Prediction in Multiple Linear Regression (MLR)

Prediction of the response variable

1. The $1-\alpha$ prediction limits for mean response $\boldsymbol{E\{Y_h\}}$ corresponding to X_h are

$$\hat{Y}_h \pm t(1-\alpha/2\,; n-p)s\{\hat{Y}_h\} \qquad s^2\{\hat{Y}_h\}=X_h'\Sigma\{b\}X_h$$

Where $\Sigma\{b\} = MSE(X'X)^{-1}$

2. The $1-\alpha$ prediction limits for single response $\boldsymbol{Y_{h(new)}}$ corresponding to X_h are

$$\hat{Y}_h \pm t(1-\alpha/2; n-p)s\{\text{pred}\} \qquad s^2\{\text{pred}\} = MSE + s^2\{\hat{Y}_h\} = MSE(1+\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h)$$

3. The $1-\alpha$ prediction limits for means of m new responses $\boldsymbol{at}$ X_h are

$$\hat{Y}_h \pm t(1-\alpha/2; n-p)s\{\text{predmean}\} \qquad s^2\{\text{predmean}\} = \frac{MSE}{m} + s^2\{\hat{Y}_h\} = MSE\left(\frac{1}{m} + \mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h\right)$$

4. The simultaneous $1-\alpha$ prediction limits for g mean observations $\boldsymbol{at}$ X_h (Bonferroni procedure) are

$$\hat{Y}_h \pm Bs\{\text{pred}\} \qquad B = t(1-\alpha/2g; n-p)$$

The Dwaine Studios example: The Dwaine operates studios that specialize in portraits of children. The company is considering whether sales (Y) in a community can be predicted from the number of persons aged 16 or younger in the community (X1) and the per capita disposable personal income in the community (X2). Data is Dwaine.csv **n=21**

1). Estimate the 95% CI for the mean response when X1=65.4 and X2=17.6

The $1-\alpha$ confidence limits for $E\{Y_h\}$ are:

$$\hat{Y}_h \pm t(1-\alpha/2; n-p)s\{\hat{Y}_h\}$$

$dfE =$ =18 $t(0.975,\ dfE) = 2.101$

$s^2\{\hat{Y}_h\}$=$X_h'\Sigma\{b\}X_h =$ $= 7.656$

The CI, $\hat{Y}_h \pm ts\{\hat{Y}_h\} =$

$= (185.3, 196.9)$

We are 95% confident that the average sale will be between 185 and 197 when the population is 65.4 unit and personal income is 17.6 unit.

```
ci.reg(dwa.mod, new, type='m',alpha=0.05)
```

The Dwaine Studios example: The Dwaine operates studios that specialize in portraits of children. The company is considering whether sales (Y) in a community can be predicted from the number of persons aged 16 or younger in the community (X1) and the per capita disposable personal income in the community (X2). Data is Dwaine.csv **n=21**

2). Estimate the 95% CI for single resposne $\boldsymbol{at}$ X_h =(65.4, 17.6) are

$s^2\{pred\}$=MSE+$X_h'\Sigma\{b\}X_h =$

$$= 128.82$$

The CI, $\hat{Y}_h \pm ts\{\hat{Y}_h\} =$

$$= (167.3, 214.9)$$

```
ci.reg(dwa.mod, new, type='n',alpha=0.05)
```

The Dwaine Studios example: The Dwaine operates studios that specialize in portraits of children. The company is considering whether sales (Y) in a community can be predicted from the number of persons aged 16 or younger in the community (X1) and the per capita disposable personal income in the community (X2). Data is Dwaine.csv **n=21**

3). Estimate the 95% CI for the mean of m(e.g, 2) new observations at the same X_h =(65.4, 17.6) are

$$s^2\{\text{predmean}\} = \frac{MSE}{m} + s^2\{\hat{Y}_h\} = MSE\left(\frac{1}{m} + \mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h\right)$$

=68.24

The CI, $\hat{Y}_h \pm ts\{predmean\} =$

$$= (173.75\,, 208.46)$$

```
ci.reg(dwa.mod, new, type='nm', m=2, alpha=0.05)
```

