Key

Chapter 3 Test Review

Multiple Choice Questions

1. The baggage handling services of On-Time Airlines is interested in how many baggage handlers they need on duty at various times of the day to ensure that passengers do not wait an unreasonable amount of time for their baggage. An airport executive performed a study and found that there is correlation between the number of passengers arriving at given times and the number of bag handlers needed. She sampled various times during the day and different days of the week including weekends. She recorded the number of passengers arriving within any 1-hour time block. The computer output from the regression equation analysis is shown below.

Predicted Baggage Handlers = 2.86 + 0.00408 (number of passengers)

Predictor	Coef	StDev	T	P
Constant	2.860	1.324	2.16	0.083
Passengers	0.004081	0.001168	3.49	0.017
S = 1.562	R-sq = 70.9%		R-Sq(adj) = 65.1%	

What is the value of the correlation coefficient for the number of baggage handlers and number of arriving passengers?

$\sqrt{.709} = .842$

A) -0.842
B) 0.651
C) 0.709
D) 0.842
E) 1.562

2. Some AP Statistics students were interested in finding out if there was a relationship between the number of hours of study for a chapter test and the score on the test. On the basis of the number of hours their classmates studies for the chapter 3 test and the scores on the test (out of 100%), the least-squares regression line calculated was $\hat{y} = 72.53 + 5.88x$, where x is the number of hours studied and $\hat{y}$ is the predicted score on the test. Which statement correctly interprets the meaning of the slope of this regression line?

A) For each additional hour studied, the predicted score on the test increases by 73%.
B) For each additional hour studied, the predicted score on the test increases by 6%.
C) For each additional percent of increase on the test, the predicted score on the test increases approximately 73%.
D) For each additional percent of increase on the test, the predicted score on the test increases approximately 6%.
E) We cannot use this regression equation, since cause-effect has not been proven.

3. The traffic safety officer of a local police force was trying to see if there was an association between the number of cars that did not use a main intersection in town because of the traffic light and the number of tickets written for speeding on the alternate route. The correlation between these two variables was found to be 0.58. Which of the following statements is true?

A) About 58% of the variation in the number of speeding tickets can be explained by the linear relationship between the number of speeding tickets issued and the number of cars that did not use the main intersection in town.

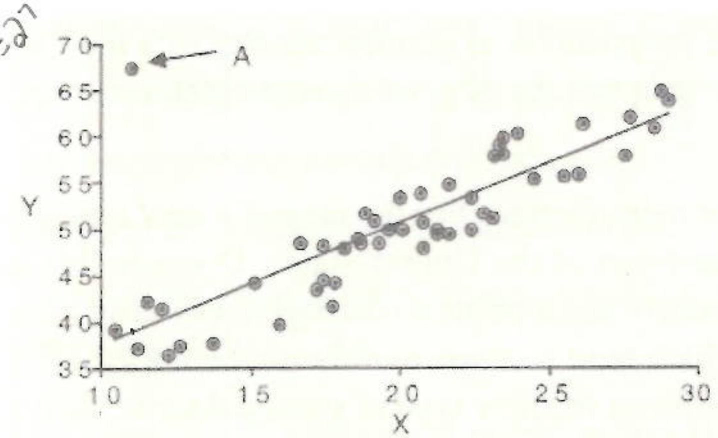
B) Any potential linear relationship between the number of cars not using the main intersection in town and the number of speeding tickets written on the alternate route would be positive.

C) If one used the main intersection through town, one is 58% more likely to receive a ticket than using the alternate route.

D) Since the correlation is not close to 1, there cannot be a linear relationship between the number of cars not using the main intersection in town and the number of speeding tickets written on an alternate route.

4. What is the approximate residual of the data point "A" on the scatterplot with the least-squares regression line shown below?

67-40=27



A) 11

B) 29

C) 39

D) 58

E) 68

5. A study was conducted of the relationship between the number of hours of television a student watched in the 24-hour period before a statistics examination and the score of the exam. The following is a computer printout from a least-squares regression analysis.

