Departamento de Engenharia de produção
EESC Escola de Engenharia de São Carlos
USP
Grupo de Pesquisa em Gestão da Qualidade e Mudança
Research Group on Quality and Change Management

Tópico 02 – Métodos Estatísticos para a Melhoria da Qualidade

Disciplina: SEP-280
Controle da Qualidade de Processos de Fabricação

Research Group Leaders:
Luiz C. R. Carpinetti, Associate Professor
Mateus C. Gerolamo, Assistant Professor

Document version:
August, 2015

- Estatística é um conjunto de técnicas úteis para a tomada de decisão sobre um processo ou população.

- Estatística é a língua na qual engenheiros, operários, compradores, administradores e outros integrantes da empresa se comunicam sobre qualidade.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 25

Parte I: Modelando a Qualidade do Processo

Modelando a Qualidade do Processo

- Esta parte do curso aborda o uso da metodologia estatística no controle e aprimoramento da qualidade e possui dois objetivos:

  - Primeiro, mostrar como ferramentas simples de estatística descritiva podem ser usadas para expressar quantitativamente a variação em uma característica da qualidade quando uma amostra dos dados sobre esta característica está disponível (em geral, uma amostra é um subconjunto de dados selecionados de um processo ou população maior);

  - Segundo, introduzir as **distribuições de probabilidade** e mostrar como elas fornecem meios para modelar ou descrever características da qualidade de um processo.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Parte I: Modelando a Qualidade do Processo

DESCREVENDO A VARIAÇÃO

Descrevendo a Variação

- Duas unidades produzidas por um processo de fabricação nunca são idênticas...

→ alguma variação é inevitável!

- Estatística é a ciência de analisar dados e tirar conclusões, levando em conta a variação nos dados.

- Há vários métodos gráficos que são úteis para resumir e apresentar dados. Um dos gráficos mais úteis é o **gráfico ramo-e-folhas**.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

O Gráfico Ramo-e-Folhas

- Suponha que os dados sejam representados por $x_1, x_2, ..., x_n$ e que cada número x_i consista de pelo menos dois dígitos.

- Para construir um ramo-e-folhas, dividimos cada número x_i em duas partes:
  - um ramo, formado por um ou mais dígitos iniciais; e
  - uma folha, formada pelos dígitos restantes.

Descrevendo a Variação

O Gráfico Ramo-e-Folhas (continuação)

- Por exemplo, se os dados representarem percentuais, variando de 0 a 100, de peças defeituosas em lotes de semicondutores, então podemos dividir o valor 76 no ramo 7 e na folha 6.

- Em geral, devemos escolher relativamente poucos ramos em comparação ao número de observações. Normalmente é melhor escolher entre 5 e 20 ramos.

- Escolhido o número de ramos, eles são listados à esquerda de uma margem no desenho e ao lado de cada ramo, todas as folhas correspondentes aos valores dos dados observados são listadas na ordem de ocorrência no conjunto de dados.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Making a stem and leaf plot

Data

23	71
58	71
62	72
62	80
63	82
65	82
67	82

stem	
2	3
3	
4	
5	8
6	2 2 3 5 7
7	1
8	

Descrevendo a Variação

- O Diagrama de Ramos e Folhas
- *Stem-and-leaf display*

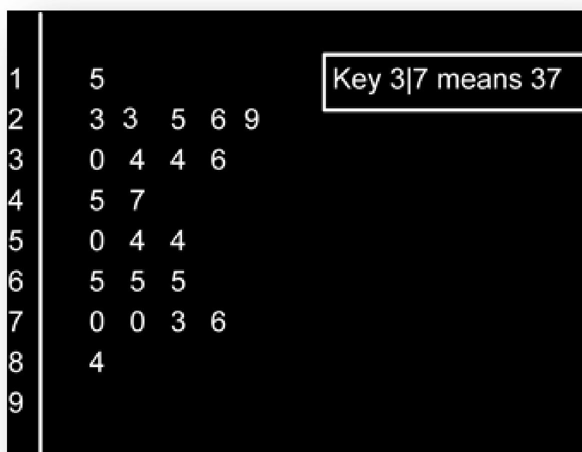
https://www.khanacademy.org/math/pre-algebra/applying-math-reasoning-topic/reading_data/v/u08-l1-t2-we3-stem-and-leaf-plots

https://en.wikipedia.org/wiki/Stem-and-leaf_display


みなとみらい線標準時刻表
Train Departures of Minatomirai Line
みなとみらい 元町・中華街 方面
for Minatomirai, Motomachi·Chūkagai
平日 Weekdays
土曜日・休日 Saturdays, Sundays and Holidays
時刻表の見方

Variações do Gráfico Ramo-e-Folhas

- Um **gráfico ramo-e-folhas ordenado** tem suas folhas organizadas por ordem de grandeza, como exibido na figura a seguir.



