

Versatilidades do Box Plot

Autor: Edson R. Montoro

O Box Plot ou Gráfico de Caixa, criado por John Tukey, tem inúmeras aplicações além de ser uma ferramenta muito útil e de fácil construção e interpretação.

Existem muitas variações do Box Plot, por ele ser uma ferramenta simples e muito visual as vezes ocorrem certos exageros em se colocar muitas informações sobre ele; o recomendado é que ele seja o mais simples possível, com as informações necessárias para realmente mostrar somente o que se deseja; sem poluição visual, pois senão ele perde toda a sua força.

Neste artigo mostraremos os tipos mais básicos de Box Plot e alguns exemplos de aplicação.

1. Box Plot Simples

Basicamente, o Box Plot (Figura 1) mostra a distribuição dos resultados experimentais e é composto dos seguintes valores: “Menor Valor”, “1º Quartil”, “2º Quartil” (ou Mediana), “3º Quartil” e “Maior Valor”.

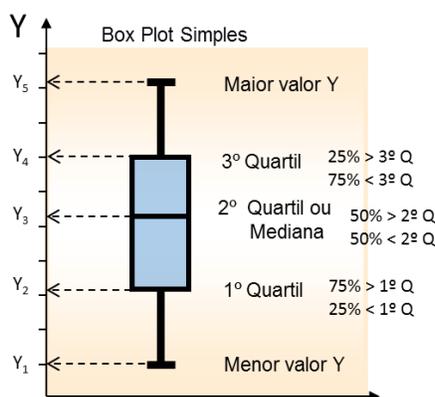


Figura 1 – Interpretação do Box Plot simples.

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro
edsonmontoro@ermontoro.com
 +55 (35) 99161-8141

Flavio A. Montoro
flaviomontoro@ermontoro.com
 +55 (16) 98102-6651

www.ermontoro.com

Estas estatísticas são facilmente calculadas pelas Fórmulas 1, 2 e 3, que produzem a posição da respectiva estatística, que tem o resultado experimental correspondente.

$$\text{Posição do 1}^\circ \text{ quartil} = \frac{1}{4}(n + 1) \quad (1)$$

$$\text{Posição do 2}^\circ \text{ quartil} = \frac{2}{4}(n + 1) \quad (2)$$

$$\text{Posição do 3}^\circ \text{ quartil} = \frac{3}{4}(n + 1) \quad (3)$$

Normalmente quando temos poucas medidas de uma variável aleatória não conseguimos construir um Histograma (necessita-se de pelo menos 50 dados para um bom Histograma); para visualizar a distribuição desses resultados, utilizamos o Box Plot.

Uma comparação entre o Box Plot e o Histograma pode ser visto na Figura 2.

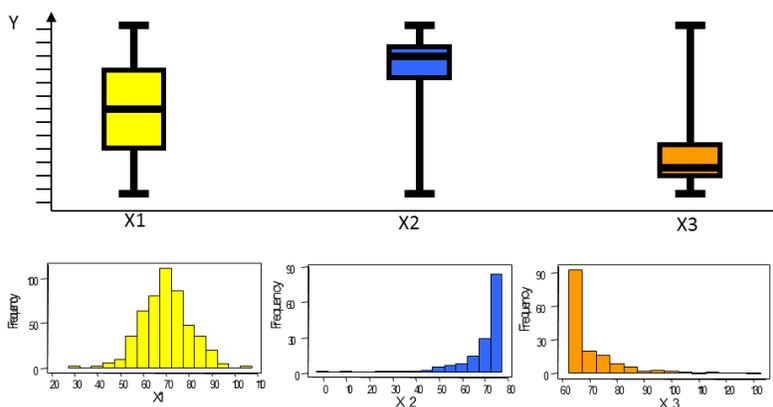


Figura 2 – Comparação do Box Plot com Histograma.

