UNIVERSIDADE DE PERNAMBUCO

Escola Politécnica de Pernambuco

Especialização em Ciência de Dados e Analytics

Estatística Computacional

Aula 2.1 – Testes de Hipóteses e ANOVA – PARTE I

Prof. Dr. Rodrigo Lins Rodrigues

rodrigolins.rodrigues@ufrpe.br

Conteúdo programático

- ✓ Introdução sobre testes de hipóteses;
- ✓ Formulação de hipóteses;
- ✓ Etapas para a construção de hipóteses;
- ✓ Testes de normalidade;
- ✓ Teste t de Student para uma amostra;
- ✓ Teste para duas amostras aleatórias independentes;
- ✓ Teste para duas amostras aleatórias pareadas;
- ✓ Análise de Variância (*ANOVA*);
- ✓ Aplicações computacionais.

O que você aprenderá ?

- ✓Entender **o que é um teste de hipótese** e seus objetivos;
- ✓Entender os **diferentes tipos** de testes de hipóteses que existem;
- ✓Compreender as **suposições** inerentes a cada teste;
- ✓Classificar um teste de hipótese como **paramétrico** e **não-paramétrico**;
- ✓Calcular teste de hipóteses de forma **manual** e **computacional**.

Introdução a Testes de Hipóteses

Hipótese

O que você acha que seja
uma **Hipótese?**

Hipótese

- É uma afirmação "*especulatória*" sobre as características de uma população.

- *Exemplo:*
  - ✓ *...Sedentarismo é um dos fatores que mais influenciam na diabetes...*

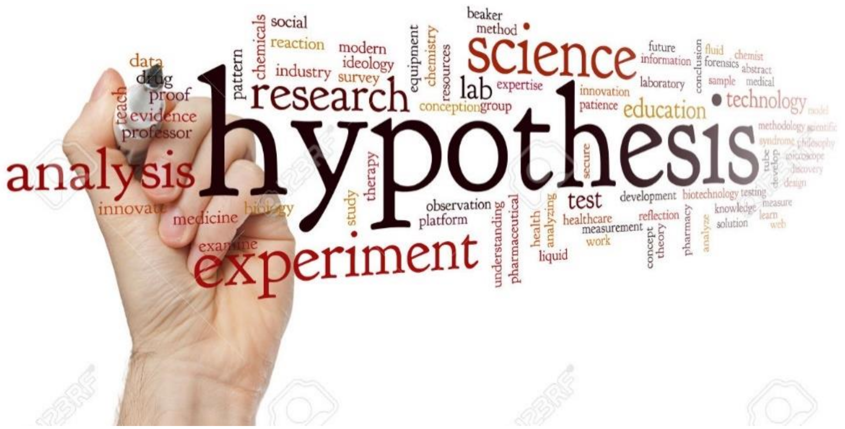
Testes de Hipóteses

O que você acha que seja
um **Teste de Hipótese?**

Testes de Hipóteses

- É um procedimento estatístico para **verificar a veracidade** ou **falsidade** de uma determinada afirmação*;*

- E realizado com base em uma **característica da população**.



Testes de Hipóteses

- Trata-se de uma técnica para se fazer a **inferência estatística** sobre uma população a partir de **uma amostra;**

- Toda hipótese tem como objetivo testar **parâmetros populacionais**;

- É baseado em uma **amostra representativa da população**;

Testes de Hipóteses

- São aplicados em situações em que se **conhece a distribuição dos dados**;

- É necessário **pressuposto de normalidade**;

- São testes **mais robustos** do que os testes não-paramétricos;

- Servem para **testar parâmetros populacionais**, tais como: média, variância e proporção;

Teste de Hipóteses

- Um teste de hipótese se divide em:

  - ✓ **A hipótese nula (H_0):** simboliza a situação atual – a que todos assumem como sendo verdadeira até que se prove o contrário.

  - ✓ **A hipótese alternativa (H_1):** representa a situação alternativa.

Teste de Hipóteses

- Antes de realizar qualquer teste de hipóteses é importante ficar atendo a alguns pontos:

  ✓ É necessário fazer um **bom processo de amostragem** da variável;

  ✓ É necessário verificar se já existem **informações prévias** sobre aquela variável;

  ✓ É importante identificar a **distribuição da variável**.

Distribuições de Probabilidade

Distribuição de probabilidade

- Existem um **grande número de distribuições** de probabilidades para as variáveis aleatórias discretas e contínuas;

- Algumas distribuições discretas que são:
  - ✓Uniforme,
  - ✓Bernoulli;
  - ✓Binomial;
  - ✓...

Distribuição Contínua

- São variáveis com valores pertencentes ao intervalo dos números reais;

- Exemplos de variáveis aleatórias contínuas:
  - ✓Renda,
  - ✓salário,
  - ✓tempo de duração de um equipamento,
  - ✓comprimento de uma peça.

Distribuição Contínua

- Dentre as várias distribuições de probabilidade contínuas, a mais famosa é **a distribuição normal**;

- Apresenta **grande aplicação em pesquisas** científicas e tecnológicas;

- Grande parte das variáveis contínuas de interesse prático **seguem essa distribuição** ou podem **ser aproximadas**.

Distribuição Contínua

- Função de Densidade:
  - ✓A função densidade de probabilidade da distribuição normal é dada por:

$$\mathrm{f}(x) = \frac{1}{\sigma\sqrt{2\pi}} . exp^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, x \in R.$$

Onde:

μ e σ são a média e o desvio padrão, respectivamente, da distribuição de probabilidade.

π corresponde a 3,1415 e exp é uma função exponencial.

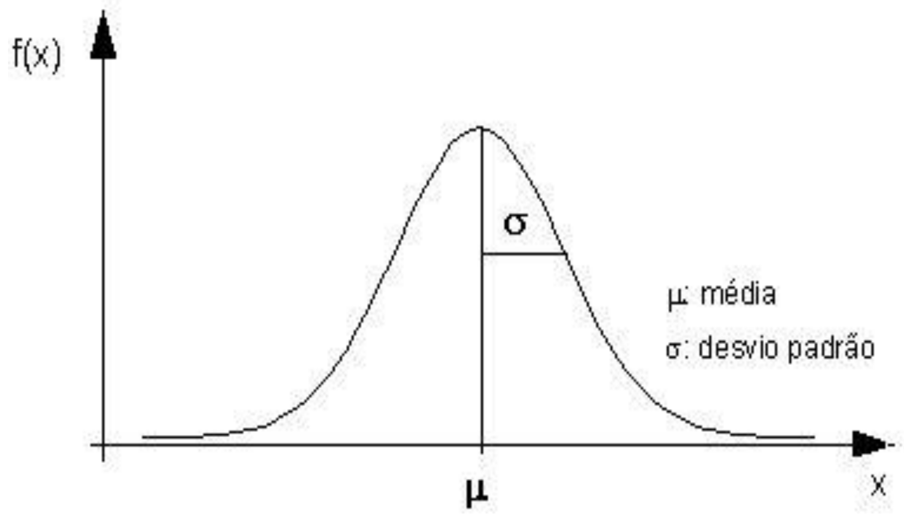
Distribuição Contínua

- **As principais propriedades da distribuição normal são:**

✓**é simétrica** em relação ao ponto x = μ (50% abaixo e 50% acima da média);

