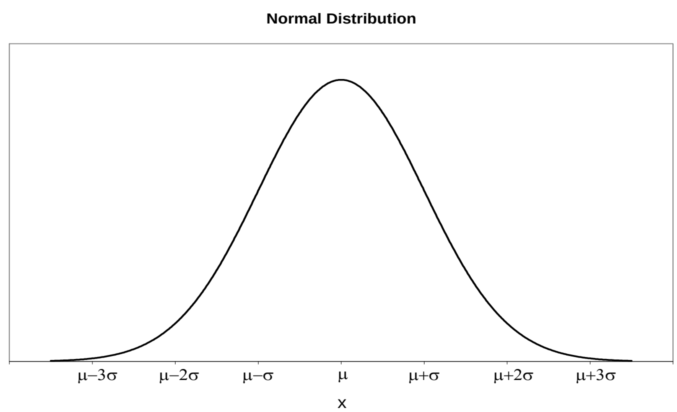
The Multivariate Normal Distribution

Probability Density Function of Univariate Normal Distribution

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

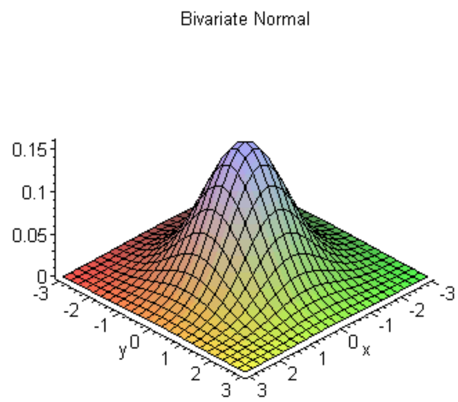
Graph of the Univariate Normal Distribution



Probability Density Function of Multivariate Normal Distribution

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

Graph of the Bivariate Normal Distribution



$$\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_p^2 \end{bmatrix}$$

$\mu_j = \mathrm{E}\left(x_j\right) =$ mean of the j^{th} response variable

$\sigma_j^2 =$ variance of x_j, the j^{th} response variable

$\sigma_{ij} =$ covariance between x_i and x_j

Let $\mathrm{y} = \mathbf{a}'\mathbf{x}$

where $\mathbf{a}' = \begin{pmatrix} a_1 & a_2 & \cdots & a_p \end{pmatrix}$

is a vector of constants

$$\mathbf{x} \sim \mathrm{N_p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

iff

$$\mathrm{y} \sim \mathrm{N}(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$$

Assessing Multivariate Normality

Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_N \sim iid\, \mathrm{N}_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$

$$\bar{\mathbf{X}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i$$

$$\mathbf{S} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)'$$

Assessing Multivariate Normality

- To assess whether the data arise from a multivariate normal distribution, a quantile plot can be used.
- The procedure is to compute a quantity from each multivariate observation, such that this quantity follows a known probability distribution when the data follow a multivariate normal distribution.
- Then a Q-Q plot of this quantity against the quantiles of the reference distribution will plot as a straight line.

Assessing Multivariate Normality

- For each multivariate observation, compute the squared Mahalanobis distance between that observation and the sample mean vector:

$$\mathrm{D}_i^2 = \left(\mathbf{X}_i - \overline{\mathbf{X}}\right)' \mathbf{S}^{-1} \left(\mathbf{X}_i - \overline{\mathbf{X}}\right)$$

Mahalanobis Distance

- D^2 is the multivariate analog of the square of the standard score for a single variable, z^2, which measures the distance from the mean in standard deviation units.
- D^2 measures the distance from the mean vector in relation to the variance-covariance matrix, which takes into account the precision of the variables as well as their intercorrelations.

Mahalanobis Distance

With p variables, $D^2 \sim \chi_p^2$

Therefore, a Q-Q plot of the ordered distance values

$$D_{(i)}^2$$

against the corresponding quantiles of χ_p^2 should yield a straight line.

Assessing Multivariate Normality

- The SAS Macro cqplot (contained in the file cqplot.sas) can be used to produce a Q-Q plot to assess multivariate normality.