

2000 AP® STATISTICS FREE-RESPONSE QUESTIONS

STATISTICS

SECTION II

Part A

Questions 1-5

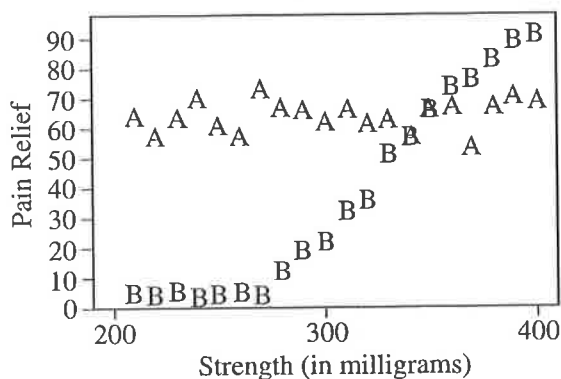
Spend about 65 minutes on this part of the exam.

Percent of Section II grade—75

Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanation.

1. Two pain relievers, A and B, are being compared for relief of postsurgical pain. Twenty different strengths (doses in milligrams) of each drug were tested. Eight hundred postsurgical patients were randomly divided into 40 different groups. Twenty groups were given drug A. Each group was given a different strength. Similarly, the other twenty groups were given different strengths of drug B. Strengths used ranged from 210 to 400 milligrams. Thirty minutes after receiving the drug, each patient was asked to describe his or her pain relief on a scale of 0 (no decrease in pain) to 100 (pain totally gone).

The strength of the drug given in milligrams and the average pain rating for each group are shown in the scatterplot below. Drug A is indicated with A's and drug B with B's.



2

2000 AP[®] STATISTICS FREE-RESPONSE QUESTIONS

(a) Based on the scatterplot, describe the effect of drug A and how it is related to strength in milligrams.

(b) Based on the scatterplot, describe the effect of drug B and how it is related to strength in milligrams.

(c) Which drug would you give and at what strength, if the goal is to get pain relief of at least 50 at the lowest possible strength? Justify your answer based on the scatterplot.

1999 STATISTICS
SECTION II

Part A

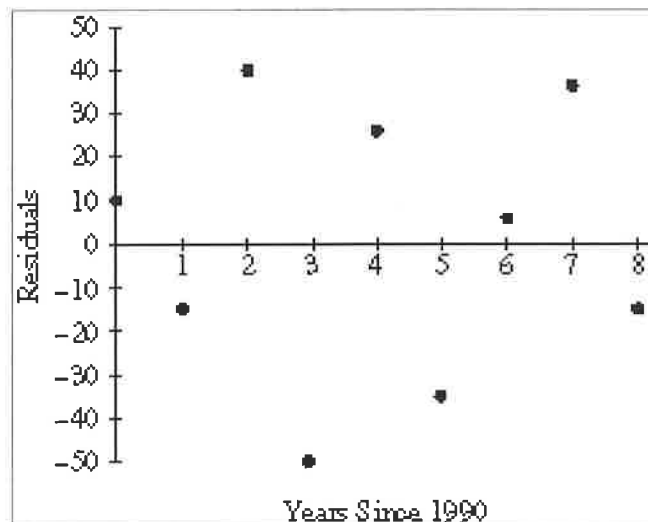
Questions 1-5

Spend about 65 minutes on this part of the exam.

Percent of Section II grade—75

Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanation.

1. Lydia and Bob were searching the Internet to find information on air travel in the United States. They found data on the number of commercial aircraft flying in the United States during the years 1990-1998. The dates were recorded as years since 1990. Thus, the year 1990 was recorded as year 0. They fit a least squares regression line to the data. The graph of the residuals and part of the computer output for their regression are given below.



