Advanced Mathematical Statistics: Assignment 1

Chris Hayduk

September 19, 2019

Problem 3.4.

a) $\mathbf{y}_1 = \begin{bmatrix} 3.497900 \\ 2.485475 \\ 1.782875 \\ 1.725450 \\ 1.645575 \\ 1.469800 \end{bmatrix}$

Then we have,

$$\begin{aligned}
\text{proj}_{\mathbf{1}}\, \mathbf{y_1} &= \frac{\mathbf{y_1'1}}{\mathbf{1'1}}\mathbf{1} \\
&= \frac{12.607075}{6}\mathbf{1} \\
&= (2.10117917)\mathbf{1} \\
&= \begin{bmatrix} 2.10117917 \\ 2.10117917 \\ 2.10117917 \\ 2.10117917 \\ 2.10117917 \\ 2.10117917 \end{bmatrix}
\end{aligned}$$

b) We have $\overline{x}_1 = \frac{1}{6}(\sum_{i=1}^{6} x_{1_i}) = 2.10117917$

Thus,

$$\begin{aligned}
\mathbf{d}_1 = \mathbf{y}_1 - \overline{x}_1\mathbf{1} &= \mathbf{y}_1 - (2.10117917)\mathbf{1} \\
&= \begin{bmatrix} 1.39672083 \\ 0.38429583 \\ -0.31830417 \\ -0.37572917 \\ -0.45560417 \\ -0.63137917 \end{bmatrix}
\end{aligned}$$

Furthermore, $L_{\mathbf{d}_1} = \sqrt{\mathbf{d}_1'\mathbf{d}_1} = 1.7167461$

The sample standard deviation is given by

$$\begin{aligned}
\sqrt{\frac{1}{6}\sum_{i=1}^{6}(y_{1_i} - \overline{x}_1)^2} &= \sqrt{\frac{1}{6}\mathbf{d}_1'\mathbf{d}_1} \\
&= \frac{1}{\sqrt{6}}\sqrt{\mathbf{d}_1'\mathbf{d}_1} \\
&= \frac{1}{\sqrt{6}}L_{\mathbf{d}_1} \\
&= \frac{1}{\sqrt{6}}(1.7167461) \\
&\approx 0.70085866
\end{aligned}$$

Thus, the sample standard deviation of $\mathbf{y}_1$ is given by $\frac{1}{\sqrt{6}}L_{\mathbf{d}_1}$.

c) $L_{\mathbf{y}_1 - \overline{x}_1\mathbf{1}} = L_{\mathbf{d}_1} = \sqrt{\mathbf{d}_1'\mathbf{d}_1} = 1.7167461$

$L_{\mathbf{y}_1} = \sqrt{\mathbf{y}_1'\mathbf{y}_1} = 5.425582$
$L_{\overline{x}_1\mathbf{1}} = \sqrt{\overline{x}_1\mathbf{1}'\overline{x}_1\mathbf{1}} = 5.14681682$

d) $\mathbf{y}_2 = \begin{bmatrix} 0.623 \\ 0.593 \\ 0.512 \\ 0.500 \\ 0.463 \\ 0.395 \end{bmatrix}$

Then we have,

$$\begin{aligned}
\text{proj}_{\mathbf{1}}\,\mathbf{y_2} &= \frac{\mathbf{y_2'1}}{\mathbf{1'1}}\mathbf{1} \\
&= \frac{3.086}{6}\mathbf{1} \\
&= (0.51433)\mathbf{1} \\
&= \begin{bmatrix} 0.51433 \\ 0.51433 \\ 0.51433 \\ 0.51433 \\ 0.51433 \\ 0.51433 \end{bmatrix}
\end{aligned}$$

We have $\overline{x}_2 = \frac{1}{6}(\sum_{i=1}^{6} x_{2_i}) = 0.51433$

Thus,

$$\mathbf{d}_2 = \mathbf{y}_2 - \overline{x}_2\mathbf{1} = \mathbf{y}_2 - (0.51433)\mathbf{1}$$
$$= \begin{bmatrix} 0.10867 \\ 0.07867 \\ -0.00233 \\ -0.01433 \\ -0.05133 \\ -0.11933 \end{bmatrix}$$

Furthermore, $L_{\mathbf{d}_2} = \sqrt{\mathbf{d}_2'\mathbf{d}_2} = 0.187305$

The sample standard deviation is given by

$$\begin{aligned}\sqrt{\frac{1}{6}\sum_{i=1}^{6}(y_{2_i} - \overline{x}_2)^2} &= \sqrt{\frac{1}{6}\mathbf{d}_2'\mathbf{d}_2} \\ &= \frac{1}{\sqrt{6}}\sqrt{\mathbf{d}_2'\mathbf{d}_2} \\ &= \frac{1}{\sqrt{6}}L_{\mathbf{d}_2} \\ &= \frac{1}{\sqrt{6}}(0.187305) \\ &\approx 0.0764671\end{aligned}$$

Thus, the sample standard deviation of $\mathbf{y}_2$ is given by $\frac{1}{\sqrt{6}}L_{\mathbf{d}_2}$.

