



**AP[®] Statistics
2008 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 5,400 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[®], the PSAT/NMSQT[®], and the Advanced Placement Program[®] (AP[®]). 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.

© 2008 The College Board. All rights reserved. College Board, AP Central, Advanced Placement Program, AP, SAT, and the acorn logo are registered trademarks of 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.

Visit the College Board on the Web: www.collegeboard.com.
AP Central is the online home for AP teachers: apcentral.collegeboard.com.

**AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)**

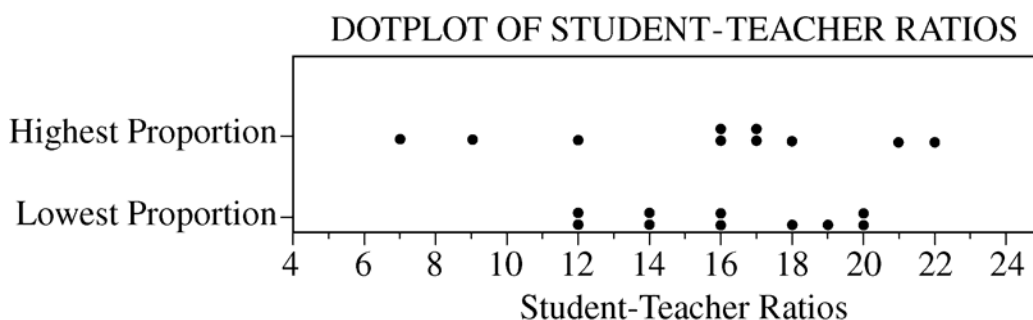
Question 1

Intent of Question

The primary goals of this question were to assess a student’s ability to (1) construct a dotplot that can be used to compare two distributions and to make appropriate comparisons of the center and spread of the two distributions and (2) understand and explain why in some situations inferential procedures that generalize from samples to larger populations are not appropriate.

Solution

Part (a):



A similarity is that the two distributions are centered in approximately the same place. A difference is that the distribution for the schools with the lowest proportions of students meeting the standards is less variable.

Part (b):

The two groups of schools are not random samples from two populations of interest. One group is the population of all 10 schools with the highest proportion of students meeting the standards, and the other group is the population of all 10 schools with the lowest proportion of students meeting the standards. Therefore, inferential procedures that generalize from samples to larger populations are not appropriate.

Scoring

Part (a) is divided into three sections. Section 1 is related to the dotplots, section 2 is related to the centers of the distributions, and section 3 is related to the spreads. Each is scored as either essentially correct (E) or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the dotplots are given with appropriate labels and scales (minor errors in the placement of the dots are okay).

Incorrect (I) if the labels or scales are missing *OR* if a histogram with dots or a scatterplot is given.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 1 (continued)

Section 2 is scored as follows:

Essentially correct (E) if the response states that the two distributions have approximately the same center (no numerical values are required) *OR* if the response states that the means are slightly different.

Incorrect (I) if measures of center are not given *OR* if they are given, but there is no *comparison* of the center of the two distributions.

Section 3 is scored as follows:

Essentially correct (E) if the response states that the spread of the two distributions differs (no numerical values are required).

Incorrect (I) if measures of spread are not given *OR* if they are given, but there is no *comparison* of the spread of the two distributions.

Note: Usually distributions are described in terms of shape, center, and spread. However, with small data sets, shape is difficult to judge, so no comparison of shape is required.

Part (b) is considered as section 4 and is scored as either essentially correct (E), partially correct (P), or incorrect (I).

Section 4 is scored as follows:

Essentially correct (E) if the response states that the data are not samples from some larger population *OR* that they are not random samples but instead are those with the highest and lowest proportions of students meeting a standard, and therefore inference is not appropriate. The response must not include any other reason (such as small sample sizes or the shape of the distribution).

Partially correct (P) if the response states only that the schools are not randomly selected. The response must not include any other reason (such as small sample sizes or the shape of the distribution).

Incorrect (I) if the response states that the sample sizes are too small *OR* if the response refers to the shape of the distribution or to some other criterion.

Note: If there is information relating to part (a) in the explanation in part (b), it may be used in the scoring of part (a).

