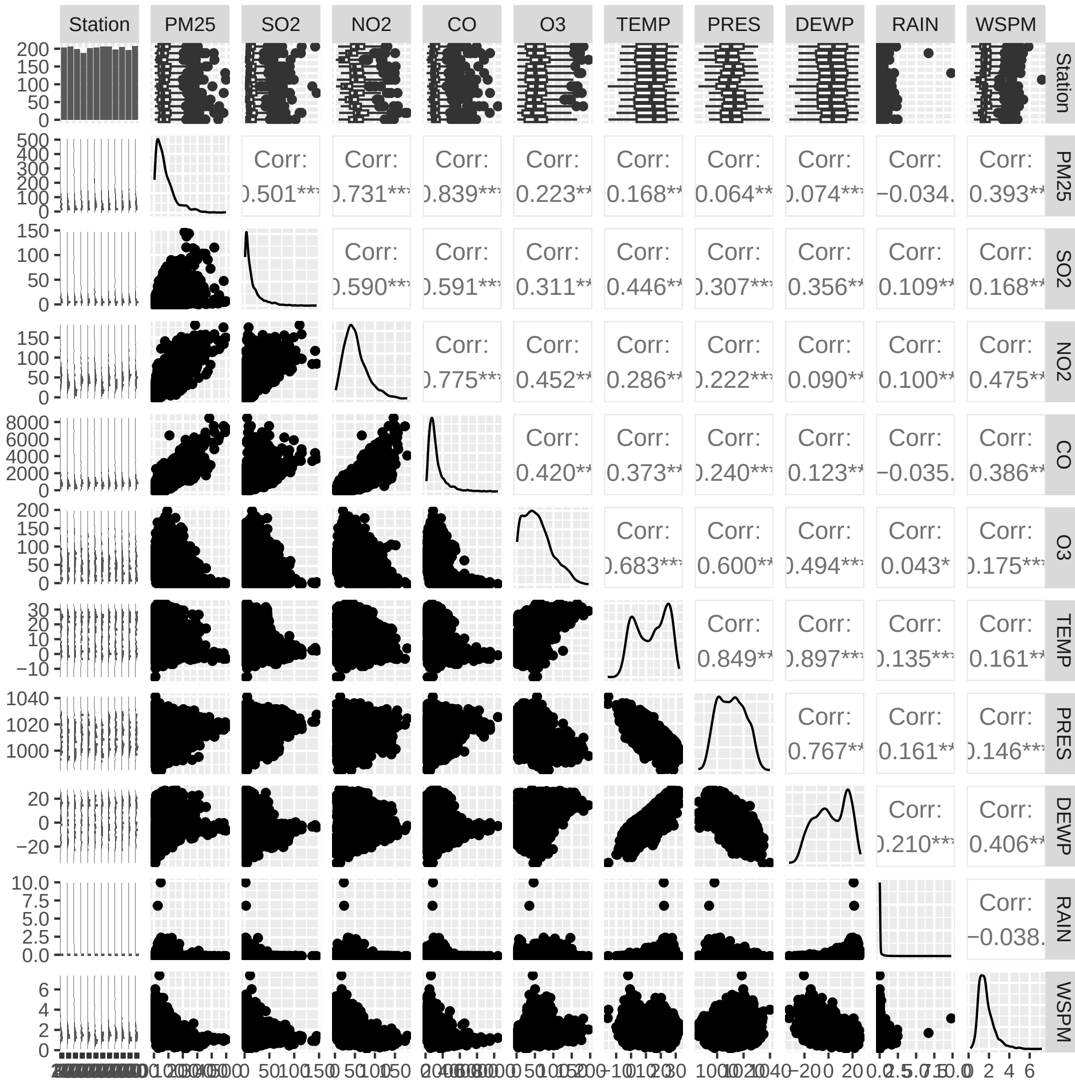
Sketch of Solutions
Compulsory HWA1, Multivariate Analysis, 2023HT

Shaobo Jin

1. (3p) Download the data set HWA1.RData from the Studium page, and complete the following tasks. A short description of the data set can be found at Studium.

