STAT 501 Homework 5

Multinomial

March 22, 2018

1. (a) Let $J_{\boldsymbol{\varphi}}(\boldsymbol{\omega};\boldsymbol{\lambda}) = \frac{\partial \boldsymbol{\varphi}(\boldsymbol{\omega};\boldsymbol{\lambda})}{\partial \boldsymbol{\omega}}$ be the Jacobian Matrix of $\boldsymbol{\varphi}$ with respect to $\boldsymbol{\omega}$ for a given $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_p)$, then since $\boldsymbol{\varphi} = \begin{bmatrix}\varphi(\omega_1;\lambda_1) & \varphi(\omega_2;\lambda_2) & \cdots & \varphi(\omega_p;\lambda_p)\end{bmatrix}^T$, we have

$$J_{\boldsymbol{\varphi}}(\boldsymbol{\omega}) = \operatorname{diag}\left(\{\varphi'(\omega_j;\lambda_j)\}_{j=1}^p\right),$$

where

$$\varphi'(\omega_j;\lambda_j) = \omega_j^{\lambda_j - 1}$$

Then we have

$$|\det J_{\boldsymbol{\varphi}}(\boldsymbol{\omega})| = \prod_{j=1}^p \omega_j^{\lambda_j - 1}$$

For sample $\boldsymbol{X}_i$, we have

$$\begin{aligned}\log f(\boldsymbol{X}_i) &= \log f_{\boldsymbol{U}_i^{(\lambda)}}(\boldsymbol{\varphi}(\boldsymbol{X}_i;\boldsymbol{\lambda})) + \log\left(|\det J_{\boldsymbol{\varphi}}(\boldsymbol{\omega})|\right) \\ &= -\frac{p}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}_{(\boldsymbol{\lambda})}| - \frac{1}{2}(\boldsymbol{\varphi}(\boldsymbol{X}_i;\boldsymbol{\lambda}) - \boldsymbol{\mu}_{(\boldsymbol{\lambda})})^T\boldsymbol{\Sigma}_{(\boldsymbol{\lambda})}^{-1}(\boldsymbol{\varphi}(\boldsymbol{X}_i;\boldsymbol{\lambda}) - \boldsymbol{\mu}_{(\boldsymbol{\lambda})}) + \sum_{j=1}^p(\lambda_j - 1)\log X_{ij}\end{aligned}$$

where X_{ij} is the j-th element of $\boldsymbol{X}_i$. Then the log likelihood given $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ and $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_m$ is

$$\begin{aligned}&\ell(\boldsymbol{\mu}_{(\boldsymbol{\lambda})}, \boldsymbol{\Sigma}_{(\boldsymbol{\lambda})}; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m, \boldsymbol{\lambda}) \\ =& -\frac{(n+m)p}{2}\log(2\pi) - \frac{n+m}{2}\log|\boldsymbol{\Sigma}_{(\boldsymbol{\lambda})}| \\ &- \frac{1}{2}\sum_{i=1}^n(\boldsymbol{\varphi}(\boldsymbol{X}_i;\boldsymbol{\lambda}) - \boldsymbol{\mu}_{(\boldsymbol{\lambda})})^T\boldsymbol{\Sigma}_{(\boldsymbol{\lambda})}^{-1}(\boldsymbol{\varphi}(\boldsymbol{X}_i;\boldsymbol{\lambda}) - \boldsymbol{\mu}_{(\boldsymbol{\lambda})}) + \sum_{i=1}^n\sum_{j=1}^p(\lambda_j - 1)\log X_{ij} \\ &- \frac{1}{2}\sum_{k=1}^m(\boldsymbol{\varphi}(\boldsymbol{Y}_k;\boldsymbol{\lambda}) - \boldsymbol{\nu}_{(\boldsymbol{\lambda})})^T\boldsymbol{\Sigma}_{(\boldsymbol{\lambda})}^{-1}(\boldsymbol{\varphi}(\boldsymbol{Y}_k;\boldsymbol{\lambda}) - \boldsymbol{\nu}_{(\boldsymbol{\lambda})}) + \sum_{k=1}^m\sum_{l=1}^p(\lambda_l - 1)\log Y_{kl}\end{aligned}$$