Existem softwares que já constroem o Box Plot automaticamente, como o *Action*, *Minitab*, *Statgraphics*, o *JMP* entre outros; mas pode-se contruir uma planilha em Excel para os cálculos necessários.

Para calcular o primeiro, o segundo e o terceiro quartil (Q1, Q2 ou mediana, Q3); deve-se primeiro ordenar os dados em ordem crescente e depois aplicar as fórmulas (1), (2) e (3) vistas anteriormente.

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro
edsonmontoro@ermontoro.com
 +55 (35) 99161-8141

Flavio A. Montoro
flaviomontoro@ermontoro.com
 +55 (16) 98102-6651

www.ermontoro.com

Podemos ver um exemplo dos cálculos usando os dados de duas variáveis aleatórias com 10 valores cada ($n_1 = n_2 = 10$) apresentados na Tabela 1.

Tabela 1 – Variável Resposta: tempo de processamento (min).

	1	2	3	4	5	6	7	8	9	10
X1	45	183	207	226	243	251	268	272	290	345
X2	102	125	145	156	188	192	215	228	251	279

Para calcular a mediana (segundo quartil), tanto de X1 quanto de X2, é utilizada a Fórmula 2:

$$\text{Mediana (segundo quartil)} = M = \frac{n + 1}{2} = \frac{10 + 1}{2} = 5,5$$

Como a posição é a 5,5ª, o valor da Mediana deve estar entre o 5º e o 6º valor, que conforme o exemplo: para X1 são respectivamente, 243 e 251; então a Mediana será a média entre estes dois valores, 247. Já para X2, será a média entre 188 e 192, que é 190.

O 1º Quartil calculado pela Fórmula 1:

$$Q1 \text{ (primeiro quartil)} = \frac{(1)n + 2}{4} = \frac{10 + 2}{4} = 3$$

é o valor que ocupa a 3ª posição, que no exemplo apresentado, para X1 é 207 e para X2, 145. Já para o 3º Quartil, calculado pela Fórmula 3:

$$Q3 \text{ (terceiro quartil)} = \frac{(3)n + 2}{4} = \frac{(3)10 + 2}{4} = 8$$

é o valor da 8ª posição, que para X1 é 272 e para X2, 228. Esses resultados podem ser vistos na Figura 3.

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro

edsonmontoro@ermontoro.com

+55 (35) 99161-8141

Flavio A. Montoro

flaviomontoro@ermontoro.com

+55 (16) 98102-6651

www.ermontoro.com

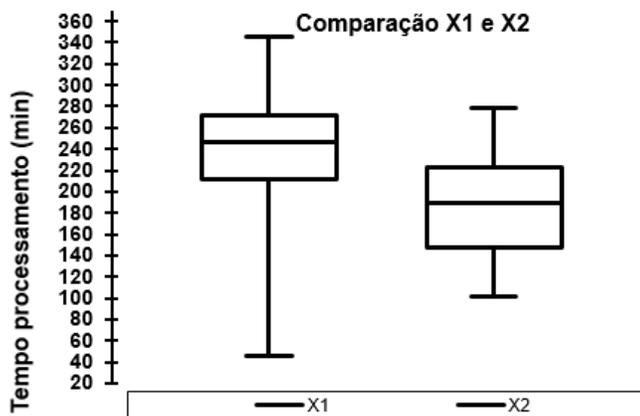


Figura 3 - Comparação de variáveis com Box Plot.

2. Box Plot chanfrado

O Box Plot chanfrado (Figura 4) inclui a informação do Intervalo de Confiança de 95% para a Mediana; o que quer dizer, que é uma estimativa da mediana por intervalo, isto é, o valor real da mediana, com 95% de certeza, deve estar dentro deste intervalo.

