STATISTICS-BUSINESS 756 (Johnson)
PRACTICE EXAM - March 2004

1. Given the $n = 4$ observations

$$\mathbf{X} = \begin{bmatrix} 1 & 7 & 5 & 7 \\ 3 & 6 & 4 & 3 \end{bmatrix}$$

   (a) Evaluate the T^2 statistic and carry out the test of

$$H_0 : \boldsymbol{\mu} = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \qquad H_1 : \boldsymbol{\mu} \neq \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

   (b) Specify the 95 % confidence region for $\boldsymbol{\mu}$. Evaluate the axes of the ellipsoid and specify their half lengths.

2. Let $n = 6$ bivariate observations have the values

$$\boldsymbol{x}_1 = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \quad \boldsymbol{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \boldsymbol{x}_3 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad \boldsymbol{x}_4 = \begin{bmatrix} 1 \\ 5 \end{bmatrix} \quad \boldsymbol{x}_5 = \begin{bmatrix} -2 \\ 2 \end{bmatrix} \quad \boldsymbol{x}_6 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

   A computer calculation gives

$$\bar{\boldsymbol{x}} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \text{ and } \boldsymbol{S} = \begin{bmatrix} 3.2 & 1 \\ 1 & 2.4 \end{bmatrix}$$

   (a) Find sample mean of $y_j = x_{1j} - x_{2j}$ for $j = 1, 2, 3, 4, 5, 6$.

   (b) Find sample variance of the y_j.

   (c) Find the sample covariance between y_j and $w_y = x_{1j} + x_{2j}$.

3. Let $\boldsymbol{X}_1, ..., \boldsymbol{X}_{50}$ be a random sample of size 50 from a $p-$variate normal distribution having mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Specify completely the distribution of

   (a) $\bar{\boldsymbol{X}}$

   (b) $(\boldsymbol{X}_1 - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu})$

   (c) $n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$

   (d) $n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$

4. Data on gasoline trucks on x_1 fuel, x_2 repair, and x_3 capital costs per mile are given in the table below along with the standardized variables and d^2 distances.

   Use these values and the normal scores and d^2 plot to comment on normality and possible outliers.

