EXERCÍCIO PARA CASA

Um pesquisador realizou um levantamento obtendo 84 observações sobre uma grandeza com média 65,00. Após, verificou que duas destas observações, com medidas 95,52 e 105,82, estavam comprometidas. Deseja-se eliminar estas duas medidas e calcular o valor da média.

Sabendo-se que a variância de todas as observações é 41,590, qual vai ser o valor da nova variância (conjunto de observações sem as medidas comprometidas)?

EXERCÍCIO PARA CASA

$$N_T = 84$$

$$\mu_T = \frac{\sum N_T}{N_T} = 65{,}00$$

$$\sum N_T = \mu_T * N_T = 65{,}00 * 84 = 5460$$

$$\sum N_T = \sum N_P + 95{,}52 + 105{,}82$$

$$\sum N_P = \sum N_T - 95{,}52 - 105{,}82$$

$$\sum N_P = 5460 - 95{,}52 - 105{,}82 = 5258{,}66$$

$$N_P = 82$$

$$\mu_P = \frac{\sum N_P}{N_P} = \frac{5258{,}66}{82}$$

$$\boldsymbol{\mu_P = 64,13}$$

EXERCÍCIO PARA CASA

$$\sigma_T^2 = \frac{\sum(X_T - \mu_T)^2}{N_T - 1} = 41{,}590$$

$$\mu_T = \frac{\sum X_T}{N_T}$$

$$\sum(X_T - \mu_T)^2 = (X_{T1} - \mu_{T1})^2 + (X_{T2} - \mu_{T2})^2 + \ldots + (X_{Tn} - \mu_{Tn})^2$$

$$\sigma_T^2 = \frac{\sum X_T^2 - 2\mu_T \sum X_T + \sum \mu_T^2}{N_T - 1} = \frac{\sum X_T^2 - 2\mu_T(\mu_T N_T) + (\mu_T^2 N_T)}{N_T - 1} = \frac{\sum X_T^2 - \mu_T^2 N_T}{N_T - 1}$$

$$\sum X_T^2 = [\sigma_T^2 * (N_T - 1)] + \mu_T^2 N_T = 41{,}590 * 83 + 65{,}00^2 * 84 = 358351{,}97$$

$$\sum X_P^2 = \sum X_T^2 - 95{,}52^2 - 105{,}82^2 = 338030{,}0272$$

$$\sigma_P^2 = \frac{\sum X_P^2 - \mu_P^2 N_P}{N_P - 1} = \frac{338030{,}0272 - 64{,}13^2 * 82}{82 - 1} = \mathbf{9{,}780}$$

EXERCÍCIO PARA CASA

Em uma instituição bancária, o salário médio dos 100 empregados do sexo masculino é de R$ 1.500,00, com desvio padrão de R$100,00. O salário médio dos 150 empregados do sexo feminino é de R$ 1.000,00, com desvio padrão de R$200,00. A variância em (R$) dos dois grupos reunidos é de: (BACEN, 2005)

a) 25.600,00; b) 28.000,00; c) 50.000,00; d) 62.500,00; e) 88.000,00

EXERCÍCIO PARA CASA

Observe-se que se está falando de população e não amostra. Segue-se que:

$$\sigma^2 = \frac{\sum(X-\mu)^2}{N} \qquad\qquad \mu = \frac{\sum X}{N}$$

$$\sigma^2 = \frac{\sum X^2 - 2\mu \sum X + \sum \mu^2}{N} = \frac{\sum X^2 - 2\mu (\mu N) + (\mu^2 N)}{N} = \frac{\sum X^2 - \mu^2 N}{N}$$

Ou ainda: $\sum X^2 = N * (\sigma^2 + \mu^2)$

Tem-se:

$N_M = 100$	$N_F = 150$
$\mu_M = 1500{,}00$	$\mu_F = 1000{,}00$
$\sigma_M = 100$	$\sigma_F = 200$

EXERCÍCIO PARA CASA

Calcula-se: $\sum X_M^2 = N_M * (\sigma_M^2 + \mu_M^2) = 226000000$

$\sum X_F^2 = N_F * (\sigma_F^2 + \mu_F^2) = 156000000$

Para o conjunto total da população tem-se:

$\sum X^2 = \sum X_M^2 + \sum X_F^2 = N * (\sigma^2 + \mu^2)$ onde: $N = N_M + N_F$

A média da população total pode ser calculada:

$\mu = \frac{\sum X}{N} \Rightarrow \sum X = N * \mu \Rightarrow \sum X_M + \sum X_F = (N_M + N_F) * \mu$

$\mu_M * N_M + \mu_F * N_F = (N_M + N_F) * \mu$

$$\mu = \frac{\mu_M * N_M + \mu_F * N_F}{(N_M + N_F)} = 1200$$

EXERCÍCIO PARA CASA

$$(N_M + N_F) * (\sigma^2 + \mu^2) = N_M * (\sigma_M^2 + \mu_M^2) + N_F * (\sigma_F^2 + \mu_F^2)$$

onde: $$\sigma^2 = \frac{N_M * (\sigma_M^2 + \mu_M^2) + N_F * (\sigma_F^2 + \mu_F^2) - \mu^2(N_M + N_F)}{(N_M + N_F)}$$

$$\sigma^2 = \frac{226000000 + 156000000 - [1200^2 * (250)]}{250}$$

$$\sigma^2 = 88000$$

opção: e

UNIVERSIDADE FEDERAL DO PARANÁ

SETOR DE CIÊNCIAS DA TERRA

DEPARTAMENTO DE GEOMÁTICA

AJUSTAMENTO 1 – GA106 A ENGENHARIA CARTOGRÁFICA E DE AGRIMENSURA

Prof. Dr. Mario Ernesto Jijón Palma
Departamento de Geomática
Setor de Ciências da Terra
Universidade Federal do Paraná - UFPR

INTRODUÇÃO

Seja uma população finita de 100 elementos, que apresenta média μ e variância σ^2.

