STATS 206
Applied Multivariate Analysis
Lecture 2: Multivariate Normal Distribution

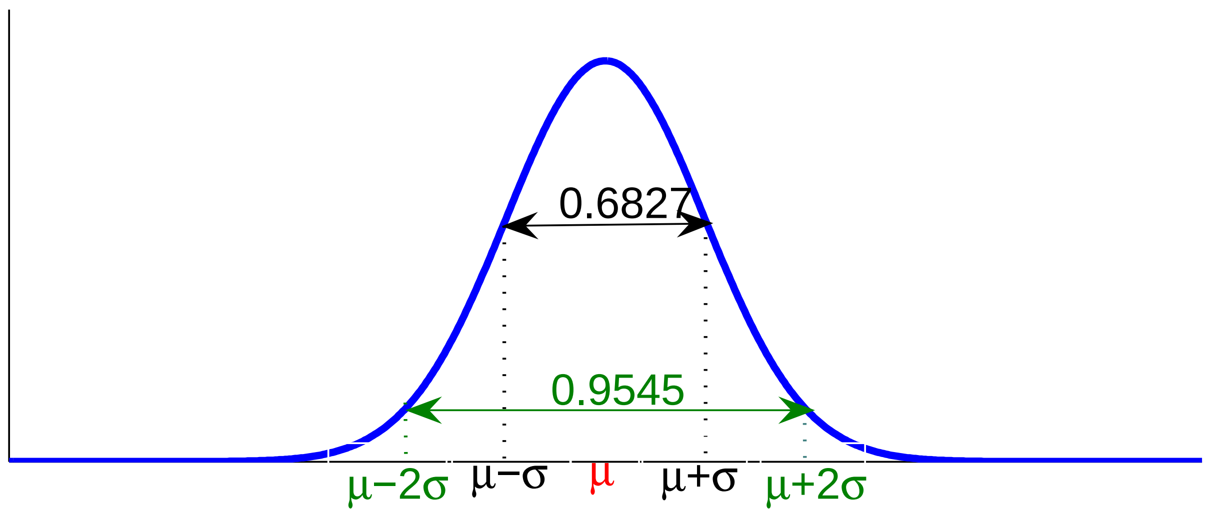
Prathapasinghe Dharmawansa

Department of Statistics
Stanford University
Autumn 2013

Univariate Normal Density

- P.d.f. of the univariate normal (Gaussian) distribution $N(\mu, \sigma^2)$

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



- Standard univariate normal/Gaussian density: $N(0, 1)$

Multivariate Normal Density

- Generalization for observations $\mathbf{x}_{p\times 1}$ of random vector $\mathbf{X}_{p\times 1}$:

$$\begin{array}{ccc} \text{univariate} & & \text{multivariate} \\ \left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu) & \rightarrow & (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \\ \frac{1}{\sqrt{2\pi\sigma^2}} = \frac{1}{(2\pi)^{1/2}(\sigma^2)^{1/2}} & \rightarrow & \frac{1}{\sqrt{|2\pi\boldsymbol{\Sigma}|}} = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \end{array}$$

$\boldsymbol{\mu}$: mean vector of $\mathbf{X} = [X_1, \ldots, X_p]'$, $\boldsymbol{\Sigma}$: covariance matrix of $\mathbf{X}$

- The p-dimensional multivariate normal density $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $\mathbf{X}$

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}}, \quad -\infty < x_i < \infty, \forall i.$$

Example: Bivariate Normal Density ($p = 2$)

- Let $\mathbf{X}_{2\times 1} = [X_1 \ X_2]'$

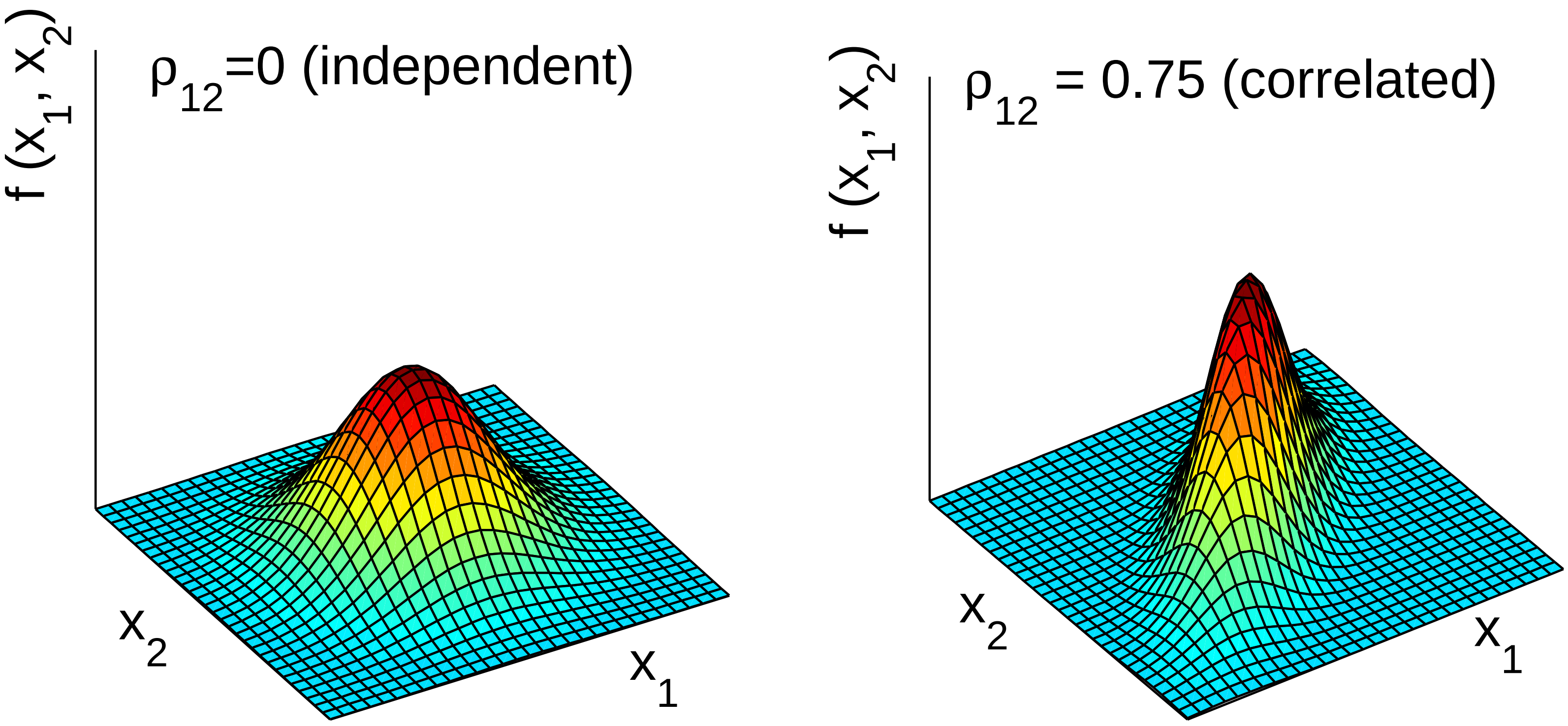
$$\mathsf{E}(\mathbf{X}) = [\mathsf{E}(X_1) \ \ \mathsf{E}(X_2)]' = [\mu_1 \ \mu_2]', \quad \sigma_{ii} = \mathsf{E}(X_i - \mu_i)^2, i = 1, 2$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}, \ \ \sigma_{12} = \mathsf{E}(X_1 - \mu_1)(x_2 - \mu_2), \ \ \rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$$

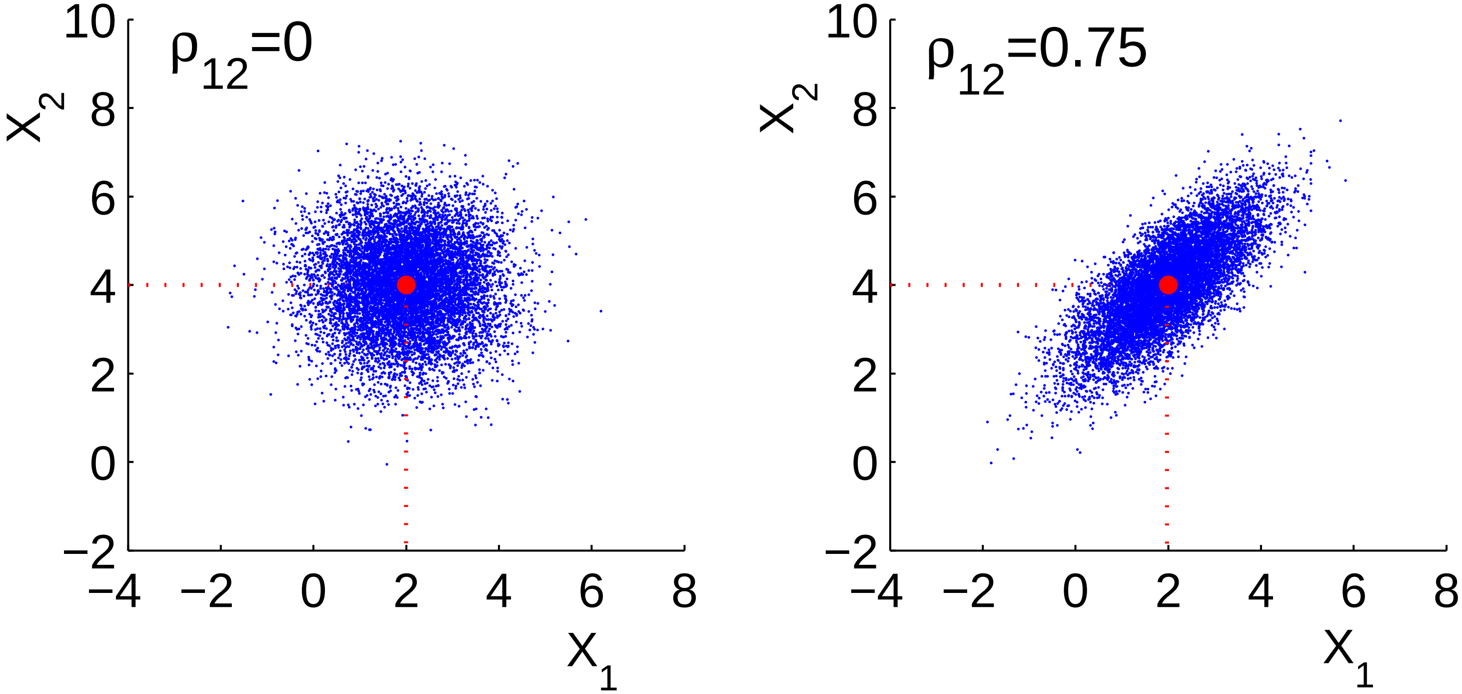
It can be shown that:

$$\begin{aligned} f(x_1, x_2) =& \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}} \\ & \times \exp\left\{ -\frac{1}{2(1-\rho_{12}^2)} \left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 \right.\right. \\ & \left.\left. -2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right) \right] \right\} \end{aligned}$$

