

Homepage

Página de Rosto

Índice Geral



Página 1 de 55

Voltar

Full Screen

Fechar

Desistir

Estatística Aplicada

Rita Gaio

Departamento de Matemática - FCUP

argaio@fc.up.pt

May 14, 2020

Homepage

Página de Rosto

Índice Geral



Página 2 de 55

Voltar

Full Screen

Fechar

Desistir

Índice Geral

1	Testes paramétricos no contexto de uma e duas amostras	5
1.1	Teste- t : teste de hipóteses sobre a média de uma distribuição	6
1.1.1	Relação entre intervalos de confiança para a média e o teste t	8
1.1.2	Exemplo 1	10
1.1.3	Exemplo 2	13
1.1.4	Exemplo 3	15
1.2	Amostras emparelhadas	17
1.2.1	Exemplo 1	18
1.2.2	Exemplo 2	21
1.3	Teste hip. sobre a diferença de médias de duas variáveis normais independentes (variâncias iguais e desconhecidas) - Teste t	25
1.4	Teste de hipóteses sobre a diferença de médias de duas variáveis normais independentes (variâncias desconhecidas) - Teste t'	27
1.4.1	Exemplo 1	29
1.4.2	Exemplo 2	33
1.4.3	Exemplo 3	34
1.5	Teste de hipóteses sobre uma proporção	38
1.5.1	Exemplo 1	42
1.5.2	Exemplo 2	44
1.5.3	Exemplo 3	46
1.6	Teste de hipóteses sobre a diferença de duas proporções	48
1.6.1	Exemplo 1	50

1.7

Teste de hipóteses sobre a razão de variâncias de duas distribuições normais

51

1.7.1

Exemplo 1

53

1.7.2

Exemplo 2

55



Homepage

Página de Rosto

Índice Geral

◀◀ ▶▶

◀ ▶

Página 4 de 55

Voltar

Full Screen

Fechar

Desistir

Chapter 1

Testes paramétricos no contexto de uma e duas amostras

[Homepage](#)[Página de Rosto](#)[Índice Geral](#)[Página 5 de 55](#)[Voltar](#)[Full Screen](#)[Fechar](#)[Desistir](#)

1.1. Teste- t : teste de hipóteses sobre a média de uma distribuição

Problema: testar se a média de uma variável aleatória X contínua que segue uma distribuição normal é diferente, menor ou maior do que um dado valor μ_0 . Para além do desconhecimento da média μ , também se desconhece a variância σ^2 .

- **Dados:** a.a. x_1, \dots, x_n de X
- (i) $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$ ^a
- (ii) $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$
- (iii) $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$
- **Estatística de teste:** sob H_0 ,

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$$

onde

- $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ é a média amostral
- $S = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2}$ é o desvio padrão amostral

A designação de **teste- t** para este teste de hipóteses resulta da distribuição da estatística de teste usada.

^aNa verdade, tem-se $H_0 : \mu = \mu_0$, σ^2 desconhecida, $H_1 : \mu \neq \mu_0$, σ^2 desconhecida; e analogamente nas situações (ii) e (iii).

- **Decisão:** Rejeitar H_0 com nível de significância α se

(i) $t > t_{1-\alpha/2}(n-1)$ ou $t < t_{\alpha/2}(n-1)$
(equivalentemente ^a $T^2 \geq F_{1-\alpha}(1, n-1)$)

(ii) $t \leq t_{\alpha}(n-1)$

(iii) $t \geq t_{1-\alpha}(n-1)$

- Pode-se provar (Efron, 1969) que a quase-simetria da população de base garante que o comportamento do teste não se altera grandemente relativamente ao que se passa com populações normais. Para poder aplicar o teste- t em populações não normais, há portanto que garantir que a população de base não é desvairadamente assimétrica. Isso pode ser conseguido através de uma análise gráfica (histograma e boxplot) dos dados.

Nos casos mais difíceis, pode-se usar um teste de normalidade para avaliar o ajustamento a uma distribuição normal. Aqui há contudo que ter em atenção que **efetuar vários testes sobre os mesmos dados inflaciona o nível de significância de uma forma não trivial!**

- Sabe-se que

* n grande $\implies Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \stackrel{a}{\sim} N(0, 1)$

* n grande $\implies t(n-1) \approx N(0, 1)$

portanto se n for *suficientemente grande* e a avaliação gráfica da distribuição empírica não excluir uma quase-simetria, pode realizar-se um teste- t para a média.

^aRelembre-se que $Y \sim t(\nu) \implies Y^2 \sim F(1, \nu)$.

Instruções em R:

```
> t.test(x,  
  alternative = c("two.sided", "less", "greater"),  
  mu = 0,  
  conf.level = 0.95,  
  ...)
```

onde

- x é a variável (contínua) em causa
- alternative refere-se à formulação da hipótese alternativa: two.sided (por omissão), less ou greater.
- mu=... refere-se à formulação da hipótese nula
- conf.level=... indica o nível de confiança do intervalo de confiança para a média.

Homepage

Página de Rosto

Índice Geral



Página 7 de 55

Voltar

Full Screen

Fechar

Desistir

1.1.1. Relação entre intervalos de confiança para a média e o teste t

Dado que, na secção sobre intervalos de confiança, só vimos intervalos centrados, iremos estabelecer apenas a relação entre intervalos de confiança centrados e testes de hipóteses bilaterais. Existe uma relação análoga para intervalos de confiança unilaterais e testes de hipóteses unilaterais.

Considere-se então um teste de hipóteses bilateral para a média de uma população

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

na situação em que a v.a. X segue uma distribuição normal. Se o valor da estatística de teste na amostra, t , cair na região de rejeição de H_0 , então

$$t < -t_{1-\alpha/2}(n-1) \quad \text{ou} \quad t > t_{1-\alpha/2}(n-1).$$

Suponhamos $t < -t_{1-\alpha/2}(n-1)$. Então, por definição de t ,

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} < -t_{1-\alpha/2}(n-1)$$

o que dá, através de manipulações algébricas simples,

$$\mu_0 > \bar{x} + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1).$$

Se supusermos $t > t_{1-\alpha/2}(n-1)$ obteremos, de forma análoga,

$$\mu_0 < \bar{x} - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1).$$

Dado que o intervalo a $100(1-\alpha)\%$ de confiança para a média de uma população normal é

$$IC_{100(1-\alpha)\%} = \left(\bar{x} - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1), \bar{x} + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) \right).$$

resulta da dedução anterior que:

- a hipótese nula pode ser rejeitada com um nível de significância α (permitindo portanto concluir que $\mu \neq \mu_0$) se e só se o intervalo a $100(1-\alpha)\%$ de confiança para a média da população não contém o valor μ_0 .
- a hipótese nula não pode ser rejeitada com um nível de significância α (não permitindo portanto concluir que $\mu \neq \mu_0$) se e só se o intervalo a $100(1-\alpha)\%$ de confiança para a média da população contém o valor μ_0 .

Observações:

1. Dependência do tamanho da amostra

A estatística de teste anterior é

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \sqrt{n} \frac{\bar{X} - \mu_0}{S} \sim t(n-1).$$

Observe-se que quanto maior for o tamanho amostral, n , maior será o valor da estatística de teste, mantendo fixos os valores dos restantes parâmetros.

Em particular, se quisermos efectuar um teste de hipóteses para a média, $H_0 : \mu = \mu_0$, em duas amostras com os mesmos valores de \bar{x} e s , a amostra maior corresponderá a um maior valor de t e portanto teremos aí uma maior probabilidade de rejeitar H_0 do que na amostra mais pequena.

Resulta daqui que estudos com uma diferença $\bar{x} - \mu_0$ relativamente pequena podem de facto apresentar diferenças significativas apenas porque o tamanho da amostra é grande. Nessas situações, deve ser prestada uma atenção especial à interpretação dos resultados e será essencial considerar o intervalo de confiança correspondente para perceber até que ponto é que a diferença entre a média amostral e o valor de referência é significativa.

2. Porque não usar apenas intervalos de confiança?

Depois de ilustradas as vantagens dos intervalos de confiança relativamente aos testes de hipóteses, esta é uma questão que pode naturalmente surgir. A resposta passa pelo valor- p fornecido pelos testes de hipóteses. Esse valor não pode ser obtido de um intervalo de confiança, cujo nível de confiança tem de ser estabelecido à priori pelo investigador.

3. Generalização

Um nível de significância de α para um teste de hipóteses retribui um intervalo de confiança correspondente a $(1 - \alpha)\%$ de confiança. O resultado do teste de hipóteses é significativo sempre que o intervalo de confiança não contém o valor especificado na hipótese nula (muitas vezes o zero, mas não sempre!).

Há contudo situações em que não é possível determinar os desvios padrão das estatísticas de teste e os intervalos de confiança são contruídos a partir das variâncias assintóticas. Nesta situação, as propriedades estatísticas dos intervalos de confiança assim contruídos são menos interessantes do que o resultado ("isolado") do teste de hipóteses.

3. Recomendação

A propósito de testes de hipóteses e intervalos de confiança, podemos sempre seguir a recomendação de Knapp (1999):

use a hypothesis test if you have a hypothesis to test; otherwise, use interval estimation.

1.1.2. Exemplo 1

Realizou-se uma amostra aleatória de 30 suspensões celulares. A tabela ao lado contém os valores do consumo de oxigénio (em ml) observado para cada uma das células durante o período de incubação. Estudos anteriores referem que a quantidade média de oxigénio consumido numa população de células sem anomalias no período de tempo considerado é de 12 ml.

