



Introdução à datascience com R

Cleuton Sampaio

Lição 5: Inferência

Inferência estatística

Inferência estatística é um procedimento para inferir parâmetros de uma população a partir de estatísticas amostrais.

- **População:** o conjunto de todos os elementos ou resultados sob investigação;
- **Parâmetro:** Uma característica da população, exemplos: Média e variância;
- **Amostra:** Uma parte ou extrato da população para fins de estudos estatísticos;
- **Estatística:** Uma variável aleatória retirada de uma amostra, com fins de estimativa de um parâmetro da população correspondente.

Exemplos

Em uma escola, a média final das notas dos alunos do quinto ano do ensino fundamental do ano de 2015 foi de 6,50, com desvio padrão de 2,5. É uma população, com os parâmetros:

- μ (média): 6,50;
- σ (desvio padrão): 2,5.

É muito difícil termos esta situação onde conhecemos exatamente os parâmetros da população. Vamos supor uma situação interessante...

Em 2016, partiu-se do suposto que a média final dos alunos será de 6,50, com desvio padrão de 2,5, como em 2015. Como podemos confirmar isso?

Podemos pegar amostras e conferir suas estatísticas. Por exemplo, pegamos 10 alunos da turma 1, 10 alunos da turma 2 e 10 alunos da turma 3 e criamos uma amostra, com as estatísticas:

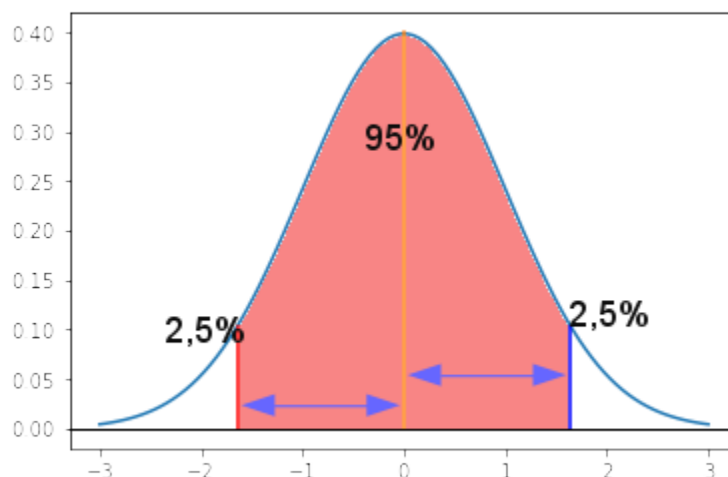
- \bar{x} (média amostral): 5,25;
- s (desvio padrão amostral): 1,8;

A pergunta a ser tratada pela inferência estatística é:

O quanto a estatística da média se aproxima da média da população?

Intervalo de confiança

De acordo com o Teorema do Limite Central, vamos assumir que a distribuição de probabilidades de determinada variável aleatória seja normal, então, o intervalo de confiança é dado pelos limites do **nível de confiança**:



Em um nível de confiança de 95%, temos esta certeza de que a média da população estará na área delimitada na figura, com 5% de chance de estar fora desta área, seja na esquerda ou na direita, com 2,5% de chances cada uma.

Margem de erro

Você deve ouvir falar muito em "margem de erro", pois os jornalistas falam constantemente sobre isto em suas notícias. Reveja a figura anterior. Temos a expectativa da média, que é a linha central (laranja) e a margem de erro, para cima ou para baixo, demarcada pelas setas azuis.

Procedimentos para inferir a média

1) Determinar o tipo de distribuição a ser utilizada:

- Conhece a variância? O tamanho da amostra é maior que 30? Use a distribuição normal (estatística Z);
- Use a distribuição T de Student (estatística T);

2) Escolha o nível de confiança:

- 99%: O valor do parâmetro está contido em 99% dos intervalos de confiança das amostras;
- 95%: O valor do parâmetro está contido em 99% dos intervalos de confiança das amostras;

A contrapartida do nível de confiança é o nível de significância. Se usamos 95%, existe um nível de significância de 5%, mas veremos isso daqui a pouco.

