MAST30025 Week 10 Lab

#Question 1 #Part a: Find s2 (Correct)

#Re-typing the code from Week 8's Lab  
setwd("~/Desktop/UNIMELB 2021 Material/UNIMELB S1 2021 (Currently)/MAST30025/Tutorials /Tutorials/Rfile/data")  
filters = read.csv("filters.csv")  
  
#More to add how to convert into a factor! #Slide 63 Less than full rank model!  
filters$type = factor(filters$type)

#Attempt 1  
  
#y matrix  
y = filters$life  
  
#X matrix  
n = dim(filters)  
k = length(levels(filters$type))  
X = matrix(0,n,k+1)  
#Add all ones in the first column  
X[,1] = 1  
#Add all ones for each tao factor  
  
#2nd column  
X[filters$type==1,2] = 1  
#3rd column  
X[filters$type==2,3] = 1  
#4th column  
X[filters$type==3,4] = 1  
#5th column  
X[filters$type==4,5] = 1  
#6th column  
X[filters$type==5,6] = 1

y

## [1] 261 186 239 243 296 270 221 188 167 224 178 147 201 146 96 173 280 100 600  
## [20] 301 608 283 193 159 160 135 455 402 457 559

X

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 1 1 0 0 0 0  
## [2,] 1 1 0 0 0 0  
## [3,] 1 1 0 0 0 0  
## [4,] 1 1 0 0 0 0  
## [5,] 1 1 0 0 0 0  
## [6,] 1 1 0 0 0 0  
## [7,] 1 0 1 0 0 0  
## [8,] 1 0 1 0 0 0  
## [9,] 1 0 1 0 0 0  
## [10,] 1 0 1 0 0 0  
## [11,] 1 0 1 0 0 0  
## [12,] 1 0 1 0 0 0  
## [13,] 1 0 0 1 0 0  
## [14,] 1 0 0 1 0 0  
## [15,] 1 0 0 1 0 0  
## [16,] 1 0 0 1 0 0  
## [17,] 1 0 0 1 0 0  
## [18,] 1 0 0 1 0 0  
## [19,] 1 0 0 0 1 0  
## [20,] 1 0 0 0 1 0  
## [21,] 1 0 0 0 1 0  
## [22,] 1 0 0 0 1 0  
## [23,] 1 0 0 0 1 0  
## [24,] 1 0 0 0 1 0  
## [25,] 1 0 0 0 0 1  
## [26,] 1 0 0 0 0 1  
## [27,] 1 0 0 0 0 1  
## [28,] 1 0 0 0 0 1  
## [29,] 1 0 0 0 0 1  
## [30,] 1 0 0 0 0 1

#Actual X matrix should works out exactly above  
X1 = matrix(0,30,6)  
X1[,1] = 1  
for (i in 1:5){X1[filters$type==i,i+1] = 1}  
X1

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 1 1 0 0 0 0  
## [2,] 1 1 0 0 0 0  
## [3,] 1 1 0 0 0 0  
## [4,] 1 1 0 0 0 0  
## [5,] 1 1 0 0 0 0  
## [6,] 1 1 0 0 0 0  
## [7,] 1 0 1 0 0 0  
## [8,] 1 0 1 0 0 0  
## [9,] 1 0 1 0 0 0  
## [10,] 1 0 1 0 0 0  
## [11,] 1 0 1 0 0 0  
## [12,] 1 0 1 0 0 0  
## [13,] 1 0 0 1 0 0  
## [14,] 1 0 0 1 0 0  
## [15,] 1 0 0 1 0 0  
## [16,] 1 0 0 1 0 0  
## [17,] 1 0 0 1 0 0  
## [18,] 1 0 0 1 0 0  
## [19,] 1 0 0 0 1 0  
## [20,] 1 0 0 0 1 0  
## [21,] 1 0 0 0 1 0  
## [22,] 1 0 0 0 1 0  
## [23,] 1 0 0 0 1 0  
## [24,] 1 0 0 0 1 0  
## [25,] 1 0 0 0 0 1  
## [26,] 1 0 0 0 0 1  
## [27,] 1 0 0 0 0 1  
## [28,] 1 0 0 0 0 1  
## [29,] 1 0 0 0 0 1  
## [30,] 1 0 0 0 0 1