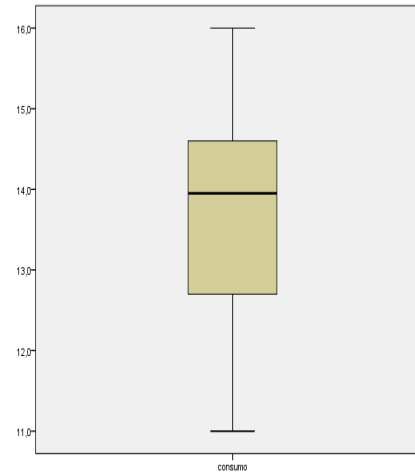
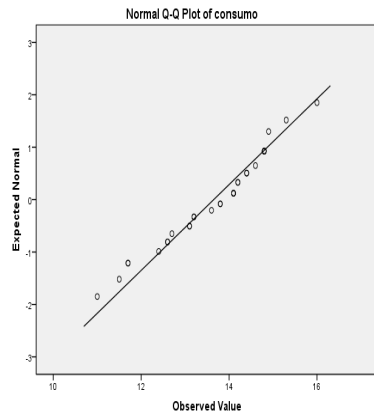
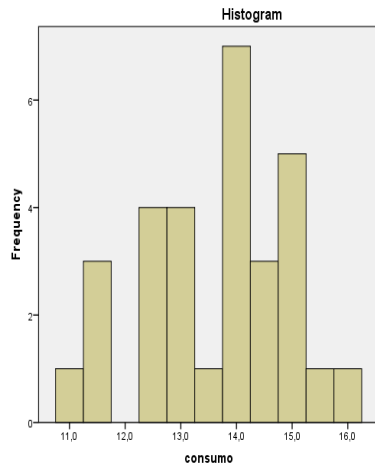
Problema: Para um nível de significância de .05, os dados apresentados fornecem evidência estatística suficiente para concluir que o consumo médio de oxigénio na população de células de onde a amostra foi retirada é diferente de 12 ml? Ou será que as diferenças encontradas entre o consumo médio amostral e os 12 ml referidos em H_0 são devidas ao acaso da amostragem?

- Dados: a.a. de pop. aprox. normal (ver gráficos seguintes)
- $H_0 : \mu = 12$
 $H_1 : \mu \neq 12$
- Estatística de teste:
sob H_0 , $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$
- Decisão: Rejeitar H_0 com nível de significância α se $t > t_{1-\alpha/2}(n-1)$ ou $t < -t_{1-\alpha/2}(n-1)$.

Célula	O ₂ (ml)	Célula	O ₂ (ml)
1	14.2	16	13.8
2	14.2	17	14.4
3	11.5	18	16.0
4	14.9	19	15.3
5	13.8	20	14.8
6	13.2	21	14.4
7	12.6	22	14.1
8	13.1	23	11.7
9	14.8	24	14.1
10	13.1	25	13.6
11	14.1	26	14.8
12	11.7	27	12.4
13	12.6	28	11.0
14	13.2	29	14.6
15	12.7	30	14.8
Total	$n = 30$	Gr. lib. (ν)	$\nu = 29$
Média (\bar{x})	13.65	$t = \frac{\bar{x} - 12}{s/\sqrt{30}}$	$t = 7.395$
D. padrão (s)	1.222	$t_{0.975}(29)$	2.045
$p = 2P(t \geq 7.395) < 0.001$			

Resposta: Com um nível de significância de 0.05, pode-se rejeitar H_0 e concluir que o consumo médio de oxigénio na população de onde a amostra foi extraída é diferente de 12.

Em particular, erraremos menos de (aprox.) 5 em 100 vezes ao afirmar que o consumo médio de oxigénio na população de células é diferente de 12 ml.



Uma análise gráfica dos dados não parece contrariar a hipótese de a amostra provir de uma população que segue uma distribuição normal.

```
> consumo <- c(14.2, 13.8, 14.2, 14.4, 11.5, 16.0, 14.9, 15.3, 13.8, 14.8,
13.2, 14.4, 12.6, 14.1, 13.1, 11.7, 14.8, 14.1, 13.1, 13.6, 14.1, 14.8,
11.7, 12.4, 12.6, 11.0, 13.2, 14.6, 12.7, 14.8)
> t.test(consumo, mu=12)
```

One Sample t-test

```
data: consumo
t = 7.3948, df = 29, p-value = 3.791e-08
alternative hypothesis: true mean is not equal to 12
95 percent confidence interval:
 13.19365 14.10635
sample estimates:
mean of x
 13.65
```

O output indica um valor para a estatística de teste de $t = 7.395$, $T \sim t(29)$, correspondente a um valor de prova de $p = 0.000$ (para 3 casas decimais). Para um nível de significância de 0.05, isto permite rejeitar a hipótese nula e afirmar que a média (populacional) do consumo de oxigénio é diferente de 12 ml. O erro mínimo associado a esta afirmação é de 3.8×10^{-8} .

Nota: O teste t permite efectuar inferências sobre a média da população, e não sobre a média da amostra. De facto, a média da amostra fica totalmente definida a partir da própria amostra (no exemplo anterior vale 13.43 ml), e não precisaria de nenhum teste para averiguar se 13.43 é ou não diferente de 12. Portanto é considerado um erro muito grave de compreensão fazer

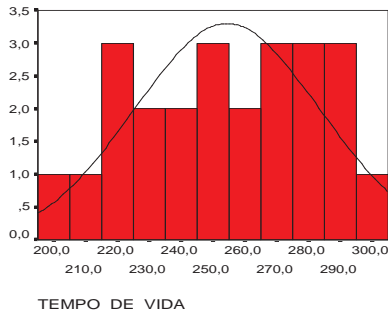
$$H_0 : \bar{x} = 12.$$

1.1.3. Exemplo 2

Sabe-se que o tempo de vida de uma determinada bactéria é igual ou superior a 250 horas. Um laboratório interessado em desenvolver um antibiótico para essa bactéria gostaria de afirmar que o tempo médio de vida da bactéria é superior a 250 horas. Para isso, os seus técnicos mediram os tempos de vida (em horas) de 24 das bactérias em causa. Os resultados foram os apresentados na tabela ao lado.

Problema: Assumindo que a amostra utilizada pelo laboratório é aleatória, pode o laboratório concluir o que pretende, para um nível de significância de 0.05?

- Dados: a.a. de pop. aprox. normal
- $H_0 : \mu = 250$
 $H_1 : \mu > 250$
- Estatística de teste:
Sob H_0 , $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$
- Decisão: Rejeitar H_0 com nível de significância α se $t > t_{1-\alpha}(n-1)$



Bact.	T. vida	Bact.	T. vida
1	271	13	198
2	219	14	225
3	253	15	262
4	224	16	291
5	264	17	211
6	268	18	243
7	230	19	275
8	284	20	282
9	216	21	288
10	253	22	236
11	295	23	252
12	272	24	294
Total	Média	D. padrão	
$n = 24$	$\bar{x} = 254.42$	$s = 28.984$	
Gr. lib. (ν)	$t = \frac{\bar{x} - 250}{s/\sqrt{24}}$	$t_{0.9}(23)$	$t_{0.95}(23)$
$\nu = 23$	$t = 0.747$	1.319	1.7139
$p = P(t \geq 0.747) = 0.231$			

O histograma e o boxplot (página seguinte) não evidenciam nenhuma assimetria marcada na distribuição empírica.

Resposta: Para um nível de significância de 0.05, os dados não apresentam evidência estatística suficiente para afirmar que o tempo médio de vida das bactérias (na população) é superior a 250 horas.

```

> length(t.vida)
[1] 24

> mean(t.vida)
[1] 254.4167

> sd(t.vida)
[1] 28.98413

> est.teste <- (mean(t.vida)-250)/(sd(t.vida)/sqrt(24))
> est.teste
[1] 0.7465175

> t.test(t.vida,
          mu=250,
          alternative="greater")

One Sample t-test

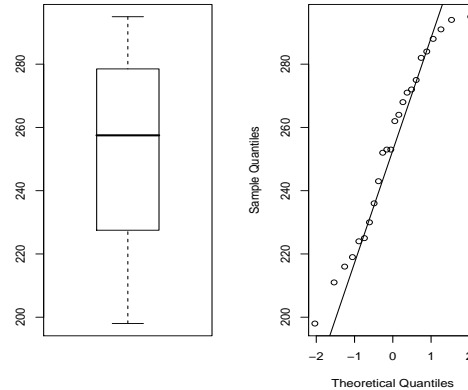
data:  t.vida
t = 0.7465, df = 23, p-value = 0.2315
alternative hypothesis:
      true mean is greater than 250
95 percent confidence interval:
 244.2768      Inf
sample estimates:
mean of x
 254.4167

```

Observe-se que o intervalo de confiança para a média a 95% é também ele unilateral:

$$(244.3, +\infty).$$

A decisão de não se poder rejeitar H_0 reflecte-se também no facto de o intervalo de confiança conter o valor 250.



Nota sobre a escolha da região de rejeição:
A distribuição da estatística de teste é válida apenas quando assumimos H_0 , isto é

$$T = \frac{\bar{X} - 250}{S/\sqrt{n}} \sim t(n-1).$$

Temos agora de pensar no que aconteceria aos valores desta estatística T na situação em que a média populacional fosse superior a 250. Assumindo a mesma dispersão, tenderíamos a obter valores para \bar{X} superiores a 250, e portanto a valores *grandes* para T . Sabendo que a região de rejeição tem área igual a α , isso significa então que a cauda direita da distribuição $t(n-1)$ corresponderia aos valores amostrais mais elevados sob a hipótese $\mu > 250$. Sendo uma região "boa" para H_1 , essa deve ser portanto a região de rejeição de H_0 .

