STAT6038 Week 8 Lecture Notes

xarskii

2017-04-26

1 Wednesday Lecture

1.1 Estimation and Prediction using Multiple (MR) Regression models

Estimate of Y given new values of the X variables

$$\begin{aligned}\hat{Y}|\boldsymbol{x_0} &= \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \cdots + \hat{\beta}_k x_{0k} \\ &= \boldsymbol{x}_0^T \hat{\boldsymbol{\beta}}\end{aligned}$$

$$\text{where } \boldsymbol{x_0} = \begin{pmatrix} 1 \\ x_{01} \\ x_{02} \\ x_{03} \\ \vdots \\ x_{0k} \end{pmatrix}, \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}.$$

$$\begin{aligned}\operatorname{Var}(\hat{Y}|x_0) &= \operatorname{Var}(\boldsymbol{x_0}^T \hat{\boldsymbol{\beta}}) \\ &= \boldsymbol{x_0}^T \operatorname{Var}(\hat{\boldsymbol{\beta}}) (\boldsymbol{x_0}^T)^T \\ &= \sigma^2 \boldsymbol{x_0}^T (X^T X)^{-1} \boldsymbol{x_0}\end{aligned}$$

So, a $100(1-\alpha)\%)$ confidence interval for $E[Y|\boldsymbol{x_0}]$ is $\hat{Y}|\boldsymbol{x_0} \pm t_{n-p}(1-\frac{\alpha}{2}) s \sqrt{\boldsymbol{x_0}^T (X^T X)^{-1} \boldsymbol{x_0}}$ (of SLR $\hat{Y}|x^* \pm t_{n-2}(1-\frac{\alpha}{2}) s \sqrt{\frac{(x^*-\bar{x})^2}{\text{SS}_x}}$)

And, a $100(1-\alpha)\%$ prediction interval for $Y|\boldsymbol{x_0}$ is $\hat{Y}|\boldsymbol{x_0} \pm t_{n-p}(1-\frac{\alpha}{2}) s \sqrt{1 + \boldsymbol{x_0}^T (X^T X)^{-1} \boldsymbol{x_0}}$

$\rightarrow$ again, we leave the implementation of these formulae to R and use the `predict()` function.

1.2 Problem of Multiple Comparisons

Forming a 95% interval estimate (prediction or confidence interval) is directly related to a two-sided hypothesis test.

- both types of inference are forms of comparisons.
  In forming 3 intervals, we have made 3 comparisons, all at the 95% confidence level.
  Is our overall confidence 95%?
  No, with m comparisons, it is closer to

$$\begin{aligned} P(\text{all "tests" accepted})) &= 1 - P(\text{at least one test is rejected}) \\ &\geq 1 - \sum_{i=1}^{m} P(\text{each test is rejected}) \quad \text{(relies on Boole's inequality)} \\ &= 1 - m\alpha \end{aligned}$$

  See Faraway text, pg 87.

In this instance, $m = 3$ comparisons, each at $\alpha = 0.05$.
So, our overall confidence is $\simeq 1 - 3(0.05) = 0.85$ i.e. 85%.
If we know in advance (a priori) that we are going to make $m = 3$ comparisons, we could solve $1 - m\alpha = 0.95 \implies m\alpha = 1 - 0.095 \implies m\alpha = 0.05 \implies \alpha = \frac{0.05}{m} = 0.0167$ i.e. do the 3 "tests", all at the $\alpha = 0.0167$ level of significance or $(1-\alpha)100\% = 0.9833 \times 100\%$, i.e. 98.3% confidence intervals.
This $(1 - \alpha/m)$ correction is called the *Bonferroni* (1936) correction.