The Dwaine Studios example: The Dwaine operates studios that specialize in portraits of children. The company is considering whether sales (Y) in a community can be predicted from the number of persons aged 16 or younger in the community (X1) and the per capita disposable personal income in the community (X2). Data is Dwaine.csv **n=21**

A supplement question to 4). Estimate the 95% simultaneous CI for predicting mean responses when (X1, X2)=(65.4, 17.6) and (66, 20)

$s^2\{\hat{Y}_h\}=X_h'\Sigma\{b\}X_h$ Where $B = t\left(1-\frac{0.05}{2g};18\right) = 2.445$

$$X_h'\Sigma\{b\}X_h = \begin{pmatrix} 1 & 65.4 & 17.6 \\ 1 & 66.0 & 20.0 \end{pmatrix} \begin{pmatrix} 3602.03467 & 8.74593958 & -241.4229923 \\ 8.74594 & 0.04485151 & -0.6724426 \\ -241.42299 & -0.67244260 & 16.5157558 \end{pmatrix} \begin{pmatrix} 1.0 & 1 \\ 65.4 & 66 \\ 17.6 & 20 \end{pmatrix} = \begin{pmatrix} 7.65517 & 20.22547 \\ 20.22547 & 126.00603 \end{pmatrix}$$

$s^2\{\hat{Y}_h\} = X_h'\Sigma\{b\}X_h = 7.655$ when (X1, X2)=(65.4, 17.6)

$s^2\{\hat{Y}_h\} = X_h'\Sigma\{b\}X_h =$ 126.006 when (X1, X2)=(66, 20)

The simultaneous CI, $\hat{Y}_h \pm ts\{\hat{Y}_h\}$ = $190.0 \pm 2.445 * \sqrt{7.655} = (184.13 \quad 197.66\,)$

$\hat{Y}_h \pm ts\{\hat{Y}_h\} = 214.45 \pm 2.445 * \sqrt{126.006} = (\,187, \qquad 241.9)$

```
ci.reg(dwa.mod, new, type='gn',alpha=0.05)
```

The Dwaine Studios example: The Dwaine operates studios that specialize in portraits of children. The company is considering whether sales (Y) in a community can be predicted from the number of persons aged 16 or younger in the community (X1) and the per capita disposable personal income in the community (X2). Data is Dwaine.csv **n=21**

4). Estimate the 95% simultaneous CI for predicting single responses when (X1, X2)=(65.4, 17.6) and (66, 20)

$$\hat{Y}_h \pm Bs\{\text{pred}\}$$ Where $B = t\left(1 - \frac{0.05}{2g}; 18\right) = 2.445$

$s^2\{pred\}$=MSE+$X_h'\Sigma\{b\}X_h$ $=$ $121.1626 + 7.656 = 128.82$ when (X1, X2)=(65.4, 17.6)

$s^2\{pred\}$=MSE+$X_h'\Sigma\{b\}X_h$ $=$ $= 247.16$ when (X1, X2)=(66, 20)

$$X_h'\Sigma\{b\}X_h = \begin{pmatrix} \quad \end{pmatrix} \begin{pmatrix} 3602.03467 & 8.74593958 & -241.4229923 \\ 8.74594 & 0.04485151 & -0.6724426 \\ -241.42299 & -0.67244260 & 16.5157558 \end{pmatrix} \begin{pmatrix} 1.0 & 1 \\ 65.4 & 66 \\ 17.6 & 20 \end{pmatrix} = \begin{pmatrix} 7.65517 & 20.22547 \\ 20.22547 & 126.00603 \end{pmatrix}$$

The simultaneous CI, $\hat{Y}_h \pm ts\{\hat{Y}_h\}$ = $= (163.4, 218.9)$

$\hat{Y}_h \pm ts\{\hat{Y}_h\}$ = $= (176, 252.9)$

```
ci.reg(dwa.mod, new, type='gn',alpha=0.05)
```

Simultaneous confidence intervals for g mean response, at different X_h levels

1. Use the Working-Hotelling method

$$\hat{Y}_h \pm Ws\{\hat{Y}_h\}$$

Where $W^2 = pF(1-\alpha; p, n-p)$

2. Use the Bonferroni method

$$\hat{Y}_h \pm Bs\{\hat{Y}_h\}$$

where:

$$B = t(1-\alpha/2g; n-p)$$