STT43: Multivariate Data Analysis **Term: Spring 2026**

Homework #1 (due 01/25/26)

Instructor: Dr. Guanqun Cao (caoguanq@msu.edu)
Grader: Andrews Boahen (boahenan@msu.edu)

All problems on this homework are from the required textbook Applied Multivariate Statistical Analysis (6th edition, by Johnson and Wichern).

Problems from Chapter 1

Problem 1.6

The data is provided in Table 1.5

Solution

R codes provided for both questions

a) **marginal dot diagrams for all the variables.**

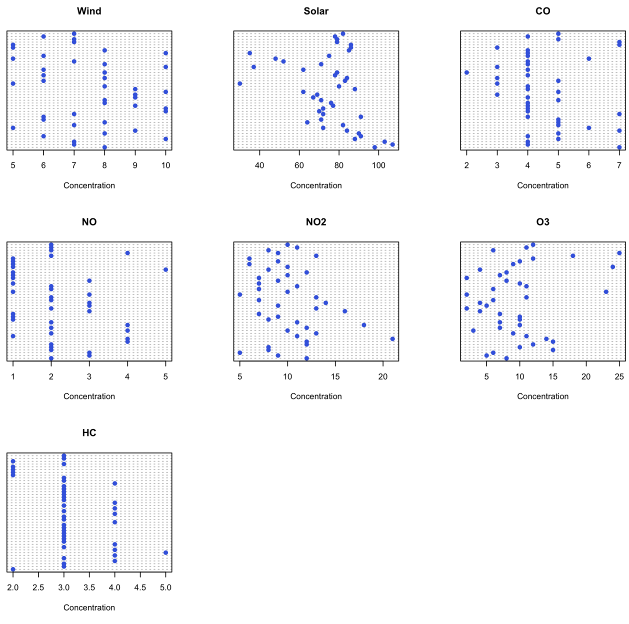

Figure 1: Dot plot for all the variables

```r
library(readr)
library(dplyr)
air <- read_csv("air_pollution_table.csv") #specify path to this file if your working directory is different
numeric_cols <- names(air)[sapply(air, is.numeric)] #column names
par(mfrow = c(3,3))    # adjust dimensions depending on how many numeric columns you have
for (col_name in numeric_cols) {
  dotchart(air[[col_name]],
           xlab = "Concentration",
           main = col_name,
           pch = 19, color = "royalblue")}
```

b) **Construction of $\bar{x}, S_n$ and R arrays.**
let denote the column names respectively by $x_1, x_2, \ldots, x_7$. We have:

```r
xbar <- colMeans(air)
xbar
Sn <- cov(air)
Sn
R <- cor(air)
R
```

$$\bar{X} = [7.476190, 75.095238, 4.523810, 2.166667, 10.142857, 9.500000, 3.119048]'$$

$$S_n = \begin{bmatrix} 2.499419 & -2.802555 & -0.304297 & -0.471545 & -0.533101 & -2.097561 & 0.185830 \\ -2.802555 & 272.185830 & 3.729384 & -0.772358 & 3.522648 & 28.365854 & -0.353078 \\ -0.304297 & 3.729384 & 1.475029 & 0.691057 & 2.216028 & 2.609756 & 0.131243 \\ -0.471545 & -0.772358 & 0.691057 & 1.166667 & 1.170732 & -0.719512 & 0.199187 \\ -0.533101 & 3.522648 & 2.216028 & 1.170732 & 10.954704 & 2.609756 & 0.933798 \\ -2.097561 & 28.365854 & 2.609756 & -0.719512 & 2.609756 & 30.207317 & 0.475610 \\ 0.185830 & -0.353078 & 0.131243 & 0.199187 & 0.933798 & 0.475610 & 0.448897 \end{bmatrix}$$

and

$$R = \begin{bmatrix} 1.000000 & -0.107449 & -0.158481 & -0.276140 & -0.101880 & -0.241401 & 0.175438 \\ -0.107449 & 1.000000 & 0.186125 & -0.043342 & 0.064511 & 0.312829 & -0.031942 \\ -0.158481 & 0.186125 & 1.000000 & 0.526793 & 0.551283 & 0.390970 & 0.161288 \\ -0.276140 & -0.043342 & 0.526793 & 1.000000 & 0.327479 & -0.121202 & 0.275242 \\ -0.101880 & 0.064511 & 0.551283 & 0.327479 & 1.000000 & 0.143464 & 0.421094 \\ -0.241401 & 0.312829 & 0.390970 & -0.121202 & 0.143464 & 1.000000 & 0.129158 \\ 0.175438 & -0.031942 & 0.161288 & 0.275242 & 0.421094 & 0.129158 & 1.000000 \end{bmatrix}$$

