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

Descrevendo a Variação

O Gráfico Ramo-e-Folhas (continuação)

- Para ilustrar a construção de um gráfico ramo-e-folhas, considere os dados da Tabela a seguir que representam a produção semanal de uma fábrica de semicondutores.

Semana	Produção	Semana	Produção	Semana	Produção	Semana	Produção
1	48	11	59	21	68	31	75
2	53	12	54	22	65	32	85
3	49	13	47	23	73	33	81
4	52	14	49	24	88	34	77
5	51	15	45	25	69	35	82
6	52	16	64	26	83	36	76
7	63	17	79	27	78	37	75
8	60	18	65	28	81	38	91
9	53	19	62	29	86	39	73
10	64	20	60	30	92	40	92

Tabela: Produção Semanal de uma Unidade de Fabricação de Semicondutores

fonte: Montgomery, D. (2004). Introdução ao Controle Estatístico da Qualidade, 4a Edição, LTC - Livros Técnicos e Científicos Editora S.A. Rio de Janeiro-RJ. (pág. 28)

- Para construir o ramo-e-folhas, selecione como ramos os valores 4, 5, 6, 7, 8, e 9.

O Gráfico Ramo-e-Folhas (continuação)

- O ramo-e-folhas resultante é mostrado na figura a seguir. Uma análise do gráfico revela que a distribuição da produção tem uma forma aproximadamente simétrica, com um único pico.

Ramo	Folha	Frequência
4	8 9 7 9 5	5
5	3 2 1 2 3 9 4	7
6	3 0 4 4 5 2 0 8 5 9	10
7	9 3 8 5 7 6 5 3	8
8	8 3 1 6 5 1 2	7
9	2 1 2	3

Figura: Gráfico Ramo-e-Folhas para os Dados da Produção Semanal de Semicondutores

- O gráfico ramo-e-folhas permite-nos determinar rapidamente aspectos importantes dos dados que não são óbvios a partir da tabela original.

- Por exemplo, o gráfico ramo-e-folhas fornece uma noção da forma, da distribuição ou variabilidade, e da tendência central ou centro dos dados.

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.

Ramo	Folha	Frequência
4	5 7 8 9 9	5
5	1 2 2 3 3 4 9	7
6	0 0 2 3 4 4 5 5 8 9	10
7	3 3 5 5 6 7 8 9	8
8	1 1 2 3 5 6 8	7
9	1 2 2	3

Figura: Gráfico Ramo-e-Folhas Ordenado para os Dados da Produção Semanal de Semicondutores

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:

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

- Exemplo:

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

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

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

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

Descrevendo a Variação

O Gráfico Ramo-e-Folhas (continuação)

- Exercício: Considere o exemplo anterior considerando os dados da Tabela a seguir que representam a produção semanal de uma fábrica de semicondutores.

- Calcule:
  - A mediana;
  - O décimo percentil;
  - O primeiro quartil;
  - O terceiro Quartil.

Semana	Produção	Semana	Produção	Semana	Produção	Semana	Produção
1	48	11	59	21	68	31	75
2	53	12	54	22	65	32	85
3	49	13	47	23	73	33	81
4	52	14	49	24	88	34	77
5	51	15	45	25	69	35	82
6	52	16	64	26	83	36	76
7	63	17	79	27	78	37	75
8	60	18	65	28	81	38	91
9	53	19	62	29	86	39	73
10	64	20	60	30	92	40	92

Tabela: Produção Semanal de uma Unidade de Fabricação de Semicondutores
fonte: Montgomery, D. (2004). Introdução ao Controle Estatístico da Qualidade, 4a Edição, LTC - Livros Técnicos e Científicos Editora S.A. Rio de Janeiro-RJ. (pág. 28)

Variações do Gráfico Ramo-e-Folhas (continuação)

- No exemplo dos semicondutores, como temos um número de observações de 40 (semanas), a mediana é calculada da seguinte forma:

Semana	15	13	1	3	14	5	4	6	2	9	12	11	8	20	19	7	10	16	18	22	21	25	23	39	31	37	36	34	27	17	28	33	35	26	32	29	24	38	30	40
Posto	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Produção	45	47	48	49	49	51	52	52	53	53	54	59	60	60	62	63	64	64	65	65	68	69	73	73	75	75	76	77	78	79	81	81	82	83	85	86	88	91	92	92

$$\text{Mediana} = \frac{\text{Valor Posto } n/2 + \text{valor posto } n/2 + 1}{2} = \frac{65 + 68}{2} = 131/2, \text{ logo mediana} = 66{,}5$$

