Mind on Statistics

Chapter 14

Sections 14.1 - 14.3

1. Which expression is a regression equation for a simple linear relationship in a population?
   A. $\hat{y} = b_0 + b_1 x$
   B. $\hat{y} = 44 + 0.60 x$
   C. $E(Y) = \beta_0 + \beta_1 x$
   D. $E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2$

KEY: C

2. One use of a regression line is
   A. to determine if any x-values are outliers.
   B. to determine if any y-values are outliers.
   C. to determine if a change in x causes a change in y.
   D. to estimate the change in y for a one-unit change in x.

KEY: D

3. The slope of the regression line tells us
   A. the average value of x when $y = 0$.
   B. the average value of y when $x = 0$.
   C. how much the average value of y changes per one unit change in x.
   D. how much the average value of x changes per one unit change in y.

KEY: C

4. A regression line is used for all of the following *except* one. Which one is not a valid use of a regression line?
   A. to *estimate* the average value of y at a specified value of x.
   B. to *predict* the value of y for an individual, given that individual's x-value.
   C. to *estimate* the change in y for a one-unit change in x.
   D. to determine if a change in x *causes* a change in y.

KEY: D

5. Which statement is *not* one of the assumptions made about a simple linear regression model for a population?
   A. The variance of Y is the same for every value of x.
   B. The distribution of Y follows a normal distribution at every value of x.
   C. The average (or mean) of Y changes linearly with x.
   D. The deviations (or residuals) follow a normal distribution with mean $\beta_0 + \beta_1 x$.

KEY: D

6. Which choice is not an appropriate description of $\hat{y}$ in a regression equation?
   A. Estimated response
   B. Predicted response
   C. Estimated average response
   D. Observed response

KEY: D

7. Which choice is not an appropriate term for the x variable in a regression equation?
   A. Independent variable
   B. Dependent variable
   C. Predictor variable
   D. Explanatory variable
KEY: B

8. What is the *best* way to determine whether or not there is a statistically significant linear relationship between two quantitative variables?
   A. Compute a regression line from a sample and see if the sample slope is 0.
   B. Compute the correlation coefficient and see if it is greater than 0.5 or less than −0.5.
   C. Conduct a test of the null hypothesis that the population slope is 0.
   D. Conduct a test of the null hypothesis that the population intercept is 0.
KEY: C

9. In a statistics class at Penn State University, a group working on a project recorded the time it took each of 20 students to drink a 12-ounce beverage and also recorded body weights for the students. Which of these statistical techniques would be the most appropriate for determining if there is a statistically significant relationship between drinking time and body weight? (Assume that the necessary conditions for the correct procedure are met.)
   A. Compute a chi-square statistic and test to see if the two variables are independent.
   B. Compute a regression line and test to see if the slope is significantly different from 0.
   C. Compute a regression line and test to see if the slope is significantly different from 1.
   D. Conduct a paired difference t-test to see if the mean difference is significantly different from 0.
KEY: B

10. To determine if there is a statistically significant relationship between two quantitative variables, one test that can be conducted is
   A. a t-test of the null hypotheses that the slope of the regression line is zero.
   B. a t-test of the null hypotheses that the intercept of the regression line is zero.
   C. a test that the correlation coefficient is less than one.
   D. a test that the correlation coefficient is greater than one.
KEY: A

11. The r^2 value is reported by a researcher to be 49%. Which of the following statements is correct?
   A. The explanatory variable explains 49% of the variability in the response variable.
   B. The explanatory variable explains 70% of the variability in the response variable.
   C. The response variable explains 49% of the variability in the explanatory variable.
   D. The response variable explains 70% of the variability in the explanatory variable.
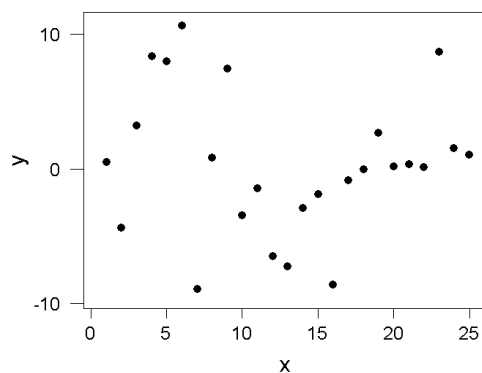KEY: A

12. Shown below is a scatterplot of *y* versus *x*.
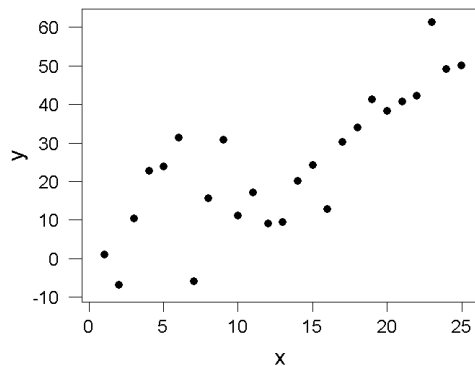
Which choice is most likely to be the approximate value of r^2, the proportion of variation in *y* explained by the linear relationship with *x*?
   A.   0%
   B.   5%
   C.   63%
   D.   95%
KEY: C

13. Shown below is a scatterplot of *y* versus *x*.

Which choice is most likely to be the approximate value of r^2, the proportion of variation in *y* explained by the linear relationship with *x*?
   A.   0%
   B.   63%
   C.   95%
   D.   99%
KEY: A

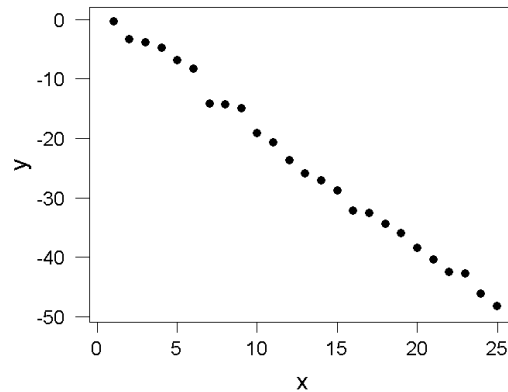14. Shown below is a scatterplot of *y* versus *x*.

