

Estatística Aplicada à Medicina II

Lupércio F. Bessegato

UFJF – Departamento de Estatística
2013



Roteiro

1. Comparação de Grupos
2. Medida de Efeito
3. Noções de Regressão Linear

Comparação de 2 Grupos

Introdução

- Busca por novos métodos ou procedimentos que melhorem aqueles já existentes
 - √ Drogas, tratamentos, etc.
- É preciso comparar as técnicas usuais com as técnicas alternativas

Comparação entre Grupos

- Objetivo:
 - √ Verificar a superioridade de um tratamento ou a equivalência entre eles.
- Consideração importante:
 - √ Variabilidade entre os indivíduos
 - √ “Não se conhece *a priori* a reação de cada indivíduo, prescreve-se o tratamento que apresenta, em média, os melhores resultados.”

Resultados

- Extrapolações devem ser cuidadosas
 - √ Cuidados no planejamento (tanto de ensaios clínicos quanto de estudos observacionais)
 - √ Controle dos fatores que afetam a resposta

Procedimento de Comparação

1. Clara identificação dos grupos a serem comparados
 - √ Caso-controle: doentes e não doentes
 - √ Coorte: expostos e não expostos

2. Definir a variável resposta de comparação

- √ Contínua:
 - Nível de colesterol
 - Pressão sanguínea
- √ Dicotômica:
 - % exposição ao fator de risco em cada grupo
 - % doentes em cada grupo

Pareamento

- Pode-se observar características (ou medir valores) de variável resposta em pares de indivíduos
 - √ No planejamento, cuida-se para que as características em cada par sejam as mais próximas possíveis
- Vantagem:
 - √ Os indivíduos no par são semelhantes, exceto no que se refere ao tratamento recebido

Teste Estatístico

- O tipo de teste estatístico depende de:
 - √ Tipo de variável resposta
 - √ Tipo de planejamento

Conceitos Fundamentais

Exemplo 6.1 – Eficácia AZT

- (Fisch et al., 1987) Ensaio clínico aleatorizado sobre eficácia de AZT para prolongar a vida de pacientes com AIDS

	Situação		Total
	Vivo	Morto	
AZT	144	1	145
Placebo	121	16	137
Total	265	17	282

	Situação		Total
	Vivo	Morto	
AZT	144	1	145
Placebo	121	16	137
Total	265	17	282

$$\text{AZT} = \frac{144}{145} = 0,993$$

$$\text{Placebo} = \frac{121}{137} = 0,883$$

- Comparação:
 - √ % sobreviventes em cada grupo
- Diferença entre proporções parece indicar que AZT tem efeito de prolongar a vida
- Antes de aceitar esta conclusão deve-se afastar o acaso como explicação alternativa

Pergunta

- Será que o resultado ocorreu por mero acaso ou por ser o AZT de fato uma droga efetiva?

Hipóteses

- Serão testadas:
 - √ Hipótese nula (H_0)
 - √ Hipótese alternativa (H_1)

Hipótese Nula

- É usual fixar como hipótese de interesse a inexistência de diferença entre os tratamentos
 - √ Coloca-se o ônus da prova da efetividade no tratamento novo
 - √ Hipótese conservativa
- $H_0: p_C = p_T$

Hipótese Alternativa

- Para cada situação existem muitas hipóteses alternativas adequadas
- Em geral, se utiliza:
 $H_1: p_C \neq p_T$

Critério de Decisão

- Construir um critério baseado em H_0 .
- Basear-se em estatística de teste
- Critério genérico e intuitivo:
 √ A estatística de teste mede a discrepância entre o que foi observado na amostra e o que seria esperado se H_0 fosse verdadeira

- Grande distância (medida por distribuição de probabilidade) indica que H_0 não é verdadeira
 √ Rejeita-se H_0 em favor de H_1 se o valor da estatística de teste é “grande”
 √ Deve-se comparar esse valor com alguma distribuição de probabilidade (depende de cada caso)

Erros do Teste

Conclusão do Teste	Situação Real	
	H_0 é Verdadeira	H_0 é Falsa
Não rejeitar H_0	Decisão correta	Erro tipo II
Rejeitar H_0	Erro tipo I	Decisão correta

- $\alpha = P\{\text{Erro tipo I}\}$:
 √ Percentil de uma distribuição de probabilidade
 √ Busca-se um critério de decisão que torne este erro pouco provável
 √ α é denominado nível de significância do teste

$\beta = P\{\text{Erro tipo II}\}$

- √ Para um tamanho fixo da amostra, não há como controlar simultaneamente ambos os erros
- √ Convencionou-se que o erro mais sério é o erro tipo I
- √ Calcula-se tamanho da amostra que reduza β a níveis aceitáveis
- $1 - \beta = \text{Poder do teste}$
- √ $P\{\text{Rejeição de } H_0 \text{ quando } H_0 \text{ é falsa}\}$
- √ Capacidade de um teste identificar diferenças que realmente existem

Erros Tipo I e II

- Qual dos dois tipos de erro é mais grave, devendo ser evitado?

Exemplo 1

- H_0 : Réu é inocente vs. H_1 : Réu é culpado
 - √ Erro tipo I: Condenar inocente
 - √ Erro tipo II: Absolver culpado
 - √ Qual o erro mais grave?

Exemplo 2

- H_0 : Medicamento NEW não é melhor que AZT
- H_1 : Medicamento NEW é melhor que AZT
 - √ Erro tipo I:
 - √ NEW é considerado melhor que AZT, quando ele não é o mais eficaz
 - √ Erro tipo II:
 - √ NEW não é considerado melhor que o AZT, quando ele é mais eficaz
- Qual o erro mais grave?

Erros Tipo I e II

- Devido às diretrizes sugeridas para especificar H_0 e H_1 , em geral:
um erro tipo I é **mais grave** que um erro tipo II

Probabilidade de Significância (Valor-p)

- Expressa a conclusão final de um teste
- É a probabilidade de se obter resultados mais extremos do observado, sob a hipótese de igualdade dos grupos
 - √ $p = P\{\text{distâncias iguais ou superiores sob } H_0\}$
 - √ Qual a probabilidade de outra amostra apresentar resultados mais favoráveis (distâncias maiores) para rejeitar H_0 ?

- Se o p-valor é pequeno há duas conjecturas:
 - √ H_0 é verdadeira e ocorreu um evento extremamente raro
 - √ H_0 não deve ser verdadeira, pois a conjectura inicial não parece ser plausível
- Quanto menor o p-valor, maior a evidência para se rejeitar H_0
 - √ De maneira geral $p \leq 0,05$ indica que há diferenças significativas entre grupos

Exemplo – Lançamento Moeda

- Você lança uma moeda 20 vezes e obtém 16 caras e 4 coroas.
 - √ H_0 : a moeda é honesta ($p_c = p_k = 0,50$)
 - √ Qual é a chance de obter 16 ou mais caras (ou 4 ou menos caras) lançando uma moeda honesta?
 - √ Resp.: $2 \times 0,59\% = 1,18\%$
 - É o mesmo que obter 4 coroas ou menos

- 200 alunos lançarão uma moeda 20 vezes e anotarão a quantidade de caras
 - √ Espera-se obter pelo menos 16 caras (ou coroas) em 1,18% dos experimentos.
 - √ Não seria surpresa encontrar um mais experimento com pelo menos 16 caras (ou coroas)
 - √ Esse é um caso de comparação múltipla!

Exemplo – Temperatura Corporal

- Muitos autores afirmam que a temperatura corporal normal média é 37° C.
 - A temperatura média de uma amostra com 130 pessoas é 36,82°C
- Se a média populacional é realmente 37°C, qual é a chance de uma amostra deste tamanho apresentar uma diferença absoluta entre a média populacional e a média amostral de 0,18°C ou mais?

√ Resp.: valor-p: 0,0000018

- Há uma chance muito pequena de a média de amostra com 130 pessoas estar mais afastada da média hipotética que o valor observado
- Isso não prova H_0 é falsa.
 - √ Conclusão: há uma pequena chance de obter os dados observados, caso H_0 seja verdadeira

- Perguntas relevantes para conclusão:
 - √ Esta é uma única amostra de 130 pessoas?
 - √ Ou foram escolhidas as 130 menores temperaturas em uma base de dados muito grande?
 - √ Ou as temperaturas foram medidas diversas vezes no mesmo sujeito, anotando-se o menor dos valores?(nos últimos casos os valores não são representativos)

- Outro ponto importante com p-valores:
 - √ Sim, é provável que haja diferença estatística.
 - √ Essa diferença é clinicamente relevante?

- P-valor de uma amostra menor:
 - √ Média de temperatura de amostra com 12 pessoas é 36,76°C
 - √ Pergunta: Se a média verdadeira da população é 37°C, qual é a chance de que o valor absoluto da diferença entre a média amostral e a média hipotética da população seja 0,24°C ou mais?
 - √ Resp.: 0,0687
 - √ A diferença não é estatisticamente significativa
 - √ Isso não prova que a média da população seja 37°C
 - √ Pode-se dizer que se H_0 for verdadeira, os dados observados não parecem ser surpreendentes

Exemplo – Antibióticos em Cirurgias

- Aplicação de Cloranfenicol em feridas cirúrgicas para redução de infecção.
 - √ Alocação aleatória de pacientes cirúrgicos para receber ou para receber uma pomada antibiótica ou como pomada sem medicação ativa
 - √ 6,6% de infecções nos pacientes que receberam antibiótico e 11,0% nos pacientes que receberam placebo
 - √ p-valor da comparação das 2 taxas: 0,010

- Para interpretar p-valor deve-se definir H_0
 - √ H_0 : o risco de infecção é o mesmo em pacientes que receberam ou não a medicação
- Pergunta respondida pelo p-valor?
 - √ Se H_0 é verdadeira, qual é a chance de que uma amostra aleatória de mesmo tamanho apresente uma diferença nas taxas de infecção iguais ou maiores que a diferença observada neste estudo?
 - √ Resp: 1 em 100

Comentários

- Em geral, a hipótese nula (hipótese testada) é oposta à hipótese que o pesquisador espera ser verdadeira (hipótese de interesse)
- Calculam-se probabilidades de obtenção de resultados que não foram obtidos realmente
- Não é feita uma amostragem de uma população conhecida. O cálculo do p-valor baseia-se em uma população hipotética

Resposta Dicotômica: Amostras Independentes

Comparação de 2 Grupos

- Resposta dicotômica:
 - √ Ocorrência ou não de evento de interesse (sucesso ou fracasso)
 - √ Desenvolvimento de doença, presença de certo atributo, etc.
- Comparação das probabilidades de ocorrência do evento (ou atributo) nos grupos
 $H_0: p_1 = p_2$ vs. $H_1: p_1 \neq p_2$

Alguns Tipos de Testes

- Teste qui-quadrado (χ^2)
- Teste qui-quadrado com correção de continuidade
- Teste exato de Fisher
- Teste Z para comparação de proporções

Teste Qui-Quadrado (χ^2)

Grupo	Ocorrência do Evento		Total
	Sim	Não	
I	a	b	$n_1 = a + b$
II	c	d	$n_2 = c + d$
Total	$m_1 = a + c$	$m_2 = b + d$	$N = n_1 + n_2$

