QUIZ 1: Applied Multivariate Analysis

Day / Date : Monday, 07-03-2022 from 10:45
Submitted on Monday, 07-03-2022 at 15.00 via iriawan.nur@gmail.com with the file name format as “Quiz1_AMA_NRP_Name.ZIP” or “Quiz1_AMA_NRP_Name.RAR”

Lecturer: Prof. Drs. NUR Iriawan, M.I.Kom., Ph.D.

---

1. **(Proportion 50%)** Give the answer as 'Right' or 'Wrong' to each of the questions below and provide an explanation of your answer choices, at least three lines.

A. UNIVARIATE NORMAL (UVN)

a. The difference in the location of the two data with UVN distribution can be tested using the hypothesis

$H_0: \mu_1 = \mu_2,$

$H_1: \mu_1 \neq \mu_2,$

and the statistics test is t-Student, that is $t_{df;2,5\%}$.

b. The difference in variance of the two data with UVN distribution can be tested using the hypothesis

$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1,$

$H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1,$

and the statistics test is $\chi^2_{df;5\%}$.

a. The failure of *Goodness of Fit* (GoF) *test* on the UVN data is only caused by the *skewness* of the data which is not zero.

b. *Goodness of Fit* (GoF) *test* on the normality of the data suspected of having UVN distribution can be done using a *t*-test.

c. Data normality test using the Kolmogorov-Smirnov method can be used the following hypothesis

$H_0: D_{max} = 0,$

$H_1: D_{max} \neq 0,$

where D_{max} is the statistics test of the greatest difference between the empirical cumulative distribution (CDF) of data and the normal distribution CDF.

d. The reference for rejection of the hypothesis in each test must be $\alpha = 5\%$.

e. Testing the similarity of the mean on as many as k data groups must be carried out using a hypothetical ANOVA $H_0: \mu_1=\mu_2=\ldots=\mu_k$ and $H_1: \mu_1\neq\mu_2\neq\ldots\neq\mu_k$.

B. MATRIKS

a. Suppose that we know the Variance-co-variance matrix, $\mathbf{\Sigma}$, from a database supposedly MVN distribution, name it as $\mathbf{X} \sim N_3(\boldsymbol{\mu}_X, \mathbf{\Sigma}_X)$ as follows

$$\mathbf{\Sigma} = \begin{bmatrix} 0,3 & 0,2 & 0,1 \\ a & 0,5 & 0,15 \\ b & c & 0,6 \end{bmatrix},$$

then its correlation matrix would be as follows

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & d & e \\ 0,5164 & 1 & f \\ 0,2357 & 0,2739 & 1 \end{bmatrix}$$

b. The value of 'c' in the Variance-co-variance matrix in problem 1.B.a. is $0{,}15$
c. By using the correct correlation matrix according to your answer in question 1.B.a., then the value of 'e' in the correlation matrix is $0{,}2357$.
d. No matter how many rows of data are entered (say a number of n data) in each field in a database containing 3 fields, both the Variance-co-variance matrix and the correlation matrix will be of the order $(n \times 3)$.
e. The eigenvalues of a Variance-co-variance matrix in problem 1.B.a. will consist of 3 different eigenvalues and if added together will be 3.
f. By using the correct correlation matrix according to your answer in question 1.B.a., the eigenvalues are as follows (1,7016; 0,8166; 0,4817).