Predictor	Coef	Stdev	t-ratio	p
Constant	2939.93	20.55	143.09	0.000
Years	233.517	4.316	54.11	0.000

$s = 33.43$

- Is a line an appropriate model to use for these data? What information tells you this?
- What is the value of the slope of the least squares regression line?
Interpret the slope in the context of this situation.
- What is the value of the intercept of the least squares regression line?
Interpret the intercept in the context of this situation.
- What is the predicted number of commercial aircraft flying in 1992?
- What was the actual number of commercial aircraft flying in 1992?

GO ON TO THE NEXT PAGE

2012 AP[®] STATISTICS FREE-RESPONSE QUESTIONS

STATISTICS

SECTION II

Part A

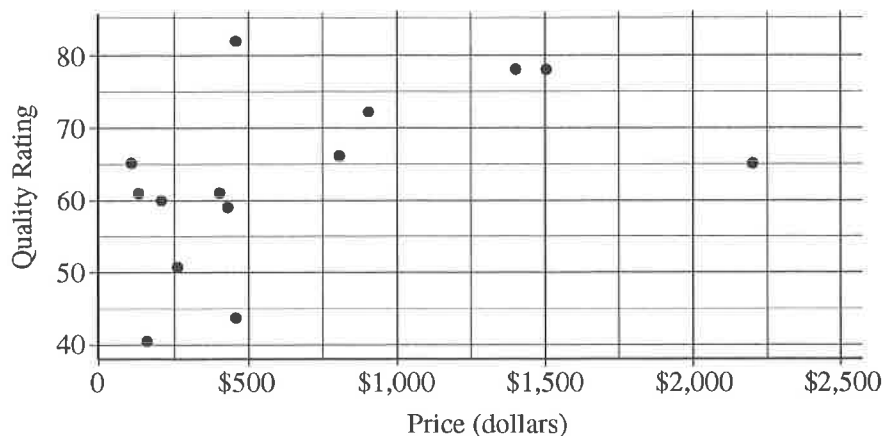
Questions 1-5

Spend about 65 minutes on this part of the exam.

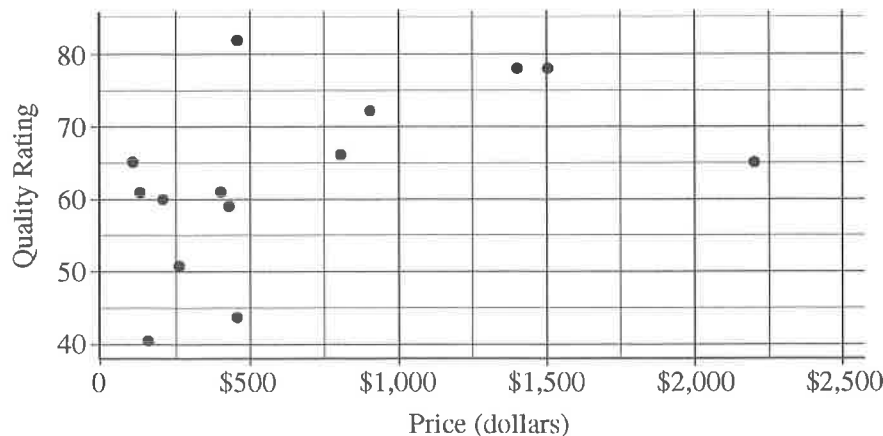
Percent of Section II score—75

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.
- (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.
- (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.



© 2012 The College Board.

Visit the College Board on the Web: www.collegeboard.org.

GO ON TO THE NEXT PAGE.

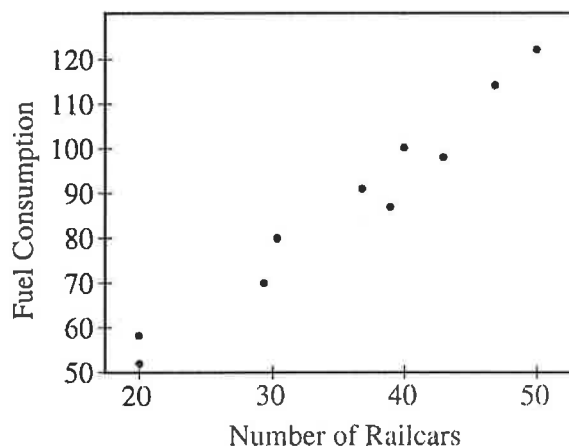
2005 AP[®] STATISTICS FREE-RESPONSE QUESTIONS

3. The Great Plains Railroad is interested in studying how fuel consumption is related to the number of railcars for its trains on a certain route between Oklahoma City and Omaha.