- Se não há diferenças:

$$\frac{a}{n_1} = \frac{c}{n_2} = \frac{a+c}{n_1+n_2} = \frac{m_1}{N} \implies a = \frac{m_1 n_1}{N} \quad c = \frac{m_1 n_2}{N}$$

$$\frac{b}{n_1} = \frac{d}{n_2} = \frac{b+d}{n_1+n_2} = \frac{m_2}{N} \implies b = \frac{m_2 n_1}{N} \quad d = \frac{m_2 n_2}{N}$$

Valores Observados e Esperados

- Valores Observados (O_i):
√ Denotados por a, b, c e d
- Valores Esperados (E_i):
√ Calculados sob a hipótese de igualdade
- Se as proporções são iguais, a diferença entre eles não deve ser grande.
- Medida da diferença (Pearson):

$$X^2 = \sum_{i=1}^4 \frac{(O_i - E_i)^2}{E_i} = \frac{N(ad - bc)^2}{m_1 m_2 n_1 n_2}$$

Exemplo

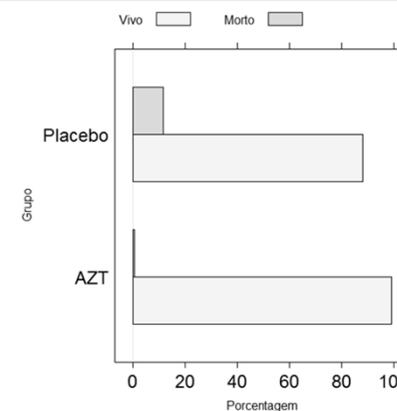
- Ex. 6.2 – AZT (cont.):

	Situação		Total
	Vivo	Morto	
AZT	144	1	145
Placebo	121	16	137
Total	265	17	282

$$E_1 = 145 \times \frac{265}{282} = 136,26$$

Grupo	Ocorrência do Evento		Total
	Vivo	Morto	
AZT	136,26	8,74	145
Placebo	128,74	8,26	137
Total	265	17	282

- Gráfico de barras:



```
> tabela.azt
      Situacao
grupo  Vivo  Morto
AZT    99.3103448 0.6896552
Placebo 88.3211679 11.6788321
```

- Cálculo da estatística X^2 :

i	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	144	136,26	7,74	59,91	0,44
2	121	128,74	-7,74	59,91	0,47
3	1	8,74	-7,74	59,91	6,85
4	16	8,26	7,74	59,91	7,25
Total	282	282	0	239,64	15,01

$\sqrt{X^2} = 15,01$

- O valor encontrado é ou não um valor grande?

- Critério para decisão:

√ Sob H_0 :

- valores grandes ocorrem com pouca frequência
- $X^2 \sim \chi_1^2$

- Cálculo p-valor:

$$p = P\{\chi_1^2 \geq 15,01\} = 0,000107$$

√ Usar tabela ou pacote computacional (Excel, R, etc.)

```
> 1-pchisq(15.01, 1)
[1] 0.000106943
```

- p-valor da estatística de teste (X^2) é baixo

√ $P = 0,000107$

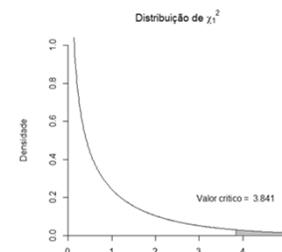
- Conclusão do teste:

√ Podemos afirmar com grande certeza que o AZT tem efeito de prolongar a vida de pacientes com AIDS

- Outra abordagem:

√ O nível de significância do teste é $\alpha = 5\%$

$$P\{\chi_1^2 \geq \text{valor crítico}\} = 0,05$$



√ X^2 (estatística de teste) > valor crítico

- Rejeita-se a hipótese de igualdade entre os grupos de tratamento e controle

- No exemplo: $X^2=15,01 > 3,84$
- Conclusão do teste:
 - √ Decidimos, com 95% de certeza, que há evidências do efeito do AZT

Exemplo 6.3 – Fatores de Risco para AVC

- (Welin et al., 1987) Fatores de risco para AVC em homens de meia idade.
 - √ Homens nascidos em Gotemburgo, em 1913, em dias múltiplos de 3.
 - √ Total: 789 homens
 - Sofreram AVC: 57 homens
 - Não sofreram: 732 homens
 - √ Duração do estudo: 1967 a 1986

• Resultados do estudo:

Fator	AVC		Valor-p
	Sim	Não	
AVC como causa principal da morte			
na mãe	29,8	11,2	0,0005
no pai	7,0	7,5	0,720
AVC como causa principal ou contribuindo para a causa da morte			
na mãe	29,8	14,2	0,002
no pai	7,0	9,2	0,590
Doença coronariana	7,0	6,3	0,830
Sinais eletrocardiográficos			
hipertrofia do ventrículo esquerdo	3,5	1,0	0,080
doença coronariana	10,5	6,5	0,230

• Conclusão:

- √ Diferenças altamente significativas:
 - AVC como causa principal ou contribuindo para causa da morte na mãe.

Teste Qui-Quadrado (χ^2) com Correção de Continuidade

$$X_c^2 = \frac{N(|ad - bc| - \frac{N}{2})^2}{m_1 m_2 n_1 n_2}$$

- Yates (1934) demonstrou que a expressão é mais apropriada para comparação com χ_1^2
 - Denominada com correção de continuidade ou com correção de Yates

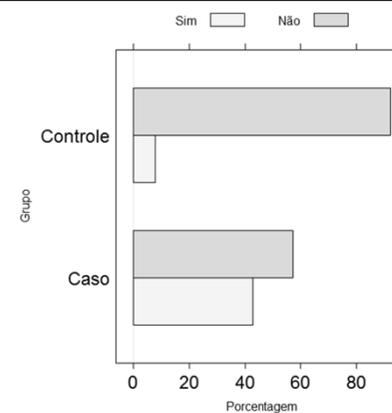
Exemplo 6.4

- (Shapiro et al., 1979) Estudo de associação entre contraceptivos orais e infarto do miocárdio
 - √ Pacientes com idade entre 30 e 34 anos

Grupo	Uso recente		Total
	Sim	Não	
Casos	9	12	21
Controles	33	390	423
Total	42	402	444

- Proporção de uso recente de contraceptivo:
 - √ Caso: $9/21 = 0,43$
 - √ Controle: $33/423 = 0,08$
- Será que o resultado não ocorreu por acaso?

- Gráfico de barras:



```
> tabela.infarto
      Uso_Recente
grupo      Sim      Não
Caso      42.857143 57.142857
Controle  7.801418 92.198582
```

- Estatística de teste:

$$\sqrt{a} = 9 \quad b = 12$$

$$\sqrt{c} = 33 \quad d = 390 \quad N = 444$$

$$X_c^2 = \frac{N(|ad - bc| - \frac{N}{2})^2}{m_1 m_2 n_1 n_2} = \frac{444(|(9)(390) - (12)(33)| - \frac{444}{2})^2}{(42)(402)(21)(423)}$$

$$= 24,760$$

- p-valor:

$$p = P \{ \chi_1^2 \geq 24,760 \} = 0,000001$$

```
> 1-pchisq(24.75965,1)
[1] 6.494284e-07
```

- Saída R (teste com e sem correção)

```
> X2.infarto
Pearson's Chi-squared test

data: amostra.infarto
X-squared = 28.7068, df = 1, p-value = 8.421e-08
```

```
> X2c.infarto
Pearson's Chi-squared test with Yates' continuity correction

data: amostra.infarto
X-squared = 24.7597, df = 1, p-value = 6.494e-07
```

- Conclusão:

√ Podemos afirmar, com alto grau de certeza, que existe associação entre o uso de contraceptivos orais e infarto de miocárdio para pacientes entre 30 e 34 anos.

Uso do Teste χ^2

- Dificuldade técnica na aplicação do teste χ^2
 - √ A distribuição χ_1^2 não é mais apropriada quando o valor esperado em alguma casela for **menor que 5**
 - √ Casela com contagem zero
- Restrição válida para os testes χ^2 com e sem correção de continuidade
- Solução:
 - √ Usar o Teste Exato de Fisher

Teste Exato de Fisher

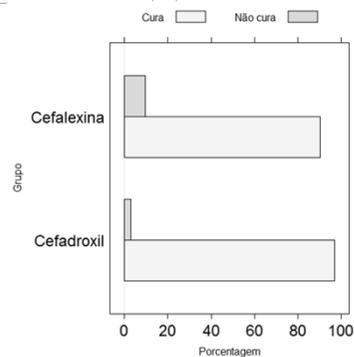
- É a versão exata do teste χ^2
- A maioria dos pacotes estatísticos dispõe dele
- Não serão discutidas as etapas de sua construção
 - √ Embora simples, são muito técnicas
- √ Cálculos de probabilidade estão relacionados com distribuição hipergeométrica

Exemplo 6.5

- (Pasternak et al., 1992) Eficácia e segurança de dois antibióticos em tratamento de pneumonia de origem comunitária em adultos
 - √ Cefadroxil e cefalexina
 - √ Avaliação resposta terapêutica:
 - Evolução quadro clínico e exame radiológico
 - Admissão no estudo e no 10º dia de tratamento

• Comparação medicamentos quanto à cura:

> amostra.pneumo			> tabela		
grupo	Situacao		grupo	Situacao	
	Cura	Não cura		Cura	Não cura
Cefadroxil	31	1	Cefadroxil	96.875000	3.125000
Cefalexina	28	3	Cefalexina	90.322581	9.677419



• Teste exato de Fisher – Saída R:

```
> exato.pneumo

Fisher's Exact Test for Count Data

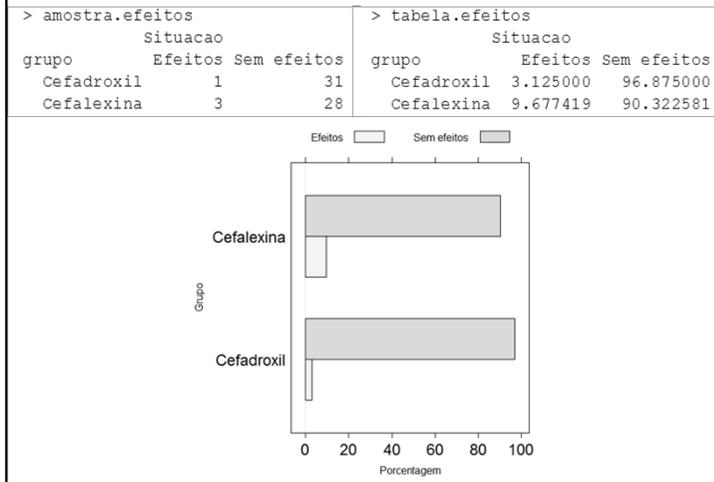
data: amostra.pneumo
p-value = 0.3547
alternative hypothesis: true odds ratio is not equal to 1
sample estimates:
odds ratio
3.263013
```

√ $H_0: p_1 = p_2$ vs. $H_1: p_1 \neq p_2$

√ Para estudos caso-controle pode-se usar OR:

$$OR = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}$$

• Comparação quanto a efeitos adversos:



• Teste exato de Fisher – Saída R:

```
Fisher's Exact Test for Count Data

data: amostra.efeitos
p-value = 0.3547
alternative hypothesis: true odds ratio is not equal to 1
sample estimates:
odds ratio
0.3064652
```

$$\sqrt{H_0: p_1 = p_2 \text{ vs. } H_1: p_1 \neq p_2}$$

• Conclusão:

- √ Não há diferenças entre os dois tratamentos
- Em termos de cura e de efeitos colaterais

