AP® Statistics
2005 Scoring Guidelines

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 4,700 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three and a half 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.

Visit the College Board on the Web: www.collegeboard.com.
AP Central is the official online home for the AP Program and Pre-AP: apcentral.collegeboard.com.

Question 1

Solution

Part (a):

The mean (40.45 cal/kg) and median (41 cal/kg) daily caloric intake of ninth-grade students in the rural school are higher than the corresponding measures of center, mean (32.6 cal/kg) and median (32 cal/kg), for ninth-graders in the urban school. There is also more variability or spread in the daily caloric intake for students in the rural school (Range=19, SD=6.04, IQR=10) than in the daily caloric intake for students in the urban school (Range=16, SD=4.67, IQR=7). The shapes of the two distributions are also different. The distribution of daily caloric intake for rural students is more uniformly distributed (symmetric) between 32 cal/kg and 51 cal/kg while the distribution of daily caloric intake for urban students appears to be skewed toward the larger values.

Part (b):

No, the samples include students from only one rural and one urban high school so it is not reasonable to generalize the findings from these two schools to all rural and urban ninth-grade students in the United States.

Part (c):

Since we are assuming that students keep accurate records, Plan II will do a better job of comparing the daily caloric intake of adolescents living in rural areas with the daily caloric intake of adolescents living in urban areas. Both plans take body weight into account by converting to food consumed per kilogram of body weight. Plan II includes a 7-day period (possibly days in school and days at home on the weekend), and there are differences in caloric intake among days. It would therefore be better to average over the 7-day period rather than considering only the food consumed in one day, as is the case with Plan I. Plan II would provide a more precise estimate of the average daily intake.

Scoring

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

Part (a) is essentially correct (E) if the student correctly compares center, shape, and spread of the two distributions. Specific numerical values are not required.

Part (a) is partially correct (P) if the student correctly compares any two of the three characteristics (center, shape, or spread) of the two distributions.

Part (a) is incorrect (I) if the student correctly compares no more than one characteristic.

Question 1 (continued)

Part (b) is essentially correct (E) if the student realizes that these findings cannot be generalized because the students were selected from only one rural and one urban high school.

Part (b) is incorrect (I) if the student argues that these findings:

- cannot be generalized with no explanation given; OR
- cannot be generalized with an invalid explanation (e.g., the response indicates that the student sample size (n=20) is not big enough); OR
- can be generalized because the randomly selected students from these two schools may represent all urban and rural ninth-grade students in the United States.

Part (c) is essentially correct (E) if Plan II is chosen and a correct justification involving day-to-day variability is provided.

Part (c) is partially correct (P) if Plan II is chosen, but a weak statistical justification that includes a discussion of day-to-day variability is provided.

Part (c) is incorrect (I) if:

- Plan I is chosen; OR
- Plan II is chosen and a correct justification is not provided.

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)

Two parts essentially correct and zero parts partially correct
OR
One part essentially correct and two parts partially correct

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

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

Question 2

<u>Solution</u>

Part (a):

The expected number of telephone lines in use by the technical support center at noon is:

$$E(X) = 0 \times 0.35 + 1 \times 0.2 + 2 \times 0.15 + 3 \times 0.15 + 4 \times 0.1 + 5 \times 0.05$$
$$= 1.6$$

Part (b):

We would expect the average based on 1,000 days to be closer to 1.6 than the first average based on 20 days. Both averages have the same expected value (1.6), but the variability for sample averages based on 1,000 days is smaller than the variability for sample averages based on 20 days.

Part (c):

The median of X is 1.

x	$P(X \le x)$	$P(X \ge x)$
0	0.35	1.0
1	0.55	0.65
2	0.70	0.45
3	0.85	0.30
4	0.95	0.15
5	1.0	0.05

OR

The median of X is 1 because $P(X \le 1) = 0.55 \ge 0.50$ and $P(X \ge 1) = 0.65 \ge 0.50$.