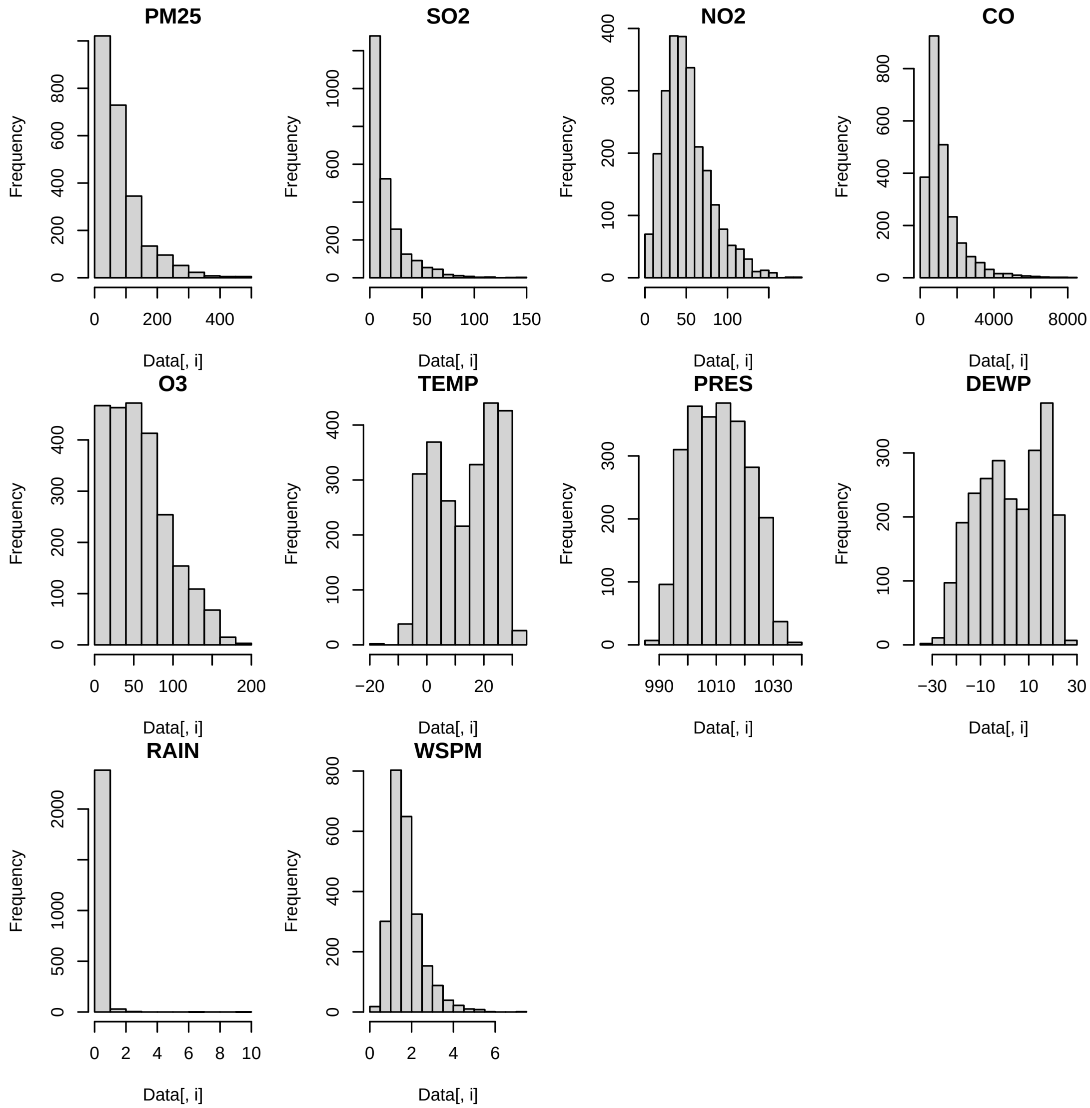
   (a) Apply different graphical methods for visualization.
   Solution: The matrix of scatter plots are shown below.

```
#########################################################
####   Chapter 4: Normal
#########################################################
library(ggplot2)
library(GGally)
load("HWA1.RData")
ggpairs(Data, progress = FALSE) # Suppress message
```



