AP® STATISTICS
2012 SCORING GUIDELINES

Question 1

Intent of Question

The primary goals of this question were to assess students' ability to (1) describe a nonlinear association based on a scatterplot; (2) describe how an unusual observation may affect the appropriateness of using a linear model for bivariate numeric data; (3) implement a decision-making criterion on data presented in a scatterplot.

Solution

Part (a):

The data show a weak but positive association between price and quality rating for these sewing machines. The form of the association does not appear to be linear. Among machines that cost less than $500, there appears to be very little association between price and quality rating. But the machines that cost more than $500 do generally have better quality ratings than those that cost less than $500, which causes the overall association to be positive.

Part (b):

The sewing machine that most affects the appropriateness of using a linear regression model is the one that costs about $2,200 and has a quality rating of about 65. Although the other four sewing machines costing more than $500 generally have higher quality ratings than those costing under $500, their prices and quality ratings follow a trend that suggests that quality ratings may not continue to increase with higher prices, but instead may approach a maximum possible quality rating. The $2,200 sewing machine is the most expensive of all but has a relatively low quality rating, which is consistent with a nonlinear model that approaches a maximum possible quality rating and then perhaps decreases. If a linear model were fit to all of the data, this one machine would substantially pull the regression line toward it, resulting in a poor overall fit of the line to the data.
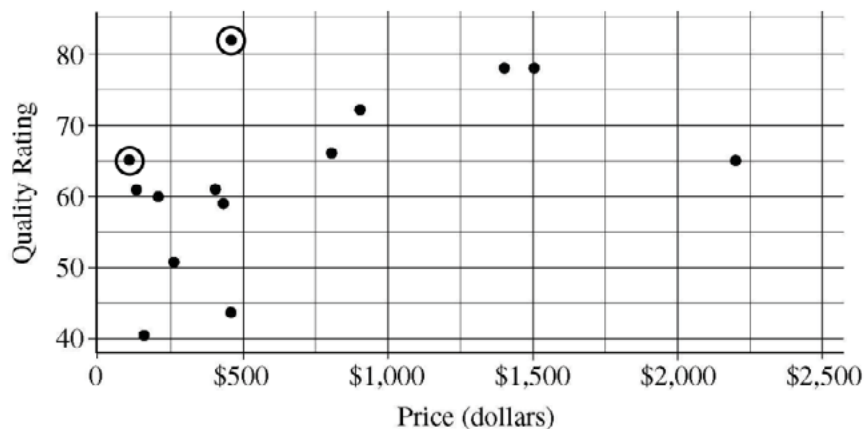
Part (c):

According to Chris's criterion, there are two sewing machine models that he will consider buying:
1. The model that costs a bit more than $100 and has a quality rating of 65.
2. The model that costs a bit below $500 and has a quality rating of 81 or 82.

The data points corresponding to these two machines have been circled on the scatterplot below.

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 correctly describes three aspects of association: direction (positive), strength (weak or moderate), and form (curved or nonlinear), *AND* describes the association in context.

Partially correct (P) if the response correctly describes two aspects of association in context
  OR
if the response describes all three aspects of association without context.

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 identifies the correct point with reasonable approximations to the price and quality values *AND* gives either of the following two explanations:
1. The point in conjunction with the entire collection of points appears to have a curved (or nonlinear) form.
2. A linear model that includes all the points would result in a poor overall fit to the data, largely owing to the presence and influence of the identified point.

AP® STATISTICS
2012 SCORING GUIDELINES

Question 1 (continued)

Partially correct (P) if the response identifies the correct point with reasonable approximations to the price and quality values *AND* gives a weak explanation of why the point affects the reasonableness of a linear model. The following are examples of weak explanations.
1. The point is an outlier.
2. Removal of the point makes the pattern more linear.
3. The point does not follow the linear pattern of the others.
4. A sewing machine this expensive should have a higher quality rating.
5. There is a much cheaper sewing machine with the same quality rating as this one.
6. The point has considerable influence on the parameters of the least squares regression line.

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

Part (c) is scored as follows:

Essentially correct (E) if the correct two points are circled *AND* no other points are circled.

Partially correct (P) if the correct two points are circled *AND* one or two other points are circled.
   OR
if only one of the two correct points is circled *AND* at most one other point is circled.

