STAT 666 HW #5

Maggie Buffum

June 2, 2019

Problem 8.2

In the normal linear regression model $Y_i = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_k X_{i,k} + \epsilon_i$, $i = 1, \ldots, N$, obtain a test statistic for $H : \beta_j = d$ where $0 \le j \le k$ and d is a constant.

We can rewrite the above hypothesis as $H : \mathbf{C}'\beta = \mathbf{d}$, where $\mathbf{C}'$ is a row vector with k elements that are each 0 except for the jth element, which is a one.

We know that

$$F(H) = \frac{Q/s}{SSE/(N-r)} \sim F_{(s,N-r)}$$

where r is the rank of $\mathbf{X}$ ($k-1$ in this case), s is the rank of $\mathbf{C}$ (1), $SSE = \mathbf{y}'(\mathbf{I} - \mathbf{P}_X)\mathbf{y}$, and $Q = (\hat{\beta}_j - d)'[(\mathbf{X}'\mathbf{X})^{-1}]^{-1}(\beta_j^0 - d)$. We need to find Q. Knowing that $\mathbf{C}$ has only one nonzero element, $\mathbf{C}'\mathbf{G}\mathbf{C}$ is just the jjth element of the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix; let's call this element g_{jj}. Therefore,

$$Q = \frac{(\hat{\beta}_j - d)^2/s}{g_{jj}SSE/(N-r)}$$

and

$$F(H) = \frac{(\hat{\beta}_j - d)^2}{g_{jj}SSE/(N-k-1)} \sim F_{(1,N-k-1)}$$

Problem 8.6

Consider the independent regressions

$$Y_{k,i} = \alpha_k + \beta(X_{k,i} - \bar{X}_k) + \epsilon_{k,i},\ \ i = 1, \ldots, n_k,\ \ k = 1, 2,$$

were $\epsilon_{k,i}$ are iid $N(0, \sigma^2)$ variables, and $\bar{X}_k = \sum_{i=1}^{n_k} X_{k,i}/n_k$.

(a) Estimate α_1, α_2 and β, and thus the vertical distance between the two lines, measured parallel to the y-axis.

The given model suggests the following $\mathbf{X}$ matrix:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & X_{1,1} - \bar{X}_{1.} \\ \vdots & \vdots & \vdots \\ 1 & 0 & X_{1,n_1} - \bar{X}_{1.} \\ 0 & 1 & X_{2,1} - \bar{X}_{2.} \\ \vdots & \vdots & \vdots \\ 0 & 1 & X_{2,1} - \bar{X}_{2.} \end{bmatrix}$$

We can estimate $\beta' = \begin{bmatrix}\alpha_1 & \alpha_2 & \beta\end{bmatrix}$ as $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. First let's find $(\mathbf{X}'\mathbf{X})^{-1}$:

$$
\begin{aligned}
(\mathbf{X}'\mathbf{X})^{-1} &= \left(\begin{bmatrix} n_1 & 0 & \sum_{i=1}^{n_1}(X_{1,i}-\bar{X}_1) \\ 0 & n_2 & \sum_{i=1}^{n_2}(X_{2,i}-\bar{X}_2) \\ \sum_{i=1}^{n_1}(X_{1,i}-\bar{X}_1) & \sum_{i=1}^{n_2}(X_{2,i}-\bar{X}_2) & \sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i}-\bar{X}_k)^2 \end{bmatrix}\right)^{-1} \\
&= \left(\begin{bmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & \sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i}-\bar{X}_k)^2 \end{bmatrix}\right)^{-1} \\
&= \begin{bmatrix} \frac{1}{n_1} & 0 & 0 \\ 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & \frac{1}{\sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i}-\bar{X}_k)^2} \end{bmatrix}
\end{aligned}
$$

Now,

$$
\begin{aligned}
(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' &= \begin{bmatrix} \frac{1}{n_1} & 0 & 0 \\ 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & \frac{1}{\sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i}-\bar{X}_k)^2} \end{bmatrix}\begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 \\ X_{1,1}-\bar{X}_1 & \dots & X_{1,n_1}-\bar{X}_1 & X_{2,1}-\bar{X}_2 & \dots & X_{2,n_2}-\bar{X}_2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n_1} & \dots & \frac{1}{n_1} & 0 & \dots & 0 \\ 0 & \dots & 0 & \frac{1}{n_2} & \dots & \frac{1}{n_2} \\ \frac{X_{1,1}-\bar{X}_1}{\sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i}-\bar{X}_k)^2} & \dots & \frac{X_{1,n_1}-\bar{X}_1}{\sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i}-\bar{X}_k)^2} & \frac{X_{2,1}-\bar{X}_2}{\sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i}-\bar{X}_k)^2} & \dots & \frac{X_{2,n_2}-\bar{X}_2}{\sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i}-\bar{X}_k)^2} \end{bmatrix}
\end{aligned}
$$

