A1 Random Vectors and Matrices

Problem 1

Three variables were measured (in milliequivalents per 100 g) at different locations. The variables are

y_1 = available soil calcium
y_2 = exchangeable soil calcium
y_3 = turnip green calcium

Location Number	y_1	y_2	y_3
1	35	3.5	2.80
2	35	4.9	2.70
3	40	30.0	4.38
4	10	2.8	3.21
5	6	2.7	2.73
6	20	2.8	2.81
7	35	4.6	2.88
8	35	10.9	2.90
9	35	8.0	3.28
10	30	1.6	3.20

The dataset is given below. Find the mean vector, the sample covariance matrix, and the sample correlation matrix.

Problem 2

The results of an experiment in which subjects respond to "probe words" at five positions in a sentence. The variables are response times for the jth probe word, $y_j, \quad j = 1,\cdots,5$. The data matrix is given below.

```
51  36  50  35  42;
27  20  26  17  27;
37  22  41  37  30;
42  36  32  34  27;
27  18  33  14  29;
43  32  43  35  40;
41  22  36  25  38;
38  21  31  20  16;
36  23  27  25  28;
26  31  31  32  36;
29  20  25  26  25;
```

1. Find the mean $\bar{z}$ and variance s_z^2 for the linear combination $z = 3y_1 - 2y_2 + 4y_3 - y_4 + y_5$
2. Find the mean $\bar{w}$ and variance s_w^2 for the linear combination $w = y_1 + 3y_2 - y_3 + y_4 - 2y_5$
3. Find the sample covariance of z and w, s_{zw}.
4. Find the sample correlation between z and w, r_{zw}.
5. Consider the linear combinations defined by $\mathbb{z} = \mathbb{A}\mathbb{y}$, where

$$\mathbb{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & -3 & 1 & -2 & -1 \\ -1 & -2 & 1 & -2 & 3 \end{bmatrix}$$

Determine $\bar{\mathbb{z}}$, $\mathbb{S}_z$, $\mathbb{R}_z$.

Problem 3

In an experiment, five variables are recorded for normal patients and diabetics. The data matrix for normal patients is given in the table below. The five variables are

$$
\begin{aligned}
y_1 &= \text{ relatve weight} \\
y_2 &= \text{ fasting plasma glucose} \\
x_1 &= \text{ glucose intolerance} \\
x_2 &= \text{ insulin response to oral glucose} \\
x_3 &= \text{ insulin resistance}
\end{aligned}
$$

Determine the mean vector and covariance matrix

Consider the two linear combinations

$$
\begin{aligned}
u &= 2y_1 - y_2 = \mathbf{a}'\mathbf{y} \\
v &= 2x_1 - 3x_2 + x_3 = \mathbf{b}'\mathbf{x}.
\end{aligned}
$$

Determine the covariance and correlation of u and v.

Consider the linear combinations

$$
\begin{aligned}
u_1 &= y_1 + y_2 & \qquad u_2 &= y_1 - y_2 \\
v_1 &= x_1 + x_2 + x_3 & \qquad v_2 &= x_1 - 2x_2 + 2x_3
\end{aligned}
$$

Determine the covariance of $\mathbf{u}$ and $\mathbf{v}$.

Patient Number	y_1	y_2	x_1	x_2	x_3
1	.81	80	356	124	55
2	.95	97	289	117	76
3	.94	105	319	143	105
4	1.04	90	356	199	108
5	1.00	90	323	240	143
6	.76	86	381	157	165
7	.91	100	350	221	119
8	1.10	85	301	186	105
9	.99	97	379	142	98
10	.78	97	296	131	94
11	.90	91	353	221	53
12	.73	87	306	178	66
13	.96	78	290	136	142
14	.84	91	371	200	93
15	.74	86	312	208	68
16	.98	80	393	202	102
17	1.10	90	364	152	76
18	.85	99	359	185	37
19	.83	85	296	116	60
20	.93	90	345	123	50
21	.95	91	378	136	47
22	.74	88	304	134	50
23	.95	95	347	184	91
24	.97	91	327	192	124
25	.72	92	386	279	74

Problem 4

Show that the centring matrix $\left(\mathbf{I}-\frac{1}{n}\mathbf{J}\right)$ centers $\mathbf{Y}$: If $\mathbf{Y}_c$ denotes the centred data matrix, show that $\mathbf{S}=\frac{1}{n-1}\mathbf{Y}_C'\mathbf{Y}_c$.

Problem 5

If the data matrix $\mathbb{Y}=\left(y_{ij}\right)$ is standardised to $\mathbb{Z}=\left(z_{ij}\right)$, where $z_{ij}=\left(y_{ij}-\bar{y}_j\right)/s_j$, show that the covariance matrix for the z's is equal to the correlation matrix for the y's.

$$\mathbb{S}_z=\frac{1}{n-1}\mathbb{Z}'\mathbb{Z}=\mathbb{R}$$

Problem 6

If a random vector $\mathbf{y}$ has mean μ and covariance $\mathbf{\Sigma}$, show that

$$\Sigma=E\left(\mathbf{y}\mathbf{y}'\right)-\mu\mu'$$

Problem 7

Show that $\operatorname{cov}(\mathbf{x}-\mathbf{a})=\operatorname{cov}(\mathbf{x})$, where $\mathbf{a}$ is constant. If the random vectors $\mathbf{x}$ and $\mathbf{y}$ are independent and are of the same size, show that

$$\operatorname{cov}\left(\mathbf{x}+\mathbf{y}\right)=\operatorname{cov}(\mathbf{x})+\operatorname{cov}\left(\mathbf{y}\right)$$
$$\operatorname{cov}\left(\mathbf{x}-\mathbf{y}\right)=\operatorname{cov}(\mathbf{x})+\operatorname{cov}\left(\mathbf{y}\right)$$

Problem 8

Define the standardisation $\mathbf{z}_i=\mathbf{S}^{-1/2}\left(\mathbf{y_i}-\bar{\mathbf{y}}\right) \qquad i=1,\ldots,n$. Show that $\bar{\mathbf{z}}=0$ and $\mathbf{S}_z=\mathbf{I}$.

Problem 9

If a sample $\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_n$ is transformed to $\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_n$ where $\mathbf{z}_i=\mathbf{A}\mathbf{y}_i$, show that the $k\times 1$ sample mean vector and $k\times k$ sample covariance matrix of the $\mathbf{z}$'s are given by

$$\bar{\mathbf{z}}=\mathbf{A}\bar{\mathbf{y}}$$
$$\mathbf{S}_z=\mathbf{A}\mathbf{S}\mathbf{A}'$$

Problem 10

Let $\mathbf{x}$ be an $n \times 1$ vector of random variables and let $\mathbf{A}$ be an $n \times n$ symmetric matrix. If $E(\mathbf{x}) = \mu$ and $\text{cov}(\mathbf{x}) = \mathbf{\Sigma}$, show that

$$E\left(\mathbf{x}'\mathbf{A}\mathbf{x}\right) = \text{tr}\left(\mathbf{A}\mathbf{\Sigma}\right) + \mu'\mathbf{A}\mu$$