3. The Multivariate Normal Distribution

3.1 Introduction

- A generalization of the familiar bell shaped normal density to several dimensions plays a fundamental role in multivariate analysis

- While real data are never ***exactly*** multivariate normal, the normal density is often a useful approximation to the "true" population distribution because of a ***central limit*** effect.

- One advantage of the multivariate normal distribution stems from the fact that it is mathematically tractable and "nice" results can be obtained.

To summarize, many real-world problems fall naturally within the framework of normal theory. The importance of the normal distribution rests on its dual role as both population model for certain natural phenomena and approximate sampling distribution for many statistics.

3.2 The Multivariate Normal density and Its Properties

- Recall that the univariate normal distribution, with mean μ and variance σ^2, has the probability density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)/\sigma]^2/2} \qquad -\infty < x < \infty$$

- The term

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$$

- This can be generalized for $p \times 1$ vector **x** of observations on serval variables as

$$(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

  The $p \times 1$ vector $\boldsymbol{\mu}$ represents the expected value of the random vector $\boldsymbol{X}$, and the $p \times p$ matrix Σ is the variance-covariance matrix of $\boldsymbol{X}$.

- A p-dimensional normal density for the random vector $\boldsymbol{X}' = [X_1, X_2, \ldots, X_p]$ has the form

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})/2}$$

where $-\infty < x_i < \infty, i = 1, 2, \ldots, p$. We should denote this p-dimensional normal density by $N_p(\boldsymbol{\mu}, \Sigma)$.

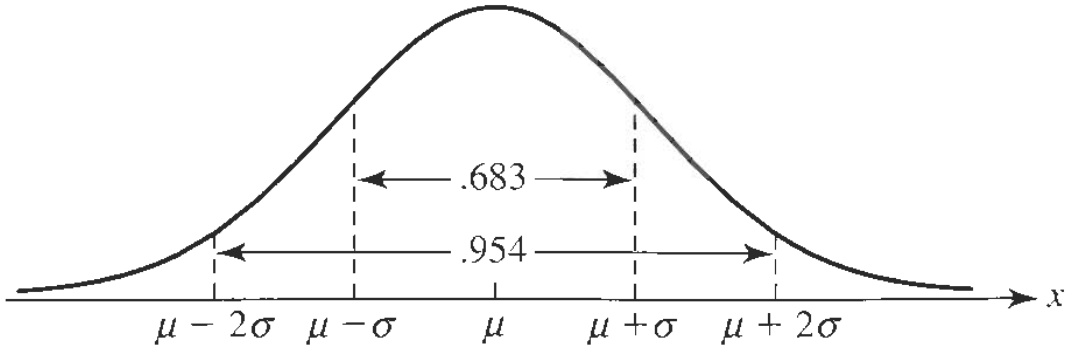


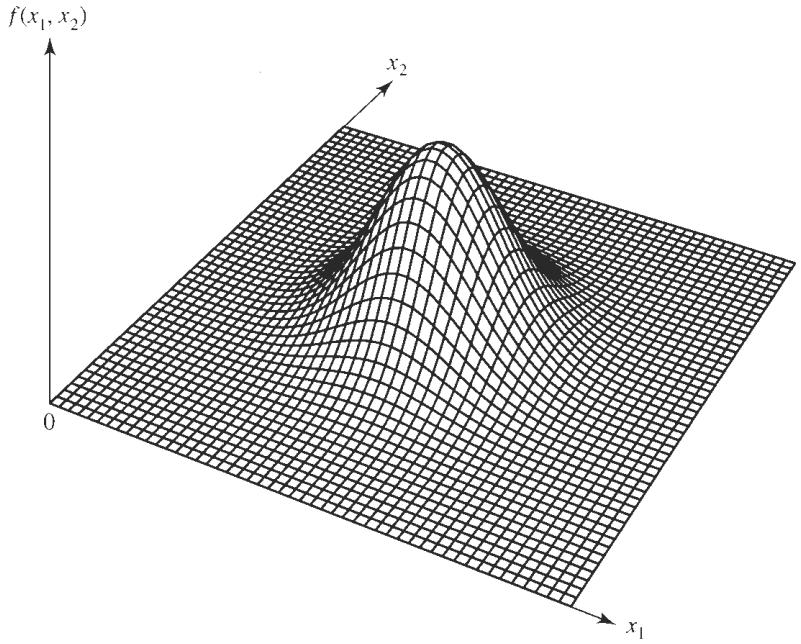
Figure 4.1 A normal density with mean μ and variance σ^2 and selected areas under the curve.

Example 3.1 (Bivariate normal density) Let us evaluate the $p = 2$ variate normal density in terms of the individual parameters $\mu_1 = \mathrm{E}(X_1), \mu_2 = \mathrm{E}(X_2), \sigma_{11} = \mathrm{Var}(X_1), \sigma_{22} = \mathrm{Var}(X_2)$, and $\rho_{12} = \sigma_{12}/(\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}) = \mathrm{Corr}(X_1, X_2)$.

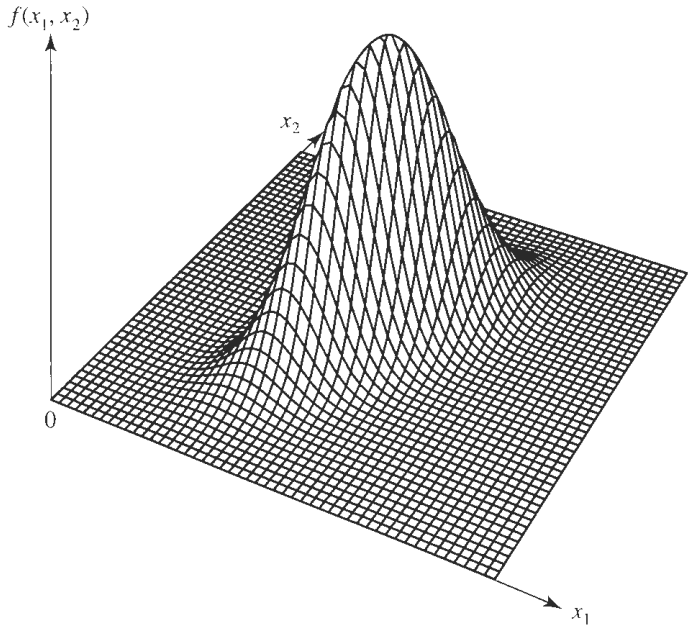
Result 3.1 If Σ is positive definite, so that Σ^{-1} exists, then

$$\Sigma\mathbf{e} = \lambda\mathbf{e} \quad \text{implies} \quad \Sigma^{-1}\mathbf{e} = \frac{1}{\lambda}\mathbf{e}$$

so $(\lambda, \mathbf{e})$ is an eigenvalue-eigenvector pair for Σ corresponding to the pair $(1/\lambda, \mathbf{e})$ for Σ^{-1}. Also Σ^{-1} is positive definite.



(a)



(b)

Figure 4.2 Two bivariate normal distributions. (a) $\sigma_{11} = \sigma_{22}$ and $\rho_{12} = 0$. (b) $\sigma_{11} = \sigma_{22}$ and $\rho_{12} = .75$.

Constant probability density contour

$= \{$ all $\mathbf{x}$ such that $(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c^2\}$

$=$ surface of an ellipsoid centered at $\boldsymbol{\mu}$.

Contours of constant density for the p-dimensional normal distribution are ellipsoids defined by $\mathbf{x}$ such the that

$$(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c^2$$

These ellipsoids are centered at $\boldsymbol{\mu}$ and have axes $\pm c\sqrt{\lambda_i}\mathbf{e}_i$, where $\Sigma\mathbf{e}_i = \lambda_i$ for $i = 1, 2, \ldots, p$.

Example 4.2 (Contours of the bivariate normal density) Obtain the axes of constant probability density contours for a bivariate normal distribution when $\sigma_{11} = \sigma_{22}$

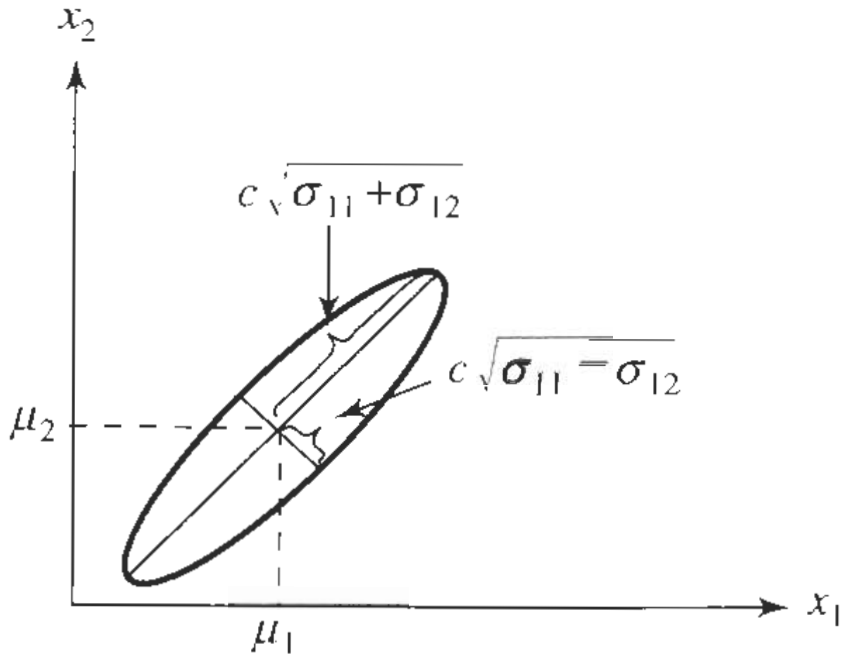


Figure 4.3 A constant-density contour for a bivariate normal distribution with $\sigma_{11} = \sigma_{22}$ and $\sigma_{12} > 0$ (or $\rho_{12} > 0$).

The solid ellipsoid of $\mathbf{x}$ values satisfying

$$(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) \leq \chi^2_p(\alpha)$$

has probability $1-\alpha$ where $\chi^2_p(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with p degrees of freedom.

