Chapter 4

4.1 (a) We are given $p = 2$, $\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, $\boldsymbol{\Sigma} = \begin{bmatrix} 2 & -.8 \times \sqrt{2} \\ -.8 \times \sqrt{2} & 1 \end{bmatrix}$ so $|\,\boldsymbol{\Sigma}\,| = .72$ and

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \frac{1}{.72} & \frac{\sqrt{2}}{.9} \\ \frac{\sqrt{2}}{.9} & \frac{2}{.72} \end{bmatrix}$$

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)\sqrt{.72}} \exp\left(-\frac{1}{2} \left[\frac{1}{.72}(x_1 - 1)^2 + \frac{2\sqrt{2}}{.9}(x_1 - 1)(x_2 - 3) + \frac{2}{.72}(x_2 - 3)^2 \right] \right)$$

(b)

$$\frac{1}{.72}(x_1 - 1)^2 + \frac{2\sqrt{2}}{.9}(x_1 - 1)(x_2 - 3) + \frac{2}{.72}(x_2 - 3)^2$$

4.2 (a) We are given $p = 2$, $\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$, $\boldsymbol{\Sigma} = \begin{bmatrix} 2 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1 \end{bmatrix}$ so $|\,\boldsymbol{\Sigma}\,| = 3/2$

and

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \frac{2}{\sqrt{3}} & -\frac{\sqrt{2}}{3} \\ -\frac{\sqrt{2}}{3} & \frac{2}{\sqrt{3}} \end{bmatrix}$$

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)\sqrt{3/2}} \exp\left(-\frac{1}{2} \left[\frac{2}{3}x_1^2 - \frac{2\sqrt{2}}{3}x_1(x_2 - 2) + \frac{4}{3}(x_2 - 2)^2 \right] \right)$$

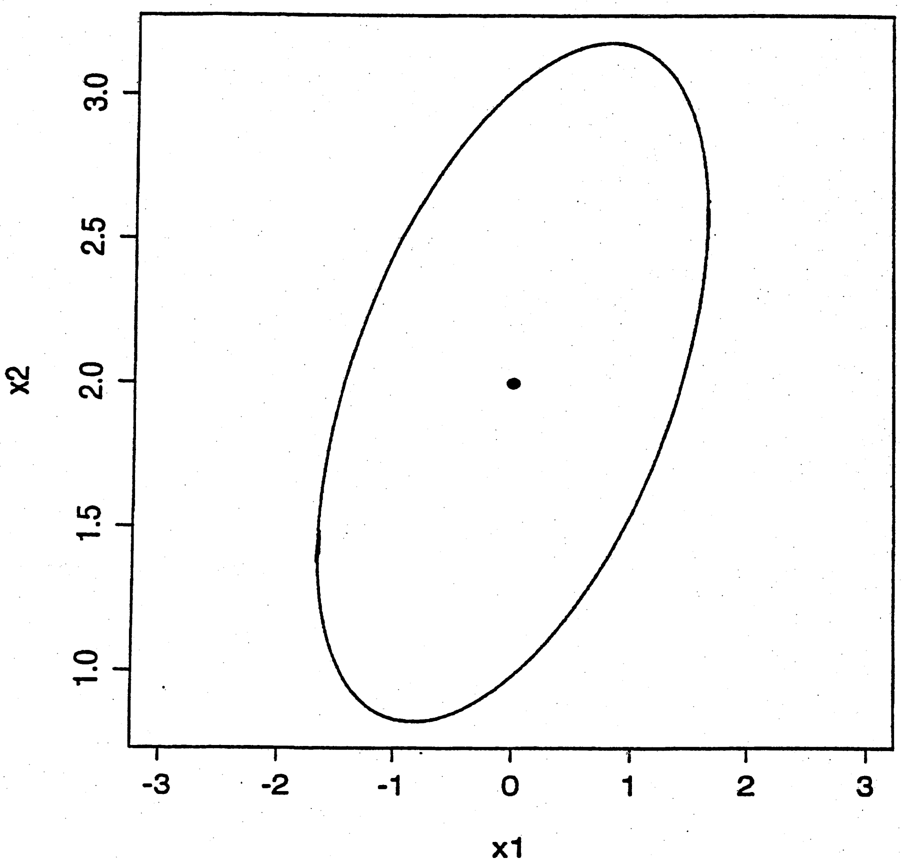
(b)

$$\frac{2}{3}x_1^2 - \frac{2\sqrt{2}}{3}x_1(x_2 - 2) + \frac{4}{3}(x_2 - 2)^2$$

(c) $c^2 = \chi_2^2(.5) = 1.39$. Ellipse centered at $[0, 2]'$ with the major axis having half-length $\sqrt{\lambda_1}\, c = \sqrt{2.366}\sqrt{1.39} = 1.81$. The major axis lies in the direction $\boldsymbol{e} = [.888, .460]'$. The minor axis lies in the direction $\boldsymbol{e} = [-.460\,, .888]'$ and has half-length $\sqrt{\lambda_2}\, c = \sqrt{.634}\sqrt{1.39} = .94$.

Constant density contour that contains

50% of the probability



4.3 We apply Result 4.5 that relates zero covariance to statistical independence

a) No, $\sigma_{12} \neq 0$

b) Yes, $\sigma_{23} = 0$

c) Yes, $\sigma_{13} = \sigma_{23} = 0$

d) Yes, by Result 4.3, $(X_1+X_2)/2$ and X_3 are jointly normal and their covariance is $\frac{1}{2}\sigma_{13}+\frac{1}{2}\sigma_{23} = 0$.

e) No, by Result 4.3 with $A = \begin{bmatrix} 0 & 1 & 0 \\ -\frac{5}{2} & 1 & -1 \end{bmatrix}$, form $A \Sigma A'$ to see that the covariance is 10 and not 0.

4.4 a) $3X_1 - 2X_2 + X_3$ is $N(13,9)$

b) Require Cov $(X_2, X_2 - a_1X_1 - a_3X_3) = 3 - a_1 - 2a_3 = 0$. Thus any $\underset{\sim}{a}' = [a_1, a_3]$ of the form $\underset{\sim}{a}' = [3-2a_3, a_3]$ will meet the requirement. As an example, $\underset{\sim}{a}' = [1,1]$.

4.5 a) $X_1|x_2$ is $N(\frac{1}{\sqrt{2}}(x_2-2), \frac{3}{2})$

b) $X_2|x_1, x_3$ is $N(-2x_1-5, 1)$

c) $X_3|x_1, x_2$ is $N(\frac{1}{2}(x_1+x_2+3), \frac{1}{2})$

4.6 (a) X_1 and X_2 are independent since they have a bivariate normal distribution with covariance $\sigma_{12} = 0$.

(b) X_1 and X_3 are dependent since they have nonzero covariance $\sigma_{13} = -1$.

(c) X_2 and X_3 are independent since they have a bivariate normal distribution with covariance $\sigma_{23} = 0$.

(d) X_1, X_3 and X_2 are independent since they have a trivariate normal distribution where $\sigma_{12} = 0$ and $\sigma_{32} = 0$.

(e) X_1 and $X_1 + 2X_2 - 3X_3$ are dependent since they have nonzero covariance

$$\sigma_{11} + 2\sigma_{12} - 3\sigma_{13} = 4 + 2(0) - 3(-1) = 7$$

4.7 (a) $X_1|x_3$ is $N(1 + .5(x_3 - 2)\ , 3.5)$

(b) $X_1|x_2, x_3$ is $N(1 + .5(x_3 - 2)\ , 3.5)$. Since X_2 is independent of X_1, conditioning further on x_2 does not change the answer from Part a).

4.15 First,

$$\sum_{j=1}^{n} (\bar{\mathbf{x}}-\boldsymbol{\mu})(\mathbf{x}_j-\bar{\mathbf{x}})' = (\bar{\mathbf{x}}-\boldsymbol{\mu})[\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})']$$

$$= (\bar{\mathbf{x}}-\boldsymbol{\mu})(\sum_{j=1}^{n} \mathbf{x}_j-n\bar{\mathbf{x}})'$$

$$= (\bar{\mathbf{x}}-\boldsymbol{\mu})(n\bar{\mathbf{x}}-n\bar{\mathbf{x}})'$$

$$= 0$$

Also,

$$\sum_{j=1}^{n} (\mathbf{x}_j-\bar{\mathbf{x}})(\bar{\mathbf{x}}-\boldsymbol{\mu})' = [\sum_{j=1}^{n}(\bar{\mathbf{x}}-\boldsymbol{\mu})(\mathbf{x}_j-\bar{\mathbf{x}})']' = 0' = 0.$$

4.16 (a) By Result 4.8, with $c_1 = c_3 = 1/4$, $c_2 = c_4 = -1/4$ and $\boldsymbol{\mu}_j = \boldsymbol{\mu}$ for $j = 1, ..., 4$ we have $\sum_{j=1}^{4} c_j\boldsymbol{\mu}_j = 0$ and $(\sum_{j=1}^{4} c_j^2)\, \Sigma = \frac{1}{4}\Sigma$. Consequently, $\boldsymbol{V}_1$ is $N(0, \frac{1}{4}\Sigma)$. Similarly, setting $b_1 = b_2 = 1/4$ and $b_3 = b_4 = -1/4$, we find that $\boldsymbol{V}_2$ is $N(0, \frac{1}{4}\Sigma)$.