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 1 (continued)

4 Complete Response

All four sections essentially correct

3 Substantial Response

Three sections essentially correct

2 Developing Response

Two sections essentially correct

1 Minimal Response

One section essentially correct

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 2

Intent of Question

The primary goals of this question were to assess a student's ability to (1) recognize an unbiased estimator and explain why the estimator is unbiased and (2) compare two estimators with respect to center and variability.

Solution

Part (a):

Statistics A, C, and D appear to be unbiased. This is indicated by the fact that the mean of the estimated sampling distribution for each of these statistics is about 75, the value of the true population parameter.

Part (b):

Statistic A would be a better choice because it appears to be unbiased. Although the variability of the two estimated sampling distributions is similar, statistic A would produce estimates that tend to be closer to the true population parameter value of 75 than would statistic B.

Part (c):

Statistic C would be a better choice because it has smaller variability. Although both statistic C and statistic D appear to be unbiased, statistic C would produce estimates that tend to be closer to the true population parameter value of 75 than would statistic D.

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the response contains the following two components.

Component 1: Identifies statistics A, C, and D as the unbiased estimators.

Component 2: Clearly demonstrates an understanding of the meaning of the term *unbiased*. That is, states that the mean (or center) of each distribution is about 75. No other characteristic (e.g., shape, spread) should be mentioned in the response unless it clearly is discounted as a criterion for being unbiased.

Partially correct (P) if the response contains just one of these components. That is, the response either identifies statistics A, C, and D as the unbiased estimators but gives a weak or no explanation or includes some discussion of another characteristic (e.g., shape, spread) as of some importance in judging bias *OR* demonstrates clear understanding of the meaning of the term *unbiased* but identifies only one or two of statistics A, C, and D as unbiased estimators.

Incorrect (I) otherwise.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 2 (continued)

Part (b) is scored as follows:

Essentially correct (E) if the response gives a clear explanation that statistic A is the better choice because it is centered at 75. (It is not necessary that the term *unbiased* be used as long as the response clearly addresses the center. It is not necessary to mention that statistics A and B both have about the same variability.)

Note: Students who were penalized in part (a) for mentioning other characteristics should not be repenalized in part (b).

Partially correct (P) if the response shows an understanding that centering is an important issue in determining which statistic is better but either does not apply the concept correctly *OR* communicates the explanation poorly.

Incorrect (I) otherwise.

Part (c) is scored as follows:

Essentially correct (E) if the response gives a clear explanation that statistic C is the better choice based on the fact that it has less variability. (It is not necessary to mention that statistics A and B are both centered at the parameter.)

Partially correct (P) if the response shows an understanding that variability is an important issue in determining which statistic is better but either does not apply the concept correctly *OR* communicates the explanation poorly.

Incorrect (I) otherwise.

4 Complete Response

All three parts essentially correct

3 Substantial Response

OR Two parts essentially correct and the other part partially correct

OR

Part (a) and one other part essentially correct and the remaining part incorrect

2 Developing Response

OR Part (a) essentially correct and one of the other parts partially correct

OR

All three parts partially correct

OR

Part (a) partially correct and one of parts (b) and (c) essentially correct

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 2 (continued)

1 Minimal Response

- OR* One part essentially correct and the other parts incorrect
- OR* Two parts partially correct and the other part incorrect
- Part (a) incorrect, one of the other parts essentially correct, and one part partially correct

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 3

Intent of Question

The primary goals of this question were to assess a student's ability to (1) calculate the proper sample size for a given margin of error and (2) relate that value to practical restrictions given in the problem.

Solution

Part (a):

The required sample size for 95% confidence and a margin of error (E) of 2 is

$$n = \frac{z^2 \sigma^2}{E^2} = \frac{(1.96)^2 (12)^2}{2^2} = 138.3$$

So, use $n = 139$.

Note: It is also acceptable to use $z = 2$ in the formula for sample size, in which case $n = 144$.

The cost of carrying out the study would be \$13,900 (or \$14,400), which exceeds the amount budgeted for the study. Sufficient funds are not available.