Descrevendo a Variação

Variações do Gráfico Ramo-e-Folhas **(continuação)**

- Esta versão do gráfico facilita a obtenção dos **percentis** dos dados.
- Em geral, o 100 *k*º percentil é um valor tal que pelo menos 100 *k*% dos dados são iguais ou inferiores a este valor e pelo menos 100 (1-*k*)% são iguais ou superiores a este valor.
- O quinquagésimo percentil da distribuição dos dados é chamado de **mediana amostral $\tilde{x}$**.
- A mediana pode ser pensada como o valor que divide a amostra ao meio, com metade das observações sendo menor que a mediana e a outra metade, maior.

Descrevendo a Variação

Variações do Gráfico Ramo-e-Folhas (continuação)

- Se *n*, o número de observações, é ímpar, é fácil achar a mediana.
  - Primeiro, ordene as observações em ordem crescente;
  - Então, a mediana será a observação cujo posto é:

$$\text{Posição (posto) da Mediana} = \frac{(n-1)}{2} + 1$$

Posição → .... *1* ....*2* ... .*3* ... .*4* ... .*5* ... .*6* ... .*n*

Pos. 1	**Pos. 2**	**Pos. 3**	**Pos. 4**	**Pos. 5**	**Pos. 6**	**Pos. 7**
45	45	49	53	58	58	59

- Exemplo:

$$\text{Posição (posto) da Mediana} = \frac{(7-1)}{2} + 1 = \text{posição 4,}$$ logo mediana = 53

Variações do Gráfico Ramo-e-Folhas (continuação)

- Se *n*, o número de observações, é par, a mediana é a média $(n/2)^{a}$ e da $(n/2 + 1)^{a}$ observações na lista ordenada.

- Exemplo:

Posição → .... 1 ....2 ....3 ....4 ....5 ... .6 ... .7.................n

Pos. 1	Pos. 2	Pos. 3	Pos. 4	Pos. 5	Pos. 6	Pos. 7	Pos. 8
45	45	49	53	58	58	59	59

$$\text{Mediana} = \frac{\text{Valor Posto } n/2 + \text{valor posto } n/2 + 1}{2} = \frac{53 + 58}{2} = 111/2, \text{ logo mediana} = 55{,}5$$

Variações do Gráfico Ramo-e-Folhas (continuação)

- O primeiro e o terceiro quartis são ocasionalmente representados pelos símbolos Q1 e Q3, respectivamente, e o **intervalo interquartil** IIQ = Q3 – Q1 é, às vezes, usado como medida de variabilidade.

Descrevendo a Variação

Variações do Gráfico Ramo-e-Folhas (continuação)

- Embora o ramo-e-folhas seja uma excelente forma de visualizar a variabilidade nos dados, ele não leva em conta a **ordem** temporal das observações.

- O tempo é, muitas vezes, um importante fator que contribui para a variabilidade nos problemas de melhoria da qualidade.

- Poderíamos, simplesmente, plotar os valores dos dados *versus* o tempo; tal gráfico é chamado **gráfico de séries temporais** ou **gráfico de linha ou sequencial**.

- No entanto, uma abordagem bastante útil seria combinar o gráfico de linha com ramo-e-folhas para produzir o gráfico **digidot**.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Variações do Gráfico Ramo-e-Folhas (continuação)

- A figura mostra o gráfico digidot para os dados da produção de semicondutores (do nosso exemplo).

Leaf	Stem
8	9e
6	9s
45	9f
2333	9t
0010000	9z
99998	8e
66677	8s
45	8f
23	8t
1	8z

Time series plot

Variações do Gráfico Ramo-e-Folhas (continuação)