Teste de Z para Comparação de Proporções

- Teste para equivalência entre 2 tratamentos
- √ $H_0: p_0 = p_1$ vs. $H_1: p_0 \neq p_1$
- √ Variável resposta: sucesso ou fracasso
- √ Teste aproximado
- √ Requer amostras grandes
- $n_1 \hat{p}_1$ e $n_2 \hat{p}_2 > 5$

- Estatística de teste:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

- Estatística de teste alternativa

$$Z^* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\sqrt{\hat{p}} = \frac{a+c}{n_1+n_2} \text{ e } a \text{ e } c \text{ contam sucessos na amostra}$$

- Com amostras grandes, Z (Z^*) $\sim N(0, 1)$

- Calcula-se o p-valor do teste considerando-se $2 \times P\{\text{valores mais favoráveis rejeição } H_0\}$

- Abordagem alternativa:

√ Rejeita-se H_0 se $|Z| > z_{\alpha/2}$ ou $|Z^*| > z_{\alpha/2}$

√ $z_{\alpha/2}$: percentil de ordem $(1 - \alpha/2)$ da $N(0, 1)$

- O teste Z e χ^2 são equivalentes para amostras grandes

Exemplo 6.6

- Comparação de drogas contra náusea

√ 400 marinheiros

√ Pílula A: 200 marinheiros (152 não enjoaram)

√ Pílula B: 200 marinheiros (132 não enjoaram)

- Parâmetros populacionais:

√ p_i : proporção marinheiros que não enjoaram no grupo i , $i = A, B$

- Estimativas amostrais:

$$\hat{p}_A = \frac{152}{200} = 0,76$$

$$\hat{p}_B = \frac{132}{200} = 0,66$$

$$\hat{p} = \frac{152 + 132}{200 + 200} = 0,71$$

- Estatísticas de Teste

$$Z = \frac{0,76 - 0,66}{\sqrt{\frac{0,76(1-0,76)}{200} + \frac{0,66(1-0,66)}{200}}} = 2,22$$

$$Z^* = \frac{0,76 - 0,66}{\sqrt{0,71(1-0,71)\left(\frac{1}{200} + \frac{1}{200}\right)}} = 2,20$$

- Cálculo do p-valor (estatística Z):

```
> 1-pnorm(2.22)
[1] 0.01320938
> 2*0.01320938
[1] 0.02641876
```

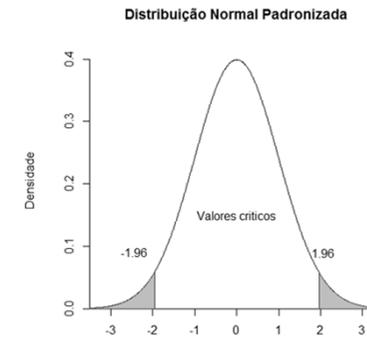
- Cálculo do p-valor (estatística Z*):

```
> 1-pnorm(2.22)
[1] 0.01320938
> 2*0.01320938
[1] 0.02641876
```

- Valor crítico para nível de significância $\alpha=5\%$

✓ Rejeita-se H_0 para $|Z| > 1,96$ (ou $|Z^*| > 1,96$)

- Montagem da região de rejeição:



✓ Rejeita-se H_0 quando a estatística Z (Z^*) situa-se na região sombreada

- Conclusão:

✓ As duas pílulas não são igualmente efetivas.

✓ Há indicações de que a pílula A oferece maior proteção contra náusea, comparada à pílula B.

**Resposta Dicotômica:
Amostras Paredas**

Amostra Pareadas

- Exemplo
 - √ Dois patologistas examinaram, separadamente o material de 100 tumores
 - √ Resposta: benigno ou maligno
 - √ Os patologistas diferem em seus critérios de decisão?

Diagnósticos de B	Diagnósticos de A		Total
	Vivo	Morto	
Malignos	9	1	10
Benignos	9	81	90
Total	18	82	100

- Se H_0 é verdadeira os dois grupos se equivalem (discrepâncias observadas devem-se ao acaso)
 - √ Pares que produziram sucesso ou fracasso não contém informação para discriminar p_1 e p_2
 - √ Se r e s tem valores semelhantes sob H_0 , espera-se que a metade do número de discordâncias
 - √ Deve-se rejeitar H_0 se a distância entre os valores esperados e observados for grande.

Teste de McNemar

- Hipóteses:
 - √ $H_0: p_1 = p_2$ vs. $H_1: p_1 \neq p_2$
 - √ p_1 e p_2 : probabilidade de sucesso nos grupos controle e tratamento

Controle	Tratamento		Total
	Sucesso	Fracasso	
Sucesso	k	r	n_1
Fracasso	s	l	n_2
Total	m_1	m_2	N

- Estatística de McNemar:

$$X_{McN}^2 = \frac{\left(|r - \frac{r+s}{2}| - \frac{1}{2}\right)^2}{\frac{r+s}{2}} + \frac{\left(|s - \frac{r+s}{2}| - \frac{1}{2}\right)^2}{\frac{r+s}{2}}$$

$$= \frac{(|r - s| - 1)^2}{r + s}$$

- Distribuição amostral da estatística de teste
 - √ $X_{McN}^2 \sim \chi_1^2$

Exemplo 6.7

- Quimioterapia e câncer de mama

√ Grupos:

- A: tratamento quimioterápico durante 6 meses (início na 1ª. Semana após mastectomia)
- B: tratamento quimioterápico na 1ª. Semana

√ Pares de pacientes com aproximadamente mesma idade e condições clínicas

(grupos devem ser tão comparáveis quanto possível)

Tratamento	Sobrevivência		Total
	Sim	Não	
A	526	95	621
B	515	106	621
Total	1041	201	1242

$$\hat{p}_A = \frac{526}{621} = 0,847$$

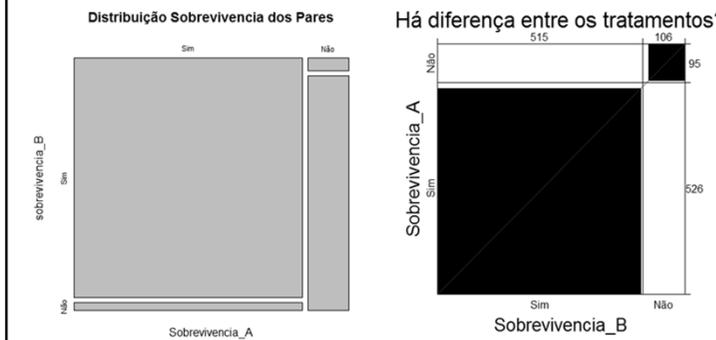
$$\hat{p}_B = \frac{515}{621} = 0,830$$

- O teste X^2 não é aplicável pois as amostras não são independentes

- Contagem concordâncias/discordâncias

Sobrevivência Tratamento A	Sobrevivência Tratamento B		Total
	Sim	Não	
Sim	510	16	526
Não	5	90	95
Total	515	106	621

```
> tabela.mama
      Sobrevivencia_B
Sobrevivencia_A Sim Não
Sim 82.126 2.576
Não 0.805 14.493
```



- Cálculo estatística de McNemar

Sobrevivência Tratamento A	Sobrevivência Tratamento B		Total
	Sim	Não	
Sim	510	16	526
Não	5	90	95
Total	515	106	621

$$X_{McN}^2 = \frac{(|r - s| - 1)^2}{r + s} = \frac{(|16 - 5| - 1)^2}{16 + 5} = 4,76$$

```
amostra.mama<- matrix(c(510, 16, 5, 90), 2, 2)
mcnemar.test(amostra.mama)

McNemar's Chi-squared test with continuity correction

data: amostra.mama
McNemar's chi-squared = 4.7619, df = 1, p-value = 0.0291
```

- p-valor= 0,02919
- Há evidências estatísticas de diferença entre os tratamentos

Resposta Contínua: Amostras Independentes

Objetivo

- Comparação de dois grupos
 - √ Doentes e não doentes
 - √ Tratamento e placebo, etc.
- Resposta contínua:
 - √ Pressão sistólica, nível de colesterol, etc.
- Teste
 - √ Igualdade das médias das respostas dos dois grupos

Teste t para Comparação de Médias

- Comparação das médias de dois grupos
 - √ Resposta contínua
 - √ $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$
- Suposição:
 - √ Variável resposta tem distribuição gaussiana com mesmo desvio-padrão (variância)
 - √ Parâmetros populacionais desconhecidos: μ_1, μ_2 e σ

Procedimento

- Coleta-se amostra 1
 - √ Tamanho n_1
 - √ Elementos amostrais: $(x_{11}, x_{12}, \dots, x_{1n_1})$
 - √ Calculam-se: (\bar{x}_1, s_1)
- Coleta-se a amostra 2:
 - √ Tamanho n_2
 - √ Elementos amostrais: $(x_{21}, x_{22}, \dots, x_{2n_2})$
 - √ Calculam-se: (\bar{x}_2, s_2)

- Estatística de teste:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{DP(\bar{X}_1 - \bar{X}_2)}$$

- √ Numerador: maior a diferença maior a chance de os grupos serem realmente diferentes
- √ Denominador: maior a variabilidade, maior a dificuldade de perceber diferenças entre efeitos médios
- Critério de decisão:
 - √ Rejeita-se H_0 para $|T|$ grande

- Distribuição da estatística de teste:

$$T \sim t_{n_1+n_2-2}$$

- √ Distribuição t de student com n_1+n_2-2 graus de liberdade
- Estimativas das variâncias :
 - √ Melhor estimativa da variância populacional (σ^2):

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- √ Melhor estimativa da variância da estatística de teste:

$$Var(\bar{X}_1 - \bar{X}_2) = s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

- Teste t:

- √ Valor observado da estatística de teste:

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

- √ Cálculo do p-valor (teste bilateral):

$$p = 2 \times \Pr \{T > |t_{obs}|\}$$

- √ Região de rejeição de H_0 :

- Valor crítico para nível de significância α :

$$t_{crit} = t_{n_1+n_2-2, (1-\alpha/2)}$$

- Critério de rejeição de H_0 :

$$|t_{obs}| > t_{crit}$$

Exemplo 6.8

- (Camargo et al., 1994)
 - √ Níveis séricos de frutossamina:
 - √ Objetivo do estudo:
 - Estabelecer valores normais em homens e mulheres
 - √ Planejamento do estudo:
 - Grupo 1: 42 indivíduos normais (21 mulheres, de 22 a 71 anos e 21 homens, de 24 a 81 anos)
 - Grupo 2: 36 gestantes, de 17 a 37 anos
 - Idade gestacional: 17 a 37 semanas

• Valores amostrais

Variável	Mulheres (n = 21)		Homens (n = 21)		Gestantes (n = 36)	
	\bar{x}_1	s ₁	\bar{x}_2	s ₂	\bar{x}	s
Idade (anos)	39	14	43	14	25	7
Glicose (mg/dL)	95	20	84	24	28	5
Frutossamina (mmol/L)	2,70	0,26	2,99	0,32	103	17
Triglicéridios (mg/dL)	145	135	124	62	2,40	0,6

• Comparações de médias:

√ Grupos: Homens e Mulheres

Variável	Grupos		p-valor
	Mulheres	Homens	
Idade (anos)	39	43	0,36
Glicose – jejum (mg/dL)	95	84	0,11
Frutossamina (mmol/L)	2,70	2,99	0,003
Triglicéridios (mg/dL)	145	124	0,52

√ Há diferenças das médias de frutossamina por sexo

• Frutossamina–Estabelecimento faixa de referência :