(b) Again by Result 4.8, we know that $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are jointly multivariate normal with covariance

$$(\sum_{j=1}^{4} b_j c_j)\, \Sigma = \left(\frac{1}{4}(\frac{1}{4}) + \frac{-1}{4}(\frac{1}{4}) + \frac{1}{4}(\frac{-1}{4}) + \frac{-1}{4}(\frac{-1}{4}) \right) \Sigma = 0$$

That is,

$$\begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \end{bmatrix} \quad \text{is distributed} \quad N_{2p}\left(0, \begin{bmatrix} \frac{1}{4}\Sigma & 0 \\ 0 & \frac{1}{4}\Sigma \end{bmatrix}\right)$$

so the joint density of the $2p$ variables is

$$f(\boldsymbol{v}_1, \boldsymbol{v}_2) = \frac{1}{(2\pi)^p |\frac{1}{4}\Sigma|} \exp\left(-\frac{1}{2}[\, \boldsymbol{v}_1', \boldsymbol{v}_2' \,] \begin{bmatrix} \frac{1}{4}\Sigma & 0 \\ 0 & \frac{1}{4}\Sigma \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_1 \end{bmatrix} \right)$$

$$= \frac{1}{(2\pi)^p |\frac{1}{4}\Sigma|} \exp\left(-\frac{1}{8}(\boldsymbol{v}_1' \Sigma^{-1}\boldsymbol{v}_1 + \boldsymbol{v}_2' \Sigma^{-1}\boldsymbol{v}_2) \right)$$

4.17 By Result 4.8, with $c_1 = c_2 = c_3 = c_4 = c_5 = 1/5$ and $\boldsymbol{\mu}_j = \boldsymbol{\mu}$ for $j = 1, ..., 5$ we find that $\boldsymbol{V}_1$ has mean $\sum_{j=1}^{5} c_j\boldsymbol{\mu}_j = \boldsymbol{\mu}$ and covariance matrix $(\sum_{j=1}^{5} c_j^2)\, \Sigma = \frac{1}{5}\Sigma$.

Similarly, setting $b_1 = b_3 = b_5 = 1/5$ and $b_2 = b_4 = -1/5$ we find that $\boldsymbol{V}_2$ has mean $\sum_{j=1}^{5} b_j\boldsymbol{\mu}_j = \frac{1}{5}\boldsymbol{\mu}$ and covariance matrix $(\sum_{j=1}^{5} b_j^2)\, \Sigma = \frac{1}{5}\Sigma$.

Again by Result 4.8, we know that $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ have covariance

$$(\sum_{i=1}^{4} b_j c_j)\, \Sigma = \left(\frac{1}{5}(\frac{1}{5}) + \frac{-1}{5}(\frac{1}{5}) + \frac{1}{5}(\frac{1}{5}) + \frac{-1}{5}(\frac{1}{5}) + \frac{1}{5}(\frac{1}{5}) \right) \Sigma = \frac{1}{25}\Sigma$$

4.18 By Result 4.11 we know that the maximum likelihood estimates of $\underset{\sim}{\mu}$ and Σ are $\underset{\sim}{\bar{x}} = [4,6]'$ and

$$\frac{1}{n}\sum_{j=1}^{n}(\underset{\sim}{x}_j-\underset{\sim}{\bar{x}})(\underset{\sim}{x}_j-\underset{\sim}{\bar{x}})' = \frac{1}{4}\left\{\left(\begin{bmatrix}3\\6\end{bmatrix}-\begin{bmatrix}4\\6\end{bmatrix}\right)\left(\begin{bmatrix}3\\6\end{bmatrix}-\begin{bmatrix}4\\6\end{bmatrix}\right)'+\left(\begin{bmatrix}4\\4\end{bmatrix}-\begin{bmatrix}4\\6\end{bmatrix}\right)\left(\begin{bmatrix}4\\4\end{bmatrix}-\begin{bmatrix}4\\6\end{bmatrix}\right)'\right.$$

$$\left.+\left(\begin{bmatrix}5\\7\end{bmatrix}-\begin{bmatrix}4\\6\end{bmatrix}\right)\left(\begin{bmatrix}5\\7\end{bmatrix}-\begin{bmatrix}4\\6\end{bmatrix}\right)'+\left(\begin{bmatrix}4\\7\end{bmatrix}-\begin{bmatrix}4\\6\end{bmatrix}\right)\left(\begin{bmatrix}4\\7\end{bmatrix}-\begin{bmatrix}4\\6\end{bmatrix}\right)'\right\}$$

$$= \frac{1}{4}\left\{\begin{bmatrix}-1\\0\end{bmatrix}\begin{bmatrix}-1 & 0\end{bmatrix}+\begin{bmatrix}0\\-2\end{bmatrix}\begin{bmatrix}0 & -2\end{bmatrix}+\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix}+\begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}0 & 1\end{bmatrix}\right\}$$

$$= \frac{1}{4}\begin{bmatrix}2 & 1\\1 & 6\end{bmatrix}$$

4.19 a) By Result 4.7 we know that $(\underset{\sim}{X}_1-\underset{\sim}{\mu})'\,\Sigma^{-1}(\underset{\sim}{X}_1-\underset{\sim}{\mu}) \sim \chi^2_6$

b) From (4-23), $\underset{\sim}{\bar{X}} \sim N_6(\underset{\sim}{\mu},\frac{1}{20}\Sigma)$. Then $\underset{\sim}{\bar{X}}-\underset{\sim}{\mu} \sim N_6(\underset{\sim}{0},\frac{1}{20}\Sigma)$ and finally $\sqrt{20}\,(\underset{\sim}{\bar{X}}-\underset{\sim}{\mu}) \sim N_6(\underset{\sim}{0},\Sigma)$

c) From (4-23), 19S has a Wishart distribution with 19 d.f.

4.20 $B(19S)B'$ is a 2×2 matrix distributed as $W_{19}(\cdot\,|\,B\Sigma B')$ with 19 d.f. where

a) $B\Sigma B'$ has

(1,1) entry $= \sigma_{11} + \frac{1}{4}\sigma_{22} + \frac{1}{4}\sigma_{33} - \sigma_{12} - \sigma_{13} + \frac{1}{2}\sigma_{23}$

(1,2) entry $= -\frac{1}{2}\sigma_{14} + \frac{1}{4}\sigma_{24} + \frac{1}{4}\sigma_{34} - \frac{1}{2}\sigma_{15} + \frac{1}{4}\sigma_{25} + \frac{1}{4}\sigma_{35} + \sigma_{16} - \frac{1}{2}\sigma_{26} - \frac{1}{2}\sigma_{36}$

(2,2) entry $= \sigma_{66} + \frac{1}{4}\sigma_{55} + \frac{1}{4}\sigma_{44} - \sigma_{46} - \sigma_{56} + \frac{1}{2}\sigma_{45}$

b)

$$B\Sigma B' = \begin{bmatrix}\sigma_{11} & \sigma_{13}\\ \sigma_{31} & \sigma_{33}\end{bmatrix}$$

4.21 (a) $\overline{\boldsymbol{X}}$ is distributed $N_4(\boldsymbol{\mu}, n^{-1}\boldsymbol{\Sigma})$

(b) $\boldsymbol{X}_1 - \boldsymbol{\mu}$ is distributed $N_4(\mathbf{0}, \boldsymbol{\Sigma})$ so $(\boldsymbol{X}_1 - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu})$ is distributed as chi-square with p degrees of freedom.