Descrevendo a Variação

Variações do Gráfico Ramo-e-Folhas (continuação)

- O **décimo percentil** é a observação com posto (0,1)*(*n*) + 0,5 = (0,1)*(40) + 0,5 = 4,5 → no meio da quarta e da quinta observação:

Semana	15	13	1	3	14	5	4	6	2	9	12	11	8	20	19	7	10	16	18	22
Posto	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Produção	45	47	48	49	49	51	52	52	53	53	54	59	60	60	62	63	64	64	65	65

Semana	21	25	23	39	31	37	36	34	27	17	28	33	35	26	32	29	24	38	30	40
Posto	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Produção	68	69	73	73	75	75	76	77	78	79	81	81	82	83	85	86	88	91	92	92

$$\text{Décimo Percentil} = \frac{49 + 49}{2} = 49$$

Variações do Gráfico Ramo-e-Folhas (continuação)

- O **primeiro quartil** é a observação com posto (0,25)*(*n*) + 0,5 = (0,25)*(40) + 0,5 = 10,5 → no meio da décima e da décima primeira observação:

Semana	15	13	1	3	14	5	4	6	2	9	12	11	8	20	19	7	10	16	18	22	21	25	23	39	31	37	36	34	27	17	28	33	35	26	32	29	24	38	30	40
Posto	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Produção	45	47	48	49	49	51	52	52	53	53	54	59	60	60	62	63	64	64	65	65	68	69	73	73	75	75	76	77	78	79	81	81	82	83	85	86	88	91	92	92

$$\text{Primeiro Quartil} = \frac{53 + 54}{2} = 107/2 = 53{,}5$$

Descrevendo a Variação

Variações do Gráfico Ramo-e-Folhas (continuação)

- O **terceiro quartil** é a observação com posto (0,75)*(*n*) + 0,5 = (0,75)*(40) + 0,5 = 70,5 → no meio da trigésima e da trigésima primeira observação:

Semana	15	13	1	3	14	5	4	6	2	9	12	11	8	20	19	7	10	16	18	22
Posto	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Produção	45	47	48	49	49	51	52	52	53	53	54	59	60	60	62	63	64	64	65	65

Semana	21	25	23	39	31	37	36	34	27	17	28	33	35	26	32	29	24	38	30	40
Posto	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Produção	68	69	73	73	75	75	76	77	78	79	81	81	82	83	85	86	88	91	92	92

$$\text{Terceiro Quartil} = \frac{79 + 81}{2} = 160/2 = 80$$

Descrevendo a Variação

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.

- Para os dados da produção de semicondutores (exemplo), o intervalo interquartil é:
  - IIQ = Q3 – Q1 = 80 – 53,5 = 26,5

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**.

Descrevendo a Variação

Variações do Gráfico Ramo-e-Folhas (continuação)

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

Folha	Ramo	Frequência
	10	
2 1 2	9	3
2 1 5 6 1 3 8	8	7
3 5 6 7 5 8 3 9	7	8
9 5 8 0 2 5 4 4 0 3	6	10
4 9 3 2 1 2 3	5	7
5 9 7 9 8	4	5

