Math 455 Class 22 March 21

Leverages or "Hat Values".

Q: How much does the value of Y_i change the result of our regression analysis?

This is a broad question and there are quantitative versions of this in Ch. 6.

One quantitative version:

$\hat{Y}_i$ is a function of $X_1, \ldots, X_n, Y_1, \ldots, Y_n$.

(thinking here in the 1-variable case.)

We could compute $\dfrac{\partial \hat{Y}_i}{\partial Y_i}$

It turns out that this is easy!

$\hat{Y} = HY$ where H is the hat matrix.

So the answer is $\dfrac{\partial \hat{Y}_i}{\partial Y_i} = h_{ii}$ [ith diagonal entry of matrix H]

Note: Textbook writes h_i for h_{ii}.

This is a measure of the potential influence of Y_i on $\hat{Y}_i$.

Note that H depends only on X.

$$\left[H = X (X^T X)^{-1} X^T \right]$$

Intuitive version: if the X_i-value associated to Y_i is far from the mean X-value, then this point has a lot of "leverage".

R version: We can ask for

> hatvalues(lmod)

Properties of the hat values:

$$\sum_{i=1}^{n} h_i = p \quad = \# \text{ parameters.}$$

$$0 \leq h_i \leq 1.$$

Which points are worth investigating?

Rule of thumb: look at points with $h_i > \frac{2p}{n}$.

Note: in "experimental design" we could try to arrange our Xs so that no h_{ii} values are too large.

Note: "high leverage" and "outlier" are not the same concept.

outlier: a point that does not fit well with the model.

"high leverage": h_i unusually large $\left(> \frac{2p}{n} ?\right)$

See fig. 6.9 p. 86 for illustration of difference.

Also: "influential point": a point whose removal from the data set would change the fitted model a lot. Ex: 3rd diagram in 6.9.

Why is this important?

You do not want the result of your regression to be highly dependent on a few points.

If it is, then you might effectively have 3 data points and not 10,000 data points.

Next concept associated to leverages:

"standardized residuals".

Vector e of residuals: $e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$

$$e = Y - \hat{Y} = Y - HY = (I - H)Y$$

We can compute $\text{Var}(e) = \text{Cov}((I-H)Y, (I-H)Y)$

$$= (I-H)\sigma^2 I (I-H)^T = (I-H)\sigma^2 I (I-H)$$

$$= (I-H)^2 \sigma^2 I = \sigma^2 (I-H)$$

So $\text{Var}(e_i) = \sigma^2(1-h_i)$

We can "cook up" e_i to have mean 0 and variance 1:

$$r_i = \frac{e_i}{\hat{\sigma}\sqrt{1-h_i}}$$ ← standardized residual.

$$\hat{\sigma}^2 = \frac{RSS}{n-p}$$

We can test whether standardized residuals "look right" with a QQ plot.

(These are obtained with "rstandard" in R.)

Q: What can we do about a point that we suspect has high influence / is an outlier / etc?

A: Leave it out and see what happens.

This idea is called the "Jackknife".

There are often convenient formulas for computing what happens. This saves the trouble of doing n regressions, one for each missing point.

Notation: in the textbook subscript (i) denotes something with the i^{th} row of data removed.

Example: $\hat{Y}$ – vector of fitted values.

$\hat{Y}_{(i)}$ – remove the i^{th} row, then

then compute $\hat{\beta}_{0(i)}, \hat{\beta}_{1(i)}$ etc.

then compute a new $\hat{Y}$ (including the i^{th} row of data not used in finding $\hat{\beta}_{0(i)}$, etc.

so we can look at a difference $\hat{Y} - \hat{Y}_{(i)}$.

A vector measuring how much things change when we remove i^{th} row of data.

The "Cook Statistics" are a normalized version of the length of this vector.

$$D_i = \frac{\left(\hat{Y} - \hat{Y}_{(i)}\right)^T \left(\hat{Y} - \hat{Y}_{(i)}\right)}{p\,\hat{\sigma}^2}$$

leverages = hat values

$$\text{ICBST}: \quad D_i = \frac{1}{p} r_i^2 \frac{h_i}{1-h_i}.$$

D_i: Cook statistics

r_i: standardized residuals.

D_i is a measure of "influence" that's "actual influence".

h_i is a measure of "leverage"

or "potential influence".

The 4th standard diagnostic plot obtained with plot(lmod)

is exactly standardized residuals r_i vs. leverage h_i.

With funny curved lines.

Rule of thumb: points with large Cook's distance should be investigated ($D_i > \frac{1}{2}$?)

These points are in the top-right & bottom-right corners of the plot.

One more idea: influential points can hide one another.

If we have many points which are approximately the same, the could have low individual influence but large collective influence.