MAST30025 Week 12 and 13 Lab R Code

#REVISE THEM ON YOUR OWN TIME! #Question 1: #Part a:Estimate the difference between treatment effects, and test if it is significantly different from 0.

mydata = data.frame(response = c(7.5, 9.6, 8.4, 10.6, 9.9, 10.6,9.5, 9.7, 10.8, 11.9, 10.0, 12.9),treatment = factor(rep(c(1,2), c(6,6)))) #Ensure you convert these values as factors because we do not want R to assume it is at full rank!  
summary(lm(response ~ treatment, mydata))

##   
## Call:  
## lm(formula = response ~ treatment, data = mydata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.93333 -1.05000 0.08333 1.11667 2.10000   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.4333 0.5312 17.757 6.84e-09 \*\*\*  
## treatment2 1.3667 0.7513 1.819 0.0989 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.301 on 10 degrees of freedom  
## Multiple R-squared: 0.2486, Adjusted R-squared: 0.1735   
## F-statistic: 3.309 on 1 and 10 DF, p-value: 0.09893

#The estimated difference between treatment effects is 1.3667. The p-value for the test of τ2 − τ1 = 0 (against a general alternative) is 0.0989, which is borderline significant (but insignificant at a 5% level).

#Part b:Now suppose that it is discovered that the response can be affected by the season, and that the data was collected over a period of six months, in the order given by the table. That is, a month was spent collecting each row of the table. # We re-express the experiment by blocking: each month (row of the table) is considered one block, and we model the data as an additive two-factor model (the factors being the treatment and the block). Using this model, repeat your analysis. Is the estimate different? Is the p-value different?

mydata$block = factor(rep(1:6, 2))  
model = lm(response ~ treatment + block, mydata)  
summary(model)

##   
## Call:  
## lm(formula = response ~ treatment + block, data = mydata)  
##   
## Residuals:  
## 1 2 3 4 5 6 7 8   
## -0.31667 0.63333 -0.51667 0.03333 0.63333 -0.46667 0.31667 -0.63333   
## 9 10 11 12   
## 0.51667 -0.03333 -0.63333 0.46667   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.8167 0.5692 13.734 3.67e-05 \*\*\*  
## treatment2 1.3667 0.4302 3.176 0.02464 \*   
## block2 1.1500 0.7452 1.543 0.18343   
## block3 1.1000 0.7452 1.476 0.19994   
## block4 2.7500 0.7452 3.690 0.01414 \*   
## block5 1.4500 0.7452 1.946 0.10926   
## block6 3.2500 0.7452 4.361 0.00728 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7452 on 5 degrees of freedom  
## Multiple R-squared: 0.8768, Adjusted R-squared: 0.7289   
## F-statistic: 5.93 on 6 and 5 DF, p-value: 0.03493

anova(model)

## Analysis of Variance Table  
##   
## Response: response  
## Df Sum Sq Mean Sq F value Pr(>F)   
## treatment 1 5.6033 5.6033 10.0900 0.02464 \*  
## block 5 14.1567 2.8313 5.0984 0.04910 \*  
## Residuals 5 2.7767 0.5553   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#The estimated difference between treatment effects is still 1.3667. It has not changed be- cause the new design is a complete block design. The p-value for the test of τ2 − τ1 = 0 (against a general alternative) is now 0.0246, which is significant. The blocking has reduced the variability of our estimate and thus increased the power of our test.

#Question 4b:Give the design matrix X^A for a model with block and treatment effects (and an overall mean).

XA = matrix(0, nrow = 12, ncol = 11)  
XA[,1] = 1  
XA[cbind(1:12,rep(1:6,each=2)+1)] = 1  
XA[cbind(1:12,c(1,2,1,3,1,4,2,3,2,4,3,4)+7)] = 1  
XA

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

#Question 4c: Using this model, estimate τ1 − τ2, the difference between the first two treatment effects, and its variance. Write the variance estimate as s2cT (XAT XA)cc for a suitable c.

library(MASS)  
library(Matrix)  
y = c(1.245, 1.804, 2.468, 6.664, 5.573, -0.560,7.880, 10.469, 0.457, -3.621, -4.291, -9.384)  
bA = ginv(t(XA) %\*% XA) %\*% t(XA) %\*% y  
cA = c(0, 0, 0, 0, 0, 0, 0, 1, -1, 0, 0)  
n = 12  
rA = rankMatrix(XA)[1]  
rA

