CollegeBoard AP

AP® Statistics
2010 Scoring Guidelines

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.com.
AP Central is the official online home for the AP Program: apcentral.collegeboard.com.

Question 1

Intent of Question

The primary goals of this question were to assess students' ability to (1) apply terminology related to designing experiments; (2) construct an appropriate plot that could be used to investigate the fit of a linear model; (3) decide, from a graphical display, whether a linear regression model is appropriate for a set of data.

Solution

Part (a):

  i.   The treatments are the different concentrations of garlic in the food granules. Specifically, there are five treatments: 0 percent, 2 percent, 10 percent, 25 percent and 50 percent.
  ii. The experimental units are the birds (starlings), each placed in an individual cage.
  iii. The response is the number of food granules consumed by the bird.

Part (b):

  i.   The following scatterplot results from these data.
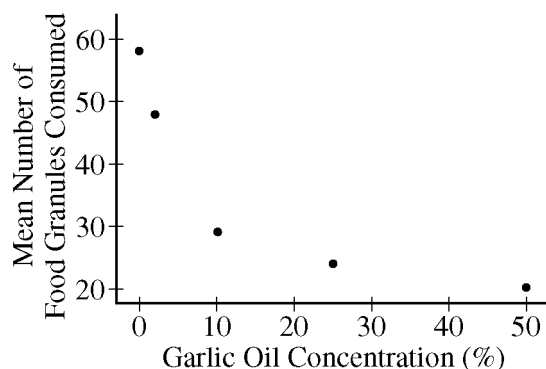
  ii. The curved pattern in this scatterplot reveals that a linear regression model would not be appropriate for modeling the relationship between these variables.

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the student correctly identifies all three subparts—the treatments, the experimental units and the response that will be measured.

Partially correct (P) if the student identifies two subparts correctly.

Incorrect (I) otherwise.

Notes
- In subpart ii, it is acceptable to identify the experimental units as the cages.
- In subpart iii, it is not correct to identify the response as the *mean* number of granules consumed.

Part (b) is scored as follows:

Essentially correct (E) if the student produces a correct graph (a reasonable scatterplot or residual plot with correct labels and scales) and then concludes, based on one or more features of the graph, that the pattern of the relationship does not appear to be linear.

Partially correct (P) if the student produces a correctly shaped graph but concludes that the data are linear *OR* if the student produces an incorrectly shaped graph but makes a reasonable conclusion based on one or more features of the graph.

Incorrect (I) if the student does not include a graph *OR* if the student produces an incorrect graph and does not make a reasonable conclusion based on one or more features of the graph.

Note: Any of the following will result in an incorrect graph.
- Incorrect scale
- Reversed axes
- Missing label(s)
- Other types of graph (histogram, bar graph, etc.)

4   Complete Response

   Both parts essentially correct

3   Substantial Response

   One part essentially correct and one part partially correct

2   Developing Response

   One part essentially correct and one part incorrect
   OR
   Both parts partially correct

1   Minimal Response

   One part partially correct and one part incorrect

Question 2

Intent of Question

The primary goals of this question were to assess students' ability to (1) describe a sampling distribution of a sample mean; (2) set up and perform a normal probability calculation based on the sampling distribution.

Solution

Part (a):

The sampling distribution of the sample mean song length has mean $\mu_{\bar{X}} = \mu = 3.9$ minutes and standard deviation $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{1.1}{\sqrt{40}} \approx 0.174$ minutes. The central limit theorem (CLT) applies in this case because the sample size ($n = 40$) is fairly large, especially with the population of song lengths having a roughly symmetric distribution. Thus, the sampling distribution of the sample mean song length is approximately normal.
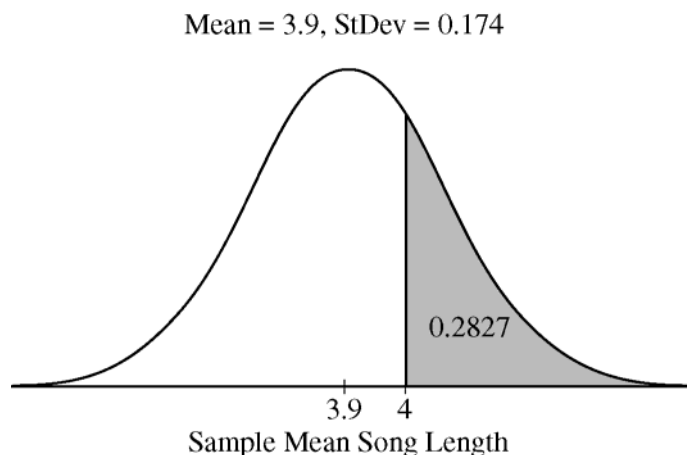
Part (b):

