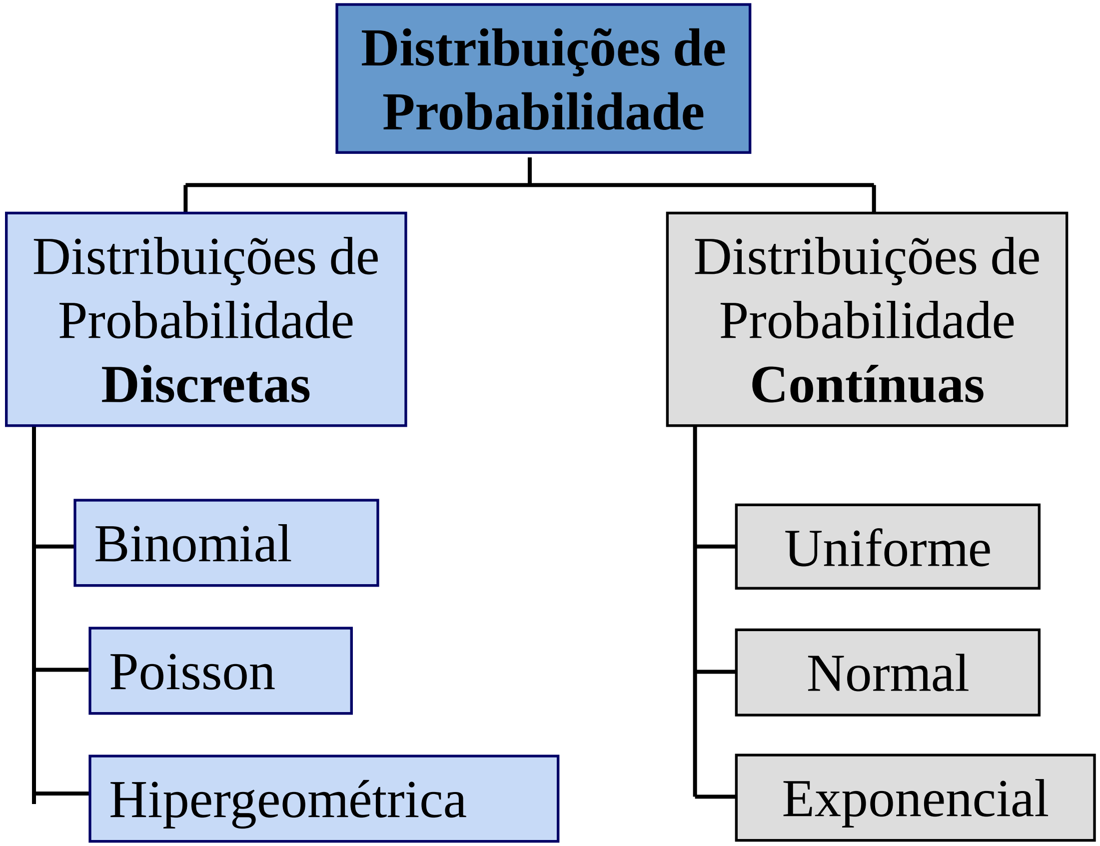
Probabilidade e Estatística

Aula 6 – Distribuições Contínuas (Parte 02)

Leitura obrigatória: Devore, Capítulo 4

Distribuições de Probabilidade



Uniforme

X é v.a. com distribuição uniforme contínua ($X \sim U[a, b]$) tem fdp:

Definição!

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{caso contrário} \end{cases}$$

Em que a = mínimo valor possível para X

b = máximo valor possível para X

<u>Eventos com mesmo comprimento são equiprováveis!</u>

Uniforme

Propriedades:

A **média** de uma distribuição uniforme é:

$$\mu = \frac{a+b}{2}$$

O **desvio-padrão** de uma distribuição uniforme é:

$$\sigma = \sqrt{\frac{(b-a)^2}{12}}$$

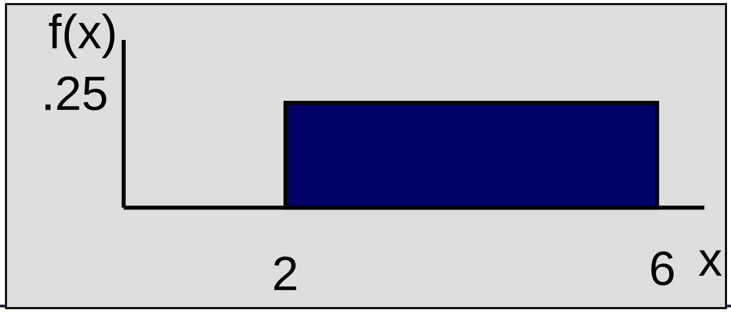
Uniforme

Execício: Seja $X \sim U[2,6]$, isto é:

$$f(x) = \frac{1}{6-2} = 0.25 \;\; para \;\; 2 \le x \le 6$$

a) $P(2.5 < X < 3.75) = ?$

b) Determine a média e o desvio-padrão de X.



$$\mu = \frac{a+b}{2} = \frac{2+6}{2} = 4$$

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(6-2)^2}{12}} = 1.1547$$

Uniforme

Exercício: O peso líquido, em libras, de um pacote com herbicida químico é uniforme para $49.75 < x < 50.25$ libras.

a) Determine a média e a variância do peso dos pacotes.

b) Determine a função de distribuição acumulada do peso dos pacotes.

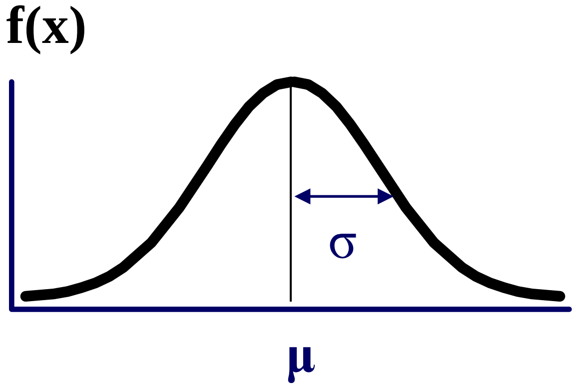
c) Determine $P(X < 50.1)$.