(c) Using Part a),

$$(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'(n^{-1}\boldsymbol{\Sigma})^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) = n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})$$

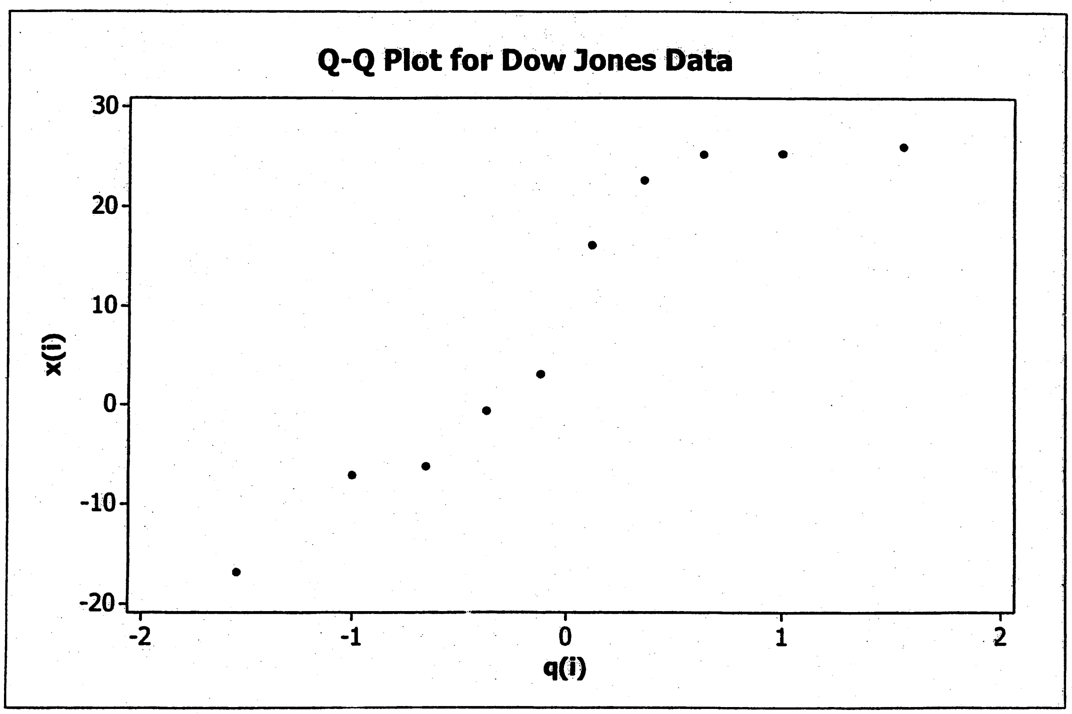
is distributed as chi-square with p degrees of freedom.

(d) Approximately distributed as chi-square with p degrees of freedom. Since the sample size is large, $\boldsymbol{\Sigma}$ can be replaced by S.

4.22 a) We see that $n = 75$ is a sufficiently large sample (compared with p) and apply Result 4.13 to get $\sqrt{n}(\underset{\sim}{X}-\underset{\sim}{\mu})$ is approximately $N_p(\underset{\sim}{0},\Sigma)$ and that $\bar{\underset{\sim}{X}}$ is approximately $N_p(\underset{\sim}{\mu},\frac{1}{n}\Sigma)$.

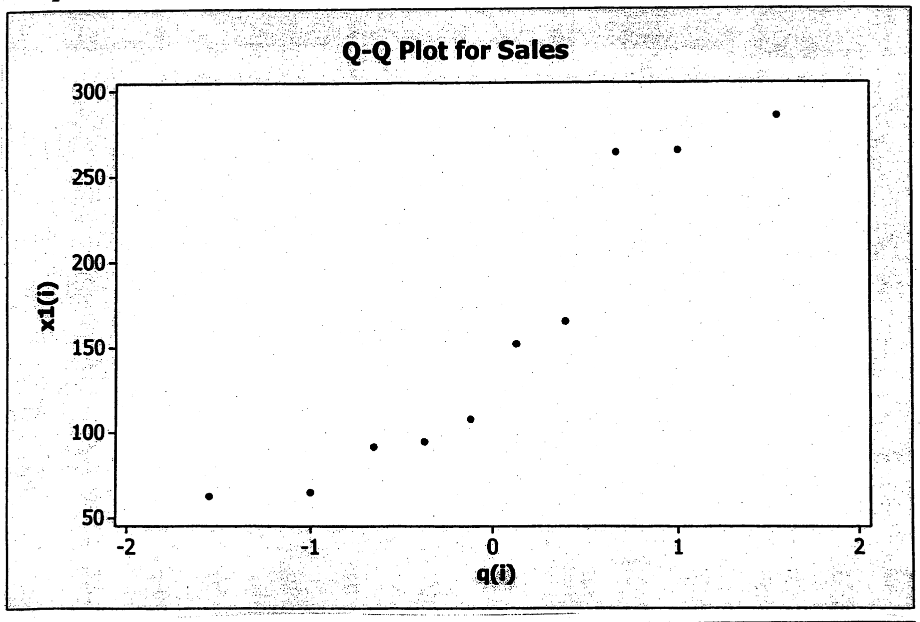
b) By (4-28) we conclude that $\sqrt{n}(\underset{\sim}{X}-\underset{\sim}{\mu})'S^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})$ is approximately χ_p^2.

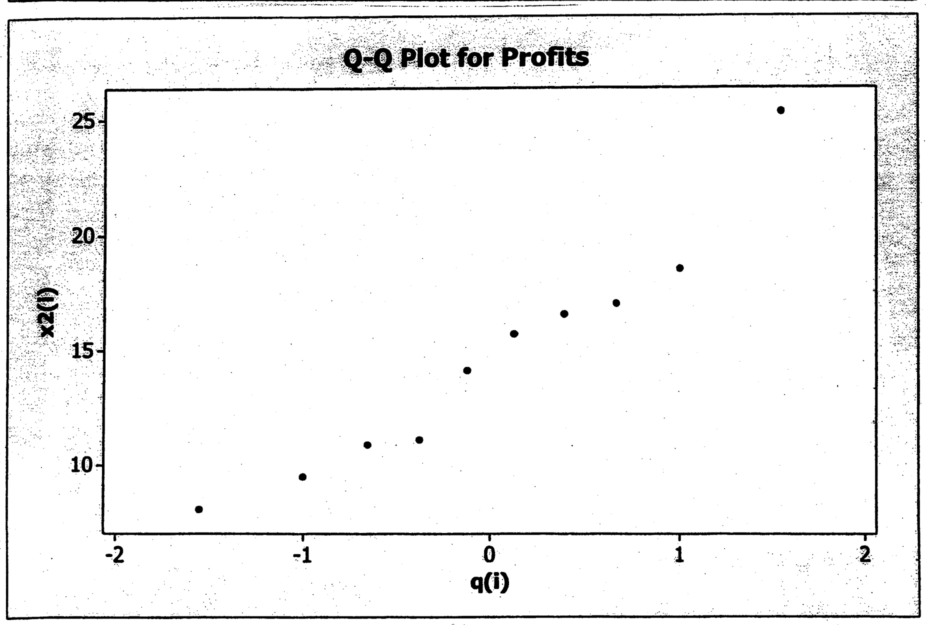
4.23 **(a)** The *Q-Q* plot shown below is not particularly straight, but the sample size $n = 10$ is small. Difficult to determine if data are normally distributed from the plot.



(b) $r_Q = .95$ and $n = 10$. Since $r_Q = .95 > .9351$ (see Table 4.2), cannot reject hypothesis of normality at the 10% level.

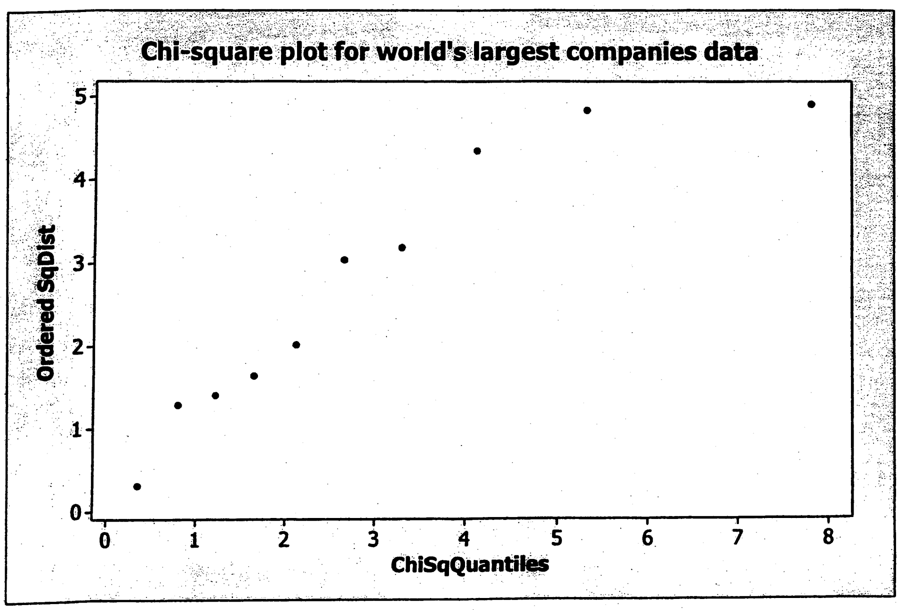
4.24 **(a)** *Q-Q* plots for sales and profits are given below. Plots not particularly straight, although *Q-Q* plot for profits appears to be "straighter" than plot for sales. Difficult to assess normality from plots with such a small sample size ($n = 10$).





(b) The critical point for $n = 10$ when $\alpha = .10$ is .9351. For sales, $r_Q = .940$ and for profits, $r_Q = .968$. Since the values for both of these correlations are greater than .9351, we cannot reject normality in either case.

