,000 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®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of equity and excellence, and that commitment is embodied in all of its programs, services, activities, and concerns.

For further information, visit www.collegeboard.com

Copyright © 2003 College Entrance Examination Board. All rights reserved. College Board, Advanced Placement Program, AP, AP Vertical Teams, APCD, Pacesetter, Pre-AP, SAT, Student Search Service, and the acorn logo are registered trademarks of the College Entrance Examination Board. AP Central is a trademark owned by the College Entrance Examination Board. PSAT/NMSQT is a registered trademark jointly owned by the College Entrance Examination Board and the 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.

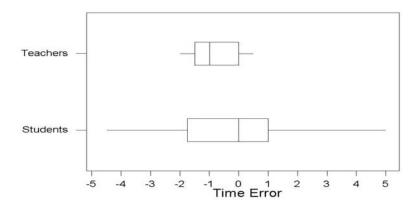
For the College Board's online home for AP professionals, visit AP Central at apcentral.collegeboard.com.

Question 1

Solution

Part (a):

Students	Teachers
Min -4.5	-2.0
Q ₁ -1.75 (or -0.5)	-1.5
Median 0	-1.0
Q_3 1.0 (or 0.5)	0
Max 5.0	0.5
no outliers	no outliers



Part (b):

The teachers' watches tend to have times that are closer to the true time. Although the teachers' watches tend to be slow, the times are less variable than the student times and so more teachers' watches had times that were close to the true time.

Part (c):

Testing this pair of hypotheses will not answer the teacher's question. The mean amount of time by which student watches differ from the true time could be zero even when student watch times differ greatly from the true time if large positive differences are offset by large negative differences.

Question 1 (cont'd)

Scoring:

Each part is essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is essentially correct (E) if parallel boxplots are correctly drawn (using either version of students' quartiles), includes correct labels (students vs. teachers) and shows a common scale. If only one group is labeled, we'll infer the correct label for the other group. A label on the scale is not needed, but should be correct if included. The five-number summary is not required, but may be used to compensate for an incorrectly placed point on a boxplot.

Part (a) is partially correct (P) if there is only one error (e.g., omitting labels and/or scale, showing outliers, highly inconsistent scale, one misplaced point, an incorrect method of calculating quartiles, or reversing labels). Minor arithmetic errors will be overlooked.

NOTE: Two separate boxplots are acceptable as parallel plots providing the scales for the two plots are the same so that comparisons can be easily made.

Part (b) is essentially correct (E) if teachers are picked based on some form of variability (e.g., range, compactness, IQR, spread, min-to-max, etc.).

Part (b) is partially correct (P) if either "Teachers" are selected with an incorrect explanation or no explanation, OR "Students" are selected based on center.

Note: If Q1 = -0.5 and Q3 = 0.5 are being used for the students, then a choice of students based on IQR OR a choice of teachers based on range is scored essentially correct (E). An argument for students based on center stills scores as partially correct (P). A choice of one group over the other without indicating which measure of variability has been used is incorrect (I), but considering both issues without making a choice is an example of parallel solutions – each of which is essentially correct (E).

Part (c) is essentially correct (E) if the response says that testing these hypotheses will not answer the question posed and a plausible explanation is given. Plausible explanations should focus on things such as: positives and negatives canceling out, individual times versus group average, or variability. Reference to the boxplots is not expected.

Part (c) is partially correct (P) if the answer given is "No" and either there is no explanation <u>and</u> part (b) was scored as essentially correct (E),

OR

There is an incomplete or incorrect explanation that addresses the connection (or lack of connection) between the given hypotheses and the teacher's question. [Note: Answers that discuss conditions needed to *perform* the test do not address the issue of connection. Such responses are scored as incorrect (I), as are responses that attempt to change the teacher's question.]

Question 1 (cont'd)

4 Complete Response (3E)

All three parts essentially correct

3 Substantial Response (2E 1P)

Two parts essentially correct and 1 part partially correct

2 Developing Response (2E 0P or 1E 2P or 3P)

Two parts essentially correct and no parts partially correct OR
One part essentially correct and 2 parts partially correct
OR
two parts partially correct

1 Minimal Response (1E 1P or 1E 0P or 0E 2P)

One part essentially correct and either 0 or 1 parts partially correct $\ensuremath{\mathsf{OR}}$

No parts essentially correct and 2 parts partially correct

Question 2

Solution

Part (a):

p, θ , or π = proportion of all cars of the specified make and model that have the defect.

H_o:
$$p = 0.05$$
 OR $p \le 0.05$
H_a: $p > 0.05$

Part (b):

Definition

Type I error: The law firm believes that the proportion of cars that have the defect is greater than 0.05, when in fact it is not.