Gráfico temporal ou Gráfico de linha

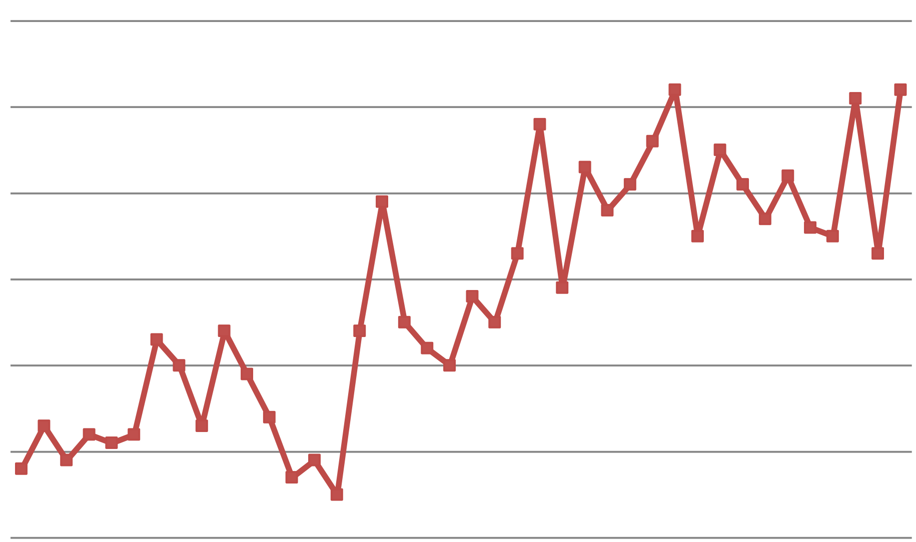


Figura: Gráfico Digidot para os Dados da Produção Semanal de Semicondutores

Descrevendo a Variação

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.
- Mais especificamente, a produção nas 20 primeiras semanas está bem abaixo da produção das últimas 20 semanas.
- Alguma coisa deve ter sido mudada no processo (ou ter sido deliberadamente mudada pelo pessoal da operação ou pelos engenheiros), levando a uma melhora na produção.

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

Descrevendo a Variação

A Distribuição de Frequências e o Histograma

- Imagine que sejam feitas 125 observações referentes ao diâmetro interno de anéis de piston usados em motores de automóveis.

- Os dados foram coletados em 25 amostras de cinco observações cada.

- Pela tabela com dados, observa-se que há uma variabilidade no diâmetro dos anéis.

- Entretanto, é muito difícil ver qualquer *padrão* de variabilidade ou *estrutura* nos dados, com as observações organizadas em forma de tabela.

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

A Distribuição de Frequências e o Histograma

	Observações				
Número da Amostra	1	2	3	4	5
1	74,030	74,002	74,019	73,992	74,008
2	73,995	73,992	74,001	74,011	74,004
3	73,988	74,024	74,021	74,005	74,002
4	74,002	73,996	73,993	74,015	74,009
5	73,992	74,007	74,015	73,989	74,014
6	74,009	73,994	73,997	73,985	73,993
7	73,995	74,006	73,994	74,000	74,005
8	73,985	74,003	73,993	74,015	73,988
9	74,008	73,995	74,009	74,005	74,004
10	73,998	74,000	73,990	74,007	73,995
11	73,994	73,998	73,994	73,995	73,990
12	74,004	74,000	74,007	74,000	73,996
13	73,983	74,002	73,998	73,997	74,012
14	74,006	73,967	73,994	74,000	73,984
15	74,012	74,014	73,998	73,999	74,007
16	74,000	73,984	74,005	73,998	73,996
17	73,994	74,012	73,986	74,005	74,007
18	74,006	74,010	74,018	74,003	74,000
19	73,984	74,002	74,003	74,005	73,997
20	74,000	74,010	74,013	74,020	74,003
21	73,988	74,001	74,009	74,005	73,996
22	74,004	73,999	73,990	74,006	74,009
23	74,010	73,989	73,990	74,009	74,014
24	74,015	74,008	73,993	74,000	74,010
25	73,982	73,984	73,995	74,017	74,013

Tabela: Diâmetro Interno (mm) de Anéis de Pistons

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, Cap. 02, p. 30

Descrevendo a Variação

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.
- Por exemplo, uma distribuição de frequência de dados dos anéis de piston é exibido na tabela a seguir...

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

Descrevendo a Variação

A Distribuição de Frequências e o Histograma

- Nesta tabela podemos ver que um anel tinha diâmetro entre 73,965mm e 73,970mm; os diâmetros de oito anéis estavam entre 73,980mm e 73,985mm; e assim por diante...

Diêmtro do Anel *x* , (mm)			Contagem	Frequência	Frequência Acumulada	Frequência Relativa	Frequência Relativa Acumulada
73,965	$\leq x <$	73,970	I	1	1	0,008	0,008
73,970	$\leq x <$	73,975		0	1	0,000	0,008
73,975	$\leq x <$	73,980		0	1	0,000	0,008
73,980	$\leq x <$	73,985	~~IIII~~ III	8	9	0,064	0,072
73,985	$\leq x <$	73,990	~~IIII~~ ~~IIII~~	10	19	0,080	0,152
73,990	$\leq x <$	73,995	~~IIII~~ ~~IIII~~ ~~IIII~~ IIII	19	38	0,152	0,304
73,995	$\leq x <$	74,000	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ III	23	61	0,184	0,488
74,000	$\leq x <$	74,005	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ II	22	83	0,176	0,664
74,005	$\leq x <$	74,010	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ II	22	105	0,176	0,840
74,010	$\leq x <$	74,015	~~IIII~~ ~~IIII~~ III	13	118	0,104	0,944
74,015	$\leq x <$	74,020	IIII	4	122	0,032	0,976
74,020	$\leq x <$	74,025	II	2	124	0,016	0,992
74,025	$\leq x <$	74,030	I	1	125	0,008	1,000
			total	125		1,000	