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 two parts partially correct
OR
One part essentially correct and one part partially correct
(*BUT* see the exception noted with an asterisk below)
OR
All three parts partially correct

1   Minimal Response

One part essentially correct and two parts incorrect
OR
*Part (c) essentially correct, part (b) partially correct, and part (a) incorrect
OR
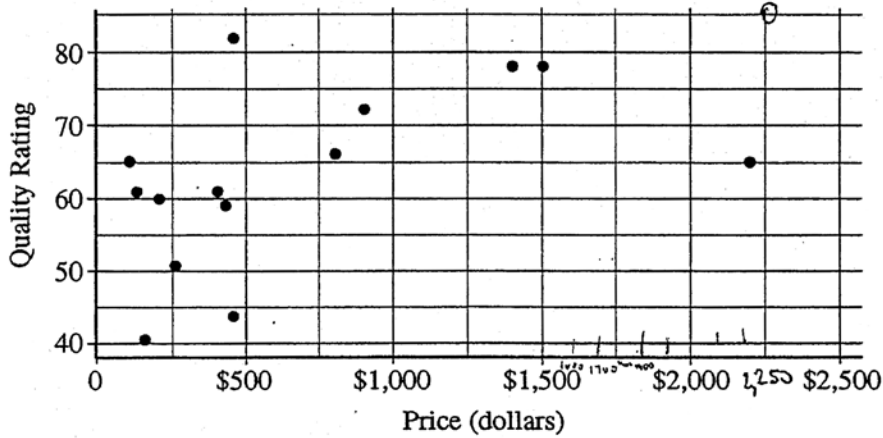Two parts partially correct and one part incorrect

STATISTICS
SECTION II
Part A
Questions 1-5
Spend about 65 minutes on this part of the exam.
Percent of Section II score—75

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. The scatterplot below displays the price in dollars and quality rating for 14 different sewing machines.

(a) Describe the nature of the association between price and quality rating for the sewing machines.

Strength, positive, linear ✓

The association between price and quality rating as depicted by the Scatterplot above appears to be a positive association that is moderate in strength. The association also appears to be slightly curved in nature. In general, as the price in dollars increases, so does the quality rating, although there are some exceptions to this statement such as the quality rating dropping from approximately 51 to 44 as the price in dollars moved up. Because of this slightly curved model, we may infer that a linear model would not be appropriate for this data.
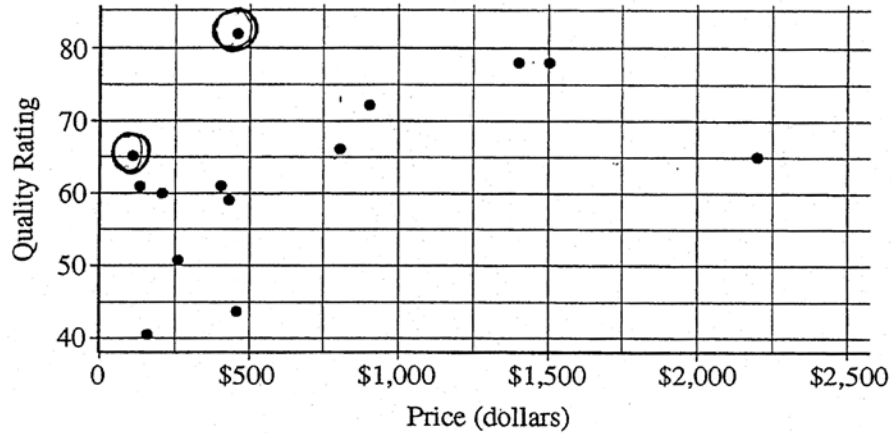
GO ON TO THE NEXT PAGE.

-6-

(b) One of the 14 sewing machines substantially affects the appropriateness of using a linear regression model to predict quality rating based on price. Report the approximate price and quality rating of that machine and explain your choice.

The approximate **price** and quality rating of this sewing machine is $2200 and 65, respectively. This point is clearly an influential point, meaning that if it were removed, the data would be greatly affected. This also contributes to the curved nature because it does not follow the rule that when price increases, quality rating does as well. If it did follow this rule, we would estimate this point at around $2200 to have a quality rating around 85.