2 Thursday Lecture

`predict(..., interval="prediction")` vs. `predict(..., interval="confidence")`

2.1 Residual Diagnostics

Raw residual for hte i^{th} observation

$$e_i = Y_i - \hat{Y}_i$$

$\rightarrow$ these are estimates of the errors ϵ_i i.e. $e_i = \hat{\epsilon}_i$
Note the assumptions about the errors, $\text{Var}(\epsilon) = \sigma^2 I$, but (see earlier) $\text{Var}(e) = \sigma^2(I - H)$ where H is the hat matrix.
So, the **standardized** (internally Studentised) residuals for the i^{th} observation are:

$$r_i = \frac{e_i - 0}{\sqrt{\sigma^2(1 - h_{ii}}} \simeq \frac{e_i}{\sqrt{\text{MS}_{\text{error}}}(1 - h_{ii})}$$

where h_{ii} is the hat value (leverage) of the i^{th} observation, and we use $\hat{\sigma}^2 = s^2 = \text{MS}_{\text{error}}$ to estimate the unknown σ^2.
So

$$r_i = \frac{e_i}{s\sqrt{1-h_{ii}}}$$

As σ^2 is estimated these are approximated distributed as a Student's t distribution.

2.2 Residual Plots (Ian's preferred plots)

1. Main residual plot
   Standardized (internally Studentized) residuals (r_i) against the fitted values ($\hat{Y}_i$)
   $\rightarrow$ we should check this for every model we fit.
   $\rightarrow$ why? checks the key assumptions of independence and constant variance.

2. Normal quantile plot (of the standardized residuals)
   Default `plot(model, which=2)` works fine
   $\rightarrow$ only bother checking once **plot 1** is okay
   $\rightarrow$ checks assumption of normality
   $\rightarrow$ could add 45° line for comparison (`abline(0, 1, lty=2)`)

3. Outlier/Influence plot
   My preference is a bar plot of Cook's distances.
   `plot(model, which=4)`
   $\rightarrow$ only really need if there is some indication that outliers and/or influential points might be a problem on **plots 1 and/or 2**.
   $\rightarrow$ can further investigate - check leverage values (bar plot)
   $\rightarrow$ could also use `plot(model, which=5)` (but ignore the arbitrary cut-offs for Cook's distance)
   $\rightarrow$ also, we could perform a test ...

3 Friday Lecture

3.1 Deletion Residual

Also called **PRESS** residuals, "Prediction Sum of Squares".

$$e_{i,-i} = Y_i - \hat{Y}_{i,-i}$$

where $\hat{Y}_{i,-i}$ is the fitted value for the i^{th} observation based on a model which has been fitted to the data with the i^{th} observation deleted (or excluded).

Surely this involves fitting a model (so we can calculate $e_{i,-i}$ for each $i = 1, 2, \ldots, n$)?

No, as it can be shown that

$$e_{i,-i} = \frac{e_i}{1 - h_{ii}}$$

and these deletion residuals have $\text{Var}(e_{i,-i}) = \frac{\sigma^2}{1-h_{ii}}$.

So, if we standardize the deletion residuals

$$\frac{e_{i,-i} - 0}{\sqrt{\sigma^2/(1 - h_{ii}}} \simeq \frac{e_{i,-i}}{\sqrt{s^2/(1 - h_{ii}}} = \frac{e_i}{(1 - h_{ii})s\sqrt{1/(1 - h_{ii})}} = \frac{e_i}{s\sqrt{1 - h_{ii}}} = r_i$$

Same standardized (internally Studentized) residuals as before!

But the internally Studentized residuals

$$r_i = \frac{e_i}{s\sqrt{1 - h_{ii}}}$$

are called "internally" Studentized as the estimate of σ^2 used is based on a model which uses all data (including the current or i^{th} observation).

Again, we can derive as estimate of σ^2 that excludes the current observation without to fit the entire model to a new reduced data set – it turns out