The MLEs are then

$$\begin{aligned}\hat{\boldsymbol{\mu}}_{(\boldsymbol{\lambda})} &= \frac{1}{n}\sum_{i=1}^n\boldsymbol{\varphi}(\boldsymbol{X}_i;\boldsymbol{\lambda}) \\ \hat{\boldsymbol{\nu}}_{(\boldsymbol{\lambda})} &= \frac{1}{m}\sum_{k=1}^m\boldsymbol{\varphi}(\boldsymbol{Y}_k;\boldsymbol{\lambda}) \\ \hat{\boldsymbol{\Sigma}}_{(\boldsymbol{\lambda})} &= \frac{1}{n+m}\left(\sum_{i=1}^n(\boldsymbol{\varphi}(\boldsymbol{X}_i;\boldsymbol{\lambda}) - \hat{\boldsymbol{\mu}}_{(\boldsymbol{\lambda})})(\boldsymbol{\varphi}(\boldsymbol{X}_i;\boldsymbol{\lambda}) - \hat{\boldsymbol{\mu}}_{(\boldsymbol{\lambda})})^T + \sum_{k=1}^m(\boldsymbol{\varphi}(\boldsymbol{Y}_k;\boldsymbol{\lambda}) - \hat{\boldsymbol{\nu}}_{(\boldsymbol{\lambda})})(\boldsymbol{\varphi}(\boldsymbol{Y}_k;\boldsymbol{\lambda}) - \hat{\boldsymbol{\nu}}_{(\boldsymbol{\lambda})})^T\right)\end{aligned}$$

Plugging in the MLE to the log likelihood and dropping the constants, we have the maximized log likelihood as a function of $\boldsymbol{\lambda}$ as below

$$\begin{aligned}\hat{\ell}(\boldsymbol{\lambda}) &= -\frac{n+m}{2}\log|\hat{\boldsymbol{\Sigma}}_{(\boldsymbol{\lambda})}| + \sum_{i=1}^{n}(\boldsymbol{\lambda}-1)^T\log\boldsymbol{X}_i + \sum_{k=1}^{m}(\boldsymbol{\lambda}-1)^T\log\boldsymbol{Y}_k \\ &= -\frac{n+m}{2}\log|\hat{\boldsymbol{\Sigma}}_{(\boldsymbol{\lambda})}| + (\boldsymbol{\lambda}-1)^T\left(\sum_{i=1}^{n}\log\boldsymbol{X}_i + \sum_{k=1}^{m}\log\boldsymbol{Y}_k\right)\end{aligned}$$

(b) i. We generate all the possible $\boldsymbol{\lambda}$'s and calculate the maximized MLE for each of them. The code is shown below and the result is

$$\boldsymbol{\lambda} = (4, 0, 0, 2, 4, 4)$$

```r
#Problem 1
colleges <- read.table("Colleges.txt", head = T, sep = "\t")

phi <- function(omega, lambda){
  if(lambda == 0)
    return(log(omega))
  else
    return((omega^lambda - 1)/lambda)
}

Phi <- function(Omega, Lambda){
  p <- length(Omega)
  res <- rep(0, p)
  for(i in 1:p)
    res[i] <- phi(Omega[i], Lambda[i])
  return(res)
}

X <- as.matrix(colleges[colleges$School_Type == "Lib Arts",
 ↪  -c(1,2)])
Y <- as.matrix(colleges[colleges$School_Type == "Univ",
 ↪  -c(1,2)])

Phi_X <- apply(X, MARGIN = 1, Phi, Lambda = lambdas.grid[400,])
mu <- apply(Phi_X, MARGIN = 1, mean)
Phi_Y <- apply(Y, MARGIN = 1, Phi, Lambda = lambdas.grid[400,])
nu <- apply(Phi_Y, MARGIN = 1, mean)
Sigma <- array(rep(0, length(mu)^2), dim = c(length(mu),
 ↪  length(mu)))
for (i in 1:ncol(Phi_X)){
  Sigma <- Sigma + (Phi_X[,i] - mu) %*% t(Phi_X[,i] - mu)
}
for (i in 1:ncol(Phi_Y)){
  Sigma <- Sigma + (Phi_Y[,i] - nu) %*% t(Phi_Y[,i] - nu)
}
Sigma <- Sigma/(ncol(Phi_X) + ncol(Phi_Y))