1.1.4. Exemplo 3

Uma amostra de 242 mulheres americanas com idade superior a 30 anos apresenta um valor médio de consumo energético diário de 6992.6 kJ e um desvio padrão de 1176.83 kJ.

Problema: Pode-se concluir, com um nível de significância de 0.01, que o consumo energético diário médio nas mulheres americanas é diferente do valor recomendado de 7725 kJ?

- Dados: a.a. de tamanho grande e razoavelmente simétrica

O histograma sugere que muito provavelmente a amostra não provém de uma distribuição normal. Contudo, dado que o tamanho amostral já é razoável e o histograma e o box-plot sugerem uma distribuição razoavelmente simétrica, prosseguimos com o teste t .

- $H_0 : \mu = 7725$

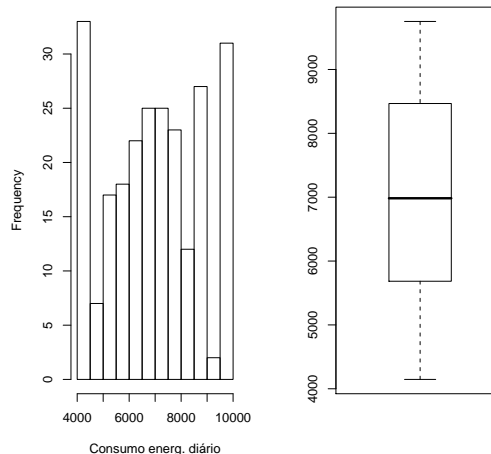
$$H_1 : \mu \neq 7725$$

- Estatística de teste:

$$\text{Sob } H_0, \quad T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \stackrel{a}{\sim} N(0, 1) \approx t(n-1)$$

- Decisão: Rejeitar H_0 com nível de significância α se $t > t_{1-\alpha/2}(n-1)$ ou $t < -t_{1-\alpha/2}(n-1)$.

Total	$n = 242$
Média (\bar{x})	6 992.6
D. padrão (s)	1 176.83
$t = \frac{6992.6 - 7725}{1176.83/\sqrt{242}}$	$t = -9.681$
$t_{0.995}(241)$	3.596



X : v.a. que representa o consumo energético diário na população de mulheres americanas com idade superior a 30 anos

$$t = \frac{6992.6 - 7725}{1176.83/\sqrt{242}} \approx -9.681$$

$$T \sim t(241)$$

$$p = 2P(T \leq -9.681) < 0.001$$

O intervalo (centrado) a 99% confiança é (6796.2, 7189.1) kJ.

Resposta: Com um nível de significância de .01, pode-se concluir que o consumo energético diário médio das mulheres americanas com idade superior a 30 anos é diferente do valor recomendado de 7725 kJ. O intervalo de confiança sugere que seja inferior a 7725 kJ.

```
> t.test(y, mu=7725, conf.level=0.99)
```

One Sample t-test

```
data: y  
t = -9.681, df = 241, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 7725  
99 percent confidence interval:  
 6796.228 7189.057  
sample estimates:  
mean of x  
 6992.643
```

[Homepage](#)[Página de Rosto](#)[Índice Geral](#)

Página 16 de 55

[Voltar](#)[Full Screen](#)[Fechar](#)[Desistir](#)

1.2. Amostras emparelhadas

Situação 1: Numa investigação pretende-se comparar o efeito de um novo tratamento, digamos A, com um outro já existente, digamos B, sobre uma determinada população de doentes. Se o tratamento A é testado numa qualquer amostra aleatória de doentes e o tratamento B numa outra qualquer amostra aleatória, eventuais diferenças entre as respostas de cada um dos grupos podem ser devidas a outros factores alheios ao fenómeno em estudo e serem erradamente consideradas efeito desse fenómeno. Suponhamos, por exemplo, que os tratamentos A e B em causa são muito dependentes das dietas alimentares que os acompanham. Se as amostras provierem de populações com dietas diferentes, esse factor "dieta" pode estar a funcionar como *factor de confundimento*. O mesmo se poderia passar com a idade, sexo, grupo sanguíneo, ...

Situação 2: Quer-se testar o efeito de um novo tratamento sobre os indivíduos doentes. Se se escolherem aleatoriamente dois grupos de indivíduos doentes, um deles funcionar como controlo e se submeter o outro ao novo tratamento, eventuais diferenças encontradas nas respostas podem ser devidas a um efeito de confundimento semelhante ao anterior (por exemplo, se o grupo sujeito ao tratamento consistir de indivíduos mais novos, as diferenças encontradas entre os grupos podem ser devidas à idade e não ao tratamento em si).

Uma solução para o tipo de problemas levantados consiste em escolher aleatoriamente duas amostras, ditas **emparelhadas**, que consistam de indivíduos o mais "semelhantes" possível excepto em relação ao factor que está a ser testado (nos exemplos anteriores: o tratamento). Isto faz com que eventuais diferenças encontradas nas respostas sejam imputáveis à diferença de tratamentos e não a fenómenos alheios ao factor em estudo (**factores de confundimento**).

Estatisticamente, duas amostras **emparelhadas** correspondem a n observações do par de v.a. (X, Y) nas seguintes condições:

- os pares de observações $(X_1, Y_1), \dots, (X_n, Y_n)$ são independentes
- as observações de cada par são realizadas sob condições idênticas.

Um conselho geral em planeamento experimental é: emparelhe o que puder, aleatorize a atribuição dos tratamentos às unidades experimentais quando não puder emparelhar. O emparelhamento mais simples consiste em cada unidade experimental funcionar também como unidade de controlo. ^a

^aDinis Pestana, Velosa, *Introdução à Probabilidade e à Estatística*, Fundação Calouste Gulbenkian, 2008.

1.2.1. Exemplo 1

Um grupo de investigadores conduziu um estudo para a determinação da perda de peso, composição do corpo, distribuição da gordura pelo corpo, e taxa de metabolismo basal, em indivíduos obesos antes e depois de 12 semanas de tratamento com uma dieta muito pouco calórica. A experiência consistiu de 37 indivíduos obesos, 17 homens e 20 mulheres. Os pesos (em Kg) obtidos para as mulheres, antes e depois do tratamento com a dieta pouco calórica, foram os apresentados na tabela ao lado.

Problema: Os resultados obtidos apresentam evidência suficiente para concluir, com um nível de significância de 0.005, que o tratamento para a perda de peso em mulheres obesas é eficiente?

- Dados: a.a. de tamanho n de um par de variáveis aleatórias emparelhadas (X, Y) tal que $D = X - Y$ segue uma distribuição normal
- $H_0 : \mu_X = \mu_Y \iff \mu_X - \mu_Y = 0 \iff \mu_D = 0$
 $H_1 : \mu_X \neq \mu_Y$
- Estatística de teste: Sob H_0 ,
 $T = \frac{\bar{D} - \mu_0}{s_D / \sqrt{n}} \sim t(n-1)$,
 $S_D^2 = \frac{1}{n-1} \sum_{k=1}^n ((X_k - Y_k) - (\bar{X} - \bar{Y}))^2$
- Decisão: Rejeitar H_0 com nível de significância α se $t > t_{1-\alpha/2}(n-1)$ ou $t < -t_{1-\alpha/2}(n-1)$.

Obs.	Antes (X)	Depois (Y)	Diferença (D)
1	91.2	79.0	12.2
2	103.3	86.7	16.6
3	97.9	88.2	9.7
4	99.3	86.4	12.9
5	105.2	95.8	9.4
6	100.1	80.4	19.7
7	94.7	81.2	13.5
8	101.7	81.4	20.3
9	78.5	70.7	7.8
10	91.5	83.7	7.8
11	100.0	78.1	21.9
12	97.9	74.3	23.6
13	87.8	73.5	14.3
14	89.7	69.2	20.5
15	92.6	74.3	18.3
16	101.1	74.3	26.8
17	80.4	68.3	12.1
18	96.7	90.4	6.3
19	106.9	88.2	18.7
20	115.8	98.4	17.4
Totais	$n_X = 20$ $\bar{x} = 96.6$ $s_X = 8.78$	$n_Y = 20$ $\bar{y} = 81.1$ $s_Y = 8.56$	$n = 20$ $\bar{d} = 15.5$ $s_D = 5.77$
$t = \frac{15.5-0}{5.77/\sqrt{20}}$	$t = 12.015$ $t_{.95}(8)$ 1.729	$\nu = 19$ $t_{.995}(8)$ 2.861	$t_{.9975}(8)$ 3.174

