



ENGENHARIA DA QUALIDADE A ENG 09008

AULA 2 REVISÃO DE ESTATÍSTICA

PROFESSORES:

**CARLA SCHWENGBER TEN CATEN
ROGÉRIO FEROLDI MIORANDO**

Introdução

Em um ambiente industrial, os dados devem formar a base para as decisões e ações.

Uma vez que os dados brutos tenham sido coletados, eles devem ser tabulados e convertidos em “informação” através do uso de métodos estatísticos.

Coleta de dados

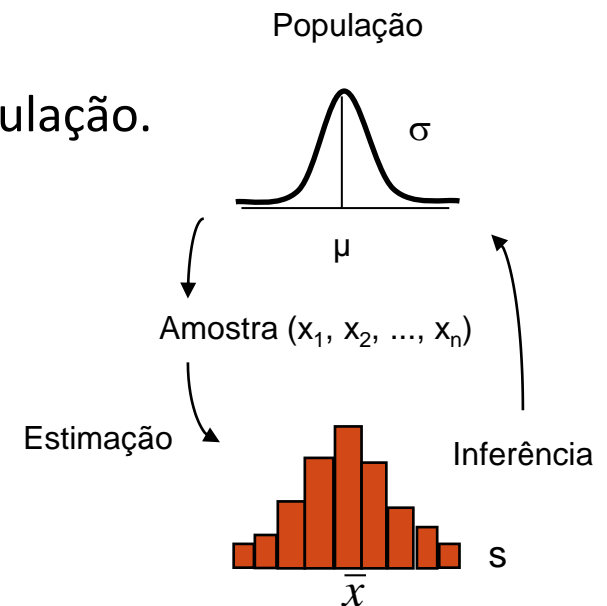
População: corresponde ao sistema ou ao todo que se quer descrever. É um conjunto de elementos com características comuns.

- **Censo:** inspeciona todos os elementos de uma população.

- **Parâmetros:** valor desconhecido associado a uma característica (média = μ , variância = σ^2)

- **Amostra:** é uma parte representativa da população.

- **Estimador:** função que estima o valor de um Parâmetro baseando-se nas observações (média = \bar{x} , variância = s^2)



|| Estratificação de dados

Trabalha-se com dados classificados em agrupamentos (camadas ou estratos)

- Tempo: os resultados são diferentes de manhã, à tarde ou a noite?
- Local: os resultados são diferentes nas linhas de produção?
- Tipo: os resultados obtidos são diferentes entre os fornecedores?
- Indivíduo: é possível comparar os operadores?

Tipos de dados

Atributo – é resultado da contagem de peças/defeitos que não atendem determinada especificação gerando dados discretos

- Percentual $p = \frac{\text{num. de defeituosos}}{\text{num total de peças}}$
 - Conforme = 0, Não-conforme = 1 logo, $0 < p < 1$
- Taxa= número de defeitos/meio contínuo
- taxa = 0,1,2,3,4,... ∞

Variáveis – é resultado de um sistema de medição gerando dados contínuos: infinitos valores possíveis entre dois extremos

- Tempo (1h:35min),
- Pressão (1.013,105 KPa) ,
- Dimensão (16,54 mm),
- Temperatura (23,5°C)

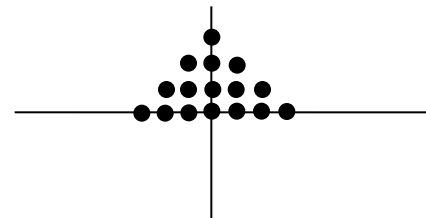
||| Análise de dados

- 1) Medidas de tendência central
- 2) Medidas de variabilidade
- 3) Histograma
- 4) Boxplot
- 5) Distribuição de probabilidade Normal
- 6) Gráfico de normalidade

1) Medidas de tendência central

A tendência central é uma medida do centro de um conjunto de dados segundo uma regra estabelecida a priori (média aritmética, geométrica, harmônica, ponderada, etc.)

- Média aritmética
- Mediana
- Moda



1) Medidas de tendência central

Média aritmética $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

- Anotamos a temperatura de uma pessoa de 1 em 1 hora, durante 8 horas. Qual a média da temperatura?
- Valores observados: 37, 37, 38, 39, 37, 39, 39°C.
- O tamanho da amostra é $n = 7$

$$\bar{x} = \frac{37 + 37 + 38 + 39 + 37 + 39 + 39}{7} = 38^{\circ}\text{C}$$

A média amostral é bom um estimador da média populacional. Quanto maior n melhor a estimativa.

1) Medidas de tendência central

Mediana $\tilde{x} = \begin{cases} x_{((n+1)/2)} & n \text{ ímpar} \\ \frac{x_{(n/2)} + x_{(n/2+1)}}{2} & n \text{ par} \end{cases}$

- Ela é não é influenciada pelos dados atípicos
- Deve-se ordenar os dados em ordem crescente
- Qual a mediana da temperatura?
- Valores observados: 37, 37, 38, 39, 37, 39, 39°C.
- Valores ordenados: 37, 37, 37, 38, 39, 39, 39°C
- $n = 7$ é ímpar – mediana valor central $\tilde{x} = 38^\circ\text{C}$

1) Medidas de tendência central

Moda: observação que ocorre com mais frequência

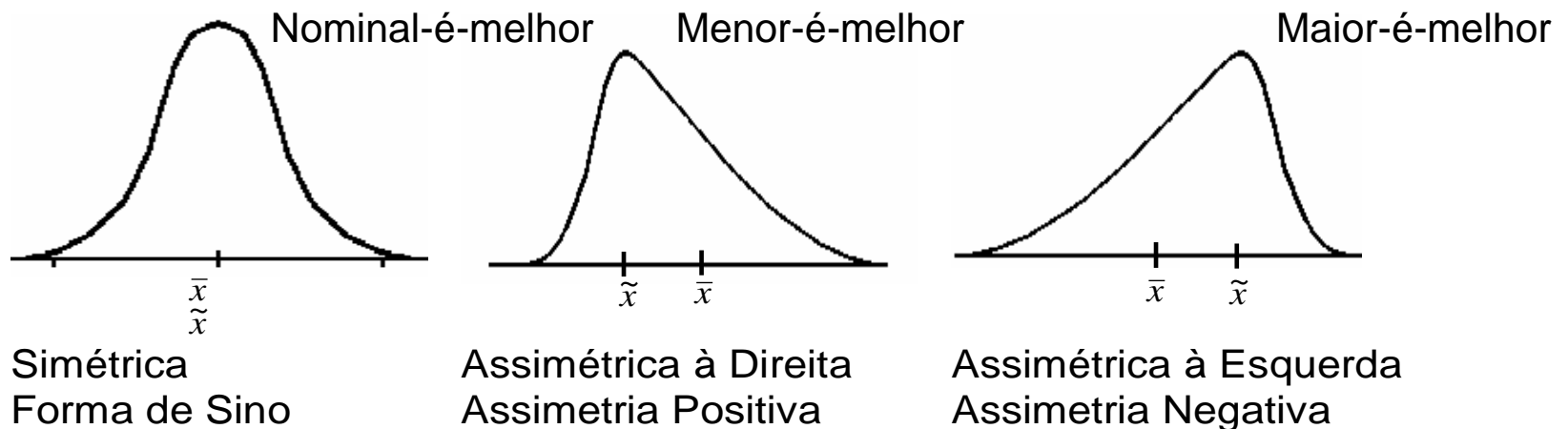
- Qual a moda da temperatura?
- Valores observados: 37, 37, 37, 38, 39, 39, 39°C
- Duas modas: 37 e 39°C

1) Medidas de tendência central

Relação entre média e mediana → fornece a forma da dispersão

A	Distribuição simétrica	10 12 14 16 18	$\bar{x} = 14 = \tilde{x} = 14$
B	Distribuição assimétrica à direita	10 12 14 16 23	$\bar{x} = 15 > \tilde{x} = 14$
C	Distribuição assimétrica à esquerda	05 12 14 16 18	$\bar{x} = 13 < \tilde{x} = 14$

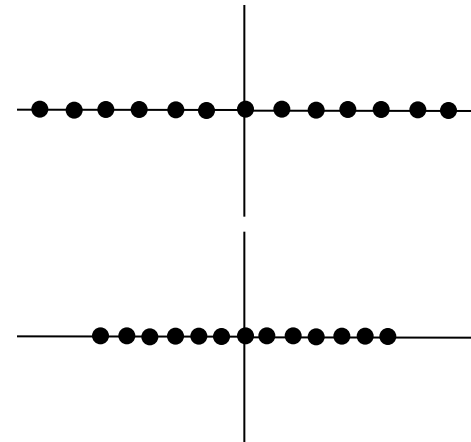
Mediana tem maior robustez a dados atípicos do que a média



2) Medidas de variabilidade

Observações individuais apresentam dispersão em torno do valor médio. Isto chama-se variabilidade dos dados

- Amplitude
- Quartil
- Desvio-padrão
- Coeficiente de Variação



2) Medidas de variabilidade

Amplitude: $R = X_{\max} - X_{\min}$

Exemplo: 8,5 8,7 8,9 10,1 10,5 10,7 11,5 11,9

$$R = 11,9 - 8,5 = 3,4$$

- A amplitude é fácil de calcular e fornece uma idéia da magnitude da faixa de variação dos dados.
- Não informa a respeito da dispersão dos valores que caem entre os dois extremos.
- Ela é influenciada pelos dados atípicos
- Quando $n < 10$ pode resultar em uma medida de variação bastante satisfatória.