Predictor	Coeff	StDev	T	P
Constant	93.052	3.426	27.16	0.000
Hours	-3.2319	0.7819	-4.13	0.001
s = 7.843	R-Sq = 55.0%		R-Sq(adj) = 51.7%	

$\sqrt{.55} = -.742$

Which of the following gives the correct value and interpretation of the correlation coefficient for the linear relationship between the test score and the number of hours of television watched?

A) Correlation = -0.742. The linear relationship between the test score and the number of hours of TV watched is moderate and negative.
B) Correlation = 0.550. Fifty-five percent of the variation in test scores is explained by the number of hours of TV watched.
C) Correlation = 7.416. There is a relatively weak linear relationship between the test score and the number of hours of TV watched.
D) Correlation = 0.742. About 74% of the data points lie on the least-squares regression line.
E) Correlation = -0.742. For every additional hour of television watched, the average test score dropped by about three-fourths of a point.

6. The least-squares regression line has been computed to predict the yield of a certain variety of roses from the number of seeds planted. The equation is $\hat{y} = -1.05 + 0.385x$. What does the model predict the yield will be for 18 seeds planted?

$\hat{y} = -1.05 + 0.385(18)$

A) About 6
B) About 18
C) About 50
D) About 690
E) There are not enough seeds to yield any plants

7. A linear relationship exists between the amount of money spent (in thousands of dollars) on advertising and the amount (in thousands of dollars) of sales for a particular shoe manufacturer. The least-squares regression line was calculated to be $\hat{y} = 25.2 + 6.2x$, were x is the money spent on advertising and $\hat{y}$ is the amount is sales. What is the estimated increase in sales (in thousands of dollars) for every $3,000 spent on advertising?

$\hat{y} = 25.2 + 6.2(3) = 43.8$

A) 6.2
B) 18.6
C) 21.5
D) 25.2
E) 43.8

$43.8 - 25.2 = 18.6$

8. The equation of a least-squares regression line is $\hat{y} = 3.34x - 7.012$. One of the points in the scatterplot was (5, 10). Which is the residual for this x-value?

A) -10.388
B) -0.312
C) 0.312
D) 9.688
E) 10.388

$\hat{y} = 3.34(5) - 7.012 = 9.688$

$10 - 9.688 = .312$

9. The cost of production of a product at a manufacturing plant can be modeled by the equation $\hat{C} = 7200 + 585n$, where n is the number of units of the product produced in a month and $\hat{C}$ is the estimated total cost of producing the product in dollars. What is the estimated increase in cost that corresponds to an increase in production of 100 units per month?

A) \$585
B) \$7,200
C) \$7,785
D) \$58,500
E) \$65,700

$\hat{C} = 7200 + 585(100) = 65700$

$65700 - 7200 = 58500$

10. A set of residuals is created for a least-squares regression model. A negative residual indicates that the regression model

A) Only predicts negative values
B) Has overpredicted the explanatory variable
C) Has overpredicted the response variable
D) Has underpredicted the explanatory variable
E) Has underpredicted the response variable

Free Response Questions

1. The summary statistics for the number of strikeouts per year for Major League pitcher Roger Clemens are displayed below (through the 2004 season).

N	Min.	Q_1	Median	Q_3	Max	Mean	Standard Deviation
21	74	165.5	209	248.5	292	205.571	55.88

A) How would you determine if there are any outliers in the data?

I would determine any outliers by seeing if there are any values that are more than 1.5 IQR from the nearest quartile. That is, the outliers are any values less than $Q_1 - 1.5IQR$ or greater than $Q_3 + 1.5IQR$.

B) If in the next year, Roger Clemens' number of strikeouts is an outlier, how many strikeouts would be considered an outlier in the data? Show statistical methods to compute the outlier.

