CollegeBoard AP

**AP® Statistics
2011 Scoring Guidelines
Form B**

The College Board

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 College Board is composed of more than 5,700 schools, colleges, universities and other educational organizations. Each year, the College Board serves seven million students and their parents, 23,000 high schools, and 3,800 colleges through major programs and services in college readiness, college admission, guidance, assessment, financial aid and enrollment. Among its widely recognized programs are the SAT®, the PSAT/NMSQT®, the Advanced Placement Program® (AP®), SpringBoard® and ACCUPLACER®. 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.org.
AP Central is the official online home for the AP Program: apcentral.collegeboard.com.

AP® STATISTICS
2011 SCORING GUIDELINES (Form B)

Question 1

Intent of Question

The primary goals of this question were to assess students' ability to (1) describe and use a procedure for estimating medians from histograms; (2) use graphical displays to compare two different distributions; (3) use graphical and numerical information to compare the means for two groups.

Solution

Part (a):

The median is the value with half of the P-T ratios at or below it and half of the values at or above it. For n observations in a group, use $\dfrac{n+1}{2}$ to find the position of the median in the ordered list of observations.

For states west of the Mississippi ($n = 24$) the median falls between the 12th and 13th value in the ordered list, and both the 12th and 13th values fall in the interval 15–16. For states east of the Mississippi ($n = 26$) the median falls between the 13th and 14th value in the ordered list, and both of these values also fall in the interval 15–16.

From the histogram, cumulative frequencies for the two groups are shown in the table below.

Interval	West	East
12–13	1	2
13–14	$1+4=5$	$2+4=6$
14–15	$1+4+6=11$	$2+4+4=10$
15–16	$1+4+6+3=14$	$2+4+4+11=21$

Thus, the median P-T ratio for both groups is at least 15 students per teacher and at most 16 students per teacher.

Part (b):

The shapes of the two histograms are different. The histogram for states that are west of the Mississippi River is unimodal and skewed to the right, whereas the histogram for states that are east of the Mississippi River is unimodal and nearly symmetric.

As noted in part (a), the medians of the two distributions are about the same, between 15 and 16 for both distributions.

The histograms also show that there is more variability in the P-T ratios for states that are west of the Mississippi River. Although the greatest and least values for each group are not known, the range can be approximated. The range for the west is at most $22 - 12 = 10$, and the range for the east is at most $19 - 12 = 7$.

Part (c):

The medians of the two distributions are about the same, as determined in part (a). The distribution of P-T ratios for states that are west of the Mississippi River is skewed to the right, indicating that the mean will probably be higher than the median. The rough symmetry for the east group indicates that the mean will be close to the median. Thus, the mean for the west group will probably be greater than the mean for the east group.

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if a correct estimation method is described and appropriate estimates (values between 15 and 16, inclusive) are provided.

Partially correct (P) for any of the following:
- The response describes a correct estimation method, but the estimates are not provided.
- The response describes a method that conveys the idea of median as the middle value but is not entirely correct (for example, it describes the 12th value rather than the average of the 12th and 13th values), *AND* it provides reasonable estimates.
- The response gives an incomplete description of the method *AND* provides reasonable estimates.
- The response shows work only on the histograms *AND* correct estimates are provided, *BUT* no verbal explanation of the method is given.

Incorrect (I) if the response fails to meet the criteria for E or P.

Part (b) is scored as follows:

Essentially correct (E) if appropriate comparative statements are made for the centers, the shapes, and the spreads of the two groups.

Note: The shape of the east histogram can be described as skewed, *approximately* symmetric, or *approximately* normal. However, if the shape is described as symmetric or normal, part (b) cannot be scored an E.

Partially correct (P) if all three comparative statements are not made, but correct information regarding all three characteristics (center, shape, and spread) is provided for both groups, *OR* if only two of the three comparative statements are made.

Note: If a comparative statement about medians is made in part (a) or in part (c), this can count for a comparison of center in part (b).

Incorrect (I) if at most one comparative statement is made *AND* the response does not include correct information about all three characteristics (center, shape, and spread) for both groups.

Part (c) is scored as follows:

Essentially correct (E) if the response indicates that the west group has a higher mean than the east group and provides a justification based on the relationship between means and medians for distributions with different shapes.