Bivariate Normal Density Function: Figures ($\sigma_{11} = \sigma_{22}$)



Bivariate Normal Data Samples



Both figures: $n = 10^4$ data samples ((X_1, X_2) pairs)

Data generated with $\mu_1 = 2, \mu_2 = 4, \sigma_{11} = \sigma_{22} = 1$

Constant Probability Density Contour

- Constant prob. density contour $=$ surface of an ellipsoid centered at $\boldsymbol{\mu}$

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

  – $\boldsymbol{\Sigma}_{p\times p}$: positive definite. Recall:

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}', \quad \mathbf{P} = [\mathbf{e}_1\ \mathbf{e}_2\ \ldots\ \mathbf{e}_p], \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & \ldots & 0 \\ 0 & \ddots & 0 \\ 0 & \ldots & \lambda_p \end{bmatrix}$$

$$\Longrightarrow \boldsymbol{\Sigma}^{-1} = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}', \quad \boldsymbol{\Lambda}^{-1} = \begin{bmatrix} \lambda_1^{-1} & \ldots & 0 \\ 0 & \ddots & 0 \\ 0 & \ldots & \lambda_p^{-1} \end{bmatrix}$$

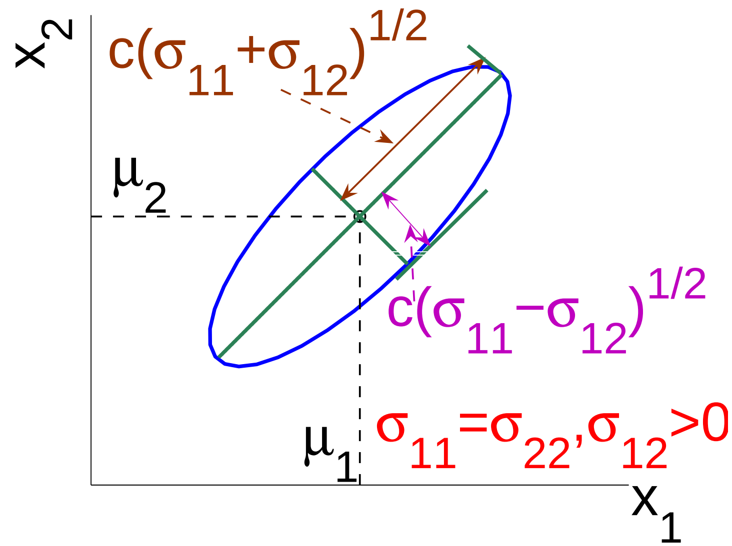
  – $(\mathbf{x}-\boldsymbol{\mu})$ must be aligned with one of the eigenvectors $(\mathbf{e}_i)$
    $\Longrightarrow$ Ellipsoids centered at $\boldsymbol{\mu}$ with axes $\pm c\sqrt{\lambda_i}\mathbf{e}_i, i = 1, \ldots, p$

Contours of Bivariate Normal Density

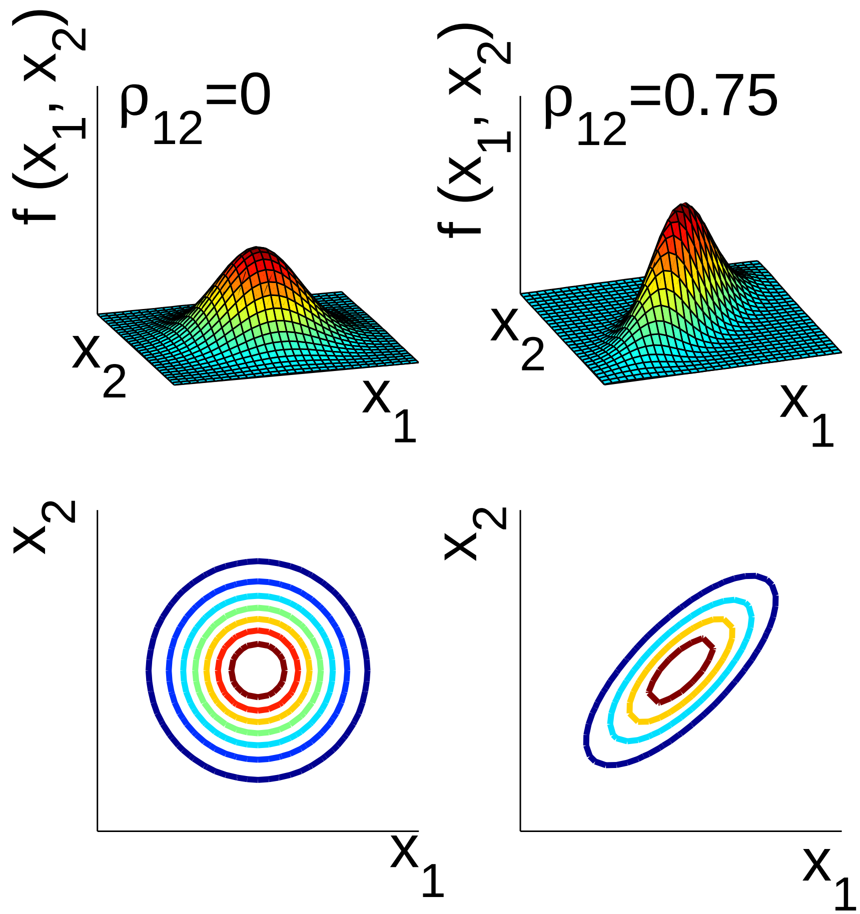
- For the bivariate normal density with $\underline{\sigma_{11} = \sigma_{22}}$,

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \Longrightarrow \begin{cases} \lambda_1 = \sigma_{11} + \sigma_{12}, \mathbf{e}_1 = [\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}}]' \\ \lambda_2 = \sigma_{11} - \sigma_{12}, \mathbf{e}_2 = [\frac{1}{\sqrt{2}} \quad \frac{-1}{\sqrt{2}}]' \end{cases}$$

Axes: $\pm c\sqrt{\sigma_{11} + \sigma_{12}}\, [\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}}]'$, $\quad \pm c\sqrt{\sigma_{11} - \sigma_{12}}\, [\frac{1}{\sqrt{2}} \quad \frac{-1}{\sqrt{2}}]'$



Contours of Bivariate Normal Density: Figure



Both figures: $\sigma_{11} = \sigma_{22}$

Properties of Multivariate Normal Distribution (1)

$\mathbf{X}_{p\times 1} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Longrightarrow$ linear combinations of X: normally distributed

- Linear combinations of components of $\mathbf{X}$
  - Given vector $\mathbf{a}_{p\times 1}$, $\mathbf{a}'\mathbf{X} = \sum_{i=1}^{p} a_i X_i \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ (univariate).
    Conversely: $\forall \mathbf{a}, \mathbf{a}'\mathbf{X} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}) \Longrightarrow \mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- Linear (affine) transformation of $\mathbf{X}$ (affine if $\mathbf{d} \neq \mathbf{0}$; see below)
  - Given (full-row rank) matrix $\mathbf{A}_{q\times p}$ and vector $\mathbf{d}_{q\times 1}$

$$\mathbf{A}\mathbf{X} + \mathbf{d} \sim N_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{d}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$$

Properties of Multivariate Normal Distribution (2)

$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Longrightarrow$ all subsets of $\mathbf{X}$: normally distributed

- In another word:

$$\mathbf{X}_{p\times 1} = \left[\begin{array}{c} \underset{(q\times 1)}{\mathbf{X}_1} \\ \hdashline \underset{((p-q)\times 1)}{\mathbf{X}_2} \end{array}\right] \qquad \boldsymbol{\mu}_{p\times 1} = \left[\begin{array}{c} \underset{(q\times 1)}{\boldsymbol{\mu}_1} \\ \hdashline \underset{((p-q)\times 1)}{\boldsymbol{\mu}_2} \end{array}\right]$$

$$\boldsymbol{\Sigma}_{p\times p} = \left[\begin{array}{c:c} \underset{(q\times q)}{\boldsymbol{\Sigma}_{11}} & \underset{(q\times(p-q))}{\boldsymbol{\Sigma}_{12}} \\ \hdashline \underset{((p-q)\times q)}{\boldsymbol{\Sigma}_{21}} & \underset{((p-q)\times(p-q))}{\boldsymbol{\Sigma}_{22}} \end{array}\right] \qquad (q \le p)$$

$$\Longrightarrow \mathbf{X}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$$

Properties of Multivariate Normal Distribution (3)