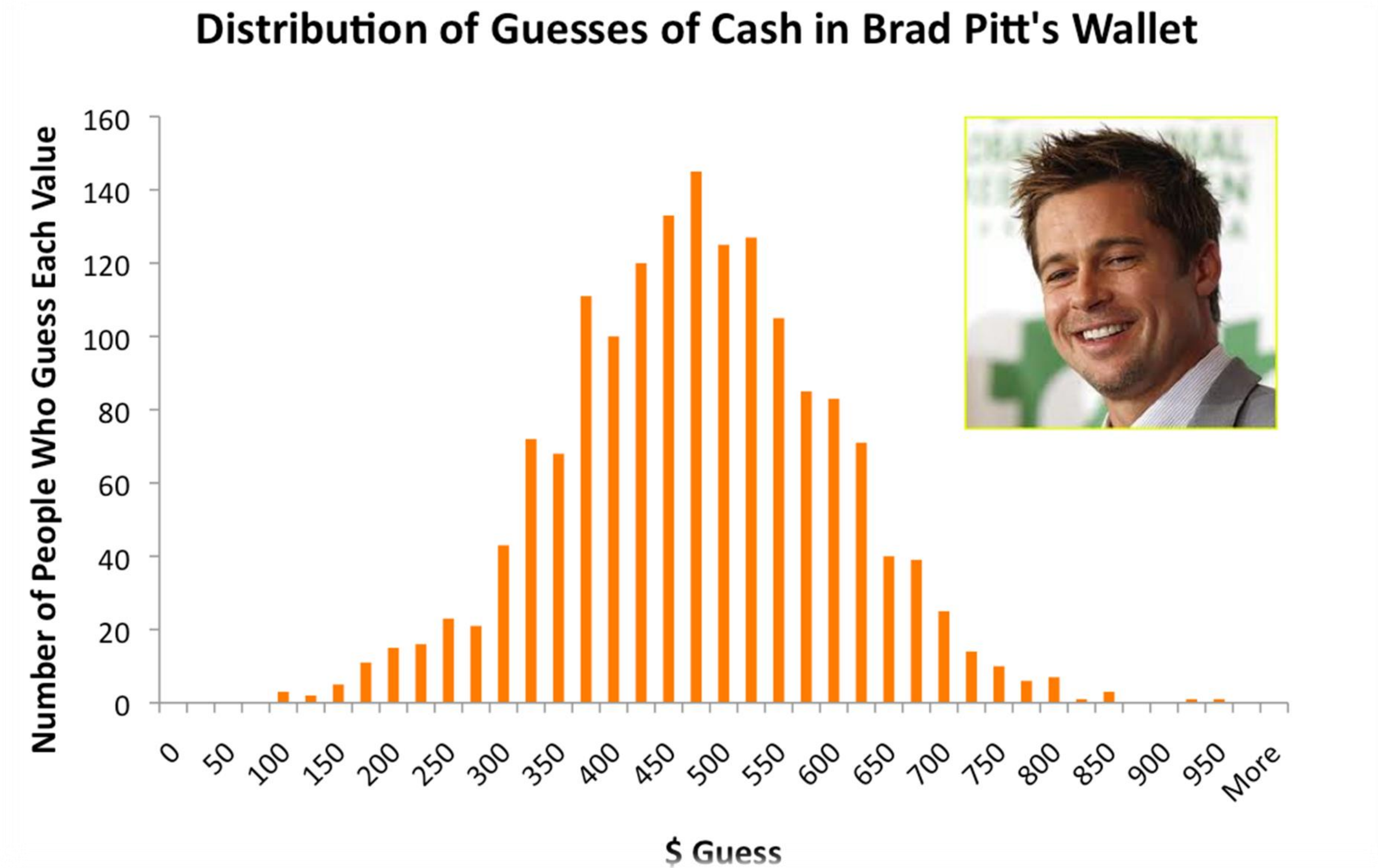
- O gráfico anterior simplesmente indica que o tempo é uma importante fonte de variabilidade no processo de produção.

A Distribuição de Frequências e o Histograma

- Uma **distribuição de frequências** consiste em uma reorganização dos dados por grandeza.
- É um sumário dos dados ainda mais compacto do que o gráfico ramo-e-folhas.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Descrevendo a Variação



A Distribuição de Frequências e o Histograma

- O gráfico das frequências observadas *versus* a variável de interesse é chamado **histograma**!
- A altura de cada barra é igual à frequência de ocorrência da variável na faixa (intervalo) especificada.
- O **histograma** permite enxergarmos alguns aspectos no processo que a inspeção dos dados brutos da tabela original não mostra.

A Distribuição de Frequências e o Histograma

- O **histograma** é uma representação visual dos dados na qual podemos ver mais facilmente três propriedades:
  - Forma;
  - Posição ou tendência central; e
  - Espalhamento ou dispersão.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Descrevendo a Variação

A Distribuição de Frequências e o Histograma

- Algumas diretrizes são úteis na construção de **histogramas**. Quando os dados são numerosos, agrupá-los em classes ou celas, como no exemplo dos anéis de piston, é bastante útil. Em geral:
  - Use entre 4 e 20 classes – escolher o número de classes aproximadamente igual à raiz quadrada do número de observações costuma ser uma boa opção;
  - Use classes de mesmo comprimento;
  - Inicie o limite inferior da primeira classe ligeiramente abaixo do menor valor dos dados.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Descrevendo a Variação

A Distribuição de Frequências e o Histograma

- Agrupar os dados em classes condensa os dados originais e, como resultado, algum detalhe é perdido.

- Assim, quando o número de observações é relativamente pequeno ou quando as observações assumem poucos valores, o histograma pode ser construído a partir da distribuição de frequência dos dados não-agrupados.

- Outra alternativa é usar o gráfico ramo-e-folhas; uma vantagem do ramo-e-folhas é que as observações individuais são preservadas, enquanto que nos histograma elas são perdidas.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Resumo Numérico dos Dados

- Além das ferramentas gráficas, é também importante usar medidas numéricas de tendência central e dispersão.

