Using the partial F test to test $H_0 : \beta_3 = 0$ against $H_A : \beta_3 \neq 0$ for the cement data

1. Calculate the error sum of squares and mean squared error for the full model.

```
> data=read.csv("http://chase.mathstat.dal.ca/~bsmith/stat3340/Data/cement.csv")
>   #read data
>  X=as.matrix(cbind(rep(1,13),data[,-c(1,2)])) #construct the X matrix
>  y=matrix(data[,2],byrow=T,ncol=1)   #construct Y
>   n=dim(X)[1]   #number of observations
>   k=dim(X)[2]-1 #number of predictor variables in full model
>   r=1           #number of parameters set to 0 under the null hypothesis
> beta=solve(t(X)%*%X)%*%t(X)%*%y  #calculate least squares estimates
> SSE=t(y)%*%y-t(beta)%*%t(X)%*%y  #get the error SS for full model
> SSE

         [,1]
[1,] 47.86364

> MSE=SSE/(n-k-1)                    #MSE under full model
> MSE

         [,1]
[1,] 5.982955
```

2. Calculate the error sum of squares for the reduced model, with $\beta_3 = 0$.

```
> X2=X[,c(1:3,5)]   #construct the X matrix for the reduced model
> beta2=solve(t(X2)%*%X2)%*%t(X2)%*%y  #calculate least squares estimates
>                                      #for the reduced model
> SSE2=t(y)%*%y-t(beta2)%*%t(X2)%*%y   #get the error SS for the reduced model
> SSE2

         [,1]
[1,] 47.97273
```

3. Calculate the observed value of the test statistic and the p-value.

```
> Fobs=((SSE2-SSE)/r)/MSE    #observed value of test statistic
> Fobs

           [,1]
[1,] 0.01823346

> pv=1- pf(Fobs,r,n-k-1)     #p-value for the test
> pv

          [,1]
[1,] 0.8959227
```

4. Check results using the anova command and the output from two calls to the "lm" procedure, one call for the full model, and one call for the reduced model.

```
> lm.out=lm(y~x_1+x_2+x_3+x_4,data=data)
> lm.out2=lm(y~x_1+x_2+x_4,   data=data)
> anova(lm.out,lm.out2)

Analysis of Variance Table

Model 1: y ~ x_1 + x_2 + x_3 + x_4
Model 2: y ~ x_1 + x_2 + x_4
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1      8 47.864
2      9 47.973 -1  -0.10909 0.0182 0.8959
```

Example 2: testing $H_0 : \beta_3 = \beta_4 = 0$ against the alternative that at least one of β_2 or β_3 are non-zero.

1. ```
   > r2=2          #number parameters set to 0 under H_0
   > X3=X[,c(1:3)] #construct the X matrix for the reduced model
   > beta3=solve(t(X3)%*%X3)%*%t(X3)%*%y  #calculate least squares estimates
   >                                      #for the reduced model
   > SSE3=t(y)%*%y-t(beta3)%*%t(X3)%*%y   #get the SSE for the reduced model
   > Fobs2=((SSE3-SSE)/r2)/MSE    #observed value of test statistic
   > pv2=1- pf(Fobs2,r2,n-k-1)    #p-value for the test
   > c(SSE3, Fobs2, pv2)

   [1] 57.9044832  0.8391208  0.4668465
   ```

2. Check using the builtin "anova" function.

   ```
   > lm.out3=lm(y~x_1+x_2,   data=data)
   > anova(lm.out,lm.out3)

   Analysis of Variance Table

   Model 1: y ~ x_1 + x_2 + x_3 + x_4
   Model 2: y ~ x_1 + x_2
     Res.Df    RSS Df Sum of Sq      F Pr(>F)
   1      8 47.864
   2     10 57.904 -2   -10.041 0.8391 0.4668
   ```
   ```