√ Suposta distribuição gaussiana dos dados

√ Critério: média \pm 2 desvios-padrão

• Faixas de referência

Grupo	Dados Amostrais			Frutossamina (mmol/L)	
	\bar{x}	s	2 s	Inferior	Superior
Homens	2,99	0,32	0,64	2,35	3,63
Gestantes	2,70	0,26	0,52	2,18	3,22
Não gestantes	2,40	0,22	0,44	1,96	2,84

• Comparações de médias:

√ Grupos: Gestantes e Não Gestantes

Variável	Grupos		p-valor
	Gestantes	Não Gestantes	
Idade (anos)	25	39	<0,001
Glicose – jejum (mg/dL)	103	95	0,11
Frutosamina (mmol/L)	2,40	2,70	<0,001
Triglicéridios (mg/dL)	4,1	145	0,52

√ Há diferenças das médias de frutosamina em presença ou não de gravidez

• Memória de cálculo – Idade:

√ (Mulheres x Homens):

```
> source("apresentacaoFunc.txt")
> teste.t(amostra=c(39, 14, 21), amostra2=c(43,14,21), var=
"Idade",
+ grupos=c("Mulheres", "Homens"))
```

```
Teste t - Amostras Independentes
Estatísticas Amostrais
Grupo
Mulheres Homens
Média      39      43
Desvio     14      14
Tamanho    21      21
Diferença médias: -4
Variancia combinada: 196
Erro padrão diferença: 4.320494

dados:      Idade
Grupos:     Mulheres e Homens
tobs = -0.9258 , gl = 40 , p-valor = 0.36009
Hipótese alternativa: médias são diferentes.
```

• Memória de cálculo – Glicose em jejum:

√ (Mulheres x Homens):

```
> teste.t(amostra=c(95, 20, 21), amostra2=c(84,24,21), var=
"Glicose",
+ grupos=c("Mulheres", "Homens"))
```

```
Teste t - Amostras Independentes
Estatísticas Amostrais
Grupo
Mulheres Homens
Média      95      84
Desvio     20      24
Tamanho    21      21
Diferença médias: 11
Variancia combinada: 488
Erro padrão diferença: 6.817345

dados:      Glicose
Grupos:     Mulheres e Homens
tobs = 1.6135 , gl = 40 , p-valor = 0.11449
Hipótese alternativa: médias são diferentes.
```

• Memória de cálculo – Frutosamina:

√ (Mulheres x Homens):

```
> teste.t(amostra=c(2.70, 0.26,
21), amostra2=c(2.99,0.32,21), var= "Frutosamina",
+ grupos=c("Mulheres", "Homens"))
```

```
Teste t - Amostras Independentes
Estatísticas Amostrais
Grupo
Mulheres Homens
Média      2.70    2.99
Desvio     0.26    0.32
Tamanho    21.00   21.00
Diferença médias: -0.29
Variancia combinada: 0.085
Erro padrão diferença: 0.08997354

dados:      Frutosamina
Grupos:     Mulheres e Homens
tobs = -3.2232 , gl = 40 , p-valor = 0.00252
Hipótese alternativa: médias são diferentes.
```

- Memória de cálculo – Triglicerídios:
 $\sqrt{}$ (Mulheres x Homens)

```
> teste.t(amostra=c(145, 135, 21), amostra2=c(124, 62, 21),
+ grupos=c("Mulheres", "Homens"), var= "Triglicerídios")
```

Teste t - Amostras Independentes

Estatísticas Amostrais

Grupo	Grupo	
	Mulheres	Homens
Média	145	124
Desvio	135	62
Tamanho	21	21

Diferença médias: 21

Variância combinada: 11034.5

Erro padrão diferença: 32.41766

dados: Triglicerídios

Grupos: Mulheres e Homens

tobs = 0.6478 , gl = 40 , p-valor = 0.52082

Hipótese alternativa: médias são diferentes.

- Memória de cálculo – Idade:
 $\sqrt{}$ (Gestantes x Não Gestantes):

```
> teste.t(amostra=c(25, 7, 36), amostra2=c(39, 14, 21), var=
"Idade",
+ grupos=c("Gestantes", "Não Gestantes"))
```

Teste t - Amostras Independentes

Estatísticas Amostrais

Grupo	Grupo	
	Gestantes	Não Gestantes
Média	25	39
Desvio	7	14
Tamanho	36	21

Diferença médias: -14

Variância combinada: 102.4545

Erro padrão diferença: 2.779343

dados: Idade

Grupos: Gestantes e Não Gestantes

tobs = -5.0372 , gl = 55 , p-valor = 1e-05

Hipótese alternativa: médias são diferentes.

- Memória de cálculo – Glicose em jejum:
 $\sqrt{}$ (Gestantes x Não Gestantes):

```
> teste.t(amostra=c(103, 17, 36), amostra2=c(95, 20, 21), var=
"Glicose",
+ grupos=c("Gestantes", "Não Gestantes"))
```

Teste t - Amostras Independentes

Estatísticas Amostrais

Grupo	Grupo	
	Gestantes	Não Gestantes
Média	103	95
Desvio	17	20
Tamanho	36	21

Diferença médias: 8

Variância combinada: 329.3636

Erro padrão diferença: 4.983269

dados: Glicose

Grupos: Gestantes e Não Gestantes

tobs = 1.6054 , gl = 55 , p-valor = 0.11414

Hipótese alternativa: médias são diferentes.

- Memória de cálculo – Frutosamina:
 $\sqrt{}$ (Gestantes x Não Gestantes)

```
> teste.t(amostra=c(2.40, 0.22, 36), amostra2=c(2.70, 0.26,
21), var= "Frutosamina",
+ grupos=c("Gestantes", "Não Gestantes"))
```

Teste t - Amostras Independentes

Estatísticas Amostrais

Grupo	Grupo	
	Gestantes	Não Gestantes
Média	2.40	2.70
Desvio	0.22	0.26
Tamanho	36.00	21.00

Diferença médias: -0.3

Variância combinada: 0.05538182

Erro padrão diferença: 0.06461899

dados: Frutosamina

Grupos: Gestantes e Não Gestantes

tobs = -4.6426 , gl = 55 , p-valor = 2e-05

Hipótese alternativa: médias são diferentes.

Exemplo 6.9

- (Rocha, 1995)
 - √ Comparação Tianeptina com Placebo
 - Fármaco antidepressivo
 - √ Ensaio clínico aleatorizado, duplo-cego
 - √ Pacientes de Belo Horizonte, Rio e Campinas
 - √ Quantificação depressão:
 - escala de Montgomery-Asberg (MADRS)
 - √ Escores obtidos para cada paciente:
 - 7, 14, 21, 28 e 42 dias após início do ensaio

- Classificação da escala MADRS:

Score	Classificação
Zero a 6	Ausência de depressão
7 a 19	Depressão leve
20 a 34	Depressão moderada
35 a 60	Depressão grave

- √ Um escore não tem teoricamente a distribuição normal, pois, **não** é uma variável contínua
- √ Recomenda-se teste de normalidade para verificar a razoabilidade desse pressuposto

- Escore final na escala MADRS
 - √ Pacientes dos dois grupos, admitidos em BH

Grupo	Escore							
	Placebo	6	33	21	26	10	29	33
	37	15	2	21	7	26	13	
Tianeptina	10	8	17	4	17	14	9	4
	21	3	7	10	29	13	14	2

- Valores observados na amostra

Placebo			Tianeptina		
n_1	\bar{x}_1	s_1	n_2	\bar{x}_2	s_2
15	20,53	11,09	16	11,37	7,26

- Cálculo variância amostral combinada:

$$s_p^2 = \frac{(15-1)(11,09)^2 + (16-1)(7,26)^2}{15+16-2} = (9,31)^2$$

- Cálculo estatística T_{obs} :

$$t_{obs} = \frac{20,53 - 11,37}{9,31 \sqrt{\left(\frac{1}{15} + \frac{1}{16}\right)}} = 2,74$$

- Determinação p-valor:

$$p = 2 \times \Pr \{T > |2,74|\} = 0.01040262$$

- Valor crítico para teste com nível $\alpha = 0,05$

$$t_{crit} = t_{(15+16-2), (1-0.05/2)} = 2.04523$$

- Conclusão do teste:

- √ Há evidências amostrais de diferença significativa entre os dois grupos comparados
- √ Recomenda-se inspeção adicional sobre a eficácia da tianeptina

- Saída computacional – R:

√ Análise descritiva dos dados:

Média

```
> by(escores, grupo, mean)
grupo: Placebo
[1] 20.53333
-----
grupo: Tianeptina
[1] 11.375
```

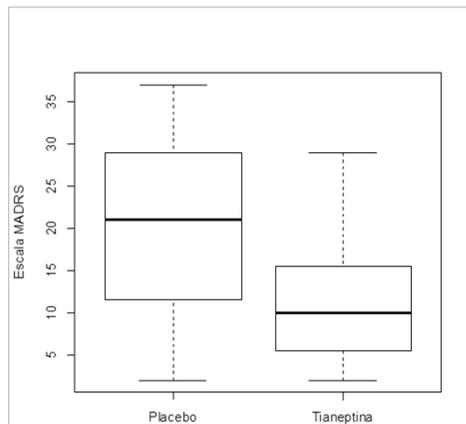
Desvio-padrão

```
> by(escores, grupo, sd)
grupo: Placebo
[1] 11.08968
-----
grupo: Tianeptina
[1] 7.256032
```

Tamanho

```
> by(escores, grupo, length)
grupo: Placebo
[1] 15
-----
grupo: Tianeptina
[1] 16
```

- Box-plot dos grupos



- Teste t:

√ Supondo variâncias iguais

```
> t.test(escores ~ grupos, var.equal = T, conf.level = 0.95)

Two Sample t-test
data:  escores by grupos
t = 2.7383, df = 29, p-value = 0.01045
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
 mean in group Placebo mean in group Tianeptina
 20.53333                11.37500
```

Teste de Igualdade de Variâncias

- Suposição de igualdade de variâncias dos dois grupos pode ser testada
- Há vários testes paramétricos e não-paramétricos disponíveis nos pacotes:
 - √ Teste F
 - √ Teste de Bartlett
 - √ Teste de Levene

• Teste F:

√ $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_1: \sigma_1^2 \neq \sigma_2^2$

√ Suposição:

- Amostras provenientes de população normal

√ Estatística de teste:

$$F = \frac{s_1^2}{s_2^2}$$

√ Distribuição estatística de teste:

$$F \sim F_{(n_1-1), (n_2-1)}$$

√ p-valor:

$$p = 2 \times \Pr \{ F_{(n_1-1), (n_2-1)} > F_{obs} \}$$

- No exemplo, de comparação da tianeptina:

√ Saída – R:

```
> var.test(escores ~grupos)

F test to compare two variances

data:  escores by grupos
F = 2.3358, num df = 14, denom df = 15, p-value = 0.1146
alternative hypothesis: true ratio of variances is not equal to 1
sample estimates:
ratio of variances
2.335821
```

- Pelo teste F, H_0 não é rejeitada, indicando homogeneidade de variâncias nos dois grupos
 - √ Graficamente, sugere-se que a variabilidade dos dois grupos não é a mesma.

