homework7

beverly chigarira

11/11/2020

#Question 1a. Creating matrices to create the transformation for each coded variable from the given table  
  
#y = xt  
  
x = cbind("x1", "x2", "x3", "x4") #x  
x

## [,1] [,2] [,3] [,4]  
## [1,] "x1" "x2" "x3" "x4"

matrixy = cbind(50,70,10,20,1,2,100,200)  
y =t(matrixy)  
y #y

## [,1]  
## [1,] 50  
## [2,] 70  
## [3,] 10  
## [4,] 20  
## [5,] 1  
## [6,] 2  
## [7,] 100  
## [8,] 200

tmatrix = cbind(c(-1,0,0,0),c(1,0,0,0),c(0,-1,0,0),c(0,1,0,0),c(0,0,-1,0),c(0,0,1,0),c(0,0,0,-1),c(0,0,0,1)) #t  
tmatrix

## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]  
## [1,] -1 1 0 0 0 0 0 0  
## [2,] 0 0 -1 1 0 0 0 0  
## [3,] 0 0 0 0 -1 1 0 0  
## [4,] 0 0 0 0 0 0 -1 1

#Question 1b. Fitting the linear model  
# model-> y = B0 + B1X1 + B2X2 + B3X3+ B4X4 + E  
X = cbind(c(rep(1,16)),c(-1,-1,-1,-1,-1,-1,-1,-1,1,1,1,1,1,1,1,1),c(-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1,1,1),c(rep(-1,1),1),c(1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1))  
print("X")

## [1] "X"

X

## [,1] [,2] [,3] [,4] [,5]  
## [1,] 1 -1 -1 -1 1  
## [2,] 1 -1 -1 1 1  
## [3,] 1 -1 -1 -1 -1  
## [4,] 1 -1 -1 1 -1  
## [5,] 1 -1 1 -1 1  
## [6,] 1 -1 1 1 1  
## [7,] 1 -1 1 -1 -1  
## [8,] 1 -1 1 1 -1  
## [9,] 1 1 -1 -1 1  
## [10,] 1 1 -1 1 1  
## [11,] 1 1 -1 -1 -1  
## [12,] 1 1 -1 1 -1  
## [13,] 1 1 1 -1 1  
## [14,] 1 1 1 1 1  
## [15,] 1 1 1 -1 -1  
## [16,] 1 1 1 1 -1

matrixY =cbind(22.2,24.5,24.4,25.9,19.4,24.1,25.2,28.4,22.1,19.6,23.5,16.5,14.2,12.7,19.3,16.0)  
Y =t(matrixY)   
print("Y")

## [1] "Y"

Y

## [,1]  
## [1,] 22.2  
## [2,] 24.5  
## [3,] 24.4  
## [4,] 25.9  
## [5,] 19.4  
## [6,] 24.1  
## [7,] 25.2  
## [8,] 28.4  
## [9,] 22.1  
## [10,] 19.6  
## [11,] 23.5  
## [12,] 16.5  
## [13,] 14.2  
## [14,] 12.7  
## [15,] 19.3  
## [16,] 16.0

alphamatrix = cbind(1,-1,1,-1,1)  
alpha =t(alphamatrix)  
alpha

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

beta\_hat= solve(t(X)%\*%X)%\*%(t(X)%\*%Y)  
beta\_hat

## [,1]  
## [1,] 21.1250  
## [2,] -3.1375  
## [3,] -1.2125  
## [4,] -0.1625  
## [5,] -1.2750

sse = (t(Y)%\*%Y)-t(beta\_hat)%\*%(t(X)%\*%Y)  
sse

## [,1]  
## [1,] 98.8125

s=sqrt(sse/(16-3))  
s

## [,1]  
## [1,] 2.756984

#Therefore the fitted line is, using the calculated Beta\_hat values is:  
#Y(hat) = 21.1 - 3.14x - 1.21x^2 - 0.163x^3 - 1.27x^4

#Question 1C.Hypothesis test for B1=0, B2=0, B3=0, B4=0  
  
#Ho: B1 = 0  
#Ha: B1 not equal 0  
#B1 = 21.2/(s\*sqrt(t(alpha)%\*%solve(t(X)%\*%X)%\*%alpha))  
#B1

#to find SSE for reduced model  
  
print("XB1")