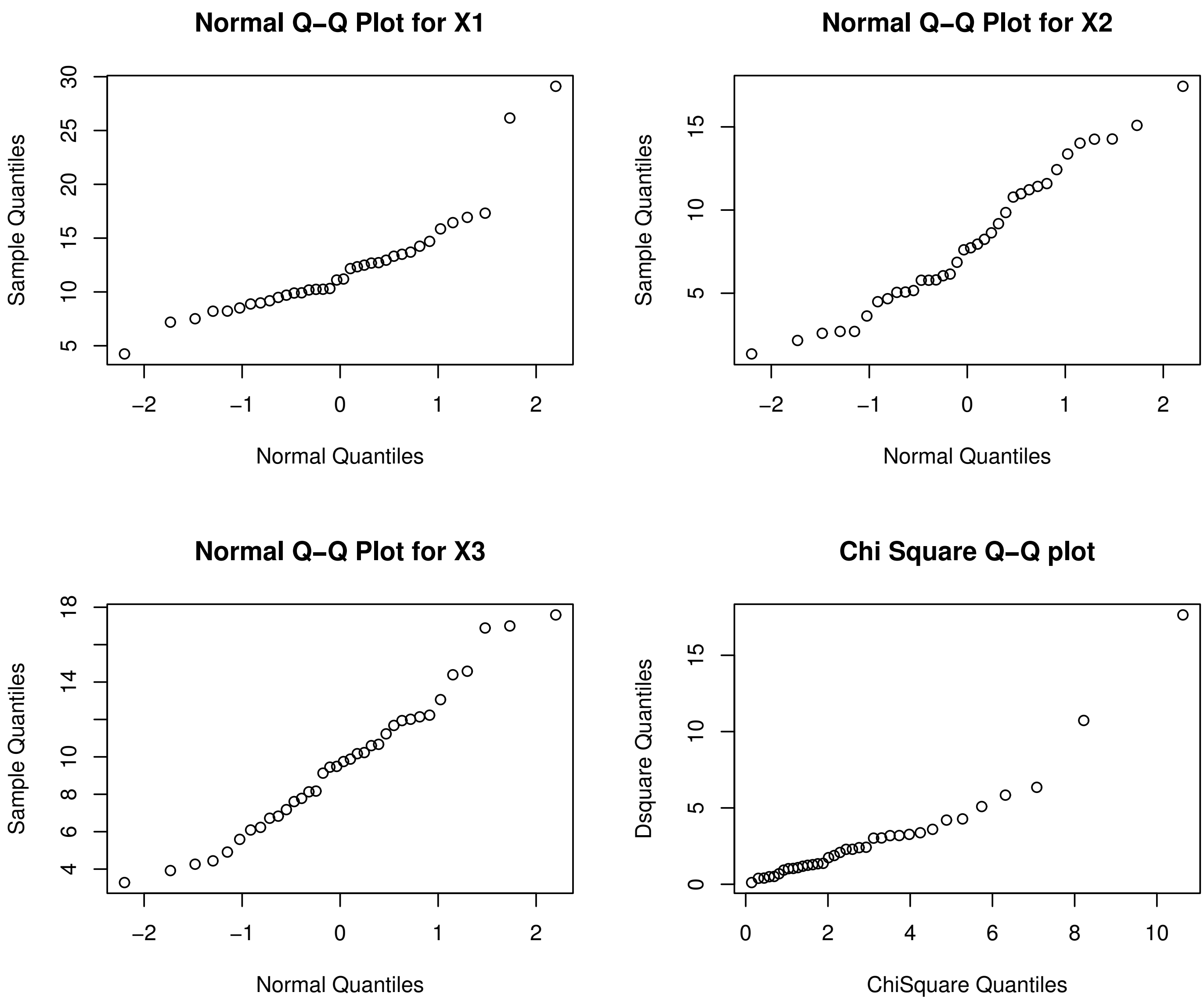

Figure 1: Q-Q Plots

5. Observations on the length of time since Bachelor's degree and current salary were obtained on men and women graduates of a program. The data for men are

$$n_1 = 50, \quad \overline{\boldsymbol{x}}_1 = \begin{bmatrix} 8 \\ 22 \end{bmatrix} \text{ and } \boldsymbol{S}_1 = \begin{bmatrix} 4 & 1 \\ 1 & 10 \end{bmatrix}$$

and those for women are

$$n_2 = 50, \quad \overline{\boldsymbol{x}}_2 = \begin{bmatrix} 7 \\ 25 \end{bmatrix} \text{ and } \boldsymbol{S}_2 = \begin{bmatrix} 6 & 3 \\ 3 & 4 \end{bmatrix}$$

(a) Give the Bonferroni simultaneous 95% confidence intervals for the mean differences.

[Hint: $t_{98}(.0125) = 2.2764$ and $t_{98}(.025) = 1.9845$]

(b) Give the simultaneous 95% confidence intervals, that hold for all linear combinations, specialized to the mean differences.

(c) Evaluate T^2 to test for equality. Use $\alpha = .05$.

6. Given the sample covariance matrix $\boldsymbol{S} = \begin{bmatrix} 6 & 2 \\ 2 & 9 \end{bmatrix}$:

(a) Determine the first principal component.

(b) Evaluate the first principal component for the observation $\boldsymbol{x} = [4, 2]$.

(c) Find the proportion of variance explained by the second principal component.

(d) Find the sample correlation between first component x_1 and the first principal component.

(e) Given the sample mean vector $\overline{\boldsymbol{x}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ graph the new axes for the scatter plot based on the principal components. (Since data are not given, just draw the original and the new axes.)