4.25 The chi-square plot for the world's largest companies data is shown below. The plot is reasonably straight and it would be difficult to reject multivariate normality given the small sample size of $n = 10$. Information leading to the construction of this plot is also displayed.



$$\bar{\mathbf{x}} = \begin{bmatrix} 155.6 \\ 14.7 \\ 710.9 \end{bmatrix} \qquad \mathbf{S} = \begin{bmatrix} 7476.5 & 303.6 & -35576 \\ 303.6 & 26.2 & -1053.8 \\ -35576 & -1053.8 & 237054 \end{bmatrix}$$

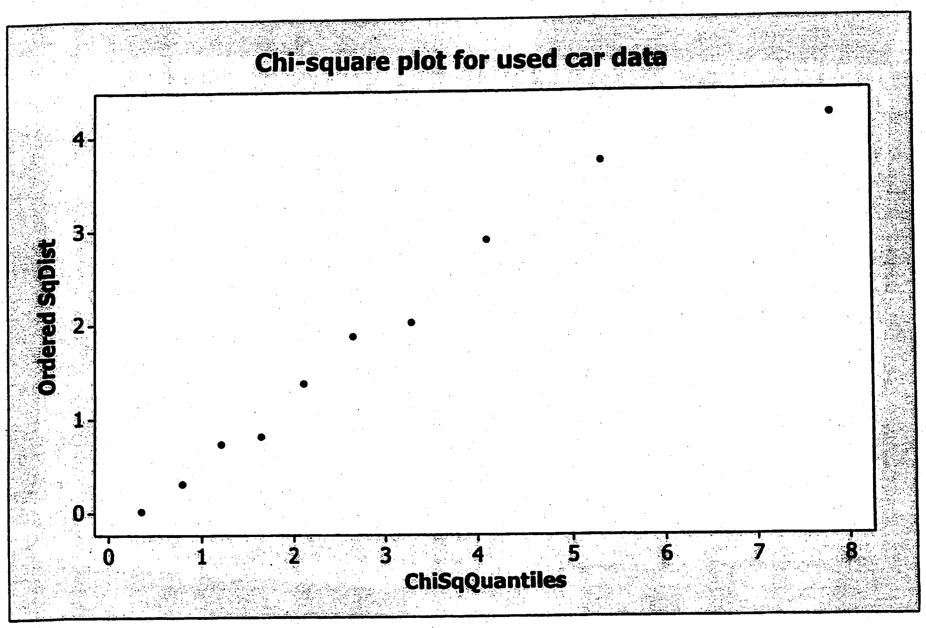
Ordered SqDist	Chi-square Quantiles
.3142	.3518
1.2894	.7978
1.4073	1.2125
1.6418	1.6416
2.0195	2.1095
3.0411	2.6430
3.1891	3.2831
4.3520	4.1083
4.8365	5.3170
4.9091	7.8147

4.26 **(a)** $\bar{\mathbf{x}} = \begin{bmatrix} 5.20 \\ 12.48 \end{bmatrix}$, $\mathbf{S} = \begin{bmatrix} 10.6222 & -17.7102 \\ -17.7102 & 30.8544 \end{bmatrix}$, $\mathbf{S}^{-1} = \begin{bmatrix} 2.1898 & 1.2569 \\ 1.2569 & .7539 \end{bmatrix}$

Thus d_j^2 = 1.8753, 2.0203, 2.9009, .7353, .3105, .0176, 3.7329, .8165, 1.3753, 4.2153

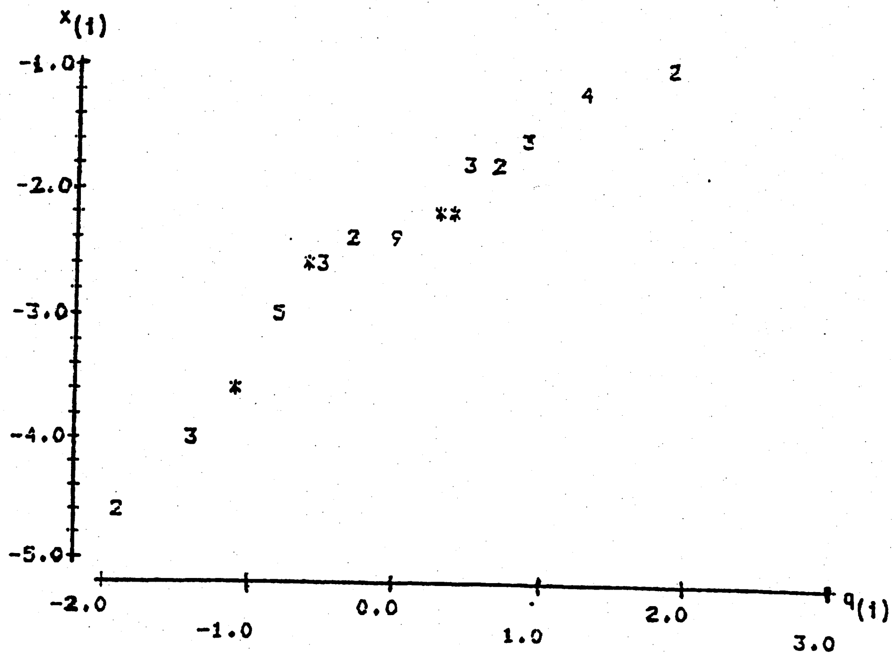
(b) Since $\chi_2^2(.5) = 1.39$, 5 observations (50%) are within the 50% contour.

(c) The chi-square plot is shown below.



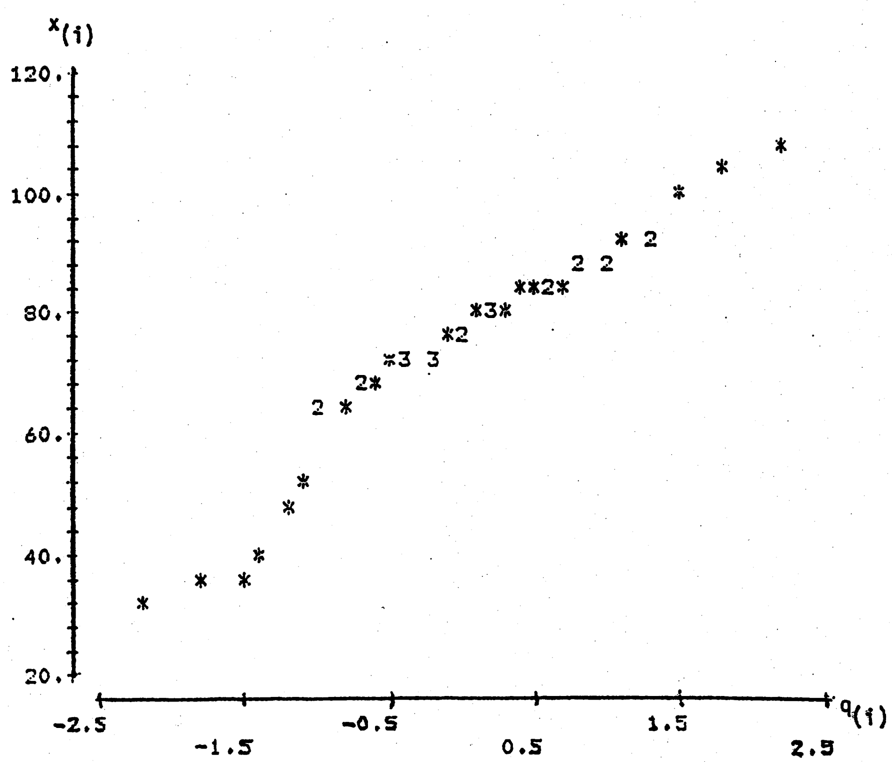
(d) Given the results in parts (b) and (c) and the small number of observations (n = 10), it is difficult to reject bivariate normality.

4.27 Q-Q plot is shown below.



The Q-Q plot is reasonably straight. $r_Q = .978$ (λ=0)
For λ = 1/4, $r_Q = .993$ so λ = 1/4 is a little better choice for the normalizing transformation.

4.28 Q-Q plot is shown below.



Since r_Q = .970 < .973 (See Table 4.2 for n = 40 and α = .05), we would reject the hypothesis of normality at the 5% level.

4.29

(a).

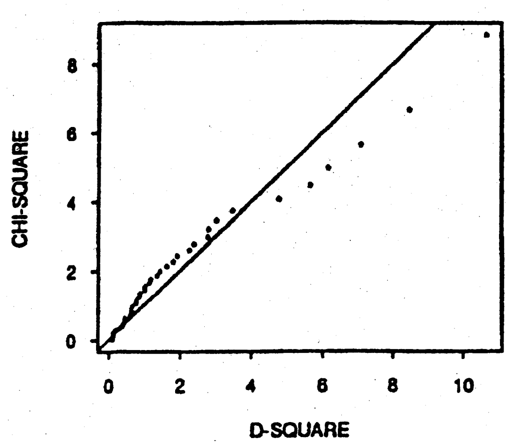
$$\bar{\mathbf{x}} = \begin{pmatrix} 10.046719 \\ 9.4047619 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} 11.363531 & 3.126597 \\ & 30.978513 \end{pmatrix}.$$

Generalized distances are as follows;

0.4607	0.6592	2.3771	1.6283	0.4135	0.4761	1.1849
10.6392	0.1388	0.8162	1.3566	0.6228	5.6494	0.3159
0.4135	0.1225	0.8988	4.7647	3.0089	0.6592	2.7741
1.0360	0.7874	3.4438	6.1489	1.0360	0.1388	0.8856
0.1380	2.2489	0.1901	0.4607	1.1472	7.0857	1.4584
0.1225	1.8985	2.7783	8.4731	0.6370	0.7032	1.8014

(b). The number of observations whose generalized distances are less than $\chi_2(0.5) = 1.39$ is 26. So the proportion is 26/42=0.6190.

