MAST30025 Week 9 Lab

#Question 1: Reusing the data from Q5 from the last 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  
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) #converts all zeros for all elements in the first row and the first column in a 6 by 6 matrix  
XtXc[2:6,2:6] = solve((t(X)%\*%X)[2:6,2:6]) #Selects 2nd to 6th row and 2nd to 6th column respectively!  
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

#For Question 1a and 1b if they are estimable?

onea = c(1,-1,0,0,0,1)  
round(onea%\*%XtXc%\*%t(X)%\*%X)

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

#They are not estimable according to Theorem 6.10

oneb = c(0,1,0,-0.5,-0.5,0)  
round(oneb%\*%XtXc%\*%t(X)%\*%X)

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

#They are estimable according to Theorem 6.10

#Part c: In the week 8 Lab you were asked to find two solutions to the normal equations. Verify that they produce the same estimate of tao4 - tao5?

#Attempt 1  
dtao4tao5 = c(0,0,0,0,1,-1) #difference of tao 4 and tao 5  
  
b1 = XtXc%\*%t(X)%\*%y  
dtao4tao5%\*%b1

## [,1]  
## [1,] -4

#Attempt 1  
library(MASS)  
dtao4tao5 = c(0,0,0,0,1,-1) #difference of tao 4 and tao 5  
XtXc2 = ginv(t(X)%\*%X)  
b2 = XtXc2%\*%t(X)%\*%y  
dtao4tao5%\*%b2

## [,1]  
## [1,] -4

#They produce the same estimate. According to Theorem 6.11, that regardless of any conditional inverse you are using produces exactly the same estimate. Thus this estimator is unique. Derived from the lectures!

#Part d: Do your two solutions produce the same estimate of 2mew + tao1?

#Attempt 1  
twomewtaoone = c(2,1,0,0,0,0) # 2mew + tao1   
  
twomewtaoone%\*%b1

## [,1]  
## [1,] 249.1667

#Attempt 1  
twomewtaoone = c(2,1,0,0,0,0) # 2mew + tao1   
  
twomewtaoone%\*%b2

## [,1]  
## [1,] 469.3889

#They are not the same. Using Theorem 6.11 the estimator is NOT UNIQUE!

#Part e:

#Part i:

tt <- c(1,1,0,0,0,0)  
t(tt) %\*% XtXc %\*% t(X) %\*% X

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

tt%\*%b1

## [,1]  
## [1,] 249.1667

#Part ii:

tt <- c(0,0,1,-1,0,0)  
round(t(tt) %\*% XtXc %\*% t(X) %\*% X)

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

tt%\*%b1

## [,1]  
## [1,] 21.5

#Part iii:

tt <- c(0,-.2,-.2,-.2,.8,-.2)  
round(t(tt) %\*% XtXc %\*% t(X) %\*% X)

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

tt%\*%b1

## [,1]  
## [1,] 93.06667

#Part iv:

tt <- c(5,1,1,1,1,1)  
round(t(tt) %\*% XtXc %\*% t(X) %\*% X)

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

tt%\*%b1

## [,1]  
## [1,] 1321.333

#Part f:

filters$type <- factor(filters$type)  
model <- lm(life ~ type, data = filters)  
mu <- model$coefficients + c(0, 1, 1, 1, 1)\*model$coefficients[1]  
names(mu)[1] <- "type1"  
(mubar <- mean(mu))

## [1] 264.2667

mu - mubar

## type1 type2 type3 type4 type5   
## -15.10000 -76.76667 -98.26667 93.06667 97.06667

contrasts(filters$type) <- contr.sum(5)  
model2 <- lm(life ~ type, data = filters)  
summary(model2)

##   
## Call:  
## lm(formula = life ~ type, data = filters)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -226.333 -61.458 -2.833 39.625 250.667   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 264.27 22.59 11.700 1.23e-11 \*\*\*  
## type1 -15.10 45.17 -0.334 0.7410   
## type2 -76.77 45.17 -1.699 0.1017   
## type3 -98.27 45.17 -2.175 0.0393 \*   
## type4 93.07 45.17 2.060 0.0499 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 123.7 on 25 degrees of freedom  
## Multiple R-squared: 0.3468, Adjusted R-squared: 0.2423   
## F-statistic: 3.319 on 4 and 25 DF, p-value: 0.026

-sum(model2$coefficients[-1]) # mu\_5 - bar(mu)

## [1] 97.06667