MA50259: Statistical Design of Investigations

Lab sheet 3: Estimability in designs of less than full rank

In this practical you will learn more of how to analyse an experimental design that does not have full rank and also the basics of the corresponding hypothesis testing. Take the time to run each of the following commands and analyse the displayed results to understand what the code is doing.

Estimation in the treatment effects model

Consider the treatment effects model of a completely randomized design (CRD) with $t = 3$ levels and $r_i = 4$ replicates

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \sigma^2\, \boldsymbol{I})$ and

$$\boldsymbol{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \end{pmatrix}, \; \boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \; \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}, \; \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{34} \end{pmatrix}$$

1. Load the corresponding packages to use in this tutorial

```
library(tidyverse)
library(MASS)
```

2. As done in Lab 2, simulate the response with a normal distribution. Also, construct the design matrix $\boldsymbol{X}$.

```
t<-3; # the number of treatmenet levels
r<-4; # the number of replicates
n<-t*r
levels<-c("level 1","level 2","level 3");
fact<-gl(t,r,labels=levels)
crd <- tibble(  treatment=fact )
set.seed(13579)
mu<-500 # reference value
tau<-c(50,0,-30) # treatment effects= differences wrt to mu
sd<-10 # overall standard deviation
means<-mu+tau %>% rep(each=r) # vector of means
y<-rnorm(n,mean=means,sd=sd)
crd$response<-y
# construct the design matrix X
Z <- model.matrix(~ fact-1)
X<-cbind(1,as.matrix(Z))
```

```
colnames(X)<-c("reference","effect 1","effect 2","effect 3")
X
```

```
##    reference effect 1 effect 2 effect 3
## 1          1        1        0        0
## 2          1        1        0        0
## 3          1        1        0        0
## 4          1        1        0        0
## 5          1        0        1        0
## 6          1        0        1        0
## 7          1        0        1        0
## 8          1        0        1        0
## 9          1        0        0        1
## 10         1        0        0        1
## 11         1        0        0        1
## 12         1        0        0        1
```

3. **Definition:** A generalized inverse of the symmetric matrix $\boldsymbol{X^T X}$ is a square matrix $\boldsymbol{G}$ satisfying

$$\boldsymbol{X^T X G X^T X = X^T X}$$

The command `ginv` in the package `MASS` computes one type of generalised inverse called the Moore-Penrose generalised inverse (here we call it $\boldsymbol{G}_1$). Another type of generalized inverse can be calculated for our specific case of $\boldsymbol{X^T X}$ as follows

$$\boldsymbol{G}_2 = \begin{pmatrix} 0 & 0 \\ 0 & (\boldsymbol{Z^T Z})^{-1} \end{pmatrix}$$

where $\boldsymbol{Z}$ is the design matrix of the means model. Compute two solutions, $\tilde{\boldsymbol{\beta}}_1$ and $\tilde{\boldsymbol{\beta}}_2$ to the normal equations

$$\boldsymbol{X^T X \beta = X^T y}$$

in our example using these two types of generalised inverse. That is, compute $\tilde{\boldsymbol{\beta}}_1 = \boldsymbol{G}_1 \boldsymbol{X^T y}$ and $\tilde{\boldsymbol{\beta}}_2 = \boldsymbol{G}_2 \boldsymbol{X^T y}$

```
ginv1<-ginv(t(X)%*%X)
ginv2<-rbind(0,cbind(0,solve(t(Z)%*%Z)))
beta1<-ginv1%*%t(X)%*%y
beta1
```

```
##           [,1]
## [1,] 380.13679
## [2,] 159.19078
## [3,] 131.24750
## [4,]  89.69851
```

```
beta2<-ginv2%*%t(X)%*%y
beta2
```

```
##                 [,1]
##               0.0000
## factlevel 1 539.3276
## factlevel 2 511.3843
## factlevel 3 469.8353
```

Note the two solutions are very different!