Partially correct (P) for any of the following:
- The response indicates that the west group has a higher mean, *BUT* the justification is not based on the relationship between means and medians for distributions with different shapes — for example, the justification is not based on the answers to parts (a) and (b).
- The response includes correct statements about the relative sizes of the mean and median for each group, *BUT* it does not explicitly compare the means of the two groups.
- The response provides a justification based on the relationship between the mean and median for a skewed distribution *BUT* concludes that the mean will be smaller than the median for a right-skewed distribution.

Incorrect (I) if either group is selected with no justification.

4   Complete Response

>   All three parts essentially correct

3   Substantial Response

>   Two parts essentially correct and one part partially correct

2   Developing Response

>   Two parts essentially correct and one part incorrect
>
> *OR*
>
>   One part essentially correct and one or two parts partially correct
>
> *OR*
>
>   Three parts partially correct

1   Minimal Response

>   One part essentially correct and two parts incorrect
>
> *OR*
>
>   Two parts partially correct and one part incorrect

Question 2

Intent of Question

The primary goals of this question were to assess students' ability to (1) distinguish an experiment from an observational study; (2) critique statistical information, in particular whether or not researchers are justified in making a specific conclusion based on the given information; (3) recognize and describe a potential problem with a study that lacks random assignment or blinding.

Solution

Part (a):

The study was an experiment because treatments (D-cycloserine or placebo) were imposed by the researchers on the people with acrophobia.

Part (b):

No, the experiment was designed to compare the D-cycloserine group with a control group that received the placebo. The researchers can conclude that the D-cycloserine pill and two therapy sessions show significantly more improvement than a placebo and two therapy sessions. However, there is no basis for comparison with another group of people with acrophobia who received eight therapy sessions and no pill.

Part (c):

One example is that if the therapists were allowed to choose who received the placebo and who received D-cycloserine, they might assign the people with more severe acrophobia to one of the groups and the people with less severe acrophobia to the other group. Thus, the improvement after only two therapy sessions could be related to the initial severity of the acrophobia rather than to the effects of D-cycloserine.

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 scored as follows:

Essentially correct (E) if the response indicates that this was an experiment, *AND* the explanation clearly communicates that two treatments were imposed.

Partially correct (P) if the response indicates that this was an experiment, *BUT* the explanation does not clearly communicate that two treatments were imposed.

Note: If the response indicates that this was an experiment because there was random assignment *to treatments*, this implies imposition of treatments and is scored as E. If the response does not clearly state that the random assignment is *to treatments*, this is scored as P.

Incorrect (I) if the response indicates that this is an observational study *OR* if the explanation is missing or incorrect.

Part (b) is scored as follows:

Essentially correct (E) if the response says "no" *AND* clearly explains why this is not reasonable based on the fact that there was no experimental group that received eight therapy sessions and no pill.

Incorrect (I) if the response provides an answer with an incorrect or no justification.

Part (c) is scored as follows:

Essentially correct (E) if the response indicates that this method of assignment might create experimental groups that differ in some systematic way other than the treatment *AND* provides a justification that describes the potential confounding,
    OR
if the response indicates that if the therapists know who was in which group, it may influence the therapists' behavior when dealing with or evaluating the people with acrophobia.

Partially correct (P) for any of the following:
- The response indicates that the assignment might create experimental groups that differ in some systematic way, *BUT* does not provide an explanation of the potential confounding.
- The response makes a general statement that the lack of random assignment could lead to confounding *BUT* does not provide an example in context.
- The response indicates that the therapists know who is in which group *BUT* does not give a reason as to why this might lead to a misleading conclusion.
- The response makes a general statement that failure to blind may lead to bias *BUT* does not provide an example in context.

Incorrect (I) if the response fails to meet the criteria for E or P.

Note: If the response discusses incorrect conclusions that might result from having the *people with acrophobia* (rather than the *therapists*) choose their own treatments, the response is scored as I, because such a response does not address the question asked.

4   **Complete Response**

         All three parts essentially correct

3   **Substantial Response**

         Two parts essentially correct and one part partially correct

2   **Developing Response**

         Two parts essentially correct and one part incorrect
     OR
         One part essentially correct and one or two parts partially correct.

1   **Minimal Response**

One part essentially correct and two parts incorrect

OR

One or two parts partially correct and one part incorrect

Question 3

Intent of Question

The primary goals of this question were to assess students' ability to (1) describe a situation as a series of Bernoulli trials; (2) calculate probabilities of events involving Bernoulli trials; (3) recognize whether or not an event is likely to occur.