C. MULTIVARIATE NORMAL (MVN)

a. A database table consisting of 3 fields, each of which is a continuous data type, and each field meets UVN properties, it is certain that the combination of all these fields will build an MVN.
b. Suppose that we have data as $\mathbf{X} \sim \text{MVN}(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ or $\mathbf{X} \sim N_p(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ is a matrix with size of $n \times p$ or n rows and p column, then each of the p vectors that compose $\mathbf{X}$ is UVN distribution.
c. Rejection of the NULL hypothesis in the GoF test for normality of data from one field of a database table with 10 fields, then all the database fields will only be multivariable data.
d. If the variance-co-variance matrix is known as in problem 1.B.a., then the database consists of 3 fields and the fields are correlated with each other, so that the data in the database table is only included as a multivariable category.
e. It is known that two tables in the database have the same dimensions with a matrix structure as $\mathbf{X} \sim N_p(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ and $\mathbf{Y} \sim N_p(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$. After testing the mean difference, with the hypothesis

   H_0: $\boldsymbol{\mu}_X = \boldsymbol{\mu}_Y$,

   H_1: $\boldsymbol{\mu}_X \neq \boldsymbol{\mu}_Y$,

   and the test state that H_0 is not rejected. It can be concluded that the two tables have data from the same population.

2. **(Proportion 15%)** Supposing that x_1 and x_2 state that x_1 = long tail of bird and x_2 = bird's wing length (both in mm). Suppose it is known

$$\mathbf{X} = (X_1, X_2) \sim N_2\left(\begin{pmatrix}190\\270\end{pmatrix}, \begin{pmatrix}115 & 100\\100 & 120\end{pmatrix}\right).$$

What is the distribution of the wingspan of a bird that has a tail length of 175 mm?

3. **(Proportion 10%)** Supposing that $\mathbf{X} \sim N_p(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ and $\mathbf{Y} \sim N_p(\mu_Y, \Sigma_Y)$ with $\mathbf{X}$ and $\mathbf{Y}$ are independent. What is the distribution of $\mathbf{X} - \mathbf{CY}$ where $\mathbf{C}$ is a constant matrix of $p \times p$ dimension?
4. **(Proportion 5%)** Supposing that $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as an i.i.d with $|\boldsymbol{\Sigma}|$ is given. Determine the distribution of $\overline{\mathbf{X}}$.

5. **(Proportion 20%)** Given a table with 4 fields from a database, name the fields as $\mathbf{X_1}$, $\mathbf{X_2}$, $\mathbf{X_3}$, and $\mathbf{X_4}$. There are 25-line data (see Table 1) already entered and stored in the table. The questions are
   a. If you have to test the data is MVN distributed or not, then write down the steps you will do.
   b. Test whether the data in the table has an MVN distribution? (YES / NO) Give your reasons.
   c. If your answer is in question 5.b. is 'YES', determine the mean and variance-co-variance of the MVN.
   (**Hint**: You are allowed to use the MINITAB or R or Python program packages).

Table 1: Data 4 fields in a table in a database.

no	x1	x2	x3	x4
1	9.886691	10.73365	11.54236	12.6119
2	10.13369	11.12835	5.547425	11.99222
3	10.50622	13.75112	11.29901	15.91793
4	12.43918	23.52018	10.98086	15.3776
5	8.235376	15.22296	7.69249	12.17379
6	7.60488	19.90873	7.941369	16.85299
7	12.99666	18.09095	3.527439	14.6378
8	11.83583	19.05017	-1.60881	17.62717
9	10.02178	11.46859	8.725119	8.817317
10	8.715971	11.1656	4.721776	9.479828
11	9.402685	17.19948	4.210214	18.78757
12	9.134275	14.73866	0.04253	12.27579
13	8.655501	15.29451	3.817219	16.56064
14	7.579231	10.7614	-1.61793	13.76621
15	10.7929	15.28713	6.836963	11.16098
16	9.951115	17.72267	1.019018	17.80411
17	11.87285	14.5267	5.936778	15.93847
18	9.91373	18.73225	14.73007	17.64422
19	7.245187	11.59925	10.8359	8.262903
20	9.580398	9.713942	2.700122	10.48844
21	10.14906	21.06418	7.003014	21.39829
22	10.84155	16.84561	4.333997	13.62105
23	13.75297	13.39317	1.127618	16.63828
24	12.135	19.49732	9.917877	19.95227
25	12.88637	10.86221	6.712351	16.1719