ORIGIN ≡ 1

Multivariate Data: Linear Combinations

The concept of linear combination underlies much of linear algebra and multivariate statistics. For more complete information, see http://en.wikipedia.org/wiki/Linear_combination, RA Johnson & DW Wichern *Applied Multivariate Statistical Analysis 4th Edition* 1998, or AC. Rencher *Methods of Multivariate Analysis* 1995.

Read Data:

M := READPRN("C:\DATA\Multivariate\iris.txt")

M := submatrix(M, 1, 150, 2, 5)

M =

	1	2	3	4
1	5.1	3.5	1.4	0.2
2	4.9	3	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1
11	5.4	3.7	1.5	0.2
12	4.8	3.4	1.6	0.2
13	4.8	3	1.4	0.1
14	4.3	3	1.1	0.1
15	5.8	4	1.2	0.2
16	5.7	4.4	1.5	0.4
17	5.4	3.9	1.3	0.4
18	5.1	3.5	1.4	0.3
19	5.7	3.8	1.7	0.3
20	5.1	3.8	1.5	0.3
21	5.4	3.4	1.7	0.2
22	5.1	3.7	1.5	0.4
23	4.6	3.6	1	0.2
24	5.1	3.3	1.7	0.5

Prototype in R:

Anderson's Iris dataset - 50 cases for each of 3 species:

```
#LINEAR COMBINATIONS OF VARIABLES
#USE BUILT-IN iris DATA:
iris
M=subset(iris,select=c("Sepal.Length","Sepal.Width",
  "Petal.Length","Petal.Width"))
M

> M
    Sepal.Length Sepal.Width Petal.Length Petal.Width
1            5.1         3.5          1.4         0.2
2            4.9         3.0          1.4         0.2
3            4.7         3.2          1.3         0.2
4            4.6         3.1          1.5         0.2
5            5.0         3.6          1.4         0.2
6            5.4         3.9          1.7         0.4
7            4.6         3.4          1.4         0.3
8            5.0         3.4          1.5         0.2
9            4.4         2.9          1.4         0.2
10           4.9         3.1          1.5         0.1
…
51           7.0         3.2          4.7         1.4
52           6.4         3.2          4.5         1.5
53           6.9         3.1          4.9         1.5
54           5.5         2.3          4.0         1.3
55           6.5         2.8          4.6         1.5
56           5.7         2.8          4.5         1.3
57           6.3         3.3          4.7         1.6
58           4.9         2.4          3.3         1.0
59           6.6         2.9          4.6         1.3
60           5.2         2.7          3.9         1.4
…
101          6.3         3.3          6.0         2.5
102          5.8         2.7          5.1         1.9
103          7.1         3.0          5.9         2.1
104          6.3         2.9          5.6         1.8
105          6.5         3.0          5.8         2.2
106          7.6         3.0          6.6         2.1
107          4.9         2.5          4.5         1.7
108          7.3         2.9          6.3         1.8
109          6.7         2.5          5.8         1.8
110          7.2         3.6          6.1         2.5
…
```

$X := M^T$ < **transpose of M taken to conform with usual definitions below**

Linear Combinations of Variables:

Linear combinations of objects (rows) are single numbers (scalars) resulting from multiplying each element within an object vector by an associated constant, called *linear coeffients*, and summing the result. Linear coefficients may be any arbitrary scalar number, but in ueful statistical work are often calculated to maximize or minimize some statistical property. Using matrix algebra, this is accomplished by multiplying the row vector of linear coefficients (i.e., transpose) by the column vector for each object.