2) Medidas de variabilidade

Quartis

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. Ela não é influenciada pelos dados atípicos

- **1º quartil ou quartil inferior (Q1)** = valor aos 25% da amostra ordenada
- **2º quartil ou mediana (Q2)** = valor até ao qual se encontra 50% da amostra ordenada
- **3º quartil ou quartil superior (Q3)** = valor a aos 75% da amostra ordenada

2) Medidas de variabilidade

Exemplo

Amostra: 36, 40, 7, 41, 15, 39

Amostra ordenada: 7, 15, 36, 39, 40, 41

$$Q1 = 15$$

$$Q2 = (39+36)/2 = 37,5$$

$$Q3 = 40$$

Intervalo inter-quartil: $Q3 - Q1$ ($40 - 15 = 25$)

Regra para descobrir os quartis

- 1) use a mediana para dividir os dados ordenados em duas metades, não inclua a mediana nas metades
- 2) o quartil inferior (ou superior) é a mediana da metade inferior (ou superior).

2) Medidas de variabilidade

Variância

Quadrado da distância de todos os valores x_i em relação a sua média

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

Desvio-padrão

A raiz quadrada da variância (é expresso na unidade original dos dados)

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

2) Medidas de variabilidade

Nem sempre se conhece a variância e o desvio padrão populacional. Desta forma, deve-se usar um estimador a partir de uma amostra

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} \quad \text{populacional}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} \quad \text{populacional}$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad \text{amostral}$$

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad \text{amostral}$$

A correção de Bessel (eliminando 1 grau de liberdade para $n < 30$) torna a variância amostral um estimador da variância populacional não-viesado.

2) Medidas de variabilidade

Exemplo

Amostra: 10 12 14 16 18 (*meses*)

A média é 14 cm, a variância e o desvio-padrão são:

$$s^2 = \frac{(10-14)^2 + (12-14)^2 + (14-14)^2 + (16-14)^2 + (18-14)^2}{5-1} = 9,98 \text{ meses}^2$$

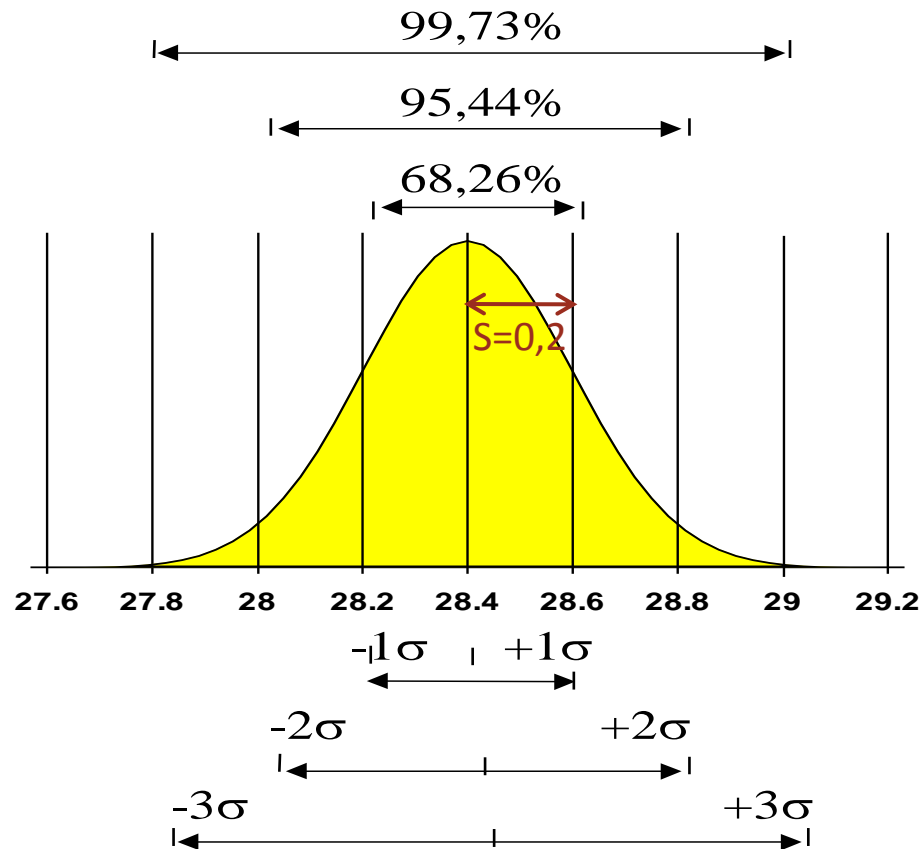
$$s = \sqrt{9,98 \text{ cm}^2} = 3,16 \text{ meses}$$

Os desvios de cada valor em relação à média totalizam zero pois a média é o valor central

A média e o desvio padrão possuem a mesma unidade de medida

2) Medidas de variabilidade

Seja um processo com média 28,4 e desvio-padrão $S = 0,2$



2) Medidas de variabilidade

Coeficiente de variação $CV = \frac{s}{\bar{x}} \times 100$

- Um desvio padrão pode ser considerado grande ou pequeno dependendo da ordem de grandeza da média da variável.
- Quanto menor o CV mais homogêneo é o conjunto de dados.
- Medida adimensional, útil para comparar resultados de amostras cujas unidades podem ser diferentes.

2) Medidas de variabilidade

Exemplo

Duas turmas de Eng. da Qualidade obtiveram as seguintes notas:

Turma B: Média = 50, Desvio Padrão = 5

Turma C: Média = 70, Desvio Padrão = 7

Qual das turmas é relativamente mais precisa?

$$CV B = (5 / 50) * 100 = 10\%$$

$$CV C = (7 / 70) * 100 = 10\%$$

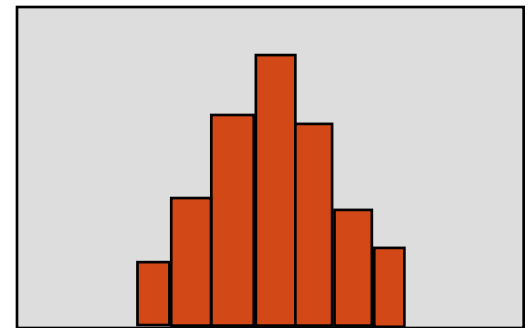
As duas turmas são igualmente precisas (homogênea).

3) Histograma

O histograma é um gráfico de barras cujo eixo horizontal representa a variação total da característica de qualidade subdividida em vários intervalos.

Para cada um destes intervalos é construída uma barra vertical proporcional ao número de observações na amostra pertencente ao respectivo intervalo.

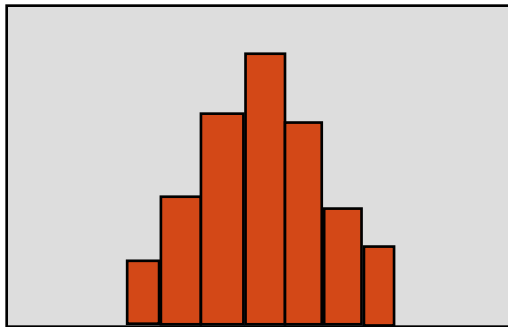
A construção de histogramas tem carácter preliminar em qualquer estudo e é um importante indicador da distribuição de dados .



