Boxplot

Você está aqui: [Home](http://marcosfs2006.googlepages.com/home) > [Usando R em Auditoria](http://marcosfs2006.googlepages.com/usandoremauditoria) > **Boxplot**

O boxplot é um gráfico que possibilita representar a distribuição de um conjunto de dados com base em alguns de seus parâmetros descritivos, quais sejam: a mediana (q2), o quartil inferior (q1), o quartil superior (q3) e do intervalo interquartil (IQR = q3 - q1).

A figura a seguir apresenta o boxplot, destacando suas principaiscaracterísticas:



A linha central da caixa marca a mediana do conjunto de dados. A parte inferior da caixa é delimitada pelo quartil inferior (q1) e a parte superior pelo quartil superior (q3). As hastes inferiores e superiores se estendem, respectivamente, do quartil inferior até o menor valor não inferior a q1 - 1.5IQR e do quartil superior até o maior valor não superior a q3 + 1.5IQR.  Os valores inferiores a q1 - 1.5IQR e superiores a q3 + 1.5IQR são representados individualmente no gráfico sendo estes valores caracterizados como*outliers*.

As quantidades q1 - 1.5IQR e q3 + 1.5IQR delimitam, respectivamente, as cercas inferior e superior e constituem limites para além dos quais, como visto, os dados passam a ser considerados *outliers*.

O boxplot permite avaliar a simetria dos dados, sua dispersão e a existência ou não de outliers nos mesmos, sendo especialmente adequado para a comparação de dois ou mais conjuntos de dados correspondentes às categorias de uma variável qualitativa.

Para a construção do boxplot será utilizado um conjunto de dados apresentado na página 233 do livro Análise Financeira de Balanços - Abordagem Básica e Gerencial (6a Edição) do Dante C. Matarazzo, relativo aos índices de liquidez corrente de empresas solventes e insolventes. O conjunto de dados chama-se [ilc.txt](http://marcosfs2006.googlepages.com/ilc.txt).

O boxplot é construído com a função boxplot() do R.

> empresas <- read.table("ilc.txt", header=TRUE)
> str(empresas)
'data.frame':   99 obs. of  2 variables:
 $ ilc     : num  1.86 1.89 1.93 1.94 1.97 2.85 2.02 2.1 3.02 1.43 ...
 $ situacao: Factor w/ 2 levels "insolvente","solvente": 2 2 2 2 2 2 2 2 2 2 ...

Os dados possem duas variáveis (ilc e situacao) e 99 registros ou observações.
A variável ilc registra os índices de liquidez corrente das empresas avaliadas.
Os registros iniciais da base de dados pode ser visualizada da seguinte forma:

> head(empresas)
   ilc situacao
1 1.86 solvente
2 1.89 solvente
3 1.93 solvente
4 1.94 solvente
5 1.97 solvente
6 2.85 solvente

Iremos agora construir um boxplot para comparar a distribuição dos índices de liquidez corrente das empresas solventes e insolventes. O boxplot pode ser feito da seguinte forma:

> boxplot(ilc ~ situacao, data=ilc, main="Solventes vs. Insolventes", xlab="Situação da Empresa", ylab="Índice de Liquidez Corrente")



O gráfico acima apresenta a distribuição da variável ilc segundo as categorias da variável situacao. Observando o gráfico, verifica-se que as empresas classificadas como solventes possem índices de liquidez corrente em geral maiores que os índices das empresas classificadas como insolventes.