Multivariate Analysis - H.2

Max Machalek

03 October 2019

```r
library(car)

library(yacca)
library(CCP)
# data as loaded into a csv:
# 72 26 9 10 11 70
# 63 76 7 85 22 93
# 96 31 7 83 63 73
# 96 98 6 82 75 97
# 84 94 6 36 77 97
# 66 10 5 28 24 75
# 31 40 9 64 23 75
# 45 14 2 19 15 50
# 42 18 6 33 13 70
# 79 74 4 23 14 90
# 39 12 2 37 13 70
# 54 35 3 23 74 53
# 60 75 5 45 58 83
# 63 45 5 22 67 53

# making data readable for rmd to knit .pdf
dat <- read.csv("datv.csv", sep = ",", header = FALSE)
colnames(dat) <- c("career","supervisor","finance",
"variety","feedback","autonomy")

Y <- dat[,1:3]
X <- dat[,4:6]
```

H.2.1

(a) Give the overall F-test results and R # values for the 3 univariate multiple linear regressions of each measure of employee satisfaction (Y) given all job characteristics (X). For each of the 3 model fits, interpret the R # values and explain which satisfaction responses have linear associations in the populaton with the job characteristics at the 0.05 level using the overall F-test.

Supervisor is the only response with an overall F-test that indicates, at the 0.05 level, a linear association in the population between the predictors (variety, feedback, autonomy), and any of the responses. Because the p-value for the overall F-tests for the career and finance models was not below 0.05, the results of the tests do not provide sufficient evidence that those responses have a linear association in the population between any of the predictors. Career had a notably low p-value though (0.07), and a moderate adjusted R-squared of 0.34.

H_0: $\beta_{x1} = \beta_{x2} = \beta_{x3} = 0$

All predictor regression coefficients in the model for the given univariate response are 0.
H_a: β_{x1} or β_{x2} or $\beta_{x3} \neq 0$

Not all predictor regression coefficients in the model for the given univariate response are 0.

$\alpha = 0.05$

<u>Career as response:</u>
F-statistic: 3.218
p-value: 0.06994

We have insufficient evidence at the 0.05 level to reject the null hypothesis, and are unable to claim that there is a linear association in the population between career and any of variety, feedback, or autonomy (p = 0.07). It is worth noting that while the p-value was not less than 0.05, it was very close to it.

R-squared: 0.4912, Adjusted = 0.3385.

About 49% of the variation in the career response is explained by the first order linear regression model including variety, feedback, and autonomy.

Supervisor as response:
F-statistic: 12.34
p-value: 0.001067

We have sufficient evidence at the 0.05 level to reject the null hypothesis and claim that there is a linear association in the population between supervisor and at least one of variety, feedback, or autonomy (p = 0.00107).

R-squared: 0.7874, Adjusted = 0.7236

Almost 79% of the variation in supervisor is explained by the first order linear regression model including variety, feedback, and autonomy.

Finance as response:
F-statistic: 0.874
p-value: 0.4866

We have insufficient evidence at the 0.05 level to reject the null hypothesis and claim that there is a linear association in the population between finance and at least one of variety, feedback, or autonomy (p = 0.4866).

R-squared: 0.2077, Adjusted = -0.02994

When considering an un-adjusted R-squared, 20% of the variation in finance is explained by the first order linear regression model including variety, feedback, and autonomy. The adjusted R-squared becomes negative, indicating that the model including all three of these predictors is not appropriate.

```
# career as response
fitCareer <- lm(Y$career~X$variety + X$feedback + X$autonomy)
summary(fitCareer)

##
## Call:
## lm(formula = Y$career ~ X$variety + X$feedback + X$autonomy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.0017  -9.4248  -0.3442   7.6479  28.2023
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.13428   23.42280   0.134   0.8962
## X$variety   -0.07783    0.21807  -0.357   0.7286
## X$feedback   0.36484    0.17407   2.096   0.0625 .
## X$autonomy   0.65943    0.34497   1.912   0.0850 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 16.6 on 10 degrees of freedom
## Multiple R-squared:  0.4912, Adjusted R-squared:  0.3385
## F-statistic: 3.218 on 3 and 10 DF,  p-value: 0.06994
```

```
# supervisor as response
fitSupervisor <- lm(Y$supervisor~X$variety + X$feedback + X$autonomy)
summary(fitSupervisor)
```

```
##
## Call:
## lm(formula = Y$supervisor ~ X$variety + X$feedback + X$autonomy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.299  -7.806   2.913  12.669  17.227
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -82.2718    23.0723  -3.566  0.00513 **
## X$variety    -0.1363     0.2148  -0.634  0.54006
## X$feedback    0.5273     0.1715   3.075  0.01174 *
## X$autonomy    1.5164     0.3398   4.463  0.00121 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.35 on 10 degrees of freedom
## Multiple R-squared:  0.7874, Adjusted R-squared:  0.7236
## F-statistic: 12.34 on 3 and 10 DF,  p-value: 0.001067
```

```
# finance as response
fitFinance <- lm(Y$finance~X$variety + X$feedback + X$autonomy)
summary(fitFinance)
## Call:
## lm(formula = Y$finance ~ X$variety + X$feedback + X$autonomy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3943 -0.8783 -0.2499  0.7670  4.2989
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.35168    3.16314   0.743    0.474
## X$variety    0.02639    0.02945   0.896    0.391
## X$feedback  -0.00976    0.02351  -0.415    0.687
## X$autonomy   0.03133    0.04659   0.672    0.517
##
## Residual standard error: 2.242 on 10 degrees of freedom
## Multiple R-squared:  0.2077, Adjusted R-squared:  -0.02994
## F-statistic: 0.874 on 3 and 10 DF,  p-value: 0.4866
```

