563 HW4

Amber Wang, Daniel Chen, Xilong Liang

3/16/2022

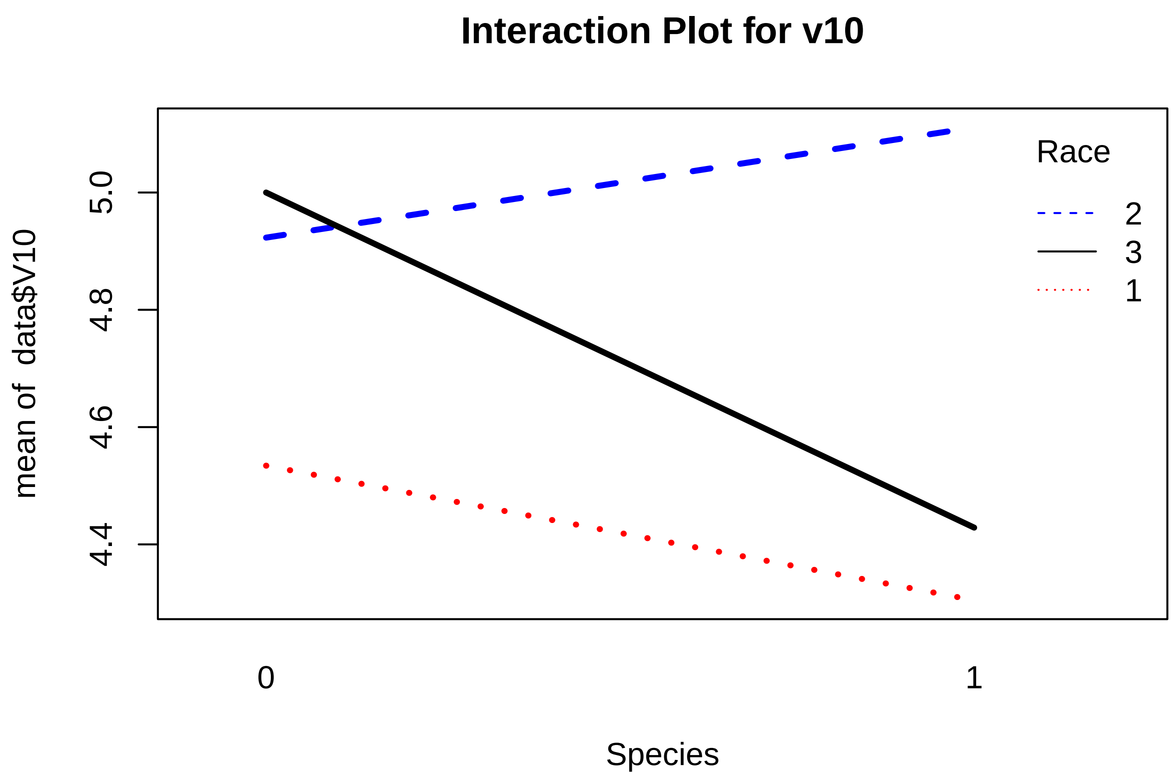
```
library(MASS)
library(biotools)
```

```
## ---
## biotools version 4.2
```

```
library(DiscriMiner)
library(klaR)
library(car)
```

```
## Loading required package: carData
```

```
library("readxl")
setwd("~/Downloads")
data <- read_excel("ohiocrimehm.xls")
data = na.omit(data)
#First, we fit a two-way MANOVA model. For two categorical factors, make interaction plots for each of

interaction.plot(data$V70, data$V71, data$V10,
  lwd = 3, col = c("red", "blue", "black"), trace.label = "Race",
  xlab = "Species", main = "Interaction Plot for v10")
```




```r
interaction.plot(data$V70, data$V71, data$V12,
  lwd = 3, col = c("red", "blue", "black"), trace.label = "Race",
  xlab = "Species", main = "Interaction Plot for v12")
```

```r
#V70 and V71 are both having interaction, when we set V10 and V12 as response variable. The interaction
```

```r
##2
options(contrasts = c("contr.sum", "contr.poly"))

model <- lm(cbind(V10, V12) ~ V70*V71,
                data = data)

#Multivariate and univariate results
summary(Anova(model, type = 3), univariate = T)
```

```
##
## Type III MANOVA Tests:
##
## Sum of squares and products for error:
##          V10       V12
## V10 815.5805  465.0591
## V12 465.0591 1349.2870
##
## ------------------------------------------
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##          V10      V12
## V10 211.1186 202.5865
## V12 202.5865 194.3992
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.2238320 71.37428      2    495 < 2.22e-16 ***
## Wilks             1 0.7761680 71.37428      2    495 < 2.22e-16 ***
## Hotelling-Lawley  1 0.2883809 71.37428      2    495 < 2.22e-16 ***
## Roy               1 0.2883809 71.37428      2    495 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: V70
##
## Sum of squares and products for the hypothesis:
##            V10         V12
## V10 0.39455168 0.051264474
## V12 0.05126447 0.006660842
##
## Multivariate Tests: V70
##                  Df test stat  approx F num Df den Df  Pr(>F)
## Pillai            1 0.0005540 0.1371941      2    495 0.87183
## Wilks             1 0.9994460 0.1371941      2    495 0.87183
## Hotelling-Lawley  1 0.0005543 0.1371941      2    495 0.87183
## Roy               1 0.0005543 0.1371941      2    495 0.87183
##
## ------------------------------------------
```

