CE003
Estatística II

Silvia Shimakura

silvia.shimakura@ufpr.br

Laboratório de Estatística e Geoinformação

Estatística Inferencial

Estimação, Intervalos de Confiança, Testes de hipóteses

Estatística Inferencial

- Populações X Amostras
- Parâmetros X Estimativas
- Estimativas: Pontuais ou Intervalares
- Testes de Hipóteses

Teoria Elementar da Amostragem

- Teoria da amostragem
  - Retira informação sobre a **população** a partir de **amostras**
  - Estimativas pontuais ou intervalares
  - Testes de Hipóteses
- Números e amostras aleatórias
  - As **conclusões** da teoria de amostragem e da inferência estatística serão **válidas** se as amostras forem **representativas** da população
  - Um método para obter amostras representativas é a **amostragem aleatória simples**

Definições

Parâmetro: é uma medida numérica, em geral desconhecida, que descreve uma característica de interesse da população.

São representados, geralmente, por letras gregas tais como, μ (média populacional) e σ (desvio padrão populacional)

Usaremos a letra p para representar a proporção populacional.

Definições

Estatística: é qualquer valor calculado a partir dos dados amostrais.

Exemplos:

- $\bar{X}$ (média amostral)
- **S** (desvio padrão amostral)
- $\hat{p}$ (proporção amostral)

A **estatística** é uma variável aleatória pois:

- é uma quantidade incerta (antes de obter a amostra não sabemos seu valor)
- seu valor varia de amostra para amostra

Estimador x Estimativa

Estimador e Estimativa: uma estatística destinada a estimar um parâmetro é chamada estimador.

Dada uma amostra, o valor assumido pelo estimador é chamado de estimativa ou valor estimado do parâmetro.

Exemplo

Consideremos uma população formada por 5 alunos

A população de alunos tem:

- idade média μ=20,4 anos
- desvio padrão σ=1,36 anos
- 40% dos alunos são homens, ou seja, a proporção de homens é p=0,40
- A média e o desvio padrão das idades e a proporção de homens descrevem a população de alunos, portanto são **parâmetros**.

Identificação dos alunos	Idade (anos)	Sexo
1	22	f
2	19	f
3	19	m
4	20	f
5	22	m

Inferência estatística

Os procedimentos básicos de inferência estatística compreendem duas metodologias:

- **Estimação:** usamos o resultado amostral para estimar o valor desconhecido do parâmetro
- **Teste de hipóteses:** usamos o resultado amostral para avaliar se uma afirmação sobre o parâmetro (uma hipótese) é sustentável ou não.

Estimação

Embora tenhamos acesso a todos os dados de sexo e idade dos 5 alunos, vamos recorrer a amostragem para estimar μ usando $\bar{X}$, a idade média amostral, e p usando $\hat{p}$, a proporção amostral de homens.

Vamos tomar todas as amostras possíveis de tamanho 2, que podem ser selecionadas da população dos 5 alunos, por amostragem aleatória simples com reposição.

Para cada amostra i, podemos calcular $\bar{X}_i$, uma estimativa para a idade média e $\hat{p}_i$, uma estimativa da proporção de homens.

Amostra i	Alunos selecionados	Dados amostrais	$\bar{X}_i$	$\hat{p}_i$
1	1 e 1	22 f, 22 f	22,0	0,0
2	1 e 2	22 f, 19 f	20,5	0,0
3	1 e 3	22 f, 19 m	20,5	0,5
4	1 e 4	22 f, 20 f	21,0	0,0
5	1 e 5	22 f, 22 m	22,0	0,5
6	2 e 1	19 f, 22 f	20,5	0,0
7	2 e 2	19 f, 19 f	19,0	0,0
8	2 e 3	19 f, 19 m	19,0	0,5
9	2 e 4	19 f, 20 f	19,5	0,0
10	2 e 5	19 f, 22 m	20,5	0,5
11	3 e 1	19 m, 22 f	20,5	0,5
12	3 e 2	19 m, 19 f	19,0	0,5
13	3 e 3	19 m, 19 m	19,0	1,0
14	3 e 4	19 m, 20 f	19,5	0,5
15	3 e 5	19 m, 22 m	20,5	1,0
16	4 e 1	20 f, 22 f	21,0	0,0

Estimativa de p

Estimativa de μ

Temos estimativas variadas de μ e p

Amostra i	Alunos selecionados	Dados amostrais	$\bar{X}_i$	$\hat{p}_i$
1	1 e 1	22 f, 22 f	22,0	0,0
2	1 e 2	22 f, 19 f	20,5	0,0
3	1 e 3	22 f, 19 m	20,5	0,5
4	1 e 4	22 f, 20 f	21,0	0,0
5	1 e 5	22 f, 22 m	22,0	0,5
6	2 e 1	19 f, 22 f	20,5	0,0
7	2 e 2	19 f, 19 f	19,0	0,0
8	2 e 3	19 f, 19 m	19,0	0,5
9	2 e 4	19 f, 20 f	19,5	0,0
10	2 e 5	19 f, 22 m	20,5	0,5
11	3 e 1	19 m, 22 f	20,5	0,5
12	3 e 2	19 m, 19 f	19,0	0,5
13	3 e 3	19 m, 19 m	19,0	1,0
14	3 e 4	19 m, 20 f	19,5	0,5
15	3 e 5	19 m, 22 m	20,5	1,0
16	4 e 1	20 f, 22 f	21,0	0,0

Estimativa de p

Estimativa de μ

Média de $\bar{X}$:

$\sum \bar{X}_i/25=20,4=\mu$

Média de $\hat{p}$:

$\sum \hat{p}_i/25=0,4=p$

Em média acertamos os valores dos parâmetros!

Amostra i	Alunos selecionados	Dados amostrais	$\bar{X}_i$	$\hat{p}_i$
1	1 e 1	22 f, 22 f	22,0	0,0
2	1 e 2	22 f, 19 f	20,5	0,0
3	1 e 3	22 f, 19 m	20,5	0,5
4	1 e 4	22 f, 20 f	21,0	0,0
5	1 e 5	22 f, 22 m	22,0	0,5
6	2 e 1	19 f, 22 f	20,5	0,0
7	2 e 2	19 f, 19 f	19,0	0,0
8	2 e 3	19 f, 19 m	19,0	0,5
9	2 e 4	19 f, 20 f	19,5	0,0
10	2 e 5	19 f, 22 m	20,5	0,5
11	3 e 1	19 m, 22 f	20,5	0,5
12	3 e 2	19 m, 19 f	19,0	0,5
13	3 e 3	19 m, 19 m	19,0	1,0
14	3 e 4	19 m, 20 f	19,5	0,5
15	3 e 5	19 m, 22 m	20,5	1,0
16	4 e 1	20 f, 22 f	21,0	0,0

Estimativa de p

Estimativa de μ

Para as variâncias de $\bar{X}$ e $\hat{p}$, temos um outro resultado interessante.

Denotando tamanho da amostra por n, podemos mostrar também que a variância de $\bar{X}$ é:
$\sigma^2/n=1{,}36^2/2=0{,}92$

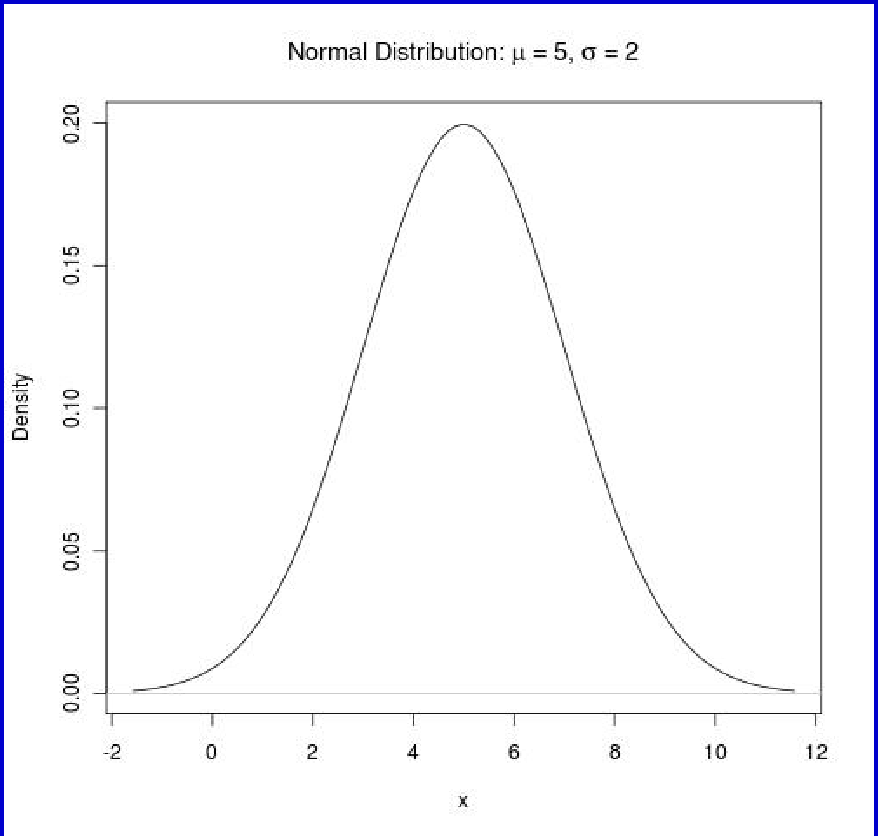
e a variância de $\hat{p}$ é:

$p(1-p)/n=(0{,}4*0{,}6)/2=0{,}12$

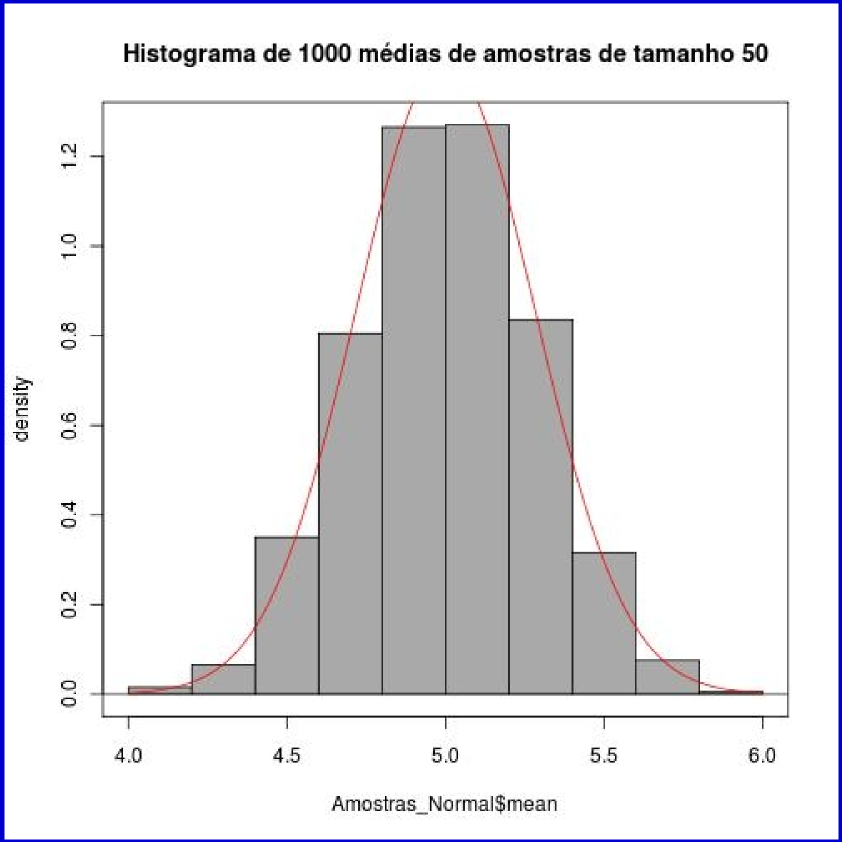
Teorema Central do Limite

- Valores estatísticos amostrais
  - Valores estatísticos obtidos de amostras são eles próprios variáveis
  - Assim, podem ser definidas distribuições a valores estatísticos amostrais
- Teorema central do limite
  - As **médias de amostras** de tamanho n retiradas de uma população normal **têm sempre uma distribuição normal**
  - As médias de amostras de tamanho n retiradas de uma população não normal têm uma distribuição que **tende para a normal à medida que n aumenta** (geralmente, a partir de $n \geq 30$ é já uma boa aproximação da normal)

Exemplo: TCL



Normal Distribution: μ = 5, σ = 2
Density
0.00
0.05
0.10
0.15
0.20
-2
0
2
4
6
8
10
12
x



Histograma de 1000 médias de amostras de tamanho 50
density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
4.0
4.5
5.0
5.5
6.0
Amostras_Normal$mean

Teorema Central do Limite (cont.)

- A distribuição das médias amostrais tende para uma distribuição **$N(\mu, \sigma/\sqrt{n})$**

- Erro Padrão
  - **Erro Padrão** é o desvio padrão das estatísticas amostrais
  - Assim, o **Erro Padrão da Média=$\sigma/\sqrt{n}$** uma vez que é o desvio padrão das médias amostrais

Teoria da Estimação Paramétrica

- Estimação Paramétrica
  - Um dos problemas da estatística inferencial é a estimação de parâmetros populacionais, também designada por **Estimação Paramétrica**
- Estimação
  - **Pontual**
  - **Intervalar**

Teoria da Estimação Paramétrica

- Intervalos de Confiança para parâmetros populacionais
- Intervalos de Confiança (IC) para a Média

$$\left(\bar{X} \pm z \frac{\sigma}{\sqrt{n}} \right)$$

  - z é um valor da distribuição normal padrão
  - No caso do IC 95% ⇨ z = 1,96
  - No caso do IC 99% ⇨ z = 2,58

Intervalos de Confiança para a Média

- **Interpretação**

O intervalo ***$\mu \pm 1{,}96\ (\sigma/\sqrt{n})$*** contém 95% das possíveis médias amostrais, então, há uma probabilidade de 95% da média da nossa amostra estar dentro deste intervalo

Assim sendo, pode-se afirmar analogamente que 95% dos intervalos definidos por ***Média amostral ± 1,96 $(\sigma/\sqrt{n})$*** cobrem a média da população (μ)

O intervalo ***Média amostral ± 1,96 $(\sigma/\sqrt{n})$*** é chamado de **Intervalo de Confiança a 95% para a Média**

Distribuição t de Student e Teste de Hipóteses

Distribuição t de Student, Teste de Hipóteses, Teste t para uma média, teste t para a diferença entre duas médias e teste t para dados pareados

Distribuição t de Student

- Tendo em conta o Teorema Central do Limite, temos que:

$$\left(\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right) \sim N(0,1)$$

- Este resultado assume que σ é conhecido mas na prática não é.

Distribuição t de Student

- Para resolver este problema Gossett (1908), com o pseudonimo de Student, propôe uma distribuição que utiliza o desvio padrão da amostra ao invés do desvio padrão da população

$$t = \left(\frac{\bar{X} - \mu}{s / \sqrt{n}} \right)$$

- Se a variável em estudo segue uma distribuição normal, então t segue uma distribuição t de Student com n-1 graus de liberdade

Distribuição t de Student

- É semelhante à distribuição normal, mas com uma maior dispersão em torno do valor central
- Esta distribuição tem uma forma diferente em função do tamanho da amostra (n)
- À medida que n aumenta a distribuição tende para uma distribuição normal

Distribuição t de Student

- Assim, se não conhecermos o desvio padrão da população o **Intervalo de Confiança de 95% para a Média** poderá ser calculado do seguinte modo:

$$\left(\bar{X} \pm t_{(n-1\,;\,0{,}05)} \frac{s}{\sqrt{n}}\right)$$

Distribuição t de Student

Intervalo de Confiança a 95% para a Média:

Erro Padrão

IC 95% = Média da amostra ± $t_{(n-1)}$ $(s/\sqrt{n})$

Valor apropriado da distribuição t com (n-1) graus de liberdade

Exemplo:

Estatística descritiva (n=462)

			Estatística	Erro Padrão
Peso da criança ao nascer	Média		3263,23	25,752
	Intervalo de confiança a 95% para a média	Limite inferior	3212,62	
		Limite superior	3313,83	

IC 95% = 3263,23 ± $t_{(462-1)}$ (25,752)

IC 95% = 3263,23 ± 1,965 (25,752) = [3212,62; 3313,83]

Testes de Hipóteses

- Utilizando a mesma estrutura teórica que nos permite calcular Intervalos de Confiança podemos **testar hipóteses** sobre um parâmetro populacional

Exemplo: Queremos testar a hipótese de que a altura média de uma certa população é 160 cm. Numa amostra aleatória de 9 pessoas a altura média amostral foi 170 cm com desvio padrão amostral de 10 cm.

Qual é a probabilidade de se obter uma média amostral tão distante, ou ainda mais distante, da hipótese inicial de 160 cm?

Testes de Hipóteses

- Utilizando a mesma estrutura teórica que nos permite calcular Intervalos de Confiança podemos **testar hipóteses** sobre um parâmetro populacional

Exemplo: Queremos testar a hipótese de que a altura média de uma certa população é 160 cm. Numa amostra aleatória de 9 pessoas a altura média amostral foi 170 cm com desvio padrão amostral de 10 cm.

Qual é a probabilidade de se obter uma média amostral tão distante, ou ainda mais distante, da hipótese inicial de 160 cm?

Se essa probabilidade for muito baixa, podemos rejeitar a hipótese inicial.