OR

For \$12,000, the company can afford to use a sample size of 120. The margin of error (with 95% confidence) expected for this sample size is

$$1.96 \frac{12}{\sqrt{120}} = 2.15 \quad \text{OR, using 2 in place of 1.96,} \quad 2 \frac{12}{\sqrt{120}} = 2.19$$

It is not possible to estimate the true mean stopping distance to within 2 feet with 95% confidence for \$12,000.

Part (b):

To be within 2 feet of the true mean with 95% confidence, 139 observations are required. The budget of \$12,000 only allows 120 observations to be taken. Therefore, the company will not be able to meet the regulatory agency's requirements with the allocated budget.

OR

Given the fact that \$12,000 will allow 120 observations to be taken, the margin of error will be 2.15. Therefore, the company will not be able to meet the regulatory agency's requirements with the allocated budget.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 3 (continued)

Scoring

There are two components to this question: component 1 consists of the calculations; component 2 consists of the conclusion/consequences. Components 1 and 2 are each scored as essentially correct (E), partially correct (P), or incorrect (I).

Component 1 is scored as follows:

Essentially correct (E) if:

- (1) A correct approach to calculating either the margin of error or the sample size is used.
AND
- (2) The required sample size (139) is correctly computed **OR** the margin of error that could be achieved for \$12,000 is correctly computed (2.15).

Partially correct if:

One or more minor errors are made in the formula and/or calculation. Examples of minor errors are the following:

- A square or square root sign is missing or is included erroneously
- A multiplier of 1.645 is used instead of 1.96
- The margin of error is rounded to 2, so the student thinks there are sufficient funds
- The sample size is rounded down instead of up, or left as 138.3
- The student thinks the \$12,000 will only allow 12 cars to be tested

Incorrect if:

One or more major errors are made. Examples of major errors are the following:

- Not including any multiplier in the margin of error formula
- Thinking the full width rather than half the width of the interval is the margin of error
- Stating that n must be greater than 30 (to have a “large sample”)

Component 2 is scored as follows:

Essentially correct (E) if:

- (1) A correct conclusion that the budget is insufficient (based on the sample size or margin of error calculation) is stated and clearly explained. (If the sample size or margin of error calculation is incorrect, the conclusion may be that the budget is sufficient.)
AND
- (2) The consequence that the car manufacturer will not be able to meet the requirement of the regulatory agency is stated. (If the sample size or margin of error calculation is incorrect, the conclusion may be that the manufacturer will be able to meet the requirement of the regulatory agency.)

Note: One would expect the conclusion (1) to be given in part (a) and the consequence (2) to be given in part (b); if both are given in part (a), then there must be an additional consequence in context specified in part (b) for component 2 to be essentially correct.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 3 (continued)

Partially correct (P) if only one of (1) and (2) is correctly stated.

Incorrect (I) if neither (1) nor (2) is correctly stated.

4 Complete Response

Both components essentially correct

3 Substantial Response

One component essentially correct and one component partially correct

2 Developing Response

One component essentially correct and the other component incorrect
OR

Both components partially correct

1 Minimal Response

One component partially correct and the other component incorrect

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 4

Intent of Question

The primary goals of this question were to assess a student's ability to (1) design an experiment to compare two treatments and (2) identify the associated potential Type I and Type II errors and decide which of these two would be more serious.

Solution

Part (a):

Approach 1: Paired Design

Each subject will receive both treatments, with a suitable length of time between treatments. The order of the treatments will be randomly assigned to the subjects. For example, for each patient flip a coin to determine which treatment will be administered first. Measure diastolic blood pressure, then have the subject sit quietly for 20 minutes in either a noise-free environment or in a room where soothing music is played, depending on which treatment was selected at random (based on the coin flip). At the end of the 20 minutes, measure diastolic blood pressure again and compute its change (*after – before*). After a suitable period of time, repeat with the other treatment.

When the data have been collected, the difference (*music – noise-free*) in the change in diastolic blood pressure will be computed for each subject, and then a paired *t*-test will be run to see if the mean difference is significantly different from zero.