Part (d):

The probability histogram is clearly skewed to the right (or toward the larger values) so the mean (1.6) is larger than the median (1), as is typical for a right-skewed distribution.

Question 2 (continued)

Scoring

Parts (a) and (c) are combined as one computational part. Each part is scored as essentially correct (E), partially correct (P), or incorrect (I).

Collectively parts (a) and (c) are essentially correct (E) if both parts are calculated correctly, with the exception of minor arithmetic errors.

Collectively parts (a) and (c) are partially correct (P) if one of the two parts is calculated correctly, with the exception of minor arithmetic errors.

Collectively parts (a) and (c) are incorrect (I) if both parts are calculated incorrectly.

Note: Unsupported answers in parts (a) and (c) are scored as incorrect.

Part (b) is essentially correct (E) if the student:

1. States the new estimate based on 1,000 days should be closer to the expected value of 1.6; OR the new estimate will increase, or decrease if the answer in part (a) is less than 1.25.

AND

2. Provides justification by stating the variability for sample averages based on 1,000 days will be smaller than the variability for sample averages based on 20 days; OR as the sample size increases the sample average approaches the expected value of *X*.

Part (b) is partially correct (P) if the student provides one of the two items above.

Part (d) is essentially correct (E) if the student states that since the distribution is skewed to the right, the mean is greater than the median; OR since the mean is greater than the median, the distribution is skewed to the right.

Note: There must be evidence that the student looked at the given distribution.

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

- States that since the mean is greater than the median, the distribution is skewed to the right (with no evidence that the student looked at the given distribution); OR
- Compares the two measures of center by referring to the inappropriate or incomplete shape of the distribution (e.g., “skewed to the left” or “skewed”); OR
- Makes a correct statement about the measures of center and the shape without connecting the two.

Question 2 (continued)

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

- Compares the two measures of center without mentioning the shape of the distribution; OR
- Correctly describes the shape without correct conclusions about the relative location of the mean and median; OR
- Makes multiple "generic" statements about the relationship of mean, median, and shape with no reference to the given distribution.

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)

Two parts essentially correct and zero parts partially correct
OR
One part essentially correct and two parts partially correct

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

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

Question 3

Solution

Part (a):

Yes, the linear model is appropriate for these data. The scatterplot shows a strong, positive, linear association between the number of railcars and fuel consumption, and the residual plot shows a reasonably random scatter of points above and below zero.

Part (b):

According to the regression output, fuel consumption will increase by 2.15 units for each additional railcar. Since the fuel consumption cost is \$25 per unit, the average cost of fuel per mile will increase by approximately $\$25 \times 2.15 = \$53.75$ for each railcar that is added to the train.

Part (c):

The regression output indicates that $r^2 = 96.7\%$ or 0.967. Thus, 96.7% of the variation in the fuel consumption values is explained by using the linear regression model with number of railcars as the explanatory variable.

Part (d):

No, the data set does not contain any information about fuel consumption for any trains with more than 50 cars. Using the regression model to predict the fuel consumption for a train with 65 railcars, known as extrapolation, is not reasonable.

Scoring

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

Part (a) is essentially correct (E) if the model is deemed appropriate AND the explanation clearly indicates:
- There is a linear pattern in the scatterplot; OR
- There is no pattern in the residual plot.

Part (a) is partially correct (P) if the:
- Model is deemed appropriate AND the student refers to the scatterplot or residual plot but fails to state the relevant characteristic of the plot; OR
- Student refers to the relevant characteristic of the scatterplot or residual plot without deeming model appropriate.

Part (a) is incorrect (I) if the student:
- States that the model is appropriate without an explanation; OR
- States that the model is inappropriate; OR
- Makes a decision based only on numeric values from the computer output.

Question 3 (continued)

Part (b) is essentially correct (E) if the point estimate for the slope (2.15 or 2.1495) and the fuel consumption cost per unit ($25) are used to calculate the correct point estimate ($53.75 or $53.7375 ≈ $53.74).