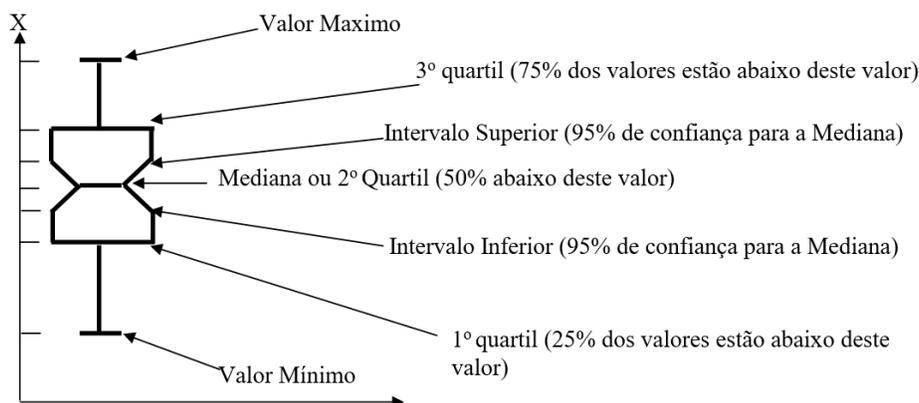


Figura 4 – Interpretação do Box Plot chanfrado.

A utilização deste tipo de Box Plot é em comparações estatísticas como se fosse um Teste de Hipótese “visual”. Se os chanfros de dois ou mais Box Plots coincidirem, podemos dizer que não existe diferença significativa entre as medianas, a um nível de significância de 5%.

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro
edsonmontoro@ermontoro.com
 +55 (35) 99161-8141

Flavio A. Montoro
flaviomontoro@ermontoro.com
 +55 (16) 98102-6651

www.ermontoro.com

Se as variáveis podem ser consideradas como uma boa aproximação para o modelo de distribuição Normal, podemos aproximar essa conclusão também para as médias.

A seguir, veremos três exemplos de aplicação do Box Plot Chanfrado.

2.1. Exemplo 1: comparação da variabilidade e da tendência central de vários equipamentos.

Como se pode observar na Figura 5, não existe diferença significativa entre os equipamentos B e C, pois há uma coincidência entre os respectivos chanfros do intervalo de confiança, já o A é diferente destes dois com relação à mediana pois o chanfro do intervalo de confiança não coincide.

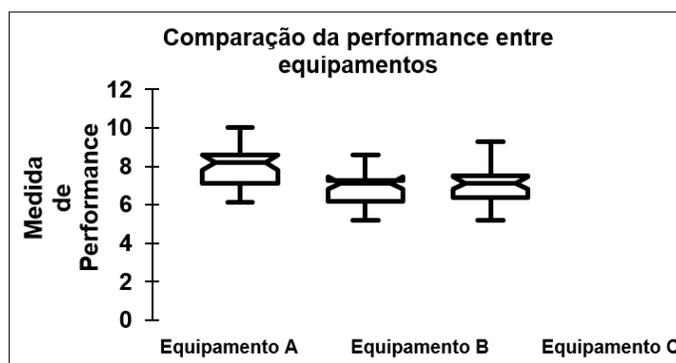


Figura 5 – Comparação usando Box Plot.

Quanto à variabilidade, não dá para afirmar que exista diferença significativa entre os três equipamentos, pois as alturas dos Box Plots são muito parecidas entre si.

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro
edsonmontoro@ermontoro.com
 +55 (35) 99161-8141

Flavio A. Montoro
flaviomontoro@ermontoro.com
 +55 (16) 98102-6651

www.ermontoro.com

2.2. Exemplo 2: comparação da performance de uma variável aleatória ao longo do tempo.

Como se verifica na Figura 6, a variabilidade diminuiu significativamente, é visível que mês após mês a altura do Box Plot, apresentando os valores experimentais diminuíram. Em janeiro, o range de variação é de aproximadamente de 1 a 15, enquanto que em abril é de 7 a 9.

