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

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

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 new observations $\boldsymbol{at}$ X_h (Bonferroni procedure) are

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

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

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 predicting mean Y when X1=65.4 and X2=17.6

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

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

$dfE = n - p =$ 21-3=18 $\quad t\,(0.975; 18) = 2.101$

$$s^2\{\hat{Y}_h\} = X_h' s^2\{b\} X_h = \begin{pmatrix} 1 & 65.4 & 17.6 \end{pmatrix} \begin{pmatrix} 3602.03467 & 8.74593958 & -241.4229923 \\ 8.74594 & 0.04485151 & -0.6724426 \\ -241.42299 & -0.67244260 & 16.5157558 \end{pmatrix} \begin{pmatrix} 1.0 \\ 65.4 \\ 17.6 \end{pmatrix} = 7.656$$

The CI, $\hat{Y}_h \pm ts\{\hat{Y}_h\}$ =191.1 $\pm\ 2.101\sqrt{7.656} = (185.3, 196.9)$

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). The $1-\alpha$ prediction limits for one new observations ***at*** X_h =(65.4, 17.6) are

$$\hat{Y}_h \pm t(1-\alpha/2; n-p)s\{\text{pred}\} \qquad where\ t\ (0.975; 18) = 2.101$$

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

The CI, $\hat{Y}_h \pm ts\{\hat{Y}_h\}$ =191.1 ±2.101 $\left(\sqrt{128.82}\ \right) = (167.3, 214.9)$

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

………..that the next sale will be between 167 and 215 ………..

3). The $1-\alpha$ prediction limits for m(e.g, 2) new observations at the same X_h =(65.4, 17.6) are

$$\hat{Y}_h \pm t(1-\alpha/2; n-p)s\{\text{predmean}\} \qquad where\ t\ (0.975; 18) = 2.101$$

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

$= MSE\left(\frac{1}{2}\right) + 7.656$ = 121.1626/2+7.656 =68.24

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

The CI, $\hat{Y}_h \pm ts\{predmean\}$ =191.1 ±2.101 $\left(\sqrt{68.24}\ \right) = (173.75\,, 208.46)$

………..that the average of next 2 sales will be between 174 and 208 ………..

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

$s^2\{pred\}$=MSE+$X_h' s^2\{b\} X_h$ $=$ $121.1626\ + 126.006\ = 247.16$ when (X1, X2)=(66, 20)

$$X_h' s^2\{b\} X_h = \begin{pmatrix} 1 & 65.4 & 17.6 \\ 1 & 66.0 & 20.0 \end{pmatrix} \begin{pmatrix} 3602.03467 & 8.74593958 & -241.4229923 \\ 8.74594 & 0.04485151 & -0.6724426 \\ -241.42299 & -0.67244260 & 16.5157558 \end{pmatrix} \begin{pmatrix} 1.0 & 1 \\ 65.4 & 66 \\ 17.6 & 20 \end{pmatrix} = \begin{pmatrix} 7.65517 & 20.22547 \\ 20.22547 & 126.00603 \end{pmatrix}$$

The simultaneous CI, $\hat{Y}_h \pm ts\{\hat{Y}_h\}$ = $191.1 \pm 2.445\left(\sqrt{128.82}\right) = (163.4, 218.9)$

$\hat{Y}_h \pm ts\{\hat{Y}_h\}$ = $214.45 \pm 2.445\left(\sqrt{247.16}\right) = (176, 252.9)$

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

……….that the next two sales will be between 163 and 219 when (X1, X2)=(65.4, 17.6);
and between 176 and 253 when (X1, X2)=(66, 20)

Simultaneous confidence intervals for several mean response

1. Use the Working-Hotelling confidence region bounds for several X_h vectors of interest:

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

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

2. Use Bonferroni simultaneous confidence region intervals for g interval estimates:

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

where:

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