Part (b) is partially correct (P) if only the point estimate for the slope (2.15 or 2.1495) is stated with a supporting calculation or interpretation.

Part (c) is essentially correct (E) if the student states:

- 96.7% of the variation in fuel consumption is explained by the linear regression model; OR
- 96.7% of the variation in fuel consumption is explained by the number of railcars.

Part (c) is partially correct (P) if the student makes one of the above statements using R-Sq(adj) = 96.3%.

Part (d) is essentially correct (E) if the student states that this is unreasonable due to extrapolation.

Part (d) is partially correct (P) if the student states this is:

- Unreasonable but provides a weak explanation; OR
- Reasonable even though it is considered a slight extrapolation.

Note: Any answer appearing without supporting work is scored as incorrect (I).

Each essentially correct (E) response counts as 1 point, each partially correct (P) response counts as ½ point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

Note: If a response is in between two scores (for example, 2 ½ points), use a holistic approach to determine whether to score up or down depending on the strength of the response and communication.

Question 4

Solution

This question is divided into four parts.

Part (a): State a correct pair of hypotheses.

Let p = the proportion of boxes of this brand of breakfast cereal that include a voucher for a free video rental.

$$H_0 : p = 0.2$$
$$H_a : p < 0.2$$

Part (b): Identify a correct test (by name or by formula) and check appropriate conditions.

One-sample z-test for a proportion *OR* $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$

Conditions:

1. $np_0 = 65 \times 0.2 = 13 > 10$ and $n(1 - p_0) = 65 \times 0.8 = 52 > 10$.
2. It is reasonable to assume that the company produces more than $65 \times 10 = 650$ boxes of this cereal ($N > 10n$).
3. The observations are independent because it is reasonable to assume that the 65 boxes are a random sample of all boxes of this cereal.

Part (c): Use correct mechanics and calculations, and provide the p-value (or rejection region).

The sample proportion is $\hat{p} = \dfrac{11}{65} = 0.169$. The test statistic is $z = \dfrac{0.169 - 0.2}{\sqrt{\dfrac{0.2(1 - 0.2)}{65}}} = -0.62$ and the p-value is

$P(Z < -0.62) = 0.2676$.

Part (d): State a correct conclusion, using the result of the statistical test, in the context of the problem.

Since the p-value $= 0.2676$ is larger than any reasonable significance level (e.g., $\alpha = 0.05$), we cannot reject the company's claim. That is, we do not have statistically significant evidence to support the student's belief that the proportion of cereal boxes with vouchers is less than 20 percent.

Question 4 (continued)

Scoring

The question is divided into four parts. Each part is scored as essentially correct (E) or incorrect (I).

Part (a) is essentially correct (E) if the student states a correct pair of hypotheses.

Notes:

1. Since the proportion was defined in the stem, standard notation for the proportion (p *or* π) need not be defined in the hypotheses.
2. Nonstandard notation must be defined correctly.
3. A two-sided alternative is incorrect for this part.

Part (b) is essentially correct (E) if the student identifies a correct test (by name or by formula) and checks for appropriate conditions.

Notes:

1. $np_0 > 5$ and $n(1 - p_0) > 5$ are OK as long as appropriate values are used for n and p_0.
2. Since students cannot check the actual population size, they do not need to mention it.
3. The stem of the problem indicates this is a random sample so it (or a discussion of independence) does not need to be repeated in the solution.

Part (c) is essentially correct (E) if no more than one of the following errors is present in the student's work:

- Undefined, nonstandard notation is used; OR
- The correct z-value $= -0.62$ is given with no setup for the calculation; OR
- The incorrect z-value $= -0.67$ is given because $\hat{p}$ was used in the calculation of the standard error. For this incorrect z-value, the p-value $= 0.2514.$; OR
- The incorrect z-value is calculated because of a minor arithmetic error.

