

Stat 104 – Homework 3 Solution

Assignment:

- Complete the following problems from the text: 3.8, 3.16, 3.36, 3.40, and 3.60.

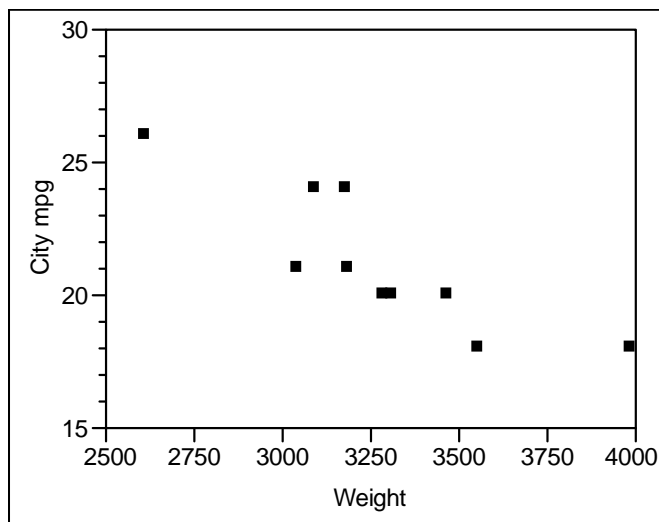
If you have questions on these problems, please see me.

- 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 (pounds).

Vehicle name	Ford Focus	Chevy Malibu	Ford Taurus	Kia Optima	Nissan Altima	Toyota Camry	Mitsubishi Diamante	VW Jetta	Lexus ES	Cadillac Deville
Weight, x	2606	3174	3306	3279	3039	3086	3549	3179	3460	3984
City mpg, y	26	24	20	20	21	24	18	21	20	18

Note: $SS(x) = 1170579.6$, $SS(y) = 63.6$, $SS(xy) = -7366.4$

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



- Compute the mean and standard deviation for the Weight. Round final answers to 4 decimal places.

Mean Weight = 3266.2 lbs, Standard deviation of Weight = 360.6444 lbs

- Compute the mean and standard deviation for the City mpg. Round final answers to 4 decimal places.

Mean City mpg = 21.2 mpg, Standard deviation of City mpg = 2.6583 mpg

- d) The correlation between the Weight and the City mpg is $r = -0.8537$. Explain in words what this correlation means.

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 6 decimal places.

$$b_1 = \frac{SS(xy)}{SS(x)} = \frac{-7366.4}{1170579.6} = -0.006293$$

or

$$b_1 = r \frac{s_y}{s_x} = -0.8537 \left(\frac{2.6583}{360.6444} \right) = -0.006293$$

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

For every 1 pound added to the weight of a vehicle, the City mpg decreases 0.006293 mpg, on average.

So a vehicle that is 1000 lbs heavier will get approximately 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 3 decimal places.

$$b_0 = \bar{y} - b_1 \bar{x} = 21.2 - (-0.006293)(3266.2) = 41.754 \text{ mpg}$$

- h) Give the equation of the least squares regression line. Use this equation to predict the City mpg for a vehicle that weighs 3306 pounds.

$$\text{Predicted City mpg} = 41.754 - 0.006293 * (\text{Weight})$$

$$\text{Predicted City mpg} = 41.754 - 0.006293 * (3306) = 20.949 \text{ mpg}$$

3. Data were collected on the selling price of homes (\$100) and the living area of the home (ft²) for a sample of 50 homes. The data are available at <http://www.public.iastate.edu/~wrstephe/stat104.html>. 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)

Sdample 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.

- 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 R² 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.

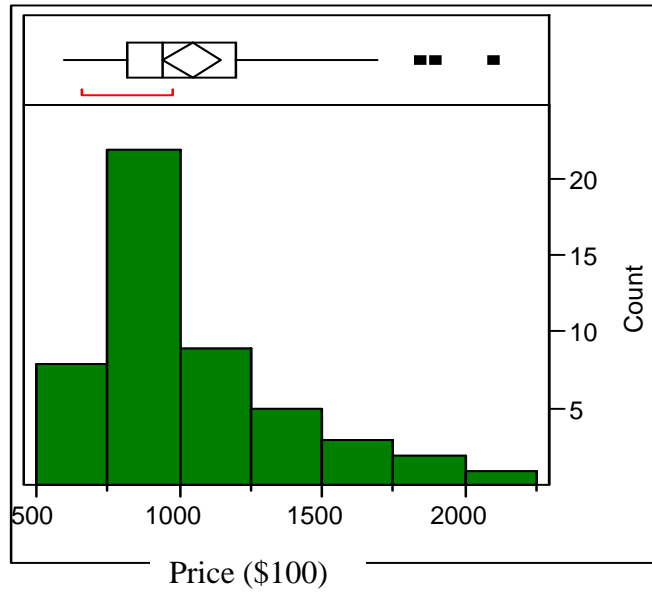
$$\text{predicted price} = -50.90 + 0.6884\text{area}$$

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

R² = 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

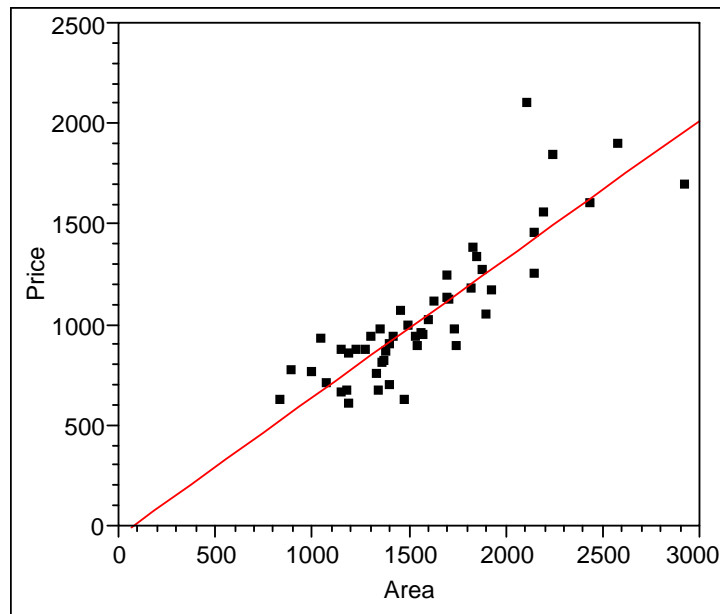


Quantiles

100.0%	maximum
75.0%	quartile
50.0%	median
25.0%	quartile
0.0%	minimum

Moments

2100.0	Mean	1044.14
942.5	Std Dev	346.6578
817.5	N	50



Linear Fit

$$\text{Price} = -50.89871 + 0.6884178 \text{ Area}$$

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