## [1] "XB1"

XB1=cbind(c(-1,-1,-1,-1,-1,-1,-1,-1,1,1,1,1,1,1,1,1))  
XB1

## [,1]  
## [1,] -1  
## [2,] -1  
## [3,] -1  
## [4,] -1  
## [5,] -1  
## [6,] -1  
## [7,] -1  
## [8,] -1  
## [9,] 1  
## [10,] 1  
## [11,] 1  
## [12,] 1  
## [13,] 1  
## [14,] 1  
## [15,] 1  
## [16,] 1

print("YB1")

## [1] "YB1"

Y=cbind(c(22.2,24.5,24.4,25.9,19.4,24.1,25.2,28.4,22.1,19.6,23.5,16.5,14.2,12.7,19.3,16.0))  
Y

## [,1]  
## [1,] 22.2  
## [2,] 24.5  
## [3,] 24.4  
## [4,] 25.9  
## [5,] 19.4  
## [6,] 24.1  
## [7,] 25.2  
## [8,] 28.4  
## [9,] 22.1  
## [10,] 19.6  
## [11,] 23.5  
## [12,] 16.5  
## [13,] 14.2  
## [14,] 12.7  
## [15,] 19.3  
## [16,] 16.0

#sse\_=t(Y)%\*%Y-t(beta\_hat)%\*%t(X)%\*%Y  
#sse\_c  
#print("beta\_hat")  
## [1] "beta\_hat"  
beta\_hat1=solve(t(XB1)%\*%XB1)%\*%(t(XB1)%\*%Y)  
beta\_hat1

## [,1]  
## [1,] -3.1375

print("sse\_r")

## [1] "sse\_r"

sse\_r1=t(Y)%\*%Y-t(beta\_hat1)%\*%t(XB1)%\*%Y  
sse\_r1

## [,1]  
## [1,] 7289.017

#To find F statistic  
k=4 # k is the number of independent variables in the complete model.  
g=3 # g is the number of variables in the reduced model.  
n=16 # n is the number of observations.  
print("F\_Statistic")

## [1] "F\_Statistic"

## [1] "F\_Statistic"  
F\_statistic=((sse\_r1-sse)/(k-g))/(sse/(n-k-1))  
F\_statistic

## [,1]  
## [1,] 800.4276

#the F\_critical value is given by "qf(1-alpha,numerator degree of freedom, de  
#nominator degree of freedom)  
print("F\_alpha")

## [1] "F\_alpha"

F\_alpha=qf(1-0.01,k-g,n-k-1)  
F\_alpha

## [1] 9.646034

#Since the F\_statistic is greater than the F\_alpha, we fail to reject null hypothesis at alpha=0.01 and conclude that Beta\_hat1 is not equal to 0

#Question 1C.Hypothesis test for B1=0, B2=0, B3=0, B4=0  
  
#Ho: B2 = 0  
#Ha: B2 not equal 0  
#B1 = 21.2/(s\*sqrt(t(alpha)%\*%solve(t(X)%\*%X)%\*%alpha))  
#B1  
#to find SSE for reduced model  
print("XB2")

## [1] "XB2"

XB2=cbind(c(-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1,1,1))  
XB2

## [,1]  
## [1,] -1  
## [2,] -1  
## [3,] -1  
## [4,] -1  
## [5,] 1  
## [6,] 1  
## [7,] 1  
## [8,] 1  
## [9,] -1  
## [10,] -1  
## [11,] -1  
## [12,] -1  
## [13,] 1  
## [14,] 1  
## [15,] 1  
## [16,] 1

print("YB2")

## [1] "YB2"

Y=cbind(c(22.2,24.5,24.4,25.9,19.4,24.1,25.2,28.4,22.1,19.6,23.5,16.5,14.2,12.7,19.3,16.0))  
Y

## [,1]  
## [1,] 22.2  
## [2,] 24.5  
## [3,] 24.4  
## [4,] 25.9  
## [5,] 19.4  
## [6,] 24.1  
## [7,] 25.2  
## [8,] 28.4  
## [9,] 22.1  
## [10,] 19.6  
## [11,] 23.5  
## [12,] 16.5  
## [13,] 14.2  
## [14,] 12.7  
## [15,] 19.3  
## [16,] 16.0