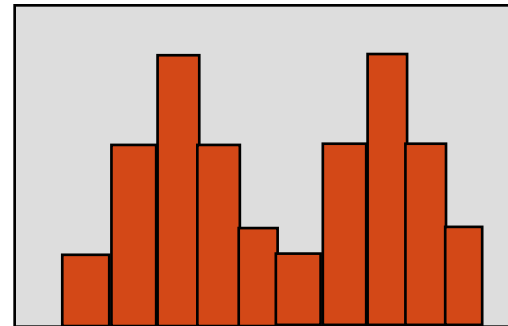
3) Histograma

Tipos de Histograma

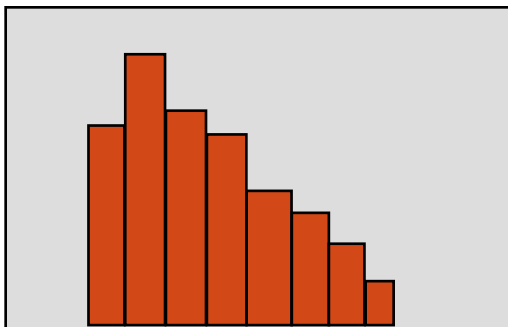
Forma de Sino



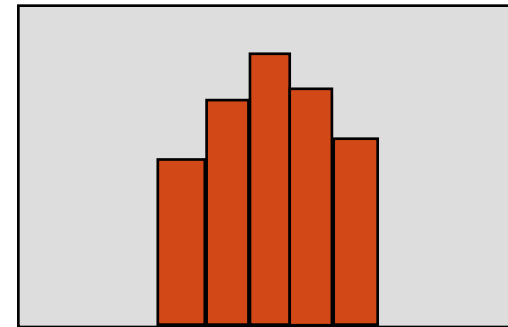
Bi-modal



Assimétrico

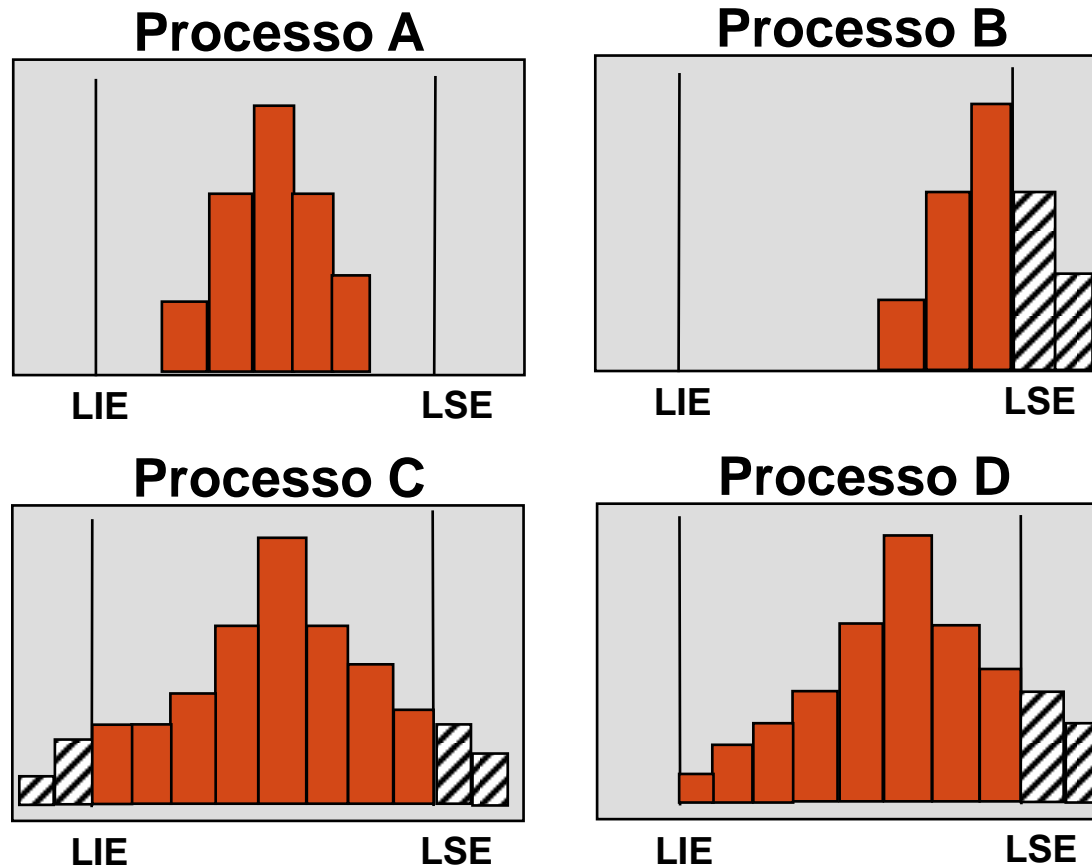


Truncado



3) Histograma

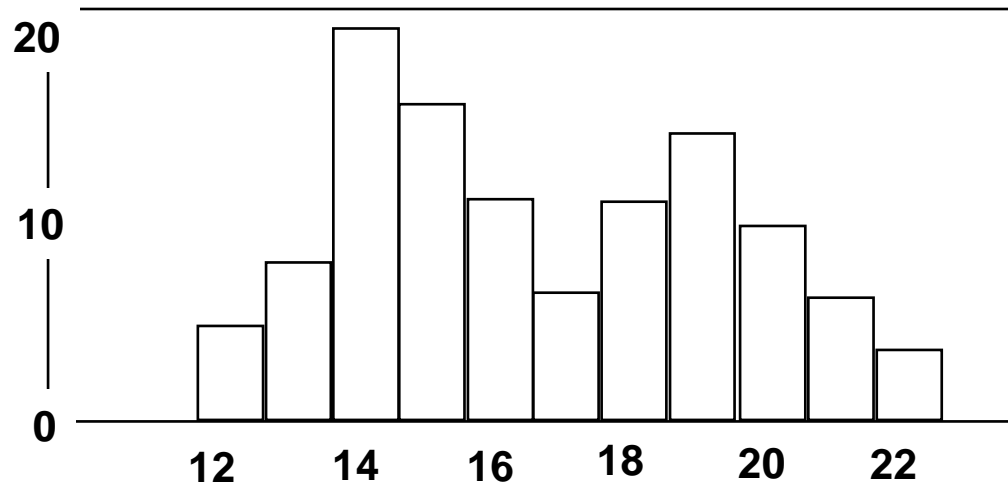
Comparação com as especificações



3) Histograma

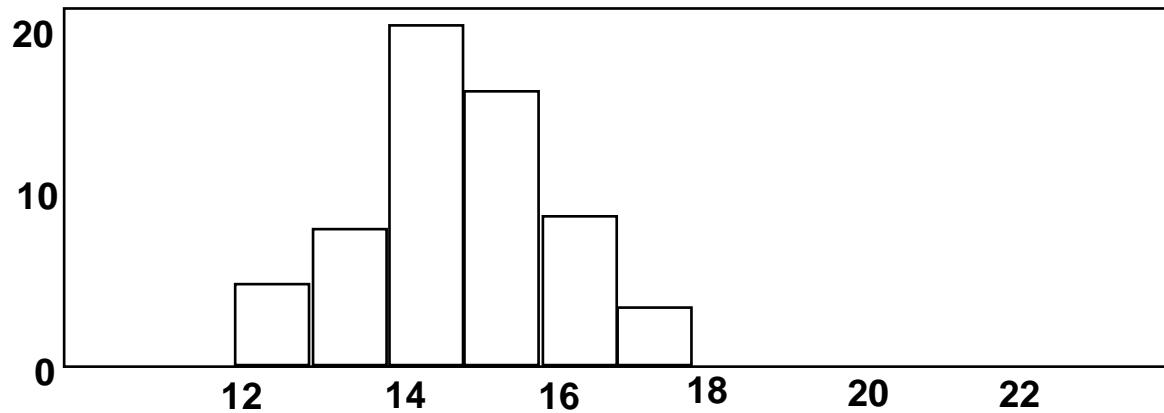
Estratificação

Muitas vezes identifica-se distribuições diferentes para níveis distintos dos fatores estratificados (mistura de populações quando se apresentam bimodais)

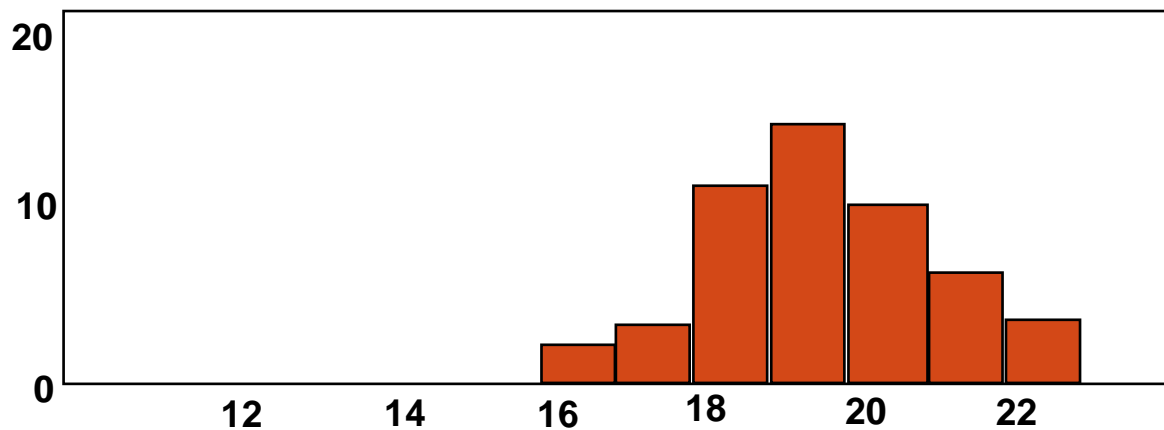


3) Histograma

Estratificação



Fornec. A



Fornec. B

3) Histograma

Exemplo

A Tabela abaixo apresenta 50 observações de tempos de atendimento em minutos numa central telefônica (em ordem crescente).