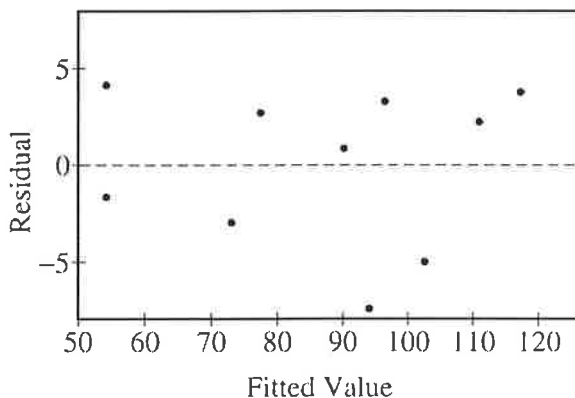
A random sample of 10 trains on this route has yielded the data in the table below.

Number of Railcars	Fuel Consumption (units/mile)
20	58
20	52
37	91
31	80
47	114
43	98
39	87
50	122
40	100
29	70

A scatterplot, a residual plot, and the output from the regression analysis for these data are shown below.



RESIDUALS VERSUS THE FITTED VALUES



2005 AP[®] STATISTICS FREE-RESPONSE QUESTIONS

6

The regression equation is
 Fuel Consumption = $10.7 + 2.15$ Railcars

Predictor	Coef	StDev	T	P
Constant	10.677	5.157	2.07	0.072
Railcar	2.1495	0.1396	15.40	0.000

$S = 4.361$ $R\text{-Sq} = 96.7\%$ $R\text{-Sq}(\text{adj}) = 96.3\%$

- Is a linear model appropriate for modeling these data? Clearly explain your reasoning.
- Suppose the fuel consumption cost is \$25 per unit. Give a point estimate (single value) for the change in the average cost of fuel per mile for each additional railcar attached to a train. Show your work.
- Interpret the value of r^2 in the context of this problem.
- Would it be reasonable to use the fitted regression equation to predict the fuel consumption for a train on this route if the train had 65 railcars? Explain.

- Some boxes of a certain brand of breakfast cereal include a voucher for a free video rental inside the box. The company that makes the cereal claims that a voucher can be found in 20 percent of the boxes. However, based on their experiences eating this cereal at home, a group of students believes that the proportion of boxes with vouchers is less than 0.2. This group of students purchased 65 boxes of the cereal to investigate the company's claim. The students found a total of 11 vouchers for free video rentals in the 65 boxes.

Suppose it is reasonable to assume that the 65 boxes purchased by the students are a random sample of all boxes of this cereal. Based on this sample, is there support for the students' belief that the proportion of boxes with vouchers is less than 0.2? Provide statistical evidence to support your answer.

- A survey will be conducted to examine the educational level of adult heads of households in the United States. Each respondent in the survey will be placed into one of the following two categories:

- Does not have a high school diploma
- Has a high school diploma

The survey will be conducted using a telephone interview. Random-digit dialing will be used to select the sample.

- For this survey, state one potential source of bias and describe how it might affect the estimate of the proportion of adult heads of households in the United States who do not have a high school diploma.
- A pilot survey indicated that about 22 percent of the population of adult heads of households do not have a high school diploma. Using this information, how many respondents should be obtained if the goal of the survey is to estimate the proportion of the population who do not have a high school diploma to within 0.03 with 95 percent confidence? Justify your answer.
- Since education is largely the responsibility of each state, the agency wants to be sure that estimates are available for each state as well as for the nation. Identify a sampling method that will achieve this additional goal and briefly describe a way to select the survey sample using this method.

