STAT 524

HW1

Satoshi Ido
34788706

09/05/2023

1.1

1.1. Consider the seven pairs of measurements (x_1, x_2) plotted in Figure 1.1:

x_1	3	4	2	6	8	2	5
x_2	5	5.5	4	7	10	5	7.5

Calculate the sample means $\bar{x}_1$ and $\bar{x}_2$, the sample variances s_{11} and s_{22}, and the sample covariance s_{12}.

$$\bar{x}_1 = \frac{1}{7}(3+4+2+6+8+2+5) = \frac{30}{7} \cong \underline{4.286}$$

$$\bar{x}_2 = \frac{1}{7}(5+5.5+4+7+10+5+7.5) = \frac{44}{7} \cong \underline{6.286}$$

$$S_{11} = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2$$

$$= \frac{1}{6}\{(3-4.286)^2 + (4-4.286)^2 + (2-4.286)^2 + (6-4.286)^2$$
$$+ (8-4.286)^2 + (2-4.286)^2 + (5-4.286)^2\}$$

$$\cong \frac{29.4286}{6} \cong \underline{4.90}$$

$$S_{22} = \frac{1}{6}\{(5-6.286)^2 + (5.5-6.286)^2 + (4-6.286)^2$$
$$+ (7-6.286)^2 + (10-6.286)^2 + (5-6.286)^2$$
$$+ (7.5-6.286)^2\}$$

$$\cong \frac{24.9286}{6} \cong \underline{4.15}$$

$$S_{12} = \frac{1}{7-1}\sum_{i=1}^{7}(x_{i1} - \bar{x}_{.1})(x_{i2} - \bar{x}_{.2})$$

$$= \frac{1}{6}\{(3-4.286)(5-6.286) + (4-4.286)(5.5-6.286)$$
$$+ (2-4.286)(4-6.286) + (6-4.286)(7-6.286)$$
$$+ (8-4.286)(10-6.286) + (2-4.286)(5-6.286)$$
$$+ (5-4.286)(7.5-6.286)\}$$

$\approx \frac{25.929}{6}$ $= 4.321$

1.2.

1.2. A morning newspaper lists the following used-car prices for a foreign compact with age x_1 measured in years and selling price x_2 measured in thousands of dollars:

x_1	1	2	3	3	4	5	6	8	9	11
x_2	18.95	19.00	17.95	15.54	14.00	12.95	8.94	7.49	6.00	3.99

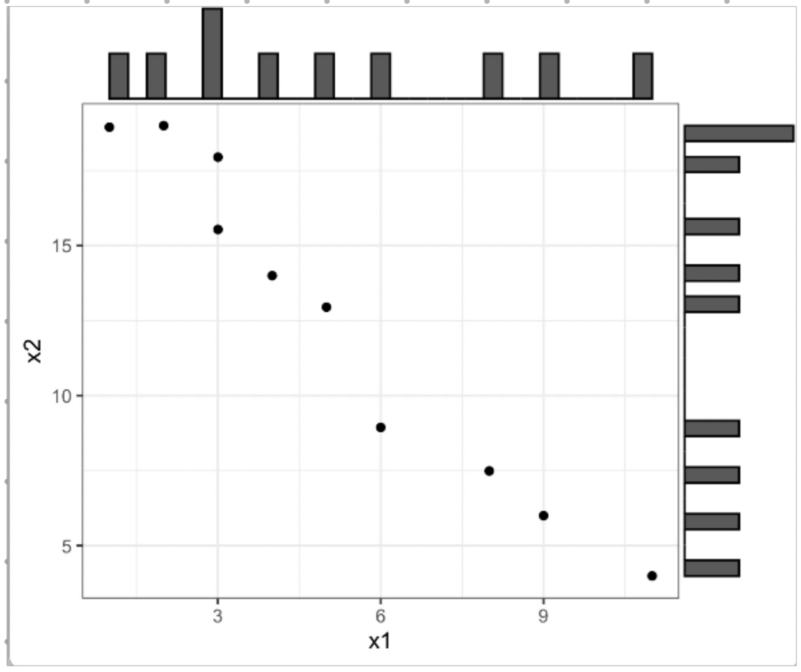
(a) Construct a scatter plot of the data and marginal dot diagrams.