- Suponha que $x_1, x_2, \ldots, x_n$ sejam as observações em uma amostra. A medida de tendência central mais importante é a média amostral:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Descrevendo a Variação

Resumo Numérico dos Dados

- Note que a média amostral $\bar{x}$ é simplesmente a média aritmética das *n* observações.
- Em comparação ao histograma, note que a média amostral é o ponto no qual o histograma se "equilibra". Então, e média amostral corresponde ao centro de massa dos dados da amostra.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Descrevendo a Variação

Resumo Numérico dos Dados

- A variabilidade nos dados amostrais é medida pela **variância amostral**:

$$S^2 = \frac{\sum_{i=1}^{n} (Xi - \overline{X})^2}{(n-1)}$$

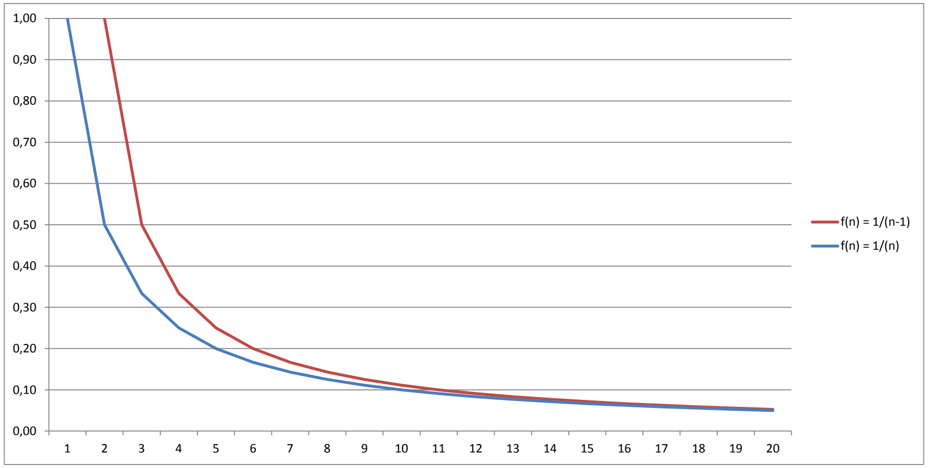
- Note que a variância amostral é simplesmente a soma dos quadrados dos desvios de cada observação em relação à média amostral $\bar{x}$, dividida pelo tamanho da amostra menos um.

Descrevendo a Variação

Por que dividir por (n-1)?

- Dividimos o desvio padrão amostral por n – 1, porque há apenas n – 1 valores independentes.
- Ou seja, dada uma média, apenas n – 1 valores podem ser associados a qualquer número, antes que o último valor seja determinado.
- Além disso, se s^2 fosse definido como a divisão por n, ele sistematicamente subestimaria o valor de σ^2, o que é compensado pela diminuição do denominador. Note que quanto menor o tamanho da amostra, maior o fator de compensação.

$$s^2 = \frac{\sum_{i=1}^{n}(Xi - \overline{X})^2}{(n-1)}$$



Descrevendo a Variação

Resumo Numérico dos Dados

- A unidade da variância amostral é o quadrado da unidade original dos dados.
- Muitas vezes isso é conveniente e difícil de interpretar e, assim, nós usualmente preferimos usar a raiz quadrada de S^2, chamada **desvio-padrão amostral S**, como uma medida de variabilidade:

$$S = \sqrt{\frac{\sum_{i=1}^{n} (Xi - \overline{X})^2}{(n-1)}}$$

Resumo Numérico dos Dados

- A principal vantagem do desvio-padrão amostral é que ele é expresso na unidade de medida original.

- A maior variabilidade de uma amostra de dados é refletida em um maior desvio-padrão amostral.

Descrevendo a Variação

Resumo Numérico dos Dados

- No entanto, imagine duas amostras:
  - A1 = (x_1 = 1; x_2 = 3; x_3 = 5), resultando $\overline{x}$ = 3; e
  - A2 = (x_1 = 101; x_2 = 103; x_3 = 105), resultando $\overline{x}$ = 103.

- Se calcularmos o desvio padrão-amostral para as duas amostras, teremos $S_{A1} = S_{A2} = 2$.
- Ou seja, comparando as duas amostras, vê-se que ambas têm a mesma variabilidade ou dispersão em torno da média, e é por isso que elas têm o mesmo desvio padrão.

- Isso nos leva a um ponto importante: **o desvio-padrão não reflete a magnitude dos dados amostrais, reflete apenas a dispersão em torno da média**.

