

shane.hutton@vanderbilt.edu
phillips 360 during class time tomorrow

40 questions

Outlier

Know cut off scores for the following:

- Cooks d
- Hat diagonal
- Rstudent
- Dfbetas
- Dffits

Don't need to know formulas for dfbetas or dffits just know how to use values

Look at SAS output or index plot

Know what leverage, discrepancy, local and global measures of influence. Know what they go with

Model selection

- All possible regression
- Forward and backward selection
- Stepwise
- Hierarchical- idea behind it, increment in r^2

Know how tolerance is computed

Know what AIC and SBC are and how they are used

Stepwise- does forward selection first but then can go back. Probably the optimal procedure.

If you want to find largest r^2 you can possibly get, use all possible regressions

Adjusted r^2 is the best metric most likely

Remember the more variables the higher the r^2 so that's why adjusted is important.

Box cox procedure – way to optimize a transformation. Optimization procedure.

Of x and y or x or y

Tukey's ladder of powers- be familiar with the possible transformation

Heteroscedasticity

- Weighted least squares, why we use them. Parameter estimates will change. We use it to account for heteroscedasticity

Logistic Regression

- Logit- using our dependent variable- log odds
- Can interpret model in terms of logit but doesn't make sense
- So we back transform into odds and odds ratios
- Know both interpretations
- Can plug in values to predict like any other model, will have to back transform for it to make sense.
- Odds of 1 is .5 probability
- Types of questions – interpretation – of parameter estimates
- Don't get r^2 get deviant statistics

Other Material

Robust

- model can handle mild violations without getting messed up
- Not robust means we have to be strict about any violations

Brown-Forsythe test

- using a formal test to look at heteroscedasticity- split into low and high section, can split into x or \hat{y} .
- Find median of each group or residuals
- Find deviation of each resid from median

Differences between multiple and simple regression

- F test and t test are equivalent in simple linear regression
- f test is a little different
- Standardized coefficient = r in slr but not in multiple unless they are uncorrelated
- Hold at 0 for interaction
- Hold constant for multiple or control for

Review matrix algebra

Assumptions of error terms

Independence

Normality- not need for OLS but is needed for inference tests