Juan Villegas

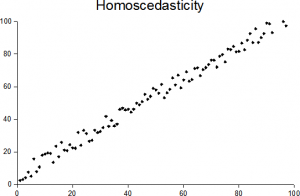
Final Exam

Answer 4 of the following. Each one is worth 25 points.

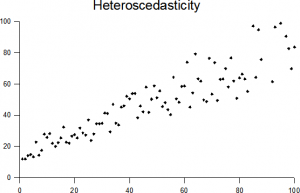
1. Why do Economist engage in quasi experiments instead of experiments? What’s the difference between them?
   1. The difference between a quasi-experiment and a normal experiment is how individuals are selected for groups. For example, in a normal experiment candidates might be randomly selected to be in certain groups. In a quasi-experiment, the groups are chosen strategically to see cause and effect. An example of this can be from the homework that I had a while ago. The experiment was: A group of students were given high speed internet. The students were divided by a couple of factors. For instance: athletes where one group, Engineer majors where another group, art majors where another and so on. This Quasi-experiment measures outcomes, it has treatments, and experimental units, but there was no random assignment.
   2. Economists engage in these experiments because they have more control over the experiment they are conducting and can see what directly might be associated with something. I believe this reduces the chances of lurking/outside variables affecting experiments.
2. In what instances would you use a logit regression and what does it tell you?
   1. You use logit regression when you have dummy variables. In other words: When your results can be binary (yes or no…. 1 or 0… male or female and so on). Logit regression is used to see if 2 binary variables (usually categorical) are more significant than one another. For example, in the textbook we had examples like when we compared: someone’s salary that was a male vs female, having a college degree vs no college degree, etc…. Logit regression tells you the log odds of something happening… You can use formulas to know the odds and even the probability.
3. What is multicollinearity? How do you test for it and why is it a problem?
   1. Multicollinearity happens when 2 or more independent variables (predictor variables) are highly correlated (+-.7) with one another in a regression model. This means that an independent variable can be predicted from another independent variable using a regression model.
   2. You can test for multicollinearity using statistical software like R, STATA, or even some python modules. Some examples of code can be the corr function (STAT) or you can create a heatmap on Python/R that shows you which variables can be correlated with what.
   3. Multicolinearity wont have much of an impact on a models accuracy but, multicollinearity does affect the variance/R^2 of a model. It can cause overfitting in our model and can cause our model to become untrustworthy

5. What’s the difference between homoskedasticity and heteroskedasticity? Which one do you have to correct for and how do you correct for it?

Homoscedasticity: If we created a linear regression model with a LSRL on it, homoscedastic would mean that the residuals from the LRSL are all about the same distance away from the LSRL.



Heteroscedasticity: If we created a linear regression model with a LSRL on it, heteroscedastic would mean that the residuals from the LRSL are all wildly scattered.



We correct Heteroscedasticity because it makes coefficient estimates less precise. Low precision on coefficient estimates increases the likelihood that the more uncorrected from the actual estimation.

You can correct heteroscedasticity by transforming data (which means to manipulate X or Y variables to pull the data how you would like) OR compute the weighted least squares (WLS) estimator using a hypothesized specification for the variance.