Which choice is most likely to be the approximate value of r^2, the proportion of variation in *y* explained by the linear relationship with *x*?

A.   −99.5%
B.   2.0%
C.   50.0%
D.   99.5%

KEY: D

Questions 15 to 18: A regression equation is determined that describes the relationship between average January temperature (degrees Fahrenheit) and geographic latitude, based on a random sample of cities in the United States. The equation is:

$$\text{Temperature} = 110 - 2(\text{Latitude}).$$

15. Estimate the average January temperature for a city at Latitude = 45.
    A.   10 degrees
    B.   20 degrees
    C.   30 degrees
    D.   45 degrees

KEY: B

16. How does the estimated temperature change when latitude is increased by one?
    A.   It goes up 2 degrees.
    B.   It goes up 108 degrees.
    C.   It goes up 110 degrees.
    D.   It goes down 2 degrees.

KEY: D

17. Based on the equation, what can be said about the association between temperature and latitude in the sample?
    A.   There is a positive association.
    B.   There is no association.
    C.   There is a negative association.
    D.   The direction of the association can't be determined from the equation.

KEY: C

18. Suppose that the latitudes of two cities differ by 10. What is the estimated difference in the average January temperatures in the two cities?
    A.   2 degrees
    B.   10 degrees
    C.   20 degrees
    D.   90 degrees

KEY: C

Questions 19 to 22: Based on a representative sample of college men, a regression line relating y = ideal weight to x = actual weight, for men, is given by

Ideal weight = 53 + 0.7 × Actual weight

19. For a man with actual weight = 200 pounds, his ideal weight is predicted to be
    A. 153 pounds.
    B. 193 pounds.
    C. 200 pounds.
    D. 253 pounds.
KEY: B

20. If a man weighs 200 pounds but his ideal weight is 210 pounds, then his residual is
    A. 10 pounds.
    B. −10 pounds.
    C. 17 pounds.
    D. −17 pounds.
KEY: C

21. In this situation, if a man has a residual of 10 pounds it means that
    A. his predicted ideal weight is 10 pounds more than his stated ideal weight.
    B. his predicted ideal weight is 10 pounds less than his stated ideal weight.
    C. his predicted ideal weight is 10 pounds more than his actual weight.
    D. his predicted ideal weight is 10 pounds less than his actual weight.
KEY: B

22. In this context, the slope of +0.7 indicates that
    A. all men would like to weigh 0.7 pounds more than they do.
    B. on average, men would like to weigh 0.7 pounds more than they do.
    C. on average, as ideal weight increases by 1 pound actual weight increases by 0.7 pounds.
    D. on average, as actual weight increases by 1 pound ideal weight increases by 0.7 pounds.
KEY: D

Questions 23 to 29: The relation between y = ideal weight (lbs) and x =actual weight (lbs), based on data from $n = 119$ women, resulted in the regression line $\hat{y} = 44 + 0.60\,x$

23. The slope of the regression line is _____
    A. 119
    B. 44
    C. 0.60
    D. None of the above
KEY: C

24. The intercept of the regression line is _____
    A. 119
    B. 44
    C. 0.60
    D. None of the above
KEY: B

25. The estimated ideal weight for a women who weighs 118 pounds is _____
    A. 120.0 pounds.
    B. 118.0 pounds.
    C. 114.8 pounds.
    D. None of the above
KEY: C

26. What is the interpretation of the value 0.60 in the regression equation for this question?
    A. The proportion of women whose ideal weight is greater than their actual weight.
    B. The estimated increase in average actual weight for an increase on one pound in ideal weight.
    C. The estimated increase in average ideal weight for an increase of one pound in actual weight.
    D. None of the above
KEY: C

27. What is the interpretation of the value 44 in the regression for this question?
    A. The slope of the regression line.
    B. The difference between the actual and ideal weight for a woman who weighs 100 pounds.
    C. The ideal weight for a woman who weighs 0 pounds.
    D. None of the above
KEY: D

28. If a woman weighs 100 pounds and her ideal weight is just that, 100 pounds, then her residual is
    A. −4 pounds
    B. 4 pounds
    C. 44 pounds
    D. None of the above
KEY: A

29. If a woman weighs 120 pounds and her ideal weight is just that, 120 pounds, then her residual is
    A. −4 pounds
    B. 4 pounds
    C. 44 pounds
    D. None of the above
KEY: B

Questions 30 to 34: A representative sample of 190 students resulted in a regression equation between y = left hand spans (cm) and x = right hand spans (cm). The least squares regression equation is $\hat{y} = 1.46 + 0.938\,x$. The error sum of squares (SSE) was 76.67, and total sum of squares (SSTO) was 784.8.

30. What is the estimated standard deviation for the regression, s?
    A. 0.6352
    B. 0.6369
    C. 0.6386
    D. None of the above
KEY: C

31. For a student with a right hand span of 26 cm, what is the estimated left hand span?
    A. 24.39
    B. 25.85
    C. 29.60
    D. None of the above
KEY: B

32. For a student with a right and left hand span of 26 cm, what is the value of the residual?
    A. −0.152
    B. 0.152
    C. 25.848
    D. 26
KEY: B

33. Use the empirical rule to find an interval that describes the left hand spans of approximately 95% of all individuals who have a right hand span of 26 cm.
    A. (23.12, 25.66)
    B. (24.57, 27.12)
    C. (28.33, 30.87)
    D. None of the above
KEY: B

34. What is the value of r^2, the proportion of variation in left hand spans explained by the linear relationship with right hand spans.
    A. 9.8%
    B. 90.2%
    C. 95.0%
    D. None of the above
KEY: B

Questions 35 to 39: The data from a representative sample of 43 male college students was used to determine a regression equation for y = weight (lbs) and x = height (inches). The least squares regression equation was $\hat{y} = -318 + 7.00\,x$. The error sum of squares (SSE) was 23617; the total sum of squares (SSTO) = 34894.