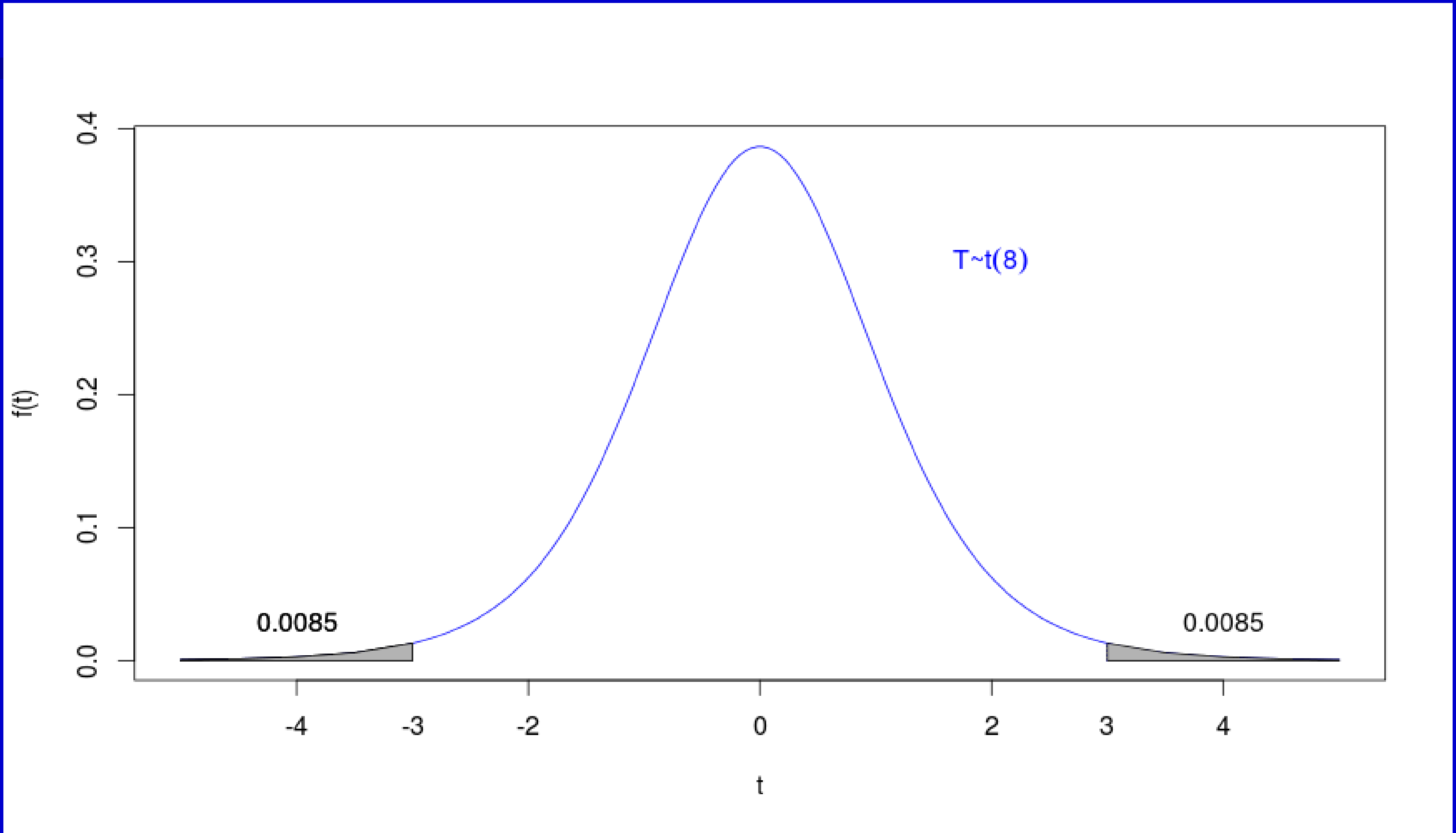
Teste t para uma média

H_0: μ=160cm x H_A: μ≠160cm

$$n=9 \quad \bar{X}=170\,cm \quad s=10\,cm$$

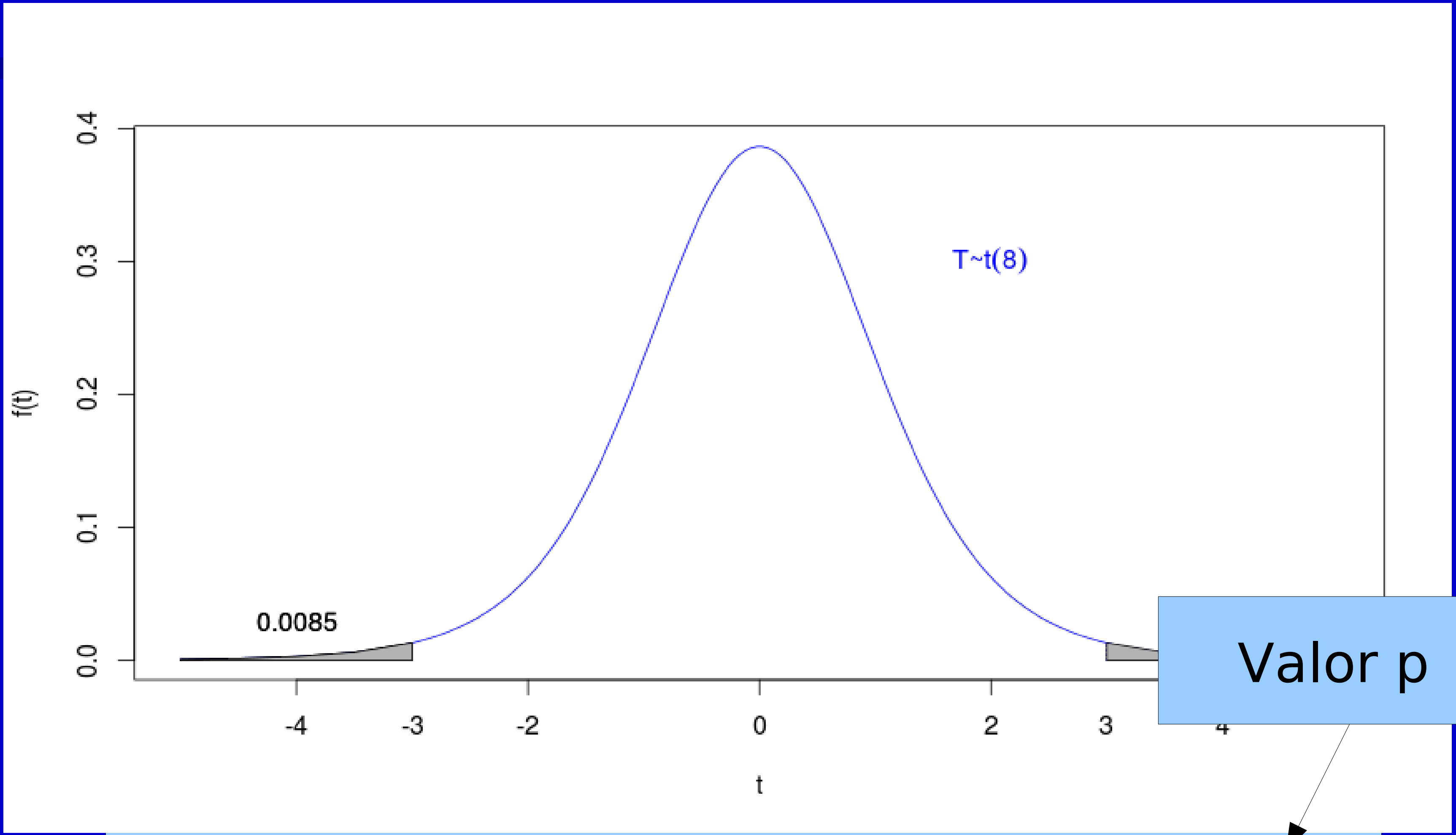
$$T=\left|\frac{\bar{X}-\mu}{s/\sqrt{n}}\right|=\left|\frac{170-160}{10/\sqrt{9}}\right|=3 \;\sim\; t_{(9-1)}=t_8$$

