Statistics 101 – Homework 5

Solution

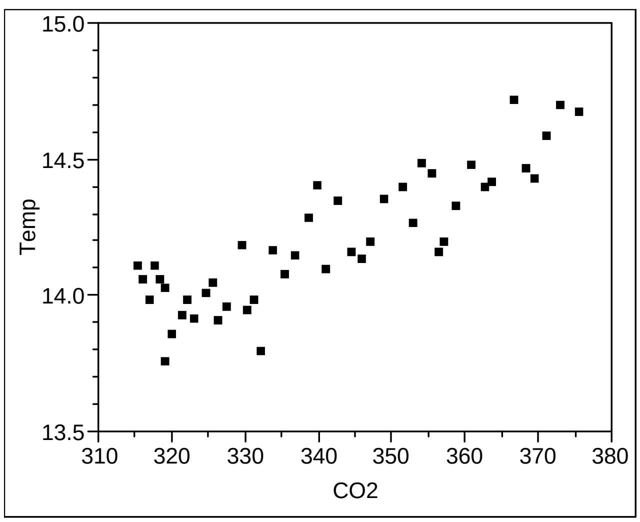
Assignment:

1. Do the following problems from the text, *Intro Stats*, 3rd Edition. If you have an earlier edition of the text, check with someone who has the 3rd Edition to make sure you do the correct problems.

   a) Chapter 8 – problems 1, 3, 5, 7, 9, 13, 14, 21, 22, and 45.
   b) Chapter 9 – problems 3, 4, 9, 11, and 12.

   If you have questions about the problems in the book, see your course instructor.

2. A problem on Homework 4 looked at the relationship between global temperatures and the amount of carbon dioxide in the atmosphere. Annual atmospheric carbon dioxide (CO2) concentrations measured as parts per million by volume (ppmv) are derived from air samples collected at Mauna Loa Observatory in Hawaii. Additionally the annual average global temperatures (degrees Celsius) are recorded. Below are a plot and summaries of the data.



CO2 (ppmv), x		Temperature (C), y		
Mean, $\bar{x}$	341.1	Mean, $\bar{y}$	14.2	Correlation
Std Dev, s_x	18.33	Std Dev, s_y	0.242	$r = +0.864$

   a) Calculate the estimate of the slope coefficient for the line of best fit.

$$b_1 = r\frac{s_y}{s_x} = +0.864\frac{0.242}{18.33} = 0.0114\ C/ppmv$$

b) Give an interpretation of the estimate of the slope coefficient within the context of the problem.

For each additional 1 ppmv of CO_2 in the atmosphere, the global temperature goes up 0.0114 C, on average.

c) Calculate the estimate of the y-intercept for the line of best fit.

$$\boldsymbol{b_0 = \bar{y} - b_1\bar{x} = 14.2 - 0.0114(341.1) = 14.2 - 3.88854 = 10.31\,C}$$

d) Why is it inappropriate to give an interpretation of the estimate of the y-intercept given the context of the problem?

It doesn't make sense to have no CO_2 in the atmosphere.

e) Give the prediction equation for predicting global temperature from CO_2 concentration.

Predicted Temperature = 10.31 + 0.0114*CO2

f) In August 2008, the CO_2 concentration was 386 ppmv. Using the prediction equation in e), predict the average global temperature.

Predicted Temperature = 10.31 + 0.0114(386) = 10.31 + 4.40 = 14.71 C

g) How does your prediction compare to the actual average global temperature of 14.54 C?

The predicted temperature is 0.17 C higher than the actual temperature.

h) According to these data, is the increase in CO_2 causing the increase in average global temperature? Explain briefly.

Not necessarily. This is an observational study and there may be other factors that have not been considered.

3. On homework 4 you used JMP to look at the relationship between the horsepower and city mpg for a sample of 57 cars selected from the 2004 model year. In this exercise we will look at using horsepower to predict the city mpg for those cars. Remember horsepower is the explanatory variable and city mpg is the response. You can find the JMP data table on the course website.

