Stat501\_Homework5

Kelby Kies

3/24/2021

# Question 1:

## Part b.)

### i.) Because the variables in each of the two populations may not be multivariate normally distributed, we will find the which transforms the data such that they are so. For a grid of values, where each component takes values in {0, 1/4, 1/3, 1/2, 1, 2, 3, 4}, find the which maximizes the joint likelihood of (from among the grid) give the observations. [10 points]

library(car)  
# Read in the data  
colleges <- read.table('~/Desktop/stat\_501/Colleges.txt', sep = '\t', header = T)  
  
# Box Cox Function  
box\_cox <- function(w, lambda, eps = 1e-03)   
 { if (abs(lambda) < eps)  
 log(w)  
 else   
 ((w^lambda) - 1)/lambda  
}  
  
# Log Likelihood function  
llhd <- function(lambda, x, y)  
 {  
 # Calculate means mu and v:  
 mu <- mean(box\_cox(x, lambda))  
 v <- mean(box\_cox(y, lambda))  
 sigma\_x <- var(box\_cox(x, lambda))  
 sigma\_y <- var(box\_cox(y, lambda))  
   
 length(x)/2 \* log(sigma\_x) + length(y)/2 \* log(sigma\_y) + ((lambda - 1) \* (sum(log(x)) + sum(log(y))))  
}  
  
  
library(dplyr)  
X\_df <- dplyr::filter(colleges, colleges$School\_Type == 'Lib Arts') %>% select(SAT, Acceptance, X..Student, Top.10., X.PhD, Grad.)  
Y\_df <- dplyr::filter(colleges, colleges$School\_Type == 'Univ') %>% select(SAT, Acceptance, X..Student, Top.10., X.PhD, Grad.)  
  
  
final\_grid <- data.frame(lamda\_values = c(0, 1/2, 1/3, 1/4, 1, 2, 3, 4),   
 SAT = c(llhd(lambda = 0, x = X\_df$SAT, y = Y\_df$SAT), llhd(lambda = 1/4, x = X\_df$SAT, y = Y\_df$SAT),   
 llhd(lambda= 1/3, x = X\_df$SAT, y = Y\_df$SAT), llhd(lambda = 1/2, x = X\_df$SAT, y = Y\_df$SAT),   
 llhd(lambda = 1, x = X\_df$SAT, y = Y\_df$SAT), llhd(lambda = 2, x = X\_df$SAT, y = Y\_df$SAT),   
 llhd(lambda = 3, x = X\_df$SAT, y = Y\_df$SAT), llhd(lambda = 4, x = X\_df$SAT, y = Y\_df$SAT)),  
   
 Acceptance = c(llhd(lambda = 0, x = X\_df$Acceptance, y = Y\_df$Acceptance), llhd(lambda = 1/4, x = X\_df$Acceptance, y = Y\_df$Acceptance), llhd(lambda= 1/3, x = X\_df$Acceptance, y = Y\_df$Acceptance), llhd(lambda = 1/2, x = X\_df$Acceptance, y = Y\_df$Acceptance), llhd(lambda = 1, x = X\_df$Acceptance, y = Y\_df$Acceptance), llhd(lambda = 2, x = X\_df$Acceptance, y = Y\_df$Acceptance), llhd(lambda = 3, x = X\_df$Acceptance, y = Y\_df$Acceptance), llhd(lambda = 4, x = X\_df$Acceptance, y = Y\_df$Acceptance)),  
   
 X..Student = c(llhd(lambda = 0, x = X\_df$X..Student, y = Y\_df$X..Student), llhd(lambda = 1/4, x = X\_df$X..Student, y = Y\_df$X..Student), llhd(lambda= 1/3, x = X\_df$X..Student, y = Y\_df$X..Student), llhd(lambda = 1/2, x = X\_df$X..Student, y = Y\_df$X..Student), llhd(lambda = 1, x = X\_df$X..Student, y = Y\_df$X..Student), llhd(lambda = 2, x = X\_df$X..Student, y = Y\_df$X..Student), llhd(lambda = 3, x = X\_df$X..Student, y = Y\_df$X..Student), llhd(lambda = 4, x = X\_df$X..Student, y = Y\_df$X..Student)),  
   