2014 AP[®] STATISTICS FREE-RESPONSE QUESTIONS

STATISTICS

SECTION II

Part B

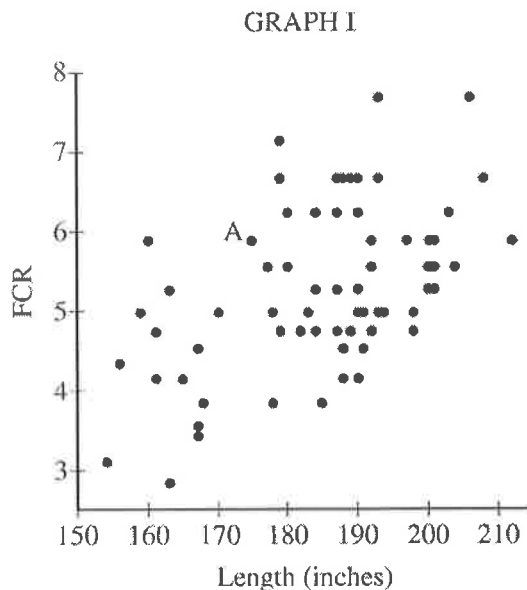
Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

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.

6. Jamal is researching the characteristics of a car that might be useful in predicting the fuel consumption rate (FCR); that is, the number of gallons of gasoline that the car requires to travel 100 miles under conditions of typical city driving. The length of a car is one explanatory variable that can be used to predict FCR. Graph I is a scatterplot showing the lengths of 66 cars plotted with the corresponding FCR. One point on the graph is labeled A.



Jamal examined the scatterplot and determined that a linear model would be a reasonable way to express the relationship between FCR and length. A computer output from a linear regression is shown below.