O que acontece ao se retirar um conjunto amostral de 30 elementos?

- Valores da média $\bar{x}$ e variância s^2 amostral **diferentes** da população.
- Porém estes são os *estimadores* → estimam a população

Quando :

Tamanho da amostra → tamanho da população

$$\bar{x} \rightarrow \mu$$

$$s^2 \rightarrow \sigma^2$$

INTRODUÇÃO

Porém como as amostras são tomadas de forma aleatória, **então média e variância variam de amostra a amostra.**

Exemplo:

População

- 100 elementos $\mu = 26{,}1$ $\sigma^2 = 17{,}5$

Amostra 1

- 30 elementos $\bar{x} = 26{,}9$ $s^2 = 28{,}1$

Amostra 2

- 30 elementos $\bar{x} = 27{,}9$ $s^2 = 2{,}9$

Amostra 3

- 30 elementos $\bar{x} = 27{,}2$ $s^2 = 23{,}0$

Qual destas três amostras é mais confiável? Qual é mais precisa? Qual é mais acurada?

INTRODUÇÃO

Se analisar a média:
Amostra 1
Porém é a que possui maior variância (maior dispersão)

Exemplo:

População

- 100 elementos $\mu = 26{,}1$ $\sigma^2 = 17{,}5$

Amostra 1

- 30 elementos $\bar{x} = 26{,}9$ $s^2 = 28{,}1$

Amostra 2

- 30 elementos $\bar{x} = 27{,}9$ $s^2 = 2{,}9$

Amostra 3

- 30 elementos $\bar{x} = 27{,}2$ $s^2 = 23{,}0$

Qual destas três amostras é mais confiável?
Qual é mais precisa? Qual é mais acurada?

INTRODUÇÃO

Se analisar a precisão:
Amostra 2
Maior confiança (menor dispersão)

Exemplo:

População
- 100 elementos $\mu = 26{,}1$ $\sigma^2 = 17{,}5$

Amostra 1
- 30 elementos $\bar{x} = 26{,}9$ $s^2 = 28{,}1$

Amostra 2
- 30 elementos $\bar{x} = 27{,}9$ $s^2 = 2{,}9$

Amostra 3
- 30 elementos $\bar{x} = 27{,}2$ $s^2 = 23{,}0$

Qual destas três amostras é mais confiável?
Qual é mais precisa? Qual é mais acurada?

INTRODUÇÃO

A solução é usar funções de distribuição de probabilidade para avaliar os estimadores (média e variância) dada um determinado nível de significância (α).

Ou seja, busca-se a resposta para:
quão bom são os estimadores?

As distribuições podem ser:

- Distribuição de probabilidade de uma **v.a. discreta**
- Distribuição de probabilidade de uma **v.a. contínua**
- Distribuição de probabilidade **acumulada,** para os casos de v.a. contínua

Distribuição de probabilidade de uma v.a. discreta

Seja X uma v.a. discreta, isto é, que assume valores em associação com números inteiros $x_1, x_2, \ldots, x_n$.

Associemos a cada x_i um número $p(x_i)$ representativo da sua probabilidade.

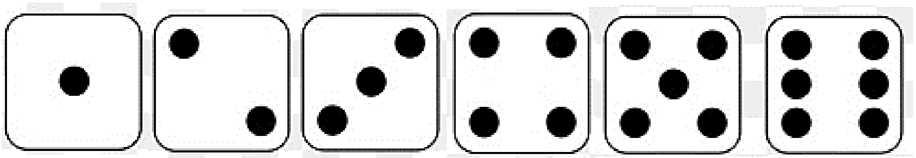
$$\boldsymbol{p(x_i) = P(X = x_i)}$$

Tal que:

- $0 \leq p(x_i) \leq 1$
- $\sum_{i=0}^{n} x_i = 1$
- $P(a \leq X \leq b) = \sum x_i \qquad \forall \; i \; tal \; que \; a \leq X \leq b$