 Top.10. = c(llhd(lambda = 0, x = X\_df$Top.10., y = Y\_df$Top.10.), llhd(lambda = 1/4, x = X\_df$Top.10., y = Y\_df$Top.10.), llhd(lambda= 1/3, x = X\_df$Top.10., y = Y\_df$Top.10.), llhd(lambda = 1/2, x = X\_df$Top.10., y = Y\_df$Top.10.), llhd(lambda = 1, x = X\_df$Top.10., y = Y\_df$Top.10.), llhd(lambda = 2, x = X\_df$Top.10., y = Y\_df$Top.10.), llhd(lambda = 3, x = X\_df$Top.10., y = Y\_df$Top.10.), llhd(lambda = 4, x = X\_df$Top.10., y = Y\_df$Top.10.)),  
 X.PhD = c(llhd(lambda = 0, x = X\_df$X.PhD, y = Y\_df$X.PhD), llhd(lambda = 1/4, x = X\_df$X.PhD, y = Y\_df$X.PhD),   
 llhd(lambda= 1/3, x = X\_df$X.PhD, y = Y\_df$X.PhD), llhd(lambda = 1/2, x = X\_df$X.PhD, y = Y\_df$X.PhD),   
 llhd(lambda = 1, x = X\_df$X.PhD, y = Y\_df$X.PhD), llhd(lambda = 2, x = X\_df$X.PhD, y = Y\_df$X.PhD),   
 llhd(lambda = 3, x = X\_df$X.PhD, y = Y\_df$X.PhD), llhd(lambda = 4, x = X\_df$X.PhD, y = Y\_df$X.PhD)),  
   
 Grad. = c(llhd(lambda = 0, x = X\_df$Grad., y = Y\_df$Grad.), llhd(lambda = 1/4, x = X\_df$Grad., y = Y\_df$Grad.),   
 llhd(lambda= 1/3, x = X\_df$Grad., y = Y\_df$Grad.), llhd(lambda = 1/2, x = X\_df$Grad., y = Y\_df$Grad.),   
 llhd(lambda = 1, x = X\_df$Grad., y = Y\_df$Grad.), llhd(lambda = 2, x = X\_df$Grad., y = Y\_df$Grad.),   
 llhd(lambda = 3, x = X\_df$Grad., y = Y\_df$Grad.), llhd(lambda = 4, x = X\_df$Grad., y = Y\_df$Grad.))  
 )  
  
print(final\_grid)

## lamda\_values SAT Acceptance X..Student Top.10. X.PhD Grad.  
## 1 0.0000000 -510.7680 -230.46020 -580.85791 -304.48270 -343.4347 -340.79900  
## 2 0.5000000 -332.3428 -141.12268 -324.20421 -198.01712 -231.9765 -230.70999  
## 3 0.3333333 -272.8660 -111.26682 -238.56475 -162.50616 -194.8109 -194.00698  
## 4 0.2500000 -153.9098 -51.44105 -67.14815 -91.45079 -120.4604 -120.59105  
## 5 1.0000000 202.9797 128.92997 448.24228 121.97942 102.7428 99.73459  
## 6 2.0000000 916.8519 493.41091 1484.00564 549.98067 549.8008 540.72272  
## 7 3.0000000 1630.8471 862.15749 2524.92693 979.37929 997.6497 982.13307  
## 8 4.0000000 2344.9632 1234.28314 3569.00345 1410.00438 1446.1830 1423.93455

Printed above is my final grid that contains the calculated log likelihood function for each lambda value. I am not sure if I did this correctly, but I wasn’t sure how to optimize the log liklihood function so that I am trying every different value (0,1/4,1/3,1/2,1,2,3,4) for each different position in the vector.

Based on what I have here it looks like the MLE is

### ii.) With the transformed data, compare the mean values of the (transformed) SAT, % acceptance, cost per student, per cent of students in top 10 per cent of HS graduating class, per cent faculty with Ph.D.s and graduation rate, for the liberal arts vis-a-vis public universities? Are any of these means equal? [10 points]

# Transform the data  
trans\_SAT <- box\_cox(w = colleges$SAT, lambda = 4)  
trans\_Accept <- box\_cox(w = colleges$Acceptance, lambda = 4)  
trans\_XStudent <- box\_cox(w = colleges$X..Student, lambda = 4)  
trans\_top10 <- box\_cox(w = colleges$Top.10., lambda = 4)  
trans\_XPhD <- box\_cox(w = colleges$X.PhD, lambda = 4)  
trans\_Grad <- box\_cox(w = colleges$Grad., lambda = 4)  
  
transformed\_dta <- data.frame(cbind(colleges$School\_Type, as.numeric(trans\_SAT), as.numeric(trans\_Accept), as.numeric(trans\_XStudent), as.numeric(trans\_top10), as.numeric(trans\_XPhD), as.numeric(trans\_Grad)))  
#   
# lib\_colleges <- filter(transformed\_dta, transformed\_dta$V1 == 'Lib Arts')  
# Univ <- filter(transformed\_dta, transformed\_dta$V1 != 'Lib Arts')  
# compare the means of the variables

### iii.) Setting the False Discovery Rate at q = 0.05,which of the six variables have a significant difference between the liberal arts colleges and public universities. Interpret the results. 10 points]

# Add a group varibale  
colleges2 <- colleges %>% mutate(group = ifelse(colleges$School\_Type == "Lib Arts", 0, 1))  
  
fit.lm <- lm(cbind(trans\_SAT, trans\_Accept, trans\_XStudent, trans\_top10, trans\_XPhD, trans\_Grad)~group , data = colleges2)  
  
fit.Manova <- Manova(fit.lm)  
#summary(fit.Manova)

