

## Scatterplots and Regression

Math 106, Spring 2010

The federal government maintains a large facility in Hanford WA at which various activities related to nuclear reactors and bombs are carried out. The facility occupies land adjacent to the Columbia River. Over the years there has been leaks from open pit storage of radioactive wastes into the Columbia River. Data concerning cancer downstream from Hanford in Oregon has been collected from nine counties in Oregon, and can be compared with data on radioactivity levels. Some of the data from a study conducted in 1959-64 is contained in the minitab worksheet on P-drive, called **hanford.mtw**. The index of exposure (C1) was based on several factors, including distance from Hanford, and average distance of the population from water frontage on the Columbia. Column C2 gives the Cancer Mortality per 100,000 person-years and column C3 gives the name of the county.

1. Make a scatterplot of the data. Which variable is the explanatory variable?
2. Is the association between the variables positive or negative?
3. Find the least squares regression line for predicting cancer deaths from the index of exposure. For each of the exposure indexes, compute the predicted value of cancer mortality and the associated residual.
4. Interpret the value of the slope in the least squares line. i.e., explain what this slope says about the change in cancer death rates for different exposure indexes.
5. Plot the residuals versus the index of exposure. What does the plot indicate about the adequacy of the linear fit?
6. Make another scatterplot of the data and include the least squares line on the plot.
7. Suppose you lived in a county with radioactive contamination index of exposure equal to 5. Use the least squares line to predict the cancer mortality in your home county.
8. Compute the correlation coefficient  $r$  between index of exposure and cancer mortality.
9. Create two new variables  $x^* = 10x$  and  $y^* = y/10$ . This can be done easily by using the Calculator in Minitab to create:

$$c4 = 10 * c1$$

$$c5 = c2/10$$

10. Make a scatterplot of the transformed indexes and mortality rates. Does this plot have the same appearance as the plot you constructed in part a?
11. Is the correlation coefficient for the transformed values the same as the correlation coefficient for the original values?
12. Does the slope of the least squares line of  $y^*$  on  $x^*$  have the same slope as the regression line of  $y$  on  $x$ ?