and finally,

$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

$$
\begin{aligned}
&= \begin{bmatrix} \frac{1}{n_1} & \dots & \frac{1}{n_1} & 0 & \dots & 0 \\ 0 & \dots & 0 & \frac{1}{n_2} & \dots & \frac{1}{n_2} \\ \frac{X_{1,1}-\bar{X}_1}{\sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i}-\bar{X}_k)^2} & \dots & \frac{X_{1,n_1}-\bar{X}_1}{\sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i}-\bar{X}_k)^2} & \frac{X_{2,1}-\bar{X}_2}{\sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i}-\bar{X}_k)^2} & \dots & \frac{X_{2,n_2}-\bar{X}_2}{\sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i}-\bar{X}_k)^2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n_1}\sum_{i=1}^{n_1} y_i \\ \frac{1}{n_2}\sum_{i=1}^{n_2} y_i \\ \frac{\sum_{i=1}^{n_1}(X_{1,i}-\bar{X}_1)y_i+\sum_{i=1}^{n_2}(X_{2,i}-\bar{X}_2)y_i}{\sum_{i=1}^{n_1}(X_{1,i}-\bar{X}_1)^2+\sum_{i=1}^{n_2}(X_{2,i}-\bar{X}_2)^2} \end{bmatrix} \\
&= \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \frac{S_{XY,1}+S_{XY,2}}{S_{XX,1}=S_{XX,2}} \end{bmatrix}
\end{aligned}
$$

The two lines are obviously parallel. The perpendicular distance between the two lines is given by $D = Y_1 - Y_2$, where $Y_1 = \alpha_1 + \beta(X - \bar{X}_1)$ and $Y_2 = \alpha_2 + \beta(X - \bar{X}_2)$ such that

$$
D = (\alpha_1 + \beta(X - \bar{X}_1)) - (\alpha_2 + \beta(X - \bar{X}_2)) = (\alpha_1 - \alpha_2) + \beta(\bar{X}_1 - \bar{X}_2),
$$

which is estimated by $\hat{D} = (\bar{Y}_1 - \bar{Y}_2) + \hat{\beta}(\bar{X}_1 - \bar{X}_2)$.

(b) Construct a 95% C.I. for this distance.

To construct the confidence interval of the distance, we need to evaluate the simultaneous confidence intervals given in Result 7.3.3:

$$\mathbf{c}'\hat{\beta} \pm \hat{\sigma}\left(s(\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c})F_{s,N-r,\alpha}\right)^{1/2}$$

Consider that in the hypothesis test for nonzero distance, $H : \mathbf{C}'\beta = 0$, $\mathbf{C}' = \begin{bmatrix}1 & -1 & (\bar{X}_1 - \bar{X}_2)\end{bmatrix}$ and $\beta' = \begin{bmatrix}\alpha_1 & \alpha_2 & \beta\end{bmatrix}$. From part (a) we know $(\mathbf{X}'\mathbf{X})^{-1}$. Therefore, we evaluate $\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$ as

$$\begin{aligned}\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} &= \begin{bmatrix}1 & -1 & (\bar{X}_1 - \bar{X}_2)\end{bmatrix}\begin{bmatrix}\frac{1}{n_1} & 0 & 0 \\ 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & \frac{1}{S_{XX,1}+S_{XX,2}}\end{bmatrix}\begin{bmatrix}1 \\ -1 \\ (\bar{X}_1 - \bar{X}_2)\end{bmatrix} \\ &= \frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{X}_1 - \bar{X}_2)^2}{S_{XX,1} + S_{XX,2}}\end{aligned}$$

Finally, noting that the rank of $\mathbf{C}$ (s) is 1 and the rank of $\mathbf{X}$ (r) is 3, the 95% confidence interval for the estimate of the distance between the two lines is

$$\hat{D} \pm \hat{\sigma}\left[\left(\frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{X}_1 - \bar{X}_2)^2}{S_{XX,1} + S_{XX,2}}\right)F_{1,n_1+n_2-3,0.05}\right]^{1/2}$$