# Question 2:

library(sas7bdat)  
library(car)  
#source('~/Desktop/stat\_501/manova.R')  
psych\_sas7bdat <- read.sas7bdat('~/Desktop/stat\_501/hw5/psych.sas7bdat', debug=FALSE)  
psych\_sas7bdat$PROG <- as.factor(psych\_sas7bdat$PROG)

## Part a.) Fit a linear model to the above and all the variables. Ignore interactions for now. Assume that the first level in the categorical variable has no additional effect (i.e. ) in the contrast. Summarize the results. [10 points]

# Assume that the first level in PROG has no additional effect in the contrast.   
psych\_sas7bdat$PROG <- C(object = psych\_sas7bdat$PROG, contr = contr.treatment(n = 3, base = 2))  
psych\_lm\_a <- lm(cbind(LOCUS\_OF\_CONTROL, SELF\_CONCEPT, MOTIVATION) ~ READ + WRITE + SCIENCE + PROG, data = psych\_sas7bdat)  
psych\_lm\_a

##   
## Call:  
## lm(formula = cbind(LOCUS\_OF\_CONTROL, SELF\_CONCEPT, MOTIVATION) ~   
## READ + WRITE + SCIENCE + PROG, data = psych\_sas7bdat)  
##   
## Coefficients:  
## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## (Intercept) -1.496970 -0.095858 -0.950513   
## READ 0.012505 0.001308 0.009674   
## WRITE 0.012145 -0.004293 0.017535   
## SCIENCE 0.005761 0.005306 -0.009001   
## PROG1 -0.127795 -0.276483 -0.360329   
## PROG3 0.123875 0.146876 0.259367

psych\_lm\_a\_manova <- Manova(psych\_lm\_a)  
summary(psych\_lm\_a\_manova)

##   
## Type II MANOVA Tests:  
##   
## Sum of squares and products for error:  
## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## LOCUS\_OF\_CONTROL 218.85624 34.14870 35.93761  
## SELF\_CONCEPT 34.14870 282.04029 77.83401  
## MOTIVATION 35.93761 77.83401 344.36143  
##   
## ------------------------------------------  
##   
## Term: READ   
##   
## Sum of squares and products for the hypothesis:  
## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## LOCUS\_OF\_CONTROL 4.1681596 0.43586639 3.2244794  
## SELF\_CONCEPT 0.4358664 0.04557875 0.3371853  
## MOTIVATION 3.2244794 0.33718531 2.4944504  
##   
## Multivariate Tests: READ  
## Df test stat approx F num Df den Df Pr(>F)   
## Pillai 1 0.0235748 4.764416 3 592 0.0027266 \*\*  
## Wilks 1 0.9764252 4.764416 3 592 0.0027266 \*\*  
## Hotelling-Lawley 1 0.0241440 4.764416 3 592 0.0027266 \*\*  
## Roy 1 0.0241440 4.764416 3 592 0.0027266 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## ------------------------------------------  
##   
## Term: WRITE   
##   
## Sum of squares and products for the hypothesis:  
## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## LOCUS\_OF\_CONTROL 4.725243 -1.6704333 6.822473  
## SELF\_CONCEPT -1.670433 0.5905193 -2.411831  
## MOTIVATION 6.822473 -2.4118306 9.850527  
##   
## Multivariate Tests: WRITE  
## Df test stat approx F num Df den Df Pr(>F)   
## Pillai 1 0.0526060 10.95734 3 592 5.1862e-07 \*\*\*  
## Wilks 1 0.9473940 10.95734 3 592 5.1862e-07 \*\*\*  
## Hotelling-Lawley 1 0.0555271 10.95734 3 592 5.1862e-07 \*\*\*  
## Roy 1 0.0555271 10.95734 3 592 5.1862e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## ------------------------------------------  
##   
## Term: SCIENCE   
##   
## Sum of squares and products for the hypothesis:  
## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## LOCUS\_OF\_CONTROL 0.9224864 0.8495491 -1.441248  
## SELF\_CONCEPT 0.8495491 0.7823788 -1.327294  
## MOTIVATION -1.4412481 -1.3272945 2.251736  
##   
## Multivariate Tests: SCIENCE  
## Df test stat approx F num Df den Df Pr(>F)   
## Pillai 1 0.0165945 3.329911 3 592 0.019305 \*  
## Wilks 1 0.9834055 3.329911 3 592 0.019305 \*  
## Hotelling-Lawley 1 0.0168745 3.329911 3 592 0.019305 \*  
## Roy 1 0.0168745 3.329911 3 592 0.019305 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## ------------------------------------------  
##   
## Term: PROG   
##   
## Sum of squares and products for the hypothesis:  
## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## LOCUS\_OF\_CONTROL 5.029620 8.290863 12.25844  
## SELF\_CONCEPT 8.290863 14.218385 20.61640  
## MOTIVATION 12.258441 20.616397 30.18084  
##   
## Multivariate Tests: PROG  
## Df test stat approx F num Df den Df Pr(>F)   
## Pillai 2 0.1086487 11.35496 6 1186 2.2795e-12 \*\*\*  
## Wilks 2 0.8914383 11.67076 6 1184 9.8057e-13 \*\*\*  
## Hotelling-Lawley 2 0.1216850 11.98597 6 1182 4.2255e-13 \*\*\*  
## Roy 2 0.1208775 23.89346 3 593 1.3102e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Based on the Manova summary, all of our predictor variables have a significant effect on predicting locus of control, self-concept and motivation of high school students at the 0.05 level.The PROG and WRTIE variables seem to have stronger effect at the 0.0001 level, while the SCIENCE variable has the weakest effect at the 0.05 level, although still significant.

