Math 455 Class 25 March 23

Today: More about Chapter 7: "Problems with the Predictors"

We discussed "errors-in-variables": what if our Y or X is measured with errors? (Section 7.1) two ways: simulation in R and theoretical analysis as in Math 447.

A third angle: Suppose we want to answer the question: "what would the regression coefficients be if we had the true values of X (or Y)?"

Idea: we have no way to get rid of the errors. BUT we can increase the errors. So we can look at various error sizes and try to extrapolate to zero error.

R Example later.

Next: Section 7.2: changes of scale.

Simple example: we try to predict weight (Y) (in kg) from height (in meters) (X).

We get some regression output, including

$\hat{\beta}_1$, R^2, $\hat{\sigma}^2$, etc.

Suppose we change our X to be measured in cm.

$X_{new} = 100\, X_{old}$.

How does the regression output change?

We can ask the same type of question for Y:

if Y is measured in kg and Y_{new} is measured in grams, $Y_{new} = 1000\, Y_{old}$.

The answer is in the textbook, p. 104.

it's only 3 lines.

Linear (or technically affine) changes of variable: $Y_{new} = \dfrac{(Y_{old} + a)}{b}$ $X_{new} = \dfrac{(X_{old} + a)}{b}$

we understand what will happen.

Section 7.3 : Collinearity:

We have generally assumed that X^TX is invertible.

Equivalent: the predictors are linearly independent.

The opposite of linear independence is "collinearity."

In pure math, either you have linear independence (or invertibility) or you don't.

In applied math, is the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1/100 \end{pmatrix}$ invertible? It's "almost-not-invertible": it's very close to $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ which is NOT invertible. It's inverse is $\begin{pmatrix} 1 & 0 \\ 0 & 100 \end{pmatrix}$

and so any formula involving the inverse may produce large numbers.

This might mean, for example, that std. errors of $\hat{\beta}_i$ are very large, making our p-values high and coeff. "insignificant".

This problem is very common.

Two big ideas from 7.3 "Collinearity".

① How do we detect this phenomenon?

② What do we do about it?

For question ①:

(A1): We look at a regression output and see that estimates of $\hat{\beta}_i$ are insignificant (technically, difference from 0 is insignificant)

Maybe the R^2 is large, say 0.7+.

This is a symptom of collinearity.

(A2): We look at the correlation matrix of the predictors and see high pairwise correlations.

(A3): Variance inflation factors.

High VIF (bigger than 5 or 10, depending on context) indicates collinearity.

(A4): Condition Number: $K = \sqrt{\frac{|\lambda_p|}{|\lambda_1|}}$

(highest & lowest eigenvalues of X^TX)

Condition # > 30 suggests collinearity problems.

These methods are arranged in order:

increasing sophistication, increasing opaqueness.

Example: (A4) can detect approximate collinearity between any linear combination of the predictors and any other linear combination. e.g. $2X_1 - X_2$ is "close" to $3X_3 + 6X_4 - 5X_5$.

But: the number K doesn't directly tell you what this collinearity is, only that it exists.

(A2) Correlation matrix of predictors:

If can detect that X_1 is almost a linear function of X_2, e.g.

$Cor(X_1, X_2) = 0.99$: Explicit, but can't detect the stuff that A4 can.

Big question (2) about collinearity:

What can we do about this?

- Option 0: Do nothing.
- Option 1: Drop some variables.

  If X_1 & X_2 have high correlation, drop one of them..

- Option 2: Create an index out of variables that are highly correlated. Include only the index as an explanatory variable and not the components (variables) used to create it.

  (Maybe this index already exists?)

- Option 3: "Penalized" Regression.

  Examples: LASSO, Ridge, Elastic Net.

  These methods penalize larger coefficients and create bias. But they reduce the variance of coeff. estimates. (Chapter 11)

- Option 4: Principal Components or Partial Least Squares (Ch. 11).

  Replace old predictors with linear combinations, that avoid collinearity and drop the "unimportant" ones.

  It helps if we have an interpretation for the principal components, or if we don't care about interpretability.

- Option 5: Change our problem to one that doesn't have this issue.