Distribuição de probabilidade de uma v.a. discreta

$$p(x_i) = P(X = x_i)$$

X = 

P(X = 1) = 1/6

P(X = 2) = 1/6

P(X = 3) = 1/6

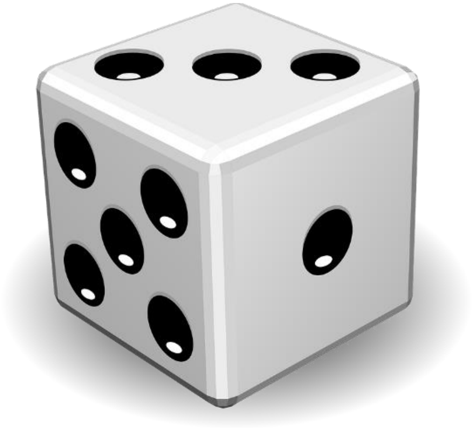
P(X = 4) = 1/6

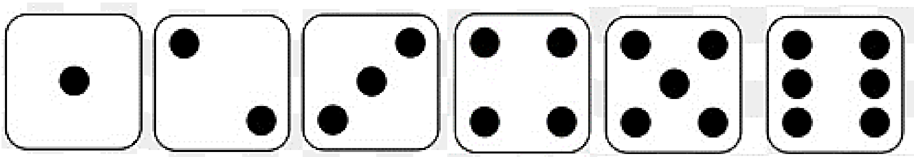
P(X = 5) = 1/6

P(X = 6) = 1/6

Distribuição de probabilidade de uma v.a. discreta

$$p(x_i) = P(X = x_i)$$



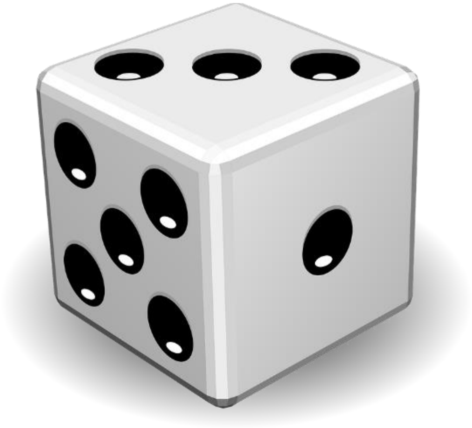
X = 

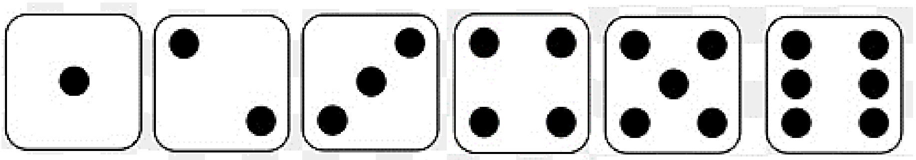
P(X = 3) = 1/6

- $0 \leq p(x_i) \leq 1$ → $0 \leq \frac{1}{6} \leq 1$
- $\sum_{i=0}^{n} x_i = 1$ → $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$

Distribuição de probabilidade de uma v.a. discreta

$$p(x_i) = P(X = x_i)$$



X = 

Probabilidade de obter um número par

- $P(a \leq X \leq b) = \sum x_i$

- $P(2 \leq X \leq 6) = \frac{3}{6} = \frac{1}{2}$

Distribuição de probabilidade de uma v.a. contínua

Seja X uma v.a. contínua.

A probabilidade "pontual" associada à variável discreta é substituída pela densidade de probabilidade $\varphi(x)$ relativa a um **intervalo infinitésimo**.

$$\boldsymbol{\varphi(x)dx = P(x \leq X \leq x + dx)}$$

ou

$$\varphi(x) = \lim \frac{P(x \leq X \leq x + dx)}{\Delta x}$$

Tal que:

- $\varphi(x) \geq 0$
- $\int_{-\infty}^{+\infty} \varphi(x)\, dx = 1$
- $P(a \leq X \leq b) = \int_a^b \varphi(x)\, dx$

Distribuição de probabilidade de uma v.a. contínua

Seja X uma v.a. contínua $\Rightarrow$ densidade de probabilidade $\varphi(x)$

$$\boldsymbol{\varphi(x)dx = P(x \leq X \leq x + dx)}$$

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

$$\mu = 170\ cm \qquad \sigma = 10\ cm$$

- $\varphi(x) \geq 0$

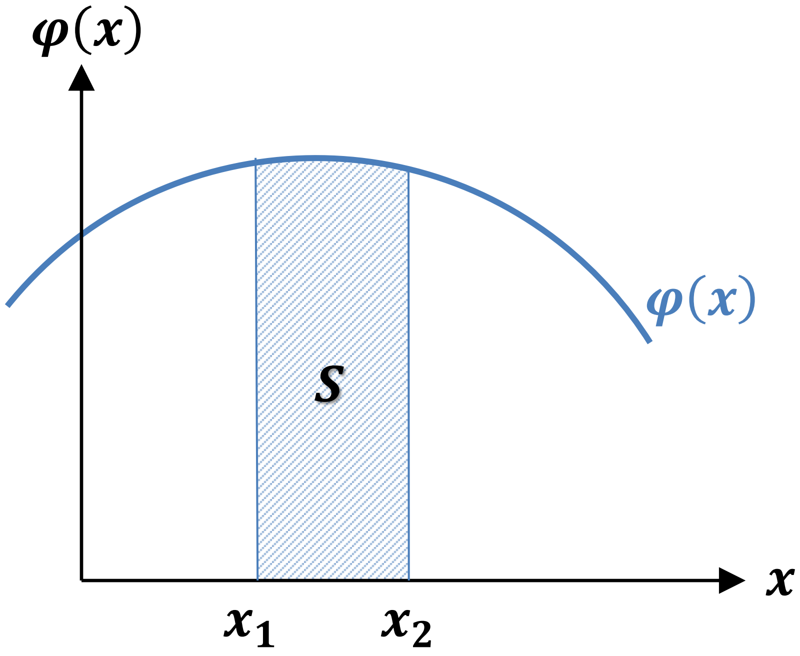
$$\varphi(160) = \frac{1}{10\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{160-170}{10})^2}$$

Distribuição de probabilidade de uma v.a. contínua

A função $\varphi(x)$ que estabelece a correspondência entre um valor da v.a. contido no intervalo elementar e a densidade de probabilidade é denominada **função densidade de probabilidade (fdp)**.