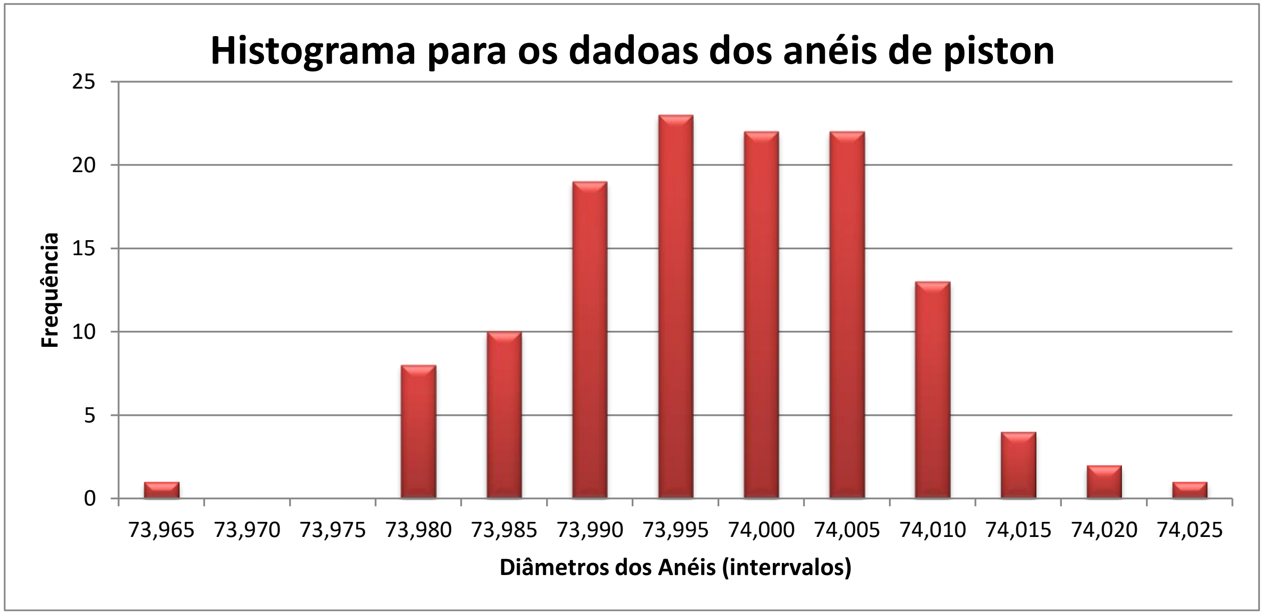
Tabela: Distribuição de Frequência para os Diâmetros Internos (mm) dos Anéis de Pistons

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, Cap. 02, p. 30

Descrevendo a Variação

A Distribuição de Frequências e o Histograma

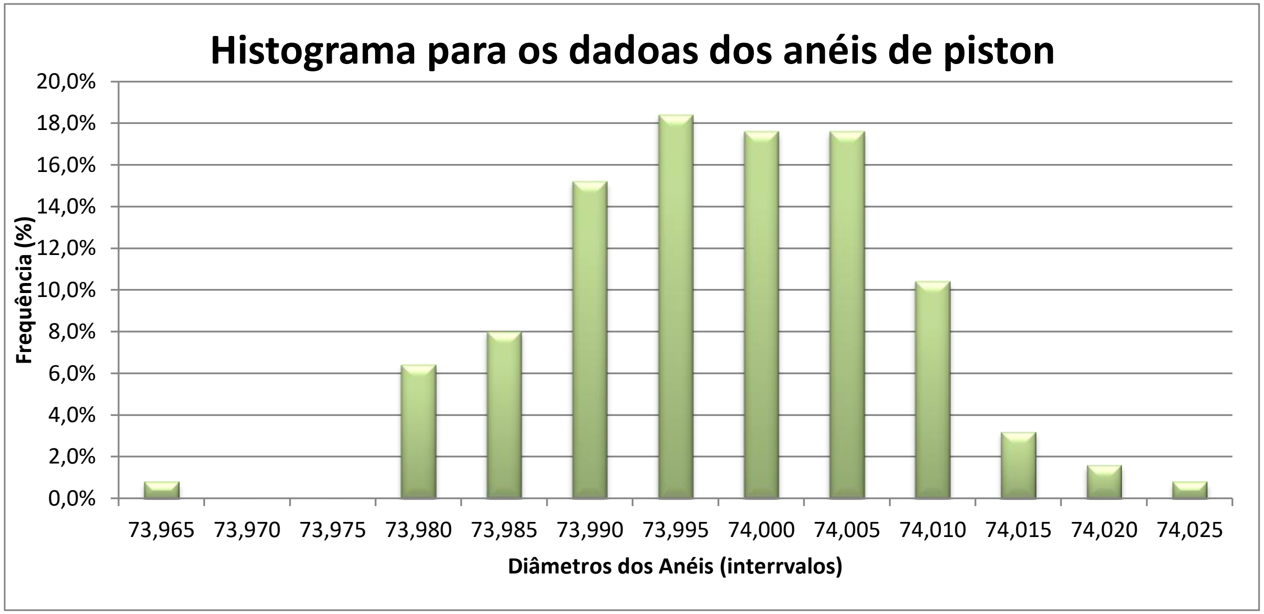
- O gráfico das frequências observadas *versus* o diâmetro do anel é exibido na figura é chamado **histograma**!
- A altura de cada barra é igual à frequência de ocorrência do diâmetro do anel na faixa (intervalo) especificada.