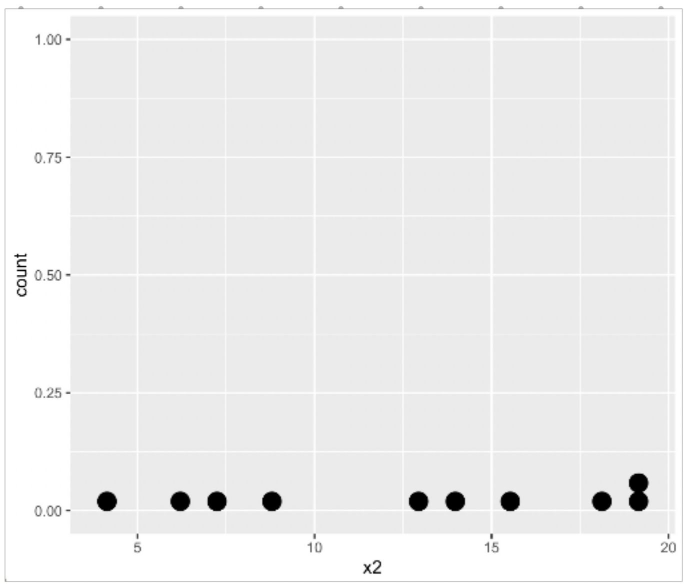
(b) Infer the sign of the sample covariance s_{12} from the scatter plot.

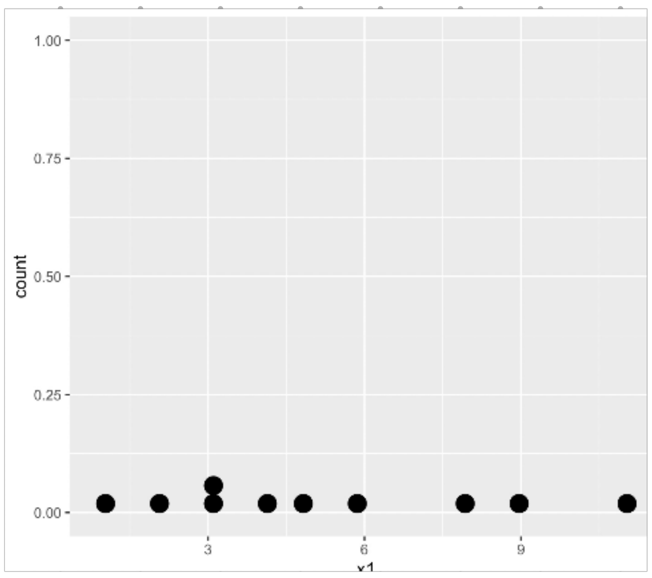
(c) Compute the sample means $\bar{x}_1$ and $\bar{x}_2$ and the sample variances s_{11} and s_{22}. Compute the sample covariance s_{12} and the sample correlation coefficient r_{12}. Interpret these quantities.

(d) Display the sample mean array $\bar{\mathbf{x}}$, the sample variance-covariance array $\mathbf{S}_n$, and the sample correlation array $\mathbf{R}$ using (1-8).

a)







(b) Since two data are negatively correlatted, sign should be negative (−).

