

HOMEWORK 1

(Due: December 14, 2020, Monday – 23.59)

- You should work on your own. Please feel free to get help from me, but not from anyone else. Let me know if my wording in the questions is not clear. Therefore, absolutely, no late homework will be accepted.
- **I will open a forum on ODTUClass. Please write your questions in this forum.**
- Please use R Markdown to do your homework. Then, produce a **word** or **html** file using it. If you have any question or problem, please let me know.
- You will submit your homework as a zip file including both your manual solution and R output file to ODTUClass.

1. Let, $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$

- a) Show that A is positive definite.
- b) Determine the eigenvalues and the normalized eigenvectors of A .
- c) Write the spectral decomposition of A .
- d) Find $A^{1/2}$.

2. Let X be $N_3(\mu, \Sigma)$ with $\mu' = [-3, 1, 4]$ and $\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$.

a) Are $\frac{X_1 + X_2}{2}$ and X_3 independent?

b) Find the distribution of $X_2 - \frac{5}{2}X_1 - X_3$.

c) Find the conditional distribution of X_2 , given that $X_1 = x_1$ and $X_3 = x_3$.

3. Please answer each part separately.

a) State whether the given information is true. If not please declare the mistake.

- For the detection of outliers, under the dimension $p > 1$, we can accept the

observation as outlier if $|Z_{ij}| > 1.5$ while $Z_{ij} = \frac{y_{ij} - \bar{y}}{s_i} \sim N(0,1)$ with

$$y_i \sim N(\mu, \sigma^2).$$

- The Wishart distribution with $(n - p)$ degrees of freedom is the generalization of the chi-square distribution when the dimension $p > 1$ and sample size n .
- A singular and symmetric $(p \times p)$ matrix Z can be represented as $Z = \lambda_1 e_1 e_1' + \dots + \lambda_p e_p e_p'$ where $\lambda_1, \dots, \lambda_p$ are eigenvalues of Z and e_1, \dots, e_p are the corresponding normalized eigenvectors of Z .
- If the p -dimensional random vector X has multivariate normal distribution, its marginals can deviate from univariate normal distribution.

b) If your multivariate data deviate from normality assumption, what would be your strategy to make an analysis from the available data? Only list your statements.

4. Please use **R Studio** in this question.

Consider **USairpollution** data set in **MVA** package. Then, please answer the following question.

- a)** Obtain mean vector and Var-Cov and Correlation matrix of the dataset. Please make a comment.
- b)** Calculate 5 number summaries of the variables in the dataset with a one line code. Interpret them.
- c)** Construct a scatterplot matrix of the data and comment on the result.
- d)** Please generate a research question and answer it with one appropriate visual tool.
- e)** Use the bivariate boxplot on the scatterplot of each pair of variables in the air pollution data to identify any outliers. Calculate the correlation between each pair of variables using all the data and the data with any identified outliers removed. Comment on the results.
- f)** Please check multivariate normality of the data by using both visual and formal ways. Don't forget to state your hypothesis.
- g)** Compare the chi-plots with the corresponding scatterplots for each pair of variables in the air pollution data. Do you think that there is any advantage in the former?
- h)** Identify the outliers via adjusted mahalanobis distance and compare the result with part e.