For normal random variables, "uncorrelated" $\Longrightarrow$ "independent"

$$\text{If } \left[\begin{array}{c} \mathbf{X}_1 \\ \hdashline \mathbf{X}_2 \end{array}\right] \sim N_{q_1+q_2}\left(\left[\begin{array}{c} \boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2 \end{array}\right], \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right]\right)$$

then $\mathbf{X}_1, \mathbf{X}_2$ independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$

- Note: for general random variables, we have only

$$\mathbf{X}_1, \mathbf{X}_2 \text{ independent} \Longrightarrow \text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{0}$$

- Furthermore, for normal $\mathbf{X}_1, \mathbf{X}_1$

If $\mathbf{X}_1, \mathbf{X}_2$ independent, $\mathbf{X}_1 \sim N_{q_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}), \mathbf{X}_2 \sim N_{q_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$

$$\text{Then } \left[\begin{array}{c} \mathbf{X}_1 \\ \hdashline \mathbf{X}_2 \end{array}\right] \sim N_{q_1+q_2}\left(\left[\begin{array}{c} \boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2 \end{array}\right], \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \hdashline \mathbf{0}' & \boldsymbol{\Sigma}_{22} \end{array}\right]\right)$$

Properties of Multivariate Normal Distribution (4)

$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Longrightarrow$ conditional dist. of a subset: normally distributed

- Let $\mathbf{X} = \left[\begin{array}{c} \underset{(q\times 1)}{\mathbf{X}_1} \\ \hdashline \underset{((p-q)\times 1)}{\mathbf{X}_2} \end{array} \right] \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $q \le p$

$$\boldsymbol{\mu} = \left[\begin{array}{c} \boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2 \end{array} \right], \boldsymbol{\Sigma} = \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array} \right], |\boldsymbol{\Sigma}_{22}| > 0$$

Then the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$

$$\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2 \sim N_q \left(\underbrace{\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)}_{\text{mean}}, \underbrace{\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}}_{\text{Covariance}} \right)$$

The covariance <u>does NOT</u> depend on the value of $\mathbf{x}_2$.

Properties of Multivariate Normal Distribution (5)

$$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ with } |\boldsymbol{\Sigma}| > 0$$

- (χ_p^2: chi-square distribution with p degrees freedom)

$$(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi_p^2$$

- ($\chi_p^2(\alpha)$: the upper $(100\alpha)^{\text{th}}$ percentile of χ_p^2)

$$P\left\{(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \le \chi_p^2(\alpha)\right\} = 1 - \alpha$$

i.e., the $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ assigns prob. $(1 - \alpha)$ to the solid ellipsoid

$$\left\{\mathbf{x} : (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \le \chi_p^2(\alpha)\right\}$$

(Proof on next page)

Proving $(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi_p^2 \quad (\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}))$

1. $\chi_p^2 \triangleq$ the dist. of $\sum_{i=1}^p Z_i^2 \qquad Z_1, \ldots, Z_p$: i.i.d. $N(0,1)$
2. $|\boldsymbol{\Sigma}| > 0 \Longrightarrow$ let $\mathbf{B} = \boldsymbol{\Sigma}^{1/2}$; here $\mathbf{B} = \mathbf{B}'$, $\mathbf{B}\mathbf{B}' = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$

   Also let $\mathbf{A} = \mathbf{B}^{-1} = \boldsymbol{\Sigma}^{-1/2}$; here $\mathbf{A} = \mathbf{A}'$, $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \mathbf{A}\mathbf{B}\mathbf{B}'\mathbf{A}' = \mathbf{I}$
3. Let $\mathbf{Z} \triangleq \mathbf{A}(\mathbf{X}-\boldsymbol{\mu})$ [$\mathbf{Z}$ normal since $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (previous properties)]

$$
\begin{aligned}
\mathsf{E}(\mathbf{Z}) &= \mathbf{A}\mathsf{E}(\mathbf{X}-\boldsymbol{\mu}) = \mathbf{0} \\
\mathsf{Cov}(\mathbf{Z}) &= \mathsf{E}(\mathbf{Z}\mathbf{Z}') = \mathsf{E}\left\{\mathbf{A}(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\mathbf{A}'\right\} \\
&= \mathbf{A}\mathsf{E}\left\{(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\right\}\mathbf{A}' = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \mathbf{I} \\
&\Longrightarrow \mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I}) \;\; (Z_i \sim N(0,1), \forall i)
\end{aligned}
$$

$$
(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \stackrel{2.}{=} (\mathbf{X}-\boldsymbol{\mu})'\mathbf{A}'\mathbf{A}(\mathbf{X}-\boldsymbol{\mu}) \stackrel{3.}{=} \mathbf{Z}'\mathbf{Z} = \sum_{i=1}^{p} Z_i^2 \stackrel{1.,3.}{\sim} \chi_p^2
$$

Properties of Multivariate Normal Distribution (6)

$\mathbf{X}_1, \ldots, \mathbf{X}_n$ mutually independent, $\mathbf{X}_j \sim N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$ (same $\boldsymbol{\Sigma}, \forall j$)

$$\mathbf{V}_1 = c_1\mathbf{X}_1 + \ldots + c_n\mathbf{X}_n = \sum_{i=1}^{n} c_i\mathbf{X}_i, \quad \mathbf{V}_2 = \sum_{i=1}^{n} b_i\mathbf{X}_i$$

- $\mathbf{V}_1 \sim N_p\left(\sum_{j=1}^n c_j\boldsymbol{\mu}_j, \left(\sum_{j=1}^n c_j^2\right)\boldsymbol{\Sigma}\right)$
  (Note: $\sum_{j=1}^n c_j^2 = \|\mathbf{c}\|^2, \quad \mathbf{c} = [c_1, \ldots, c_n]'$; similar for $\mathbf{V}_2$)

- $\mathbf{V}_1, \mathbf{V}_2$ are jointly normal with covariance matrix

$$\begin{aligned}\text{Cov}\left(\begin{bmatrix}\mathbf{V}_1\\ \mathbf{V}_2\end{bmatrix}\right) &= \text{E}\left(\begin{bmatrix}\mathbf{V}_1 - \text{E}(\mathbf{V}_1)\\ \mathbf{V}_2 - \text{E}(\mathbf{V}_2)\end{bmatrix}\left[(\mathbf{V}_1 - \text{E}(\mathbf{V}_1))' \;\; (\mathbf{V}_2 - \text{E}(\mathbf{V}_2))'\right]\right)\\ &= \begin{bmatrix}\|\mathbf{c}\|^2\boldsymbol{\Sigma} & (\mathbf{b}'\mathbf{c})\boldsymbol{\Sigma}\\ (\mathbf{b}'\mathbf{c})\boldsymbol{\Sigma} & \|\mathbf{b}\|^2\boldsymbol{\Sigma}\end{bmatrix}\end{aligned}$$

$\Longrightarrow \mathbf{V}_1, \mathbf{V}_2$ independent if $\mathbf{b}'\mathbf{c} = \sum_{j=1}^n b_jc_j = 0$

Maximum Likelihood (ML) Estimation

- Likelihood function $L(\boldsymbol{\theta})$:
  - $L(\boldsymbol{\theta})$: multivariate data summarization (as a function of $\boldsymbol{\theta}$)
  - $\boldsymbol{\theta}$: (unknown) non-random (deterministic) population parameters

- $\widehat{\boldsymbol{\theta}}$ (if it exists): a *maximum likelihood estimate* if for all possible $\boldsymbol{\theta}$

$$L(\widehat{\boldsymbol{\theta}}) \geq L(\boldsymbol{\theta})$$

Equivalently

$$\ln L(\widehat{\boldsymbol{\theta}}) \geq \ln L(\boldsymbol{\theta})$$

Note: Property of $\ln$ (natural log): monotonic increasing

Multivariate Normal Likelihood

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: a random sample from the population $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $\mathbf{X}_1, \ldots, \mathbf{X}_n$: i.i.d.

$$\Longrightarrow \text{ joint density } = \prod_{j=1}^{n} f(\mathbf{x}_j) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{\sum_{j=1}^{n}(\mathbf{x}_j-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j-\boldsymbol{\mu})}{2}}$$

- Substituting the observed values $\mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_n = \mathbf{x}_n$ into the joint density $\Longrightarrow$ $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: a function of the population parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{\sum_{j=1}^{n}(\mathbf{x}_j-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j-\boldsymbol{\mu})}{2}}$$

Note: $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ has the same expression as the joint density

Maximum Likelihood Estimation of μ and Σ

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: a random sample from the population $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

ML estimator of $\boldsymbol{\mu} \Longrightarrow \widehat{\boldsymbol{\mu}} = \overline{\mathbf{X}} = \dfrac{\mathbf{X}_1 + \ldots + \mathbf{X}_n}{n}$

ML estimator of $\boldsymbol{\Sigma} \Longrightarrow \widehat{\boldsymbol{\Sigma}} = \dfrac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})' = \dfrac{n-1}{n}\mathbf{S}$

- $\mathbf{S}$ defined above:

$$\mathbf{S} \triangleq \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})', \qquad \mathrm{E}(\mathbf{S}) = \boldsymbol{\Sigma} \text{ (unbiased)}$$

From now on, we refer to $\mathbf{S}$ as the sample covariance matrix

Maximum Likelihood Estimation of μ and Σ

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: a random sample from the population $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

ML estimator of $\boldsymbol{\mu} \Longrightarrow \widehat{\boldsymbol{\mu}} = \overline{\mathbf{X}} = \dfrac{\mathbf{X}_1 + \ldots + \mathbf{X}_n}{n}$ (a random vector)

ML estimator of $\boldsymbol{\Sigma} \Longrightarrow \widehat{\boldsymbol{\Sigma}} = \dfrac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})' = \dfrac{n-1}{n}\mathbf{S}$

(a random matrix)

- ML estimators and ML estimates:
  - ML estimators: random variables
  - ML estimate of $\boldsymbol{\mu} = \overline{\mathbf{x}}$ (the observed values of ML estimator $\widehat{\boldsymbol{\mu}}$)
    ML estimate of $\boldsymbol{\Sigma} = \frac{1}{n}\sum(\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})'$ (the observed value of ML estimator $\widehat{\boldsymbol{\Sigma}}$)

Maximum Likelihood Estimation of μ and Σ

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: a random sample from the population $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\text{ML estimator of } \boldsymbol{\mu} \Longrightarrow \widehat{\boldsymbol{\mu}} = \overline{\mathbf{X}} = \frac{\mathbf{X}_1 + \ldots + \mathbf{X}_n}{n}$$

$$\text{ML estimator of } \boldsymbol{\Sigma} \Longrightarrow \widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})' = \frac{n-1}{n}\mathbf{S}$$

- The maximum of the likelihood is

$$L(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = \frac{1}{(2\pi)^{np/2}|\widehat{\boldsymbol{\Sigma}}|^{n/2}} e^{-np/2}$$

Note that it depends on the sample only through $|\widehat{\boldsymbol{\Sigma}}|$.

