## 3/4/2010 36-402/608 ADA-II H. Seltman Breakout #15 Results

Summary: To perform a canonical correlation analysis in R, use cancor(X, Y) which produces a list with components "cor" (correlations of the canonical variable pairs), "xcoef" and "ycoef" (coefficients transforming data to canonical variables), and "xcenter" and "ycenter" (original X and Y column means). There is no special print() or other methods, but running cancor() does a special printout. The built in function does not compute p-values; use p.asym() (or p.perm() in package "CCP" to get the p-values.

The "University Data Set" at the UCI Machine Learning Repository has records on college and university characteristics, presumably collected in the 1980s. We will look at data on 242 institutions, considering several objective measures of who is admitted as the explanatory variables. Our outcomes are three (subjective) quality measures. We will use CCA to find dimensions of university characteristics that predict (correlate with) the measures of quality.

| names(univ)[xvar]                     |                      |                         |                         |  |  |
|---------------------------------------|----------------------|-------------------------|-------------------------|--|--|
| <pre># [1] "maleFemaleRatio"</pre>    | "sat.verbal"         | "sat.math"              | "percent.financial.aid" |  |  |
| <pre># [5] "percent.admittance"</pre> | "percent.enrolled"   |                         |                         |  |  |
| names(univ)[yvar]                     |                      |                         |                         |  |  |
| <pre># [1] "academics.scale"</pre>    | "social.scale"       | "quality.of.life.scale" |                         |  |  |
|                                       |                      |                         |                         |  |  |
| <pre>round(cor(univ[,xvar],univ</pre> | [yvar]),2)           |                         |                         |  |  |
| # aca                                 | ademics.scale social | l.scale qual            | lity.of.life.scale      |  |  |
| <pre># maleFemaleRatio</pre>          | -0.04                | -0.01                   | 0.11                    |  |  |
| # sat.verbal                          | -0.03                | -0.04                   | 0.16                    |  |  |
| # sat.math                            | 0.02                 | 0.03                    | 0.08                    |  |  |
| <pre># percent.financial.aid</pre>    | -0.12                | -0.13                   | -0.14                   |  |  |
| <pre># percent.admittance</pre>       | 0.11                 | -0.05                   | -0.11                   |  |  |
| <pre># percent.enrolled</pre>         | 0.07                 | 0.07                    | 0.06                    |  |  |

