Today's Agenda

- Multicollinearity Ex Discussion
- Spring Interpretation of Multicollin
- Finish up Modeling Probs
  - ↳ Non Normal Resids
  - ↳ Non Linearity (recap)
- More on Interactions

$$\hat{\beta} = VD^{-1}U^TY = V\left(\begin{matrix} \frac{1}{d_1} & & \\ & \ddots & \\ & & \frac{1}{d_p} \end{matrix} \right)$$

Modeling Problems

- Non Normality of Residuals

Let's say you fit a regression, plot the resids w/ a histogram & it looks not Normal (QQplot also).

Good

- $\hat{\beta}$ is still BLUE

Bad

- Normality was needed for inference
  - ↳ t-tests
  - ↳ F-tests
  - ↳ CIs
  - ↳ PIs

Good

n is rly big

If $n >>> 0$ then

a Centre Limit Theorem type result

gives us the fact that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{D}} N(0, Var(\hat{\beta}))$$

$$\Updownarrow$$

$$\boxed{\hat{\beta} \overset{\circ}{\sim} N(\beta, Var(\hat{\beta}))}$$

For this reason, if n is large this assumption is the "least important" to obey.

How to check Non-Normality?

- Histogram of Residuals
- QQ Plot
- Tests for Normality: Anderson Darling, Shapiro Wilks, Ryan Joiner, Kolmogorov-Smirnov

H_0: resids are Normal $\quad e \overset{=}{\sim} F$

H_1: resids exhibit significant deviation from Normality $\quad e \overset{\neq}{\sim} F$

Skew:

True parameter: $E\left(\left(\frac{e_i}{\sigma}\right)^3\right)$

sample skewness $\frac{1}{n}\sum_i \left(\frac{e_i}{\hat{\sigma}}\right)^3$

Kurtosis:

True parameter $E\left(\left(\frac{e_i}{\sigma}\right)^4\right)$

sample kurtosis $\frac{1}{n}\sum_i \left(\frac{e_i}{\hat{\sigma}}\right)^4$

Omnibus K-Squared Normality Test

$K^2_{stat} = [\underbrace{Z_1(g_1)}_{\text{skewness stat}}]^2 + [\underbrace{Z_2(g_2)}_{\text{kurtosis}}]^2$ $\quad Z_1$ & Z_2 special transformations

$K^2_{stat} > \chi^2_2(1-\alpha) \Rightarrow$ reject H_0

Jarque-Bera Test:

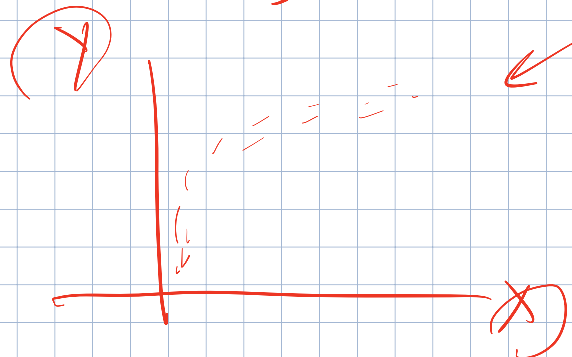
Combining Skewness & Kurtosis

$$JB_{stat} = \frac{n-p}{6}\left(\underset{\text{skew}}{(g_1)^2} + \frac{1}{4}\underset{\text{kurtosis}}{(g_2)^2}\right)$$

Under H_0: $JB_{stat} \sim \chi^2_2$

Idea: If you have non Normal resids & you don't trust the asymptotic CI, consider using a bootstrap CI.