Maximum Likelihood Estimation of μ and Σ

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: a random sample from the population $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

ML estimator of $\boldsymbol{\mu} \Longrightarrow \widehat{\boldsymbol{\mu}} = \overline{\mathbf{X}} = \dfrac{\mathbf{X}_1 + \ldots + \mathbf{X}_n}{n}$

ML estimator of $\boldsymbol{\Sigma} \Longrightarrow \widehat{\boldsymbol{\Sigma}} = \dfrac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})' = \dfrac{n-1}{n}\mathbf{S}$

- General *invariance property* for all ML Estimators:

$$\boldsymbol{\theta} \xrightarrow{\text{ML}} \widehat{\boldsymbol{\theta}} \qquad \Longrightarrow \qquad h(\boldsymbol{\theta}) \xrightarrow{\text{ML}} h(\widehat{\boldsymbol{\theta}})$$

- For $\mathbf{X}_1, \ldots, \mathbf{X}_n$: random sample from the normal population $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$\overline{\mathbf{X}}, \mathbf{S}$: sufficient statistics

$\Longleftrightarrow$ all info. about $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ in the sample is contained in $\overline{\mathbf{x}}$ and $\mathbf{S}$

Sampling Distribution of $\overline{\mathbf{X}}$ and $\mathbf{S}$

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: random sample $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\overline{\mathbf{X}} = \frac{\mathbf{X}_1 + \ldots + \mathbf{X}_n}{n}$$

$$\mathbf{S} = \frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'$$

- $\overline{\mathbf{X}} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$ (Recall: $\mathrm{E}(\overline{\mathbf{X}}) = \boldsymbol{\mu}$, $\mathrm{Cov}(\overline{\mathbf{X}}) = \frac{1}{n}\boldsymbol{\Sigma}$, normal sample)
- $(n-1)\mathbf{S}$: Wishart distribution with $(n-1)$ degrees of freedom
- $\overline{\mathbf{X}}$ and $\mathbf{S}$ are independent.

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

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: indep. observ. (each: mean $\boldsymbol{\mu}$, finite covariance $\boldsymbol{\Sigma}$)

- (Law of large numbers) $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ ($Y_i's$: indep. observ., mean μ)

  $\overline{Y}$ converges to μ in prob., $n \to \infty$

- (Law of large numbers applied here)
  - $\overline{\mathbf{X}}$ converges to $\boldsymbol{\mu}$ in prob., $n \to \infty$
  - $\mathbf{S}$ converges to $\boldsymbol{\Sigma}$ in prob., $n \to \infty$

Note: here $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are not necessarily having a multivariate normal distribution!!

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

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: indep. observ. (each: mean $\boldsymbol{\mu}$, finite covariance $\boldsymbol{\Sigma}$)

- (Central Limit Theorem)

$$\sqrt{n}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \to_d N_p(\mathbf{0}, \boldsymbol{\Sigma})(n \text{ large})$$

  - Recall: $(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2_p \quad (\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}))$
  - Consequently, $n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \to_d \chi^2_p$, n large

- Further using the law of large numbers for $\mathbf{S}$ (non-singular $\boldsymbol{\Sigma}$ assumed)

$$n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \to_d \chi^2_p$$

Assessing the Assumption of Normality

Evaluating Univariate Normality $(p = 1)$

- Quantile-quantile (Q-Q) plots (available in many comp. softwares)
  - Order the observation to get $x_{(1)}, \ldots, x_{(n)}$

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$$

  - Get the values of $p_{(j)} = (j - 1/2)/n, \;\; j = 1, \ldots, n$
  - Calculate the standard normal quantiles $q_{(1)}, \ldots, q_{(n)}$

$$\int_{-\infty}^{q_{(j)}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = p_{(j)}$$

  - Plot $x_{(i)}$ versus $q_{(i)}, \forall i$
  - Examine the straightness

The Straightness of a Q-Q Plot

- Test based on correlation coefficient (r_Q) between $(q_{(j)}, x_{(j)})$

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

Reject the hypothesis of normality if $r_Q <$ its appropriate value

Table 4.2 Critical Points for the Q-Q plot
Corr. Coefficient Test for Normality

Sample size	Significance level α		
n	0.01	0.05	0.10
5	0.8299	0.8788	0.9032
10	0.8801	0.9198	0.9351
⋮	⋮	⋮	⋮

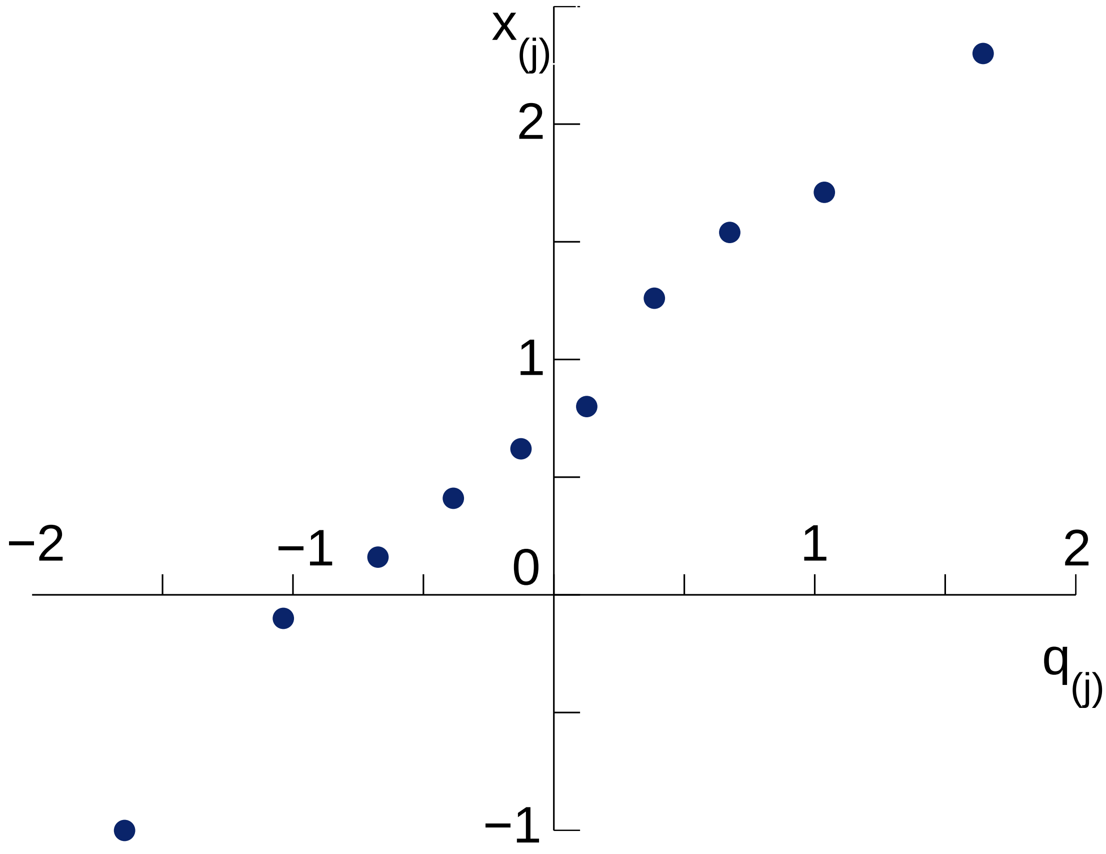
Example 1: The Sample (i)

Consider the following $n = 10$ observations:

Ordered observ. $x_{(j)}$	Prob. levels $(j - \frac{1}{2})/n$	Std. normal quantiles $q_{(j)}$
-1.00	0.05	-1.645
-0.10	0.15	-1.036
0.16	0.25	-0.674
0.41	0.35	-0.385
0.62	0.45	-0.125
0.80	0.55	0.125
1.26	0.65	0.385
1.54	0.75	0.674
1.71	0.85	1.306
2.30	0.95	1.645

This table is taken from Example 4.9 in the textbook.