Part (c) is incorrect (I) if:

- Inference for a lower tail alternative is based on either two-tails p-value $= 0.535$ or the upper tail p-value $= 0.734$; OR
- An unsupported z-value other than -0.62 or -0.67 is given; OR
- The correct z-value $= -0.62$ is given but equated to an incorrect formula.

Question 4 (continued)

Notes:

1. Students using a rejection region approach should have critical values appropriate for a lower tail test, e.g., for $\alpha = 0.05$ the rejection region is $z < -1.645$.
2. Other possible correct mechanics include:

   - Exact Binomial
     X ~ Binomial(n=65, p=0.2). The exact *p*-value is $P(X \leq 11) = 0.33$.

   - Normal Approximation to Binomial (with or without a continuity correction)
     X is approximately Normal(13, 3.225). The approximate *p*-value using the continuity correction is $P\left(Z \leq \frac{11 + 0.5 - 13}{3.225}\right) = P(Z \leq -0.4651) = 0.3209$.

   - Confidence interval approach – provided there is a reasonable interpretation tied to a significance level. For example if $\alpha = 0.05$, and $p = 0.20$ is within a 95% upper confidence bound (0, 0.2457) or a two-tailed 90% confidence interval (0.0927, 0.2457).

Part (d) is essentially correct (E) if the student states a correct conclusion in the context of the problem, using the result of the statistical test.

Notes:

1. If both an α and a *p*-value (or critical value) are given, the linkage is implied.
2. If no α 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.
3. If the *p*-value in part (c) is incorrect but the conclusion is consistent with the computed *p*-value, part (d) can be considered as essentially correct (E).
4. If a student accepts the null hypothesis and concludes the proportion really is 0.20, this part is incorrect (I).

Each essentially correct (E) response counts as 1 point, each partially correct (P) response counts as ½ point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

Note: If a response is in between two scores (for example, 2 ½ points), use a holistic approach to determine whether to score up or down depending on the strength of the response and communication.

Question 5

Solution

Part (a):

Since random-digit dialing will be used, individuals without phones will not be included in the sample. People without a high school diploma are more likely to have lower-paying jobs and therefore may not be able to afford a telephone. Thus, the estimated proportion of adult heads of households in the United States without a high school diploma may be less than the true population proportion.

Part (b):

The sample size necessary to estimate the proportion of the population that does not have a high school diploma, p, within 0.03 with 95% confidence is:

$$0.03 = z^*\left(\sqrt{\frac{p^*\left(1-p^*\right)}{n}}\right), \quad or$$

$$0.03 = 1.96\left(\sqrt{\frac{0.22(1-0.22)}{n}}\right), \quad so$$

$$n = 0.22(0.78)\left[\frac{1.96}{0.03}\right]^2 = 732.4651\ .$$

Thus, 733 respondents would be needed.

Part (c):

To achieve this additional goal, the agency should use stratified random sampling by taking samples within each state. Each state would be a stratum. Within each state, a random sample of adult heads of households would be selected and surveyed. The sample size within each state will be based on the desired precision. Data from the individual states should be combined to obtain the national estimate.

Question 5 (continued)

Scoring

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

Part (a) is essentially correct (E) if the student:

- Provides a possible source of sampling or nonsampling bias is provided and linked to whether or not a person has a high school diploma; AND
- Describes the impact on the survey correctly.

Note:

Other potential sources of bias:

1. Wording of the questions or the tone of the interviewers may make callers less likely to reveal that they do not have a high school diploma (nonsampling bias), forcing the estimate to be too low.
2. Calls may be made during the day when individuals with diplomas will be at work, leading to an overreporting of individuals without a high school diploma (sampling bias). Thus, the estimate will be too high.
3. Higher-educated people may be more likely to have Caller ID and not answer the phone when they see the call is from a survey firm. Thus, the estimate will be too high.
4. Since people do not like responding to cold calling (consider the size of the national Do Not Call list), the response rates might be so low that the survey results are not useful at all. The impact of this bias on the estimate is not clear.

