STAT 441: Lecture 23
Canonical correlations

Find directions in the first and second sample, respectively, that exhibit maximal correlation

And then second maximal, third maximal...

Summarize correlations between two samples

Sample correlation coefficient

Recall:

$$\rho_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

Properties?

The prescription

Consider two data matrices:

X composed of lines $x_i^\top$, and

Y composed of lines $y_i^\top$, $i = 1, \ldots, n$

(the numbers p and q of variables, their columns, may be different, but the number of rows, the number of datapoints, is the same).

We seek (nonzero) a_1 and b_1 such that the (sample) correlation coefficient of Xa_1 and Yb_1 is maximal. Once found, we may continue: seek a_2 and b_2 again maximizing correlation of Xa_2 and Yb_2, but now such that a_2 is orthogonal to a_1 and b_2 to b_1.

Continuing this, we may seek nonzero a_j, orthogonal to all previous a_i and nonzero b_j orthogonal to all previous i_i such that the correlation of Xa_j and Yb_j is maximal. Of course, this is possible only if $j \leqslant p$ and $j \leqslant q$; hence we can repeat the above only min$\{p, q\}$ times.

Remarks

Note that the correlation coefficient change only by the sign of c if either a_i or b_i is replaced by replaced by $c a_i$ or $c b_i$, respectively.

Therefore, we do not have to worry about negative correlations - just take the opposite a_i or b_i and they become positive. Then, when we are multiplying either a_i or b_i by positive constants, the correlations remain the same. Therefore, the only side condition on a_i and b_i is that they are nonzero - from the mathematical point of view, everything else is well-posed.

From the numerical point of view, however, it may be practical to scale a_i and b_i in some convenient way; *it is usually done that the variances of resulting linear combinations of the data* - canonical variates - *are* 1.

Canonical variates come in pairs, and given what was mentioned above, there is $\min\{p, q\}$ of these pairs. In a special case when p or q is equal to one, there is only one pair; the maximized correlation coefficient is called *multiple correlation coefficient*.

Solution: variance matrices again

Let

$$\begin{pmatrix} S_{XX} & S_{XY} \\ S_{YX} & S_{YY} \end{pmatrix}$$

be the variance-covariance matrix of the data matrix $\begin{pmatrix} X & Y \end{pmatrix}$.

Maximal correlation between Xa and Yb,

maximized over $a \neq 0$ and $b \neq 0$,

is $\sqrt{\lambda}$, where λ is the largest eigenvalue of both

$$\begin{matrix} S_{XX}^{-1} S_{XY} S_{YY}^{-1} S_{YX} \\ S_{YY}^{-1} S_{YX} S_{XX}^{-1} S_{XY} \end{matrix} \text{ and } \begin{matrix} a \\ b \end{matrix} \text{ are the corresponding eigenvectors.}$$

Example: sons

Length and breadth, respectively, of the head of the first and second son

	L1	B1	L2	B2
1	191	155	179	145
2	195	149	201	152
3	181	148	185	149
4	183	153	188	149
5	176	144	171	142
6	208	157	192	152
7	189	150	190	149
8	197	159	189	152
9	188	152	197	159
10	192	150	187	151
11	179	158	186	148
12	183	147	174	147
13	174	150	185	152
14	190	159	195	157
15	188	151	187	158
16	163	137	161	130
17	195	155	183	158
18	186	153	173	148
19	181	145	182	146
20	175	140	165	137
21	192	154	185	152
22	174	143	178	147
23	176	139	176	143
24	197	167	200	158
25	190	163	187	150

