

Estatística e Informática

Aula 08 - Variáveis Aleatórias Contínuas

Alan Rodrigo Panosso alan.panosso@unesp.br

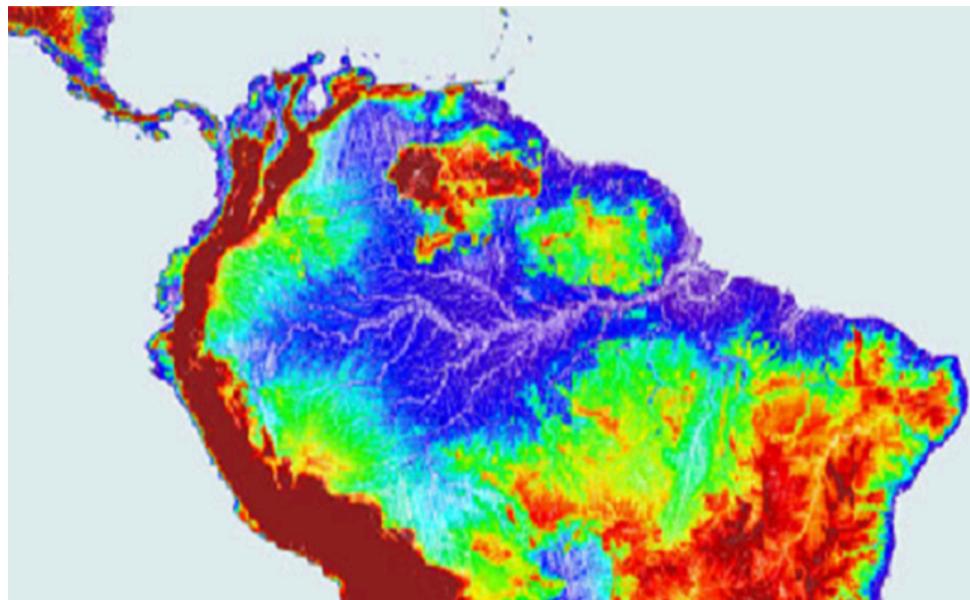
Departamento de Engenharia e Ciências Exatas FCAV/UNESP

(25-04-2024)

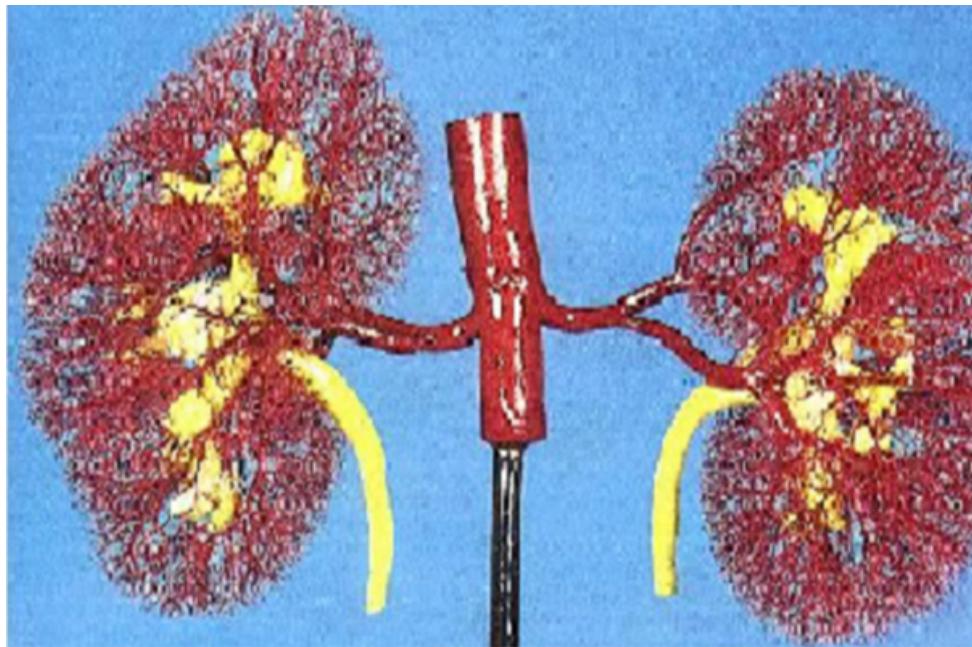
Distribuições teóricas de probabilidades de Variáveis Aleatórias Contínuas

Seríamos capazes de medir, com precisão:

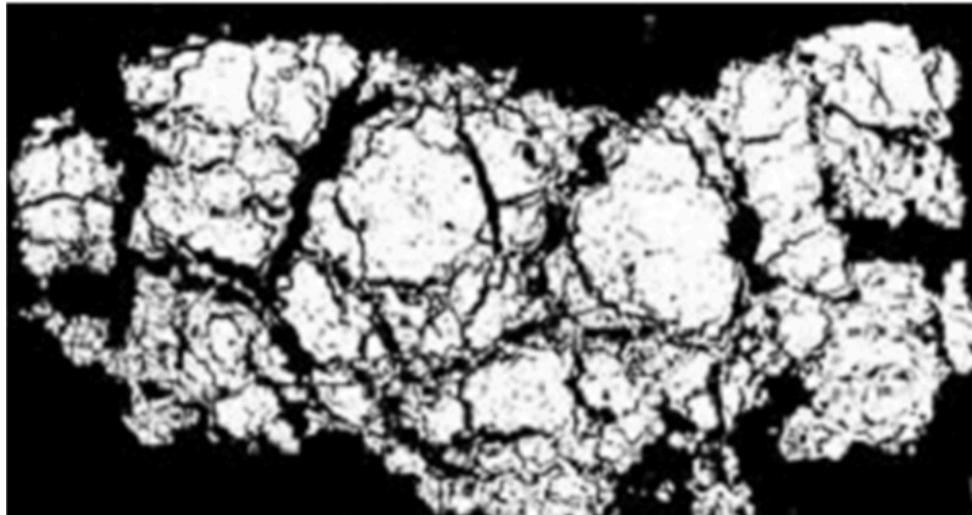
Extensão total da bacia hidrográfica do rio Amazonas?



