# AP® STATISTICS
# 2016 SCORING GUIDELINES

## Question 6

### Intent of Question

The primary goals of this question were to assess a student's ability to (1) use a scatterplot to comment on a report about the relationship between two variables and interpret the slope for the least-squares regression line summarizing this relationship; (2) describe the relationship between two variables in a scatterplot when a categorical variable is introduced and compare a characteristic of the distribution of a variable for different categories of individuals in a scatterplot; and (3) describe how the associations between two variables for each category of individuals in a scatterplot differ from the overall association in the same scatterplot.

### Solution

**Part (a):**

The scatterplot supports the newspaper report about number of semesters needed to complete an academic program and starting salary because it shows a positive association between these two variables.

**Part (b):**

The slope is 1.1594. For each additional semester needed to complete an academic program, the predicted starting salary increases by €1,159.40.

**Part (c):**

For the business majors alone, there is a strong, negative, linear association between number of semesters and starting salary. Business majors who need a greater number of semesters to complete an academic program tend to have lower starting salaries.

**Part (d):**

Business majors have the lowest median starting salary at around €38,000, followed by physics majors at around €48,000, and then chemistry majors with the highest median starting salary at around €55,000.

**Part (e):**

The newspaper report should be modified to account for major. Overall, majors that take longer to complete tend to have higher starting salaries, with chemistry the highest, physics the next highest, and business the lowest. However, within a major, students who take a greater number of semesters tend to have lower starting salaries.

### Scoring

This question is scored in three sections. Section 1 consists of parts (a) and (b), section 2 consists of parts (c) and (d), and section 3 consists of part (e). Sections 1, 2, and 3 are scored as essentially correct (E), partially correct (P), or incorrect (I).

# AP® STATISTICS
# 2016 SCORING GUIDELINES

## Question 6 (continued)

**Section 1** is scored as follows:

Essentially correct (E) if the response includes the following five components:

1. In part (a) the response addresses the positive association.
2. In part (a) the response uses the positive association to justify that the scatterplot supports the newspaper report.
3. In part (b) the response correctly identifies the numerical value of the slope from the computer output.
4. In part (b) the response interprets the slope as the change in starting salary for each additional semester, in context.
5. In part (b) the interpretation of slope includes nondeterministic language (e.g., "predicted starting salary" or equivalent) when interpreting the slope.

Partially correct (P) if the response includes three or four of the five components.

Incorrect (I) if the response does not meet the criteria for E or P.

*Notes*:

- In part (a) the response can use phrases such as "positive association (correlation, relationship)," "increasing relationship," or describe a positive association (e.g., "the starting salaries are higher when there is a greater number of semesters") to satisfy component 1. However, describing the relationship between only two points does not satisfy this component.
- In part (a) comments about linearity and strength should be ignored, even if the response implies these are required (e.g., "Yes, because the relationship is strong, positive, and linear.").
- In part (a) responses that answer "no," "maybe," "kind of," "somewhat," or equivalent do not satisfy component 2.
- In part (a) a response that says "no, because the three clusters of points each have a negative association" does not satisfy component 1 or component 2.
- In part (a) no context is required, but variable names are required in part (b) to satisfy component 4.
- In part (b) a response that states the equation $\hat{y} = 34.018 + 1.1594x$ does not satisfy component 3 unless the slope is specifically identified or used in the interpretation.
- In part (b) a response that incorrectly identifies the numerical value of the slope can still satisfy components 4 and 5 using the incorrect value.
- In part (b) the 1-unit increase in number of semesters must be stated or implied to satisfy component 4 (e.g., for each semester, for every semester). A response that states or implies an unspecified number of semesters does not satisfy this component (for example, as semesters increase).
- In part (b) examples of nondeterministic language include "predicted starting salary," "expected starting salary," "estimated starting salary," "typical starting salary," "average starting salary," "starting salary, on average," "our model predicts," and so on. However, "about," "approximately," and "according to the model" do not satisfy component 5.
- In part (b) no units are required for the change in predicted starting salary (which means it is OK to say 1.1594 euros, 1159.40, 1.1594, 1,159.40 dollars, and so on).
- In all parts it is acceptable if a response refers to salary rather than starting salary.

# AP® STATISTICS
# 2016 SCORING GUIDELINES

## Question 6 (continued)

**Section 2** is scored as follows:

Essentially correct (E) if the response includes the following five components:

1. In part (c) the response states that the association is negative.
2. In part (c) the response states that the association is strong *OR* linear *OR* both.
3. In part (c) the response refers to both variables (semesters, salary) in context.
4. In part (d) the response correctly compares the three majors.
5. In part (d) the response provides reasonable values for the median salaries or refers to "median starting salaries" when describing a characteristic of the graph.