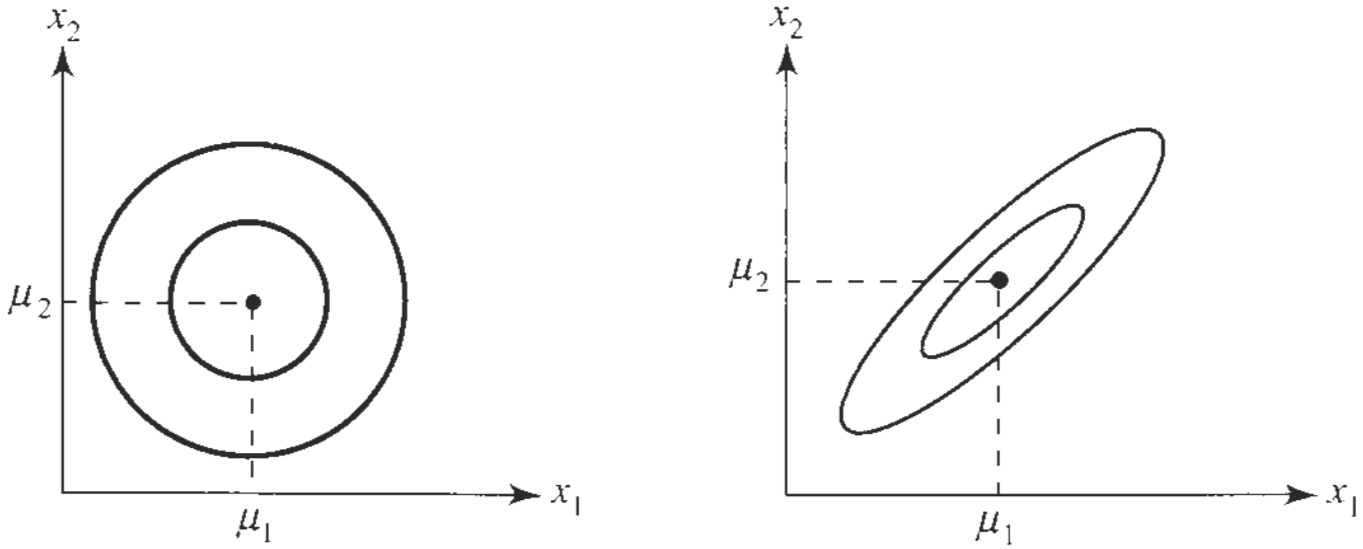


Figure 4.4 The 50% and 90% contours for the bivariate normal distributions in Figure 4.2.

Additional Properties of the Multivariate Normal Distribution

The following are true for a normal vector $\boldsymbol{X}$ having a multivariate normal distribution:

1. Linear combination of the components of $\boldsymbol{X}$ are normally distributed.

2. All subsets of the components of $\boldsymbol{X}$ have a (multivariate) normal distribution.

3. Zero covariance implies that the corresponding components are independently distributed.

4. The conditional distributions of the components are normal.

Result 3.2 If $\boldsymbol{X}$ is distributed as $N_p(\boldsymbol{\mu}, \Sigma)$, then any linear combination of variables $\mathbf{a}'\boldsymbol{X} = a_1X_1 + a_2X_2 + \cdots + a_pX_p$ is distributed as $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\Sigma\mathbf{a})$. Also if $\mathbf{a}'\boldsymbol{X}$ is distributed as $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\Sigma\mathbf{a})$ for every $\mathbf{a}$, then $\boldsymbol{X}$ must be $N_p(\boldsymbol{\mu}, \Sigma)$.

Example 3.3 (The distribution of a linear combination of the component of a normal random vector) Consider the linear combination $\mathbf{a}'\boldsymbol{X}$ of a multivariate normal random vector determined by the choice $\mathbf{a}' = [1, 0, \ldots, 0]$.

Result 3.3 If $\boldsymbol{X}$ is distributed as $N_p(\boldsymbol{\mu}, \Sigma)$, the q linear combinations

$$\mathbf{A}_{(q\times p)}\mathbf{X}_{p\times 1} = \begin{bmatrix} a_{11}X_1 + \cdots + a_{1p}X_p \\ a_{21}X_1 + \cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots + a_{qp}X_p \end{bmatrix}$$

are distributed as $N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}')$. Also $\boldsymbol{X}_{p\times 1} + \mathbf{d}_{p\times 1}$, where $\mathbf{d}$ is a vector of constants, is distributed as $N_p(\boldsymbol{\mu} + \mathbf{d}, \Sigma)$.

Example 3.4 (The distribution of two linear combinations of the components of a normal random vector) For $\boldsymbol{X}$ distributed as $N_3(\boldsymbol{\mu}, \Sigma)$, find the distribution of

$$\begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \mathbf{A}\boldsymbol{X}$$

Result 3.4 All subsets of $\boldsymbol{X}$ are normally distributed. If we respectively partition $\boldsymbol{X}$, its mean vector $\boldsymbol{\mu}$, and its covariance matrix Σ as

$$\boldsymbol{X}_{(p\times 1)} = \begin{bmatrix} X_1 \\ (q\times 1) \\ \cdots\cdots \\ X_2 \\ (p-q)\times 1 \end{bmatrix} \quad \boldsymbol{\mu}_{(p\times 1)} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ (q\times 1) \\ \cdots\cdots \\ \boldsymbol{\mu}_2 \\ (p-q)\times 1 \end{bmatrix}$$

and

$$\Sigma_{(p\times p)} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ (q\times 1) & (q\times(p-q)) \\ \cdots\cdots & \cdots\cdots \\ \Sigma_{21} & \Sigma_{22} \\ ((p-q)\times q) & ((p-q)\times(p-q)) \end{bmatrix}$$

then $\boldsymbol{X}_1$ is distributed as $N_q(\boldsymbol{\mu}_1, \Sigma_{11})$.

Example 3.5 (The distribution of a subset of a normal random vector) If $\boldsymbol{X}$ is distributed as $N_5(\boldsymbol{\mu}, \Sigma)$, find the distribution of $[X_2, X_4]'$.

Result 3.5

(a) If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent, then $\mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) = 0$, a $q_1 \times q_2$ matrix of zeros, where $\boldsymbol{X}_1$ is $q_1 \times 1$ random vector and $\boldsymbol{X}_2$ is $q_2 \times 1$. random vector

(b) If $\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$ is $N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$, then $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent if and only if $\Sigma_{12} = \Sigma_{21} = 0$.

(c) If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent and are distributed as $N_{q_1}(\boldsymbol{\mu}_1, \Sigma_{11})$ and $N_{q_2}(\boldsymbol{\mu}_2, \Sigma_{22})$, respectively, then $\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$ has the multivariate normal distribution

$$N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}\right)$$

Example 3.6 (The equivalence of zero covariance and independence for normal variables) Let $\boldsymbol{X}_{3\times 1}$ be $N_3(\boldsymbol{\mu}, \Sigma)$ with

$$\Sigma = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Are X_1 and X_2 independent ? What about (X_1, X_2) and X_3 ?

Result 3.6 Let $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$ be distributed as $N_p(\boldsymbol{\mu}, \Sigma)$ with $\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$, $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$, and $|\Sigma_{22}| > 0$. Then the conditional distribution of $\boldsymbol{X}_1$, given that $\boldsymbol{X}_2 = \mathbf{x}_2$ is normal and has

$$\text{Mean} = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

and

$$\text{Covariance} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Note that the covariance does not depend on the value $\mathbf{x}_2$ of the conditioning variable.

Example 3.7 (The conditional density of a bivariate normal distribution) Obtain the conditional density of X_1, give that $X_2 = x_2$ for any bivariate distribution.

Result 3.7 Let $\boldsymbol{X}$ be distributed as $N_p(\boldsymbol{\mu}, \Sigma)$ with $|\Sigma| > 0$. Then

(a) $(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$ is distributed as χ^2_p, where χ^2_p denotes the chi-square distribution with p degrees of freedom.

(b) The $N_p(\boldsymbol{\mu}, \Sigma)$distribution assign probability $1 - \alpha$ to the solid ellipsoid $\{\mathbf{x} : (\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \leq \chi^2_p(\alpha)\}$, where $\chi^2_p(\alpha)$ denote the upper (100α)th percentile of the χ^2_p distribution.

Result 3.8 Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be mutually independent with $\boldsymbol{X}_j$ distributed as $N_p(\boldsymbol{\mu}_j, \Sigma)$. (Note that each $\boldsymbol{X}_j$ has the *same* covariance matrix Σ.) Then

$$\mathbf{V}_1 = c_1\boldsymbol{X}_1 + c_2\boldsymbol{X}_2 + \cdots + c_n\boldsymbol{X}_n$$

is distributed as $N_p\left(\sum_{j=1}^{n} c_j\boldsymbol{\mu}_j, (\sum_{j=1}^{n} c_j^2)\Sigma\right)$. Moreover, $\mathbf{V}_1$ and $\mathbf{V}_2 = b_1\boldsymbol{X}_1 + b_2\boldsymbol{X}_2 + \cdots + b_n\boldsymbol{X}_n$ are jointly multivariate normal with covariance matrix

$$\begin{bmatrix} (\sum_{j=1}^{n} c_j^2)\Sigma & \mathbf{b}'\mathbf{c}\Sigma \\ \mathbf{b}'\mathbf{c}\Sigma_{21} & (\sum_{j=1}^{n} b_j^2)\Sigma \end{bmatrix}$$

Consequently, V_1 and V_2 are independent if $\mathbf{b}'\mathbf{c} = \sum_{j=1}^{n} c_j b_j = 0$.

Example 3.8 (Linear combinations of random vectors) Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$ and $\boldsymbol{X}_4$ be independent and identically distributed 3×1 random vectors with

$$\boldsymbol{\mu} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

(a) find the mean and variance of the linear combination $\mathbf{a}'\boldsymbol{X}_1$ of the three components of $\boldsymbol{X}_1$ where $a = [a_1 \ a_2 \ a_3]'$.

(b) Consider two linear combinations of random vectors

$$\frac{1}{2}\boldsymbol{X}_1 + \frac{1}{2}\boldsymbol{X}_2 + \frac{1}{2}\boldsymbol{X}_3 + \frac{1}{2}\boldsymbol{X}_4$$

and

$$\boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{X}_3 - 3\boldsymbol{X}_4.$$

Find the mean vector and covariance matrix for each linear combination of vectors and also the covariance between them.

