

## STAT 516 - Spring 2009 - Homework 2

Due: Monday, February 9, 2009

1) (1 points) Concept Questions: In each case remember to correct the statement if it is false.

Page 326-327: 14, 15

2) (7 points) Page 328: 7 a-b using SAS.

- c) Check the assumptions that can be verified using the residual vs. predicted plot and the residual q-q plot.
- d) Find r-squared and briefly interpret it in terms of the problem.
- e) Find the 95% prediction interval for an average temperature of 80, and briefly interpret it.
- f) What about the way this data was gathered might make you doubt that the errors are independent, and why might a plot of the residuals versus day help you check?

The data can be found on page 330, and also on the textbook's web page as fw07p07. Be sure to present copies of the relevant portions of the SAS output and any code you used in the program editor window. (You can put the code in a really small font-size to save paper.)

3) (2 points) In the example on the course web-page for checking the assumptions, we saw that taking the logarithm of the y-variable can sometimes fix the assumption that the variances are not equal. Instead of taking only the logarithm of y, some people recommend that you take the logarithm of both y and x. In the housing price example, this would give the equation:

Model Equation				
L_price	=	3.9468	+	1.2774 L_size

where  $L\_price$  is the natural log of the price (in \$1,000s) and  $L\_size$  is the natural log of the size (in \$1,000 sq.ft). We could even add the error to the equation and get:

$$\ln(\text{price}) = 3.9468 + 1.2774 \ln(\text{size}) + \epsilon$$

Take the exponent of this equation (with the error term in it) to give the equation you would use for predicting price in terms of size, and simplify it. What about this simplified equation makes you think that using the logarithm of both sides might be helpful?

(An old math text or a search on google should turn up some rules for dealing with exponents and logarithms if you are feeling rusty!)