✓tem **formato de curvatura** (como mostrado na figura);

✓as três medidas de posição (**média, mediana e moda) localiza-se no ponto máximo** da curva;



Formulando um teste de hipóteses

Formulação do teste

- A formulação de uma hipótese deve ser baseada no problema de pesquisa;

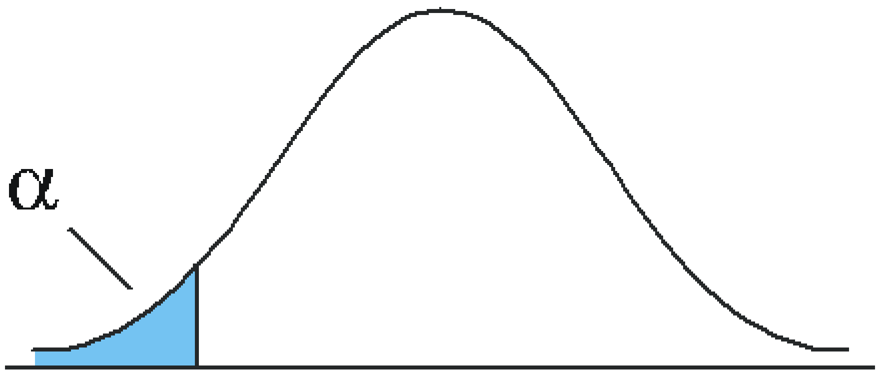
- Existem três formas de formular hipótese:
  - ✓Unilateral a esquerda;
  - ✓Unilateral a direita;
  - ✓Bilateral;

Formulação do teste

Unilateral à **esquerda***:*

✓A região crítica está na cauda esquerda da distribuição e corresponde ao nível de significância;

$$\begin{cases} H_o: \mu = 50 \\ H_1:: \mu > 50 \end{cases}$$

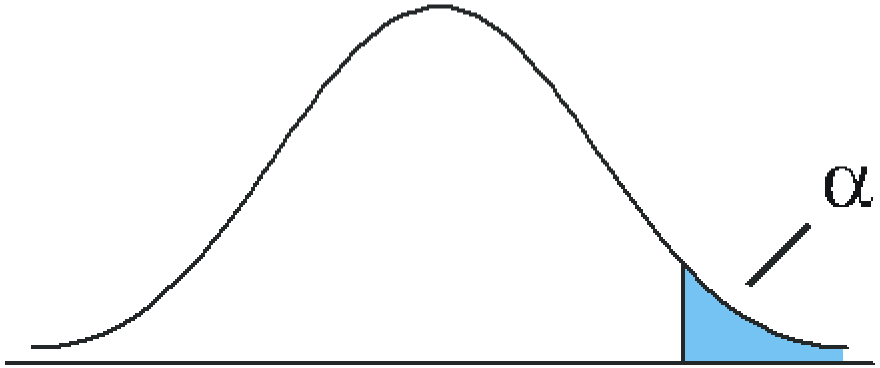


Formulação do teste

Unilateral à **direita***:*

✓A região crítica está na cauda direita da distribuição e corresponde ao nível de significância.

$$\begin{cases} H_o: \mu = 50 \\ H_1:: \mu < 50 \end{cases}$$

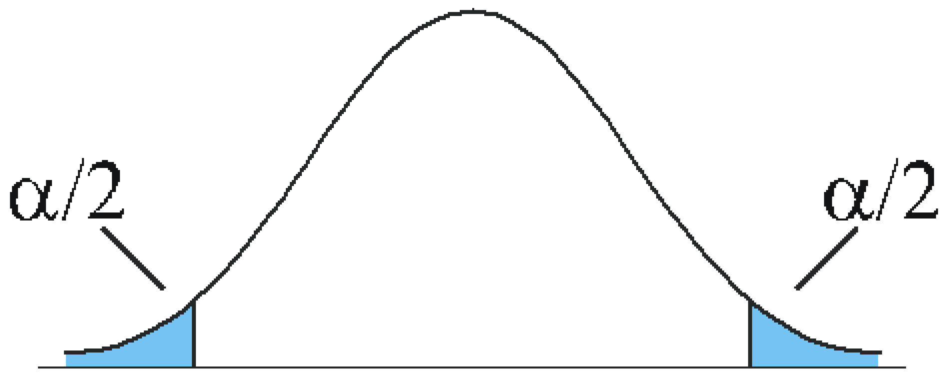


Formulação do teste

Hipótese **Bilateral**

- ✓É representada por duas caudas de tamanhos iguais, nas extremidades esquerda e direita da curva.

$$\begin{cases} H_o : : \mu = 50 \\ H_1 :: \mu \neq 50 \end{cases}$$



Formulação do teste

- Se o valor da estatística do teste cair na região crítica, rejeita-se Ho.
- Ao contrário, quando aceitamos, dizemos que não houve evidência amostral significativa no sentido de permitir a rejeição de Ho.

Formulação do teste

- **Tipos de erros:**
  - ✓ Como a decisão de uma hipótese é baseada com base na amostra, dois tipos de erros podem ocorrer:

