Multiple Regression I

KNNL – Chapter 6

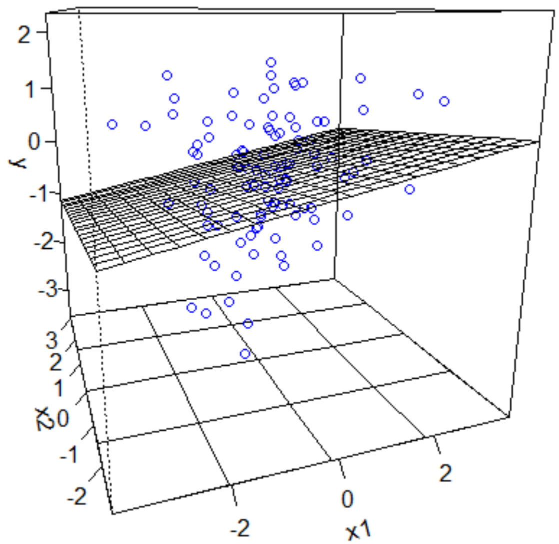
Models with Multiple Predictors

- Most practical problems have more than one potential predictor variable
- The goal is to determine effects (if any) of each predictor, controlling for the others
- Can include polynomial terms to allow for nonlinear relations
- Can include product terms to allow for interactions when effect of one variable depends on level of another variable
- Can include "dummy" variables for categorical predictors

First-Order Model with Two Numeric Predictors

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

$$E\left\{\varepsilon_i\right\} = 0 \quad \Rightarrow \quad E\left\{Y\right\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad \text{Plane in 3-dimensions}$$

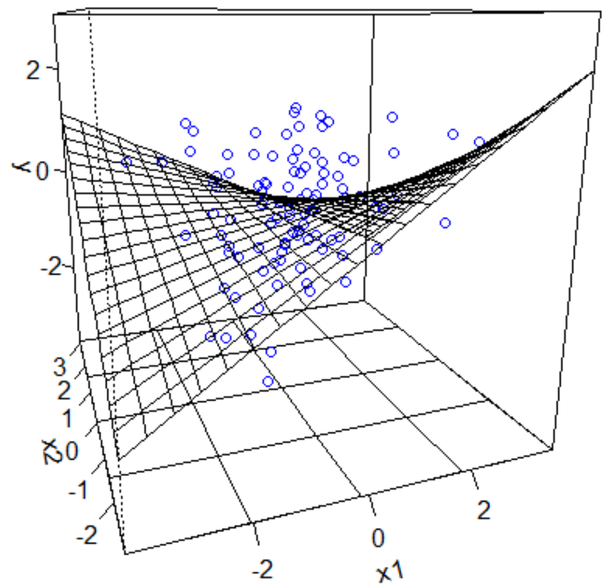


Interpretation of Regression Coefficients

- Additive: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \equiv$ Mean of Y @ X_1, X_2
- $\beta_0 \equiv$ Intercept, Mean of Y when $X_1 = X_2 = 0$
- $\beta_1 \equiv$ Slope with Respect to X_1 (effect of increasing X_1 by 1 unit, while holding X_2 constant)
- $\beta_2 \equiv$ Slope with Respect to X_2 (effect of increasing X_2 by 1 unit, while holding X_1 constant)
- These can also be obtained by taking the partial derivatives of E{Y} with respect to X_1 and X_2, respectively
- Interaction Model: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$
- When $X_2 = 0$: Effect of increasing X_1 by 1: $\beta_1(1)+\beta_3(1)(0)= \beta_1$
- When $X_2 = 1$: Effect of increasing X_1 by 1: $\beta_1(1)+\beta_3(1)(1)= \beta_1+\beta_3$
- The effect of increasing X_1 depends on level of X_2, and vice versa

Interaction Model with Two Numeric Predictors

$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$
The geometric surface is called a "saddle"