#Actual X matrix should works out exactly above  
t(X)%\*%X

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 30 6 6 6 6 6  
## [2,] 6 6 0 0 0 0  
## [3,] 6 0 6 0 0 0  
## [4,] 6 0 0 6 0 0  
## [5,] 6 0 0 0 6 0  
## [6,] 6 0 0 0 0 6

#If you want it to convert into a matrix form!  
#finding the inverse  
XtXc = matrix(0,6,6)  
XtXc[2:6,2:6] = solve((t(X)%\*%X)[2:6,2:6])  
XtXc

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000  
## [2,] 0 0.1666667 0.0000000 0.0000000 0.0000000 0.0000000  
## [3,] 0 0.0000000 0.1666667 0.0000000 0.0000000 0.0000000  
## [4,] 0 0.0000000 0.0000000 0.1666667 0.0000000 0.0000000  
## [5,] 0 0.0000000 0.0000000 0.0000000 0.1666667 0.0000000  
## [6,] 0 0.0000000 0.0000000 0.0000000 0.0000000 0.1666667

library(MASS)  
#Finding another conditional inverse (Open the MASS Package)  
XtXc2 = ginv(t(X)%\*%X)  
XtXc2

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0.02314815 0.00462963 0.00462963 0.00462963 0.00462963 0.00462963  
## [2,] 0.00462963 0.13425926 -0.03240741 -0.03240741 -0.03240741 -0.03240741  
## [3,] 0.00462963 -0.03240741 0.13425926 -0.03240741 -0.03240741 -0.03240741  
## [4,] 0.00462963 -0.03240741 -0.03240741 0.13425926 -0.03240741 -0.03240741  
## [5,] 0.00462963 -0.03240741 -0.03240741 -0.03240741 0.13425926 -0.03240741  
## [6,] 0.00462963 -0.03240741 -0.03240741 -0.03240741 -0.03240741 0.13425926

#Finding the least squares estimator   
b = XtXc2%\*%t(X)%\*%y  
b

## [,1]  
## [1,] 220.22222  
## [2,] 28.94444  
## [3,] -32.72222  
## [4,] -54.22222  
## [5,] 137.11111  
## [6,] 141.11111

#Finding s2  
e = y - X%\*%b  
SSRes = sum(e^2)  
s2 = SSRes/(30-5)  
s2

## [1] 15304.2

#Part b: 95% Confidence Interval for the difference in lifespan between filter types 3 and 4.(correct)

tt = c(0,0,0,1,-1,0)  
ta = qt(0.975,30-5)  
halfwidth = ta\*sqrt(s2\*t(tt)%\*%XtXc2%\*%tt)  
c(tt%\*%b-halfwidth,tt%\*%b+halfwidth)

## [1] -338.43399 -44.23268

#Part c: Show that the hypothesis that the filters all have the same lifespan is testable. (Nearly there!)

#Using theorm 6.10  
#Type 1 and 2  
tt1 = c(0,1,-1,0,0,0)  
round(tt1%\*%XtXc2%\*%t(X)%\*%X)

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0 1 -1 0 0 0

#Using theorm 6.10  
#Type 1 and 3  
tt2 = c(0,1,0,-1,0,0)  
round(tt2%\*%XtXc2%\*%t(X)%\*%X)

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0 1 0 -1 0 0

#Using theorm 6.10  
#Type 1 and 4  
tt3 = c(0,1,0,0,-1,0)  
round(tt3%\*%XtXc%\*%t(X)%\*%X)

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0 1 0 0 -1 0

#Using theorm 6.10  
#Type 1 and 5  
tt4 = c(0,1,0,0,0,-1)  
round(tt%\*%XtXc2%\*%t(X)%\*%X)

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0 0 0 1 -1 0

#Using theorm 6.10  
#Type 2 and 3  
tt5 = c(0,0,1,-1,0,0)  
round(tt5%\*%XtXc2%\*%t(X)%\*%X)

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0 0 1 -1 0 0

#Using theorm 6.10  
#Type 2 and 4  
tt6 = c(0,0,1,0,-1,0)  
round(tt6%\*%XtXc2%\*%t(X)%\*%X)

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0 0 1 0 -1 0

#Using theorm 6.10  
#Type 2 and 5  
tt7 = c(0,0,1,0,0,-1)  
round(tt7%\*%XtXc2%\*%t(X)%\*%X)