Descrevendo a Variação

A Distribuição de Frequências e o Histograma

- O gráfico das frequências observadas *versus* o diâmetro do anel é exibido na figura é chamado **histograma**!
- A altura de cada barra é igual à frequência de ocorrência do diâmetro do anel na faixa (intervalo) especificada.



Descrevendo a Variação

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

- Para os dados do exemplo dos diâmetros dos anéis de piston, podemos ver que a distribuição é:
  - aproximadamente simétrica e unimodal (em forma de monte);
  - com tendência central bem próxima de 74 mm

- A variabilidade nos diâmetros dos anéis é aparentemente alta, uma vez que alguns anéis são tão pequenos quanto 73,967 mm e outros tão grandes quanto 74,030 mm.

- Assim, 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

- Algumas <u>diretrizes</u> 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

Descrevendo a Variação

Resumo Numérico dos Dados

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

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

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

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

Descrevendo a Variação

Resumo Numérico dos Dados

- Note que a média amostral $\bar{x}$ é simplesmente a média aritmética das *n* observações.

- A média amostral para os dados dos anéis de piston é:

$$\bar{x} = \frac{\sum_{i=1}^{125} x_i}{125} = \frac{9250,147}{125} = 74,001\ mm$$

- 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.

Descrevendo a Variação

Resumo Numérico dos Dados

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

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

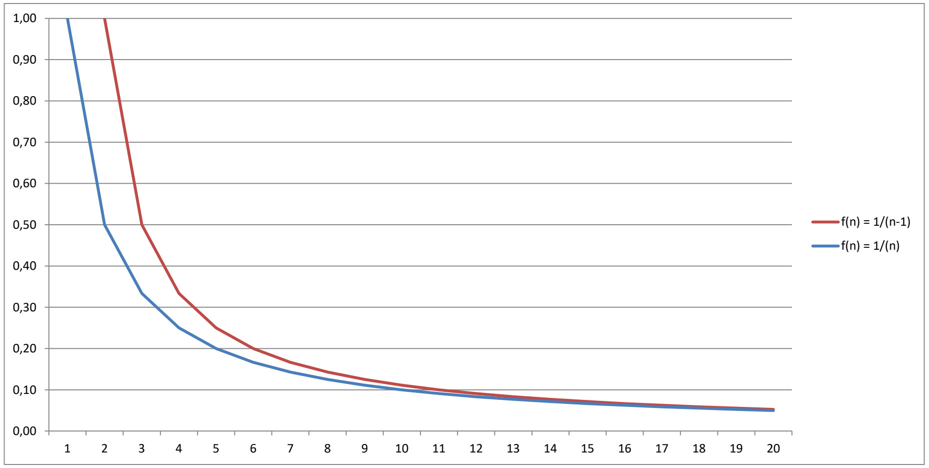
- Note que a variância amostral é simplesmente a soma dos quadrados dos desvios de cada observação em relação à média amostral $\bar{x}$, dividida pelo tamanho da amostra menos um.

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:

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

Descrevendo a Variação

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.

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

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*)

Descrevendo a Variação

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.

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

O Diagrama de Caixa *(Box Plot)*

- Os dados da tabela representam amostras dos diâmetros (em mm) de orifícios em um conjunto de 12 nervuras do bordo da asa de um avião de transporte comercial.