Teste t para uma média



$$P(T<-3)+P(T>3)=2\times 0{,}0085=0{,}017$$

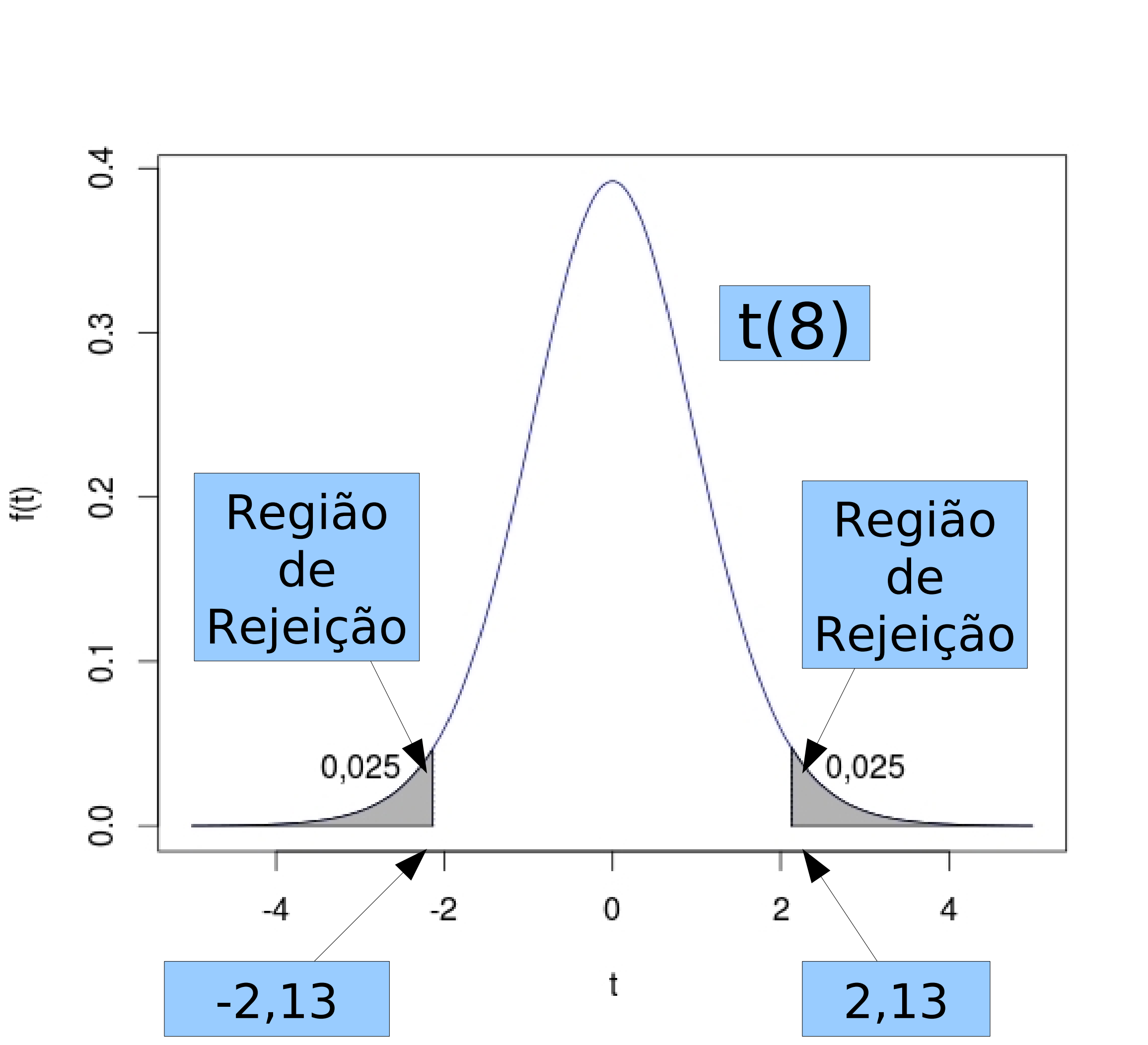
Teste t para uma média



Valor p

$$P(T<-3)+P(T>3)=2\times 0{,}0085=0{,}017$$

Teste t para uma média



Teste t para uma média

- Suposição:
  - Distribuição normal ou aproximadamente normal da variável de interesse

Teste t para uma média

1. Especificar H_0 e H_A

$H_0: \mu = \mu_0$ $\qquad$ $H_A: \mu \neq \mu_0$

2. Escolher o nível de significância (α = 5%)
3. Calcular a estatística de teste

$$T = \left| \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \right|$$

4. Comparar o valor de T com uma distribuição de t com n-1 graus de liberdade
5. Calcular o valor de p e comparar com α
6. Descrever os resultados e conclusões estatísticas

Tipos de Erros

- Erro tipo I (α)

Probabilidade de rejeitar a H0 quando H0 é verdadeira

- Erro tipo II (β)

Probabilidade de não rejeitar a H0 quando H0 é falsa

Exemplo:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Birthweight	462	3263,23	553,516	25,752

Valor de p

H0: μ = 3500 g; HA: μ ≠ 3500 g

One-Sample Test

	Test Value = 3500					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Birthweight	-9,194	461	,000	-236,77	-287,38	-186,17

Teste t para a diferença entre duas médias

1. Especificar H_0 e H_A

$H_0: \mu_1 = \mu_2$ $\quad$ $H_A: \mu_1 \neq \mu_2$

$H_0: \mu_1 - \mu_2 = 0$ $\quad$ $H_A: \mu_1 - \mu_2 \neq 0$

2. Escolher o nível de significância (α = 0,05 ou 5%)

3. Calcular a estatística e a estatística de teste

Média das duas amostras

$t = [(\text{Média 1} - \text{Média 2}) - (\mu_1 - \mu_2)] / [s_{(\text{Média 1} - \text{Média 2})}]$

4. Comparar o valor de t com uma distribuição de t com $(n_1 + n_2 - 2)$ graus de liberdade

5. Calcular o valor de p

6. Comparar p e α

7. Descrever os resultados e conclusões estatísticas

Teste t para a diferença entre duas médias

- Suposições:
  - Distribuição normal ou aproximadamente normal da variável nos dois grupos
  - Independência entre os grupos

Exemplo:

Group Statistics

	Premature birth?	N	Mean	Std. Deviation	Std. Error Mean
Birthweight	No	401	3367,13	442,718	22,108
	Yes	59	2558,98	697,190	90,766

Valor de p

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Birthweight	Equal variances assumed	22,954	,000	12,014	458	,000	808,15	67,268	675,959	940,344
	Equal variances not assumed			8,651	65,053	,000	808,15	93,420	621,582	994,722

Teste t para a diferença entre duas médias

Group Statistics

	Sex of baby	N	Mean	Std. Deviation	Std. Error Mean
Birthweight	Male	250	3290,02	580,145	36,692
	Female	212	3231,63	519,954	35,711

Valor de p

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Birthweight	Equal variances assumed	1,265	,261	1,130	460	,259	58,39	51,663	-43,138	159,913
	Equal variances not assumed			1,140	458,577	,255	58,39	51,201	-42,229	159,005

Exemplo: Birthweight (cont.)

- Dados>Modificação de variáveis...>Converter variável numérica...
- Estatísticas>Variâncias>Teste de Levene
- Estatísticas>Médias>Teste t para amostras independentes

Rcmdr: Convertendo variável numérica

Rcmdr: Teste de Levene

Rcmdr: Teste t para amostras independentes

Teste t para dados pareados

1. Especificar H_0 e H_A

$H_0: \mu_d = 0$ $H_A: \mu_d \neq 0$

2. Escolher o nível de significância (α = 0,05 ou 5%)

3. Calcular a estatística e a estatística de teste

Média das duas amostras

t = (Média das diferenças - μ_d) / $s_{(\text{diferenças})}$

4. Comparar o valor de t com uma distribuição de t com (n-1) graus de liberdade

5. Calcular o valor de p

6. Comparar p e α

7. Descrever os resultados e conclusões estatísticas

Teste t para dados pareados

- Assume-se
  - Distribuição normal ou aproximadamente normal das diferenças
  - Dependência (correlação) entre os grupos

Teste t para dados pareados

- Exemplo:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Score na escala de depressão antes do tratamento	62,10	10	7,249	2,292
	Score na escala de depressão depois do tratamento	55,80	10	11,545	3,651

Valor de p

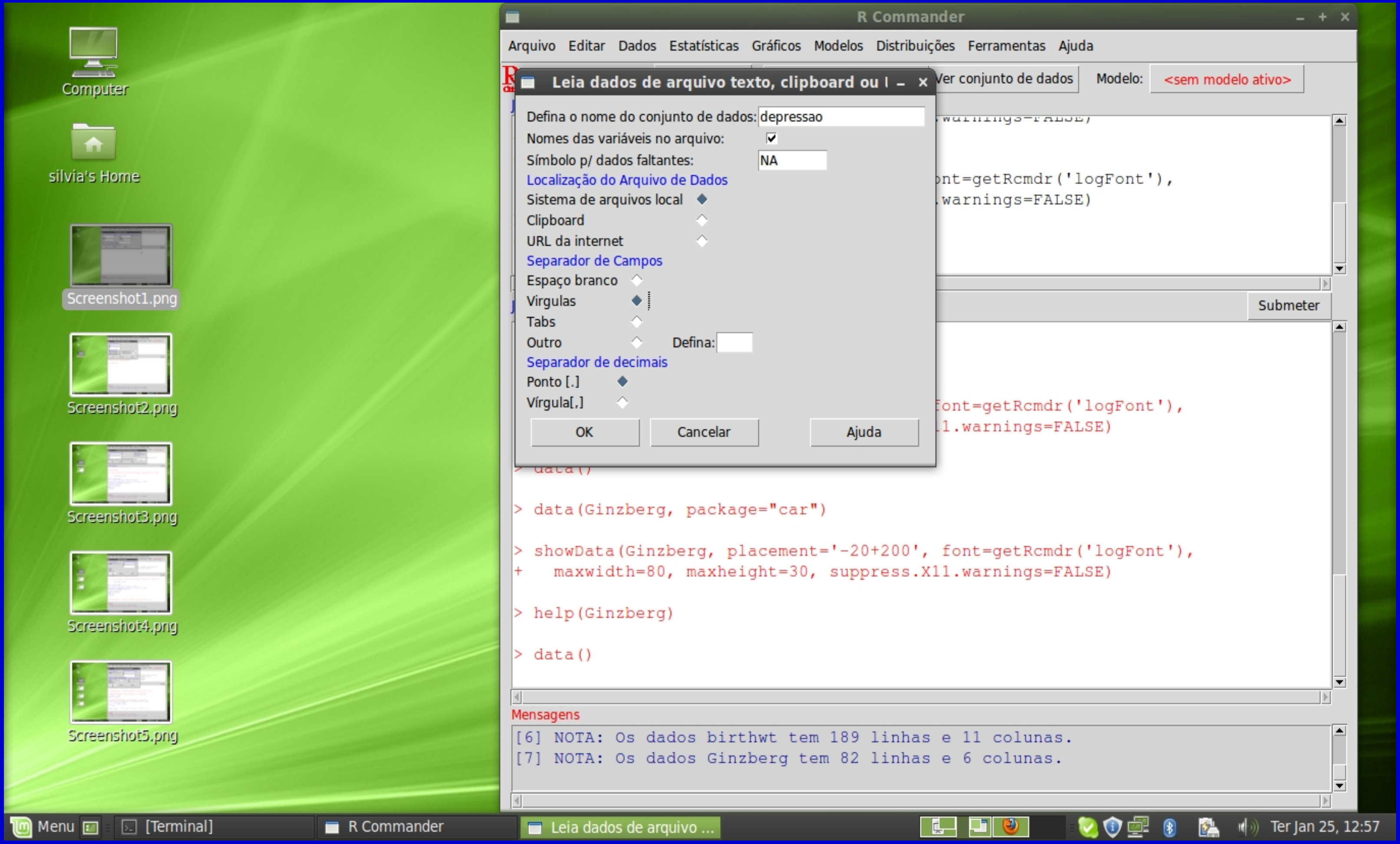
Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	Score na escala de depressão antes do tratamento - Score na escala de depressão depois do tratamento	6,30	9,298	2,940	-,35	12,95	2,143	9	,061