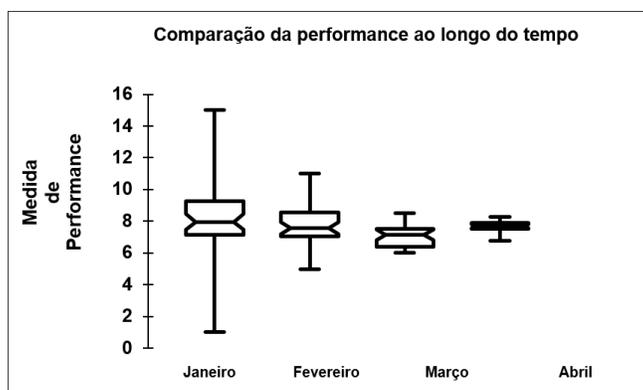


Figura 6 – Comparação de performance usando Box Plot.

2.3. Exemplo 3: monitoramento de processo.

Este exemplo é muito interessante. Numa planta química, os operadores executavam várias medições de nível e toda vez que este era maior que um limite pré definido (no caso 17 cm), eles tinham que executar uma tarefa manual um pouco trabalhosa.

Por falta de orientação sobre a importância do controle dessa variável de processo, toda vez que o nível era maior que 17 cm, alguns operadores anotavam o valor da medição como 17 cm, e deixavam a tarefa para o próximo turno. Isso acarretava um descontrole no processo, gerando desperdícios e causando impacto no controle de outras variáveis do processo.

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro
edsonmontoro@ermontoro.com
 +55 (35) 99161-8141

Flavio A. Montoro
flaviomontoro@ermontoro.com
 +55 (16) 98102-6651

www.ermontoro.com

Os dados de um período foram plotados usando Box Plot, e se percebeu que a grande maioria das medições eram 17 cm ou menor, com poucos valores maiores que 17 cm; o que pode ser facilmente observado nos Box Plots da Figura 7.

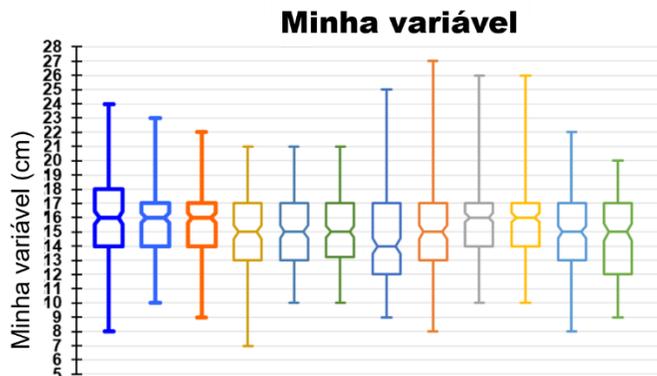


Figura 7 – Identificação de problemas usando Box Plot.

Após essa análise, foi feito um Kaizen envolvendo alguns operadores para melhorar a tarefa deixando-a mais simples. Após a mudança todos os operadores passaram por um treinamento para ficar bem claro a importância do controle dessa variável de processo. Obviamente os desperdícios foram eliminados, e os ganhos computados.

3. Detecção de Outliers

Uma outra importante aplicação do Box Plot é a detecção de *outliers*, isto é, um valor estranho, que provavelmente não pertença à população.

A distância entre o primeiro e o terceiro quartil é chamada de Range Interquartílico (RIQ) e contém praticamente 50% dos dados observados. Se um valor exceder 1,5 vezes este valor (RIQ), para cima ou para baixo, pode ser considerado como um *outlier*; veja o exemplo na Figura 8.