| University Data  |   |  |   |   |                                   |                      |                           |                                |  |  |
|--|---|--|---|---|-----------------------------------|----------------------|---------------------------|--------------------------------|--|--|
| 0  | 0                                       | 0                                      | 0   | 0                                       | 0                                 |                      | 0                         | <u>– 6      </u>               |  |  |
| 0 60000 <b>60000</b>   | <b></b>                                 | <b>\$\$\$\$\$</b> \$\$\$\$\$           | <b>60000</b>  | <b>6</b> 0 <b>670000</b>                | <b>@</b>                          | de 🗑 🔴 🕲 🖉           | e 🔊 🌑 🖉                   | 5 4 5 6                        |  |  |
|  |   | 2000                                   |   |   | 008 de 000                        |                      |                           | ality.of.life <sub>3</sub> sca |  |  |
| <b>60</b> 0 0  | <b>6</b>                                | <b>600 0</b>                           | <b>@B</b> o @ 0   | 68°00 0 0                               | <b>66</b> ℃ o                     | ଁ ଭ ଭ ଡ଼ି            | ଁ ଡିଡି                    | 123 2-                         |  |  |
|  | 0.0000                                  | 60000000000000000000000000000000000000 |   | 990 098<br>1990 098                     | 0                                 | <u>.</u>             | <u>ခြစ် ထ</u> ရ           |                                |  |  |
| 89   | මාර රු                                  |  | / <b>((()</b> ) <b>(((</b> )) <b>((</b> )) <b>(</b> ( <b>(</b> )) <b>((</b> ()) <b>(</b> ( <b>(</b> )) <b>((</b> ()) <b>(</b> ( <b>(</b> )) <b>((</b> ())) <b>((</b> ()) <b>((</b> ())) <b>((</b> |   |                                   |                      | - 4 3 4 9                 | 58 8                           |  |  |
| 0000   | Boge o                                  | 00 00 m                                |   |   |                                   | $\geq$               | -social.scale             |                                |  |  |
|  | 0000 00 0000                            |  |   | <b>39988</b> 8 9800                     | ആള് ആളം മം<br>പ                   |                      | 1 2 3 2 -                 | 122                            |  |  |
|  | ©©© _ 00 ~                              |  |   |   |                                   | a a a a              | <u>2</u> 5 -<br>      1 - |                                |  |  |
| 00 0   | 8.000                                   | <b>00</b> 00                           | ବନ ୦୦୦  | 8 00 8                                  | 000                               | -5111                | 0000                      | <u>58 0 8</u>                  |  |  |
| <b>େ ≎</b> ⊛େ  | ہ 8 8م                                  | 🐌 oo 🔍                                 | ®oo o°,   | <b>9680</b> 9 9                         | o ∕°o                             | - 4 3 4 5            | o 🔗 🗞 o 🍕                 | 🛢းစတီ 🏉 🔮 🦉                    |  |  |
| <b>C</b> © Co  | <b>639</b> 900 800                      | <b>68000000</b>                        | 0000 C  | <b>600 800 800</b>                      | <b>9939 99</b> 80                 | ademics.scal         | 899                       | 📕 🔍 🔍 📕                        |  |  |
| 00 00 80 00 CO   |   | <b>1899-199</b> 0 o                    |   | <b>68 099 (A</b>                        | <b>666 6</b> 00                   | 1232-                | § 9 Ø Ø Ø                 | 800                            |  |  |
| 0000 00  | 80%                                     | ളം ആശ്ര ര                              | <u>ം ക്കെ ഉള്ള</u><br>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~   | <u>6</u> 6,980 0                        | တ္က တြင်္                         |                      |                           | <mark>} 0_0_0_0</mark> 0       |  |  |
|  | 8 0 8                                   | ືສັດ້ວ                                 |   | ິດີເວີ                                  | - 15 10 15                        | ິຊີຊັ                | ິຊັ້                      | າຊີ້ຊີ ຊື່                     |  |  |
| 80,000   | 80 80 08<br>60 00 08                    |  | ക്കാരം റ  | ၀ ရ <b>၀</b> ၀ စီ /                     | ercent.enrollec                   | ക്ക്റ്റ്             |                           |                                |  |  |
|  |   |  |   |   | 5 5 -                             | õ 📲 🖉 õ 🦯            | ခ်ဳိးမ်ိဳး ၀ နီ           |                                |  |  |
| Barrie Barr | ൙ൟ                                      |  |   |   |                                   |                      |                           |                                |  |  |
|  | 8                                       | Q                                      | ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  | 20 1 1                                  |                                   | 00                   | 00                        |                                |  |  |
| ရွိ ၀၀ ၂၀  | ക്കം കം കു                              | ૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢ  | ଞ୍ଚେକ୍ଷ୍ଟ୍ର୍ରେନ୍ଦ୍ର   | - 15 10 15 2                            | ଞ୍ଚି <b>ଌ</b> ୍ଦିର୍ଦ୍ଦ <b>୦</b> ୨ | 6 6 6                | န္ခ်ီစစ                   | မီစွာဝ 💐 🧳 ငှ                  |  |  |
|  | Ša Sa °∕a                               |  |   | rcent.admittan                          | <u></u>                           | ଦ୍ ହ ହ ହ ର ୁ         | õ 🙎 🖕                     | 🖗 🎍 🕵 /                        |  |  |
|  |   |  |   | 5 10 5 —                                |                                   |                      |                           |                                |  |  |
| <u>6000000000000000000000000000000000000</u>   |   | Contraction of the second              | <b>Company</b>  | <u> </u>                                | 0 0 00 0                          | <b></b> .            | <b></b>                   |                                |  |  |
| 80000  | e 8 °                                   |  | - 20      <br>- 15 <sup>10</sup> <sup>15</sup> <sup>20</sup>  | ၜၟၜႍၜၴၴႍႍႍၜၜ                            |                                   | ິຊີຊີຊີວິ            | ° % °                     |                                |  |  |
|  | ດຢູ່ດ <b>ຍ</b> ັດດິດ<br>ຊີວິຊີ ຫວິດ     |  | cent financial.   |   |                                   |                      |                           |                                |  |  |
|  |   |  | 0 5 10 5 <u>–</u>   |   |                                   |                      |                           | છે. જે 🖉 🖉                     |  |  |
| A CAR S  |   | 8 9999 8                               |   |   |                                   | 8                    |                           |                                |  |  |
|  | 8 o ø                                   |  | <b>6</b> 00000  | 8 <sup>0</sup> 0 <sup>8</sup> 0         | °° °                              | 8800                 | 00                        | 8 8 0                          |  |  |
|  | ିଛି 🖋 😚                                 | 50 30 40 50 60<br>40                   |   |   |                                   | ိန္က်န္တိုင္စစ       |                           | <b>8</b> ° 8 <b>2</b>          |  |  |
| 500 O  | ၜႍႃၜၟၟၟၜၟၜၜ                             | - sat.math -                           | 88°°  |   | ၜၘၜႍႍႝၓၴၜႍၜႍႍႜ                    |                      |                           | 88                             |  |  |
|  | <b>్</b> రంగ                            | 0 10 20 30 20 -<br>10 -                |   |   | യും തുര                           |                      |                           |                                |  |  |
|  |   |  |   | 00 00 00 00 00 00 00 00 00 00 00 00 00  | <b>00 ~ 00</b>                    |                      | 87 <b>60 69 69 (</b>      |                                |  |  |
|  | 50<br>40 30 40 50 6                     | ୖୄୖୖ                                   | പ്പോയം  | <u>ૢૢૢૢૢૢૢૢૢૢૢૢૢ</u> ૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢ | <b>2</b> 00000                    |                      | žð 4                      |                                |  |  |
|  | — 40<br>— sæt.verløål —                 |  | စိုစိုစိုစို  |   | ക്രെപ്                            | 80.00                | മ്ത്ര ഉ                   |                                |  |  |
|  | 0 10 20 30 10 -                         |  |   | ൙ൣൕൕ                                    |                                   | ရှိ 🖉 🖉 ၀ 🧃          | ୁଇଛିଡି୦                   | နီးစစ္တို 🦓                    |  |  |
|  |   | <b>8</b> 88° 00                        | 80000   | 86088                                   | စ်နိုင်ဆီထွ                       |                      |                           |                                |  |  |
| - 30   | 80000                                   | 0 <sup>8</sup> 0000                    | 8000  | n <sup>8</sup> ° ° °                    | <b>8</b> 000                      | 880                  | 00                        | <u> </u>                       |  |  |
| 20 30  | 80 A - 0                                | ଁୄୢୖ୶ୖ                                 | ိုုလိုုလိုု   |   | 8 8 8 %                           | × ě o                | 8 9                       | <u>۲</u> ۲۲۲۲۲۲۲۲۲۲۲۲          |  |  |
| aleFemaleRat   | 8 8 8 8 9 9 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | Mar Son                                | ള്ക്ക്റ്റിക്ക്  | <b>ૢૢૢૢૢૢૢૢૢૢૢૢૢૢૢ</b><br>૱             |                                   | <u>ି</u> ଭୂଁ କୁ କୁ ଁ | 88.8                      |                                |  |  |
| 0 10 10 -  |   |  |   |   |                                   |                      |                           |                                |  |  |
|  | 60000                                   |  | o espande   | 00000000000000000000000000000000000000  |                                   | p 💕 🐻 🚳 🦉            | 👸 🐻 🔞 🍆 👸                 | 📲 🙆 🖉 🖉 🖉                      |  |  |

