Some Exercises before MIDTERM
MATH 347

1.

a). To check the normality of a one-dimensional variable, we can use "Q-Q norm" plot (plot of quantiles of the data against quantiles of a standard normal distribution). How do you evaluate based on this plot whether the normality assumption is correct or not?

b). Similarly, to check the normality of multi-variate data, we may try "Q-Q $\chi^2$" plot. How do you evalute whether the normality assumption is correct or not based on this plot?

c). For one-dimensional variable with 10 observations, we compute the correlation of its quantiles with the quantiles of a standard normal distribution, denoted by $r_Q$, and find that $r_Q = .9250$. Will you reject the normality hypothesis at significance level 0.05? Will you reject the normality hypothesis at significance level 0.01?

2. Let $X = \binom{x_1}{x_2}$ be a two-variate normal random variable with mean $\mu = \binom{1}{0}$ and variance $\mathbf{\Sigma} = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}$.

a). Find the distribution of $2x_1 + 3x_2$.

b). Find the distribution of $(X - \mu)'\mathbf{\Sigma}^{-1}(X - \mu)$.

c). Find the conditional distribution of $x_1$ given $x_2 = 1$

3. It is known adult tigers living in north China have average body weight of 50 $kg$ and length of 120 $cm$. A zoologist wants to investigate similarities between tigers living in north China and south China. For this purpose, data on 10 wild adult south China tigers are collected and, using first component to denote body weight and second to denote body length, the sample average and variances are

$$\bar{X} = \begin{pmatrix} 45 \\ 130 \end{pmatrix}, \qquad S = \begin{pmatrix} 25 & 40 \\ 40 & 100 \end{pmatrix}$$

From this data, do you find strong evidence against the belief that tigers living in northern China and south China are of same species?

a). Use $T^2$ test to answer the above question.

b). Construct Bonferroni simultaneuous confidence intervals of 95% confidence level for the average body weight and the average body length of adult south China tigers.

4. We want to study whether the new airport is different from the old airport in terms of the pollutions caused by airplane traffic. The pollutions are measured by two

standard: air pollution and noise pollution. For simplicity, we consider the following data of three random samples collected from the old airport and three random samples collected from the new airport:

$$\text{The old airport}: \quad X_{11} = \begin{pmatrix} 10 \\ 2 \end{pmatrix} \quad X_{12} = \begin{pmatrix} 9 \\ 3 \end{pmatrix} \quad X_{13} = \begin{pmatrix} 11 \\ 2 \end{pmatrix};$$

$$\text{The new airport}: \quad X_{21} = \begin{pmatrix} 8 \\ 3 \end{pmatrix} \quad X_{22} = \begin{pmatrix} 9 \\ 4 \end{pmatrix} \quad X_{23} = \begin{pmatrix} 7 \\ 2 \end{pmatrix}.$$

Here the first component is air pollution index and the second is noise pollution index. Based on this data, what is your answer to the question? (Assume the variances of the sampled pollution are same for the two airports.)

5. To collect data on household income and expense, totally 150 households are randomly selected from three cities; SZ, SH and BJ, 50 from each. Their sample mean and variance from are:

$$\text{SZ}: \quad \bar{X}_1 = \begin{pmatrix} 90 \\ 40 \end{pmatrix} \quad S_1^2 = \begin{pmatrix} 25 & 0 \\ 0 & 25 \end{pmatrix}$$

$$\text{SH}: \quad \bar{X}_2 = \begin{pmatrix} 100 \\ 50 \end{pmatrix} \quad S_2^2 = \begin{pmatrix} 16 & 9 \\ 9 & 9 \end{pmatrix}$$

$$\text{BJ}: \quad \bar{X}_3 = \begin{pmatrix} 80 \\ 60 \end{pmatrix} \quad S_3^2 = \begin{pmatrix} 19 & 1 \\ 1 & 16 \end{pmatrix}$$

Based on this data, is there significance evidence indicating the residents of the three cities have difference household incomes and expenses?

**SOLUTIONS**

1. a). close to a straight line $\Longleftrightarrow$ normality.

close to a straight line through origin $\Longleftrightarrow$ normality with mean 0

close to a straight line at 45° through origin $\Longleftrightarrow$ standard normality.

b) close to a straight line at at 45° through origin $\Longleftrightarrow$ normality. (Note that the quantiles are based on $(x_i - \bar{x})'S^{-1}(x_i - \bar{x}), 1 \le i \le n$.

c) From table 4.2, we reject the normality at significance level 0.05 if $r_Q < 0.9198$. Since $r_Q = .925 > 0.9198$, we should accept the normality assumption at significance level 0.05. Likewise, we should also accept the normality assumption at significance level 0.01.

d) We should use B-C transformation with $\lambda < 1$.

3. a) $N(2, 28)$

b) $\chi_2^2$-distribution. The mean is 2.

c) mean is $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) = 3/4$. variance is $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = 3/4$.