```

```r
logXsum <- apply(log(X), MARGIN = 2, sum)
logYsum <- apply(log(Y), MARGIN = 2, sum)

mloglik <- -(1/2) * (ncol(Phi_X) + ncol(Phi_Y)) *
 ↪ log(abs(det(Sigma))) + t(lambdas.grid[400,]-1)%*%(logXsum +
 ↪ logYsum)

maxloglik <- function(lambda, X, Y){
  Phi_X <- apply(X, MARGIN = 1, Phi, Lambda = lambda)
  mu <- apply(Phi_X, MARGIN = 1, mean)
  Phi_Y <- apply(Y, MARGIN = 1, Phi, Lambda = lambda)
  nu <- apply(Phi_Y, MARGIN = 1, mean)
  Sigma <- array(rep(0, length(mu)^2), dim = c(length(mu),
   ↪ length(mu)))
  for (i in 1:ncol(Phi_X)){
    Sigma <- Sigma + (Phi_X[,i] - mu) %*% t(Phi_X[,i] - mu)
  }
  for (i in 1:ncol(Phi_Y)){
    Sigma <- Sigma + (Phi_Y[,i] - nu) %*% t(Phi_Y[,i] - nu)
  }
  Sigma <- Sigma/(ncol(Phi_X) + ncol(Phi_Y))

  logXsum <- apply(log(X), MARGIN = 2, sum)
  logYsum <- apply(log(Y), MARGIN = 2, sum)

  mloglik <- -(1/2) * (ncol(Phi_X) + ncol(Phi_Y)) *
   ↪ log(abs(det(Sigma))) + t(lambda - 1)%*%(logXsum +
   ↪ logYsum)

  return(mloglik)
}

library(gtools)
lambdas <- c(0, 1/4, 1/3, 1/2, 1, 2, 3, 4)
lambdas.grid <- permutations(n = length(lambdas), r = 6, v =
 ↪ lambdas, repeats.allowed = T)

maxlogliks <- apply(X = lambdas.grid, MARGIN = 1, FUN =
 ↪ function(lambda) maxloglik(lambda, X = X, Y = Y))

max.lambda <- lambdas.grid[which.max(maxlogliks), ]
```

ii. We transform the data with Box-Cox transformation with the optimal $\boldsymbol{\lambda}$ and we tested if the tranformed data are multivarite normal.

```r
U <- apply(X, MARGIN = 1, Phi, Lambda = max.lambda)
V <- apply(Y, MARGIN = 1, Phi, Lambda = max.lambda)

colleges.tranformed <- cbind(colleges$School_Type,
 ↪ as.data.frame(rbind(t(U), t(V))))
names(colleges.tranformed) <- names(colleges)[-1]

```

```r
source("testnormality.R")
testnormality(X =
 ↪  colleges.tranformed[colleges.tranformed$School_Type == "Lib
 ↪  Arts", -1])
testnormality(X =
 ↪  colleges.tranformed[colleges.tranformed$School_Type ==
 ↪  "Univ", -1])
```

The test results (p-values) are **0.7551009** and **0.8752648**, so we can say that the transformed data are pretty multivariate normal. Then we did t-tests to compare the means of each of the components between the liberal arts and public universities.

```r
tp_value<-function(X, cl){
  class <- levels(cl)
  return(t.test(X[cl == class[1]], X[cl == class[2]], var.equal
   ↪  = T)$p.value)
}

p_vals <- sapply(colleges.tranformed[,-1], tp_value, cl =
 ↪  as.factor(colleges.tranformed$School_Type))
```

The p-values are

```
         SAT   Acceptance   X..Student      Top.10.        X.PhD        Grad.
2.799562e-01 9.546224e-02 3.220655e-07 2.853453e-05 9.288646e-03 7.517852e-01
```

From the results, SAT, acceptance rate graduation rate have big p-values and we can conclude that they are equal between liberal arts and universities.

iii. We adjusted the p-values with FDR at 0.05.

```r
p.adjust(p_vals[order(p_vals)], method = "fdr")
```

And the adjusted p-values are

```
  X..Student      Top.10.        X.PhD   Acceptance          SAT        Grad.
1.932393e-06 8.560360e-05 1.857729e-02 1.431934e-01 3.359474e-01 7.517852e-01
```

From the result, percent faculties with PhDs has a adjusted p-value less then 0.05, so cost per student, percent of students in top 10 percent of HS graduating class and percent of faculties with PhDs are significant different between liberal arts and universities.