General Linear Regression Model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

$$\Rightarrow \quad Y_i = \beta_0 + \sum_{k=1}^{p-1} \beta_k X_{ik} + \varepsilon_i$$

$$\Rightarrow \quad Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \qquad \text{where: } X_{i0} \equiv 1$$

$$E\{\varepsilon_i\} = 0 \quad \Rightarrow \quad E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{p-1} X_{p-1} \quad \text{(Hyperplane in } p\text{-dimensions)}$$

$$p - 1 = 1 \quad \Rightarrow \quad \text{Simple linear regression}$$

Normality, independence, and constant variance for errors:

$$\varepsilon_i \sim NID\left(0, \sigma^2\right) \quad \Rightarrow \quad Y_i \sim N\left(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1}, \sigma^2\right) \quad \sigma\left\{Y_i, Y_j\right\} = 0 \quad \forall\, i \neq j$$

Special Types of Variables/Models - I

- *p*-1 distinct numeric predictors (attributes)
  - Y = Sales, X_1=Advertising, X_2=Price

- Categorical Predictors – Indicator (Dummy) variables, representing *m*-1 levels of a m level categorical variable
  - Y = Salary, X_1=Experience, X_2=1 if College Grad, 0 if Not

- Polynomial Terms – Allow for bends in the Regression
  - Y=MPG, X_1=Speed, X_2=Speed^2

- Transformed Variables – Transformed Y variable to achieve linearity $Y'=\ln(Y)$ $Y'=1/Y$

Special Types of Variables/Models - II

- Interaction Effects – Effect of one predictor depends on levels of other predictors
  - Y = Salary, X_1=Experience, X_2=1 if Coll Grad, 0 if Not, $X_3=X_1X_2$
  - $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$
  - Non-College Grads ($X_2 = 0$):
  - $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2(0) + \beta_3 X_1(0) = \beta_0 + \beta_1 X_1$
  - College Grads ($X_2 = 1$):
  - $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2(1) + \beta_3 X_1(1) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1$
- Response Surface Models
  - $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2^2 + \beta_5 X_1 X_2$
- Note: Although the Response Surface Model has polynomial terms, it is linear wrt Regression parameters

Matrix Form of Regression Model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad i = 1,\ldots,n$$

Matrix Form:

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{n\times p}{\mathbf{X}} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}$$

$$\underset{p\times 1}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \qquad \underset{n\times 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad \mathbf{E}\left\{\underset{\mathbf{n\times 1}}{\varepsilon}\right\} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \sigma^{\mathbf{2}}\left\{\underset{n\times 1}{\varepsilon}\right\} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}$$

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\, \underset{p\times 1}{\beta} + \underset{n\times 1}{\varepsilon} \quad \Rightarrow \quad \mathbf{E}\left\{\underset{\mathbf{n\times 1}}{\mathbf{Y}}\right\} = \mathbf{E}\left\{\underset{\mathbf{n\times p}}{\mathbf{X}}\, \underset{\mathbf{p\times 1}}{\beta} + \underset{\mathbf{n\times 1}}{\varepsilon}\right\} = \underset{n\times p}{\mathbf{X}}\, \underset{p\times 1}{\beta} \qquad \sigma^{\mathbf{2}}\left\{\underset{n\times 1}{\mathbf{Y}}\right\} = \sigma^2 \mathbf{I}$$

Least Squares Estimation of Regression Coefficients

Goal: Minimize: $Q = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_{p-1} X_{i,p-1} \right)^2$

$\Rightarrow$ Obtain Estimates of $\beta_0, \beta_1, \ldots, \beta_{p-1}$ that minimize $Q \quad \Rightarrow \quad b_0, b_1, \ldots, b_{p-1}$

Normal Equations: $\underset{p\times p}{\mathbf{X'X}} \underset{p\times 1}{\mathbf{b}} = \underset{p\times 1}{\mathbf{X'Y}} \quad \Rightarrow \quad \underset{p\times 1}{\mathbf{b}} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix} = \left(\mathbf{X'X} \right)^{\mathbf{-1}} \mathbf{X'Y}$

Maximum Likelihood also leads to the same estimator $\mathbf{b}$:

$$L\left(\beta, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_{p-1} X_{i,p-1} \right)^2 \right]$$

since maximizing L involves minimizing $\sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_{p-1} X_{i,p-1} \right)^2$