```
##
## Term: V71
##
## Sum of squares and products for the hypothesis:
##          V10       V12
## V10 1.917898 1.3260897
## V12 1.326090 0.9168963
##
## Multivariate Tests: V71
##                  Df test stat  approx F num Df den Df  Pr(>F)
## Pillai            1 0.0023719 0.5884476      2    495 0.55558
## Wilks             1 0.9976281 0.5884476      2    495 0.55558
## Hotelling-Lawley  1 0.0023776 0.5884476      2    495 0.55558
## Roy               1 0.0023776 0.5884476      2    495 0.55558
##
## ------------------------------------------
##
## Term: V70:V71
##
## Sum of squares and products for the hypothesis:
##              V10        V12
## V10 0.0004718059 0.02218056
## V12 0.0221805584 1.04275336
##
## Multivariate Tests: V70:V71
##                  Df test stat  approx F num Df den Df  Pr(>F)
## Pillai            1 0.0009384 0.2324634      2    495 0.79267
## Wilks             1 0.9990616 0.2324634      2    495 0.79267
## Hotelling-Lawley  1 0.0009392 0.2324634      2    495 0.79267
## Roy               1 0.0009392 0.2324634      2    495 0.79267
##
##  Type III Sums of Squares
##              df        V10        V12
## (Intercept)   1 2.1112e+02 1.9440e+02
## V70           1 3.9455e-01 6.6608e-03
## V71           1 1.9179e+00 9.1690e-01
## V70:V71       1 4.7181e-04 1.0428e+00
## residuals   496 8.1558e+02 1.3493e+03
##
##  F-tests
##                V10   V12
## (Intercept) 128.39 71.46
## V70           0.24  0.00
## V71           1.17  0.34
## V70:V71       0.00  0.38
##
##  p-values
##             V10        V12
## (Intercept) < 2.22e-16 3.1717e-16
## V70         0.62446    0.96055
## V71         0.28067    0.56180
## V70:V71     0.98649    0.53612
```

```r
#multivariate results: based on the multivariate results, under a significance level of alpha = 0.05, t

anova(lm(V10 ~ V70*V71, data = data))
```