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

- Provides a possible source of sampling or nonsampling bias linked to whether or not a person has a high school diploma, with an unreasonable description of the impact of the survey; OR
- Provides a possible source of bias but the bias is improperly named; however, the impact on the survey is consistent with the description.

Part (a) is incorrect (I) if the source of bias is not reasonable for this survey.

Part (b) is essentially correct (E) if:

- The appropriate critical value, margin of error, and a standard deviation based on a value of p (0.22 or 5) are used to calculate the number of necessary respondents; AND
- Work is shown; AND
- The numeric response is rounded up.

Note:

Other possible essentially correct (E) solutions for part (b):

- The sample size necessary to estimate the proportion of the population that does not have a high school diploma, p, within 0.03 with 95% confidence is:

$$n = p^*\left(1 - p^*\right)\left[\frac{z^*}{m}\right]^2 = 0.22(0.78)\left[\frac{1.96}{0.03}\right]^2 = 732.4651.$$ So, we would need 733 respondents.

Question 5 (continued)

- Conservative approximation $(p^* = 0.5)$:

$$n = \left[\frac{z^*}{2m}\right]^2 = \left[\frac{1.96}{2 \times 0.03}\right]^2 = 1067.1111$$, so we need 1,068 respondents.

- Wilson estimate:

$$n + 4 = p^*\left(1 - p^*\right)\left[\frac{z^*}{m}\right]^2 = 0.22(0.78)\left[\frac{1.96}{0.03}\right]^2 = 732.4651$$, so we need 729 respondents.

- Wilson estimate with conservative approach $(p^* = 0.5)$:

$$n + 4 = \left[\frac{z^*}{2m}\right]^2 = \left[\frac{1.96}{2 \times 0.03}\right]^2 = 1067.1111$$, so we need 1,064 respondents.

Part (b) is partially correct (P) if work is shown and no more than one of the following occur:

- An incorrect critical value is used in the calculation; OR
- An incorrect margin of error is used in the calculation; OR
- An incorrect standard deviation is used in the calculation; OR
- Numeric response is rounded down or is not an integer.

Part (b) is incorrect (I) if a solution is provided with no justification or an incorrect formula is used to justify the calculation.

Part (c) is essentially correct (E) if stratified random sampling is used to select a random sample from each state. The student must indicate that:

- The states are the strata. (The student must use the phrase “strata” or “stratified”); AND
- A random sample is taken in each state.

Part (c) is partially correct (P) if only one of the two items necessary for an essentially correct score is provided.

Part (c) is incorrect (I) if the student suggests that random digit dialing is continued until large enough samples are obtained for all 50 states.

Question 5 (continued)

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)

Two parts essentially correct and zero parts partially correct
OR
One part essentially correct and two parts partially correct

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

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

Question 6

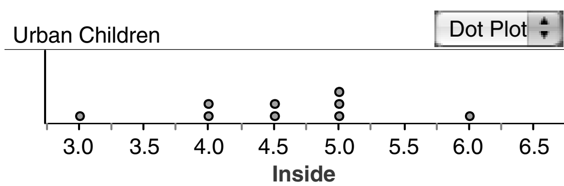
Solution

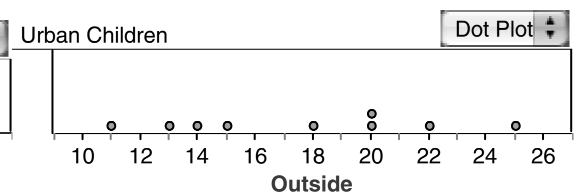
Part (a):

Step 1: State and check appropriate conditions for two-sample *t*-confidence interval.

State conditions: The two samples are selected randomly and independently from the two populations. The population distributions of the amount of lead on the dominant hand for both groups of children (those who could be sent to play inside and those who could be sent to play outside) are normal.

Check conditions: The procedure described in the stem is equivalent to taking a random sample from the population of children in urban day-care centers who could be assigned to play inside and an independent random sample from the population of children in urban day-care centers who could be assigned to play outside. The symmetry and lack of outliers in the dotplots below indicate that the normal assumption is reasonable for both populations of children.