Parte I: Modelando a Qualidade do Processo

O DIAGRAMA DE CAIXA (*BOX PLOT*)

O Diagrama de Caixa *(Box Plot)*

- O **diagrama de caixa (**em inglês, ***box plot*****)** é um gráfico que exibe simultaneamente vários aspectos importantes dos dados, tais como:
  - tendência central ou posição;
  - dispersão ou variabilidade;
  - afastamento da simetria e identificação de observações muito afastadas da maior parte dos dados *(outliers).*

O Diagrama de Caixa *(Box Plot)*

Na estatística descritiva, um **quartil** é qualquer um dos três valores que divide o conjunto ordenado de dados em quatro partes iguais, e assim cada parte representa 1/4 da amostra ou população.
(ver: http://pt.wikipedia.org/wiki/Quartil)

- O diagrama de caixa exibe:
  - os três quartis,
  - e o mínimo e o máximo dos dados em uma caixa retangular, alinhada vertical ou horizontalmente

- A caixa cobre o intervalo interquartil com:
  - a linha esquerda (ou inferior) posicionada no primeiro quartil Q1, e
  - a linha direita (ou superior) posicionada no terceiro quartil Q3.

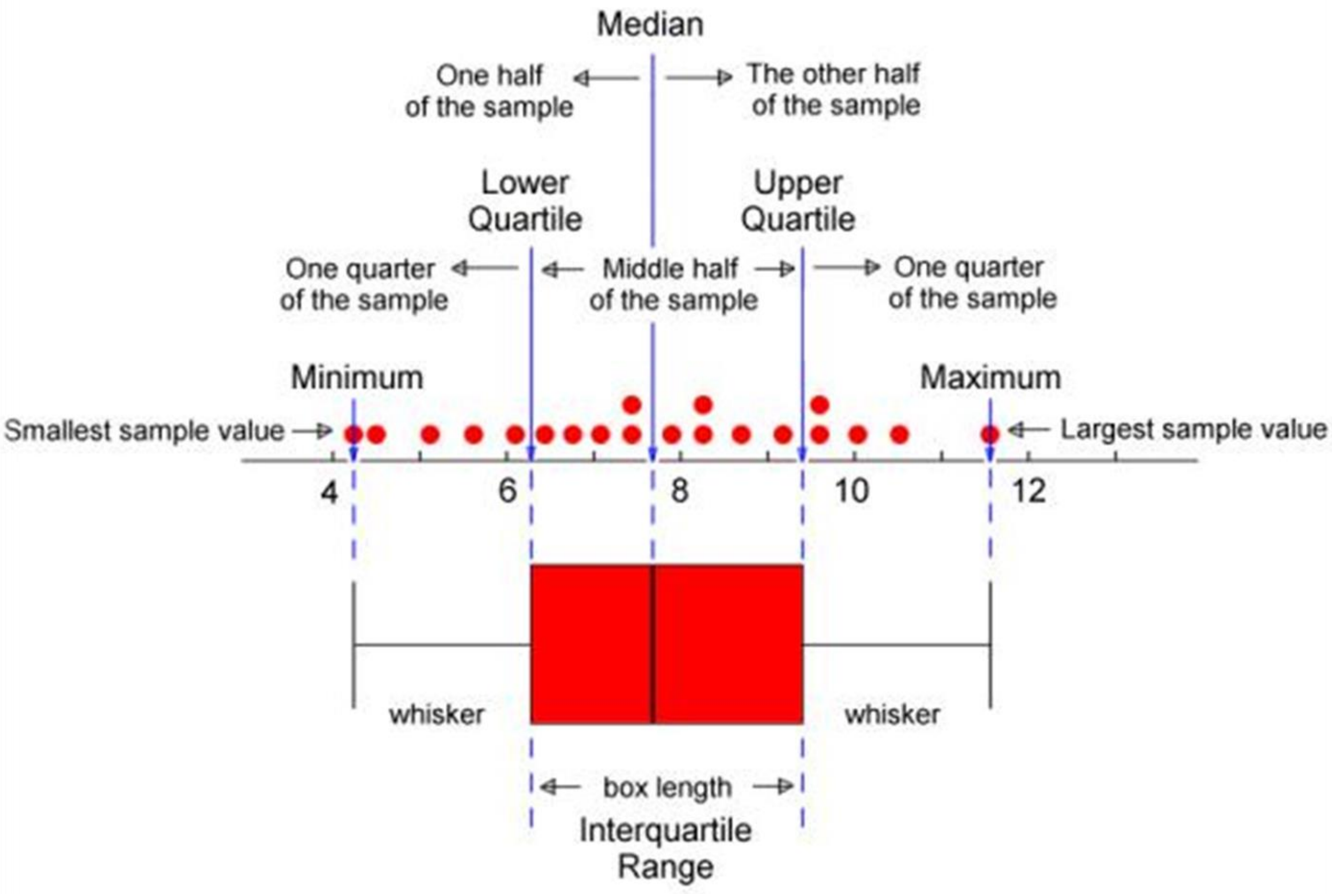
Descrevendo a Variação

O Diagrama de Caixa *(Box Plot)*

- Uma linha é traçada ao longo da caixa na posição do segundo quartil (que representa o quinquagésimo percentil ou mediana) $Q2 = \tilde{x}$.

- Em ambos os extremos da caixa uma linha se estende até os valores extremos.

