Math 455 Class 24 March 26

Topics to be discussed

6.3 Added variable plots and Partial residual plots.

7.1 Errors-in-variables.

7.2 Changes of Scale

7.3 Collinearity.

We'll start with 7.2, Changes of Scale.

⚠ Question on MT2 about this.

Standard Example: we change units:

Farenheit to Celsius

mph to meters/second

etc.

~~Noti~~ Notice that these changes are of the form "$ax+b$".

This means that the effect on our regression analysis is very predictable.

There is a passage on p. 104 that explains what will happen.

One thing that we learn from p. 104:

statistical tests, p-values, R^2 DO NOT change.

Q: Why would we do this?

A: To make the model easier to understand.

Humans deal better with ~~small~~ numbers that are moderately sized. p. 103:

Better to have $\hat{\beta} = 3.51$ than

$\hat{\beta} = 0.000000351$

---

p. 104:

Rescaling $x_i \mapsto \dfrac{x_i + a}{b}$

t- & F-tests $\hat{\sigma}$, R^2 UNCHANGED.

HOWEVER : $\hat{\beta}_i$ changes to $b\hat{\beta}_i$

Notice that $(\hat{\beta}_i)_{new} \cdot (x_i)_{new}$

$$= (b\hat{\beta}_i)\left(\frac{x_i + a}{b}\right) = \hat{\beta}_i x_i + a$$

Notice $\hat{\beta}_i x_i$ is $(\hat{\beta}_i)_{old} (x_i)_{old}$.

The a will result in a change to $\hat{\beta}_0$.

So the "effect" of X_i does not change.

$$\boxed{\begin{array}{l}\text{Now: rescaling } y \longmapsto \dfrac{y+a}{b} \\ \text{the } t\text{- and } F\text{-tests are unchanged but} \\ \hat{\sigma} \text{ and the } \hat{\beta}_i \text{ will be rescaled by } b.\end{array}}$$

Notice that $\hat{\sigma}$ is measured in the same units as Y, and $\hat{\beta}_i x_i$ are measured in the same units as Y.

Also note that the $\frac{a}{b}$ part will affect $\hat{\beta}_0$.

7.3 Collinearity

A matrix may be very close to being not invertible, e.g. $\begin{pmatrix} 1 & 0 \\ 0 & 0.001 \end{pmatrix}$.

Two vectors may be very close to being linearly dependent.

Example: $(10, 10, 10, 0, 10, 10, 10)$ X_1

$(10, 10, 10, 1, 10, 10, 10)$ X_2

Difference: $(0, 0, 0, 1, 0, 0, 0)$. $= X_2 - X_1$

Having numbers like this lets you fit the 4^{th} data point in a special way with $(\text{coeff}) \cdot (X_2 - X_1)$.

If we have lots of other variables, fit the rest of the data with the other variables and then get the 4^{th} data point right with $(\text{coeff})(X_2 - X_1)$

This is likely to be a coincidental effect, due to the quirks of our data set, rather than a fundamental feature of the object of study.

We want to have a way of detecting when this kind of thing is present in our data.

We also want to know what to do about this.

3 methods of detection of collinearity.

(Collinearity means something like "almost linear dependence".)

1. Correlation matrix of the predictors.

If $Corr(X_3, X_4) = 0.99$

X_3 & X_4 are almost linearly dependent.

Issue with this: what if $2X_3 + 4X_1 - 6X_2$ and X_4 are "almost linearly dependent"?

We won't see this from the Corr matrix.

Reminder: if X_3 and X_4 <u>actually</u> <u>are</u> linearly dependent, $Corr(X_3, X_4) = 1$.

2. We can do a regression

$$X_i \sim X_1 + X_2 + \cdots + X_{i-1} + X_{i+1} + \cdots + X_p$$

We get from this an R^2,

let's call this R_i^2.

If $R_i^2 \approx 1$, then X_i is almost equal to a linear combination of the other X_is.

- Again we have a collinearity problem.

Notice that this addresses a problem with method #1. (Corr matrix of predictors)

R has an automatic method of checking for this: we don't have to do all these regressions.

Method: "Variance Inflation Factors".

the "Variance Inflation Factor" (VIF) associated to X_j is $\dfrac{1}{1-R_j^2}$

This is a factor in $Var(\hat{\beta}_j)$

There is an R command (VIF) to get this.

So you can just run vif

Ideally all vifs should be ≈ 1.

Rule of thumb: $\text{vif} > 10$ may be an issue.

Notice a problem with method 2:

(*) What if one linear comb of predictors is $\approx$ another linear combination?

Method 2 will not see this:

<u>Method 3</u>: Condition number.

Find the eigenvalues of $X^T X$.

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0.$$

$$K = \sqrt{\frac{\lambda_1}{\lambda_p}} \quad \text{is called the "condition number".}$$

Method 3: deals with problem (*).

$K \geq 1$.

Rule of thumb. $K \geq 30$ is a problem.

This means the matrix in the new basis is like $\begin{pmatrix} 30 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$