(c) Chris is interested in buying one of the 14 sewing machines. He will consider buying only those machines for which there is no other machine that has both higher quality and lower price. On the scatterplot reproduced below, circle all data points corresponding to machines that Chris will consider buying.
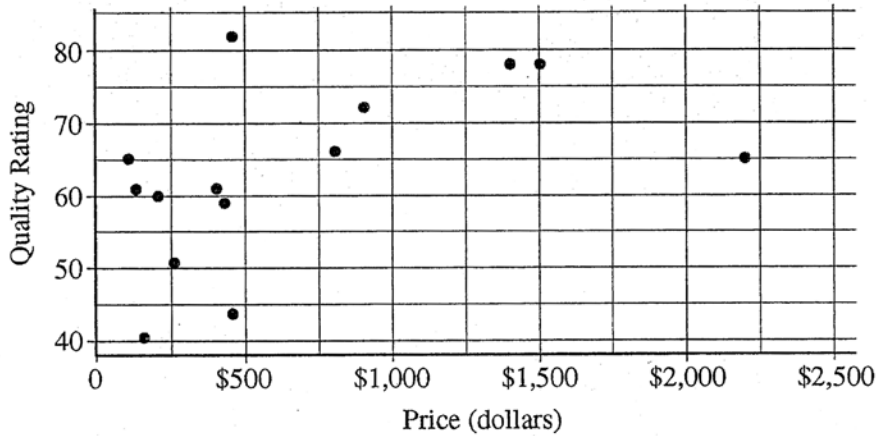
GO ON TO THE NEXT PAGE.

-7-

STATISTICS
SECTION II
Part A
Questions 1-5
Spend about 65 minutes on this part of the exam.
Percent of Section II score—75

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. The scatterplot below displays the price in dollars and quality rating for 14 different sewing machines.

(a) Describe the nature of the association between price and quality rating for the sewing machines.

The association seems to have a weak positive correlation between price and quality rating of the sewing machines. There are several influential points, but some may increase the sum of the squared residuals, therefore creating a weaker association (lower R). However, the weak positive correlation indicates that an increase in price will increase the quality rating.

The low value of R^2 also tells that very little percent of variation in the dependent variable can be explained by the price.

GO ON TO THE NEXT PAGE.

(b) One of the 14 sewing machines substantially affects the appropriateness of using a linear regression model to predict quality rating based on price. Report the approximate price and quality rating of that machine and explain your choice.

Price : ≈ 2200 dollars

Quality : ≈ 65

This point denotes a substantial reason why not to use a linear regression model because a rapid increasing linear trend is shown with the removal of the point. However, with the outlier, the trend is more along the lines of a curve regression rather than a linear regression
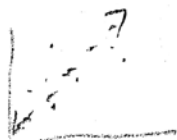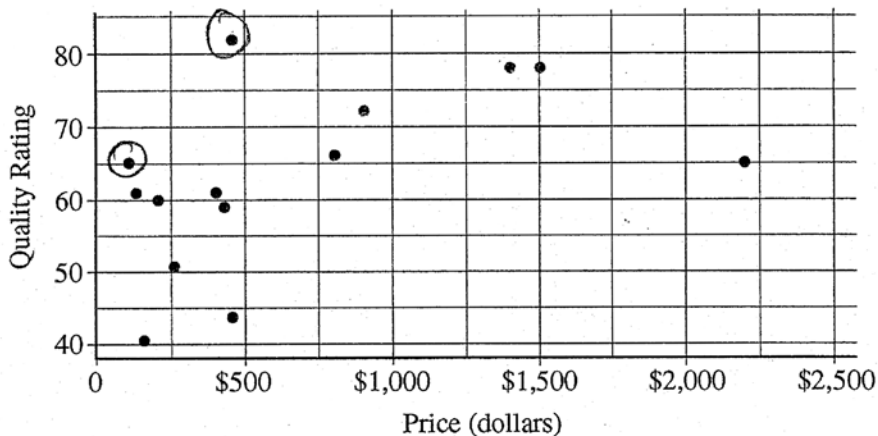
without   with