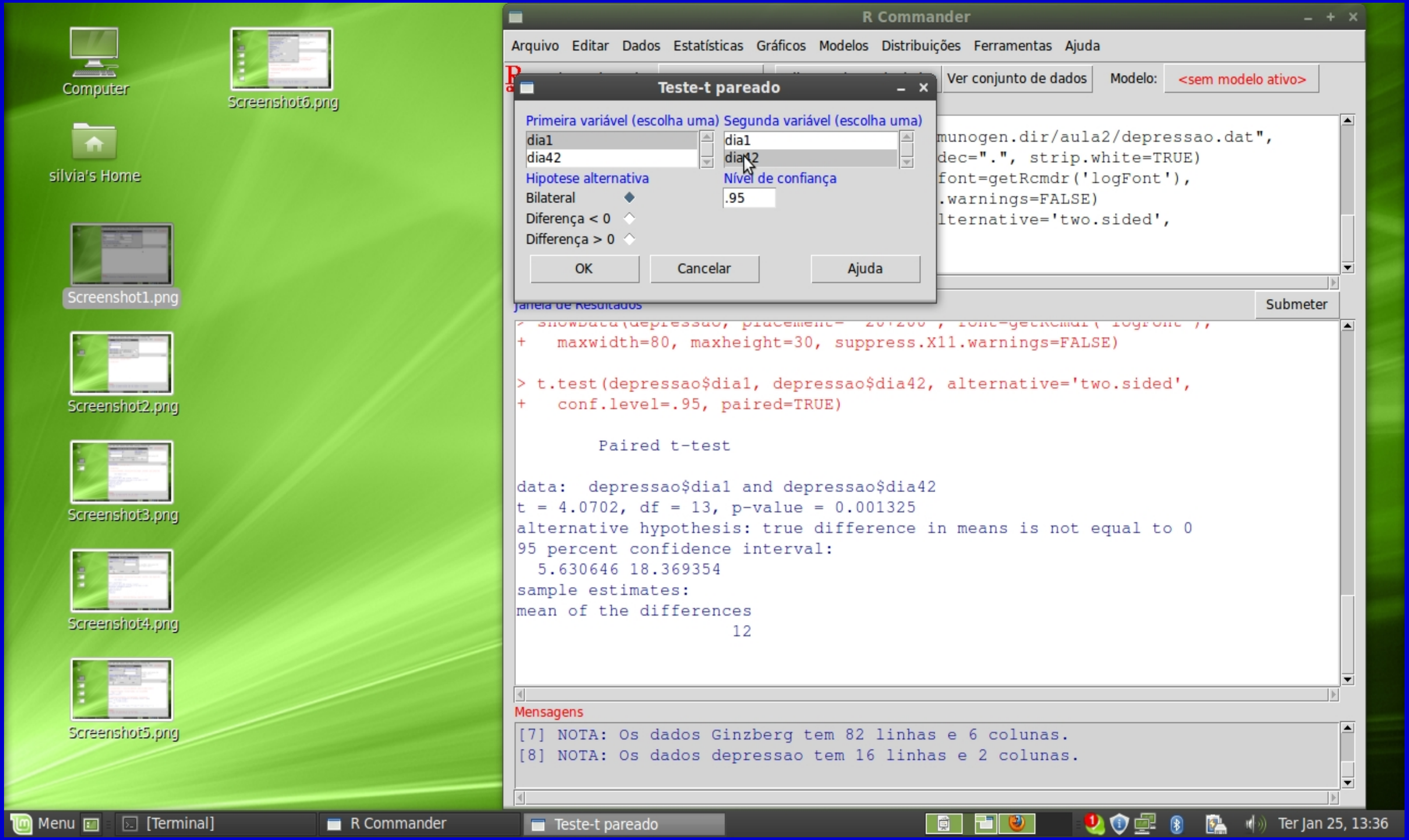
Exemplo: Escores de depressão

- Dados>Importar arquivos de dados>de arquivo texto...
- Estatísticas>Médias>Teste t (dados pareados)

Rcmdr: Lendo banco de dados de arquivo texto



Rcmdr: Teste t para dados pareados



ANOVA

Análise de variância

ANOVA

- Comparação de médias de 2 grupos

  Teste t

  $H_0: \mu_1=\mu_2$ Erro tipo I (α) = 1-0,95 = 0,05

- Mais de 2 grupos:

Ex: $H_0: \mu_1 =\mu_2 =\mu_3$

(1) $H_0: \mu_1=\mu_2$ (2) $H_0: \mu_1=\mu_3$ (3) $H_0: \mu_2=\mu_3$

Erro tipo I = $1-0{,}95^3$ = 0,14

- Comparação de médias de mais de 2 grupos

  ANOVA

$$H_0: \mu_1 =\mu_2 =\mu_3 =\ldots =\mu_k$$

ANOVA

- Considere um conjunto de k grupos, com n_i indivíduos cada um, um total de n indivíduos, uma média de cada grupo x_i e uma média comum X

Ex: Considere os pesos (em kg) de 3 grupos de indivíduos de grupos étnicos diferentes (caucasianos, latinos e asiáticos).

Grupo 1: 80; 75; 82; 68; 76; 86; 78; 90; 85; 64 → x_1= 78,40 kg

Grupo 2: 65; 84; 63; 54; 86; 62; 73; 64; 69; 81 → x_2= 70,10 kg

Grupo 3: 58; 59; 61; 63; 71; 53; 54; 72; 61; 57 → x_3= 60,90 kg

X=69,80 kg k = 3

n_1=10 n_2=10 n_3=10 n = 30

ANOVA

- Fontes de variação:
  - **Intra-grupos** - Variabilidade das observações em relação à média do grupo

    - **Within group SS**

(sum of squares)

$$\sum_{i=1}^{k}\left[\sum_{j=1}^{n_i}\left(x_{ij}-\overline{X}_i\right)^2\right]$$

    - **Within group DF**

(degrees of freedom)

$$\sum_{i=1}^{k}\left(n_i-1\right)=n-k$$

    - **Within group MS**

(mean square = variance)

$$\frac{\text{Withingroup SS}}{\text{Withingroup DF}}$$

ANOVA

- Fontes de variação:
  - **Entre-grupos** - Variabilidade entre os grupos. Dependente da média do grupo em relação à média conjunta
    - **Between group SS** $$\sum_{i=1}^{k} n_i \left(\overline{X}_i - \overline{X} \right)^2$$
    - **Between group DF** $$k-1$$
    - **Between group MS** $$\frac{\text{Between group SS}}{\text{Between group DF}}$$

ANOVA

- A variabilidade observada num conjunto de dados deve-se a:
  - Variação em relação à média do grupo - Within group MS
  - Variação da média do grupo em relação à média comum - Between group MS

ANOVA

- Prova-se que se $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$, então, Between MS e Within MS serão ambas estimativas de σ^2 - a variância comum aos k grupos - logo, Between MS $\approx$ Within MS
- Se pelo contrário $\mu_1 \neq \mu_2 \neq \mu_3 \neq \ldots \neq \mu_k$, então, Between MS será maior que Within MS
- Assim, para testar a Hipótese nula

