Ch. 7 Book Review Key

Chapter 7 Test Review

For Exercises 1–3, fill in the blank.

1. The difference between an actual data value and a predicted value is called a(n) residual error.
2. Using a model to make predictions beyond the range of the data is called extrapolation
3. The process that a calculator or computer uses to find a linear model for data for which the sum of the squares of the residuals is minimized is linear regression

For Exercises 4–6, use the following information.

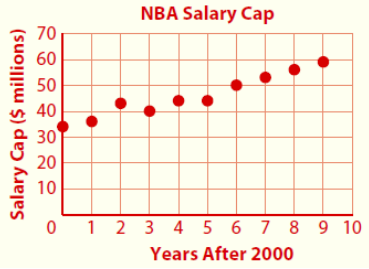
The table below shows changes in the National Basketball Association (NBA) salary cap over a 10-year period. A team whose payroll is above the salary cap must pay a "luxury tax" to the NBA.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Salary Cap ($millions)	34	36	43	40	44	44	50	53	56	59

SOURCE: NATIONAL BASKETBALL ASSOCIATION

4. Make a scatter plot of the data. (use calc: Stat → Edit $L_1 = x$, $L_2 = y$)
5. Draw a line that appears to best describe the data. Find its equation. (use calc: Stat → Calc → LinReg(ax+b))
6. Explain the meaning of the slope in this context. The salary cap increases by 2.75 million dollars per year.
7. Explain how a residual plot can be used to help you determine whether a model does a good job of describing data. A residual plot shows the errors that are not explained by the model. They should be randomly scattered above and below the x-axis (0) and should be small for the model to be considered "good."

4. Sample scatter plot:



5. Sample answer:
$S = 2.75y + 33.25$

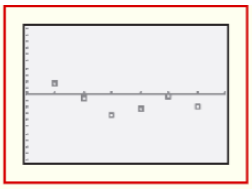
8. The table shows the total value of counterfeit goods sold on the Internet from 2003 to 2008.

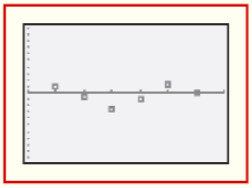
Year	2003	2004	2005	2006	2007	2008
Value of Counterfeit Sales ($billions)	45.5	62.4	78.9	98.9	119.7	137.0

SOURCE: MARKMONITOR/*THE BOSTON GLOBE*

Two people estimated linear models for forecasting the future value V of counterfeit sales. The first model is $V_1 = 19t - 13$, and the second is $V_2 = 18.5t - 11$. (Time t is measured in years after 2000.)

Make residual plots for the two models. Explain what they show about the models.

① 

② 

using your calc, type in $L_1 = 3,4,5,\ldots$ and $L_2 = 45.5, 62.4, \ldots$ Then in y_1, put $19t-13$. You will need to calculate residuals and then create scatterplot using L_1 and L_4(residuals). Repeat for $18.5t - 11$

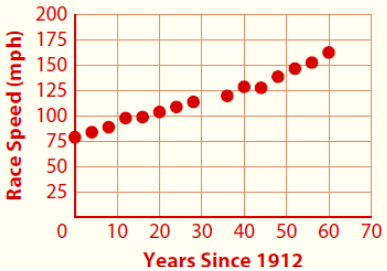
Both models have relatively small residuals; however the majority of residuals are neg in #1 and more evenly scattered pos. and neg. in #2 making model 2 the better model. They both have somewhat of a "S" pattern, meaning there may be a better model than a linear one.

For Exercises 9–13, use the following information.

The table on the next page shows the winners of the Indianapolis 500 auto race and their average speed every four years, from 1912 to 1972. (No race was run in 1944 due to World War II.)

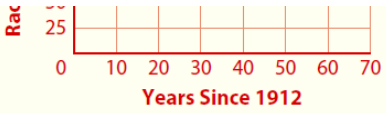
9. Make a scatter plot of the data and find the equation of a linear regression model. Let S represent speed and t the number of years since 1912 as the independent variable.
10. Explain the meaning of the slope and

9. $S = 1.298t + 78.231$



equation of a linear regression model. Let S represent speed and t the number of years since 1912 as the independent variable.