Fitted Values and Residuals

Fitted Values: $\underset{n\times 1}{\hat{\mathbf{Y}}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}$ Residuals: $\underset{n\times 1}{\mathbf{e}} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times p}{\mathbf{X}}\,\underset{p\times 1}{\mathbf{b}} = \mathbf{X}\left(\mathbf{X'X}\right)^{\mathbf{-1}}\mathbf{X'Y} = \mathbf{HY} \qquad \mathbf{H} = \mathbf{X}\left(\mathbf{X'X}\right)^{\mathbf{-1}}\mathbf{X'} \quad \mathbf{H} = \mathbf{H'} = \mathbf{HH}$$

$$\underset{n\times 1}{\mathbf{e}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times p}{\mathbf{X}}\,\underset{p\times 1}{\mathbf{b}} = \mathbf{Y} - \mathbf{X}\left(\mathbf{X'X}\right)^{\mathbf{-1}}\mathbf{X'Y} = \left(\mathbf{I}-\mathbf{H}\right)\mathbf{Y} \qquad \left(\mathbf{I}-\mathbf{H}\right) = \left(\mathbf{I}-\mathbf{H}\right)' = \left(\mathbf{I}-\mathbf{H}\right)\left(\mathbf{I}-\mathbf{H}\right)$$

$$\sigma^2\left\{\hat{\mathbf{Y}}\right\} = \sigma^2\left\{\mathbf{HY}\right\} = \mathbf{H}\sigma^2\left\{\mathbf{Y}\right\}\mathbf{H'} = \sigma^2\mathbf{H} \qquad \mathbf{s}^2\left\{\hat{\mathbf{Y}}\right\} = MSE\left(\mathbf{H}\right)$$

$$\sigma^2\left\{\mathbf{e}\right\} = \sigma^2\left\{\left(\mathbf{I}-\mathbf{H}\right)\mathbf{Y}\right\} = \left(\mathbf{I}-\mathbf{H}\right)\sigma^2\left\{\mathbf{Y}\right\}\left(\mathbf{I}-\mathbf{H}\right)' = \sigma^2\left(\mathbf{I}-\mathbf{H}\right) \qquad \mathbf{s}^2\left\{\mathbf{e}\right\} = MSE\left(\mathbf{I}-\mathbf{H}\right)$$

Inferences Regarding Regression Parameters

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \qquad \mathbf{E}\{\boldsymbol{\varepsilon}\} = \mathbf{0} \qquad \boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\} = \sigma^2\mathbf{I} \quad \Rightarrow \quad \mathbf{E}\{\mathbf{Y}\} = \mathbf{X\beta} \qquad \boldsymbol{\sigma}^2\{\mathbf{Y}\} = \sigma^2\mathbf{I}$$

$$\mathbf{E}\{\mathbf{b}\} = \mathbf{E}\left\{(\mathbf{X'X})^{-1}\mathbf{X'Y}\right\} = (\mathbf{X'X})^{-1}\mathbf{X'E}\{\mathbf{Y}\} = (\mathbf{X'X})^{-1}\mathbf{X'X\beta} = \boldsymbol{\beta}$$

$$\boldsymbol{\sigma}^2\{\mathbf{b}\} = \begin{bmatrix} \sigma^2\{b_0\} & \sigma\{b_0,b_1\} & \cdots & \sigma\{b_0,b_{p-1}\} \\ \sigma\{b_1,b_0\} & \sigma^2\{b_1\} & \cdots & \sigma\{b_1,b_{p-1}\} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma\{b_{p-1},b_0\} & \sigma\{b_{p-1},b_1\} & \cdots & \sigma^2\{b_{p-1}\} \end{bmatrix} \qquad \mathbf{s}^2\{\mathbf{b}\} = \begin{bmatrix} s^2\{b_0\} & s\{b_0,b_1\} & \cdots & s\{b_0,b_{p-1}\} \\ s\{b_1,b_0\} & s^2\{b_1\} & \cdots & s\{b_1,b_{p-1}\} \\ \vdots & \vdots & \ddots & \vdots \\ s\{b_{p-1},b_0\} & s\{b_{p-1},b_1\} & \cdots & s^2\{b_{p-1}\} \end{bmatrix}$$