The probability that the total airtime of 40 randomly selected songs exceeds the available time (that is, the probability that the total airtime of 40 randomly selected songs is greater than 160 minutes) is equivalent to the probability that the sample mean length of the 40 songs is greater than $\dfrac{160}{40} = 4.0$ minutes.

According to part (a), the distribution of the sample mean length $\bar{X}$ is approximately normal. Therefore,

$$P(\bar{X} > 4.0) \approx P\left(Z > \frac{4.0 - 3.9}{0.174}\right) = P(Z > 0.57) = 1 - 0.7157 = 0.2843.$$

(The calculator gives the answer as 0.2827.)

The approximate sampling distribution of the sample mean song length and the desired probability are displayed below.

Mean = 3.9, StDev = 0.174

0.2827

3.9   4

Sample Mean Song Length

Part (b) (*alternative*):

An equivalent approach is to note that the sampling distribution of the total airtime, *T*, for the 40 songs is approximately normal, with mean $40(3.9) = 156$ minutes and standard deviation $\sqrt{40}(1.1) \approx 6.96$ minutes. The *z*-score for a total airtime of 160 minutes is then $z = \dfrac{160 - 156}{6.96} \approx 0.57$, and the calculation proceeds as above.

<u>Scoring</u>

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

Part (a) is scored as follows:

Essentially correct (E) if the student correctly provides all three components of the sampling distribution: shape (*approximately* normal), center (mean 3.9) and spread (standard deviation $\dfrac{1.1}{\sqrt{40}} \approx 0.174$).

Partially correct (P) if the student correctly provides only two of the three components.

Incorrect (I) if the student correctly provides only one or none of the components.

Notes
- Describing the sampling distribution as normal instead of approximately normal does not earn credit for the shape component.
- To earn credit for the spread component, the response must show how the standard deviation is calculated.
- If a response contains incorrect notation or terminology, it can at best be scored as partially correct (P).

Part (b) is scored as follows:

Essentially correct (E) if the student sets up and performs a correct normal probability calculation.

Partially correct (P) if the student sets up the normal probability calculation correctly but does not carry it through correctly *OR* sets up an incorrect but plausible calculation (for example, by using an incorrect standard deviation) but carries it through correctly.

Incorrect (I) if the student does not set up or perform the normal probability calculation correctly.

Notes
- A student can earn a score of essentially correct (E) in part (b) even with incorrect parameter values in part (a) by providing a correct calculation that uses the mean and standard deviation from part (a).
- Calculator syntax: An answer containing "normalcdf(…)" with no additional work or labeling is at best partially correct (P). If an appropriate sketch with the mean and standard deviation correctly labeled accompanies the calculator command, *OR* if the mean and standard deviation used in the calculator command are clearly identified in part (a) or part (b), then the response should be scored as essentially correct (E).
- If a student uses the sampling distribution of the total amount of time, *T*, needed to play the 40 randomly selected songs to do the probability calculation, the student must show how the standard deviation is calculated — unless this value is carried forward from part (a) — for the response to be scored as essentially correct (E). For example,

$$\sigma_T = \sqrt{40}\,\sigma_X = \sqrt{40}(1.1) \approx 6.96 \qquad OR \qquad \sigma_T = 40\sigma_{\bar{X}} = 40(0.174) \approx 6.96.$$

4   Complete Response

        Both parts essentially correct

3   Substantial Response

        One part essentially correct and one part partially correct

2   Developing Response

        One part essentially correct and one part incorrect

OR

        Both parts partially correct

1   Minimal Response

        One part partially correct and one part incorrect

Intent of Question

The primary goals of this question were to assess students' ability to (1) interpret the meaning of a
confidence level; (2) use a confidence interval to test the plausibility of a claim about the value of a
population parameter; (3) perform a sample size calculation related to a confidence interval.