35. What is the estimated standard deviation for the regression, s?
    A. 24.0
    B. 23.7
    C. 23.4
    D. None of the above
KEY: A

36. For a male student with a height of 70 inches, what is the estimated weight?
    A. 165
    B. 170
    C. 172
    D. None of the above
KEY: C

37. For a male student with a height of 70 inches and a weight of 200 lbs, what is the value of the residual?
    A. −28
    B. 28
    C. 172
    D. 200
KEY: B

38. Use the empirical rule to find an interval that describes the weights of approximately 95% of male college students who are 70 inches tall.
    A. (122.6, 217.4)
    B. (118.2, 211.8)
    C. (124, 220)
    D. None of the above
KEY: C

39. What is the proportion of variation in weight that is explained by the linear relationship with height?
    A. 32.3%
    B. 56.9%
    C. 67.7%
    D. None of the above
KEY: A

Questions 40 to 43: Data from a sample of 10 student is used to find a regression equation relating y = score on a 100-point exam to x = score on a 10-point quiz. The least squares regression equation is $\hat{y} = 35 + 6\,x$. The standard error of the slope is 2. The following hypotheses are tested:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

40. What is the value of the t-statistic for testing the hypotheses?
    A. -2.0
    B. 2.0
    C. 3.0
    D. 0
KEY: C

41. What is the p-value for the test? (Table A.3 or its equivalent needed.)
    A. 0.009
    B. 0.018
    C. 0.080
    D. None of the above
KEY: B

42. What is a 95% confidence interval for β_1? (Table A.2 or its equivalent needed.)
    A. (1.38, 10.62)
    B. (1.48, 10.52)
    C. (1.54, 10.46)
    D. None of the above
KEY: A

43. What is a 90% confidence interval for β_1? (Table A.2 or its equivalent needed.)
    A. (2.08, 9.92)
    B. (2.28, 9.72)
    C. (2.34, 9.66)
    D. None of the above
KEY: B

Questions 44 to 47: A representative sample of $n = 12$ male college students is used to find a regression equation for y = weight (lbs) and x = height (inches). The least squares regression equation is $\hat{y} = 30 + 2x$. The standard error of the estimated slope is 1. The following hypotheses will be tested:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

44. What is the value of the t-statistic for testing these hypotheses?
    A. 1.0
    B. 1.5
    C. 2.0
    D. 30

KEY: C

45. What is the p-value for the test? (Table A.3 or its equivalent needed.)
    A. 0.074
    B. 0.162
    C. 0.226
    D. None of the above

KEY: A

46. What is a 90% confidence interval for β_1 ? (Table A.2 or its equivalent needed.)
    A. (0.35, 6.65)
    B. (−1.32, 5.32)
    C. (0.19, 3.81)
    D. (0.00, 4.00)

KEY: C

47. What is the p-value for testing the following hypothesis about the correlation coefficient ρ ?
    $$H_0: \rho = 0$$
    $$H_a: \rho \neq 0$$
    A. 0.074
    B. 0.162
    C. 0.226
    D. None of the above

KEY: A

Questions 48 to 51: Grades for a random sample of students who have taken statistics from a certain professor over the past 20 year were used to estimate the relationship between y = grade on the final exam and x = average exam score (for the three exams given during the term).

```
The regression equation is
Final = 16.6 + 0.784 ExamAvg

Predictor        Coef       StDev           T        P
Constant       16.609       4.246        3.91    0.000
ExamAvg       0.78357     0.05593       14.01    0.000

S = 9.801       R-Sq = 52.4%     R-Sq(adj) = 52.2%

Analysis of Variance

Source             DF          SS          MS          F        P
Regression          1       18855       18855     196.29    0.000
Residual Error    178       17097          96
Total             179       35952

     Fit   StDev Fit         95.0% CI              95.0% PI
  75.377      0.731    ( 73.935,  76.818)    ( 55.982,  94.771)
```

48. The results for a test of $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$ show that
    A. the null hypothesis can be rejected because $t = 3.91$ and the p-value $= 0.000$.
    B. the null hypothesis can be rejected because $t = 14.01$ and the p-value $= 0.000$.
    C. the null hypothesis cannot be rejected because $t = 3.91$ and the p-value $= 0.000$.
    D. the null hypothesis cannot be rejected because $t = 14.01$ and the p-value $= 0.000$.
KEY: B

49. The estimate of the population standard deviation σ is given by
    A. SSE = 17097
    B. MSE = 96
    C. StDev = 4.246
    D. S = 9.801
KEY: D

50. The "Fit" information shown at the end of the output is for ExamAvg = 75. From this, we can conclude that
    A. the probability is about 0.95 that a randomly selected student with an exam average of 75 will score between 74 and 77 on the final exam.
    B. the final exam score for a randomly selected student with an exam average of 75 is likely to be between 56 and 95.
    C. about 95% of the students with an exam average of 75 will score between 74 and 77 on the final exam.
    D. the average final exam score for students with an exam average of 75 is likely to be between 56 and 95.
KEY: B

51. The two values used to determine r^2 are
    A. 17097 and 35952
    B. 96 and 17097
    C. 1 and 178
    D. 178 and 179
KEY: A

52.   For the regression line $\hat{y} = b_0 + b_1 x$, explain what the values b_0 and b_1 represent.

KEY: The term b_0 is the estimated intercept for the regression line, and b_1 is the estimated slope. The intercept is the value of $\hat{y}$ when $x = 0$. The slope b_1 tells us how much of an increase (or decrease) there is for $\hat{y}$ when x increases by one unit.