Normal ou Gaussiana

- A distribuição normal é a mais importante de todas as distribuições.
- Também é chamada de distribuição Gaussiana.
- Muitas populações numéricas possuem distribuição que pode ser aproximada por uma normal.
  - Altura, peso, etc
  - Erros de medida em experimentos
  - Notas em testes.
- Mesmo que as variáveis aleatórias não sejam distribuídas como uma normal:
  - a soma e média de v.a. tem uma distribuição aproximadamente normal sob certas condições (Teorema do Limite Central)

Normal: propriedades

- **Formato de Sino**
- **Simétrica**
- **Média, Mediana e Moda são iguais**
- **Teoricamente o domínio da variável vai de -∞ a +∞**



Média
= Mediana
= Moda

Normal: fdp

A função densidade de probabilidade de uma v.a. $X \sim N(\mu, \sigma)$ é:

Definição!

$$\boldsymbol{f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$$

Em que $e = 2.71828$

$\pi = 3.14159$

μ = média populacional

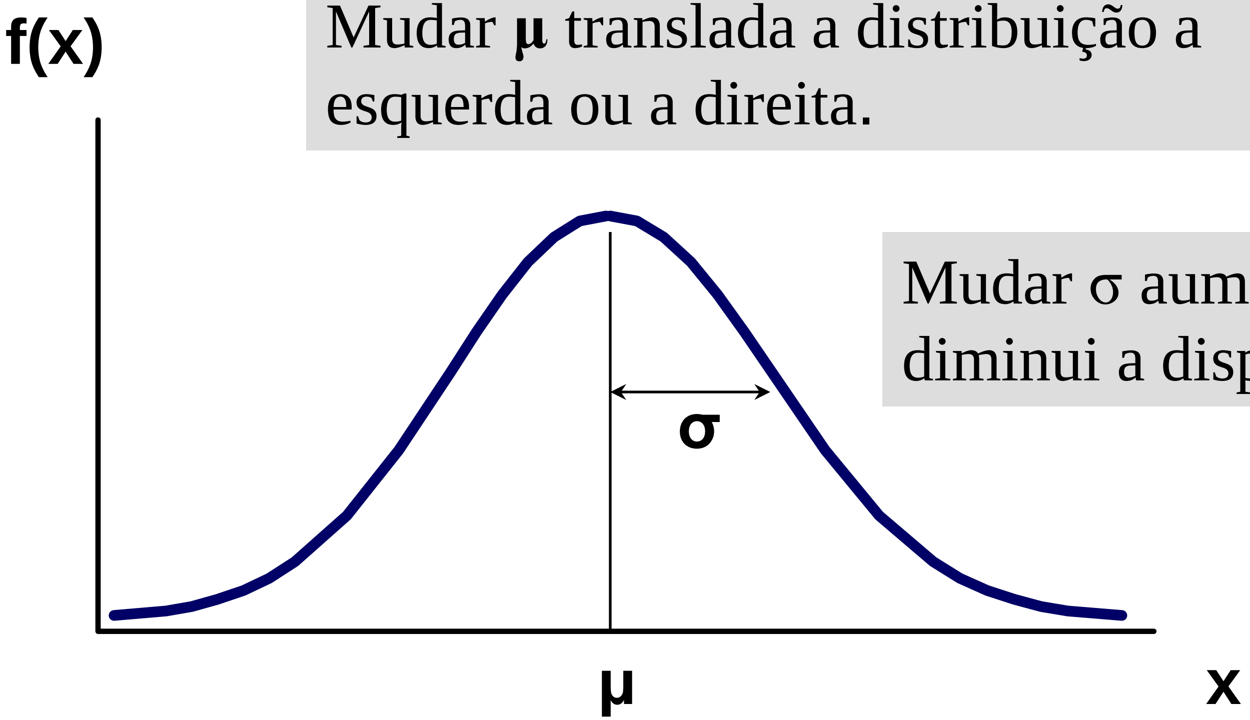
σ = desvio-padrão populacional

x = qualquer valor da variável contínua $(-\infty, \infty)$

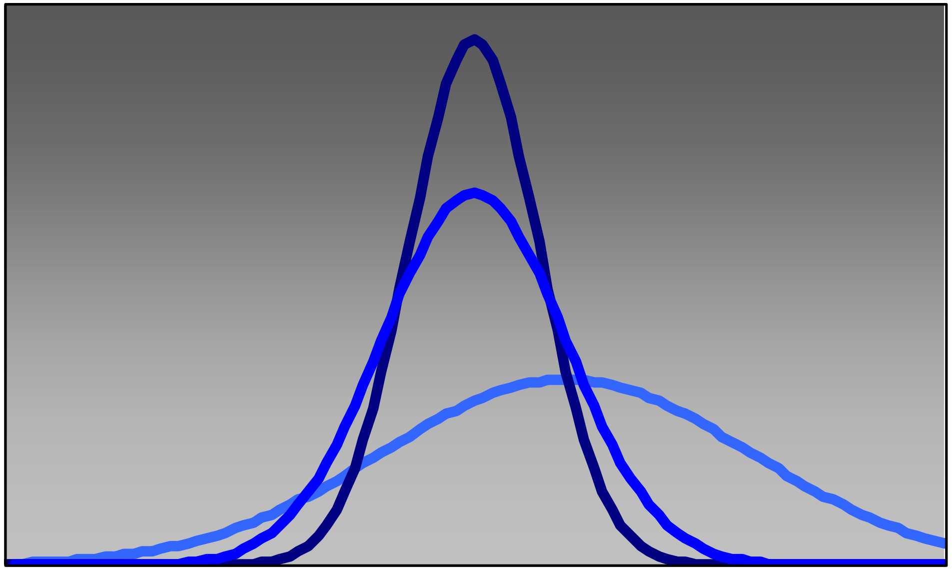
Normal: formato

Mudar **μ** translada a distribuição a esquerda ou a direita.