• Teste F – Comentários

- √ O teste F é extremamente sensível à não-normalidade dos dados
- √ Para níveis de significância altos (pelo menos 0,05) e layouts balanceados, o teste F é relativamente robusto
- √ São usados para testar diferenças entre as médias de 3 ou mais grupos
- Alternativas ao teste F:
 - √ Teste de Bartlett
 - √ Teste de Levine

- Teste de Bartlett:

√ Usado para testar homogeneidade de variâncias de k amostras

√ $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ versus

$H_1: \sigma_i^2 \neq \sigma_j^2$, para pelo menos um par (i, j)

√ Estatística de teste (T)

√ Distribuição estatística de teste (sob H_0): $T \sim \chi_{k-1}^2$

√ p-valor:

$$p = \Pr \{ \chi_{k-1}^2 > T_{obs} \}$$

√ O teste de Bartlett é **sensível** a desvios da normalidade

- No exemplo, de comparação da tianeptina:

√ Saída – R:

```
> bartlett.test(escores-grupos)
Bartlett test of homogeneity of variances
data:  escores by grupos
Bartlett's K-squared = 2.4699, df = 1, p-value = 0.116
```

- Pelo teste de Bartlett, H_0 não é rejeitada, indicando que as variâncias populacionais dos dois grupos são iguais

- Teste de Levine:

√ Usado para testar homogeneidade de variâncias de k amostras

√ $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ versus

$H_1: \sigma_i^2 \neq \sigma_j^2$, para pelo menos um par (i, j)

√ Estatística de teste (W)

√ Distribuição estatística de teste (sob H_0):

$$W \sim F_{(k-1), (N-k)}$$

√ p-valor:

$$p = \Pr \{ F_{(k-1), (N-k)} > W_{obs} \}$$

√ O teste de Levene é mais robusto a desvios da normalidade

- No exemplo, de comparação da tianeptina:

√ Saída – R:

```
> leveneTest(escores-grupos)
Levene's Test for Homogeneity of Variance (center = median)
Df F value Pr(>F)
group 1 4.1212 0.05161
29
```

- O teste de Levene é limítrofe no nível de significância de 5%, indicando que evidência de homocedasticidade não é tão clara para esse conjunto de dados.

- No caso de dúvidas sobre a homogeneidade das variâncias dos dois grupos recomenda-se comparar resultados para duas possibilidades:

√ Variâncias iguais:

- teste t

√ Variâncias diferentes:

- teste t com correção de Welch

- Variâncias iguais:

```
> t.test(escores ~ grupos, var.equal = T, conf.level = 0.95)

Two Sample t-test
data:  escores by grupos
t = 2.7383, df = 29, p-value = 0.01045
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean in group Placebo mean in group Tianeptina
20.53333                11.37500
```

- Variâncias diferentes:

```
> t.test(escores ~ grupos, conf.level = 0.95) # Correção de Welch

Welch Two Sample t-test
data:  escores by grupos
t = 2.7019, df = 23.9, p-value = 0.01248
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean in group Placebo mean in group Tianeptina
20.53333                11.37500
```

- Conclusão:

√ Nos dois casos considerados (variâncias iguais ou diferentes) conclui-se pela diferença significativa entre os dois grupos comparados

√ Apesar de um escore MADRS não ter distribuição normal, pode-se verificar que esse pressuposto é razoável

√ Recomenda-se o uso do teste de Mann-Whitney (não paramétrico)

- É mais apropriado pois a escala usada para medir depressão é ordinal

Exemplo 6.10 – Continuação

- Fatores de risco para AVC

(continuação Exemplo 6.3)

√ Agregadas diversas variáveis quantitativas

√ Homens nascidos em Gotemburgo, em 1913, em dias múltiplos de 3.

√ Total: 789 homens

– Sofreram AVC: 57 homens

– Não sofreram: 732 homens

√ Duração do estudo: 1967 a 1986

- Valores médios dos fatores estudados
 ✓ p-valor obtido por meio de teste t de comparação de médias

Fator	AVC		p-valor
	Sim	Não	
Fumo	2,9	2,6	0,16
Pressão sistólica (mmHg)	151,9	143,2	0,004
Pressão diastólica (mmHg)	96,8	90,1	0,0001
Índice de massa corporal (kg/m ²)	25,7	24,9	0,08
Medida do quadril (cm)	94,8	93,6	0,23
Índice de obesidade abdominal	0,95	0,93	0,0004

- Valores médios dos fatores estudados
 (continuação)

Fator	AVC		p-valor
	Sim	Não	
Colesterol (mg/dL)	275	269	0,40
Glicose (mg/dL)	64,8	67,9	0,60
Fibrinogênio (g/L)	3,6	3,3	0,01
Hemoglobina (g/dL)	14,9	14,8	0,40
Hematócrito (%)	44,3	44,0	0,46
Capacidade vital (L)	4,3	4,5	0,03

- Conclusão:
 ✓ São *possíveis* fatores de risco para o AVC em homens:
 - Pressão sistólica
 - Pressão diastólica
 - Índice de obesidade abdominal
 - Fibrinogênio
 - Capacidade vital
- Comentário:
 ✓ Utilizadas análises multivariadas na verificação final da influência destes fatores na ocorrência de AVC

Importante

- Pressuposto para aplicação do teste t:
 ✓ Os dois grupos devem ter a **mesma variabilidade**
- Isso nem sempre acontece na prática!

Teste Z para Comparação de Médias

- Pode ser utilizado para comparação de grupos com **variabilidades diferentes**
 - √ Não é necessária nenhuma suposição sobre σ_1^2 e σ_2^2 (podem ser iguais ou diferentes)
 - √ Usado no caso de amostras grandes (n_1 e $n_2 \geq 30$)

- Hipóteses:
 - √ $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$

- Estatística de teste:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Distribuição da estatística de teste (sob H_0):

- √ Resultado assintótico

$$Z \approx N(0, 1)$$

- Teste z:

- √ Valor observado da estatística de teste:

$$z_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- √ Cálculo do p-valor (teste bilateral):

$$p = 2 \times \Pr \{Z > |z_{obs}|\}$$

- √ Região de rejeição de H_0 :

- Valor crítico para nível de significância α :

$$z_{crit} = z_{1-\alpha/2}$$

- Critério de rejeição de H_0 :

$$|z_{obs}| > z_{crit}$$

Exemplo 6.11

- (Conahan et al., 1973) Efeito do halotano em cirurgias cardíacas:

- √ Tratamentos:

- Morfina: pequeno efeito sobre atividade cardíaca
- Halotano: gás anestésico de inalação

- √ Possíveis efeitos colaterais:

- Depressão do sistema respiratório e cardiovascular, sensibilização a arritmias, lesões hepáticas

- √ Estudo comparativos destes agentes anestésicos

- Pacientes alocados aleatoriamente a cada grupo
- Efeitos dos dois tratamentos na pressão sanguínea são iguais?

- Resultados amostrais:

Halotano			Morfina		
n_1	\bar{x}_1	s_1	n_2	\bar{x}_2	s_2
61	66,9	12,2	61	73,2	14,4

- Estatística de teste:

$$z_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{66,9 - 73,2}{\sqrt{\frac{(12,2)^2}{61} + \frac{(14,4)^2}{61}}} = -2,61$$

- Determinação p-valor:

$$p = 2 \times \Pr\{Z > |-2,61|\} = 0.009054222$$

```
> 2*(1-pnorm(2.61))
[1] 0.009054222
```

- Região de rejeição de H_0 :

√ Valor crítico para nível de significância $\alpha = 0,05$:

$$z_{crit} = z_{0,975} = 1.959964$$

```
> qnorm(.975)
[1] 1.959964
```

√ Critério de decisão:

$$|z_{obs}| = 2,61 > z_{crit} = 1,96$$

√ Decisão:

- Há evidências amostrais para se rejeitar H_0 .

Resposta Contínua: Amostras Pareadas

Pareamento

- Muito utilizado em planejamento de estudos da área de saúde
- Situações:
 - √ Medidas tomadas no mesmo indivíduo em ocasiões diferentes
 - √ Delineamentos em que um caso é pareado com um controle, para controlar fatores de confusão
- As amostras são pareadas e não independentes

- Exemplo de pareamento:
 - √ Amostras pareadas na situação de uma medida basal e a outra após o tratamento
 - √ Efeito do tratamento refletido na diferença de medidas dentro de cada par

Teste t para Amostras Pareadas

- Comparação de dois grupos:
 - √ $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$
 - √ $H_0: \mu_1 - \mu_2 = 0$ vs. $H_1: \mu_1 - \mu_2 \neq 0$
 - √ $H_0: \mu_d = 0$ vs. $H_1: \mu_d \neq 0$ (média da diferença)
- Pares de observações:
 - $(x_{11}, x_{21}), (x_{12}, x_{22}) \dots (x_{1n}, x_{2n})$
- Diferenças de cada par:
 - $d_1 = x_{11} - x_{21}, d_2 = x_{12} - x_{22}, \dots, d_n = x_{1n} - x_{2n}$

- Média e desvio-padrão das diferenças amostrais:

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$$

$$s_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n - 1}}$$

- Estatística de teste:

$$T_p = \frac{\bar{D}}{\frac{s_d}{\sqrt{n}}}$$

- Distribuição estatística de teste (sob H_0)

$$T_p \sim t_{n-1}$$

- √ Cálculo do p-valor (teste bilateral):

$$p = 2 \times \Pr \{T_p > |t_{p_{obs}}|\}$$

- √ Região de rejeição de H_0 :

- Valor crítico para nível de significância α :

$$t_{p_{crit}} = t_{n-1, (1-\alpha/2)}$$

- Critério de rejeição de H_0 :

$$|t_{p_{obs}}| > t_{p_{crit}}$$

Exemplo 6.12

- Programa para redução do nível de colesterol:
 - √ Estudo com objetivo de avaliar efetividade de dieta combinada com programa de exercícios
 - √ Amostra com 12 indivíduos
 - Nível colesterol medido no início e no final do programa

- Níveis de colesterol – Início e Final:

		Diferença	Desvio	
Início	Final	$d = x_1 - x_2$	$d - \bar{d}$	$(d - \bar{d})^2$
201	200	1	-19,16	367,36
231	236	-5	-25,16	633,36
221	216	5	-15,16	230,03
260	233	27	6,83	46,69
228	224	4	-16,16	261,36
237	216	21	0,83	0,69
326	296	30	9,83	96,69
235	195	40	19,83	393,36
240	207	33	12,83	164,69
267	247	20	-0,16	0,03
284	210	74	53,83	2898,03
201	209	-8	-28,16	793,36

- Diferenças – valores observados:

$$\bar{d} = 20,17$$

$$s_d = 23,13$$

- Cálculo estatística T_{obs} :

$$t_{p_{obs}} = \frac{20,17}{\frac{23,13}{\sqrt{12}}} = 3,02$$

- Determinação p-valor:

$$p = 2 \times \Pr \{T > |3,02|\} = 0.012$$

- Valor crítico para teste com nível $\alpha = 0,05$

$$t_{crit} = t_{(12-1), (1-0.05/2)} = 2.200985$$

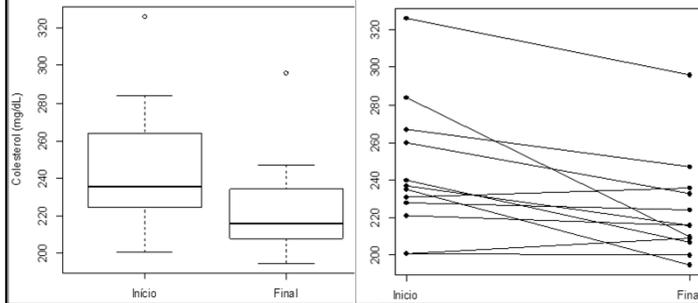
- Decisão:

√ Rejeita-se H_0

- Conclusão:

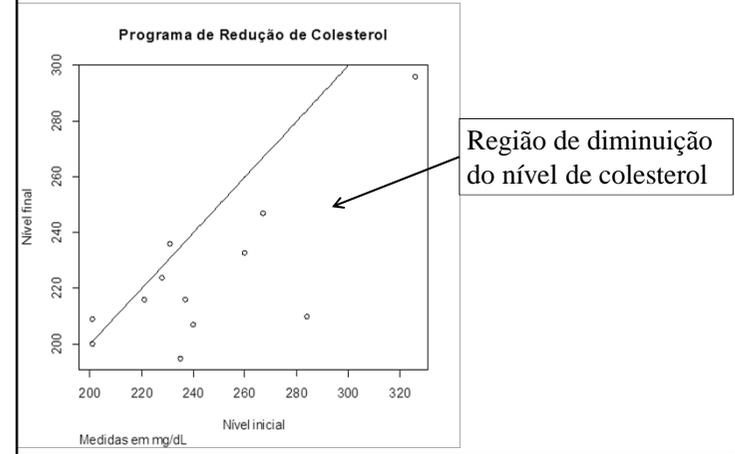
√ Há evidências amostrais de que, em média, o programa altera o nível de colesterol

- Análise gráfica do emparelhamento:



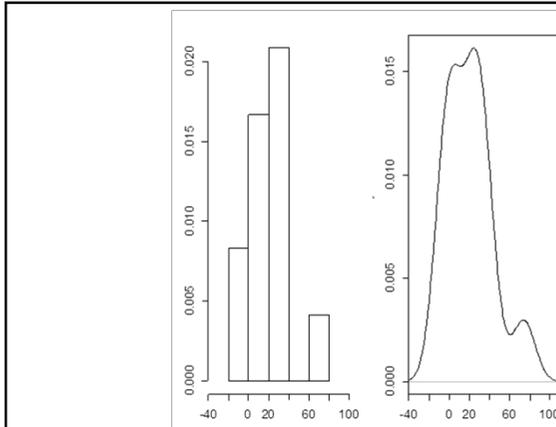
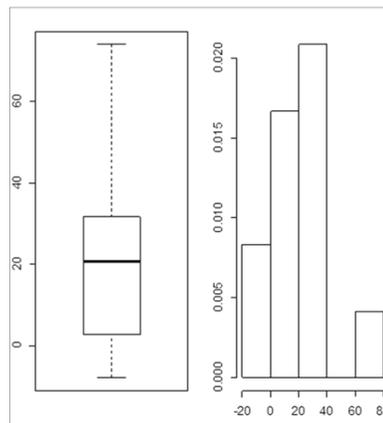
√ Essa comparação pode ser efetuada?

- Uma outra visualização:



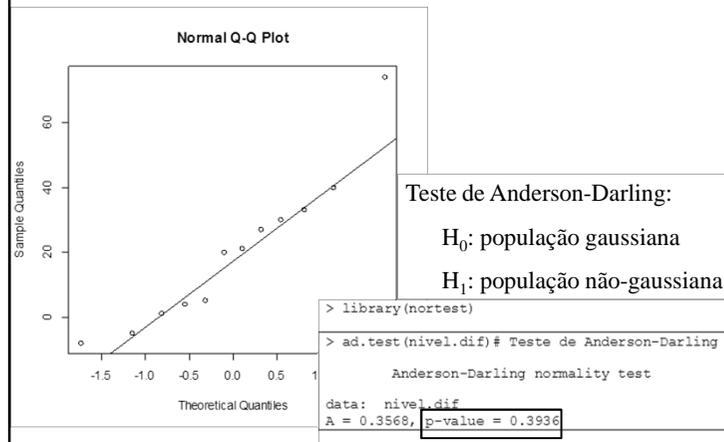
- Análise descritiva das diferenças:

Diferenças	
n_1	x_1
Média	20,17
Desvio-padrão	23,13
Tamanho (n)	12



- A população das diferenças é normal (gaussiana)?

- Verificação de normalidade:



- Teste t emparelhado:

```
> t.test(inicio, final, paired=T, alternative="two.sided") # teste independente

Paired t-test

data: inicio and final
t = 3.0201, df = 11, p-value = 0.01165
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean of the differences
20.16667
```

√ Há evidências amostrais para se rejeitar H₀

- Conclusão:

√ Em média, o programa altera o nível de colesterol

Testes Não-paramétricos

- Testes com resposta contínua (amostras independentes e pareadas):

√ Efeito do tratamento é descrito por variável contínua com distribuição gaussiana

√ Estatísticas de teste usam estimativas dos parâmetros da distribuição gaussiana

√ São denominados *testes paramétricos*.

Testes Não-paramétricos

- Não estão condicionados a qualquer distribuição de probabilidades dos dados em análise
 - √ São designados por “*distribution-free tests*”.
- São boas opções para situações em que ocorrem violações dos pressupostos básicos necessários para aplicação de um teste paramétrico
 - √ Exemplos:
 - Testar diferença de dois grupos quando a distribuição subjacente é assimétrica
 - Dados coletados em um escala ordinal

• Alguns testes não-paramétricos

- √ Teste de Mann-Whitney
 - Amostras independentes
- √ Teste de Wilcoxon
 - Amostras pareadas
- √ Desenvolvidos para comparar grupos em situações em que a distribuição variável de interesse não é conhecida ou tem comportamento não-gaussiano
- √ Cálculos simples, mas trabalhosos
- √ Baseados em postos (rank) das observações
 - Número de ordem das observações ordenadas
 - Quando há empates, toma-se a média dos postos que seriam atribuídos às observações, caso os empates não existissem

Teste de Mann-Whitney

- Comparação de dois grupos independentes:
 - √ Obtém-se os postos de todas as observações com se elas fossem uma única amostra
 - √ Estatística de teste:

$$MW = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - T$$
 - √ n_1 e n_2 : tamanhos das amostras dos dois grupos ($n_1 \leq n_2$)
 - √ T: total de postos do grupo menor

• Tomada de decisão:

- √ Valor da estatística MW é comparado com percentil de distribuição especial

• Alternativa:

- √ Para amostras com pelo menos 10 em cada grupo
- √ Teste pode ser baseado na estatística $T \approx N(\mu_T, \sigma_T)$

$$\mu_T = \frac{n_P(n_P + n_G + 1)}{2} \quad Z = \frac{T - \mu_T}{\sigma_T} \sim N(0, 1)$$

$$\sigma_T = \sqrt{\frac{n_G \mu_T}{6}}$$

- n_P e n_G : tamanhos grupos menor e maior, respectivamente

Exemplo 6.13

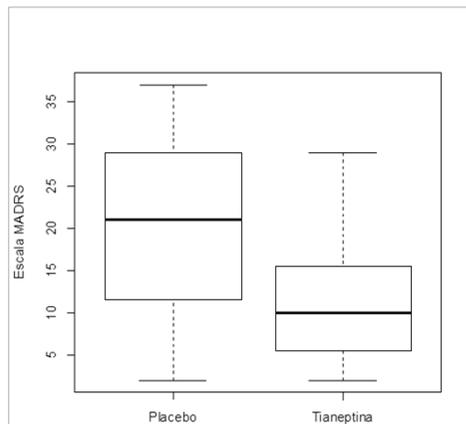
- (Continuação exemplo 6.9)
 - √ Comparação Tianeptina com Placebo
 - Fármaco antidepressivo
 - √ Ensaio clínico aleatorizado, duplo-cego
 - √ Pacientes de Belo Horizonte, Rio e Campinas
 - √ Quantificação depressão:
 - escala de Montgomery-Asberg (MADRS)
 - √ Escores obtidos para cada paciente:
 - 7, 14, 21, 28 e 42 dias após início do ensaio

- Classificação da escala MADRS:

Score	Classificação
Zero a 6	Ausência de depressão
7 a 19	Depressão leve
20 a 34	Depressão moderada
35 a 60	Depressão grave

- √ Um escore não tem teoricamente a distribuição normal, pois, **não** é uma variável contínua

- Box-plot dos grupos



- Escore final na escala MADRS

- √ Pacientes dos dois grupos, admitidos em BH

Grupo	Escore							
Placebo	6	33	21	26	10	29	33	29
	37	15	2	21	7	26	13	
Tianeptina	10	8	17	4	17	14	9	4
	21	3	7	10	29	13	14	2

- √ $n_p = 15$ e $n_G = 16$

- Postos atribuídos aos dados agregados

Placebo		Tianeptina		Placebo		Tianeptina	
Escore	Posto	Escore	Posto	Escore	Posto	Escore	Posto
2	1,5	2	1,5	15	18		
		3	3			17	19,5
		4	4,5			17	19,5
		4	4,5	21	22	21	22
6	6			21	22		
7	7,5	7	7,5	26	24,5		
		8	9	26	24,5		
		9	10	29	27	29	27
		10	12	29	27		
10	12	10	12	33	29,5		
13	14,5	13	14,5	33	29,5		
		14	16,5	37	31		
		14	16,5				
				T = 296,5			

- Estatística de teste (W):

$$MW = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - T$$

$$= (15)(16) + \frac{(15)(15 + 1)}{2} - 296,5 = 63,5$$

- Resultado do teste (saída R):

```
> wilcox.test(tianeptina, placebo, data=farmaco.data, correct=T, paired=F)

Wilcoxon rank sum test with continuity correction

data: tianeptina and placebo
W = 63.5, p-value = 0.02655
alternative hypothesis: true location shift is not equal to 0
```

√ O p-valor do teste de Mann-Whitney indica diferença entre os dois grupos comparados

- Cálculo alternativo (n_1 e $n_2 > 10$)

√ Parâmetros estatística de teste:

$$\mu_T = \frac{n_P(n_P + n_G + 1)}{2} = \frac{15(15 + 16 + 1)}{2} = 240$$

$$\sigma_T = \sqrt{\frac{n_G \mu_T}{6}} = \sqrt{\frac{(16)(240)}{6}} = 25,298$$

√ Estatística de teste:

$$Z = \frac{296,5 - 240}{25,298} = 2,233$$

√ Cálculo p-valor:

$$p = 2 \times \Pr \{Z > |2,233|\} = 0.0255$$

- Estatística T

(Teste paramétrico executado anteriormente – ex. 6.9)

- Variâncias iguais:

```
> t.test(escores ~ grupos, var.equal = T, conf.level = 0.95)

Two Sample t-test

data:  scores by grupos
t = 2.7383, df = 29, p-value = 0.01045
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
 mean in group Placebo mean in group Tianeptina
 20.53333                11.37500
```

- Variâncias diferentes:

```
> t.test(escores ~ grupos, conf.level = 0.95) # Correção de Welch

Welch Two Sample t-test

data:  scores by grupos
t = 2.7019, df = 23.9, p-value = 0.01248
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
 mean in group Placebo mean in group Tianeptina
 20.53333                11.37500
```

- Resultados obtidos na comparação dos grupos:

	Não-paramétricos		Paramétricos	
	Mann-Whitney		Teste t – Independente	
	Corrigida	Aproximada	σ^2 iguais	(cor. Welch)
Estatística	MW = 63,5	Z = 2,233	T = 2,738	T = 2,702
Distribuição	Tabelada	N(0,1)	t_{29}	t_{29}
p-valor	0,0266	0,0255	0,01045	0,01248

- √ O p-valor do teste de Mann-Whitney é mais que o dobro do p-valor obtido pelos testes t
- √ Uma possível explicação é que o teste não-paramétrico é menos eficiente
- √ Para ter resultados mais próximos do teste paramétrico, deveria ser utilizada amostra maior

- Conclusão do teste:

√ Há evidências amostrais de diferença significativa entre os dois grupos comparados

Teste de Wilcoxon

- Usado para comparar dois tratamentos quando os dados são pareados
- Procedimento:
 1. Calcular a diferença entre as observações por par
 2. Ignorar o sinal das diferenças, atribuindo postos
 3. Calcular T^+ : soma dos postos dos d_i 's positivos
 4. Calcular T^- : soma dos postos dos d_i 's negativos

- Soma de todos os postos: $\frac{n(n+1)}{2}$

- Relação entre as estatísticas: $T^- = \frac{n(n+1)}{2} - T^+$

- Estatística de teste: T^+ .