→

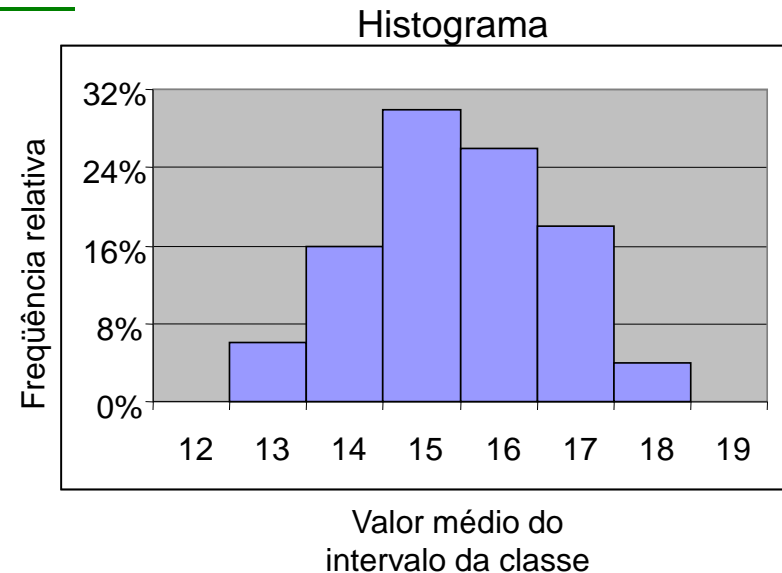
12,58	12,97	13,45	13,53	13,59	13,61	13,62	13,78	13,97	14,21
14,47	14,51	14,53	14,58	14,65	14,78	14,83	14,97	15,06	15,13
15,17	15,23	15,29	15,37	15,40	15,45	15,51	15,62	15,67	15,73
15,83	15,98	16,01	16,11	16,17	16,23	16,35	16,43	16,49	16,52
16,67	16,83	16,97	17,05	17,13	17,22	17,3	17,48	17,8	18,47

↓

3) Histograma

Histograma de freqüências relativas.

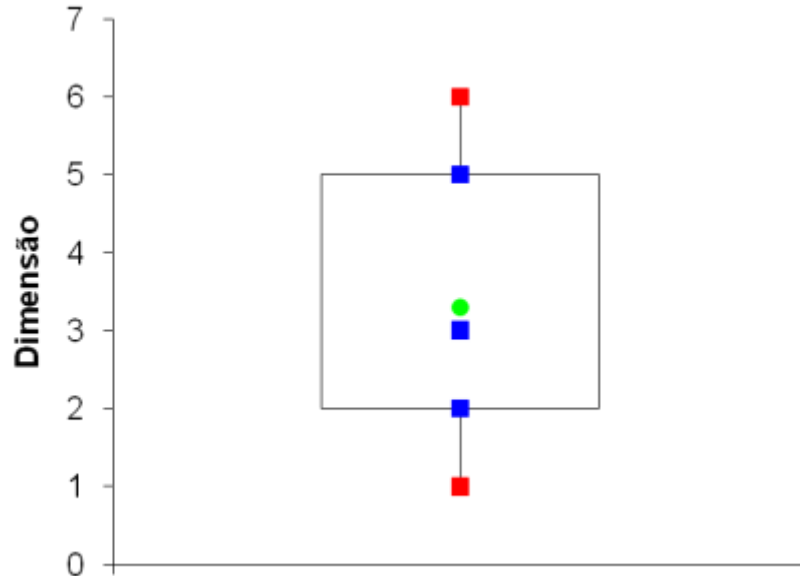
Intervalos de classe	Freqüência absoluta	Freqüência relativa
12,50 a 13,50	3	6%
13,51 a 14,50	8	16%
14,51 a 15,50	15	30%
15,51 a 16,50	13	26%
16,51 a 17,50	9	18%
17,51 a 18,50	2	4%



4) BoxPlot

Gráfico que apresenta a variabilidade de um conjunto de dados através de 6 medidas

Exemplo: 1, 2, 2, 3, 4, 5, 6



Valor máximo = 6

Q3 = 5

\bar{x} = média = 3,3

Q2 = Mediana = 3

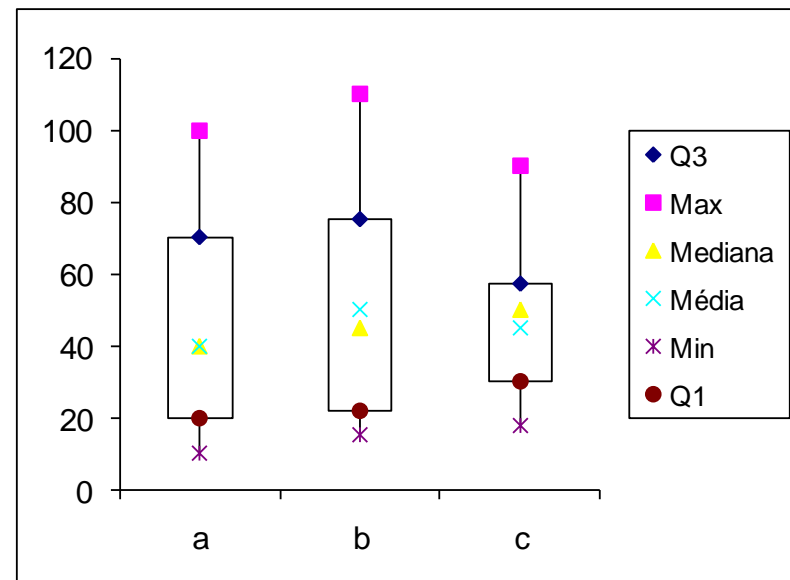
Q1 = 2

Valor mínimo = 1

4) BoxPlot

Útil para comparar dispersão, tendência central e pontos extremos de diversas populações (processos) sem fazer suposição quanto a distribuição estatística. Também pode indicar assimetria.

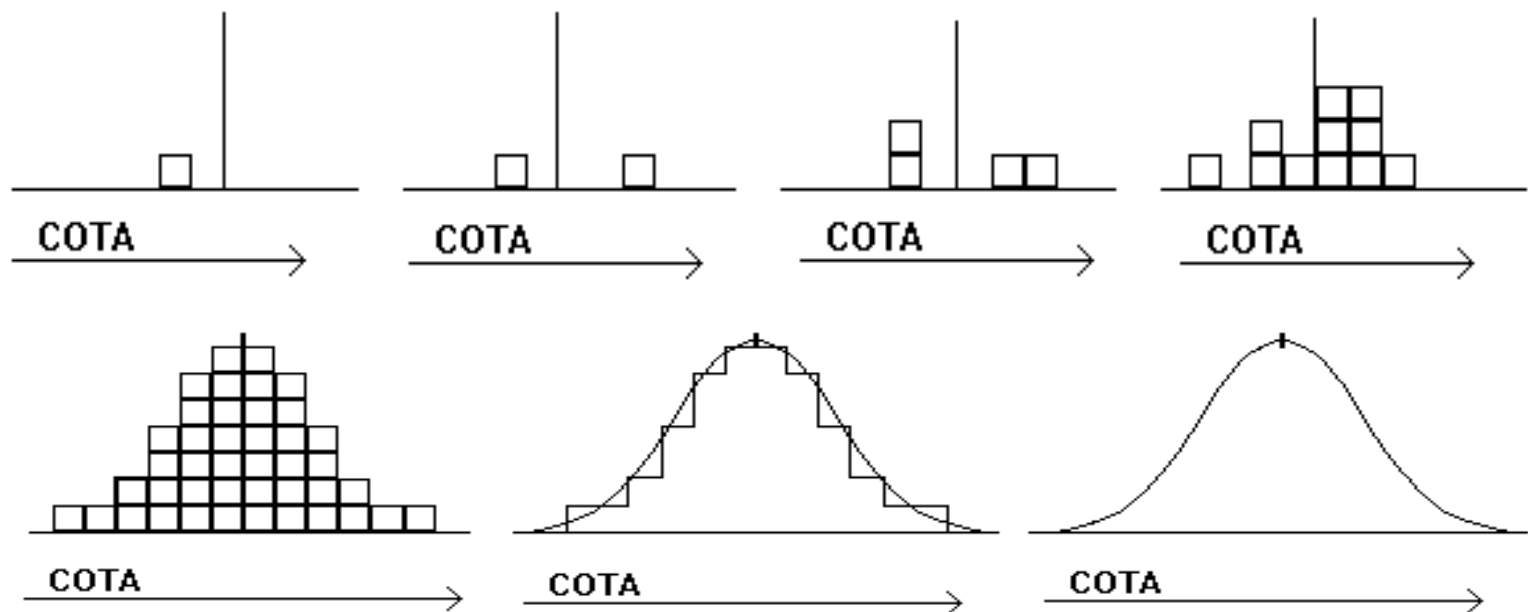
	a	b	C
Q3	70	75	57
Max	100	110	90
Mediana	40	45	50
Média	40	40	50
Min	10	15	18
Q1	20	22	30



5) Distribuição de probabilidades

Devido à variabilidade inerente do processo, as medidas individuais são diferentes, mas em grupo elas tendem a formar um padrão.