4. **Definition:** A linear combination $\boldsymbol{\lambda^T \beta}$ of the parameter vector $\boldsymbol{\beta}$ is said to be **estimable** if it can be estimated unbiasedly with a linear combination of the observations, that is, if there exists a vector $\boldsymbol{a}$ of length n such that

$$E[\boldsymbol{a^T Y}] = \boldsymbol{\lambda^T \beta}, \qquad \text{for all } \boldsymbol{\beta}$$

- It is easy to check that $\boldsymbol{\lambda^T\beta}$ is estimable if and only if there exists a vector $\boldsymbol{a}$ of length n such that $\boldsymbol{\lambda = X^T a}$, that is, if $\boldsymbol{\lambda}$ belongs to the space generated by the rows of $\boldsymbol{X}$. Which ones of the following linear combinations are estimable?
- $\boldsymbol{\lambda_1^T\beta} = \mu + \tau_1$
- $\boldsymbol{\lambda_2^T\beta} = \tau_1 - \tau_2$
- $\boldsymbol{\lambda_3^T\beta} = \mu - \tau_1$
- $\boldsymbol{\lambda_4^T\beta} = \tau_1$

```
lambda1<-c(0,1,-1,0)
lambda2<-c(1,1,0,0)
lambda3<-c(1,-1,0,0)
lambda4<-c(0,1,0,0)
# you can check for estimability using the rank
# of the extended matrix
qr(rbind(X,lambda1))$rank
```

```
## [1] 3
```

```
qr(rbind(X,lambda2))$rank
```

```
## [1] 3
```

```
# the 2 above are estimable since the rank does not change
# this means the lambda vector belongs to the row space of X
qr(rbind(X,lambda3))$rank
```

```
## [1] 4
```

```
qr(rbind(X,lambda4))$rank
```

```
## [1] 4
```

```
# the last 2 above are NOT estimable since the rank increases
# this means the lambda vector does not belong to the row space of X
```

5. If $\boldsymbol{\lambda^T\beta}$ is estimable, then we can estimate it with any solution to the normal equations $\tilde{\boldsymbol{\beta}}$ by simply using

$$\widehat{\boldsymbol{\lambda^T\beta}} := \boldsymbol{\lambda^T\tilde{\beta} = \lambda^T G X^T y = a^T X G X^T y}$$

and the estimator does not depend on the particular choice of generalised inverse $\boldsymbol{G}$. Verify that this last statement is true for the linear combinations $\boldsymbol{\lambda_1^T\beta} = \mu + \tau_1$ and $\boldsymbol{\lambda_2^T\beta} = \tau_1 - \tau_2$ using the two solutions of the normal equations found in question 3.

```
lambda1<-c(1,1,0,0)
lambda2<-c(0,1,-1,0)



t(lambda1)%*%ginv1%*%t(X)%*%y
```

```
##          [,1]
## [1,] 539.3276
```

```
t(lambda1)%*%ginv2%*%t(X)%*%y
```

```
##          [,1]
## [1,] 539.3276
```

```
t(lambda2)%*%ginv1%*%t(X)%*%y
```

```
##          [,1]
## [1,] 27.94328
```

```
t(lambda2)%*%ginv2%*%t(X)%*%y
```

```
##          [,1]
## [1,] 27.94328
```

6. The reason why the 2 estimators above: one based on $\boldsymbol{G}_1$ and the other based on $\boldsymbol{G}_2$, do not depend on the particular choice of generalised inverse is because the product

$$\boldsymbol{XGX^T}$$

is invariant to the choice of generalised inverse $\boldsymbol{G}$. Verify this product does not depend on the choice of $\boldsymbol{G}$ using the two generalised inverses $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ computed above, that is, verify that

