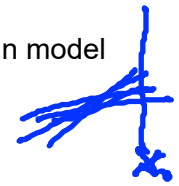
Stat 414 – Day 10
Heteroscedasticity-Consistent Standard Errors (12.2)

Previously

- Basic regression model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$
  - $E(Y|x_i) = \beta_0 + \beta_1 x_i$; $V(Y|x_i) = V(\epsilon_i) = \sigma^2$; $Cov(\epsilon_i, \epsilon_j) = 0$
  - One way to estimate $\hat{\sigma} = \Sigma(y_i - \hat{y}_i)^2/(n-p-1)$ where p = # slopes in model
  - $SE(\hat{\beta}_1) = \frac{\hat{\sigma}}{s_x\sqrt{n-1}}$ or $V(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{s_x^2(n-1)}$ *aka* $\hat{\sigma}^2(X^TX)^{-1}$
  - $SE(\hat{y}) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_o-\bar{x})^2}{s_x^2(n-1)}}$ *aka* $\hat{\sigma}^2(X(X'X)^{-1}X')$
- "Fixes" for normality, linearity, and unequal variance include transformations, polynomial models, weighted regression
  - One of the main reasons for dealing with the heteroscedasticity is otherwise our estimates of the standard errors of our slope coefficients may be off, which impacts our p-values and confidence intervals.

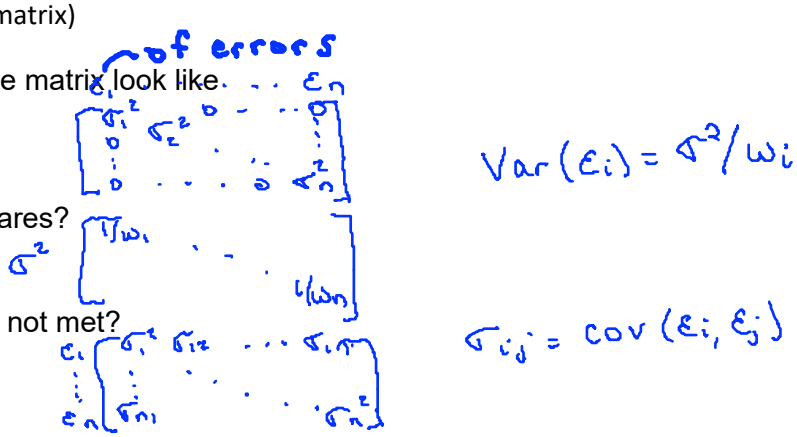


In matrix notation, our assumptions about the random errors can be written:

$$V(\epsilon|X) = \sigma^2 \boldsymbol{I} = \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix}$$ (n x n matrix)

(a) What does this variance-covariance matrix look like

- if we have heteroscedasticity?
- if we are using weighted least squares?
- if the independence assumption is not met?



Key Idea: A more general form of the variance-covariance matrix for the estimated coefficients is $Var(\hat{\beta}) = (X^TX)^{-1}X^TCov(\epsilon)\,X\,(X^TX)^{-1}$ with WLS simplifies to $\hat{\sigma}(X^TWX)^{-1}$.

Example 1: The Salaries dataset within the carData package consists of nine-month salaries collected from 397 collegiate professors in the U.S. during 2008 to 2009. In addition to salaries, the professor's rank, sex, discipline, years since Ph.D., and years of service were also collected.

(a) Run a basic regression model to predict salary from years since Ph.D. and years of service.

(b) Use matrix multiplication in R to confirm the standard error calculations.

(c) Use vcov to verify the standard errors of the coefficients.

Key Idea: The diagonal elements of the variance-covariance matrix (of the estimated coefficients) gives us the standard errors.

(d) Run a model that puts more weight on early-career professors and less weight on late-career professors. How do the models compare?

Which model is better? Then one that has the right variance structure! But we often don't know that in advance… "The best we can do is apply theory and judgment in a thoughtful way."

When you don't know the weights to use, some options include

- estimating them from the residuals (see iteratively reweighted least squares; IRWLS)
- heteroscedasticity-consistent standard errors aka "robust standard errors"
  White (1980) proposed estimating the "meat" with $(X^T \hat{U} X)$ where $\hat{U}$ is a diagonal matrix of the squared residuals (aka HC0). The main idea is you have taken into account the heteroscedasticity without having to know about or model the functional form of the heteroscedasticity or use "arbitrary" transformations.

Apply the Huber-White adjustment (Heteroscedasticity-Consistent) to the standard errors, and rerun the significance tests with the "corrected" standard errors.

```
#library(lmtest)
#library(sandwich)
vcovHC(model1, "HC1") # HC1 gives us the White-Huber standard errors
coeftest(model1, vcov = vcovHC(model1, type = "HC1"))   #updates the significance tests
```

(e) How do the standard errors of the slope coefficients change? Does the statistical significance of any of the variables change? (If not, then can claim analysis was not being affected by the heterogeneity.)

Example 2: The data in MMConcept.txt are from a study of 7th grade student GPAs (pre-2003). Of interest is whether subscales of the Piers-Harris self-concept scale predict GPA after adjusting for IQ and sex.

(a) Fit the model to predict GPA from IQ, sex, and the behavior, popularity, and anxiety subscales. What do you conclude? Is there evidence of heterogeneity? Does transforming GPA help?

(b) Carry out hypothesis tests for the coefficients using a standard error estimator that does not assume homogeneity. What do you conclude?

Reminders:

- When heteroscedasticity is discovered, we should not simply ask "What can I do to make the problem go away?" without also asking "What does heteroscedasticity tell me about the process I am studying?" (Hayes & Cai, 2007). Keep in mind that non-constant variance could be due to a misspecified model (e.g., missing key predictors, interactions, or non-linear effects).
- Weighted least squares is a special case of *generalized least squares*, but we need to learn a few other things first…