Example 1: Q-Q plot (ii)



Example 1: Test based on correlation coefficient (r_Q)

Here : $\sum_{j=1}^{10}(x_{(j)} - \overline{x})q_{(j)} = 8.5849 \ (\overline{x} = 0.7700, \overline{q} = 0)$

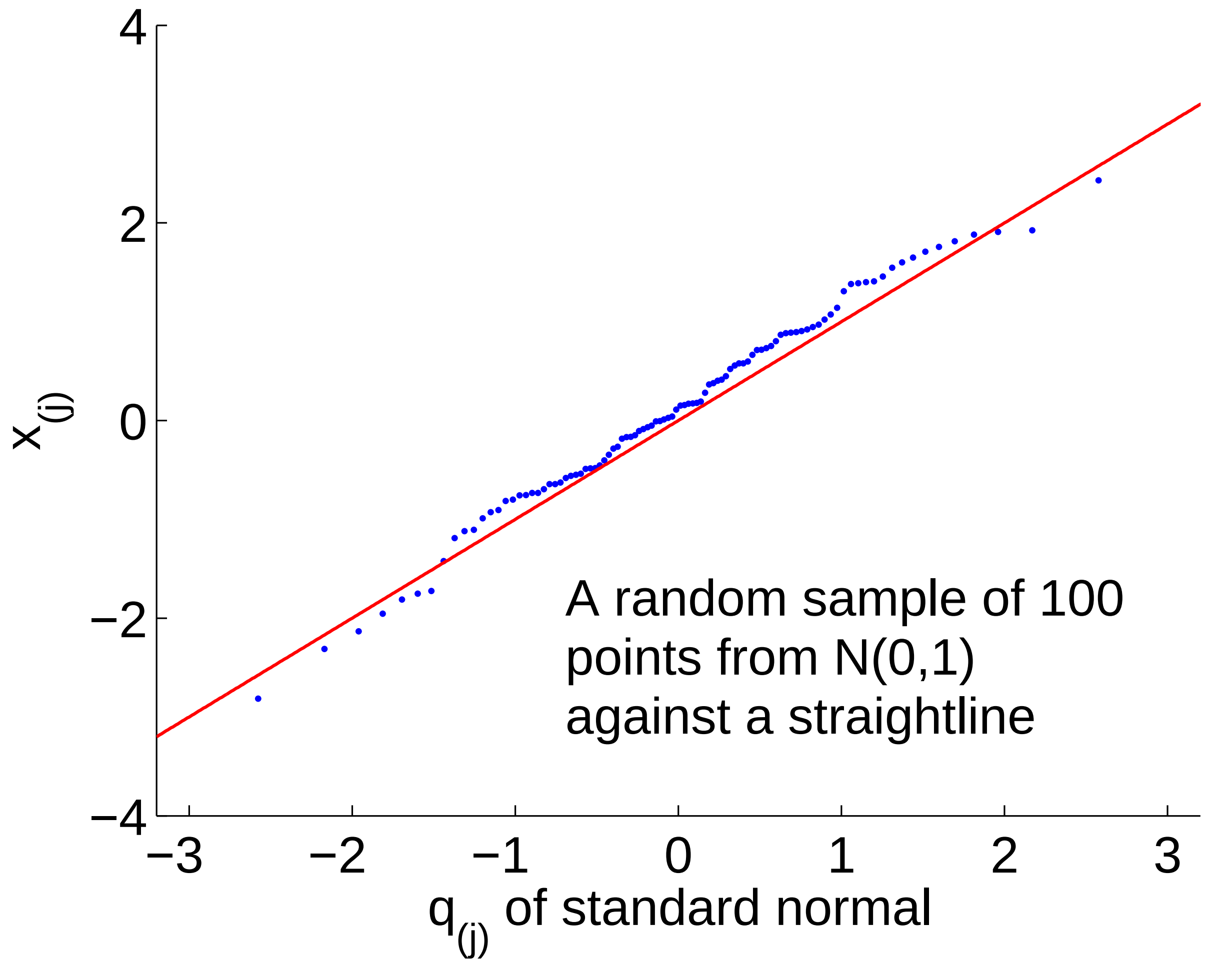
$$\sum_{j=1}^{10}(x_{(j)} - \overline{x})^2 = 8.4724, \ \sum_{j=1}^{10} q_{(j)}^2 = 8.7979$$

$$\Longrightarrow r_Q = \frac{8.5849}{\sqrt{8.4724 \times 8.7979}} = 0.9944 > 0.9351 \ (\alpha = 10\%)$$

Table 4.2 Critical Points for the Q-Q plot
Corr. Coefficient Test for Normality

Sample size	Significance level α		
n	0.01	0.05	0.10
5	0.8299	0.8788	0.9032
10	0.8801	0.9198	0.9351
⋮	⋮	⋮	⋮

Example 2: Q-Q Plot (Random Sample from $N(0,1)$)



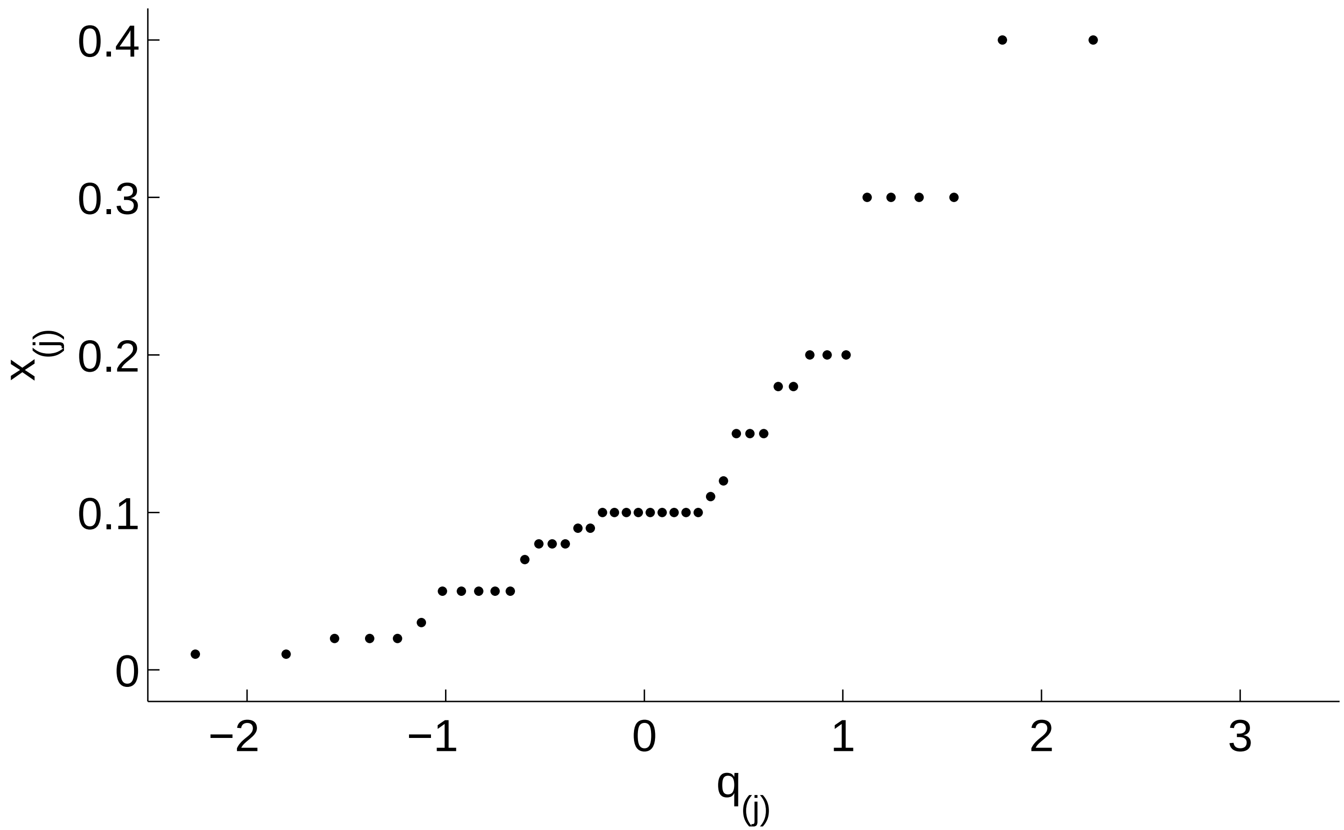
Example 3: Sample (Radiation Data) (i)

Consider the observations from the following table :