Linear Fit

$$\text{FCR} = -1.595789 + 0.0372614 * \text{Length}$$

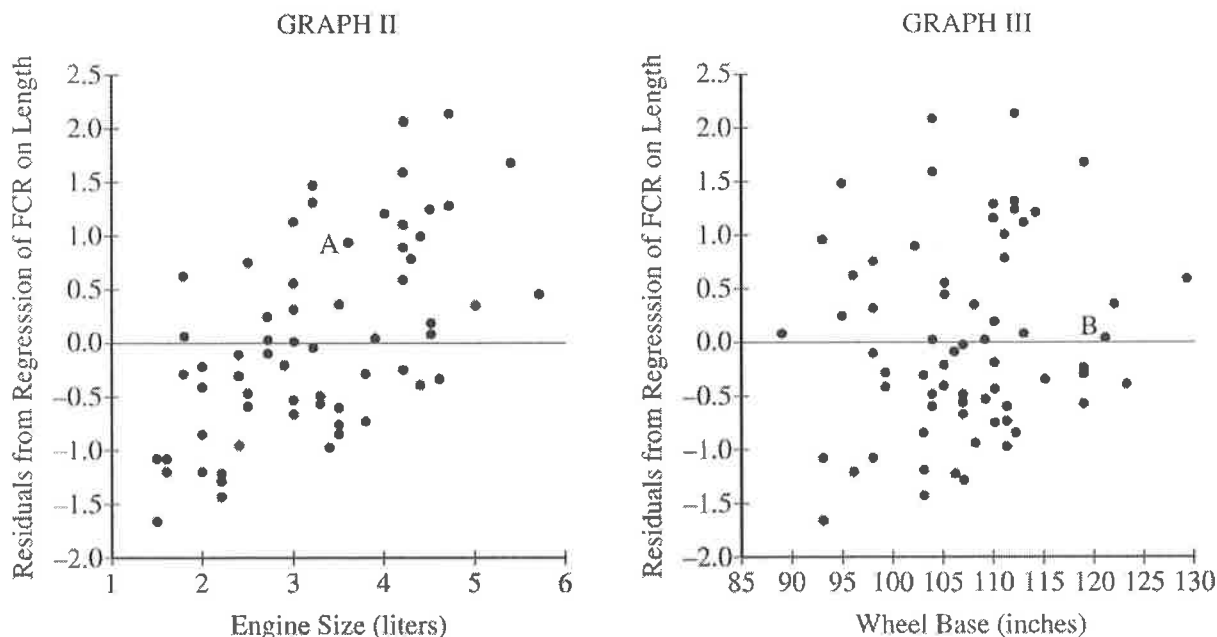
Summary of Fit

RSquare	0.250401
Root Mean Square Error	0.902382
Observations	66

- (a) The point on the graph labeled A represents one car of length 175 inches and an FCR of 5.88. Calculate and interpret the residual for the car relative to the least squares regression line.

2014 AP[®] STATISTICS FREE-RESPONSE QUESTIONS

Jamal knows that it is possible to predict a response variable using more than one explanatory variable. He wants to see if he can improve the original model of predicting FCR from length by including a second explanatory variable in addition to length. He is considering including engine size, in liters, or wheel base (the length between axles), in inches. Graph II is a scatterplot showing the engine size of the 66 cars plotted with the corresponding residuals from the regression of FCR on length. Graph III is a scatterplot showing the wheel base of the 66 cars plotted with the corresponding residuals from the regression of FCR on length.



- (b) In graph II, the point labeled A corresponds to the same car whose point was labeled A in graph I. The measurements for the car represented by point A are given below.

FCR	Length (inches)	Engine Size (liters)	Wheel Base (inches)
5.88	175	3.6	93

- (i) Circle the point on graph III that corresponds to the car represented by point A on graphs I and II.
- (ii) There is a point on graph III labeled B. It is very close to the horizontal line at 0. What does that indicate about the FCR of the car represented by point B?

2014 AP[®] STATISTICS FREE-RESPONSE QUESTIONS

- (c) Write a few sentences to compare the association between the variables in graph II with the association between the variables in graph III.
- (d) Jamal wants to predict FCR using length and one of the other variables, engine size or wheel base. Based on your response to part (c), which variable, engine size or wheel base, should Jamal use in addition to length if he wants to improve the prediction? Explain why you chose that variable.

STOP

END OF EXAM

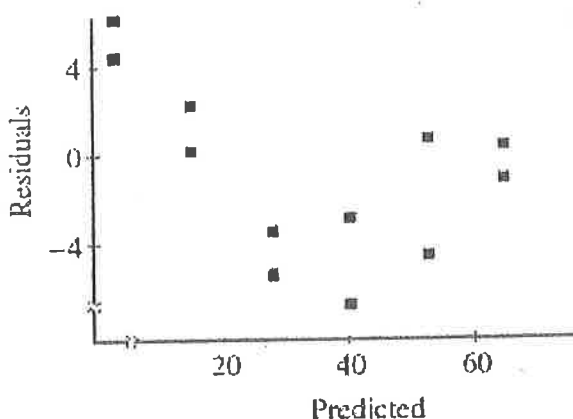
4. In a study of the application of a certain type of weed killer, 14 fields containing large numbers of weeds were treated. The weed killer was prepared at seven different strengths by adding 1, 1.5, 2, 2.5, 3, 3.5, or 4 teaspoons to a gallon of water. Two randomly selected fields were treated with each strength of weed killer. After a few days, the percentage of weeds killed on each field was measured. The computer output obtained from fitting a least squares regression line to the data is shown below. A plot of the residuals is provided as well.

Dependent variable is: percent killed

R squared = 97.2% R squared (adjusted) = 96.9%

$s = 4.505$ with $14 - 2 = 12$ degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	8330.16	1	8330.16	410
Residual	243.589	12	20.2990	
Variable	Coefficient	s.e. of Coeff	t-ratio	Prob
Constant	-20.5893	3.242	-6.35	≤ 0.0001
No. Teaspoons	24.3929	1.204	20.3	≤ 0.0001



- (a) What is the equation of the least squares regression line given by this analysis? Define any variables used in this equation.
- (b) If someone uses this equation to predict the percentage of weeds killed when 2.6 teaspoons of weed killer are used, which of the following would you expect?
- ☐ The prediction will be too large.
 - ☐ The prediction will be too small.
 - ☐ A prediction cannot be made based on the information given on the computer output.
- Explain your reasoning.