University Data

Question 1: What do you observe on the original scales?

The correlations are all fairly weak (and remember these are "r" values, so the highest  $r^2$  here is  $0.16^2 = 0.026$ . Academics is slightly positively correlated with percent admittance and negatively with percent receiving aid, while QOL is positively associated with sat verbal and negatively with percent aid and percent admittance. Social life is slightly negatively correlated with % financial aid. We hope higher correlation will come from CCA.

```
cc=cancor(univ[,xvar],univ[,yvar])
```

```
print(round(cc$cor,3))
# 0.271 0.257 0.111
print(round(cc$xcoef,4))
#
                            [,1]
                                    [,2]
                                            [,3]
                                                     [,4]
                                                             [,5]
                                                                     [,6]
# maleFemaleRatio
                         0.0022
                                 0.0031 -0.0001 -0.0021 -0.0098
                                                                  0.0005
# sat.verbal
                                 0.0043 -0.0043 0.0022 0.0022 -0.0002
                         0.0010
```

```
# sat.math
                         0.0007 -0.0017
                                         0.0026 -0.0057
                                                         0.0011
                                                                 0.0014
                                        0.0034 0.0002
# percent.financial.aid -0.0128 0.0070
                                                         0.0005
                                                                 0.0069
# percent.admittance
                        -0.0034 -0.0107 -0.0138 -0.0025 -0.0036
                                                                 0.0003
# percent.enrolled
                         0.0096 -0.0062 0.0063 0.0104 -0.0009
                                                                 0.0134
print(round(cc$ycoef,3))
                        [,1]
                               [,2]
                                      [,3]
#
                       0.025 -0.059 -0.047
#academics.scale
#social.scale
                       0.021 -0.020 0.041
#quality.of.life.scale 0.046 0.035 -0.015
require(CCP)
p.asym(cc$cor, nrow(univ), length(xvar), length(yvar))
#Wilks' Lambda, using F-approximation (Rao's F):
                                       df2
#
               stat
                       approx df1
                                             p.value
          0.8551364 1.2977101
                               18 410.6072 0.1846153
#1 to 3:
#2 to 3:
          0.9226508 1.1993408
                               10 292.0000 0.2908825
#3 to 3:
         0.9877572 0.4554981
                                4 147.0000 0.7682613
```

## Question 2: What tells you that we haven't found any interesting new scales? If you pretend that the first p-value is 0.00185, what conclusions would you reach?

Even though we have found some higher correlation values, the "1 to 3" p-value retains the null hypothesis of no correlations at all in the data. If the first CVs were correlated we would explain it as a smaller than average financial aid percent coupled with a larger than average percent enrolled is associated with a high mean rating (with extra emphasis on QOL).

The Greek air pollution example show how CCA can pull out meaningful variables such as the set of pollutants associated with two types of combustion, the separate ozone pollution and the weather patterns that are correlated with these.