Suponhamos que os dados do problema tinham sido inseridos no software SPSS, no ficheiro TesteAmEmparelhadas.Ex1.sav. Para efectuarmos o teste de hipóteses, temos primeiro de importar o ficheiro para o R e só depois proceder com as instruções adequadas.

```
> library(foreign)

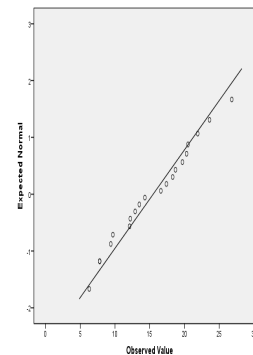
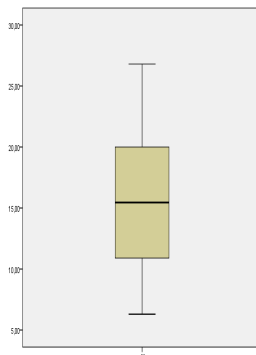
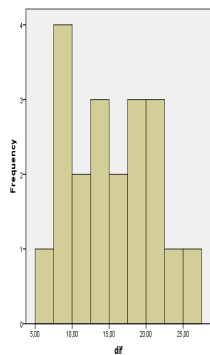
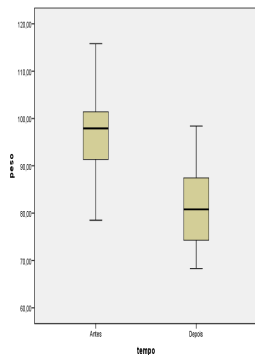
> base <- read.spss("TesteAmEmparelhadas_Ex1.sav",
  to.data.frame=TRUE,
  use.value.labels=FALSE)

> base
  indiv antes depois
1      1  91.2   79.0
2      2 103.3   86.7
3      3  97.9   88.2
4      4  99.3   86.4
5      5 105.2   95.8
6      6 100.1   80.4
7      7  94.7   81.2
8      8 101.7   81.4
9      9  78.5   70.7
10     10  91.5   83.7
11     11 100.0   78.1
12     12  97.9   74.3
13     13  87.8   73.5
14     14  89.7   69.2
15     15  92.6   74.3
16     16 101.1   74.3
17     17  80.4   68.3
18     18  96.7   90.4
19     19 106.9   88.2
20     20 115.8   98.4
```

```
> base$dif <- base$antes - base$depois

> names(base)
[1] "indiv" "antes" "depois" "dif"

> boxplot(base$antes, base$depois,
  names=c("antes", "depois"))
> hist(dif, main="")
> boxplot(dif, ylab="dif")
> qqnorm(dif, main="")
> qqline(dif, main="")
```



```
> t.test(base$dif, mu=0, conf.level=0.995)
```

One Sample t-test

```
data: base$dif
t = 12.0154, df = 19, p-value = 2.536e-10
alternative hypothesis: true mean is not
                        equal to 0
99.5 percent confidence interval:
 11.39849 19.58151
sample estimates:
mean of x
 15.49
```

Outra instrução absolutamente equivalente seria

```
> t.test(base$antes, base$depois, paired=TRUE,
mu=0, conf.level=0.995)
```

Paired t-test

```
data: base$antes and base$depois
t = 12.0154, df = 19, p-value = 2.536e-10
alternative hypothesis: true difference in means
                        is not equal to 0
99.5 percent confidence interval:
 11.39849 19.58151
sample estimates:
mean of the differences
 15.49
```

Aqui, é essencial escrever "paired=TRUE", caso contrário o teste realizado é o teste t para duas amostras independentes, que é totalmente diferente do teste t para amostras emparelhadas.

Resposta: Com um nível de significância de 0.005, pode-se rejeitar H_0 e concluir que o tratamento para a perda de peso é eficiente. O intervalo de confiança a 99.5% para a perda média de peso é (11.40, 19.58).

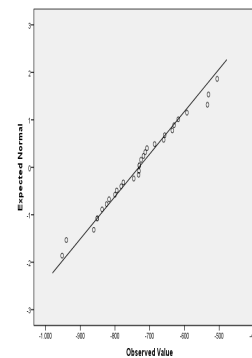
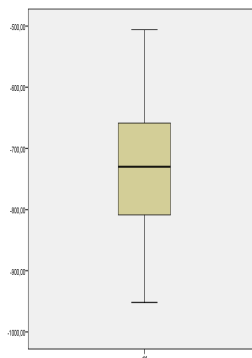
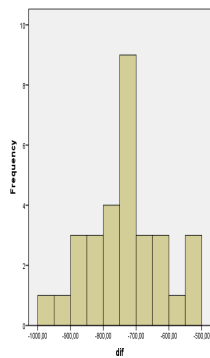
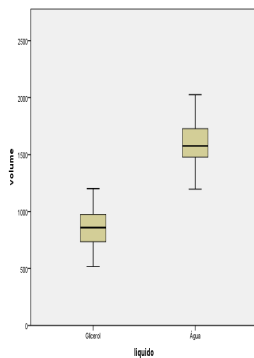
1.2.2. Exemplo 2

Foi efectuado um estudo para analisar o efeito da hiperhidratação enriquecida com glicerol na resistência ao esforço físico prolongado. A experiência consistiu de 31 indivíduos ciclistas, com idades compreendidas entre os 22 e os 40 anos, que pedalavam pelo menos 120Km por semana. As entradas da tabela ao lado representam o volume (em ml) de urina obtida dos indivíduos depois do exercício: após a ingestão de glicerol e após a ingestão de água.

Problema: Com base neste estudo, é possível concluir, com um nível de significância de 0.05, que o glicerol altera o volume de urina obtido?

- Dados: a.a. de tamanho n de um par de variáveis aleatórias emparelhadas (X, Y) tal que $D = X - Y$ segue uma distribuição normal
- $H_0 : \mu_X = \mu_Y \iff \mu_D = 0$
 $H_1 : \mu_X \neq \mu_Y$
- Estatística de teste:
Sob H_0 , $T = \frac{\bar{D} - \mu_0}{s_D / \sqrt{n}} \sim t(n-1)$
- Decisão: Rejeitar H_0 com nível de significância α se $t > t_{1-\alpha/2}(n-1)$ ou $t < -t_{1-\alpha/2}(n-1)$.

Obs.	Glicerol (X)	Água (Y)	D
1	787	1739	-952
2	949	1674	-725
3	732	1325	-593
4	865	1716	-851
5	859	1654	-795
6	686	1523	-837
7	935	1553	-618
8	849	1355	-506
9	798	1580	-782
10	889	1620	-731
11	1132	1878	-746
12	586	1316	-730
13	1192	1852	-660
14	517	1457	-940
15	666	1197	-531
16	784	1502	-718
17	700	1500	-800
18	986	1643	-657
19	1043	1860	-817
20	907	1537	-630
21	1003	1864	-861
22	646	1359	-713
23	725	1576	-851
24	1202	2026	-824
25	886	1521	-635
26	769	1303	-534
27	1065	1797	-732
28	962	1692	-730
29	739	1515	-776
30	764	1450	-686
31	1034	1742	-708
Totais	$n_X = 31$ $\bar{y} = 1591.2$	$n_Y = 31$ $\bar{d} = -731.3$	$n = 31$ $s_D = 111.58$
$t = \frac{-731.3 - 0}{111.58 / \sqrt{31}}$	$t = -36.491$ $t_{.95}(30)$ 1.697	$\nu = 30$ $t_{.975}(30)$ 2.042	



Os gráficos sugerem a não rejeição da hipótese de que a variável aleatória diferença segue uma distribuição normal. Avança-se então para a realização de um teste-t de comparação de médias em amostras emparelhadas.

```
> base <- read.spss("TesteMediasEmp2.sav", to.data.frame=TRUE, use.value.labels=FALSE)
> names(base)
[1] "Indiv"      "Glicerol"   "Agua"       "VAR00006"  "VAR00005"  "dif"        "volume"     "liquido"

> t.test(base$Glicerol, base$Agua, paired=TRUE)

      Paired t-test

data:  base$Glicerol and base$Agua
t = -36.4906, df = 30, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -772.1845 -690.3317
sample estimates:
mean of the differences
      -731.258
```

Resposta: O teste é significativo: com um nível de significância de 0.05 pode-se rejeitar H_0 e concluir que o glicerol altera o volume médio de urina obtido. O intervalo de confiança a 95% para a diferença entre os volumes médios de urina após ingestão de glicerol e após ingestão de água é $[-772.2, -690.3]$, sugerindo que o glicerol diminui o volume médio de urina obtido.

Outra instrução possível de usar, muito comum no **R**, seria a seguinte

```
> t.test(base$volume ~ base$liquido, paired=TRUE)

Paired t-test

data:  base$volume by base$liquido
t = -36.4906, df = 30, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -772.1845 -690.3317
sample estimates:
mean of the differences
      -731.258
```

Notar a forma como os dados têm de estar introduzidos para usar esta instrução (página seguinte): a variável contínua em causa está toda numa coluna, e o factor que caracteriza cada uma das amostras (portanto uma variável binária) está noutra coluna.

[Homepage](#)[Página de Rosto](#)[Índice Geral](#)[<<](#)[>>](#)[<](#)[>](#)[Página 23 de 55](#)[Voltar](#)[Full Screen](#)[Fechar](#)[Desistir](#)

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help												
	Indiv	Glicerol	Agua	VAR00006	VAR00005	dif	volume	liquido	var	var	var	
21	21	1003	1864	.	.	-861,00	1003	1				
22	22	646	1359	.	.	-713,00	646	1				
23	23	725	1576	.	.	-851,00	725	1				
24	24	1202	2026	.	.	-824,00	1202	1				
25	25	886	1521	.	.	-635,00	886	1				
26	26	769	1303	.	.	-534,00	769	1				
27	27	1065	1797	.	.	-732,00	1065	1				
28	28	962	1692	.	.	-730,00	962	1				
29	29	739	1515	.	.	-776,00	739	1				
30	30	764	1450	.	.	-686,00	764	1				
31	31	1034	1742	.	.	-708,00	1034	1				
32	1739	2				
33	1674	2				
34	1325	2				
35	1716	2				
36	1654	2				
37	1523	2				
38	1553	2				
39	1355	2				
40	1580	2				
41	1620	2				
42	1878	2				
43	1316	2				
1												
Data View Variable View												

[Homepage](#)
[Página de Rosto](#)
[Índice Geral](#)

[Página 24 de 55](#)
[Voltar](#)
[Full Screen](#)
[Fechar](#)
[Desistir](#)

1.3. Teste hip. sobre a diferença de médias de duas variáveis normais independentes (variâncias iguais e desconhecidas) - Teste t

Problema: testar se a diferença de médias $\mu_Y - \mu_X$ de duas populações normais e independentes, é diferente, menor ou maior do que um dado valor μ_0 . Para além do desconhecimento das médias também se desconhecem as variâncias, σ_X^2 e σ_Y^2 , mas pode ser assumido que $\sigma_X^2 = \sigma_Y^2$ ($= \sigma^2$).