(c). CHI-SQUARE PLOT FOR (X1 X2)

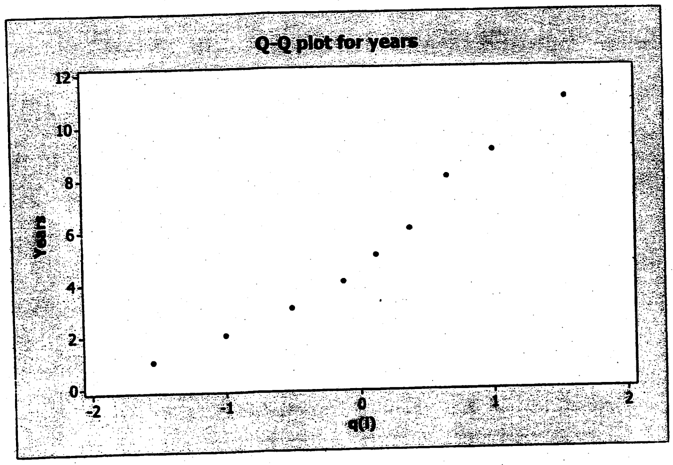


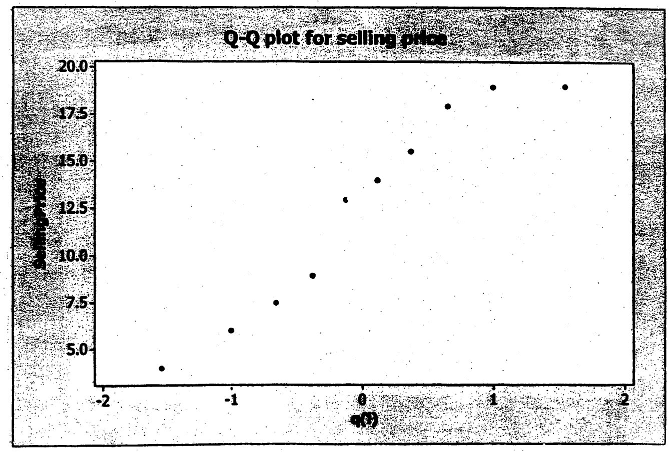
4.30 **(a)** $\hat{\lambda}_1 = 0.5$ but $\hat{\lambda}_1 = 1$ (i.e. no transformation) not ruled out by data. For $\hat{\lambda}_1 = 1$, $r_Q = .981 > .9351$ the critical point for testing normality with $n = 10$ and $\alpha = .10$. We cannot reject the hypothesis of normality at the 10% level (and, consequently, not at the 5% level).

(b) $\hat{\lambda}_1 = 1$ (i.e. no transformation). For $\hat{\lambda}_1 = 1$, $r_Q = .971 > .9351$ the critical point for testing normality with $n = 10$ and $\alpha = .10$. We cannot reject the hypothesis of normality at the 10% level (and, consequently, not at the 5% level).

(c) The likelihood function $l(\lambda_1, \lambda_2)$ is fairly flat in the region of $\lambda_1 = 1, \lambda_2 = 1$ so these values are not ruled out by the data. These results are consistent with those in parts (a) and (b).

Q-Q plots follow.





4.31

The non-multiple-sclerosis group:

	X_1	X_2	X_3	X_4	X_5
r_Q	0.94482*	0.96133*	0.95585*	0.97574*	0.94446*
Transformation	$X_1^{-0.5}$	$X_2^{-3.5}$	$(X_3+0.005)^{0.4}$	$X_4^{-3.4}$	$(X_5+0.005)^{0.32}$

$*$: significant at 5 % level (the critical point = 0.9826 for n=69).

The multiple-sclerosis group:

	X_1	X_2	X_3	X_4	X_5
r_Q	0.97137	0.97209	0.79523*	0.97869	0.84135*
Transformation	–	–	$(X_3+0.005)^{0.26}$	–	$(X_5+0.005)^{0.21}$

$*$: significant at 5 % level (the critical point = 0.9640 for n=29).

Transformations of X_3 and X_4 do not improve the approximation to normality very much because there are too many zeros.

4.32

	X_1	X_2	X_3	X_4	X_5	X_6
r_Q	0.98464*	0.94526*	0.9970	0.98098*	0.99057	0.92779*
Transformation	$(X_1+0.005)^{-0.59}$	$X_2^{-0.49}$	–	$X_4^{0.25}$	–	$(X_5+0.005)^{0.51}$

$*$: significant at 5 % level (the critical point = 0.9870 for n=98).

4.33

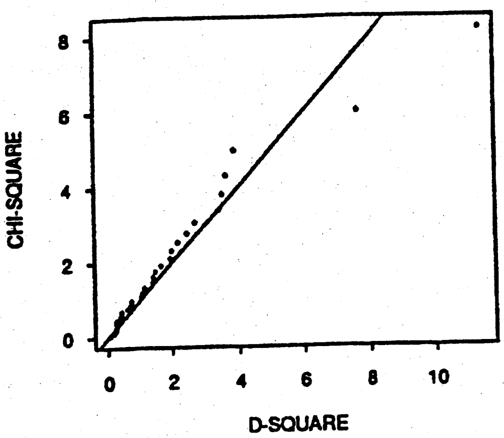
Marginal Normality:

	X_1	X_2	X_3	X_4
r_Q	0.95986*	0.95039*	0.96341	0.98079

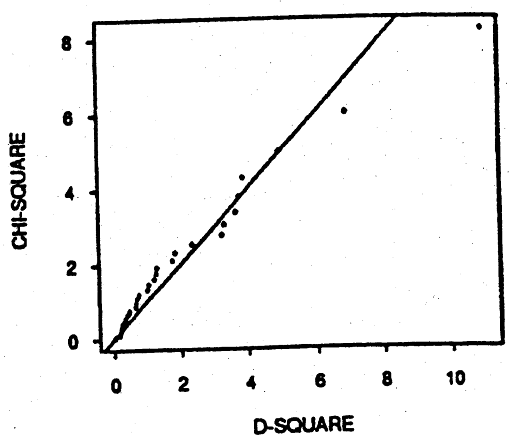
$*$: significant at 5 % level (the critical point = 0.9652 for n=30).

Bivariate Normality: the χ^2 plots are given in the next page. Those for (X_1, X_2), (X_1, X_3), (X_3, X_4) appear reasonably straight.

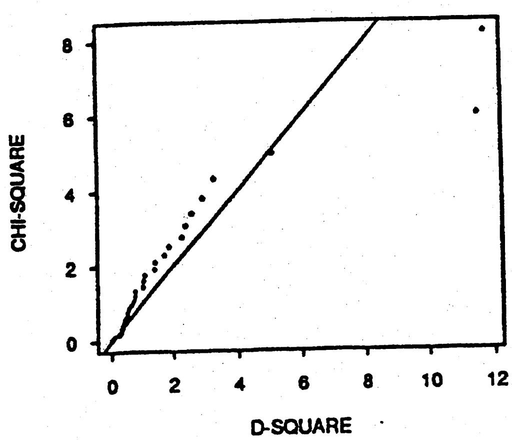
CHI-SQUARE PLOT FOR (X1,X2)



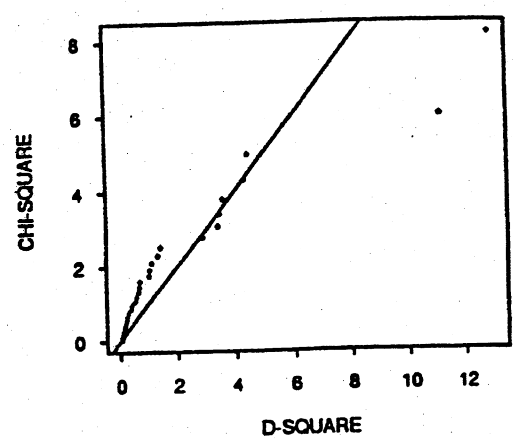
CHI-SQUARE PLOT FOR (X1,X3)



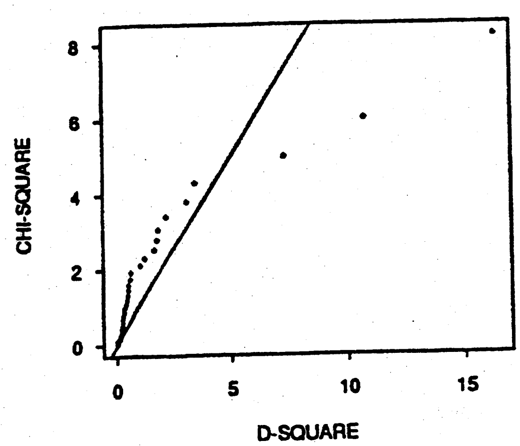
CHI-SQUARE PLOT FOR (X1,X4)



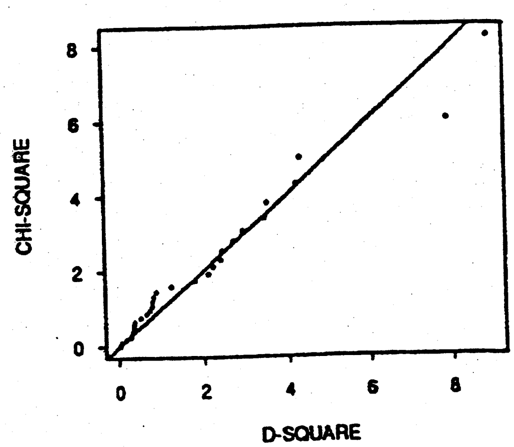
CHI-SQUARE PLOT FOR (X2,X3)



CHI-SQUARE PLOT FOR (X2,X4)



CHI-SQUARE PLOT FOR (X3,X4)



4.34

Marginal Normality:

	X_1	X_2	X_3	X_4	X_5	X_6
r_Q	0.95162*	0.97209	0.98421	0.99011	0.98124	0.99404

*: significant at 5 % level (the critical point = 0.9591 for n=25).