3) Calcule o valor da estatística e da margem de erro:

- A média da amostra;
- Calcule o valor dos limites do intervalo de confiança;
- Calcule a estatística (média) da amostra;
- Calcule o valor da margem de erro, considerando o valor do limite, a média amostral, o desvio padrão e o tamanho da amostra.

Exemplo:

O script está no repositório, dentro da pasta "lesson5/inferencia.R".

Em determinada cidade, a prefeitura disse que as famílias ali residentes teriam, em média, 2 filhos em idade escolar.

Precisamos fazer um estudo estatístico, e utilizar a inferência para estimar o parâmetro "média" da quantidade de filhos em idade escolar por família residente na cidade. Vamos dizer que coletamos dados de 50 residências, espalhadas igualmente pela cidade:

```
filhos <- c(3, 2, 2, 2, 1, 2, 2, 2, 2, 3, 1, 2, 1, 2, 2, 0, 1, 2, 2, 1, 2, 1, 1,
           2, 2, 3, 2, 1, 2, 3, 4, 1, 1, 2, 1, 1, 1, 2, 3, 0, 2, 2, 1, 2, 3, 3,
           2, 2, 3, 2)
```

Vamos calcular as estatísticas básicas:

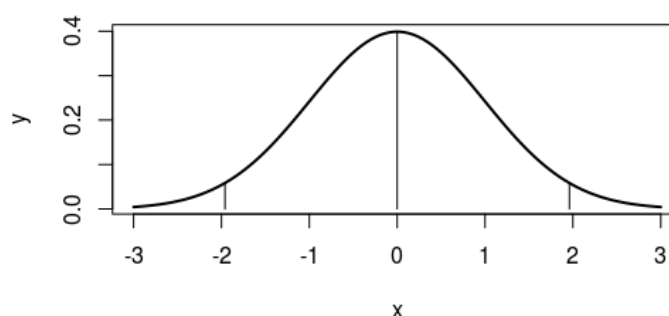
```
[1] "Média: 1.84"
[1] "Desvio: 0.817162836918577"
[1] "Assimetria: 0.072175428198334"
[1] "Curtose: 3.09005206807707"
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00   1.00   2.00   1.84   2.00   4.00
```

Para inferirmos um parâmetro da população, como a média, precisaremos supor uma distribuição de probabilidades, para utilizarmos como modelo estatístico da população. De acordo com o Teorema Central do Limite, em amostras grandes a distribuição de probabilidades se aproxima da normal. Podemos usar a distribuição normal se soubermos a variância da população, o que geralmente desconhecemos, então, usamos a distribuição T de Student.

Para simplificar, neste primeiro exemplo, vou usar a distribuição normal, porém, como a variância da população é desconhecida, o mais correto seria utilizar a distribuição T de Student.

Agora, vamos estabelecer um intervalo de confiança para a estimativa da média. Podemos usar 90%, 95% ou 99%, por exemplo. Vou optar por 95%. Então, eu tenho que calcular a área na distribuição normal, onde a média deixa de ser válida, que seria: 0,025, para ambos os lados.

Preciso encontrar os pontos no eixo x (z-scores) que delimitam essas áreas. Eles seriam: -1,95996398454 e 1,95996398454.



Temos que "pular" 25% da esquerda e ir até 97,5%, deixando os últimos 25% livres. Estes dois pontos delimitam o intervalo de confiança de 95%. Para outros níveis de confiança use:

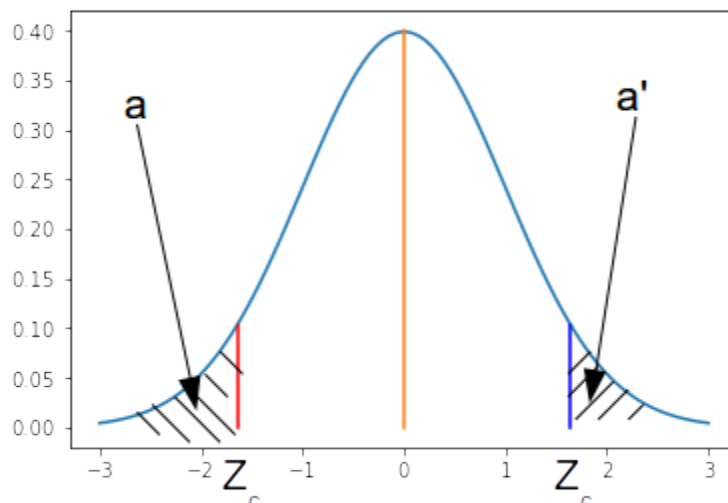
- 90%: Pulamos 5% (0,05) da esquerda e pegamos até 95% (0,95);
- 99%: Pulamos meio por cento (0,005) e pegamos até 99,05% (0,995);

Intervalo de confiança

Temos um parâmetro da população, por exemplo, a média. Não sabemos a média real da população, então, como podemos estimar estatisticamente?

Primeiro, selecionamos uma amostra. Depois, estabelecemos o **nível de confiança** que queremos ter. Este nível indica a probabilidade de que o valor real do parâmetro esteja dentro de um intervalo que calcularemos.

Os níveis de confiança mais utilizados são: 90%, 95% e 99%. Representam a área (debaixo do "sino" da distribuição) na qual o parâmetro deve ser encontrado.



Na figura acima, vemos uma distribuição normal padrão (média zero e desvio padrão 1), e vemos duas áreas (a e a'), localizadas nas pontas extremas. Se a média estiver dentro de alguma dessas áreas, então está fora do intervalo de confiança. Esperamos que a média esteja na área central da distribuição.

Para calcular os pontos no eixo "x", que são conhecidos como "z-scores" (desvios padrões a partir da média), usamos uma tabela de distribuição normal.

Vamos supor que queiramos 95% de confiança, então, temos que calcular as áreas a e a', lembrando que a área total sob o "sino" da distribuição tem o valor de 1. Se queremos calcular a área que corresponde a 95% de probabilidade (no meio) podemos pegar as áreas dos extremos. Cada área de cada extremo é calculada subtraindo-se 95% de 1, como temos que dividir em metades, usamos a fórmula:

$$a = \frac{(1 - \text{nível})}{2}$$

Calculando, temos a e $a' = 0,025$ cada uma. Para saber os pontos Z_c (os z-scores) temos duas opções:

1. Buscamos a área em uma tabela da distribuição normal padrão (existem várias na internet);
2. Calculamos com R.

Considerando um intervalo de confiança de 95%, podemos obter os valores críticos de Z com uma função R.

Como vamos utilizar R, é melhor irmos nos acostumando. Eis como calcular diretamente o z-score:

```
qnorm(0.025)
```

```
qnorm(0.975)
```

Os pontos Z_c , para 95%, serão, respectivamente: -1,95996398454 e 1,95996398454.

Margem de erro

Agora, precisamos calcular a margem de erro. Esta margem é correspondente ao intervalo de confiança que desejamos e representa o erro amostral. Para calculá-la podemos calcular com a fórmula:

$$E = \frac{Z_c * \sigma}{\sqrt{n}}$$

Não temos o desvio padrão da população (sigma), mas, para simplificar, vamos usar o da própria amostra (na verdade, estamos usando o "erro padrão", em vez do desvio padrão).

Por exemplo, sabendo que $Z_c = 1,960$, $s = 0.82$ e $n = 50$, a margem de erro é: 0,23.

Finalmente, podemos calcular o limite inferior e o limite superior do nosso intervalo de confiança: média \pm margem de erro.

Estatística T

Quando a variância da população é desconhecida e usamos a variância amostral, ou quando temos um tamanho de amostra pequeno (menor que 30 elementos), usamos a distribuição **T de Student** como aproximação da distribuição de probabilidades do fenômeno.

A distribuição T de Student introduz um novo conceito: Graus de liberdade, que é a quantidade de observações independentes de um fenômeno contidas em uma amostra. É calculado como o tamanho da amostra menos 1.