Questions 53 and 54: A regression line relating y =student's height (inches) to x = father's height (inches) for $n = 70$ college males is $\hat{y} = 15 + 0.8\ x$.

53.   What is the estimated height of a son whose father's height is 70 inches?

KEY: $\hat{y} = 71$ inches.

54.   If the son's actual height is 68 inches, what is the value of the residual?

KEY: The residual is -3.00 inches.

Questions 55 and 56: A linear regression analysis of the relationship between y = daily hours of TV watched and x = age is done using data from $n = 50$ adults. The error sum of squares is SSE = 1,000. The total sum of squares is SSTO = 5,000.

55.   What is the estimated standard deviation of the regression, s?

KEY: $s = 4.56$

56.   What is the value of r^2, the proportion of variation in daily hours of TV watching explained by the linear relationship with x = age?

KEY: 80%

Questions 57 and 58: A linear regression analysis of the relationship between y = grade point average and x = hours studied per week is done using data from $n = 10$ students. The error sum of squares is SSE = 100 and the total sum of squares is SSTO = 900.

57.   What is the value of s = estimated standard deviation for the regression?

KEY: $s = 3.54$

58.   What is the value of r^2, the proportion of variation in grade point average explained by the linear relationship with x = hours studied per week?

KEY: 88.9%

Questions 59 to 61: A regression line relating y = hours of sleep the previous day to x =hours studied the previous day is estimated using data from $n = 10$ students. The estimated slope $b_1 = -0.30$. The standard error of the slope is s.e.$(b_1) = 0.20$.

59.   What is the value of the test statistic for the following hypothesis test about β_1, the population slope?

$$H_0:\ \beta_1 = 0$$
$$H_a:\ \beta_1 \neq 0$$

KEY: $t = -1.50$

60.   What is the value of the p-value for the test in question 59?

KEY: p-value $= 0.172$

61.   What is a 90% confidence interval for β_1, the population slope?

KEY: $(-0.67, 0.07)$

Questions 62 to 64: A regression line relating y =grade point average to x = hours studied per week is estimated using data for $n = 5$ students. The estimated slope is $b_1 = 0.02$. The standard error of the slope is s.e.$(b_1) = 0.01$.

62. What is the value of the test statistic for the following hypothesis test about β_1, the population slope?
$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

KEY: $t = 2.00$

63. What is the value of the p-value for the test in question 62?
KEY: p-value $= 0.140$

64. What is a 95% confidence interval for β_1, the population slope?
KEY: $(-0.012 , 0.052)$

Questions 65 to 73: Data has been obtained on the house size (in square feet) and the selling price (in dollars) for a sample of 100 homes in your town. Your friend is saving to buy a house and she asks you to investigate the relationship between house size and selling price and to develop a model to predict the price from size.

65. Identify the response and the explanatory variable.
KEY: In this study the response variable is selling price and the explanatory variable is house size.

66. Consider the scatterplot below.
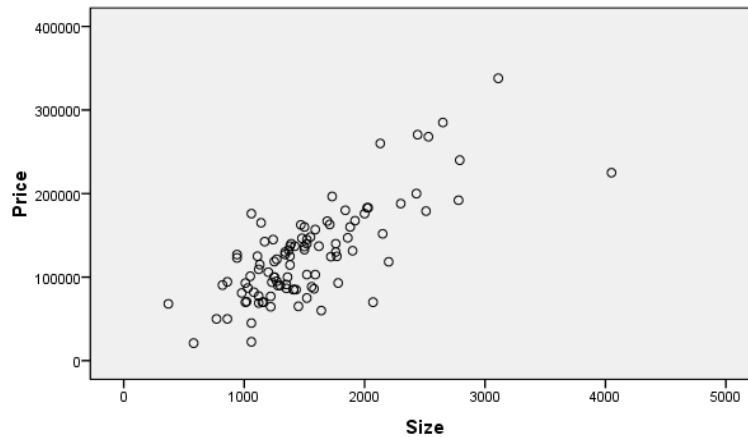
Does a linear relationship between price and size seem reasonable? If so, what appears to be the direction of the relationship?
KEY: Yes, there appears to be a positive, linear relationship.

67. Some of the regression output is provided below.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.761[a]	.580	.575	36730.206

a. Predictors: (Constant), Size

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9161.159	10759.786		.851	.397
	Size	77.008	6.626	.761	11.622	.000

a. Dependent Variable: Price

Give the value of r^2 and interpret it in the context of the problem.

KEY: An r^2 value of 0.580 indicates that about 58% of the variation in selling prices can be explained by the linear relationship between price and size.

68. Give the least square regression equation for predicting price from house size.

KEY: $\hat{y} = 9161.159 + 77 \times x$ where $\hat{y}$ is the predicted selling price (in dollars) and x is the observed size of house (in square feet).

69. What is the estimated change in the average selling price of a house for an increase in house size of 100 square feet?

KEY: 100*77 = 7700, so the selling price is expected to increase by $7700.

70. Predict the price for the 60th observed house, which had a home size of 2430 square feet.

KEY: $\hat{y} = 9161.159 + 77 \times x = 9161.159 + 77 \times 2430 = 196271.2$ or $196,271.20

71. The observed selling price for that house was $200,000. Compute the residual value.

KEY: Residual = 200,000 − 196271.2 = 3728.8 or $3,728.80.

72. Write the null and alternative hypotheses for assessing if there is a significant positive linear relationship between price and size for the population of all houses in your town.

KEY: $H_0: \beta_1 = 0$ versus $H_a: \beta_1 > 0$

73. Report the corresponding p-value and state the conclusion in context of this problem.

KEY: There is sufficient evidence to say there is a significant positive linear relationship between the price and size for the population of houses represented by this sample.