3.3 Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation

The Multivariate Normal Likelihood

- Joint density function of all $p \times 1$ observed random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$

$$
\begin{aligned}
&\left\{ \begin{array}{c} \text{Joint density} \\ \text{of } \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n \end{array} \right\} \\
&= \prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-(\mathbf{x}_j - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_j - \boldsymbol{\mu})/2} \right\} \\
&= \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} e^{-\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_j - \boldsymbol{\mu})/2} \\
&= \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} e^{-\mathrm{tr}\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'\right)\right]/2}
\end{aligned}
$$

- ***Likelihood***
  When the numerical values of the observations become available, they may be substituted for the $\mathbf{x}_j$ in the equation above. The resulting expression, now considered as a function of $\boldsymbol{\mu}$ and Σ for the fixed set of observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, is called the *likelihood*.

- ***Maximum likelihood estimation***
  One meaning of best is to select the parameter values that maximize the joint density evaluated at the observations. This technique is called *maximum likelihood estimation*, and the maximizing parameter values are called *maximum likelihood estimates.*

Result 3.9 Let $\mathbf{A}$ be a $k \times k$ symmetric matrix and $\mathbf{x}$ be a $k \times 1$ vector. Then

(a) $\mathbf{x}'\mathbf{A}\mathbf{x} = \text{tr}(\mathbf{x}'\mathbf{A}\mathbf{x}) = \text{tr}(\mathbf{A}\mathbf{x}\mathbf{x}')$

(b) $\text{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$, where the λ_i are the eigenvalues of $\mathbf{A}$.

Maximum Likelihood Estimate of μ and Σ

Result 3.10 Given a $p \times p$ symmetric positive definite matrix $\mathbf{B}$ and a scalar $b > 0$, it follows that

$$\frac{1}{|\Sigma|^b} e^{-\text{tr}(\Sigma^{-1}\mathbf{B})/2} \leq \frac{1}{|B|^b}(2b)^{pb} e^{-bp}$$

for all positive definite $\Sigma_{p\times p}$, with equality holding only for $\Sigma = (1/2b)\mathbf{B}$.

Result 3.11 Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a random sample from a normal population with mean $\boldsymbol{\mu}$ and covariance Σ. Then

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{X}} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{X}_j - \bar{X})(\boldsymbol{X}_j - \bar{X})' = \frac{n-1}{n}\mathbf{S}$$

are the ***maximum likelihood estimators*** of $\boldsymbol{\mu}$ and Σ, respectively. Their observed value $\bar{\mathbf{x}}$ and $(1/n)\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'$, are called the ***maximum likelihood estimates*** of $\boldsymbol{\mu}$ and Σ.

Invariance Property of Maximum likelihood estimators

Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimator of $\boldsymbol{\theta}$, and consider the parameter $h(\boldsymbol{\theta})$, which is a function of $\boldsymbol{\theta}$. Then the maximum likelihood estimate of

$$h(\boldsymbol{\theta}) \text{ is given by } h(\hat{\boldsymbol{\theta}}).$$

For example

1. The maximum likelihood estimator of $\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$ is $\hat{\boldsymbol{\mu}}\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$, where $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{X}}$ and $\hat{\Sigma} = \frac{n-1}{n}\mathbf{S}$ are the maximum likelihood estimators of $\boldsymbol{\mu}$ and Σ respectively.

2. The maximum likelihood estimator of $\sqrt{\sigma_{ii}}$ is $\sqrt{\hat{\sigma}_{ii}}$, where

$$\hat{\sigma}_{ii} = \frac{1}{n}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2$$

is the maximum likelihood estimator of $\sigma_{ii} = \mathrm{Var}(X_i)$.

Sufficient Statistics

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a random sample from a multivariate normal population with mean μ and covariance Σ. Then

$$\bar{\boldsymbol{X}} \text{ and } \mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{X}_j - \bar{\boldsymbol{X}})(\boldsymbol{X}_j - \bar{\boldsymbol{X}})' \text{ are sufficient statistics}$$

- The importance of sufficient statistics for normal populations is that all of the information about $\boldsymbol{\mu}$ and Σ in the data matrix $\boldsymbol{X}$ is contained in $\bar{\boldsymbol{X}}$ and $\mathbf{S}$, regardless of the sample size n.

- This generally is not true for nonnormal populations.

- Since many multivariate techniques begin with sample means and covariances, it is prudent to check on the adequacy of the multivariate normal assumption.

- If the data cannot be regarded as multivariate normal, techniques that depend solely on $\bar{\boldsymbol{X}}$ and $\mathbf{S}$ may be ignoring other useful sample information.

3.4 The Sampling Distribution of $\bar{X}$ and S

- **The univariate case** $(p = 1)$
  - $\bar{X}$ is normal with mean μ =(population mean) and variance

$$\frac{1}{n}\sigma^2 = \frac{\text{population variance}}{\text{sample size}}$$

  - For the sample variance, recall that $(n-1)s^2 = \sum_{j=1}^{n}(X_j - \bar{X})^2$ is distributed as σ^2 times a chi-square variable having $n-1$ degrees of freedom (d.f.).
  - The chi-square is the distribution of a sum squares of independent standard normal random variables. That is, $(n-1)s^2$ is distributed as $\sigma^2(Z_1^2 + \cdots + Z_{n-1}^2) = (\sigma Z_1)^2 + \cdots + (\sigma Z_{n-1})^2$. The individual terms σZ_i are independently distributed as $N(0, \sigma^2)$.

- ***Wishart distribution***

$$\begin{aligned} W_m(\cdot|\Sigma) &= \text{Wishart distribution with } m \text{ d.f.} \\ &= \text{distribution of } \sum_{j=1}^{n} \mathbf{Z}_j \mathbf{Z}_j' \end{aligned}$$

where $\mathbf{Z}_j$ are each independently distributed as $N_p(0, \Sigma)$.

- **Properties of the Wishart Distribution**

1. If $\mathbf{A}_1$ is distributed as $W_{m_1}(\mathbf{A}_1|\Sigma)$ independently of $\mathbf{A}_2$, which is distributed as $W_{m_2}(\mathbf{A}_2|\Sigma)$, then $\mathbf{A}_1 + \mathbf{A_2}$ is distributed as $W_{m_1+m_2}(\mathbf{A}_1 + \mathbf{A}_2|\Sigma)$. That is, the the degree of freedom add.
2. If $\mathbf{A}$ is distributed as $W_m(\mathbf{A}|\Sigma)$, then $\mathbf{CAC}'$ is distributed as $W_m(\mathbf{CAC}'|\mathbf{C}\Sigma\mathbf{C}')$.

- **The Sampling Distribution of $\bar{X}$ and S**

  Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a random sample size n from a p-variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Then

  1. $\bar{\boldsymbol{X}}$ is distributed as $N_p(\boldsymbol{\mu}, \frac{1}{n}\Sigma)$.
  2. $(n-1)\mathbf{S}$ is distributed as a Wishart random matrix with $n-1$ d.f.
  3. $\bar{\boldsymbol{X}}$ and $\mathbf{S}$ are independent.

4.5 Large-Sample Behavior of $\bar{X}$ and S

Result 3.12 (Law of Large numbers) Let $Y_1, Y_2, \ldots, Y_n$ be independent observations from a population with mean $\mathrm{E}(Y_i) = \mu$, then

$$\bar{Y} = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$$

converges in probability to μ as n increases without bound. That is, for any prescribed accuracy $\varepsilon > 0$, $P[-\varepsilon < \bar{Y} - \mu < \varepsilon]$ approaches unity as $n \to \infty$.

Result 3.13 (The central limit theorem) Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be independent observations from any population with mean $\boldsymbol{\mu}$ and finite covariance Σ. Then

$$\sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \text{ has an approximate } N_p(0, \Sigma) \text{distribution}$$

for large sample sizes. Here n should also be large relative to p.

Large-Sample Behavior of $\bar{X}$ and $\mathbf{S}$

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be independent observations from a population with mean $\boldsymbol{\mu}$ and finite (nonsingular) covariance Σ. Then

$$\sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \text{is approximately } N_p(0, \Sigma)$$

and

$$n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \text{ is approximately } \chi^2_p$$

for $n - p$ large.

3.6 Assessing the Assumption of Normality

- Most of the statistical techniques discussed assume that each vector observation $\boldsymbol{X}_j$ comes from a multivariate normal distribution.

- In situations where the sample size is large and the techniques dependent solely on the behavior of $\bar{\boldsymbol{X}}$, or distances involve $\bar{X}$ of the form $n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\mathbf{S}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$, the assumption of normality for the individual observations is less crucial.

- But to some degree, the ***quality*** of inferences made by these methods depends on how closely the true parent population resembles the multivariate normal form.

Therefore, we address these questions:

1. Do the marginal distributions of the elements of $\boldsymbol{X}$ appear to be normal ? What about a few linear combinations of the components X_j ?

2. Do the scatter plots of observations on different characteristics give the elliptical appearance expected from normal population ?

3. Are there any "wild" observations that should be checked for accuracy ?

Evaluating the Normality of the Univariate Marginal Distributions

- Dot diagrams for smaller n and histogram for $n > 25$ or so help reveal situations where one tail of a univariate distribution is much longer than other.

- If the histogram for a variable X_i appears reasonably symmetric , we can check further by counting the number of observations in certain interval, for examples
  A univariate normal distribution assigns probability 0.683 to the interval

$$(\mu_i - \sqrt{\sigma_{ii}}, \mu_i + \sqrt{\sigma_{ii}})$$

  and probability 0.954 to the interval

$$(\mu_i - 2\sqrt{\sigma_{ii}}, \mu_i + 2\sqrt{\sigma_{ii}})$$

  Consequently, with a large same size n, the observed proportion $\hat{p}_{i1}$ of the observations lying in the interval $(\bar{x}_i - \sqrt{s_{ii}}, \bar{x}_i + \sqrt{s_{ii}})$ to be about 0.683, and the interval $(\bar{x}_i - 2\sqrt{s_{ii}}, \bar{x}_i + 2\sqrt{s_{ii}})$ to be about 0.954

Using the normal approximating to the sampling of $\hat{p}_i$, observe that either

$$|\hat{p}_{i1} - 0.683| > 3\sqrt{\frac{(0.683)(0.317)}{n}} = \frac{1.396}{\sqrt{n}}$$

or

$$|\hat{p}_{i2} - 0.954| > 3\sqrt{\frac{(0.954)(0.046)}{n}} = \frac{0.628}{\sqrt{n}}$$

would indicate departures from an assumed normal distribution for the ith characteristic.

