

STAT 625 F2022 Assignment 2 - Due on Dec. 9, Friday, 2022, 11:59pm

2022-11-28

Contents

General Policies [Total 60 marks]	2
Problem 4. Exercise 6.27 on pp. 350-351 of JW.	3

General Policies [Total 60 marks]

- Your assignment solutions should be created in "RStudio" with R Markdown to include the R codes and should be saved as an R Markdown file and a PDF file generated from the R Markdown file. You should name your files as "Lastname-Firstname-Stat635-A3.Rmd" and "Lastname-Firstname-Stat635-A3.pdf".
- Test your R codes before submission to make sure it can be executed successfully in "RStudio".
- For each assignment, submit only one PDF file and the associated Rmd file to D2L, and the same PDF to gradescope.ca. Only the PDF file in gradescope is graded. The Rmd file may be used to test your R programs when needed.
- If Monte-Carlo methods are used, you must fix the random seed in your R code by using `set.seed(2022)`.
- For all numerical problems, summarize the computer generated results in tables or figures and interpret them using your own words, then draw conclusions from them. **Show all your work and spell out the details.**
- Late submission is not acceptable.
- For guidelines on how to write a good assignment, go to D2L to read two sample assignments, **Bad-assignment-example.pdf** and **Good-assignment-example.pdf**, one is bad and the other is good, you are expected to do a good one.

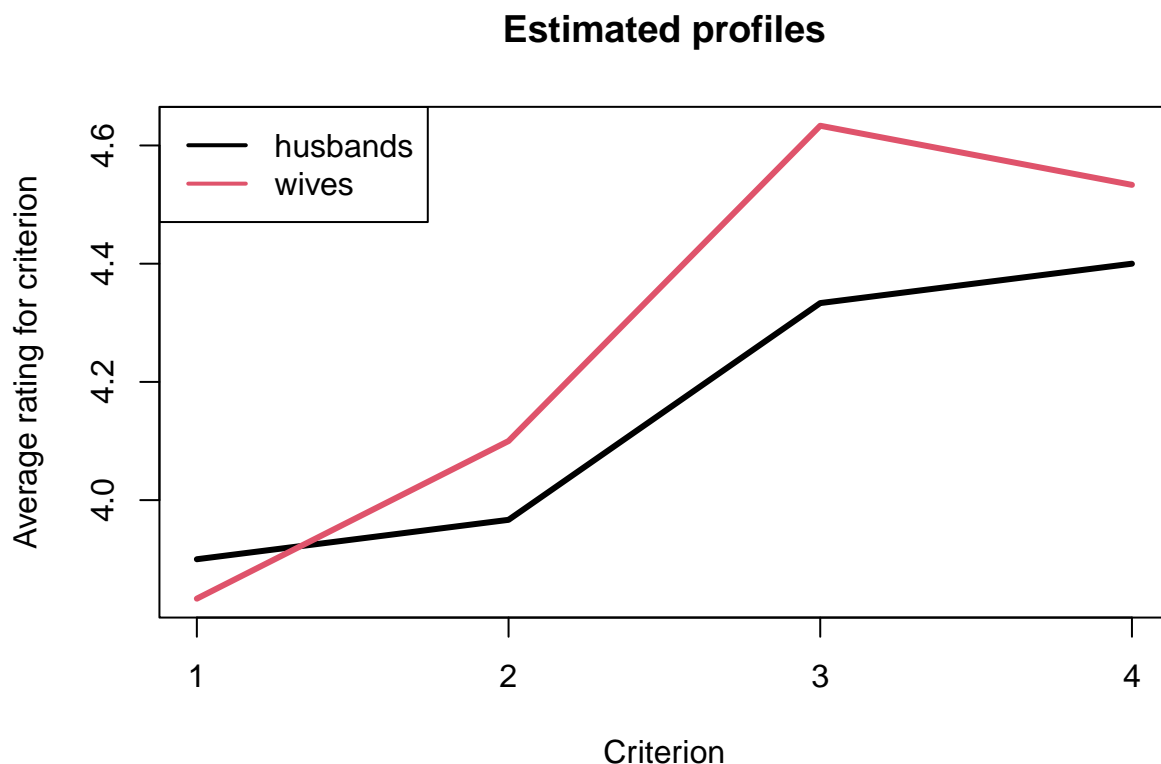
Note: Materials are based on Dobson and Barnett (D&B or DB), 3rd Edition, 2008.

Problem 4. Exercise 6.27 on pp. 350-351 of JW.

a) Plot the mean vectors for husbands and wives as sample profiles

Read the data and draw the profile plot

```
#####  
# Profile analysis of spouses rating each other; data in Table 6.14  
data = read.table("C:/Users/Hao Nan Wang/Desktop/schedule/ucalgary/STAT 625/T6-14.dat")  
tail(data,1)  
  
##      V1 V2 V3 V4   V5  
## 60   4  4  5  5 Wife  
  
X = as.matrix(data[,1:4])  
spouse = c(rep(1,30), rep(2,30)) # 1 = Husband rating wife, 2 = Wife rating husband  
spouse = as.factor(spouse)  
xbar1 = colMeans(X[spouse==1,])  
xbar2 = colMeans(X[spouse==2,])  
# Plot of estimated profiles:  
matplot(1:4, cbind(xbar1,xbar2), type = 'l', xaxp = c(1,4,3),  
lty = 1, lwd = 3, col= 1:2, xlab = "Criterion", ylab = "Average rating for criterion",  
main = "Estimated profiles")  
legend("topleft", c("husbands", "wives"), lty = 1, lwd = 2, col= 1:2)
```