Mudar σ aumenta ou diminui a dispersão.



Normal: formato



Ao variarmos os parâmetros μ e σ, obtemos distribuições normais diferentes

Ver planilha do khanacademy.com

Normal: FDA

Para calcular a probababilidade de X pertencer a um intervalo $[a, b]$, fazemos: $P(a \leq X \leq b) = F(b) - F(a)$.

Por definição, a função distribuição acumulada de uma v.a. $X \sim N(\mu, \sigma)$ é:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{(v-\mu)}{\sigma}\right)^2} dv$$

PROBLEMA: Esta integral **não possui uma expressão analítica**.

SOLUÇÃO: Calcular numericamente (computacionalmente) → Tabelar os valores! Mas vamos precisar fazer uma tabela para cada combinação de parâmetros (μ, σ)???

Normal Padrão

- Vamos precisar fazer uma tabela para cada combinação de parâmetros (μ, σ)? Não! Vamos precisar de somente uma tabela, a da normal padrão!
  - A distribuição normal padrão tem média 0 e desvio-padrão 1.

- Qualquer distribuição normal (com qualquer média e desvio-padrão) pode ser transformada em uma **distribuição normal padrão (Z)**.

- Para calcular probabilidades de X, precisamos transformar X unidades em Z unidades.

Normal Padrão

Teorema: Seja $X \sim N(\mu, \sigma)$. O escore-Z associado a X, tem distribuição normal padrão!

$$\boldsymbol{Z = \frac{X - \mu}{\sigma} \sim N(0, 1).}$$

Obs: provar que a distribuição de Z é normal é uma tarefa muito avançada para este curso, mas é fácil provar que a média de Z é 0 e o desvio-padrão é 1:

$$E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}(E(X) - \mu) = 0 \qquad e \qquad V\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}V(X) = 1$$

Normal Padrão: fdp

Função densidade para a distribuição normal padrão, $Z \sim N(0,1)$:

Definição!

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

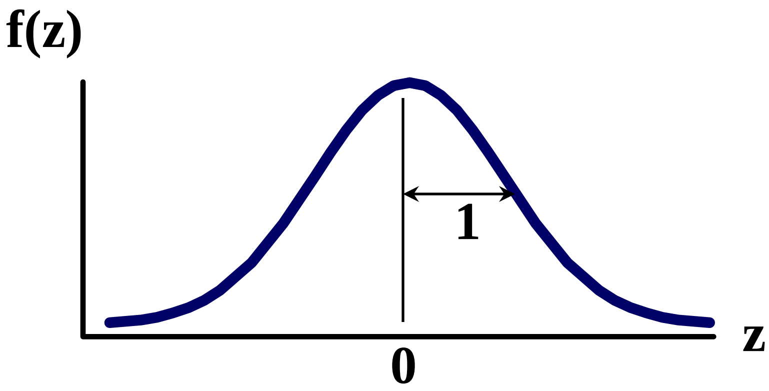
Em que e = 2.71828

$\pi = 3.14159$

z = qualquer valor real, ou seja, $z \in (-\infty, \infty)$

Normal Padrão: formato

- Também conhecida como distribuição “Z”
- Média é 0
- Desvio-padrão é 1



Valores acima da média são valores com $z > 0$. Valores abaixo da média são valores com $z < 0$.

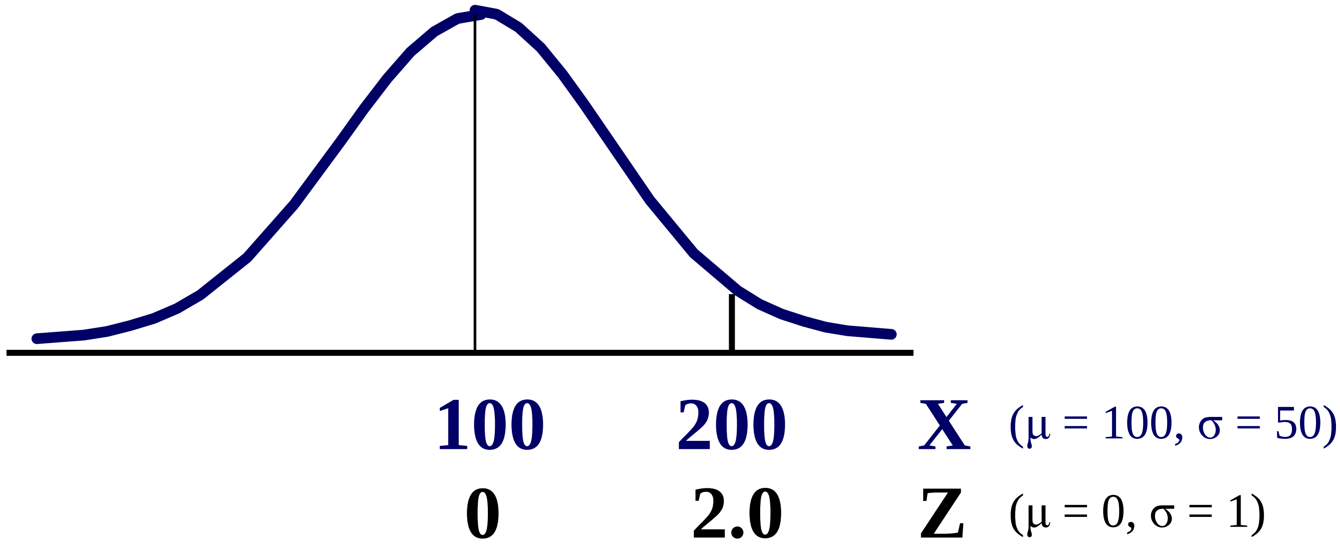
Normal: exemplo

- Se X tem distribuição normal com média 100 e desvio-padrão 50, o valor Z para X=200 é:

$$Z = \frac{X - \mu}{\sigma} = \frac{200 - 100}{50} = 2.0$$

- Isto significa que X = 200 está dois desvios-padrão acima (2 incrementos de 50 unidades) da média de 100.