O Diagrama de Caixa *(Box Plot)*



Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Descrevendo a Variação

O Diagrama de Caixa *(Box Plot)*

- Este gráfico mostra que a distribuição dos diâmetros dos orifícios não é exatamente simétrica em torno de um valor central porque as linhas esquerda e direita e as partes da caixa à direita e à esquerda não têm o mesmo comprimento.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Descrevendo a Variação

O Diagrama de Caixa *(Box Plot)*

- Os diagramas de caixa são muito úteis para a comparação gráfica de conjunto de dados, uma vez que eles têm um forte impacto visual e são fáceis de compreender.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Descrevendo a Variação

O Diagrama de Caixa *(Box Plot)*

<u>Exemplo</u>:

- A figura a seguir exibe diagramas de caixa comparativos para um índice de qualidade para produtos de três fábricas.
- A inspeção deste gráfico mostra que há muita variabilidade na planta 2 e que as plantas 2 e 3 precisam melhorar seu desempenho.

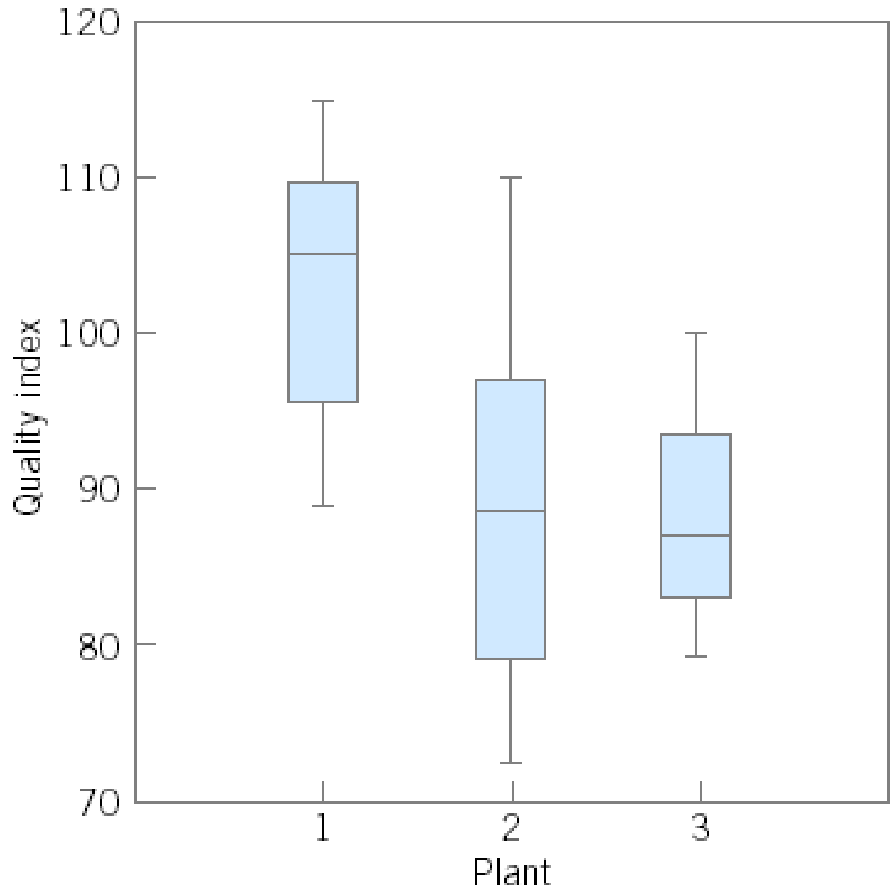


Figure 2-8 Comparative box plots of a quality index for products produced at three plants.

Parte I: Modelando a Qualidade do Processo

DISTRIBUIÇÕES DE PROBABILIDADE INTRODUÇÃO

Descrevendo a Variação

Distribuições de Probabilidade

- O histograma (ou ramo-e-folhas ou diagrama de caixa) é usado para descrever os dados de uma **amostra**.
- Uma **amostra** é um conjunto de medidas selecionado de uma **população** maior.
- Com o uso de *métodos estatísticos*, podemos analisar **amostras** e tirar conclusões sobre a **população**.

Descrevendo a Variação

Distribuições de Probabilidade (continuação)

- Uma **distribuição de probabilidade** é um modelo matemático que relaciona o valor da variável com a probabilidade de ocorrência daquele valor na população.