(c) $\bar{x}_1 = \frac{1}{10}(1+2+3+3+4+5+6+8+9+11) \simeq \underline{5.2}$

$\bar{x}_2 = \frac{1}{10}(18.95 + 19.00 + 17.95 + 15.54 + 14.00 + 12.95 + 8.94 + 7.49 + 6.00 + 3.99) \simeq \underline{12.48}$

$$S_{11} = \frac{1}{9}\sum_{i=1}^{10}(x_{i1} - \bar{x}_{\cdot 1})^2$$

$$= \frac{1}{9}\{(1-5.2)^2 + (2-5.2)^2 + (3-5.2)^2 + (3-5.2)^2 + (4-5.2)^2 + (5-5.2)^2 + (6-5.2)^2 + (8-5.2)^2 + (9-5.2)^2 + (11-5.2)^2\}$$

$$= \underline{10.62}$$

$$S_{22} = \frac{1}{9}\sum_{i=1}^{10}(x_{i2} - \bar{x}_{\cdot 2})^2$$

$$= \frac{1}{9}\{(18.95-12.48)^2 + (19-12.48)^2 + (17.95-12.48)^2 + (15.54-12.48)^2 + (14-12.48)^2 + (12.95-12.48)^2 + (8.94-12.48)^2 + (7.49-12.48) + (6-12.48)^2 + (3.99-12.48)^2\}$$

$$= \frac{1}{9} \times 277.7 = \underline{30.85}$$

$$S_{12} = \frac{1}{9} \sum_{i=1}^{10} (x_{i1} - \bar{x}_{.1})(x_{i2} - \bar{x}_{.2})$$

$$= \frac{1}{9} \{ (-4.2)(6.47) + (-3.2)(6.5) + (-2.2)(5.47)$$
$$+ (-2.2)(3.06) + (-1.2)(1.5) + (-0.2)(0.47)$$
$$+ (0.8)(-3.54) + (2.8)(-4.99) + (3.8)(-6.48)$$
$$+ (5.8)(-8.49) \}$$

$$= \frac{1}{9}(-159.4)$$

$$= \underline{-17.67}$$

$$r_{12} = \frac{-17.67}{\sqrt{10.62}\sqrt{30.85}} = \underline{-0.976}$$

Since $S_{11} < S_{22}$, the x_2 data is scattered more.

r_{12} is negative, so x_1 and x_2 data correlate negatively.

r_{12} is close to -1, hence x_1 and x_2 data correlate strongly.

(d)

$$\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 5.2 \\ 12.48 \end{bmatrix}$$

$$S_n = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} 10.62 & -17.67 \\ -17.67 & 30.85 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix} = \begin{bmatrix} 1 & -0.976 \\ -0.976 & 1 \end{bmatrix}$$

1.6

1.6. The data in Table 1.5 are 42 measurements on air-pollution variables recorded at 12:00 noon in the Los Angeles area on different days. (See also the air-pollution data on the web at www.prenhall.com/statistics.)

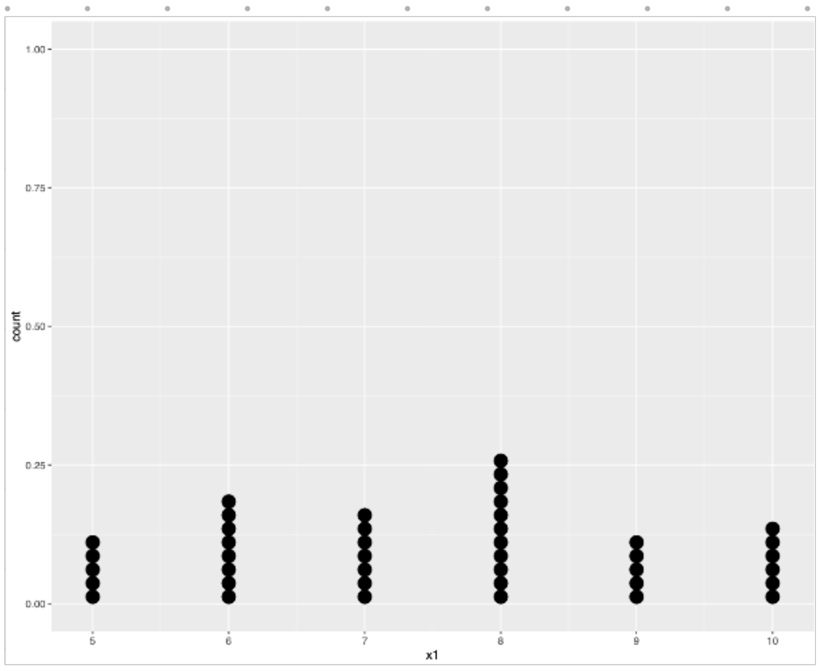
(a) Plot the marginal dot diagrams for all the variables.

