Computer Homework 7

The objective of this homework is to investigate the effect of scaling of variables on linear regression models. We have seen, that if a variable X has mean \( \mu \) and variance \( \sigma^2 \), then the transformed variable \( Z = (X-\mu)/\sigma \) has mean 0 and variance 1. Absent a better label, call this the "Z transformation."

1. Use the data in the Stata data file mls.dta to construct the "Z transformed" values of SP and SQFT. Denote these variables ZSP and ZSQFT, respectively. The Stata command line is

\[
\text{egen zsp=std(sp)}
\]

where egen is short for “extensions to generate.”

Estimate the following models:

- Model A: \( SP = \alpha + \beta \text{ SQFT} + \varepsilon \)
- Model B: \( SP = \alpha + \beta \text{ ZSQFT} + \varepsilon \)
- Model C: \( ZSP = \alpha + \beta \text{ ZSQFT} + \varepsilon \)

The observational subscript has been omitted to simplify notation.

2. a. Why is the estimate of \( \alpha \) in model B equal to the sample mean of SP?
   b. Why is the estimate of \( \beta \) in model B equal to the effect of a one standard-deviation increase in SQFT space in model A?

3. a. Why is the estimate of \( \alpha \) in model C essentially zero?
   b. Why is the estimate of \( \beta \) in model C equal to the square root of the coefficient of determination (R^2)?

4. Compare the t-statistics for \( \beta \) across models. Compare the coefficient of determination across models. Is there any substantive difference between these models?

Since some observations of a "Z transformed" variable will take negative values, it cannot be used prior to a log transformation. As an alternative, standardize SP and SQFT by dividing each variable by its standard deviation. Denote these variables by SSP and SSQFT, respectively. The Stata command lines

\[
\text{summarize sp} \\
\text{generate ssp=sp/r(sd)}
\]

will generate SSP. Results saved when a summarize command is executed may be recalled using \( \text{r(keyword)} \), where \text{keyword} indicates the result to be retrieved, and \text{sd} is the \text{keyword} for the standard deviation. Among the other keywords available are \text{mean}, \text{min}, and \text{max}.

Estimate the following models:

- Model D: \( \ln(SP) = \alpha + \beta \ln(SQFT) + \varepsilon \)
- Model E: \( \ln(SSP) = \alpha + \beta \ln(SSQFT) + \varepsilon \)

5. Why is it that only the estimate of \( \alpha \) changes?