  H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ calcula-se a estatística F

$$F = \frac{\text{Between group MS}}{\text{Within group MS}}$$

ANOVA

- A estatística F tem uma distribuição teórica conhecida - Distribuição F - dependente dos graus de liberdade Between DF e Within DF
- O cálculo da estatística F e seu enquadramento na distribuição adequada permite-nos conhecer um valor de p - probabilidade de obter um F tão ou mais extremo que o calculado se a hipótese nula for verdadeira
- O valor de p é subsequentemente comparado com o grau de significância (α) à partida estabelecido e
  - **Se $p \leq \alpha$, rejeita-se a H_0 => Existem diferenças estatisticamente significativas entre as médias dos grupos**
  - **Se $p > \alpha$, aceita-se a H_0 => Não existem diferenças estatisticamente significativas entre as médias dos grupos**

ANOVA

- Suposições:
  - Normalidade
  - Igualdade das variâncias dos grupos

- Funciona melhor se:
  - Igual tamanho dos grupos
  - Igualdade dos grupos exceto na variável de interesse

Exemplo:

Descriptives

Peso do indivíduo (Kg)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Caucasiano	10	78,40	8,06	2,55	72,64	84,16	64	90
Latino	10	70,10	10,61	3,35	62,51	77,69	54	86
Asiático	10	60,90	6,38	2,02	56,33	65,47	53	72
Total	30	69,80	10,98	2,00	65,70	73,90	53	90

Test of Homogeneity of Variances

Peso do indivíduo (Kg)

Levene Statistic	df1	df2	Sig.
1,862	2	27	,175

ANOVA

Valor de p

ANOVA

Peso do indivíduo (Kg)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1532,600	2	766,300	10,534	,000
Within Groups	1964,200	27	72,748		
Total	3496,800	29			

Exemplo: Peso x raça

- Crie banco de dados do exemplo acima numa planilha e salve como txt
- Converter grupo em fator
- Realizar teste de Levene
- Fazer a Anova

peso	grupo
80	1
75	1
82	1
68	1
76	1
86	1
78	1
90	1
85	1
64	1
65	2
84	2
63	2
54	2
86	2
62	2
73	2

Testes Não Paramétricos

Mann-Whitney Test; Wilcoxon Signed Ranks Test; Kruskal-Wallis Test

Mann-Whitney Test

- Análogo ao teste t para a diferença entre duas médias
- Quando as assumpções necessárias para a utilização do teste t não são cumpridas (normalidade e igualdade de variâncias) tem que se optar pelos testes análogos não paramétricos
- Não faz assumpções sobre a distribuição da variável
- Faz uso das posições ordenadas dos dados (ranks) e não dos valores da variável obtidos

Mann-Whitney Test

- Ex: Para investigar se os mecanismos envolvidos nos ataques fatais de asma provocados por alergia à soja são diferentes dos mecanismos envolvidos nos ataques fatais de asma típica compararam-se o número de células T CD3+ na submucosa de indivíduos destes dois grupos.

Mann-Whitney Test

- Ex: situações possíveis (dois grupos A e B de 5 elementos cada um):

A	A	A	A	A	B	B	B	B	B		A	B	A	B	A	B	A	B	A	B
1º	2º	3º	4º	5º	6º	7º	8º	9º	10º		1º	2º	3º	4º	5º	6º	7º	8º	9º	10º

A e B diferentes

Não há diferenças entre A e B

- São calculadas as seguintes estatísticas:

R_1= soma das posições no grupo 1

R_2= soma das posições no grupo 2

Mann-Whitney Test

- A maior destas estatísticas é comparada com uma distribuição adequada (distribuição da estatística U ou aproximação normal)
- Obtem-se um valor de p - probabilidade de se obter uma estatística tão ou mais extrema do que a verificada caso a hipótese nula seja verdadeira
- O valor de p é subsequentemente comparado com o grau de significância (α) à partida estabelecido e
  - **Se $p \leq \alpha$, rejeita-se a H_0 => Existem diferenças estatisticamente significativas relativamente à distribuição da variável entre os grupos**
  - **Se $p > \alpha$, aceita-se a H_0 => Não existem diferenças estatisticamente significativas relativamente à distribuição da variável entre os grupos**

Mann-Whitney Test

- Exemplo:

Ranks

	Grupo	N	Mean Rank	Sum of Ranks
Número de células T CD3+ na submucosa (células/mm2)	Grupo de alergia à soja	7	4,57	32,00
	Grupo de asma típica	10	12,10	121,00
	Total	17		

Valor de p

Test Statistics[b]

	Número de células T CD3+ na submucosa (células/mm2)
Mann-Whitney U	4,000
Wilcoxon W	32,000
Z	-3,033
Asymp. Sig. (2-tailed)	,002
Exact Sig. [2*(1-tailed Sig.)]	,001[a]

a. Not corrected for ties.

b. Grouping Variable: Grupo

Wilcoxon Signed Ranks Test

- Análogo do teste t para pares emparelhados ou teste t para a diferença entre 2 médias de grupos dependentes
- Ex: Num ensaio de um fármaco antidepressivo obtêm-se os seguintes scores numa escala de depressão, antes e depois do tratamento:

Wilcoxon Signed Ranks Test

- Posicionam-se os valores absolutos das diferenças de forma ascendente e atribui-se o sinal da diferença à posição
- Calculam-se as seguintes estatísticas:

T+ = soma das posições com sinal positivo

T- = soma das posições com sinal negativo

- Utiliza-se a menor destas estatísticas, sendo esta comparada com uma distribuição adequada (distribuição da estatística T ou aproximação normal)

Wilcoxon Signed Ranks Test

- Obtem-se um valor de p - probabilidade de se obter uma estatística tão ou mais extrema do que a verificada caso a hipótese nula seja verdadeira
- O valor de p é subsequentemente comparado com o grau de significância (α) à partida estabelecido e
  - **Se $p \leq \alpha$, rejeita-se a H_0 => Existem diferenças estatisticamente significativas relativamente à distribuição da variável entre os grupos**
  - **Se $p > \alpha$, aceita-se a H_0 => Não existem diferenças estatisticamente significativas relativamente à distribuição da variável entre os grupos**

Wilcoxon Signed Ranks Test

- Exemplo:

Ranks

		N	Mean Rank	Sum of Ranks
Score na escala de depressão depois do tratamento - Score na escala de depressão antes do tratamento	Negative Ranks	7[a]	6,43	45,00
	Positive Ranks	3[b]	3,33	10,00
	Ties	0[c]		
	Total	10		

a. Score na escala de depressão depois do tratamento < Score na escala de depressão antes do tratamento

b. Score na escala de depressão depois do tratamento > Score na escala de depressão antes do tratamento

c. Score na escala de depressão antes do tratamento = Score na escala de depressão depois do tratamento

Valor de p

Test Statistics[b]

	Score na escala de depressão depois do tratamento - Score na escala de depressão antes do tratamento
Z	-1,786[a]
Asymp. Sig. (2-tailed)	,074

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Kruskal-Wallis Test

- Análogo da Análise de Variância (ANOVA) para a comparação das médias de 3 ou mais grupos
- Ex: Pesos em Kg de 3 grupos de indivíduos de grupos étnicos diferentes (caucasianos, latinos e asiáticos).

Grupo 1: 80; 75; 82; 68; 76; 86; 78; 90; 85; 64

Grupo 2: 65; 84; 63; 54; 86; 62; 73; 64; 69; 81

Grupo 3: 58; 59; 61; 63; 71; 53; 54; 72; 61; 57

Organizam-se todos os valores por ordem crescente de modo a cada valor ter uma posição atribuída

Kruskal-Wallis Test

- Calcula-se a estatística:

- **N** = nº total de indivíduos; $\mathbf{n_i}$ = nº de indivíduos no grupo i e $\mathbf{R_i}$ = soma das posições no grupo i
- Esta estatística será comparada com uma distribuição adequada (distribuição de Qui-quadrado com k-1 graus de liberdade)

Kruskal-Wallis Test

- Obtem-se um valor de p - probabilidade de se obter uma estatística tão ou mais extrema do que a verificada caso a hipótese nula seja verdadeira
- O valor de p é subsequentemente comparado com o grau de significância (α) à partida estabelecido e
  - **Se $p \leq \alpha$, rejeita-se a H_0 => Existem diferenças estatisticamente significativas relativamente à distribuição da variável entre os grupos**
  - **Se $p > \alpha$, aceita-se a H_0 => Não existem diferenças estatisticamente significativas relativamente à distribuição da variável entre os grupos**

Kruskal-Wallis Test

- Exemplo:

Ranks

	Grupo étnico	N	Mean Rank
Peso do indivíduo (Kg)	Caucasiano	10	22,40
	Latino	10	16,20
	Asiático	10	7,90
	Total	30	

Valor de p

Test Statistics[a,b]

	Peso do indivíduo (Kg)
Chi-Square	13,675
df	2
Asymp. Sig.	,001

a. Kruskal Wallis Test

b. Grouping Variable: Grupo étnico

Tabelas de Contingência e Teste Qui-quadrado

Tabelas de contingência; teste qui-quadrado; teste exato de Fisher; correção de Yates; teste de McNemar; teste qui-quadrado para tendências

Tabelas de Contingência

- Forma de representar a relação entre duas variáveis categóricas. Distribuição das frequências das categorias de uma variável em função das categorias de uma outra variável.