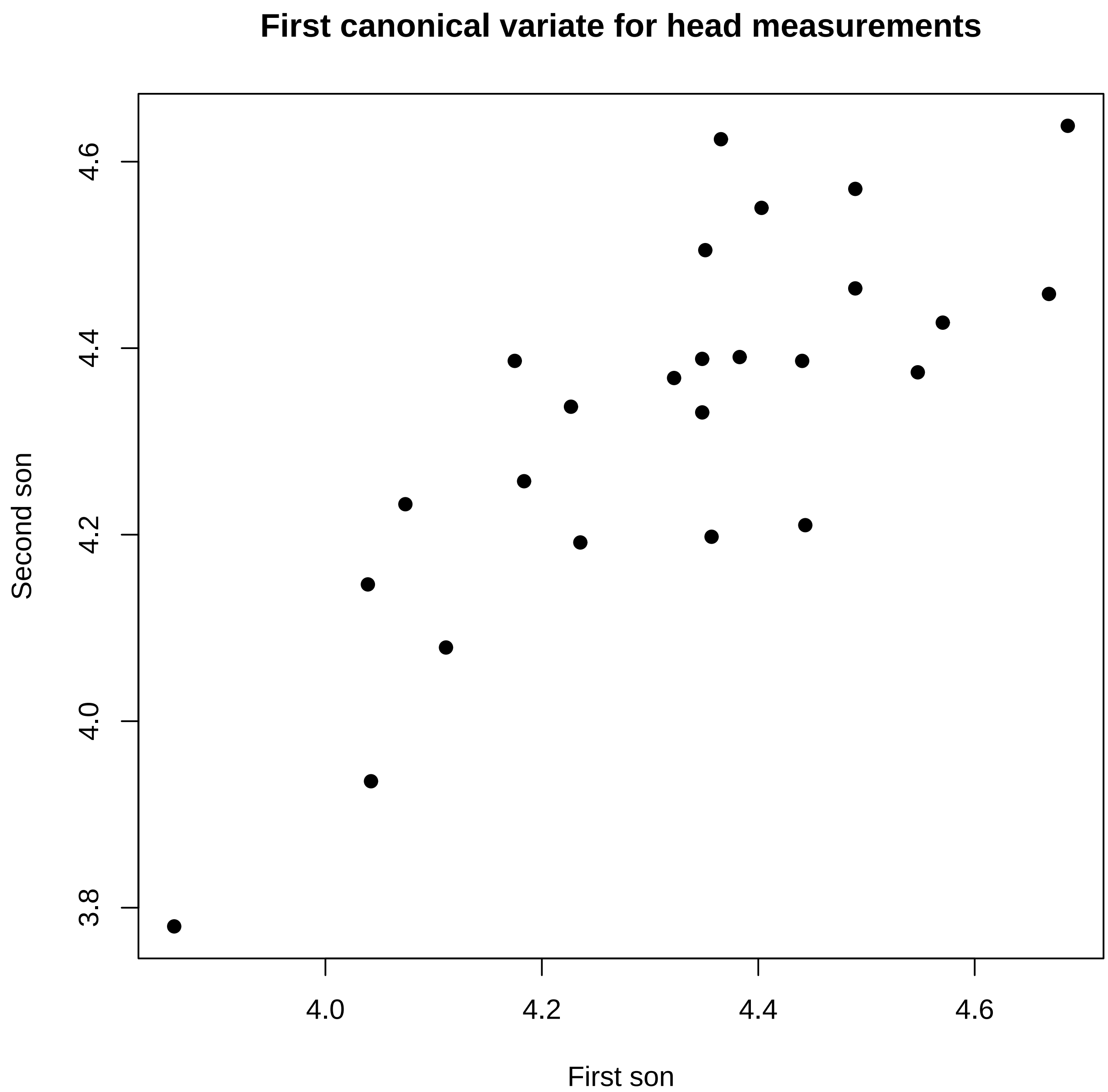
The result

```
> sons <- read.table("sons.d")
> cancor(sons[,1:2],sons[,3:4])
$cor
[1] 0.7885079 0.0537397
$xcoef
         [,1]        [,2]
L1 0.01154653 -0.02857148
B1 0.01443910  0.03816093
$ycoef
         [,1]        [,2]
L2 0.01025573 -0.03595605
B2 0.01637533  0.05349758
$xcenter
    L1     B1
185.72 151.12
$ycenter
    L2     B2
183.84 149.24
```

Canonical variates?

```
> sons.cc <- cancor(sons[,1:2],sons[,3:4])
> canvarx <- as.matrix(sons[,1:2]) %*% sons.cc$xcoef[,1]
> canvary <- as.matrix(sons[,3:4]) %*% sons.cc$ycoef[,1]
> plot(canvarx, canvary, pch=16,
+ xlab='First son',ylab='Second son')
```

And the plot

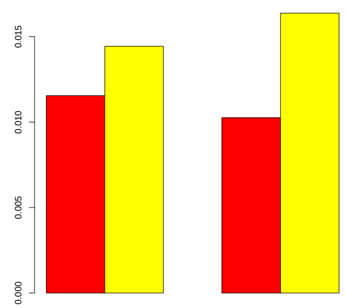

First canonical variate for head measurements
Second son
First son
3.8
4.0
4.2
4.4
4.6
4.0
4.2
4.4
4.6

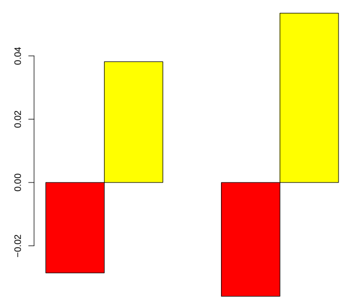
What else?

```
> var(sons)
         L1       B1        L2       B2
L1 95.29333 52.86833  69.66167 46.11167
B1 52.86833 54.36000  51.31167 35.05333
L2 69.66167 51.31167 100.80667 56.54000
B2 46.11167 35.05333  56.54000 45.02333

$xcoef
         [,1]        [,2]
L1 0.01154653 -0.02857148
B1 0.01443910  0.03816093

$ycoef
         [,1]        [,2]
L2 0.01025573 -0.03595605
B2 0.01637533  0.05349758
```





Appendix: some useful math

Eigenvalue and eigenvector: there is $a \neq 0$ such that $Qa = \lambda a$.

QS and SQ (if can be multiplied) have the same *nonzero* eigenvalues.

If Q is symmetric and positive semidefinite, then

$$\max_{a \neq 0} \frac{a^\top Q a}{a^\top a} = \max_{\|a\|=1} a^\top Q a = \lambda$$

where λ is the largest eigenvalue of Q and the maximum is attained at its corresponding eigenvector a.

This was used for principal components; for Fisher discriminants and canonical correlations, we have similar tricks.

Appendix: more useful math

If Q is as above and S is positive definite, then

$$\max_{a \neq 0} \frac{a^\top Q a}{a^\top S a} = \lambda$$

where λ is the largest eigenvalue of $S^{-1}Q$ and the maximum is attained at the corresponding eigenvector a of the latter matrix.

In particular, if S is positive definite, then $\max_{a \neq 0} \frac{(a^\top z)^2}{a^\top S a} = zS^{-1}z$ and is attained when a is proportional to $S^{-1}z$.

If S, T are positive definite and Q as above, then

$$\max_{\substack{a \neq \mathbf{0} \\ b \neq \mathbf{0}}} \frac{(a^\top Q b)^2}{(a^\top S a)(b^\top T b)} = \lambda$$

where λ is the largest eigenvalue of both $S^{-1}QT^{-1}Q^\top$ and $T^{-1}Q^\top S^{-1}Q$ and a, b are their corresponding eigenvectors, respectively.