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0 0 1 0 0 -1

#Using theorm 6.10  
#Type 3 and 4  
tt8 = c(0,0,0,1,-1,0)  
round(tt8%\*%XtXc2%\*%t(X)%\*%X)

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0 0 0 1 -1 0

#Using theorm 6.10  
#Type 3 and 5  
tt9 = c(0,0,0,1,0,-1)  
round(tt9%\*%XtXc2%\*%t(X)%\*%X)

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0 0 0 1 0 -1

#Using theorm 6.10  
#Type 4 and 5  
tt10 = c(0,0,0,0,1,-1)  
round(tt10%\*%XtXc2%\*%t(X)%\*%X)

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0 0 0 0 1 -1

#Suprising they are all testable! #For all having the same lifespan that is testable!

#Actual Solution!

#Using theorm 6.10  
C = matrix(c(0,1,-1,0,0,0,0,0,1,-1,0,0,0,0,0,1,-1,0,0,0,0,0,1,-1),4,6,byrow=TRUE)  
round(C%\*%XtXc2%\*%t(X)%\*%X)

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0 1 -1 0 0 0  
## [2,] 0 0 1 -1 0 0  
## [3,] 0 0 0 1 -1 0  
## [4,] 0 0 0 0 1 -1

#Where C is at full rank! From Slide 6 Inference for the less than full rank model!

#Part d: Test this hypothesis, using matrix theory!

#Attempt 1  
numer = t(C%\*%b)%\*%solve(C%\*%XtXc%\*%t(C))%\*%C%\*%b/4 #Missing Bracket from  
#ginv(t(X)%\*%X)  
#Theorem 7.3 derived from the lectures, it is important!  
Fstat = numer/s2  
pf(Fstat,4,25,lower=F)

## [,1]  
## [1,] 0.02599945

#We can reject the null hypothesis that all filters have the same lifespan.

#Part e: Test the same hypothesis using the linearHypothesis function from the car package. (ASK ABOUT THE C2 MATRIX)

library(car)

## Loading required package: carData

model = lm(filters$life~filters$type,data = filters)  
C2 = matrix(c(0,1,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,1),4,5,byrow=TRUE)  
linearHypothesis(model,C2,rep(0,4))

## Linear hypothesis test  
##   
## Hypothesis:  
## filters$type2 = 0  
## filters$type3 = 0  
## filters$type4 = 0  
## filters$type5 = 0  
##   
## Model 1: restricted model  
## Model 2: filters$life ~ filters$type  
##   
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 29 585770   
## 2 25 382605 4 203165 3.3188 0.026 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Question 4 #Part a: #Module 7 slide 23 reading from the Tennis ball example! #Test the hypothesis that the work plan has no effect on the absenteeism.

#Attempt 1  
n = 275  
r = 3  
XtXc = diag(c(0,1/c(100,85,90)))  
b = c(0,9,6.2,10.1)  
s2 = 110.15  
C = matrix(c(0,1,-1,0,0,0,1,-1),2,4,byrow = T) #Depends on your C matrix also add byrow = T  
Fstat = (t(C%\*%b)%\*%solve(C%\*%XtXc%\*%t(C))%\*%C%\*%b/2/s2) #Theorem 7.3 understand your rank of hypothesis!  
pf(Fstat,2,n-r,lower=F)

## [,1]  
## [1,] 0.04228613

#We reject the null hypothesis, conclude that work plan has no effect on the absenteeism.

#Part b: Test the hypothesis that work plans 1 and 3 have the same rate of absenteeism

#Attempt 1  
n = 275  
r = 3  
XtXc = diag(c(0,1/c(100,85,90)))  
b = c(0,9,6.2,10.1)  
s2 = 110.15  
C1 = matrix(c(0,1,0,-1),1,4,byrow = T) #Depends on your C matrix  
Fstat = (t(C1%\*%b)%\*%solve(C1%\*%XtXc%\*%t(C1))%\*%C1%\*%b/s2)  
pf(Fstat,1,n-r,lower=F)

## [,1]  
## [1,] 0.471315

#We do not reject the null hypothesis, conclude that work plans 1 and 3 do not have the same rate of absenteeism.