```
## Analysis of Variance Table
##
## Response: V10
##            Df Sum Sq Mean Sq F value  Pr(>F)
## V70         1   5.21  5.2062  3.1662 0.07579 .
## V71         1   7.01  7.0128  4.2649 0.03943 *
## V70:V71     1   0.00  0.0005  0.0003 0.98649
## Residuals 496 815.58  1.6443
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lm(V12 ~ V70*V71, data = data))
```

```
## Analysis of Variance Table
##
## Response: V12
##            Df  Sum Sq Mean Sq F value  Pr(>F)
## V70         1   13.92 13.9225  5.1179 0.02411 *
## V71         1    0.03  0.0258  0.0095 0.92253
## V70:V71     1    1.04  1.0428  0.3833 0.53612
## Residuals 496 1349.29  2.7203
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#univaraite results: when "V10" is the response variable, only "V71" is significant under a significanc
```

```r
#3
#3
#RUN LINE BELOW TO MAKE SURE CONTRASTS ARE SET CORRECTLY -  this returns things to their default level
options(contrasts = c("contr.treatment", "contr.poly"))

#Make sure TRTCOMB is a factor
data$V72 <- as.factor(data$V72)

#Fit two way MANOVA model
dataaMAOV2 <- lm(cbind(V10, V12) ~ V72,
                data = data)
#Fit one way ANOVA model just for ASQRT
dataaASQRT <- lm(V10 ~ V72,
                data = data)

contrasts(data$V72)
```

```
##   2 3 4 5 6 7 8
## 1 0 0 0 0 0 0 0
## 2 1 0 0 0 0 0 0
## 3 0 1 0 0 0 0 0
## 4 0 0 1 0 0 0 0
## 5 0 0 0 1 0 0 0
## 6 0 0 0 0 1 0 0
## 7 0 0 0 0 0 1 0
## 8 0 0 0 0 0 0 1
```

```r
levels(data$V72)
```

```
## [1] "1" "2" "3" "4" "5" "6" "7" "8"
```

```r
#Get multivariate contrast for 1 vs 2 - notice that 1 is the reference level
linearHypothesis(dataaMAOV2, "V722+V723+V724+V725- V726 - V727- V728= 0")
```

```
##
## Sum of squares and products for the hypothesis:
##             V10        V12
## V10 0.004578635 0.07473129
## V12 0.074731289 1.21974469
##
## Sum of squares and products for error:
##          V10      V12
## V10 801.7941  463.602
## V12 463.6020 1341.792
##
## Multivariate Tests:
##                  Df test stat  approx F num Df den Df  Pr(>F)
## Pillai            1 0.0010615 0.2608767      2    491 0.77048
## Wilks             1 0.9989385 0.2608767      2    491 0.77048
## Hotelling-Lawley  1 0.0010626 0.2608767      2    491 0.77048
## Roy               1 0.0010626 0.2608767      2    491 0.77048
```

```r
linearHypothesis(dataaASQRT, "V722+V723+V724+V725- V726 - V727- V728= 0")
```

```
## Linear hypothesis test
##
## Hypothesis:
## V722  + V723  + V724  + V725 - V726 - V727 - V728 = 0
##
## Model 1: restricted model
## Model 2: V10 ~ V72
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    493 801.80