## Part b.) Refit the model but after dropping the dependent variables on the test scores of writing and science. Summarize the results. [8 points]

psych\_lm\_b <- lm(cbind(LOCUS\_OF\_CONTROL, SELF\_CONCEPT, MOTIVATION) ~ READ + PROG, data = psych\_sas7bdat)  
  
psych\_lm\_b\_manova <- Manova(psych\_lm\_b)  
summary(psych\_lm\_b\_manova)

##   
## Type II MANOVA Tests:  
##   
## Sum of squares and products for error:  
## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## LOCUS\_OF\_CONTROL 225.90750 33.57873 41.58308  
## SELF\_CONCEPT 33.57873 283.15145 74.86532  
## MOTIVATION 41.58308 74.86532 354.80754  
##   
## ------------------------------------------  
##   
## Term: READ   
##   
## Sum of squares and products for the hypothesis:  
## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## LOCUS\_OF\_CONTROL 33.616976 3.1957148 20.255356  
## SELF\_CONCEPT 3.195715 0.3037927 1.925525  
## MOTIVATION 20.255356 1.9255254 12.204532  
##   
## Multivariate Tests: READ  
## Df test stat approx F num Df den Df Pr(>F)   
## Pillai 1 0.1439298 33.28946 3 594 < 2.22e-16 \*\*\*  
## Wilks 1 0.8560702 33.28946 3 594 < 2.22e-16 \*\*\*  
## Hotelling-Lawley 1 0.1681286 33.28946 3 594 < 2.22e-16 \*\*\*  
## Roy 1 0.1681286 33.28946 3 594 < 2.22e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## ------------------------------------------  
##   
## Term: PROG   
##   
## Sum of squares and products for the hypothesis:  
## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## LOCUS\_OF\_CONTROL 5.652949 8.49380 13.37923  
## SELF\_CONCEPT 8.493800 13.90261 20.98700  
## MOTIVATION 13.379228 20.98700 32.35107  
##   
## Multivariate Tests: PROG  
## Df test stat approx F num Df den Df Pr(>F)   
## Pillai 2 0.1121308 11.78009 6 1190 7.2868e-13 \*\*\*  
## Wilks 2 0.8880617 12.10849 6 1188 3.0304e-13 \*\*\*  
## Hotelling-Lawley 2 0.1258310 12.43630 6 1186 1.2626e-13 \*\*\*  
## Roy 2 0.1240839 24.60996 3 595 5.0622e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

After refitting the linear model and removing the effect for the WRITE and SCIENCE variable, we see that both PROG and READ have a strong effect on the locus of control, self-concept and motivation at the 0.0001 significance level. By removing these variable, the READ effect became stronger at predicting locus of control, self-concept and motivation.

## Part c.) Is there a significant evidence that the writing and science test scores are related to the psychological profiles? [2 points]

#LRT  
library(stats)  
anova(psych\_lm\_a, psych\_lm\_b, test = "Wilks")

## Analysis of Variance Table  
##   
## Model 1: cbind(LOCUS\_OF\_CONTROL, SELF\_CONCEPT, MOTIVATION) ~ READ + WRITE +   
## SCIENCE + PROG  
## Model 2: cbind(LOCUS\_OF\_CONTROL, SELF\_CONCEPT, MOTIVATION) ~ READ + PROG  
## Res.Df Df Gen.var. Wilks approx F num Df den Df Pr(>F)   
## 1 594 0.45201   
## 2 596 2 0.46105 0.93285 6.9794 6 1184 2.618e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Here we see that the WRITE and SCIENCE terms are needed to significantly improve the model based on the 0.05 significance level because the P-value here is .Also the Wilks statistic is very large and close to 1. We will only reject the null if the wilk’s statistic is small. Therefore we fail to reject the null hypothesis on the 0.05 significance level.

