Multivariate Contrasts

David Lieberman, Yavuz Ramiz Çolak, Ryo Tamaki, Liana Wang

April 5th, 2019

Challenge Question

In this pset, we will do the challenge question.

Environmental Attitude Survey:

For four questions, the scale is (1=Strongly Agree, 5=Strongly Disagree) v10 : Modern science will solve our environmental problems with little change to our way of life v12 : Almost everything we do in modern life harms the environment. V25 : There is no point in doing what I can for the environment unless others do the same v47 : Poorer countries should be expected to make less effort than richer countries to protect the environment

One question has a different scale : v38 : In general, do you think that a rise in the world's temperature caused by the greenhouse effect is (1=Extremely Dangerous for you and family, 5=Not dangerous for you and family)

In addition, the following variables were measured V3 : Country (Japan, US, Mexico, Netherlands) V200 : Gender (1=male, 2=female) V201 : Age in Years V202 : Marital Status (1=Married, 2=Widowed, 3=Divorced, 4=Separated, 5=Single never married) V204 : Years of education V205 : Highest Degree obtained V246 : Political Views (1=Extremely Liberal (Communist), 5=Extremely Conservative (Fascist))

```
envsurvey <-  read.csv("http://www.reuningscherer.net/stat660/data/Environmental_Survey_MANOVA.csv", header=T)
envsurvey <- envsurvey[1:6]
envsurvey
```

```
##                   V3 V10 V12 V25 V38 V47
## 1   USA-United States   4   4   2   4   4
## 2   USA-United States   4   3   2   3   4
## 3   USA-United States   2   5   5   2   5
## 4   USA-United States   3   2   2   2   3
## 5   USA-United States   3   4   4   2   4
## 6   USA-United States   4   1   4   1   4
## 7   USA-United States   4   4   4   3   4
## 8   USA-United States   2   3   4   3   3
## 9   USA-United States   2   4   4   1   4
## 10  USA-United States   4   4   5   2   3
## 11  USA-United States   3   4   2   1   2
## 12  USA-United States   5   4   3   3   2
## 13  USA-United States   2   2   4   1   4
## 14  USA-United States   3   2   4   3   4
## 15  USA-United States   2   4   4   3   4
## 16  USA-United States   2   3   5   3   4
## 17  USA-United States   4   2   4   3   4
## 18  USA-United States   4   4   3   2   4
## 19  USA-United States   5   2   5   3   4
## 20  USA-United States   5   2   4   1   1
## 21  USA-United States   2   2   2   3   4
## 22  USA-United States   2   2   2   3   2
## 23  USA-United States   4   2   4   2   2
## 24  USA-United States   4   4   4   3   5
## 25  USA-United States   1   2   4   3   2
## 26  USA-United States   2   4   5   4   4
## 27  USA-United States   4   2   4   3   5
## 28  USA-United States   2   4   2   3   5
## 29  USA-United States   4   5   5   1   5
## 30  USA-United States   4   3   3   2   2
## 31  USA-United States   4   2   4   1   4
## 32  USA-United States   4   4   4   3   4
## 33  USA-United States   1   2   3   4   2
## 34  USA-United States   2   1   4   2   4
## 35  USA-United States   3   2   3   3   5
## 36  USA-United States   2   2   2   1   2
## 37  USA-United States   2   4   2   3   5
## 38  USA-United States   3   3   3   3   4
## 39  USA-United States   3   2   2   2   1
## 40  USA-United States   2   2   4   4   2
## 41  USA-United States   1   1   4   1   3
## 42  USA-United States   3   2   4   2   3
## 43  USA-United States   4   2   1   1   5
## 44  USA-United States   4   2   4   2   4
## 45  USA-United States   4   4   2   3   5
## 46  USA-United States   2   2   4   3   3
## 47  USA-United States   3   4   3   3   2
## 48  USA-United States   3   3   3   4   3
## 49  USA-United States   2   4   4   3   2
## 50  USA-United States   3   3   4   3   3
## 51  USA-United States   3   3   4   1   4
## 52  USA-United States   2   1   1   2   2
```

```r
multicontrast <- function(contrast, data, grouping){

  # Groups
  groups <- as.vector(as.matrix(unique(grouping)))

  # Multivariate Means for Each Variable in Each Group
  M <- apply(data, 2, function(y) tapply(y, grouping, mean))
  M <- M[match(groups, row.names(M)),]

  # Counts for Each Group
  N <- table(grouping)
  N <- N[match(groups, row.names(N))]

  # Calculate Weighted Sum of Squared Weights
  SW <- sum(contrast^2 / N)

  # Calculate SSCP Between (Hypothesis Matrix)
  C_hat <- colSums(contrast*M)
  SSCP_between <- (cbind(C_hat) %*% rbind(C_hat)) / SW

  # Calculate SSCP Within (Error Matrix)
  SSCP_within_each_group <- list()
  for (i in seq_along(groups)){
    X <- subset(data, grouping == groups[i])
    deviations <- matrix(0, nrow(X), ncol(X))
    for (j in 1:ncol(X)){
      deviations[,j] <- X[,j] - M[i,j]
    }
    SSCP_within_each_group[[i]] <- t(deviations) %*% deviations
  }
  SSCP_within <- Reduce("+", SSCP_within_each_group)

  # Calculate Wilks' Lambda
  lambda <- det(SSCP_within) / det(SSCP_between + SSCP_within)

  # Calculate Degrees of Freedom
  p <- ncol(data) # no.variables
  m <- nrow(data) - length(groups)  # no.observations - no.groups

  df1 <- p
  df2 <- m-p+1

  # Calculate approx. F
  F.stat <- (1 - lambda) / lambda * df2 / df1

  # Calculate p-value from F distribution
  p.value <- 1 - pf(F.stat, df1, df2)

  # Output
  out <- c(lambda, F.stat, df1, df2, p.value)
  names(out) <- c("Wilks", "approx.F", "df1", "df2", "p.value")
  return(out)
}
```

```
as.matrix(unique(envsurvey[1])) # Just to see the countries' order for the contrasts vector
```

```
##     V3
## 1   "USA-United States"
## 201 "NL-Netherlands"
## 401 "J-Japan"
## 601 "MEX-Mexico"
```

Japan vs US

```
multicontrast(c(1,0,-1,0), envsurvey[2:6], envsurvey[1])
```

```
##       Wilks    approx.F         df1         df2     p.value
##   0.7960165  40.5908514   5.0000000 792.0000000   0.0000000
```

US vs Rest of the World

```
multicontrast(c(3,-1,-1,-1), envsurvey[2:6], envsurvey[1])
```

```
##        Wilks     approx.F          df1          df2      p.value
## 9.410041e-01 9.930834e+00 5.000000e+00 7.920000e+02 3.157384e-09
```

*We check our results by the SAS output presented in the lecture notes. We see that our R function gives the same result as the already-existing SAS function.

Bonus: Adjusted p-values

We also calculated the BONUS question of the original pset.

```
pvalues <- as.numeric(c(
multicontrast(c(3,-1,-1,-1), envsurvey[2:6], envsurvey[1])[5],
multicontrast(c(1,0,-1,0), envsurvey[2:6], envsurvey[1])[5]
))

p.adjust(pvalues, method="bonferroni")
```

```
## [1] 6.314768e-09 0.000000e+00
```

```
p.adjust(pvalues, method="holm")
```

```
## [1] 3.157384e-09 0.000000e+00
```

```
p.adjust(pvalues, method="hochberg")
```

```
## [1] 3.157384e-09 0.000000e+00
```