Linear Coefficients:

$$a := \begin{pmatrix} 2 \\ 3 \\ -1 \\ 5 \end{pmatrix} \qquad b := \begin{pmatrix} 0 \\ 2 \\ 1 \\ 4 \end{pmatrix}$$

<- Vectors containing arbitrary constants

```
> #LINEAR COEFFICIENTS:
> a=c(2,3,-1,5)
> a
[1]  2  3 -1  5
> b=c(0,2,1,4)
> b
[1] 0 2 1 4
```

$$a^T = (2 \quad 3 \quad -1 \quad 5)$$
$$b^T = (0 \quad 2 \quad 1 \quad 4)$$

<- transpose

Linear Combinations:

$lc_a := a^T \cdot X$

$lc_a =$

	1	2	3	4	5	6	7	8	9	10
1	20.3	18.4	18.7	18	20.4	22.8	19.5	19.7	17.1	18.1

$lc_b := b^T \cdot X$

$lc_b =$

	1	2	3	4	5	6	7	8	9	10
1	9.2	8.2	8.5	8.5	9.4	11.1	9.4	9.1	8	8.1

```
#LINEAR COMBINATIONS:
lca=t(a)%*%t(M)
lca
lcb=t(b)%*%t(M)
lcb
```

Linear combinations are single numbers, one for each case, resulting from multiplying each variable within each case by the corresponding linear coefficient in vectors a or b and summing the result. For the iris dataset, there are 150 linear combinations.

Statistics of Original Variables:

Original Variables:

```
#STATISTICS OF ORIGINAL VARIABLES:
Xbar=mean(M)

> Xbar
Sepal.Length  Sepal.Width Petal.Length  Petal.Width 
    5.843333     3.057333     3.758000     1.199333 
> S=cov(M)
```

$n := \text{rows}(M) \qquad n = 150$

$p := \text{cols}(M) \qquad p = 4$

$i := 1..n \qquad j := 1..p$

$l_i := 1$

$$X_{bar} := \frac{1}{n} \cdot X \cdot l \qquad X_{bar} = \begin{pmatrix} 5.843333 \\ 3.057333 \\ 3.758 \\ 1.199333 \end{pmatrix}$$

$I := \text{identity}(n)$

$$S := \frac{1}{n-1} \cdot X \cdot \left(I - \frac{1}{n} \cdot l \cdot l^T \right) \cdot X^T \qquad S = \begin{pmatrix} 0.6856935 & -0.042434 & 1.2743154 & 0.5162707 \\ -0.042434 & 0.1899794 & -0.3296564 & -0.1216394 \\ 1.2743154 & -0.3296564 & 3.1162779 & 1.2956094 \\ 0.5162707 & -0.1216394 & 1.2956094 & 0.5810063 \end{pmatrix}$$

```
> S
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    0.6856935  -0.0424340    1.2743154   0.5162707
Sepal.Width    -0.0424340   0.1899794   -0.3296564  -0.1216394
Petal.Length    1.2743154  -0.3296564    3.1162779   1.2956094
Petal.Width     0.5162707  -0.1216394    1.2956094   0.5810063
```

Statistics of Linear Combinations:

Mean:

$lca_{bar} := mean\left(lc_a^T\right)$ $\qquad lca_{bar} = 23.0973$

$lcb_{bar} := mean\left(lc_b^T\right)$ $\qquad lcb_{bar} = 14.67$

```
#STATISTICS OF LINEAR COMBINATIONS:
#MEAN
lcabar=mean(lca)
lcbbar=mean(lcb)
#VARIANCE
lcavar=var(t(lca))
lcbvar=var(t(lcb))
#COVARIANCE
lccov=cov(t(lca),t(lcb))
Result=c(lcabar,lcbbar,lcavar,lcbvar,lccov)

> Result
[1] 23.09733 14.67000 12.18563 20.27232 14.44267
```

Variance:

$lca_{var} := \frac{n}{n-1} \cdot var\left(\left(lc_a\right)^T\right)$ $\qquad lca_{var} = 12.18563$

$lcb_{var} := \frac{n}{n-1} \cdot var\left(\left(lc_b\right)^T\right)$ $\qquad lcb_{var} = 20.27232$

Covariance:

$$lc_{cov} := \frac{1}{n-1} \cdot \sum_i \left[\left(lc_a^T\right)_i - lca_{bar}\right] \cdot \left[\left(lc_b^T\right)_i - lcb_{bar}\right]$$