## Part d.) From the model in your results in (c) above, test simultaneously for whether there is a difference in psychological profiles between Program 1 and 2 and between Program 2 and 3. [10 points]

# Decide which model from the LRT in partc   
print("Here is my Beta Matrix:")

## [1] "Here is my Beta Matrix:"

print(coef(psych\_lm\_a))

## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## (Intercept) -1.496969664 -0.095857801 -0.950512536  
## READ 0.012504619 0.001307614 0.009673547  
## WRITE 0.012145048 -0.004293428 0.017535449  
## SCIENCE 0.005761477 0.005305940 -0.009001453  
## PROG1 -0.127795079 -0.276483394 -0.360329390  
## PROG3 0.123875431 0.146875797 0.259366647

# For that model, SIMULtaneously test whether there is a difference in psychological profiles between 1.) prog 1 and 2, 2.) prog 2 & 3  
# Test by testing for a diff in beta coefficinets : C\* Beta  
  
C <- matrix(c(0,0,0,0,1,0,0,0,0,0,1,-1), ncol = 6, by = T)  
C

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

psych\_lm\_a\_hyp <- linearHypothesis(model = psych\_lm\_a, hypothesis.matrix = C)  
psych\_lm\_a\_hyp

##   
## Sum of squares and products for the hypothesis:  
## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## LOCUS\_OF\_CONTROL 5.029620 8.290863 12.25844  
## SELF\_CONCEPT 8.290863 14.218385 20.61640  
## MOTIVATION 12.258441 20.616397 30.18084  
##   
## Sum of squares and products for error:  
## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## LOCUS\_OF\_CONTROL 218.85624 34.14870 35.93761  
## SELF\_CONCEPT 34.14870 282.04029 77.83401  
## MOTIVATION 35.93761 77.83401 344.36143  
##   
## Multivariate Tests:   
## Df test stat approx F num Df den Df Pr(>F)   
## Pillai 2 0.1086487 11.35496 6 1186 2.2795e-12 \*\*\*  
## Wilks 2 0.8914383 11.67076 6 1184 9.8057e-13 \*\*\*  
## Hotelling-Lawley 2 0.1216850 11.98597 6 1182 4.2255e-13 \*\*\*  
## Roy 2 0.1208775 23.89346 3 593 1.3102e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Here I have set up a linearhypothesis() function that is testing to see if there is a difference in psychological profiles (Locus, movitvation, self-concept) between the groups Prog 1&2 and 2&3 simultaneously. If I did this correctly, then there is a significant difference in profiles between the 3 programs based on the 0.001 significance level.

## Part e.) Test the null hypothesis that the coefficient for the written test scores with locus of control as the outcome is equal to the corresponding coefficient with self concept as the outcome. [10 points]

print("Here is my Beta Matrix:")

## [1] "Here is my Beta Matrix:"

print(coef(psych\_lm\_a))

## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## (Intercept) -1.496969664 -0.095857801 -0.950512536  
## READ 0.012504619 0.001307614 0.009673547  
## WRITE 0.012145048 -0.004293428 0.017535449  
## SCIENCE 0.005761477 0.005305940 -0.009001453  
## PROG1 -0.127795079 -0.276483394 -0.360329390  
## PROG3 0.123875431 0.146875797 0.259366647

print("Here is my C Matrix:")

## [1] "Here is my C Matrix:"

C\_e <- matrix(c(0,0,1,0,0,0), nrow = 1, by = T)  
  
print("Here is my M Matrix:")

## [1] "Here is my M Matrix:"

M\_e <- matrix(c(1,-1,0), nrow = 3, by = T)  
  
part\_e\_answer <- linearHypothesis(model = psych\_lm\_a, hypothesis.matrix = C\_e, P = M\_e)  
part\_e\_answer

##   
## Response transformation matrix:  
## [,1]  
## LOCUS\_OF\_CONTROL 1  
## SELF\_CONCEPT -1  
## MOTIVATION 0  
##   
## Sum of squares and products for the hypothesis:  
## [,1]  
## [1,] 8.656629  
##   
## Sum of squares and products for error:  
## [,1]  
## [1,] 432.5991  
##   
## Multivariate Tests:   
## Df test stat approx F num Df den Df Pr(>F)   
## Pillai 1 0.0196182 11.88638 1 594 0.00060546 \*\*\*  
## Wilks 1 0.9803818 11.88638 1 594 0.00060546 \*\*\*  
## Hotelling-Lawley 1 0.0200107 11.88638 1 594 0.00060546 \*\*\*  
## Roy 1 0.0200107 11.88638 1 594 0.00060546 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Here I have set up a linearhypothesis() function that is testing if the coefficient for the written test scores with locus of control as the outcome is equal to the corresponding coefficient with self concept as the outcome. Here we can see that there is a significant difference in these coefficients when tested at the significance level 0.001.