#sse\_=t(Y)%\*%Y-t(beta\_hat)%\*%t(X)%\*%Y  
#sse\_c  
#print("beta\_hat")  
## [1] "beta\_hat"  
beta\_hat2=solve(t(XB2)%\*%XB2)%\*%(t(XB2)%\*%Y)  
beta\_hat2

## [,1]  
## [1,] -1.2125

print("sse\_r2")

## [1] "sse\_r2"

sse\_r2=t(Y)%\*%Y-t(beta\_hat2)%\*%t(XB2)%\*%Y  
sse\_r2

## [,1]  
## [1,] 7422.997

#To find F statistic  
k=4 # k is the number of independent variables in the complete model.  
g=3 # g is the number of variables in the reduced model.  
n=16 # n is the number of observations.  
print("F\_Statistic")

## [1] "F\_Statistic"

## [1] "F\_Statistic"  
F\_statistic2=((sse\_r2-sse)/(k-g))/(sse/(n-k-1))  
F\_statistic2

## [,1]  
## [1,] 815.3425

#the F\_critical value is given by "qf(1-alpha,numerator degree of freedom, de  
#nominator degree of freedom)  
print("F\_alpha")

## [1] "F\_alpha"

F\_alpha2=qf(1-0.01,k-g,n-k-1)  
F\_alpha2

## [1] 9.646034

#Since the F\_statistic is greater than the F\_alpha, we fail to reject null hypothesis at alpha=0.01 and conclude that Beta\_hat2 is not equal to 0

#Question 1C.Hypothesis test for B1=0, B2=0, B3=0, B4=0  
  
#Ho: B3 = 0  
#Ha: B3 not equal 0  
  
#to find SSE for reduced model  
print("XB3")

## [1] "XB3"

XB3=cbind(c(-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1))  
XB3

## [,1]  
## [1,] -1  
## [2,] 1  
## [3,] -1  
## [4,] 1  
## [5,] -1  
## [6,] 1  
## [7,] -1  
## [8,] 1  
## [9,] -1  
## [10,] 1  
## [11,] -1  
## [12,] 1  
## [13,] -1  
## [14,] 1  
## [15,] -1  
## [16,] 1

print("YB3")

## [1] "YB3"

Y=cbind(c(22.2,24.5,24.4,25.9,19.4,24.1,25.2,28.4,22.1,19.6,23.5,16.5,14.2,12.7,19.3,16.0))  
Y

## [,1]  
## [1,] 22.2  
## [2,] 24.5  
## [3,] 24.4  
## [4,] 25.9  
## [5,] 19.4  
## [6,] 24.1  
## [7,] 25.2  
## [8,] 28.4  
## [9,] 22.1  
## [10,] 19.6  
## [11,] 23.5  
## [12,] 16.5  
## [13,] 14.2  
## [14,] 12.7  
## [15,] 19.3  
## [16,] 16.0

#sse\_=t(Y)%\*%Y-t(beta\_hat)%\*%t(X)%\*%Y  
#sse\_c  
#print("beta\_hat")  
## [1] "beta\_hat"  
beta\_hat3=solve(t(XB3)%\*%XB3)%\*%(t(XB3)%\*%Y)  
beta\_hat3

## [,1]  
## [1,] -0.1625

print("sse\_r3")

## [1] "sse\_r3"

sse\_r3=t(Y)%\*%Y-t(beta\_hat3)%\*%t(XB3)%\*%Y  
sse\_r3

## [,1]  
## [1,] 7446.097

#To find F statistic  
k=4 # k is the number of independent variables in the complete model.  
g=3 # g is the number of variables in the reduced model.  
n=16 # n is the number of observations.  
print("F\_Statistic")

## [1] "F\_Statistic"

## [1] "F\_Statistic"  
F\_statistic3=((sse\_r3-sse)/(k-g))/(sse/(n-k-1))  
F\_statistic3

## [,1]  
## [1,] 817.9141

#the F\_critical value is given by "qf(1-alpha,numerator degree of freedom, de  
#nominator degree of freedom)  
print("F\_alpha")

## [1] "F\_alpha"

F\_alpha3=qf(1-0.01,k-g,n-k-1)  
F\_alpha3

## [1] 9.646034

#Since the F\_statistic is greater than the F\_alpha, we fail to reject null hypothesis at alpha=0.01 and conclude that Beta\_hat3 is not equal to 0