- Plots are always useful devices in any data analysis. Special plots called $Q-Q$ *plots* can be used to assess the assumption of normality.

  Let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ represent these observations after they are ordered according to magnitude. For a standard normal distribution, the quantiles $q_{(j)}$ are defined by the relation

$$P[Z \leq q_{(j)}] = \int_{-\infty}^{q(j)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = p_{(j)} = \frac{j - \frac{1}{2}}{n}$$

  Here $p_{(j)}$ is the probability of getting a value less than or equal to $q_{(j)}$ in a single drawing from a standard normal population.

- The idea is to look at the pairs of quantiles $(q_{(j)}, x_{(j)})$ with the same associated cumulative probability $(j - \frac{1}{2})/n$. If the data arise from a normal population, the pairs $(q_{(j)}, x_{(j)})$ will be approximately linear related, since $\sigma q_{(j)} + \mu$ is nearly expected sample quantile.

Example 3.9 (Constructing a Q-Q plot) A sample of $n = 10$ observation gives the values in the following table:

Ordered observations $x_{(j)}$	Probability levels $(j - \frac{1}{2})/n$	Standard normal quantiles $q_{(j)}$
−1.00	.05	−1.645
−.10	.15	−1.036
.16	.25	−.674
.41	.35	−.385
.62	.45	−.125
.80	.55	.125
1.26	.65	.385
1.54	.75	.674
1.71	.85	1.036
2.30	.95	1.645

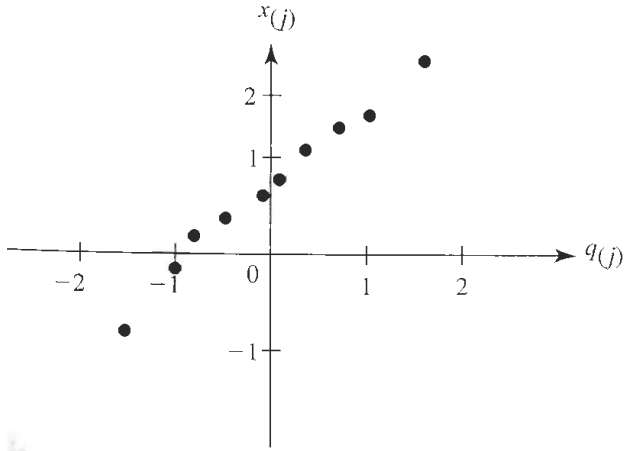


Figure 4.5 A Q–Q plot for the data in Example 4.9.

The steps leading to a Q-Q plot are as follows:

1. Order the original observations to get $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ and their corresponding probability values $(1 - \frac{1}{2})/n, (2 - \frac{1}{2})/n, \ldots, (n - \frac{1}{2})/n$;
2. Calculate the standard quantiles $q_{(1)}, q_{(2)}, \ldots, q_{(n)}$ and
3. Plot the pairs of observations $(q_{(1)}, x_{(1)}), (q_{(2)}, x_{(2)}), \ldots, (q_{(n)}, x_{(n)})$, and examine the "straightness" of the outcome.

Example 4.10 (A Q-Q plot for radiation data) The quality -control department of a manufacturer of microwave ovens is required by the federal government to monitor the amount of radiation emitted when the doors of the ovens are closed. Observations of the radiation emitted through closed doors of $n = 42$ randomly selected ovens were made. The data are listed in the following table.

Table 4.1 Radiation Data (Door Closed)

Oven no.	Radiation	Oven no.	Radiation	Oven no.	Radiation
1	.15	16	.10	31	.10
2	.09	17	.02	32	.20
3	.18	18	.10	33	.11
4	.10	19	.01	34	.30
5	.05	20	.40	35	.02
6	.12	21	.10	36	.20
7	.08	22	.05	37	.20
8	.05	23	.03	38	.30
9	.08	24	.05	39	.30
10	.10	25	.15	40	.40
11	.07	26	.10	41	.30
12	.02	27	.15	42	.05
13	.01	28	.09		
14	.10	29	.08		
15	.10	30	.18		

Source: Data courtesy of J. D. Cryer.

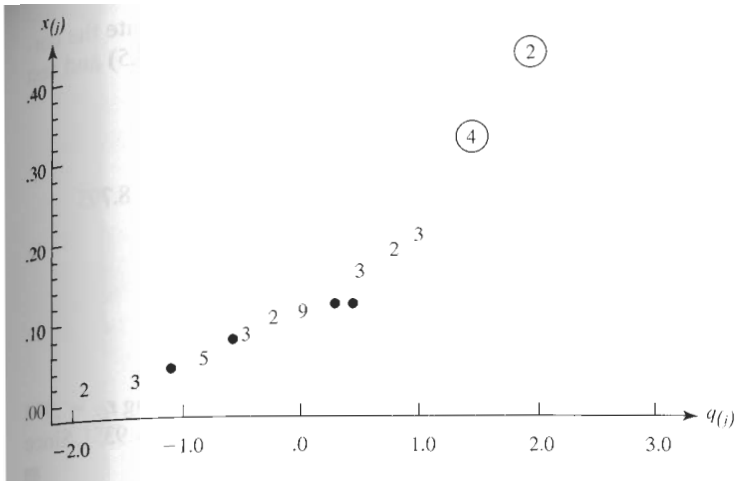


Figure 4.6 A Q–Q plot of the radiation data (door closed) from Example 4.10. (The integers in the plot indicate the number of points occupying the same location.)

The straightness of the Q-Q plot can be measured ba calculating the correlation coefficient of the points in the plot. The correlation coefficient for the Q-Q plot is defined by

$$r_Q = \frac{\sum_{j=1}^{n} (x_{(j)} - \bar{x})(q_{(j)} - \bar{q})}{\sqrt{\sum_{j=1}^{n} (x_{(j)} - \bar{x})^2} \sqrt{\sum_{j=1}^{n} (q_{(j)} - \bar{q})^2}}$$

and a powerful test of normality can be based on it. Formally we reject the hypothesis of normality at level of significance α if r_Q fall *below* the appropriate value in the following table

Table 4.2 Critical Points for the Q–Q Plot Correlation Coefficient Test for Normality

Sample size n	Significance levels α		
	.01	.05	.10
5	.8299	.8788	.9032
10	.8801	.9198	.9351
15	.9126	.9389	.9503
20	.9269	.9508	.9604
25	.9410	.9591	.9665
30	.9479	.9652	.9715
35	.9538	.9682	.9740
40	.9599	.9726	.9771
45	.9632	.9749	.9792
50	.9671	.9768	.9809
55	.9695	.9787	.9822
60	.9720	.9801	.9836
75	.9771	.9838	.9866
100	.9822	.9873	.9895
150	.9879	.9913	.9928
200	.9905	.9931	.9942
300	.9935	.9953	.9960

Example 3.11 (A correlation coefficient test for normality) Let us calculate the correlation coefficient r_Q from Q-Q plot of Example 3.9 and test for normality.

Linear combinations of more than one characteristic can be investigated. Many statistician suggest plotting

$$\hat{\mathbf{e}}_1'\mathbf{x}_j \quad \text{where} \quad \mathbf{S}\hat{\mathbf{e}_1} = \hat{\lambda}_1\hat{\mathbf{e}_1}$$

in which $\hat{\lambda}_1$ is the largest eigenvalue of $\mathbf{S}$. Here $\mathbf{x}_j' = [x_{j1}, x_{j2}, \ldots, x_{jp}]$ is the jth observation on p variables $X_1, X_2, \ldots, X_p$. The linear combination $\hat{\mathbf{e}}_p\mathbf{x}_j$ corresponding to the smallest eigenvalue is also frequently singled out for inspection

Evaluating Bivariate Normality

- By Result 3.7, the set of bivariate outcomes $\mathbf{x}$ such that

$$(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) \leq \chi_2^2(0.5)$$

  has probability 0.5.

- Thus we should expect *roughly* the same percentage, 50%, of sample observations lie in the ellipse given by

$$\{\text{all } \mathbf{x} \text{ such that } (\mathbf{x}-\hat{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}-\hat{\mathbf{x}}) \leq \chi_2^2(0.5)\}$$

  where $\boldsymbol{\mu}$ is replaced by $\hat{\mathbf{x}}$and Σ^{-1} by its estimate $\mathbf{S}^{-1}$. If not, the normality assumption is suspect.

Example 3.12 (Checking bivariate normality) Although not a random sample, data consisting of the pairs of observations (x_1 = sales, x_2 = profits) for the 10 largest companies in the world are listed in the following table. Check if (x_1, x_2) follows bivariate normal distribution.

The World's 10 Largest Companies[1]

Company	x_1 = sales (billions)	x_2 = profits (billions)	x_3 = assets (billions)
Citigroup	108.28	17.05	1,484.10
General Electric	152.36	16.59	750.33
American Intl Group	95.04	10.91	766.42
Bank of America	65.45	14.14	1,110.46
HSBC Group	62.97	9.52	1,031.29
ExxonMobil	263.99	25.33	195.26
Royal Dutch/Shell	265.19	18.54	193.83
BP	285.06	15.73	191.11
ING Group	92.01	8.10	1,175.16
Toyota Motor	165.68	11.13	211.15

[1]From www.Forbes.com partially based on *Forbes* "The Forbes Global 2000," April 18, 2005.

- A somewhat more formal method for judging normality of a data set is based on the squared generalized distances

$$d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})$$

- When the parent population is multivariate normal and both n and $n - p$ are greater than 25 or 30, each of the squared distance $d_1^2, d_2^2, \ldots, d_n^2$ should behave like a chi-square random variable.

- Although these distances are not independent or exactly chi-square distributed, it is helpful to plot them as if they were. The resulting plot is called a ***chi-square plot*** or ***gamma plot***, because the chi-square distribution is a special case of the more general gamma distribution. To construct the chi-square plot

  1. Order the square distance in the equation above from smallest to largest as $d_{(1)}^2 \leq d_{(2)}^2 \leq \cdots \leq d_{(n)}^2$.
  2. Graph the pairs $(q_{c,p}((j - \frac{1}{2})/n), d_{(j)}^2)$, where $q_{c,p}((j - \frac{1}{2})/n)$ is the $100(j - \frac{1}{2})/n$ quantile of the chi-square distribution with p degrees of freedom.

