Multivariate Analysis

Multivariate data

x_{jk} = measurement of the kth variable on the jth item

Consequently, n measurements on p variables can be displayed as follows:

	Variable 1	Variable 2	$\cdots$	Variable k	$\cdots$	Variable p
Item 1:	x_{11}	x_{12}	$\cdots$	x_{1k}	$\cdots$	x_{1p}
Item 2:	x_{21}	x_{22}	$\cdots$	x_{2k}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
Item j:	x_{j1}	x_{j2}	$\cdots$	x_{jk}	$\cdots$	x_{jp}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
Item n:	x_{n1}	x_{n2}	$\cdots$	x_{nk}	$\cdots$	x_{np}

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix}$$

Descriptive Statistics

Let $x_{11}, x_{21}, \ldots, x_{n1}$ be n measurements on the first variable. Then the arithmetic average of these measurements is

$$\bar{x}_1 = \frac{1}{n} \sum_{j=1}^{n} x_{j1}$$

In general, mean for the kth variable is given as,

$$\bar{x}_k = \frac{1}{n} \sum_{j=1}^{n} x_{jk} \qquad k = 1, 2, \ldots, p$$

In general, sample variance for the kth variable is given as,

$$s_k^2 = \frac{1}{n} \sum_{j=1}^{n} (x_{jk} - \bar{x}_k)^2 \qquad k = 1, 2, \ldots, p$$

The *sample covariance*

$$s_{ik} = \frac{1}{n} \sum_{j=1}^{n} (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k) \qquad i = 1, 2, \ldots, p, \quad k = 1, 2, \ldots, p$$

The sample correlation coefficient for the ith and kth variables is

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}} = \frac{\sum_{j=1}^{n} (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)}{\sqrt{\sum_{j=1}^{n} (x_{ji} - \bar{x}_i)^2}\sqrt{\sum_{j=1}^{n} (x_{jk} - \bar{x}_k)^2}}$$

Arrays of Basic Descriptive Statistics

Sample means $\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$

Sample variances and covariances $\mathbf{S}_n = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$

Sample correlations $\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}$

- Paper is manufactured in continuous sheets several feet wide. Because of the orientation of fibers within the paper, it has a different strength when measured in the direction produced by the machine than when measured across, or at right angles to, the machine direction. Table below shows the measured values of

x_1 = density (grams/cubic centimeter)

x_2 = strength (pounds) in the machine direction

x_3 = strength (pounds) in the cross direction

Paper-Quality Measurements

Specimen	Density	Strength	
		Machine direction	Cross direction
1	.801	121.41	70.42
2	.824	127.70	72.47
3	.841	129.20	78.20
4	.816	131.80	74.89
5	.840	135.10	71.21
6	.842	131.50	78.39
7	.820	126.70	69.02
8	.802	115.10	73.10
9	.828	130.80	79.28
10	.819	124.60	76.48
11	.826	118.31	70.25
12	.802	114.20	72.88
13	.810	120.30	68.23
14	.802	115.70	68.12
15	.832	117.51	71.62
16	.796	109.81	53.10
17	.759	109.10	50.85
18	.770	115.10	51.68
19	.759	118.31	50.60
20	.772	112.60	53.51
21	.806	116.20	56.53
22	.803	118.00	70.70
23	.845	131.00	74.35
24	.822	125.70	68.29
25	.971	126.10	72.10
26	.816	125.80	70.64
27	.836	125.50	76.33
28	.815	127.80	76.75
29	.822	130.50	80.33
30	.822	127.90	75.68
31	.843	123.90	78.54
32	.824	124.10	71.91
33	.788	120.80	68.22
34	.782	107.40	54.42
35	.795	120.70	70.41
36	.805	121.91	73.68
37	.836	122.31	74.93
38	.788	110.60	53.52
39	.772	103.51	48.93
40	.776	110.71	53.67
41	.758	113.80	52.42

- The data in Table below are 42 measurements on air-pollution variables recorded at 12:00 noon in the city area on different days.

  (a) Plot Scatter Plot Matrix with all the variables.

  (b) Construct the mean, covariance and correlation matrix and interpret.

Wind (x_1)	Solar radiation (x_2)	CO (x_3)	NO (x_4)	NO_2 (x_5)	O_3 (x_6)	HC (x_7)
8	98	7	2	12	8	2
7	107	4	3	9	5	3
7	103	4	3	5	6	3
10	88	5	2	8	15	4
6	91	4	2	8	10	3
8	90	5	2	12	12	4
9	84	7	4	12	15	5
5	72	6	4	21	14	4
7	82	5	1	11	11	3
8	64	5	2	13	9	4
6	71	5	4	10	3	3
6	91	4	2	12	7	3
7	72	7	4	18	10	3
10	70	4	2	11	7	3
10	72	4	1	8	10	3

9	77	4	1	9	10	3
8	76	4	1	7	7	3
8	71	5	3	16	4	4
9	67	4	2	13	2	3
9	69	3	3	9	5	3
10	62	5	3	14	4	4
9	88	4	2	7	6	3
8	80	4	2	13	11	4
5	30	3	3	5	2	3
6	83	5	1	10	23	4
8	84	3	2	7	6	3
6	78	4	2	11	11	3
8	79	2	1	7	10	3
6	62	4	3	9	8	3
10	37	3	1	7	2	3
8	71	4	1	10	7	3
7	52	4	1	12	8	4
5	48	6	5	8	4	3
6	75	4	1	10	24	3
10	35	4	1	6	9	2
8	85	4	1	9	10	2
5	86	3	1	6	12	2
5	86	7	2	13	18	2
7	79	7	4	9	25	3
7	79	5	2	8	6	2
6	68	6	2	11	14	3
8	40	4	3	6	5	2