Approach 2: Matched Pairs Design

Measure diastolic blood pressure for each of the 100 subjects and then form 50 pairs based on these readings by pairing the two with the highest diastolic blood pressure, then the two with the next highest, and so on. For each pair, toss a coin to determine which member of the pair will be assigned to group 1, and then assign the other member of the pair to group 2. For group 1, measure diastolic blood pressure, then have the subjects sit quietly in a noise-free environment for 20 minutes, and then measure diastolic blood pressure again and compute its change (*after – before*). For group 2, the plan is the same, except that they will sit for 20 minutes in a room where soothing music is played between blood pressure measurements.

When the data have been collected, the difference (*music – noise-free*) in the change in diastolic blood pressure will be computed for each pair, and then a paired *t*-test will be run to see if the mean difference is significantly different from zero.

Approach 3: Completely Randomized Design (This is not as good a choice as the two previous approaches, but because of the large number of subjects available for each treatment group, it is considered an acceptable solution.)

Assign the 100 patients numbers from 00 to 99. From a random number table, select 50 unique numbers; the patients with the selected values will form group 1; the remaining 50 patients will form group 2. For group 1, measure diastolic blood pressure, then have the subjects sit quietly in a noise-free environment for 20 minutes, and then measure diastolic blood pressure again. For group 2, the plan is the same, except that they will sit for 20 minutes in a room where soothing music is played between blood pressure measurements.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 4 (continued)

When the data have been collected, the change in diastolic blood pressure will be computed for each subject, and then a two-sample t -test will be run to see if there is a significant difference between the mean change attributable to *music* and the mean change attributable to a *noise-free* environment.

Part (b):

Type I error: Concluding that soothing music does reduce mean diastolic blood pressure compared to sitting quietly, when in fact it does not. The consequence of this type of error is that the clinic will offer music therapy when it is not effective.

Type II error: Soothing music does reduce diastolic blood pressure compared to sitting quietly, but we fail to detect this and conclude that it does not. The consequence of this type of error is that the clinic will choose not to offer music therapy when it would have been effective.

Which type of error is more serious? A case can be made for either type of error, and the student can take either side as long as a reasonable justification is given. For example, the student can say a Type I error is more serious because it will cost the clinic money with no benefit, or the student can say that a Type II error is more serious because the clinic will miss an opportunity to improve the health and well-being of its patients.

Scoring

Part (a) is divided into two sections: section 1 is the randomization, and section 2 is the experimental runs. Section 1 is scored as essentially correct (E), partially correct (P), or incorrect (I). Section 2 is scored as essentially correct (E) or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the design includes an appropriate randomization of patients to treatment groups or randomization of the order of treatments in the paired design. The description of the randomization should be sufficiently clear that it could be duplicated by the reader.

Partially correct (P) if the student states that the patients should be randomized to treatment groups or that there should be randomization of the order of treatment in the paired design but does not specify how this randomization is to be accomplished *OR* if there are flaws in the randomization *OR* if the randomization could not be duplicated by the reader.

Incorrect (I) otherwise.

Note: It is acceptable to first block by variables such as gender or age if the student then correctly uses one of the above approaches.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 4 (continued)

Section 2 is scored as follows:

Essentially correct (E) if treatments are applied and blood pressures are measured after treatment. (Although it would be better if the student suggests measuring blood pressure before and after treatment, it is sufficient that the student measures blood pressure only after treatment; however, measuring blood pressure both before and after treatment is considered a plus. Additionally, a statement that a comparison of the two groups will be made based on the change in blood pressure is not required but is also considered a plus.)

Incorrect (I) otherwise.

Part (b) is divided into two sections: section 1 consists of the identification of the errors, and section 2 consists of the consequences. Each section is scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the response gives correct descriptions of Type I and Type II errors in context.

Partially correct (P) if the response gives a correct description of one type of error in context and the description of the other type of error has a minor flaw.

Incorrect (I) otherwise.

Section 2 is scored as follows:

Essentially correct if:

- (1) The response describes the consequences for each of Type I and Type II errors.
- AND*
- (2) It states which type of error is more serious and gives a reason to support the selection made.

Partially correct if only one of (1) and (2) is correctly stated.

Incorrect (I) otherwise.

Notes:

- If Type I and Type II errors are reversed but the description of the errors and the consequences are correct for this reversal, give credit for part (b), section 2, but not for part (b), section 1.
- If only one type of error and its consequences are described, give credit for one section of part (b).