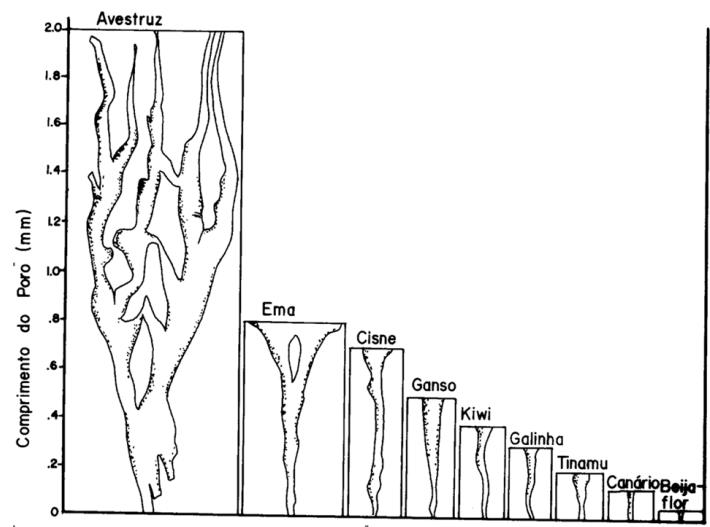
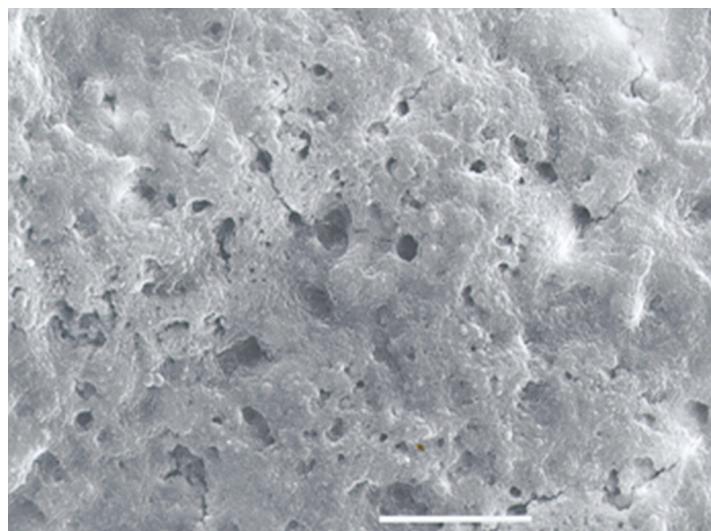
Sistema nefrágico, seríamos capazes de medir com exatidão o volume total desse sistema?



Superfície específica de um agregado do solo?



O comprimento e volume total de poros na casca de um ovo de galinha?



Variável Quantitativa

é aquela que apresenta como possíveis realizações (valores) números resultantes de uma contagem ou mensuração, podendo ser:

Contínua

- Os possíveis valores formam um intervalo de números reais e que resultam, normalmente, de uma mensuração. Exemplos: peso, altura, produção de leite, pressão arterial, teor de nitrogênio no solo ou na planta.

Essencialmente, dizemos que X é uma variável aleatória contínua, se X puder tomar todos os valores em algum intervalo:

$$a \leq x \leq b$$

onde a e b podem ser $-\infty$ e $+\infty$, respectivamente.

Variável Aleatória Contínua

Sendo resultado de um processo de mensuração, o seu valor pode ser pensado como pertencendo a um intervalo ao redor do valor efetivamente observado. Por exemplo:



Quando dizemos que a altura de uma criança é 125, estamos medindo sua altura usando como unidade de medida o *cm*, e portanto o valor observado é, na realidade, um valor entre 124,5 cm e 125,5 cm, por exemplo.

Um vídeo bastante interessante do **Khan Academy Brasil** apresentando a diferença entre as variáveis aleatórias discrete e contínuas, segue abaixo:

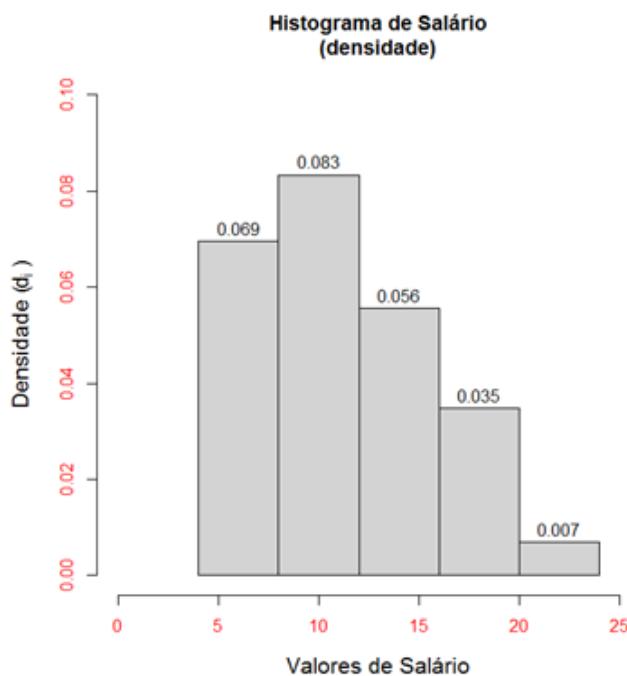


[Link do vídeo](#)