Solution

Part (a):

Let Y denote the number of flights Sam must make until he receives his first upgrade. The random variable Y follows a geometric distribution with $p = 0.1$. The probability that Sam's upgrade will occur after his third flight is calculated below.

$$
\begin{aligned}
P(Y \geq 4) &= 1 - P(Y \leq 3) \\
&= 1 - \left[P(Y = 1) + P(Y = 2) + P(Y = 3) \right] \\
&= 1 - \left[0.1 + 0.9 \times 0.1 + (0.9)^2 \times 0.1 \right] \\
&= 1 - \left[0.1 + 0.09 + 0.081 \right] \\
&= 0.729
\end{aligned}
$$

Part (b):

Let p denote the probability that Sam will be upgraded to first class on a particular flight. Let X denote the number of upgrades Sam will receive in 20 flights. The random variable X follows a binomial distribution with $n = 20$ independent trials and $p = 0.1$. The probability that Sam will be upgraded exactly 2 times in his next 20 flights is calculated as follows.

$$
P(X = 2) = \binom{20}{2} (0.1^2)(0.9^{18})
$$
$$
\approx 0.2852
$$

Part (c):

Let X denote the number of upgrades Sam will receive in 104 flights. The random variable X follows a binomial distribution with $n = 104$ independent trials and $p = 0.1$. Thus,

$$
\begin{aligned}
P(X > 20) &= 1 - P(X \leq 20) \\
&\approx 1 - 0.9986 \\
&\approx 0.0014.
\end{aligned}
$$

Because this probability is so small, it is very unlikely that Sam would receive more than 20 upgrades in 104 flights if the airline's claim is correct. This would be expected to happen less than 1 percent of the time, indicating that one should be surprised if Sam receives more than 20 upgrades during the next year.

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the response includes a correct probability and shows supporting work.

Note: An alternative solution to part (a) is:

$$P(\text{first upgrade after third flight}) =$$
$$P(\text{no upgrade on (flight 1 or flight 2 or flight 3)}) =$$
$$P(\text{no upgrade on flight 1 and no upgrade on flight 2 and no upgrade on flight 3}) =$$
$$0.9 \times 0.9 \times 0.9 = 0.729$$

Partially correct (P) if the response includes a correct calculation of a related probability such as $P(Y=3) = (0.9)(0.9)(0.1)$ or $P(Y=4) = (0.9)(0.9)(0.9)(0.1)$, with supporting work, *OR* the response includes a correct probability with insufficient supporting work.

Note: The calculator command, 1 – geometcdf(0.1,3) with no other justification should be scored no higher than P, unless the value 0.1 is labeled as p somewhere in the response, in which case a score of E can be given.

Incorrect (I) if an answer is provided with no supporting work, *OR* an unreasonable probability (greater than 1 or less than 0) is provided.

Part (b) is scored as follows:

Essentially correct (E) if the response includes a correct probability and shows supporting work.

Partially correct (P) if the response includes supporting work but makes a calculation mistake that results in a reasonable incorrect probability. For example, $0.1^2 \times 0.9^{18}$ should be scored as partially correct, *OR* if the response includes a correct probability with insufficient supporting work.

*Note*s
- The calculator command, binomialpdf(20,0.1,2), with no other justification should be scored no higher than P, unless somewhere in the response, the value 0.1 is labeled as p and the value of 20 is labeled as n. In such a case, a score of E can be given for this part.
- If the response in part (a) was downgraded to P for failure to identify the geometric parameter, do not downgrade part (b) for this same oversight.

Incorrect (I) if an answer is provided with no supporting work, *OR* an unreasonable probability (greater than 1 or less than 0) is provided.

Part (c) is scored as follows:

Essentially correct (E) if the response includes an answer to the question that is linked to a correct probability.

Partially correct (P) for any of the following:
- The response includes an answer to the question *AND* provides a justification based on expected value.
- The response includes an answer to the question that is linked to an incorrect but reasonable probability.
- The response gives a correct probability *AND* includes a correct answer to the question but fails to link the two.
- The response gives a correct probability *BUT* fails to answer the question.

Incorrect (I) if the response fails to meet the criteria for E or P.