Solution

Part (a):

The 95 percent confidence level means that if one were to repeatedly take random samples of the same
size from the population and construct a 95 percent confidence interval from each sample, then in the
long run 95 percent of those intervals would succeed in capturing the actual value of the population
proportion of households in the county that own at least one dog.

Part (b):

No. The 95 percent confidence interval 0.417 ± 0.119 is the interval $(0.298, 0.536)$. This interval
includes the value 0.39 as a plausible value for the population proportion of households in the county
that own at least one dog. Therefore, the confidence interval does not provide evidence that the
proportion of dog owners in this county is different from the claimed national proportion.

Part (c):

The sample proportion is 0.417, and the margin of error is 0.119. Determining the sample size requires

solving the equation $0.119 = 1.96 \times \sqrt{\dfrac{0.417 \times (1 - 0.417)}{n}}$ for n.

Thus, $n = \dfrac{1.96^2 \times 0.417 \times (1 - 0.417)}{0.119^2} \approx 65.95,$ so the humane society must have selected 66 households

for its sample.

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 student provides a correct interpretation of the confidence level in the
context of the study. A correct interpretation can take one of two approaches:

1. Based on the concept of repeated sampling, the response must fulfill the following three
    requirements:
    *   Mentions repeated sampling or "in the long run" or "using this method"
    *   Mentions that 95 percent of the intervals will capture the population proportion
    *   Includes the context of the study

2. Based on probability, the response must state that there is a 0.95 probability that a random sample *selected in the future* will produce an interval that captures the actual value of the population proportion of households in the county that have at least one dog.

Partially correct (P) if the student provides an interpretation of the confidence level that includes two of the three components required for the repeated sampling interpretation *OR* provides a correct probability interpretation, but not in context.

Incorrect (I) if the student attempts to interpret a particular confidence *interval* rather than the confidence *level* (for example, by saying that we are 95 percent confident that an interval that has been obtained includes the population proportion of households in the county that have at least one dog) *OR* provides an interpretation of the confidence level that mentions at most one of the three components required for the repeated sampling interpretation.

Part (b) is scored as follows:

Essentially correct (E) if the student correctly states that because 0.39 (or "the claimed value") is in the computed interval, the interval does not provide evidence that the proportion of dog owners in the county is different from the claimed national proportion.

Partially correct (P) if the student indicates that the goal is to check whether the claimed value of 0.39 is in the computed interval but makes errors in implementation. Examples of errors include the following:
- The student notes that 0.39 is within the interval but does not draw a correct conclusion.
- The student makes an arithmetic error in computing the endpoints of the interval, but the conclusion is consistent with the computed interval.

    OR

The student correctly notes that 0.39 is in the interval and then concludes that 0.39 *is* the population proportion for the county.

Incorrect (I) if the student does not recognize how to check whether the claim is consistent with the confidence interval.

Part (c) is scored as follows:

Essentially correct (E) if the student provides a correct equation with correct numerical values substituted, as well as a correct integer solution.

Partially correct (P) if the student provides a correct equation (and substitutions) but makes one or more of the following errors:
- Does not complete the calculation or completes the calculation incorrectly
- Uses 0.5 or 0.39 rather than 0.417 as the sample proportion
- Uses an incorrect but plausible z^* value
- Reports the answer as a non-integer value
- Gives the calculated value of n as a lower bound for the required sample size

Incorrect (I) otherwise.

Notes
- It is acceptable to use $z^* = 2$ instead of 1.96.
- It is acceptable for the response to round up or down to get an integer answer.