(b) Carry out different tools for assessing normality, ignoring the variable *Station*. Is it plausible to assume normality?
Solution: We first compute histograms for each variable separately.

```
## Histogram
par(mfrow = c(3, 4), mar = c(4.1, 4.1, 1, 1))
for(i in 2 : 11){
  hist(Data[, i], main = names(Data)[i])
}
```



You can also compute the kernel density plots or the QQ plots. But we skip them here. There are various tests that you can use to test normality. For example, we use the Shapiro test to test univariate normality as follows.

```
for(i in 2 : 11){
  print(shapiro.test(Data[, i]))
}

##
##  Shapiro-Wilk normality test
##
## data:  Data[, i]
## W = 0.8248, p-value < 2.2e-16
##
##
##  Shapiro-Wilk normality test
```

```
##
## data:  Data[, i]
## W = 0.73399, p-value < 2.2e-16
##
##
##  Shapiro-Wilk normality test
##
## data:  Data[, i]
## W = 0.94716, p-value < 2.2e-16
##
##
##  Shapiro-Wilk normality test
##
## data:  Data[, i]
## W = 0.76999, p-value < 2.2e-16
##
##
##  Shapiro-Wilk normality test
##
## data:  Data[, i]
## W = 0.95227, p-value < 2.2e-16
##
##
##  Shapiro-Wilk normality test
##
## data:  Data[, i]
## W = 0.93507, p-value < 2.2e-16
##
##
##  Shapiro-Wilk normality test
##
## data:  Data[, i]
## W = 0.98021, p-value < 2.2e-16
##
##
##  Shapiro-Wilk normality test
##
## data:  Data[, i]
## W = 0.95534, p-value < 2.2e-16
##
##
##  Shapiro-Wilk normality test
##
## data:  Data[, i]
```

```
## W = 0.17399, p-value < 2.2e-16
##
##
##  Shapiro-Wilk normality test
##
## data:  Data[, i]
## W = 0.8966, p-value < 2.2e-16
```

Univariate normality will be rejected for every variable. Hence, it is not surprising to see that the multivariate test also suggests non-normality.

```
library(MVN)
mvn(Data[, -1], mvnTest = "mardia")$multivariateNormality
```

```
##              Test        Statistic p value Result
## 1 Mardia Skewness 141087.639306458       0     NO
## 2 Mardia Kurtosis 801.732954871278       0     NO
## 3             MVN             <NA>    <NA>     NO
```

When we test univariate and multivariate normality, we only used one test. You are also welcome to use other tests.

(c) Compute the sample mean and sample covariance matrix of the whole data set in R, ignoring the variable *Station*.
Solution: The sample mean can be computed as follows.

```
colMeans(Data[, -1])
```

```
##         PM25          SO2          NO2           CO           O3          TEM
## 8.065123e+01 1.565051e+01 5.109601e+01 1.237770e+03 5.680153e+01 1.350092e+0
##         PRES         DEWP         RAIN         WSPM
## 1.010832e+03 2.519436e+00 6.121196e-02 1.714251e+00
```