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} \varphi(x)\, dx = \text{área } S$$

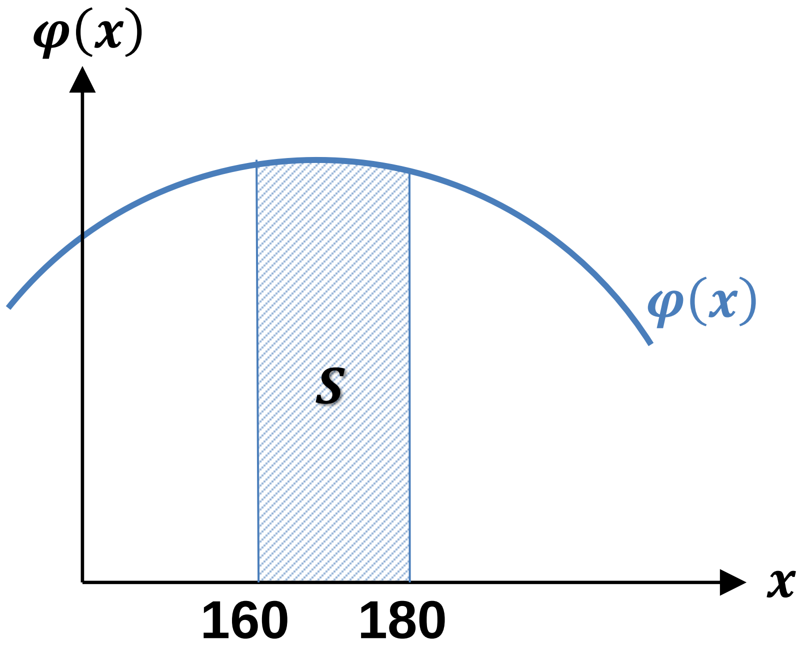
- $\int_{-\infty}^{+\infty} \varphi(x)\, dx = 1$



Distribuição de probabilidade de uma v.a. contínua

- $P(a \leq X \leq b) = \int_a^b \varphi(x)\, dx$

- $P(160 \leq X \leq 180) = \int_{160}^{180} \varphi(x)\, dx$ **≈ 0.6826**



Função de distribuição de probabilidade acumulada

Seja $\varphi(x)$ a função de **distribuição acumulada (fda)** de uma v.a. X, tal que:

$$\varphi(x) = P(X \leq x) = \int_{-\infty}^{x} \varphi(u)\, du$$

contínuo

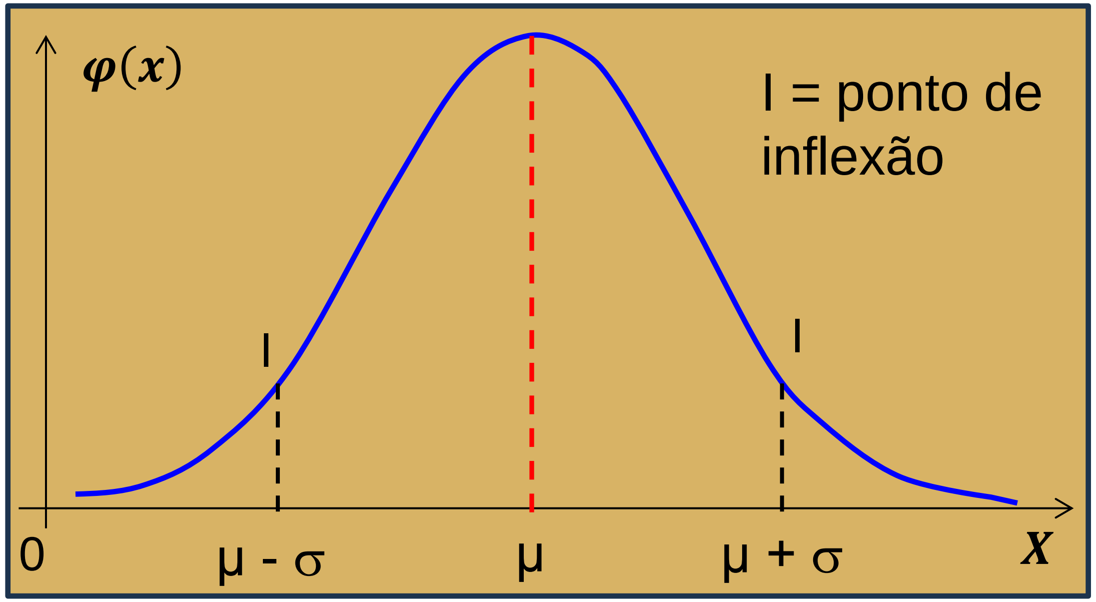
$$P(X \leq x) = \sum_{x_i \leq x} p(x)$$

discreto

Distribuição Normal

- ✓ A **distribuição normal** é uma das mais importantes da estatística e aparece frequentemente em nas aplicações tecnológicas e científicas.
- ✓ Ela tem a famosa forma de "**curva em sino**" e é descrita por dois parâmetros:
  - •**Média (μ)**: indica o centro da distribuição.
  - •**Desvio padrão (σ)**: indica a dispersão dos dados.