(c) Chris is interested in buying one of the 14 sewing machines. He will consider buying only those machines for which there is no other machine that has both higher quality and lower price. On the scatterplot reproduced below, circle all data points corresponding to machines that Chris will consider buying.
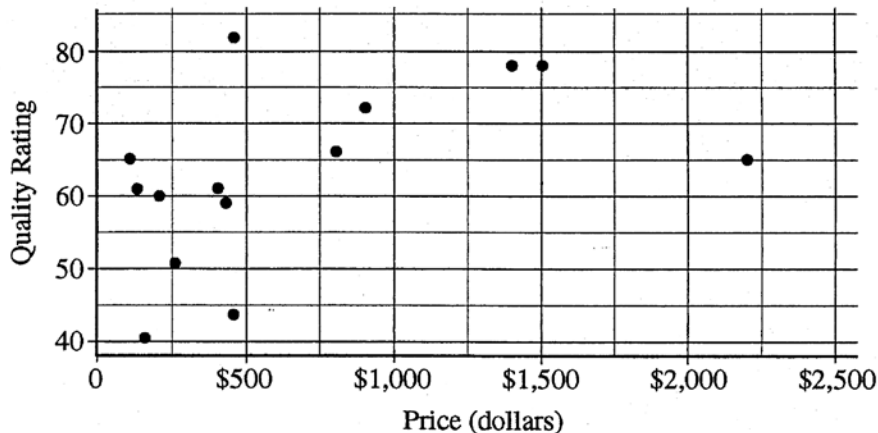
GO ON TO THE NEXT PAGE.

© 2012 The College Board.
Visit the College Board on the Web: www.collegeboard.org.

STATISTICS
SECTION II
Part A
Questions 1-5
Spend about 65 minutes on this part of the exam.
Percent of Section II score—75

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. The scatterplot below displays the price in dollars and quality rating for 14 different sewing machines.

(a) Describe the nature of the association between price and quality rating for the sewing machines.

The scatter Plot shows a somewhat linear relationship between price and quality rating. The majority of the observations falls between $0-$500 with an average quality rating of 60. As you go beyond the $0-$500 range, the quality rating goes up. There is an outlier in the $2000-$2500 range. This will affect the regression line since it is an outlier in the x direction. If the outlier were removed The chart would be a better Predictor of quality rating given any price from $0-$2500.
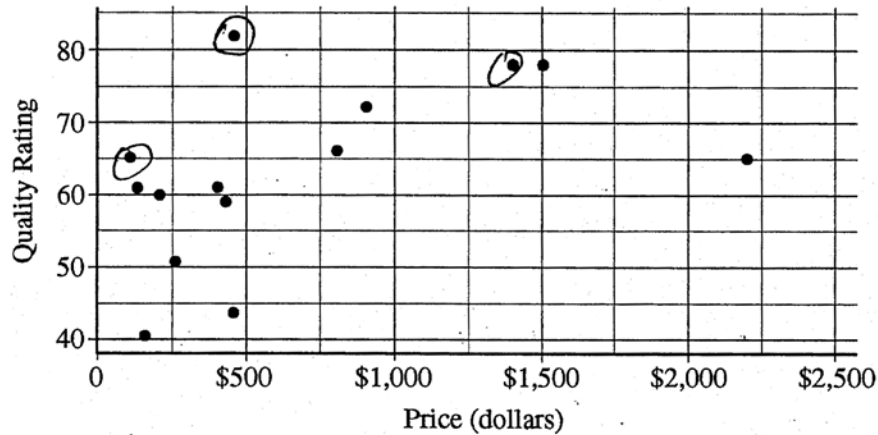
GO ON TO THE NEXT PAGE.

(b) One of the 14 sewing machines substantially affects the appropriateness of using a linear regression model to predict quality rating based on price. Report the approximate price and quality rating of that machine and explain your choice.

The Price of the influential Point is approximately $2,200 with a quality rating of 65. This Point affects the regression line by changing the Slope of the line. If the Point were to be kept ine the graph, than the Slope of the regression line will not predict the quality ratings as accurately.

(c) Chris is interested in buying one of the 14 sewing machines. He will consider buying only those machines for which there is no other machine that has both higher quality and lower price. On the scatterplot reproduced below, circle all data points corresponding to machines that Chris will consider buying.

GO ON TO THE NEXT PAGE.

© 2012 The College Board.
Visit the College Board on the Web: www.collegeboard.org.

Question 1

Overview

The primary goals of this question were to assess students' ability to (1) describe a nonlinear association based on a scatterplot; (2) describe how an unusual observation may affect the appropriateness of using a linear model for bivariate data; (3) implement a decision-making criterion on data presented in a scatterplot.