Podemos seguir o mesmo raciocínio que usamos com a distribuição normal, só que agora, os "z-scores" serão chamados de "t-scores" e temos que considerar os graus de liberdade da amostra.

Vamos imaginar o mesmo exemplo anterior (da quantidade de filhos por família em idade escolar), só que diminuindo a amostra para 12 famílias, será que a média continuará sendo considerada válida?

- Tamanho = 50;
- Média = 1,84;
- Desvio = 0,82;
- Graus de liberdade = 49 (tamanho da amostra menos 1);
- Nível de confiança = 95%
- $T_c = \pm 2,00957523449$
- Margem de erro = 0,232235108375;
- Intervalo = 1,60776489162 e 2,07223510838

Sempre que não soubermos a variância da população (a maioria das vezes), independentemente do tamanho da amostra, é recomendável usar a distribuição T de Student, como aproximação da distribuição normal.

Testes de hipóteses

Teste de hipótese é uma formalização do que vimos anteriormente. Na verdade, podemos definir como uma afirmação ou conjectura sobre um parâmetro (ou parâmetros) da distribuição de probabilidades de característica de um fenômeno (ou variável aleatória).

Tudo fica melhor com um exemplo, não? Então, vamos lá...

Uma fábrica de laticínios vende peças de queijo pelo menos 60 kg (em média) cada uma, com pouca variação. A fábrica lhe contratou para validar estatisticamente este parâmetro (peso), pois havia reclamações que, em alguns lotes, havia peças com peso bem menor que o especificado. Como você faria isto? Você poderia pegar um lote, aleatoriamente, e verificar estatisticamente.

Vamos pensar que temos duas hipóteses:

- Os queijos pesam 60kg;
- Os queijos pesam menos que 60kg.

Quando falamos sobre testes de hipótese, há sempre uma hipótese que desejamos testar, e o fazemos buscando observações que nos permitam rejeitá-la. Atenção: não se trata de semântica:

Devemos buscar evidências que nos permitam rejeitar a hipótese que desejamos testar!

Bem, voltando ao teste de hipótese, temos geralmente duas hipóteses antagônicas, e a rejeição de uma implica a não rejeição da outra. Chamamos a hipótese principal de "nula", ou " H_0 ", e a antagônica de "alternativa" ou " H_1 ". Em nosso caso:

- H_0 : Todos os queijos têm peso médio maior ou igual de 60 kg;
- H_1 : Há queijo(s) com peso **significativamente** menores.