$$\boldsymbol{\sigma}^2\{\mathbf{b}\} = \boldsymbol{\sigma}^2\left\{(\mathbf{X'X})^{-1}\mathbf{X'Y}\right\} = (\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\sigma}^2\{\mathbf{Y}\}\left((\mathbf{X'X})^{-1}\mathbf{X'}\right)' = \sigma^2(\mathbf{X'X})^{-1}\mathbf{X'X}(\mathbf{X'X})^{-1} = \sigma^2(\mathbf{X'X})^{-1}$$

$$\mathbf{s}^2\{\mathbf{b}\} = MSE(\mathbf{X'X})^{-1}$$

$$\frac{b_k - \beta_k}{s\{b_k\}} \sim t_{n-p} \qquad (1-\alpha)100\% \text{ CI for } \beta_k \equiv b_k \pm t\left(1-\frac{\alpha}{2}; n-p\right)s\{b_k\}$$

Test of $H_0: \beta_k = 0 \quad H_A: \beta_k \neq 0$ Test Statistic: $t^* = \dfrac{b_k}{s\{b_k\}}$

Rejection Region: $|t^*| \geq t\left(1-\dfrac{\alpha}{2}; n-p\right)$ P-value=$2\Pr\left(t(n-p) \geq |t^*|\right)$

Simultaneous $(1-\alpha)100\%$ CI^s for $g \leq p$ β^s: $b_k \pm t\left(1-\dfrac{\alpha}{2g}; n-p\right)s\{b_k\}$

Analysis of Variance – Sums of Squares

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{n\times 1}{\hat{\mathbf{Y}}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \mathbf{Xb} = \mathbf{HY} \qquad \underset{n\times 1}{\overline{\mathbf{Y}}} = \begin{bmatrix} \overline{Y} \\ \overline{Y} \\ \vdots \\ \overline{Y} \end{bmatrix} = \frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \left(\frac{1}{n}\right)\mathbf{JY}$$

$$SSTO = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \left(\mathbf{Y}-\overline{\mathbf{Y}}\right)'\left(\mathbf{Y}-\overline{\mathbf{Y}}\right) = \mathbf{Y}'\left(\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{Y}$$

$$SSE = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \left(\mathbf{Y}-\hat{\mathbf{Y}}\right)'\left(\mathbf{Y}-\hat{\mathbf{Y}}\right) = \mathbf{Y}'\left(\mathbf{I}-\mathbf{H}\right)\mathbf{Y} = \mathbf{Y'Y} - \mathbf{b'X'Y} \qquad MSE = \frac{SSE}{n-p}$$

$$SSR = \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2 = \left(\hat{\mathbf{Y}}-\overline{\mathbf{Y}}\right)'\left(\hat{\mathbf{Y}}-\overline{\mathbf{Y}}\right) = \mathbf{Y}'\left(\mathbf{H}-\left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{Y} = \mathbf{b'X'Y} - \mathbf{Y}'\left(\frac{1}{n}\right)\mathbf{JY} \qquad MSR = \frac{SSR}{p-1}$$

$$E\left\{MSE\right\} = \sigma^2$$

$$E\left\{MSR\right\} = \sigma^2 + \sum_{k=1}^{p-1}\beta_k^2 SS_{kk} + \sum_{k=1}^{p-1}\sum_{k'\neq k}\beta_k\beta_{k'}SS_{kk'} \qquad SS_{kk'} = \sum_{i=1}^{n}\left(X_{ik}-\overline{X}_k\right)\left(X_{ik'}-\overline{X}_{k'}\right)$$

$$E\left\{MSR\right\} \geq E\left\{MSE\right\} \qquad E\left\{MSR\right\} = E\left\{MSE\right\} \Leftrightarrow \beta_1 = \ldots = \beta_{p-1} = 0$$