## [1] 9

s2A = sum((y - XA %\*% bA)^2)/(n-rA)  
s2A

## [1] 1.57602

t(cA) %\*% bA

## [,1]  
## [1,] -0.1675

t(cA) %\*% ginv(t(XA) %\*% XA) %\*% cA

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

s2A \* t(cA) %\*% ginv(t(XA) %\*% XA) %\*% cA

## [,1]  
## [1,] 1.57602

#Part d: Give the design matrix XB for a model with just treatment effects (and an overall mean).

XB = matrix(0, nrow = 12, ncol = 5)  
XB[,1] = 1  
XB[cbind(1:12,c(1,2,1,3,1,4,2,3,2,4,3,4)+1)] = 1  
XB

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

#Part e: Using this model, estimate τ1 − τ2, the difference between the first two treatment effects, and its variance. Write the variance estimate as s2cT (XBT XB)cc for a suitable c.

bB = ginv(t(XB) %\*% XB) %\*% t(XB) %\*% y  
cB = c(0, 1, -1, 0, 0)  
rB = rankMatrix(XB)[1]  
rB

## [1] 4

s2B = sum((y - XB %\*% bB)^2)/(n-rB)  
s2B

## [1] 24.85392

t(cB) %\*% bB

## [,1]  
## [1,] -0.285

t(cB) %\*% ginv(t(XB) %\*% XB) %\*% cB

## [,1]  
## [1,] 0.6666667

s2B \* t(cB) %\*% ginv(t(XB) %\*% XB) %\*% cB

## [,1]  
## [1,] 16.56928

#Part f: Show that when going from model A (BIBD) to model B (CRD) the term cT (XT X)cc decreases, but s2 increases markedly. What does this indicate? See above. This indicates that the blocks are effective in their intended purpose (reducing variance).

#Part g: Is your estimate for τ1 − τ2 the same or different for the two models? Why? The estimates are different. This is because in the BIBD, the blocks are not orthogonal to the treatments.

#Quesiton 6 #Part a: Which design is a block design! SOLUTION: The first design is a complete block design: each treatment appears exactly once in each block.

#Part b: Write down the design matrix for each design. Hence show that τ2 − τ1 is estimable in each case.

Xa = matrix(0, 12, 8)  
Xa[,1] = 1  
Xa[1:3,2] = 1  
Xa[4:6,3] = 1  
Xa[7:9,4] = 1  
Xa[10:12,5] = 1  
Xa[,6] = c(1,0,0, 0,1,0, 1,0,0, 0,0,1)  
Xa[,7] <- c(0,1,0, 1,0,0, 0,0,1, 0,1,0)  
Xa[,8] <- c(0,0,1, 0,0,1, 0,1,0, 1,0,0)  
Xb <- Xa  
Xb[,6] = c(1,1,0, 0,0,0, 0,0,1, 1,0,0)  
Xb[,7] = c(0,0,1, 1,1,0, 0,0,0, 0,1,0)  
Xb[,8] = c(0,0,0, 0,0,1, 1,1,0, 0,0,1)  
Xa

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

Xb

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

library(MASS)  
t = c(0,0,0,0,0,-1,1,0)  
round(t(t) %\*% ginv(t(Xa) %\*% Xa) %\*% t(Xa) %\*% Xa, 10)

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

round(t(t) %\*% ginv(t(Xb) %\*% Xb) %\*% t(Xb) %\*% Xb, 10)

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

#Part c: For each design, in terms of the unknown error variance σ2, what is the variance of the estimator for τ2 − τ1, the difference between the first two treatment effects? # Based on this, which design is better?

t(t) %\*% ginv(t(Xa) %\*% Xa) %\*% t

## [,1]  
## [1,] 0.5

t(t) %\*% ginv(t(Xb) %\*% Xb) %\*% t

## [,1]  
## [1,] 0.6666667

#The variance of the estimator is (1/2)σ2 for the first design and (2/3)σ2 for the second design. Clearly we prefer the CBD, as it gives the smaller variance.