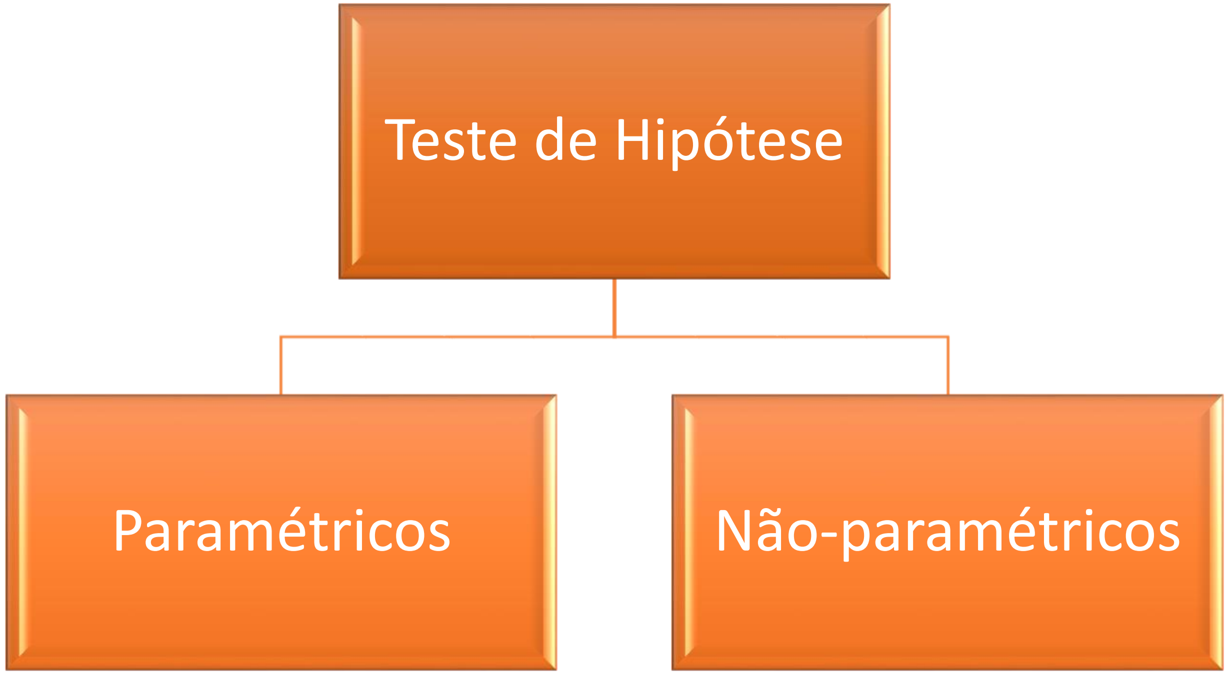
Possibilidades envolvidas em um teste de hipóteses

Realidade \ Decisão	Aceitar H_0	Rejeitar H_0
H_0 é verdadeira	**Decisão correta** $1 - \alpha$ = P(Aceitar H_0 / H_0 é V) = P(H_0 / H_0)	**Erro do Tipo I** α = P(Erro do tipo I) = P(Rejeitar H_0 / H_0 é V) = Nível de significância do teste = P(H_1 / H_0)
H_0 é falsa	**Erro do Tipo II** β = P(Erro do tipo II) = = P(Aceitar H_0 / H_0 é falsa) = P(Aceitar H_0 /H_1 é V) = P(H_0 /H_1)	**Decisão correta** $1 - \beta$ = P(Rejeitar H_0 / H_0 é falsa) = P(H_1 / H_1) = Poder do teste.

Teste Paramétrico

Teste Paramétrico

- Os testes de hipóteses se dividem em dois tipos:



Teste Paramétrico

- Os testes paramétricos envolvem **parâmetros populacionais**;

- Nos **testes não-paramétricos**, as hipóteses são formuladas sobre **características qualitativas** da população;

- São valores fixos geralmente desconhecidos e representados por letras gregas:
  - ✓ (μ) média populacional;
  - ✓ (σ) desvio-padrão populacional;
  - ✓ (σ^2) variância populacional;
  - ✓ (p) proporção populacional...

Teste Paramétrico

- Os métodos paramétricos são então aplicados para dados quantitativos e exigem suposições fortes sobre sua validação:

  i. As observações devem ser independentes;
  ii. A amostra deve ser retirada de população com **distribuição normal**;
  iii. Para testes de comparação de médias emparelhadas as populações devem ter variâncias iguais;
  iv. As variáveis devem ser medidas em escalas intervalar.

Pressuposto de normalidade

- Teste de normalidade:
  - ✓ Pressuposto que os dados seguem **distribuição normal** ou gaussiana;
  - ✓ É possível ter **indícios gráficos**, sobre a distribuição dos dados, no entanto, somente os testes de aderência podem comprovar se os dados seguem normalidade;
  - ✓ Os dois **principais testes** de normalidade são:
    - ✓ Kolmogorov-Smirnov
    - ✓ Shapiro-Wilk

Pressuposto de normalidade

- É muito importante saber **verificar a normalidade** de uma variável;

- Grande parte das análises estatísticas inferenciais **tem pressuposto de normalidade**;

- Inicialmente é **importante plotar o gráfico** de histograma e posteriormente aplicar um teste.

Verificando normalidade

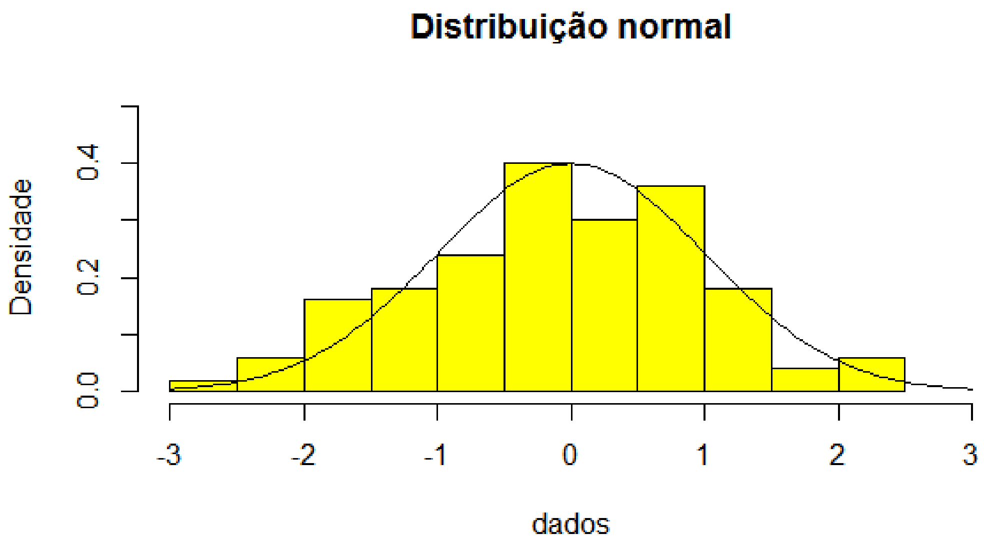
- Gerando números aleatórios com distribuição normal:

```r
#Gerando dados com distribuição normal
dados <-rnorm(100)

#Grafico de distribuicao normal
hist(dados,probability=TRUE,col="yellow", main="Distribuição normal",
     ylab="Densidade", ylim=c(0,0.5),xlim=c(-3,3))
curve(dnorm(x),add=T)

#Realizando teste de normalidade
shapiro.test(dados)
```

Verificando normalidade

- Resultado:

```
        Shapiro-Wilk normality test

data:  dados
W = 0.99144, p-value = 0.7796
```



Obs: Se o *p-valor* for maior do que 0,05, então a variável tem distribuição normal.