Medidas	1	2	3	4	5	6	7	8	9	10	11	12
Valores (mm)	120,5	120,9	120,3	121,3	120,4	120,2	120,1	120,5	120,7	121,1	120,9	120,8

Tabela: Diâmetro dos Orifícios (em mm) das Nervuras do Bordo da Asa

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, Cap. 02, p. 32

- Podemos, como primeiro passo, ordenar as observações por ordem crescente dos valores dos diâmetros observados em cada amostra.

Descrevendo a Variação

O Diagrama de Caixa *(Box Plot)*

- Note que a mediana dos dados está a meio caminho da sexta e sétima observações ordenadas.

Medidas	7	6	3	5	1	8	9	12	2	11	10	4
Posição ou Posto	1	2	3	4	5	6	7	8	9	10	11	12
Valores (mm) ordenados	120,1	120,2	120,3	120,4	120,5	120,5	120,7	120,8	120,9	120,9	121,1	121,3

$$\text{Mediana } (\tilde{x}) = \frac{\text{Valor Posto } n/2 + \text{valor posto } n/2 + 1}{2} = \frac{120{,}5 + 120{,}7}{2} \rightarrow \text{logo mediana } \tilde{x} = 120{,}6$$

- Podemos, também, calcular os primeiro e terceiro quartis...

O Diagrama de Caixa *(Box Plot)*

- Cálculo do primeiro quartil Q1:
  - posto (0,25)*(*n*) + 0,5 = (0,25)*(12) + 0,5 = 3,5

Medidas	7	6	3	5	1	8	9	12	2	11	10	4
Posição ou Posto	1	2	3	4	5	6	7	8	9	10	11	12
Valores (mm) ordenados	120,1	120,2	120,3	120,4	120,5	120,5	120,7	120,8	120,9	120,9	121,1	121,3

$$\text{Primeiro Quartil} = \frac{120{,}3 + 120{,}4}{2} = 240{,}7/2 = 120{,}35$$

O Diagrama de Caixa *(Box Plot)*

- Cálculo do terceiro quartil Q3:
  - posto (0,75)*(*n*) + 0,5 = (0,75)*(12) + 0,5 = 9,5

Medidas	7	6	3	5	1	8	9	12	2	11	10	4
Posição ou Posto	1	2	3	4	5	6	7	8	9	10	11	12
Valores (mm) ordenados	120,1	120,2	120,3	120,4	120,5	120,5	120,7	120,8	120,9	120,9	121,1	121,3

$$\text{Terceiro Quartil} = \frac{120{,}9 + 120{,}9}{2} = 241{,}8/2 = 120{,}9$$

Descrevendo a Variação

O Diagrama de Caixa *(Box Plot)*

- O diagrama de caixa é apresentado na figura a seguir:

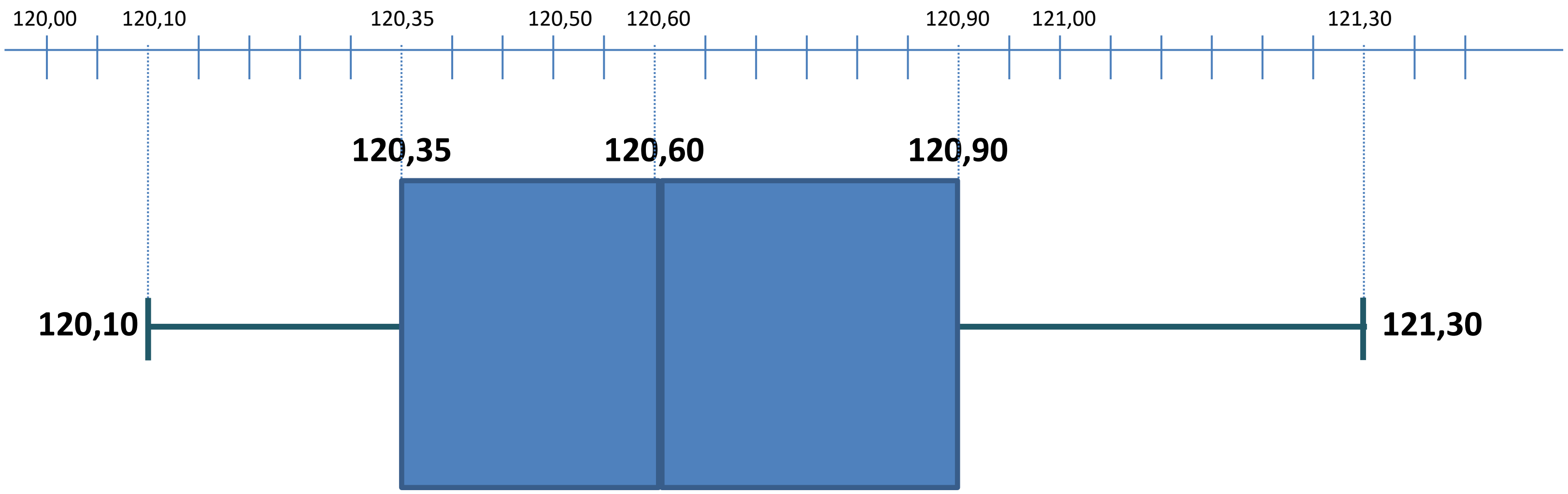


Figura: Diagrama de Caixa para os Diâmetros de Orifícios (em mm) das Nervuras do Bordo da Asa