O teste assume portanto três hipóteses: **independência, normalidade e homoscedasticidade**. O teste não é válido para populações que não sejam independentes nem para populações que se saiba terem variâncias diferentes. **Para amostras grandes e com distribuições moderadamente assimétricas, o teste é robusto a desvios da normalidade.**

- **Dados:** amostras aleatórias x_1, \dots, x_{n_X} e y_1, \dots, y_{n_Y} , de tamanhos n_X e n_Y , respectivamente, de cada uma das populações.
- (i) $H_0 : \mu_Y - \mu_X = \mu_0$, $H_1 : \mu_Y - \mu_X \neq \mu_0$
(ii) $H_0 : \mu_Y - \mu_X = \mu_0$, $H_1 : \mu_Y - \mu_X > \mu_0$
(iii) $H_0 : \mu_Y - \mu_X = \mu_0$, $H_1 : \mu_Y - \mu_X < \mu_0$
- **Estatística de teste:** sob H_0 ,

$$T = \frac{\bar{Y} - \bar{X} - \mu_0}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \sim t(n-2)$$

onde

- $n = n_X + n_Y$
- $S_p^2 = \frac{(n_X-1)S_X^2 + (n_Y-1)S_Y^2}{n_X + n_Y - 2}$ é a variância amostral combinada das duas amostras (*pooled sample variance*).

Notar que

$$\text{Var}(\bar{X} - \bar{Y}) = \frac{\sigma^2}{n_X} + \frac{\sigma^2}{n_Y}$$

e que S_p^2 é um estimador de σ^2 .

- Decisão: Rejeitar H_0 com **nível de significância α** se
 - (i) $t \geq t_{1-\alpha/2}(n-2)$ ou $t \leq t_{\alpha/2}(n-2)$
(equivalentemente, $T^2 \geq F_{1-\alpha}(1, n-2)$)
 - (ii) $t \geq t_{1-\alpha}(n-2)$
 - (iii) $t \leq t_{\alpha}(n-2)$

- Instruções em R:

```
# Instrução 1
t.test(x, y = NULL,
       alternative = c("two.sided",
                      "less", "greater"),
       mu = 0,
       paired = FALSE,
       var.equal = TRUE,
       conf.level = 0.95, ...)

OU

## Instrução 2
t.test(formula, data, subset, na.action,
       ...)
```

- Na Instrução 1, x e y representam as observações de cada uma das amostras
- Na Instrução 2, a fórmula a usar deverá ser do tipo

$$\text{resposta} \sim \text{fator}$$

onde *resposta* é o vector (contínuo) consistindo de ambas as amostras e *fator* é a variável binária que identifica a amostra.

As variáveis *resposta* e *fator* correspondem a duas colunas distintas no data frame.

Observações:

- Para populações normais e independentes, as hipóteses nula e alternativa do teste t podem ser pensadas da seguinte forma:

$H_0 : \mu_X = \mu_Y; \sigma_X^2$ e σ_Y^2 desconhecidas, mas pode-se assumir que $\sigma_X^2 = \sigma_Y^2$

$H_1 : \mu_X \neq \mu_Y; \sigma_X^2$ e σ_Y^2 desconhecidas, mas pode-se assumir que $\sigma_X^2 = \sigma_Y^2$

Há autores ^a que afirmam que, na prática, é muito difícil encontrar duas populações (normais e independentes) de indivíduos que reajam de forma diferente, simultaneamente em relação à localização e escala, devido ao efeito de um tratamento.

- A hipótese de homogeneidade de variâncias poderia conduzir à realização de um teste F de igualdade de variâncias anterior à realização do teste t de igualdade de médias. Tal procedimento é errado de acordo com a formulação do teste t dada acima, mas também porque conduz a uma inflação do erro de tipo I cometido.

^aSawilowsky S., "Fermat, Schubert, Einstein and Beherns-Fisher: The Probable Difference Between Two Means When $\sigma_1^2 \neq \sigma_2^2$ ", *Journal of Modern Applied Statistical Methods*, 2002, Vol.1, No.2, 461-472.

1.4. Teste de hipóteses sobre a diferença de médias de duas variáveis normais independentes (variâncias desconhecidas) - Teste t'

Problema: testar se a diferença de médias $\mu_Y - \mu_X$ de duas populações normais e independentes, é diferente, menor ou maior do que, um dado valor μ_0 ^a. Para além do desconhecimento das médias também se desconhecem as variâncias, σ_X^2 e σ_Y^2 .

O teste não é válido para populações que não sejam independentes.

Para amostras *grandes* e com distribuições *moderadamente* assimétricas, o teste é robusto a desvios da normalidade - nessa situação, a estatística de teste segue aproximadamente a distribuição normal reduzida.

- **Dados:** a.a. x_1, \dots, x_{n_X} e y_1, \dots, y_{n_Y} , de tamanhos n_X e n_Y , respectivamente, de cada uma das populações
- (i) $H_0 : \mu_Y - \mu_X = \mu_0$, $H_1 : \mu_Y - \mu_X \neq \mu_0$
- (ii) $H_0 : \mu_Y - \mu_X = \mu_0$, $H_1 : \mu_Y - \mu_X > \mu_0$
- (iii) $H_0 : \mu_Y - \mu_X = \mu_0$, $H_1 : \mu_Y - \mu_X < \mu_0$
- **Estatística de teste:** sob H_0 ,

$$T = \frac{(\bar{X} - \bar{Y}) - \mu_0}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} \simeq t(\nu^*)^b$$

^aO teste em causa é usualmente designado por **Teste de Welch-Satterwaite**; podem também se usadas as designações **Teste de Satterwaite-Fisher**, **Teste de Welch**, ou **Teste de Welch-Aspin**.

^bo problema da determinação da distribuição exacta desta estatística de teste é uma questão em aberto, conhecida como **problema de Behrens-Fisher**

onde o grau de liberdade ν^* é aproximadamente

$$\nu^* = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{1}{n_X-1} \left(\frac{s_X^2}{n_X}\right)^2 + \frac{1}{n_Y-1} \left(\frac{s_Y^2}{n_Y}\right)^2}.$$

- **Decisão:** represente-se por ν o maior inteiro que não excede ν^* .

Rejeitar H_0 com **nível de significância α** se

- (i) $t \geq t_{1-\alpha/2}(\nu)$ ou $t \leq t_{\alpha/2}(\nu)$
(equivalentemente, $T^2 \geq F_{1-\alpha}(1, \nu)$)
- (ii) $t \geq t_{1-\alpha}(\nu)$
- (iii) $t \leq t_{\alpha}(\nu)$

- **Instruções em R:**

```
## Instrução 1
t.test(x, y = NULL,
       alternative = c("two.sided",
                      "less", "greater"),
       mu = 0, paired = FALSE, conf.level = 0.95,
       ...)

ou

## Instrução 2
t.test(formula, data, subset, na.action,
       ...)
```

(por omissão, a instrução `t.test` dá lugar ao teste de Welch-Satterwaite).

Quanto às instruções anteriores:

- na Instrução 1, x e y representam as observações de cada uma das amostras
- na Instrução 2, a fórmula a usar deverá ser do tipo

$$resposta \sim fator$$

onde *resposta* é o vector (contínuo) consistindo de ambas as amostras e *fator* é a variável binária que identifica a amostra.

As variáveis *resposta* e *fator* correspondem a duas colunas distintas no data frame.

Observação: Assumindo que cada uma das amostras provém de uma distribuição normal, tem-se

$$\begin{aligned}\bar{X} &\sim N\left(\mu_X, \frac{\sigma_X^2}{n_X}\right) \\ \bar{Y} &\sim N\left(\mu_Y, \frac{\sigma_Y^2}{n_Y}\right).\end{aligned}$$

Dado que

$$\begin{aligned}E(\bar{X} - \bar{Y}) &= E(\bar{X}) - E(\bar{Y}) = \mu_X - \mu_Y \\ \text{Var}(\bar{X} - \bar{Y}) &= \text{Var}(\bar{X}) + \text{Var}(\bar{Y}) = \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\end{aligned}$$

(a soma na fórmula das variâncias resulta das amostras serem independentes), vem

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right).$$

O que acontece agora na maior parte das situações é que as variâncias σ_X^2 e σ_Y^2 das populações são desconhecidas.