## 2    492 801.79  1 0.0045786 0.0028 0.9577
```

since for p value is 0.9577 so the model1 is better since model2 is not significant.

```r
#4'

options(contrasts = c("contr.sum", "contr.poly"))

### Default is: options(contrasts = c("contr.treatment", "contr.poly"))

#Fit the model
dataaMod2 <- lm(cbind(V10, V12) ~ V70*V71+V64,
                data = data)

#Multivariate results and univariate results with with type 3 Sum of squares
summary(Anova(dataaMod2, type = 3), univariate = T)
```

```
##
## Type III MANOVA Tests:
##
```

```
## Sum of squares and products for error:
##          V10       V12
## V10 803.9033  464.5396
## V12 464.5396 1349.2639
##
## ------------------------------------------
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##          V10      V12
## V10 139.3511 149.0269
## V12 149.0269 159.3745
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.1697794 50.51128      2    494 < 2.22e-16 ***
## Wilks             1 0.8302206 50.51128      2    494 < 2.22e-16 ***
## Hotelling-Lawley  1 0.2044991 50.51128      2    494 < 2.22e-16 ***
## Roy               1 0.2044991 50.51128      2    494 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: V70
##
## Sum of squares and products for the hypothesis:
##           V10        V12
## V10 0.5702664 0.06587120
## V12 0.0658712 0.00760875
##
## Multivariate Tests: V70
##                  Df test stat approx F num Df den Df  Pr(>F)
## Pillai            1 0.0008215 0.203073      2    494 0.81629
## Wilks             1 0.9991785 0.203073      2    494 0.81629
## Hotelling-Lawley  1 0.0008222 0.203073      2    494 0.81629
## Roy               1 0.0008222 0.203073      2    494 0.81629
##
## ------------------------------------------
##
## Term: V71
##
## Sum of squares and products for the hypothesis:
##          V10       V12
## V10 1.364720 1.1053034
## V12 1.105303 0.8951987
##
## Multivariate Tests: V71
##                  Df test stat  approx F num Df den Df Pr(>F)
## Pillai            1 0.0017625 0.4361062      2    494 0.6468
## Wilks             1 0.9982375 0.4361062      2    494 0.6468
## Hotelling-Lawley  1 0.0017656 0.4361062      2    494 0.6468
## Roy               1 0.0017656 0.4361062      2    494 0.6468
```

```