## Part f.) Now, test the null hypothesis that the coefficient for science scores for locus of control is equal to the corresponding coefficient for science for the self concept variable, and that the coefficient for the written scores for locus of control is equal to the coefficient for the written scores for self concept. [10 points]

C\_f <- matrix(c(0, 0, 1, 1, 0, 0), nrow = 1, by = T)  
M\_f <- matrix(c(1,-1,0), nrow = 3, by = T)  
  
final\_answerf <- linearHypothesis(model = psych\_lm\_a, hypothesis.matrix = C\_f, P = M\_f)  
final\_answerf

##   
## Response transformation matrix:  
## [,1]  
## LOCUS\_OF\_CONTROL 1  
## SELF\_CONCEPT -1  
## MOTIVATION 0  
##   
## Sum of squares and products for the hypothesis:  
## [,1]  
## [1,] 5.579309  
##   
## Sum of squares and products for error:  
## [,1]  
## [1,] 432.5991  
##   
## Multivariate Tests:   
## Df test stat approx F num Df den Df Pr(>F)   
## Pillai 1 0.0127330 7.660925 1 594 0.0058189 \*\*  
## Wilks 1 0.9872670 7.660925 1 594 0.0058189 \*\*  
## Hotelling-Lawley 1 0.0128972 7.660925 1 594 0.0058189 \*\*  
## Roy 1 0.0128972 7.660925 1 594 0.0058189 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Here I have set up a linearhypothesis() function that is testing the following: 1.) if the coefficient for science scores for locus of control is equal than the science score for selc concept 2.) if the coefficient for the written test scores with locus of control is equal to the coefficient for the written scores for self concept Here we can see that there is a significant difference in both of these coefficient comparisons at the significance level 0.01.

## Part g.) Depending on the results from (c), fit a linear model with all interactions included. Interpret the results. [10 points]

psych\_lm\_g <- lm(cbind(LOCUS\_OF\_CONTROL, SELF\_CONCEPT, MOTIVATION) ~ (READ + WRITE + SCIENCE + PROG)^4, data = psych\_sas7bdat)  
psych\_lm\_g

##   
## Call:  
## lm(formula = cbind(LOCUS\_OF\_CONTROL, SELF\_CONCEPT, MOTIVATION) ~   
## (READ + WRITE + SCIENCE + PROG)^4, data = psych\_sas7bdat)  
##   
## Coefficients:  
## LOCUS\_OF\_CONTROL SELF\_CONCEPT MOTIVATION  
## (Intercept) 1.499e+00 5.752e+00 1.885e+00  
## READ -5.327e-02 -1.416e-01 -5.367e-03  
## WRITE -7.331e-02 -1.211e-01 -8.957e-02  
## SCIENCE -1.921e-02 -7.847e-02 -5.774e-02  
## PROG1 2.117e+00 -5.757e+00 -7.157e+00  
## PROG3 -3.913e+00 -1.466e+01 -1.063e+01  
## READ:WRITE 1.678e-03 2.751e-03 1.359e-03  
## READ:SCIENCE 6.145e-04 2.035e-03 1.241e-04  
## READ:PROG1 2.440e-02 7.830e-02 8.935e-02  
## READ:PROG3 1.033e-01 2.688e-01 5.194e-02  
## WRITE:SCIENCE 9.947e-04 1.610e-03 1.746e-03  
## WRITE:PROG1 -2.540e-02 1.715e-01 1.347e-01  
## WRITE:PROG3 1.115e-01 3.244e-01 3.070e-01  
## SCIENCE:PROG1 -1.798e-01 1.047e-01 1.333e-01  
## SCIENCE:PROG3 5.088e-02 2.382e-01 2.383e-01  
## READ:WRITE:SCIENCE -1.946e-05 -3.785e-05 -1.913e-05  
## READ:WRITE:PROG1 -7.181e-04 -2.794e-03 -1.636e-03  
## READ:WRITE:PROG3 -2.376e-03 -5.758e-03 -3.116e-03  
## READ:SCIENCE:PROG1 2.059e-03 -8.331e-04 -1.742e-03  
## READ:SCIENCE:PROG3 -1.394e-03 -4.155e-03 -1.425e-03  
## WRITE:SCIENCE:PROG1 3.192e-03 -3.481e-03 -2.355e-03  
## WRITE:SCIENCE:PROG3 -1.630e-03 -5.053e-03 -5.995e-03  
## READ:WRITE:SCIENCE:PROG1 -3.581e-05 4.475e-05 2.697e-05  
## READ:WRITE:SCIENCE:PROG3 3.439e-05 8.617e-05 5.943e-05

summary(psych\_lm\_g)