2. (a) we fit the model with no interactions and summarize the fitted model and the MANOVA for the model.

```r
library(sas7bdat)
psych <- read.sas7bdat("psych.sas7bdat")

#a) Fit the linear model
library(car)
psych$PROG <-as.factor(psych$PROG)
levels(psych$PROG)
fit.lm <- lm(cbind(LOCUS_OF_CONTROL,SELF_CONCEPT,MOTIVATION) ~ READ
 ↪  + WRITE + SCIENCE + PROG, data = psych)
summary(fit.lm)
```

```r
fit.manova <- Manova(fit.lm)
summary(fit.manova)
```

Summary of the fitted model:

```
Response LOCUS_OF_CONTROL :

Call:
lm(formula = LOCUS_OF_CONTROL ~ READ + WRITE + SCIENCE + PROG,
    data = psych)

Residuals:
    Min      1Q  Median      3Q     Max
-1.9560 -0.3889 -0.0219  0.3725  1.9039

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.624765   0.157005 -10.348  < 2e-16 ***
READ         0.012505   0.003718   3.363 0.000819 ***
WRITE        0.012145   0.003391   3.581 0.000370 ***
SCIENCE      0.005761   0.003641   1.582 0.114109
PROG2        0.127795   0.063955   1.998 0.046150 *
PROG3        0.251671   0.068470   3.676 0.000259 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.607 on 594 degrees of freedom
Multiple R-squared:  0.1868,        Adjusted R-squared:  0.1799
F-statistic: 27.28 on 5 and 594 DF,  p-value: < 2.2e-16


Response SELF_CONCEPT :

Call:
lm(formula = SELF_CONCEPT ~ READ + WRITE + SCIENCE + PROG, data = psych)

Residuals:
     Min       1Q   Median       3Q      Max
-2.38183 -0.46594  0.00604  0.46063  2.28836

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.372341   0.178234  -2.089 0.037127 *
READ         0.001308   0.004220   0.310 0.756801
WRITE       -0.004293   0.003850  -1.115 0.265214
SCIENCE      0.005306   0.004133   1.284 0.199765
PROG2        0.276483   0.072602   3.808 0.000155 ***
PROG3        0.423359   0.077728   5.447 7.52e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.6891 on 594 degrees of freedom
Multiple R-squared:  0.05404,        Adjusted R-squared:  0.04607
F-statistic: 6.786 on 5 and 594 DF,  p-value: 3.629e-06


Response MOTIVATION :

Call:
lm(formula = MOTIVATION ~ READ + WRITE + SCIENCE + PROG, data = psych)

Residuals:
     Min       1Q   Median       3Q      Max
-2.31821 -0.50736 -0.03076  0.51596  2.33499
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.310842   0.196944  -6.656 6.41e-11 ***
READ         0.009674   0.004664   2.074   0.0385 *
WRITE        0.017535   0.004254   4.122 4.29e-05 ***
SCIENCE     -0.009001   0.004567  -1.971   0.0492 *
PROG2        0.360329   0.080224   4.492 8.50e-06 ***
PROG3        0.619696   0.085887   7.215 1.65e-12 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.7614 on 594 degrees of freedom
Multiple R-squared:   0.15,        Adjusted R-squared:  0.1428
F-statistic: 20.96 on 5 and 594 DF,  p-value: < 2.2e-16
```

Summary of 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
```

(b) we refit the model after dropping writing and science.

```
#b) Refit the model without test scores of writing and science.
fit.lm1 <- lm(cbind(LOCUS_OF_CONTROL,SELF_CONCEPT,MOTIVATION)~ READ
 ↪  + PROG, data = psych)
summary(fit.lm1)
fit.manova1 <- Manova(fit.lm1)
summary(fit.manova1)
```

Summary of the refitted model:

```
Response LOCUS_OF_CONTROL :

Call:
lm(formula = LOCUS_OF_CONTROL ~ READ + PROG, data = psych)

Residuals:
     Min       1Q   Median       3Q      Max
-2.16592 -0.40135 -0.03684  0.38493  1.90217

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.268584   0.134844  -9.408  < 2e-16 ***
READ         0.023665   0.002513   9.418  < 2e-16 ***
PROG2        0.117619   0.064820   1.815 0.070098 .
PROG3        0.263111   0.069339   3.795 0.000163 ***
```

```
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.6157 on 596 degrees of freedom
Multiple R-squared:  0.1606,        Adjusted R-squared:  0.1563
F-statistic:    38 on 3 and 596 DF,  p-value: < 2.2e-16


Response SELF_CONCEPT :

Call:
lm(formula = SELF_CONCEPT ~ READ + PROG, data = psych)

Residuals:
     Min       1Q   Median       3Q      Max
-2.39289 -0.48196  0.00952  0.46987  2.35663

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.369454   0.150964  -2.447 0.014681 *
READ         0.002250   0.002813   0.800 0.424230
PROG2        0.275973   0.072570   3.803 0.000158 ***
PROG3        0.417687   0.077628   5.381 1.07e-07 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.6893 on 596 degrees of freedom
Multiple R-squared:  0.05031,       Adjusted R-squared:  0.04553
F-statistic: 10.52 on 3 and 596 DF,  p-value: 9.411e-07