N(**μ**, **σ²**)



Exemplos:

Altura de pessoas

Notas de alunos em uma prova

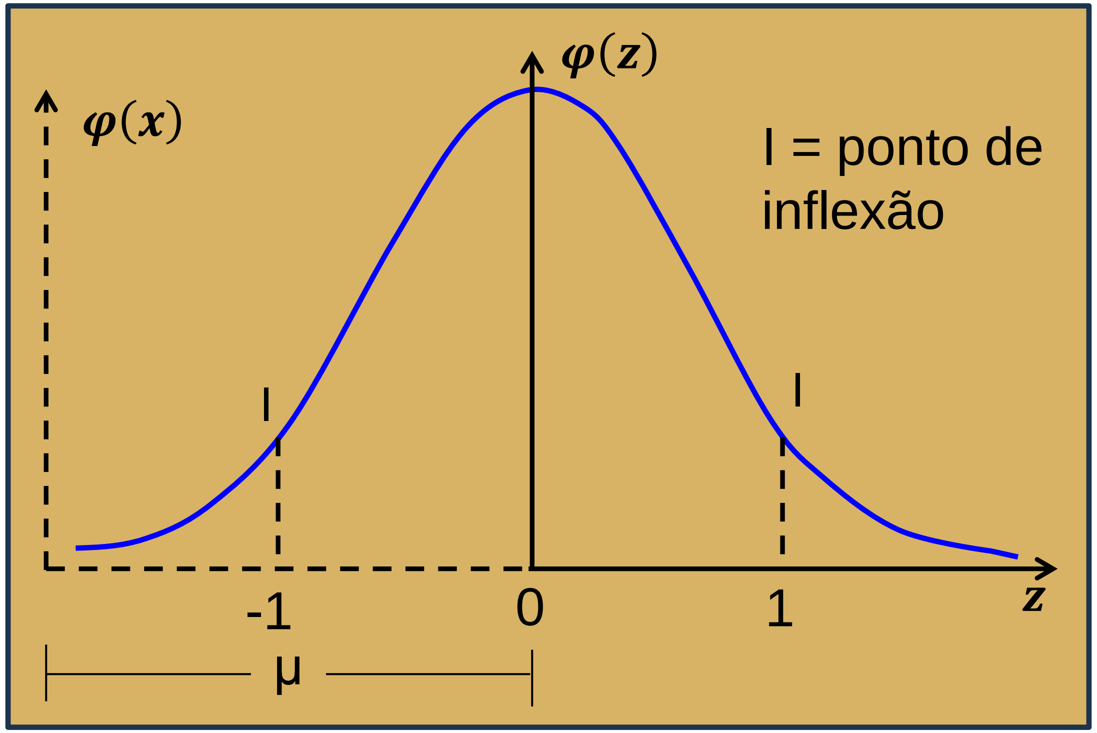
Erros de medição em experimentos

Distribuição Normal Reduzida

✓ A **distribuição normal reduzida** é uma versão padronizada da distribuição normal e segue a distribuição N(0,1)