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 (1) assess students' ability to calculate an expected value and a standard deviation; (2) recognize the applicability of a binomial distribution and perform a relevant binomial probability calculation (or recognize the applicability of a normal approximation and use it to perform a relevant probability calculation); (3) suggest an appropriate sampling method to achieve a given goal.

Solution

Part (a):

Because the population size is so large compared with the sample size ($\frac{297,354}{2,000} \approx 148.7$ times the sample size), far greater than the usual standard of 10 or 20 times larger, we can use the binomial probability distribution even though this is technically sampling without replacement. The parameters of this binomial distribution are the sample size, n, which has a value of $n = 2,000,$ and the proportion of new car buyers who bought model E, p, which has a value of $p = \frac{2,323}{297,354} \approx 0.0078.$ The expected value of the number of model E buyers in a simple random sample of 2,000 is therefore $n \times p = 2,000 \times 0.0078 \approx 15.62.$ The variance is $n \times p \times (1-p) = 2,000 \times 0.0078 \times (1-0.0078) \approx 15.50,$ so the standard deviation is $\sqrt{15.50} \approx 3.94.$

Part (b):

For the reason given in part (a), the binomial distribution with $n = 2,000$ and $p \approx 0.0078$ can be used here. The probability that the sample would contain fewer than 12 owners of model E is calculated from the binomial distribution to be $\sum_{x=0}^{11} \binom{2,000}{x} (0.0078)^x (0.9922)^{2,000-x} \approx 0.147$. This probability is small enough that the result (fewer than 12 owners of model E in the sample) is not likely, but this probability is also not small enough to consider the result very unlikely.

This binomial probability can also be evaluated using a normal approximation. This is reasonable because $n \times p = (2,000) \times (0.0078) = 15.6$ is larger than 10 and $n(1-p) = (2,000) \times (0.9922) = 1,984.4$ is much larger than 10. Using the mean and standard deviation from part (a) gives

$$P(X \le 11) \approx P\left(Z < \frac{12.0 - 15.62}{3.94}\right) = P(Z < -0.92) = 0.179.$$

Part (c):

Stratified random sampling addresses the concern about the number of owners for models D and E. By stratifying on car model and then taking a simple random sample of at least 12 owners from the population of owners for each model, the company can ensure that at least 12 owners are included in the sample for each model while maintaining a total sample size of 2,000. For example, the company could select simple random samples of sizes 755, 647, 560, 22 and 16 for models A, B, C, D and E, respectively, to make the sample size approximately proportional to the size of the owner population for each model.

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 correctly addresses the following two components:
- Calculation of the expected number of owners, showing a proper method for the calculation and providing the correct numerical value
- Calculation of the standard deviation for the number of owners, indicating recognition of the appropriate binomial distribution and providing the calculation and the correct numerical value

Partially correct (P) if the response contains only one of the two components listed above *OR* displays correct formulas for both the expected value and the standard deviation of a binomial distribution but fails to show both of the correct numerical values.

Incorrect (I) if the response provides only numerical values without showing how they were calculated.

Part (b) is scored as follows:

Essentially correct (E) if the student does any of the following:
- Recognizes the applicability of the binomial distribution, identifies the correct parameters, sets up the relevant probability calculation, and completes the calculation correctly
- Uses a normal probability approximation, identifying the relevant mean and standard deviation, and shows a correct calculation of the probability
- Provides an argument based on an appropriate *z*-score, or the number of standard deviations away from the mean, with a reasonable conclusion about likeliness

Partially correct (P) if the student does any of the following:
- Recognizes the applicability of the binomial distribution and identifies the correct parameters *BUT* sets up an incorrect cumulative binomial probability calculation
- Recognizes the applicability of the binomial distribution and shows the calculation correctly *BUT* does not identify the correct parameters in either part (a) or part (b)
- Recognizes the applicability of the normal approximation and identifies the correct parameters *BUT* incorrectly calculates the *z*-score or probability

Incorrect (I) otherwise.