False Linearity



If nonlinearity exists:

① transform Y (watch out → affects heteroskedasticity)

② transform X

③ transform Y & X (potentially log form)

$$\tilde{Y} = \begin{bmatrix} \log Y_1 \\ \vdots \\ \log Y_n \end{bmatrix}$$

$$\log Y = \beta_0 + \beta_1 \log X$$

$$\Updownarrow$$

$$\tilde{X} = \begin{bmatrix} 1 & \log X_1 \\ \vdots & \vdots \\ 1 & \log X_n \end{bmatrix}$$

$$e^{\log Y} = e^{\beta_0 + \beta_1 \log X}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$\Updownarrow$$

$$Y = e^{\beta_0} e^{\beta_1 \log X}$$

$$= e^{\beta_0} e^{\log X^{\beta_1}}$$

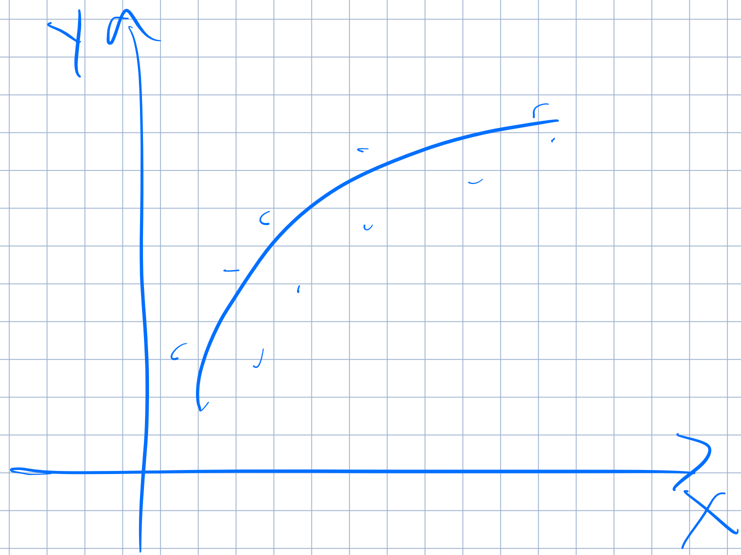
$$= e^{\beta_0} X^{\beta_1}$$

$$\tilde{\beta_0} = e^{\beta_0}$$

$$= \tilde{\beta_0} X^{\beta_1}$$

Watch out for interpretations $\longrightarrow$

sometimes hard after transformation

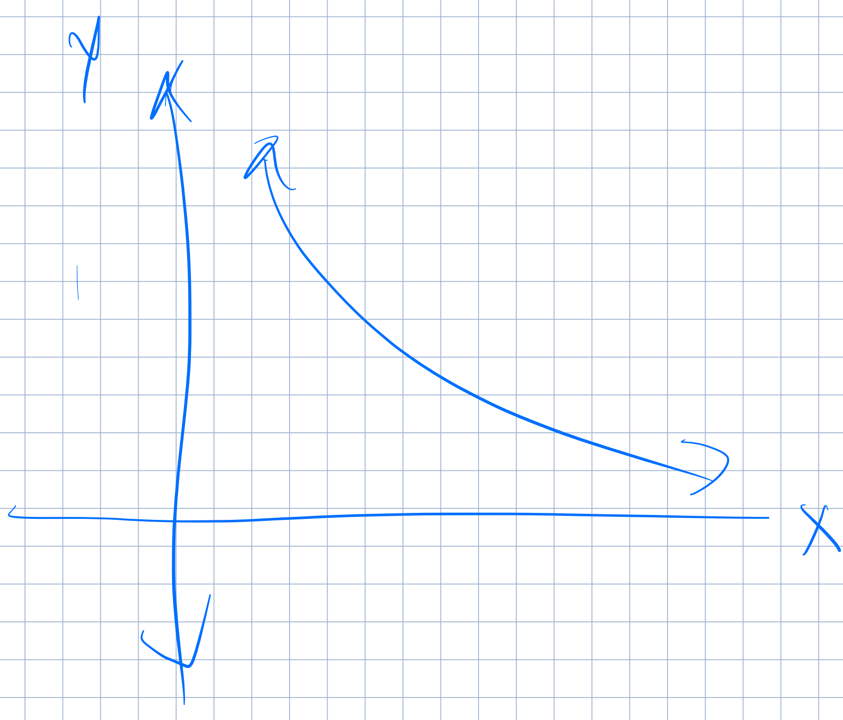


$f(x) = \log(x)$

$f(x) = \sqrt{x}$

$f(x) = x^2$

$f(x) = e^x$



$f(x) = 1/x$

$f(x) = e^{-x} = \frac{1}{e^x}$