Math 455 Class 30 April 9

Proposed midterm dates: Votes

Wed	Apr. 23	3
Fri	Apr 25	3
Mond	Apr. 28	(6)

Proposal: Monday April 28

Today: Generalized Least Squares and weighted least squares. (In Ch. 8 "Problems with the error.")

We assumed $\varepsilon \sim N(0, \sigma^2 I)$

(the errors are iid normal with variance σ^2)

To some extent we can adapt to this not being true. It is enough that we know that the errors have a multivariate normal distribution and that we know the structure of the variance-covariance matrix.

Knowing the "structure" means that we have

$$\text{Var}(\varepsilon) = \sigma^2 \Sigma$$ where σ is unknown but Σ is known.

The standard case is $\Sigma_{n\times n} = I$.

The simplest case of this is Σ is a diagonal matrix: this means the errors are uncorrelated but have different variances.

$$\Sigma = \begin{pmatrix} 1/w_1 & & 0 \\ & \ddots & \\ 0 & & 1/w_n \end{pmatrix}$$

these w's are called "weights".

and this special case is called "weighted least squares".

The idea is this: if one error has large variance (or small w_i) this error should be counted LESS.

The way we do this is we count this error with a multiple w_i – a small number.

So instead of $\sum_{i=1}^{n} e_i^2$ we

minimize $\sum w_i e_i^2$

Q: How would you ever know the matrix Σ, or even the numbers w_i in the diagonal (weighted least squares) case?

A: Imagine that you have measurements taken from 2 types of sensors, and one type is more reliable than the other. We can up-weight the reliable and down-weight the unreliable.

Another example: each data point is an average, but the size of the data set averaged differs from data point to data point.

Example in text: French Provincial Election

Data: Data is % of voters in the province who voted for a particular candidate. If the number of voters is small, this is a less reliable estimate of the percentage.

Q: Why do we write $\varepsilon \sim N(0, \sigma^2 I)$

rather than the ε_i are iid normal with common variance σ^2?

A: This means the same thing.

But the matrix notation is more compact.

For example (Faraway p. 33)

$$\hat{\beta} = (X^T X)^{-1} X^T y \sim N(\beta, (X^T X^{-1})\sigma^2)$$

is one line, but in Wackerly, et. al, this occupies many pages.

Also, when we "generalize least squares", we can see that we are just replacing I with Σ.

Key idea of Generalized Least squares:

Using linear algebra we can construct an equivalent problem in ordinary least squares where the iid normal assumption holds.

Solve this equivalent problem and then use linear algebra to translate this to a solution of our original problem.

All of this will be done for you by R.

for the Weighted Least Squares (WLS) case all you have to do is say

" lmod ← lm(Y ~ X1+X2, weights = ...)

For generalized least squares (nonzero error correlations) you need the gls command in the nlme package.