Narrador - Ciência Todo Dia

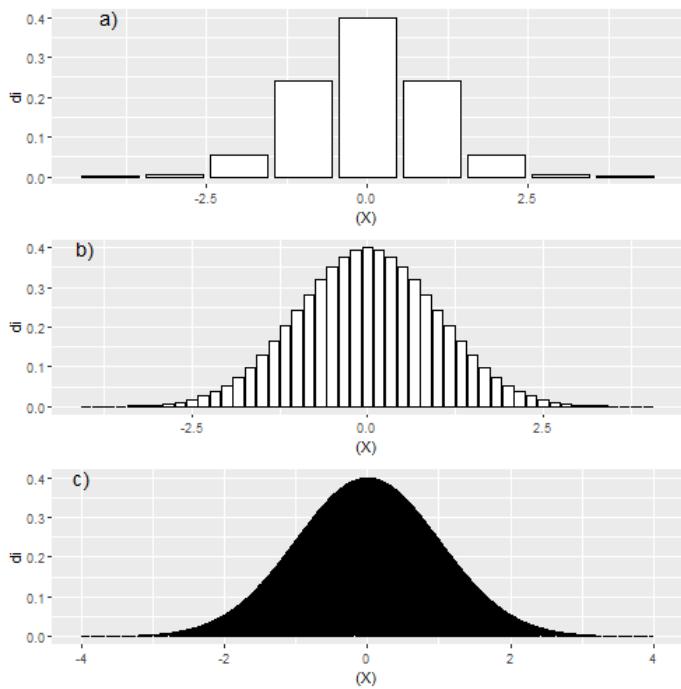
Variável Aleatória Contínua

Relembrando o conceito do histograma a respeito da densidade de frequência (di)



A altura do retângulo em cada intervalo de classe (Δ_i) é proporcional à densidade de frequência (fi/Δ_i) do intervalo, de modo que a área do retângulo seja igual $\Delta_i \times \frac{fi}{\Delta_i} = fi$.

Ou seja, com um número suficientemente grande de observações, diminuindo-se os intervalos de classe, o histograma tende ficar cada vez menos irregular, até aproximar da forma de uma curva bem mais suave.

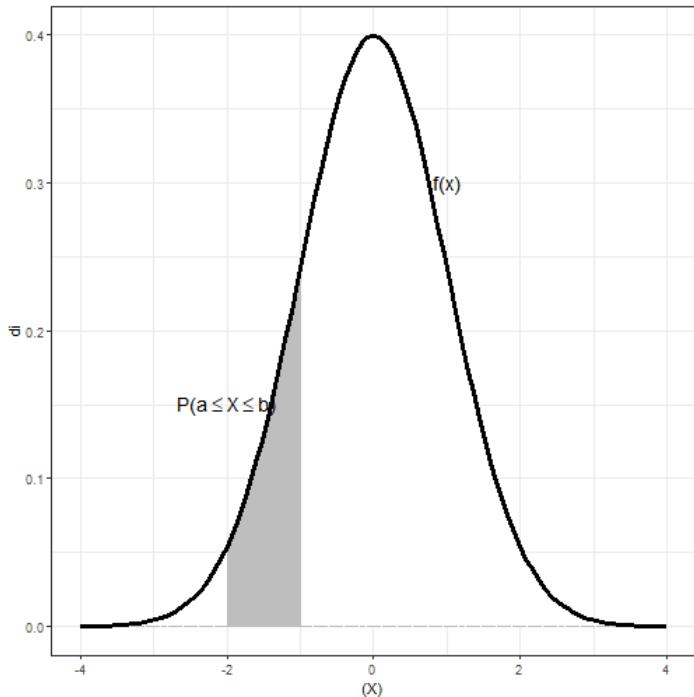


Como a probabilidade é interpretada como a frequência relativa de um evento em uma longa série de ensaios independentes, a curva obtida como a forma limite dos histogramas (c) representa a maneira pela qual a probabilidade total (1) é distribuída em relação à amplitude dos possíveis valores da v.a. X .

A função matemática $f(x)$, cujo gráfico produz tal curva é chamada **função densidade de probabilidade (f. d. p.)** da v.a. contínua X .

PROPRIEDADES

- a área total sob a curva é igual a 1 ;
- $P(a \leq X \leq b) =$ área sob a curva entre os pontos a e b ;
- $f(x) \geq 0$ (não negativa) ;
- $P(X = x_i) = 0$.



Estando $f(x)$ de uma variável aleatória contínua X especificada, o problema de se calcular $P(a \leq X \leq b)$ vem a ser o cálculo da área sob a curva.