- Em outras palavras, podemos visualizar os resultados do parâmetro de um processo como uma **variável aleatória**, porque ele assume diferentes valores na população de acordo com algum mecanismo aleatório e, assim, a distribuição de probabilidade desse parâmetro do processo descreve a probabilidade de ocorrência de qualquer valor de tal parâmetro na população.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Distribuições de Probabilidade (continuação)

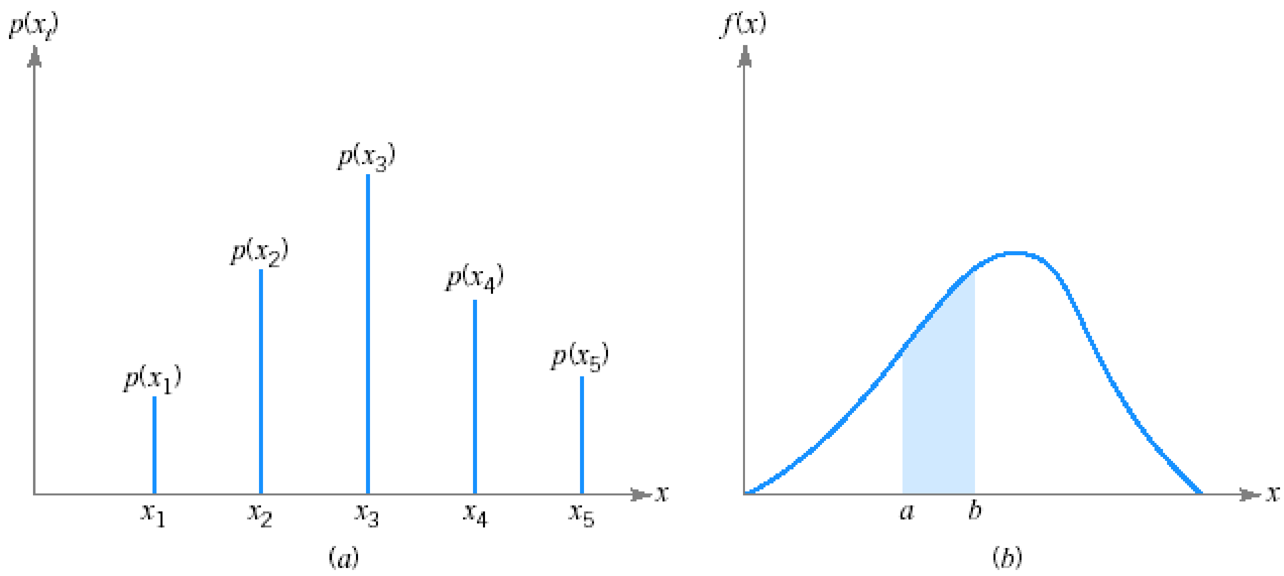
- Há dois tipos de distribuição de probabilidade:
  - **Distribuições contínuas** – quando a variável sendo medida é expressa em uma escala contínua, sua distribuição de probabilidade é chamada *distribuição contínua*. Ex. medida do diâmetro de uma peça.
  - **Distribuições discretas** – quando o parâmetro sendo medido só pode assumir certos valores, tais como os inteiros 0, 1, 2, ..., a distribuição de probabilidade é chamada *distribuição discreta*. Ex. número de defeitos em um produto.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Descrevendo a Variação

Distribuições de Probabilidade (continuação)

- Exemplos de distribuições discreta e contínua são apresentadas nos gráficos a seguir, respectivamente.



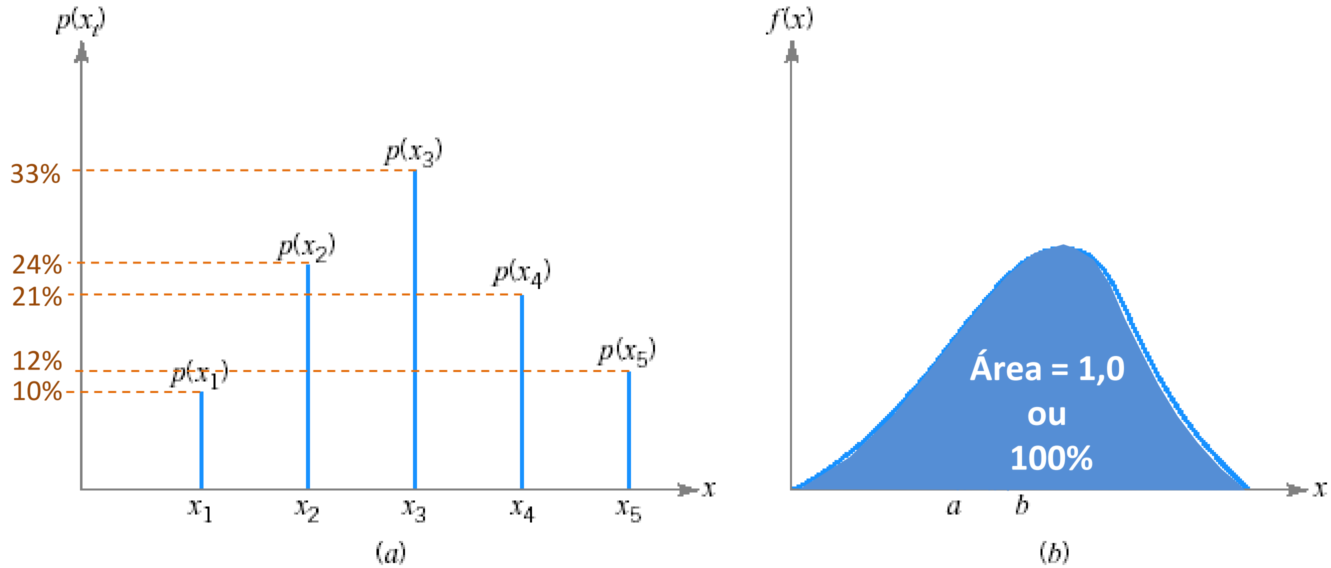
Gráficos: Distribuições de Probabilidade: (a) Caso Discreto e (b) Caso Contínuo

