Today: Some Ch 3 exercises
Analysis of a Multivariate Normal Data Set –Ch 4

Ex 3.14 This uses a small data set X = (9 1) (5 3) (1 2) 3 cases x 2 variables

$$X = \begin{pmatrix} 9 & 1 \\ 5 & 3 \\ 1 & 2 \end{pmatrix}$$

And, $X_1 =$ (9 5 1)'
$X_2 =$ (1 3 2)'

The linear combinations c'X' = $-X_1 + 2X_2$ and
b'X'= $2X_1 + 3X_2$ are defined in the exercise.

In a) we are to compute the means, variances and covariances of these new rvs.

The observed values of c'X' and b'X' are (-7 1 3) and (21 19 8) resp. and clearly the means are –1 and 16 respectively. Sample Variance of c'X' is [(36+4+16)/2] = 28 Sample Variance of b'X' is [(25+9+64)/2] = 49 and Cov (b'X', c'X') = [(-30+6-32)/2] = -28.

This is obvious using the given linear combinations – but for a matrix way to do it …

$$(-1\ 2)\begin{pmatrix} 9 & 5 & 1 \\ 1 & 3 & 2 \end{pmatrix} = (-7\ \ 1\ \ 3) \quad \text{and}$$

$$(\ 2\ 3)\begin{pmatrix} 9 & 5 & 1 \\ 1 & 3 & 2 \end{pmatrix} = (21\ 19\ \ 8)$$

To get the sample mean for c'X' use

$$(1/n)c'X'\begin{pmatrix}1\\1\\1\end{pmatrix} = (1/3)(-7\ 1\ 3)\begin{pmatrix}1\\1\\1\end{pmatrix} = -1$$

Similarly the mean of b'X'=16

Now write

$$Y = \begin{pmatrix} c'X' \\ b'X' \end{pmatrix} = \begin{pmatrix} -7 & 1 & 3 \\ 21 & 19 & 8 \end{pmatrix}$$

We need Sample Cov(Y)

which is part b) Do the means and cov from the means and cov of X (and then using b' and c') b)

See (3-36) p 142. Cov(Y) = c'Sb

See (3-27) p 140. $S = (1/n-1)\, X'(I-(1/n)11')X = (1/2) \begin{pmatrix} 9 & 5 & 1 \\ 1 & 3 & 2 \end{pmatrix} \left(I - (1/3)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}(1\ 1\ 1)\right) \begin{pmatrix} 9 & 1 \\ 5 & 3 \\ 1 & 2 \end{pmatrix}$

$= \begin{pmatrix} 16 & -2 \\ -2 & 1 \end{pmatrix}$ which is S

Substituting in (3-36) $Cov\,(c'X', b'X') = (-1\ 2)\begin{pmatrix} 16 & -2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix} = (-20\ 4)\begin{pmatrix} 2 \\ 3 \end{pmatrix} = -40 + 12 = -28$

We could similarly show Var(c'X') = c'Sc = 28 and Var (b'X') = b'Sb=

3. 15 is just a repeat of 3.14 with n=p=3. means 12 and –1, vars 12 and 43, cov -3

3.16 Σ_V is the covariance of vector V, by def $= E(V-\mu_V)(V-\mu_V)'$

(Note V is a column vector of length p.)

$= E(VV') - \mu_V\mu_V'$ (by usual expansion)

so $E(VV') = \Sigma_V + \mu_V\mu_V'$

Next topic – Multivariate Normal – Ch 4

Data p 187 p=4 variables on a sample of lumber.

First: A look at Outliers

How do compute the d^2 ?

Order of computations

S p x p

S^{-1} p x p

(1/n) $X'\, 1_n$ which is $\overline{X}$ 1 x p

Subt $\overline{X}$ from X n x p

Transpose $(X-\overline{X})$ p x n

Mult $(X-\overline{X})'\, S^{-1}\, (X-\overline{X})$ n x n

Take diagonal. 1 x n

Outliers: univariate, multivariate.

Multivariate outliers are not necessarily univariate outliers. See row 16 of Vib*Static1.

Univariate Outliers are not necessarily multivariate outliers. See row 9 of Vib*Static1.

More about p-Normal:

Density p 150 - number of parameters

Contours - connection with statistical distance
– connection with eigenanalysis Result 4.1 p 153 and bottom of page

Bivariate Case p 151

Distribution of Lin Combination of univariate (possibly correlated) normals p 157

Chi-Square Distribution of Quadratic Forms based on Σ^{-1} p 163-4

Normal Q-Q Plots pp 178ff

Next time – rest of Ch 4 – and exercises.