##
## ------------------------------------------
##
## Term: V64
##
## Sum of squares and products for the hypothesis:
##            V10        V12
## V10 11.6771558 0.51952728
## V12  0.5195273 0.02311424
##
## Multivariate Tests: V64
##                  Df test stat approx F num Df den Df   Pr(>F)
## Pillai            1 0.0172947 4.346962      2    494 0.013445 *
## Wilks             1 0.9827053 4.346962      2    494 0.013445 *
## Hotelling-Lawley  1 0.0175990 4.346962      2    494 0.013445 *
## Roy               1 0.0175990 4.346962      2    494 0.013445 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: V70:V71
##
## Sum of squares and products for the hypothesis:
##             V10        V12
## V10  0.05140557 -0.2284059
## V12 -0.22840587  1.0148558
##
## Multivariate Tests: V70:V71
##                  Df test stat  approx F num Df den Df  Pr(>F)
## Pillai            1 0.0012614 0.3119656      2    494 0.73215
## Wilks             1 0.9987386 0.3119656      2    494 0.73215
## Hotelling-Lawley  1 0.0012630 0.3119656      2    494 0.73215
## Roy               1 0.0012630 0.3119656      2    494 0.73215
##
##  Type III Sums of Squares
##              df        V10        V12
## (Intercept)   1 139.351107 1.5937e+02
## V70           1   0.570266 7.6088e-03
## V71           1   1.364720 8.9520e-01
## V64           1  11.677156 2.3114e-02
## V70:V71       1   0.051406 1.0149e+00
## residuals   495 803.903348 1.3493e+03
##
##  F-tests
##               V10   V12
## (Intercept) 85.80 58.47
## V70          0.35  0.00
## V71          0.84  0.33
## V64          7.19  0.01
## V70:V71      0.03  0.37
##
##  p-values
##             V10        V12
```

```
## (Intercept) < 2.22e-16 1.0828e-13
## V70         0.5537396  0.9578857
## V71         0.3597515  0.5668520
## V64         0.0075752  0.9266669
## V70:V71     0.8588647  0.5420241
```

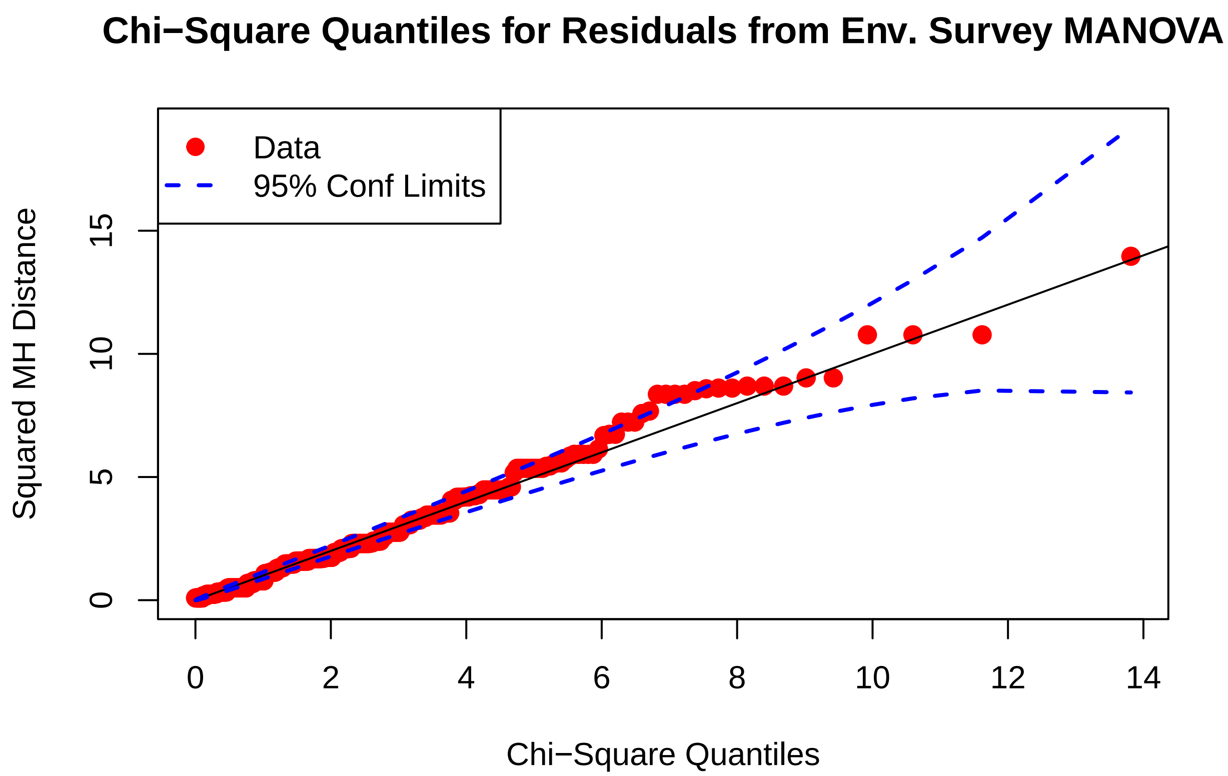
```
#Fit the model
dataaMod3 <- lm(V10 ~ V70*V71+V64,
                data = data)

#Multivariate results and univariate results with with type 3 Sum of squares
summary(Anova(dataaMod3, type = 3), univariate = T)
```

```
##      Sum Sq             Df            F value             Pr(>F)
##  Min.   :  0.0514   Min.   :  1.00   Min.    : 0.03165   Min.   :0.000000
##  1st Qu.:  0.7689   1st Qu.:  1.00   1st Qu.: 0.35114   1st Qu.:0.007575
##  Median :  6.5209   Median :  1.00   Median : 0.84032   Median :0.359752
##  Mean   :159.4863   Mean   : 83.33   Mean    :18.84362   Mean   :0.355986
##  3rd Qu.:107.4326   3rd Qu.:  1.00   3rd Qu.: 7.19016   3rd Qu.:0.553740
##  Max.   :803.9033   Max.   :495.00   Max.    :85.80484   Max.   :0.858865
##                                      NA's    :1          NA's   :1
```

