Stat 104 – Homework 3
Solution

Assignment:

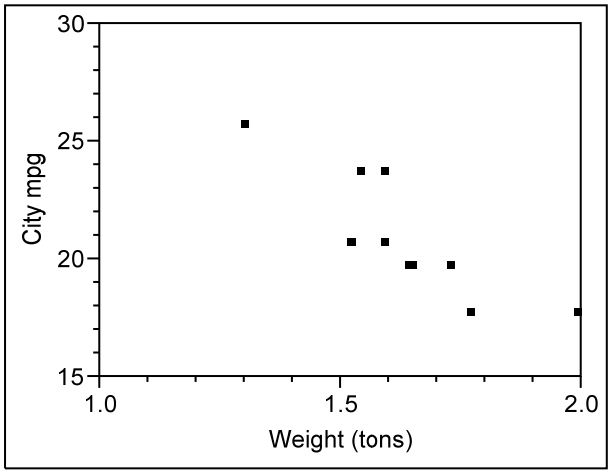
1. Complete the following problems from the text: 3.1, 3.5, 3.9, 3.11, 3.16, 3.27, 3.28, 3.33, and 3.47.

   If you have questions about the solutions to these problems, see your instructor.

2. The December 2003 issue of Kiplinger's Personal Finance published data on the 2004 model year cars and light trucks (vehicles). A random sample of 10 vehicles was taken from the data published. Below are the names of the vehicles, the EPA city mileage (mpg) ratings and the weights (tons).

Vehicle name	Ford Focus	Chevy Malibu	Ford Taurus	Kia Optima	Nissan Altima	Toyota Camry	Mitsubishi Diamante	VW Jetta	Lexus ES	Cadillac Deville
Weight, *x*	1.30	1.59	1.65	1.64	1.52	1.54	1.77	1.59	1.73	1.99
City mpg, *y*	26	24	20	20	21	24	18	21	20	18

a) Plot the data. Use Weight as the explanatory variable, *x*, and City mpg as the response, *y*.



As the weight of a vehicle increases the city mpg tends to decrease, a negative linear relationship.

b) Compute the mean and standard deviation for the Weight. Round final answers to 3 decimal places.

Mean Weight = 1.632 tons, Standard deviation of Weight = 0.180 tons

c) Compute the mean and standard deviation for the City mpg. Round final answers to 3 decimal places.

Mean City mpg = 21.2 mpg, Standard deviation of City mpg = 2.658 mpg

d) Using $\sum(x-\bar{x})(y-\bar{y}) = -3.674$, show that the correlation between Weight and City mpg is $r = -0.853$. Explain in words what this correlation means.

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{(n-1)s_x s_y} = \frac{-3.674}{9(0.180)(2.658)} = \frac{-3.674}{4.306} = -0.853$$

There is a fairly strong negative linear relationship between the weight of a vehicle and the city mpg a vehicle gets. As weight goes up, city mpg tends to go down in a linear fashion.

e) Compute the estimate of the slope for the least squares regression line. Round final answer to 2 decimal places.

$$b = r\frac{s_y}{s_x} = -0.853\left(\frac{2.658}{0.180}\right) = -12.60$$

f) Give an interpretation of the estimated slope within the context of the problem.

For every 1 ton added to the weight of a vehicle, the City mpg decreases 12.6 mpg, on average.

So a vehicle that is 1000 lbs heavier (½ ton) with get 6.3 mpg less in city driving, on average.

g) Compute the estimate of the intercept for the least squares regression line. Round final answer to 2 decimal places.

$$a = \bar{y} - b\bar{x} = 21.2 - (-12.6)(1.632) = 41.76 \text{ mpg}$$

Note: There is not a meaningful interpretation of this value within the context of the problem because a vehicle cannot weight 0 tons.

h) Give the equation of the least squares regression line. Use this equation to predict the City mpg for the Taurus that weighs 1.65 tons. Give the residual for the Taurus.

Predicted City mpg = 41.76 – 12.6 * (Weight)

Predicted City mpg = 41.76 – 12.6 * (1.65) = 20.97 mpg

Residual = y – ŷ = 20 – 20.97 = – 0.97 mpg

3. Data were collected on the selling price of homes ($100) and the living area of the home (ft^2) for a sample of 50 homes. The data are available by following the link to (all) sections of Stat 104 from www.stat.iastate.edu/courses/. Follow the instructions in the JMP Guide to download/open the data set. Use JMP to look at the distribution of selling price and the relationship between living area and selling price. Use the JMP output to help you answer the questions below. Be sure to attach the JMP output to your assignment.

a. Describe the distribution of selling price of homes. Make sure to include in your description the five number summary, the mean and standard deviation, and the shape of the histogram. Are there any outliers? If so, what are the selling prices?