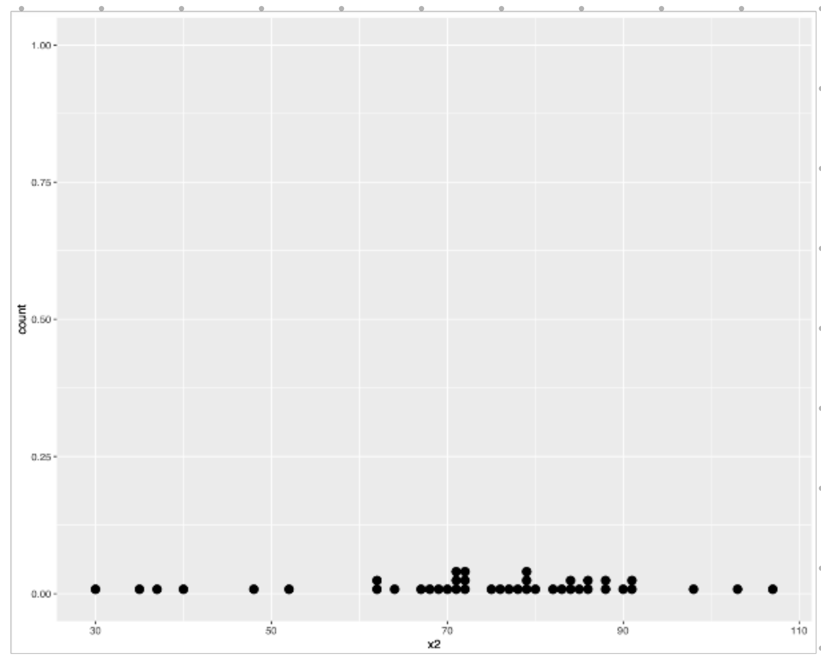
(b) Construct the $\bar{\mathbf{x}}$, $\mathbf{S}_n$, and $\mathbf{R}$ arrays, and interpret the entries in $\mathbf{R}$.