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, Cap. 02, p. 32

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

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.

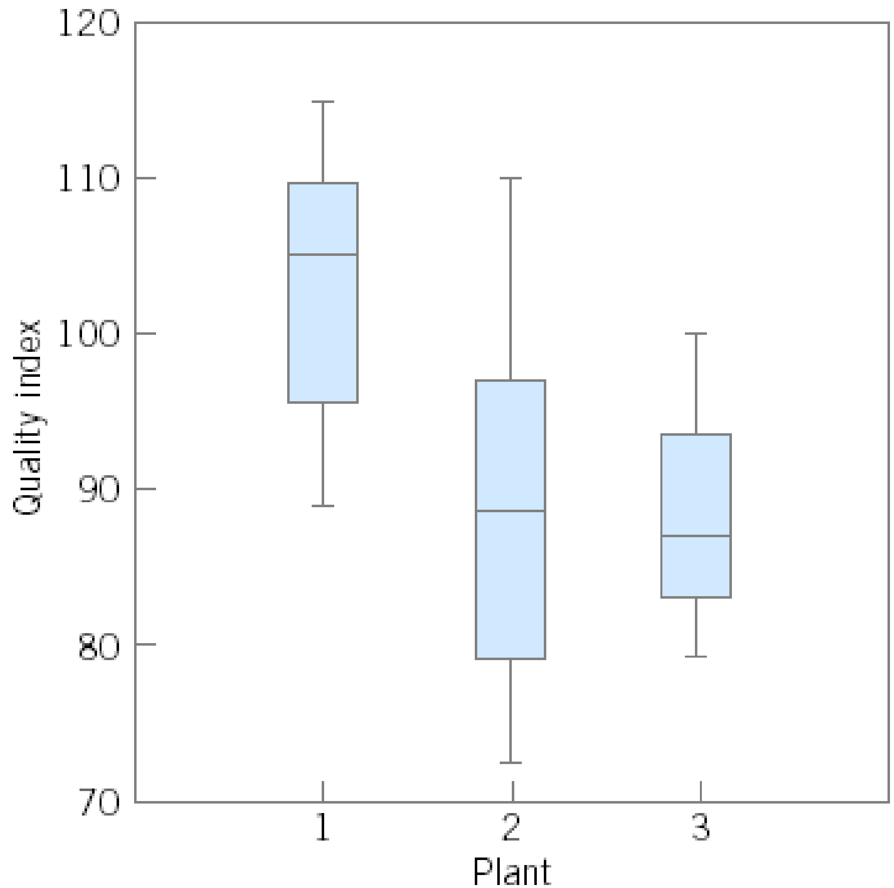


Figure 2-8 Comparative box plots of a quality index for products produced at three plants.

Parte I: Modelando a Qualidade do Processo

DISTRIBUIÇÕES DE PROBABILIDADE INTRODUÇÃO

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**.

Distribuições de Probabilidade (introdução)