Region of the United States * Race of Respondent Crosstabulation

			Race of Respondent			
			White	Black	Other	Total
Region of the United States	North East	Count	582	82	15	679
		% within Region of the United States	85,7%	12,1%	2,2%	100,0%
		% within Race of Respondent	46,0%	40,2%	30,6%	44,8%
		% of Total	38,4%	5,4%	1,0%	44,8%
	South East	Count	307	94	14	415
		% within Region of the United States	74,0%	22,7%	3,4%	100,0%
		% within Race of Respondent	24,3%	46,1%	28,6%	27,4%
		% of Total	20,2%	6,2%	,9%	27,4%
	West	Count	375	28	20	423
		% within Region of the United States	88,7%	6,6%	4,7%	100,0%
		% within Race of Respondent	29,7%	13,7%	40,8%	27,9%
		% of Total	24,7%	1,8%	1,3%	27,9%
Total		Count	1264	204	49	1517
		% within Region of the United States	83,3%	13,4%	3,2%	100,0%
		% within Race of Respondent	100,0%	100,0%	100,0%	100,0%
		% of Total	83,3%	13,4%	3,2%	100,0%

Teste Qui-quadrado

- Quando estamos perante duas variáveis categóricas podemos usar o teste qui-quadrado para testar a hipótese da existência de uma associação entre as variáveis na população.
- As hipóteses nula e alternativa que serão testadas são:
  – H_0: Não existe uma associação entre as categorias de uma variável e as da outra variável na população ou as proporções de indivíduos nas categorias de uma variável não variam em função das categorias da outra variável na população
  – H_A: Existe uma associação entre as categorias de uma variável e as da outra variável na população ou as proporções de indivíduos nas categorias de uma variável variam em função das categorias da outra variável na população

Teste Qui-quadrado

- Podem-se apresentar os dados numa tabela de contingência $r \times c$ (r - nº de linhas; c - nº de colunas). As entradas da tabela são frequências e cada célula contem o nº de indivíduos que pertencem simultaneamente àquela linha e coluna.
- Calcula-se as frequências esperadas caso a hipótese nula fosse verdadeira. A frequência esperada numa determinada célula é o produto do total da linha e do total da coluna dividido pelo total global.
- Baseada na estatística de teste (χ^2): discrepância entre as **frequências observadas** e as **frequências esperadas**, caso a H_0 seja verdadeira, em cada célula da tabela. Se a discrepância for grande é improvável que a hipótese nula seja verdadeira.

Teste Qui-quadrado

- A estatística de teste calculada (χ^2) tem a seguinte forma genérica:

  O - frequência observada na célula e E - frequência esperada na célula, caso a H_0 seja verdadeira.
- A tabela de contingência tem a seguinte forma genérica:

Teste Qui-quadrado

- A estatística de teste segue a Distribuição de Qui-quadrado com $(r-1)\times(c-1)$ graus de liberdade.
- O cálculo da estatística χ^2 e seu enquadramento na distribuição adequada permite-nos conhecer um valor de p (probabilidade de obter um χ^2 tão ou mais extremo que o calculado se a hipótese nula for verdadeira)
- O valor de p é comparado com o grau de significância (α):
  - **Se $p \leq \alpha$, rejeita-se a H_0 =>** Existe uma associação entre as categorias de uma variável e as da outra variável na população **ou** as proporções de indivíduos nas categorias de uma variável variam em função das categorias da outra variável na população
  - **Se $p > \alpha$, não rejeita-se a H_0 =>** Não existe evidência suficiente de uma associação entre as categorias de uma variável e as da outra variável na população

Teste Qui-quadrado

- **Ex:** Num ensaio clínico compara-se a eficácia de um Medicamento X (n=30 indivíduos) em relação ao placebo (n=32 indivíduos) na melhoria do estado clínico dos doentes 6 meses após o tratamento (melhorado, agravado, falecido).

Estado clínico 6 meses após o tratamento * Tramento efectuado Crosstabulation

			Tramento efectuado		
			Placebo	Medicamento X	Total
Estado clínico 6 meses após o tratamento	Melhorado	Count	9	17	26
		Expected Count	13,4	12,6	26,0
	Agravado	Count	12	9	21
		Expected Count	10,8	10,2	21,0
	Falecido	Count	11	4	15
		Expected Count	7,7	7,3	15,0
Total		Count	32	30	62
		Expected Count	32,0	30,0	62,0

E_{11}= (26*32)/62= 13,4

E_{12}= (26*30)/62= 12,6

E_{21}= (21*32)/62= 10,8

E_{22}= (21*30)/62= 10,2

E_{31}= (15*32)/62= 7,7

E_{32}= (15*30)/62= 7,3

Teste Qui-quadrado

- Ex: (continuação)

Valor de p

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6,099[a]	2	,047
Likelihood Ratio	6,264	2	,044
Linear-by-Linear Association	5,947	1	,015
N of Valid Cases	62		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 7,26.

Teste Qui-quadrado

- $p = 0{,}047$ Logo, $p < \alpha$ => Rejeita-se a H_0.
- Existem uma associação entre o estado clínico 6 meses após o tratamento (melhorado, agravado, falecido) e o tipo de tratamento efectuado (placebo ou medicamento X) **ou** Existem diferenças estatisticamente significativas quanto ao estado clínico 6 meses após o tratamento entre

Teste Qui-quadrado

- Assume-se:
  - **Independência dos grupos**

Caso as variáveis em análise sejam dependentes deverá ser usado o ***Teste de McNemar***.

  - **Pelo menos 80% das frequências esperadas têm valores ≥ 5**

No caso de existirem mais de 20% de células com valores esperados <5 deve ***reduzir-se a tabela***, através da fusão de colunas ou linhas (esta fusão deve fazer sentido no contexto da análise que está a ser feita), até ter pelo menos 80% das frequências esperadas com valor ≥ 5.

Se numa tabela de 2×2 (corresponde à fusão máxima possível) existir uma ou mais frequências esperadas com valor <5, então deverá ser usado o ***Teste Exato de Fisher***.

Teste Qui-quadrado

- Teste Exato usado em tabelas de 2×2 (faz o cálculo das probabilidades exatas e não faz uso da distribuição de qui-quadrado como aproximação para o cálculo de probabilidades).
- Utiliza-se no caso de uma tabela de contingência de 2×2, uma ou mais frequências esperadas < 5.
- Ex: num outro ensaio clínico comparou-se a mortalidade no grupo tratado com placebo e tratado com o medicamento X e obtiveram-se os seguintes resultados:

Teste Exato de Fisher

Valor de p

Mortalidade 6 meses após o tratamento * Tramento efectuado Crosstabulation

			Tramento efectuado		
			Placebo	Medicamento X	Total
Mortalidade 6 meses após o tratamento	Vivo	Count	24	29	53
		Expected Count	27,4	25,6	53,0
	Morto	Count	8	1	9
		Expected Count	4,6	4,4	9,0
Total		Count	32	30	62
		Expected Count	32,0	30,0	62,0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5,858[b]	1	,016		
Continuity Correction[a]	4,242	1	,039		
Likelihood Ratio	6,606	1	,010		
Fisher's Exact Test				,027	,017
Linear-by-Linear Association	5,763	1	,016		
N of Valid Cases	62				

a. Computed only for a 2x2 table

b. 2 cells (50,0%) have expected count less than 5. The minimum expected count is 4,35.

Correção de Yates

- Correção para a continuidade em tabelas de 2×2:

Valor de p

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5,858[b]	1	,016		
Continuity Correction[a]	4,242	1	,039		
Likelihood Ratio	6,606	1	,010		
Fisher's Exact Test				,027	,017
Linear-by-Linear Association	5,763	1	,016		
N of Valid Cases	62				

a. Computed only for a 2x2 table

b. 2 cells (50,0%) have expected count less than 5. The minimum expected count is 4,35.

Teste de McNemar

- Análogo ao teste qui-quadrado mas para variáveis dependentes.

Teste de McNemar

Ex:

Tosse antes do tratamento * Tosse depois do tratamento Crosstabulation

			Tosse depois do tratamento		
			Ausente	Presente	Total
Tosse antes do tratamento	Ausente	Count	44	0	44
		Expected Count	34,8	9,2	44,0
	Presente	Count	5	13	18
		Expected Count	14,2	3,8	18,0
Total		Count	49	13	62
		Expected Count	49,0	13,0	62,0

Valor de p

Chi-Square Tests

	Value	Exact Sig. (2-sided)
McNemar Test		,063[a]
N of Valid Cases	62	

a. Binomial distribution used.

