**MA231 – Spring 2009 Homework 4 Due 3/2/2009**
**Solutions a.k.a. Homework Deming**

**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) A study found that the correlation between the educational level of husbands and wives in a certain town was 0.50; both averaged 12 years of schooling completed, with an SD of 3 years.

a. Predict the educational level of a man whose wife has completed 18 years of schooling.

$$\hat{y} = a + bx$$
$$b = r\frac{s_Y}{s_X} = 0.5\frac{3}{3} = 0.5$$
$$a = \overline{y} - b\overline{x} = 12 - 0.5(12) = 6$$
$$\hat{y} = 6 + 0.5x$$

So, the predicted educational level for a man whose wife has 18 years of schooling is 15 years.

b. Predict the educational level of a woman whose husband has completed 15 years of schooling.

Note that we're changing the roles of the explanatory and response variables. Since the mean and SD is the same for both men and women, the regression line does not change.

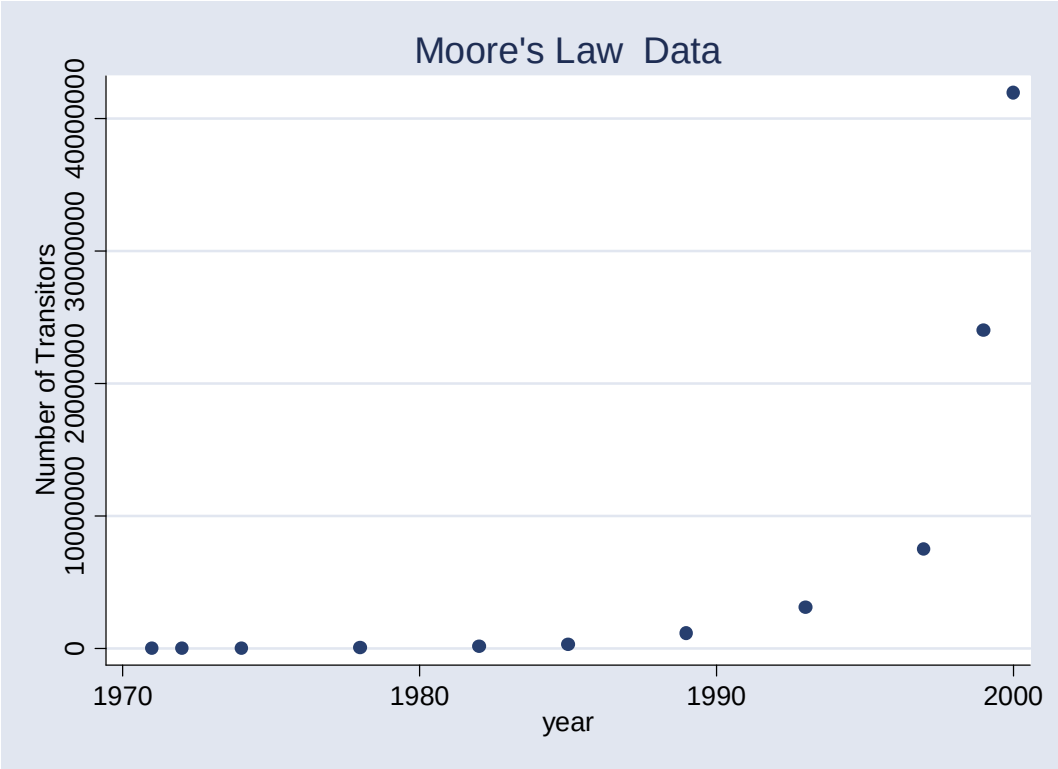
This gives a predicted value of the woman's education to be 13.5 years for a man with 15 years of education.

c. Apparently, well-educated women marry men who are less well educated than themselves, but men marry women with even less education? This seems paradoxical. How can this be possible?

Notice that the woman described in part (a) is 3 SD's above the mean. This is quite extreme (she's in the 97.5$^{th}$ percentile), and thus is less likely to marry someone as extreme as she is. The man that has an educational level of 15 years is less extreme (only 1 SD above the mean), and thus is expected to marry someone above average and not too dissimilar from himself. This is a classic example of regression to the mean.

(2) In 1964, Gordon Moore, one of the founders of Intel Corporation, predicted that the number of transistors on an integrated circuit chip would double every 18 months; this become known as Moore's Law. The Stata dataset *moore.dta*, contains the dates (*year*) and number of transistors (*transistors*) for Intel microprocessors from 1971-2000. The *Stata* dataset is on the course webpage.

a. Use Stata to produce a scatter plot showing how the number of transistors changes with time. Does the relationship appear to be linear?