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

onde:

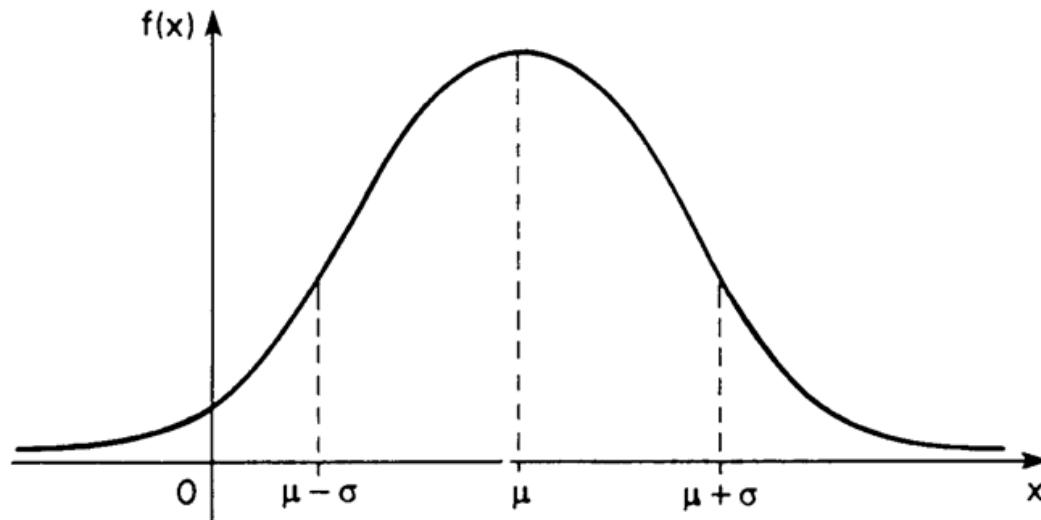
$$P(a \leq X \leq B) = P(a \leq X < B) = P(a < X \leq B) = P(a < X < B)$$

Distribuições Normal

Definição

Uma v.a. X tem distribuição normal com parâmetros μ e σ^2 , $-\infty < \mu < +\infty$ e $0 < \sigma^2 < \infty$, e sua *f. d. p.* dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty$$



Propriedades

i) Os parâmetros μ e σ^2 representam, respectivamente, a média e a variância da distribuição, isto é:

Onde:

i) $E(X) = \mu$ e $Var(X) = \sigma^2$;

ii) $f(x) \rightarrow 0$ quando $x \rightarrow \pm\infty$;

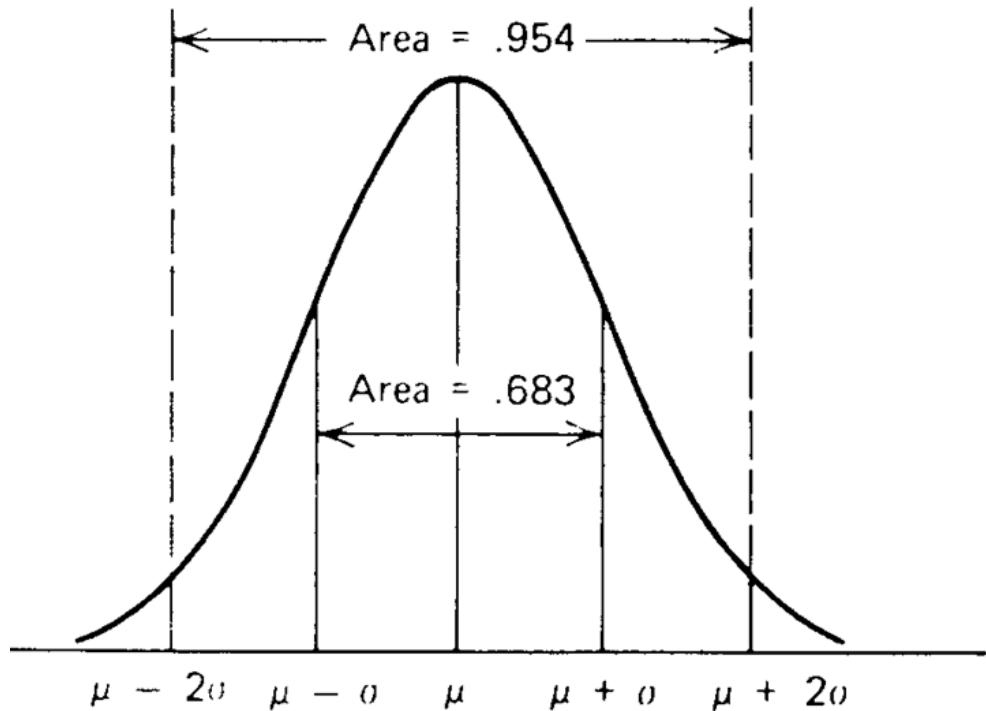
iii) $\mu - \sigma$ e $\mu + \sigma$ são os pontos de inflexão de $f(x)$;

iv) $x = \mu$ é o ponto de máximo de $f(x)$ e o valor máximo é $\frac{1}{\sigma\sqrt{2\pi}}$;

v) $f(x)$ é simétrica ao redor de $x = \mu$, isto é, $f(\mu + x) = f(\mu - x)$, para todo $-\infty < x < +\infty$;

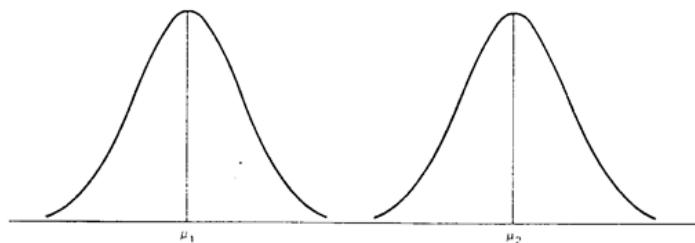
vi) média = moda = mediana.