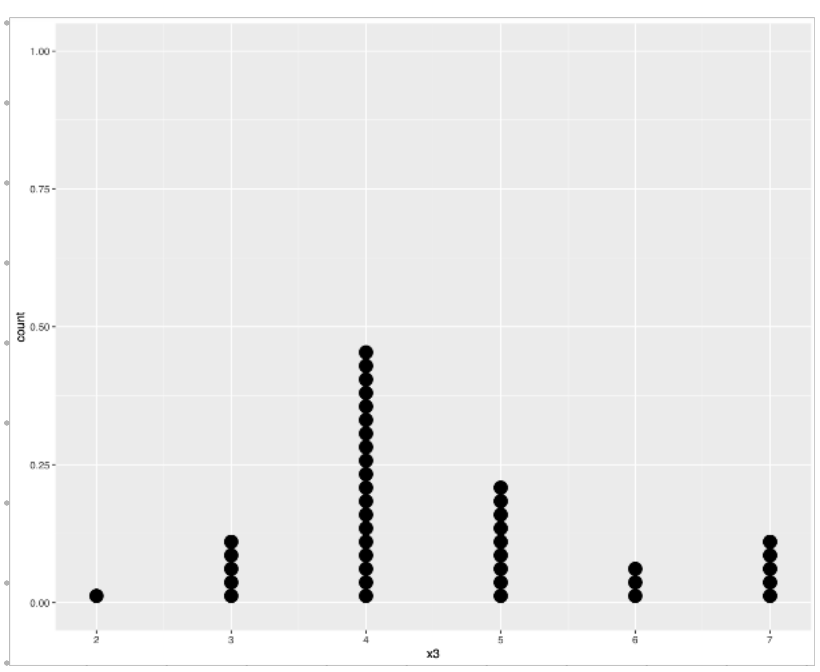
Table 1.5 Air-Pollution Data

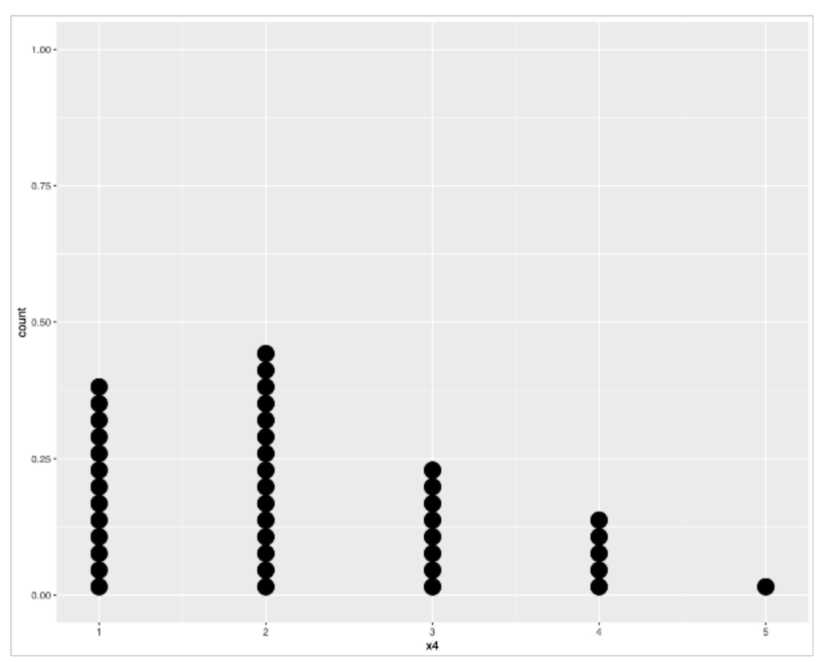
Wind (x_1)	Solar radiation (x_2)	CO (x_3)	NO (x_4)	NO_2 (x_5)	O_3 (x_6)	HC (x_7)
8	98	7	2	12	8	2
7	107	4	3	9	5	3
7	103	4	3	5	6	3
10	88	5	2	8	15	4
6	91	4	2	8	10	3
8	90	5	2	12	12	4
9	84	7	4	12	15	5
5	72	6	4	21	14	4
7	82	5	1	11	11	3
8	64	5	2	13	9	4

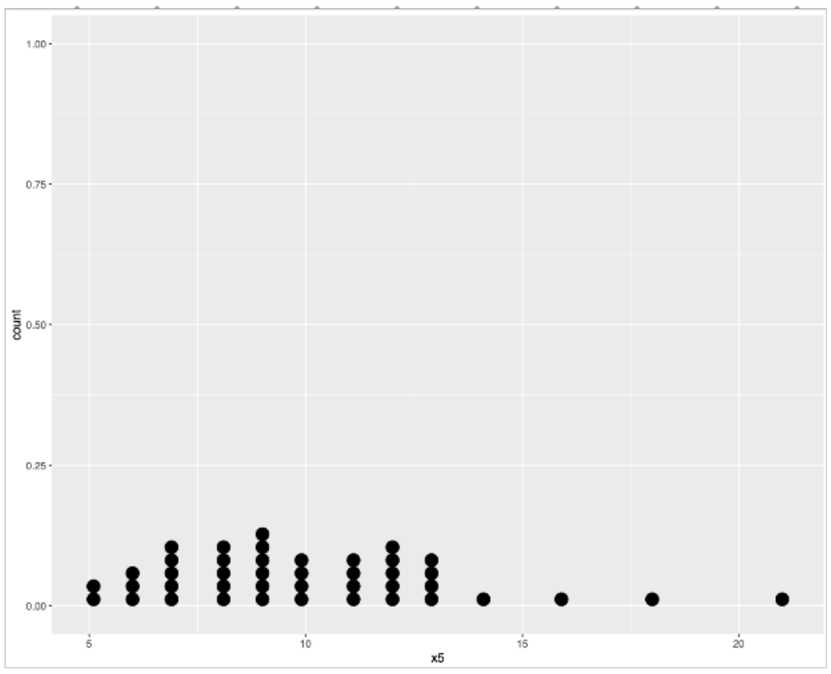
(a) Marginal dot diagrams are below:

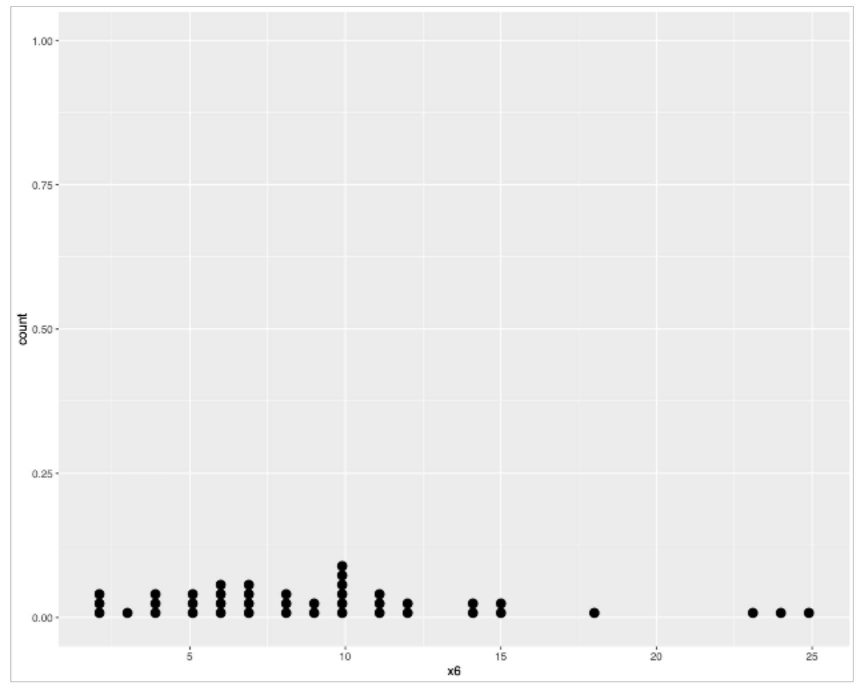


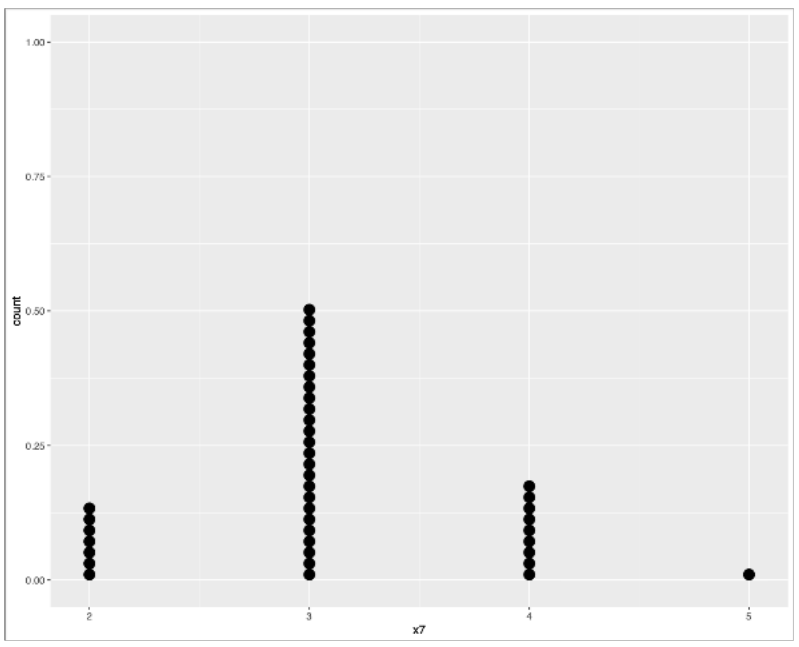












(d)

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 7.50 \\ 73.86 \\ 4.55 \\ 2.19 \\ 10.04 \\ 9.40 \\ 3.09 \end{bmatrix}$$

$$S_n = \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} & S_{16} & S_{17} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} & S_{26} & S_{27} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} & S_{36} & S_{37} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} & S_{46} & S_{47} \\ S_{51} & S_{52} & S_{53} & S_{54} & S_{55} & S_{56} & S_{57} \\ S_{61} & S_{62} & S_{63} & S_{64} & S_{65} & S_{66} & S_{67} \\ S_{71} & S_{72} & S_{73} & S_{74} & S_{75} & S_{76} & S_{77} \end{bmatrix}$$

$$
= \begin{bmatrix}
2.5 & & & & & & \\
-2.78 & 300.5 & & & & & \\
-0.378 & 3.9 & 1.52 & & & & \\
-0.46 & -1.39 & 0.67 & 1.18 & & & \\
-0.585 & 6.76 & 2.31 & 1.09 & 11.36 & & \\
-2.23 & 30.79 & 2.82 & -0.81 & 3.13 & 30.98 & \\
0.17 & 0.62 & 0.14 & 0.17 & 1.04 & 0.59 & 0.48
\end{bmatrix}
$$

Since this is a covariance matrix, it is symmetric, meaning the data in upper right hand side is the same as that of bottom left hand side. Hence no need to write down the upper right hand side.

$$
R = \begin{bmatrix}