- Empates:

√ São retiradas da análise as diferenças nulas e o tamanho da amostra é diminuído

√ Dois ou mais d_i 's podem ter a mesma magnitude

- Atribui-se a média dos postos caso tivessem diferido

- Distribuição da estatística de teste:

- √ Para amostras pequenas ($n \leq 25$):
 - p-valor é obtido através de tabela específica
- √ Para amostras grandes ($n > 25$)

$$Z = \frac{T^+ - \mu_{T^+}}{\sigma_{T^+}} \underset{as.}{\sim} N(0, 1) \quad \mu_{T^+} = \frac{n(n+1)}{4}$$

$$\sigma_{T^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

- Não há suposição alguma sobre a forma da distribuição subjacente dos dados
 - √ Teste é mais poderoso para distribuições simétricas

Exemplo 6-14

- Evolução do tratamento com tianeptina:
 - √ Dados dos pacientes admitidos em BH
 - √ Escores no primeiro dia (x_1) e no último dia (x_{42})
 - √ Escore MADRS é uma medida em escala ordinal
 - Teste não-paramétrico é mais recomendado
 - O uso do teste t só se justifica como aproximação
 - √ Dois pacientes (14 e 15) não completaram o estudo

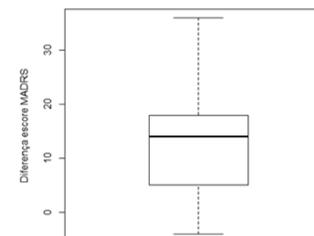
- Escores do grupo tianeptina dia 1 e dia 42

Nº	x_1	x_{42}	$d=x_{42}-x_1$	Nº	x_1	x_{42}	$d=x_{42}-x_1$
1	24	6	-18	9	35	37	2
2	46	33	-13	10	30	15	-15
3	26	21	-5	11	38	2	-36
4	44	26	-18	12	38	21	-17
5	27	10	-17	13	31	7	-24
6	34	29	-5	14	27	*	*
7	33	33	0	15	34	*	*
8	25	29	4	16	32	26	-6

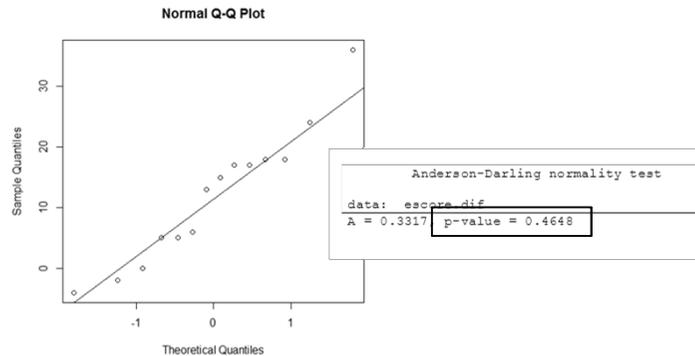
- √ Estatísticas descritivas das diferenças:

```
> mean(primeiro)
[1] 33.07143
> mean(ultimo)
[1] 21.07143
> mean(escore.dif)
[1] 12
> sd(escore.dif)
[1] 11.03142
```

- √ Gráfico das diferenças:



• Teste de normalidade das diferenças:



- ✓ Não rejeita a hipótese de normalidade
- ✓ Escore não é uma variável contínua!

• Teste t emparelhado:

```
> t.test(primeiro, ultimo, paired=T, alternative="two.sided")

Paired t-test

data: primeiro and ultimo
t = 4.0702, df = 13, p-value = 0.001325
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 5.630646 18.369354
sample estimates:
mean of the differences
12
```

- ✓ Rejeita-se a hipótese de igualdade das médias dos grupos
- ✓ Teste t só se justifica como aproximação, pois o escore MADRS é uma medida em escala ordinal
 - Mesmo assim, hipótese de normalidade mostrou-se razoável
- ✓ Mais apropriado uso de teste não-paramétrico

• Posto atribuído ao valor absoluta da diferença:

d	Posto	Nº	d	Posto	Nº
0	Elim.	7	15	7	10
2	1	9	17	8,5	5
4	2	8	17	8,5	12
5	3,5	3	18	10,5	1
5	3,5	6	18	10,5	4
6	5	16	24	12	13
13	6	2	36	13	11

S = 105

- ✓ Não se aplica o procedimento para amostras grandes (aproximação normal)

• Posto das diferenças +:

```
sort(diff.rank.sign[diff.rank.sign > 0])
[1] 2 1
```

• Estatística T⁺: 3

- ✓ Estatística V (R)

• Posto das diferenças -:

```
sort(diff.rank.sign[diff.rank.sign < 0])
[1] 3.5 3.5 5.0 6.0 7.0 8.5 8.5 10.5 10.5 12.0 13.0
```

• Estatística T⁻: 88

- Saída do R:

```
> wilcox.test(ultimo, primeiro, data=tianeptina.data, correct=F, exact = F,  
+ paired=T)  
  
Wilcoxon signed rank test  
data: ultimo and primeiro  
V = 3, p-value = 0.00295  
alternative hypothesis: true location shift is not equal to 0
```

- √ O p-valor referente ao teste de Wilcoxon é mais alto que o obtido pelo teste t para amostras pareadas ($p = 0,001$)
- √ Ao nível de significância de 5% há indicações de alteração dos níveis de depressão para pacientes que fizeram uso de tianeptina

- Conclusão:

- √ Ao nível de significância de 5% há indicações de alteração dos níveis de depressão para pacientes que fizeram uso de tianeptina

Importante

- Pode haver discrepância entre os resultados do teste de igualdade de variâncias
- Pode haver diferenças entre os resultados do teste paramétrico e não paramétrico

Verificação de Adequação do Modelo – Distribuição Gaussiana

Verificação de Adequação de Modelo

- Verificação se a variável estudada pode ser descrita adequadamente por distribuição gaussiana:

√ Exame gráfico (histograma e/ou box-plot)

- Gráfico deve apresentar razoável simetria e forma aproximada de sino

√ Verificação das proporções de dados em determinados intervalos

$\bar{x} \pm s \approx 68\%$ das observações

$\bar{x} \pm 2s \approx 95\%$ das observações

$\bar{x} \pm 3s \approx$ todas as observações

- Critérios mais precisos:

√ Q-Q plot

- Disponível nos pacotes estatísticos
- Se houver boa aderência dos dados ao modelo gaussiano, as observações estarão alinhadas em torno de uma reta destacada no gráfico

√ Testes de aderência:

- Teste de Anderson-Darling
- Teste de Shapiro-Wilk
- Teste de Kolmogorov-Smirnov

√ Hipóteses dos testes:

- H_0 : Amostra provém de distribuição normal (teórica)
- H_1 : Amostra não provém de distribuição normal

Exemplo 5-12

- Pressão sistólica de estudantes

√ Amostra com 60 estudantes do sexo masculino

√ Medidas de pressão sistólica

- Estatísticas descritivas da amostra:

```
> mean(sistolica)
[1] 122.5333
> sd(sistolica)
[1] 12.45394
```

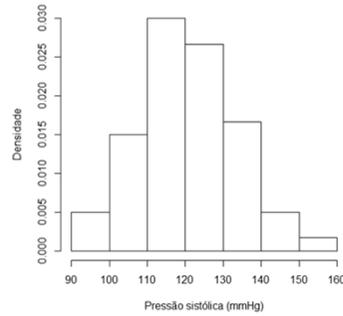
- Proporção de dados em intervalos

Intervalo	Esperado	Observado
(110,1; 135,5)	68,3%	63,3%
(97,6; 147,4)	95,4%	95,0%
(85,2; 159,9)	99,7%	100%

√ Não há grandes discrepâncias entre as porcentagens e as probabilidades teóricas

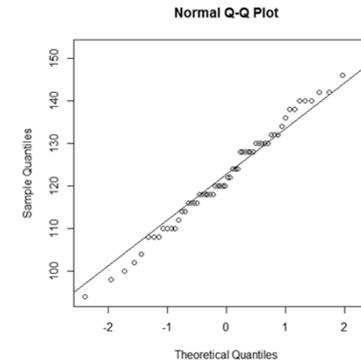
√ Modelo de Gauss para esta variável é razoável

- Histograma:



√ Histograma simétrico indicando possibilidade de normalidade dos dados

- Q-Q plot:



√ Gráfico indica uma excelente aderência dos dados ao modelo gaussiano

- Testes formais de aderência:

- √ Anderson-Darling:

```
> ad.test(sistolica)# Teste de Anderson-Darling
Anderson-Darling normality test
data:  sistolica
A = 0.253, p-value = 0.7237
```

- √ Shapiro-Wilk

```
> shapiro.test(sistolica)# Teste de Shapiro-Wilk
Shapiro-Wilk normality test
data:  sistolica
W = 0.9905, p-value = 0.9233
```

√ Evidências amostrais apontam adequação do modelo gaussiano aos dados

Considerações Práticas sobre Testes de Hipóteses

p-valor

- Decisão final sobre H_0 tomada comparando-se p -valor com valor pré-fixado.
 - √ Usualmente 0,05
- Resultados do teste
 - √ $p < 0.05$: resultado estatisticamente significativo
 - √ $p < 0,01$: resultado altamente significativo
 - √ Outros valores de p : resultado não significantes
- Testes são denominados **testes de significância**

- Pontos de corte são arbitrários:

- √ Não se deve dar importância exagerada a eles!
- √ Resultados de dois estudos em que os p -valores sejam 0,045 e 0,055
 - Esses valores devem levar a conclusões muito parecidas e não a conclusões opostas
- √ Necessária uma medida do efeito de uma intervenção ou exposição

Hipóteses Bilaterais vs. Unilaterais

- Hipóteses alternativas consideradas:
 - √ $H_1: p_1 \neq p_2$, para proporções
 - √ $H_1: \mu_1 \neq \mu_2$, para médias
 - √ Hipóteses bilaterais
 - √ Qualquer dos dois grupos pode ter proporção (ou média) maior do que o outro
 - √ p -valor: probabilidade de se obter diferença igual ou mais extrema em qualquer direção

- Hipóteses alternativas podem ser desmembradas:

- √ $H_1: p_1 > p_2$ ou $p_1 < p_2$.
- √ $H_1: \mu_1 > \mu_2$ ou $\mu_1 < \mu_2$.
- √ Hipóteses unilaterais
- √ As comparações são estabelecidas em uma determinada direção