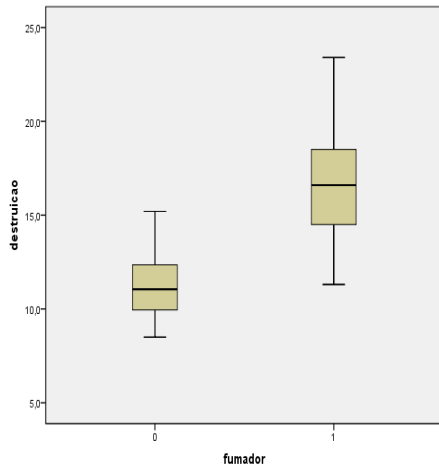
- Quando as amostras são *pequenas*, tem de ser usar a aproximação de Welch-Satterwaite, tal como descrito na estatística de teste anterior.
- Quando as amostras são *grandes*, podem substituir-se as variâncias populacionais pelas variâncias amostrais obtendo uma distribuição normal assintótica:

$$\bar{X} - \bar{Y} \simeq N\left(\mu_X - \mu_Y, \frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right).$$

1.4.1. Exemplo 1

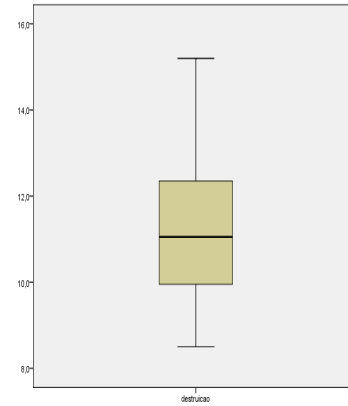
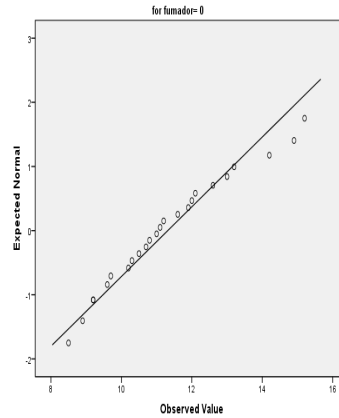
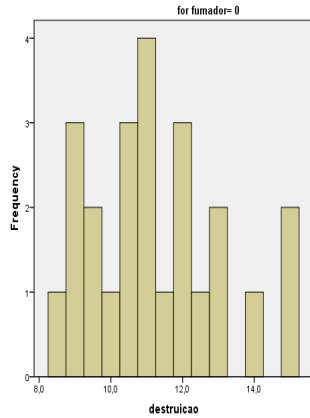
Um grupo de investigadores identificou três locais nos pulmões que permitem avaliar o grau de destruição destes órgãos. Para determinar se o tabaco é prejudicial aos pulmões, realizaram análises aos três locais de interesse nos pulmões em 24 indivíduos não-fumadores e em 29 indivíduos fumadores (os dados foram conseguidos entre os óbitos de um hospital, falecidos por doença não relacionada com problemas respiratórios). Os resultados para um desses locais nos pulmões são apresentados na tabela seguinte.

Problema: Para um nível de significância de 0.01, as diferenças observadas entre os dois grupos são devidas ao acaso da amostragem ou o facto de um indivíduo ser ou não fumador afecta de facto o grau de destruição dos tecidos?

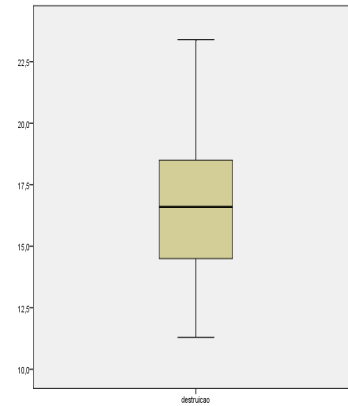
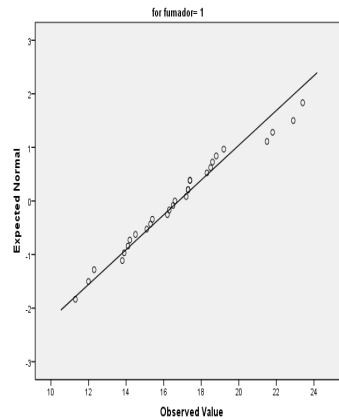
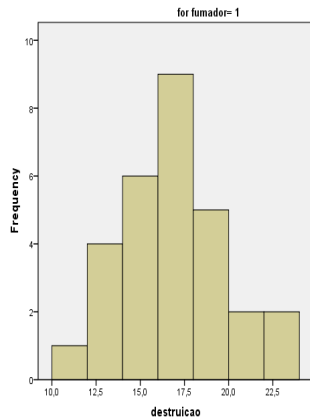


Grau de destruição para não-fumadores	Grau de destruição para fumadores
12.1	16.6
12.0	13.9
10.8	11.3
11.0	14.5
9.7	17.4
14.9	15.3
8.5	14.2
13.0	12.3
12.6	18.6
14.2	12.0
11.6	15.1
13.2	16.5
15.2	21.8
11.9	16.3
8.9	23.4
9.2	18.8
10.2	17.2
10.5	18.5
10.7	16.2
11.2	13.8
9.6	17.3
9.2	14.1
10.3	17.4
11.1	21.5
	22.9
	19.2
	17.3
	15.4
	18.3
$n_X = 24$	$n_Y = 29$

Nível de destruição em não fumadores



Nível de destruição em fumadores



- Dados: a.a. de duas pop. indep. e aprox. normais
- $H_0 : \mu_X = \mu_Y$ (equivalentemente, $\mu_X - \mu_Y = 0$)
 $H_1 : \mu_X \neq \mu_Y$

- Estatística de teste:

$$T = \frac{(\bar{X} - \bar{Y}) - \mu_0}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}} \approx t(\nu^*) \approx t(\nu) \text{ onde}$$

$$\nu^* = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{1}{n_X - 1} \left(\frac{s_X^2}{n_X}\right)^2 + \frac{1}{n_Y - 1} \left(\frac{s_Y^2}{n_Y}\right)^2}$$

e ν é o maior inteiro que não excede ν^* .

- Decisão: Rejeitar H_0 com nível de significância α se $t > t_{1-\alpha/2}(\nu)$ ou $t < -t_{1-\alpha/2}(\nu)$

$$\text{valor-}p = 2P(T \leq -8.011)$$

A conclusão do teste de hipóteses confirma-se através do intervalo de confiança a 99% para a diferença de médias entre as duas populações,

$$IC_{99\%} = (-7.32, -3.64)$$

dado que este intervalo não contém o 0.

Observações	X (não fum.)	Y (fum.)
1	12.1	16.6
2	12.0	13.9
3	10.8	11.3
4	11.0	14.5
5	9.7	17.4
6	14.9	15.3
⋮	⋮	⋮
23	10.3	17.4
24	11.1	21.5
25		22.9
26		19.2
27		17.3
28		15.4
29		18.3
Totais	$n_X = 24$	$n_Y = 29$
Médias	$\bar{x} = 11.32$	$\bar{y} = 16.80$
D. padrão	$s_X = 1.839$	$s_Y = 3.079$
$t = -8.011$ $t_{0.99}(\nu)$ 2.410	$\nu^* = 46.765$ $t_{0.995}(\nu)$ 2.687	$\nu = 46$

Resposta: As diferenças observadas para o nível de destruição entre fumadores e não fumadores são significativas, para um nível de significância de 0.01.

```

> base <- read.spss("TesteMediasInd1.sav", to.data.frame=TRUE)
Warning message:
In read.spss("TesteMediasInd1.sav", to.data.frame = TRUE) :
  TesteMediasInd1.sav: Unrecognized record type 7, subtype 18 encountered in system file

> names(base)
[1] "indiv"      "destruicao" "fumador"
> dim(base)
[1] 53  3

> t.test(base$destruicao ~base$fumador, conf.level=0.99)

      Welch Two Sample t-test

data:  base$destruicao by base$fumador
t = -8.0107, df = 46.765, p-value = 2.542e-10
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -7.316691 -3.643079
sample estimates:
mean in group 0 mean in group 1
    11.31667      16.79655

```


1.4.2. Exemplo 2

O objectivo de um estudo realizado por Sairam *et al* (2003) foi identificar o papel de algumas doenças e factores de risco no desenvolvimento da trombose. Um dos passos deste estudo consistiu na comparação entre os níveis de anticorpos IgG em indivíduos com e sem trombose. A tabela ao lado apresenta os resultados obtidos.

Problema: Com base nestes resultados, e para um nível de significância de 0.01, estude o efeito da doença sobre os valores de IgG.

- Dados: a.a. de duas pop. grandes e independentes apresentando uma distribuição (empírica) que se assume aproximadamente simétrica.

$$\begin{aligned} H_0 : \mu_X &= \mu_Y \\ H_1 : \mu_X &\neq \mu_Y \end{aligned}$$

- Estatística de teste:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}} \underset{a}{\sim} N(0, 1)$$

- Decisão: Rejeitar H_0 com nível de significância α se $z > N_{1-\alpha/2}(0, 1)$ ou $z < N_{\alpha/2}(0, 1)$

Níveis de IgG (ml/unidade)
em indivíduos com e sem trombose

Grupo	Tamanho	Média	D. Padrão
Com trombose (X)	153	59.01	44.89
Sem trombose (Y)	174	46.61	34.85
$z = 2.762$	$N_{.95}(0, 1)$ 1.65	$N_{.99}(0, 1)$ 2.33	$N_{.995}(0, 1)$ 2.58

A apresentação unicamente da tabela obriga a que os cálculos sejam feitos *à mão*.

O valor- p do teste é

$$p = 2 P(Z \geq 2.762) \approx 0.0057.$$

Resposta: O teste é significativo para um nível de significância de .01, o que mostra que os valores médios de IgG em indivíduos com e sem trombose são diferentes.

Exercício: Repetir o teste para $n_X = 53$ e $n_Y = 54$, supondo que existe normalidade das variáveis em causa. Obtém-se $z = 1.59$ ($p = 0.112$) pelo que o teste não é significativo. Justificar as diferenças relativamente ao que se encontrou acima.

1.4.3. Exemplo 3

Retiraram-se amostras de tecido da gengiva a um grupo de 22 mulheres com implantes dentários e a um grupo controlo de 20 mulheres. Cada amostra de tecido esteve numa solução salina durante 24 horas, ao fim das quais se procedeu à determinação do nível de interleucina-6 (pg/ml/10g do tecido), como indicador da inflamação do tecido. Os resultados obtidos foram os seguintes

Grupo experimental	236	421	52	129	76	29	96
X_1	98	226	472	664	157	84	135
	75	264	64	73	136	174	
Grupo controlo	22	29	16	39	47	176	290
X_2	132	165	127	65	34	36	64
	24	74	95	78	59	154	110
	73						

Problema: Os dados apresentados fornecem evidência suficiente para, com um nível de significância de .05, concluir que os implantes dentários podem provocar inflamações dos tecidos adjacentes?