Type II error: The law firm is not convinced that the proportion of cars that have the defect is greater than 0.05, when in fact it is.

Consequence

Type I error:

The firm will not recover its expenses, resulting in a loss to the firm.

Type II error:

The firm will miss an opportunity to make money on this case.

OR

Definition and Consequence

Type I error:

The firm will take the case when they should not have. The firm will not recover its expenses, resulting in a loss to the firm.

AND

Type II error:

The firm will refuse to handle the suit when it really should have. The firm will miss an opportunity to make money on this case.

Question 2 (cont'd)

Scoring

This problem is scored in three components. Part (a) is one component and part (b) is broken into two components for scoring purposes. Each component is then scored as either essentially correct (E), partially correct (P), or incorrect (I).

Component 1 (part a) is essentially correct (E) if it

- 1. correctly defines the parameter of interest AND
- 2. gives correct null and alternative hypotheses.

Component 1 is partially correct (P) if only one of the two elements above is present.

Component 2 (from part b) is essentially correct (E) if

- 1. correctly defines type I and type II errors for the stated hypotheses in part (a) AND
- 2. the definition of the two types of errors is in the context of the problem and consistent with the hypotheses given in part (a).

Component 2 is partially correct (P) if

The definitions of type I and type II errors are correct but not in context.

OR

The definitions are in context, but type I and type II errors are reversed.

OR

The definition of only one error is given correctly and in context.

Component 3 (from part b) is essentially correct (E) if

- 1. it discusses the consequences of each type of error AND
- 2. the discussion of consequences is consistent with the definition of type I and type II errors given in component 2.

Component 3 is partially correct (P) if

The consequences of only one type of error are discussed correctly.

OR

The student identifies consequences of two types of error, but the consequences are associated with the wrong type of error (given the definitions of type I and type II errors in component 2).

OR

The student correctly identifies for both errors that they take the case or that they do not take the case, but does not give economic consequences.

Question 2 (cont'd)

4 Complete Response (3E)

All three components essentially correct

3 Substantial Response (2E 1P)

Two components essentially correct and 1 component partially correct

2 Developing Response (2E 0P or 1E 2P or 3P)

Two components essentially correct and no components partially correct OR

One component essentially correct and 2 components partially correct OR

Three components partially correct

1 Minimal Response (1E 1P or 1E 0P or 0E 2P)

One component essentially correct and either 0 or 1 components partially correct OR

No components essentially correct and 2 components partially correct

Question 3

Solution

Part (a):

 $P(necksize < 14 \ or \ necksize \ge 18)$

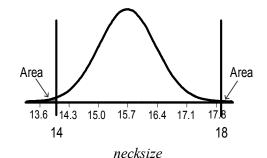
$$= P(necksize < 14) + P(necksize \ge 18)$$

$$= P\bigg(z < \frac{14 - 15.7}{0.7}\bigg) + P\bigg(z \ge \frac{18 - 15.7}{0.7}\bigg)$$

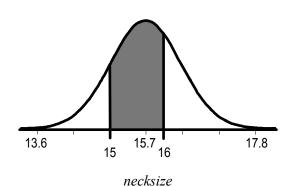
$$= P(z < -2.429) + P(z \ge 3.286)$$

$$= 0.00758 + 0.00051$$

=0.00809



Part (b):



$$P(15 \le necksize < 16)$$

$$= P\left(\frac{15 - 15.7}{0.7} \le z < \frac{16 - 15.7}{0.7}\right)$$

$$= P(-1.000 \le z < 0.429)$$

$$= 0.50723$$

Part (c):

X = number of customers who request size M

X is binomial with n = 12 customers and p = 0.5072

$$P(X = 4) = {}_{12}C_4(0.5072)^4(0.4928)^8 = 495(0.06618)(0.00348) = 0.1139$$

Question 3 (cont'd)

Scoring

Each part is scored as either essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is essentially correct (E) if the response

- 1. recognizes the need to look at neck sizes below 14 and above 18
- 2. correctly computes the two tail probabilities (except for minor arithmetic or transcription errors) and adds those probabilities

Part (a) is partially correct (P) if the response

considers only neck sizes below 14 (or above 18) but computes that corresponding tail area correctly OR

recognizes the need to look at neck sizes below 14 and above 18 but does not compute both tail probabilities correctly

OR

recognizes the need to look at neck sizes below 14 and above 18 but approximates tail probabilities using the Empirical Rule

OR

computes the proportion of customers that will find the store <u>carries</u> their size (i.e., 1 -correct answer)

OR

States the correct answer (0.0081) without supporting work

NOTE: A normal curve with correct regions shaded showing both correct end points (14 and 18) and the mean and the standard deviation may be used for element 1.