Bivariate Normality: Omitted.

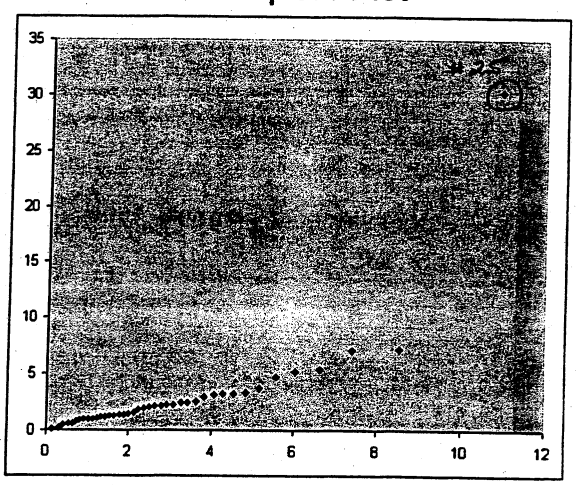
4.35 Marginal normality:

	X_1 (Density)	X_2 (MachDir)	X_3 (CrossDir)
r_Q	.897*	.991	.924*

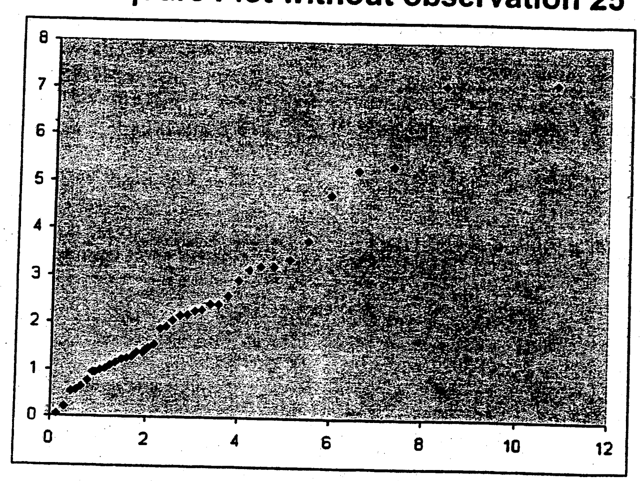
* significant at the 5% level; critical point = .974 for $n = 41$

From the chi-square plot (see below), it is obvious that observation #25 is a multivariate outlier. If this observation is removed, the chi-square plot is considerably more "straight line like" and it is difficult to reject a hypothesis of multivariate normality. Moreover, r_Q increases to .979 for density, it is virtually unchanged (.992) for machine direction and cross direction (.926).

Chi-square Plot



Chi-square Plot without observation 25



4.36 Marginal normality:

	100m	200m	400m	800m	1500m	3000m	Marathon
r_Q	.983	.976*	.969*	.952*	.909*	.866*	.859*

* significant at the 5% level; critical point = .978 for $n = 54$

Notice how the values of r_Q decrease with increasing distance. As the distance increases, the distribution of times becomes increasingly skewed to the right.

The chi-square plot is not consistent with multivariate normality. There are several multivariate outliers.

4.37 Marginal normality:

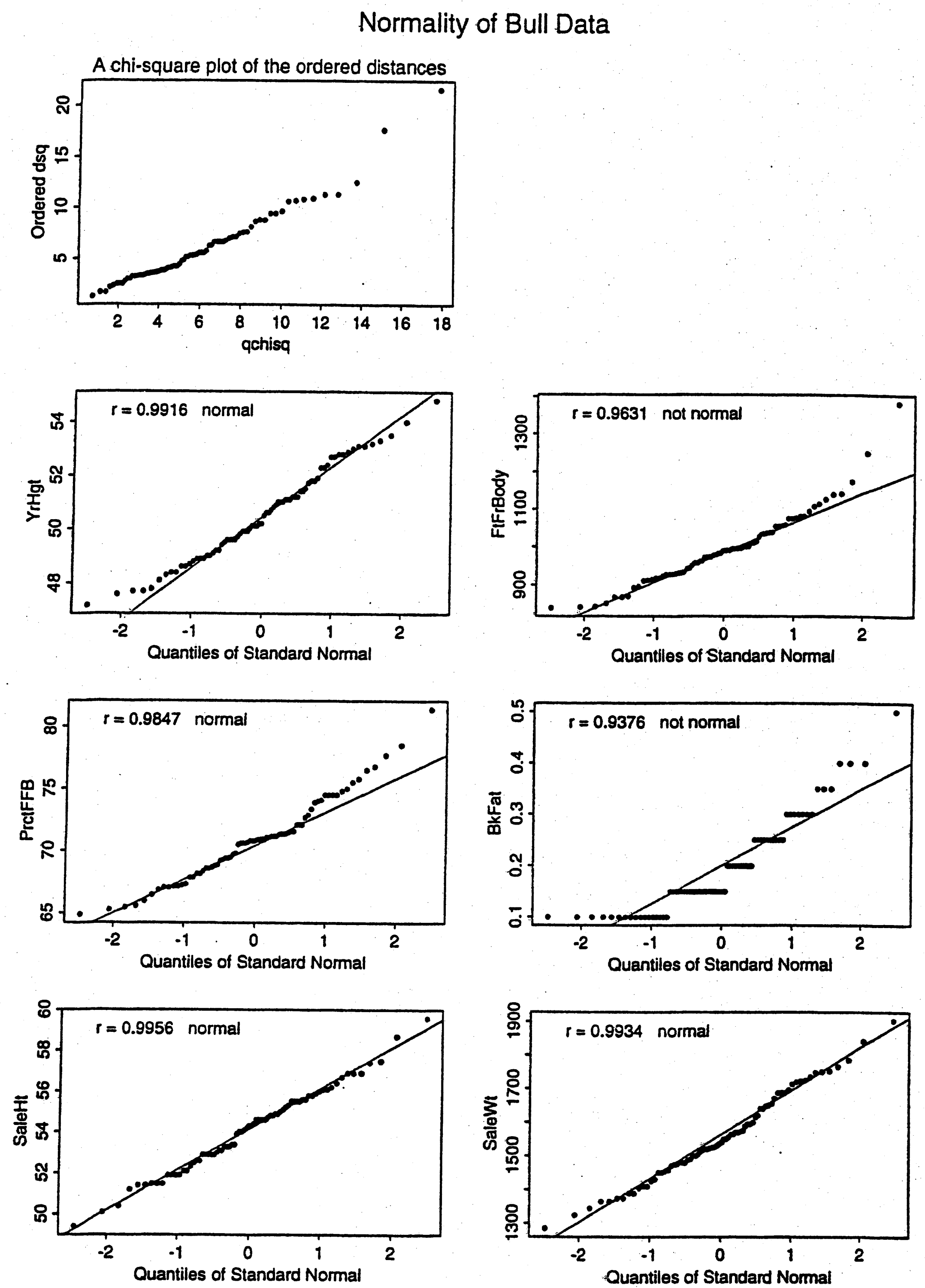
	100m	200m	400m	800m	1500m	3000m	Marathon
r_Q	.989	.985	.984	.968*	.947*	.929*	.921*

* significant at the 5% level; critical point = .978 for $n = 54$

As measured by r_Q, times measured in meters/second for the various distances are more nearly marginally normal than times measured in seconds or minutes (see Exercise 4.36). Notice the values of r_Q decrease with increasing distance. In this case, as the distance increases the distribution of times becomes increasingly skewed to the left.

The chi-square plot is not consistent with multivariate normality. There are several multivariate outliers.