$IQR = 248.5 - 165.5 = 83$

$165.5 - 1.5(83) = 41$

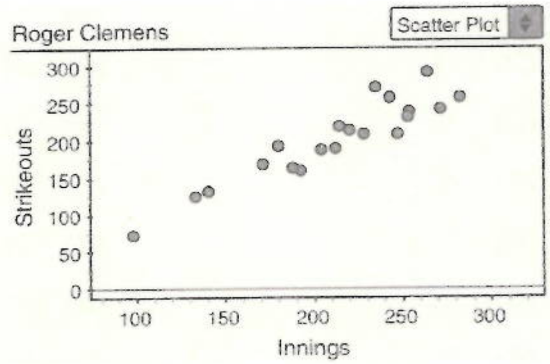
$248.5 + 1.5(83) = 343$

Outliers would be any values less than 41 or greater than 343.

C) Would a year with 240 strikeouts be considered an unusually great year? Give statistical support for your conclusion.

240 strikeouts is below the third quartile of Roger Clemen's yearly totals. This means that in at least 25% of his seasons, he had more than 240 strike-outs. There would be nothing unusually great about a year with 240 strikeouts.

Because of the wide range of strikeouts, there is interest about whether strikeout totals are related to the number of innings pitched in a season. Below is a scatterplot and least-squares regression model for strikeouts versus innings pitched.



The regression equation is

Strikeouts = − 19.5 + 1.05 Innings

Predictor	Coef	Stdev	t-ratio	p
Constant	−19.51	21.63	−0.90	0.378
Innings	1.05202	0.09863	10.67	0.000
s = 21.69	R-sq = 85.7%		R-sq(adj) = 84.9%	

D) Interpret the slope of the least-squares regression line in the context of this situation.

The slope is 1.05202 strikeouts per inning. For each additional inning pitched, Roger will strike out on average 1.05202 additional batters.

E) Interpret the value of r^2 in the context of this situation.

85.7% of the variability in strikeouts from year to year can be explained by its linear relationship with the number of innings he pitches.

F) If Clemens's three lowest strikeout years were removed, would the strength of the linear relationship increase or decrease? Justify your answer.

The strength of the relationship would decrease. These three points, having the smallest innings, are the most influential on the correlation and regression line. The rest of the points form an elliptical shape and are further away from the regression line.

2. In a study designed to investigate the relationship between mathematical reasoning of children aged 10 to 19, a single spatial reasoning test was administered to groups of 25 randomly selected children at each age level. The table below shows the mean scores of the 25 students in each group.

Group	1	2	3	4	5	6	7	8	9	10
Age	10	11	12	13	14	15	16	17	18	19
Mean Scores	65	63	70	73	81	79	81	80	83	85

Summary Statistics	Mean Age	SD Age	Mean Scores	SD Mean Scores	Correlation
	14.5	3.03	76	7.75	0.929

A) Find the equation of the least-squares regression line.

$b_1 = .929\left(\frac{7.75}{3.03}\right) = 2.376$

$b_0 = 76 - 2.376(14.5) = 41.548$

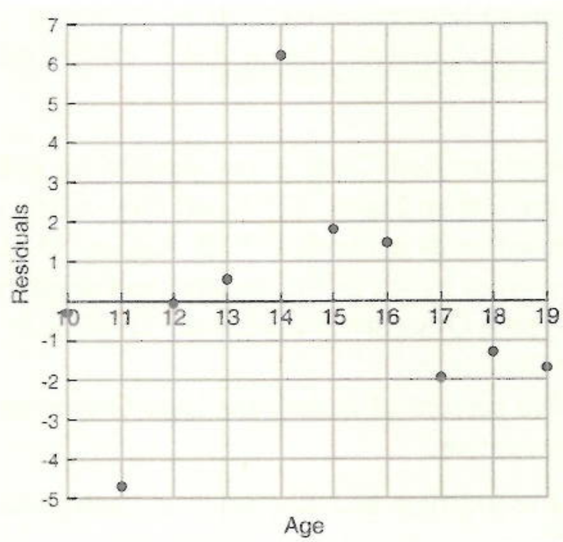
$\hat{y} = 41.548 + 2.376x$

B) Calculate and interpret the value of r^2.

$r^2 = (.929)^2 = .863$

86.3% of the variation in the mean scores can be explained by the linear relationship with the mean age.

C) The plot of the residuals is shown below. Comment on the appropriateness of the linear model.



The residuals plot shows a curved pattern. Such a pattern suggests that a linear model is not appropriate.