Part (b) is essentially correct (E) if

- 1. the appropriate probability is illustrated using a normal curve in which the end points are identified and the mean and standard deviation are implied
- 2. the required probability is correctly computed (except for minor arithmetic errors)

Part (b) is partially correct (P) if only one of the above elements is correct.

NOTES:

- (1) If part (a) was not essentially correct because the student interchanged the mean and standard deviation, and the same values for mean and standard deviation are used in part (b), then part (b) can be considered essentially correct if the probability calculated is correct for the mean and standard deviation used.
- (2) A reasonable approximation using the Empirical Rule in part (b) is only acceptable if the computation in part (a) is done correctly (i.e., without using the Empirical Rule).

Question 3 (cont'd)

Part (c) is essentially correct (E) if

- 1. the student recognized the setting as binomial
- 2. the probability calculated in part (b) is used for p
- 3. work is shown that is, the correct values for *n* and *x* are given and the desired probability calculated, or the binomial formula is correctly evaluated.

Part (c) is partially correct (P) if

the student recognizes the situation as binomial and identifies p from part (b) but does not compute the desired probability

OR

the student computes the probability as either $(0.5072)^4 (0.4928)^8$ or $\binom{12}{4} (0.5072)^4$

OR

the student gives the correct probability of 0.1139 but work is not shown

NOTE: Rounding the probability in part (b) for use in part (c) is acceptable.

4 Complete Response (3E)

All three parts essentially correct

3 Substantial Response (2E 1P)

Two parts essentially correct and one part partially correct

2 Developing Response (2E 0P or 1E 2P or 3P)

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 (1E 1P or 1E 0P or 0E 2P)

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

No parts essentially correct and two parts partially correct

Copyright © 2003 by College Entrance Examination Board. All rights reserved.

Available at apcentral.collegeboard.com.

Question 4

Solution

Part (a):

For example, a deadline in the department where the group of volunteers works has been moved back, lowering the stress levels of those working in the department. If the volunteers from this department were all in the same treatment group, this <u>change</u> in <u>stress level</u> could mistakenly be attributed to the treatment.

Without random assignment of volunteers to the two programs, it is possible that the two treatment groups could differ in some way that affects the outcome of the experiment. Randomization "evens out" the possible effects of potentially confounding variables.

Part (b):

Without the control group, the company could compare the two treatments, but would not be able to say whether the observed reduction in stress was attributable to participation in the programs. For example, a change in the work environment during this period might have reduced the stress level of all employees. The addition of a control group would enable the company to assess the magnitude of the mean reduction attributable to each treatment, as opposed to just determining if the two programs differ.

Part (c):

It is not reasonable to generalize the findings of this study to all employees, because

the participants in this experiment were volunteers and volunteers may not be representative of the population

OR

the participants were not randomly selected from the company employees.

Question 4 (cont'd)

Scoring

Each component is scored as either essentially correct (E), partially correct (P), or incorrect (I).

Part (a) has two components: the example, and the randomization.

• The example is scored as essentially correct (E) if it contains each of the elements in the table below:

Elements	Sample statements		
1. Identify a plausible example of a problem	"Because a deadline has been moved back"		
2. Relate the identified problem to the change in stress level (the response)	"lowering the stress levels of those working in the department. This change in stress level"		
3and state that the identified problem effects can not be distinguished from the difference in treatment effects	"could mistakenly be attributed to the treatment." (Note: A construction such as "can't tell the difference" is OK here.)		

The example is scored as partially correct (P) if the response contains 2 of the 3 components.

• The randomization is scored as essentially correct (E) if the student gives a reason for the necessity of random assignment. Possibilities include:

clearly stating in context that randomization is relied upon to create comparable groups

clearly stating in context that randomization controls for the effects of potentially confounding variables or reduces bias. (Both "Avoiding" bias and "Eliminating" bias are incorrect (I).)

The randomization is scored as partially correct (P) if the statement about randomization is not in context or is poorly communicated.

Note: Constructions such as "split up" and "divided into" can be interpreted to indicate randomization.

Question 4 (cont'd)

Part (b) is scored as essentially correct (E) if the student

- 1. indicates that a control group does provide additional information AND
- 2. explains that the control group allows the company to determine if either or both treatments are effective in reducing stress

OR

explains that the control group provides a baseline for comparison

Part (b) is scored as partially correct (P) if the student indicates there is additional information, even if the student's explanation is incorrect.