$$\boldsymbol{XG_1X^T} = \boldsymbol{XG_2X^T}$$

```
round(X%*%ginv1%*%t(X),2)
```

```
##       1    2    3    4    5    6    7    8    9   10   11   12
## 1  0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## 2  0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## 3  0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## 4  0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## 5  0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00
## 6  0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00
## 7  0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00
## 8  0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00
## 9  0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25
## 10 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25
## 11 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25
## 12 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25
```

```
X%*%ginv2%*%t(X)
```

```
##       1    2    3    4    5    6    7    8    9   10   11   12
## 1  0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## 2  0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## 3  0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## 4  0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## 5  0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00
## 6  0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00
## 7  0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00
## 8  0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25 0.00 0.00 0.00 0.00
## 9  0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25
## 10 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25
## 11 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25
## 12 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.25 0.25 0.25 0.25
```

```
# checks they are the same
# we use 'round' to save output space
round(X%*%ginv1%*%t(X)-X%*%ginv2%*%t(X),2)
```

```
##    1 2 3 4 5 6 7 8 9 10 11 12
## 1  0 0 0 0 0 0 0 0 0  0  0  0
```

```
## 2  0 0 0 0 0 0 0 0 0  0  0  0
## 3  0 0 0 0 0 0 0 0 0  0  0  0
## 4  0 0 0 0 0 0 0 0 0  0  0  0
## 5  0 0 0 0 0 0 0 0 0  0  0  0
## 6  0 0 0 0 0 0 0 0 0  0  0  0
## 7  0 0 0 0 0 0 0 0 0  0  0  0
## 8  0 0 0 0 0 0 0 0 0  0  0  0
## 9  0 0 0 0 0 0 0 0 0  0  0  0
## 10 0 0 0 0 0 0 0 0 0  0  0  0
## 11 0 0 0 0 0 0 0 0 0  0  0  0
## 12 0 0 0 0 0 0 0 0 0  0  0  0
```

7. For the treatment effects model in a CRD, there can be at most t linearly independent estimable linear combinations, in other words, the space of estimable linear combinations is of dimension t, the number of levels of the treatment effect. Given two estimable linear combinations $\boldsymbol{\lambda_1^T\beta}$ and $\boldsymbol{\lambda_2^T\beta}$, they are said to be linearly independent if $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are linearly independent. Verify that $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ in question 5 are linearly independent. Here, the dimension is $t = 3$ so find a third linearly independent combination $\boldsymbol{\lambda_3^T\beta}$

```
aux2<-rbind(lambda1,lambda2)
# if linearly independent the rank should be 2
qr(aux2)$rank
```

```
## [1] 2
```

```
lambda_3<-c(0,0,1,-1)
aux3<-rbind(lambda1,lambda2,lambda3)
# if new vector is linearly independent to the 2 above
# then the rank should be 3 now
qr(aux3)$rank
```

```
## [1] 3
```

```
# note there are many vectors like this
lambda_3<-c(0,1,1,-1)
aux3<-rbind(lambda1,lambda2,lambda3)
qr(aux3)$rank
```

```
## [1] 3
```

8. Any linear model of less than full rank (such as the treatment effects model in a CRD) can always be reparametrised into a linear model of full rank. Given t linearly independent estimable combinations $\boldsymbol{\lambda_1^T\beta}, \ldots, \boldsymbol{\lambda_t^T\beta}$. Let $\boldsymbol{\Lambda^T} = (\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_t)$ be the matrix of dimension $(t+1) \times t$ whose columns are $\{\boldsymbol{\lambda}_i\}$. Let γ be a vector of length $t+1$ such that the square matrix $\boldsymbol{U}^T = (\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_t, \gamma) = (\boldsymbol{\Lambda} \,|\, \gamma)$ is non-singular with inverse $\boldsymbol{W} = \boldsymbol{U}^{-1} = (\boldsymbol{W}_1 \,|\, \eta)$. If $\boldsymbol{W}$ diagonalises $\boldsymbol{X^TX}$, that is

$$\boldsymbol{W}^T\boldsymbol{X^TXW} = diag(v_1, \ldots, v_t, 0)$$

for some constants $v_1, \ldots, v_t$. Then $\boldsymbol{X\eta} = \boldsymbol{0}$ and we can write