Precisamos pesar os queijos da amostra e calcular algumas estatísticas amostrais:

```
lote <- c(58.5, 60.1, 60.02, 57.4, 60.3, 55.4, 58.2, 59.8, 54.3, 60.4, 60.7,  
         60.1, 55.6, 57.1, 60.0, 60.7, 60.3, 56.7, 57.9, 59.01)  
  
print(summary(lote))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
54.30	57.32	59.40	58.63	60.15	60.70

Podemos cometer 2 tipos de erro com esse problema, e precisamos, de alguma forma, estabelecer limites para esses erros:

- Erro **tipo I**: rejeitar a hipótese nula, mesmo sendo verdadeira;
- Erro **tipo II**: falhar em rejeitar a hipótese nula, mesmo com evidências.

Temos que estabelecer limites para rejeição da hipótese nula, como fizemos no estudo de intervalos de confiança. Para começar, temos que ver qual teste vamos usar:

- **Teste Z**: Baseado na distribuição normal (teorema do limite central);
- **Teste T**: Baseado na distribuição T de Student (para amostras menores que 30 elementos ou variância da população desconhecida).

Neste caso, temos 20 elementos e desconhecemos a variância da população, logo, seria mais prudente usar a distribuição T de Student.

Lembra-se dos intervalos de confiança? É a mesma coisa: precisamos estabelecer uma região no "sino" da distribuição normal, que confirme a hipótese nula.

Chamamos de " α " (alfa) a probabilidade de cometermos um erro do tipo I: rejeitar a hipótese nula, sendo ela verdadeira. E chamaremos de " β " (beta) a probabilidade de cometermos um erro de tipo II: não rejeitar a hipótese nula, mesmo sendo falsa.

Devemos ver o que é mais importante para nós, ou seja, qual o tipo de erro mais significativo para nós. Geralmente, trabalhamos mais com o erro tipo I, logo, consideramos mais a significância da probabilidade α embora possamos trabalhar com ambas (α e β).

Atenção: α é uma probabilidade, logo, é uma área! Os pontos no eixo das abscissas são os desvios da média (normalizada), as áreas são os vários valores que "y" (a variável esperada) pode assumir.

Antes de continuar, gostaria de esclarecer a diferença entre valor das observações e probabilidade de ocorrência. Isso sempre foi muito difícil para mim, quando eu comecei a estudar estatística a sério.

O que estamos querendo testar

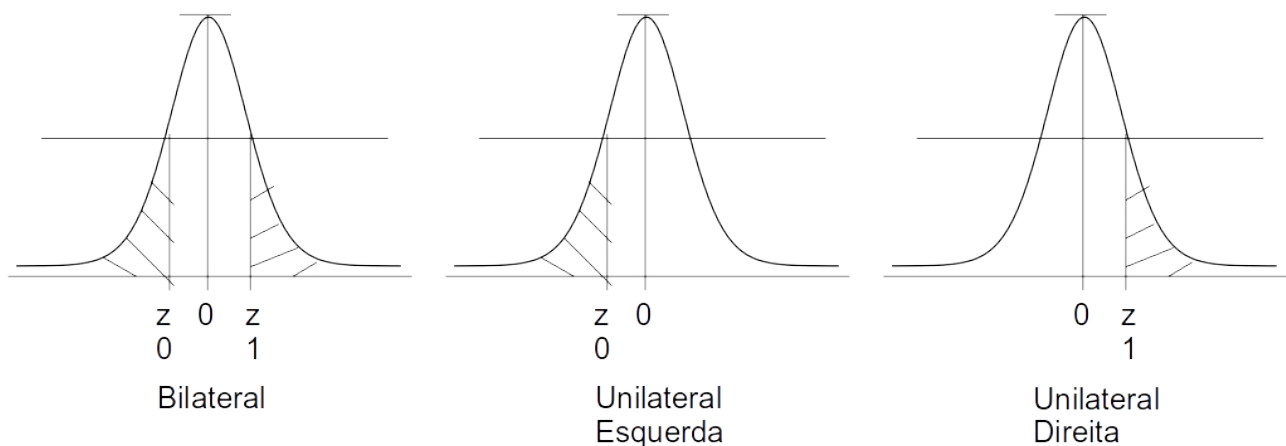
Nunca devemos perder isto de vista. Queremos saber se, baseados na amostra, temos evidências estatísticas de que os queijos produzidos pela fábrica pesam em torno de 60 kg. Vamos verificar a probabilidade de haver médias de pesos de queijos mais extremas que as da amostra, ou seja, muito menores que a da amostra. Houve relatos de que há queijos pesando bem menos que 60 kg, logo, precisamos saber qual é a probabilidade de haver queijos com pesos médios muito diferentes de 60 kg na vida real.

Tipos de teste

Quando definimos hipóteses, temos que selecionar as regiões do gráfico onde a hipótese nula pode ser rejeitada. Isso pode ocorrer de 2 formas:

- Bilateral ou bicaudal: Quando valores em ambas as caudas possam rejeitar H_0 ;
- Unilateral ou unicaudal (à direita ou esquerda): Quando apenas valores em uma das duas caudas pode rejeitar H_0 .

É mais fácil visualizar...



- Bilateral: $H_0: \mu = k$; $H_1: \mu \neq k$;
- Unilateral à esquerda: $H_0: \mu = k$; $H_1: \mu < k$;
- Unilateral à direita: $H_0: \mu = k$; $H_1: \mu > k$;

Qual é o teste do nosso exemplo?

- H_0 : peso = 60;
- H_1 : peso < 60.

Então, seria unilateral à esquerda, pois as observações com peso < 60 poderiam nos forçar a rejeitar a hipótese nula. Agora, precisamos pensar em um nível de significância, que será a área das probabilidades, a partir da qual, rejeitaremos a hipótese nula.

Podemos pensar em probabilidades: 1%, 5% ou 10%, que são os valores mais comuns. Se optarmos por 5% podemos demarcar essa área traçando uma reta que corta o eixo das abscissas no valor (Z ou T) correspondente. Para encontrar esse ponto, basta pesquisar na tabela da distribuição T de Student (ou normal padrão, se for o caso).

Para sabermos qual é a coordenada desse ponto (P(observação)) no eixo das abscissas, precisamos pesquisar na tabela da distribuição T de Student, ou então usamos uma função em R:

- $n = 20$;
- graus de liberdade = 19;
- $\bar{x} = 58,66$;
- $s = 1,97$;

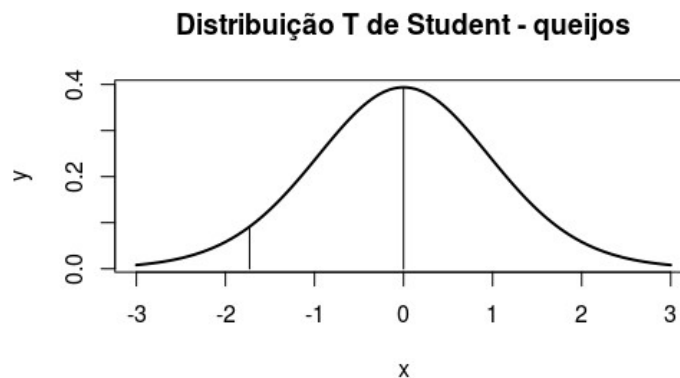
Se pesquisarmos na tabela da distribuição T de Student, vemos que isso corresponde ao valor: 1,7291. Como nosso teste é à partir da cauda esquerda, então esse valor será negativo. Em R podemos obter isso desta forma:

```
graus = length(lote) - 1 # Graus de liberdade
print(paste('Graus de liberdade:', graus))
tq1 <- qt(0.05, df=graus) # teste unicaudal à esquerda por isso usamos 0.05 para 95%
```

Lembre-se que temos " $n - 1$ " graus de liberdade, pois estamos utilizando a Estatística T! A função "qt()" retorna o valor T correspondente à 5% (teste unicaudal à esquerda).

Podemos pensar em um gráfico agora, certo? A próxima figura mostra isso. Para gerá-la, usamos o seguinte código em R:

```
x <- seq(-3, 3, length=500)
y <- dt(x, df=graus)
plot(x, y, type="l", lwd=2, main = 'Distribuição T de Student - queijos')
lines(c(0, 0), c(0, dt(0, df=graus)))
lines(c(tq1, tq1), c(0, dt(tq1, df=graus)))
```



Se o valor da média estiver na região demarcada pela primeira linha, rejeitaremos H_0 .

Agora, precisamos calcular o “t-score” da nossa média usando a fórmula:

$$t = \frac{(\bar{x} - \mu)}{(s / \sqrt{n})}$$
$$t = \frac{(58,66 - 60)}{1,97 / \sqrt{20}}$$

Em R:

```
ta <- (mean(lote) - 60) / (sd(lote) / sqrt(length(lote)))
```

Chegamos ao t-score: -3,120903. Bem, nosso ponto crítico era: -1,7291 e o nosso t- score é menor do que ele: -3.12, portanto, podemos **rejeitar** a hipótese nula.

valor-p

Outra maneira de constatar isso é comparar a área de erro tipo I, ou área de significância (α) com a área onde está nosso t-score, que é conhecida como “valor-p” (ou **p-value**). Basta procurar a área correspondente ao t-score na tabela ou então usar R:

```
pvalue <- pt(ta, 19)
if (pvalue < 0.05) {
  print('Rejeitamos a hipótese nula pelo valor-p')
} else {
```