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 4 (continued)

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.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 5

Intent of Question

The primary goals of this question were to assess a student's ability to (1) describe the distribution of the difference of two normal random variables and (2) use this distribution to find a probability and to find a value given its location in the distribution.

Solution

Part (a):

X is normally distributed with $\mu = 170$ and $\sigma = 20$, and Y is normally distributed with $\mu = 200$ and $\sigma = 10$. The distribution of $Y - X$ has mean and standard deviation:

$$\begin{aligned}\mu_{Y-X} &= \mu_Y - \mu_X = 200 - 170 = 30 \\ \sigma_{Y-X} &= \sqrt{\sigma_Y^2 + \sigma_X^2} = \sqrt{10^2 + 20^2} = \sqrt{500} = 22.36\end{aligned}$$

The distribution of $Y - X$ is normal with mean 30 and standard deviation 22.36 (or, variance 500).

Part (b):

The train from Bullsnake will have to wait when $Y - X$ is positive:

$$P(Y - X > 0) = P\left(z > \frac{0 - 30}{22.36}\right) = P(z > -1.34) = 0.9099$$

(Calculator: 0.9082408019 or 0.9078172963, if z is not rounded.)

The proportion of days that the train will have to wait is about 0.91.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 5 (continued)

Part (c):

Let D denote the delay that will be needed for the train leaving Bullsnake. With the additional constant delay,

$X + D$ is normally distributed with $\mu_{X+D} = 170 + D$ and $\sigma_{X+D} = \sigma_X = 20$

Y is normally distributed with $\mu_Y = 200$ and $\sigma_Y = 10$

Thus, the difference $Y - (X + D)$ is normally distributed with

$$\mu_{Y-(X+D)} = \mu_Y - \mu_{(X+D)} = 200 - (D + 170) = 30 - D$$

$$\sigma_{Y-(X+D)} = \sigma_{Y-X} = 22.36$$

The combined delay and travel time ($X + D$) for the Bullsnake train must be *less than* the travel time for the Diamondback train (Y) with probability 0.01. That is, $P(Y - (X + D) > 0) = 0.01$, so we need

$$\frac{0 - \mu_{Y-(X+D)}}{\sigma_{Y-(X+D)}} = \frac{0 - (30 - D)}{22.36} = 2.33$$

Solving for D , the train from Bullsnake should be delayed by 82.099 minutes.

OR, with alternative notation:

Let X' denote the combined delay and travel time from Bullsnake to Copperhead, and let Y represent the travel time to Copperhead for the Diamondback train. The distribution of $Y - X'$ also is normal (because D is constant), with mean $\mu_{Y-X'} = \mu_Y - \mu_{X'}$ and standard deviation $\sigma_{Y-X'} = \sigma_{Y-X} = 22.36$.

The combined delay and travel time for the Bullsnake train (X') must be *less than* the time for the Diamondback train (Y) with probability 0.01. That is, $P(Y - X' > 0) = 0.01$, and we need

$$z = \frac{0 - \mu_{Y-X'}}{\sigma_{Y-X'}} = \frac{0 - \mu_{Y-X'}}{22.36} = 2.33$$

Solving, $\mu_{Y-X'} = \mu_Y - \mu_{X'} = -52.099$, so the mean travel time for the Diamondback train (Y) should be 52.099 minutes less than the mean combined travel and delay time for the Bullsnake train X' . The mean travel time for the Diamondback train is now 30 minutes more than the mean travel time for the Bullsnake train, so the train from Bullsnake should be delayed by $52.099 + 30 = 82.099$ minutes.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 5 (continued)

Scoring

Part (a) is divided into two sections: section 1 and section 2. Section 1 is scored as essentially correct (E) or incorrect (I). Section 2 is scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the response states that the distribution of $Y - X$ is normal.

Incorrect (I) otherwise.

Section 2 is scored as follows:

Essentially correct (E) if the response correctly computes the mean and standard deviation *AND* shows some work for the calculation of the standard deviation. May contain a minor arithmetic error.

Partially correct (P) if the response correctly states the values of the mean and standard deviation.