**Sample: 1A
Score: 4**

In part (a) the student correctly describes the direction, strength, and form of the association between the price and quality rating of the sewing machines and does so in context. The direction and strength are correctly described in the first sentence, where the student writes that the association "appears to be a positve [*sic*] association that is moderate in strength." The form of the association is correctly described in the second sentence, where the student writes that the association "appears to be slightly curved in nature." Although further description of the direction and strength of the association is not required for the response in part (a) to be scored as essentially correct, it is a strength of this response that the next sentence describes what it means for the association to be positive, using natural language rather than technical terms: "In general, as the price in dollars increases, so does the quality rating." And although further elaboration on the form of the association is not required for the response in part (a) to be scored as essentially correct, it is a strength of this response that the student concludes by saying that "[b]ecause of this slightly curved modil [*sic*], we may infer that a linear model would not be appropriate for this [*sic*] data." Part (a) was scored as essentially correct. In part (b) the student identifies the correct point ("approximate price and quality rating of this sewing machine is [*sic*] $2200 and 65, respectively"). The second sentence alone would have been considered a weak explanation of why the point substantially affects the appropriateness of using a linear model, because it says only that "if it were removed, the data would be greatly affected" but does not describe *how* the data or a model for the data would be affected. However, the following sentence does describe how the data point, in conjunction with the other points, makes the pattern look curved: "This also contributes to the curved nature because it does not follow the rule that when price increases, quality rating does as well." And although the response to part (b) would have been scored as essentially correct even without the last sentence, it is a strength of the response that the student concludes by estimating what the quality rating of the most expensive machine should have been in order for it to follow a linear trend. Part (b) was scored as essentially correct. In part (c) the correct two points are circled, and no other points are circled, so part (c) was scored as essentially correct. Because all three parts were scored as essentially correct, the response earned a score of 4.

**Sample: 1B
Score: 3**

In part (a) the student begins by writing, "[t]he association seems to have a weak positive correlation between price and quality rating of the sewing machines." This is a correct description of the strength and direction of the association, and it is given in context. Most of the remaining statements are either further elaborations of direction ("an increase in price will increase the quality rating") or of strength ("very little percent of variation in the dependent variable can be explained by the price"). Although such elaborations are not required, they show greater understanding of statistical principles than is demonstrated in the first sentence alone, thus making the response stronger. Influential points are mentioned, but they are not given as exceptions to a linear form; indeed, the response in part (a) includes no description of the form of the association. Because the strength and direction of the association are described in context, but the

form of the association is not described, the response in part (a) was scored as partially correct. In part (b) the student identifies the correct point ("Price: ≈ 2200 dollars Quality: ≈ 65") and explains that the reason it substantially affects the appropriateness of a linear regression model is that "with the outlier, the trend is more along the lines of a curve regression rather than a linear regression." The explanation is correct, so part (b) was scored as essentially correct. In part (c) the correct two points are circled, and no other points are circled, so part (c) was scored as essentially correct. Because two parts were scored as essentially correct and one part was scored as partially correct, the response received a score of 3.

**Sample: 1C
Score: 2**

In part (a) the response correctly describes, in context, the direction and form of the association between price and quality rating for the sewing machines, but the response does not address the strength of the association. In the first sentence the student describes the form of the association as "somewhat linear." Addressing possible curvature would have been a stronger response, and "linear," without the word "somewhat," would not have been acceptable; but "somewhat linear" is acceptable as a description of the form of the association. The second sentence does not address the association of price and quality rating and was considered extraneous. The student then writes, "As you [move] beyond the $0–$500 range, the quality rating goes up," which is an accurate description of the direction of the association. The rest of the response in part (a) addresses an outlier but does not do so in a way that contributes meaningfully to a description of the association between price and quality rating. The response in part (a) includes a correct description of the form and direction of the association but does not address the strength, so part (a) was scored as partially correct. In part (b) the student identifies the correct point ("approximately $2,200 with a quality rating of 65") and then states, "This point affects the regression line by changing the slope of the line," which makes it clear that the student is thinking specifically about the *least squares* regression line, not about linear models in general. However, a description of the influence of an outlier on the least squares regression line is acceptable if the response describes how the least squares regression model on the full set of data would result in a poor overall fit, owing to the influence of the outlier. The response contains such a description in the last sentence: "If the point were to be kept in the graph, than [*sic*] the slope of the regression line will not predict the quality ratings as accurately." Part (b) was scored as essentially correct. In part (c) the correct two points are circled, but an additional point is circled as well, so part (c) was scored as partially correct. Because one part was scored as essentially correct and two parts were scored as partially correct, the response earned a score of 2.