Kyle Beitz
STAT 4300
10/03/2019

Homework 2

X[, 1] = variety, X[, 2] = feedback, X[, 3] = autonomy

> smlm

```
Career:      Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.13428   23.42280   0.134   0.8962
X[, 1]      -0.07783    0.21807  -0.357   0.7286
X[, 2]       0.36484    0.17407   2.096   0.0625
X[, 3]       0.65943    0.34497   1.912   0.0850
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: Estimate Std. Error t value Pr(>|t|)
(Intercept) -82.2718    23.0723  -3.566  0.00513
X[, 1]       -0.1363     0.2148  -0.634  0.54006
X[, 2]        0.5273     0.1715   3.075  0.01174
X[, 3]        1.5164     0.3398   4.463  0.00121
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:     Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.35168    3.16314   0.743    0.474
X[, 1]       0.02639    0.02945   0.896    0.391
X[, 2]      -0.00976    0.02351  -0.415    0.687
X[, 3]       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
```

- **H.2.1:**
  (a) Overall F-Test
    i. Career, j = 1:
      I. We can see that response career (Y) has an R^2 value of 49%. This states that 49% of the variability in career is explained by the predictor's variety, feedback and autonomy.
      II. H_0 : $\beta_{k1} = 0$ for all k = 1,2,3, H_1 : $\beta_{k1} \neq 0$ for some k = 1,2,3
      III. Since F(3,10) = 3.22 with a p-value > 0.05, we fail to reject H_0 at the 0.05 level. We cannot declare $\beta_{k1} \neq 0$ for some k =

1,2,3. Thus, there is no evidence of a linear association in the population between control and at least one of the predictors variety, feedback and autonomy.

ii. Supervisor, $j = 2$:
   I. We can see that response career (Y) has an R^2 value of 79%. This states that 79% of the variability in career is explained by the predictor's variety, feedback and autonomy.
   II. $H_0 : \beta_{k2} = 0$ for all $k = 1,2,3$, $H_1 : \beta_{k2} \neq 0$ for some $k = 1,2,3$
   III. Since $F(3,10) = 12.34$ with a p-value < 0.05, we can reject H_0 at the 0.05 level in favor of H_1. We will declare $\beta_{k2} \neq 0$ for some $k = 1,2,3$. Thus, there is a linear association in the population between supervisor and at least one of the predictors variety, feedback and autonomy.

iii. Finance, $j = 3$:
   I. We can see that response career (Y) has an R^2 value of 21%. This states that 21% of the variability in career is explained by the predictor's variety, feedback and autonomy.
   II. $H_0 : \beta_{k3} = 0$ for all $k = 1,2,3$, $H_1 : \beta_{k3} \neq 0$ for some $k = 1,2,3$
   III. Since $F(3,10) = 0.874$ with a p-value > 0.05, we fail to reject H_0 at the 0.05 level. We cannot declare $\beta_{k3} \neq 0$ for some $k = 1,2,3$. Thus, there is no evidence of a linear association in the population between control and at least one of the predictors variety, feedback and autonomy.

```
Term: X[, 1] = variety

             career supervisor    finance
career      35.09822   61.45477 -11.902832
supervisor  61.45477  107.60342 -20.841109
finance    -11.90283  -20.84111   4.036599

                 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[, 2] = feedback

              career supervisor     finance
career     1210.46476 1749.33620 -32.3801474
supervisor 1749.33620 2528.10097 -46.7950541
finance     -32.38015  -46.79505   0.8661747

                 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

Term: X[, 3] = autonomy

              career supervisor    finance
career     1006.86284  2315.3990  47.831668
supervisor 2315.39903  5324.5313 109.994522
finance      47.83167   109.9945   2.272274

                 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
```

(b) Multivariate test:

i. Variety, k = 1:

I. $H_0 : \beta_{1j} = 0$ for all j = 1,2,3, $H_1 : \beta_{1j} \neq 0$ for some j = 1,2,3

II. Since F(3,8) = 0.39 with a p-value = 0.76, we fail to reject H_0 at the 0.05 level and not in favor of H_1. We cannot declare $\beta_{1j} \neq 0$ for some j = 1,2,3. Thus, there is no evidence of a linear association in the population between variety and at least one of the responses career, supervisor and finance.

ii. Feedback, k = 2:

        I. $H_0 : \beta_{2j} = 0$ for all j = 1,2,3, $H_1 : \beta_{2j} \neq 0$ for some j = 1,2,3
        II. Since $F(3,8) = 4.44$ and a p-value = 0.04, we can reject H_0 at the 0.05 level in favor of H_1. We can declare $\beta_{2j} \neq 0$ for some j = 1,2,3. Thus, there is a linear association in the population between feedback and at least one of the responses career, supervisor and finance.

iii. Autonomy, k = 3:

        I. $H_0 : \beta_{3j} = 0$ for all j = 1,2,3, $H_1 : \beta_{3j} \neq 0$ for some j = 1,2,3
        II. Since $F(3,8) = 7.35$ and a p-value = 0.01, we can reject H_0 at the 0.05 level in favor of H_1. We can declare $\beta_{3j} \neq 0$ for some j = 1,2,3. Thus, there is a linear association in the population between autonomy and at least one of the responses career, supervisor and finance.

```
F Test for Canonical Correlations (Rao's F Approximation)
         Corr         F   Num df  Den df Pr(>F)
CV 1 0.919412       Inf 0.000000  9.4971     NA #Reject NA = 0#
CV 2 0.418649 2.282996 1.000000 10.0000 0.1617 #Fail to reject#
CV 3 0.113366 0.029198 4.000000 18.0000 0.9982 #fail to reject#

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

	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

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

- **H.2.2:**

(a) From the data above we can see that the first canonical correlation value is 0.92. This value represents that there is a strong correlation between are sets of X and Y.

(b) Test results of canonical correlations:

i. $H_0 : p(1) = p(2) = p(3) = 0$, $H^{(1)}_1 : p(1) \neq 0, p(2) \neq 0, \text{or } p(3) \neq 0$

ii. (1) Reject $H^{(1)}_0$, (2) Fail to Reject $H^{(2)}_0$, (3) Fail to Reject $H^{(3)}_0$

iii. We declare 1 population canonical correlation not 0 at the 0.05 level.

(c) The first canonical variate u_1 consists of the combination of career (-0.30) and supervisor (-0.79). The variate u_1 is higher for employee satisfaction with lower career and supervisor. The first canonical variable v_1 consists of feedback (-0.55) and autonomy (-0.84). The variate v_1 is higher for job characteristics with lower feedback and autonomy.

(d) The first canonical variable $V_{(1)}$, associated with a negative of feedback and autonomy) has a strong negative sample correlation with supervisor (-0.89). The first canonical variable $U_{(1)}$ (associated with negative of career and supervisor) has moderate negative sample correlations with feedback (-0.57) and a strong correlation with autonomy (-0.77)

R Code:

```
> datv = scan()
1: 72 26 9 10 11 70
7: 63 76 7 85 22 93
13: 96 31 7 83 63 73
19: 96 98 6 82 75 97
25: 84 94 6 36 77 97
31: 66 10 5 28 24 75
37: 31 40 9 64 23 75
43: 45 14 2 19 15 50
49: 42 18 6 33 13 70
55: 79 74 4 23 14 90
61: 39 12 2 37 13 70
67: 54 35 3 23 74 53
73: 60 75 5 45 58 83
79: 63 45 5 22 67 53
85:
Read 84 items

> dat = matrix(datv,14,6,byrow=T)
> Y = as.matrix(dat[,1:3]);
> colnames(Y) = c("career","supervisor","finance")
> X = as.matrix(dat[,4:6]);
> colnames(X) = c("variety","feedback","autonomy")
> n = nrow(dat); p = ncol(Y); m = ncol(X);
> R = cor(dat)
> mlm = lm(Y~X[,1]+X[,2]+X[,3])
> smlm = summary(mlm)
> MA = Manova(mlm,test='Wilks')
> SA = summary(MA)

Y = as.matrix(dat[,1:3]);
> Y = as.matrix(dat,[1:3]);
> X = as.matrix(dat[,4:6]);
> n = row(Y); p = ncol(Y); m = ncol(X)
> R = cor(dat); R2 = R^2
> smlm.Y = summary(lm(Y~X))
> smlm.X = summary(lm(X~Y))
> cc = cca(X,Y,xscale=T,yscale=T)
> Ryv = cc$ycrosscorr
> Rxu = cc$xcrosscorr
```