10. Explain the meaning of the slope and S-intercept.



Slope: Every year the speed increases by 1.298 mph.

S-intercept: Indicates the model predicts a speed of 78.231 mph for 1912 – very close to the actual speed for that year.

Indianapolis 500 Auto Race Results

Year	Number of Years Since 1912	Winner	Winner's Average Race Speed (mph)
1912	0	Joe Dawson	79
1916	4	Dario Resta	84
1920	8	Gaston Cheverolet	89
1924	12	L. L. Corum & J. Boyer	98
1928	16	Louis Meyer	99
1932	20	Fred Frame	104
1936	24	Louis Meyer	109
1940	28	Wilbur Shaw	114
1944	32	—	—
1948	36	Mauri Rose	120
1952	40	Troy Ruttman	129
1956	44	Pat Flaherty	128
1960	48	Jim Rathmann	139
1964	52	A. J. Foyt	147
1968	56	Bobby Unser	153
1972	60	Mark Donohue	163

SOURCE: WWW.INDIANAPOLISMOTORSPEEDWAY.COM

11. A. J. Foyt won the Indianapolis 500 in 1977 with an average speed of 161.3 mph. How well does the linear regression model predict this value?

12. For what year does the model predict that the average speed would be 200 mph?

13. Dario Franchitti won the 2010 race with an average speed of 161.6 mph. Is the model a good predictor for this year? Why or why not?

(11) 1977 is year 65 for the model. The model predicts $1.298(65)+78.231 \approx 162.6$ mph which is off by only 1.3 mph. The prediction is extrapolated, but no far beyond the data, so its fairly accurate

(12) $1.298t + 78.321 = 200$

$t \approx 94$ years

year 2006

(13) $1.298(98) + 78.321 = 205.4$

The model is <u>not</u> a good predictor for this year, b/c it involves extrapolation far beyond the data. The accuracy would not be expected.

For Exercises 14–19, use the following information.

The 11 members of a college women's golf team play a practice round, then the next day they play a round in competition on the same course. Their scores are shown in the table. (A golf score is the total number of strokes required to complete the course, so low scores are better.)

Player	1	2	3	4	5	6	7	8	9	10	11
Practice	89	90	87	95	86	81	105	83	88	91	79
Competition	94	85	89	89	81	76	89	87	91	88	80

14. Make a scatter plot of competition score vs practice score.

15. Describe the relationship between practice and competition scores. Is there a positive or negative relationship? Explain why you would expect the scores to have a relationship like the one you observe.

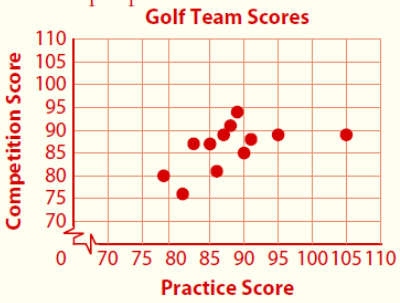
16. Find a linear model for the relationship.

17. One point on the scatter plot is clearly an outlier. A good golfer can have a bad round, or a weak golfer can have a good round. Can you tell from the given data whether the unusual point is produced by a good player or by a poor player?

18. Remove the outlier and find a new linear model for the remaining data.

19. Another golf team member shot a 95 in practice. Predict her score in competition using each of your models. Which do you think is more reliable?

14. Sample plot:



15) There is a positive relationship. Better golfers should have lower scores for both practice and competition.

16) $C = 0.410p + 50$

17) It is not possible to tell from the given info. whether the player is good or poor

18) $C = 0.754p + 20.5$

19) w/ outlier: $C = .410(95) + 50$
$C \approx 89$

w/o outlier $C = .754(95) + 20.5$
$C \approx 92$

The 92 prediction should be more reliable, assuming this golfer is typical of most of the other team members.