Response MOTIVATION :

Call:
lm(formula = MOTIVATION ~ READ + PROG, data = psych)

Residuals:
     Min       1Q   Median       3Q      Max
-2.29569 -0.50487 -0.01792  0.52213  2.37701

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.100395   0.168990  -6.512 1.58e-10 ***
READ         0.014259   0.003149   4.528 7.20e-06 ***
PROG2        0.355326   0.081235   4.374 1.44e-05 ***
PROG3        0.640053   0.086898   7.366 5.90e-13 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.7716 on 596 degrees of freedom
Multiple R-squared:  0.1242,        Adjusted R-squared:  0.1198
F-statistic: 28.18 on 3 and 596 DF,  p-value: < 2.2e-16
```

Summary of 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
```

(c) We compare the model in (a) and (b). The reduced model in (b) does not have writing and science in the dependent variables.

```
anova(fit.lm, fit.lm1, 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
```

From the result we can see writing and science are significant and thus they are related to the psychological profiles.

(d) We use likelihood ratio test to test any differences Program 1 and 2 and between Program 2 and 3 with model in (a) as a full model. Then we adjusted p-value with Bonferoni and FDR.

```r
#c)
anova(fit.lm, fit.lm1, test = "Wilks")
#d)Test simultaneously
p_values <- c(0,0)

fit.lm.P12 <- lm(cbind(LOCUS_OF_CONTROL,SELF_CONCEPT,MOTIVATION) ~ READ + WRITE + SCIENCE + PROG, data = psych[psych$PROG %in% c(1,2),])
fit.lm.P12_reduced <- lm(cbind(LOCUS_OF_CONTROL,SELF_CONCEPT,MOTIVATION) ~ READ + WRITE + SCIENCE, data = psych[psych$PROG %in% c(1,2),])
test_P12 <- anova(fit.lm.P12, fit.lm.P12_reduced, test = "Wilks")
p_vals[1] <- test_P12$`Pr(>F)`[2]

fit.lm.P23 <- lm(cbind(LOCUS_OF_CONTROL,SELF_CONCEPT,MOTIVATION) ~ READ + WRITE + SCIENCE + PROG, data = psych[psych$PROG %in% c(2,3),])
fit.lm.P23_reduced <- lm(cbind(LOCUS_OF_CONTROL,SELF_CONCEPT,MOTIVATION) ~ READ + WRITE + SCIENCE, data = psych[psych$PROG %in% c(2,3),])
test_P23 <- anova(fit.lm.P23, fit.lm.P23_reduced, test = "Wilks")
p_values[2] <- test_P23$`Pr(>F)`[2]

bon_adj<-p.adjust(p_values,method="bonferroni")
bon_adj
fdr_adj<-p.adjust(p_values,method="fdr")
fdr_adj
```

The adjusted pvalues are

```r
# Bonferroni
0.000000000 0.001529967

# FDR
0.0000000000 0.0007649833
```

Both adjusted p-values are small, so the profiles between Program 1 and 2 and between Program 2 and 3 are different.

(e) To compare this two parameters, with $\tau_1 = 0$, the multiplying matrices are

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}, M = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

```r
#e)
C <- matrix(c(0,0,1,0,0,0), nrow = 1)
M <- matrix(c(1,-1,0), ncol = 1)
newfit <- linearHypothesis(fit.lm, hypothesis.matrix = C, P = M)
print(newfit)
```

Thus the result is

```
 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
```

From the result we can see with small p-value, the coefficient for the written test scores with locus of control as the outcome is different from the corresponding coefficient with self concept as the outcome.

(f) The multiplying matrices are

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, M = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

```
#f)
C1 <- matrix(c(0,0,1,0,0,0,0,0,0,1,0,0), nrow = 2, byrow = T)
M <- matrix(c(1,-1,0), ncol = 1)
newfit1 <- linearHypothesis(fit.lm, hypothesis.matrix = C1, P = M)
print(newfit1)
```

The result is

```
 Response transformation matrix:
                  [,1]
LOCUS_OF_CONTROL    1
SELF_CONCEPT       -1
MOTIVATION          0

Sum of squares and products for the hypothesis:
          [,1]
[1,] 9.302364

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            2 0.0210508 6.386518      2    594 0.0018021 **
Wilks             2 0.9789492 6.386518      2    594 0.0018021 **
Hotelling-Lawley  2 0.0215034 6.386518      2    594 0.0018021 **
Roy               2 0.0215034 6.386518      2    594 0.0018021 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

From the result we can see with small p-value, the coefficient for science scores for locus of control is different to the corresponding coefficient for science for the self concept variable, or that the coefficient for the written scores for locus of control is different to the coefficient for the written scores for self concept.

(g) We fit the model with all interactions of dependent variables in model (a).