Step 2: Identify the two-sample *t*-interval (by name or formula) and provide correct mechanics

Identification: Two-sample *t*-confidence interval for the difference of two means

OR

$$\bar{x}_{in} - \bar{x}_{out} \pm t^* \sqrt{\frac{s_{in}^2}{n_{in}} + \frac{s_{out}^2}{n_{out}}}$$

Mechanics: Using the summary statistics

	N	Mean	StDev
Inside	9	4.56	0.846
Outside	9	17.56	4.61

and the conservative $df = \min(n_1 - 1, n_2 - 1) = 8$, the 95% confidence interval for the difference of the two means is:

$$(4.56 - 17.56) \pm 2.306\sqrt{\frac{(0.846)^2}{9} + \frac{(4.61)^2}{9}} \quad \text{or} \quad (-16.60, -9.40) \text{ mcgs.}$$

Question 6 (continued)

OR

Using fractional $df = 8.537$, the 95% C.I. for $(\mu_{in} - \mu_{out})$ is (–16.57, –9.43) mcgs.

Step 3: Interpret the confidence interval in context.

We are 95 percent confident that the difference between the mean amount of lead on the dominant hands of the population of urban day-care children after an hour of play inside and the mean amount of lead on the dominant hands of the population of urban day-care children after an hour of play outside at an urban day-care center is between –16.60 and –9.40 mcgs.

OR

Because this interval does not include zero, we can conclude that there is a significant difference in the mean amount of lead on the hands of the two different groups of urban day-care children after one hour of play. On average, urban day-care children who play outside have higher amounts of lead on their hands.

Part (b)

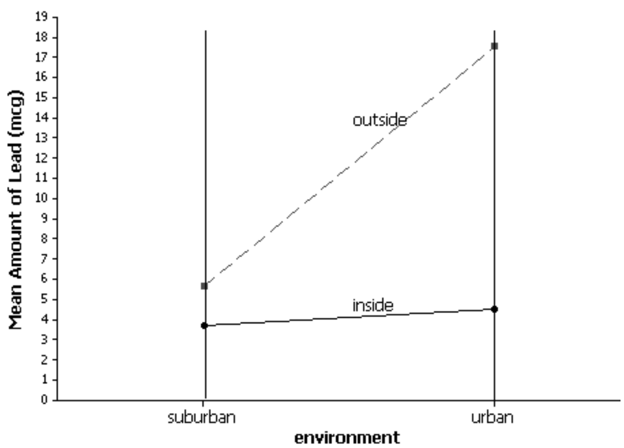


Table of Means

	Suburban	Urban
Inside	3.75	4.56
Outside	5.65	17.56

Question 6 (continued)

Part (c)

Inside/Outside: For both environments (urban and suburban), the mean amount of lead on the dominant hands of children who play outside is higher than the mean amount of lead on the dominant hands of children who play inside. This can be justified by comparing means OR confidence intervals OR interpreting the graph. All four endpoints of the two confidence intervals (inside minus outside) are negative. The graph clearly shows that the line connecting the two "outside" means for the different environments is above the line that connects the "inside" means.

Suburban/Urban: For both settings (inside and outside), the amount of lead on the dominant hand of urban children is higher on average than the amount of lead on the dominant hand of suburban children. This can be justified by comparing means or interpreting the graph: both lines slant upward to the right, which indicates an increase from suburban to urban both for children who played inside and for children who played outside.

Relationship: The magnitude of the difference in the mean amount of lead between children playing inside and children playing outside depends on the environment. Equivalently, the graph shows that the means for the urban environment are much farther apart than the means for the suburban environment.

OR

Whether the children play inside or outside makes a bigger difference in the urban environment than in the suburban environment. This is shown by the graph or the fact that the endpoints for the urban confidence interval are farther away from zero than the corresponding endpoints of the suburban confidence interval (and the intervals do not overlap), indicating that the difference is larger for the urban environment.

