MAST30025 Assignment 3 2009 R Code

#Question 1, #Creating a data frame first #Part a: Find a conditional inverse t(X)%\*%X, using the algorrithm given in the lecture slides.

yield = c(43,45,47,46,48,33,37,38,35,54,54,57)  
fertiliser = c(1,1,1,1,1,2,2,2,2,3,3,3)  
typefert = factor(fertiliser)

tomato = data.frame(yield,typefert)  
tomato

## yield typefert  
## 1 43 1  
## 2 45 1  
## 3 47 1  
## 4 46 1  
## 5 48 1  
## 6 33 2  
## 7 37 2  
## 8 38 2  
## 9 35 2  
## 10 54 3  
## 11 54 3  
## 12 57 3

#Response variable  
y = tomato$yield  
#Our Design Matrix  
X = matrix(0,12,4)  
X[,1]=1  
for (i in 1:3){X[typefert == i, i +1] = 1}  
X

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

#Finding the Condition Inverse according from the lectures!  
XtXc = matrix(0,4,4)  
XtXc[2:4,2:4] = solve((t(X)%\*%X)[2:4,2:4])  
XtXc

## [,1] [,2] [,3] [,4]  
## [1,] 0 0.0 0.00 0.0000000  
## [2,] 0 0.2 0.00 0.0000000  
## [3,] 0 0.0 0.25 0.0000000  
## [4,] 0 0.0 0.00 0.3333333

#Part b:Find two solutions to the normal equations, using only the conditional inverse you found in question 1a,

#First from the tutorials we find another Conditional inverse using the ginv function  
library(MASS)  
library(Matrix)  
XtXc2 = ginv(t(X)%\*%X)  
XtXc2

## [,1] [,2] [,3] [,4]  
## [1,] 0.048958333 0.001041667 0.01354167 0.034375  
## [2,] 0.001041667 0.148958333 -0.06354167 -0.084375  
## [3,] 0.013541667 -0.063541667 0.17395833 -0.096875  
## [4,] 0.034375000 -0.084375000 -0.09687500 0.215625

b = XtXc%\*%t(X)%\*%y  
b

## [,1]  
## [1,] 0.00  
## [2,] 45.80  
## [3,] 35.75  
## [4,] 55.00

b2 = XtXc2%\*%t(X)%\*%y  
b2

## [,1]  
## [1,] 34.1375  
## [2,] 11.6625  
## [3,] 1.6125  
## [4,] 20.8625

I = diag(c(rep(1,4)))  
I - XtXc%\*%t(X)%\*%X

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

# One solution to the normal equation is b.

# Another solution to the normal equation is b2 = b + (I - XtXc %*% t(X) %*% X) %*% z, Where z is an arbitrary 4*1 vector.

b2 = b + (I - XtXc %\*% t(X) %\*% X) %\*% as.vector(c(1,0,0,0))  
b2

## [,1]  
## [1,] 1.00  
## [2,] 44.80  
## [3,] 34.75  
## [4,] 54.00

#Part c: Is μ + τ1 - τ2 + τ3 estimable?

library(MASS)  
library(Matrix)  
tt = c(1,1,-1,1)  
tt%\*%ginv(t(X)%\*%X)%\*%t(X)%\*%X

## [,1] [,2] [,3] [,4]  
## [1,] 1 1 -1 1

#Yes it is estimable because it is a linear combination of μ + τ1, μ + τ2 and μ + τ3, which are all elements of Xβ and therefore estimable!

#Part d: Find a 95% confidence interval for the estimable quantity τ2 - τ3.

#Attempt 1  
library(Matrix)  
n = 12  
SSRes = sum((y-X%\*%b)^2)  
s2 = SSRes/(n-rankMatrix(X)[1])  
tt = c(0,0,1,-1) #treatment difference of fertiliser 2 and 3  
ta = qt(0.975,n-rankMatrix(X)[1])  
s = sqrt(s2)  
  
#Creating our Confidence Interval  
halfwidth = ta\*s\*sqrt(t(tt)%\*%ginv(t(X)%\*%X)%\*%tt)  
c(tt%\*%b-halfwidth,tt%\*%b+halfwidth)

## [1] -22.68384 -15.81616

#Part e: Test the hypothesis that fertiliser has no effect on yield.

#Attempt 1  
library(Matrix)  
C = matrix(c(1,0,-1,0,1,0,0,-1),2,4)  
m = rankMatrix(C)[1]  
num = t(C%\*%b)%\*%solve(C%\*%ginv(t(X)%\*%X)%\*%t(C))%\*%C%\*%b  
Fstat = (num/m)/s2  
Fstat

## [,1]  
## [1,] 1835.089

pf(Fstat,m,n-rankMatrix(X)[1], lower = F)

## [,1]  
## [1,] 1.770976e-12

#We reject the null hypothesis

#Actual solution!

library(Matrix)  
C = matrix(c(0,1,-1,0,0,0,1,-1),2,4)  
m = rankMatrix(C)[1]  
num = t(C%\*%b)%\*%solve(C%\*%ginv(t(X)%\*%X)%\*%t(C))%\*%C%\*%b  
Fstat = (num/m)/s2  
Fstat

## [,1]  
## [1,] 5164.716

pf(Fstat,m,n-rankMatrix(X)[1], lower = F)

## [,1]  
## [1,] 1.694514e-14

#Question 2. #Part a) What information do you gather about the parameters of the model in equation 1, from the coefficients of the fitted R model?

#rot = -10.222 - 3.333*oxygen.f2 - #3.333*oxygen.f3 + 11.556\*temp.f16 + error(ijk)

#Intercept corresponds to mew + tao1 + beta1, oxygen.f2 corresponds tao2-tao1, oxygen.f3 corresponds tao3-tao1 and temp.f16 is beta2-beta1!

#Part b) Should we accept or reject the hypothesis that oxygen level has no effect on rot? #Since the p value is 0.4481124 we do not reject the null hypothesis.

#Part c) Should we accept or reject the hypothesis that temperature has no effect on rot? #Since the p value is 0.0002849 we reject the null hypothesis.

#Part d) What hypothesis is the linear.hypothesis function testing, given in terms of the parameters of the model in equation 1? #The hypothesis is testing tao2-tao1 = beta2 - beta1; in other words, whether the difference in effect between the temperature levels is equal to the difference in effect between the first two oxygen levels.

#Part e) Using R, test for the presense of interaction between the factors

potato = data.frame(potato = c(13,11,3,10,4,7,15,2,7,26,19,24,15,22,18,20,24,8),oxygen = factor(c(1,1,1,2,2,2,3,3,3,1,1,1,2,2,2,3,3,3)),temp = factor(c(1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2)))

imodel = lm(potato~temp\*oxygen, data = potato)  
anova(imodel)

## Analysis of Variance Table  
##   
## Response: potato  
## Df Sum Sq Mean Sq F value Pr(>F)   
## temp 1 600.89 600.89 20.6412 0.000674 \*\*\*  
## oxygen 2 44.44 22.22 0.7634 0.487453   
## temp:oxygen 2 16.44 8.22 0.2824 0.758816   
## Residuals 12 349.33 29.11   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#There is no interaction so the p value is 0.76