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} = (\boldsymbol{XW_1})(\boldsymbol{\Lambda\beta}) + (\boldsymbol{X\eta})(\boldsymbol{\gamma^T\beta}) + \boldsymbol{\epsilon} = \boldsymbol{Z\delta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{Z} = \boldsymbol{XW_1}$ has rank t and $\boldsymbol{\delta} = \boldsymbol{\Lambda\beta}$ so this model is full rank. The reparametrised full rank model is not unique. The commands below construct two different reparametrised full rank models both starting from the non full rank treatment effect model. Verify that the first one corresponds to the means model.

```r
# first define t=3 linearly independent  estimable combinations
lambda1<-c(1,1,0,0)
lambda2<-c(1,0,1,0)
lambda3<-c(1,0,0,1)
Lambda=rbind(lambda1,lambda2,lambda3)
# checks they are  linearly independent
qr(Lambda)$rank
```

```
## [1] 3
```

```r
# Now add a new vector gama linearly independent to all of the above
gama<-c(1,1,0,1)
#gama<-c(1,0,1,1) # alternative 1
#gama<-c(0,0,1,1) # alternative 2
# the vector gama is arbitrary as long as U below is non-singular
U<-rbind(Lambda,gama)
# check is full rank and therefore non-singular
qr(U)$rank
```

```
## [1] 4
```

```r
# compute its inverse
W<-solve(U)
round(W,3)
```

```
##      lambda1 lambda2 lambda3 gama
## [1,]       1       0       1   -1
## [2,]       0       0      -1    1
## [3,]      -1       1      -1    1
## [4,]      -1       0       0    1
```

```r
# W1 is formed with the first 3 columns of W
W1<-W[,1:3]
# eta is formed with the last column of W
eta<-W[,4]
t(X%*%eta) # should be zero vector
```

```
##      1 2 3 4 5 6 7 8 9 10 11 12
## [1,] 0 0 0 0 0 0 0 0 0  0  0  0
```

```r
Z1<-X%*%W1 # new design matrix
Z1
```

```
##    lambda1 lambda2 lambda3
## 1        1       0       0
## 2        1       0       0
## 3        1       0       0
## 4        1       0       0
## 5        0       1       0
## 6        0       1       0
## 7        0       1       0
## 8        0       1       0
## 9        0       0       1
## 10       0       0       1
## 11       0       0       1
## 12       0       0       1
```

```r
# it corresponds to the means model!!!
qr(Z1)$rank # should have full column rank
```

```
## [1] 3
```

```r
# we double check that it diagonalises X'X
# v_1=v_2=v_3=4
t(W)%*%t(X)%*%X%*%W
```

```
##         lambda1 lambda2 lambda3 gama
## lambda1       4       0       0    0
## lambda2       0       4       0    0
## lambda3       0       0       4    0
## gama          0       0       0    0
```

```r
# second reparametrisation
# uses eigenvectors to guarantee diagonalisation
# e.g. E X'X E'=diagonal matrix
# furthermore E has linearly independent columns
E<-t(eigen(t(X)%*%X)$vectors)
# we only take 3 linearly independent vectors
# since that is maximum
Lambda=E[1:3,]
# for the fourth one we (conveniently) take the last eigenvector
# but any other (linearly independent ) vector will do just as well
gama<-E[4,]
U<-rbind(Lambda,gama)
# double check is non-singular
qr(U)$rank
```

```
## [1] 4
```

```r
# compute its inverse
W<-solve(U)
# W1 is formed with the first 3 columns of W
W1<-W[,1:3]
# eta is formed with the last column of W
eta<-W[,4]
round(t(X%*%eta),3) # should be zero vector
```

```
##      1 2 3 4 5 6 7 8 9 10 11 12
## [1,] 0 0 0 0 0 0 0 0 0  0  0  0
```

```r
Z2<-X%*%W1 # new design matrix
round(Z2,2)
```

```
##
## 1  1.15 -0.58 -0.58
## 2  1.15 -0.58 -0.58
## 3  1.15 -0.58 -0.58
## 4  1.15 -0.58 -0.58
## 5  1.15 -0.21  0.79
## 6  1.15 -0.21  0.79
## 7  1.15 -0.21  0.79
## 8  1.15 -0.21  0.79
## 9  1.15  0.79 -0.21
## 10 1.15  0.79 -0.21
```

```
## 11 1.15  0.79 -0.21
## 12 1.15  0.79 -0.21
```

```
# Please note the block structure in the design matrix
# The new design matrix should have full column rank
# e.g. rank = number of columns
qr(Z2)$rank
```

```
## [1] 3
```

```
# check that it diagonalises X'X
round(t(W)%*%t(X)%*%X%*%W,3)
```

```
##             gama
##      16 0 0    0
##       0 4 0    0
##       0 0 4    0
## gama  0 0 0    0
```

```
# so now v_1=16 and v_2=v_3=4
```