Five number summary: all values are in units of $100
Minimum = 600 Q_L = 817.5 Median = 942.5 Q_U = 1194.8 Maximum = 2100

Sample mean = 1044.14 ($104,414)
Sample standard deviation = 346.66 ($34,666)

The histogram is mounded on the left and skewed to the right.

According to the box plot there are several outliers: 1844, 1900, 2100, these are the most expensive houses ($184,400, $190,000, $210,000)

b. Describe the scatter plot of selling price of homes versus living area. Give the regression equation for predicting the selling price of homes from living area. Give an interpretation of the estimated slope. Give the value of R^2 and an interpretation of this value. Finally, describe the plot of residuals versus living area and make note of any potential problems with the regression.

Above average prices are associated with above average areas and below average prices are associated with below average areas. As area increases, price tends to increase. The association has a positive direction, linear form, and is moderate to strong with the highest priced homes not necessarily the largest homes.

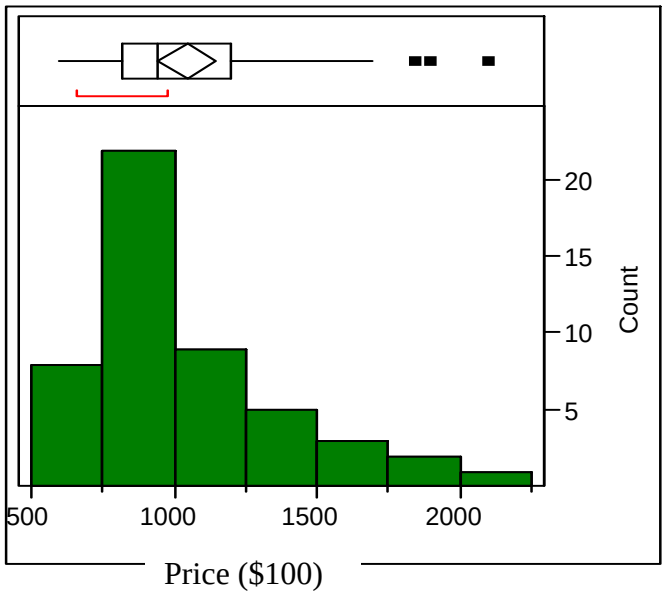
$$\textbf{predicted price} = -\textbf{50.90} + \textbf{0.6884}(\textbf{area})$$

For each additional square foot of living area, the price of the home goes up 0.6884*$100 = $68.84, on average.

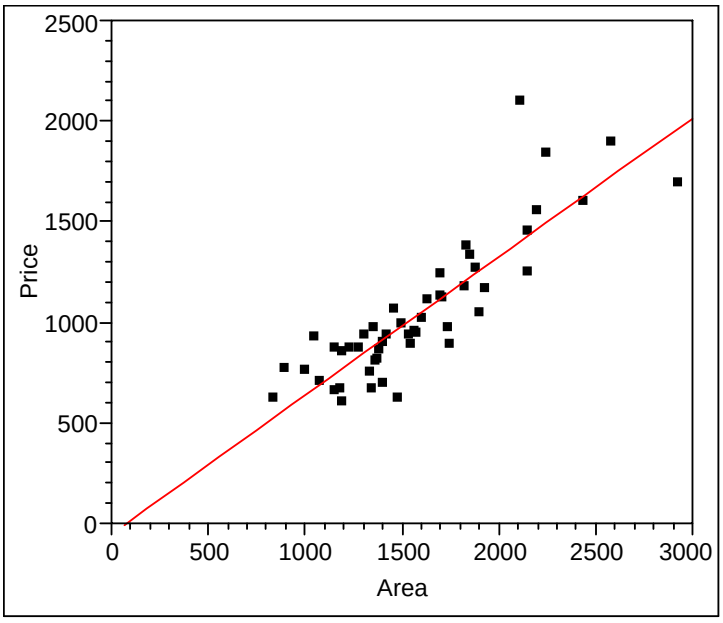
R^2 = 0.7513, 75.13% of the variation in the selling price of a home can be explained by the linear relationship with the living area.

The plot of residuals vs. living area shows basically a random scatter. There are no real problems except for a fairly large residual for the 2116 square foot home that sold for $210,000. This was the highest selling price but not the largest home.

JMP Output for Price of Homes



Five Number Summary			Sample Statistics	
100.0%	maximum	2100.0	Mean	1044.14
75.0%	quartile	1194.8	Std Dev	346.6578
50.0%	median	942.5	N	50
25.0%	quartile	817.5		
0.0%	minimum	600.0		



Linear Fit
Price = –50.89871 + 0.6884178 Area

Summary of Fit

RSquare	0.751277
RSquare Adj	0.746095
Root Mean Square Error	174.6773
Mean of Response	1044.14
Observations (or Sum Wgts)	50

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	4423826.9	4423827	144.9858
Error	48	1464583.1	30512	Prob > F
C. Total	49	5888410.0		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	–50.89871	94.23788	–0.54	0.5916
Area	0.6884178	0.057173	12.04	<.0001