PUC-RIO
INF1036 - Probabilidade Computacional

Material 7 – Variáveis Aleatórias Contínuas

Professora - Ana Carolina Letichevsky*

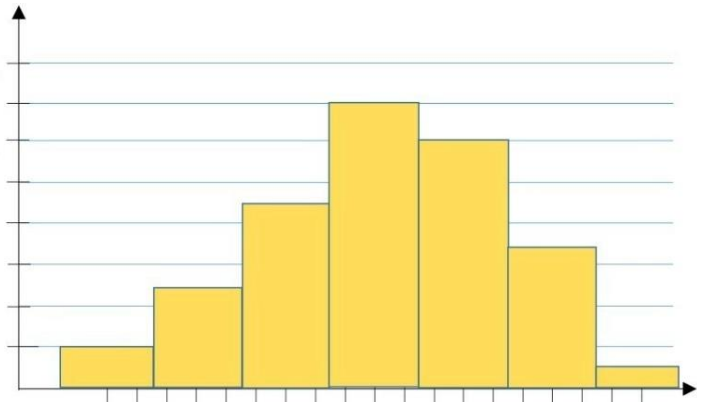
2022.1

*Material Adaptado de Professor Hélio Lopes

Distribuições discretas e contínuas

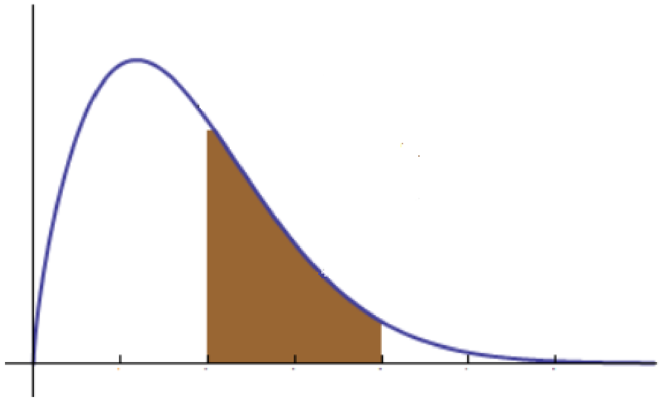
Variável aleatória discreta

Distribuição de probabilidade discreta



Variável aleatória contínua

Distribuição de probabilidade contínua



Distribuições contínuas

As **variáveis aleatórias contínuas** seguem uma natureza contínua, ou seja, elas podem assumir qualquer valor real.

As variáveis aleatórias contínuas possuem **funções de densidade de probabilidade** (**fdp**) e **distribuições acumuladas**.

A **distribuição contínua** descreve as **probabilidades** dos possíveis valores de uma variável aleatória contínua.

As **probabilidades** de **variáveis aleatórias contínuas** são definidas como a **área sob a curva** da sua **distribuição**. Assim, apenas as faixas de valores podem ter uma probabilidade diferente de zero.

A **probabilidade** de uma variável aleatória contínua assumir um determinado valor é sempre igual a zero.

Distribuição Normal

Conhecida também como **distribuição gaussiana** é, sem dúvida, a mais importante distribuição contínua.

Sua importância se deve a vários fatores, entre eles temos o **teorema central do limite**, o qual é um resultado fundamental em aplicações práticas e teóricas, pois garante que, mesmo que os dados não sejam distribuídos segundo uma **normal,** a **média dos dados converge** para uma distribuição **normal** conforme o **número de dados aumenta**.

Pode ser utilizada para modelar um grande número de fenômenos como: pressão sanguínea das pessoas; custos domésticos; tempo de vida de um equipamento, etc.

Se uma variável aleatória $\boldsymbol{X}$ tem distribuição **Normal,** então a sua função densidade de probabilidade é dada por:

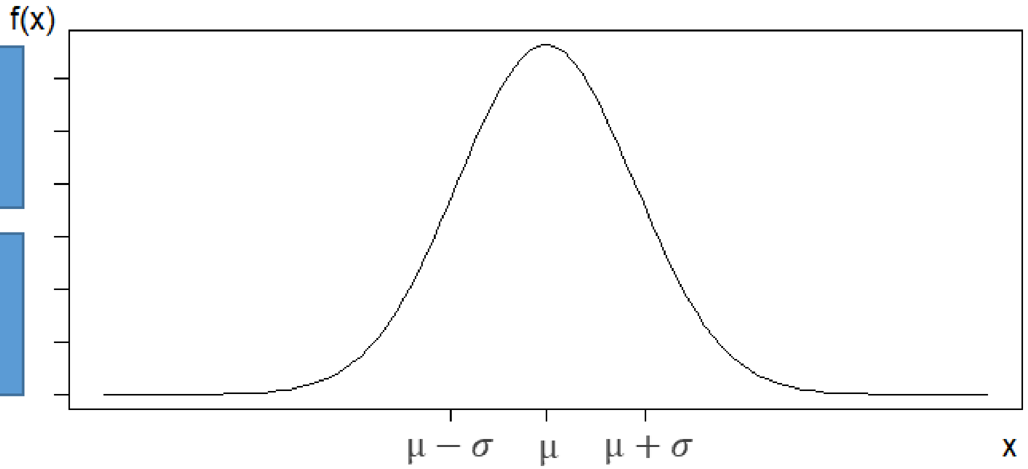
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \qquad x \in (-\infty, \infty)$$

Distribuição Normal

O gráfico da distribuição **Normal** é chamado **Curva Normal**.

A média (μ) indica a linha de simetria.

Pontos de inflexão.



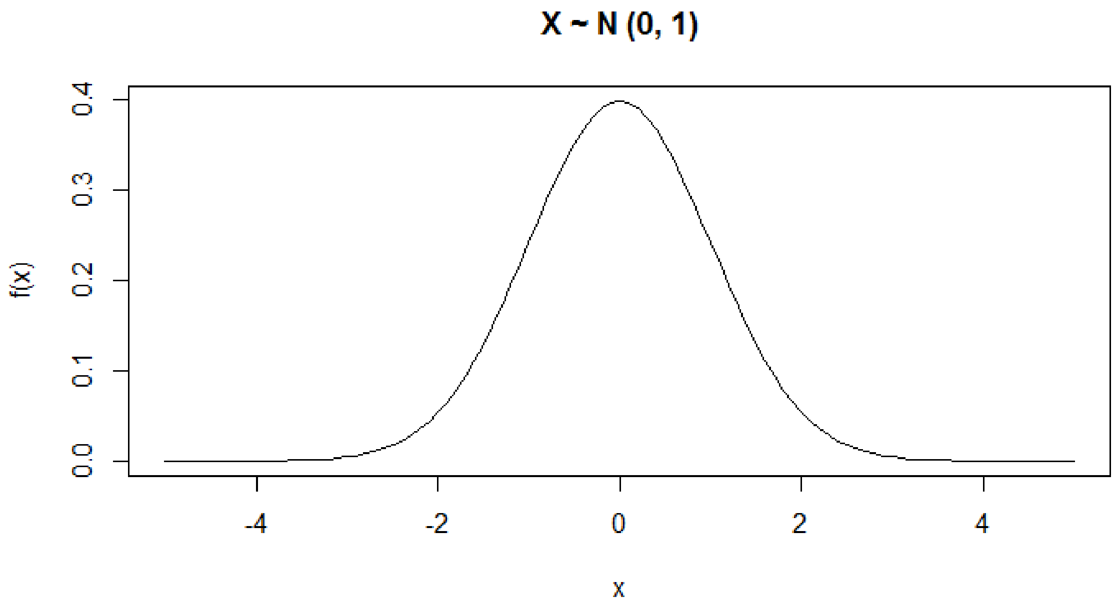
O desvio padrão (σ) indica a dispersão dos dados.