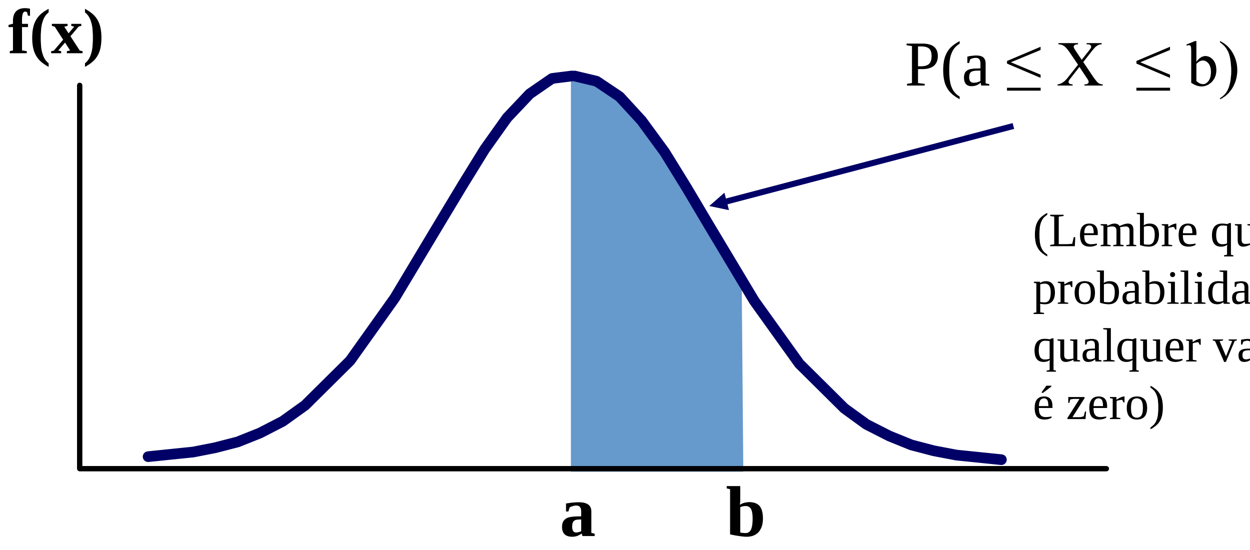
Normal: exemplo



Note que a distribuição é a mesma, apenas a escala e locação mudaram. Podemos expressar o problema em termos dos valores originais (X) ou de valores padronizados (Z).

Normal: probabilidades

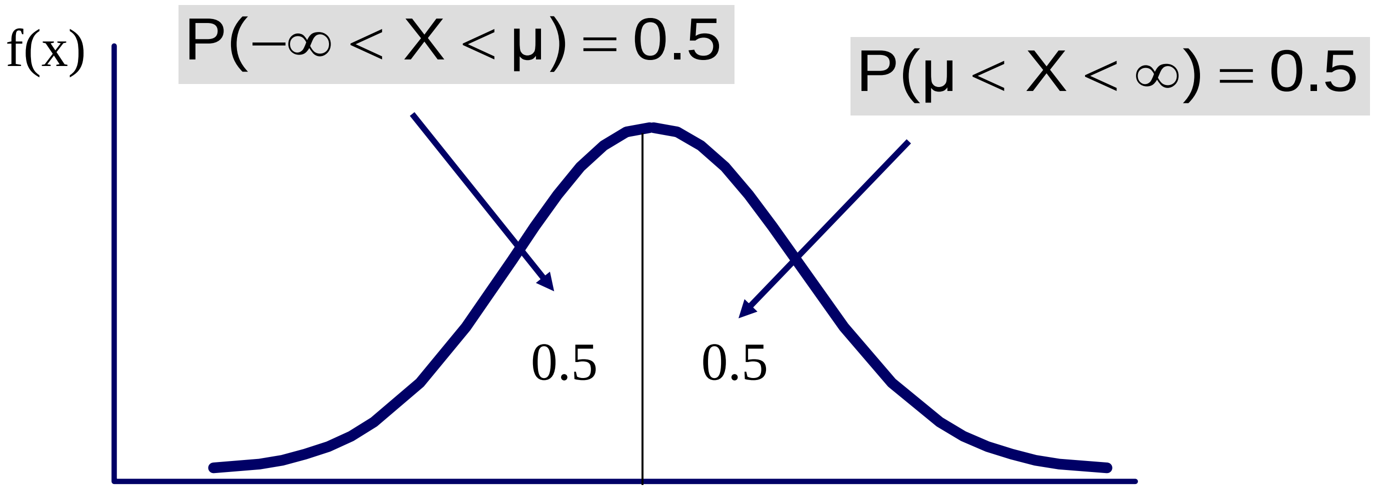
A probabilidade é medida pela área abaixo da curva:



(Lembre que a probabilidade de qualquer valor individual é zero)

Normal: probabilidades

A área total abaixo da curva é 1 e a curva é simétrica, então metade da área está abaixo da média e metade está acima da média.



$$P(-\infty < X < \infty) = 1.0$$

Normal: tabela

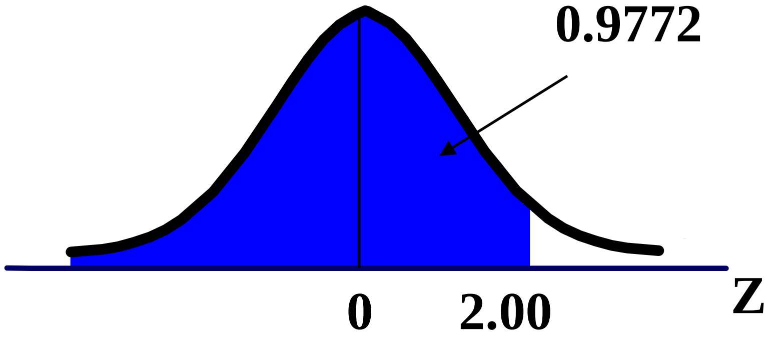
- As probabilidades de uma distribuição normal padrão estão tabeladas em qualquer livro texto.

