Properties of OLS: Property 1

- These algebraic properties are purely algebraic. They do not rely on our assumptions or interpretations!
- $\hat{y}_i$ is approximation of y_i
- The residual $\hat{u}_i = y_i - \hat{y}_i$ measures how good the approximation is.
- $\hat{u}_i$ is not always zero.
- Property 1:

$$\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i = 0.$$

Properties of OLS:Property 1

Let's verify the first one together. Remember

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i,$$

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) \\ &= \bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = 0. \end{aligned}$$

Properties of OLS:Property 1

- Property 2 : $\frac{1}{n}\sum_{i=1}^{n} x_i \hat{u}_i = 0.$
- How shall we interpret this?

$$\begin{aligned}
\sum_{i=1}^{n} (x_i - \bar{x})(\hat{u}_i - \bar{u}) &= \sum_{i=1}^{n} (x_i - \bar{x})\,\hat{u}_i - \sum_{i=1}^{n} (x_i - \bar{x})\,\bar{u} \\
&= \sum_{i=1}^{n} x_i \hat{u}_i - \sum_{i=1}^{n} \bar{x}\hat{u}_i \\
&= \sum_{i=1}^{n} x_i \hat{u}_i
\end{aligned}$$

- So

$$\sum_{i=1}^{n} x_i \hat{u}_i = 0 \Rightarrow (n-1)^{-1} \sum_{i=1}^{n} x_i \hat{u}_i = 0 \Rightarrow \mathrm{Cov}(X, \hat{u}) = 0$$

Goodness of Fit

- We want to say something about how well our model fits the data.

- In doing so, we will need to answer a few questions:
- How do we define "well"?
- In other words, what criteria we using to measure the **goodness of fit**?
- Is this a useful criteria and, if so, what is it useful for?

- To make progress, we need to define a few new concepts

Total Sum of Squares

- The first concept: **Total Sum of Squares (SST)**

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

- Measures the total variation in Y.
- Dividing by (n-1), we get our estimator of the sample variance.

$$SST = (n-1)\hat{Var}(Y)$$

- Example: if y_i is income, SST is one measure of income inequality

Explained Sum of Squares

- The second concept: **Explained Sum of Squares (SSE)**

$$SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

- Measures the part of the variation in Y that is explained by the regressor
- Example: if y_i is income and x_i is wage, the SSE is the part of income inequality that is explained by wage inequality.

Residual Sum of Squares

- The third concept: **Residual Sum of Squares (SSR)**

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2$$

- Measures the part of the variation in y that is not explained by the regressor.
- Example: if y_i is income and x_i is wage, then SSR is the part of income inequality that is **not** explained by wage inequality

SST=SSE+SSR

- Given these definitions, it's intuitive that the total variation should equal the explained variation plus the residual variation:

$$SST = SSE + SSR$$

And we can show that this is indeed the case:

$$\begin{aligned} SST &= \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \\ &= \sum_{i=1}^{n} [\hat{u}_i + (\hat{y}_i - \bar{y})]^2 \\ &= \underbrace{\sum_{i=1}^{n} \hat{u}_i^2}_{SSR} + \underbrace{2\sum_{i=1}^{n} \hat{u}_i (\hat{y}_i - \bar{y})}_{=0, why?} + \underbrace{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}_{SSE} \end{aligned}$$

SST=SSE+SSR

- So we are left wanting to show that the following expression (the sample covariance between the residuals and the fitted values) s equal to zero:

$$
\begin{aligned}
2\sum_{i=1}^{n}\hat{u}_i\left(\hat{y}_i-\bar{y}\right) &= 2\sum_{i=1}^{n}\hat{u}_i\left(\hat{\beta}_0+\hat{\beta}_1x_i-\bar{y}\right) \\
&= 2\sum_{i=1}^{n}\hat{u}_i\hat{\beta}_0+2\sum_{i=1}^{n}\hat{u}_i\hat{\beta}_1x_i-2\sum_{i=1}^{n}\hat{u}_i\bar{y} \\
&= 2\hat{\beta}_0\sum_{i=1}^{n}\hat{u}_i+2\hat{\beta}_1\sum_{i=1}^{n}\hat{u}_ix_i-2\bar{y}\sum_{i=1}^{n}\hat{u}_i
\end{aligned}
$$

- What are these terms equal to?

R-squared

- We are now ready to propose a possible measure for goodness of fit:
- R^2 = SSE/SST = [Explained Sum of Squares] / [Total Sum of Squares]
- In other words, R2 is defined as the fraction of the total sample variation in y that is explained by x
- Since SST = SSE + SSR, this implies that SSE=SST-SSR
- Plugging this into R^2, we can write R^2 another way...
- R^2= SSE/SST = (SST-SSR)/SST = 1 - SSR/SST

  = 1 - [Residual Sum of Squares] / [Total Sum of Squares]

R-squared

- These are the two extreme cases, so $0 \leq R^2 \leq 1$
- When interpreting R^2, we often multiply by 100 and interpret it as a percentage
- Example: So if R^2=0.37, we would say that 37 percent of the of the sample variation in y is explained by x

Example 2: CEOs

Source	SS	df	MS
Model	5166419.04	1	5166419.04
Residual	386566563	207	1867471.32
Total	391732982	208	1883331.64

Number of obs	=	209
F(1, 207)	=	2.77
Prob > F	=	0.0978
R-squared	=	0.0132
Adj R-squared	=	0.0084
Root MSE	=	1366.6

salary	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
roe	18.50119	11.12325	1.66	0.098	-3.428196	40.43057
_cons	963.1913	213.2403	4.52	0.000	542.7902	1383.592

- Recall this regression from the previous lecture
- What is the R^2?
- Does this seem high or low?

Cause of bad fit

- Small R^2 indicates a bad fit. Reasons for bad fit include:
- (1) there are too many variables other than X that influence Y
- (2) there may be nonlinear relationship between X and Y.
- In the former case, small R^2 is not necessarily a red flag as long as there is plausible theory that argues the other factors are uncorrelated with X
- The nonlinearity issue usually needs to be addressed.