- Dados: a.a. de 2 pop. indep. cujos logaritmos seguem distrib. aprox. normais.
- $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$
- Estatística de teste:

$$T = \frac{(\bar{X} - \bar{Y}) - \mu_0}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}} \simeq t(\nu^*) \simeq t(\nu) \text{ onde}$$

$$\nu^* = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{1}{n_X - 1} \left(\frac{s_X^2}{n_X}\right)^2 + \frac{1}{n_Y - 1} \left(\frac{s_Y^2}{n_Y}\right)^2}$$

e ν é o maior inteiro que não excede ν^* .

- Decisão: Rejeitar H_0 com nível de significância α se $t > t_{1-\alpha/2}(\nu)$ ou $t < -t_{1-\alpha/2}(\nu)$

No grupo experimental:

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
IL6	.222	20	.011	.783	20	.000

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
logIL6	.106	20	.200 [*]	.978	20	.906

a. Lilliefors Significance Correction

^{*}. This is a lower bound of the true significance.

No grupo de controlo:

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
IL6	.368	7	.005	.737	7	.009

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
logIL6	.248	7	.200 [*]	.895	7	.302

a. Lilliefors Significance Correction

^{*}. This is a lower bound of the true significance.

Homepage

Página de Rosto

Índice Geral

◀ ▶

◀ ▶

Página 34 de 55

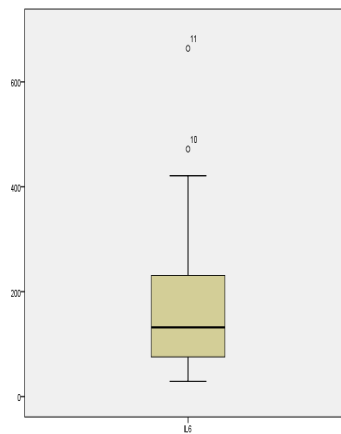
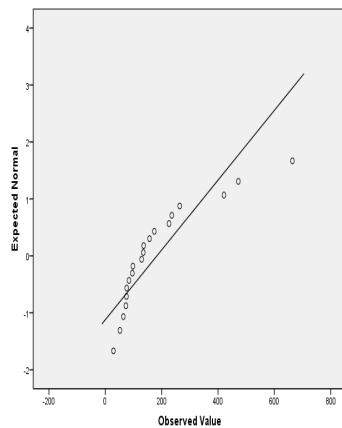
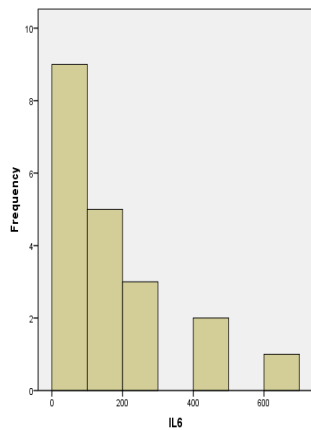
Voltar

Full Screen

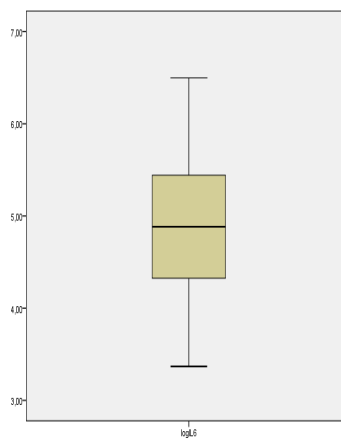
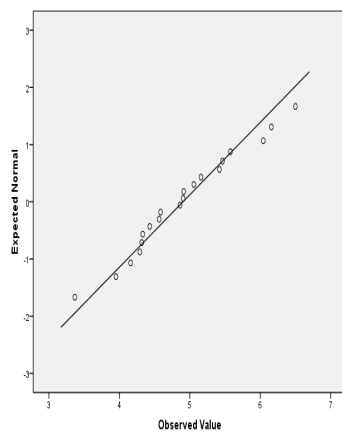
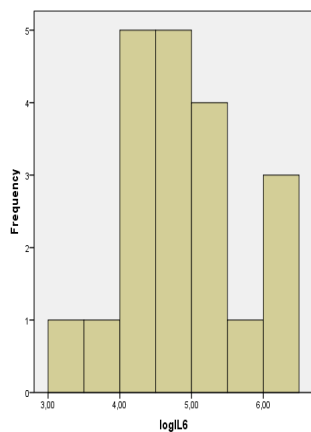
Fechar

Desistir

Nível de IL6 no grupo experimental



Nível de log(IL6) no grupo experimental



Homepage

Página de Rosto

Índice Geral

◀ ▶

◀ ▶

Página 35 de 55

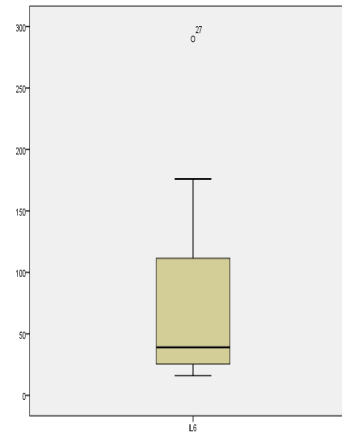
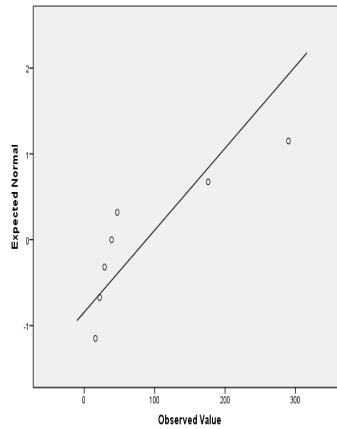
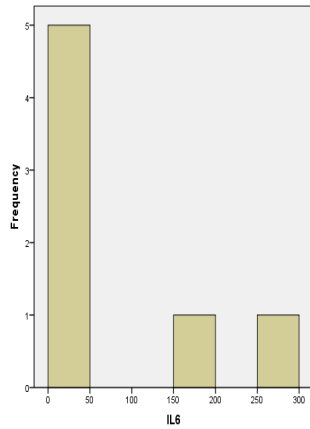
Voltar

Full Screen

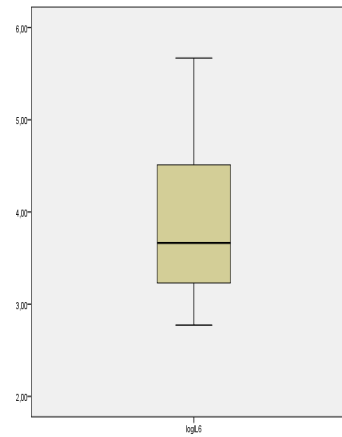
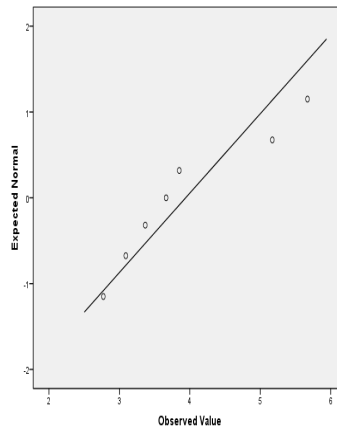
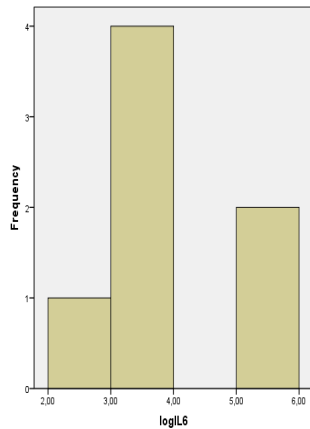
Fechar

Desistir

Nível de IL6 no grupo de controlo



Nível de log(IL6) no grupo de controlo



Reparar que o logaritmo estica a cauda esquerda e encolhe a cauda direita, fazendo com que a amostra fique menos assimétrica e se possa admitir que provém de uma distribuição normal.

```
> base <- read.spss("TesteMediasInd3.sav", to.data.frame=TRUE)
Warning message:
In read.spss("TesteMediasInd3.sav", to.data.frame = TRUE) :
  TesteMediasInd3.sav: Unrecognized record type 7, subtype 18 encountered in system file

> head(base)
  indiv IL6 controlo  logIL6
1     1 236         0 5.463832
2     2 421         0 6.042633
3     3  52         0 3.951244
4     4 129         0 4.859812
5     5  76         0 4.330733
6     6  29         0 3.367296

> t.test(base$logIL6 ~ base$controlo)

      Welch Two Sample t-test

data:  base$logIL6 by base$controlo
t = 2.9354, df = 39.302, p-value = 0.005541
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2190856 1.1893705
sample estimates:
mean in group 0 mean in group 1
   4.902374      4.198146

# alternativamente
> t.test(base$logIL6[base$controlo==0], base$logIL6[base$controlo==1])
```

Resposta: Para o nível de significância usual de 0.05 existe evidência factual para rejeitar H_0 concluindo portanto que o nível de inflamação é diferente em mulheres com implantes dentários e em mulheres sem implantes dentários. O intervalo de confiança diz-nos que, com 95% de confiança, o logIL6 é maior no grupo experimental, de 0.2 a 1.2 unidades.

