

USING THE “FIND NORMALIZING TRANSFORMATIONS” FEATURE (Corresponds to Section 13.2 in textbook)

Considerations to take into account:

1. The values found by the software are just estimates – so it is silly to try to get too precise.

Example: The software suggests exponent 0.504 for TeachTax – this is silly as an exponent, but it suggests trying exponent 0.5.

In other words: Do *not* stop with the values the software spits out; use them as starting points in your decisions.

2. One important consideration: What exponents make sense in context?

Examples:

Exponent 1 is the easiest to interpret.

Exponent 0 also can be interpreted (multiplicative rather than additive scale)

Exponent 3 makes good sense when response is volume; 2 when response is area.

Exponent $\frac{1}{2}$ might make sense when the variable is an area; $\frac{1}{3}$ might make sense when the variable is a volume.

Other exponents might make sense in particular contexts.

Integers or fractions of integers make more sense than other numbers.

3. Excessive detail *can* also lead to overfitting – remember, you just have a sample, not the population. Parsimony is always one (not the only) consideration in model building.

4. Check out possible alternatives with a Likelihood Ratio Test (LRT). (Example later.)

5. Model selection methods are fairly robust as long as predictors are linearly related – so look at the scatterplot matrix when p-values from tests are marginal.

Exercise: Find suitable transformations for the Big Mac data (using Big Mac, Bread, BusFare, TeachSal, and TeachTax) after seeing the example done in class.