$$\boldsymbol{\mu = 0}$$
$$\boldsymbol{\sigma = 1}$$

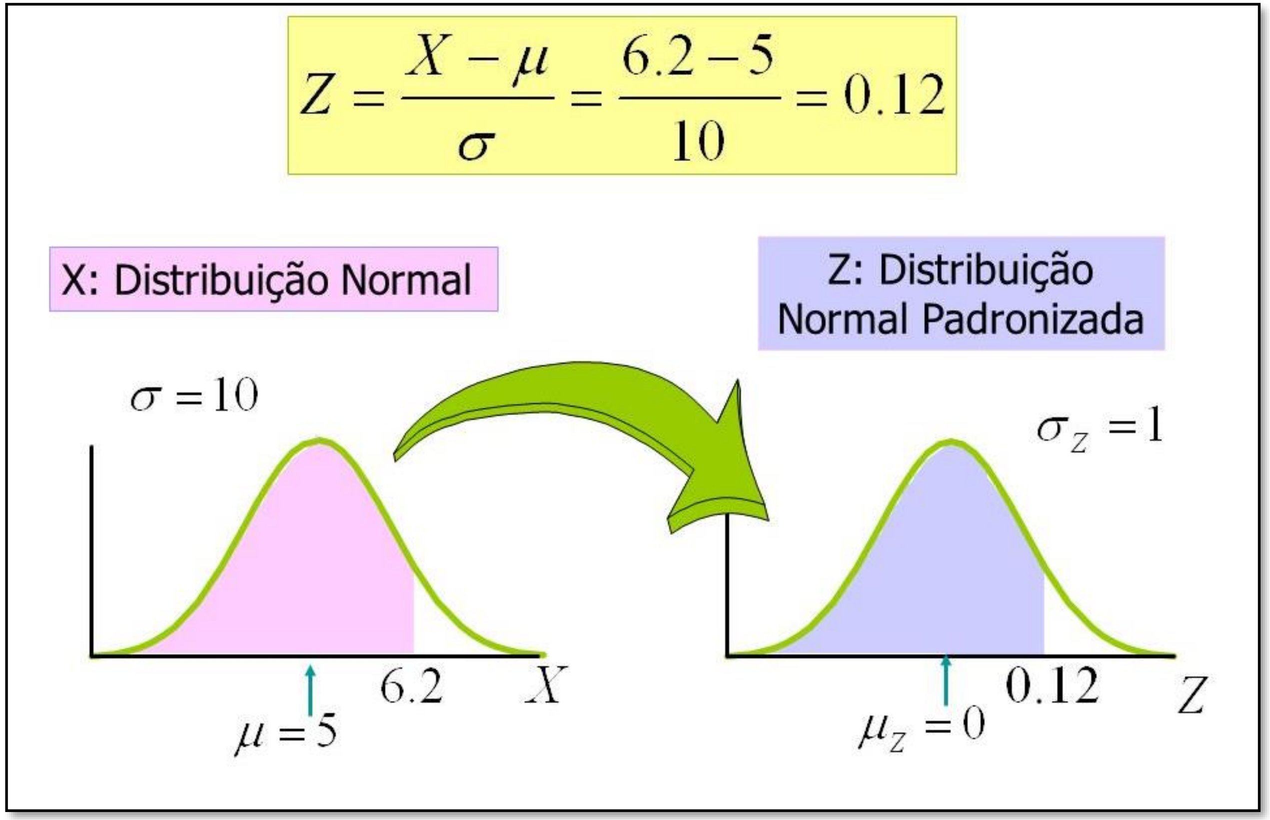
$$z = \frac{x - \mu}{\sigma}$$



Distribuição Normal Reduzida / Distribuição normal padrão

$$\mu = 0$$
$$\sigma = 1$$

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



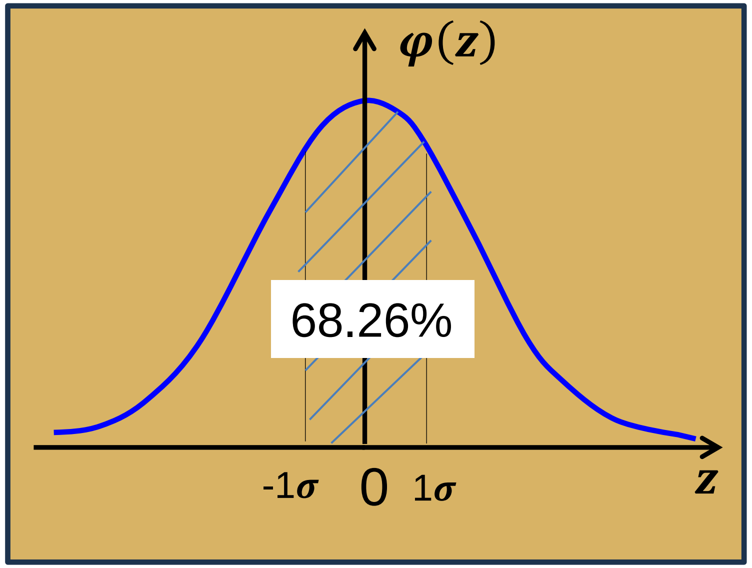
Adaptado Profª. Érica Santos Matos Baluta