4.38. Marginal and multivariate normality of bull data



XBAR	S					
	YrHgt	FtFrBody	PrctFFB	BkFat	SaleHt	SaleWt
50.5224	2.9980	100.1305	2.9600	-0.0534	2.9831	82.8108
995.9474	100.1305	8594.3439	209.5044	-1.3982	129.9401	6680.3088
70.8816	2.9600	209.5044	10.6917	-0.1430	3.4142	83.9254
0.1967	-0.0534	-1.3982	-0.1430	0.0080	-0.0506	2.4130
54.1263	2.9831	129.9401	3.4142	-0.0506	4.0180	147.2896
1555.2895	82.8108	6680.3088	83.9254	2.4130	147.2896	16850.6618

	Ordered dsq	qchisq		Ordered dsq	qchisq		Ordered dsq	qchisq
1	1.3396	0.7470	26	3.8618	4.0902	51	6.6693	6.8439
2	1.7751	1.1286	27	3.8667	4.1875	52	6.6748	6.9836
3	1.7762	1.3793	28	3.9078	4.2851	53	6.6751	7.1276
4	2.2021	1.5808	29	4.0413	4.3830	54	6.8168	7.2763
5	2.3870	1.7551	30	4.1213	4.4812	55	6.9863	7.4301
6	2.5512	1.9118	31	4.1445	4.5801	56	7.1405	7.5896
7	2.5743	2.0560	32	4.2244	4.6795	57	7.1763	7.7554
8	2.5906	2.1911	33	4.2522	4.7797	58	7.4577	7.9281
9	2.7604	2.3189	34	4.2828	4.8806	59	7.5816	8.1085
10	3.0189	2.4411	35	4.4599	4.9826	60	7.6287	8.2975
11	3.0495	2.5587	36	4.7603	5.0855	61	8.0873	8.4963
12	3.2679	2.6725	37	4.8587	5.1896	62	8.6430	8.7062
13	3.2766	2.7832	38	5.1129	5.2949	63	8.7748	8.9286
14	3.3115	2.8912	39	5.1876	5.4017	64	8.7940	9.1657
15	3.3470	2.9971	40	5.2891	5.5099	65	9.3973	9.4197
16	3.3669	3.1011	41	5.3004	5.6197	66	9.3989	9.6937
17	3.3721	3.2036	42	5.3518	5.7313	67	9.6524	9.9917
18	3.4141	3.3048	43	5.4024	5.8449	68	10.6254	10.3191
19	3.5279	3.4049	44	5.5938	5.9605	69	10.6958	10.6829
20	3.5453	3.5041	45	5.6060	6.0783	70	10.8037	11.0936
21	3.6097	3.6027	46	5.6333	6.1986	71	10.9273	11.5665
22	3.6485	3.7007	47	5.7754	6.3215	72	11.3006	12.1263
23	3.6681	3.7983	48	6.2524	6.4472	73	11.3216	12.8160
24	3.7236	3.8957	49	6.3264	6.5760	74	12.4744	13.7225
25	3.7395	3.9929	50	6.6491	6.7081	75	17.6149	15.0677
						76	21.5751	17.8649

From Table 4.2, with $\alpha = 0.05$ and $n = 76$, the critical point for the $Q-Q$ plot correlation coefficient test for normality is 0.9839. We reject the hypothesis of multivariate normality at $\alpha = 0.05$, because some marginals are not normal.

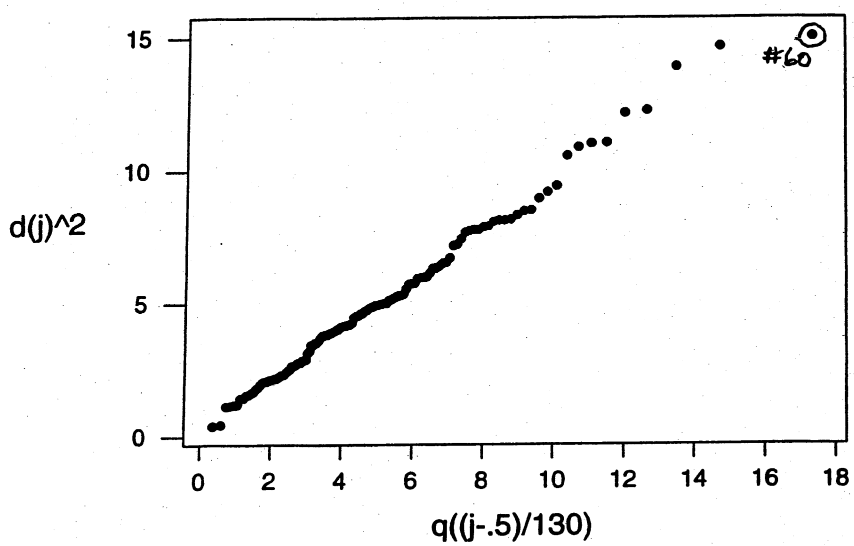
4.39 **(a)** Marginal normality:

	independence	support	benevolence	conformity	leadership
r_Q	.991	.993	.997	.997	.984*

* significant at the 5% level; critical point = .990 for n = 130

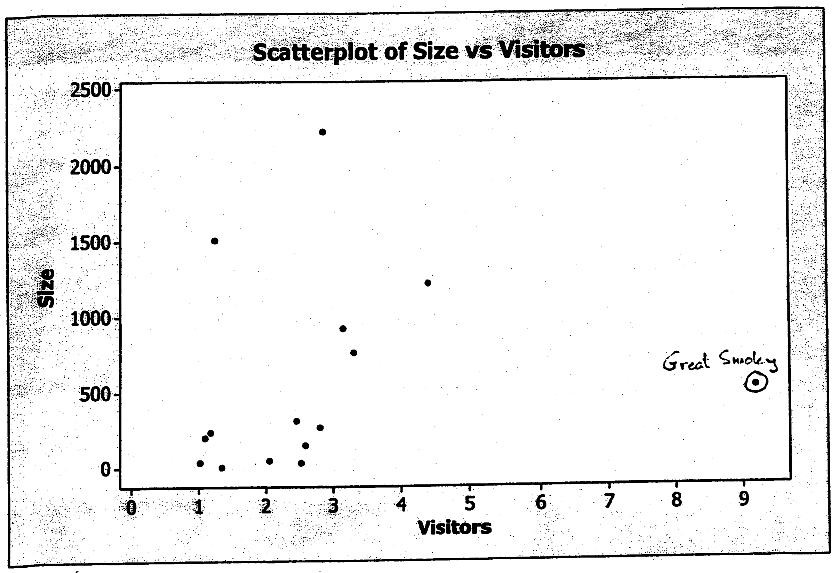
(b) The chi-square plot is shown below. Plot is straight with the exception of observation #60. Certainly if this observation is deleted would be hard to argue against multivariate normality.

Chi-square plot for indep, supp, benev, conform, leader

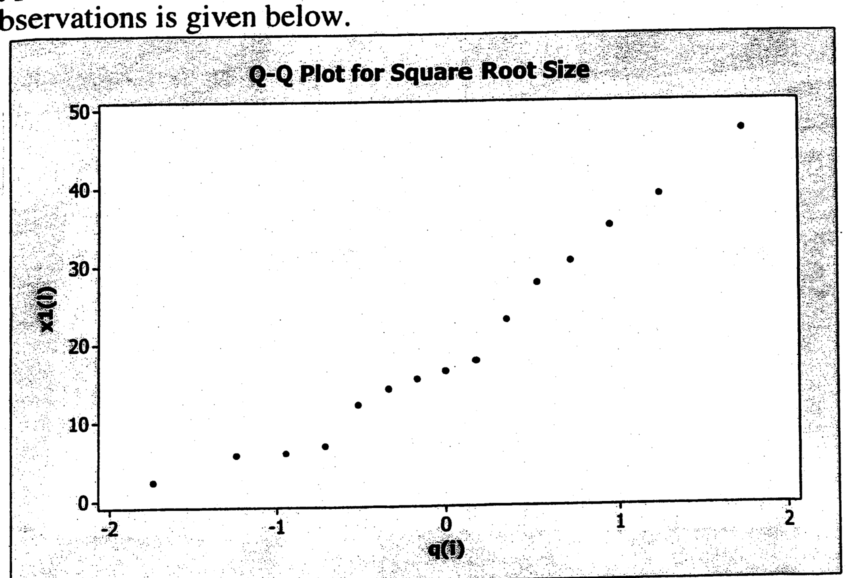


(c) Using the r_Q statistic, normality is rejected at the 5% level for leadership. If leadership is transformed by taking the square root (i.e. $\hat{\lambda} = 0.5$), r_Q = .998 and we cannot reject normality at the 5% level.