- A tabela abaixo mostra a probabilidade de valores abaixo de z:

  tabela da acumulada da normal-padrão

Exemplo:

P(Z < 2.00) = 0.9772



Normal: tabela

As colunas: listam segundas casas decimais para z.

Z	0.00	0.01	0.02 …
0.0			
0.1			
.			
.			
.			
2.0	**0.9772**		

As linhas: listam valores de z até a primeira casa decimal.

O valor nas células da tabela mostram a probabilidade de Z = – ∞ até o valor desejado de z.

P(Z < 2.00) = 0.9772

Normal: tabela

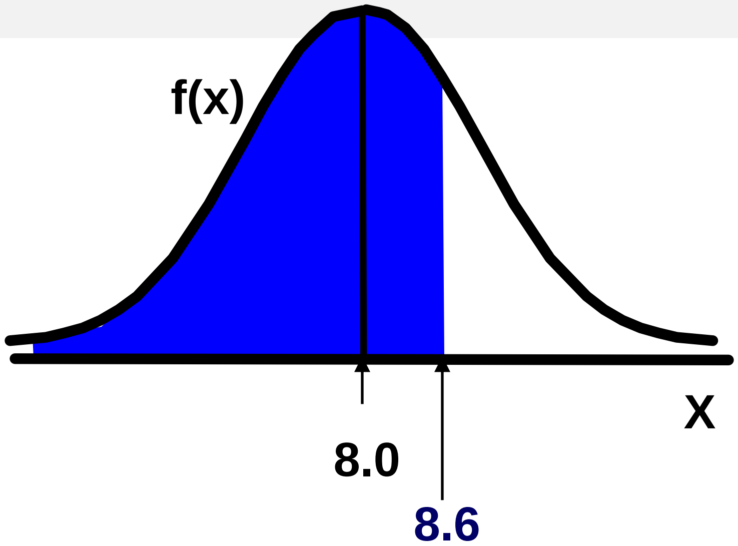
Para encontrar $P(a < X < b)$ quando X tem distribuição normal:

- Represente o problema para a curva normal de X.
- Tranforme valores de X em valores de Z.
- Use a Tabela da Normal Padrão.

Normal: tabela

Exercício: Seja X o tempo necessário em (segundos) para baixar um arquivo da internet. Suponha que X seja normal com média 8 e desvio-padrão 5

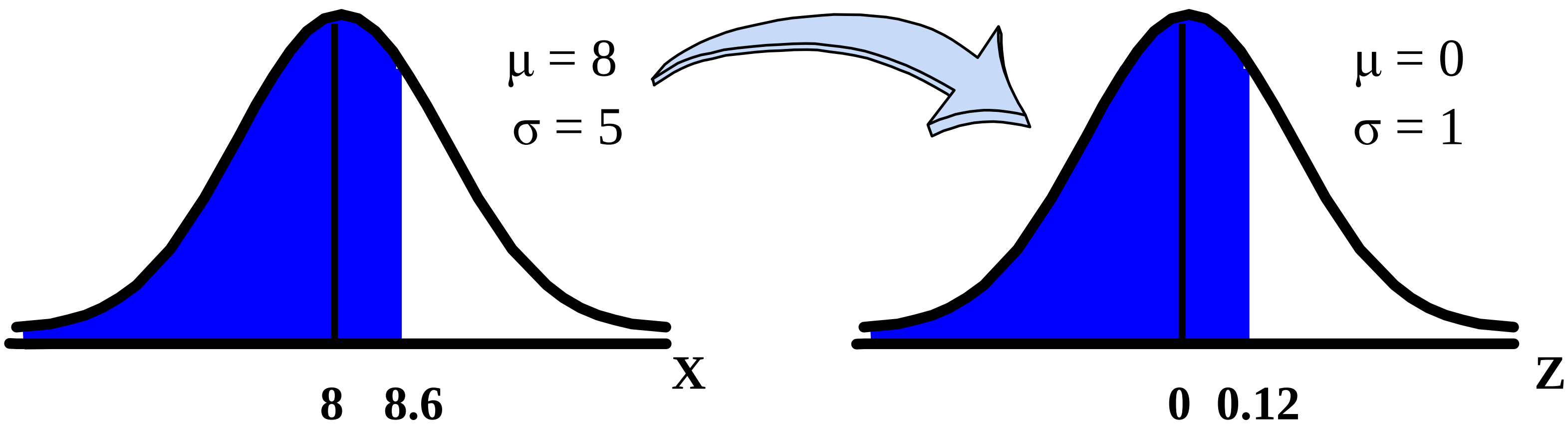
Encontre $P(X < 8.6)$.