Algumas propriedades:

- A média, a mediana e a moda são iguais.
- Apresenta formato de sino e é simétrica em relação à média.
- A curva se aproxima do eixo x, mas nunca toca.
- A área total abaixo da curva é igual a 1, sendo $0{,}5$ à direita de μ e $0{,}5$ à esquerda de μ.
- Entre $\mu - \sigma$ e $\mu + \sigma$, o gráfico curva para baixo.
- À esquerda de $\mu - \sigma$ e à direita de $\mu + \sigma$, o gráfico se curva para cima.

Distribuição Normal Padrão

É a distribuição **Normal** com média 0 e desvio padrão 1.



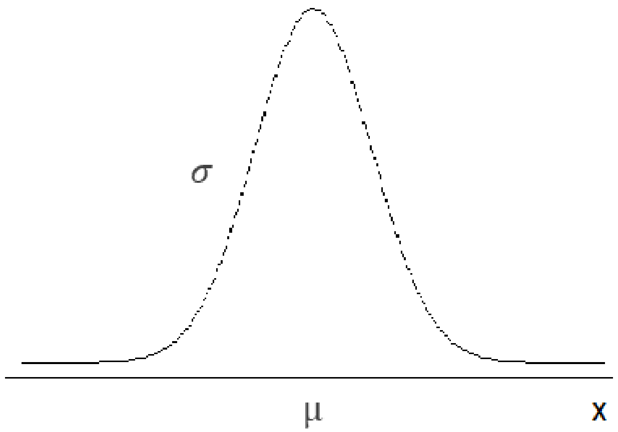
Por ser uma distribuição muito usada, existem tabelas prontas que permitem encontrar o valor da área de regiões sob a curva.

Cálculo da probabilidade em uma distribuição Normal Padrão

Se uma variável aleatória $\boldsymbol{X}$ é normalmente distribuída, para encontrar a probabilidade de $\boldsymbol{X}$ cair em um dado intervalo, basta calcular a área sob a curva normal daquele intervalo.

Transformar o valor de x em z score

Distribuição Normal Padrão



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

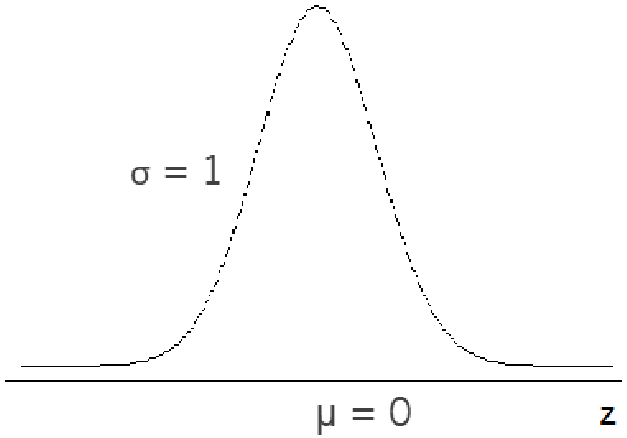


Tabela Normal Padrão fornece a área acumulada abaixo da Curva Normal Padrão.

Distribuição Normal

Exemplo 1) Uma pesquisa contratada por uma oficina mecânica indica que, para cada ida à oficina, um cliente gasta uma média de 45 minutos com um desvio padrão de 12 minutos. O tempo gasto na oficina é normalmente distribuído e representado pela variável aleatória x. Encontre o que é pedido para as seguintes situações:

a) A probabilidade de um cliente, ao entrar na oficina, demorar entre 24 e 54 minutos.

b) A probabilidade de um cliente, ao entrar na oficina, demorar mais de 39 minutos.

c) Se 15 clientes entram na oficina, quantos deles devem permanecer por mais de 39 minutos?

Distribuição Normal

Exemplo 1) Uma pesquisa contratada por uma oficina mecânica indica que, para cada ida à oficina, um cliente gasta uma média de 45 minutos com um desvio padrão de 12 minutos. O tempo gasto na oficina é normalmente distribuído e representado pela variável aleatória x. Encontre o que é pedido para as seguintes situações:

a) A probabilidade de um cliente, ao entrar na oficina, demorar entre 24 e 54 minutos.

$$z_1 = \frac{x-\mu}{\sigma} = \frac{24-45}{12} = -1{,}75 \qquad z_2 = \frac{54-45}{12} = 0{,}75$$

$$P(24 < x < 54) = P(-1{,}75 < z < 0{,}75) = 0{,}7734 - 0{,}0401 = 0{,}7333$$

b) A probabilidade de um cliente, ao entrar na oficina, demorar mais de 39 minutos.

$$z = \frac{x-\mu}{\sigma} = \frac{39-45}{12} = -0{,}5 \qquad P(x > 39) = P(z \succ -0{,}5) = 1 - 0{,}3085 = 0{,}6915$$

c) Se 15 clientes entram na oficina, quantos deles devem permanecer por mais de 39 minutos?

$$15 * P(x > 39) = 15 * 0{,}6915 = 10{,}3725 \ (\mathit{cerca\ de}\ 10\ \mathit{clientes})$$

Distribuições no R

No R, o trabalho com uma **variável aleatória** $\boldsymbol{X}$ está baseado em 4 procedimentos identificados pelas letras $\boldsymbol{p}$, $\boldsymbol{d}$, $\boldsymbol{q}$ e $\boldsymbol{r}$ que, adicionadas previamente ao código das **distribuições,** permitem o seu uso com **diferentes propósitos**:

- **rdistrib**: Gera **números aleatórios** baseados na distribuição definida. Necessita de argumentos especificando o tamanho da amostra, além dos parâmetros requeridos pela distribuição de interesse;
- **pdistrib**: Calcula a **função de probabilidade acumulada** $\boldsymbol{F(x)}$**;**
- **ddistrib**: Calcula a **distribuição de probabilidade** $\boldsymbol{f(x)}$;
- **qdistrib**: Calcula o **quantil** correspondente a uma dada probabilidade.