<u>Scoring</u>

Parts (a) and (b) are scored as essentially correct (E), partially correct (P), or incorrect (I). Part (c) is divided into two subparts. Each of these subparts is scored as essentially correct (E), partially correct (P), or incorrect (I).

Each of the three steps in part (a) is scored as acceptable or unacceptable.

Part (a) is essentially correct (E) if all three steps are acceptable. A step may be scored as acceptable even if it contains a minor error.

Part (a) is partially correct (P) if two steps are acceptable.

Part (a) is incorrect (I) if at most one step is acceptable.

Notes:

Step 1: Conditions

- A pair of dotplots, stemplots, histograms, normal probability plots, or boxplots may be provided to check the normality assumption.
- If the response uses an unacceptable procedure in Step 2, credit may be given in Step 1 if the check of conditions is consistent with the specified procedure.

Question 6 (continued)

Step 2: Identification of interval and computations

If one of the following procedures is used, Step 2 is scored as unacceptable:

- Paired *t*-procedure; OR
- Separate confidence intervals for inside and outside; OR
- Large-sample *z*-procedure: n_{in} and n_{out} are not large enough to assume normality; OR
- Pooled *t*-confidence interval. The assumption that $\sigma_{in} = \sigma_{out}$ is not reasonable; OR
- Concluding that the sample sizes are too small for inference.

Step 3: Conclusion

- Just interpreting the 95% confidence level is scored as unacceptable.
- If the response uses an unacceptable procedure in Step 2, credit may be given in Step 3 if the conclusion is acceptable for the specified procedure.

Part (b) is scored as essentially correct (E) if:

- The four points and two lines are placed correctly with a correct scale (either a numerical scale on a vertical line or the four means written next to the four points); AND
- The vertical axis is labeled ("lead" or "mcg" is sufficient) OR the two lines are labeled "inside" and "outside."

Part (b) is partially correct (P) if the four points and two lines are placed correctly but either:

- The vertical axis AND the two lines are not labeled but the numerical scale is correct; OR
- At least one label is included AND the numerical scale is incorrect.

Part (b) is incorrect (I) if the four points and two lines are not placed correctly OR there is no numerical scale.

Part (c) contains two parts. Each part is scored as essentially correct (E), partially correct (P), or incorrect (I). The first part (c1) deals with direct comparisons between environments (suburban versus urban) and settings (inside versus outside), where the comparisons must be supported using the data. The second part (c2) deals with the nature of the relationship between the means as setting and environment change.

Part (c1) is scored as essentially correct (E) if the student:

- Makes it clear that outside is greater than inside in both suburban and urban environments; AND
- Makes it clear that urban is greater than suburban in both inside and outside settings; AND
- Justifies one or both comparisons by referencing the graph, the means given in part (b), or the two confidence intervals.

Part (c1) is partially correct (P) if two of the above are included.

Part (c1) is incorrect (I) if at most one of the above is included.

Question 6 (continued)

Part (c2) is scored as essentially correct (E) if the fact that the effect of the setting depends on the environment (or the effect of the environment depends on the setting) is correctly described:

- The suburban to urban difference in means is small for children who played inside, but the urban mean is much bigger than the suburban mean for children who played outside; OR
- The outside mean is similar to the inside mean in the suburban environment, but the outside mean is much bigger than the inside mean in the urban environment.

Note: The first response also correctly compares suburban and urban. The second response also correctly compares inside and outside.

Part (c2) is partially correct (P) if the response is out of context OR communication is poor.

Part (c2) is incorrect (I) if the relationship is not described.

Note: The statement would fit if the lines had been parallel does not describe the relationship here.

Each essentially correct (E) response counts as 1 point, each partially correct (P) response counts as ½ point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

Note: If a response is in between two scores (for example, 2 ½ points), use a holistic approach to determine whether to score up or down depending on the strength of the response and communication.