Normal: tabela

Exercício - Solução: $X \sim N(8,5)$ e $P(X < 8.6)$?

$$Z = \frac{X - \mu}{\sigma} = \frac{8.6 - 8.0}{5.0} = 0.12$$



P(X < 8.6)   P(Z < 0.12)

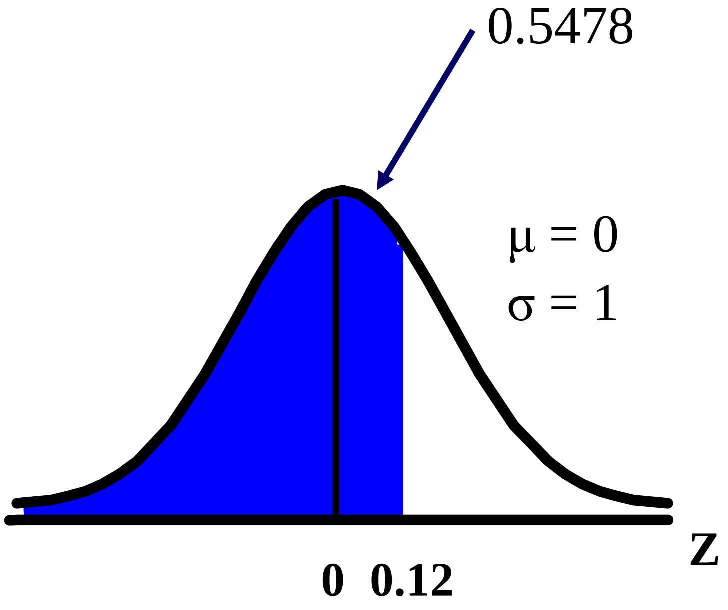
Normal: tabela

Tabela da Normal Padrão (Extrato)

Z	.00	.01	.02
0.0	.5000	.5040	.5080
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.6179	.6217	.6255

$P(X < 8.6)$

$= P(Z < 0.12)$

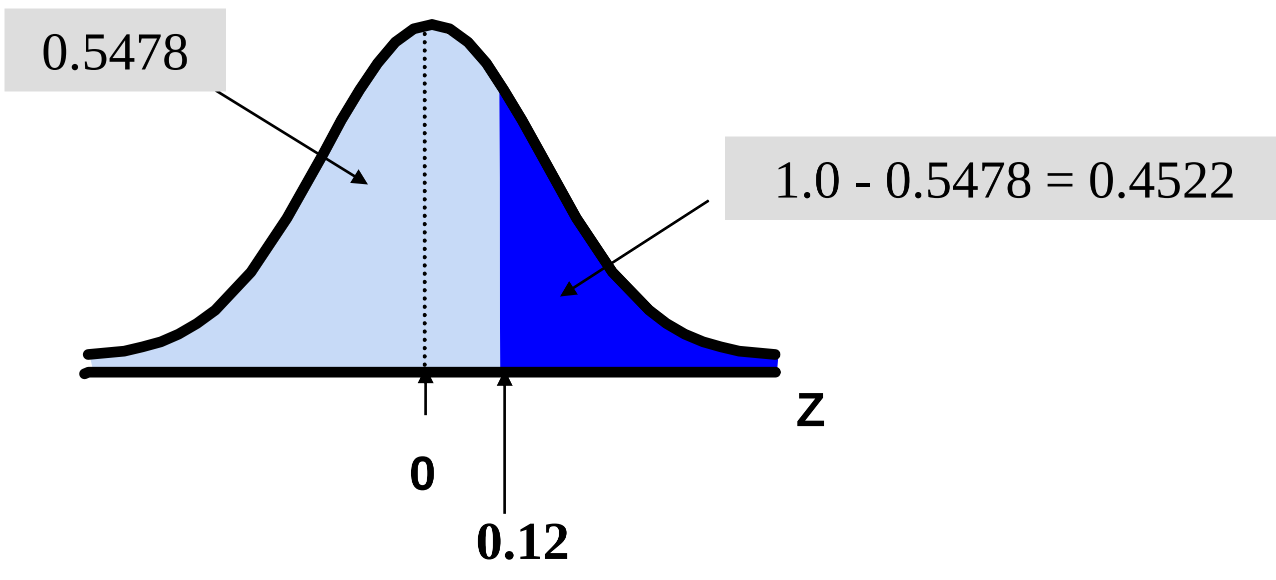


Normal: tabela

- Encontre $P(X > 8.6)$.

$$P(X > 8.6) = P(Z > 0.12) = 1.0 - P(Z \leq 0.12)$$

$$= 1.0 - 0.5478 = 0.4522$$



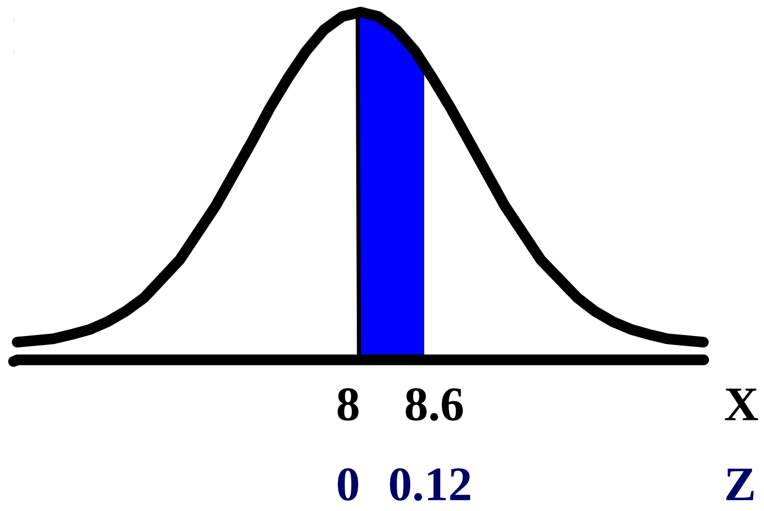
Normal: tabela

Exercício: Seja X ~ N(8, 5). Determine P(8 < X < 8.6).