#Question 1C.Hypothesis test for B1=0, B2=0, B3=0, B4=0  
  
#Ho: B4 = 0  
#Ha: B4 not equal 0  
#to find SSE for reduced model  
print("XB4")

## [1] "XB4"

XB4=cbind(c(1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1))  
XB4

## [,1]  
## [1,] 1  
## [2,] 1  
## [3,] -1  
## [4,] -1  
## [5,] 1  
## [6,] 1  
## [7,] -1  
## [8,] -1  
## [9,] 1  
## [10,] 1  
## [11,] -1  
## [12,] -1  
## [13,] 1  
## [14,] 1  
## [15,] -1  
## [16,] -1

print("YB4")

## [1] "YB4"

Y=cbind(c(22.2,24.5,24.4,25.9,19.4,24.1,25.2,28.4,22.1,19.6,23.5,16.5,14.2,12.7,19.3,16.0))  
Y

## [,1]  
## [1,] 22.2  
## [2,] 24.5  
## [3,] 24.4  
## [4,] 25.9  
## [5,] 19.4  
## [6,] 24.1  
## [7,] 25.2  
## [8,] 28.4  
## [9,] 22.1  
## [10,] 19.6  
## [11,] 23.5  
## [12,] 16.5  
## [13,] 14.2  
## [14,] 12.7  
## [15,] 19.3  
## [16,] 16.0

#sse\_=t(Y)%\*%Y-t(beta\_hat)%\*%t(X)%\*%Y  
#sse\_c  
#print("beta\_hat")  
## [1] "beta\_hat"  
beta\_hat4=solve(t(XB4)%\*%XB4)%\*%(t(XB4)%\*%Y)  
beta\_hat4

## [,1]  
## [1,] -1.275

print("sse\_r4")

## [1] "sse\_r4"

sse\_r4=t(Y)%\*%Y-t(beta\_hat4)%\*%t(XB4)%\*%Y  
sse\_r4

## [,1]  
## [1,] 7420.51

#To find F statistic  
k=4 # k is the number of independent variables in the complete model.  
g=3 # g is the number of variables in the reduced model.  
n=16 # n is the number of observations.  
print("F\_Statistic")

## [1] "F\_Statistic"

## [1] "F\_Statistic"  
F\_statistic4=((sse\_r4-sse)/(k-g))/(sse/(n-k-1))  
F\_statistic4

## [,1]  
## [1,] 815.0656

#the F\_critical value is given by "qf(1-alpha,numerator degree of freedom, de  
#nominator degree of freedom)  
print("F\_alpha")

## [1] "F\_alpha"

F\_alpha4=qf(1-0.01,k-g,n-k-1)  
F\_alpha4

## [1] 9.646034

#Since the F\_statistic is greater than the F\_alpha, we fail to reject null hypothesis at alpha=0.01 and conclude that Beta\_hat4 is not equal to 0

#Question 2  
#Ho: Mean A = MeanB = MeanC = MeanD  
#H1: At least one mean is different from others  
#Level of significance =0.05  
A =c(1.65,1.72,1.50,1.37,1.60)  
B =c(1.70,1.85,1.46,2.05,1.80)  
C =c(1.40,1.75,1.38,1.65,1.55)  
D =c(2.10,1.95,1.65,1.88,2.00)  
  
Combined\_data =data.frame(cbind(A,B,C,D))  
Stacked\_data = stack(Combined\_data)

Anova\_results = aov(values~ind, data = Stacked\_data)  
Anova\_results

## Call:  
## aov(formula = values ~ ind, data = Stacked\_data)  
##   
## Terms:  
## ind Residuals  
## Sum of Squares 0.464895 0.476800  
## Deg. of Freedom 3 16  
##   
## Residual standard error: 0.1726268  
## Estimated effects may be unbalanced

summary(Anova\_results)

## Df Sum Sq Mean Sq F value Pr(>F)   
## ind 3 0.4649 0.1550 5.2 0.0107 \*  
## Residuals 16 0.4768 0.0298   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Since P-value is 0.0107,which is less than the given significance level 0.05, so we reject the null hypothesis and conclude that there is sufficient evidence to indicate a difference in the mean amount of effluents discharged by the 4 plants at 0.05 significance level.