Interpretation
The correlation matrix reveals **moderate to weak relationships** overall (most $|\rho| < 0.4$). The most meaningful patterns are:

- **Moderate positive associations** ($\rho > 0.4$):
  * CO $\leftrightarrow$ NO_2 ($\rho \approx 0.55$)
  * CO $\leftrightarrow$ NO ($\rho \approx 0.53$)
  * NO_2 $\leftrightarrow$ HC ($\rho \approx 0.42$)
  * Solar radiation $\leftrightarrow$ O_3 ($\rho \approx 0.31$)
  * CO $\leftrightarrow$ O_3 ($\rho \approx 0.39$)

  This implies that each element in the pair of variables above tend to increase when the other element increases.
- **Weak Negative relationships** ($\rho < -0.2$):
  * Wind $\leftrightarrow$ NO ($\rho \approx -0.28$)
  * Wind $\leftrightarrow$ O_3 ($\rho \approx -0.24$)
  * HC $\leftrightarrow$ Solar ($\rho \approx -0.0.032$)

  This implies that each element in the pair of variables above tend to decrease when the other element increases. When $|\rho| \to 0$, none of the element in the pairs is linearly associated with the other.

  Same arguments can be used for all other pairs in the table.

Problem 1.8

Solution

R codes provided

```r
# install.packages("car") if needed
library(car)
Q <- c(1, 0)
P <- c(-1, -1)
# Define the inverse covariance matrix ¹
A_inv <- matrix(c(1/3,  1/9,
                  1/9,  4/27), nrow = 2, byrow = TRUE)

# We need the covariance matrix  = (¹)¹ for car::ellipse
Sigma <- solve(A_inv)

# Create the plot
plot(c(-2.5, 4.5), c(-3, 3), type = "n", asp = 1,
     xlab = "x1", ylab = "x2",
     main = "Locus: squared statistical distance = 1 from Q = (1, 0)")

# Draw the ellipse (radius² = 1 → constant = 1)
ellipse(center = Q, shape = Sigma, radius = 1,
        segments = 201, lwd = 2.5, col = "royalblue")

# Add points
points(Q[1], Q[2], pch = 19, col = "red", cex = 1.8)
text(Q[1] + 0.25, Q[2] + 0.25, "Q", col = "red", font = 2, cex = 1.4)

points(P[1], P[2], pch = 17, col = "darkgreen", cex = 1.8)
text(P[1] - 0.3, P[2] - 0.3, "P", col = "darkgreen", font = 2, cex = 1.4)

grid(lty = 3, col = "gray70")
```

Euclidean (D_1) and Statistical (D_2) distances.
$P = (-1, -1), Q = (1, 0)$, using $a_{11} = 1/3, a_{22} = 4/27, a_{12} = 1/9$, we have:

$$D_1 = \sqrt{(-1-1)^2 + (-1-0)^2} = \sqrt{5} \approx 2.2361$$

$$D_2 = \sqrt{a_{11}(-1-1)^2 + 2a_{12}(-1-1)(-1-0) + a_{22}(-1-0)^2} = \sqrt{52/27} \approx 1.3878$$