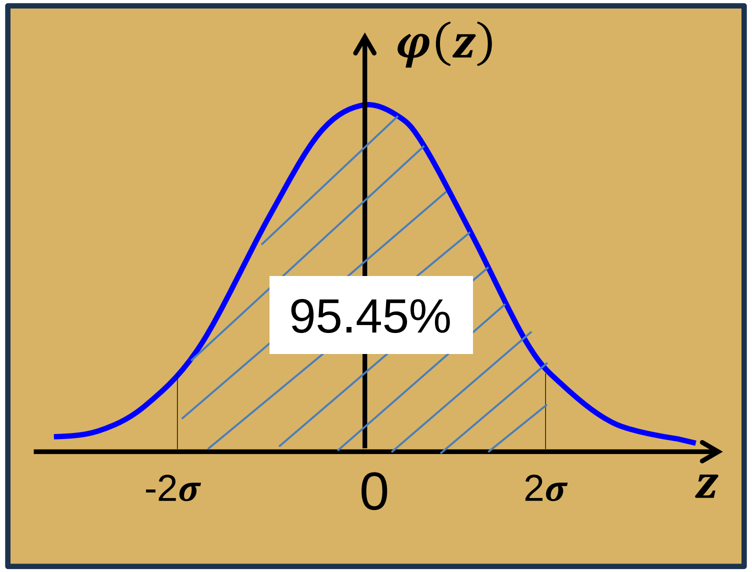
Distribuição Normal Reduzida

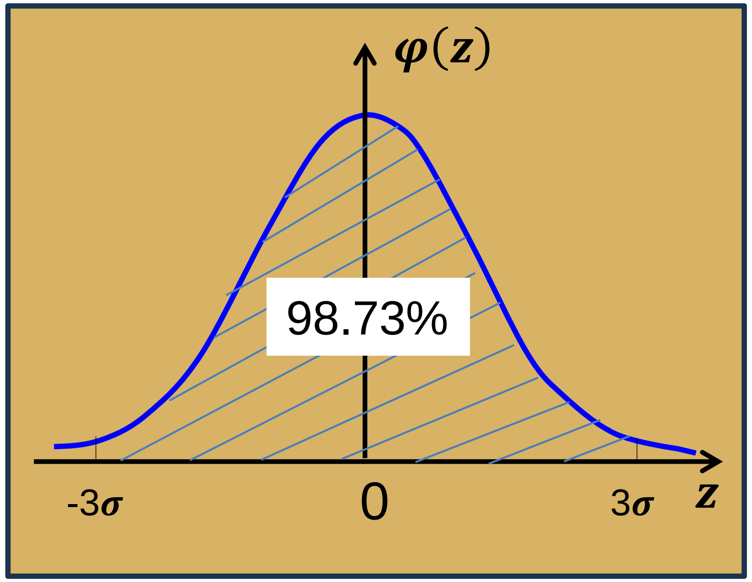
✓ Áreas sob a curva normal reduzida

$$\boldsymbol{\mu = 0}$$
$$\boldsymbol{\sigma = 1}$$

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







Exercício

A média dos diâmetros internos de uma amostra de 200 arruelas produzidas por uma certa máquina é **0,502 cm** e o desvio padrão é **0,005 cm**.

A finalidade para a qual essas arruelas são fabricadas permite a tolerância máxima, para o diâmetro de **0,496 cm** a **0,508 cm**; se isso não se verificar, as arruelas serão consideradas defeituosas.

Determinar a percentagem de arruelas defeituosas produzidas pela máquina, admitindo-se que os diâmetros **são distribuídos normalmente**.

Exercício

A média dos diâmetros internos de uma amostra de 200 arruelas produzidas por uma certa máquina é **0,502 cm** e o desvio padrão é **0,005 cm**.

A finalidade para a qual essas arruelas são fabricadas permite a tolerância máxima, para o diâmetro de **0,496 cm** a **0,508 cm**; se isso não se verificar, as arruelas serão consideradas defeituosas.

Determinar a percentagem de arruelas defeituosas produzidas pela máquina, admitindo-se que os diâmetros **são distribuídos normalmente**.

Dados

N = 200

μ = 0,502 cm

σ = 0,005 cm

$\mathbf{x_1 = 0{,}496\ cm}$

$\mathbf{x_2 = 0{,}508\ cm}$

Valores limítrofes aceitáveis
Arruelas não defeituosas

Exercício

N = 200
μ = 0,502 cm
σ = 0,005 cm

Qual o significado?

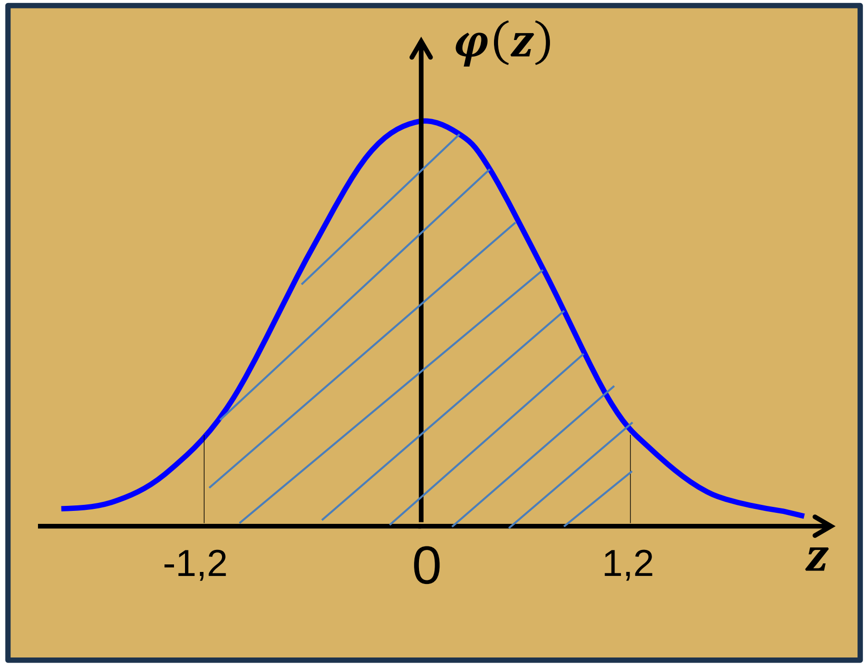
$$\mathbf{x_1} = \mathbf{0,496\ cm}$$
$$\mathbf{x_2} = \mathbf{0,508\ cm}$$

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

z_1
z_2

$$z_1 = \frac{(0{,}496 - 0{,}502)}{0{,}005} = -1{,}2$$

$$z_2 = \frac{(0{,}508 - 0{,}502)}{0{,}005} = +1{,}2$$



Exercício

Qual o significado?