Section 14.4

74. What is the main distinction between a confidence interval and a prediction interval?
    A. A confidence interval estimates the mean value of y at a particular value of x, while a prediction interval estimates the range of y values at a particular value of x.
    B. A prediction interval estimates the mean value of y at a particular value of x, while a confidence interval estimates the range of y values at a particular value of x.
    C. A confidence interval and a prediction interval are the same thing; they both estimate the mean value of y at a particular value of x.
    D. A confidence interval and a prediction interval are the same thing; they both estimate the range of y values at a particular value of x.
KEY: A

Questions 75 to 78: A sample of 19 female bears was measured for chest girth (y) and neck girth (x), both in inches. The least squares regression equation relating the two variables is $\hat{y} = 5.32 + 1.53\, x$. The standard deviation of the regression is $s = 3.692$.

75. What is a 95% confidence interval for the average (or mean) chest girth for bears with a neck girth of 20 inches? The standard error of the fit is s.e.(fit) = 0.880
    A. (27.9, 43.9)
    B. (28.1, 43.7)
    C. (34.1, 37.8)
    D. None of the above
KEY: C

76. What is a 95% prediction interval for the chest girth of a bear with a neck girth of 20 inches? The standard error of the fit, s.e.(fit), = 0.880
    A. (27.9, 43.9)
    B. (28.1, 43.7)
    C. (34.1, 37.8)
    D. None of the above
KEY: A

77. What is a 95% confidence interval for the average (or mean) chest girth for bears with a neck girth of 18 inches? The standard error of the fit, s.e.(fit), = 0.860.
    A. (24.9, 40.9)
    B. (28.1, 43.7)
    C. (31.0, 34.7)
    D. None of the above
KEY: C

78. What is a 95% prediction interval for the chest girth of a bear with a neck girth of 18 inches? The standard error of the fit, s.e.(fit), = 0.860
    A. (24.9, 40.9)
    B. (28.1, 43.7)
    C. (31.0, 34.7)
    D. None of the above
KEY: A

Questions 79 and 80: A regression line that relates y = hand span (cm) and x = height (inches) is $\hat{y} = -1 + 0.3\,x$. The sample size was $n = 5$ adults. The standard deviation of the regression is $s = 1$ cm. For height = 60 inches, the standard error of the fit is s.e.(fit) = 1 cm.

79.   What is a 90% prediction interval for the hand span of a person whose height is 60 inches?
KEY: (13.38, 20.32)

80.   What is a 90% confidence interval for the mean hand span of people who are 60 inches tall?
KEY: (14.65, 19.35)

Questions 81 to 91: A car salesman is curious if he can predict the fuel efficiency of a car (in MPG) if he knows the fuel capacity of the car (in gallons). He collects data on a variety of makes and models of cars. The scatterplot shows that a linear model is appropriate. SPSS output is provided below.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.771	.595	.589	2.711

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	819.29	1	819.29	111.44	.000
	Residual	558.76	76	7.35		
	Total	1378.05	77			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	39.67	1.51		26.31	.000
	Fuel capacity	-.87	.08	-.771	-10.56	.000

81.   How many cars were included in the study?
KEY: 78

82.   How much would you expect fuel efficiency to change for every 1-gallon increase in fuel capacity? Include the units in your answer.
KEY: We expect a *decrease* of about 0.87 miles per gallon.

83.   What is the numerical value of the correlation coefficient between fuel efficiency and fuel capacity?
KEY: $r = -0.771$

84.   What is the equation of the least squares regression line for predicting fuel efficiency from fuel capacity?
KEY: $\hat{y} = 39.67 - 0.87\,x$

85.   Suppose Bob's car holds 18 gallons of fuel. Based on the least squares regression line, what is the predicted fuel efficiency for Bob's car?
KEY: 24.01 miles per gallon

86. Suppose Bob's car holds 18 gallons of fuel and has a fuel efficiency of 28 MPG. Based on the least squares regression line, what is the residual for Bob's car?
KEY: 3.99 miles per gallon

87. To assess if there is a significant linear relationship between fuel efficiency and fuel capacity, what are the hypotheses to be tested?
KEY: $H_0:\beta_1 = 0$ versus $H_a:\beta_1 \neq 0$

88. To assess if there is a significant linear relationship between fuel efficiency and fuel capacity, what is the observed value of the test statistic and the corresponding p-value?
KEY: $t = -10.56$, p-value $= 0.000$

89. Calculate a 98% confidence interval for the average fuel efficiency for all cars that hold 18 gallons of fuel. Descriptive statistics for the two variables are provided below. Use the mean to obtain $\bar{x}$ and the variance to calculate $\sum (x_i - \bar{x})^2$. Use these two values then to calculate s.e.(fit).

Descriptive Statistics

	Mean	Std. Deviation	Variance
Fuel capacity	17.81	3.68	13.564
Fuel efficiency	24.09	4.23	17.897

KEY: (23.278, 24.742)

90. How would the 95% confidence interval for the average fuel efficiency for all cars that hold 18 gallons of fuel compare to the interval calculated in question 89?
   A. Narrower
   B. Wider
KEY: A

91. How would the 98% prediction interval for the fuel efficiency for Bob's car that holds 18 gallons of fuel compare to the interval calculated in question 89?
   A. Narrower
   B. Wider
KEY: B

Questions 92 to 100: The heights (in inches) and foot lengths (in centimeters) of 32 college men were used to develop a model for the relationship between height and foot length. The scatterplot shows that a linear model is appropriate. SPSS output is provided below.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.758	.574	.560	1.0280

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	42.74	1	42.74	40.45	.000001
	Residual	31.70	30	1.06		
	Total	74.45	31			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.25	4.33		.06	.954
	height	.38	.06	.758	6.36	.000001

92.   How much would you expect foot length to increase for each 1-inch increase in height? Include the units.
KEY: 0.38 cm.