Note: Stating that the "passage of time" reduces stress is not sufficient; the student must specify that there is a confounding variable that <u>operates</u> through time.

Part (c) is scored as essentially correct (E) if it

- indicates that it is <u>not</u> reasonable to generalize to all employees AND
- 2. gives an explanation that the participants were not randomly selected from the company employees

OR

gives an explanation tied to the use of volunteers

Note: Simply using the word "volunteer" in the explanation is not sufficient.

Part (c) is scored as partially correct (P) if the student explicitly says that it is not reasonable to generalize to all employees, even if the student's explanation is incorrect.

Part (c) is scored as incorrect (I) if the student indicates that it is reasonable to generalize to all employees.

Question 4 (cont'd)

Scoring

4 Complete Response (4E)

All four "components" are essentially correct

3 Substantial Response (3E 0P or 3E 1P or 2E 2P)

Three components essentially correct and no components partially correct OR

Three components essentially correct and 1 component partially correct OR

Two components essentially correct and 2 components partially correct

2 Developing Response (2E 0P or 2E 1 P or 1E 2P or 1E 3P or 4P)

Two components essentially correct and no components partially correct OR

Two components essentially correct and 1 component partially correct OR

One component essentially correct and 2 components partially correct OR

One component essentially correct and 3 components partially correct OR

Four components partially correct

1 Minimal Response (1E 1P or 1E 0P or 0E 2P or 0E 3P)

One component essentially correct and 1 component partially correct OR

One component essentially correct and no components partially correct OR

No components essentially correct and 2 components partially correct OR

No components essentially correct and 3 components partially correct

Question 5

Solution

Part 1: States a correct pair of hypotheses

H_o: Response and gender are independent

H_a: Response and gender are not independent

OR

H_o: There is no association between response and gender

H_a: There is an association between response and gender

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

Chi-Square test (for independence)

$$\chi^2 = \sum \frac{\left(Obs - Exp\right)^2}{Exp}$$

Conditions: Random sample and large sample size

Expected counts are

	Strongly	Somewhat	Neither	Somewhat	Strongly
	Disagree	Disagree	Agree or	Agree	Agree
			Disagree		
Male	13.5	18.0	18.0	22.5	18.0
Female	16.5	22.0	22.0	27.5	22.0

All expected counts are greater than 5 (or 10), so the sample size is large enough for the Chi-Square test to be appropriate.

(Or, all expected counts are ≥ 1 , and no more than 20% of expected counts < 5.)

Part 3: Correct mechanics, including the value of the test statistic, df, and P-value (or rejection region)

$$\chi^2 = 0.907 + 0.500 + 0.500 + 0.278 + 2.722 + 0.742 + 0.409 + 0.409 + 0.227 + 2.227 = 8.921$$

$$df = 4$$
 P-value = 0.063

(Or, using tables, 0.05 < P-value < 0.10, or rejection regions: $\alpha = 0.05$ is 9.48, $\alpha = 0.01$ is 13.27)

Question 5 (cont'd)

Part 4: Stating a correct conclusion in the context of the problem, using the result of the statistical test.

Because P-value > selected α (or because χ^2 is not in the rejection region, or because the P-value is large), fail to reject the null hypothesis. There is not sufficient evidence to conclude that response is dependent on gender (or that response and gender are not independent, or that response and gender are associated)

OR

Because results this extreme would occur about 6 times in 100 by chance alone, there is marginal evidence to reject the null hypothesis and conclude that there is an association between response and gender.

Scoring

Note that the solution has 4 parts and each part is either correct or incorrect. No partial credit is given for individual parts.

- 4 Complete Response Four parts are correct.
- 3 Substantial Response Three parts are correct.
- **Developing Response**Two parts are correct.
- 1 **Minimal Response** Only one part is correct.

Question 6

Solution

Part (a):

If the probability of a strong market is less than 0.76, vans would provide a greater expected annual return because

Graphical Argument

the expected annual return for vans is higher, as can be seen in the graph — the line for the expected return for vans is above the line for the expected return for coaches for all values of p less than 0.76

OR

Algebraic Argument

the expected return for coaches and the expected return for vans are both linear functions of the probability of strong demand (p). These functions are equal at p=0.76. For any probability less than 0.76, vans have a higher expected return than coaches. For example, since the expected return for vans when p=0.5 (\$530,000) is greater than the expected return for coaches when p=0.5 (\$285,000), the expected return for vans must be higher when p<0.76.

Part (b):

Step 1: Identifies appropriate confidence interval by name or by formula.