Partially correct (P) if the response includes three or four of the five components.

Incorrect (I) if the response does not meet the criteria for E or P.

*Notes*:

- In part (c) the response can use phrases such as "negative association (correlation, relationship)," "decreasing relationship," "inversely related," or describe a negative association (for example, "the starting salaries are smaller when there is a greater number of semesters") to satisfy component 1. However, describing the relationship between only two points does not satisfy this component.
- In part (c) responses that describe the relationship incorrectly as weak or nonlinear do not satisfy component 2, even if the other characteristic is described correctly.
- Only drawing a line on the scatterplot does not satisfy component 2.
- In part (d) a response that says "chemistry has the highest median starting salary and business has the lowest median starting salary" (or the equivalent) implies that physics is in the middle, and satisfies component 4.
- In part (d) if no values are provided for the medians, the response must use the phrase "median starting salary" at least once to satisfy component 5.
- In all parts it is acceptable if a response refers to salary rather than starting salary.

**Section 3** is scored as follows:

Essentially correct (E) if the response states that there is a negative association for each of the majors *AND* the response notes the overall positive association.

Partially correct (P) if the response states that there is a negative association for each of the majors BUT does not note the overall positive association;
*OR*
if the response states that there is a negative association for one or two specific majors (for example, for business majors) *AND* the response notes the overall positive association.

Incorrect (I) if the response does not meet the criteria for E or P.

*Notes*:

- Additional ways to note the overall positive association include stating that the original report was correct, stating that majors that take longer to complete tend to have higher starting salaries, or the equivalent.

## Question 6 (continued)

- A response that does not explicitly name the majors or refer to the three majors but makes a general statement such as "However, for an academic major, there is a negative association" satisfies the requirement to state the negative association for each of the majors.
- A response that states the original report was wrong, or incorrect, or should be retracted, etc., cannot satisfy the requirement to note the overall positive association. However, a response that says the original report might be misleading (or the equivalent) but still notes the overall positive association satisfies the requirement to note the overall positive association.
- In all parts it is acceptable if a response refers to salary rather than starting salary.

**4 Complete Response**

All three sections essentially correct

**3 Substantial Response**

Two sections essentially correct and one section partially correct

*OR*

Sections 1 and 2 partially correct and section 3 essentially correct

*OR*

Sections 1 and 3 essentially correct and section 2 incorrect

*OR*

Sections 2 and 3 essentially correct and section 1 incorrect

**2 Developing Response**

Sections 1 and 2 essentially correct and section 3 incorrect

*OR*

Section 1 essentially correct and sections 2 and 3 partially correct

*OR*

Sections 1 and 3 partially correct and section 2 essentially correct

*OR*

One section essentially correct and one section partially correct

*OR*

All three sections partially correct

**1 Minimal Response**

One section essentially correct and two sections incorrect

*OR*

Two sections partially correct and one section incorrect

*OR*

One section partially correct and two sections incorrect

6A1 of 3

6A1 of 3

(a) Does the scatterplot support the newspaper report about number of semesters and starting salary? Justify your answer.

Yes, because there is a positive correlation shown btw semesters in the program and starting salary, even if the correlation is not extremely strong.

The table below shows computer output from a linear regression analysis on the data.

| Predictor | Coef | SE Coef | T | P |
|---|---|---|---|---|
| Constant | 34.018 | 4.455 | 7.64 | 0.000 |
| Semesters | 1.1594 | 0.3482 | 3.33 | 0.003 |
| S = 7.37702 | R-Sq = 33.5% | | R-Sq(adj) = 30.5% | |

(b) Identify the slope of the least-squares regression line, and interpret the slope in context.

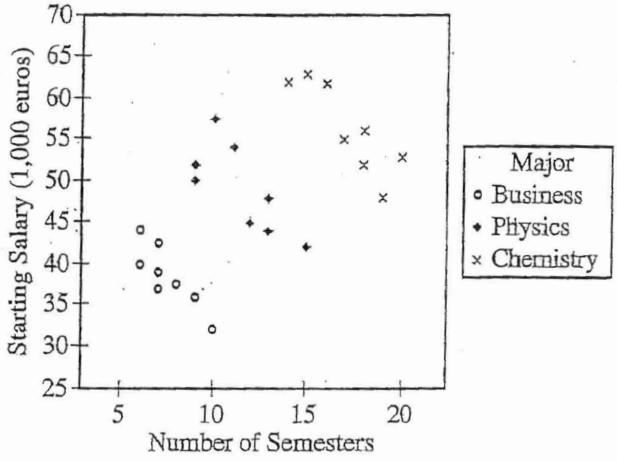
$\hat{y} = 1.1594x + 34.018$.