Quando o processo é estável, esse padrão pode ser descrito por uma distribuição de probabilidade.



5) Distribuição de probabilidades

Uma distribuição de probabilidade é um modelo matemático que relaciona um certo valor da variável em estudo com a sua probabilidade de ocorrência.

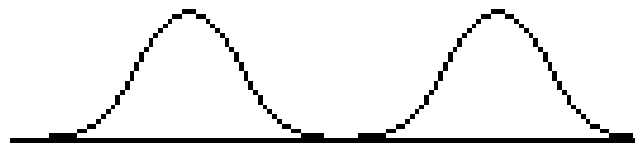
Há dois tipos de distribuição de probabilidade:

- **Distribuições Discretas:** Quando a variável que está sendo medida só pode assumir certos valores, como por exemplo os valores inteiros: 0, 1, 2, etc.
-> Distribuição Binomial, Poisson
- **Distribuições Contínuas:** Quando a variável que está sendo medida é expressa em uma escala contínua, como no caso de uma característica dimensional. -> Distribuição Normal

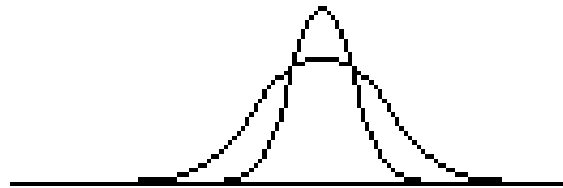
5) Distribuição de probabilidades

Uma distribuição de probabilidade pode ser caracterizada por diversos parâmetros.

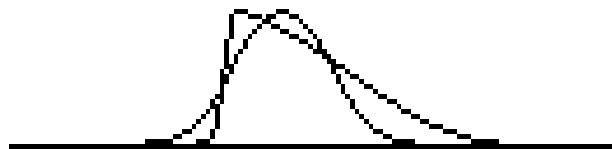
Os principais são:



Localização

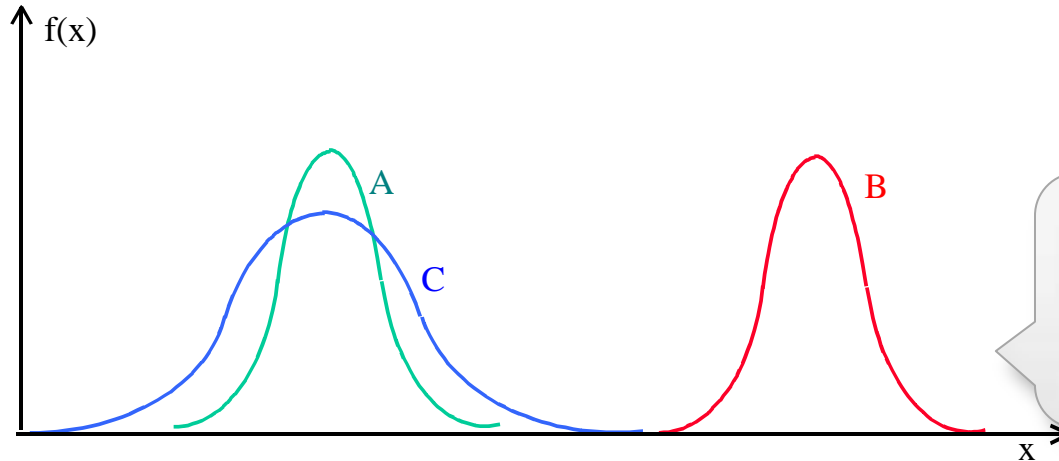


Dispersão



Forma

5) Distribuição de probabilidades

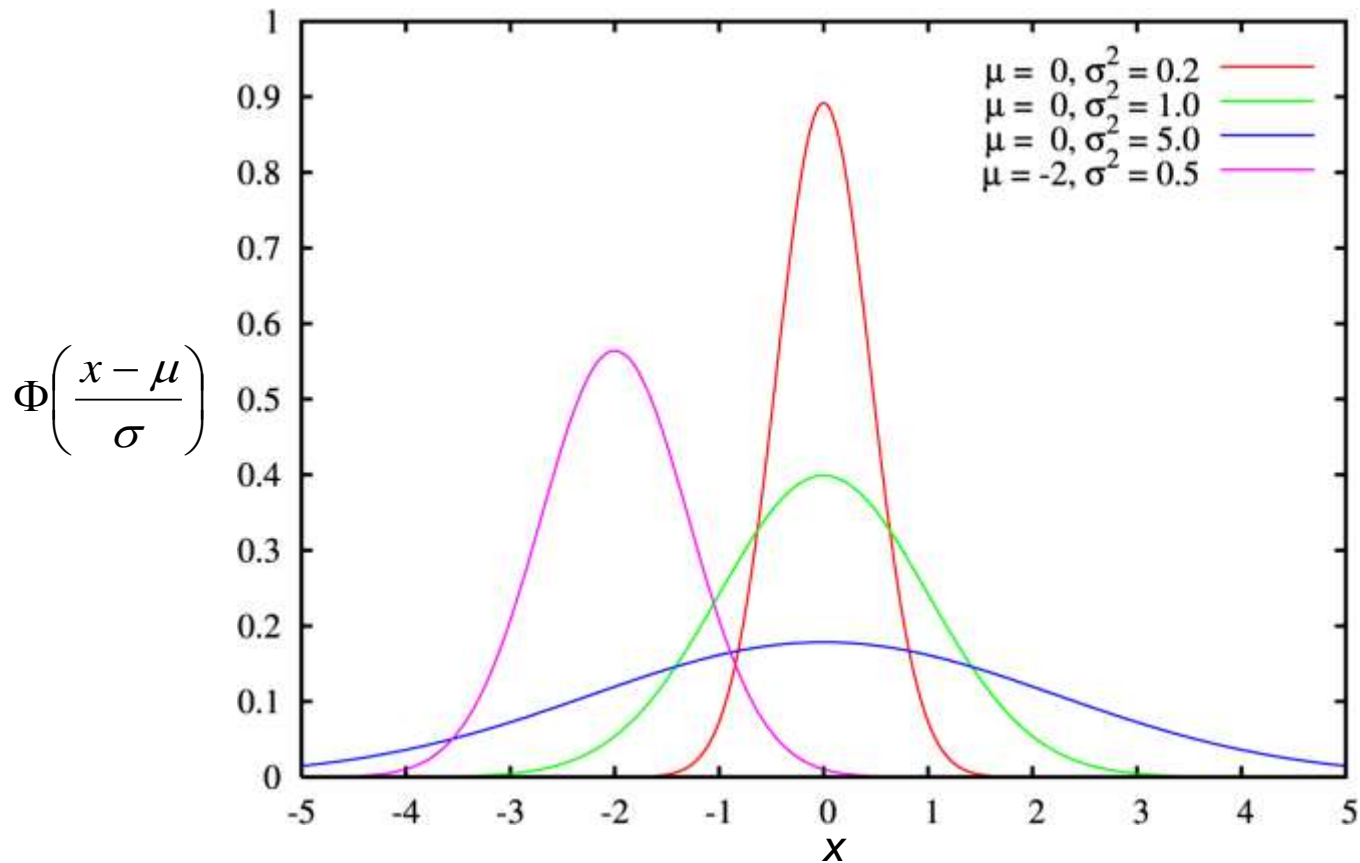


A distribuição Normal fica completamente caracterizada por dois parâmetros: a média e o desvio-padrão.

- Da distribuição A para B muda a tendência central, mas a variabilidade é constante
 - ex: mesma máquina (variabilidade), produtos com cotas (médias) diferentes
- Da distribuição A para C muda a variabilidade, mas a tendência central é constante
 - ex: mesma cota do produto (média) com máquinas diferentes (variabilidade)
- Da distribuição B para C muda a tendência central e a variabilidade - ex: produto diferente, máquina diferente

5) Distribuição de probabilidades

Função densidade de probabilidade normal acumulada para quatro diferentes conjuntos de parâmetros (μ, σ^2)



$N \sim (0, 1^2)$
A linha verde representa a distribuição normal padronizada (que está tabelada)

5) Dist. de probabilidade Normal

A distribuição Normal é completamente caracterizada por sua média e desvio-padrão... permitindo que a área sob a curva entre um ponto qualquer e a média seja função somente do número de desvios-padrão relativo a esta distância.

Como existem uma infinidade de distribuições normais (uma para cada média e desvio-padrão), transformamos a unidade estudada (peso, espessura, tempo, etc.) na unidade Z , que indica o número de desvios-padrão a contar da média.

Dessa forma, o cálculo de probabilidades (área sob a curva) pode ser realizado através de uma distribuição Normal padronizada, onde o parâmetro é a variável reduzida Z .