Incorrect (I) if the formula for the mean or standard deviation contains a conceptual error (such as not squaring the original standard deviations or subtracting the variances).

Part (b) constitutes section 3 and is scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 3 is scored as follows:

Essentially correct (E) if the response uses the distribution information from part (a) to correctly compute the desired probability. If the mean or standard deviation is computed incorrectly in part (a), those values should be used in part (b). (*Note:* If variances are incorrectly subtracted instead of added, $\sigma = 17.32$, $z = -1.73$, and the probability is $1 - 0.0418 = 0.9582$.)

Partially correct (P) if the response computes $P(Y - X < 0)$ instead of $P(Y - X > 0)$ and gets $P(Y - X < 0) = 1.0 - 0.9099 = 0.0901$.

Incorrect (I) otherwise.

Part (c) constitutes section 4 and it is scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 4 is scored as follows:

Essentially correct (E) if the response correctly concludes that the train from Bullsake should be delayed by about 82 minutes. If the mean or standard deviation is computed incorrectly in part (a), those values should be used in part (c). (*Note:* If variances are subtracted instead of added, the delay time will be $(2.33)(17.32) + 30 = 70.36$.)

Partially correct (P) if a correct line of reasoning is explored but the student fails to reach the correct answer.

Incorrect (I) otherwise.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 5 (continued)

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.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 6

Intent of Question

The primary goals of this investigative task were to assess a student's ability to (1) identify clusters in a scatterplot and deduce what such clusters represent; (2) recognize a positive linear relationship and state what it means, in context, when all points lie below the line $y = x$; (3) understand that when data are paired, a paired t -test, rather than a two-sample t -test is appropriate when testing whether the means are equal; and (4) create an appropriate graphical display that makes the dividing point between two groups of measurements visually obvious.

Solution

Part (a):

The trait that distinguishes the two groups in the scatterplot is dominant foot (left or right). All of the points in the upper left cluster represent measurements from individuals whose dominant foot is the right foot, whereas all of the points in the lower right cluster represent measurements from individuals whose dominant foot is the left foot.

Part (b):

There is a positive (linear) relationship between swelling in the dominant foot and swelling in the nondominant foot. Swelling in the dominant foot tends to be greater than swelling in the nondominant foot.

Part (c):

Component 1: States a correct pair of hypotheses in words or symbols.

$$\begin{array}{l}
 H_0 : \mu_d = 0 \quad \text{OR} \quad H_0 : \mu_D = \mu_N \quad \text{OR} \quad H_0 : \mu_D - \mu_N = 0 \\
 H_a : \mu_d \neq 0 \quad \quad \quad H_a : \mu_D \neq \mu_N \quad \quad \quad H_a : \mu_D - \mu_N \neq 0
 \end{array}$$

where

μ_d = mean difference in swelling (*dominant – nondominant*)

μ_D = mean swelling for dominant foot

μ_N = mean swelling for nondominant foot

**AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)**

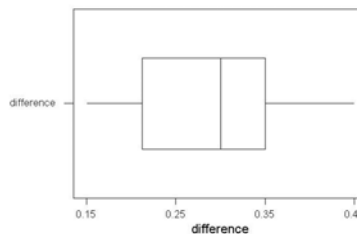
Question 6 (continued)

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

$$\text{Paired } t\text{-test } OR \quad t = \frac{\bar{x}_d - 0}{\frac{s_d}{\sqrt{n}}}$$

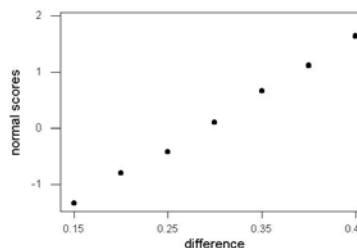
Conditions

1. We are told that this is a random sample from the population of adult female patients with Morton's neuroma (MN).
2. A boxplot of the differences (*dominant – nondominant*) has no outliers or extreme skewness. Thus, it is reasonable to believe that the population of differences is approximately normal.



OR,

A normal probability plot of the differences has no obvious departures from a linear trend, so it is reasonable to believe that the differences come from a normal distribution.