9. The error sum of squares is defined as

$$ssE := \|\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}\|^2$$

where $\tilde{\boldsymbol{\beta}}$ is a solution to the normal equations. This can be written as

$$ssE := \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}\|^2$$

and therefore is invariant to the choice of generalised inverse $\boldsymbol{G}$ since $\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\boldsymbol{T}}$ is also invariant. Verify, using our observed sample that

$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{G_1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}\|^2 - \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{G_2}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}\|^2$$

```
# enough to checkif  the vectors are the same
# if vectors are the same, the sqaured norms are the same!
d1<-y-X%*%ginv1%*%t(X)%*%y
d2<-y-X%*%ginv2%*%t(X)%*%y
# transposed and rounded to save output space
round(t(d1-d2),2)
```

```
##      1 2 3 4 5 6 7 8 9 10 11 12
## [1,] 0 0 0 0 0 0 0 0 0  0  0  0
```

10. To test the hypothesis

$$H_0 : \tau_1 = \cdots = \tau_t$$

we can compute the error sum of squares in the reduced model where H_0 is true. Note this reduced model is given by

$$\boldsymbol{Y} = \mathbf{1}\mu + \boldsymbol{\epsilon}$$

where $\mathbf{1}$ is the column vector of ones of dimension n. Clearly this model is of full rank (equal to one) and the solution to the normal equations (since $\boldsymbol{X} = \mathbf{1}$ and therefore $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} = \mathbf{1}^{\boldsymbol{T}}\mathbf{1} = n$)

$$\mathbf{1}^{\boldsymbol{T}}\mathbf{1}\mu = \mathbf{1}^{\boldsymbol{T}}\boldsymbol{y} = \sum_{i=1}^{n} y_i$$

is $\hat{\mu} = \bar{y} = \mathbf{1}^{\boldsymbol{T}}\boldsymbol{y}/n$ and the corresponding error sum of squares is

$$ssE_0 := \|\boldsymbol{y} - \mathbf{1}\hat{\mu}\|^2 = \|\boldsymbol{y} - \mathbf{1}\bar{y}\|^2$$

The test statistic commonly used is

$$F = \frac{n-t}{t-1}\frac{SSE_0 - SSE}{SSE}$$

Compute F for our data and compare to the one given using the `aov` command in R given by `aov(response treatment,crd)`.

```
SSE<-t(d1)%*%d1
SSE
```

```
##          [,1]
## [1,] 568.7192
```

```
d0<-y-mean(y)
SSE0<-t(d0)%*%d0
SSE0-SSE
```

```
##          [,1]
## [1,] 9781.762
```

```
F.statistic<-((n-t)/(t-1))*(SSE0-SSE)/SSE
F.statistic
```

```
##          [,1]
## [1,] 77.39836
```

```
aov1<-aov(response~treatment,crd)
summary(aov1)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## treatment    2   9782    4891    77.4 2.14e-06 ***
## Residuals    9    569      63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As we will see, large values of this test statistic are evidence against the null hypothesis.

In the lectures we will introduce the concept of **degrees of freedom** associated with a sum of squares such as ssE, ssE_0 or $ssE_0 - ssE$. In this particular case

- ssE has $n - t$ degrees of freedom
- ssE_0 has $n - 1$ degrees of freedom
- $ssE_0 - ssE$ has $n - 1 - (n - t) = t - 1$ degrees of freedom

The Mean squared error is defined as the corresponding sum of squares divided by its corresponding degrees of freedom so that the F statistic is simply the ratio between Mean squared error

$$\frac{ssE_0 - ssE}{t-1}$$

and the Mean squared error

$$\frac{ssE}{n-t}$$