

Elaboração de Gráficos *Box-Plot* em Planilhas de Cálculo

Marisa Veiga Capela, **Jorge M. V. Capela**

Depto de Físico-Química, IQ, UNESP

14801-970, Araraquara, SP

E-mail: marisavc@iq.unesp.br, capela@iq.unesp.br

Palavras-chave: *Planilhas eletrônicas, Box-plot, Ensino de Estatística*

Resumo: *Uma das figuras mais utilizadas para descrever graficamente uma variável em estudo é o diagrama de “extremos e quartis” (Box-plot ou Box-and-whisker plot), não contemplado pelo Assistente Gráfico de planilhas de cálculo como o Microsoft Excel. O Box-plot permite avaliar a simetria dos dados, a dispersão e a existência ou não de outliers, sendo especialmente adequado para a comparação de dois ou mais grupos. Neste trabalho, mostra-se como construir um gráfico de Box-plot empregando uma planilha do Microsoft Excel. O objetivo é demonstrar visualmente as características da variável em estudo, de forma simples e com uma fácil interpretação dos conceitos envolvidos. Foi possível concluir que o emprego da planilha eletrônica Box-plot constitui uma ferramenta útil para o ensino de conceitos estatísticos, e ganha ainda mais importância devido à popularidade e acessibilidade do Microsoft Excel, o que não ocorre com a maioria dos softwares estatísticos disponíveis no mercado.*

1 Introdução

É comum observar que estudantes de diversas áreas do conhecimento, em cuja grade curricular a Estatística é disciplina obrigatória, encontram sérias dificuldades com conceitos estatísticos importantes, restringindo sua compreensão à organização de dados numéricos, ou a cálculos de média aritmética, desvio padrão e porcentagens. Essas dificuldades não são observadas somente no ensino, mas também em atividades de profissionais que utilizam os conhecimentos dessa disciplina para analisar e interpretar dados em trabalhos científicos. As planilhas eletrônicas podem ser úteis no propósito de ampliar as aplicações da Estatística, além de facilitar a aquisição do conhecimento estatístico para um grande número de interessados. Planilhas eletrônicas, como o Excel da Microsoft ou o Calc do BOffice são ferramentas populares, de fácil acesso e oferecem uma maior interação do usuário com as funções elementares da Matemática e Estatística do que a maioria dos *softwares estatísticos* em geral [1,4,5].

Neste trabalho mostra-se como construir uma planilha de cálculo para a construção do gráfico de *Box-plot*, não contemplado pelo assistente gráfico do Microsoft Excel. Ainda que a construção de gráficos seja abordada em vários textos de Estatística, não são apresentados procedimentos para a construção de gráficos *Box-plot*.

O *Box-plot*, também conhecido como *gráfico de caixa*, é um gráfico estatístico que possibilita representar a distribuição de um conjunto de dados com base em alguns parâmetros descritivos. Existem algumas variações quanto à quantidade de estatísticas representadas nesse tipo de gráfico, mas de uma forma geral todos incluem a mediana, o 1º e o 3º quartil, os valores mínimos e máximos e eventuais *outliers* e *extremos* [6].

Na construção da planilha são fornecidos detalhes que talvez sejam desnecessários para quem tem familiaridade com a utilização de planilhas eletrônicas. Contudo, a experiência tem mostrado que muitos estudantes, ou até mesmo profissionais, não utilizam as planilhas de cálculo por desconhecerem as potencialidades deste recurso computacional.

2 Metodologia

O *Box-plot* permite avaliar a simetria dos dados, sua dispersão e a existência ou não de *outliers*, sendo especialmente adequado para a comparação de dois ou mais conjunto de dados correspondentes às categorias de uma variável.

A figura 1 ilustra um destes gráficos, onde a linha central da caixa marca a mediana do conjunto de dados e a amplitude interquartílica (AIQ) que é igual ao (3º quartil)-(1º quartil). A parte inferior da caixa é delimitada pelo 1º quartil (quartil inferior) e a parte superior pelo 3º quartil (quartil superior).

As hastes inferiores e superiores se estendem, respectivamente, do quartil inferior até o menor valor não inferior a (1º quartil) – (1.5AIQ) e do quartil superior até o maior valor não superior a (3º quartil) + (1.5AIQ). As quantidades (1º quartil) – (1.5AIQ) e (3º quartil) + (1.5AIQ) delimitam, respectivamente, as cercas inferior e superior e constituem limites para além dos quais os dados passam a ser considerados *outliers*.

