When does X have a col of $\mathbb{1}_n$?

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

$$X = \begin{bmatrix} 1 & | & | \\ 1 & | & | \\ \vdots & X_1 & X_2 \\ \vdots & | & | \\ 1 & | & | \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

vs

Intercept free Model

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

$$X = \begin{bmatrix} | & | \\ X_1 & X_2 \\ | & | \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Modeling Problems

Structural:

- Multicollinearity ☆
- Influential pts

Violating Model Assumptions:

- Heteroskedasticity
- Non-Normal Residuals
- False assumptions of linearity

Multicollinearity

Problem 2 or more predictors are highly correlated

Design Matrix:

$$X = \underbrace{\begin{pmatrix} \mathbb{1}_n & x_1 & x_2 & \cdots & x_{p-1} \end{pmatrix}}_{\text{some cols are linearly dependent...}}$$

if X doesnt have rank(X)=p, X^TX is not invertible

$$\hat{\beta} = \underline{\underline{(X^TX)^{-1}}} X^T Y$$

$$X^TX = U\Lambda U^T = U\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_p\end{pmatrix}U^T \quad \approx 0$$

$$(X^TX)^{-1} = U\Lambda^{-1}U^T = U\begin{pmatrix}1/\lambda_1 & & & \\ & 1/\lambda_2 & & \\ & & \ddots & \\ & & & 1/\lambda_p\end{pmatrix}U^T$$

$$\boxed{\frac{1}{\approx 0}}$$

For ex: lets make a simulated data set w/

$$Y_i = 1 + 2X_{1i} + 4X_{2i} + \varepsilon_i \qquad \text{true } \beta = \boxed{\begin{pmatrix}1\\2\\4\end{pmatrix}}$$

If X_1 & X_2 exhibit perfect correlation, e.g.

$$X_2 = 4X_1$$

$$Y_i = 1 + 2X_{1i} + 4(4X_{1i}) + \varepsilon_i$$

$$= 1 + 2X_{1i} + 16X_{1i} + 0X_{2i} + \varepsilon_i$$

$$= 1 + 18X_{1i} + 0X_{2i} + \varepsilon_i \qquad \text{true } \beta = \begin{pmatrix}1\\18\\0\end{pmatrix}$$

$\Downarrow$

$$Y_i = 1 + 18X_{1i} + \overset{\overset{0}{\parallel}}{X_{2i} - X_{2i}} + \varepsilon_i$$

$$= 1 + 18X_{1i} + 4X_{1i} - X_{2i} + \varepsilon_i$$

$$= 1 + 22X_{1i} - 1X_{2i} + \varepsilon_i$$

true $\beta = \begin{pmatrix} 1 \\ 22 \\ -1 \end{pmatrix}$

non-identifiability

Damage If I don't address this issue, I don't know which β I'm targeting & so:

① $\hat{\beta} = (X^TX)^{-1}X^TY$ could be impossible to calculate or very unstable b/c of dividing by ≈ 0.

② $Var(\hat{\beta}) = \sigma^2(X^TX)^{-1} \uparrow$

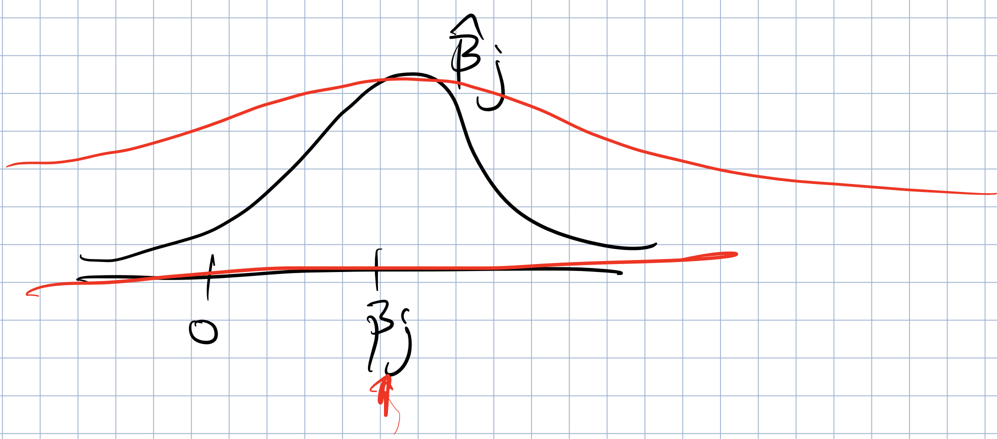
inflated variance of LS estimates.