Distribuição Normal no R

Distribuição	Sintaxe	Opções
Normal **(norm)**	**dnorm** (x, mean=0, sd=1, opções)	**x**: ponto na distribuição **mean**: média da distribuição. **sd**: desvio padrão da distribuição. **q**: ponto na distribuição **p**: probabilidade. **n**: número de observações.
	pnorm (q, mean=0, sd=1, opções)	
	qnorm (p, mean=0, sd=1, opções)	
	rnorm (n, mean=0, sd=1)	

Distribuição Normal

Exercício 1) Uma pesquisa contratada por uma oficina mecânica indica que, para cada ida à oficina, um cliente gasta uma média de 45 minutos com um desvio padrão de 12 minutos. O tempo gasto na oficina é normalmente distribuído e representado pela variável aleatória x. Encontre o que é pedido para as seguintes situações:

a) A probabilidade de um cliente, ao entrar na oficina, demorar entre 24 e 54 minutos.

```
pnorm (54, 45, 12) - pnorm (24, 45, 12)
0.7333135
```

b) A probabilidade de um cliente, ao entrar na oficina, demorar mais de 39 minutos.

```
1 - pnorm (39, 45, 12)
0.6914625
```

c) Se 15 clientes entram na oficina, quantos deles devem permanecer por mais de 39 minutos?

```
15 * (1 - pnorm (39, 45, 12))
10.37194
```

Distribuição Normal

Exemplo 2) Para uma distribuição normal padrão, calcule:

a) A função de probabilidade acumulada $F(x)$ até -1.

```
pnorm (-1)
0.1586553
```

A probabilidade de uma variável aleatória contínua assumir um determinado valor é sempre igual a zero.

b) O valor de a tal que $P(X \leq a) = 0{,}975$.

```
qnorm (0.975)
1.959964
```

c) Uma amostra de 5 elementos.

```
rnorm (5)
-0.3247033  1.0746392 -0.1851859  0.3379355 -3.2646077
```

Distribuição Normal

Exemplo 3) Qual a probabilidade de ocorrência de um valor menor que 20 em uma distribuição normal de média 50 e desvio padrão igual a 15?

```
valor.referencia <- 20
media <- 50
desvio.padrao <- 15

pnorm (valor.referencia, media, desvio.padrao)
0.02275013
```

Calcula a distribuição acumulada: $P(X \leq x)$

Qual seria o resultado esperado se o valor de referência fosse 50?

```
pnorm (50, 50, 15)

0.5
```

Qual seria o resultado esperado se o valor de referência fosse 20 e o desvio padrão 30?

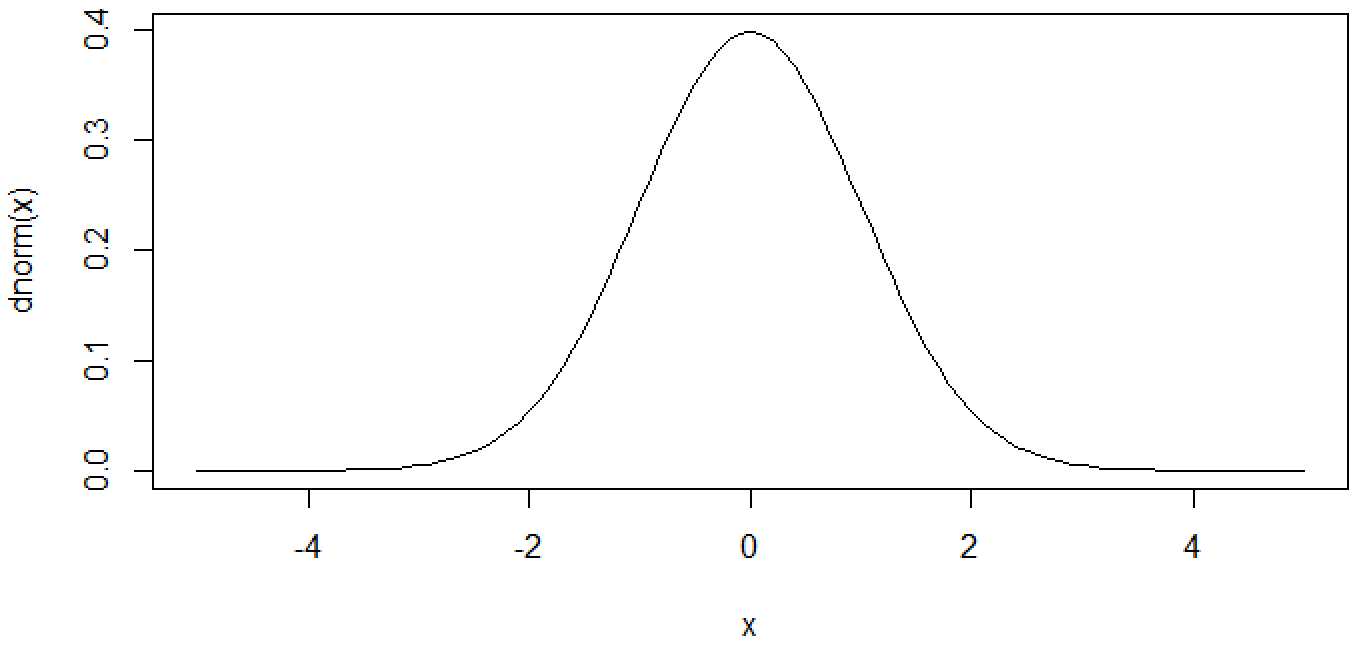
```
pnorm (20, 50, 30)

0.1586553
```

Distribuição Normal

Exemplo 4) Desenhe uma curva de uma distribuição normal padrão entre -5 e 5.

```
curve (dnorm (x), -5, 5)
```

O d em dnorm representa a função densidade de probabilidade (fdp).