$lc_{cov} = 14.44267$

Identities Demonstrated:

Mean of Linear Combinations:

$lca_{bar} = 23.097$ $\qquad a^T \cdot X_{bar} = (23.09733)$

Variance of Linear Combinations:

$lca_{var} = 12.186$ $\qquad a^T \cdot S \cdot a = (12.18563)$

Covariance of Linear Combinations:

$lc_{cov} = 14.443$ $\qquad a^T \cdot S \cdot b = (14.44267)$

```
#IDENTITIES DEMONSTRATED:
> t(a)%*%Xbar
         [,1]
[1,] 23.09733
> t(a)%*%S%*%a
         [,1]
[1,] 12.18563
> t(a)%*%S%*%b
         [,1]
[1,] 14.44267
```

^ Matrix algebra formula for calculation directly

Multiple Linear Combinations:

A := READPRN("c:/DATA/Multivariate/LC Matrix A.txt")

$$A = \begin{pmatrix} 0.599 & 1.47 & 1.7172 & 0.6062 \\ 0.4252 & 1.0342 & 2.2546 & 0.676 \\ 0.4919 & 1.3995 & 0.4426 & 0.5664 \\ 0.6929 & 0.8531 & 2.8998 & 0.613 \\ 0.8217 & 0.3827 & 2.4516 & 0.6223 \\ 0.732 & 0.5592 & 2.0467 & 2.8876 \\ 0.123 & 1.6693 & 1.5511 & 1.7048 \end{pmatrix}$$

```
#MULTIPLE LINEAR COMBINATIONS:
> A=read.table("c:/DATA/Multivariate/LC Matrix
  A.txt",header=F)
>A=as.matrix(A)
> A
         V1     V2     V3     V4
[1,] 0.5990 1.4700 1.7172 0.6062
[2,] 0.4252 1.0342 2.2546 0.6760
[3,] 0.4919 1.3995 0.4426 0.5664
[4,] 0.6929 0.8531 2.8998 0.6130
[5,] 0.8217 0.3827 2.4516 0.6223
[6,] 0.7320 0.5592 2.0467 2.8876
[7,] 0.1230 1.6693 1.5511 1.7048
```

Let A be a matrix of linear combination coefficients where each row of A represents coefficients for a single linear transformation.

$LC := A \cdot X$

Matrix LC represents the set of linear combinations (7 in this example for 150 cases).

LC =

	58	59	60	61	62	63	64	65	66	67
1	12.736	16.904	14.63	12.551	16.066	14.303	16.836	14.587	16.975	16.401
2	12.682	17.055	14.743	12.761	16.095	14.521	17.136	14.376	16.922	16.643
3	7.796	10.077	8.856	7.374	9.809	8.367	9.932	9.143	10.375	9.794
4	15.625	21.183	18.074	15.933	19.746	18.246	21.188	17.59	20.904	20.408
5	13.657	18.619	15.739	14.077	17.226	16.201	18.516	15.346	18.35	17.715
6	14.571	19.622	17.341	14.829	18.924	16.697	19.749	16.843	19.686	19.318
7	11.432	15.004	13.583	11.087	14.805	12.32	15.268	13.33	15.21	15.234

Identities Demonstrated for Multiple Linear Combinations:

Mean:

$$LC_{bar} := \frac{1}{n} \cdot LC \cdot 1$$

$$LC_{bar} = \begin{pmatrix} 15.175 \\ 14.93 \\ 9.496 \\ 18.29 \\ 15.931 \\ 17.142 \\ 13.696 \end{pmatrix} \qquad A \cdot X_{bar} = \begin{pmatrix} 15.175 \\ 14.93 \\ 9.496 \\ 18.29 \\ 15.931 \\ 17.142 \\ 13.696 \end{pmatrix}$$