## Response LOCUS\_OF\_CONTROL :  
##   
## Call:  
## lm(formula = LOCUS\_OF\_CONTROL ~ (READ + WRITE + SCIENCE + PROG)^4,   
## data = psych\_sas7bdat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.04523 -0.38841 -0.02312 0.37595 1.94187   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 1.499e+00 3.366e+00 0.445 0.656  
## READ -5.327e-02 7.656e-02 -0.696 0.487  
## WRITE -7.331e-02 7.714e-02 -0.950 0.342  
## SCIENCE -1.921e-02 6.419e-02 -0.299 0.765  
## PROG1 2.117e+00 5.410e+00 0.391 0.696  
## PROG3 -3.913e+00 6.960e+00 -0.562 0.574  
## READ:WRITE 1.678e-03 1.592e-03 1.054 0.292  
## READ:SCIENCE 6.145e-04 1.305e-03 0.471 0.638  
## READ:PROG1 2.440e-02 1.264e-01 0.193 0.847  
## READ:PROG3 1.033e-01 1.404e-01 0.735 0.462  
## WRITE:SCIENCE 9.947e-04 1.375e-03 0.724 0.470  
## WRITE:PROG1 -2.540e-02 1.198e-01 -0.212 0.832  
## WRITE:PROG3 1.115e-01 1.511e-01 0.738 0.461  
## SCIENCE:PROG1 -1.798e-01 1.151e-01 -1.562 0.119  
## SCIENCE:PROG3 5.088e-02 1.408e-01 0.361 0.718  
## READ:WRITE:SCIENCE -1.946e-05 2.572e-05 -0.756 0.450  
## READ:WRITE:PROG1 -7.181e-04 2.525e-03 -0.284 0.776  
## READ:WRITE:PROG3 -2.376e-03 2.791e-03 -0.851 0.395  
## READ:SCIENCE:PROG1 2.059e-03 2.345e-03 0.878 0.380  
## READ:SCIENCE:PROG3 -1.394e-03 2.565e-03 -0.543 0.587  
## WRITE:SCIENCE:PROG1 3.192e-03 2.326e-03 1.373 0.170  
## WRITE:SCIENCE:PROG3 -1.630e-03 2.906e-03 -0.561 0.575  
## READ:WRITE:SCIENCE:PROG1 -3.581e-05 4.290e-05 -0.835 0.404  
## READ:WRITE:SCIENCE:PROG3 3.439e-05 4.901e-05 0.702 0.483  
##   
## Residual standard error: 0.609 on 576 degrees of freedom  
## Multiple R-squared: 0.2062, Adjusted R-squared: 0.1745   
## F-statistic: 6.504 on 23 and 576 DF, p-value: < 2.2e-16  
##   
##   
## Response SELF\_CONCEPT :  
##   
## Call:  
## lm(formula = SELF\_CONCEPT ~ (READ + WRITE + SCIENCE + PROG)^4,   
## data = psych\_sas7bdat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.34313 -0.43875 -0.00875 0.45591 2.20948   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.752e+00 3.792e+00 1.517 0.1298   
## READ -1.416e-01 8.625e-02 -1.642 0.1011   
## WRITE -1.211e-01 8.690e-02 -1.394 0.1640   
## SCIENCE -7.847e-02 7.231e-02 -1.085 0.2783   
## PROG1 -5.757e+00 6.094e+00 -0.945 0.3453   
## PROG3 -1.466e+01 7.840e+00 -1.870 0.0620 .  
## READ:WRITE 2.751e-03 1.793e-03 1.535 0.1254   
## READ:SCIENCE 2.035e-03 1.470e-03 1.385 0.1666   
## READ:PROG1 7.830e-02 1.423e-01 0.550 0.5825   
## READ:PROG3 2.688e-01 1.582e-01 1.699 0.0898 .  
## WRITE:SCIENCE 1.610e-03 1.548e-03 1.040 0.2989   
## WRITE:PROG1 1.715e-01 1.350e-01 1.271 0.2044   
## WRITE:PROG3 3.244e-01 1.702e-01 1.906 0.0571 .  
## SCIENCE:PROG1 1.047e-01 1.296e-01 0.807 0.4198   
## SCIENCE:PROG3 2.382e-01 1.587e-01 1.502 0.1337   
## READ:WRITE:SCIENCE -3.785e-05 2.897e-05 -1.306 0.1920   
## READ:WRITE:PROG1 -2.794e-03 2.844e-03 -0.982 0.3263   
## READ:WRITE:PROG3 -5.758e-03 3.144e-03 -1.831 0.0676 .  
## READ:SCIENCE:PROG1 -8.331e-04 2.641e-03 -0.315 0.7525   
## READ:SCIENCE:PROG3 -4.155e-03 2.890e-03 -1.438 0.1511   
## WRITE:SCIENCE:PROG1 -3.481e-03 2.620e-03 -1.329 0.1844   
## WRITE:SCIENCE:PROG3 -5.053e-03 3.274e-03 -1.544 0.1232   
## READ:WRITE:SCIENCE:PROG1 4.475e-05 4.833e-05 0.926 0.3549   
## READ:WRITE:SCIENCE:PROG3 8.617e-05 5.521e-05 1.561 0.1191   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.686 on 576 degrees of freedom  
## Multiple R-squared: 0.09076, Adjusted R-squared: 0.05446   
## F-statistic: 2.5 on 23 and 576 DF, p-value: 0.0001512  
##   
##   
## Response MOTIVATION :  
##   
## Call:  
## lm(formula = MOTIVATION ~ (READ + WRITE + SCIENCE + PROG)^4,   
## data = psych\_sas7bdat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.31781 -0.50971 -0.00794 0.49580 2.24623   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.885e+00 4.205e+00 0.448 0.6542   
## READ -5.367e-03 9.565e-02 -0.056 0.9553   
## WRITE -8.957e-02 9.637e-02 -0.929 0.3531   
## SCIENCE -5.774e-02 8.020e-02 -0.720 0.4718   
## PROG1 -7.157e+00 6.759e+00 -1.059 0.2901   
## PROG3 -1.063e+01 8.695e+00 -1.223 0.2219   
## READ:WRITE 1.359e-03 1.988e-03 0.684 0.4945   
## READ:SCIENCE 1.241e-04 1.630e-03 0.076 0.9393   
## READ:PROG1 8.935e-02 1.579e-01 0.566 0.5716   
## READ:PROG3 5.194e-02 1.754e-01 0.296 0.7672   
## WRITE:SCIENCE 1.746e-03 1.717e-03 1.017 0.3097   
## WRITE:PROG1 1.347e-01 1.497e-01 0.899 0.3688   
## WRITE:PROG3 3.070e-01 1.887e-01 1.627 0.1043   
## SCIENCE:PROG1 1.333e-01 1.438e-01 0.927 0.3541   
## SCIENCE:PROG3 2.383e-01 1.760e-01 1.354 0.1761   
## READ:WRITE:SCIENCE -1.913e-05 3.213e-05 -0.595 0.5520   
## READ:WRITE:PROG1 -1.636e-03 3.155e-03 -0.519 0.6042   
## READ:WRITE:PROG3 -3.116e-03 3.487e-03 -0.894 0.3719   
## READ:SCIENCE:PROG1 -1.742e-03 2.929e-03 -0.595 0.5523   
## READ:SCIENCE:PROG3 -1.425e-03 3.205e-03 -0.444 0.6569   
## WRITE:SCIENCE:PROG1 -2.355e-03 2.905e-03 -0.811 0.4179   
## WRITE:SCIENCE:PROG3 -5.995e-03 3.631e-03 -1.651 0.0993 .  
## READ:WRITE:SCIENCE:PROG1 2.697e-05 5.360e-05 0.503 0.6150   
## READ:WRITE:SCIENCE:PROG3 5.943e-05 6.123e-05 0.971 0.3321   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7608 on 576 degrees of freedom  
## Multiple R-squared: 0.177, Adjusted R-squared: 0.1441   
## F-statistic: 5.385 on 23 and 576 DF, p-value: 4.564e-14