$L_{\mathbf{y}_2 - \overline{x}_2\mathbf{1}} = L_{\mathbf{d}_2} = \sqrt{\mathbf{d}_2'\mathbf{d}_2} = 0.187305$

$L_{\mathbf{y}_2} = \sqrt{\mathbf{y}_2'\mathbf{y}_2} = 1.27370$

$L_{\overline{x}_2\mathbf{1}} = \sqrt{\overline{x}_2\mathbf{1}'\overline{x}_2\mathbf{1}} = 1.25985$

e) $\cos\theta = \frac{\mathbf{d}_1'\mathbf{d}_2}{L_{\mathbf{d}_1}L_{\mathbf{d}_2}} = \frac{0.28686869}{1.7167461(0.187305)} \approx 0.89212910878$

$\implies \theta = \arccos(0.89212910878) \approx 0.4687601612$

Problem 3.7.

Problem 3.8a.

Total sample variance of $\mathbf{S}_1 = 1 + 1 + 1 = 3$.

Total sample variance of $\mathbf{S}_2 = 1 + 1 + 1 = 3$.

Although these matrices have very different entries in their off-diagonal elements, their total sample variances are still identical.

Problem 3.11.

$\mathbf{S} = \begin{bmatrix} 252.04 & -68.43 \\ -68.43 & 123.67 \end{bmatrix}$

Then $\mathbf{D}^{-1/2} = \begin{bmatrix} 0.0629891 & 0 \\ 0 & 0.0899224 \end{bmatrix}$

This yields,

$$\begin{aligned} \mathbf{D}^{-1/2}\mathbf{S} &= \begin{bmatrix} 15.875772764 & -4.310344113 \\ -6.15338932 & 11.120703208 \end{bmatrix} \\ \implies \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2} &= \begin{bmatrix} 1.000000638 & -0.3875965 \\ -0.3875965 & 1.0000322 \end{bmatrix} = \mathbf{R} \end{aligned}$$

Now define $\mathbf{D}^{1/2} = \begin{bmatrix} 15.8758 & 0 \\ 0 & 11.1207 \end{bmatrix}$

This yields,

$$\begin{aligned} \mathbf{D}^{1/2}\mathbf{R} &= \begin{bmatrix} 15.875772764 & -6.153404 \\ -4.310344113 & 11.120703208 \end{bmatrix} \\ \implies \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2} &= \begin{bmatrix} 252.04119 & -68.43016 \\ -68.43106 & 123.67001 \end{bmatrix} = \mathbf{S} \end{aligned}$$

Problem 3.12.

$$\begin{aligned} & \mathbf{S} = \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2} \\ \implies & |\mathbf{S}| = |\mathbf{D}^{1/2}||\mathbf{R}||\mathbf{D}^{1/2}| \\ \implies & |\mathbf{S}| = \left(\prod_{i=1}^{p} \sqrt{s_{ii}}\right) |\mathbf{R}| \left(\prod_{i=1}^{p} \sqrt{s_{ii}}\right) \\ \implies & |\mathbf{S}| = \left(\prod_{i=1}^{p} s_{ii}\right) |\mathbf{R}| \\ \implies & |\mathbf{S}| = (s_{11}s_{22}\cdots s_{pp})\,|\mathbf{R}| \end{aligned}$$

Problem 3.13.

The standardized variables (with mean 0 and standard deviation 1) yield the following covariance,

$$
\begin{aligned}
s_{ik} &= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_{ji}-\overline{x}_i}{\sqrt{s_{ii}}}-\overline{x}_i\right)\left(\frac{x_{jk}-\overline{x}_k}{\sqrt{s_{kk}}}-\overline{x}_k\right)\\
&= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_{ji}-\overline{x}_i}{\sqrt{s_{ii}}}\right)\left(\frac{x_{jk}-\overline{x}_k}{\sqrt{s_{kk}}}\right)\\
&= \frac{1}{n}\frac{\sum_{j=1}^{n}\left(x_{ji}-\overline{x}_i\right)\left(x_{jk}-\overline{x}_k\right)}{\sqrt{s_{ii}}\sqrt{s_{kk}}}\\
&= \frac{\sum_{j=1}^{n}\left(x_{ji}-\overline{x}_i\right)\left(x_{jk}-\overline{x}_k\right)}{\sqrt{\sum_{j=1}^{n}\left(x_{ji}-\overline{x}_i\right)^2}\sqrt{\sum_{j=1}^{n}\left(x_{jk}-\overline{x}_k\right)}} = r_{ik}
\end{aligned}
$$

Thus, $s_{ik} = r_{ik}$.