ANOVA Table, F-test, and R^2

Analysis of Variance (ANOVA) Table

Source	df	Sum of Squares	Mean Square
Regression	$p-1$	$SSR = \mathbf{b'X'Y} - \mathbf{Y'}\left(\frac{1}{n}\right)\mathbf{JY}$	$MSR = \frac{SSR}{p-1}$
Error	$n-p$	$SSE = \mathbf{Y'Y} - \mathbf{b'X'Y}$	$MSE = \frac{SSE}{n-p}$
Total	$n-1$	$SSTO = \mathbf{Y'Y} - \mathbf{Y'}\left(\frac{1}{n}\right)\mathbf{JY}$	

Test of $H_0: \beta_1 = \ldots = \beta_{p-1} = 0 \quad \left(E(Y) = \beta_0\right) \qquad H_A:$ Not all $\beta_k = 0$

Test Statistic: $F^* = \frac{MSR}{MSE}$ Rejection Region: $F^* \geq F\left(1-\alpha; p-1, n-p\right)$ P-value$=\Pr\left\{F\left(p-1, n-p\right) \geq F^*\right\}$

Coefficient of Multiple Determination: $R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$ Correlation: $R = \sqrt{R^2}$

Adjusted-$R^2 = 1 - \frac{\left[\frac{SSE}{n-p}\right]}{\left[\frac{SSTO}{n-1}\right]} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SSTO}$ places a "penalty" on models with extra predictors

Estimating Mean Response at Specific *X*-levels

Denote the new given values of X_1, …, $X_{p\text{-}1}$ by X_{h1}, …, $X_{h,p\text{-}1}$. Define the vector:

$$\boldsymbol{X}_h = \begin{bmatrix} 1 \\ X_{h1} \\ \vdots \\ X_{h,p-1} \end{bmatrix}$$

Then the estimated mean response corresponding to $\boldsymbol{X}_h$ is:

$$\hat{Y}_h = \boldsymbol{X}_h'\boldsymbol{b}$$

and its variance is:

$$\sigma^2(\hat{Y}_h) = \sigma^2\boldsymbol{X}_h'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_h$$

estimated by:

$$s^2(\hat{Y}_h) = \mathrm{MSE}\cdot\boldsymbol{X}_h'\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_h$$

which is used in the CI formula.

Predicting New Response(s) at Specific X-levels

Given set of levels of $X_1, \ldots, X_{p-1}$: $\quad X_{h1}, \ldots, X_{h,p-1}$

$$\underset{p\times 1}{\mathbf{X}_h} = \begin{bmatrix} 1 \\ X_{h1} \\ \vdots \\ X_{h,p-1} \end{bmatrix} \qquad E\{Y_h\} = \mathbf{X}_h'\boldsymbol{\beta} \qquad \hat{Y}_h = \mathbf{X}_h'\mathbf{b}$$

$$s^2\{\text{pred}\} = MSE\left(1 + \mathbf{X}_h'\left(\mathbf{X'X}\right)^{-1}\mathbf{X}_h\right)$$

mean of m observations (at same X-levels): $s^2\{\text{predmean}\} = MSE\left(\frac{1}{m} + \mathbf{X}_h'\left(\mathbf{X'X}\right)^{-1}\mathbf{X}_h\right)$

$(1-\alpha)100\%$ CI for $Y_{h(\text{new})}$: $\quad \hat{Y}_h \pm t\left(1-\frac{\alpha}{2}; n-p\right)s\{\text{pred}\}$

Scheffe: $(1-\alpha)100\%$ CI for several (g) $Y_{h(\text{new})}$: $\quad \hat{Y}_h \pm Ss\{\text{pred}\} \qquad S^2 = gF(1-\alpha; g, n-p)$

Bonferroni: $(1-\alpha)100\%$ CI for several (g) $Y_{h(\text{new})}$: $\quad \hat{Y}_h \pm Bs\{\text{pred}\} \qquad B = t\left(1-\frac{\alpha}{2g}; n-p\right)$