

Math 455 Class 26 March 31

Today: R script from Ch. 7.

Important idea: simulation & extrapolation.

Our example: data are measured with an error. What would our results look like if data were measured without error?

Assume we know the approximate size of the error. Say error $\sim N(0, \sigma^2)$, where we know σ .

Idea: introduce MORE error.

If we know how results change with increasing error, extrapolate to figure out ~~to~~ what results we expect with decreasing error.

3 ways of looking for collinearity.

1. Correlation matrix of Predictors

2. Variance Inflation Factors

3. Condition number of $X^T X$.

1. gives very specific results, but doesn't catch every sort of collinearity.

3. Catches all sorts of collinearity, but doesn't give specific results like " $\text{Cor}(X_2, X_3) = 0.95$, so we might eliminate one of them."

2. Is in between.

Usual 1st evidence of collinearity:

predictions are "good" (high R^2 or F-stat)

but no variable is significant.

2nd piece of evidence "unstable" coefficient estimates $\hat{\beta}_i$.

What can be done : 2 ideas

- ① if several variables are correlated, eliminate some of them.
- ② Replace several variables with an average or "index" that takes information from all of them
- ②' a way of creating such an index: principal components.