   Use Analyze – Fit Y by X – Fit Line to find the line of best fit between horsepower and city mpg. From the pull down menu next to city mpg, select Plot

Residuals. Use the JMP output to answer the following questions. Be sure to turn in a copy of the JMP output with your assignment.

a) Give the prediction equation for predicting city mpg given the horsepower of the car.

   Predicted City mpg = 34.032069 – 0.0655199*Horsepower

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

   For every additional 1 horsepower, the city mpg decreases, on average, 0.0655 mpg.

c) Predict the city mpg for a Toyota Camry with a 210 horsepower engine. What is the residual for this car?

   Predicted City mpg = 34.0 – 0.06552(210) = 34.0 – 13.76 = 20.24 mpg
   Residual = 21 – 20.24 = 0.76 mpg

d) Give the value of R^2 and an interpretation of this value within the context of the problem.

   R^2 = 0.569. 56.9% of the variation in City mpg can be explained by the linear relationship with horsepower.

e) What is the value of the correlation coefficient?

$$\boldsymbol{r = -\sqrt{0.569} = -0.754}$$

f) Give the value of $\boldsymbol{s_e}$. Give an interpretation of this value within the context of the problem.

$$\boldsymbol{s_e = RMSE = 2.376}$$

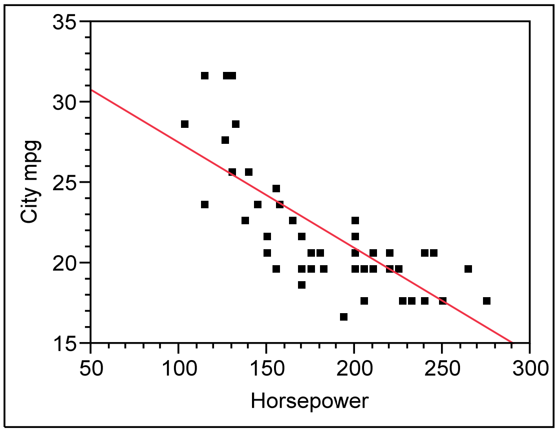
   The true City mpg values vary around the predicted values with a standard deviation of 2.376 mpg.

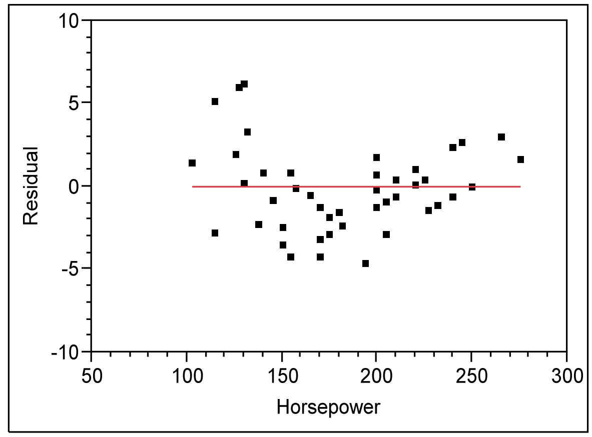
g) Describe the plot of residuals versus horsepower. What does this plot indicate about the appropriateness of the linear model relating horsepower and city mpg?

   There appears to be a slightly curved pattern. For horsepower less than 150 or more than 200 the linear model tends to under-predict. For horsepower between 150 and 200, the linear model tends to over-predict. A curved relationship may be more appropriate than the linear model.

JMP Output for the Relationship between Horsepower and City mpg

Bivariate Fit of City mpg By Horsepower





Linear Fit
City mpg = 34.032069 - 0.0655199*Horsepower

Summary of Fit

RSquare	0.568682
RSquare Adj	0.560839
Root Mean Square Error	2.376089
Mean of Response	21.96491
Observations (or Sum Wgts)	57

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	1	409.41088	409.411	72.5160
Error	55	310.51895	5.646	**Prob > F**
C. Total	56	719.92982		<.0001*

Parameter Estimates

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	34.032069	1.451587	23.44	<.0001*
Horsepower	-0.06552	0.007694	-8.52	<.0001*

Company	**Model**	**Horsepower**	**City mpg**	**Highway mpg**	**Weight**
Toyota	Camry	210	21	29	3296
Toyota	Camry	210	21	29	3362