3. a). $H_o : \mu = \binom{50}{120}$ vs. $H_a : \mu \neq \binom{50}{120}$.

$T^2 = n(\bar{x} - \mu_0)'S^{-1}(\bar{x} - \mu_0) = 100$. critical value at $\alpha = .05$ is $(n-1)p/(n-p)F_{p,n-p}(\alpha) = 10.035$. where $n = 10, p = 2$. So $T^2 = 100 > 10.035$, we should reject the null hypothesis at level 0.05 and conclude that the south China tigers are of different species from the north China tigers.

b). $t_{n-1}(\alpha/(2p)) = t_9(0.05/(2 \times 2)) = 2.821$. Then 95% Bonferrroni simultaneous C.I.s for $\mu_i$ is $\bar{x}_i \pm t_{n-1}(\alpha/(2p))\sqrt{s_{ii}/n}$:

for $\mu_1 : (40.54, 49.46)$ and for $\mu_2 : (121.08, 138.92)$.

4

$$\text{For old airport :} \quad \bar{X}_1 = \begin{pmatrix} 10 \\ 7/3 \end{pmatrix} \quad S_1 = \begin{pmatrix} 1 & -.5 \\ -.5 & 1/3 \end{pmatrix}$$

$$\text{For new airport :} \quad \bar{X}_2 = \begin{pmatrix} 8 \\ 3 \end{pmatrix} \quad S_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$S_{pooled} = \frac{n_1 - 1}{n_1 + n_2 - 2} S_1 + \frac{n_2 - 1}{n_1 + n_2 - 2} S_2 = \begin{pmatrix} 1 & 1/4 \\ 1/4 & 2/3 \end{pmatrix}$$

And the $T^2$ test of $\mu_1 = \mu_2$ v.s $H_a : \mu_1 \neq \mu_2$ is

$$T^2 = (\bar{x}_1 - \bar{x}_2)' \Big[ (\frac{1}{n_1} + \frac{1}{n_2}) S_{pooled} \Big]^{-1} (\bar{x}_1 - \bar{x}_2)$$
$$= (2 \quad -2/3) \Big[ 2/3 \begin{pmatrix} 1 & 1/4 \\ 1/4 & 2/3 \end{pmatrix} \Big]^{-1} \begin{pmatrix} 2 \\ -2/3 \end{pmatrix} = 9.38$$

The critical value is

$$\frac{(n_1+n_2-2)p}{n_1+n_2-1-p}F_{p,n_1+n_2-1-p}(\alpha) = \frac{3+3-2)2}{3+3-1-2}F_{2,3}(.05) = 25.5$$

Since $T^2 = 9.38 < 25.5$, we accept $H_0$ and conclude that the pollution of the two airports are about same.

5. $\bar{X} = \binom{90}{50}$ and

$$SS_{between} = 50\Big[\begin{pmatrix} 0 \\ -10 \end{pmatrix}(0\ -10)+\begin{pmatrix} 10 \\ 0 \end{pmatrix}(10\ 0)+\begin{pmatrix} -10 \\ 10 \end{pmatrix}(-10\ 10)\Big] = 50\begin{pmatrix} 200 & -100 \\ -100 & 200 \end{pmatrix}$$

$$SS_{within} = 49\begin{pmatrix} 25+16+19 & 0+9+1 \\ 0+9+1 & 25+9+16 \end{pmatrix}$$

And variance decomposition identity:

$$SS_{total} = SS_{between} + SS_{within}$$

$$\begin{pmatrix} 12940 & -4510 \\ -4510 & 12450 \end{pmatrix} = \begin{pmatrix} 10000 & -5000 \\ -5000 & 10000 \end{pmatrix} + \begin{pmatrix} 2940 & 490 \\ 490 & 2450 \end{pmatrix}$$

$$\mathbf{B}+\mathbf{W} = \mathbf{B}+\mathbf{W}$$

The MANOVA table: ($g = 3$, $p = 2$, $n_1 = n_2 = n_2 = 50$)

| *Source.of. .variation* | *Sum.of. .Squares* | *degree.of. freedom.* | $\Lambda^*$ |
|---|---|---|---|
| *Treatment* | $\mathbf{B}$ | $3-1=2$ | $\frac{\lvert\mathbf{W}\rvert}{\lvert\mathbf{B}+\mathbf{W}\rvert} = \frac{6962900}{140762900} = 0.0495$ |
| *Error* | $\mathbf{W}$ | $150-3=147$ | |
| *Total* | $\mathbf{B}+\mathbf{W}$ | | |

Then

$$-(n-1-(p+g)/2)\log\Lambda^* = -(150-1-(2+3)/2)log(0.0495) = 440.46$$

On the other hand, $\chi^2_{p(g-1)}(0.05) = \chi^2_4(0.05) = 9.487$ which is much small than 440.46. We should reject the null hypothesis and conclude that there exists a difference in terms of high school academic performaces for students admitted to the three universities.

Remark: Table 6.3 and table 4.2 will be provided if they are needed in exams.