1.1594 is the slope of the least squares regression line. It says ~~that~~ for every additional semester starting salary increases by 1.1594 thousand Euros (on average).

6A2 of 3

6A2 of 3

An independent researcher received the data from the newspaper and conducted a new analysis by separating the data into three groups based on the major of each person. A revised scatterplot identifying the major of each person is shown below.




(c) Based on the people in the sample, describe the association between starting salary and number of semesters for the business majors.

There is a moderately strong negative correlation btw the number of semesters in a program and starting salary for buisness majors

(d) Based on the people in the sample, compare the median starting salaries for the three majors.

Chemistry has the highest median starting salary, followed by Physics and then buisness.

6A3 of 3

6A3 of 3

(e) Based on the analysis conducted by the independent researcher, how could the newspaper report be modified to give a better description of the relationship between the number of semesters and the starting salary for the people in the sample?

Majors that typically take longer to complete tend to come with higher starting salaries, however, taking longer than average than other peers in your major tends to lead to a lower starting salary.

STOP

END OF EXAM

---

THE FOLLOWING INSTRUCTIONS APPLY TO THE COVERS OF THE SECTION II BOOKLET.

- MAKE SURE YOU HAVE COMPLETED THE IDENTIFICATION INFORMATION AS REQUESTED ON THE FRONT AND BACK COVERS OF THE SECTION II BOOKLET.
- CHECK TO SEE THAT YOUR AP NUMBER LABEL APPEARS IN THE BOX ON THE COVER.
- MAKE SURE YOU HAVE USED THE SAME SET OF AP NUMBER LABELS ON ALL AP EXAMS YOU HAVE TAKEN THIS YEAR.

6B1 of 3

(a) Does the scatterplot support the newspaper report about number of semesters and starting salary? Justify your answer.

Yes because we can see a clear positive association between numby of semesters and starting salary.

The table below shows computer output from a linear regression analysis on the data.

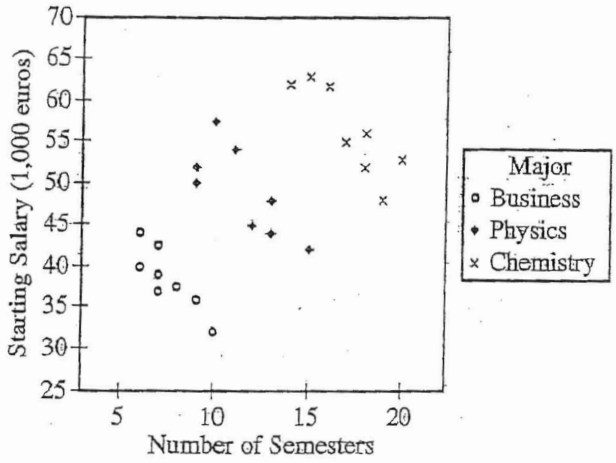
| Predictor | Coef | SE Coef | T | P |
|---|---|---|---|---|
| Constant | 34.018 | 4.455 | 7.64 | 0.000 |
| Semesters | 1.1594 | 0.3482 | 3.33 | 0.003 |
| S = 7.37702 | R-Sq = 33.5% | | R-Sq(adj) = 30.5% | |

(b) Identify the slope of the least-squares regression line, and interpret the slope in context.

The slope is 1.1594. It is predicted that as the number of semesters increases by 1, the starting salary increases by 1159,40 euros,

6B2 of 3

An independent researcher received the data from the newspaper and conducted a new analysis by separating the data into three groups based on the major of each person. A revised scatterplot identifying the major of each person is shown below.




(c) Based on the people in the sample, describe the association between starting salary and number of semesters for the business majors.

The association is negative. As the number of semesters increases, the salary decreases.

(d) Based on the people in the sample, compare the median starting salaries for the three majors.

The median salary for business majors is around 37,000 euros. The median salary for physics majors is around 48,000 euros. The median salary for chemistry majors is around 54,000 euros.



GO ON TO THE NEXT PAGE.

6B3 of 3

(e) Based on the analysis conducted by the independent researcher, how could the newspaper report be modified to give a better description of the relationship between the number of semesters and the starting salary for the people in the sample?