5) Dist. de probabilidade Normal

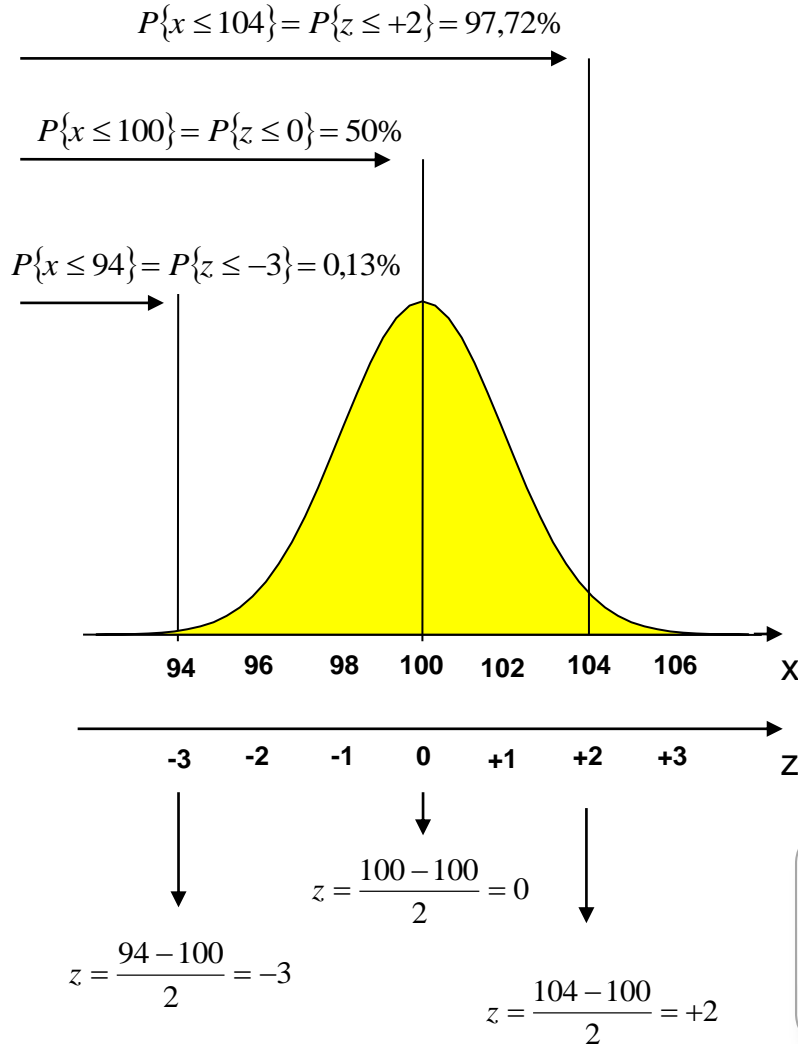
Z é chamada de variável padronizada, e a distribuição dos valores de Z é chamada de distribuição Normal padronizada.

O cálculo da variável reduzida Z faz uma transformação dos valores reais em valores codificados, descontando-se a média para eliminar o efeito de localização (tendência central) e dividindo-se pelo desvio-padrão para eliminar o efeito de escala (variabilidade).

$$Z = \frac{a - \mu}{\sigma} = \frac{LIE - \bar{x}}{S}$$

Calculada a variável Z, consulta-se a tabela Normal padronizada para identificar a probabilidade acumulada à esquerda de Z, (probabilidade de ocorrerem valores menores ou iguais ao Z consultado).

5) Distribuição de probabilidades



$$N \sim (100, 2^2)$$

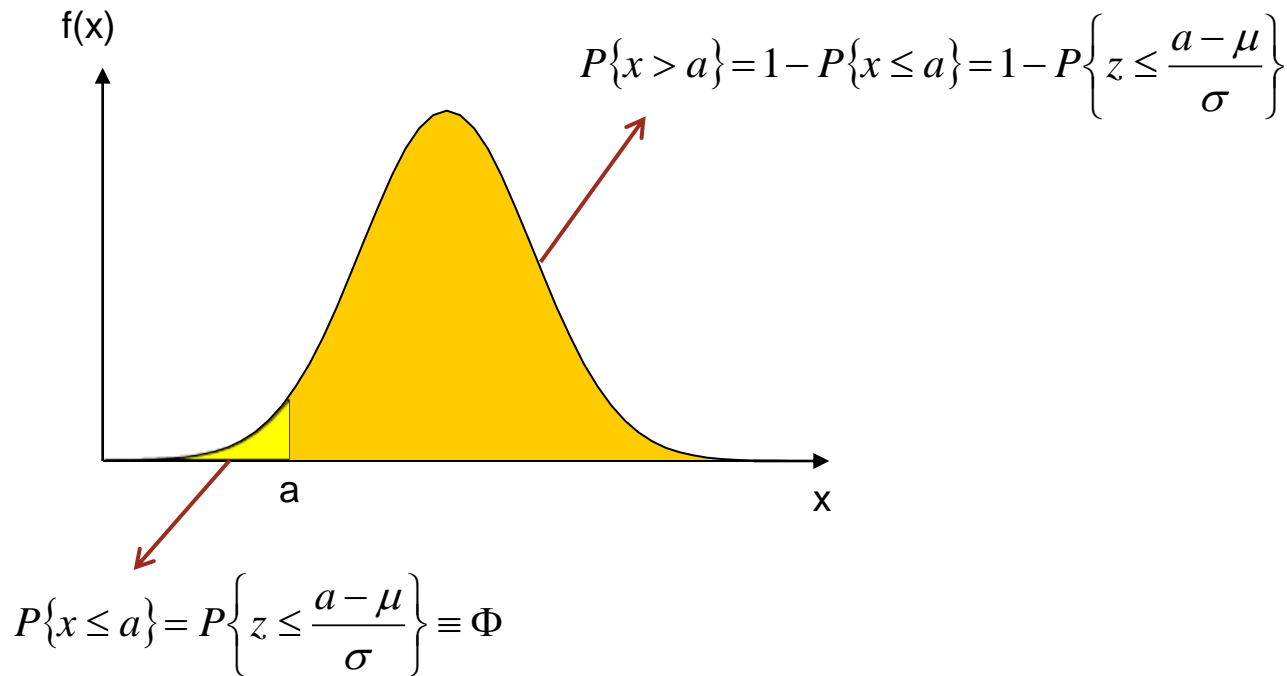
$$z = \frac{a - \mu}{\sigma}$$

$$N \sim (0, 1^2)$$

Transformação da variável x na variável padronizada z (onde as probabilidades estão tabeladas de $-\infty$ até um determinado z que é função de a).

5) Distribuição de probabilidades

Distribuição Normal $P\{x \leq a\} = F(a) = \int_{-\infty}^a \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$



5) Distribuição de probabilidades

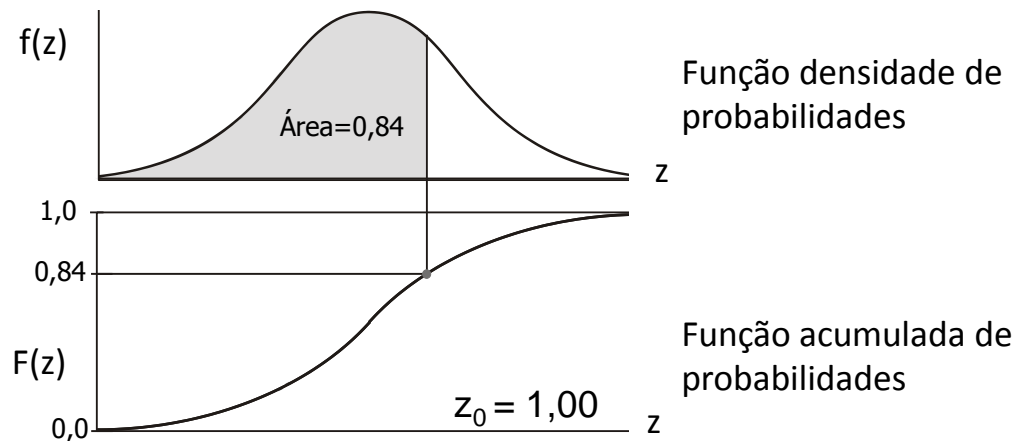
As áreas correspondentes as probabilidades da distribuição normal padrão estão tabeladas.

$$z_0 = 1,00 \rightarrow P(z \leq 1) = 0,8413$$

$$z_0 = 1,16 \rightarrow P(z \leq 1,16) = 0,8770$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147