③ Inference

$$t = \frac{\hat{\beta}_j - 0}{\widehat{SE}(\hat{\beta}_j) \uparrow} \downarrow$$

loss of statistical power

$\Rightarrow$ cannot detect signal



<u>Symptoms to look for:</u>

- Global F-test rejects H_0 but indiv t-tests all fail to reject
- adding predictors to the model increases the SE of other predictors by a lot

Ex:

$corr(X_1, X_2) = 0$

no multicol

Model	$\hat{\beta}_1$	$\hat{\beta}_2$
$Y \sim X_1$	-1	NA
$Y \sim X_1 + X_2$	-1	-5

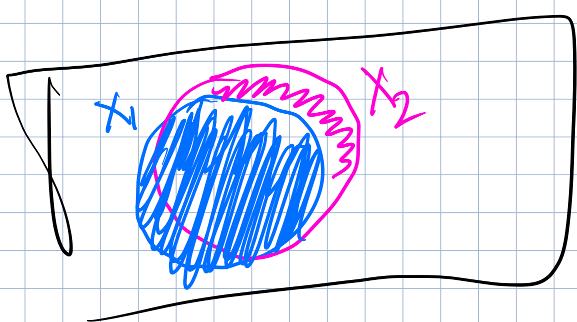
$corr(X_1, X_2) = .95$

multicoll.

Model	$\hat{\beta}_1$	$\hat{\beta}_2$
$Y \sim X_1$	-1	NA
$Y \sim X_1 + X_2$	10	-5

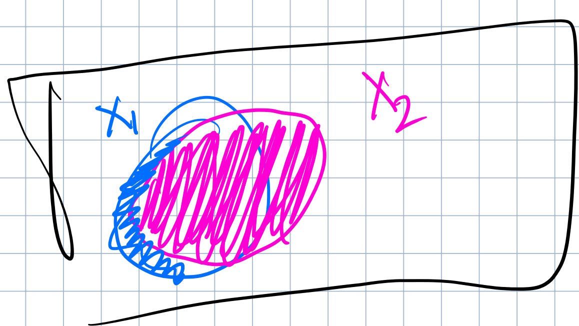
ANOVA Tables:

Type1 ANOVA table: significance changes depending on order.



SST

pred	SS	F	P	
X_1	↑	↑	↓	sig.
X_2	↓	↓	↑	Not sig.



SST

pred	SS	F	P	
X_2	↑	↑	↓	sig
X_1	↓	↓	↑	not-sig

Detection:

① Correlation Matrix (naive)

	X_1	X_2	X_3	X_4
X_1	1	.9		.05
X_2		1		
X_3			1	
X_4				1

② VIF - Variance Inflation Factor

VIF = measures how much the variance of $\hat{\beta}$ gets inflated by adding a specific predictor to the model:

$$VIF := \frac{1}{1-\underline{\underline{R_j^2}}}$$ where

R_j^2 is the coef. of det. from $\underline{\underline{X_j}} \sim \underbrace{X_1 + X_2 \cdots X_{j-1} + X_{j+1} + \cdots X_{p-1}}$

If $VIF = 1 \iff$ no corr b/w X_j & other preds.

$1 \leq VIF \leq 4 \iff$ "light"

$4 \leq VIF \leq 10 \iff$ "moderate"

$10 \leq VIF \iff$ "severe"

Solutions:

① Drop suspicious predictors

② Feature engineering & make a new predictor $X^* = g(\ldots)$ ← suspicious variables

More Advanced Methods

① Regularized Regression — Ridge / LASSO

② Dimension Reduction on X = PCA/SVD

③ Partial LS