anova(psych\_lm\_g)

## Analysis of Variance Table  
##   
## Df Pillai approx F num Df den Df Pr(>F)   
## (Intercept) 1 0.025885 5.084 3 574 0.00176 \*\*   
## READ 1 0.173281 40.104 3 574 < 2.2e-16 \*\*\*  
## WRITE 1 0.059076 12.013 3 574 1.229e-07 \*\*\*  
## SCIENCE 1 0.017404 3.389 3 574 0.01784 \*   
## PROG 2 0.111886 11.358 6 1150 2.320e-12 \*\*\*  
## READ:WRITE 1 0.002922 0.561 3 574 0.64112   
## READ:SCIENCE 1 0.003003 0.576 3 574 0.63076   
## READ:PROG 2 0.012924 1.247 6 1150 0.27970   
## WRITE:SCIENCE 1 0.003174 0.609 3 574 0.60923   
## WRITE:PROG 2 0.015347 1.482 6 1150 0.18079   
## SCIENCE:PROG 2 0.010378 1.000 6 1150 0.42393   
## READ:WRITE:SCIENCE 1 0.001600 0.307 3 574 0.82062   
## READ:WRITE:PROG 2 0.020323 1.968 6 1150 0.06740 .   
## READ:SCIENCE:PROG 2 0.006213 0.597 6 1150 0.73271   
## WRITE:SCIENCE:PROG 2 0.013614 1.314 6 1150 0.24791   
## READ:WRITE:SCIENCE:PROG 2 0.008035 0.773 6 1150 0.59108   
## Residuals 576   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Here I have printed the linear model fitted with all of the interaction terms, the summary of the linear model and the results of anova test. In referring to the anova resuls we see that the READ,WRITE & PROG terms have a significant effect at the 0.001 level. While SCIENCE is only significant at eh 0.05 level. We also see here that none of the interaction terms are found to be significant.