Example 3.13 (Constructing a chi-square plot) Let us construct a chi-square plot of the generalized distances given in Example 3.12. The order distance and the corresponding chi-square percentile for $p = 2$ and $n = 10$ are listed in the following table:

j	$d^2_{(j)}$	$q_{c,2}\left(\frac{j-\frac{1}{2}}{10}\right)$
1	.30	.10
2	.62	.33
3	1.16	.58
4	1.30	.86
5	1.61	1.20
6	1.64	1.60
7	1.71	2.10
8	1.79	2.77
9	3.53	3.79
10	4.38	5.99

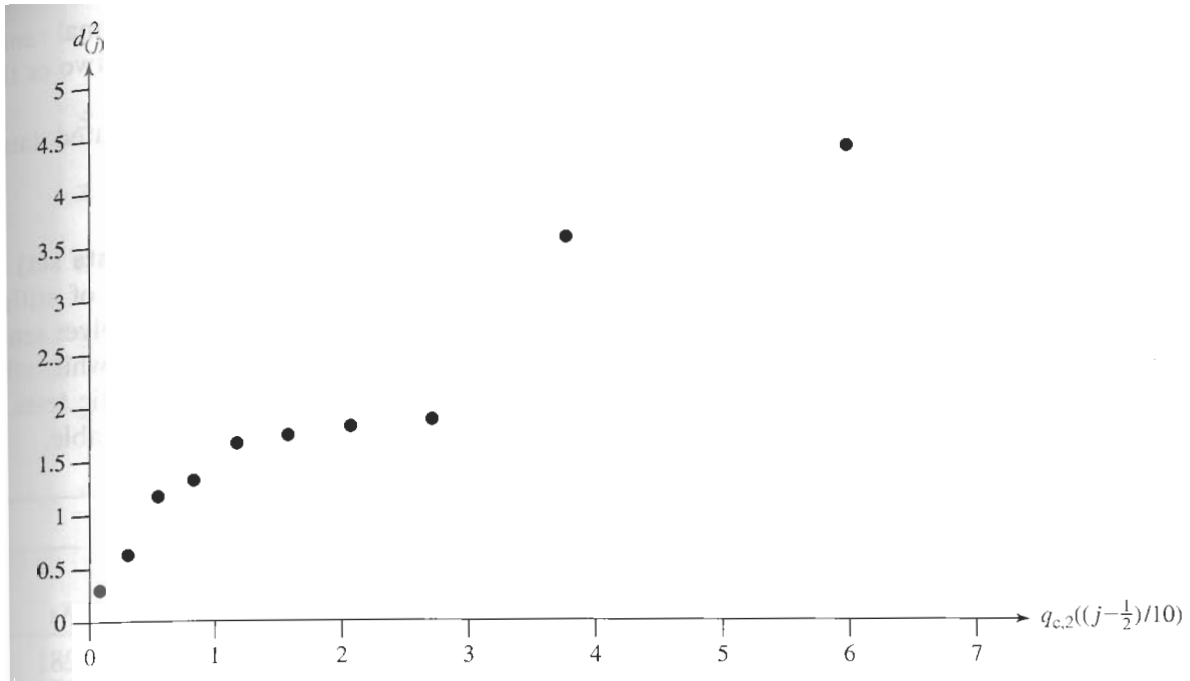


Figure 4.7 A chi-square plot of the ordered distances in Example 4.13.

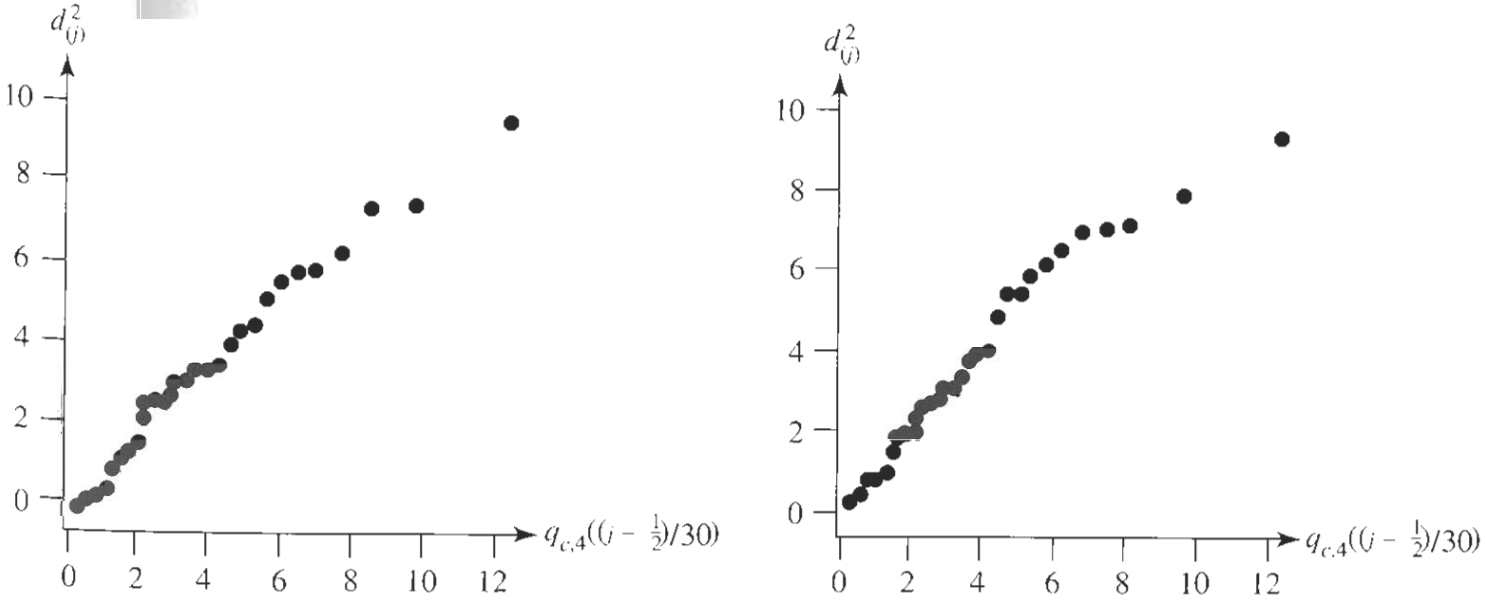


Figure 4.8 Chi-square plots for two simulated four-variate normal data sets with $n = 30$.

Example 3.14 (Evaluating multivariate normality for a four-variable data set) The data in Table 4.3 were obtained by taking four different measures of stiffness, x_1, x_2, x_3, and x_4, of each of $n = 30$ boards. the first measurement involving sending a shock wave down the board, the second measurement is determined while vibrating the board, and the last two measurements are obtained from static tests. The squared distances $d_j = (\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$ are also presented in the table

Table 4.3 Four Measurements of Stiffness

Observation no.	x_1	x_2	x_3	x_4	d^2	Observation no.	x_1	x_2	x_3	x_4	d^2
1	1889	1651	1561	1778	.60	16	1954	2149	1180	1281	16.85
2	2403	2048	2087	2197	5.48	17	1325	1170	1002	1176	3.50
3	2119	1700	1815	2222	7.62	18	1419	1371	1252	1308	3.99
4	1645	1627	1110	1533	5.21	19	1828	1634	1602	1755	1.36
5	1976	1916	1614	1883	1.40	20	1725	1594	1313	1646	1.46
6	1712	1712	1439	1546	2.22	21	2276	2189	1547	2111	9.90
7	1943	1685	1271	1671	4.99	22	1899	1614	1422	1477	5.06
8	2104	1820	1717	1874	1.49	23	1633	1513	1290	1516	.80
9	2983	2794	2412	2581	12.26	24	2061	1867	1646	2037	2.54
10	1745	1600	1384	1508	.77	25	1856	1493	1356	1533	4.58
11	1710	1591	1518	1667	1.93	26	1727	1412	1238	1469	3.40
12	2046	1907	1627	1898	.46	27	2168	1896	1701	1834	2.38
13	1840	1841	1595	1741	2.70	28	1655	1675	1414	1597	3.00
14	1867	1685	1493	1678	.13	29	2326	2301	2065	2234	6.28
15	1859	1649	1389	1714	1.08	30	1490	1382	1214	1284	2.58

Source: Data courtesy of William Galligan.

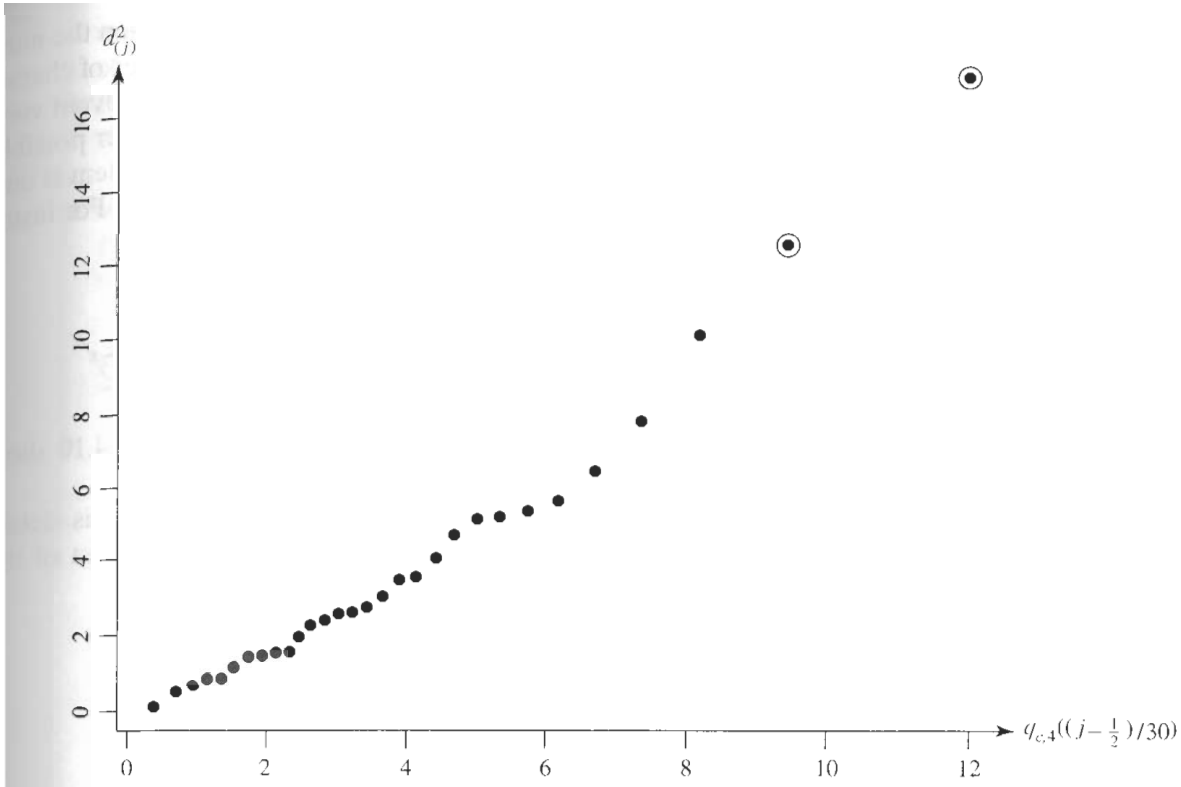


Figure 4.9 A chi-square plot for the data in Example 4.14.