GO ON TO THE NEXT PAGE

2004 AP[®] STATISTICS FREE-RESPONSE QUESTIONS (Form B)

STATISTICS

SECTION II

Part A

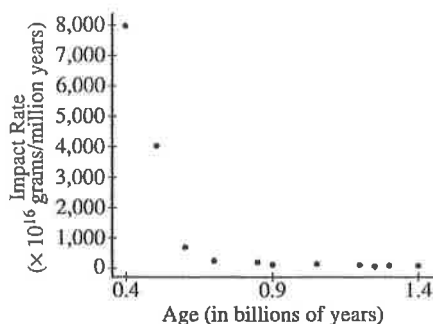
Questions 1-5

Spend about 65 minutes on this part of the exam.

Percent of Section II grade—75

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

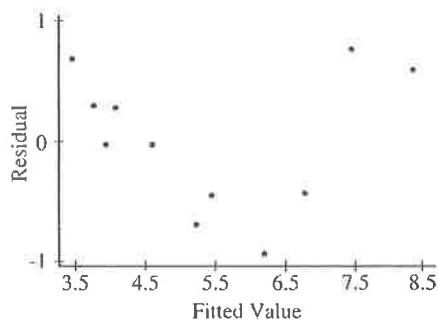
1. The Earth's Moon has many impact craters that were created when the inner solar system was subjected to heavy bombardment of small celestial bodies. Scientists studied 11 impact craters on the Moon to determine whether there was any relationship between the age of the craters (based on radioactive dating of lunar rocks) and the impact rate (as deduced from the density of the craters). The data are displayed in the scatterplot below.



- (a) Describe the nature of the relationship between impact rate and age.

Prior to fitting a linear regression model, the researchers transformed both impact rate and age by using logarithms. The following computer output and residual plot were produced.

Regression Equation: $\ln(\text{rate}) = 4.82 - 3.92 \ln(\text{age})$				
Predictor	Coef	SE Coef	T	P
Constant	4.8247	0.1931	24.98	0.000
$\ln(\text{age})$	-3.9232	0.4514	-8.69	0.000
S = 0.5977		R-Sq = 89.4%		R-Sq (adj) = 88.2%



- (b) Interpret the value of r^2 .
- (c) Comment on the appropriateness of this linear regression for modeling the relationship between the transformed variables.

1997

STATISTICS
SECTION II

Part B

Question 6

Spend about 25 minutes on this part of the exam.

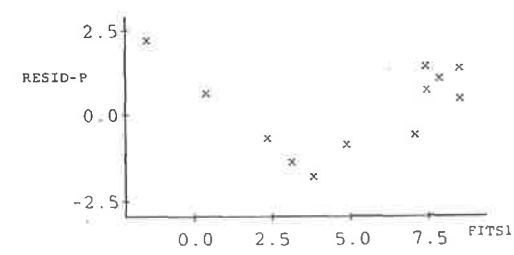
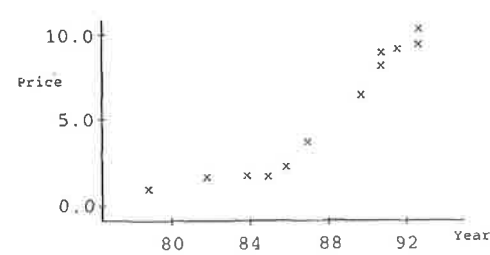
Percent of Section II grade—25

6. You are planning to sell a used 1988 automobile and want to establish an asking price that is competitive with that of other cars of the same make and model that are on the market. A review of newspaper advertisements for used cars yields the following data for 12 different cars of this make and model. You want to fit a least squares regression model to these data for use as a model in establishing the asking price for your car.