93.   What is the correlation between height and foot length?
KEY: 0.758

94.   Give the equation of the least squares regression line for predicting foot length from height.
KEY: $\hat{y} = 0.25 + 0.38\,x$

95.   Based on this model, predict the difference in the foot lengths of college men whose heights different by 10 inches. Include the units.
KEY: 3.8 cm.

96.   Suppose Max is 70 inches tall and has a foot length of 28.5 centimeters. Based on the least squares regression line, what is the value of the residual for Max?
KEY: 1.65

97.   To assess if there is a significant linear relationship between height and foot length, the hypotheses to be tested are $H_0:\beta_1 = 0$ versus $H_a:\beta_1 \neq 0$. What is the observed value of the test statistic and the corresponding p-value?
KEY: $t = 6.36$, p-value $= 0.000001$.

98.   To assess if there is a significant linear relationship between height and foot length, the hypotheses to be tested are $H_0:\beta_1 = 0$ versus $H_a:\beta_1 \neq 0$. What is the correct conclusion at the 1% significance level?
KEY: Since the p-value is so small, we would reject the null hypothesis and conclude that the linear relationship between height and foot length is indeed significant.

99.   Calculate a 95% confidence interval for the average foot length for all college men who are 70 inches tall. Descriptive statistics for the two variables are provided below. Use the mean to obtain $\bar{x}$ and note that
$$\sum (x - \bar{x})^2 = 289.87$$

Descriptive Statistics

	Mean	Std. Deviation	N
foot	27.8	1.5497	32
height	71.7	3.0579	32

KEY: (26.425, 27.275)

100. Max is 70 inches tall. If a 95% prediction interval for his foot length were calculated, the width of this interval would _____ with respect to the interval from question 99.
    A. increase
    B. decrease
KEY: A

Questions 101 to 104: Can an athlete's cardiovascular fitness (as measured by time to exhaustion on a treadmill) help to predict the athlete's performance in a 20k ski race? To address this question the time to exhaustion on a treadmill (minutes) and the time to complete a 20k ski race (minutes) were recorded for a sample of 11 athletes. The researcher hypothesizes that there would be an *inverse* linear relationship, that is, the longer the time to exhaustion, the faster the athlete's time in the 20k ski race, on average. The data were used to provide the following output.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.796	.634	.593	2.19

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	88.80	5.75		15.44	.0000001
	Time to exhaustion (min)	-2.33	.59	-.796	-3.95	.003

101. What is the least squares regression equation for predicting the race time based on the skier's treadmill exhaustion time?

KEY: predicted race time = 88.80 – 2.33(treadmill exhaustion time)

102. Based on this analysis, what percent of the variation in the ski race times is **not** accounted for by the linear regression between ski race time and time to exhaustion?

KEY: 1 – 0.634 = 0.366 so 36.6%

103. Predict the race time for a skier who had a treadmill exhaustion time of 10 minutes.

KEY: predicted race time = 88.80 – 2.33(10) = 65.5 minutes.

104. Below are two intervals, one is the 95% prediction interval for the ski time for the 7th skier and the other is the 95% confidence interval for the mean ski time for all skiers with a treadmill exhaustion time of 10 minutes. Which of the following is the prediction interval?

Interval 1: (63.9, 67.0)   Interval 2: (60.3, 70.7)

KEY: Interval 2 must be the prediction interval as it is the wider interval.

Section 14.5

105. Consider the following plot of residuals versus x for a regression analysis.
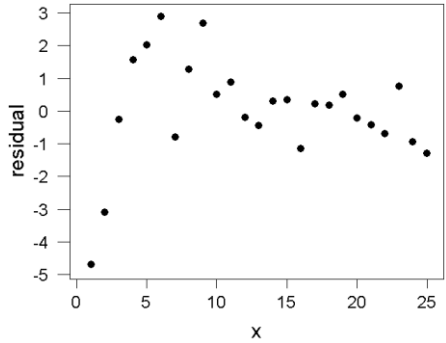
Which statement is *not* true about the regression model?
A. The sum of the residuals $= 0$.
B. The independent variable ranges from 1 to 25.
C. The residual plot shows a nonrandom pattern of residuals.
D. The residual plot shows a random pattern of residuals.

KEY: D

106. Consider the following plot of residuals versus x for a regression analysis.
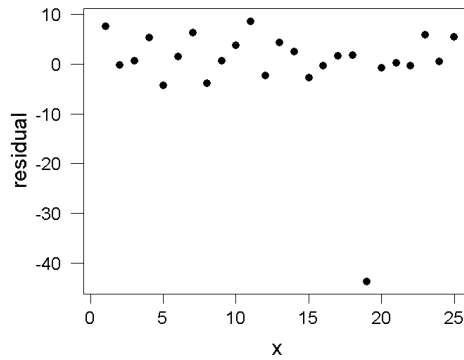
Based on the plot, what problem with the regression model or data is most noticeable?
A. The variances are not constant.
B. The mean of Y is not a linear function of X.
C. The variance of Y is not constant at each X.
D. There is an outlier in the data.

KEY: D

107. Shown below are four residual plots (residuals versus predicted values). Which one of the four does *not* show any evidence that one of the assumptions for a linear regression model is violated?