3.7 Detecting Outliers and Cleaning Data

- Outliers are best detected visually whenever this is possible
- For a single random variable, the problem is one dimensional, and we look for observations that are far from the others.
- In the bivariate case, the situation is more complicated. Figure 4.10 shows a situation with two unusual observations.
- In higher dimensions, there can be outliers that cannot be detected from the univariate plots or even the bivariate scatter plots. Here a large value of $(\mathbf{x}_j - \bar{x})'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{x})$ will suggest an unusual observation. even though it cannot be seen visually.

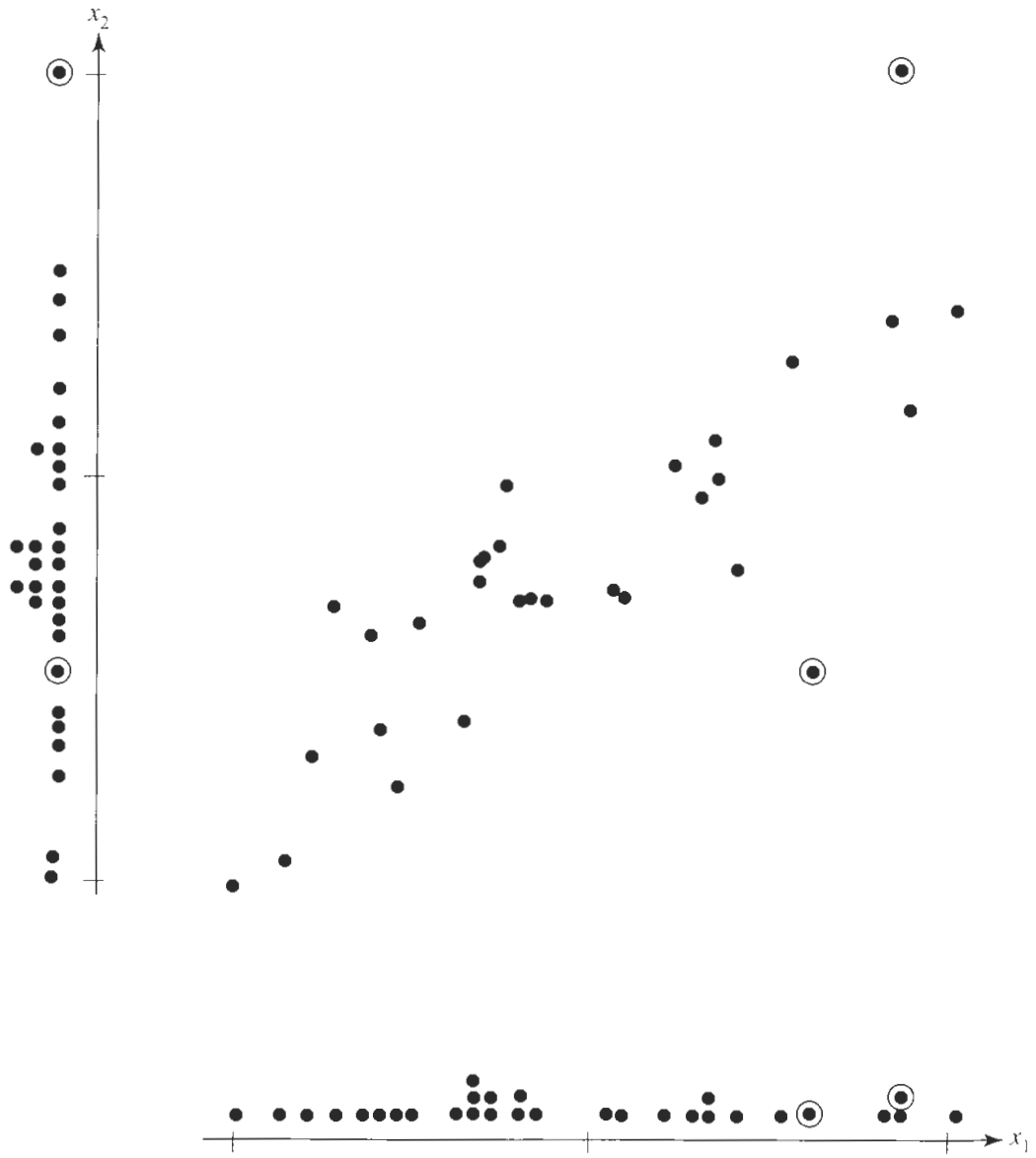


Figure 4.10 Two outliers; one univariate and one bivariate.

Steps for Detecting Outliers

1. Math a dot plot for each variable.
2. Make a scatter plot for each pair of variables.
3. Calculate the standardize variable $z_{jk} = (x_{jk} - \bar{x}_k)/\sqrt{s_{kk}}$ for $j = 1, 2, \ldots, n$ and each column $k = 1, 2, \ldots, p$. Examine these standardized values for large or small values.
4. Calculate the generalized squared distance $(\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$. Examine these distances for unusually values. In a chi-square plot, these would be the points farthest from the origin.

x_1	x_2	x_3	x_4	x_5	z_1	z_2	z_3	z_4	z_5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
1631	1528	1452	1559	1602	.06	−.15	.05	.28	−.12
1770	1677	1707	1738	1785	.64	.43	1.07	.94	.60
1376	1190	723	1285	2791	−1.01	−1.47	−2.87	−.73	(4.57)
1705	1577	1332	1703	1664	.37	.04	−.43	.81	.13
1643	1535	1510	1494	1582	.11	−.12	.28	.04	−.20
1567	1510	1301	1405	1553	−.21	−.22	−.56	−.28	−.31
1528	1591	1714	1685	1698	−.38	.10	1.10	.75	.26
1803	1826	1748	2746	1764	.78	1.01	1.23	(4.65)	.52
1587	1554	1352	1554	1551	−.13	−.05	−.35	.26	−.32
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

Example 3.15 (Detecting outliers in the data on lumber) Table 4.4 contains the data in Table 4.3, along with the standardized observations. These data consist of four different measurements of stiffness x_1, x_2, x_3 and x_4, on each $n = 30$ boards. Detect outliers in these data.

Table 4.4 Four Measurements of Stiffness with Standardized Values

x_1	x_2	x_3	x_4	Observation no.	z_1	z_2	z_3	z_4	d^2
1889	1651	1561	1778	1	−.1	−.3	.2	.2	.60
2403	2048	2087	2197	2	1.5	.9	1.9	1.5	5.48
2119	1700	1815	2222	3	.7	−.2	1.0	1.5	7.62
1645	1627	1110	1533	4	−.8	−.4	−1.3	−.6	5.21
1976	1916	1614	1883	5	.2	.5	.3	.5	1.40
1712	1712	1439	1546	6	−.6	−.1	−.2	−.6	2.22
1943	1685	1271	1671	7	.1	−.2	−.8	−.2	4.99
2104	1820	1717	1874	8	.6	.2	.7	.5	1.49
2983	2794	2412	2581	9	3.3	3.3	3.0	2.7	12.26
1745	1600	1384	1508	10	−.5	−.5	−.4	−.7	.77
1710	1591	1518	1667	11	−.6	−.5	.0	−.2	1.93
2046	1907	1627	1898	12	.4	.5	.4	.5	.46
1840	1841	1595	1741	13	−.2	.3	.3	.0	2.70
1867	1685	1493	1678	14	−.1	−.2	−.1	−.1	.13
1859	1649	1389	1714	15	−.1	−.3	−.4	−.0	1.08
1954	2149	1180	1281	16	.1	1.3	−1.1	−1.4	16.85
1325	1170	1002	1176	17	−1.8	−1.8	−1.7	−1.7	3.50
1419	1371	1252	1308	18	−1.5	−1.2	−.8	−1.3	3.99
1828	1634	1602	1755	19	−.2	−.4	.3	.1	1.36
1725	1594	1313	1646	20	−.6	−.5	−.6	−.2	1.46
2276	2189	1547	2111	21	1.1	1.4	.1	1.2	9.90
1899	1614	1422	1477	22	−.0	−.4	−.3	−.8	5.06
1633	1513	1290	1516	23	−.8	−.7	−.7	−.6	.80
2061	1867	1646	2037	24	.5	.4	.5	1.0	2.54
1856	1493	1356	1533	25	−.2	−.8	−.5	−.6	4.58
1727	1412	1238	1469	26	−.6	−1.1	−.9	−.8	3.40
2168	1896	1701	1834	27	.8	.5	.6	.3	2.38
1655	1675	1414	1597	28	−.8	−.2	−.3	−.4	3.00
2326	2301	2065	2234	29	1.3	1.7	1.8	1.6	6.28
1490	1382	1214	1284	30	−1.3	−1.2	−1.0	−1.4	2.58