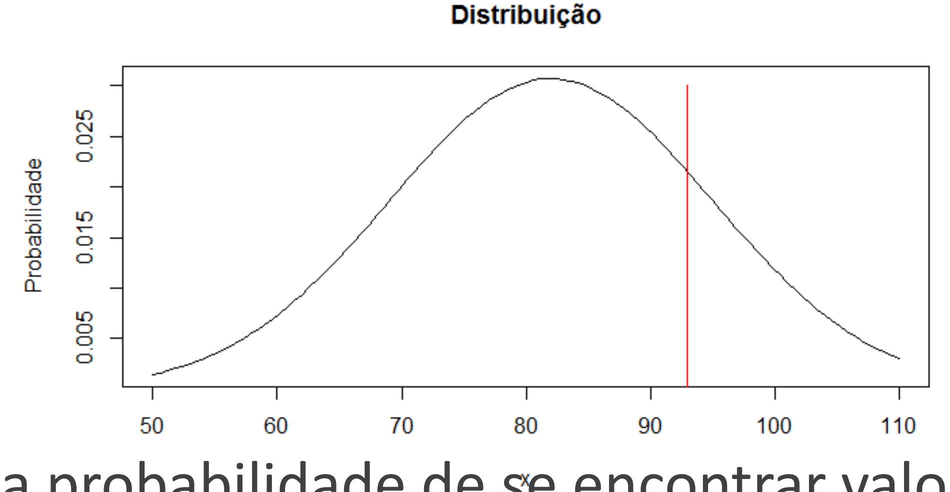


Distribuição Normal

Exercício 2) Um time de futebol coletou dados referentes ao peso de seus jogadores inscritos em um campeonato. Sabendo-se que o peso da população de jogadores segue uma distribuição normal com média 82 e desvio padrão 13, calcule:

a) A probabilidade de se encontrar um jogador com mais de 93 quilos.

```
1 - pnorm (93, 82, 13)
0.1987335
```



O valor de probabilidade encontrado corresponde exatamente à área do gráfico abaixo da curva normal e à direita da linha vermelha.

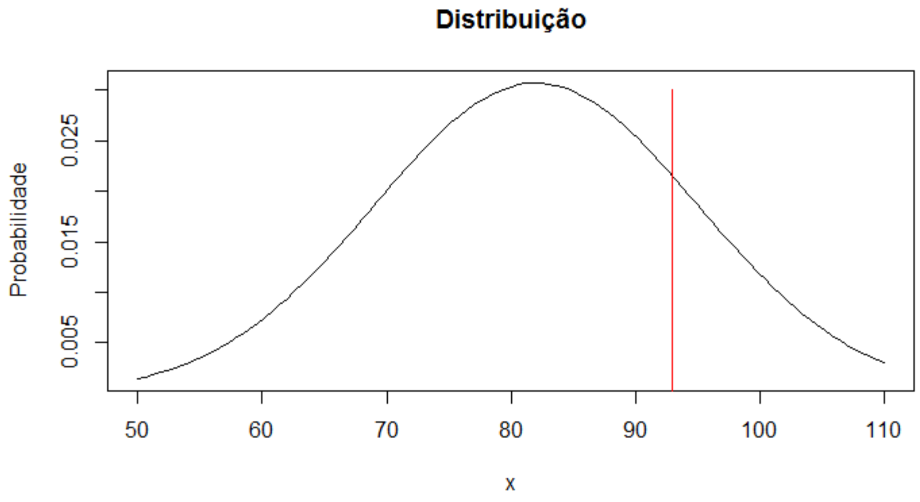
b) O valor do peso para o qual a probabilidade de se encontrar valores menores que o deste seja de 70%.

```
qnorm (0.7, 82, 13)
88.81721
```

Distribuição Normal

Exemplo 5) Plote o gráfico do exercício anterior:

```
curve (dnorm (x, 82, 13),
  50, 110, #limites do gráfico
  main = "Distribuição",
  ylab = "Probabilidade")
  lines (c (93, 93), #início e fim da linha em relação ao eixo x
  c (0, 0.03), #início e fim da linha em relação ao eixo y
  col = 2) #cor vermelha
```



Probabilidade e integrais – Distribuição normal

A probabilidade de um evento em uma distribuição contínua é uma área sob a curva da distribuição. Vamos reforçar esta ideia com um exemplo.

Seja X uma variável aleatória com distribuição $X \sim N(\mu, \sigma^2) = N(100, 100)$. Para calcular a probabilidade $P(X \leq 95)$, podemos usar o comando:

```
pnorm (95, 100, 10)
0.3085375
```

Vamos agora ver uma outra forma de resolver usando integração numérica, lembrando que a Normal tem a função de densidade dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \qquad x \in (-\infty, \infty)$$

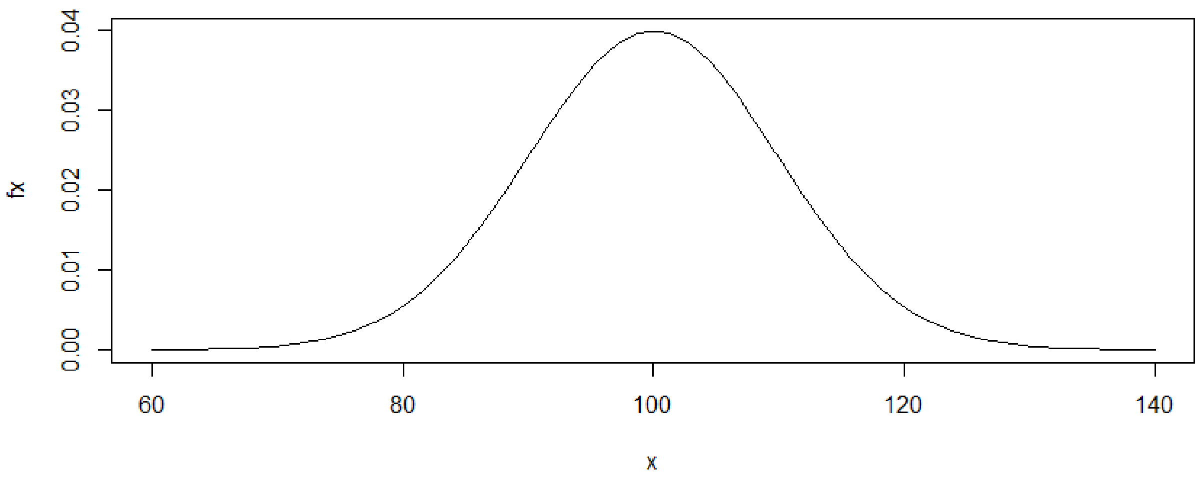
Probabilidade e integrais – Distribuição normal

Vamos escrever uma função em R que represente $f(x)$ para a distribuição Normal.

```
funcao.normal <- function (x) {
   fx <- (1 / sqrt (2 * pi * 100)) * exp ((-1 / 200) * (x - 100)^2)
   return (fx)
}
```

Graficamente temos:

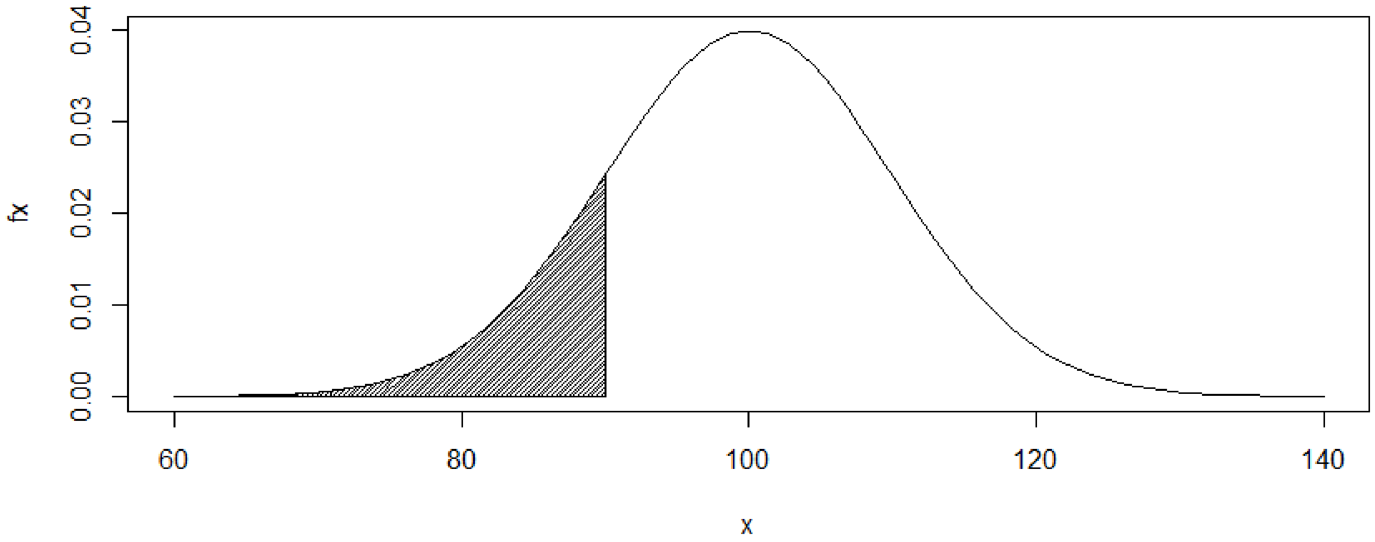
```
x <- seq (60, 140, length.out = 100)
fx <- funcao.normal (x)
plot (x, fx, type = "l")
```



Probabilidade e integrais – Distribuição normal

Para marcar no gráfico a área que corresponde à probabilidade pedida, criamos um polígono.

```
x <- seq (60, 140, length.out = 100)
fx <- funcao.normal (x)
plot (x, fx, type = "l")
vx <- c (60, 60, x[x < 90], 90, 90)
vy <- c (0, funcao.normal(60), fx[x < 90], funcao.normal(90), 0)
plot (x, fx, type = "l")
polygon (vx, vy, dens = 50)
```



Probabilidade e integrais – Distribuição normal

Para calcular a área pedida sem usar a função **pnorm(),** podemos usar a função de integração numérica **integrate()**, que diferentemente da **pnorm()** reporta ainda o erro de aproximação numérica.

```
pnorm (90, mean = 100, sd = 10)
0.1586553

integrate (funcao.normal, -Inf, 90)
0.1586553 with absolute error < 9.5e-05
```

Para $P(90 \leq X \leq 105)$ e $P(X > 90)$ teríamos:

```
pnorm(105,100,10) - pnorm(90,100,10)
0.5328072

1 - pnorm (90, 100, 10)
0.8413447
```

```
integrate (funcao.normal, 90, 105)
0.5328072 with absolute error < 5.9e-15

integrate (funcao.normal, 90, +Inf)
0.8413447 with absolute error < 1.1e-08
```

Distribuição t de Student

Uma variável aleatória contínua X tem distribuição **t de Student** com ν graus de liberdade se sua função densidade de probabilidade é dada por:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}, \qquad x \in (-\infty, \infty)$$

Grau de liberdade: para um determinado conjunto de dados correspondente ao número de valores que podem variar após terem sido impostas restrições aos valores.

Grau de liberdade também é utilizado para caracterizar uma distribuição específica em certas famílias de distribuição, como é o caso, de distribuição **t de Student**.

Tal como a distribuição **Normal**, tem a função de densidade em forma de sino e é simétrica.

Quanto maior o número de graus de liberdade, mais a distribuição **t de Student** se aproxima da distribuição **Normal**.

É muito utilizada em testes estatísticos.

Distribuições no R

No R, o trabalho com uma **variável aleatória** $\boldsymbol{X}$ está baseado em 4 procedimentos identificados pelas letras $\boldsymbol{p}$, $\boldsymbol{d}$, $\boldsymbol{q}$ e $\boldsymbol{r}$ que, adicionadas previamente ao código das **distribuições,** permitem o seu uso com **diferentes propósitos**:

- **rdistrib**: Gera **números aleatórios** baseados na distribuição definida. Necessita de argumentos especificando o tamanho da amostra, além dos parâmetros requeridos pela distribuição de interesse;
- **pdistrib**: Calcula a **função de probabilidade acumulada** $\boldsymbol{F(x)}$**;**
- **ddistrib**: Calcula a **distribuição de probabilidade** $\boldsymbol{f(x)}$;
- **qdistrib**: Calcula o **quantil** correspondente a uma dada probabilidade.

Distribuição t de Student no R

Distribuição	Sintaxe	Opções
t-Student (**t**)	**dt** (x, df, ncp=0, opções)	**x**: ponto na distribuição. **df**: grau de liberdade. **ncp**: parâmetro de não centralidade. **q**: ponto na distribuição **p**: probabilidade. **n**: número de observações.
	pt (q, df, ncp=0, opções)	
	qt (p, df, opções)	
	rt (n, df)	

Distribuição t de Student

Exercício 1) Determine o valor x de uma distribuição t de Student com 12 graus de liberdade, que garante, com probabilidade de 99% de um valor aleatório, ser abaixo de x.

```
qt (0.99, 12)
2.680998
```

Distribuição Exponencial

A variável aleatória X tem distribuição **Exponencial** com parâmetro λ, $\lambda > 0$ se tiver função densidade de probabilidade dada por:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & se\ x \geq 0 \\ 0, & se\ x < 0 \end{cases}, \qquad onde: \ E[x] = 1/\lambda \quad \mathrm{V}[x] = 1/\lambda^2$$

A variável X é igual a distância entre contagens sucessivas de um processo **Poisson**.

As relações entre as distribuições **Exponencial** e **Poisson** podem ser associadas a um processo estatístico chamado **processo de Poisson**.

Distribuições no R

No R, o trabalho com uma **variável aleatória** $\boldsymbol{X}$ está baseado em 4 procedimentos identificados pelas letras $\boldsymbol{p}$, $\boldsymbol{d}$, $\boldsymbol{q}$ e $\boldsymbol{r}$ que, adicionadas previamente ao código das **distribuições,** permitem o seu uso com **diferentes propósitos**:

- **rdistrib**: Gera **números aleatórios** baseados na distribuição definida. Necessita de argumentos especificando o tamanho da amostra, além dos parâmetros requeridos pela distribuição de interesse;

- **pdistrib**: Calcula a **função de probabilidade acumulada** $\boldsymbol{F(x)}$**;**

- **ddistrib**: Calcula a **distribuição de probabilidade** $\boldsymbol{f(x)}$;

- **qdistrib**: Calcula o **quantil** correspondente a uma dada probabilidade.