$$S_{-i} = \sqrt{\frac{(n - p)s^2 - e_i/(1 - h_{ii})}{n - p - 1}}$$

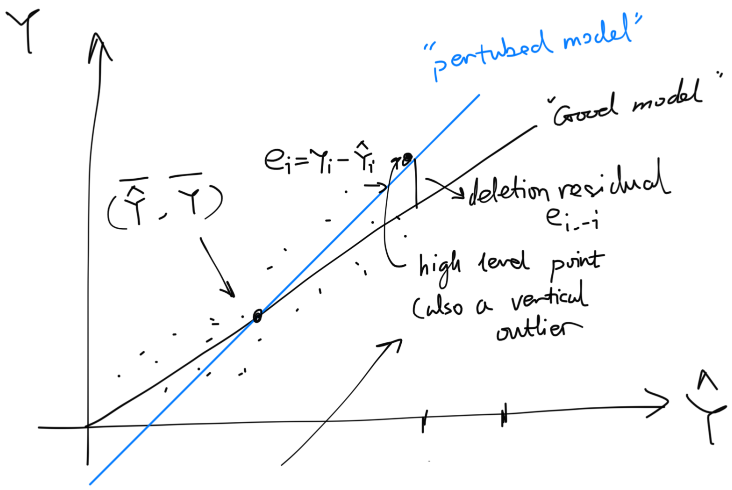
Note the new degrees of freedom used here (based on 1 less observation) as $n - p - 1$.

This gives an alternative type of standardized residuals: **the externally Studentized residuals**.

$$t_i = \frac{e_i}{s_{-i}\sqrt{1 - h_{ii}}} = \frac{e_{i,-i}}{s_{-i}/\sqrt{1 - h_{ii}}}$$

this version clearly shows that both the numerator (the deletion residual) and the denominator (a fixed function of the deletion s estimate) come from a model with the i^{th} observation excluded, hence the name "externally" Studentized residuals.

Back to the Pine example,



Example observation 20 in the full model (`pine.lm`) for the pine data.

indicator variable $I_i = \begin{cases} 0 \text{ if } i = 1, 2, \ldots, 19 \\ 1 \text{ if } i = 20 \end{cases}$

fitted model

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 I_{20}$$

if $I_{20} = 0 \implies \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$

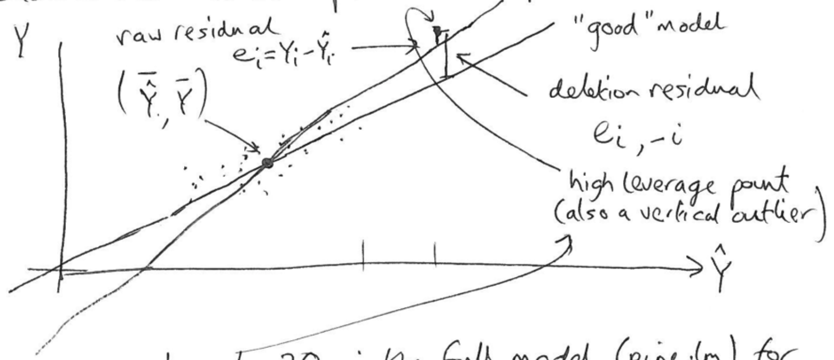
if $I_{20} = 1 \implies \hat{Y} = (\hat{\beta}_0 + \hat{\beta}_4) + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$

...

another plot

"Observation 20 failed the full model but passed the reduced model."

Back to the Pine example



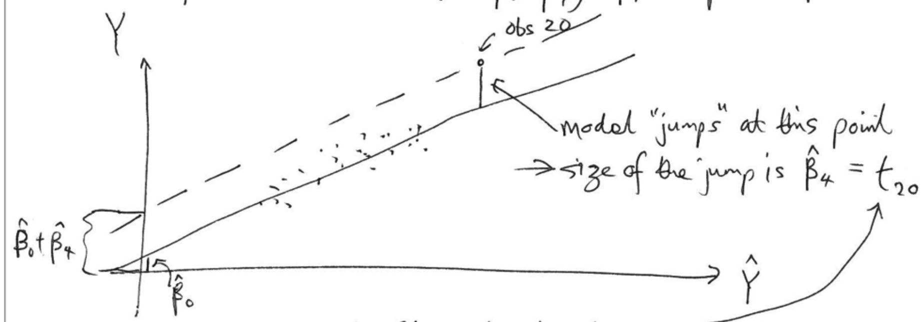
example obs 20 in the full model (pine.lm) for the pine data

indicator variable $I_{20} = \begin{cases} 0 & \text{if } i = 1, 2, \ldots 19 \\ 1 & \text{if } i = 20 \end{cases}$

fitted model $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 I_{20}$

if $I_{20} = 0 \Rightarrow \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$

if $I_{20} = 1 \Rightarrow \hat{Y} = (\hat{\beta}_0 + \hat{\beta}_4) + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$



externally Studentised residual value for observation 20