$$X \sim N(\mu; \sigma^2)$$

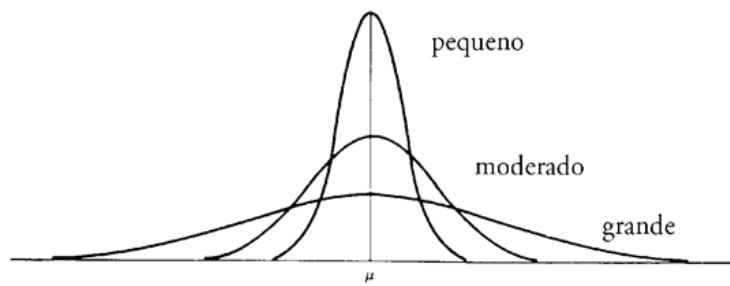


Interpretando Parâmetros

Duas distribuições normais com diferentes médias, mas com o mesmo desvio padrão (σ)

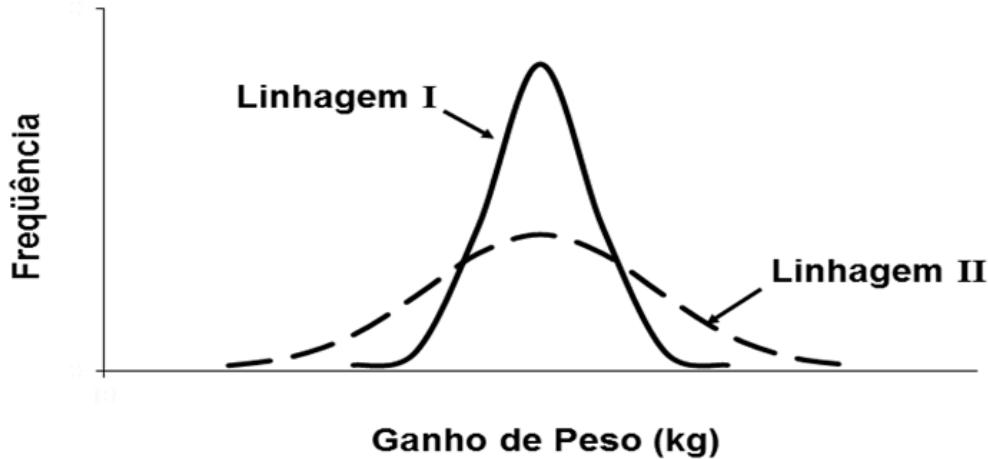


Três distribuições normais com médias iguais, mas com diferentes desvios padrões

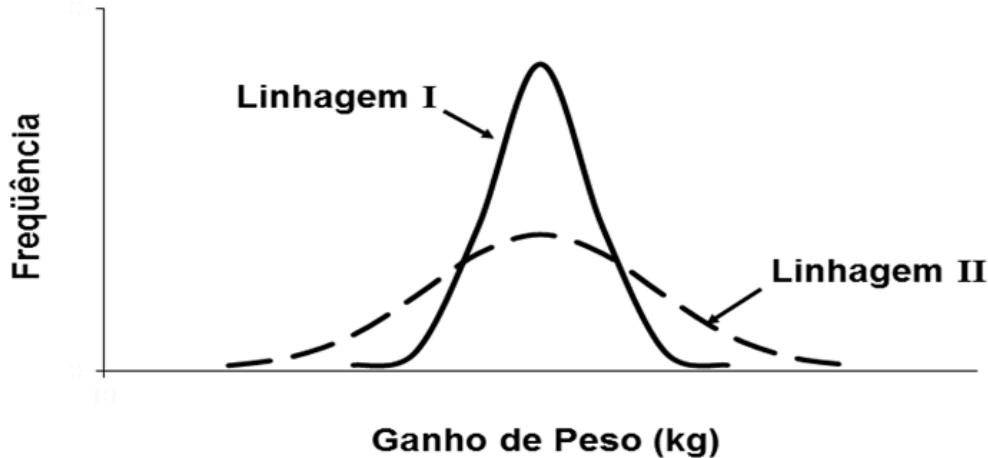


Acesse o Link para estudarmos os parâmetros da distribuição normal

<https://arpanosso.shinyapps.io/estatinfo/>



- a média aritmética e a variânci a Linhagem I são superiores às da Linhagem II.
- a média aritmética da Linhagem I é superior à da II e as variâncias são iguais.
- as médias aritméticas são iguais e a variânci a Linhagem I é superior à da II.
- as médias aritméticas são iguais e a variânci a Linhagem I é inferior à da II.
- a média aritmética e a variânci a Linhagem I são inferiores às da Linhagem II.



- () a média aritmética e a variância da Linhagem I são superiores às da Linhagem II.
- () a média aritmética da Linhagem I é superior à da II e as variâncias são iguais.
- () as médias aritméticas são iguais e a variância da Linhagem I é superior à da II.
- (x) as médias aritméticas são iguais e a variância da Linhagem I é inferior à da II.
- () a média aritmética e a variância da Linhagem I são inferiores às da Linhagem II.