A existência de *outliers*, valores extremamente altos ou extremamente baixos, pode indicar tanto dados incorretos como dados válidos que carecem de uma atenção especial. Dependendo do objetivo pode ser que justamente os *outliers* sejam os pontos de interesse da análise.

A Figura 1 também mostra a planilha do Microsoft Excel para a construção do gráfico de *Box-plot* para a comparação de quatro conjuntos de dados correspondente a uma variável. Primeiro calculam-se as estatísticas descritivas de cada grupo nas colunas C, D, E, e F, as quais contêm texto explicativo, fórmulas de cálculo ou valores introduzidos diretamente pelo usuário.

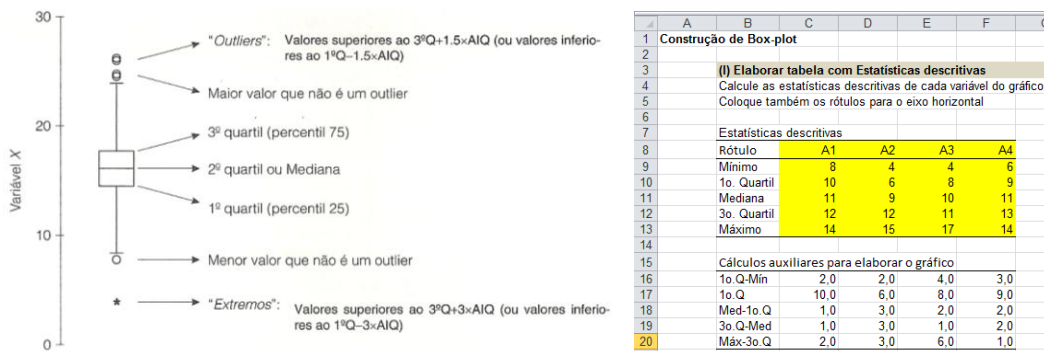


Figura 1: Gráfico ilustrativo de um Box-plot e os dados simulados da variável para o início da construção da planilha no Microsoft Excel.

Na sequência utiliza-se uma tabela com cálculos auxiliares e selecionam-se as células de B17: F19 (Figura 1) para inserir um gráfico de Colunas empilhadas (Figura 2).



Figura 2: Gráfico ilustrativo de como construir um gráfico de colunas empilhadas no Microsoft Excel.

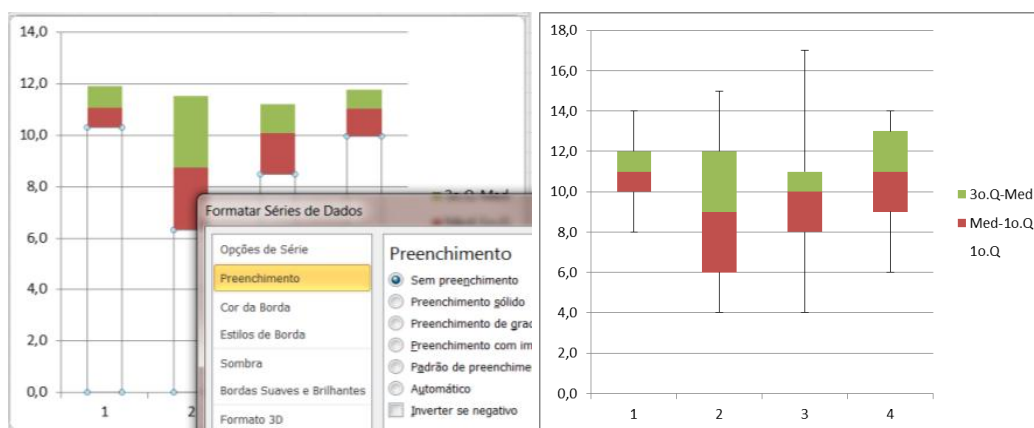


Figura 3: Sequência de procedimentos no Excel para a construção do gráfico Box-Plot .

Em seguida, na figura 3, a coluna 1^o quartil é formatada sem preenchimento e colocamos as barras de erros verticais para menos na coluna 1^o quartil igual a (1^o quartil) – (mínimo), e as barras de erros verticais para mais na coluna (3^o quartil)-(mediana) igual a (máximo) – (3^o quartil).

Finalmente, realiza-se a formatação selecionando dados, rótulos do eixo horizontal e as células de C8: F8.