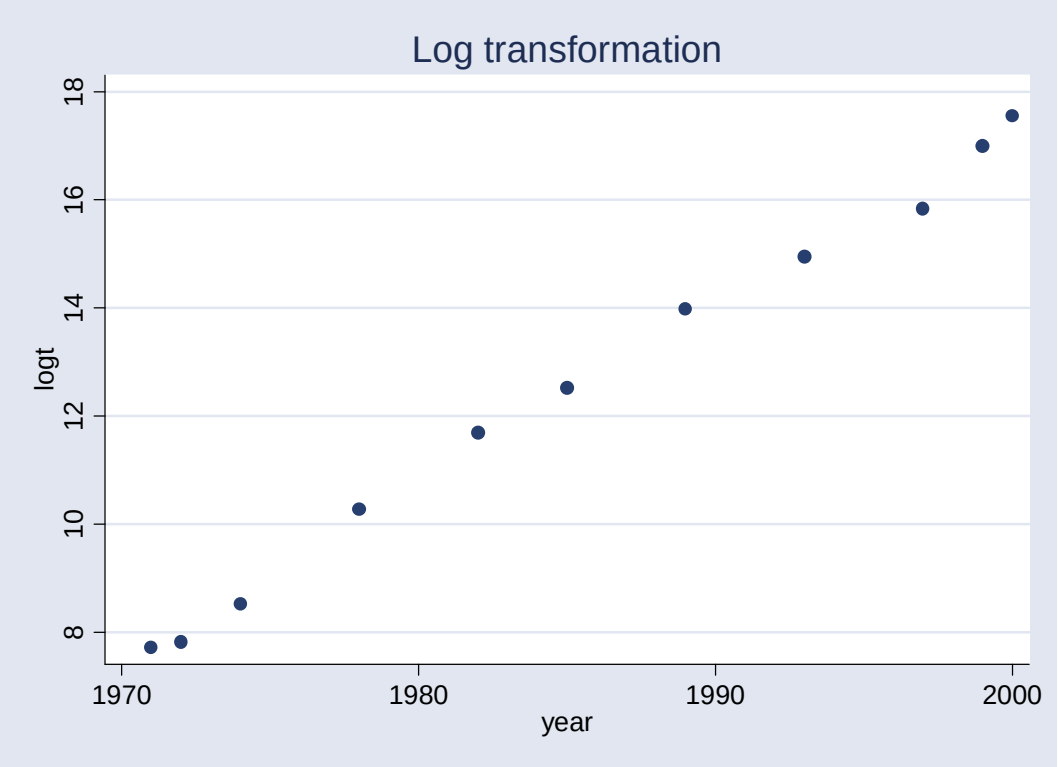

This data does not appear to have a linear relationship.

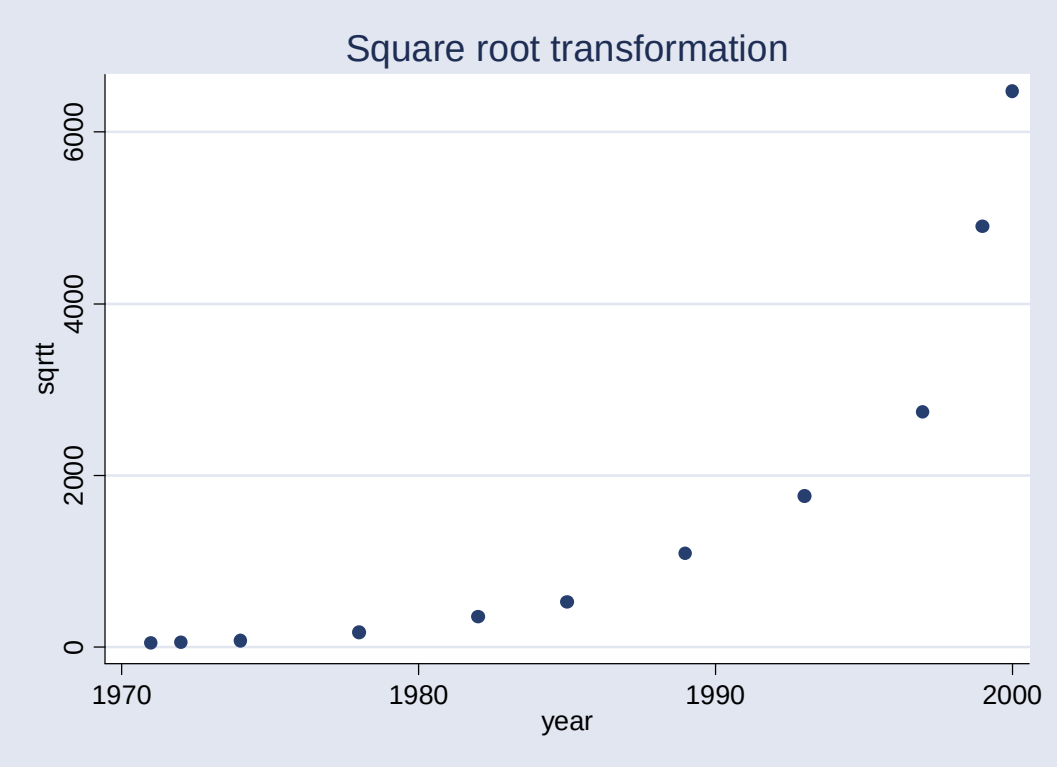
b. Consider a transformation of the number of transistors variable. Using the "circle of powers" would you recommend a suitable transformation, say $Y_{new} = (Y_{old})^P$, should be based on powers greater than 1, or powers less than 1. Give a brief explanation of your reasoning.

The data look like Quadrant IV of the Circle of Powers, so we would want to take X up or Y down. So we should base our transformation on power of P less than 1.

c. Use Stata to generate two transformations (square-root and log) of the number of transistors:
   i. P=1/2 or $Y_{new}=(Y_{old})^{0.5}$
   ii. P=0 or $Y_{new}=\log(Y_{old})$

Produce separate scatter plots for each of the transformed variables against time. Which of the two transformations seems preferable for these data?




The log transformation (P=0) has done a much better job making these data linear.