1 & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} & r_{17} \\
r_{21} & 1 & r_{23} & r_{24} & r_{25} & r_{26} & r_{27} \\
r_{31} & r_{32} & 1 & r_{34} & r_{35} & r_{36} & r_{37} \\
r_{41} & r_{42} & r_{43} & 1 & r_{45} & r_{46} & r_{47} \\
r_{51} & r_{52} & r_{53} & r_{54} & 1 & r_{56} & r_{57} \\
r_{61} & r_{62} & r_{63} & r_{64} & r_{65} & 1 & r_{67} \\
r_{71} & r_{72} & r_{73} & r_{74} & r_{75} & r_{76} & 1
\end{bmatrix}
$$

$$
= \begin{bmatrix}
1 & & & & & & \\
-0.10 & 1 & & & & & \\
-0.19 & 0.183 & 1 & & & & \\
-0.26 & -0.07 & 0.50 & 1 & & & \\
-0.11 & 0.12 & 0.56 & 0.297 & 1 & & \\
-0.25 & 0.32 & 0.41 & -0.13 & 0.17 & 1 & \\
0.16 & 0.05 & 0.17 & 0.23 & 0.45 & 0.15 & 1
\end{bmatrix}
$$

Since this is a correlation matrix, it is symmetric, meaning the data in upper right hand side is the same as that of bottom left hand side. Hence no need to write down the upper right hand side.

It is hard to compare cross-columns covariance since the units and average values vary column by column.
However, as we take a look at the correlation matrix, we can see variable "Wind (x_1)" has mostly negative correlation with other variables, yet, most correlation values are relatively low, meaning the linear relationships are mostly weak.