Plot

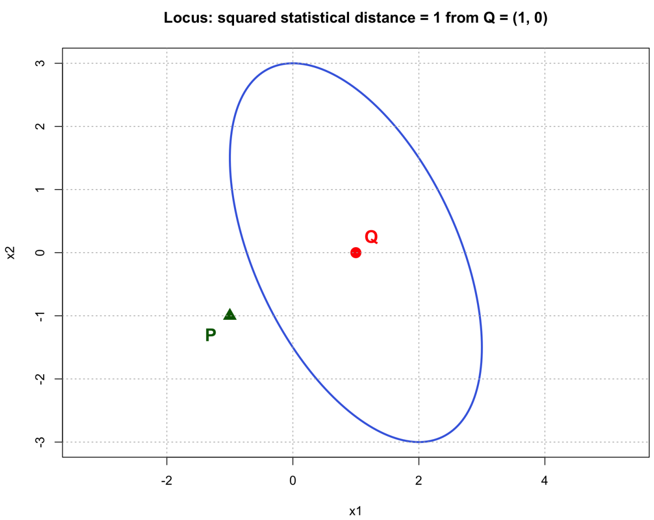

Figure 2: Locus of points that are a constant squared statistical distance 1 from the point Q.

Problems from Chapter 2

Problem 2.4

Solution

a) *Proof.*

$$\begin{aligned} (A^{-1})'A' &= (AA^{-1})' \\ &= I' \qquad \text{since } A^{-1} \text{ exists} \\ &= I \end{aligned}$$

Same idea is used to prove the right inverse. Hence $(A^{-1})' = (A')^{-1}$. $\square$

b) *Proof.*

$$\begin{aligned} AB(B^{-1}A^{-1}) &= ABB^{-1}A^{-1} \\ &= AIA^{-1} \\ &= AA^{-1} = I \end{aligned} \tag{0.1}$$

We can also use the same idea to prove left inverse. Hence, $(AB)^{-1} = B^{-1}A^{-1}$. $\square$

Problem 2.15

Solution

Proof. Let $\mathbf{x} = (x_1, x_2)'$, we observe that

$$3x_1^2 + 3x_2^2 - 2x_1x_2 = \mathbf{x}'A\mathbf{x},$$

where $A = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}$ with eigenvalues 4 and 2. A is then positive definite and hence $3x_1^2 + 3x_2^2 - 2x_1x_2$ is positive definite. $\square$

Problem 2.21

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

Solution

a) **Eigenvalues and eigenvectors of $A'A$**

We have $A'A = \begin{bmatrix} 9 & 1 \\ 1 & 9 \end{bmatrix}$ and by solving $det(A'A - \lambda I) = 0$, we get:

$$\begin{aligned} det(A'A - \lambda I) = 0 &\implies (9-\lambda)^2 - 1 = 0 \\ &\implies \lambda_1 = 8, \lambda_2 = 10. \end{aligned}$$

The corresponding (unormalized) eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are obtained by solving $(A - \lambda I)v = 0$, and we have

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

b) **Eigenvalues and eigenvectors of AA'**

We have $AA' = \begin{bmatrix} 2 & 0 & 4 \\ 0 & 8 & 0 \\ 4 & 0 & 8 \end{bmatrix}$ and by solving $det(AA' - \lambda I) = 0$, we get:

$$\begin{aligned} det(AA' - \lambda I) = 0 &\implies (8-\lambda)\begin{vmatrix} 2-\lambda & 4 \\ 4 & 8-\lambda \end{vmatrix} = 0 \\ &\implies (8-\lambda)\big[(2-\lambda)(8-\lambda) - 16\big] = 0 \\ &\implies \lambda_1 = 0, \lambda_2 = 8, \lambda_3 = 10. \end{aligned}$$

The corresponding (unormalized) eigenvectors $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ are obtained by solving $(A - \lambda I)v = 0$, and we have

$$\mathbf{v}_1 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

The nonzero eigenvalues are the same as those in part a).

c) **Singular-value decomposition of A**
The full SVD decomposition is given by:

$$\underset{(3\times 2)}{A} = \underset{(3\times 3)}{U}\underset{(3\times 2)}{\Lambda}\underset{(2\times 2)}{V'},$$

where U has m orthogonal eigenvectors of AA' as its columns, V has k orthogonal eigenvectors of $A'A$ as its columns and the 3×2 matrix Λ has (i, i) entry $\lambda_i \geq 0$, for $i = 1, 2, ..., min(m, k)$ and the other entries are zero.