Plot A

Plot B

Plot C

Plot D
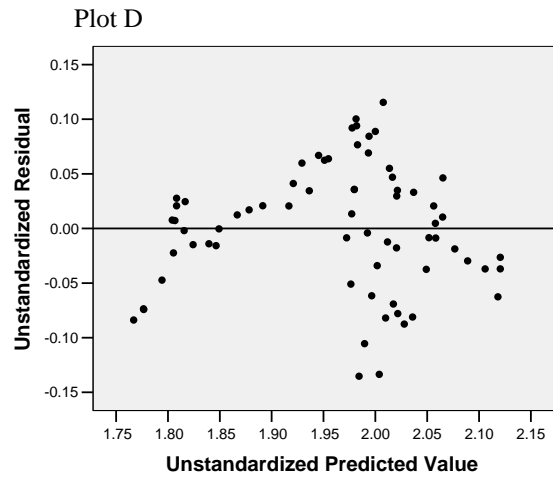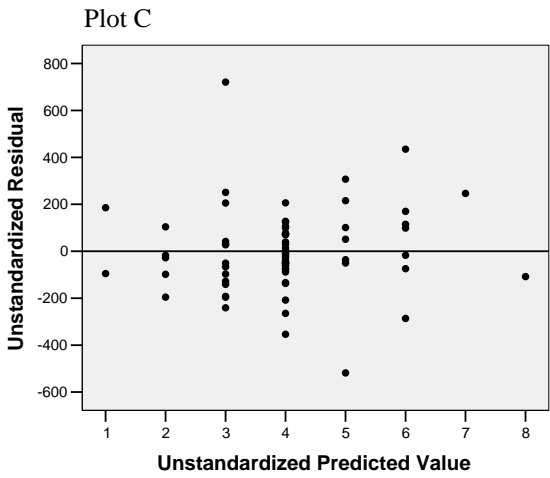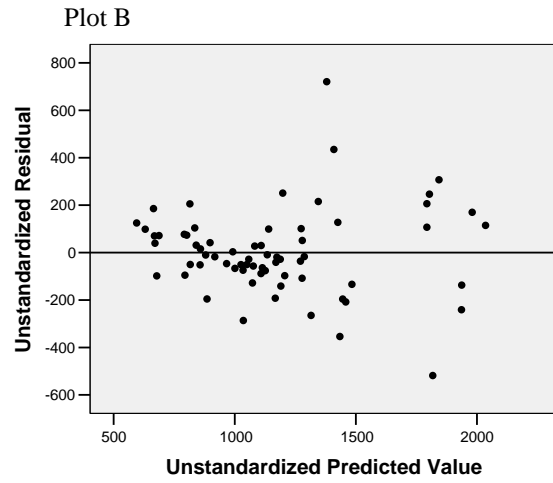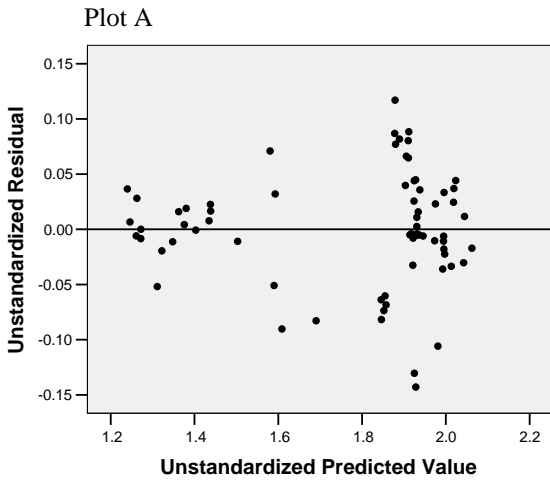
KEY: Plot C.

108. Shown below is a residual plot made as part of a regression analysis of y = ideal weight and x = actual weight for 114 women.
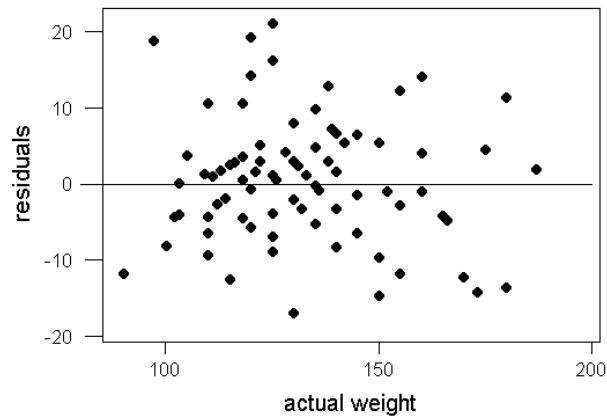
Based on the residual plot, is there any evidence that one of the assumptions for a linear regression model does not hold? Explain.

KEY: No, there is no apparent pattern nor any obvious outliers in the residual plot.

109. Consider the residual plot shown below made as part of a regression of y = fuel efficiency of a car (in MPG) and x = fuel capacity of the car (in gallons).
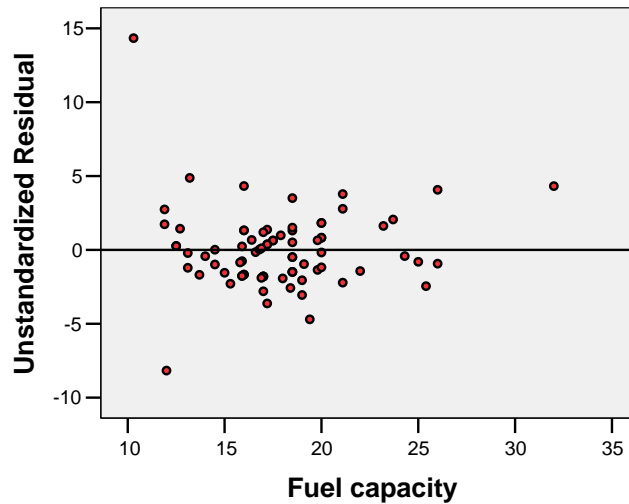
Based on the residual plot, is there any evidence that one of the assumptions for a linear regression model does not hold? Explain.

KEY: Yes. Even though there is no apparent pattern in the residual plot, there are two obvious outliers present.