Distribuição Normal Padrão (normal padronizada)

Definição

A particular distribuição normal com $\mu = 0$ e $\sigma^2 = 1$. Coincidindo com os parâmetros da variável Z :

$$Z = \frac{X - \mu}{\sigma}$$

onde $X \sim N(\mu, \sigma^2)$

Função de densidade de probabilidade:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-z^2}{2}}, -\infty < z < +\infty$$

Denotação

$$Z \sim N(0, 1)$$

Esperança e Variância

Se $X \sim N(\mu, \sigma^2)$, então a variável aleatória Z definida terá uma distribuição $N(0, 1)$.

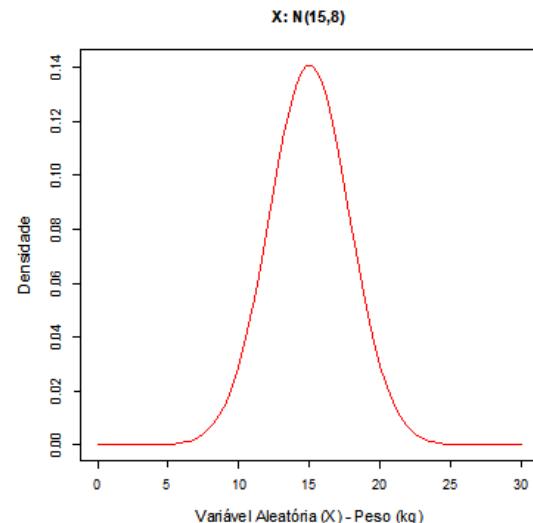
Esperança:

$$E(Z) = 0$$

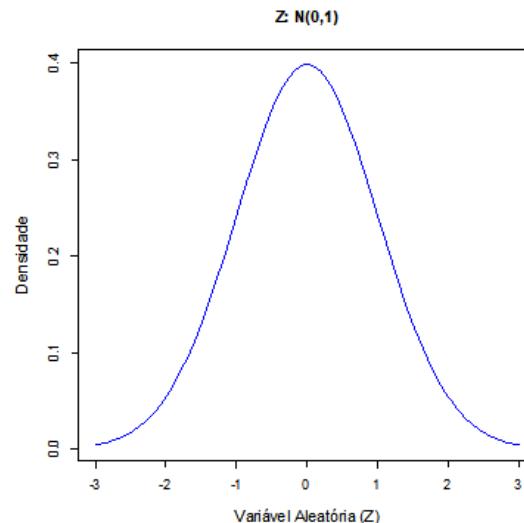
Variância:

$$Var(Z) = 1$$

A vantagem de se usar a variável Z é que as áreas, ou as probabilidades, associadas à distribuição normal padronizada são tabeladas. Assim, distribuição normal padrão é fundamental para o cálculo de probabilidades relativas a uma distribuição normal qualquer.



$$Z = \frac{X - \mu}{\sigma}$$

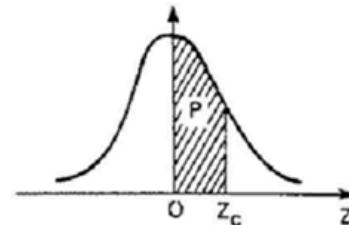



Distribuição Normal Padrão

Tabela - Normal Padrão

TÁBULA III

Distribuição normal reduzida: $N(0;1)$
Probabilidades p tais que $p = P(0 < Z < z_c)$



parte inteira e primeira decimal de z_c	SEGUNDA DECIMAL DE z_c										parte inteira e primeira decimal de z_c
	0	1	2	3	4	5	6	7	8	9	
	$p = 0$										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7

Outro vídeo bastante interessante do **Khan Academy Brasil** apresentando uma explicação sobre as ditribuição normal e a normal padrão:



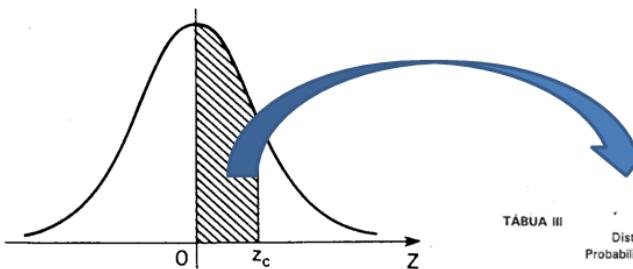
[Link do vídeo](#)

Exemplo 1

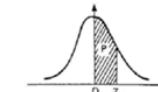
A figura abaixo ilustra a probabilidade fornecida pela tabela, ou seja,

$$P(0 \leq Z \leq z_c)$$

$$\text{Se } z_c = 1,73$$



TÁBUA III
Distribuição normal reduzida: $N(0;1)$
Probabilidades p tais que $p = P(0 < Z < z_c)$



Calcular:

- a) $P(-1,73 \leq Z \leq 0)$
- b) $P(Z \geq 1,73)$
- c) $P(Z < -1,73)$
- d) $P(Z \leq 1,73)$
- e) $P(0,47 \leq Z \leq 1,73)$

parte inteira e primeira decimal de z_c	SEGUNDA DECIMAL DE z_c										parte inteira e primeira decimal de z_c
	0	1	2	3	4	5	6	7	8	9	
0,0	0,0000	0,0399	0,0798	0,1197	0,1596	0,1994	0,2392	0,2790	0,3188	0,3586	0,0
0,1	0,0798	0,1197	0,1596	0,1994	0,2392	0,2790	0,3188	0,3586	0,3984	0,4382	0,1
0,2	0,7982	0,8371	0,8760	0,9149	0,9538	0,9927	0,6749	0,7142	0,7535	0,7928	0,2
0,3	11781	12172	12562	12952	13307	13663	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17368	17722	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21568	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	29033	29343	29651	29959	30265	30570	30874	31178	31481	31784	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33298	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35983	36214	1,0
1,1	36433	36650	36868	37076	37286	37493	37698	37890	38100	38298	1,1
1,2	38483	38686	38877	39065	39251	39433	39617	39799	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	42057	42224	42389	42554	42717	42880	43042	43202	43363	43519	1,4
1,5	43319	43448	43574	43699	43822	42943	44062	44179	44285	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7