Notes
- If the parameter values were properly identified in part (a), they do not have to be identified in part (b).
- If the response shows a correct calculation of the probability, no comment about likeliness is necessary. But such a comment is necessary if the response contains only a *z*-score without a probability or discusses standard deviations from the mean.
- With the normal calculation, it is acceptable for the response to show the probability that the normal value is below 11 or 11.5 or 12.

Part (c) is scored as follows:

Essentially correct (E) if the response describes an appropriate sampling method (e.g., stratified random sampling) that ensures all of the following:
- Total sample size of 2,000
- At least 12 owners for each of the five car models
- Random selection of owners

Partially correct (P) if the response mentions stratified random sampling but gives a weak description, or no description, of how to implement the procedure *OR* describes another appropriate sampling method but includes only two of the three components listed above.

Incorrect (I) otherwise.

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 5

<u>**Intent of Question**</u>

The primary goal of this question was to assess students' ability to set up, perform and interpret the results of a significance test. More specific goals were to assess students' ability to (1) state appropriate hypotheses; (2) identify the name of an appropriate statistical test and check appropriate assumptions/conditions; (3) compute the appropriate test statistic and *p*-value; (4) draw an appropriate conclusion, with justification, in the context of the study.

<u>**Solution**</u>

Step 1: States a correct pair of hypotheses

Let μ_B represent the population mean length of all adult fish of this species from Buy-Rite Pets, and
let μ_F represent the population mean length of all adult fish of this species from Fish Friends.

The hypotheses to be tested are H_0: $\mu_B = \mu_F$ versus H_a: $\mu_B < \mu_F$.

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

The appropriate test is a two-sample *t*-test. The first condition is that the samples are independent random samples from the two populations. This was stated in the question. The second condition is that the population distributions of fish lengths are normal. The following dotplots reveal no obvious departures from normality, so it appears reasonable to proceed with the two-sample *t*-test.
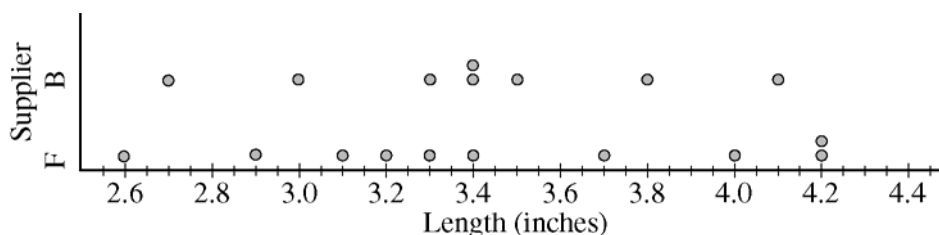
Step 3: Demonstrates correct mechanics, including the value of the test statistic, df and *p*-value (or rejection region)

The test statistic is: $t = \dfrac{\overline{x}_B - \overline{x}_F}{\sqrt{\dfrac{s_B^2}{n_B} + \dfrac{s_F^2}{n_F}}} = \dfrac{3.40 - 3.46}{\sqrt{\dfrac{0.434^2}{8} + \dfrac{0.550^2}{10}}} \approx -0.259$

With df = 15.99999, *p*-value = 0.3996.

Step 4: States a correct conclusion in the context of the problem, using the result of the statistical test

Because this *p*-value is larger than any conventional significance level (such as $\alpha = 0.10$ or $\alpha = 0.05$), we fail to reject H_0. The sample data do not provide convincing evidence to conclude that the mean length of the adult fish of the species from Fish Friends is greater than the mean length of the adult fish of the same species from Buy-Rite Pets.

Scoring

Each of steps 1, 2, 3 and 4 is scored as essentially correct (E), partially correct (P) or incorrect (I).

General Note: If a two-sample *t*-interval approach is taken without addressing the one-sided versus two-sided discrepancy, the student will lose credit for step 1 but still may earn full credit for steps 2, 3 and 4. The correct 95 percent confidence interval with df = 16 is from –0.55 to 0.43 inches.

Step 1 is scored as follows:

Essentially correct (E) if the student uses correct parameters *AND* states correct hypotheses.

Partially correct (P) if the student uses correct parameters *OR* states correct hypotheses but not both.

Incorrect (I) otherwise.