Teste Qui-quadrado para Tendências

Ex:

Grupo etário * Estado clínico 6 meses após o tratamento Crosstabulation

			Estado clínico 6 meses após o tratamento			Total
			Melhorado	Agravado	Falecido	
Grupo etário	20-35 anos	Count	14	4	3	21
		Expected Count	9,5	6,0	5,5	21,0
		% within Grupo etário	66,7%	19,0%	14,3%	100,0%
	36-50 anos	Count	13	6	3	22
		Expected Count	9,9	6,3	5,8	22,0
		% within Grupo etário	59,1%	27,3%	13,6%	100,0%
	51-65 anos	Count	6	7	7	20
		Expected Count	9,0	5,8	5,3	20,0
		% within Grupo etário	30,0%	35,0%	35,0%	100,0%
	>65 anos	Count	3	6	8	17
		Expected Count	7,7	4,9	4,5	17,0
		% within Grupo etário	17,6%	35,3%	47,1%	100,0%
Total		Count	36	23	21	80
		Expected Count	36,0	23,0	21,0	80,0
		% within Grupo etário	45,0%	28,8%	26,3%	100,0%

Teste Qui-quadrado para Tendências

Valor de p

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	14,083[a]	6	,029
Likelihood Ratio	14,681	6	,023
Linear-by-Linear Association	12,144	1	,000
N of Valid Cases	80		

a. 2 cells (16,7%) have expected count less than 5. The minimum expected count is 4,46.

Testes Qui-quadrado no R

- chisq.test()
- fisher.test()
- mcnemar.test()
- prop.trend.test()

Quadros de Síntese

Estatística; testes de hipóteses; testes de hipóteses para variáveis quantitativas; testes de hipóteses para variáveis categóricas; outros métodos

Estatística

Estatística Descritiva

Tabelas; Gráficos;
Medidas de tendência central; Medidas de dispersão

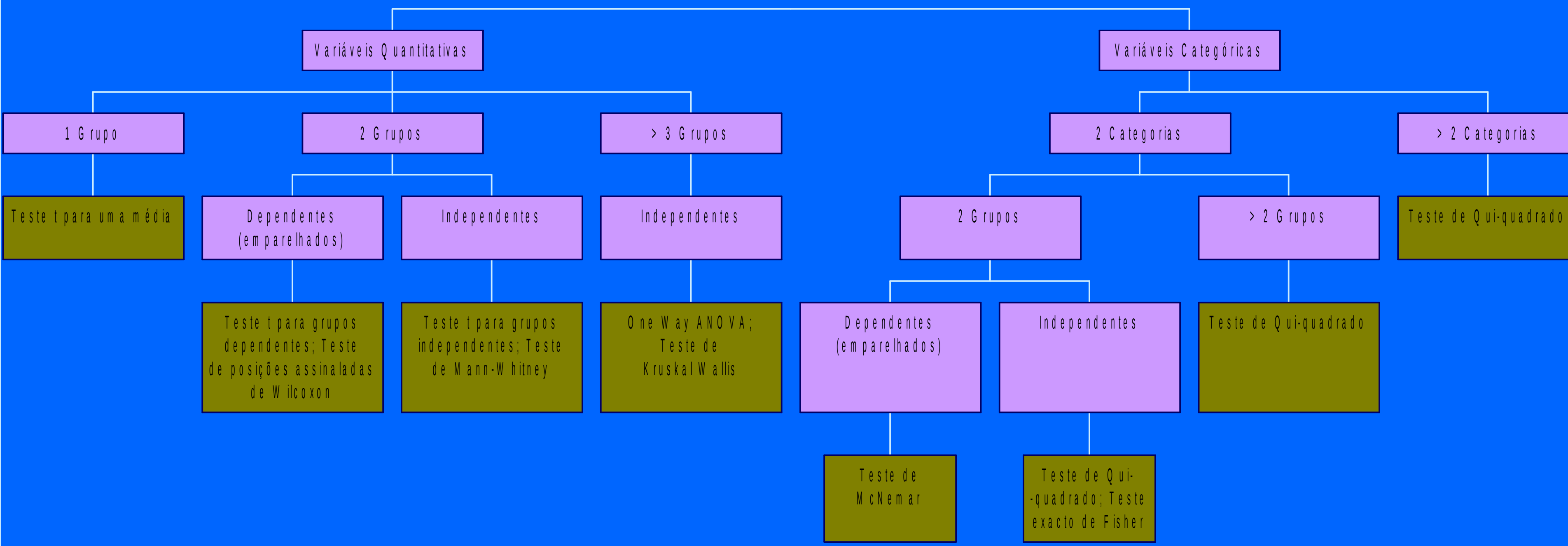
Estatística Inferencial

Estimativas pontuais;
Estimativas de intervalo;
Testes de Hipóteses

Modelação Estatística

Regressão
Linear; Quadrática
Log-linear; Logística; de Cox
Simples; Múltipla

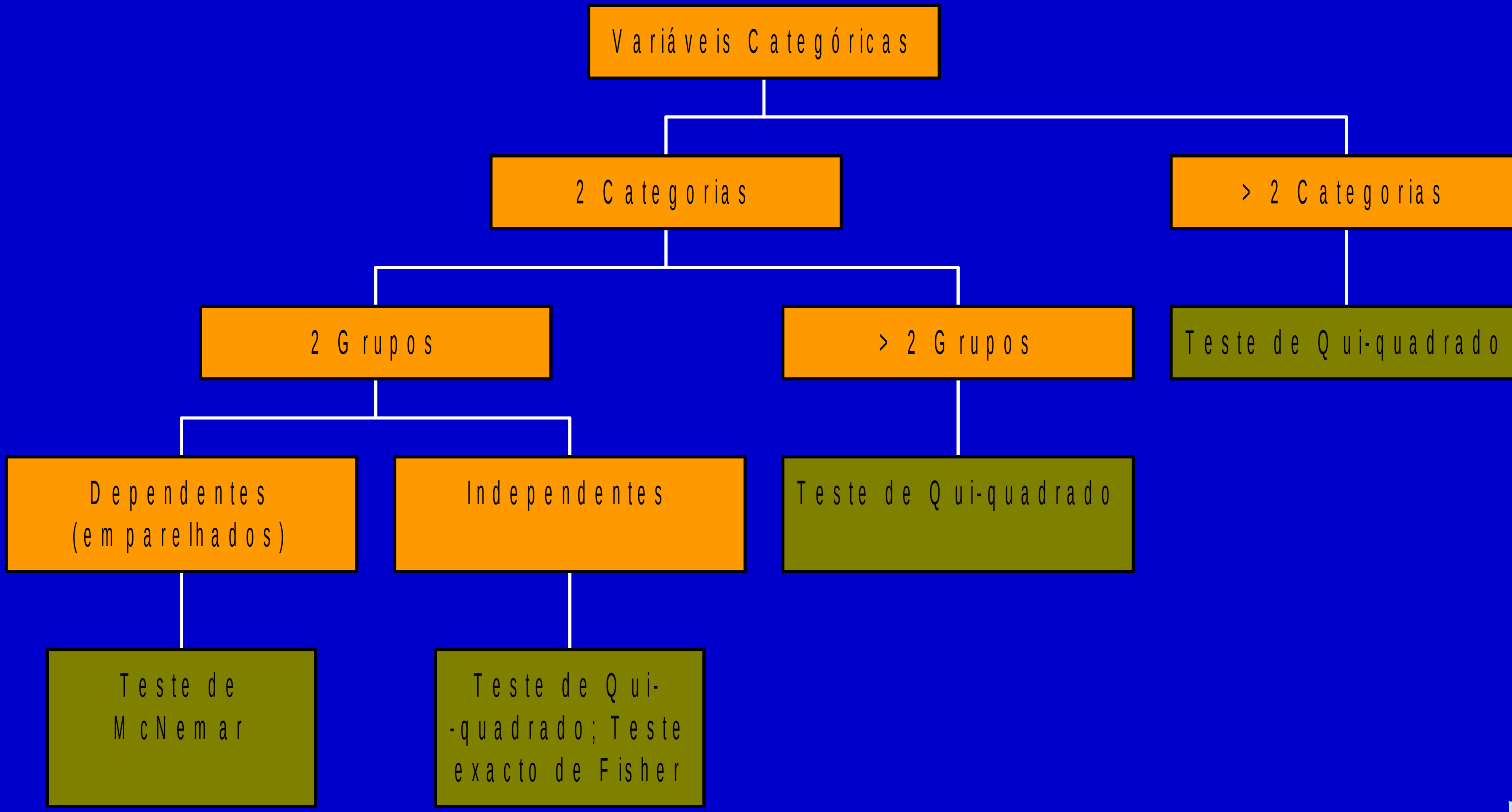
Testes de Hipóteses



Testes de Hipóteses - Variáveis Quantitativas

- Variáveis Quantitativas
  - 1 Grupo
    - Teste t para uma média
  - 2 Grupos
    - Dependentes (emparelhados)
      - Teste t para grupos dependentes; Teste de posições assinaladas de Wilcoxon
    - Independentes
      - Teste t para grupos independentes; Teste de Mann-Whitney
  - > 3 Grupos
    - Independentes
      - One Way ANOVA; Teste de Kruskal Wallis

Testes de Hipóteses - Variáveis Categóricas



Outros Métodos

Correlação

Coeficiente de correlação de Pearson; Coeficiente de correlação de Spearman

Regressão

Regressão linear simples;
Regressão linear múltipla;
Regressão logística;
Regressão de Cox

Análise de Sobrevida

Curvas de Kaplan-Meier;
Regressão de Cox

Outros

Análise de concordância;
Testes diagnósticos;
Análise de séries temporais;
Métodos Bayesianos