Distribuição Exponencial no R

Distribuição	Sintaxe	Opções
Exponencial (**exp**)	**dexp** (x, rate=1, opções) **pexp** (q, rate=1, opções) **qexp** (p, rate=1, opções) **rexp** (n, rate = 1)	**x**: ponto na distribuição. **rate (λ)**: parâmetro da distribuição. **q**: ponto na distribuição **p**: probabilidade. **n**: número de observações.

Distribuição Exponencial

Exercício 1) Uma ferramenta produzida em uma fábrica apresenta uma vida média de 90 horas. Considerando o comportamento segundo a distribuição exponencial, qual a probabilidade de essa ferramenta durar mais de 100 horas?

$$P(X > 100) = 1 - P(X \leq 100)$$

```
1 - pexp (100, rate = 1/90)
0.329193
```

A vida média é o parâmetro μ, que é o inverso de λ, portanto, λ = 1/90.

Distribuição Exponencial

Exercício 2) Uma fábrica utiliza dois métodos para a produção de sensores eletrônicos. 80% dos sensores são produzidos pelo método A, e os demais, pelo método B. A duração do sensor depende do método pelo qual ele foi produzido: os produzidos pelo método A seguem uma distribuição exponencial com parâmetro $1/90$; e os produzidos pelo método B seguem uma exponencial de parâmetro $1/110$. Qual a probabilidade de que, se escolhermos um sensor ao acaso, ele dure mais de 100 horas?

Considere os eventos C (um sensor durar mais de 100 horas), A (o sensor ter sido produzido pelo método A) e B (o sensor ter sido produzido pelo método B).

$$P(C) = P(A)P(C|A) + P(B)P(C|B) = 0{,}80P(X_A > 100) + 0{,}20P(X_B > 100)$$

```
0.8*(1 - pexp(100, rate = 1/90)) + 0.2*(1 - pexp(100, rate = 1/110))
0.3439325
```

Probabilidades e integrais

A função de densidade de probabilidade da distribuição exponencial com parâmetro λ é dada por:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & se\ x \geq 0 \\ 0, & se\ x < 0 \end{cases}$$

Seja X uma variável aleatória com distribuição exponencial de parâmetro λ = 1/600. Calcular a probabilidade $P(X \geq 300)$,usando a função de integração numérica **integrate()** e validando o resultado com as funções de distribuição do R.

```
funcao.exponencial <- function (x) {
   lambda <- 1/600
   fx <- ifelse (x < 0, 0, (lambda) * exp(-x*lambda))
   return (fx)
}
```

```
integrate (funcao.exponencial, 300, +Inf)

0.6065307 with absolute error < 7.1e-05
```

```
1 - pexp(300, rate=1/600)

0.6065307
```

Distribuição Weibull

É usada para modelar o tempo até uma falha de sistemas físicos. Os parâmetros da distribuição fornecem grande flexibilidade para modelar sistemas em que o número de falhas aumenta com o tempo, diminui com o tempo, ou permanece constante com o tempo.

Uma variável aleatória X tem distribuição **Weibull** se tiver função densidade de probabilidade dada por:

$$f(x) = \begin{cases} \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-\left(\frac{x}{\delta}\right)^{\beta}}, se\ x \geq 0 \\ 0, se\ x < 0 \end{cases}$$

Sua função de distribuição acumulada é dada por:

$$F(x) = \begin{cases} 1 - e^{-\left(\frac{x}{\delta}\right)^{\beta}}, se\ x \geq 0 \\ 0, se\ x < 0 \end{cases}$$

Distribuições no R

No R, o trabalho com uma **variável aleatória** $\boldsymbol{X}$ está baseado em 4 procedimentos identificados pelas letras $\boldsymbol{p}$, $\boldsymbol{d}$, $\boldsymbol{q}$ e $\boldsymbol{r}$ que, adicionadas previamente ao código das **distribuições,** permitem o seu uso com **diferentes propósitos**:

- **rdistrib**: Gera **números aleatórios** baseados na distribuição definida. Necessita de argumentos especificando o tamanho da amostra, além dos parâmetros requeridos pela distribuição de interesse;

- **pdistrib**: Calcula a **função de probabilidade acumulada** $\boldsymbol{F(x)}$**;**

- **ddistrib**: Calcula a **distribuição de probabilidade** $\boldsymbol{f(x)}$;

- **qdistrib**: Calcula o **quantil** correspondente a uma dada probabilidade.

Distribuição Weibull no R

Distribuição	Sintaxe	Opções
Weibull **(weibull)**	**dweibull** (x, shape, scale = 1, opções)	**x**: ponto na distribuição. **shape**: parâmetro da distribuição. **scale**: parâmetro da distribuição. **q**: ponto na distribuição **p**: probabilidade. **n**: número de observações.
	pweibull (q, shape, scale = 1, opções)	
	qweibull (p, shape, scale = 1, opções)	
	rweibull (n, shape, scale = 1)	

Distribuição Weibull

Exercício 1) O tempo de falha (em horas) de uma determinada peça em um tipo de freio de carro é satisfatoriamente modelado como uma variável aleatória de Weibull com $\beta = 1/2$ e $\delta = 6000$.

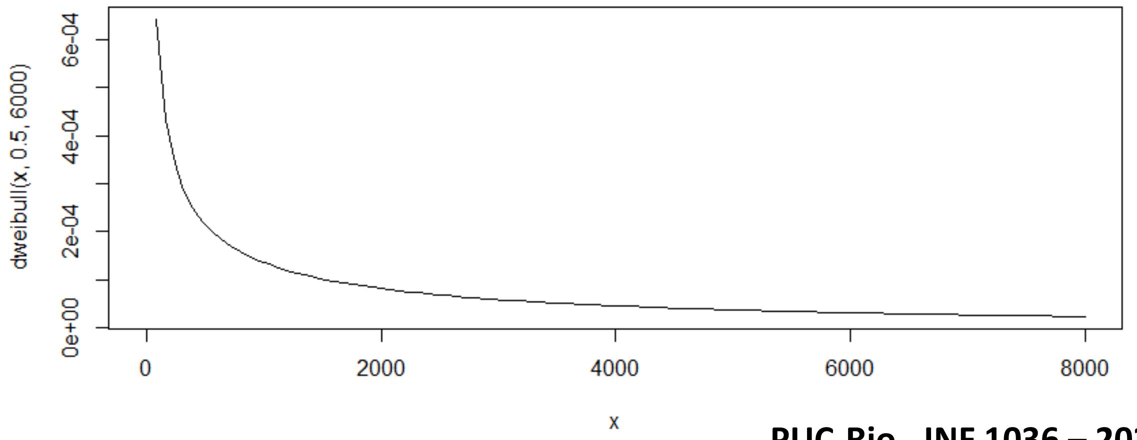
a) Determine a probabilidade de uma peça durar no mínimo 6500 h.

```
1 - pweibull (6500, 0.5, 6000)
0.3531604
```

b) Verifique a probabilidade de uma peça ter problemas nas primeiras 6500 h e plote o resultado.

```
pweibull (6500, 0.5, 6000)
0.6468396
```

```
x <- 0:8000
curve (dweibull (x, 0.5, 6000), 0, 8000)
```