4   **Complete Response**

        All three parts essentially correct

3   **Substantial Response**

        Two parts essentially correct and one part partially correct

2   **Developing Response**

        Two parts essentially correct and one part incorrect

OR

        One part essentially correct and one or two parts partially correct

OR

        Three parts partially correct

1   **Minimal Response**

        One part essentially correct and two parts incorrect

OR

        Two parts partially correct and one part incorrect

Question 4

Intent of Question

The primary goals of this question were to assess students' ability to (1) specify hypotheses for the chi-square test of independence; (2) state and check the appropriate conditions for inference; (3) interpret standard statistical output; (4) identify and describe the type of error that could have been made.

Solution

Part (a):

H_0 : There is no association between perceived effect of part-time work on academic achievement and average time spent on part-time jobs.

H_a : There is an association between perceived effect of part-time work on academic achievement and average time spent on part-time jobs.

Part (b):

The following conditions for inference are met:
1. The students were randomly selected.
2. The expected cell counts should be at least 5. The computer output indicates that all expected counts are greater than 5. The smallest expected cell count is 6.825.

Part (c):

Because the *p*-value 0.007 is less than 0.05, H_0 should be rejected. There is convincing evidence that there is an association between the perceived effect of part-time work on academic achievement and average time spent on part-time jobs.

Part (d):

Because the null hypothesis was rejected, a Type I error may have been made. A Type I error is concluding that there is an association between the perceived effect of part-time work on academic achievement and the average time spent on part-time jobs when, in reality, there is no association between the two variables.

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the response includes the following three components:
1. The statement of no association (or independence) is in the null hypothesis, and the statement of association (or dependence) is in the alternative hypothesis.
2. The hypotheses do not imply a cause-and-effect relationship.
3. Acceptable terms are used for the two variables in the hypotheses.

Partially correct (P) if the response includes exactly two of the three components above.

Incorrect (I) if the response fails to meet the criteria for E or P.

Part (b) is scored as follows:

Essentially correct (E) if the response includes *BOTH* conditions necessary for the test and indicates that *BOTH* conditions are met for these data.

Partially correct (P) if only one of the necessary conditions is included *AND* the response indicates that the condition is met for these data, *OR* both conditions are stated, *BUT* the response does not indicate that the conditions are met for these data.

Incorrect (I) if response fails to meet the criteria for E or P.

Note: If the response also includes conditions that are not required for the chi-square test, the response should be scored no higher than P for this part.

Part (c) is scored as follows:

Essentially correct (E) if the response includes a correct conclusion, in context, *AND* provides a justification based on linkage between the *p*-value and the conclusion.

Partially correct (P) if the response includes a correct conclusion, with linkage to the *p*-value, *BUT* the conclusion is not in context, *OR* the response includes a correct conclusion, in context, *BUT* linkage to the *p*-value is missing.

Incorrect (I) if response fails to meet the criteria for E or P.

Notes
- The conclusion should be scored based on the hypotheses given in the response to part (a).
- If both an α and a *p*-value are given together, the linkage between the *p*-value and the conclusion is implied. 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 size of the *p*-value.
- A response that reaches a cause-and-effect conclusion cannot earn an E, unless this was already penalized in part (a). A response that includes a cause-and-effect conclusion should be scored as P, provided that the conclusion is in context and there is linkage to the *p*-value. It should be scored as I if it lacks either context or linkage to the *p*-value.

Part (d) is scored as follows:

Essentially correct (E) if a Type I error is identified and described in the context of the question.

Partially correct (P) if a Type I error is identified and a generic description of a Type I error, without context, is provided, *OR* correct statements are provided, in context, with an incorrect error name (Type II error).

Question 4 (continued)

Incorrect (I) if a Type II error is described, *OR* no description or an incorrect description is provided.

Note: Part (d) should be scored based on the hypotheses given in the response to part (a) and the conclusion in part (c).

Each essentially correct (E) part counts as 1 point. Each partially correct (P) part counts as ½ point.

4　　**Complete Response**

3　　**Substantial Response**

2　　**Developing Response**

1　　**Minimal Response**

If a response is between two scores (for example, 2½ points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.

Question 5

Intent of Question

The primary goals of this question were to assess students' ability to (1) identify and check appropriate conditions for inference; (2) identify and carry out the appropriate inference procedure; (3) determine the sample size necessary to meet certain specifications in planning a study.

Solution

Part (a):

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

One-sample (or large-sample) interval for p (the proportion of the vaccine-eligible people in the United States who actually got vaccinated) or $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$.

