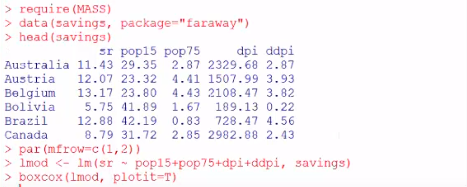
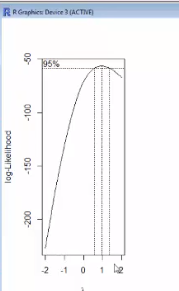
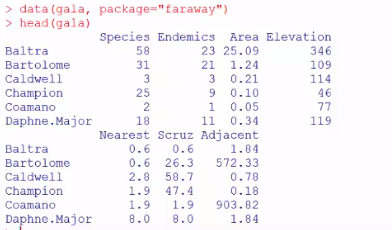
**STAT5120, Week 10: Transformations**

**Lecture 1**: Box-Cox transformations





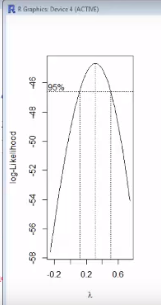


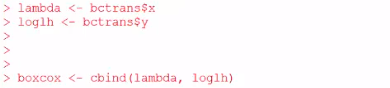


Builds a model with all variables from the ‘gala’ dataset and then builds a Box-Cox with range from -.25 to .75 and intervals of .05:

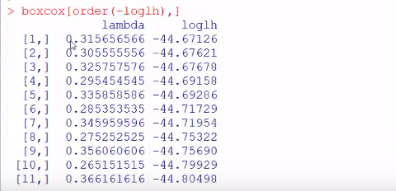


Graph shows that it maximizes somewhere around .3 and the graph does NOT straddle 1.0, indicating it may be useful:





We want the top lambda value (.3156):



Stores the lambda value in ‘l’:



Now builds a linear model raised to lambda (‘l’ not 1):



Shows summary of the new model. To recover (and translate) our model, you would need to back-transform both sides, i.e., y and the x’s by raising them to 1/λ which would be Y1/.3156 ~ X1/.3156 …

**Lecture 2**: take log of the response plus some number (α)





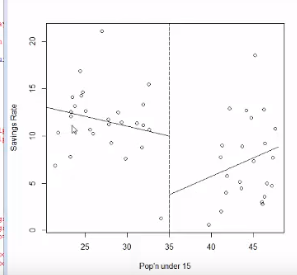
Use the logtrans() function above on the linear model

Broken Stick regression: attempts to model the data by sub-setting ranges of the data into groups in which differing slopes can be calculating. Here, he does this by building two models that sub-set the data into two groups, one with pop15 < 35 and the other with pop15 > 35:

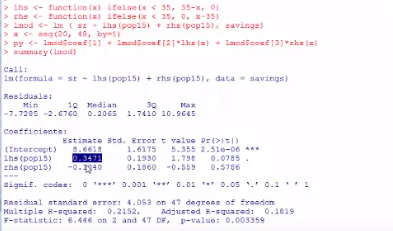


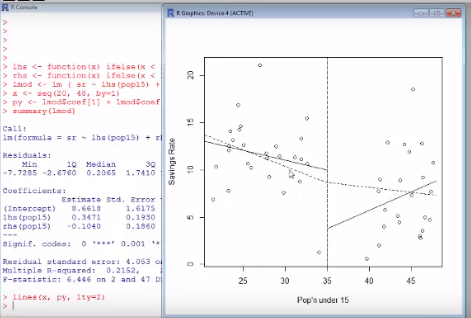






Above we see the results. But let’s say that we want to see the two lines “meet in the middle” rather than be non-contiguous. We can use another function instead:





The solid lines are separate least squares fits for the two segments and the dotted line shows if we were to regress on the constraints of the two segments together.