This handout goes through some math involved in showing what happens to the usual properties of estimates obtained by least squares regression when the error variance is not constant, as is assumed in the Classical Linear Regression Model. Note that as a byproduct of showing what happens under heteroskedasticity, you will also (1) have all the steps necessary to demonstrate that when error variance IS constant least squares estimates are unbiased and in addition and (2) you will see how to obtain the variance of the least squares estimates (i.e., the standard error) in the single regressor case.

In the simple regression model:

\[ Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \]

we know that by minimizing the sum of squared errors (deviations of data points from a straight line) we get the least squares estimate of \( \beta_1 \). I will now take several steps to re-write this formula so that we can take the variance of the estimate of \( \hat{\beta}_1 \) – so that we can find the standard error. We being with the usual equation for calculating \( \hat{\beta}_1 \) and then make a series of substitutions:

\[
\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}
\]

\[
\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(\beta_0 + \beta_1 X_i + \varepsilon_i - \bar{Y})}{\sum (X_i - \bar{X})^2}
\]

\[
\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(\beta_0 + \beta_1 X_i + \varepsilon_i - (\beta_0 + \beta_1 \bar{X}))}{\sum (X_i - \bar{X})^2}
\]

\[
\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(\beta_1((X_i - \bar{X}) + \varepsilon_i))}{\sum (X_i - \bar{X})^2}
\]
\[
\hat{\beta}_1 = \frac{\beta_1 \sum (X_i - \bar{X})^2 + \sum \varepsilon_i (X_i - \bar{X})}{\sum (X_i - \bar{X})^2}
\]

\[
\hat{\beta}_1 = \beta_1 + \frac{\sum \varepsilon_i (X_i - \bar{X})}{\sum (X_i - \bar{X})^2}
\]

These steps transform the equation so that we have isolated \( \beta_1 \) – the population value. If our estimate, \( \hat{\beta}_1 \), is unbiased, we have to show that the term after \( \beta_1 \) is equal to zero if the Gauss-Markov assumptions are met. The key ideas we will exploit are (1) that the expectation of a constant is a constant; \( E[k\alpha] = kE[\alpha] \); and (2) that for any dataset, the summations above are a known, finite quantity. The first step is to take the expectation of both sides and then use our two key ideas:

\[
E[\hat{\beta}_1] = E[\beta_1 + \frac{\sum \varepsilon_i (X_i - \bar{X})}{\sum (X_i - \bar{X})^2}]
\]

\[
E[\hat{\beta}_1] = \beta_1 + \frac{1}{\sum (X_i - \bar{X})^2} E[\sum (X_i - \bar{X}) \varepsilon_i]
\]

\[
E[\hat{\beta}_1] = \beta_1 + \frac{1}{\sum (X_i - \bar{X})^2} \sum (X_i - \bar{X}) E[\varepsilon_i]
\]

\[
E[\hat{\beta}_1] = \beta_1 + \frac{1}{\sum (X_i - \bar{X})^2} \sum (X_i - \bar{X}) E[\varepsilon_i]
\]

But since \( E[\varepsilon_i] = 0 \) (so long as we include a constant in the regression), we can see that

\[
E[\hat{\beta}_1] = \beta_1
\]

Therefore the least squares estimate of the slope coefficient is unbiased. On average, our estimate will reproduce the population value. Note that var(\( \varepsilon_i \)) has not been used here, so even if var(\( \varepsilon_i \)) is not constant across observations (i.e., if var(\( \varepsilon_i \)) = \( \sigma_i^2 \)) the least squares regression coefficients will still be unbiased.

What about the standard variance/standard error of the least squares coefficients? Begin by returning to the expression for the least squares regression estimate. We will rewrite that expression as a variance. Since the true regression coefficient and the X's are non-stochastic, their variance is zero. That helps to simplify our expression quite a bit by the time we get to the third step: we can get rid of some terms since \( \text{var}(\beta_1) = 0 \)

\[
\text{var}(\hat{\beta}_1) = \text{var}\left[\frac{\beta_1 \sum (X_i - \bar{X})^2 + \sum \varepsilon_i (X_i - \bar{X})}{\sum (X_i - \bar{X})^2}\right]
\]
$$
\text{var}(\hat{\beta}_1) = \frac{1}{\left[ \sum (X_i - \overline{X})^2 \right]} \left( \sum \text{var}(\beta_i (X_i - \overline{X})^2) + \sum \text{var}(\varepsilon_i (X_i - \overline{X})) \right)
$$

$$
\text{var}(\hat{\beta}_1) = \frac{1}{\left[ \sum (X_i - \overline{X})^2 \right]} \left( \sum (X_i - \overline{X})^2 \text{var}(\varepsilon_i) \right)
$$

Then it is the case that if $\text{var}(\varepsilon_i) = \sigma^2$ for all observations $i$, we can substitute and cancel some terms:

$$
\text{var}(\hat{\beta}_1) = \frac{1}{\left[ \sum (X_i - \overline{X})^2 \right]} \left( \sum (X_i - \overline{X})^2 \sigma^2 \right)
$$

$$
\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \overline{X})^2}
$$

This final equation is the expression for the variance of the least squares slope estimate (i.e., the standard error) in the single explanatory variable case that we have seen before. This is the equation you will find in Pindyck and Rubinfeld on page 54. Finally, this is what the Stata calculates unless you tell it otherwise. Operationally, Stata uses an estimate for $\sigma^2$ – the mean squared error (MSE) – when it calculates standard errors for the coefficients. Stata reports the “Root MSE” on the top, right-hand side of the regression output.

If, on the other hand, $\text{var}(\varepsilon_i) = \sigma_i^2$ (that is, the variance is not constant across all observations $i$), then the above expression for the variance of the least squares estimate of the regression coefficient is not correct. The correct expression requires that $\text{var}(\varepsilon_i)$ in the equation above be replaced by $\sigma_i^2$, rather than by $\sigma^2$. The “$i$” subscript indicates that the variance of the error term may change from observation to observation. This change yields the following:

$$
\text{var}(\hat{\beta}_1) = \frac{1}{\left[ \sum (X_i - \overline{X})^2 \right]} \left( \sum (X_i - \overline{X})^2 \sigma_i^2 \right)
$$

$$
\text{var}(\hat{\beta}_1) = \frac{\sum \sigma_i^2 (X_i - \overline{X})^2}{\left[ \sum (X_i - \overline{X})^2 \right]} \frac{1}{\left[ \sum (X_i - \overline{X})^2 \right]}
$$
When this formula for the variance of the error term is used to compute the standard error of the least squares estimate of the coefficients (and of course, using an analogous but more complicated expression in the multiple regression case), we have **ROBUST STANDARD ERRORS**.

But what is $\sigma^2$? If we do not know, then one possibility is to use an estimate:

$$\hat{\sigma}_i^2 = \hat{\varepsilon}_i^2$$

That is, the best estimate of the individual error variances is to square the residual – the difference between $Y_i$ and $\hat{Y}_i$. Stata can easily calculate this.

Substituting into the formula for the standard error we have:

$$\text{var}(\hat{\beta}_i) = \frac{\sum \hat{\varepsilon}_i^2 (X_i - \bar{X})^2}{\left[ \sum (X_i - \bar{X})^2 \right]^2}$$

The square root of this expression is known as the “White heteroskedastic-consistent standard error” – AKA “robust standard errors” – and is computed by most regression packages.