Notes
- If the null hypothesis is wrong, reduce the score in this step by one level (i.e., E to P, or P to I).
- If the alternative hypothesis is two-sided or in the wrong direction, the student does not get credit for the hypotheses.
- If standard symbols are used for the parameters with appropriate group labels (e.g., μ_B, μ_F), the parameter component is considered correct.
  - If *generic* standard symbols are used for the parameters (e.g., μ_1, μ_2), students must clearly identify the parameters with the suppliers.
  - If standard symbols (either with context or generic) are used for the parameters *and* the student attempts to define them, the definitions must be correct and in context, including the concept of mean.
  - If *nonstandard* symbols are used for the parameters, they must be explicitly defined in context and include the concepts of mean and population.
  - If a student does not use symbols in the hypotheses, the response can still receive an E as long as the alternative hypothesis is in the correct direction and it clearly refers to population means in context.

Step 2 is scored as follows:

Essentially correct (E) if the student correctly completes all three of the following components:
- Identifies the correct test procedure (by name or by formula)
- Checks for independent random samples
- Checks for normality

Partially correct (P) if the student correctly completes two of the three components listed above.

Incorrect (I) otherwise.

Notes

- A two-sample *z*-test is not a correct test procedure in this case, but if both conditions are checked correctly, this step is scored as partially correct.
- If a student chooses to conduct a pooled *t*-test, the equal variance condition must be addressed (e.g., by commenting on similarity of standard deviations or conducting a test for equality of variances) to get credit for choosing the appropriate test procedure.
- To get credit for the check of independent random samples, students must indicate that *more than one* random sample was taken.
- To get credit for the normality condition, students must include correct graphs of both distributions and include an appropriate comment about shape or outliers, such as "neither has outliers," "both are roughly symmetric," "no obvious departures from normality," "approximately normal," etc.
- Ignore additional conditions listed, as long as they are correct, such as "the sample sizes must be less than 10 percent of the population sizes." However, if the student includes additional *incorrect* conditions, such as $np > 10$, reduce the score in this step by one level (i.e., E to P, or P to I).

Step 3 is scored as follows:

Essentially correct (E) if the student correctly calculates both the test statistic and *p*-value.

Partially correct (P) if the student correctly calculates the test statistic but not the *p*-value *OR* omits the test statistic but correctly calculates the *p*-value.

Incorrect (I) otherwise.

Notes

- It is acceptable for students to use the conservative df (df = 7) or use the *t*-table to get a *p*-value > 0.25.
- Students who incorrectly choose a two-sample *z*-test lose credit for identifying the correct test procedure in step 2 but can earn full credit in step 3 if they provide the correct *z*-statistic ($z = -0.259$) and *p*-value (*p*-value = 0.3978).
- If the alternative hypothesis is two-sided, the *p*-value must be approximately 0.8 to get credit for the *p*-value component.
- If a student provides the correct test statistic and/or *p*-value but shows additional incorrect work, such as a wrong formula, reduce the score in this step by one level (i.e., E to P, or P to I).

Step 4 is scored as follows:

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

Partially correct (P) if the student provides a correct conclusion, with linkage to the *p*-value, but not in context *OR* provides a correct conclusion in context, but without justification based on linkage to the *p*-value.

Incorrect (I) otherwise.

Notes
- The conclusion must be about the *mean* fish lengths to get credit for context unless the student already lost credit in step 1 for neglecting to include the concept of mean.
- The conclusion must be consistent with the alternative hypothesis to get credit for context unless the student was already penalized for inconsistency with the alternative hypothesis in step 3.
- If the conclusion is consistent with an incorrect *p*-value from step 3, and also in context, with justification based on linkage to the *p*-value, then this step is scored essentially correct.
- If both a significance level α and a *p*-value are given together, the linkage between the *p*-value and the conclusion is implied. If no α level 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, such as saying: "Because the *p*-value is small, we reject the null hypothesis" or "Because the *p*-value is large, we do not reject the null hypothesis."
- If the student chooses to "retain the null hypothesis," with linkage and/or context, this should be scored partially correct (P). If the student goes on to say something equivalent to "fail to reject" (e.g., "we should not conclude the mean length of fish is greater at Fish Friends") in context, with linkage, then the response should be scored essentially correct.
- A conclusion in step 4 that is equivalent to "accept H_0" (such as "we conclude that the mean fish length is the same from both suppliers") cannot be scored essentially correct. Such a response should be scored partially correct, provided that the conclusion is in context, with justification based on linkage to the *p*-value. Such a response should be scored incorrect if it lacks either context or linkage to the *p*-value.
- If a student attempts to interpret the *p*-value, but does so incorrectly, then do not give credit for the linkage component.