Probabilidades acumuladas de ocorrência de valores abaixo de Z_0



5) Distribuição de probabilidades

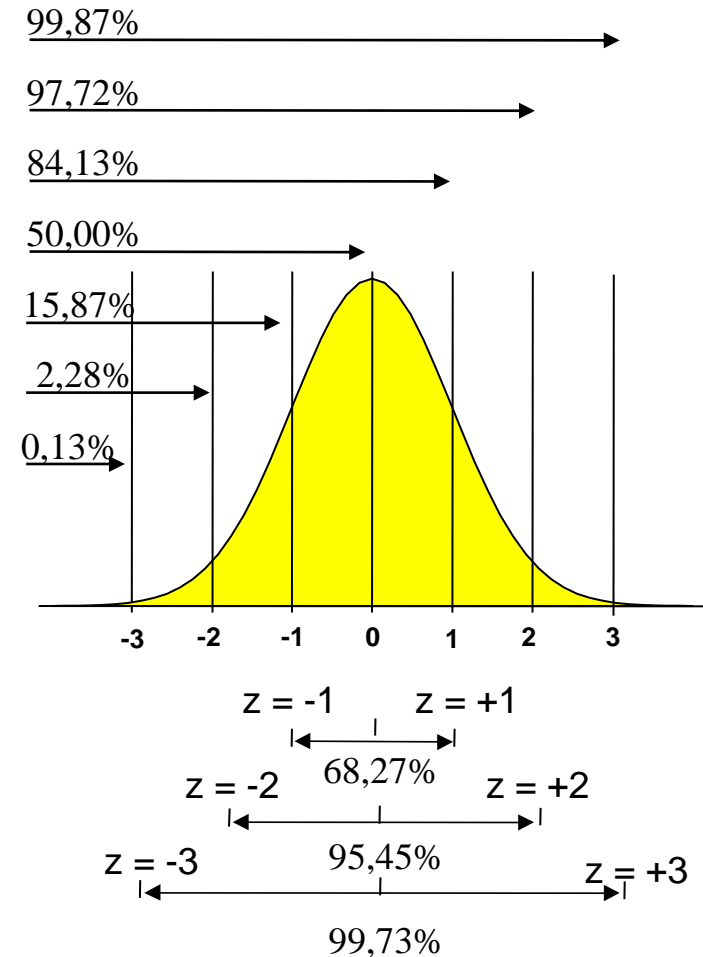
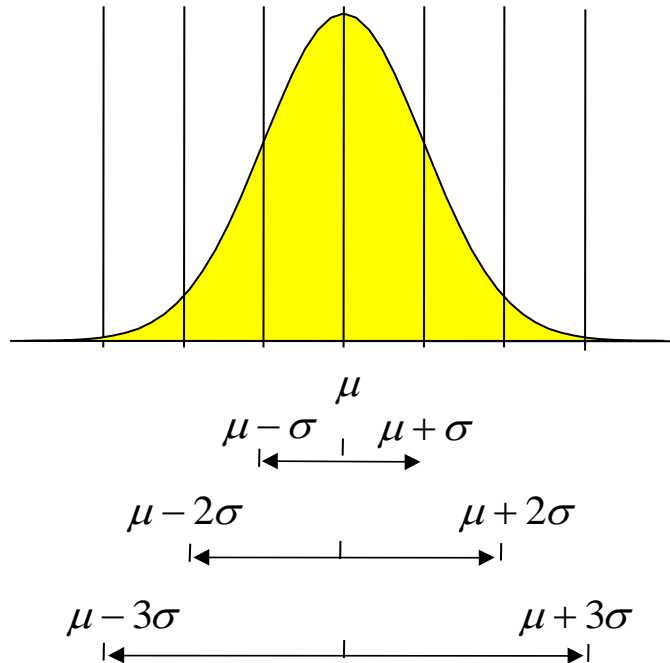
$$N \sim (\mu, \sigma)$$



$$z = \frac{a - \mu}{\sigma}$$



$$N \sim (0, 1^2)$$



5) Distribuição de probabilidades

Exemplo

A força de tensão de sacos plásticos de supermercado é normalmente distribuída com média 40 lb/in² com desvio padrão de 2 lb/in². O comprador exige que os sacos tenham resistência de pelo menos 35 lb/in².

Qual a probabilidade do produto atender a especificação?

$$P\{x > 35\} = 1 - P\{x \leq 35\}$$

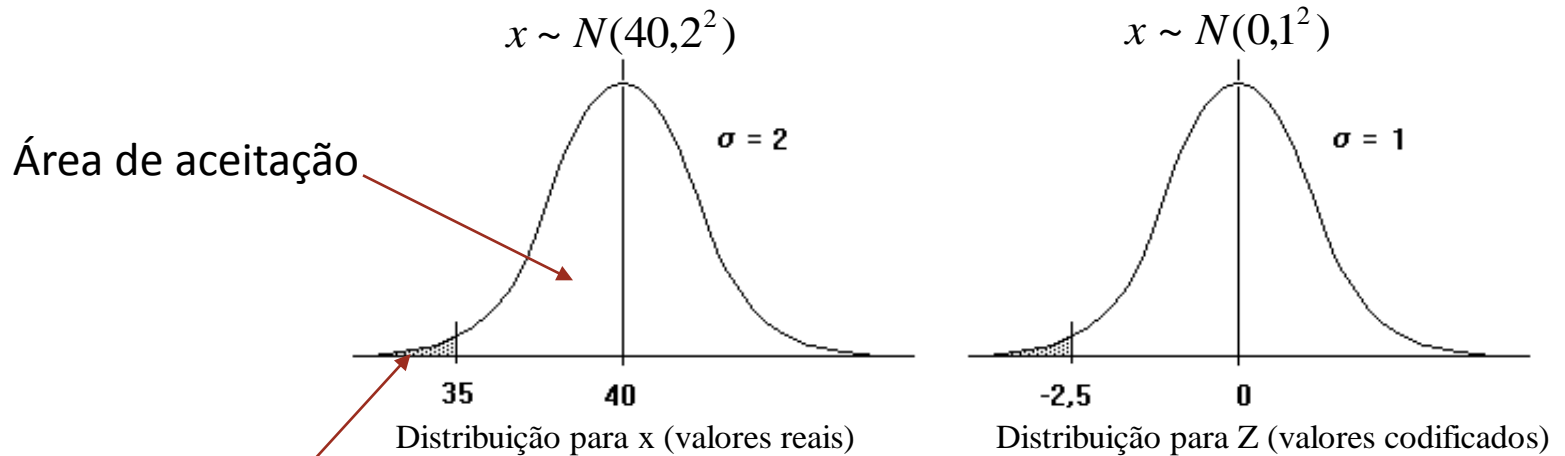
$$P\{x \leq 35\} = P\left\{z \leq \frac{35 - 40}{2}\right\} = P\{z \leq -2,5\} = \Phi(-2,5) = 0,0062$$

$$P\{x > 35\} = 1 - 0,0062 = 0,9938$$

Função no Excel
DIST.NORMP ()

5) Distribuição de probabilidades

Exemplo



Área de rejeição

$$P\{x \leq 35\} = P\{z \leq -2,5\} = \Phi(-2,5) = 0,62\%$$

$$P\{x > 35\} = 1 - P\{x \leq 35\} = 1 - P\{z \leq -2,5\} = 1 - \Phi(-2,5) = 1 - 0,0062 = 99,38\%$$

$$P\{x \leq 35\} + P\{x > 35\} = 0,62\% + 99,38\% = 100\%$$

5) Distribuição de probabilidades

O diâmetro de uma peça segue a distribuição Normal com média 25,08 e desvio padrão 0,05.

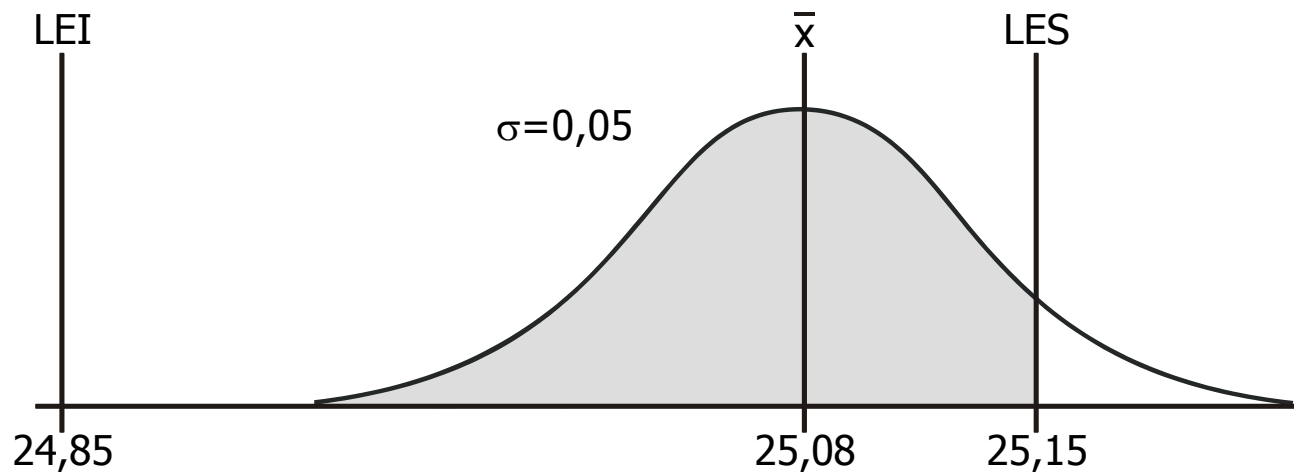
Se as especificações para esse eixo são $25,00 \pm 0,15$, determine o percentual de unidades produzidas em conformidades com as especificações.

$$\begin{aligned} P\{24,85 \leq x \leq 25,15\} &= P\{x \leq 25,15\} - P\{x \leq 24,85\} \\ &= P\left\{Z \leq \frac{25,15 - 25,08}{0,05}\right\} - P\left\{Z \leq \frac{24,85 - 25,08}{0,05}\right\} \\ &= P\{Z \leq 1,40\} - P\{Z \leq -4,60\} = 0,9192 - 0,0000 = 0,9192 \end{aligned}$$

Ou seja, 91,92% dentro das especificações e 8,08% fora das especificações.