Component 3: Performs correct mechanics, including the value of the test statistic and *p*-value (or identifies a rejection region).

$$t = \frac{\bar{x}_d - 0}{\frac{s_d}{\sqrt{n}}} = \frac{0.2875 - 0}{\frac{0.0932}{\sqrt{12}}} = 10.68 \text{ with } df = 11, \text{ so the } p\text{-value is approximately } 0$$

OR,

From the calculator, $t = 10.68166837$, $df = 11$, $p\text{-value} = 0.00000038105102$

Component 4: Draws an appropriate conclusion in context, with linkage to the *p*-value (or rejection region).

Because the *p*-value is small (or is less than an α stated by the student), reject H_0 . There is convincing evidence that the mean swelling is different for dominant and nondominant feet for women with MN.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

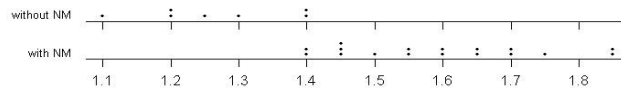
Question 6 (continued)

Part (d):

Three acceptable suggestions are given below.

Separate the 24 swelling measurements into two groups—the 17 feet with MN and the 7 feet without MN. Construct a plot that displays the two groups, such as stacked dotplots or a back-to-back stemplot. The plot below suggests that a swelling measurement of about 1.4 or higher would be a reasonable criterion for Morton’s neuroma.

Dotplot of Swelling



OR

Plot swelling measurements for only the seven individuals who do not have MN in both feet, plotting the measurements for their MN feet and their non-MN feet. The plot below suggests that a swelling measurement of about 1.4 or higher would be a reasonable criterion for MN.

Dotplot of Swelling



AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 6 (continued)

OR

Plot the 12 measurements of MN in the dominant foot, the 5 measurements of MN in the nondominant foot, and the 7 measurements of no MN in the nondominant foot. (There are no individuals in the sample who do not have MN in the dominant foot.) The plot below suggests that a swelling measurement of about 1.4 or higher would be a reasonable criterion for MN.

Dotplot for NM Dominant-No NM Nondom



Scoring

Each of the four parts is scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is scored as follows:

Essentially correct (E) if the response identifies the two groups as people with a left-dominant foot and people with a right-dominant foot. Students do not need to indicate which group is which.

Partially correct (P) if the response either identifies the two groups as people with greater swelling in the left foot and people with greater swelling in the right foot, missing the connection to foot dominance *OR* states only that “the dominant foot has more swelling than the nondominant foot” (making the connection back to the table) but without identifying the two groups.

Incorrect (I) if none of the three characterizations of the groups is present.

Part (b) is scored as follows:

Essentially correct (E) if the response contains both components below:

1. States that there is a positive (linear is optional) relationship *OR* states that as swelling in the dominant foot goes up, swelling in the nondominant foot tends to go up.
2. States that swelling in the dominant foot is greater than in the nondominant foot.

Note: One of these components may appear in part (a).

Partially correct (P) if only one of the two components is mentioned.

AP[®] STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 6 (continued)

Incorrect (I) if neither component is mentioned.

Part (c) is scored as follows:

Essentially correct (E) if all four components of the significance test are correct.

Partially correct (P) if only two or three components of the significance test are correct.

Incorrect (I) if at most one component of the significance test is correct.

Part (d) is scored as follows:

Essentially correct (E) if the response recommends a reasonable criterion for deciding whether a foot has MN based on the given data and justifies the criterion graphically. There are other possibilities for graphs than the three outlined above, but the plot should make it easy to compare the measurements in feet with and without MN. Communication must be strong.

Partially correct (P) if a reasonable recommendation based on the data is made but the graphical display is missing, made incorrectly, or is difficult to interpret *OR* if a graphical display comparing measurements for feet with MN to those without MN is provided but a recommendation based on the display is weak or missing.

Incorrect (I) if neither a reasonable conclusion nor a graph is present.

Essentially correct responses count as 1 point, and partially correct responses count as $\frac{1}{2}$ 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. Pay special attention to communication and to whether the correct test was performed in part (c), as well as to whether points were lost only because of omitted rather than incorrect elements, such as providing just one statement in part (b) and stating but not checking the condition of normality in part (c).