The sample covariance matrix can be computed as follows.

```
cov(Data[, -1])
```

```
##                PM25           SO2          NO2            CO            O3
## PM25  5108.3730788   638.4622817  1478.237572   60129.89465 -6.149024e+02
## SO2    638.4622817   317.9801921   297.713765   10571.57775 -2.134338e+02
## NO2   1478.2375724   297.7137650   800.439822   21985.61384 -4.921983e+02
## CO   60129.8946520 10571.5777467 21985.613837 1006660.48122 -1.621863e+04
## O3    -614.9024458  -213.4337529  -492.198288  -16218.62722  1.483611e+03
## TEMP  -129.5047286   -85.6234095   -87.247757   -4031.19239  2.832272e+02
## PRES    46.4232161    55.4175997    63.673500    2438.73287 -2.339092e+02
## DEWP    71.2731190   -85.4541153   -34.228655   -1663.63042  2.561879e+02
## RAIN    -0.7920767    -0.6260271    -0.914932     -11.40189  5.322628e-01
```

```
## WSPM   -21.7384293    -2.3187232   -10.393801    -300.10621  5.213447e+00
##               TEMP          PRES         DEWP          RAIN          WSPM
## PM25  -129.5047286   46.4232161   71.2731190  -0.792076689 -2.173843e+01
## SO2    -85.6234095   55.4175997  -85.4541153  -0.626027110 -2.318723e+00
## NO2    -87.2477569   63.6735002  -34.2286551  -0.914932016 -1.039380e+01
## CO   -4031.1923890 2438.7328714 -1663.6304161 -11.401891872 -3.001062e+02
## O3     283.2271836 -233.9092470  256.1878664   0.532262828  5.213447e+00
## TEMP   115.9991183  -92.6002515  130.0581696   0.468881947 -1.345915e+00
## PRES   -92.6002515  102.5880272 -104.5826720  -0.525807097  1.144950e+00
## DEWP   130.0581696 -104.5826720  181.1637828   0.913750630 -4.230421e+00
## RAIN     0.4688819   -0.5258071    0.9137506   0.104500731 -9.424985e-03
## WSPM    -1.3459148    1.1449505   -4.2304214  -0.009424985  5.994389e-01
```

(d) There are 12 stations in total. Let $\boldsymbol{S}_i$ be the sample covariance matrix of Station i and let n_i be the corresponding sample size. Suppose that the normality assumption and the independence assumption hold. What is the distribution of $\sum_i (n_i - 1)\boldsymbol{S}_i$?
Solution: If we have the normality assumption, then $(n_i - 1)\boldsymbol{S}_i \sim W_{10}(\boldsymbol{\Sigma}, n_i - 1)$. Because of the independent assumption, $\sum_i (n_i - 1)\boldsymbol{S}_i \sim W_{10}(\boldsymbol{\Sigma}, \sum_i n_i - 12)$.

2. (1p) Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two random vectors of same size. Is

$$\mathrm{var}(\boldsymbol{X}+\boldsymbol{Y}) = \mathrm{var}(\boldsymbol{X}) + 2\mathrm{cov}(\boldsymbol{X},\boldsymbol{Y}) + \mathrm{var}(\boldsymbol{Y})?$$

State also the reason.
Solution: No, unless $\mathrm{cov}(\boldsymbol{X},\boldsymbol{Y}) = \mathrm{cov}(\boldsymbol{Y},\boldsymbol{X})$. In general,

$$\mathrm{var}(\boldsymbol{X}+\boldsymbol{Y}) = \mathrm{var}(\boldsymbol{X}) + \mathrm{cov}(\boldsymbol{X},\boldsymbol{Y}) + \mathrm{cov}(\boldsymbol{Y},\boldsymbol{X}) + \mathrm{var}(\boldsymbol{Y}).$$

3. (1p) Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two random vectors, and $\boldsymbol{A}$ and $\boldsymbol{B}$ be two constant matrices. Let $\boldsymbol{S}_{XY}$ be the sample covariance matrix between $\boldsymbol{X}$ and $\boldsymbol{Y}$. Express the sample covariance matrix between $\boldsymbol{AX}$ and $\boldsymbol{BY}$ using $\boldsymbol{S}_{XY}$, $\boldsymbol{A}$, and $\boldsymbol{B}$.
Solution: The sample covariance matrix is

$$\begin{aligned}\frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{AX}_i - \boldsymbol{A}\bar{\boldsymbol{X}}_i\right)\left(\boldsymbol{BY}_i - \boldsymbol{B}\bar{\boldsymbol{Y}}_i\right)^T &= \boldsymbol{A}\frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_i\right)\left(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}_i\right)^T\boldsymbol{B} \\ &= \boldsymbol{A}\boldsymbol{S}_{XY}\boldsymbol{B}.\end{aligned}$$

4. (1p) Let $X_1 \sim N(0,1)$ and

$$X_2 = \begin{cases} -X_1, & \text{if } -1 \le X_1 \le 1, \\ X_1, & \text{otherwise.}\end{cases}$$