3 Resultados

A Figura 4 contém o resultado final da planilha de cálculo sugerida neste trabalho para a construção de um gráfico *Box-plot* para a comparação de quatro grupos em relação a uma determinada variável (resposta). O gráfico de *Box-plot* para apenas um grupo não tem muita utilidade, a não ser para a avaliação de *outliers*. Entretanto, ao comparar dois ou mais conjuntos de dados o *Box-plot* atinge sua máxima eficácia.

Observa-se na figura 4 que os dados dos grupos 1 e 4 possuem maior mediana que os dos grupos 2 e 3. Outro ponto importante na comparação visual de um *Box-plot* entre dois ou mais grupos é a medida da variabilidade dos dados. Os grupos 2 e 3 possuem uma maior dispersão dos dados que os grupos 1 e 4. Podemos observar ainda diferenças entre o 3^o e o 1^o quartil, o que também é uma medida da variabilidade dos dados.

Esta planilha vem sendo utilizada pelos autores como uma ferramenta na pesquisa e no ensino de Estatística.

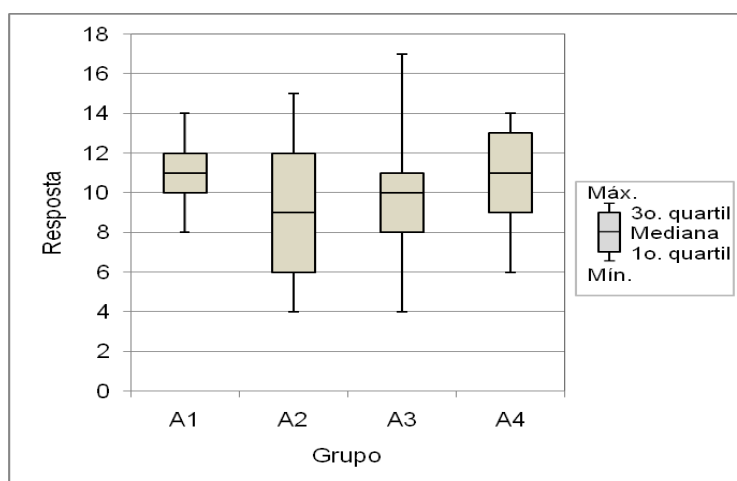


Figura 4: Resultado final, o gráfico de Box-plot para comparação de quatro grupos da variável simulada na Figura 1.

Conclusão

As planilhas eletrônicas constituem uma ferramenta muito útil para o ensino de conceitos estatísticos, pois possibilitam a modificação de alguns aspectos de um determinado raciocínio, com a visualização instantânea dos efeitos destas mudanças. Experiências de ensino de Estatística usando especificamente a planilha *Box-plot* têm sido desenvolvidas com bons resultados no curso de graduação e pós-graduação do Instituto de Química e da Faculdade de Ciências Farmacêuticas de Araraquara.

Conclui-se que a planilha de cálculo proposta possibilita, de maneira fácil e acessível, a construção ou simulação de gráficos *Box-plot*, permitindo a avaliação da simetria dos dados, da dispersão e da existência ou não de *outliers*.

5 Como proceder para adquirir a planilha eletrônica Box-Plot.

Cópias da planilha podem ser solicitadas a um dos autores do trabalho.

AGRADECIMENTOS

FAPESP (Fundação de Amparo à pesquisa do Estado de São Paulo)

PROEX (Pró-Reitoria de Extensão Universitária – UNESP) pelo apoio financeiro.

Referências

- [1] K.N Berk, P. Carey Data analysis with Microsoft Excel. Pacif Grove: Duxbury Press, 2000. 578p.
- [2] G.E.M Box; M. E. Muller, A note on the generation of random normal deviates. *Ann. Math. Statist*, V. 29., p. 610-611, 1958.
- [3] A. Grafen., R. Hails, Modern Statistics for the Life Sciences. New York: Oxford University Press, 2002.
- [4] J. C. Lapponi, Estatística usando o Excel. 4.ed. São Paulo: Editora Lapponi, 2005.
- [5] D. M. Levine, M. L. Berenson, D. Stephan, *Estatística: teoria e aplicações*. Rio de Janeiro: Livros Técnicos e Científicos Editora S.A, 2008.
- [6] J. Maroco, *Análise Estatística com a utilização do SPSS*. 3.ed. Edições Silabo , 2007. 842p.
- [7] M.F.Triola, “Introdução à Estatística”, 10ªed. Livros Técnicos e Científicos Editora, 2008.