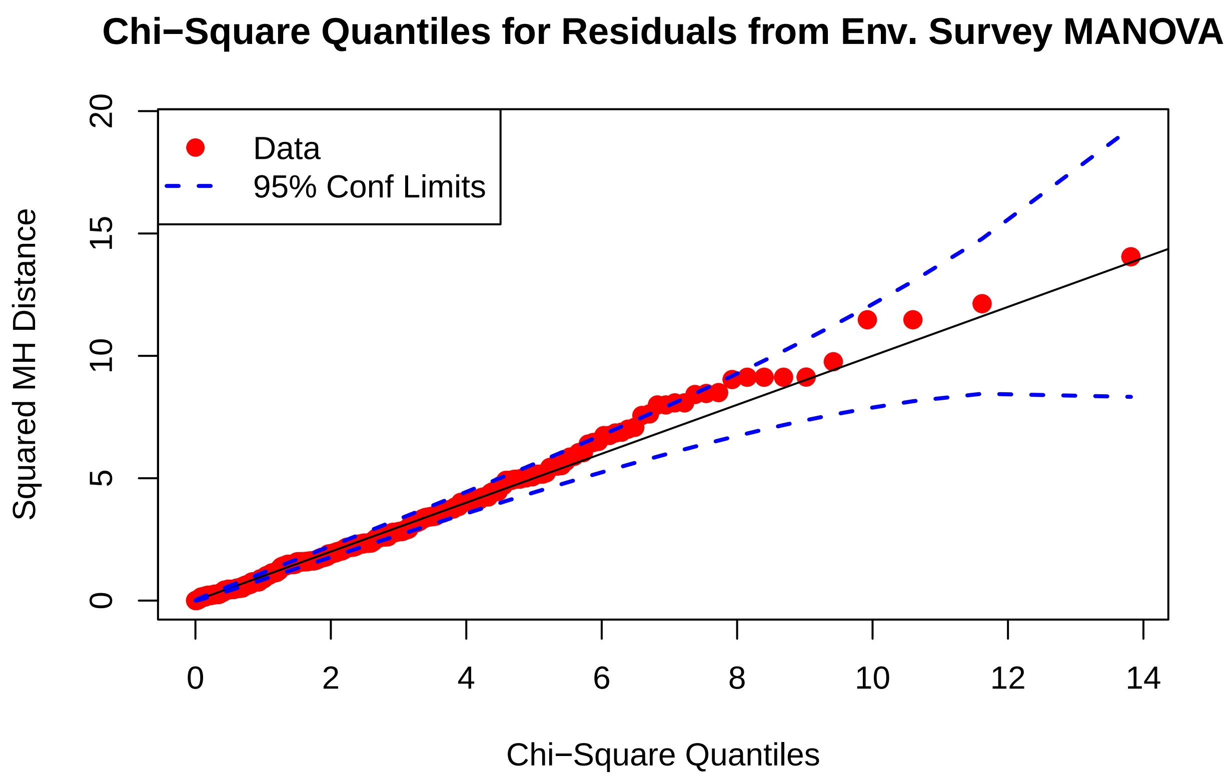
```
#Fit the model
dataaMod4 <- lm( V12~ V70*V71+V64,
                data = data)

#Multivariate results and univariate results with with type 3 Sum of squares
summary(Anova(dataaMod4, type = 3), univariate = T)
```

```
##      Sum Sq              Df            F value             Pr(>F)
##  Min.   :   0.0076   Min.   :  1.00   Min.    : 0.00279   Min.   :0.0000
##  1st Qu.:   0.2411   1st Qu.:  1.00   1st Qu.: 0.00848   1st Qu.:0.5420
##  Median :   0.9550   Median :  1.00   Median : 0.32842   Median :0.5669
##  Mean   : 251.7632   Mean   : 83.33   Mean    :11.83624   Mean   :0.5987
##  3rd Qu.: 119.7846   3rd Qu.:  1.00   3rd Qu.: 0.37232   3rd Qu.:0.9267
##  Max.   :1349.2639   Max.   :495.00   Max.    :58.46921   Max.   :0.9579
##                                       NA's    :1          NA's   :1
```

```
#5
source("http://www.reuningscherer.net/Multivariate/R/CSQPlot.r.txt")
CSQPlot(model$residuals, label = "Residuals from Env. Survey MANOVA")
```




```
CSQPlot(dataaMod2$residuals, label = "Residuals from Env. Survey MANOVA")
```




```
#since the model assumptions of residuals:
#multivariate normal distribution (which we expect will be true based on normality in each group)
```

```r
#we can tell from the plots that two models both have normality property.
```

```r
#6
#Run a multi-response permutation procedure on the survey data
#install.packages("vegan")
library(vegan)
```

```
## Loading required package: permute

##
## Attaching package: 'permute'

## The following object is masked from 'package:DiscriMiner':
##
##     getWithin

## Loading required package: lattice

## Registered S3 methods overwritten by 'vegan':
##   method      from
##   plot.rda    klaR
##   predict.rda klaR
##   print.rda   klaR

## This is vegan 2.5-7

##
## Attaching package: 'vegan'

## The following object is masked from 'package:klaR':
##
##     rda
```

```r
(mrpp1 <- mrpp(data[,c("V10", "V12")], data$V71))
```

```
##
## Call:
## mrpp(dat = data[, c("V10", "V12")], grouping = data$V71)
##
## Dissimilarity index: euclidean
## Weights for groups:  n
##
## Class means and counts:
##
##       1     2     3
## delta 2.519 2.603 3.119
## n     461   31    8
##
## Chance corrected within-group agreement A: 0.001191
## Based on observed delta 2.534 and expected delta 2.537
##
## Significance of delta: 0.215
## Permutation: free
## Number of permutations: 999
```

```r
# The MRPP test is run on response variables "V10" and "V12" and the independent variable "V71". The de
```