- A escolha de hipóteses bilaterais ou unilaterais influencia decisivamente a interpretação dos resultados
- Para testar hipóteses bilaterais baseadas na distribuição t ou distribuição gaussiana
 - √ Utiliza-se apenas uma cauda da distribuição
 - √ Normal para $\alpha = 5\%$: percentil é 1,64 e não 1,96
- Em geral o p-valor para teste bilateral é o dobro do correspondente à hipótese unilateral

- Há situações em que a hipótese unilateral é a melhor forma de formulação da questão de interesse
 - √ Estudos de comparação de inovação de procedimento padrão
 - √ Estudos sobre qualidade da água:
 - Contagens bacteriológicas estão acima do padrão aceitável
 - √ Uma das hipóteses contempladas pela hipótese bilateral é completamente inconcebível

- Argumentos contra o uso de hipóteses unilaterais
 - √ Por maior que seja a evidência de que um tratamento seja superior ao outro, nunca se tem certeza absoluta do que realmente pode acontecer
 - √ Exemplo:
 - Na comparação de uma droga com placebo, não se pode descartar a possibilidade de que os resultados sejam piores

- Aspectos a serem considerados na escolha entre hipótese bilateral e unilateral:
 - √ O tipo de hipótese deve preceder a análise dos dados
 - A escolha não deve ser influenciada pelo resultado
 - √ O teste bilateral é mais conservador do que o unilateral
 - Na maioria dos casos, testes unilaterais são vistos como uma maneira de exagerar a força dos achados
 - Ante qualquer dúvida, usar o bilateral
 - √ A distinção entre testes unilaterais e bilaterais não é de fundamental importância na interpretação
 - Deve estar claro qual foi usado
 - √ Teste bilateral é a forma padrão usado em periódicos

Conclusões a partir dos Testes de Hipóteses

- A partir de testes bilaterais não se pode concluir que um tratamento é superior a outro
 - √ O teste permite concluir sobre diferença entre tratamentos
 - √ Maneira apropriada para avaliar qual tratamento é mais eficiente
 - Quantificação dos “efeitos” dos tratamentos

• O p-valor oferece evidência probabilística

- √ p-valor grande (0,20, p.e.)
 - A evidência fornecida sugere equivalência dos tratamentos
 - Estamos mais seguros que são equivalentes
 - Não se pode excluir a possibilidade de que os tratamentos sejam diferentes
- √ p-valor pequeno (<0,001, p.e.)
 - Igualdade dos dois tratamentos não parece ser plausível
- √ p-valor de referência de 0,05:
 - Um resultado que pode ocorrer menos que uma vez em 20 quando H_0 é verdadeira.
 - Quando se rejeita H_0 , aceita-se H_1 complementar

- Se p-valor não excede ponto de corte
 - √ Não se rejeita H_0 .
 - √ Não se pode dizer que H_0 não é verdadeira, mas somente que não há evidência suficiente para que ela seja rejeitada
- Conclusão:
 - √ Não se deve fixar no critério de $p < 0,05$
 - Deve-se interpretar corretamente o resultado através do p-valor

Significância Estatística e Significância Clínica

- Resultado estatisticamente significativo não implica necessariamente em importância clínica
- Pode ser que teste que não tenha sido estatisticamente significativo, mas dados mostrem uma diferença importante na prática
 - √ Efeitos podem estar ligados ao tamanho da amostra

Fatores de Confusão

- Nem sempre as diferenças observada entre as repostas pode ser atribuída ao fator usado formar os dois grupos de comparação
 - √ Eventuais diferenças podem ser atribuídas a um outro fator (fator de confusão)
- Deve sempre ser considerada a possibilidade de existência de fatores de confusão
 - √ Respostas clínicas (sobrevida, remissão, recidiva, desaparecimento de sintomas, etc.) podem ser influenciadas por vários fatores

Métodos multivariados

- √ Métodos estatísticos para se fazer comparações válidas em presença de fatores de confusão

Apresentação dos Resultados

- Análises que envolvem testes de hipóteses:
 - √ Apresentar os valores necessários para se recalculer o p-valor
 - √ Evitar apresentação de p-valor em forma de intervalo
 - Recomendado apresentar o p-valor obtido (em duas ou três casas decimais)

Pressupostos dos Testes

- Questões importantes após a escolha do teste a ser utilizado:
 - √ Os dados analisados satisfazem os pressupostos?
 - √ Se eles não forem completamente satisfeitos, quais as consequências? São graves?

- Algumas formas de verificação dos pressupostos:
 - √ Análise descritiva dos dados
 - √ Gráficos específicos
 - √ Testes de aderência
 - √ Etc.
- Em geral, a violação da suposição altera o nível de significância
 - √ Ex.: pode-se imaginar que está testando a um nível de significância de 5% e o nível ser realmente maior

- Violação da condição (1):
 - √ Necessário usar outra análise
 - √ Ex.
 - Combinação de tabelas 2x2 (Fleiss, 1981).
- Violação da hipótese de independência (2):
 - √ Maior dificuldade na análise
 - √ Métodos adequados de análise estão agrupados no tópico Análise de Dados Dependentes

Teste Qui-quadrado – Pressupostos

- Suposições:
 1. A probabilidade de sucesso não varia de indivíduo para indivíduo dentro dos grupos de controle e de tratamento
 2. O resultado que ocorre para um paciente não influencia de nenhuma maneira o resultado de outros indivíduos

Teste t para Comparação de Médias – Pressupostos

- Suposições – Amostras independentes:
 1. Variável resposta tem distribuição gaussiana
 2. Variâncias dos dois grupos são iguais
- A hipótese de igualdade de médias é mais geral:
 - √ Equivale à hipótese de que os dois tratamentos produzem o mesmo efeito
 - √ Na prática, é o tipo de comparação que mais interessa

Consequências de Violações

- Suposição de normalidade e homocedasticidade
 - √ Se os tamanhos das amostras dos dois grupos são grandes e iguais ($n_1 = n_2$):
 - O nível de significância é válido, mesmo se as hipóteses forem violadas
 - √ Se $n_1 \neq n_2$ e os desvios-padrão não são iguais
 - Nível de significância pode ser alterado por uma quantidade que depende de quantas vezes um desvio-padrão é maior que o outro

- Violação da hipótese (1):
 - √ O teste t é robusto com relação à suposição da distribuição subjacente
 - √ Teste apropriado mesmo para variáveis que sejam um pouco assimétricas ou que efetivamente não tenham distribuição gaussiana (escalas)
- Violação da hipótese de homocedasticidade (2):
 - √ Importante ser avaliada
 - √ Violação pode resultar em conclusões incorretas

- Homocedasticidade pode ser avaliada por teste de igualdade de variância
 - √ Há vários testes paramétricos e não-paramétricos (disponíveis nos pacotes):
 - Teste F
 - Teste de Bartlett
 - Teste de Levene
- Pode-se tentar corrigir heterocedasticidade pelo uso de transformação de variável

- Além do aspecto técnico, é importante estar atento ao significado das hipóteses no caso de variâncias diferentes:
 - √ Dois tratamentos que tenham a mesma média podem ter efeitos completamente diferentes e não poderiam ser considerados equivalentes

Amostras Pareadas e Independentes

- Em geral, o pareamento melhora a precisão (menor variabilidade) das diferenças entre as médias dos tratamentos.
- O pareamento nem sempre contribui para redução da variabilidade, principalmente quando:
 - √ Os indivíduos já forem homogêneos
 - √ A heterogeneidade não puder ser atribuída a fatores identificáveis
 - √ Nesses casos, recomenda-se planejamento com amostras independentes

Pareamento

- Tipos:
 - √ Autopareamento
 - √ Pareamento natural
 - √ Pareamento artificial

Autopareamento

- Ocorre quando o indivíduo serve como seu próprio controle:
 - √ Indivíduo recebe duas drogas administradas em ocasiões diferentes
 - √ Medidas da resposta em ocasiões diferentes
 - √ Comparação de dois órgãos no mesmo indivíduo segundo alguma característica estudada

Pareamento Natural

- Formar pares tão homogêneos quanto possível, controlando os fatores que possam interferir na resposta, sendo que o pareamento aparece de forma natural
 - √ Pares de cobaias selecionados da mesma ninhada
 - √ Gêmeos univitelinos

Pareamento Artificial

- Escolhem-se indivíduos com características semelhantes ou, em geral, fatores que possam influenciar de maneira relevante a variável resposta

Pareamento – Dificuldades Práticas

- Conhecimento das características que devem ser controladas
- Formação de pares homogêneos
 - √ Exemplo: quantidade de fatores é muito grande
- Há situações em que a implementação do planejamento com amostras pareadas torna-se difícil ou mesmo impossível

Amostragem

- É questão essencial em qualquer pesquisa
- A amostra deve apresentar boa representatividade da população de interesse
- Questões chave
 - √ Como amostrar?
 - √ Quanto amostrar?
- Um dos problemas que pode acompanhar a não detecção de diferença significativa é o tamanho insuficiente da amostra

- Tipos de esquemas amostrais
 - √ Amostragem aleatória simples
 - √ Amostragem aleatória estratificada
 - √ Amostragem aleatória por conglomerados
 - √ Amostragem sistemática
 - √ Combinação entre eles

- Critérios na determinação do tamanho amostral:
 - √ Precisão
 - √ Erros associados a testes de hipóteses
 - √ Poder estatístico
 - √ Custos
 - √ Combinação desses critérios
- Há fórmulas para dimensionamento de amostras que consideram o esquema amostral e/ou o planejamento do estudo

- As estatísticas de teste de média, para amostras independentes ou emparelhadas são semelhantes às aquelas apresentadas:

√ Modificação na estatística de teste:

- Numerador a diferença das estimativas dos parâmetros é comparada com o valor específico de H_0 .

- Teste Z para comparação de proporções

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

Hipóteses mais Gerais

- As hipóteses testadas (até aqui) referem-se a:
 - √ Igualdade de médias
 - √ Igualdade de proporções dos tratamentos comparados
- Pode-se querer testar que a diferença entre as médias (ou proporções) é uma certa quantidade
 - √ $H_0: \mu_1 - \mu_2 = \mu_0$ vs. $\mu_1 - \mu_2 \neq \mu_0$.
 - √ $H_0: p_1 - p_2 = p_0$ vs. $p_1 - p_2 \neq p_0$.
 - √ Estudamos um caso especial!
($\mu_0 = 0$ ou $p_0 = 0$).

- Teste t para amostras independentes:

$$t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

- Teste Z para comparação de médias (amostras grandes):

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Teste t para amostras pareadas:

$$T_p = \frac{\bar{d} - d_0}{\frac{s_d}{\sqrt{n}}}$$

Mais de Dois Grupos

- Frequentemente o objetivo da pesquisa é comparar mais de dois grupos
 - √ Necessária extensão dos testes apresentados
 - √ Não é apropriado fazer comparações duas a duas
- Comparação de médias de mais de dois tratamentos:
 - √ Análise de variância
 - √ Comparações múltiplas

- Comparações de mais de duas proporções
 - √ Pode-se utilizar o teste qui-quadrado com algumas adaptações
- Altman (1991) é referência para esses dois tópicos

Referências

Bibliografia

- Soares, F., Siqueira, A. (Coopmed)
Introdução à Estatística Médica
- Siqueira, A. e Tibúrcio, J. D. (Coopmed)
Estatística na Área da Saúde
- Riffenburgh, R. H. (Academic Press)
Statistics in Medicine