Distribuição Gama

- É uma das distribuições mais gerais, pois diversas distribuições são casos particulares dela como, por exemplo, a exponencial, a qui-quadrado, entre outras. Essa distribuição tem como suas principais aplicações a análise de tempo de vida de produtos.

- Uma variável aleatória **X** tem distribuição **Gama** com parâmetros $\alpha > 0$ (parâmetro de forma) e $\beta > 0$ (parâmetro de taxa) se tiver função densidade de probabilidade dada por:

$$f(x) = \begin{cases} \dfrac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, se\ x \geq 0 \\ 0, se\ x < 0 \end{cases}$$

- Ela também pode ser definida por um parâmetro de escala $\theta > 0$ equivalente ao inverso do parâmetro de taxa β.

- A média da distribuição é dada por $\frac{\alpha}{\beta} = \alpha\theta$, e a sua variância é $\frac{\alpha}{\beta^2} = \alpha\theta^2$.

Distribuições no R

No R, o trabalho com uma **variável aleatória** $\boldsymbol{X}$ está baseado em 4 procedimentos identificados pelas letras $\boldsymbol{p}$, $\boldsymbol{d}$, $\boldsymbol{q}$ e $\boldsymbol{r}$ que, adicionadas previamente ao código das **distribuições,** permitem o seu uso com **diferentes propósitos**:

- **rdistrib**: Gera **números aleatórios** baseados na distribuição definida. Necessita de argumentos especificando o tamanho da amostra, além dos parâmetros requeridos pela distribuição de interesse;
- **pdistrib**: Calcula a **função de probabilidade acumulada** $\boldsymbol{F(x)}$**;**
- **ddistrib**: Calcula a **distribuição de probabilidade** $\boldsymbol{f(x)}$;
- **qdistrib**: Calcula o **quantil** correspondente a uma dada probabilidade.

Distribuição Gama no R

Distribuição	**Sintaxe**	**Opções**
Gama **(gamma)** E(X) = shape/rate Var(X) = shape/rate^2	**dgamma**(x, shape, rate = 1, opções)	**x**: ponto na distribuição. **shape**: parâmetro da distribuição. **rate**: parâmetro da distribuição. **q**: ponto na distribuição **p**: probabilidade. **n**: número de observações.
	pgamma(q, shape, rate = 1, opções)	
	qgamma(p, shape, rate = 1, opções)	
	rgamma(n, shape, rate = 1)	

Distribuição Gama

Exercício 1) A renda doméstica mensal em uma cidade é uma variável aleatória com distribuição Gama com média R$ 2.000 e desvio padrão R$ 400.

a) Calcule os parâmetros α, β e θ desta densidade.

$$\alpha\theta = 2000 \qquad \sqrt{\alpha\theta^2} = 400 \qquad \Rightarrow \qquad \alpha = 25 \quad \theta = 80 \qquad \beta = \frac{1}{80}$$

b) Calcule a probabilidade da renda média mensal de um domicílio exceder R$ 2.300.

```
1 - pgamma (2300, shape = 25, rate = 1/80)
0.2172807
```

c) Qual a mediana desta distribuição Gama?

```
qgamma (0.5, shape = 25, rate = 1/80)
1973.397
```

Distribuição Qui-quadrada

Pode ser interpretada de duas formas, como um caso particular da distribuição gamma ou como sendo a soma de normais padronizadas ao quadrado.

Uma variável aleatória contínua X tem distribuição qui-quadrado com ν graus de liberdade se sua função densidade for dada por:

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} x^{\left(\frac{\nu}{2}\right)-1} e^{\left(-\frac{x}{2}\right)}, se\ x \geq 0 \\ 0, se\ x < 0 \end{cases}$$

Distribuições no R

No R, o trabalho com uma **variável aleatória** $\boldsymbol{X}$ está baseado em 4 procedimentos identificados pelas letras $\boldsymbol{p}$, $\boldsymbol{d}$, $\boldsymbol{q}$ e $\boldsymbol{r}$ que, adicionadas previamente ao código das **distribuições,** permitem o seu uso com **diferentes propósitos**:

- **rdistrib**: Gera **números aleatórios** baseados na distribuição definida. Necessita de argumentos especificando o tamanho da amostra, além dos parâmetros requeridos pela distribuição de interesse;

- **pdistrib**: Calcula a **função de probabilidade acumulada** $\boldsymbol{F(x)}$**;**

- **ddistrib**: Calcula a **distribuição de probabilidade** $\boldsymbol{f(x)}$;

- **qdistrib**: Calcula o **quantil** correspondente a uma dada probabilidade.

Distribuição Qui-quadrada no R

Distribuição	Sintaxe	Opções
Qui-quadrada **(dchisq)**	**dchisq**(x, df, ncp=0, opções) **pchisq**(q, df, ncp=0, opções) **qchisq**(p, df, ncp=0, opções) **rchisq**(n, df, ncp=0)	**x**: ponto na distribuição. **df**: graus de liberdade. **ncp**: parâmetro de não centralidade. **q**: ponto na distribuição. **p**: probabilidade. **n**: número de observações.

Distribuição Qui-quadrada

Exercício 1) Suponha agora que X segue uma distribuição qui-quadrada com 10 graus de liberdade. Encontre x_1 e x_2 tais que $P(x_1 \leq X \leq x_2) = 0{,}95$.

Pelo exposto, temos que:

$$P(x_1 \leq X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = 0{,}95$$

Suponha que $P(X \leq x_2) = 0{,}98$, o que implica $P(X \leq x_1) = 0{,}03$ (outros valores poderiam ter sido escolhidos), logo:

$$P(x_1 \leq X \leq x_2) = 0{,}98 - 0{,}03 = 0{,}95$$

Os valores de $\boldsymbol{x_2}$ e $\boldsymbol{x_1}$.

```
qchisq (0.98, df = 10)
21.16077
```

```
qchisq (0.03, df = 10)
3.412069
```

$$P(x_1 \leq X \leq x_2) = P(X \leq 21.16077) - P(X \leq 3.412069) = 0{,}95$$

Para testar, podemos usar:

```
pchisq (21.16077, df=10) - pchisq (3.412069, df=10)
0.95
```

Teorema Central do Limite

- Considerado, pela sua relevância na troca e em aplicações, um teorema básico central de probabilidade.