Etapas para a construção de hipóteses (sem software estatístico)

- **Passo 1:** Definir a hipótese nula H_0 a ser testada e a hipótese alternativa $H_{1;}$
- **Passo 2:** Definir o nível de significância;
- **Passo 3:** Escolher uma estatística de teste adequada;
- **Passo 4:** Fixar a região crítica do teste (o valor crítico é determinado em função do nível de significância);
- **Passo 5:** Retirar uma amostra e calcular o valor observado da estatística do teste;
- **Passo 6:** Se o valor da estatística pertencer à região crítica, rejeitar H_0; caso contrário, não rejeitar H_0;

Etapas para a construção de hipóteses (com software estatístico – *p-valor*)

- **Passo 1:** Definir a hipótese nula H_0 a ser testada e a hipótese alternativa H_1.
- **Passo 2:** Definir o nível de significância.
- **Passo 3:** Escolher uma estatística de teste adequada.
- **Passo 4:** Retirar uma amostra e calcular o valor observado da estatística do teste.
- **Passo 5:** Determinar o *p-value* que corresponde à probabilidade associada ao valor observado da amostra, calculado no passo 4.
- **Passo 6:** Se o valor o *p-value* for menor do que o nível de significância α estabelecido no passo 2, rejeitar H_0; caso contrário, não rejeitar H_0.

Construindo testes de hipóteses

Construindo Testes de hipóteses

- Iremos conhecer **três tipos de testes** paramétricos:

    - ✓Teste t de Student para **uma amostra**;
    - ✓Teste t de Student para **duas amostras aleatórias** independentes;
    - ✓Teste t de Student para **amostras pareadas**.

Construindo Testes de hipóteses

- **Teste *t* de *Student* para uma amostra**

  ✓É aplicado quando não se conhece a variância populacional;

  ✓Testa se a média populacional **assume ou não um determinado valor**;

  ✓Trata-se de **testar se um valor é verdadeiro** em relação ao valor do parâmetro populacional.

Construindo Testes de hipóteses

Teste *t* de *Student* para uma amostra

(Exemplo prático - 1)

- Uma empresa está lançando um novo caminhão.
- Deseja-se testar a hipótese de que o tempo médio de pintura do novo caminhão é igual ao caminhão antigo, que é de 690 min.
- Para isto coletou-se uma amostra de 12 elementos do novo caminhão e mediu-se o tempo médio de pintura.

Caminhão Y	Tempo de pintura
1	920
2	710
3	680
4	1000
5	1010
6	850
7	880
8	990
9	1030
10	995
11	775
12	670
Média	875,833
Desvio	136,662

Construindo Testes de hipóteses

Teste *t* de *Student* para uma amostra

(Exemplo prático - 1)

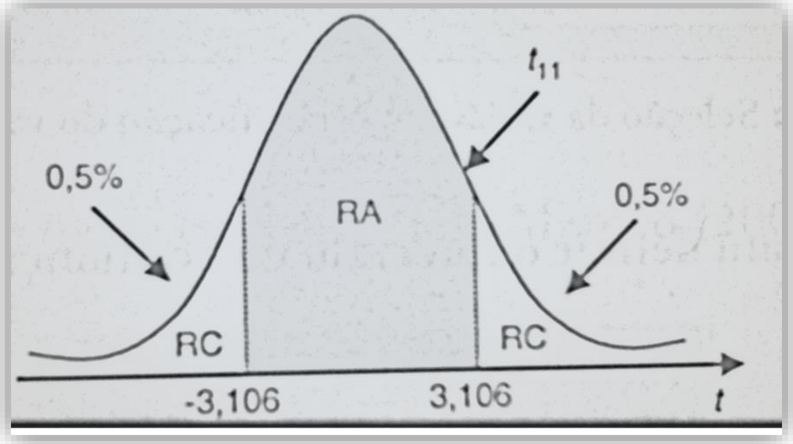
- Primeiramente, verifica-se se há normalidade dos dados por meio do teste de kolmogorov-Sminrnov ou de Shapiro-Wilk.

	Kolmogorov-Smirnov			**Shapiro-Wilk**		
	Statistic	df	Sig.	Statistic	df	Sig.
TempoPitura	.215	12	.132	.875	12	.075

- Como o nível de significância observado para os dois testes é superior a 5%, conclui-se que há normalidade dos dados. Pode-se aplicar o teste *t* para uma única amostra.

Construindo Testes de hipóteses

1. A hipótese nula H_0 afirma que o tempo médio de pintura do novo caminhão é 690 (μ = 690) e a hipótese alternativa H_0 afirma que $\mu \neq 690$;
2. O nível de significância do teste é α = 1%;
3. Escolhe-se o teste t com 11 graus de liberdade (tamanho da amostra - 1);
4. A figura abaixo representa a região crítica do teste (ver tabela)

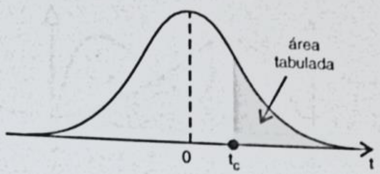


Obs: ver tabela

Construindo Testes de hipóteses

Tabela B Distribuição *t* de *Student*

$$P(T_{cal} > t_c) = \alpha$$



Valores críticos da distribuição t de Student

Graus de liberdade	Probabilidade associada na cauda superior								
ν	0,25	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	1,000	3,078	6,314	12,706	31,821	63,657	127,3	318,309	636,619
2	0,816	1,886	2,920	4,303	6,965	9,925	14,09	22,33	31,60
3	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,21	12,92
4	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,697	1,363	1,796	2,201	[illegible]	3,106	3,497	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965

Construindo Testes de hipóteses

5. Calcular o valor real da variável T:

$$T_{cal} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{875{,}833 - 690}{136{,}662/\sqrt{12}} = 4{,}710$$

6. Conclusão: como o valor da estatística pertence à região crítica, isto é, $T_{calculado} > 3{,}106$, o teste **rejeita a hipótese nula**, concluindo que o tempo médio de pintura dos caminhões do tipo Y é diferente de 690 ($\mu \neq 690$)

Construindo Testes de hipóteses

Teste *t* de *Student* para <u>duas amostras aleatórias</u> independentes

- É aplicado para **testar se as médias** de duas amostras aleatórias, extraídas da **mesma população** são ou não significativamente diferentes;

- As duas **amostras tem distribuição normal** com variâncias desconhecidas, porém, iguais;

- É pressuposto que a variabilidade das variáveis são iguais;