When considering the profile plot constructed from the sample data above, it appears there may be a parallel relationship in the population as the two samples move closely together.

b) parallelism test, coincidence test and level test

Test for parallel

```
#####  
## Test for parallel profiles, two ways  
C = rbind(c(-1, 1, 0, 0), c(0, -1, 1, 0), c(0, 0, -1, 1)) # contrast matrix  
Y = X%*%t(C)  
# Calc Tsqd:  
ybar1 = colMeans(Y[spouse==1,])  
ybar2 = colMeans(Y[spouse==2,])  
n1 = nrow(Y[spouse==1,])  
n2 = nrow(Y[spouse==2,])  
dbar = ybar1 - ybar2  
S1 = cov(Y[spouse==1,])  
S2 = cov(Y[spouse==2,])  
dfS = (n1-1) + (n2-1)  
Spooled = ((n1-1)*S1 + (n2-1)*S2)/dfS  
cov.est = (1/n1 + 1/n2)*Spooled  
Tsqd = t(dbar)%*%solve(cov.est, dbar)  
  
df1 = length(dbar)  
df2 = dfS - df1 + 1  
F = df2*Tsqd/(df1*dfS)  
pval = pf(F, df1, df2, lower.tail = 0)  
pval # .062559
```

```
##           [,1]  
## [1,] 0.06255945
```

```
# Using the R function manova():  
fit = manova(Y ~ spouse)  
summary(fit)
```

```
##           Df  Pillai approx F num Df den Df  Pr(>F)  
## spouse      1 0.12143    2.5799      3    56 0.06256 .  
## Residuals 58  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Summarize an Analysis of Variance Model  
summary.aov(fit) # Third difference in slopes is most significant
```

```
## Response 1 :  
##           Df Sum Sq Mean Sq F value Pr(>F)  
## spouse      1  0.600  0.60000   0.8758 0.3532  
## Residuals  58 39.733  0.68506  
##  
## Response 2 :  
##           Df Sum Sq Mean Sq F value Pr(>F)  
## spouse      1  0.417  0.41667   0.4792 0.4916  
## Residuals  58 50.433  0.86954  
##
```

```
## Response 3 :
##           Df Sum Sq Mean Sq F value    Pr(>F)
## spouse      1  0.4167   0.41667    5.292 0.02504 *
## Residuals   58  4.5667   0.07874
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$H_{10} : \mu_{Y1} = \mu_{Y2} \text{ against } H_{11} : \mu_{Y1} \neq \mu_{Y2}$$

From manova and manual computation, We get p-value = 0.06256, which is greater than $\alpha = 0.05$, thus, we DO NOT reject the null hypothesis at 95% significant level. This means that the parallelism hypothesis is not rejected and the husband rating wife profile is parallel to the wife rating husband profile. In other words, successive (adjacent) differences between means are equal.

Test for coincident

```
#####
## Test for coincident profiles
x1 = rowSums(X[spouse==1,]) # 1 = Husband rating wife, 2 = Wife rating husband
x2 = rowSums(X[spouse==2,])
# If var. equal is TRUE then the pooled estimate of the variance is used.
t.test(x1,x2, var.equal = T) # pval = .2207
```

```
##
## Two Sample t-test
##
## data:  x1 and x2
## t = -1.2381, df = 58, p-value = 0.2207
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.308415  0.308415
## sample estimates:
## mean of x mean of y
##      16.6      17.1
```

```
t.test(x1,x2) # Welch test; still pval = .2207
```

```
##
## Welch Two Sample t-test
##
## data:  x1 and x2
## t = -1.2381, df = 57.793, p-value = 0.2207
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.3084767  0.3084767
## sample estimates:
## mean of x mean of y
##      16.6      17.1
```

```
# Using the R function aov():
fit = aov(rowSums(X) ~ spouse)
summary(fit)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## spouse      1   3.75   3.750   1.533  0.221
## Residuals   58 141.90   2.447
```

$$H_{20} : \mu_{Y1} = \mu_{Y2} \text{ against } H_{21} : \mu_{Y1} \neq \mu_{Y2}$$

From t-test or aov() function, we can clearly see that the p-value is 0.2207, which is greater than $\alpha = 0.05$, this means that we DO NOT REJECT the null hypothesis at 95% significance level and making a conclusion that coincidence hypothesis is not rejected, which means that the wife rating husband profile and the husband rating wife profile are coincident, meaning the population means are equal for the two groups.

Test for flatness

```
#####
## Test for level profiles, using the combined sample of Y's
dbar = colMeans(Y)
df1 = length(dbar)
S = cov(Y)
cov.est = S/nrow(Y)
dfS = nrow(Y) - 1
Tsqr = t(dbar)%*%solve(cov.est, dbar)
df2 = dfS - df1 + 1
Fcalc = df2*Tsqr/(df1*dfS)
pval = pf(Fcalc, df1, df2, lower.tail = 0)
pval # .00015
```

```
##           [,1]
## [1,] 0.0001554491
```

$$H_{30} : \mu_U = 0_{(P-1) \times 1} \text{ against } H_{31} : \mu_U \neq 0$$

As we can see that the p-value is less than $\alpha = 0.05$. This means that we REJECT the null hypothesis at 95% significance level and making a conclusion that flatness hypothesis is rejected, which means that the wife rating husband profile and the husband rating wife profile are not flat, meaning it is not the case where all the means are equal.