Production Year	1990	1991	1992	1987	1993	1991	1993	1985	1984	1982	1986	1979
Asking Price (in thousands of dollars)	6.0	7.7	8.8	3.4	9.8	8.4	8.9	1.5	1.6	1.4	2.0	1.0

The computer printouts for three different linear regression models are shown below. Model 1 fits the asking price as a function of the production year, Model 2 fits the natural logarithm of the asking price as a function of the production year, and Model 3 fits the square root of the asking price as a function of the production year. Each printout also includes a plot of the residuals from the linear model *versus* the fitted values, as well as additional descriptive data produced from the least squares procedure.

Model 1



The regression equation is
Price = -58.1 + 0.719 Year.

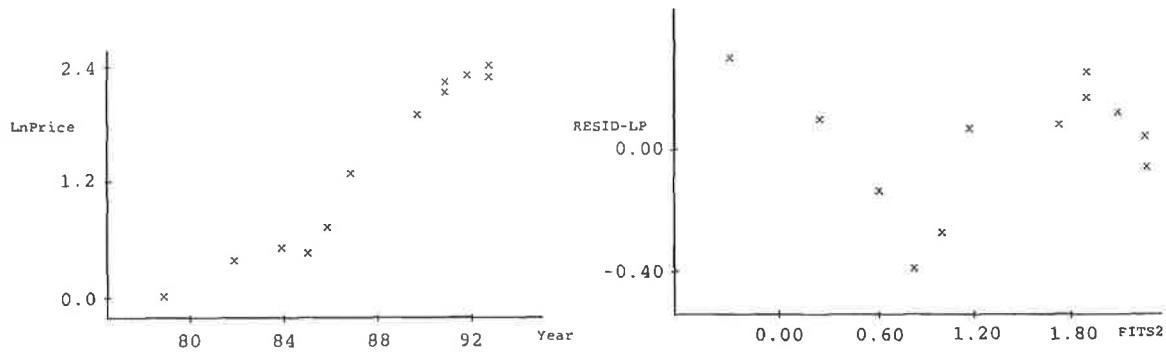
Analysis of Variance

Predictor	Coef	Stdev	t-ratio	p
Constant	-58.050	7.205	-8.06	0.000
Year	0.71900	0.08200	8.77	0.000

s = 1.255 R-sq = 88.5%

SOURCE	DF	SS	MS	F	p
Regression	1	121.10	121.10	76.88	0.000
Error	10	15.75	1.58		
Total	11	136.85			

Model 2



The regression equation is
 $\text{LnPrice} = -14.9 + 0.185 \text{ Year}$.

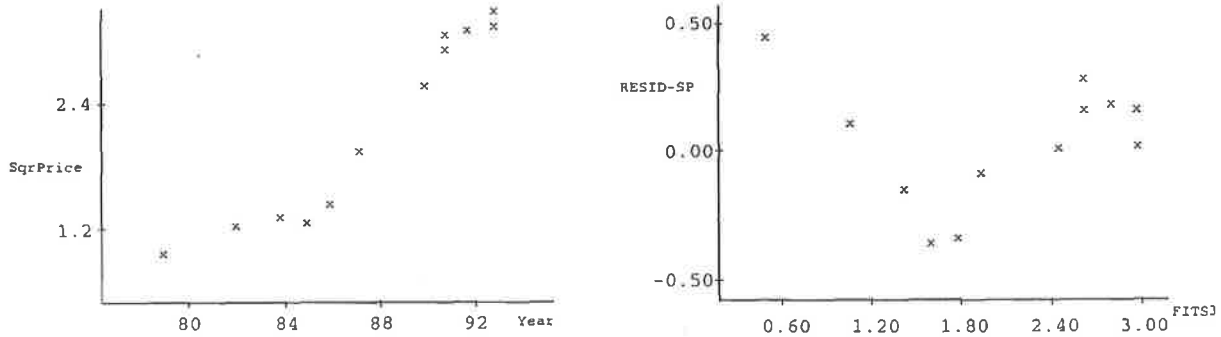
Predictor	Coef	Stdev	t-ratio	p
Constant	-14.924	1.223	-12.21	0.000
Year	0.18502	0.01392	13.30	0.000

s = 0.2130 R-sq = 94.6%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	8.0190	8.0190	176.77	0.000
Error	10	0.4536	0.0454		
Total	11	8.4726			

Model 3



The regression equation is
 $\text{SqrPrice} = -13.3 + 0.176 \text{ Year}$.