```
} print('Não rejeitamos a hipótese nula pelo valor-p')  
}
```

Nosso valor-p é 0,0028, que é bem menor que 0,05 (nível de significância), logo, podemos rejeitar a hipótese nula.

Conclusão do teste

Existem evidências com probabilidade significativa de haver médias menores que 60kg. Isso significa que devemos estudar melhor o processo produtivo, de modo a diminuir a variação de peso dos queijos.

Agora, vamos ao nosso "glossário" de R desta sessão

O script que usamos é "inferencia.R" e ele está no nosso repositório:

<https://github.com/cleuton/datascience/tree/master/R-course/lesson5>

Assimetria e curtose, vimos na sessão 4. Para calculá-las, precisamos carregar a biblioteca "moments":

```
library('moments')
```

Esta biblioteca não vem instalada no R, logo, precisamos baixar o pacote que a instalará. Por isto, se você ainda não instalou o pacote "moments", receberá um erro ao executar esta linha acima. Para instalar (se você ainda não o fez), digite isto na janela "console":

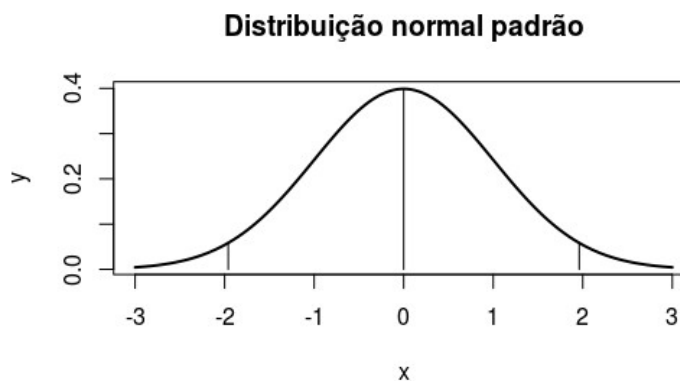
```
install.packages("moments")
```

Para calcular assimetria e curtose, usamos as funções "skewness()" (assimetria) e "kurtosis()" (curtose):

```
media = mean(filhos)  
desvio = sd(filhos)
```

```
print(paste('Média:',media))
print(paste('Desvio:',desvio))
print(paste('Assimetria:',skewness(filhos)))
print(paste('Curtose:',kurtosis(filhos)))
print(summary(filhos))
```

As vezes é "legal" plotar a distribuição normal, mostrando o intervalo de confiança, embora isto nada acrescente ao resultado, ajuda a visualizar:



A distribuição normal padrão tem média zero e desvio padrão 1. Para gerar um gráfico destes em R, geramos alguns desvios padrões, calculamos suas probabilidades e plotamos um gráfico:

```
x <- seq(-3,3,length=500)
y <- dnorm(x,mean=0, sd=1)
plot(x,y, type="l", lwd=2, main = 'Distribuição normal padrão')
```

A função "seq()" simplesmente gera uma sequência de valores, com o menor e o maior especificados. A função "dnorm()" gera a densidade das probabilidades (o eixo das ordenadas), dada a média e o desvio ("mean" e "sd"). E o comando "plot()" desenha o gráfico, com os parâmetros:

- x, y: Vetores das coordenadas dos pontos;
- type: "p" para pontos, "l" para linhas etc
(<https://www.rdocumentation.org/packages/graphics/versions/3.4.3/topics/plot>);
- lwd: Largura da linha;
- main: Título do gráfico.

Para calcular os limites do intervalo de confiança e plotar as linhas verticais, usamos este código em R:

```
# Calculando os limites do intervalo de confiança de 95%:  
z1 <- qnorm(0.025)  
z2 <- qnorm(0.975)  
lines(c(0,0),c(0,dnorm(0)))  
lines(c(z1,z1),c(0,dnorm(z1)))  
lines(c(z2,z2),c(0,dnorm(z2)))
```

A função "qnorm()" retorna qual é o Z-score (qnorm se aplica à distribuição normal) correspondente ao quartil ou percentual.

A função "lines()" desenha linhas. É muito importante entender COMO passar os parâmetros, senão, você quebrará muito sua cabeça:

```
lines(<vetor de abscissas>, <vetor de ordenadas>)
```

No nosso exemplo, o primeiro "lines()" desenha a linha da média:

```
lines(c(0,0),c(0,dnorm(0)))
```

Esta linha vai do ponto ($x = 0$, $y = 0$) até o ponto ($x = 0$, $y = \text{probabilidade da média zero}$). A função "dnorm()" retorna o valor de densidade para um z-score.

As linhas dos valores críticos são assim:

```
lines(c(z1,z1),c(0,dnorm(z1)))  
lines(c(z2,z2),c(0,dnorm(z2)))
```

A primeira linha vai do ponto ($z1,0$) até o ponto ($z1,\text{densidade de } z1$), e a segunda linha vai do ponto ($z2,0$) até o ponto ($z2, \text{densidade de } z2$).

Para calcular a margem de erro, nada fizemos de mais:

```
# Calculando a margem de erro:  
E <- z2 * (desvio / sqrt(length(filhos)))  
print(paste('Margem de erro:',E))  
minf <- media - E  
msup <- media + E  
print(sprintf('A média de filhos está entre: %f e %f com 95%% de  
confiança',minf,msup))
```

A função "sqrt()" calcula a raiz quadrada.

A única inovação aqui foi a função "sprintf()", que funciona exatamente como sua homônima, na linguagem "C". Passamos um texto contendo caracteres de formatação, e uma lista de valores, que serão inseridos dentro dos pontos delimitados, dentro do texto:

```
print(sprintf('A média de filhos está entre: %f e %f com 95%% de
confiança', minf, msup))
```

```
"A média de filhos está entre: 1.613498 e 2.066502 com 95% de confiança"
```

O caracter "%" marca um ponto de inserção, cada valor será inserido em cada caracter "%", de acordo com sua ordem nos parâmetros. Depois do "%" vem o formato:

- d – Inteiro;
- f – real;
- s – caracter;
- % - ignorar;

Você deve estar se perguntando por que coloquei dois "%%" ao lado do 95. É para a função "sprintf()" ignorar o percentual depois do número 95, tratando-o apenas como um carácter comum.

Quando passamos a usar a distribuição T de Studente, algumas funções mudam de nome:

```
# Plotando a distribuição T de Student e o intervalo de confiança:
x <- seq(-3,3,length=500)
y <- dt(x,df=graus)
plot(x,y, type="l", lwd=2, main = 'Distribuição T de Student')
```

Em vez de "dnorm()" utilizamos "dt()" passando os graus de liberdade (n-1).

Os graus de liberdade podem ser interpretados como o número de determinações independentes (dimensão da amostra) menos o número de parâmetros estatísticos a serem avaliados na população (Wikipedia). Estamos estimando o parâmetro "média", logo, subtraímos 1 do tamanho da amostra.

Deve haver uma maneira mais simples de realizar o teste de hipótese, não? E há! A função: "t.test()":

```
print(t.test(lote,alternative = "less", mu=60, conf.level = 0.95))
```

Resultado da função "t.test()":

One Sample t-test

```
data: lote
t = -3.1209, df = 19, p-value = 0.002813
alternative hypothesis: true mean is less than 60
95 percent confidence interval:
 -Inf 59.38749
sample estimates:
mean of x
```

58.6265

Sinistro, não? Todo aquele código que criamos, resumido a uma única linha. Passamos as informações e a função "t.test()" calcula e retorna o resultado do teste:

- x: Vetor da amostra. No nosso caso é o vetor "lote";
- alternative: Um texto indicando qual é o tipo de teste, ou qual é a hipótese alternativa, que, no nosso caso, é "less", ou menor que a média populacional (60 kg);
- mu: Média populacional;
- conf.level: Nível de confiança, no nosso caso 95% ou 0,95.

O resultado é bem explicativo:

- t: Valor em t-scores da média observada;
- df: Graus de liberdade;
- p-value: Valor-p. Como é 0,002813, é menor que 0,05, logo, rejeitamos a hipótese nula.