```
#g)
fit.lm2<- lm(cbind(LOCUS_OF_CONTROL,SELF_CONCEPT,MOTIVATION)~ .^2,
 ↪  data = psych)
summary(fit.lm2)
fit.manova12 <- Manova(fit.lm2)
summary(fit.manova12)
```

Summary of fitted model:

```
Response LOCUS_OF_CONTROL :

Call:
lm(formula = LOCUS_OF_CONTROL ~ (READ + WRITE + SCIENCE + PROG)^2,
    data = psych)

Residuals:
     Min       1Q   Median       3Q      Max
-2.05066 -0.39587 -0.01968  0.37354  1.91791

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1.309e+00  6.936e-01  -1.888   0.0596 .
READ           2.440e-02  2.343e-02   1.041   0.2982
WRITE          1.332e-03  2.035e-02   0.065   0.9478
SCIENCE       -6.030e-03  2.194e-02  -0.275   0.7836
PROG2         -2.042e-01  3.851e-01  -0.530   0.5961
PROG3          5.161e-01  4.441e-01   1.162   0.2457
READ:WRITE    -8.547e-05  3.966e-04  -0.216   0.8294
READ:SCIENCE  -2.226e-04  3.480e-04  -0.640   0.5227
READ:PROG2     3.740e-03  1.002e-02   0.373   0.7093
READ:PROG3     5.796e-03  1.053e-02   0.550   0.5823
WRITE:SCIENCE  3.926e-04  4.206e-04   0.934   0.3509
WRITE:PROG2   -4.924e-03  9.184e-03  -0.536   0.5921
WRITE:PROG3   -7.265e-03  1.004e-02  -0.724   0.4694
SCIENCE:PROG2  7.824e-03  9.766e-03   0.801   0.4234
SCIENCE:PROG3 -3.242e-03  1.062e-02  -0.305   0.7602
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.6095 on 585 degrees of freedom
```

```
Multiple R-squared:  0.1926,        Adjusted R-squared:  0.1733
F-statistic: 9.967 on 14 and 585 DF,  p-value: < 2.2e-16


Response SELF_CONCEPT :

Call:
lm(formula = SELF_CONCEPT ~ (READ + WRITE + SCIENCE + PROG)^2,
    data = psych)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2764 -0.4467 -0.0088  0.4775  2.2170

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.3647445  0.7811834   0.467   0.6407
READ          -0.0246541  0.0263898  -0.934   0.3506
WRITE         -0.0065516  0.0229173  -0.286   0.7751
SCIENCE        0.0115021  0.0247142   0.465   0.6418
PROG2         -0.4325538  0.4336662  -0.997   0.3190
PROG3          0.4365658  0.5001830   0.873   0.3831
READ:WRITE     0.0002084  0.0004466   0.467   0.6410
READ:SCIENCE   0.0005143  0.0003920   1.312   0.1900
READ:PROG2    -0.0095541  0.0112896  -0.846   0.3977
READ:PROG3    -0.0206186  0.0118610  -1.738   0.0827 .
WRITE:SCIENCE -0.0005729  0.0004736  -1.210   0.2270
WRITE:PROG2    0.0260726  0.0103431   2.521   0.0120 *
WRITE:PROG3    0.0251531  0.0113023   2.225   0.0264 *
SCIENCE:PROG2 -0.0032262  0.0109988  -0.293   0.7694
SCIENCE:PROG3 -0.0050141  0.0119547  -0.419   0.6751
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.6864 on 585 degrees of freedom
Multiple R-squared:  0.07568,        Adjusted R-squared:  0.05355
F-statistic: 3.421 on 14 and 585 DF,  p-value: 2.32e-05


Response MOTIVATION :

Call:
lm(formula = MOTIVATION ~ (READ + WRITE + SCIENCE + PROG)^2,
    data = psych)

Residuals:
     Min       1Q   Median       3Q      Max
-2.39626 -0.53097 -0.01044  0.50289  2.25427