Conditions: 1. Random sample
                    2. Large sample ($n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$)

The stem of the problem indicates that a random sample of vaccine-eligible people was surveyed. The number of successes (978 vaccine-eligible people who received the vaccine), and failures (1,372 vaccine-eligible people who did not receive the vaccine), are both much larger than 10, so the large-sample interval procedure can be used.

Step 2: Correct mechanics

$$\left(\frac{978}{2,350}\right) \pm 2.57583 \sqrt{\frac{0.41617(1-0.41617)}{2,350}}$$
$$0.41617 \pm 2.57583 \times 0.01017$$
$$0.41617 \pm 0.02619$$
$$\left(0.38998, 0.44236\right)$$

Step 3: Interpretation

Based on the sample, we are 99 percent confident that the proportion of the vaccine-eligible people in the United States who actually got vaccinated is between 0.39 and 0.44. Because 0.45 is not in the 99 percent confidence interval, it is not a plausible value for the population proportion of vaccine-eligible people who received the vaccine. In other words, the confidence interval is inconsistent with the belief that 45 percent of those eligible got vaccinated.

Question 5 (continued)

Part (b):

The sample-size calculation uses 0.5 as the value of the proportion in order to provide the minimum required sample size to guarantee that the resulting interval will have a margin of error no larger than 0.02.

$$n \geq \frac{(2.576)^2(0.5)(0.5)}{(0.02)^2} = \left(\frac{2.576}{2(0.02)}\right)^2 = 4{,}147.36$$

Thus, a sample of at least 4,148 vaccine-eligible people should be taken in Canada.

<u>Scoring</u>

Each step in part (a) is scored as essentially correct (E), partially correct (P), or incorrect (I); and part (b) is scored as essentially correct (E), partially correct (P), or incorrect (I).

Step 1 of part (a) is scored as follows:

Essentially correct (E) if the one-sample z-interval for a proportion is identified (either by name or formula) *AND* both conditions (of random sampling and sample size) are stated and checked.

Partially correct (P) if the response identifies the correct procedure *BUT* adequately addresses only one of the two required conditions, *OR* if the response does not identify the correct procedure *BUT* adequately addresses both required conditions.

Incorrect (I) for any of the following:
- The response identifies the correct procedure *BUT* does not adequately address either required condition,
- The correct procedure is not identified *AND* at most one of the required conditions is adequately addressed,
- An incorrect procedure is identified.

Notes
- If the formula is of the correct form, even if incorrect numbers appear in it, then the procedure may be considered correctly identified.
- Stating only that $n\hat{p}$ and $n(1-\hat{p})$ both are greater than or equal to 10 is only a statement of the sample size condition and is not sufficient for checking the condition. The response must use specific values from the question to check the condition.
- If a response includes additional inappropriate conditions, such as $n \geq 30$ or requiring a normal population, then the response can earn no more than a P for this step. However, stating and checking a condition about the size of the sample relative to the size of the population is not required but is also not inappropriate.
- Any statement of hypotheses, definitions of parameters, statements of populations, etc. should be considered extraneous. However, if these statements are included and incorrect, this should be considered poor communication in terms of holistic scoring.

Step 2 of part (a) is scored as follows:

Essentially correct (E) if a 99 percent confidence interval is correctly computed.

Partially correct (P) for any of the following:
- If a correct method (confidence interval for a proportion) is used, *BUT* an incorrect critical z-value or a t-value is used.
- 0.45 is used for the value of $\hat{p}$.
- There are errors in the calculation of the interval (unless such errors follow from an incorrect procedure in step 1).

Incorrect (I) if an incorrect method is used, such as a t-interval for a population mean, *OR* if the resulting interval is unreasonable, such as an interval with integer endpoints.

Step 3 of part (a) is scored as follows:

Essentially correct (E) if the response notes that 0.45 is not in the 99 percent confidence interval *AND* states that this is evidence against the belief that 45 percent of vaccine-eligible people had received flu-vaccine.

Partially correct (P) if a reasonable statement about the belief that 45 percent of vaccine-eligible people is made, in context, but there is no clear connection made to the confidence interval, *OR* a clear connection to the confidence interval is made, but the response includes one or both of the following two omissions:
1. The response is not in context.
2. The response does not mention the confidence level of 99 percent.