- Se refere à convergência de soma de variáveis aleatórias quantitativas para uma distribuição normal.

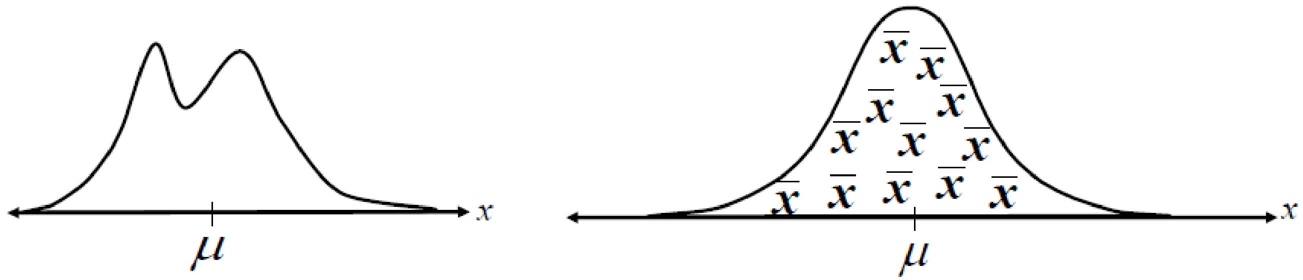
  - Suponha $x_1, \ldots x_n$ variáveis aleatórias independentes com a mesma distribuição de média μ e variável σ^2 limites. O que acontece com distribuição de soma $\mathrm{x} = x_1 + \ldots + x_n$ quando n cresce?

  - Conforme n aumenta a distribuição de $\mathrm{x} = x_1 + \ldots + x_n$, se aproxima de uma

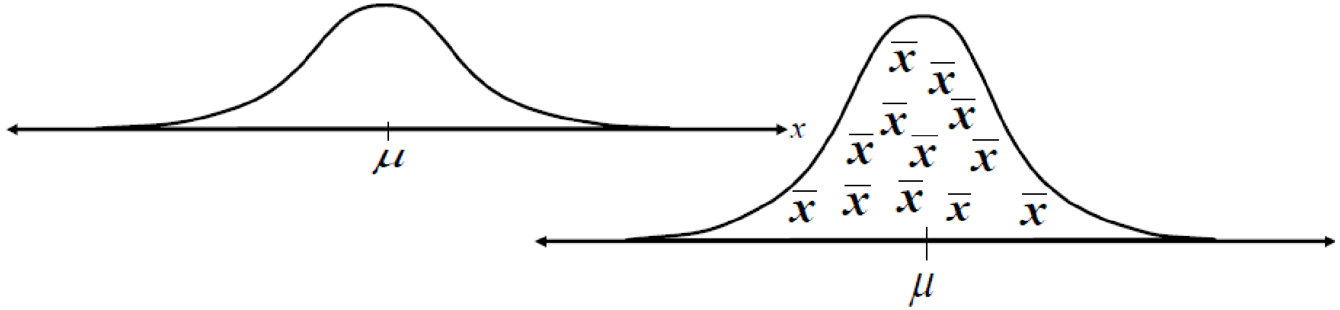
    distribuição $\lambda \sim N(\mu_x, \sigma^2 x)$, sendo $\mu_x = \frac{n}{\lambda}$ e $\sigma^2{}_x = \frac{n}{\lambda^2}$.

Teorema Central do Limite

Se amostras de tamanho $\boldsymbol{n} > \mathbf{30}$ são tiradas de qualquer população de média μ e desvio padrão σ, a distribuição da média amostral aproxima-se de uma distribuição normal. Quanto maior o tamanho da amostragem, melhor a aproximação.



Se a própria população é normalmente distribuída, a distribuição das médias amostrais é normalmente distribuída para **qualquer** tamanho de amostragem $\boldsymbol{n}$.



Teorema Central do Limite

Exemplo 1) As contas de energia dos moradores de uma cidade apresentam média de \$81 e um desvio padrão de \$16. Amostragens aleatórias de 49 contas são selecionadas dessa população, e a média de cada amostragem é determinada. Encontre a média e o erro padrão da média da distribuição amostral. Então, esboce um gráfico da distribuição das médias amostrais.

A média da distribuição de amostras é igual à média da população.

$$\mu_{\bar{x}} = \mu = 81$$

O erro padrão da média é igual ao desvio padrão populacional dividido pela raiz quadrada de n.

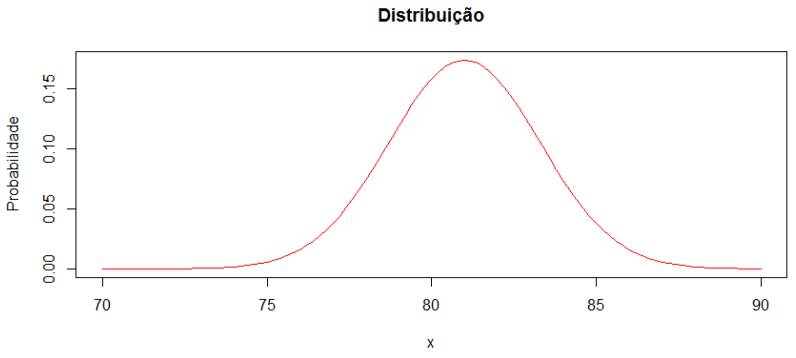
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{49}} = \frac{16}{7} = 2{,}29$$

Já que o tamanho da amostra é maior que 30, a distribuição das amostras pode ser aproximada por uma distribuição normal com:

$$\mu_{\bar{x}} = 81$$

$$\sigma_{\bar{x}} = 2{,}29$$

```
curve (dnorm (x, 81, 2.29), 70, 90,
       main = "Distribuição",
       ylab = "Probabilidade", col = 2)
```



Exemplo

- A média da distribuição de amostras é igual à média da população

$$\mu_{\bar{x}} = \mu = 81$$

- O erro padrão da média é igual ao desvio padrão populacional, dividido pela raiz quadrada de **n**.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{49}} = \frac{16}{7} = 2{,}29$$

- Já que o tamanho da amostra é maior que 30, a distribuição das amostras pode ser aproximada por uma distribuição normal com:

$$\mu_{\bar{x}} = 81$$

$$\sigma_{\bar{x}} = 2{,}29$$

```
curve (dnorm (x, 81, 2.29),
      70, 90,
      main = "Distribuição",
      ylab = "Probabilidade",
      col = 2)
```