Construindo Testes de hipóteses

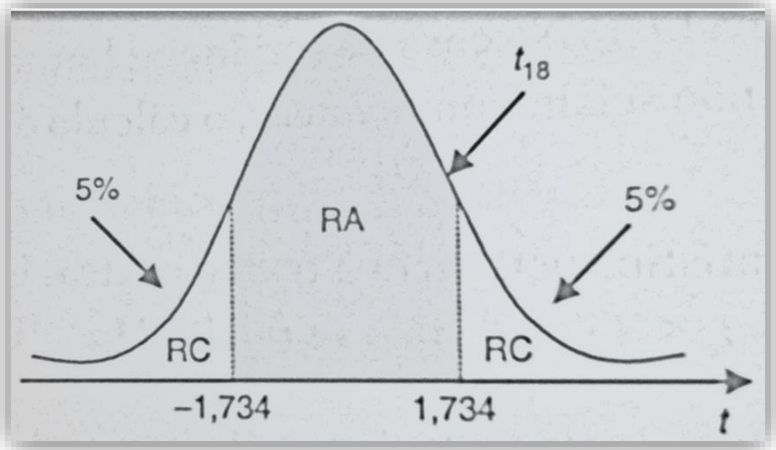
Teste *t* de *Student* para duas amostras aleatórias independentes (Exemplo prático - 2)

- Pretende-se verificar se o tempo médio de fabricação de dois produtos plásticos (x e y) é semelhante. Para cada produto, coletou-se o tempo médio de uma amostra de tamanho n = 10;

Produto	Tempo médio									
X	16	22	27	20	18	24	19	20	21	25
Y	30	25	25	28	27	33	24	22	24	29

Construindo Testes de hipóteses

1. A hipótese nula H_0 afirma que as médias populacionais são iguais, isto é, $\mathbf{H_{0:}\ \mu_x = \mu_y}$. Já a hipótese alternativa afirma que as médias populacionais são diferentes, $\mathbf{H_{1:}\ \mu_x \neq \mu_y}$;
2. O nível de significância do teste é α = 10%;
3. A variável teste escolhida é T;
4. De acordo com o teste de Levene, pode-se concluir que as variâncias são homogêneas. Logo, o número de graus de liberdade é G.L=10+10-2 = 18. A região crítica do teste é:



Construindo Testes de hipóteses

alores críticos da distribuição *t* de *Student*

Graus de liberdade ν	Probabilidade associada na cauda superior 0,25	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	1,000	3,078	6,314	12,706	31,821	63,657	127,3	318,309	636,619
2	0,816	1,886	2,920	4,303	6,965	9,925	14,09	22,33	31,60
3	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,21	12,92
4	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	0,688	[illegible]	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
		1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
						2,831	3,135	3,527	3,819
									3,792

Construindo Testes de hipóteses

5. Como as variâncias são homogêneas, o valor real da variável T é calculado a partir da equação:

$$T_{cal} = \frac{21{,}2 - 26{,}7}{\sqrt{\frac{9 \,.\, 3{,}360^2 + 9 \,.\, 3{,}335^2}{18}} \sqrt{\frac{1}{10} + \frac{1}{10}}} = -3{,}674$$

6. Conclusão: como o valor da estatística pertence à região crítica, isto é, $T_{calculado} < -1{,}734$, rejeita-se H_0, concluindo que as médias populacionais são diferentes.

Construindo Testes de hipóteses

Teste *t* de *Student* para duas amostras aleatórias relacionadas (pareadas)

- É aplicado para verificar se as médias de duas amostras relacionadas, extraídas da mesma população, são ou não significativamente diferentes;

- Além da normalidade dos dados de cada amostra, o teste exige que as variâncias de cada amostra sejam iguais entre si (homocedasticidade);

- Como exemplo temos... Imagine que queremos testar a aplicação de uma interface em dois momentos para o mesmo grupo de usuários e queremos saber se teve diferença significativa no tempo de uso para a realização de uma atividade.

Construindo Testes de hipóteses

Teste *t* de *Student* para duas amostras aleatórias relacionadas (pareadas) (Exemplo prático - 3)

- Um grupo de funcionários foi submetido a um treinamento, e o objetivo é verificar o desempenho deles antes e depois do curso. Para tanto, foram atribuídas notas para cada funcionário de 1 a 10, antes e depois do treinamento. Utilizou-se uma significância de $\alpha = 5\%$;

Momento	Notas atribuídas										
Antes	5,5	6,1	6,7	6,2	7,0	7,2	5,8	6,8	6,7	7,4	5,0
Depois	6,0	7,2	6,8	8,2	9,0	5,8	6,5	7,2	8,7	5,0	9,2

Construindo Testes de hipóteses

- Primeiramente foi necessário verificar se as hipóteses de normalidade de cada grupo e de homogeneidade das variâncias são satisfeitas;

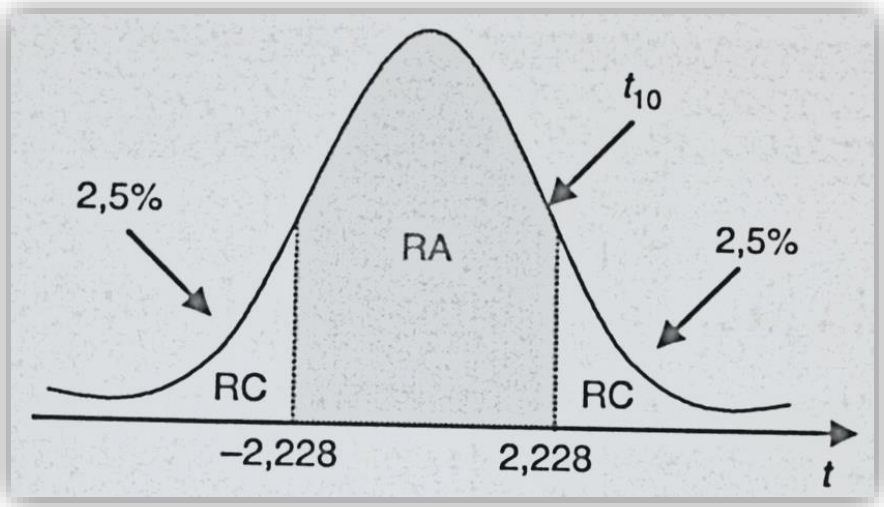
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Antes	.201	11	.200	.956	11	.723
Depois	.147	11	.200	.952	11	.675

- Verificando a homogeneidade das variâncias;

Levene Statistic	df1	df2	Sig.
3.948	1	20	.061

Construindo Testes de hipóteses

1. A hipótese nula H_0 afirma que as médias populacionais são iguais, isto é, **H_0: $\mu_x = \mu_y$**. Já a hipótese alternativa afirma que as médias populacionais são diferentes, **H_1: $\mu_x \neq \mu_y$**;
2. O nível de significância do teste é α = 5%;
3. A variável teste escolhida é T;
4. Fixar a região crítica com o auxílio da tabela T;



Construindo Testes de hipóteses

5. Calcular o valor real da variável T;

$$T_{cal} = \frac{\overline{D}}{S'_D / \sqrt{n}} = \frac{-0{,}836}{1{,}783 / \sqrt{11}} = -1{,}556$$

6. Conclusão: como o valor da estatística não pertence à região crítica, -2,228 <= T <=2,228, não se rejeita H0, concluindo que não houve melhora no desempenho dos funcionários após o treinamento.