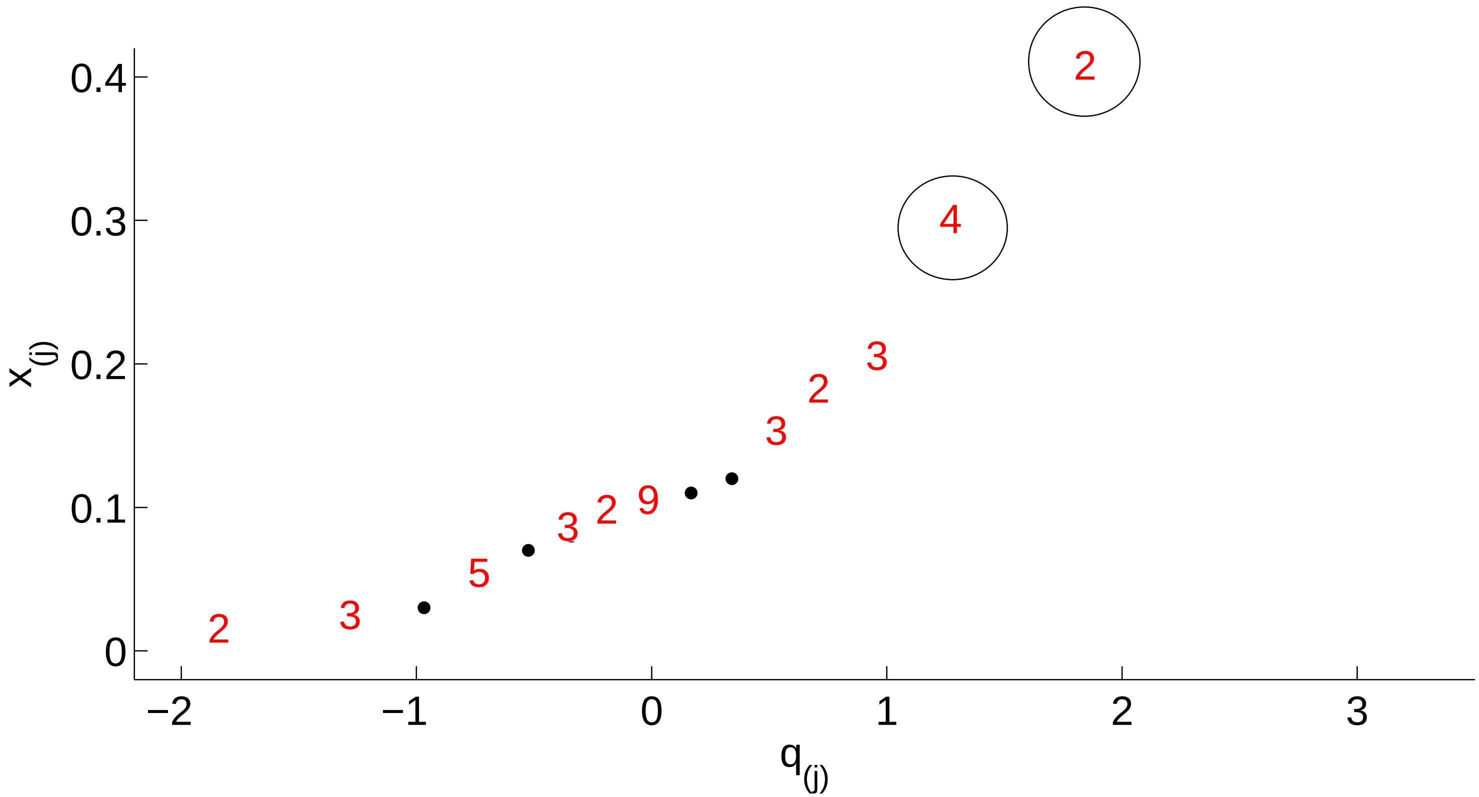
No.	Radiation	No.	Radiation	No.	Radiation
1	0.15	15	0.1	29	0.08
2	0.09	16	0.1	30	0.18
3	0.18	17	0.02	31	0.1
4	0.1	18	0.1	32	0.2
5	0.05	19	0.01	33	0.11
6	0.12	20	0.4	34	0.3
7	0.08	21	0.1	35	0.02
8	0.05	22	0.05	36	0.2
9	0.08	23	0.03	37	0.2
10	0.1	24	0.05	38	0.3
11	0.07	25	0.15	39	0.3
12	0.02	26	0.1	40	0.4
13	0.01	27	0.15	41	0.3
14	0.1	28	0.09	42	0.05

This is Table 4.1 in the textbook; the data: radiation from microwave ovens when doors are closed.

Example 3: Q-Q plot (Radiation Data) (ii)



Example 3: Q-Q plot (Radiation Data) (iii)



Assessing the Assumption of Normality

Evaluating Multivariate Normality $(p \geq 2)$

- Check univariate marginal normality
- Scatter plots
- Use a chi-square plot (χ^2 plot) (available in many comp. softwares):
  - Given $\mathbf{x}_j, j = 1, \ldots, n$, calculate $\overline{\mathbf{x}}, \mathbf{S}^{-1}$;
  - Order the squared generalized distances so that $d^2_{(1)} \leq \ldots \leq d^2_{(n)}$;

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

  - Graph the pairs $(q_{c,p}((j - \frac{1}{2})/n),\ d^2_{(j)})$

$$q_{c,p}((j - \frac{1}{2})/n) : 100(j - \frac{1}{2})/n \text{ quantile of } \chi_p^2$$

$$q_{c,p}((j - \frac{1}{2})/n) = \chi_p^2((n - j + \frac{1}{2})/n)$$

Using the χ^2 Plot

- Check whether around half of the d_j^2 are less than or equal to $q_{c,p}(0.5)$;
- Check whether the χ^2 plot is nearly a straight line having slope 1 passing through the origin.

Example 4: Company Data (i)

Consider the following data (not a random sample; $p = 2, n = 10$):

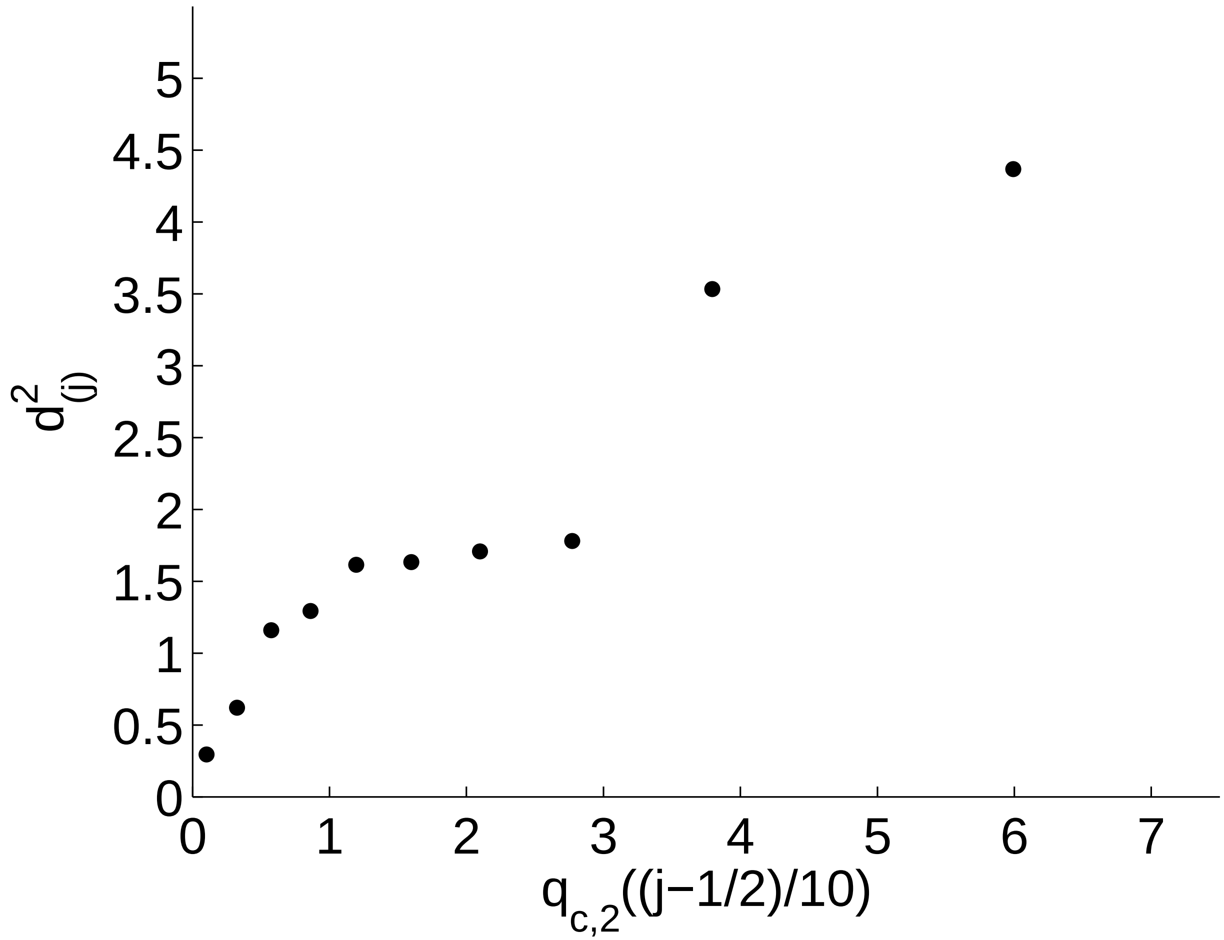
Company	x_1=sales (billions)	x_2=profits (billions)
Citigroup	108.28	17.05
General Electric	152.36	16.59
American Intl Group	95.04	10.91
Bank of America	65.45	14.14
HSBC Group	62.97	9.52
ExxonMobil	263.99	25.33
Royal Dutch/ Shell	265.19	18.54
BP	285.06	15.73
ING Group	92.01	8.10
Toyota Motor	165.68	11.13

This table is used in Example 4.13 in the textbook.

Example 4: Constructing the χ^2 Plot (ii)

j	$d^2_{(j)}$	$q_{c,2}(\frac{j-\frac{1}{2}}{10})$
1	0.30	0.1026
2	0.62	0.3250
3	1.16	0.5754
4	1.30	0.8616
5	1.61	1.1957
6	1.64	1.5970
7	1.71	2.0996
8	1.79	2.7726
9	3.53	3.7942
10	4.38	5.9915

Example 4: χ^2 Plot for the Company Data (iii)