Show that $X_1 - X_2$ is not normal and hence $\begin{bmatrix} X_1 & X_2 \end{bmatrix}^T$ is not bivariate normal.
Solution:

$$X_1 - X_2 = \begin{cases} 2X_1, & \text{if } -1 \le X_1 \le 1, \\ 0, & \text{otherwise.}\end{cases}$$

Hence,

$$P(X_1 - X_2 = 0) \quad = \quad 1 - P(-1 \le X_1 \le 1) > 0$$

which is not normal. Hence, their joint distribution cannot be bivariate normal, since any linear combination of normal should also be normal. In fact, you can show that $X_2 \sim N(0,1)$.

5. (2p) Let

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right),$$

and

$$\boldsymbol{A} \quad = \quad \begin{bmatrix} \boldsymbol{I}_{q\times q} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \boldsymbol{0} & \boldsymbol{I}_{r\times r} \end{bmatrix}.$$

(a) Find the joint distribution of $\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})$, where $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$.

Solution: Note that

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T &= \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T & \boldsymbol{I} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{0} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T & \boldsymbol{I} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.
\end{aligned}$$

The joint distribution should be normal as

$$\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu}) \quad \sim \quad N\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right)$$

as linear combinations of normal remain normal.

(b) Is $\boldsymbol{X}_2 - \boldsymbol{\mu}_2$ independent of $\boldsymbol{X}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2)$?

Solution: Note that

$$\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu}) \quad = \quad \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{X}_2 - \boldsymbol{\mu}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2) \\ \boldsymbol{X}_2 - \boldsymbol{\mu}_2 \end{bmatrix}.$$

We have shown in (a) that $\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})$ is multivariate normal. The covariance between $\boldsymbol{X}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2)$ and $\boldsymbol{X}_2 - \boldsymbol{\mu}_2$ is $\boldsymbol{0}$. Hence, they are also independent.

6. (1p) Suppose that $\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}\right)$, where the covariance matrix is positive definite. Find $\mathbb{E}(X_1X_3 \mid X_2 = a)$.

Solution: The conditional distribution is

$$\begin{aligned}\begin{bmatrix}X_1\\X_3\end{bmatrix} \mid X_2 &\sim N\left(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{x}_2,\ \boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)\\ &= N\left(\begin{bmatrix}\sigma_{11}\\\sigma_{13}\end{bmatrix}\sigma_{22}^{-1}a,\ \begin{bmatrix}\sigma_{11} & \sigma_{13}\\ \sigma_{13} & \sigma_{33}\end{bmatrix}-\begin{bmatrix}\sigma_{11}\\\sigma_{13}\end{bmatrix}\sigma_{22}^{-1}\begin{bmatrix}\sigma_{21} & \sigma_{23}\end{bmatrix}\right)\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{E}\left(X_1X_3 \mid X_2\right) &= \operatorname{cov}\left(X_1,X_3 \mid X_2\right)+\mathbb{E}\left(X_1 \mid X_2=a\right)\mathbb{E}\left(X_3 \mid X_2=a\right)\\ &= \frac{\sigma_{11}\sigma_{21}}{\sigma_{22}}+\frac{\sigma_{11}\sigma_{13}a^2}{\sigma_{22}^2}\end{aligned}$$

7. (1p) Suppose that $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ is a random sample from a $N_p\left(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}\right)$ population, where $\boldsymbol{\mu}_0$ is known. Find the MLE of $\boldsymbol{\Sigma}$.
Solution: The log-likelihood function is

$$\ell = -\frac{np}{2}\log\left(2\pi\right)-\frac{n}{2}\log\det\left(\boldsymbol{\Sigma}\right)-\frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{x}_j-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_j-\boldsymbol{\mu}_0\right)^T\right\}$$

The MLE is $\hat{\boldsymbol{\Sigma}}=n^{-1}\sum_{j=1}^{n}\left(\boldsymbol{x}_j-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_j-\boldsymbol{\mu}_0\right)^T$, by the Lemma in the slides.