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1.905e+00  8.667e-01  -2.198   0.0284 *
READ           2.528e-02  2.928e-02   0.863   0.3883
WRITE          1.199e-02  2.543e-02   0.472   0.6374
SCIENCE        1.048e-02  2.742e-02   0.382   0.7025
PROG2         -1.959e-01  4.812e-01  -0.407   0.6840
PROG3          4.859e-01  5.550e-01   0.876   0.3816
READ:WRITE     1.567e-04  4.955e-04   0.316   0.7520
READ:SCIENCE  -5.244e-04  4.349e-04  -1.206   0.2284
READ:PROG2     1.403e-02  1.253e-02   1.120   0.2630
READ:PROG3    -8.128e-03  1.316e-02  -0.618   0.5370
WRITE:SCIENCE  2.248e-05  5.255e-04   0.043   0.9659
WRITE:PROG2   -3.581e-03  1.148e-02  -0.312   0.7551
WRITE:PROG3   -6.757e-03  1.254e-02  -0.539   0.5902
SCIENCE:PROG2  5.198e-04  1.220e-02   0.043   0.9660
SCIENCE:PROG3  1.789e-02  1.326e-02   1.349   0.1779
```

```
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.7615 on 585 degrees of freedom
Multiple R-squared:  0.1626,        Adjusted R-squared:  0.1425
F-statistic: 8.113 on 14 and 585 DF,  p-value: 5.545e-16
```

Summary of MANOVA:

```
Type II MANOVA Tests:

Sum of squares and products for error:
                  LOCUS_OF_CONTROL SELF_CONCEPT MOTIVATION
LOCUS_OF_CONTROL        217.28769     34.06881   35.58755
SELF_CONCEPT             34.06881    275.58839   77.13713
MOTIVATION               35.58755     77.13713  339.25955

------------------------------------------

Term: READ

Sum of squares and products for the hypothesis:
                  LOCUS_OF_CONTROL SELF_CONCEPT MOTIVATION
LOCUS_OF_CONTROL        3.9376254   0.36706969  3.2137190
SELF_CONCEPT            0.3670697   0.03421863  0.2995864
MOTIVATION              3.2137190   0.29958635  2.6228980

Multivariate Tests: READ
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            1 0.0232352 4.622794      3    583 0.0033157 **
Wilks             1 0.9767648 4.622794      3    583 0.0033157 **
Hotelling-Lawley  1 0.0237880 4.622794      3    583 0.0033157 **
Roy               1 0.0237880 4.622794      3    583 0.0033157 **
---
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         4.716743   -1.6622303   6.745989
SELF_CONCEPT            -1.662230    0.5857876  -2.377358
MOTIVATION               6.745989   -2.3773581   9.648261

Multivariate Tests: WRITE
                 Df test stat approx F num Df den Df     Pr(>F)
Pillai            1 0.0527459 10.82106      3    583 6.2976e-07 ***
Wilks             1 0.9472541 10.82106      3    583 6.2976e-07 ***
Hotelling-Lawley  1 0.0556830 10.82106      3    583 6.2976e-07 ***
Roy               1 0.0556830 10.82106      3    583 6.2976e-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        0.8271414    0.8191007  -1.455988
SELF_CONCEPT            0.8191007    0.8111381  -1.441834
MOTIVATION             -1.4559877   -1.4418339   2.562923

```

```
Multivariate Tests: SCIENCE
                 Df test stat approx F num Df den Df  Pr(>F)
Pillai            1 0.0177120 3.504106      3    583 0.01526 *
Wilks             1 0.9822880 3.504106      3    583 0.01526 *
Hotelling-Lawley  1 0.0180314 3.504106      3    583 0.01526 *
Roy               1 0.0180314 3.504106      3    583 0.01526 *
---
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         5.080791     8.304686   12.23561
SELF_CONCEPT             8.304686    14.009223   20.37723
MOTIVATION              12.235609    20.377226   29.79404

Multivariate Tests: PROG
                 Df test stat approx F num Df den Df     Pr(>F)
Pillai            2 0.1088848 11.20832      6   1168 3.4156e-12 ***
Wilks             2 0.8911807 11.52309      6   1166 1.4745e-12 ***
Hotelling-Lawley  2 0.1220335 11.83724      6   1164 6.3763e-13 ***
Roy               2 0.1214286 23.63811      3    584 1.8864e-14 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

----------------------------------------

Term: READ:WRITE

Sum of squares and products for the hypothesis:
                 LOCUS_OF_CONTROL SELF_CONCEPT  MOTIVATION
LOCUS_OF_CONTROL       0.01725478  -0.04206559 -0.03162317
SELF_CONCEPT          -0.04206559   0.10255213  0.07709445
MOTIVATION            -0.03162317   0.07709445  0.05795642