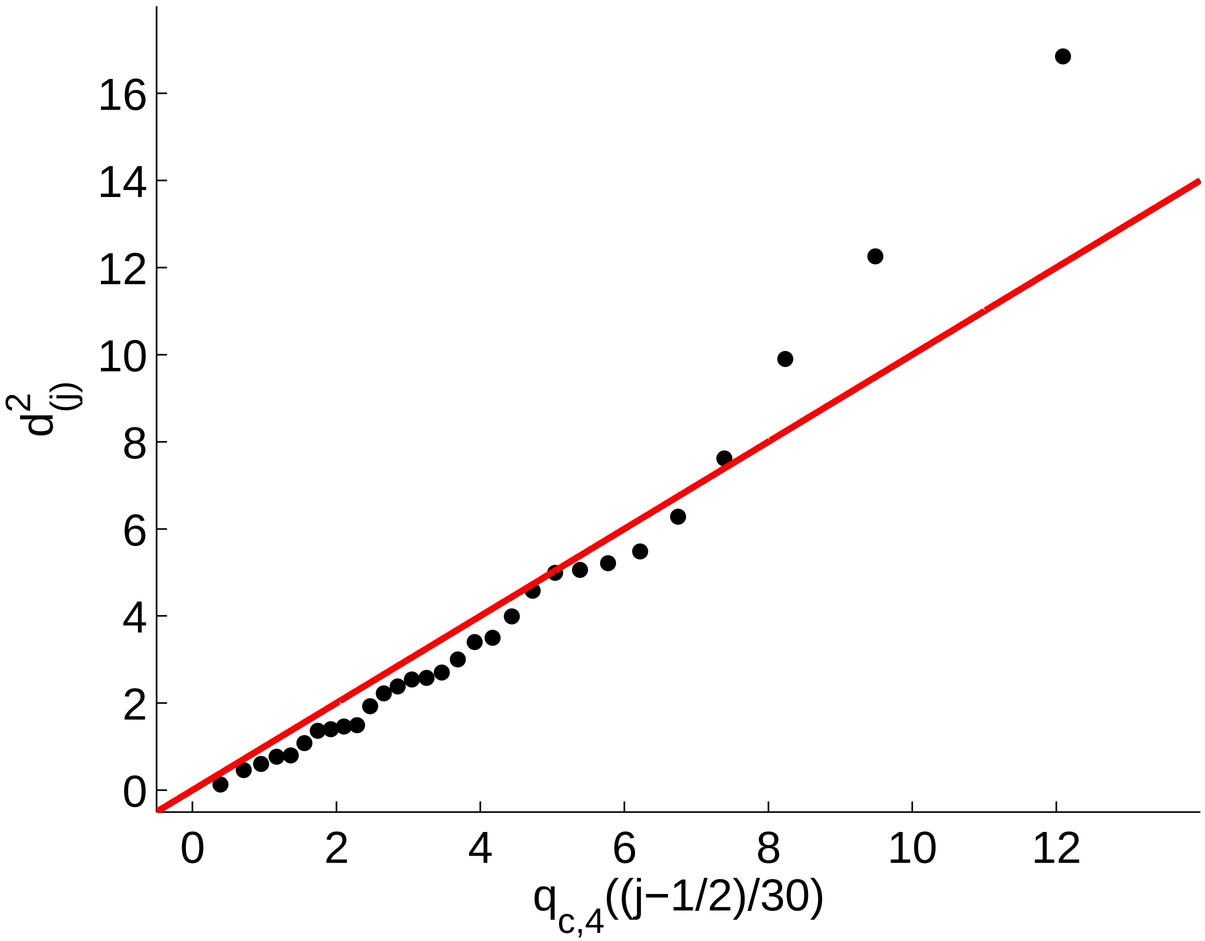


Example 5: Board Stiffness Data ($p = 4, n = 30$) (i)

Four Measurements of Stiffness											
No	x_1	x_2	x_3	x_4	d^2	No	x_1	x_2	x_3	x_4	d^2
1	1889	1651	1561	1778	0.6	16	1954	2149	1180	1281	16.85
2	2403	2048	2087	2197	5.48	17	1325	1170	1002	1176	3.5
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	1.4	20	1725	1594	1313	1646	1.46
6	1712	1712	1439	1546	2.22	21	2276	2189	1547	2111	9.9
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	0.8
9	2983	2794	2412	2581	12.26	24	2061	1867	1646	2037	2.54
10	1745	1600	1384	1508	0.77	25	1856	1493	1356	1533	4.58
11	1710	1591	1518	1667	1.93	26	1727	1412	1238	1469	3.4
12	2046	1907	1627	1898	0.46	27	2168	1896	1701	1834	2.38
13	1840	1841	1595	1741	2.7	28	1655	1675	1414	1597	3
14	1867	1685	1493	1678	0.13	29	2326	2301	2065	2234	6.28
15	1859	1649	1389	1714	1.08	30	1490	1382	1214	1284	2.58

Table 4.3 in the textbook (with d^2 included)

Example 5: χ^2 Plot for Stiffness Data (ii)



Outliers

- Outliers: Unusual observations that do not obey the pattern of the rest
  - Not always wrong (may be helpful for understanding the phenomena)
- Steps for Detecting outliers
  1. Examine the dot plot of each variable;
  2. Study the scatter plot for each pair of the variables;
  3. Examine the standardized values (z_{jk}) for large or small values;

$$z_{jk} = (x_{jk} - \overline{x}_k)/\sqrt{s_{kk}}, \forall j, k$$

  4. Examine $(\mathbf{x}_j - \overline{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_j - \overline{\mathbf{x}})$ for large values; χ^2 plots may help.

Important: For outliers in multivariate cases $(p > 2)$

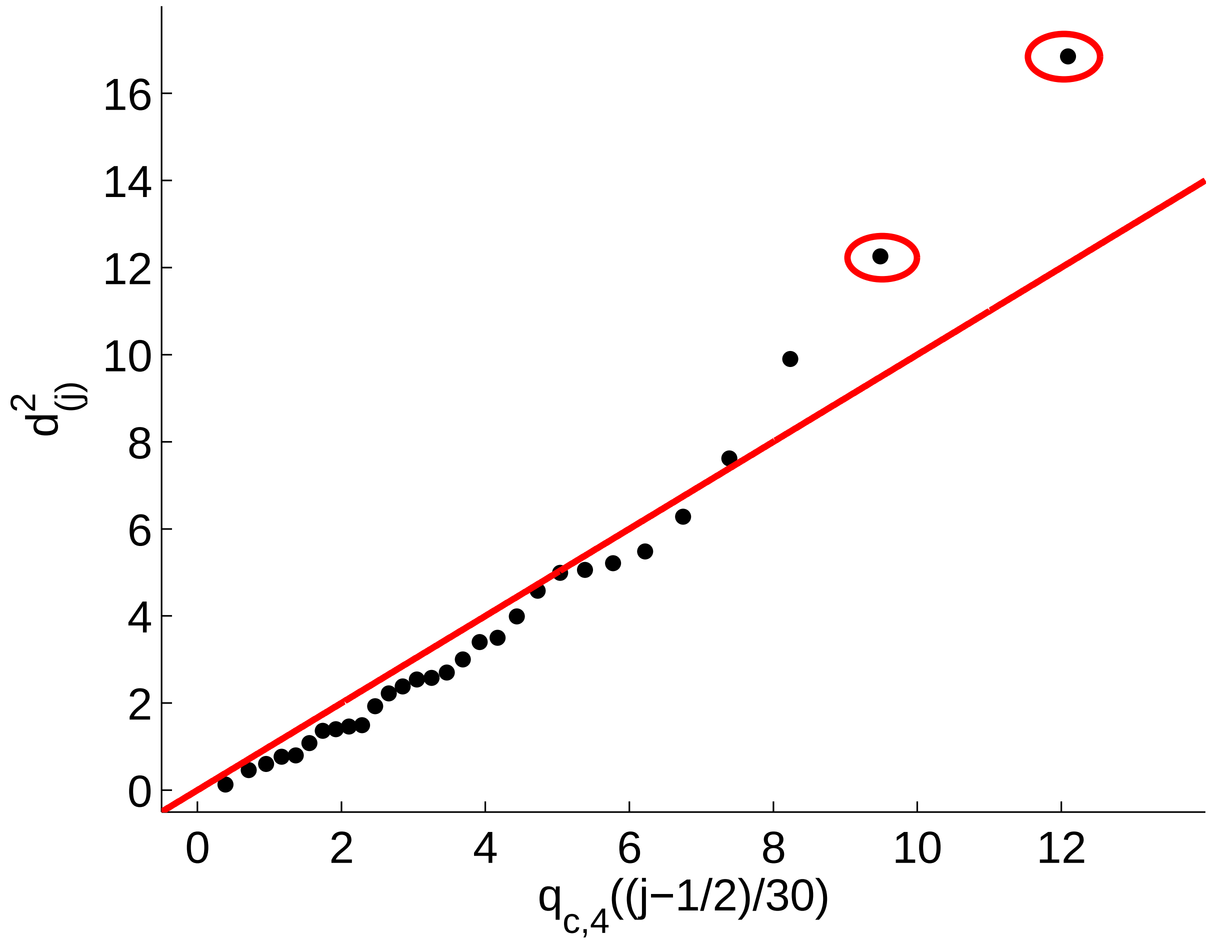
– Step 1 or Step 2 may not detect them; Step 4 is needed.

Example 6: Stiffness Data (Revisited) (i)

Four Measurements of Stiffness											
No	x_1	x_2	x_3	x_4	d^2	No	x_1	x_2	x_3	x_4	d^2
1	1889	1651	1561	1778	0.6	16	1954	2149	1180	1281	16.85
2	2403	2048	2087	2197	5.48	17	1325	1170	1002	1176	3.5
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	1.4	20	1725	1594	1313	1646	1.46
6	1712	1712	1439	1546	2.22	21	2276	2189	1547	2111	9.9
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	0.8
9	2983	2794	2412	2581	12.26	24	2061	1867	1646	2037	2.54
10	1745	1600	1384	1508	0.77	25	1856	1493	1356	1533	4.58
11	1710	1591	1518	1667	1.93	26	1727	1412	1238	1469	3.4
12	2046	1907	1627	1898	0.46	27	2168	1896	1701	1834	2.38
13	1840	1841	1595	1741	2.7	28	1655	1675	1414	1597	3
14	1867	1685	1493	1678	0.13	29	2326	2301	2065	2234	6.28
15	1859	1649	1389	1714	1.08	30	1490	1382	1214	1284	2.58

Table 4.3 in the textbook (with d^2 included)

Example 6: χ^2 Plot for Stiffness Data (Revisited) (ii)



Transformation to Near Normality

- Transformation: a reexpression of data in a different scale
- Some transformations (shown to yields nearly normal quantities):

Original Scale	—	Transformed Scale
Counts: y	—	$\sqrt{y}$
Proportions: $\widehat{p}$	—	$\text{logit}(\widehat{p}) = \frac{1}{2}\log\left(\frac{\widehat{p}}{1-\widehat{p}}\right)$
Correlations: r	—	$z(r) = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$ (Fisher's)