Incorrect (I) if the response fails to meet the criteria for E or P.

Part (b) is scored as follows:

Essentially correct (E) if an appropriate sample size is calculated and supporting work is shown.

Partially correct (P) if supporting work is shown, *BUT* the response includes one or both of the following errors:
1. 0.41617 (the sample proportion) or 0.45 is used instead of 0.5.
2. An incorrect critical z-value is used — unless the same incorrect value was used in part (a).

Incorrect (I) if the response fails to meet the criteria for E or P.

Notes
- In this situation, the formula for margin of error used to compute sample size is only an approximation of the margin of error. Because of this, we will not insist that the computed sample size be rounded up; that is, 4,147 is scored as E, as long as supporting work is shown.
- If the critical value of 2.575 is used, then the sample size should be $n \geq 4,144.14$ or $n = 4,145$ (or 4,144).
- If the final recommended sample size is not an integer, then the response can earn no more than a P.

Question 5 (continued)

Each essentially correct (E) part counts as 1 point. Each partially correct (P) part counts as ½ point.

4   **Complete Response**

3   **Substantial Response**

2   **Developing Response**

1   **Minimal Response**

If a response is between two scores (for example, 2½ points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.

Question 6

Intent of Question

The primary goals of this question were to assess students' ability to (1) interpret the slope of a regression line in context; (2) decide whether or not a model should be used for prediction; (3) describe the sampling distribution of a sample mean; (4) use the sampling distribution of a sample mean to obtain an interval of plausible values; (5) compare two different study plans and decide which one would provide a better estimator of the slope; (6) propose a different study plan to check an assumption.

Solution

Part (a):

For each additional foot that is added to the width of the grass buffer strip, an additional 3.6 parts per hundred of nitrogen is removed on average from the runoff water.

Part (b):

No. This is extrapolation beyond the range of data from the experiment. Buffer strips narrower than 5 feet or wider than 15 feet were not investigated.

Part (c):

Because the distribution of nitrogen removed for any particular buffer strip width is normally distributed with a standard deviation of 5 parts per hundred, the sampling distribution of the mean of four observations when the buffer strips are 6 feet wide will be normal with mean $33.8 + 3.6 \times 6 = 55.4$ parts per hundred and a standard deviation of $\dfrac{\sigma}{\sqrt{n}} = \dfrac{5}{\sqrt{4}} = 2.5$ parts per hundred.
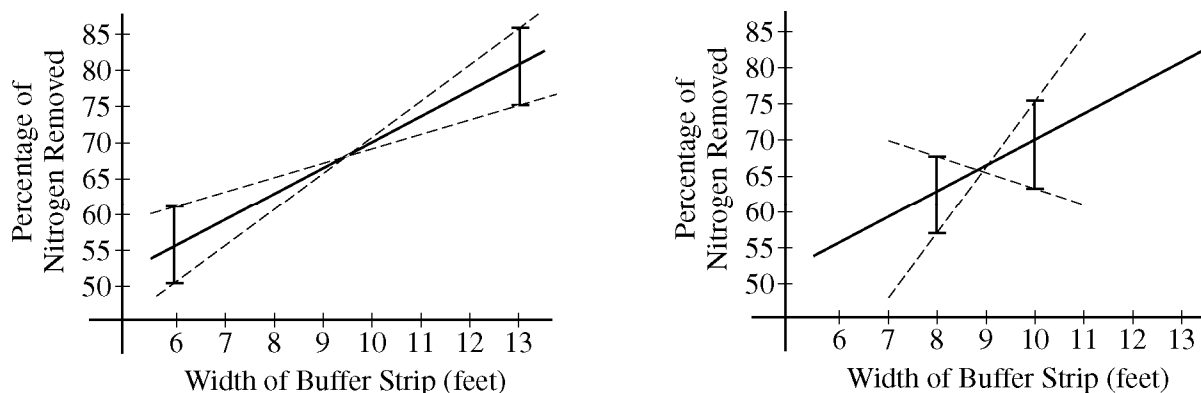
Part (d):

The distribution of the sample mean is normal, so the interval that has probability 0.95 of containing the mean nitrogen content removed from four buffer strips of width 6 feet extends from $55.4 - 1.96 \times 2.5 = 50.5$ parts per hundred to $55.4 + 1.96 \times 2.5 = 60.3$ parts per hundred.