Each essentially correct (E) step counts as 1 point. Each partially correct (P) step 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 determine whether to score up or down, depending on the overall strength of the response and communication.
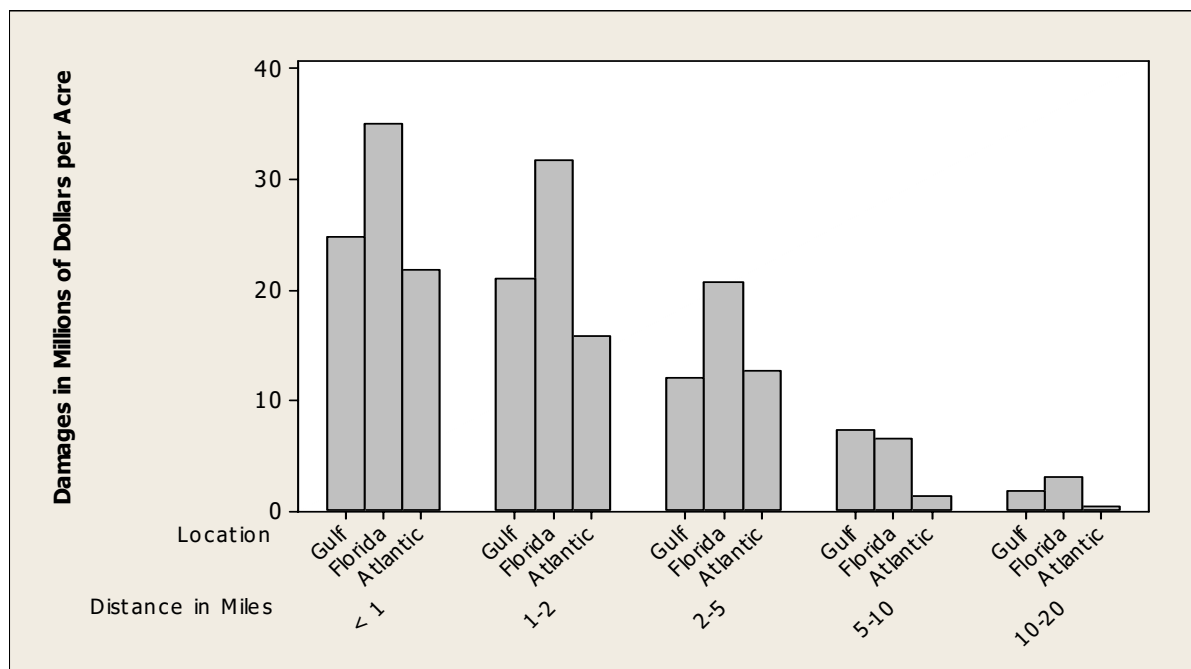
Question 6

Intent of Question

The primary goals of this investigative task were to assess students' ability to (1) produce and comment on a graphical display; (2) calculate a test statistic based on rank data; (3) use simulation results to draw an appropriate conclusion.

<u>**Solution**</u>

Part (a):

Part (b):

In all three regions (Gulf Coast, Florida, Lower Atlantic) the hurricane damage amounts tend to decrease as distance from the coast increases. For almost all given distances from the coast, the Florida region has the largest damage amounts. Also, for any given distance, the Gulf Coast and Lower Atlantic regions have similar damage amounts but with the Lower Atlantic damage amounts generally smaller.