- For other instances, use Box-Cox transformation

Box-Cox Transformation: the Univariate Case ($p = 1$)

- Transformation from x to $x^{(\lambda)}$; λ: a parameter to be chosen properly

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln x & \lambda = 0 \end{cases} \qquad \text{(continuous in } \lambda \text{ for } \underline{x > 0}\text{)}$$

- Box-Cox power λ: Given $x_1, \ldots, x_n$, choosing λ to maximize

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

$$\overline{x^{(\lambda)}} \triangleq \frac{1}{n} \sum_{j=1}^{n} x_j^{(\lambda)}$$

<u>Use graphs/tables to help determine λ</u>

Example 7: Determining λ for Radiation Data (i)

(Same data as in Example 3) Observations from the following table :

No.	Radiation	No.	Radiation	No.	Radiation
1	0.15	15	0.1	29	0.08
2	0.09	16	0.1	30	0.18
3	0.18	17	0.02	31	0.1
4	0.1	18	0.1	32	0.2
5	0.05	19	0.01	33	0.11
6	0.12	20	0.4	34	0.3
7	0.08	21	0.1	35	0.02
8	0.05	22	0.05	36	0.2
9	0.08	23	0.03	37	0.2
10	0.1	24	0.05	38	0.3
11	0.07	25	0.15	39	0.3
12	0.02	26	0.1	40	0.4
13	0.01	27	0.15	41	0.3
14	0.1	28	0.09	42	0.05

This is Table 4.1 in the textbook; the data: radiation from microwave ovens when doors are closed.

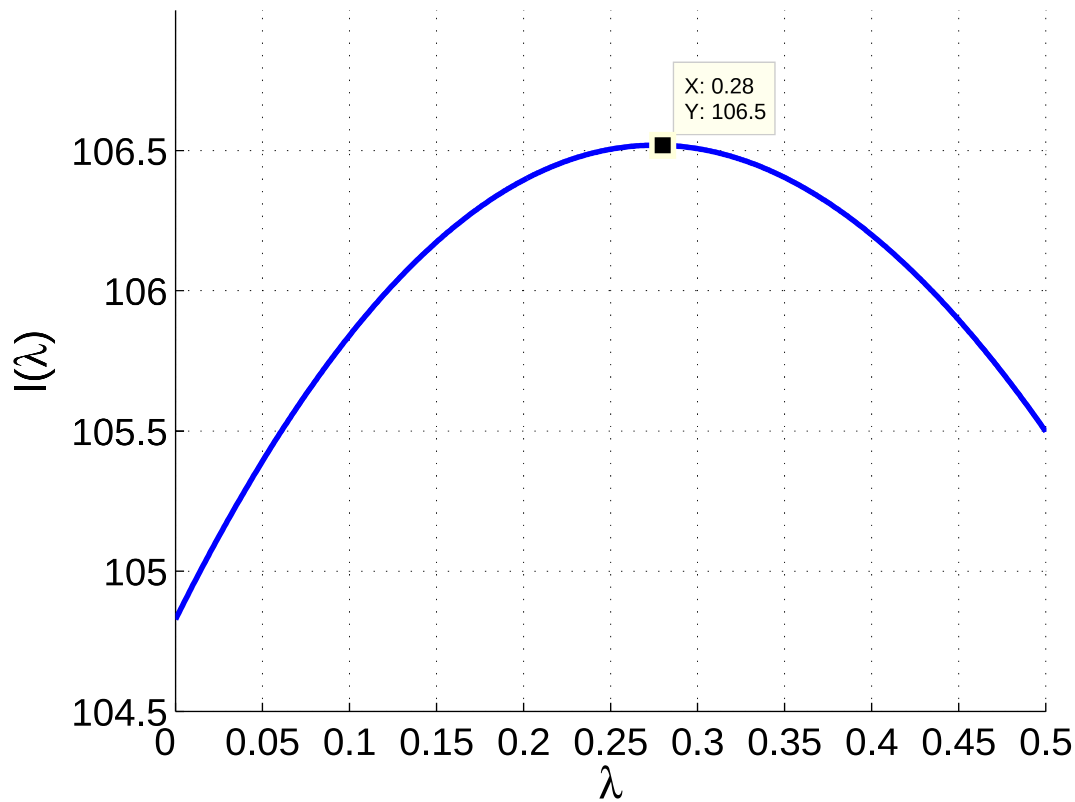
Example 7: Determining λ for Radiation Data (ii)

Table ($l(\lambda)$ vs. λ)

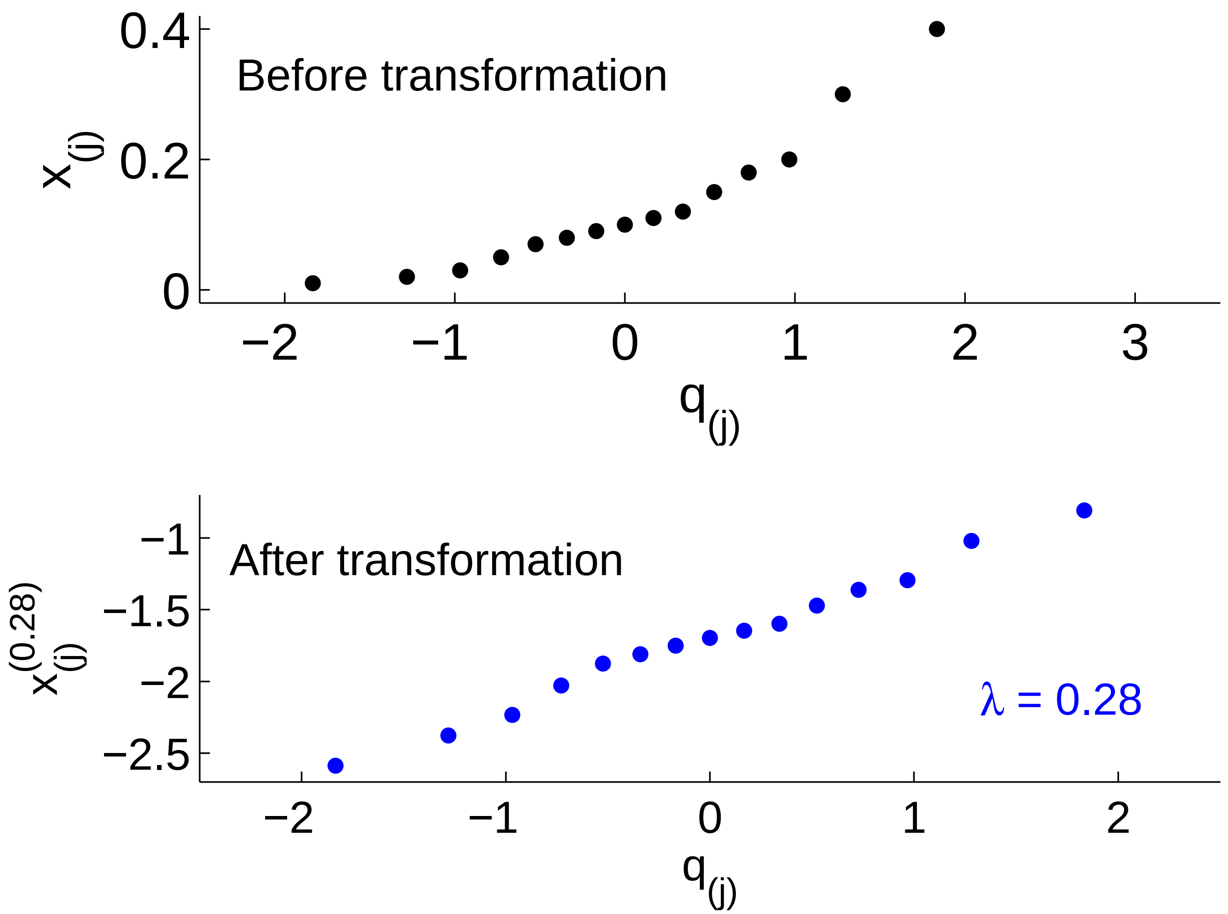
λ	$l(\lambda)$	λ	$l(\lambda)$	λ	$l(\lambda)$
-1.0	70.5227	-0.1	103.3457	0.8	101.3254
-0.9	75.6472	0.0	104.8276	0.9	99.3404
-0.8	80.4626	0.1	105.8406	1.0	97.1031
-0.7	84.9421	0.2	106.3948	1.1	94.6373
-0.6	89.0587	0.3	106.5070	1.2	91.9644
-0.5	92.7855	0.4	106.1995	1.3	89.1034
-0.4	96.0975	0.5	105.4986	1.4	86.0714
-0.3	98.9723	0.6	104.4330	1.5	82.8833
-0.2	101.3923	0.7	103.0322		

Example 7: Determining λ for Radiation Data (iii)

Graph ($l(\lambda)$ vs. λ)



Example 7: Q-Q Plot Using Transformed Data ($\lambda = 0.28$) (iv)



Box-Cox Transformation: the Multivariate Case ($p \geq 2$)

- Apply univariate Box-Cox transformation to each variable and select $\lambda_k, k = 1, \ldots, p$ individually

  $\Longleftrightarrow$ making each marginal distribution approximately normal

- Jointly select $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_p]'$ to maximize

$$l(\lambda_1, \ldots, \lambda_p) = -\frac{n}{2} \ln |\mathbf{S}(\boldsymbol{\lambda})| + \sum_{k=1}^{p} (\lambda_k - 1) \sum_{j=1}^{n} \ln x_{jk}$$

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

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

$\underline{p = 2}$: contour plots of $l(\lambda_1, \lambda_2)$ help to find the maximizing (λ_1, λ_2).