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, Cap. 02, p. 34

Descrevendo a Variação

Distribuições de Probabilidade (continuação)

- Exemplos de distribuições discreta e contínua são apresentadas nos gráficos a seguir, respectivamente.



Gráficos: Distribuições de Probabilidade: (a) Caso Discreto e (b) Caso Contínuo

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, Cap. 02, p. 34

Descrevendo a Variação

Distribuições de Probabilidade (continuação)

- A **média μ** de uma distribuição de probabilidade é uma medida da **tendência central** da distribuição, ou da sua **posição**. A média é o ponto no qual a distribuição se "equilibra" perfeitamente. Então, a média é simplesmente o centro de massa da distribuição de probabilidade.

- A média não é necessariamente o quinquagésimo percentil da distribuição (a **mediana**) e também não é necessariamente o valor mais provável da variável (o qual é chamado de **moda**). A média simplesmente determina a **posição** da distribuição.

Descrevendo a Variação

Distribuições de Probabilidade (continuação)

- A dispersão, espalhamento, ou variabilidade na distribuição é expressa pela **variância σ^2**.

- A variância é a média dos quadrados das distâncias de cada elemento da população em relação à média.

- Se $\sigma^2 = 0$, não há variabilidade na população. À medida que a variabilidade aumenta, a variância σ^2 também aumenta.

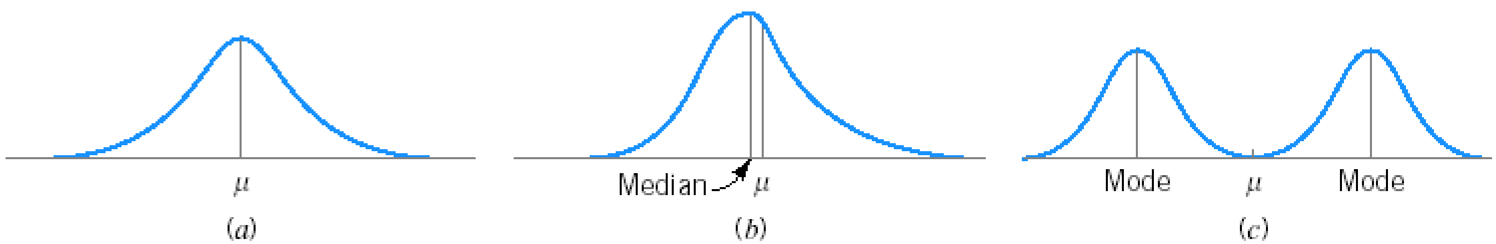
- A variância é expressa no quadrado da unidade da variável original.

Distribuições de Probabilidade (continuação)

- É costume utilizar a raiz quadrada da variância, chamada **desvio-padrão σ**.
- **O desvio-padrão é uma medida de dispersão ou espalhamento da população expressa na unidade original**. Populações com mesma média podem ter diferentes desvios-padrão.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 27-52

Distribuições de Probabilidade (continuação)



Gráficos: Média (a) Mediana (b) Moda (c) de uma Distribuição

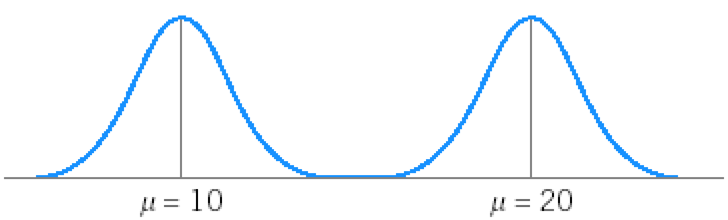


Gráfico: Duas Distribuições de Probabilidade com Médias Diferentes

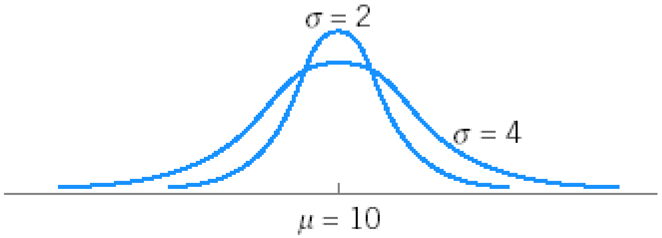


Gráfico: Duas Distribuições de Probabilidade com Mesma Média e Desvios-Padrão Diferentes

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, Cap. 02, p. 36