Part (c):

For the "10 to 20 miles" distance category: The Florida region has the most damage (3.0 million dollars per acre) and so has rank 1. The region with the second-most damage is the Gulf Coast (1.7 million dollars), obtaining rank 2. The Lower Atlantic region has the least damage (0.3 million dollars) and so has rank 3. The last columns of the table should be filled in as follows:

	10 to 20 miles	Average Rank
Gulf Coast	2	2.0
Florida	1	1.2
Lower Atlantic	3	2.8

The average ranks are computed $\dfrac{2+2+3+1+2}{5}=2.0$ for the five Gulf Coast damage ranks,

$\dfrac{1+1+1+2+1}{5}=1.2$ for the five Florida damage ranks and $\dfrac{3+3+2+3+3}{5}=2.8$ for the five Lower Atlantic damage ranks.

Part (d):

The calculated value of the test statistic Q is

$$Q = 5\left[(2.0-2)^2 + (1.2-2)^2 + (2.8-2)^2\right] = 5[0+0.64+0.64] = 6.4.$$

Part (e):

A Q value of 6.4 or larger occurred in $\dfrac{39}{1,000}=0.039$ (or 3.9 percent) of the 1,000 repetitions. All 1,000 repetitions of the simulation assumed there was no difference in the distribution of damage amounts among the three regions. This is a fairly small (approximate) p-value (less than 0.05), indicating that a test statistic as large or larger than the observed test statistic of $Q = 6.4$ would be fairly unlikely to occur by chance alone if there really was no difference among the regions for each distance category. The sample data therefore provide reasonably strong evidence that there is a difference in the distributions of hurricane damage amounts among these three regions.

Scoring

This question is scored in four sections. Section 1 consists of part (a); section 2 consists of part (b); section 3 consists of parts (c) and (d); section 4 consists of part (e). 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 a well-labeled, statistically valid graph that allows for comparing damage amounts both across regions *AND* at varying distances from the coast.

Partially correct (P) if the response includes a well-labeled, statistically valid graph that allows for comparing damage amounts across regions or at varying distances from the coast but not both *OR* if it includes a statistically valid graph that allows for both comparisons but lacks labels for either the regions or the distances (or both).

Incorrect (I) if the graph does not allow for either comparison *OR* if it is poorly labeled and allows for only one comparison.

Section 2 is scored as follows:

Essentially correct (E) if the response includes *BOTH* of the following:
- Mention of a difference — a contrast among the three regions across distance categories (e.g., "Florida has greater damage than the other two regions at most distances")
- Mention of a similarity — a statement that in all regions hurricane damage decreases with greater distance from the coast

Partially correct (P) if the response includes a valid statement about only a difference or about only a similarity.

Incorrect (I) if the response has neither a valid statement about the difference nor a valid statement about the similarity.

Note: A statement that "Florida has more damage" is not sufficient to describe a difference if "at most distances" is not specified. Phrases that imply comparisons across distances such as "generally" or "for the most part" are acceptable for the "at most distances" specification.

Section 3 is scored as follows:

Essentially correct (E) if the response includes the correct ranks in part (c) *AND* the correct test statistic calculation in part (d).

Partially correct (P) if the response includes either of the following:
- The correct ranks in part (c) but not the correct test statistic calculation in part (d)
- Incorrect ranks in part (c) but the test statistic calculated correctly in part (d), using those incorrect ranks

Incorrect (I) otherwise.

Notes
- Calculation of the average ranks need not be shown to receive credit for correctly calculating Q.
- A miscalculation of Q results in no credit for part (d), but the severity of the miscalculation may be considered later in holistic scoring.

Section 4 is scored as follows:

Essentially correct (E) if the response includes *BOTH* of the following components:
- A valid quantification of the tail probability of the test statistic Q based on the simulated Q values
- A conclusion, in context, consistent with the rarity (or likelihood) of the calculated test statistic Q relative to the simulated Q values

Partially correct (P) if the response includes only one of these two components.

Incorrect (I) if the response does not include either of these two components.

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 determine whether to score up or down, depending on the overall strength of the response and communication.