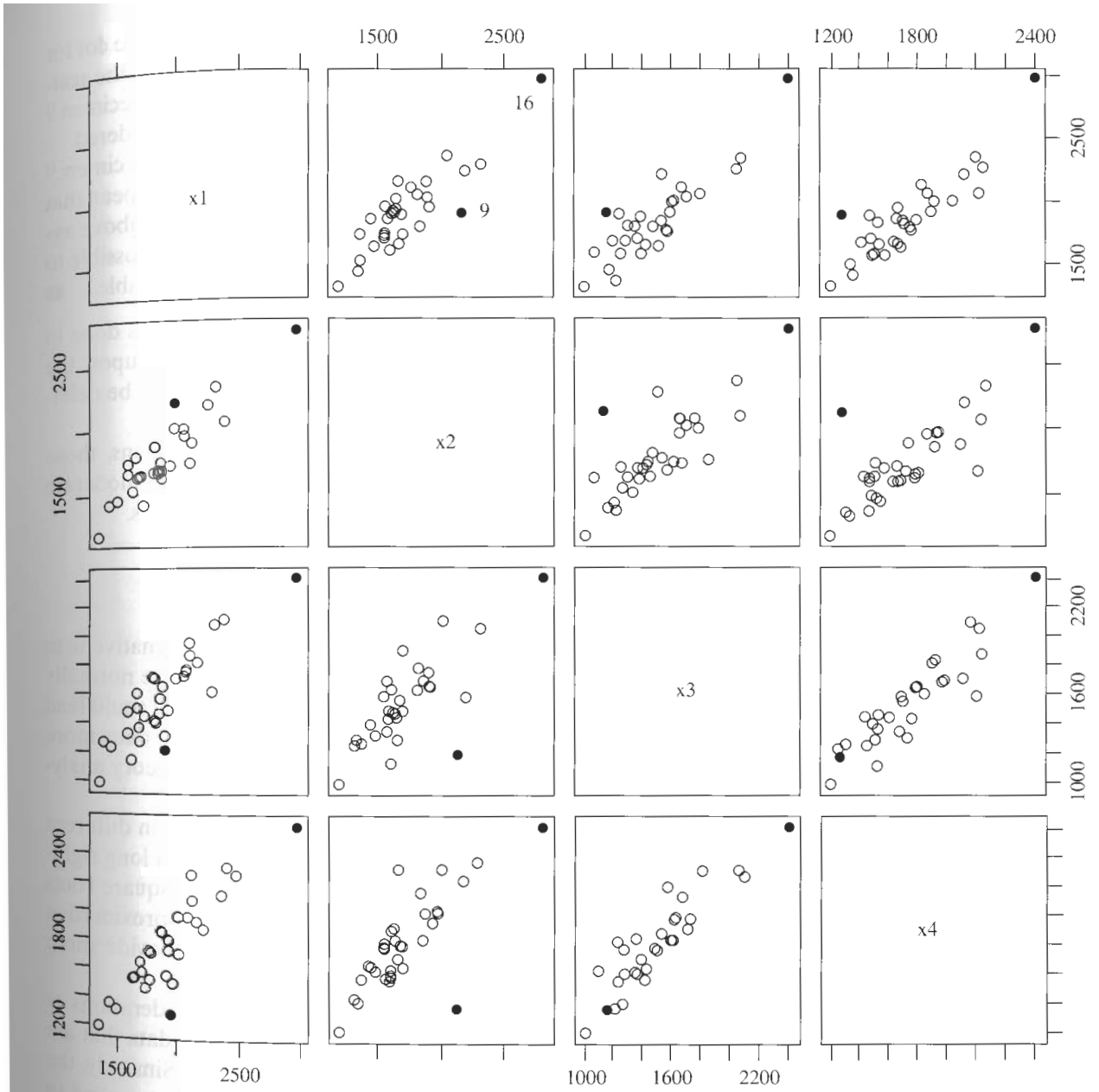


Figure 4.11 Scatter plots for the lumber stiffness data with specimens 9 and 16 plotted as solid dots.

3.8 Transformations to Near Normality

If normality is not a viable assumption, what is the next step ?

- Ignore the findings of a normality check and proceed as if the data were normality distributed. (*Not recommend*)

- Make nonnormal data more "normal looking" by considering ***transformations*** of data. Normal-theory analyses can then be carried out with the suitably transformed data.

Appropriate transformations are suggested by

1. theoretical consideration

2. the data themselves.

- **Helpful Transformations To Near Normality**

Original Scale	*Transformed Scale*
1. Counts, y	$\sqrt{y}$
2. Proportions, $\hat{p}$	logit $= \frac{1}{2}\log\left(\frac{\hat{p}}{1-\hat{p}}\right)$
3. Correlations, r	Fisher's $z(r) = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$

- **Box and Cox transformation**

$$x^{(\lambda)} = \begin{cases} \frac{x^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \ln x & \lambda = 0 \end{cases} \quad \text{or} \quad y_j^{(\lambda)} = \frac{x_j^{\lambda}-1}{\lambda\left[\left(\prod_{i=1}^{n} x_i\right)^{1/n}\right]^{\lambda-1}},\ j = 1, \ldots, n$$

Given the observations $x_1, x_2, \ldots, x_n$, the Box-Cox transformation for the choice of an appropriate power λ is the solution that maximizes the express

$$\ell(\lambda) = -\frac{n}{2}\ln\left[\frac{1}{n}\sum_{j=1}^{n}(x_j^{(\lambda)} - \bar{x^{(\lambda)}})^2\right] + (\lambda-1)\sum_{j=1}^{n}\ln x_j$$

where $\bar{x^{(\lambda)}} = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_j^{\lambda}-1}{\lambda}\right)$.

Example 3.16 (Determining a power transformation for univariate data) We gave readings of microwave radiation emitted through the closed doors of $n = 42$ ovens in Example 3.10. The Q-Q plot of these data in Figure 4.6 indicates that the observations deviate from what would be expected if they were normally distributed. Since all the positive observations are positive, let us perform a power transformation of the data which, we hope, will produce results that are more nearly normal. We must find that value of λ maximize the function $\ell(\lambda)$.

λ	$\ell(\lambda)$	λ	$\ell(\lambda)$
−1.00	70.52		
−.90	75.65	.40	106.20
−.80	80.46	.50	105.50
−.70	84.94	.60	104.43
−.60	89.06	.70	103.03
−.50	92.79	.80	101.33
−.40	96.10	.90	99.34
−.30	98.97	1.00	97.10
−.20	101.39	1.10	94.64
−.10	103.35	1.20	91.96
.00	104.83	1.30	89.10
.10	105.84	1.40	86.07
.20	106.39	1.50	82.88
.30	106.51		

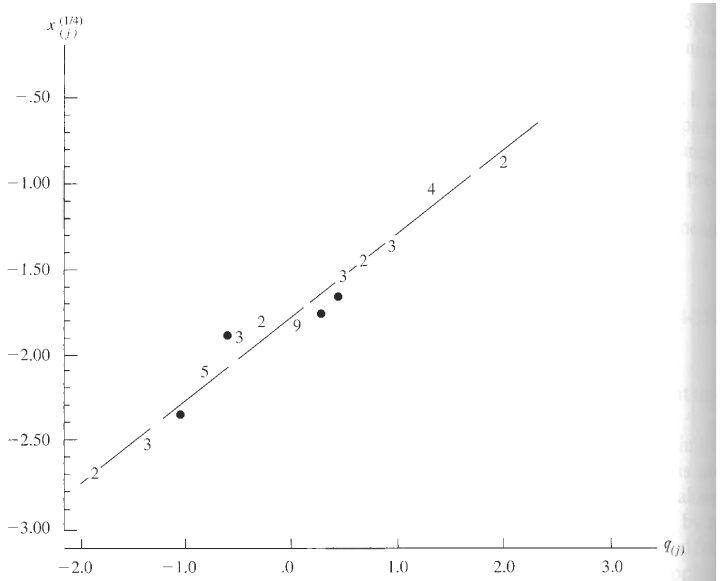


Figure 4.13 A Q–Q plot of the transformed radiation data (door closed). (The integers in the plot indicate the number of points occupying the same location.)

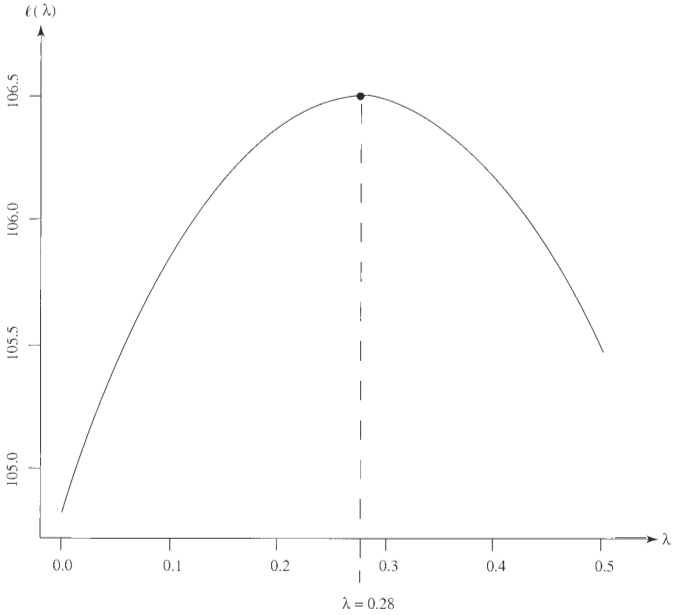


Figure 4.12 Plot of $\ell(\lambda)$ versus λ for radiation data (door closed).

Transforming Multivariate Observations

- With multivariate observations, a power transformation must be selected for each of the variables.
- Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the power transformations for the p measured characteristics. Each λ_k can be selected by *maximizing*

$$\ell(\lambda) = -\frac{n}{2}\ln\left[\frac{1}{n}\sum_{j=1}^{n}(x_{jk}^{(\lambda_k)} - \overline{x_k^{(\lambda_k)}})^2\right] + (\lambda_k - 1)\sum_{j=1}^{n}\ln x_{jk}$$

where $x_{1k}, x_{2k}, \ldots, x_{nk}$ are n observations on the kth variable, $k = 1, 2, \ldots, p$. Here

$$\overline{x_k^{(\lambda_k)}} = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_{jk}^{\lambda_k} - 1}{\lambda_k}\right)$$

- Let $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_p$ be the values that individually maximize the equation above. Then the jth transformed multivariate observation is

$$\mathbf{x}_j^{(\hat{\boldsymbol{\lambda}})} = \left[\frac{x_{j1}^{\hat{\lambda}_1} - 1}{\hat{\lambda}_1}, \frac{x_{j2}^{\hat{\lambda}_2} - 1}{\hat{\lambda}_2}, \cdots, \frac{x_{jp}^{\hat{\lambda}_p} - 1}{\hat{\lambda}_p}\right]'$$

- The procedure just described is equivalent to making each marginal distribution approximately normal. Although normal marginals are not sufficient to ensure that the joint distribution is normal, in practical applications this may be good enough.