5) Distribuição de probabilidades

Ou seja, 91,92% dentro das especificações (área cinza) e 8,08% fora das especificações.



5) Distribuição de probabilidades

No exemplo anterior tem-se cerca de 8% de unidades não-conformes, e essas unidades são invariavelmente do tipo “eixo muito largo”.

Recalcule o percentual de unidades conformes se o processo estivesse centrado em 25,00.

$$P\left\{Z \leq \frac{25,15 - 25,00}{0,05}\right\} - P\left\{Z \leq \frac{24,85 - 25,00}{0,05}\right\}$$

$$P\{Z \leq 3,0\} - P\{Z \leq -3,0\} = 0,9987 - 0,00135 = 0,9973$$

ou seja, 99,73% dentro das especificações e 0,27% fora das especificações.

|| Tópicos próxima aula

- Teorema do limite central
- Introdução ao Controle Estatístico de Processos

Próximas lâminas

- Construção do histograma
- Boxplot no Excel

Construção do histograma

	Intervalos de classe	Frequência absoluta
	12,50 a 13,50	3
	13,51 a 14,50	8
Limite inferior da classe	14,51 a 15,50	15
	15,51 a 16,50	13
	16,51 a 17,50	9
	17,51 a 18,50	2

Limite superior da classe

N observações na classe

Intervalo da classe

12,58	12,97	13,45	13,53	13,59	13,61	13,62	13,78	13,97	14,21
14,47	14,51	14,53	14,58	14,65	14,78	14,83	14,97	15,06	15,13
15,17	15,23	15,29	15,37	15,40	15,45	15,51	15,62	15,67	15,73
15,83	15,98	16,01	16,11	16,17	16,23	16,35	16,43	16,49	16,52
16,67	16,83	16,97	17,05	17,13	17,22	17,3	17,48	17,8	18,47

Construção do histograma

a) Determina-se o maior e menor valor do conjunto de dados;

Para o exemplo, Mín = 12,58 e Máx = 18,47

b) Define-se o limite inferior da primeira classe (LI), que deve ser ligeiramente inferior ao menor valor das observações

Para o exemplo, LI = 12,50

c) Define-se o limite superior da última classe (LS), que deve ligeiramente superior ao maior valor das observações;

Para o exemplo, LS = 18,50

Construção do histograma

d) Define-se o número de classes $K = \sqrt{n}$ e que deve estar compreendido entre 5 a 20.

Para o exemplo, $K = \sqrt{50} \cong 7$. Por praticidade (para que a seja = 1), foi escolhido $K = 6$.

e) Conhecido o número de classes define-se a amplitude de cada classe:

$$a = (LS - LI) / K;$$

Para o exemplo, $a = \frac{(LS - LI)}{K} = \frac{(18,50 - 12,50)}{6} = 1$

Construção do histograma

f) Conhecida a amplitude das classes, define-se os limites inferior e superior para cada classe.

Para a 1° classe: $\text{lim. inf.} = \text{LI}$; $\text{lim. sup.} = \text{LI} + a$;

Para o exemplo, $\text{lim inf} = 12,50$ e $\text{lim sup} = 12,50 + 1 = 13,50$

g) Calcula-se a freqüência de cada classe, ou seja, o número de observações pertencentes a cada classe, e completa-se a tabela de freqüência;

Para o exemplo, o número de observações pertencentes ao intervalo 12,50 a 13,50 é 3.

OBS: Use a função CONT.SE() para contar o número de observações

Na tabela ao lado, se $\text{LI}=24$ e $\text{LS}=33$, então na coluna A, $[24 < x < 33] = ?$

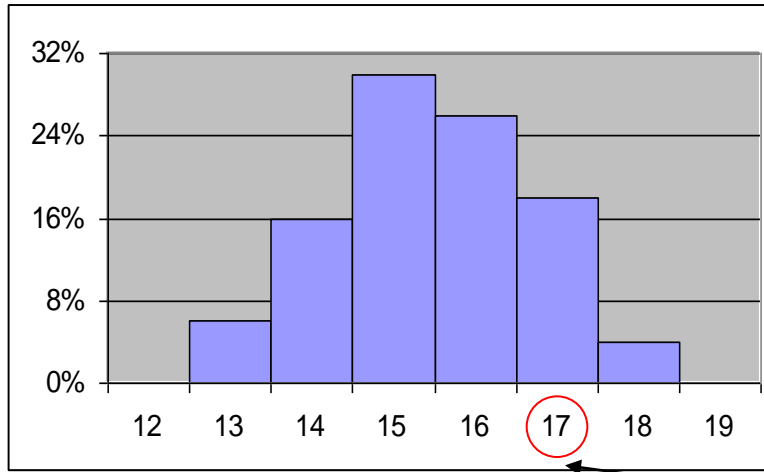
$= \text{CONT.SE}(\$A\$1:\$A\$4;">"\&C1) - \text{CONT.SE}(\$A\$1:\$A\$4;">"\&C2) = 2$

	A	B	C
1	25	LI	24
2	30	LS	33
3	35		
4	40		

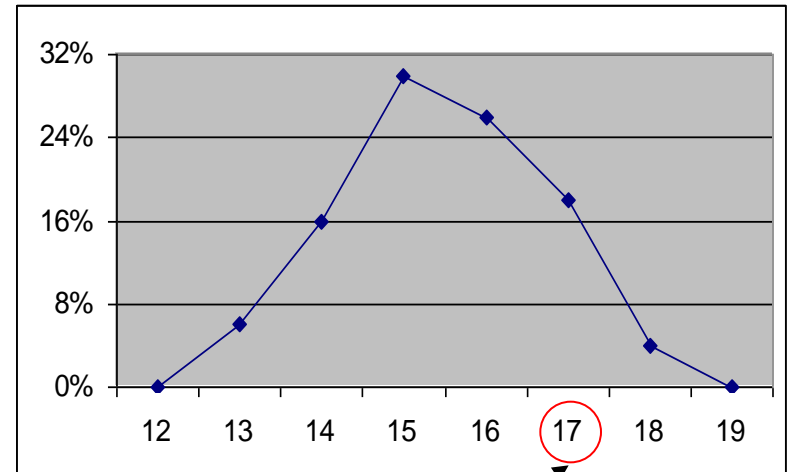


Construção do histograma

Histograma



Polígono de Frequência

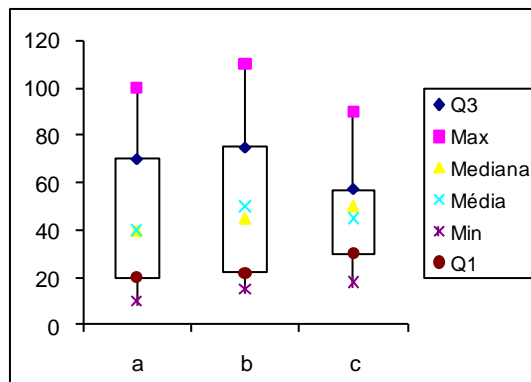


Abscissa: valor médio da classe

Boxplot no Excel

2003

- A ordem deve ser Q3, Max, Mediana, Média, Min, Q1
- Selecione todo o conjunto de dados
- Selecione Inserir Gráfico, tipo Linha com marcadores exibidos a cada valor de dado, clique Avançar
- Selecione Séries em Linha, selecione Concluir
- Selecione no gráfico uma série de dados, com o botão direito selecione Formatar Série de Dados
- Selecione a aba Padrões, na opção Linha selecione Nenhuma, repita o procedimento para as demais séries
- Selecione um dado e com botão direito selecione Formatar Série de Dados, selecione a aba Opções, selecione Linhas max/min e Barras superiores/inferiores



2007

- A ordem deve ser Q3, Max, Mediana, Média, Min, Q1
- Selecione todo o conjunto de dados
- Selecione Inserir Gráfico, tipo Linha com marcadores
- Selecione uma seqüência, clique com botão direito, Selecionar Dados, Alternar entre Linha/Coluna, OK
- Selecione no gráfico uma série de dados, com o botão direito selecione Formatar Série de Dados, Cor de Linha, Sem Linha, Fechar
- Repita este procedimento com todas as seqüências de dados
- Selecione um dado, na barra de ferramentas selecione Layout, em Análise selecione Linhas, Linhas de Máximo e Mínimo e em Barras Superiores e Inferiores selecione Barras Superiores e Inferiores

	a	b	C
Q3	70	75	57
Max	100	110	90
Mediana	40	45	50
Média	40	40	50
Min	10	15	18
Q1	20	22	30