Agora é com vocês!

- ✓Em quais situações são aplicados os testes paramétricos?

- ✓O que você entende por pressuposto de normalidade ?

- ✓Quais os principais testes de comparação de médias ?

Praticando Teste de Hipóteses

Teste de Hipóteses em R

- ✓ **Passo 1:** Definir a hipótese nula H_0 a ser testada e a hipótese alternativa H_1;
- ✓ **Passo 2:** Definir o nível de significância;
- ✓ **Passo 3:** Escolher uma estatística de teste adequada;
- ✓ **Passo 4:** Fixar a região crítica do teste (o valor crítico é determinado em função do nível de significância);
- ✓ **Passo 5:** Retirar uma amostra e calcular o valor observado da estatística do teste;
- ✓ **Passo 6:** Se o valor da estatística pertencer à região crítica, rejeitar H_0; caso contrário, aceitar H_1;

Teste de Hipóteses em R

- Scripts em R:

```r
# Teste de normalidade
shapiro.test(amostra)

# Teste para uma amostra (valor de referência)
t.test(amostra1,mu=15)

# Teste para comparação de médias para variáveis independentes
t.test(amostra1, amostra2, var.equal = TRUE)

# Teste para comparação de médias pareadas
t.test(antes,depois,paired=TRUE)
```

Teste de Hipóteses em R

Teste *t* de *Student* para uma amostra

(Exemplo prático - 1)

- Uma empresa está lançando um novo caminhão.
- Deseja-se testar a hipótese de que o tempo médio de pintura do novo caminhão é igual ao caminhão antigo, que é de 690 min.
- Para isto coletou-se uma amostra de 12 elementos do novo caminhão e realizo um teste de hipótese para um valor de referência.

Caminhão Y	Tempo de pintura
1	920
2	710
3	680
4	1000
5	1010
6	850
7	880
8	990
9	1030
10	995
11	775
12	670

Teste de Hipóteses em R

- Solução:

  ✓Formulando a hipótese para o Teste *t* de *Student* **para uma amostra**:

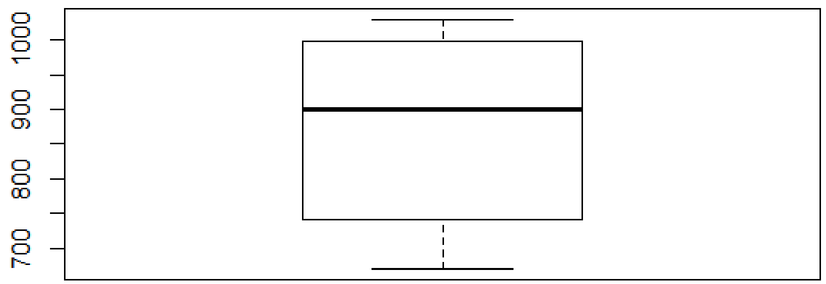
$$\begin{cases} H_{0:}\ \mu = 690 \\ H_{1:}\ \mu \neq 690 \end{cases}$$

Teste de Hipóteses em R

- Solução:

```r
# Ler os dados de um arquivo (interagindo com o usuário)
dados <- read.table(file.choose(),header=TRUE)

# Mostra boxplots da variável tempo de pintura
boxplot(dados$tempoPitura)
```

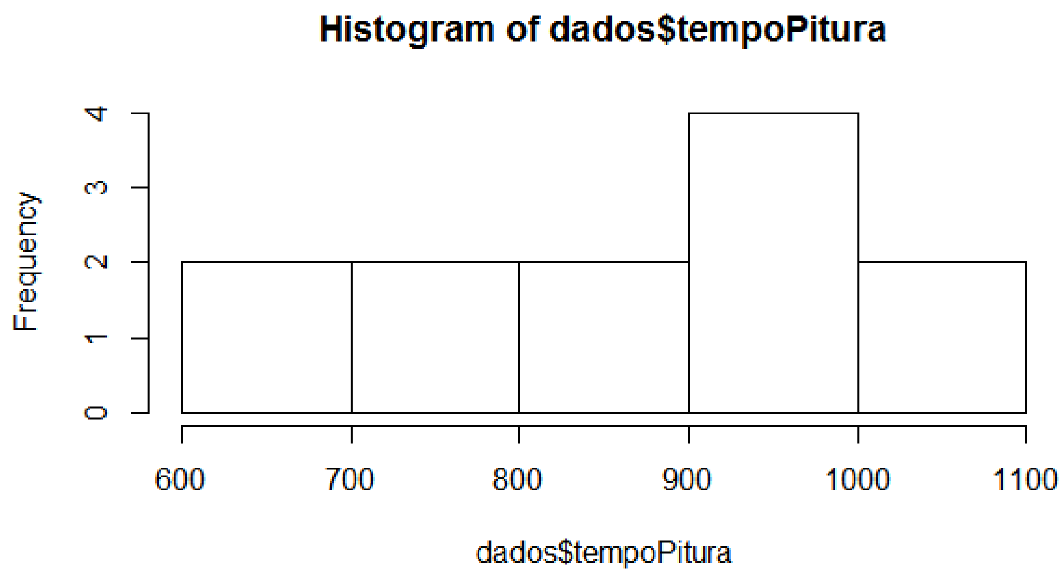
Obs: Já temos uma ideia de que a hipótese nula vai ser rejeitada, mas precisamos aplicar o teste.



Teste de Hipóteses em R

- Solução:

```r
# Plota o histograma pra verificar tendencia de normalidade
hist(dados$tempoPitura)
```



Obs: visualmente a variável não parecer ter distribuição normal. É necessário aplicar um teste de aderência (normalidade).