Covariance Matrix:

$$LC_{cov} := \frac{1}{n-1} \cdot LC \cdot \left(I - \frac{1}{n} \cdot 1 \cdot 1^T \right) \cdot LC^T$$

$$LC_{cov} = \begin{pmatrix} 13.797 & 17.107 & 5.234 & 21.908 & 19.607 & 22.818 & 14.273 \\ 17.107 & 21.378 & 6.318 & 27.437 & 24.589 & 28.601 & 17.724 \\ 5.234 & 6.318 & 2.167 & 8.024 & 7.144 & 8.359 & 5.406 \\ 21.908 & 27.437 & 8.024 & 35.247 & 31.608 & 36.719 & 22.676 \\ 19.607 & 24.589 & 7.144 & 31.608 & 28.358 & 32.928 & 20.283 \\ 22.818 & 28.601 & 8.359 & 36.719 & 32.928 & 38.458 & 23.742 \\ 14.273 & 17.724 & 5.406 & 22.676 & 20.283 & 23.742 & 14.864 \end{pmatrix}$$

$$A \cdot S \cdot A^T = \begin{pmatrix} 13.797 & 17.107 & 5.234 & 21.908 & 19.607 & 22.818 & 14.273 \\ 17.107 & 21.378 & 6.318 & 27.437 & 24.589 & 28.601 & 17.724 \\ 5.234 & 6.318 & 2.167 & 8.024 & 7.144 & 8.359 & 5.406 \\ 21.908 & 27.437 & 8.024 & 35.247 & 31.608 & 36.719 & 22.676 \\ 19.607 & 24.589 & 7.144 & 31.608 & 28.358 & 32.928 & 20.283 \\ 22.818 & 28.601 & 8.359 & 36.719 & 32.928 & 38.458 & 23.742 \\ 14.273 & 17.724 & 5.406 & 22.676 & 20.283 & 23.742 & 14.864 \end{pmatrix}$$

#IDENTITIES DEMONSTRATED FOR MULTIPLE LINEAR COMBINATIONS:
#MEAN:
LC=data.frame(t(LC))
LCbar=mean(LC)
> LCbar

```
       X1        X2        X3        X4        X5        X6        X7 
15.174710 14.930016  9.495667 18.289696 15.930966 17.141674 13.695994
```

> A%*%Xbar

```
          [,1]
[1,] 15.174710
[2,] 14.930016
[3,]  9.495667
[4,] 18.289696
[5,] 15.930966
[6,] 17.141674
[7,] 13.695994
```

> #COVARIANCE MATRIX:
> cov(LC)

```
          X1        X2       X3        X4        X5       X6       X7
X1 13.797287 17.106850 5.234495 21.908219 19.606607 22.81850 14.27279
X2 17.106850 21.378027 6.317534 27.437241 24.588570 28.60149 17.72399
X3  5.234495  6.317534 2.167349  8.023732  7.143541  8.35865  5.40558
X4 21.908219 27.437241 8.023732 35.247221 31.607558 36.71857 22.67564
X5 19.606607 24.588570 7.143541 31.607558 28.357875 32.92757 20.28327
X6 22.818500 28.601495 8.358650 36.718567 32.927570 38.45831 23.74193
X7 14.272789 17.723985 5.405580 22.675644 20.283268 23.74193 14.86373
```

> A%*%S%*%t(A)

```
          [,1]      [,2]     [,3]      [,4]      [,5]     [,6]     [,7]
[1,] 13.797287 17.106850 5.234495 21.908219 19.606607 22.81850 14.27279
[2,] 17.106850 21.378027 6.317534 27.437241 24.588570 28.60149 17.72399
[3,]  5.234495  6.317534 2.167349  8.023732  7.143541  8.35865  5.40558
[4,] 21.908219 27.437241 8.023732 35.247221 31.607558 36.71857 22.67564
[5,] 19.606607 24.588570 7.143541 31.607558 28.357875 32.92757 20.28327
[6,] 22.818500 28.601495 8.358650 36.718567 32.927570 38.45831 23.74193
[7,] 14.272789 17.723985 5.405580 22.675644 20.283268 23.74193 14.86373
```