Part (e):

If we think that the sample mean nitrogen removed at a particular buffer width might reasonably be any value in the intervals shown, a sample regression line will result from connecting any point in the interval above 6 to any point in the interval above 13. With this in mind, the dashed lines in the plots above represent extreme cases for possible sample regression lines. From these plots, we can see that there is a wider range of possible slopes in the second plot (on the right) than in the first plot (on the left). Because of this, the variability in the sampling distribution of b, the estimator for the slope of the regression line, will be smaller for the first study plan (with four observations at 6 feet and four observations at 13 feet) than it would be for the second study plan (with four observations at 8 feet and four observations at 10 feet). Therefore, the first study plan (on the left) would provide a better estimator of the slope of the regression line than the second study plan (on the right).

Part (f):

To assess the linear relationship between width of the buffer strip and the amount of nitrogen removed from runoff water, more widths should be used. To detect a nonlinear relationship it would be best to use buffer widths that were spaced out over the entire range of interest. For example, if the range of interest is 6 to 13 feet, eight buffers with widths 6, 7, 8, 9, 10, 11, 12 and 13 feet could be used.

Scoring

This question is scored in four sections. Section 1 consists of parts (a) and (b); section 2 consists of parts (c) and (d); section 3 consists of part (e); section 4 consists of part (f). Each of the four sections is scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the response includes the following two components:
1. The response in part (a) is correct, as evidenced by the correct interpretation of the slope, in context.
2. The response in part (b) is correct, as evidenced by the identification of extrapolation as the reason that the model should not be used *AND* the response is in context.

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

Incorrect (I) if the response fails to meet the criteria for E or P.

Notes
- Part (a) is incorrect if the interpretation is not in context or if the interpretation does not acknowledge uncertainty (for example, does not include "on average" or "about" or "approximately" or "predicted" when referring to the increase in nitrogen removed).
- Ideally a correct solution would also include units and make it clear that it is the approximate increase for each *additional* foot added to the buffer, but in the context of this larger investigative task, failure to do so is not sufficient to make part (a) incorrect.
- Part (b) is incorrect if extrapolation is not identified as the reason or if the response is not in context.

Section 2 is scored as follows:

Essentially correct (E) if the response includes the following two components:
1. The response in part (c) states that the sampling distribution is normal *AND* provides a correct mean and standard deviation.
2. The response in part (d) uses the correct mean and standard deviation of the sampling distribution — or incorrect values carried over from part (c) — *AND* a correct critical value (1.96 or 2) to compute a correct interval.

Partially correct (P) if only one of two components listed above is correct.

Incorrect (I) if the response fails to meet the criteria for E or P.

Notes
- Stating that the sampling distribution is *approximately* normal is acceptable for part (c).
- Part (c) is incorrect if the response does not state that the distribution is normal or if an incorrect mean or standard deviation is given.
- Part (d) is incorrect if an incorrect critical value (for example a *t* critical value) is used or if an incorrect mean or standard deviation — other than incorrect values from part (c) — is used in the computation of the interval.

Section 3 is scored as follows:

Essentially correct (E) if study plan 1 is chosen in part (e), and the response demonstrates awareness of sampling variation in the estimates of the slopes of the regression lines, *AND* this is clearly communicated in the context of the two study plans.

Partially correct (P) if study plan 1 is chosen in part (e), and the response demonstrates awareness of sampling variation in the estimates of the slopes of the regression lines, *BUT* the justification of the choice of study plan 1 is not clearly communicated.

Incorrect (I) if the response fails to meet the criteria for E or P.

Section 4 is scored as follows:

Essentially correct (E) if the response specifies another study plan that uses eight buffer strips of at least three different *specified* widths, and the response indicates in how many locations each width will be used, *OR* the response makes it clear that at least three different buffer widths will be used and indicates that the buffer widths to be used will be spread out over the range of interest.

Note: Specifying eight different widths is sufficient for an E in section 4.

Partially correct (P) if the response does not meet criteria for E, *BUT* the stated plan uses at least three different widths. Widths need not be specified for a P.

Incorrect (I) if the plan does not use at least three different widths.

Each essentially correct (E) section counts as 1 point. Each partially correct (P) section counts as ½ point.

| 4 | **Complete Response** |

| 3 | **Substantial Response** |

| 2 | **Developing Response** |

| 1 | **Minimal Response** |

If a response is between two scores (for example, 2½ points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.