It would be better if the newspaper report divided the distribution into three separate distributions so that each major is accounted for separately. In the first newspaper report, the correlation seems to be positive, but after evaluating the effect of each explanatory variable on the response variable, in the second study we see that the associations are actually negative.

STOP

END OF EXAM

---

THE FOLLOWING INSTRUCTIONS APPLY TO THE COVERS OF THE SECTION II BOOKLET.

- MAKE SURE YOU HAVE COMPLETED THE IDENTIFICATION INFORMATION AS REQUESTED ON THE FRONT AND BACK COVERS OF THE SECTION II BOOKLET.
- CHECK TO SEE THAT YOUR AP NUMBER LABEL APPEARS IN THE BOX ON THE COVER.
- MAKE SURE YOU HAVE USED THE SAME SET OF AP NUMBER LABELS ON ALL AP EXAMS YOU HAVE TAKEN THIS YEAR.

6C1 of 3

(a) Does the scatterplot support the newspaper report about number of semesters and starting salary? Justify your answer.

yes because the general direction of the spread on the scatterplot is positive, supporting the positive, linear correlation between number of semesters needed and starting salary. The newspaper states that as the number of semesters needed increases, the starting salary inc. which is not true of every point on the scatterplot, but is true of the overall positive trend.

The table below shows computer output from a linear regression analysis on the data.

| Predictor | Coef | SE Coef | T | P |
|---|---|---|---|---|
| Constant | 34.018 | 4.455 | 7.64 | 0.000 |
| Semesters | 1.1594 | 0.3482 | 3.33 | 0.003 |
| S = 7.37702 | R-Sq = 33.5% | | R-Sq(adj) = 30.5% | |

(b) Identify the slope of the least-squares regression line, and interpret the slope in context.

$\hat{y} = ax + b$

(a) = slope = 1.1594

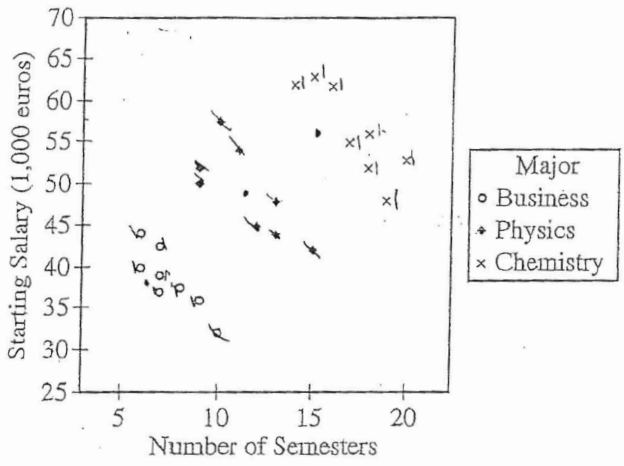
This slope means that for every 1 semester increase, the starting salary increases by 1.1594 thousands of euros.



GO ON TO THE NEXT PAGE.

6C2 of 3

An independent researcher received the data from the newspaper and conducted a new analysis by separating the data into three groups based on the major of each person. A revised scatterplot identifying the major of each person is shown below.




(c) Based on the people in the sample, describe the association between starting salary and number of semesters for the business majors.

There is a strong, negative, linear relationship between starting salary and number of semesters with the business majors. As the number of semesters increases there is a clear decrease in the starting salaries of the business majors.

(d) Based on the people in the sample, compare the median starting salaries for the three majors.

The median starting salary for business majors is the lowest of the three at about 38,000 euros. The Physics majors have the next highest median starting salary at about 47,000 euros. The chemistry majors have the highest median starting salary of about 55,000 euros.



GO ON TO THE NEXT PAGE.

6C3 of 3

(e) Based on the analysis conducted by the independent researcher, how could the newspaper report be modified to give a better description of the relationship between the number of semesters and the starting salary for the people in the sample?

The newspaper could release a new article that retracts the statements of the old article and shows the new researchers findings. They could state that stratifying the data based off of major would yield more accurate findings. The newspaper report could show 3 seperate scatterplots with each major type and show the negative inverse relationship between number of semesters and starting salary, as well as the median starting salary for each job.

STOP

END OF EXAM

---

THE FOLLOWING INSTRUCTIONS APPLY TO THE COVERS OF THE SECTION II BOOKLET.

- MAKE SURE YOU HAVE COMPLETED THE IDENTIFICATION INFORMATION AS REQUESTED ON THE FRONT AND BACK COVERS OF THE SECTION II BOOKLET.
- CHECK TO SEE THAT YOUR AP NUMBER LABEL APPEARS IN THE BOX ON THE COVER.
- MAKE SURE YOU HAVE USED THE SAME SET OF AP NUMBER LABELS ON ALL AP EXAMS YOU HAVE TAKEN THIS YEAR.