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

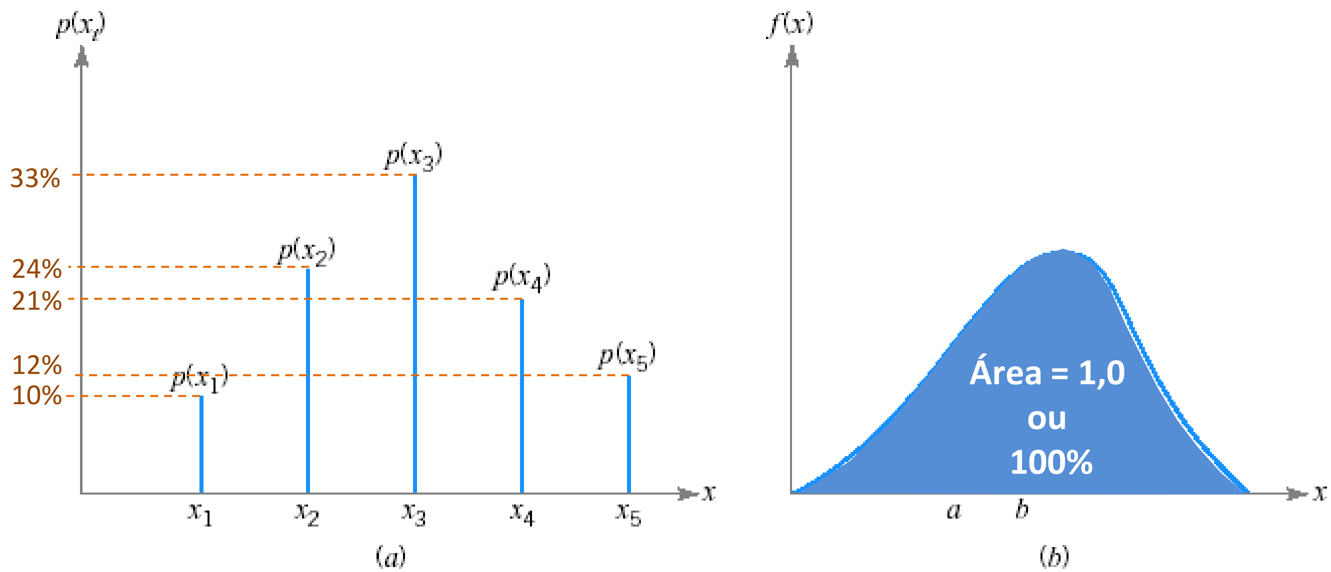
Distribuições de Probabilidade (continuação)

- 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

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

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.

Distribuições de Probabilidade (introdução)

Distribuições de Probabilidade (continuação)

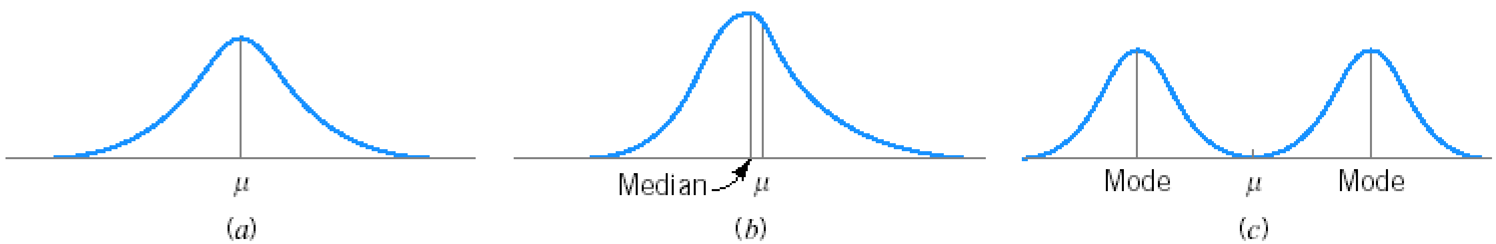
- A dispersão, espalhamento, ou variabilidade na distribuição é expressa pela **variância σ^2**.
- A variância é a média dos quadrados das distâncias de cada elemento da população em relação à média.
- Se $\sigma^2 = 0$, não há variabilidade na população. À medida que a variabilidade aumenta, a variância 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.

Distribuições de Probabilidade (continuação)



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

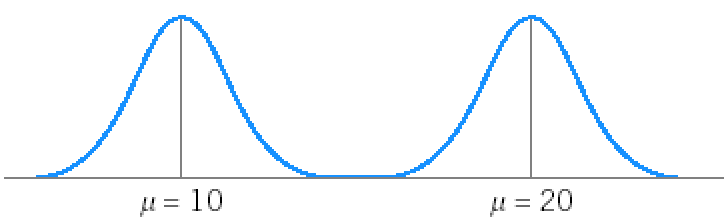


Gráfico: Duas Distribuições de Probabilidade com Médias Diferentes

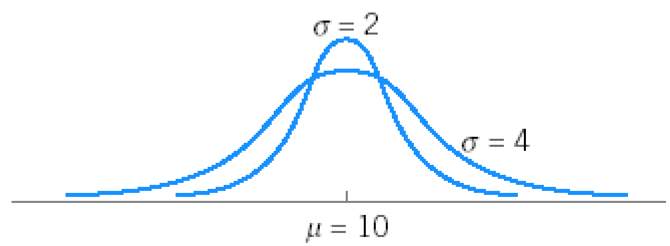


Gráfico: Duas Distribuições de Probabilidade com Mesma Média e Desvios-Padrão Diferentes

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