Calcule Z-valores:

$$Z = \frac{X - \mu}{\sigma} = \frac{8 - 8}{5} = 0$$

$$Z = \frac{X - \mu}{\sigma} = \frac{8.6 - 8}{5} = 0.12$$



P(8 < X < 8.6)

= P(0 < Z < 0.12)

Normal: tabela

Exercício - Solução:

Tabela da Normal Padrão (Extrato)

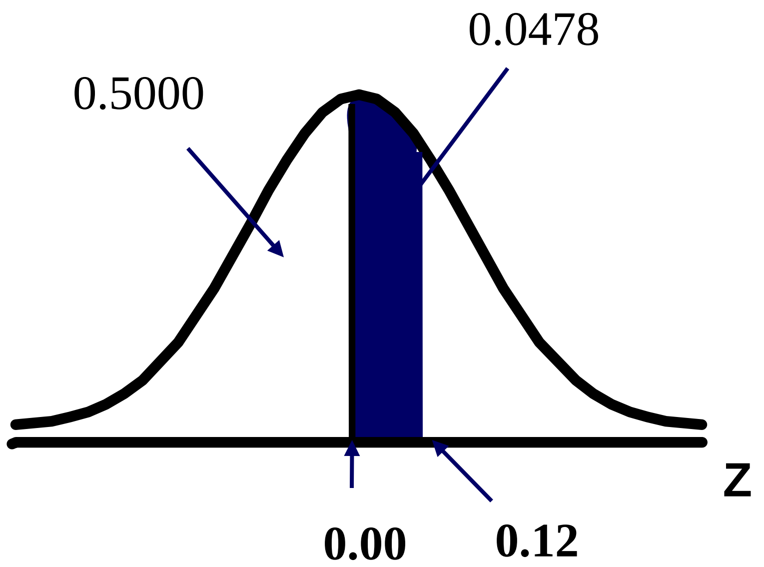
Z	.00	.01	.02
0.0	.5000	.5040	.5080
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.6179	.6217	.6255

$P(8 < X < 8.6)$

$= P(0 < Z < 0.12)$

$= P(Z < 0.12) – P(Z \leq 0)$

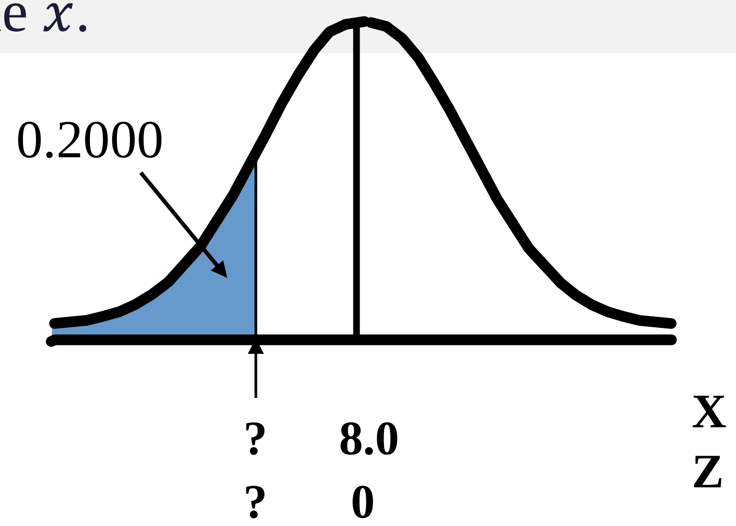
$= 0.5478 - 0.5000 = 0.0478$



Percentil

Exercício: Seja X o tempo em segundos levado para baixar um arquivo na internet. Suponha que X é uma normal com média 8 e desvio-padrão 5.

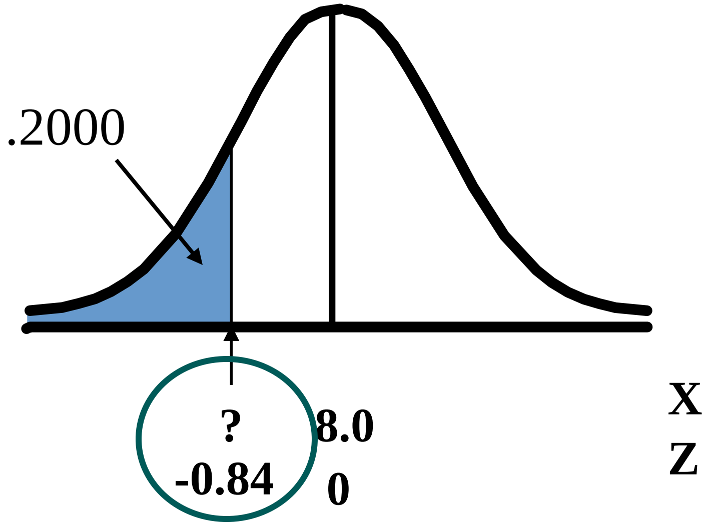
Encontre x tal que 20 % dos tempos de download sejam menores do que x.



Percentil

Exercício – Solução: Primeiro, encontre o valor de Z que corresponde a probabilidade conhecida, usando a tabela da normal padrão, ou seja, o percentil 20% de Z, $z_{0.2}$.

Z	….	.03	.04	.05
-0.9	….	.1762	.1736	.1711
-0.8	….	.2033	.2005	.1977
-0.7	….	.2327	.2296	.2266



Percentil:

- Segundo, converta o valor de Z em X, para determinar o percentil 20% de X:

$$z_{0.2} = \frac{x_{0.2} - \mu}{\sigma} \Rightarrow x_{0.2} = \mu + \sigma z_{0.2}$$

Segue:

$$x_{0.2} = 8 + 5 * (-0.84) = 3.8$$

- Então 20% dos tempos de download de uma distribuição de tempos normal com média 8 e desvio-padrão 5 estão abaixo de 3.8 segundos.

Normal: exercício

Exercício: (Prova 2011/2) Sem um sistema automático de irrigação, a altura das plantas duas semanas depois de germinar é distribuída normalmente, com uma média de 2.5 cm e um desvio-padrão de 0.5 cm.

(a) (1 ponto) Qual é a probabilidade de a altura da planta estar entre 2.0 e 3.0 cm?

(b) (1 ponto) Que altura é excedida por 90% das plantas?