(b) Obtain the multivariate tests results of the 3 job characteristic variables. Give the null and alternative statistical hypotheses for these 3 tests. Provide the statistical decisions and practical conclusions of these 3 tests at the 0.05 level.

<u>Characteristic: Variety</u>
H_0: $\beta_{j\_variety} = 0$ for all j
H_a: $\beta_{j\_variety} \neq 0$ for any j

$\alpha = 0.05$
Wilks test-statistic: 0.8728577
p-value: 0.7646

Because the test statistic is 0.873 with a p value of 0.76, we have insufficient evidence to reject the null hypothesis at the 0.05 level. We cannot claim that B(variety)j != 0 for any of the responses (j = career, supervisor, or finance). There is not evidence of a linear association in the population between variety and any of the responses (career, supervisor, finance).

<u>Characteristic: Feedback</u>
H_0: $\beta_{j\_feedback} = 0$ for all j
H_a: $\beta_{j\_feedback} \neq 0$ for any j

$\alpha = 0.05$
Wilks test-statistic: 0.3751042
p-value: 0.040722

Because the test statistic is 0.375 with a p value of 0.041, we have sufficient evidence to reject the null hypothesis at the 0.05 level and claim that B(feedback)j != 0 for some response j (j = career, supervisor, or finance). There is evidence of a linear association in the population between feedback and at least one of the responses (career, supervisor, finance).

<u>Characteristic: Autonomy</u>
H_0: $\beta_{j\_autonomy} = 0$ for all j
H_a: $\beta_{j\_autonomy} \neq 0$ for any j

$\alpha = 0.05$
Wilks test-statistic: 0.2663401
p-value: 0.010981

Because the test statistic is 0.2663 with a p value of 0.011, we have sufficient evidence to reject the null hypothesis at the 0.05 level and claim that B(autonomy)j != 0 for some response j (j = career, supervisor, or finance). There is evidence of a linear association in the population between autonomy and at least one of the responses (career, supervisor, finance).

```
Y <- as.matrix(Y) # idk why I need to do this here for it to work, but
I do
mlm <- lm(Y~X$variety + X$feedback + X$autonomy)

MA <- Manova(mlm, test = 'Wilks')

SA <- summary(MA)
SA

##
## Type II MANOVA Tests:
##
## Sum of squares and products for error:
##                 career supervisor   finance
## career     2755.4527  -444.3585  45.75380
## supervisor -444.3585  2673.6022 -24.78680
## finance      45.7538   -24.7868  50.25183
##
## ------------------------------------------
##
## Term: X$variety
##
## Sum of squares and products for the hypothesis:
##                 career supervisor    finance
## career      35.09822   61.45477 -11.902832
## supervisor  61.45477  107.60342 -20.841109
## finance    -11.90283  -20.84111   4.036599
##
## Multivariate Tests: X$variety
##                  Df test stat  approx F num Df den Df Pr(>F)
## Pillai            1 0.1271423 0.3884322      3      8 0.7646
## Wilks             1 0.8728577 0.3884322      3      8 0.7646
## Hotelling-Lawley  1 0.1456621 0.3884322      3      8 0.7646
## Roy               1 0.1456621 0.3884322      3      8 0.7646
##
## ------------------------------------------
##
## Term: X$feedback
##
## Sum of squares and products for the hypothesis:
##                  career supervisor     finance
## career     1210.46476 1749.33620 -32.3801474
## supervisor 1749.33620 2528.10097 -46.7950541
## finance     -32.38015  -46.79505   0.8661747
##
## Multivariate Tests: X$feedback
##                  Df test stat approx F num Df den Df   Pr(>F)
## Pillai            1 0.6248958 4.442469      3      8 0.040722 *
## Wilks             1 0.3751042 4.442469      3      8 0.040722 *
## Hotelling-Lawley  1 1.6659258 4.442469      3      8 0.040722 *
```

```
## Roy               1 1.6659258 4.442469      3      8 0.040722 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: X$autonomy
##
## Sum of squares and products for the hypothesis:
##                 career supervisor    finance
## career     1006.86284  2315.3990  47.831668
## supervisor 2315.39903  5324.5313 109.994522
## finance      47.83167   109.9945   2.272274
##
## Multivariate Tests: X$autonomy
##                  Df test stat approx F num Df den Df   Pr(>F)
## Pillai            1 0.7336599 7.345596      3      8 0.010981 *
## Wilks             1 0.2663401 7.345596      3      8 0.010981 *
## Hotelling-Lawley  1 2.7545985 7.345596      3      8 0.010981 *
## Roy               1 2.7545985 7.345596      3      8 0.010981 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

H.2.2

(a) Give the first canonical correlation value and explain what it represents in this application.

The first canonical correlation, CV1, is 0.919. In this context, it indicates that maximized variety and minimized feedback and autonomy are linearly associated with lower career, supervisor, and finance values.