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro

edsonmontoro@ermontoro.com

+55 (35) 99161-8141

Flavio A. Montoro

flaviomontoro@ermontoro.com

+55 (16) 98102-6651

www.ermontoro.com

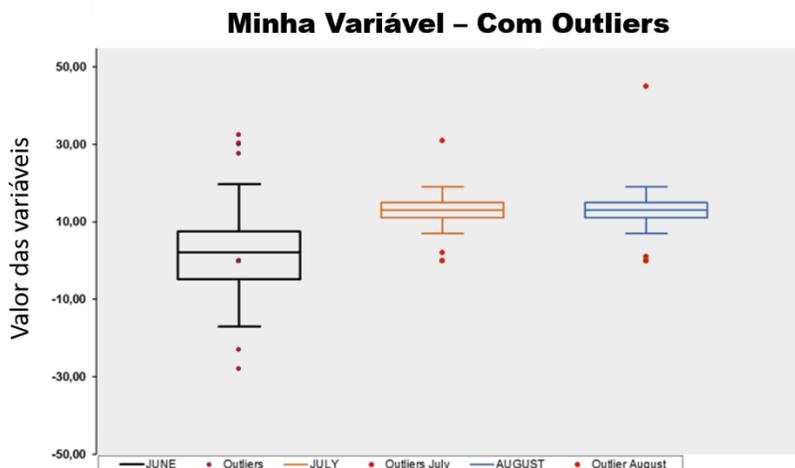


Figura 8 – Identificação de Outliers usando Box Plot.

O racional dessa técnica é que considerando esse intervalo de $\pm 1,5 \times \text{RIQ}$, isto é ($Q1 - 1,5 \times \text{RIQ}$) até ($Q3 + 1,5 \times \text{RIQ}$) praticamente corresponde aos Limites de Controle de uma carta de controle do CEP (Controle Estatístico de Processo), como podemos ver na Figura 9.

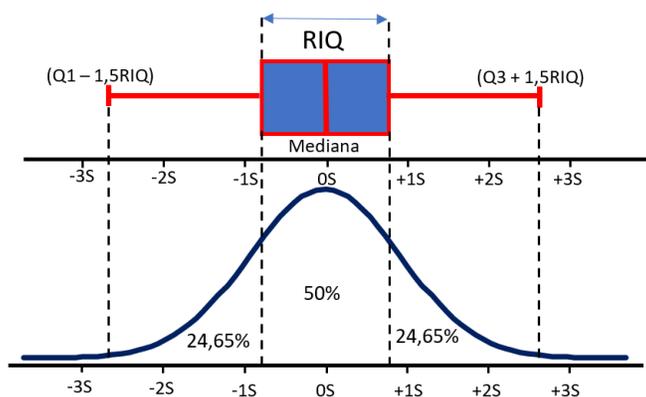


Figura 9 – Racional da identificação de Outliers usando Box Plot.

Vale a pena lembrar que não se pode simplesmente eliminar um *outlier*, temos que descobrir a sua causa antes de qualquer decisão. Pode-se aprender muito a respeito do seu processo com a identificação e análise dos *outliers*.

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro
edsonmontoro@ermontoro.com
 +55 (35) 99161-8141

Flavio A. Montoro
flaviomontoro@ermontoro.com
 +55 (16) 98102-6651

www.ermontoro.com

Referências

- (1) McGill, Robert; Tukey, John W.; Larsen, Wayne A.; Variations of Box Plots. **The American Statistician**, vol. 32 (1): pp. 12–16, February, 1978.

Sobre o Autor:

Edson R. Montoro é Diretor Técnico da ERMontoro Consultoria e Treinamento Ltda, empresa focada no desenvolvimento de pessoas e consultoria nas áreas de melhoria de processo usando Estatística Aplicada e *Lean Manufacturing*.

O autor é Químico pela UNESP (Universidade Estadual Paulista “Júlio de Mesquita Filho”) – Araraquara, MBA em Gestão Empresarial pela FGV (Fundação Getúlio Vargas), *Master Black Belt* pela Air Academy Associates, Engenheiro de qualidade pela ASQ (*America Society for Quality*) e Pós-graduação em Gerência de Produção pela UFSC (Universidade Federal de Santa Catarina).

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro

edsonmontoro@ermontoro.com

+55 (35) 99161-8141

Flavio A. Montoro

flaviomontoro@ermontoro.com

+55 (16) 98102-6651

www.ermontoro.com