Exponencial

- Usada para modelar intervalo de tempo entre ocorrência de eventos (duração ou tempo de espera)
  - Exemplos:
    - Tempo entre chegadas de caminhões para descarregar carregamento
    - Tempo entre transações em um caixa rápido
    - Vida útil de equipamentos eletrônicos ou mecânicos

- Caso particular da gama ($\alpha=1$ e $\beta=1/\lambda$)

- Se a Poisson é um bom modelo para o nº de eventos por unidade de tempo, a exponencial modela bem o tempo de espera entre ocorrência de eventos.

Exponencial

- Para uma variável aleatória, X, distribuída exponencialmente com parâmetro λ , a probabilidade do evento durar no máximo um valor x é:

Definição!

$$\boldsymbol{F(x) = P(X \leq x) = 1 - e^{-\lambda x}}$$

Em que

x = qualquer valor positivo, $x \geq 0$

λ = número médio de "chegadas" ou nº médio de eventos por unidade de tempo

Exponencial

- A função densidade de probabilidade de $X \sim Exp(\lambda)$ é:

$$\boldsymbol{f(x) = \lambda e^{-\lambda x}, para\ x \geq 0}.$$

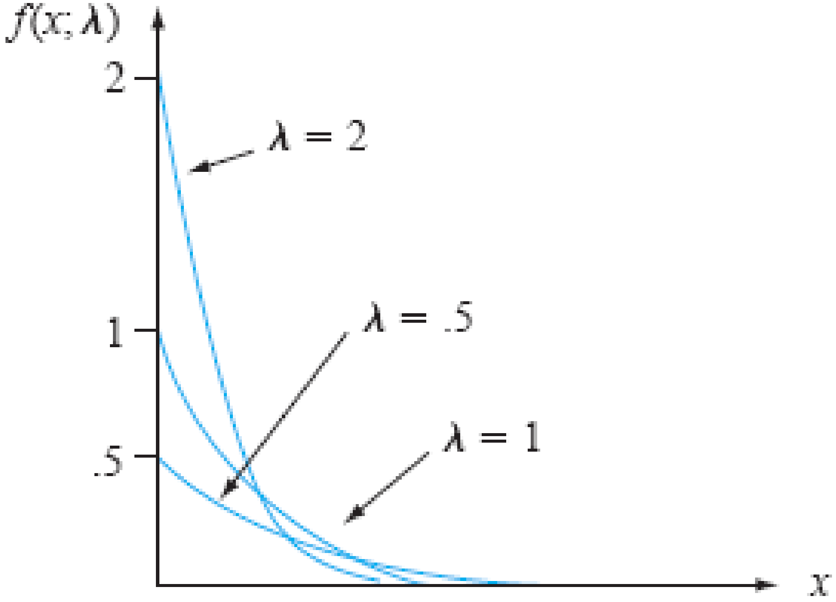
- A média, e o desvio-padrão da exponencial são dados por:

$$E(X) = \sigma(X) = 1/\lambda$$

ou seja, o tempo médio de espera é o inverso do número de chegadas (ou eventos) por unidade de tempo!

Exponencial

- A fdp de $X \sim Exp(\lambda)$ tem formato que depende de λ:



Exponencial

- Propriedade da FALHA DE MEMÓRIA:

$$P(X \geq t + t_0 | X \geq t_0) = P(X \geq t)$$

a probabilidade de o evento durar mais t unidades de tempo não depende de quanto tempo o evento já durou. Portanto, para a exponencial, o futuro não depende do passado!

$$\mathrm{P}\left(X \geq t + t_0 \middle| X \geq t_0\right) = \frac{\mathrm{P}\left(X \geq t + t_0 \bigcap X \geq t_0\right)}{\mathrm{P}\left(X \geq t_0\right)}$$

$$= \frac{\mathrm{P}\left(X \geq t + t_0\right)}{\mathrm{P}\left(X \geq t_0\right)} = \frac{1 - F\left(t + t_0\right)}{1 - F\left(t_0\right)} = \frac{e^{-\lambda\left(t + t_0\right)}}{e^{-\lambda t_0}} = e^{-\lambda t} = \mathrm{P}\left(X \geq t\right)$$

Exponencial

Exercício: Consumidores chegam ao balcão de serviços a uma taxa média de 15 por hora. Qual a probabilidade de o tempo entre a chegada de dois consumidores consecutivos ser de menos de 3 minutos? Assuma que o tempo entre chegada de consumidores tem distribuição exponencial!

Seja X = nº de chegadas por hora.

O número médio de chegadas por hora é 15, então: $X \sim Exp(\lambda = 15)$

Três minutos equivale a 0.05 horas

$$P(X < 0.05) = 1 - e^{-\lambda * 0.05} = 1 - e^{-15 * 0.05} = 0.5276$$

Exite uma probabilidade de 52.76% de que o tempo de espera entre a chegada de 2 consumidores seja menor do que 3 minutos.

Funções no Excel

- Normal padrão
  - fdp: =DIST.NORMP.N(z;falso)
  - fda: =DIST.NORMP.N(z;verdadeiro)
  - Percentil p (inversa da fda): =INV.NORMP.N(p)

- Normal
  - fdp: =DIST.NORM.N(x; média; desvio-padão; falso)
  - fda: =DIST.NORM.N(x; média; desvio-padão; verdadeiro)
  - Percentil p (inversa da fda): =INV.NORM.N(p; média; desvio-padrão)

Comandos em R

- Normal padrão
  - fdp: dnorm(x)
  - fda: pnorm(q, lower.tail = TRUE,)
  - inversa da fda: qnorm(p, lower.tail = TRUE)

- Normal com média m e desvio-padrão s
  - fdp: dnorm(x, mean = m, sd = s)
  - fda: pnorm(q, mean = m, sd = s)
  - inversa da fda: qnorm(p, mean = m, sd = s)

Resumo

Nesta parte vimos:

- Algumas distribuições continuas importantes:
  - normal, uniforme, exponencial
- Como encontrar probabilidades usando funções densidade de probabildiade e tabelas
- Quando aplicar cada distribuição