110. Consider the residual plot shown below made as part of a regression of y = foot length of college men (in centimeters) and x = height of college men (in inches).
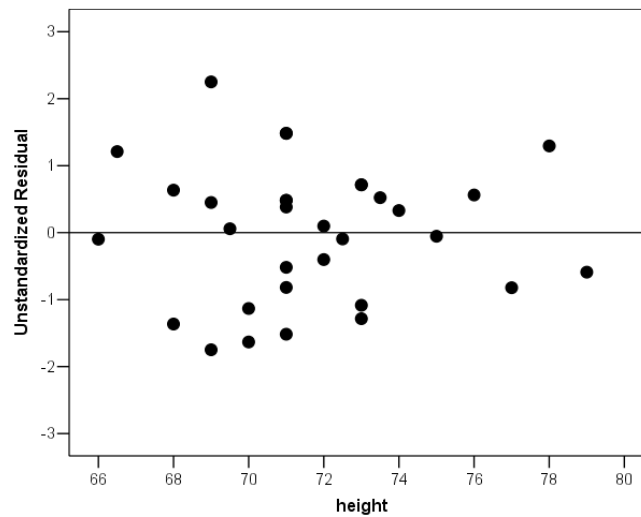
Based on the residual plot, is there any evidence that one of the assumptions for a linear regression model does not hold? Explain.

KEY: No, there is no apparent pattern nor any obvious outliers in the residual plot.

Questions 111 to 114: A consumer agency employee has been researching and testing food products for months. He is convinced that the dimensions of the packaging used for dry goods (crackers, cookies, cereal, etc) all satisfy a certain ratio. He will try to model the relationship between the minimum width of a box and the height of the box, to see if he can predict the height of the box from the minimum width (both measured in inches). Data is collected on 24 randomly selected products. The scatterplot shows that a linear model is appropriate. (Partial) SPSS output is provided below.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.389	.151	.113	3.2686

Coefficients

| Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	13.902	1.709	8.133	.0000
	x-variable	-1.157	.584	-1.982	.0601

111. What are the response and explanatory variables in this study?

KEY: Response variable: height of the box, Explanatory variable: minimum width of the box.

112. What is the equation of the least squares regression line?

KEY: $\hat{y} = 13.902 - 1.157\,x$

113. The residual plot is supplied below.
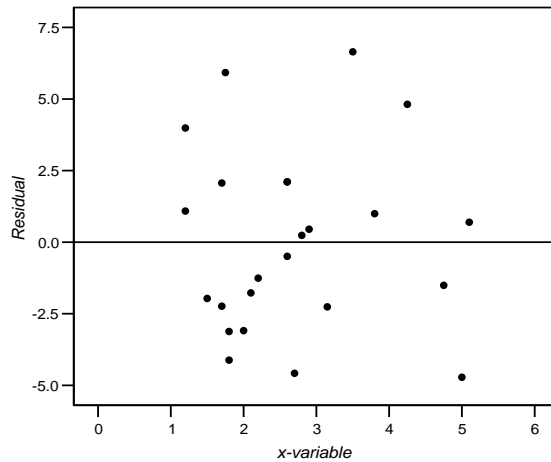
Based on the residual plot, is there any evidence that one of the assumptions for a linear regression model does not hold? Explain.

KEY: No, there is no apparent pattern nor any obvious outliers in the residual plot.

114. Is the linear relationship between minimum width and height statistically significant at the 5% significance level? Include the p-value you used to answer this question

KEY: No, the p-value for testing $H_0: \beta_1 = 0$ versus $H_a: \beta_1 \neq 0$ is 0.0601, which is not smaller than 0.05.

Questions 115 and 116: Can an athlete's cardiovascular fitness (as measured by time to exhaustion on a treadmill) help to predict the athlete's performance in a 20k ski race? To address this question the time to exhaustion on a treadmill (minutes) and the time to complete a 20k ski race (minutes) were recorded for a sample of 11 athletes. The researcher knows he is to examine the data graphically and to check various assumptions before assessing significance of the regression equation. Consider the following two plots.
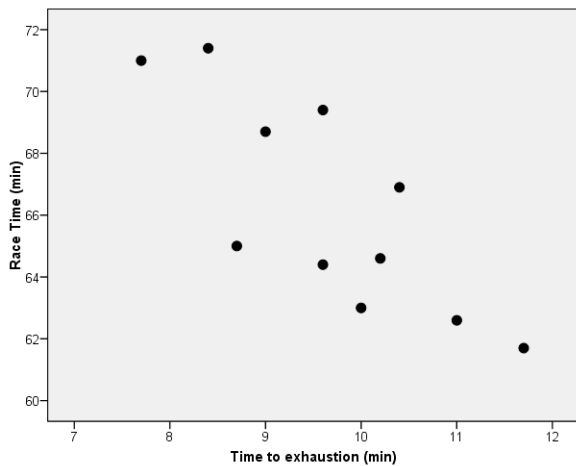
Plot A: **Plot B:**
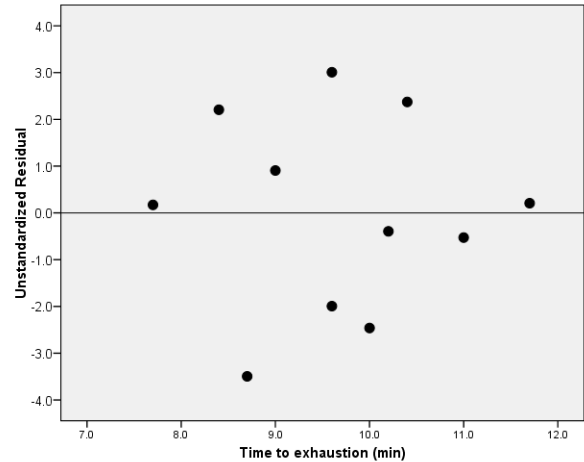
115. Give an appropriate summary regarding the form and direction shown in Plot A.

KEY: There is an approximate negative, linear association between time to exhaustion and race time.

116. Give an appropriate summary for what Plot B shows. Be sure your answer includes the assumption or data condition being assessed.

KEY: The assumption being assessed is that the true error terms have constant variance. Since in plot B the points are randomly scattered in an horizontal band around 0, this condition is reasonable.