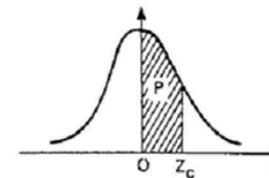
$$a) P(-1,73 \leq Z \leq 0)$$

0.5 - pnorm(-1.73)

#> [1] 0.4581849

TÁBUA III

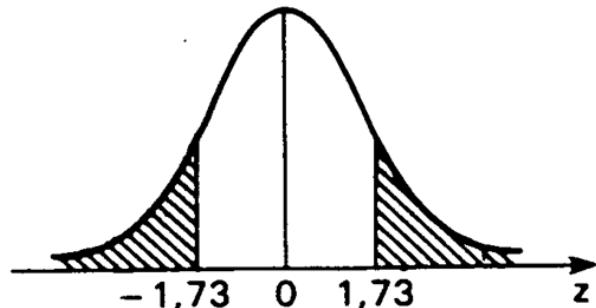
Distribuição normal reduzida: $N(0;1)$
Probabilidades p tais que $p = P(0 < Z < z_c)$



parte inteira e primeira decimal de z_c	SEGUNDA DECIMAL DE z_c									parte inteira e primeira decimal de z_c	
	0	1	2	3	4	5	6	7	8		
0,0	p = 0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1		03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2		07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3		11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4		15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5		19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38483	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7

b) $P(Z \geq 1,73)$

c) $P(Z \leq -1,73)$



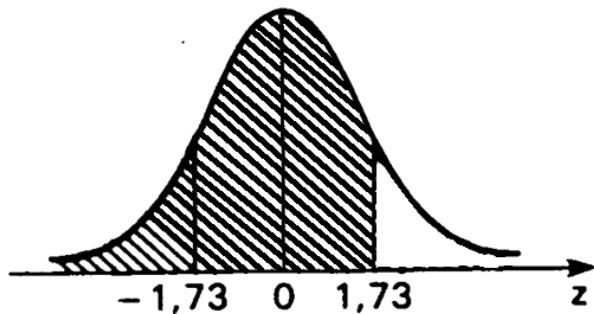
por simetria:

```
pnorm(-1.73)
```

```
#> [1] 0.04181514
```

$$\begin{aligned}P(Z \geq 1,73) &= P(Z \leq -1,73) \\&= 0,5 - P(0 \leq Z \leq 1,73) \\&= 0,5 - 0,4582 \\&= 0,0418\end{aligned}$$

d) $P(Z \leq 1,73)$

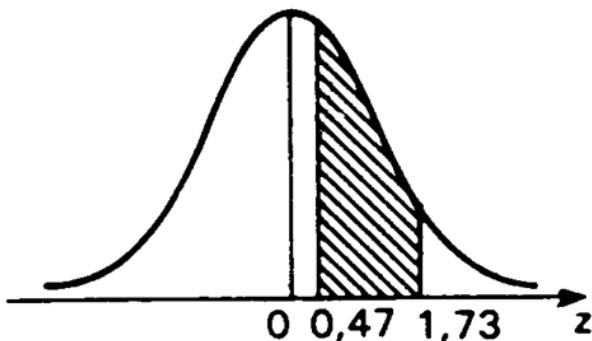


```
pnorm(1.73)
```

```
#> [1] 0.9581849
```

$$P(Z \leq 1,73) = 0,5 + P(0 \leq Z \leq 1,73) = 0,5 + 0,4582 = 0,9582$$

e) $P(0,47 \leq Z \leq 1,73)$



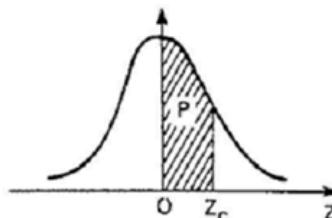
```
pnorm(1.73) - pnorm(0.47)
```

```
#> [1] 0.2773624
```

$$\begin{aligned}P(0,47 \leq Z \leq 1,73) &= P(0 \leq Z \leq 1,73) - P(0 \leq Z \leq 0,47) \\&= 0,4582 - 0,1808 = 0,2774\end{aligned}$$

TÁBUA III

Distribuição normal reduzida: $N(0;1)$
 Probabilidades p tais que $p = P(0 < Z < z_c)$



parte inteira e primeira decimal de Z_c	Se $z_c = 0,47$										parte inteira e primeira decimal de Z_c
	SEGUNDA DECIMAL DE Z_c										
	0	1	2	3	4	5	6	7	8	9	
	$p = 0$										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7

