Estimation & Prediction using Multiple (MR) Regression models

Estimate of Y given new values of the X variables

$$\hat{Y}|x_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \dots \hat{\beta}_k x_{0k}$$

$$= \underset{\sim}{x}_0^T \hat{\underset{\sim}{\beta}} \quad \text{where } \underset{\sim}{x}_0 = \begin{pmatrix} 1 \\ x_{01} \\ x_{02} \\ x_{03} \\ \vdots \\ x_{0k} \end{pmatrix}, \quad \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}$$

$$Var(\hat{Y}|x_0) = Var(\underset{\sim}{x}_0^T \hat{\underset{\sim}{\beta}})$$

$$= \underset{\sim}{x}_0^T Var(\hat{\underset{\sim}{\beta}}) (\underset{\sim}{x}_0^T)^T$$

$$= \sigma^2 \underset{\sim}{x}_0^T (X^T X)^{-1} \underset{\sim}{x}_0$$

So, a $100(1-\alpha)\%$ confidence interval for $E[Y|x_0]$ is

$$\hat{Y}|x_0 \pm t_{n-p}(1-\alpha/2)\, \underset{\uparrow}{s} \sqrt{\underset{\sim}{x}_0^T (X^T X)^{-1} \underset{\sim}{x}_0}$$

(cf SLR $\hat{Y}|x^* \pm t_{n-2}(1-\alpha/2)\, s \sqrt{\frac{(x^* - \bar{x})^2}{SS_x}}$)

And, a $100(1-\alpha)\%$ prediction interval for $Y|x_0$ is

$$\hat{Y}|x_0 \pm t_{n-p}(1-\alpha/2)\, s \sqrt{1 + \underset{\sim}{x}_0^T (X^T X)^{-1} \underset{\sim}{x}_0}$$

→ again, we leave the implementation of these formulae to R and use the predict() function

Problem of Multiple Comparisons

Forming a 95% interval estimate (prediction or confidence) 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

$$
\begin{aligned}
P(\text{all "tests" accepted}) &= 1 - P(\text{at least one test is rejected}) \\
&\geq 1 - \sum_{i=1}^{m} P(\text{each test is rejected}) \\
&\quad \text{(relies on Boole's inequality)} \\
&= 1 - m\alpha \qquad \text{[see Faraway, page 87]}
\end{aligned}
$$

In this instance $m = 3$ comparisons, each at $\alpha = 0.05$

So our overall confidence is $\simeq 1 - 3(0.05) = 0.85$ ie 85%

If we know in advance (a priori) that we are going to conduct $m = 3$ comparisons we could solve $1 - m\alpha = 0.95$

$$\Rightarrow m\alpha = 1 - 0.95 \Rightarrow m\alpha = 0.05 \Rightarrow \alpha = \frac{0.05}{m} = 0.01\dot{6}$$

ie do the 3 "tests" all at the $\alpha = 0.01\dot{6}$ level of significance or $(1-\alpha)100\% = 0.98\dot{3} \times 100\%$ ie 98.3% confidence intervals

This $(1 - \alpha/m)$ correction is called the Bonferroni (1936) correction