```
X <- as.matrix(X)
cc <- cca(X,Y, xscale = T, yscale = T)
cc

##
## Canonical Correlation Analysis
##
## Canonical Correlations:
##      CV 1      CV 2      CV 3
## 0.9194120 0.4186491 0.1133658
##
## X Coefficients:
##                CV 1       CV 2       CV 3
## variety   0.1108295  0.8095098 -0.9071112
## feedback -0.5519543 -0.7721632 -0.4194377
## autonomy -0.8402945  0.1019873  0.8296945
##
## Y Coefficients:
##                   CV 1       CV 2       CV 3
## career     -0.30284133 -0.5416140 -1.0407753
## supervisor -0.78536978  0.1305349  0.9084518
## finance    -0.05377069  0.9754208 -0.3329223
##
## Structural Correlations (Loadings) - X Vars:
##                CV 1       CV 2       CV 3
## variety  -0.4863184  0.6591844 -0.5735594
## feedback -0.6215686 -0.5452117 -0.5624915
## autonomy -0.8459188  0.4450690  0.2938283
##
## Structural Correlations (Loadings) - Y Vars:
##                  CV 1       CV 2       CV 3
## career     -0.7499135 -0.2503382 -0.6123402
## supervisor -0.9644439  0.0361891  0.2617981
## finance    -0.2873325  0.8813524 -0.3750441
##
## Aggregate Redundancy Coefficients (Total Variance Explained):
##  X | Y: 0.4345932
##  Y | X: 0.4954371
```

(b) Give the test results of the canonical correlations. Explain which population canonical correlations should be declared to be non-zero.

We should only retain the first canonical correlation. Given the p-values from our sequential hypothesis testing, we could conclude that at least one of p(1), p(2), or p(3) was non-zero. We could not conclude that at least one of p(2), or p(3) was zero. this implies that p(1) alone is non-zero (at the 0.05 level).

H_{01}: p(1) = p(2) = p(3) = 0
H_{a1}: p(1) ≠ 0 or p(2) ≠ 0 or p(3) ≠ 0

reject (p = 0.022)

H_{02}: p(2) = p(3) = 0
H_{a2}: p(2) ≠ 0 or p(3) ≠ 0

fail to reject (p = 0.745)

H_{03}: p(3) = 0
H_{a3}: p(3) ≠ 0

fail to reject (p = 0.726)

```
wilks <- F.test.cca(cc)
wilks

##
##  F Test for Canonical Correlations (Rao's F Approximation)
##
##         Corr       F  Num df Den df  Pr(>F)
## CV 1 0.91941 2.92751 9.00000  19.62 0.02233 *
## CV 2 0.41865 0.48729 4.00000  18.00 0.74497
## CV 3 0.11337 0.13019 1.00000  10.00 0.72574
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(c) Give the standardized canonical coefficients. Interpet the first canonical variate using the coefficients.

The first canonical variate v(1) the combination of variety (small contribution), feedback (large contribution), and autonomy (large contribution). This variate is higher for higher amounts of variety but lower feedback and lower autonomy.

The first canonical variate u(1) consists of career (moderate contribution), supervisor (large contribution), finance (small contribution). This variate is higher for lower values of career, supervisor, and finance.

(d) Give the cross correlations between the variables and the canonical variates (2 sets). Which variables have a linear associations with the first canonical variate? Use a cut-off of 0.5 for the magnitude (absolute value) of the correlation.

The first canonical variable U1 is associated with the negative of variety, feedback, and autonomy. It has a small negative sample correlation with variety (-0.45), a moderate negative sample correlation with feedback (-0.57), and a large negative sample correlation with autonomy (-0.78).

The first canonical variable V1 is associated with the negative of career, supervisor, and finance. It has a no sample correlation with finance (-0.26), a moderate negative sample correlation with career (-0.69), and a large negative sample correlation with supervisor (-0.89).

```
Rxu <- cc$xcrosscorr
Rxu
```

```
##                CV 1       CV 2        CV 3
## variety  -0.4471269  0.2759670 -0.06502204
## feedback -0.5714777 -0.2282524 -0.06376731
## autonomy -0.7777478  0.1863277  0.03331009
```

```
Ryv <- cc$ycrosscorr
Ryv
```

```
##                  CV 1        CV 2        CV 3
## career     -0.6894795 -0.10480386 -0.06941846
## supervisor -0.8867213  0.01515053  0.02967896
## finance    -0.2641769  0.36897740 -0.04251718
```