Problem 3.14.

a) $\mathbf{c}'\mathbf{X} = \begin{bmatrix} -7 \\ 1 \\ 3 \end{bmatrix}$

Sample mean of $\mathbf{c}'\mathbf{X} = \frac{-7+1+3}{3} = -1$

Sample variance of $\mathbf{c}'\mathbf{X} = \frac{(-7-(-1))^2+(1-(-1))^2+(3-(-1))^2}{3-1} = 28.$

$\mathbf{b}'\mathbf{X} = \begin{bmatrix} 21 \\ 19 \\ 8 \end{bmatrix}$

Sample mean of $\mathbf{b}'\mathbf{X} = \frac{21+19+8}{3} = 16$

Sample variance of $\mathbf{b}'\mathbf{X} = \frac{(21-16)^2+(19-16)^2+(8-16)^2}{3-1} = 49.$

The sample covariance is $\frac{(-7-(-1))(21-16)+(1-(-1))(19-16)+(3-(-1))(8-16)}{3-1} = -28$

b) Now, from the original data matrix, we get $\overline{\mathbf{x}} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ and $\mathbf{S} = \begin{bmatrix} 16 & -2 \\ -2 & 1 \end{bmatrix}$

Sample mean of $\mathbf{c}'\mathbf{X} = \mathbf{c}'\overline{\mathbf{x}} = \begin{bmatrix} -1 & 2 \end{bmatrix}\begin{bmatrix} 5 \\ 2 \end{bmatrix} = -1$

Sample mean of $\mathbf{b}'\mathbf{X} = \mathbf{b}'\overline{\mathbf{x}} = \begin{bmatrix} 2 & 3 \end{bmatrix}\begin{bmatrix} 5 \\ 2 \end{bmatrix} = 16$

Sample variance of $\mathbf{c}'\mathbf{X} = \mathbf{c}'\mathbf{S}\mathbf{c} = \begin{bmatrix} -1 & 2 \end{bmatrix}\begin{bmatrix} 16 & -2 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = 28$

Sample variance of $\mathbf{b}'\mathbf{X} = \mathbf{b}'\mathbf{S}\mathbf{b} = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} 16 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 49$

Sample covariance of $\mathbf{b}'\mathbf{X}$ and $\mathbf{c}'\mathbf{X} = \mathbf{b}'\mathbf{S}\mathbf{c} = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} 16 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = -28$

Problem 3.16.

We have,

$$\begin{aligned} E(\mathbf{V} - \mu_{\mathbf{V}})(\mathbf{V} - \mu_{\mathbf{V}})' &= E(\mathbf{V}\mathbf{V}') - \mu_{\mathbf{v}} E(\mathbf{V}') - E(\mathbf{V})\mu_{\mathbf{V}}' + \mu_{\mathbf{V}}\mu_{\mathbf{V}}' \\ &= E(\mathbf{V}\mathbf{V}') - \mu_{\mathbf{V}}\mu_{\mathbf{V}}' \end{aligned}$$

Thus,

$$\begin{aligned} E(\mathbf{V}\mathbf{V}') &= E(\mathbf{V} - \mu_{\mathbf{V}})(\mathbf{V} - \mu_{\mathbf{V}})' + \mu_{\mathbf{V}}\mu_{\mathbf{V}}' \\ &= \mathbf{\Sigma}_{\mathbf{V}} + \mu_{\mathbf{V}}\mu_{\mathbf{V}}' \end{aligned}$$

Problem 3.17.

We know that $P[X_1 \leq x_1, \cdots, X_p \leq x_p \text{ and } Z_1 \leq z_1, \cdots, Z_q \leq z_q] = P[X_1 \leq x_1, \cdots, X_p \leq x_p] \cdot P[Z_1 \leq z_1, \cdots, Z_q \leq z_q]$ by independence of $\mathbf{X}$ and $\mathbf{Z}$.

Fix $k \in \{1, ..., p\}$. Then, let j range over $\{1, ..., q\}$. For every $a \in \{1, ..., p\}$ such that $a \neq k$ and for every $b \in \{1, ..., q\}$ such that $b \neq j$, let x_a and z_b tend to infinity. Then we have,

$$P[X_k \leq x_k \text{ and } Z_j \leq z_j] = P[X_k \leq x_k] \cdot P[Z_j \leq z_j]$$

for every $k \in \{1, ..., p\}$ and $j \in \{1, ..., q\}$. Thus, each component of $\mathbf{X}$ is independent of each component of $\mathbf{Z}$.

Problem 3.18.

a) Sample mean of total energy consumption $= 0.766 + 0.508 + 0.438 + 0.161 = 1.873$.

   Total variance of sample $= 0.856 + 0.568 + 0.171 + 0.043 = 1.638$.

b) $\overline{x}_{1-2} = \overline{x}_1 - \overline{x}_2 = 0.258$.

   The variance of the excess petroleum consumption over natural gas consumption is given by $\text{Var}[x_1] + \text{Var}[x_2] - 2\text{Cov}[x_1, x_2] = 0.856 + 0.568 - 2(0.635) = 0.154$.