95% confidence interval for p = proportion of similar markets that will experience strong demand.

OR

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

Step 2: States and checks appropriate conditions.

Conditions: random sample, sample size < 10% of the population size, and large sample size.

AP[®] STATISTICS 2003 SCORING GUIDELINES

Question 6 (cont'd)

Check for large sample size:

$$\hat{p} = 0.65$$

$$n\hat{p} = 65$$

$$\hat{p} = 0.65$$
 $n\hat{p} = 65$ $n(1 - \hat{p}) = 35$

Both $n\hat{p}$ and $n(1-\hat{p})$ are greater than 10 (or 5), so the sample size is large enough to proceed.

Step 3: Computes mechanics correctly.

$$\hat{p} \pm z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.65 \pm 1.96 \sqrt{\frac{(0.65)(0.35)}{100}} = (0.55652, 0.74348)$$

Step 4: Interprets the confidence interval in context.

We can be 95% confident that the true proportion of similar markets that experience strong demand is between 0.56 and 0.74.

Part (c):

Based on the interval in part (b), the plausible values for the probability of strong demand are between 0.56 and 0.74. As discussed in part (a), the expected annual return is greater for vans than for coaches for all values in this range. Therefore, vans should be used for the new route.

Part (d):

The demand in Lonestar's market will either be strong or weak. We think that the demand is more likely to be strong than weak because the entire interval in part (b) is above 0.5. As shown in the table, if demand is strong, coaches will produce an annual return of \$840,000, while vans will only produce an annual return of \$610,000. Since the demand is more likely to be strong than weak and the annual return from coaches is much higher than that of vans in a strong market, coaches would be the best choice for the new route.

Question 6 (cont'd)

Scoring

Each part is scored as either essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is essentially correct (E) if the response includes the following elements:

- 1. statement that vans produce a greater expected annual return when p < 0.76
- 2. plausible explanation for choice of vans
- Part (a) is partially correct (P) if only one of the two elements above is present.
- Part (b) is essentially correct (E) if steps 2, 3, and 4 of the confidence interval solution are correct.
- Part (b) is partially correct (P) if 2 of steps 2, 3 and 4 of the confidence interval solution are correct.
 - Note 1: Step 1 of the confidence interval procedure may be omitted since it is in the stem.
 - Note 2: For step 2, response does not need to include "random sample" nor "sample size less than 10% of the population" since the problem states that this is a random sample and that the sample is taken from "several thousand markets."
 - Note 3: In step 4, a correct interpretation of the confidence level cannot substitute for a correct interpretation of the confidence interval, but it should be considered as a plus if the overall score for this problem falls between two scores.
- **Part (c)** is essentially correct (E) if the discussion is based on vans having a higher expected return for all of the plausible values for the proportion in the confidence interval.

Part (c) is partially correct (P) if

1. the justification is limited to the point estimate of 0.65 and not the interval.

OR

2. the response observes that all plausible values for *p* are less than 0.76 but does not relate these values to expected returns.

Part (d) is essentially correct (E) if

1. the justification is based on observing that, since the probability of strong demand is greater than 0.5 for all values in the confidence interval, strong demand is more likely than weak demand.

AND

2. coaches are recommended because the annual return is higher for coaches than for vans when the demand is strong.

Question 6 (cont'd)

Part (d) is partially correct (P) if

1. element 1 is included but not element 2

OR

2. elements 1 and 2 are stated but not connected (poor communication).

OR

3. response is based only on the point estimate of 0.65.

OR

4. response gives a confidence interval with higher confidence than 0.95 and says that values beyond 0.76 are plausible.

OR

5. response notes that the value of p for coaches at the lower endpoint of the confidence interval will not have negative expected return, but has the potential for much greater profits.

Part (d) is incorrect if

1. the response gives an outside the interval argument; e.g., response states that there is a 2.5% chance of getting *p* above (outside) the interval.

OR

2. the response argues that the Lonestar market may be different.

OR

3. the response indicates a need to increase sample size.

ESSENTIALLY CORRECT RESPONSES COUNT AS 1 PART AND PARTIALLY CORRECT RESPONSES COUNT AS $\frac{1}{2}$ PART.

4 Complete Response

Four parts correct

3 Substantial Response

Three parts correct

2 Developing Response

Two parts correct

1 Minimal Response

One part correct

IF A PAPER IS BETWEEN TWO SCORES (FOR EXAMPLE, 2 ½ PARTS) USE A HOLISTIC APPROACH TO DETERMINE WHETHER TO SCORE UP OR DOWN DEPENDING ON THE STRENGTH OF THE RESPONSE AND COMMUNICATION.