Teste de Hipóteses em R

- Solução:

```r
# Aplica o teste de normalidade
shapiro.test(dados$tempoPitura)
```

```
        Shapiro-Wilk normality test

data:  dados$tempoPitura
W = 0.87485, p-value = 0.07531
```

Teste de Hipóteses em R

- Solução:

```r
# Alica o teste de hipóteses
t.test(dados$tempoPitura, mu = 690)
```

```
data:  dados$tempoPitura
t = 4.7105, df = 11, p-value = 0.0006392
alternative hypothesis: true mean is not equal to 690
95 percent confidence interval:
 789.0024 962.6643
sample estimates:
mean of x
 875.8333
```

Obs: como o *p-valor* é menor do que 0,05, então rejeita-se a hipótese H_0

Teste de Hipóteses em R

Teste *t* de *Student* para duas amostras aleatórias independentes (Exemplo prático - 2)

- Pretende-se verificar se o tempo médio de fabricação de dois produtos plásticos (x e y) é semelhante com uma significância de 10%. Para cada produto, coletou-se o tempo médio de uma amostra de tamanho n = 10;

Produto	Tempo médio									
X	16	22	27	20	18	24	19	20	21	25
Y	30	25	25	28	27	33	24	22	24	29

Teste de Hipóteses em R

- Solução:

1. A hipótese nula H_0 afirma que as médias populacionais são iguais, isto é, $\mathbf{H_{0:}\ \mu_x = \mu_y}$. Já a hipótese alternativa afirma que as médias populacionais são diferentes, $\mathbf{H_{1:}\ \mu_x \neq \mu_y}$;
2. O nível de significância do teste é $\alpha = 10\%$;
3. A variável teste escolhida é T;
4. Para este exemplo vamos supor que as variâncias são homogêneas:

Teste de Hipóteses em R

- Solução:

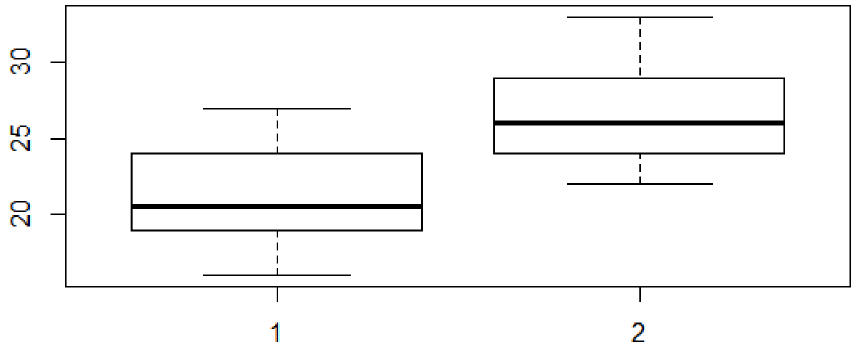
  ✓Formulando a hipótese para o teste *t* de *Student* **para duas amostras aleatórias independentes**:

$$\begin{cases} \mathbf{H_0:} \ \mu_x = \mu_y \\ \mathbf{H_1:} \ \mu_x \neq \mu_y \end{cases}$$

Teste de Hipóteses em R

- Solução:

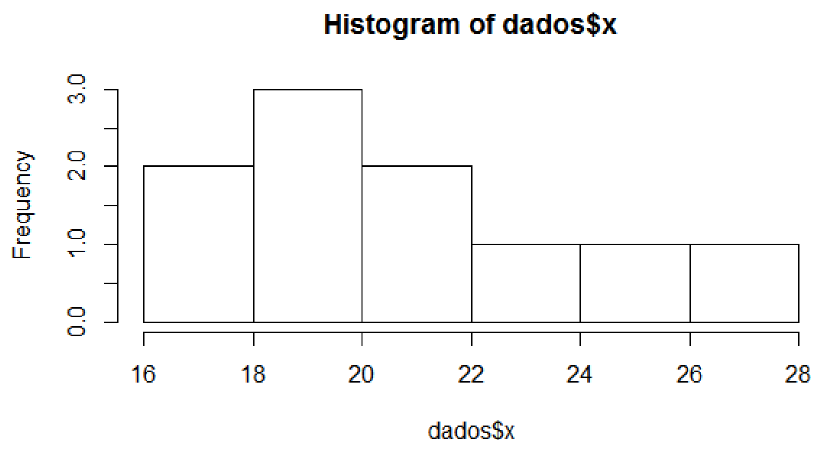
```
# Ler os dados de um arquivo (interagindo com o usuário)
dados <- read.table(file.choose(),header=TRUE)

# Mostra boxplots para as variáveis x e y
boxplot(dados$x, dados$y)
```

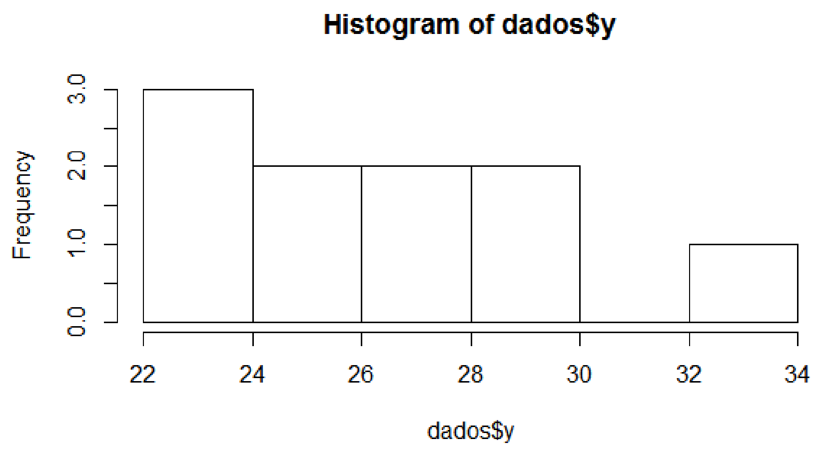


Teste de Hipóteses em R

- Solução:

```r
# Plota o hitograma para as duas variáveis
hist(dados$x)
hist(dados$y)
```





Teste de Hipóteses em R

- Solução:

```r
# Aplica o teste de normalidade
shapiro.test(dados$x)
shapiro.test(dados$y)
```

```
        Shapiro-Wilk normality test

data:  dados$x
W = 0.97856, p-value = 0.957
```

```
        Shapiro-Wilk normality test

data:  dados$y
W = 0.96042, p-value = 0.7906
```

Teste de Hipóteses em R

- Solução:

```
# Alica o teste de hipóteses
t.test(dados$x, dados$y, var.equal = TRUE, conf.level = 0.90)
```

```
        Two Sample t-test

data:  dados$x and dados$y
t = -3.6739, df = 18, p-value = 0.001737
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
 -8.095952 -2.904048
sample estimates:
mean of x mean of y 
     21.2      26.7 
```