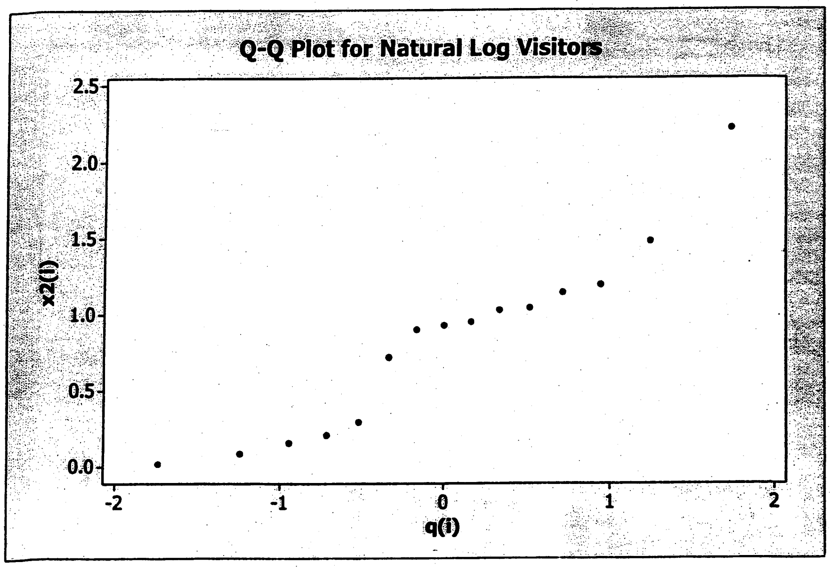
4.40 **(a)** Scatterplot is shown below. Great Smoky park is an outlier.



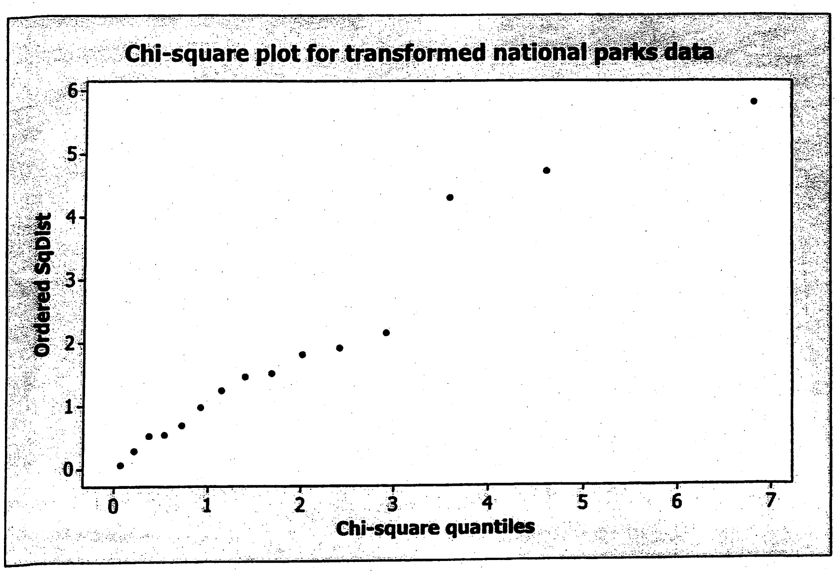
(b) The power transformation $\hat{\lambda}_1 = 0.5$ (i.e. square root) makes the size observations more nearly normal. $r_Q = .904$ before transformation and $r_Q = .975$ after transformation. The 5% critical point with $n = 15$ for the hypothesis of normality is .9389. The Q-Q plot for the transformed observations is given below.



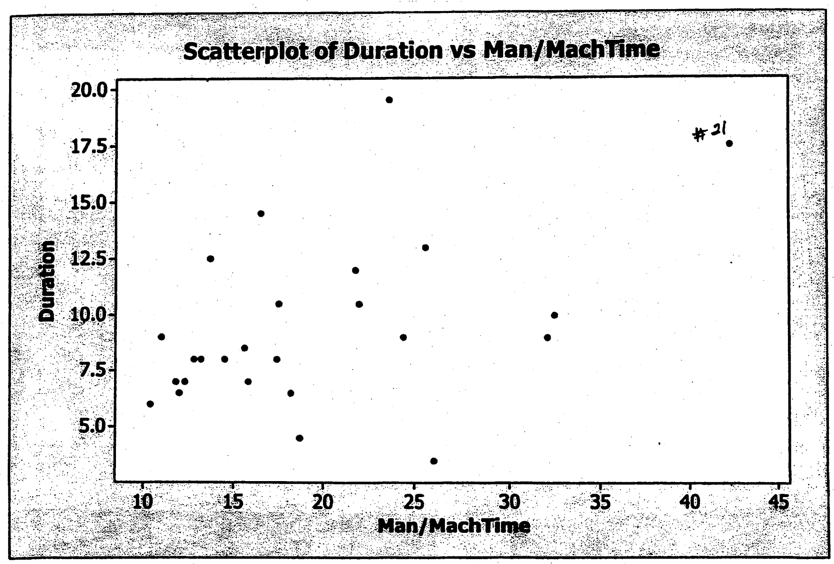
(c) The power transformation $\hat{\lambda}_2 = 0$ (i.e. logarithm) makes the visitor observations more nearly normal. $r_Q = .837$ before transformation and $r_Q = .960$ after transformation. The 5% critical point with $n = 15$ for the hypothesis of normality is .9389. The Q-Q plot for the transformed observations is given next.



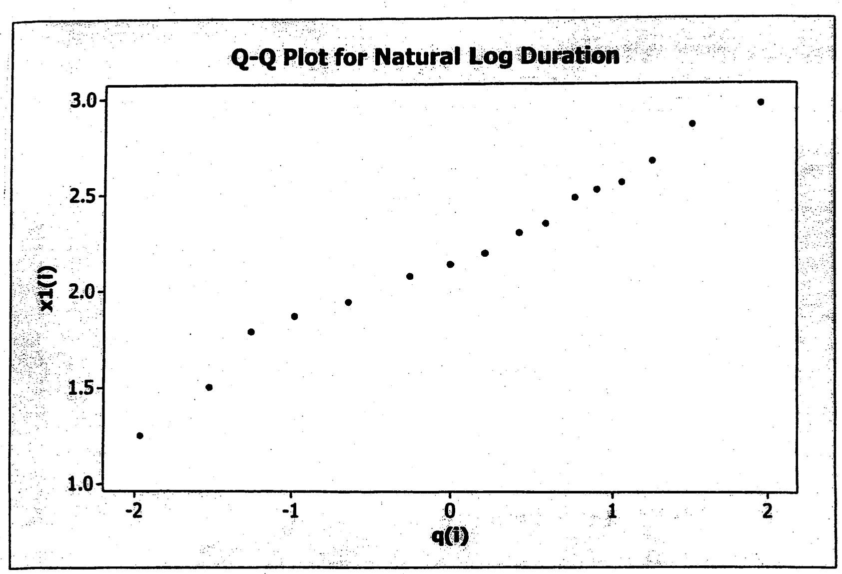
(d) A chi-square plot for the transformed observations is shown below. Given the small sample size ($n = 15$), the plot is reasonably straight and it would be hard to reject bivariate normality.



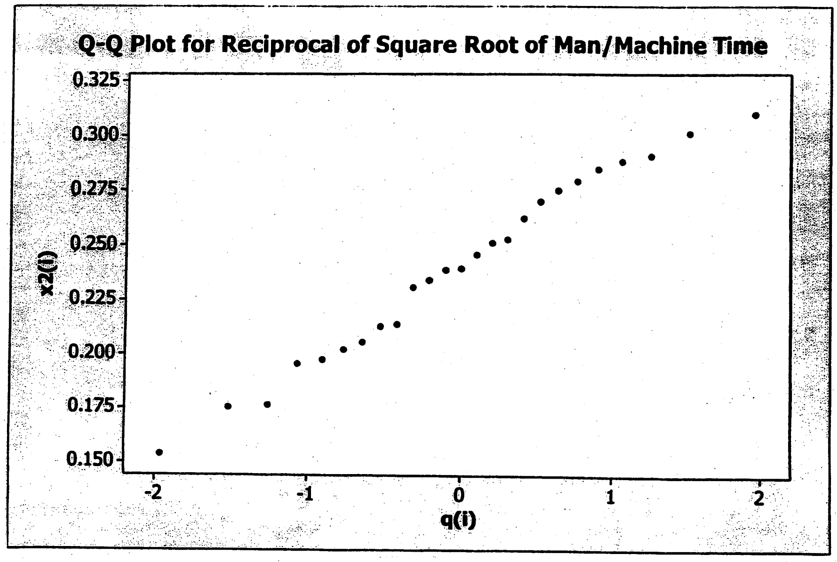
4.41 **(a)** Scatterplot is shown below. There do not appear to be any outliers with the possible exception of observation #21.



(b) The power transformation $\hat{\lambda}_1 = 0$ (i.e. logarithm) makes the duration observations more nearly normal. $r_Q = .958$ before transformation and $r_Q = .989$ after transformation. The 5% critical point with $n = 25$ for the hypothesis of normality is .9591. The *Q-Q* plot for the transformed observations is given below.



(c) The power transformation $\hat{\lambda}_2 = -0.5$ (i.e. reciprocal of square root) makes the man/machine time observations more nearly normal. $r_Q = .939$ before transformation and $r_Q = .991$ after transformation. The 5% critical point with $n = 25$ for the hypothesis of normality is .9591. The Q-Q plot for the transformed observations is given next.



(d) A chi-square plot for the transformed observations is shown below. The plot is straight and it would be difficult to reject bivariate normality.