Proporção de arruelas não defeituosas → área destacada na curva normal

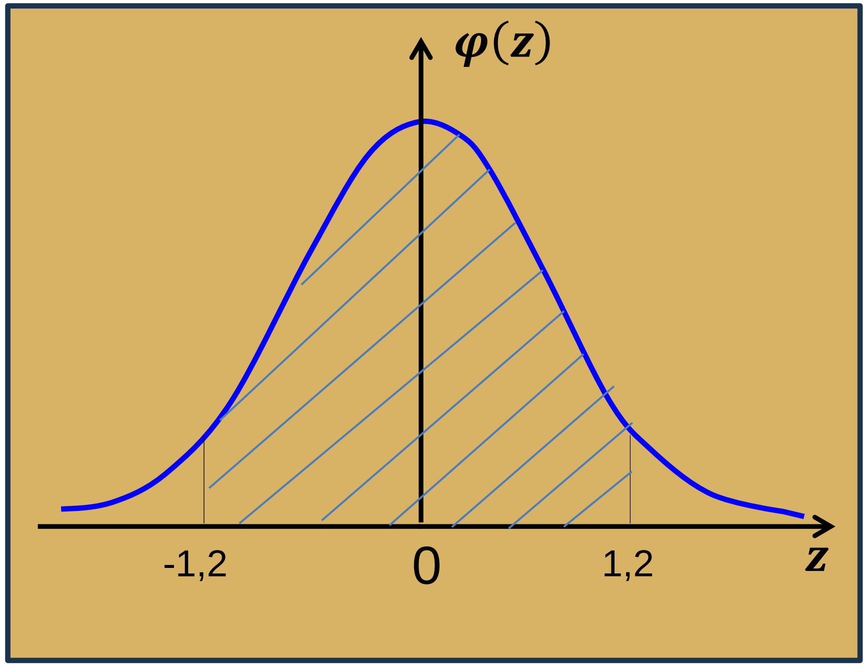
- ✓ Área compreendida entre $z = -1{,}2$ e $z = +1{,}2$
- ✓ Pela simetria → 2 vezes a área entre $z = 0$ e $z = +1{,}2$

Da tabela da distribuição normal padrão, extrai-se:

área destacada na curva normal quando z = 1,2

Distribuição Normal Padrão

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	$\boxed{0{,}3849}$	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]



Exercício

Qual o significado?

Da tabela da distribuição normal padrão, extrai-se o valor 0,3849:

$$2 \ * (0{,}3849) = \mathbf{0,7698} = \mathbf{76,98\%}$$

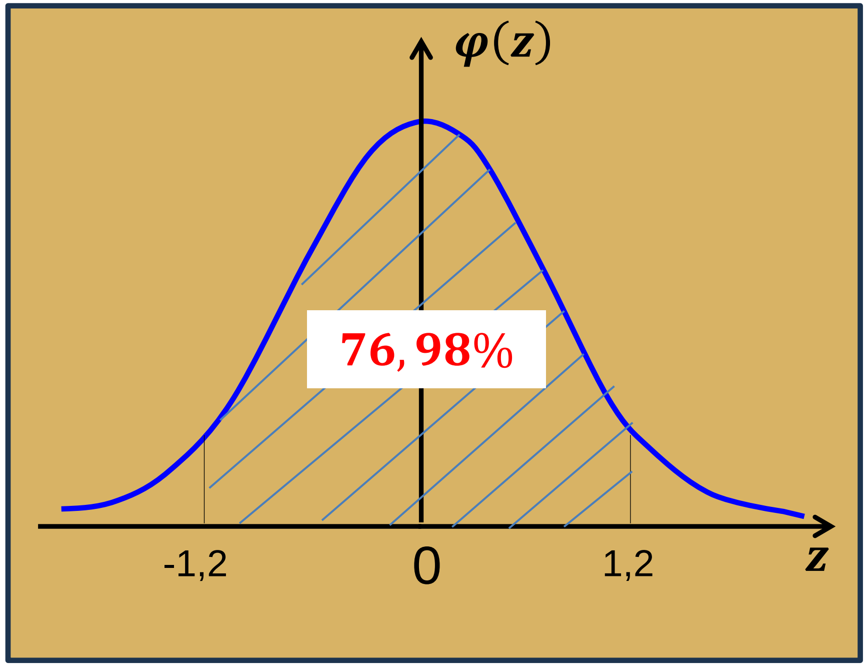
% não defeituosas

Então, a porcentagem de arruelas defeituosas:

$$= \mathbf{100\% - 76,98\%}$$

$$= \mathbf{23,02\%}$$

% defeituosas



Exercício

O peso médio de 500 estudantes do sexo masculino, de uma determinada universidade, é 75,50 kg e o desvio padrão é 7,50 kg. Admitindo-se que os pesos estão distribuídos normalmente, determinar quantos estudantes pesam:

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

a) Entre 59,75 kg e 77,75 kg;

b) Mais do que 92,75 kg;

c) Menos do que 63,75 kg;

d) Entre 63,75kg e 64,25kg;

e) 64,25 kg ou menos;

Exercício

Uma variável aleatória contínua X, que pode assumir somente valores compreendidos entre 2 e 8, inclusive, tem uma função de densidade de probabilidade dada por $a * (X + 3)$, em que a é uma constante.

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} \varphi(x)\, dx$$

a) Calcular o valor de a;

b) Determinar $P(3 < X < 5)$

c) Determinar $P(X \geq 4)$

d) Determinar $P(|X - 5| < 0{,}5)$