MA231 – Spring 2009 Homework 3 Due 2/27/2009
a.k.a. Homework Chebyshev

Homework Policies: You should give a brief and concise explanation for each question. Just writing down an answer with no explanation is usually not sufficient. If the homework requires output from Stata, incorporate that output into your written assignments. Homework is due at the *beginning* of class on the day indicated.

(1) Using the dataset *mortality.dta* posted on the course webpage,

a. Make a scatterplot of the relationship between under-5 mortality (mortality) and percent immunized against DPT (immune); then place the least squares regression line on the plot.
b. What is the interpretation of the estimates of the intercept and slope?
c. Estimate how much the under-5 mortality rate would increase/decrease if Bolivia would increase the percent immunized from 40% to 75%.

(2) M&M 2.72, p. 24

(3) In a survey of 988 men aged 18-24, the regression equation for predicting weight from height is given by,

Weight (in pounds) = -167 + 4.7 (height in inches)

What is the estimate of the intercept and what is its interpretation? Is this interpretation meaningful? Why or why not? What is the estimate of the slope and what is its interpretation? Based on the regression equation, what is the weight of a newborn infant boy measuring 16 inches? Comment on the appropriateness of this prediction.

(4) Suppose a law school finds the following relationship between LSAT scores and first-year grades (among students who complete their first year),

Mean LSAT score = 31, SD = 6 Mean first-year grade = 71, SD=12
Conditional SD of first-year grade = 9 points, r = 0.64

Assume the conditions of the standard regression model are valid (all 3).

a. What proportion of students had first-year grades over 80?
b. Consider the distribution of first-year grades among students who scored 35 on the LSAT, what is the mean of this distribution? What is the standard deviation?
c. Of the students who scored 35 on the LSAT, what proportion had first-year grades over 80?

(5) The U.S. Census Bureau divides the country into nine geographical areas (e.g., one such area is the Mountain Division comprising MT, ID, WY, CO, NM, AZ, UT, and NV). For each of these nine areas, suppose we compute the average income and average education for men aged 25-64 living there. Then we compute the correlation between these nine pairs of averages, which comes out to be $r = 0.83$. Can you necessarily conclude that there is a strong correlation between income and education for individual men aged 25-64 in the U.S.? Explain.

(6) M&M 2.90 (Page 139)

(7) The Stata file *oil.dta* contains information concerning both laboratory and field measurements of defects in the Trans-Alaska Oil Pipeline.

a. Perform the regression that relates the field (Y) to laboratory (X) measures and give the equation.
b. Provide plots of the regression line (superimposed in the data) as well as a plot of the residuals.
c. Do the first two regression assumptions appear to be valid? If not, which one(s) is/are violated?
d. Suggest a "fix," if possible, for the violation(s) you note above if any.

(8) Reverse the response and explanatory variables in the *oil.dta* dataset (i.e., make "field" your explanatory variable and"lab" your response).
a. Give the regression line equation for this. Is it the same as the line you obtained in 7(a)? Explain.
b. Plot the residuals and comment on their form.

(9) M&M 2.95 (Page 139-140) The data are in the file *mm2.71.dta*.

Famous Statistician of the Week

Who is this dude?
Pavnuty Lvovitch Chebyshev
1821-1894

Why is he cool?

Chebyshev made significant contributions to the theory of probability. In 1867 he published a paper *On mean values* which used Bienaymé's inequality to give a generalised law of large numbers. As a result of his work on this topic the inequality today is often known as the Bienaymé-Chebyshev inequality. Twenty years later Chebyshev published *On two theorems concerning probability* which gives the basis for applying the theory of probability to statistical data, generalising the central limit theorem of de Moivre and Laplace. Of this Kolmogorov wrote (see for example [1]):-

The principal meaning of Chebyshev's work is that through it he always aspired to estimate exactly in the form of inequalities absolutely valid under any number of tests the possible deviations from limit regularities. Further, Chebyshev was the first to estimate clearly and make use of such notions as "random quantity" and its "expectation (*mean*) *value".*

Courtesy of *http://www-gap.dcs.st-and.ac.uk/history/Mathematicians/Chebyshev.html*