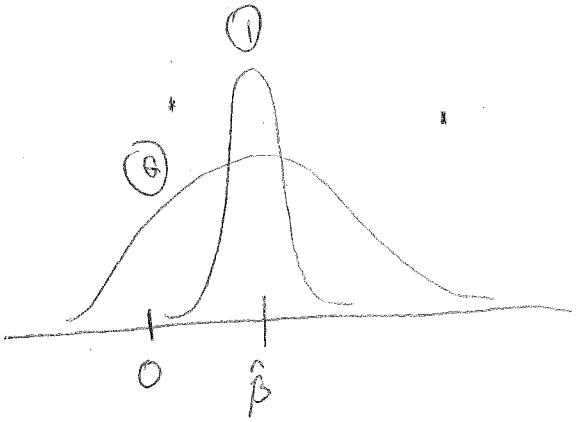
Applied Econometrics
Prof. Leo Feler
Quiz 2

Name: Key

1. Why do we care about the possibility of underestimating the variance of β_{OLS}?

We don't want to erroneously conclude that we have found an effect

Example in class was deciding that blood samples did not contain HIV when they actually did



1 says $\beta \neq 0$
2 says we can't distinguish

If our standard errors are too tight, we will be more likely to reject the null hypothesis, which can have large ramifications in a situation such as testing blood samples for HIV.

The goal is to make it as difficult as reasonably possible to prove the effect we are looking to estimate, which we do by conservatively calculating the largest standard errors (within reason).

2. Heteroskedasticity:
   a. What is heteroskedasticity?

$Var(\varepsilon) \neq \sigma^2$ → The variance of the residuals is correlated with the X_i (not constant)

---

   b. What does it imply for the variance of $\hat{\beta}_{OLS}$ if we do not correct for it?

The estimated variance of $\hat{\beta}$ tends to be too small (if + corr with X_i)

$Var(\hat{\beta}) \neq \hat{\sigma}^2(X'X)^{-1}$

With robust standard errors, $Var(\hat{\beta}) = (X'X)^{-1}X'\hat{\Sigma}X(X'X)^{-1}$

If the $\hat{\Sigma}$ is a diagonal matrix $[\sigma^2, \sigma^2, \sigma^2 \ldots, \sigma^2]$, this converges to $\hat{\sigma}^2(X'X)^{-1}$

   c. How do we test for heteroskedasticity?

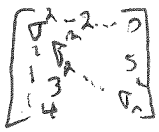
Test for heteroskedasticity with a White test, or, less optimal, a Breusch Pagan test (test stat) $NR^2 \sim \chi^2(q)$ where $q = K-1$ (df; # regressors)

You can also plot the data and look for signs of nonconstant residual variance.

3. Clustering:
   a. What is clustering?

correlation between observations within groups

$\rightarrow E(\varepsilon_i \cdot \varepsilon_j) \neq 0$



ex/ $\varepsilon_{is} = a_s + u_{is}$ (a_s: same for all students in a school; u_{is}: i.i.d)

Clustering means there are off-diagonal elements of the variance-covariance matrix that do not equal zero, and we have an idea of where, for example, within families or classes at school. But we assume there is some level of analysis, ~~such as the classroom,~~ at which the errors are i.i.d.

   b. If there is a high intraclass correlation between observations, what does this imply for the sample size you need to obtain a given level of significance once you correct for clustering? Why is this the case?

You need a larger sample because you have fewer independent observations than you think you do.

If 2 individuals are perfectly correlated, you really only have 1 observation. If intraclass correlation is 1, that is as if there were only 1 student. The higher the correlation, the fewer effective observations there are.

   c. You can only correct for clustering using the cluster command in Stata if you have more than 42 clusters. What can you do if you have less than 42 clusters to correct for clustering?

You can use bootstrapping, or possibly collapse the data by means for each grp