

Methods 2: Lab 7

Tom Hanna

University of Houston

tlhanna@uh.edu

March 29, 2021

Overview

- 1 Statistics Overview
 - Objective: Beyond simple OLS
 - Things to remember
- 2 Coding Overview
 - Things to watch out for
 - Things to remember



Nonlinear transformations

- Most relationships aren't perfectly linear
- New code lines 14-29 in script illustrate this
- Transforming variables approximates true relationship
- Log transformations: several uses, common
- Log transformations: *may* help with heteroskedasticity
- Log transformations: *may also* make heteroskedasticity worse
- Log transformations: challenging to interpret



Count variables

- Count variables violate the normality assumption
- They follow a Poisson distribution
- Poisson regression is not OLS
- Poisson regression is a GLM/MLE technique

Things to remember

- **Save the Wooldridge Table from Page 4 as its own file somewhere you can easily find it. Forever.**
- For more on R-squared and transformed variables:
<https://data.library.virginia.edu/is-r-squared-useless/>
- More on how not to use transformations:
<https://blog.usejournal.com/linear-regression-heteroskedasticity-myths-of-transformations-d27dc00db6ea>

Code hints and nudges

- \LaTeX - If you copy special characters like \$ or _ into \LaTeX , you need to precede it with a \ or you will get an error
- You may need to uncomment and run line 16

Coding details

- Health script
 - Lines 6-7 show an alternative method of plotting the first plot from the worksheet code
 - Lines 14-29 show a method using *ggplot2*
- Crab Script
 - A few labs ago, we had a problem with a non-numeric variable causing an error with the *round* command: Line 6 shows how to remove a problematic column quickly

The End