# AP® STATISTICS
# 2016 SCORING COMMENTARY

## Question 6

### Overview

The primary goals of this question were to assess a student's ability to (1) use a scatterplot to comment on a report about the relationship between two variables and interpret the slope for the least-squares regression line summarizing this relationship; (2) describe the relationship between two variables in a scatterplot when a categorical variable is introduced and compare a characteristic of the distribution of a variable for different categories of individuals in a scatterplot; and (3) describe how the associations between two variables for each category of individuals in a scatterplot differ from the overall association in the same scatterplot.

**Sample: 6A**
**Score: 4**

In part (a) the response addresses the positive association ("positive correlation") and uses the positive association to justify that the scatterplot supports the newspaper report ("Yes"), satisfying components 1 and 2 of section 1. The comment about the correlation not being extremely strong is ignored because the intention of part (a) is to address the direction of the association. In part (b) the response identifies the numerical value of the slope in the interpretation, interprets the slope correctly as the change in starting salary for each additional semester, and includes nondeterministic language in the interpretation ("on average"), satisfying components 3, 4, and 5. Because the response satisfies all five components, section 1 was scored as essentially correct. In part (c) the response states that the association is negative, states that the association is moderately strong, and uses both variable names (number of semesters, starting salary), satisfying components 1, 2, and 3 of section 2. In part (d) the response correctly compares the three majors and refers to "median starting salary," satisfying components 4 and 5. Because the response satisfies all five components, section 2 was scored as essentially correct. In part (e) the response states that there is a negative association for each major ("taking longer than average … in your major tends to lead to a lower starting salary") and notes the overall positive association ("majors that typically take longer to complete tend to come with higher starting salaries"). Section 3 was scored as essentially correct. Because all three sections were scored as essentially correct, the response earned a score of 4.

**Sample: 6B**
**Score: 3**

In part (a) the response identifies the positive association and uses the positive association to justify that the scatterplot supports the newspaper report ("Yes"), satisfying components 1 and 2 of section 1. In part (b) the response identifies the numerical value of the slope, interprets the slope correctly as the increase in starting salary as the number of semesters increases by 1, and includes nondeterministic language in the interpretation ("it is predicted that"), satisfying components 3, 4, and 5. Because the response satisfies all five components, section 1 was scored as essentially correct. In part (c) the response states that the association is negative and uses both variable names (number of semesters, starting salary), satisfying components 1 and 3 of section 2. However, the response does not satisfy component 2 because there is no description of strength or form. In part (d) the response provides reasonable values for the median salaries, satisfying component 5. However, because the response does not actually compare the three majors, it does not satisfy component 4. Because the response satisfies only three of the five components, section 2 was scored as partially correct. In part (e) the response states that there is a negative association for each major ("in the second study, we see that the associations are actually negative") and notes the overall positive association ("in the first newspaper report, the correlation seems to be positive"). Section 3 was scored as essentially correct. Because two sections were scored as essentially correct, and one section was scored partially correct, the response earned a score of 3.

# AP® STATISTICS
# 2016 SCORING COMMENTARY

## Question 6 (continued)

**Sample: 6C**
**Score: 2**

In part (a) the response identifies the positive association and uses the positive association to justify that the scatterplot supports the newspaper report ("Yes"), satisfying components 1 and 2 of section 1. In part (b) the response identifies the numerical value of the slope and interprets the slope correctly as the increase in starting salary for every one semester increase, satisfying components 3 and 4. However, because the response does not include nondeterministic language in the slope interpretation, component 5 is not satisfied. Because the response satisfies only four of the five components, section 1 was scored as partially correct. In part (c) the response states that the association is negative, states that the association is strong and linear, and uses both variable names (number of semesters, starting salary), satisfying components 1, 2, and 3 of section 2. In part (d) the response correctly compares the three majors ("business majors is the lowest … physics majors have the next highest … chemistry majors have the highest") and provides reasonable values for the median salaries, satisfying components 4 and 5. Because the response satisfies all five components, section 2 was scored as essentially correct. In part (e) the response states that there is a negative association for each major ("show 3 seperate [*sic*] scatterplots … show the negative/inverse relationship between number of semesters and starting salary"). However, because the response says that the article should be retracted, it does not satisfy the requirement to note the overall positive association. Section 3 was scored as partially correct. Because section 2 was scored essentially correct, and sections 1 and 3 were scored partially correct, the response earned a score of 2.