Multivariate Tests: READ:WRITE
                 Df test stat  approx F num Df den Df  Pr(>F)
Pillai            1 0.0006030 0.1172503      3    583 0.94998
Wilks             1 0.9993970 0.1172503      3    583 0.94998
Hotelling-Lawley  1 0.0006033 0.1172503      3    583 0.94998
Roy               1 0.0006033 0.1172503      3    583 0.94998

----------------------------------------

Term: READ:SCIENCE

Sum of squares and products for the hypothesis:
                 LOCUS_OF_CONTROL SELF_CONCEPT MOTIVATION
LOCUS_OF_CONTROL        0.1519441   -0.3510590  0.3579416
SELF_CONCEPT           -0.3510590    0.8111035 -0.8270054
MOTIVATION              0.3579416   -0.8270054  0.8432190

Multivariate Tests: READ:SCIENCE
                 Df test stat approx F num Df den Df  Pr(>F)
Pillai            1 0.0079691 1.561109      3    583 0.19773
Wilks             1 0.9920309 1.561109      3    583 0.19773
Hotelling-Lawley  1 0.0080332 1.561109      3    583 0.19773
Roy               1 0.0080332 1.561109      3    583 0.19773

----------------------------------------

Term: READ:PROG

```

```
Sum of squares and products for the hypothesis:
                 LOCUS_OF_CONTROL SELF_CONCEPT MOTIVATION
LOCUS_OF_CONTROL        0.1125807   -0.3979259 -0.1448465
SELF_CONCEPT           -0.3979259    1.4982214  0.9837472
MOTIVATION             -0.1448465    0.9837472  2.6130464

Multivariate Tests: READ:PROG
                 Df test stat approx F num Df den Df  Pr(>F)
Pillai            2 0.0134558 1.318570      6   1168 0.24565
Wilks             2 0.9865848 1.316776      6   1166 0.24646
Hotelling-Lawley  2 0.0135565 1.314981      6   1164 0.24727
Roy               2 0.0089705 1.746253      3    584 0.15639

----------------------------------------

Term: WRITE:SCIENCE

Sum of squares and products for the hypothesis:
                 LOCUS_OF_CONTROL SELF_CONCEPT   MOTIVATION
LOCUS_OF_CONTROL       0.32377319  -0.47237106  0.018537551
SELF_CONCEPT          -0.47237106   0.68916891 -0.027045483
MOTIVATION             0.01853755  -0.02704548  0.001061363

Multivariate Tests: WRITE:SCIENCE
                 Df test stat  approx F num Df den Df Pr(>F)
Pillai            1 0.0047019 0.9180441      3    583 0.4318
Wilks             1 0.9952981 0.9180441      3    583 0.4318
Hotelling-Lawley  1 0.0047241 0.9180441      3    583 0.4318
Roy               1 0.0047241 0.9180441      3    583 0.4318

----------------------------------------

Term: WRITE:PROG

Sum of squares and products for the hypothesis:
                 LOCUS_OF_CONTROL SELF_CONCEPT MOTIVATION
LOCUS_OF_CONTROL        0.1965790    -0.715953  0.1789287
SELF_CONCEPT           -0.7159530     3.246022 -0.5810130
MOTIVATION              0.1789287    -0.581013  0.1706823

Multivariate Tests: WRITE:PROG
                 Df test stat approx F num Df den Df   Pr(>F)
Pillai            2 0.0154968 1.520133      6   1168 0.167982
Wilks             2 0.9845073 1.523092      6   1166 0.167022
Hotelling-Lawley  2 0.0157323 1.526031      6   1164 0.166072
Roy               2 0.0154612 3.009774      3    584 0.029721 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

----------------------------------------

Term: SCIENCE:PROG

Sum of squares and products for the hypothesis:
                 LOCUS_OF_CONTROL SELF_CONCEPT MOTIVATION
LOCUS_OF_CONTROL       0.68533271   0.03975242 -0.8735216
SELF_CONCEPT           0.03975242   0.08316743 -0.2812081
MOTIVATION            -0.87352158  -0.28120805  1.7706648

Multivariate Tests: SCIENCE:PROG
                 Df test stat approx F num Df den Df  Pr(>F)
Pillai            2 0.0102362 1.001447      6   1168 0.42279
Wilks             2 0.9897745 1.001267      6   1166 0.42292
Hotelling-Lawley  2 0.0103204 1.001082      6   1164 0.42304
Roy               2 0.0091454 1.780306      3    584 0.14973
```

The results indicate that all p-values are larger than 0.05. Therefore, we can conclude that all possible interactions of components are non significant and we should not include interactions in the model.