Obs: como o *p-valor* é menor do que 0,05, então rejeita-se a hipótese H_0

Teste de Hipóteses em R

Teste *t* de *Student* para duas amostras aleatórias relacionadas (pareadas) (Exemplo prático - 3)

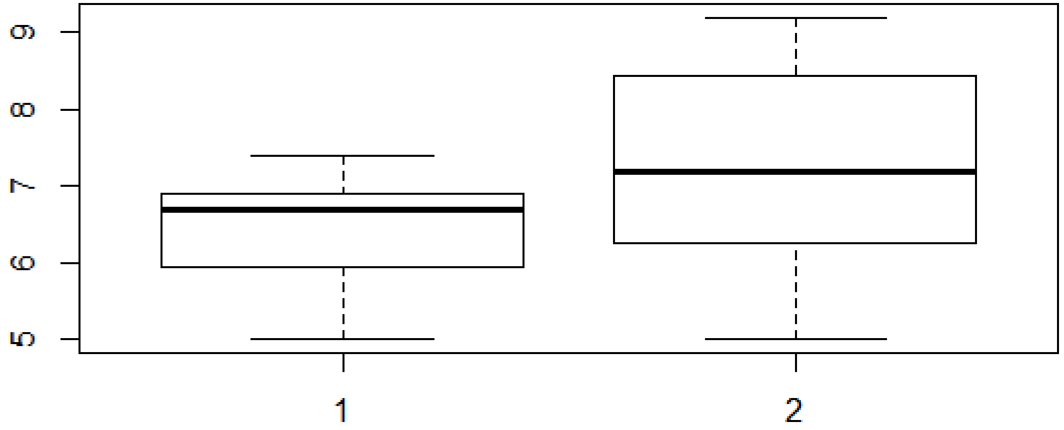
- Um grupo de funcionários foi submetido a um treinamento, e o objetivo é verificar o desempenho deles antes e depois do curso. Para tanto, foram atribuídas notas para cada funcionário de 1 a 10, antes e depois do treinamento. Utilizou-se uma significância de $\alpha = 5\%$;

Momento	Notas atribuídas										
Antes	5,5	6,1	6,7	6,2	7,0	7,2	5,8	6,8	6,7	7,4	5,0
Depois	6,0	7,2	6,8	8,2	9,0	5,8	6,5	7,2	8,7	5,0	9,2

Teste de Hipóteses em R

- Solução:

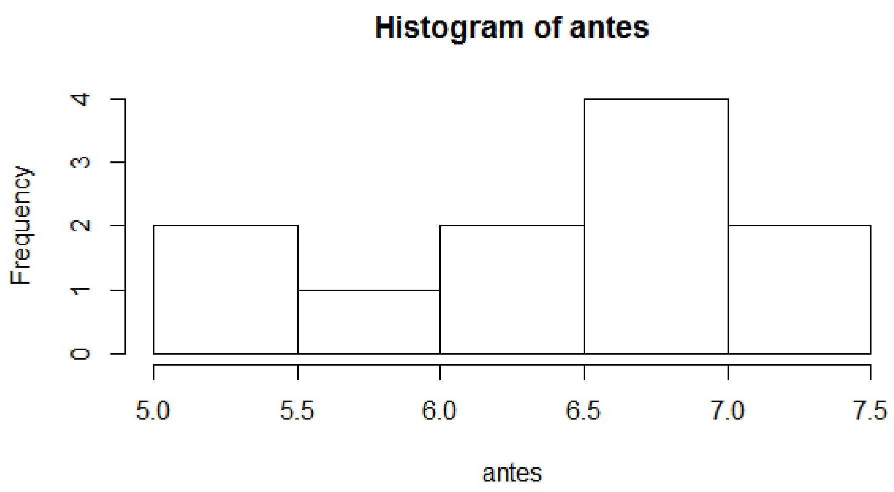
```r
# Construíndo a base de dados
antes <- c(5.5,6.1,6.7,6.2,7.0,7.2,5.8,6.8,6.7,7.4,5.0)
depois <-c(6.0,7.2,6.8,8.2,9.0,5.8,6.5,7.2,8.7,5.0,9.2)

# Mostra boxplots para as variáveis antes e depois
boxplot(antes, depois)
```

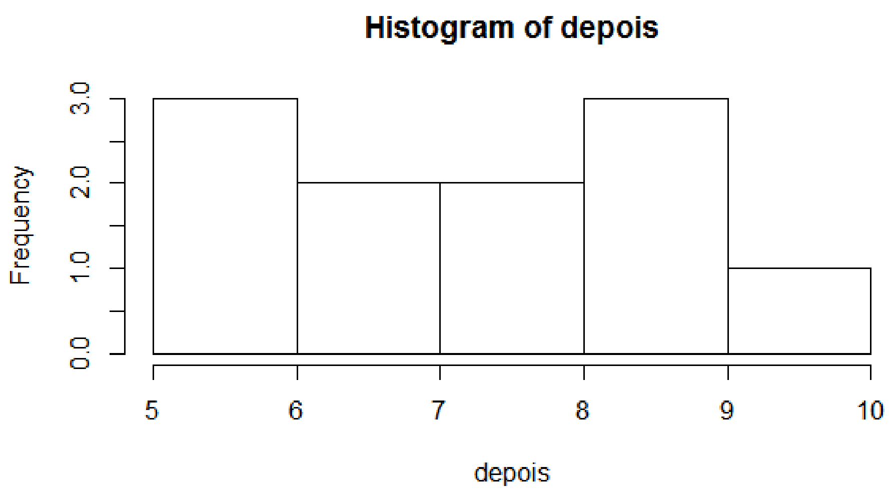


Teste de Hipóteses em R

- Solução:

```r
# Plota o hitograma para as duas variáveis
hist(antes)
hist(depois)
```





Teste de Hipóteses em R

- Solução:

```r
# Aplica o teste de normalidade
shapiro.test(antes)
shapiro.test(depois)
```

```
	Shapiro-Wilk normality test

data:  antes
W = 0.95618, p-value = 0.7234
```

```
	Shapiro-Wilk normality test

data:  depois
W = 0.95241, p-value = 0.6749
```

Teste de Hipóteses em R

- Solução:

```r
# Alica o teste de hipóteses
t.test(antes, depois, paired = TRUE, conf.level = 0.95)
```

```
        Paired t-test

data:  antes and depois
t = -1.5559, df = 10, p-value = 0.1508
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.0340969  0.3613697
sample estimates:
mean of the differences
             -0.8363636
```

Obs: como o *p-valor* é maior do que 0,05, então aceita-se a hipótese H_0

Dúvidas ?

- **Contatos:**
  - ✓Email: rodrigo.linsrodrigues@ufrpe.br
  - ✓Facebook: /rodrigomuribec