The full SVD of A is then:

$$A = \begin{bmatrix} 1/\sqrt{5} & 0 & -2/\sqrt{5} \\ 0 & 1 & 0 \\ 2/\sqrt{5} & 0 & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} \sqrt{10} & 0 \\ 0 & \sqrt{8} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Problem 2.24

We have

$$\Sigma = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Solution

a) **Inverse of Σ**

$$\Sigma^{-1} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/9 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

b) **Eigenvalues and eigenvectors of Σ**
We have

$$\begin{aligned} |\Sigma - \lambda I| = 0 &\implies (4-\lambda)(9-\lambda)(1-\lambda) = 0 \\ &\implies \lambda_1 = 1, \lambda_2 = 4, \lambda_3 = 9 \end{aligned}$$

By solving $(\Sigma - \lambda I)\mathbf{v} = 0$, we obtain the corresponding eigenectors as:

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

b) **Eigenvalues and eigenvectors of Σ^{-1}**
We have

$$|\Sigma^{-1} - \lambda I| = 0 \implies (1/4 - \lambda)(1/9 - \lambda)(1 - \lambda) = 0$$
$$\implies \lambda_1 = 1/9, \lambda_2 = 1/4, \lambda_3 = 1$$

By solving $(\Sigma^{-1} - \lambda I)\mathbf{v} = 0$, we obtain the corresponding eigenectors as:

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Problem 2.30

We have the random vector $\mathbf{X}' = [X_1, X_2, X_3, X_4]$ with mean vector $\mu'_{\mathbf{X}} = [4, 3, 2, 1]$ and covariance matrix:

$$\mathbf{\Sigma_X} = \begin{bmatrix} 3 & 0 & 2 & 2 \\ 0 & 1 & 1 & 0 \\ 2 & 1 & 9 & -2 \\ 2 & 0 & -2 & 4 \end{bmatrix}$$

Solution

Partition $\mathbf{X}$ as $\mathbf{X}^{(1)} = [X_1, X_2]'$ and $\mathbf{X}^{(2)} = [X_3, X_4]'$. Let $\mathbf{A} = \begin{bmatrix} 1 & 2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & -2 \\ 2 & -1 \end{bmatrix}$. The partitioned mean and covariance matrices are:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

$$\mathbf{\Sigma}_{11} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{\Sigma}_{12} = \begin{bmatrix} 2 & 2 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{\Sigma}_{22} = \begin{bmatrix} 9 & -2 \\ -2 & 4 \end{bmatrix}.$$

a) $E(\mathbf{X}^{(1)}) = \boldsymbol{\mu}_1 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$.

b) $E(\mathbf{A}\mathbf{X}^{(1)}) = \mathbf{A}\boldsymbol{\mu}_1 = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = 10.$

c) $\mathrm{Cov}(\mathbf{X}^{(1)}) = \mathbf{\Sigma}_{11} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$.

d) $\mathrm{Cov}(\mathbf{A}\mathbf{X}^{(1)}) = \mathbf{A}\mathbf{\Sigma}_{11}\mathbf{A}' = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 7.$

e) $E(\mathbf{X}^{(2)}) = \boldsymbol{\mu}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

f) $E(\mathbf{B}\mathbf{X}^{(2)}) = \mathbf{B}\boldsymbol{\mu}_2 = \begin{bmatrix} 1 & -2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$.

g) $\mathrm{Cov}(\mathbf{X}^{(2)}) = \mathbf{\Sigma}_{22} = \begin{bmatrix} 9 & -2 \\ -2 & 4 \end{bmatrix}$.

h) $\mathrm{Cov}(\mathbf{B}\mathbf{X}^{(2)}) = \mathbf{B}\mathbf{\Sigma}_{22}\mathbf{B}' = \begin{bmatrix} 1 & -2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 9 & -2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -2 & -1 \end{bmatrix} = \begin{bmatrix} 33 & 36 \\ 36 & 48 \end{bmatrix}$.

i) $\mathrm{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \mathbf{\Sigma}_{12} = \begin{bmatrix} 2 & 2 \\ 1 & 0 \end{bmatrix}$.

j) $\mathrm{Cov}(\mathbf{A}\mathbf{X}^{(1)}, \mathbf{B}\mathbf{X}^{(2)}) = \mathbf{A}\mathbf{\Sigma}_{12}\mathbf{B}' = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -2 & -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$.