Exemplo 2

Suponha que a variável aleatória $X \sim N(3, 16)$, calcular: $P(2 \leq X \leq 5)$

$$P(2 \leq X \leq 5) = P\left(\frac{2-3}{4} \leq \frac{X-\mu}{\sigma} \leq \frac{5-3}{4}\right) = P(-0,25 \leq Z \leq 0,5)$$

$$P(-0,25 \leq Z \leq 0,5) = P(-0,25 \leq Z \leq 0) + P(0 < Z \leq 0,5)$$

$$P(-0,25 \leq Z \leq 0,5) = 0,0987 + 0,1915 = 0,2902 \text{ ou seja,}$$

$$P(2 \leq X \leq 5) = 0,2902$$

No R:

```
pnorm(5,3,4) - pnorm(2,3,4)
```

```
#> [1] 0.2901688
```

Exemplo 3

Sabendo-se que os pesos à desmama (X) de 10.000 bezerros de um rebanho são distribuídos normalmente, com média ($\mu = 170 \text{ kg}$) e desvio padrão ($\sigma = 5 \text{ kg}$), pergunta-se:

- a) qual é o número esperado de bezerros com peso superior a 165 kg?

Solução:

$$P(X \geq 165) = P\left(\frac{X - \mu}{\sigma} > \frac{165 - 170}{5}\right) = P(Z \geq -1)$$

$$P(Z \geq -1) = P(-1 \leq Z \leq 0) + P(Z \geq 0) = P(0 \leq Z \leq 1) + 0,50$$

$$P(Z \geq -1) = 0,3413 + 0,50 = 0,8413$$

Portanto, o número esperado é $10.000 \times 0,8413 = 8413$ bezerros.

No R:

```
trunc(10000 * (1 - pnorm(165, 170, 5)))
```

```
#> [1] 8413
```

b) que peso (x_c) deve atingir um bezerro para que ele supere 80% dos pesos à desmama desse rebanho?

$$P(X \leq 170) + P(170 \leq X \leq x_c) = 0,80$$

$$0,50 + P(170 \leq X \leq x_c) = 0,80$$

$$P(170 \leq X \leq x_c) = 0,30$$

$$P(170 \leq X \leq x_c) = P\left(\frac{170 - 170}{5} \leq \frac{X - \mu}{\sigma} \leq \frac{x_c - 170}{5}\right) = 0,30$$

$$P\left(0 \leq Z \leq \frac{x_c - 170}{5}\right) = 0,30$$

Obervando a tabela

$$P(0 \leq Z \leq 0,84) = 0,30$$

Então,

$$\frac{x_c - 170}{5} = 0,84 \rightarrow x_c = 174,2 \text{ kg}$$

No R:

```
qnorm(0.80, 170, 5)
```

```
#> [1] 174.2081
```

Aproximação Normal à Binomial

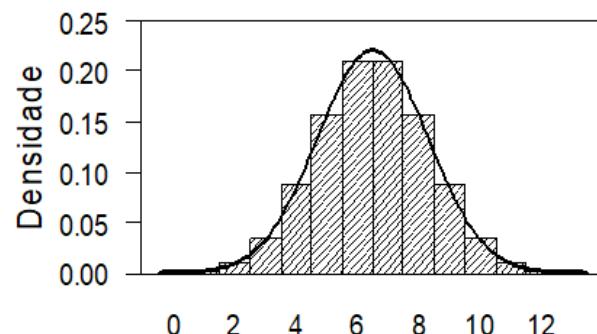
Definição

Se X tem distribuição binomial $b(n, p)$, onde n é grande e p não é muito próximo de 0 ou 1, a distribuição da variável padronizada ficará:

$$Z = \frac{X - np}{\sqrt{(np(1-p))}}$$

que é aproximadamente $N(0, 1)$.

$$P(a \leq X \leq b) = \sum_{x=a}^b {}^n \binom{x}{p} p^x (1-p)^{n-x} \cong P\left(\frac{a-np}{\sqrt{np(1-p)}} \leq Z \leq \frac{b-np}{\sqrt{np(1-p)}}\right)$$



Correção da continuidade

Tendo em vista que uma distribuição discreta (binomial) é aproximada por uma contínua (normal), a melhor aproximação é obtida calculando:

$$P(a \leq X \leq b) = P\left(\frac{\left(\frac{a-0,5}{n}\right) - p}{\sqrt{\frac{p(1-p)}{n}}} \leq Z \leq \frac{\left(\frac{b+0,5}{n}\right) - p}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

A distribuição normal pode ser recomendada para aproximar probabilidades binomiais, contanto que p seja próximo de 0,5. Quando p é muito pequeno e n é grande, a distribuição de **Poisson** é mais apropriada.

Regra prática: n pode ser assumido como "suficientemente" grande para se usar a distribuição normal, quando: $np(1 - p) \geq 3$ sendo que a aproximação melhora com o crescimento de n .

Exemplo 4

Supondo que $X \sim b(15; 0,4)$, calcule a probabilidade:

$P(7 \leq X \leq 10)$, por meio da aproximação normal à binomial.

$$P(7 \leq X \leq 10) = P\left(\frac{\left(\frac{7-0,50}{15}\right) - 0,40}{\sqrt{\frac{0,40(1-0,40)}{15}}} \leq Z \leq \frac{\left(\frac{10+0,50}{15}\right) - 0,40}{\sqrt{\frac{0,40(1-0,40)}{15}}}\right)$$

$$P(0,263 \leq Z \leq 2,368) = 0,49111 - 0,10194 = 0,389$$

no R:

```
pnorm(2.368) - pnorm(0.263)
```

```
#> [1] 0.387333
```

ou

```
pbinom(11,15,0.4) - pbinom(6,15,0.4)
```

```
#> [1] 0.3882591
```