- If not, the value $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_p$ can be obtained from the preceding transformations and iterate toward the set of values $\boldsymbol{\lambda}' = [\lambda_1, \lambda_2, \ldots, \lambda_p]$, which collectively maximizes

$$\ell(\lambda_1, \lambda_2, \ldots, \lambda_p) = -\frac{n}{2} \ln |\mathbf{S}(\boldsymbol{\lambda})| + (\lambda_1 - 1) \sum_{j=1}^{n} \ln x_{j1} + (\lambda_2 - 1) \sum_{j=1}^{n} \ln x_{j2}$$
$$+ \cdots + (\lambda_p - 1) \sum_{j=1}^{n} \ln x_{jp}$$

where $\mathbf{S}(\boldsymbol{\lambda})$ is the sample covariance matrix computed from

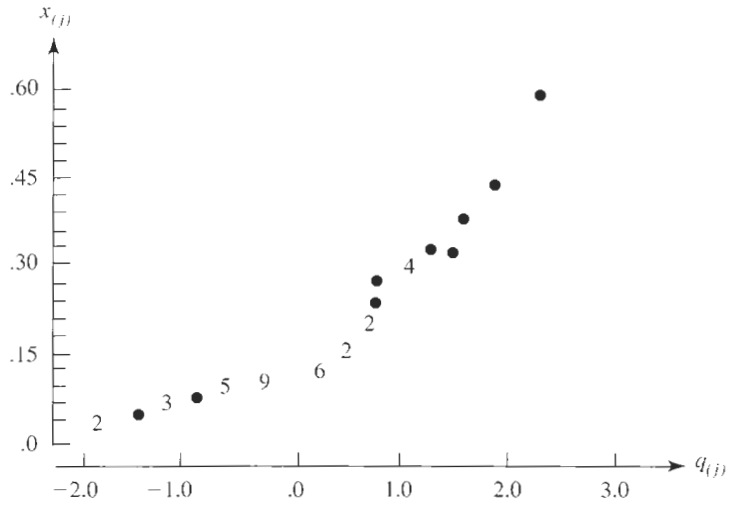
$$\mathbf{x}_j^{(\boldsymbol{\lambda})} = \left[\frac{x_{j1}^{\lambda_1} - 1}{\lambda_1}, \frac{x_{j2}^{\lambda_2} - 1}{\lambda_2}, \cdots, \frac{x_{jp}^{\lambda_p} - 1}{\lambda_p} \right]', \quad j = 1, 2, \ldots, n$$

Example 3.17 (Determining power transformations for bivariate data) Radiation measurements were also recorded though the open doors of the $n = 42$ micowave ovens introduced in Example 3.10. The amount of radiation emitted through the open doors of these ovens is list Table 4.5. Denote the door-close data $x_{11}, x_{21}, \ldots, x_{42,1}$ and the door-open data $x_{12}, x_{22}, \ldots, x_{42,2}$. Consider the joint distribution of x_1 and x_2, Choosing a power transformation for (x_1, x_2) to make the joint distribution of (x_1, x_2) approximately bivariate normal.

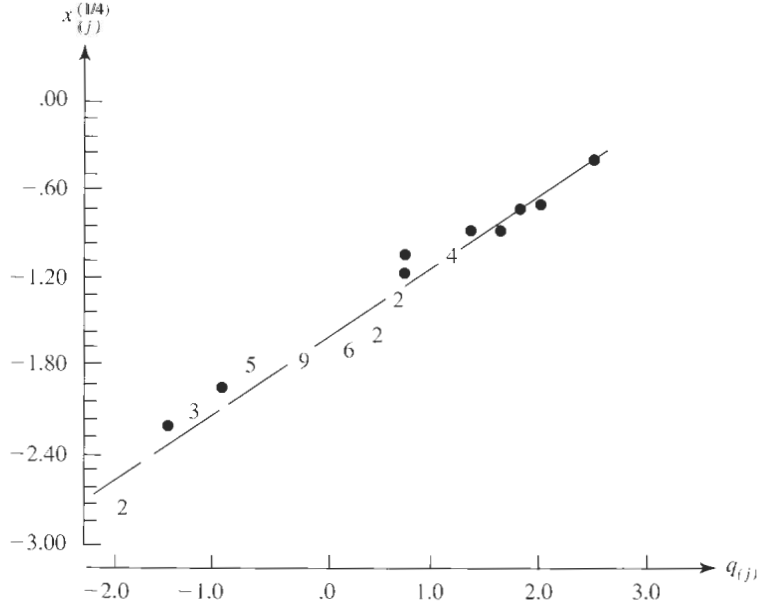
Table 4.5 Radiation Data (Door Open)

Oven no.	Radiation	Oven no.	Radiation	Oven no.	Radiation
1	.30	16	.20	31	.10
2	.09	17	.04	32	.10
3	.30	18	.10	33	.10
4	.10	19	.01	34	.30
5	.10	20	.60	35	.12
6	.12	21	.12	36	.25
7	.09	22	.10	37	.20
8	.10	23	.05	38	.40
9	.09	24	.05	39	.33
10	.10	25	.15	40	.32
11	.07	26	.30	41	.12
12	.05	27	.15	42	.12
13	.01	28	.09		
14	.45	29	.09		
15	.12	30	.28		

Source: Data courtesy of J. D. Cryer.



(a)



(b)

Figure 4.14 Q–Q plots of (a) the original and (b) the transformed radiation data (with door open). (The integers in the plot indicate the number of points occupying the same location.)

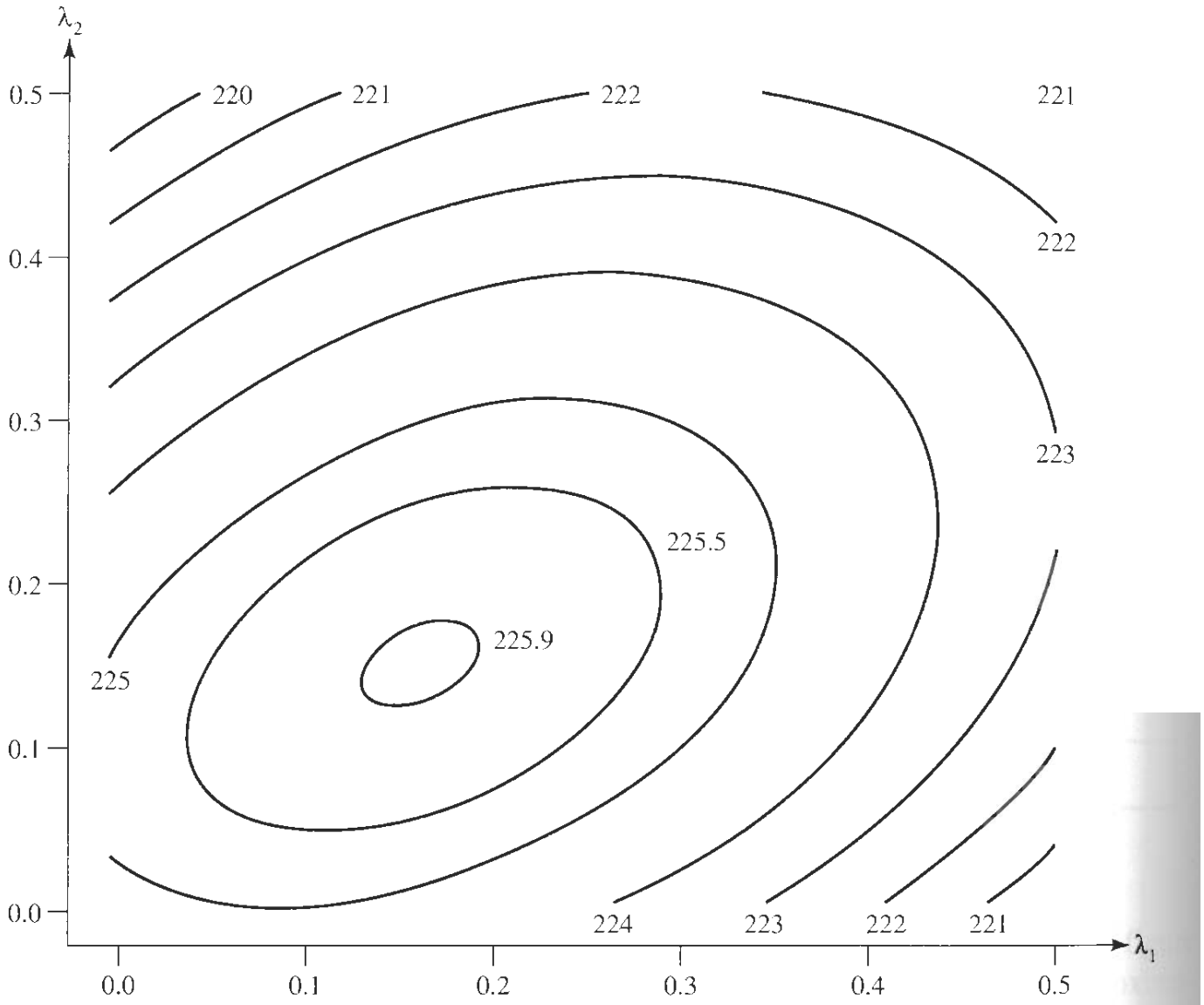


Figure 4.15 Contour plot of $\ell(\lambda_1, \lambda_2)$ for the radiation data.

If the data includes some large negative values and have a single long tail, a more general transformation should be applied.

$$x^{(\lambda)} = \begin{cases} \{(x+1)^{\lambda} - 1\}/\lambda & x \geq 0, \lambda \neq 0 \\ \ln(x+1) & x \geq 0, \lambda = 0 \\ -\{(-x+1)^{2-\lambda} - 1\}/(2-\lambda) & x < 0, \lambda \neq 2 \\ -\ln(-x+1) & x < 0, \lambda = 2 \end{cases}$$