DSA 8070 R Session 5: Comparisons of Several Mean Vectors

Whitney

September 26, 2022

Contents

Swiss Bank Notes Example

Read the data

```r
url <- "https://online.stat.psu.edu/stat505/sites/stat505/files/lesson07/swiss3.txt"
dat <- read.table(url, header = F)
```

Calculate summary statistics

```r
real <- which(dat$V1 == "real")
fake <- which(dat$V1 == "fake")
(xbar1 <- colMeans(dat[real, -1]))
```

```
##      V2      V3      V4      V5      V6      V7
## 214.969 129.943 129.720   8.305  10.168 141.517
```

```r
(xbar2 <- colMeans(dat[fake, -1]))
```

```
##      V2      V3      V4      V5      V6      V7
## 214.823 130.300 130.193  10.530  11.133 139.450
```

```r
(Sigma1 <- cov(dat[real, -1]))
```

```
##              V2          V3          V4            V5          V6            V7
## V2 0.150241414  0.05801313  0.05729293  0.0571262626  0.01445253  0.0054818182
## V3 0.058013131  0.13257677  0.08589899  0.0566515152  0.04906667 -0.0430616162
## V4 0.057292929  0.08589899  0.12626263  0.0581818182  0.03064646 -0.0237777778
## V5 0.057126263  0.05665152  0.05818182  0.4132070707 -0.26347475 -0.0001868687
## V6 0.014452525  0.04906667  0.03064646 -0.2634747475  0.42118788 -0.0753090909
## V7 0.005481818 -0.04306162 -0.02377778 -0.0001868687 -0.07530909  0.1998090909
```

```r
(Sigma2 <- cov(dat[fake, -1]))
```

```
##              V2           V3            V4          V5            V6
## V2  0.12401111  0.031515152  0.0240010101 -0.10059596  0.0194353535
## V3  0.03151515  0.065050505  0.0467676768 -0.02404040 -0.0119191919
## V4  0.02400101  0.046767677  0.0889404040 -0.01857576  0.0001323232
## V5 -0.10059596 -0.024040404 -0.0185757576  1.28131313 -0.4901919192
## V6  0.01943535 -0.011919192  0.0001323232 -0.49019192  0.4044555556
## V7  0.01156566 -0.005050505  0.0341919192  0.23848485 -0.0220707071
##              V7
## V2  0.011565657
## V3 -0.005050505
## V4  0.034191919
## V5  0.238484848
## V6 -0.022070707
## V7  0.311212121
```

```r
n1 <- length(real); n2 <- length(fake); p <- dim(dat[, -1])[2]
Sp <- ((n1 - 1) * Sigma1 + (n2 - 1) * Sigma2) / (n1 + n2 - 2)
```

Perform a two-sample Hotelling's T-Square test

```r
# Test statistic
T.squared <- as.numeric(t(xbar1 - xbar2) %*% solve(Sp * (1 / n1 + 1 / n2)) %*% (xbar1 - xbar2))
Fobs <- T.squared * ((n1 + n2 - p - 1) / ((n1 + n2 - 2) * p))
# p-value
pf(Fobs, p, n1 + n2 - p - 1, lower.tail = F)
```

```
## [1] 3.378887e-105
```

Simultaneous Confidence Intervals

```r
s1 <- diag(Sigma1); s2 <- diag(Sigma2)

xbar_diff <- xbar1 - xbar2
sp_diff <- ((n1 - 1) * s1 + (n2 - 1) * s2) / (n1 + n2 - 2)

multipler <- sqrt((p * (n1 + n2 - 2) / (n1 + n2 - p - 1)) * qf(0.95, p, n1 + n2 - p - 1))

sp <- sqrt((1 / n1 + 1 / n2) * sp_diff)
```

```r
CIs <- cbind(xbar_diff + -1 * multipler * sp, xbar_diff + 1 * multipler * sp)
CIs
```

```
##           [,1]       [,2]
## V2 -0.04432667  0.3363267
## V3 -0.51856518 -0.1954348
## V4 -0.64159649 -0.3044035
## V5 -2.69809429 -1.7519057
## V6 -1.29523310 -0.6347669
## V7  1.80719722  2.3268028
```

MANOVA: Romano-British Pottery Example

```r
dat <- read.table("pottery.txt", header = F)
out <- manova(cbind(V2, V3, V4, V5, V6) ~ V1, data = dat)
summary(out, test = "Wilks")
```

```
##           Df    Wilks approx F num Df den Df   Pr(>F)
## V1         3 0.012301   13.088     15 50.091 1.84e-12 ***
## Residuals 22
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(out)
```

```
##           Df Pillai approx F num Df den Df    Pr(>F)
## V1         3 1.5539   4.2984     15     60 2.413e-05 ***
## Residuals 22
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```