Predictor	Coef	Stdev	t-ratio	p
Constant	-13.313	1.447	-9.20	0.000
Year	0.17559	0.01647	10.66	0.000

s = 0.2520 R-sq = 91.9%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	7.2221	7.2221	113.72	0.000
Error	10	0.6351	0.0635		
Total	11	7.8572			

(a) Use Model 1 to establish an asking price for your 1988 automobile.

(b) Use Model 2 to establish an asking price for your 1988 automobile.

(c) Use Model 3 to establish an asking price for your 1988 automobile.

15

(d) Describe any shortcomings you see in these three models.

(e) Use some or all of the given data to find a better method for establishing an asking price for your 1988 automobile. Explain why your method is better.

2016 AP[®] STATISTICS FREE-RESPONSE QUESTIONS

STATISTICS

SECTION II

Part B

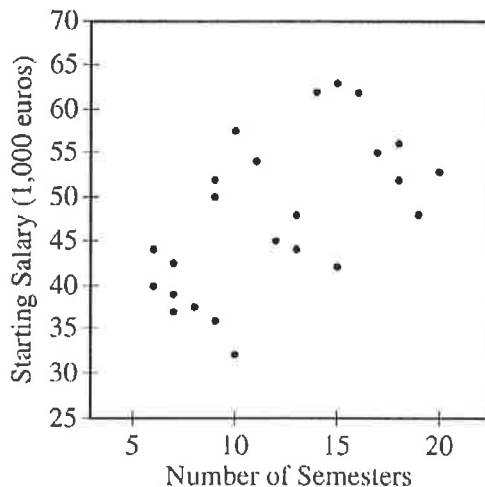
Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

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.

6. A newspaper in Germany reported that the more semesters needed to complete an academic program at the university, the greater the starting salary in the first year of a job. The report was based on a study that used a random sample of 24 people who had recently completed an academic program. Information was collected on the number of semesters each person in the sample needed to complete the program and the starting salary, in thousands of euros, for the first year of a job. The data are shown in the scatterplot below.



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

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

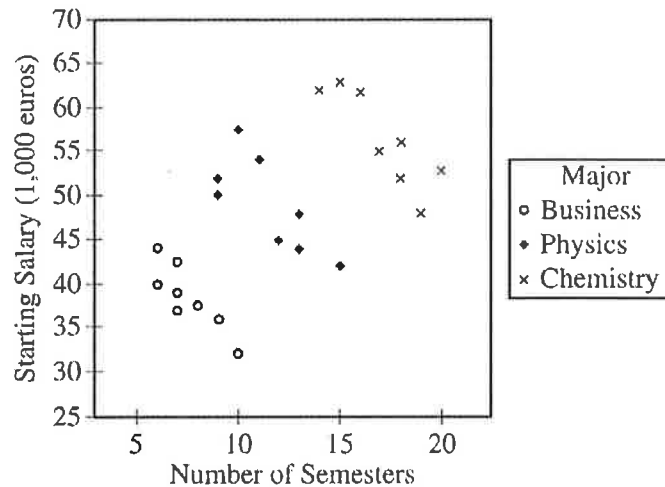
$S = 7.37702$ $R\text{-Sq} = 33.5\%$ $R\text{-Sq}(\text{adj}) = 30.5\%$

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

2016 AP[®] STATISTICS FREE-RESPONSE QUESTIONS

17

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

STOP

END OF EXAM