1.5. Teste de hipóteses sobre uma proporção

Problema: numa população infinita dividida em duas classes disjuntas, S e F , testar se a proporção de elementos da classe S é igual, menor ou maior do que um dado valor π_0 .

- **Dados:** a.a. x_1, \dots, x_n de uma **variável aleatória X binária** representando as duas **categorias em causa**

Suponhamos que o sucesso consiste em pertencer à classe S e seja π a proporção de sucessos.

A variável aleatória U que representa o **número de sucessos numa amostra de tamanho n** segue uma distribuição binomial $B(n, \pi)$.

Se n é grande e π não é muito próximo de 0 nem de 1, nomeadamente,

$$n \geq 30, \quad n\pi > 5 \quad \text{e} \quad n(1 - \pi) > 5,^a$$

pode usar-se a aproximação de De Moivre-Laplace

$$U \sim B(n, \pi) \stackrel{a}{\sim} N(n\pi, n\pi(1 - \pi)).$$

- (i) $H_0 : \pi = \pi_0, \quad H_1 : \pi \neq \pi_0$
- (ii) $H_0 : \pi = \pi_0, \quad H_1 : \pi > \pi_0$
- (iii) $H_0 : \pi = \pi_0, \quad H_1 : \pi < \pi_0$

^aEm Martin Bland - An Introduction to Medical Statistics ou Montgomery - Applied Statistics and Probability for Engineers, por exemplo

- **Estatística de teste:** sob H_0 ,

$$Z = \frac{U - n\pi_0}{\sqrt{n\pi_0(1 - \pi_0)}} = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}} \stackrel{a}{\sim} N(0, 1),$$

onde U representa o número amostral de sucessos e $P = U/n$ é a proporção amostral. (A aproximação pela distribuição normal pode usar uma *correção de continuidade de Yates*). Equivalentemente,

$$Z^2 \stackrel{a}{\sim} \chi^2(1).$$

- Decisão: Rejeitar H_0 com **nível de significância α** se:

- (i) $z > N_{1-\alpha/2}(0, 1)$ ou $z < N_{\alpha/2}(0, 1)$
(equivalentemente, $z^2 \geq \chi_{1-\alpha}^2(1)$)
- (ii) $z > N_{1-\alpha}(0, 1)$
- (iii) $z < N_{\alpha}(0, 1)$.

Observação: O teste sobre uma proporção aplica-se na maior parte das vezes à situação em que a hipótese nula afirma que as duas categorias em causa têm a mesma probabilidade de ocorrência.

Instruções em R:

```
prop.test(x,
  n,
  p = NULL,
  alternative = c("two.sided",
    "less", "greater"),
  conf.level = 0.95,
  correct = TRUE)
```

onde

- x é o número de sucessos observados
- n é o tamanho da amostra
- p é a proporção π_0 definida em H_0 (por omissão, $p = 0.5$)
- **alternative** é definida de acordo com a hipótese alternativa
- **conf.level** determina o grau de confiança do intervalo de confiança a devolver, calculado pelo método de Wald (o usual).
- **correct**= TRUE ou FALSE, caso se pretenda ou não correcção de continuidade, respectivamente. É aconselhável não fazer correcção de continuidade.

Nota: Infelizmente, quer o valor- p do teste quer o valor π_0 definido na hipótese nula são representados na instrução acima por p . É preciso ter cuidado e não os confundir.

Para amostras de tamanho pequeno ($n < 30$) ou em que falhe uma das condições $n\pi > 5$ ou $n(1-\pi) > 5$, o teste a usar considera directamente a distribuição binomial $B(n, \pi)$ - **teste exacto**.

Algumas considerações a ter em conta:

- quando a hipótese nula não é $H_0 : \pi = 0.5$, o significado do teste binomial bilateral não é consensual; diferentes estatísticos, e diferentes softwares de estatística, têm interpretações diferentes e, por vezes, fornecem resultados diferentes para os mesmos dados.

No **SPSS** (até à versão 17.0) por exemplo, não existe a possibilidade de se realizarem testes binomiais bilaterais para hipóteses nulas que não sejam $H_0 : \pi = 0.5$.

Mas nem em todas as situações é razoável assumir que os testes são unilaterais...

- Quando aplicados a amostras pequenas, os testes de hipóteses que usam distribuições discretas, como por exemplo o teste binomial para a proporção, são conservativos. Isto significa, por exemplo, que quando H_0 é verdadeira, o valor- p (ordinário) é inferior a 0.05 não exactamente em 5% das situações mas em geral em menos do que 5% das situações. Neste sentido, os valores- p de testes que usam distribuições discretas tendem a ser demasiado grandes.

Para contornar o problema, muitos estatísticos preferem usar um tipo diferente de valor- p : o valor- p intermédio (*mid p-value*)^a.

^aAgresti, *An Introduction to Categorical Data Analysis*

Para realizar o teste binomial exato partimos também da v.a. U que representa o número de sucessos observados numa amostra com n elementos, seguindo a distribuição binomial $B(n, \pi)$.

- (i) $H_0 : \pi = \pi_0$, $H_1 : \pi \neq \pi_0$
- (ii) $H_0 : \pi = \pi_0$, $H_1 : \pi > \pi_0$
- (iii) $H_0 : \pi = \pi_0$, $H_1 : \pi < \pi_0$

- **Estatística de teste:** sob H_0 ,

$$U \sim B(n, \pi)$$

u_{obs} = número de sucessos observados na amostra

- **Decisão:**

(i) o valor- p do teste corresponde à soma das probabilidades pontuais que são menores ou iguais à probabilidade pontual do número observado de sucessos.

$$\text{valor} - p = \sum_{u \in S} P(U = u | H_0)$$

onde

$$S = \{u \in U \mid P(U = u | H_0) \leq P(U = u_{obs} | H_0)\}.$$

Quanto aos testes unilaterais:

- (i) $H_0 : \pi = \pi_0$, $H_1 : \pi > \pi_0$

o valor- p corresponde à probabilidade de obter valores para U maiores ou iguais a u_{obs} , uma vez assumido H_0 ,

$$p = P(U \geq u_{obs} | H_0).$$

- (ii) $H_0 : \pi = \pi_0$, $H_1 : \pi < \pi_0$

o valor- p corresponde à probabilidade de obter valores para U menores ou iguais a u_{obs} , uma vez assumido H_0 ,

$$p = P(U \leq u_{obs} | H_0).$$

Observação:

O cálculo do valor- p do teste pode ser simplificado na situação em que $\pi_0 = 0.5$ pois, nesta situação, a distribuição binomial é simétrica^a; a probabilidade dos dois conjuntos disjuntos da condição anterior, a azul, passa a ser o dobro da probabilidade de apenas um desses conjuntos.

^aa distribuição binomial é assimétrica à esquerda se $\pi < 0.5$ e assimétrica à direita se $\pi > 0.5$.

Instruções em R: ^a:

```
binom.test(x,  
n,  
p = 0.5,  
alternative = c("two.sided",  
                "less", "greater"),  
conf.level = 0.95)
```

onde

- x é o número de sucessos observados
- n é o tamanho da amostra
- p é a proporção π_0 definida em H_0 (por omissão, $p = 0.5$)
- **alternative** é definida de acordo com a hipótese alternativa
- **conf.level** determina o grau de confiança do intervalo de confiança a devolver, calculado pelo método de Clopper-Pearson.

^apode-se escolher se o teste é unilateral ou bilateral; Os valores-p são calculados em conformidade com a unilateralidade/bilateralidade do teste.

Observação:

Suponhamos que se quer testar se numa **população finita** de tamanho N , dividida em duas classes S e F , a proporção de elementos da classe S é igual, menor ou maior do que π_0 .

Considere-se uma amostra aleatória de tamanho n **sem reposição** dessa população. Se π for a proporção de elementos da população que estão na classe S e U for a variável aleatória que representa o número de elementos observados na classe S nessa amostra, então U segue uma distribuição hipergeométrica,

$$H(N; n, \pi).$$

Se N e n forem grandes ($N > 100$ e $n \geq 30$) e também se tiver $n\pi > 5$ e $n(1 - \pi) > 5$ então pode-se usar a aproximação

$$H(n, \pi) \approx N \left(n\pi, n\pi(1 - \pi) \frac{N - n}{N - 1} \right)$$

e prosseguir como nos casos anteriores.

Homepage

Página de Rosto

Índice Geral

◀◀ ▶▶

◀ ▶

Página 41 de 55

Voltar

Full Screen

Fechar

Desistir

1.5.1. Exemplo 1

Coates *et al.* realizaram análises a 423 indivíduos consumidores de drogas injectáveis e concluíram que 18 desses indivíduos eram seropositivos.

HIV-	HIV+	n
405	18	423

Problema: Com um nível de significância de 0.05, existe evidência estatística para concluir que a proporção de indivíduos seropositivos de entre os que consomem drogas injectáveis é diferente de 5%?

- Dados: a.a. associada a uma v.a. binária que representa a serologia (positiva/negativa) para o VIH

$$H_0 : \pi = 0.05$$

$$H_1 : \pi \neq 0.05$$

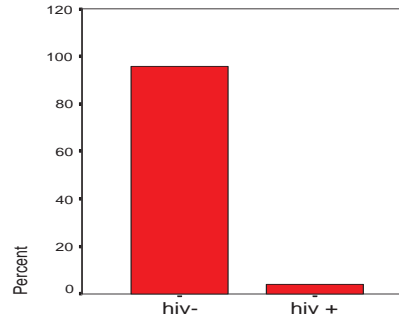
- Estatística de teste: sob H_0 ,

$$Z = \frac{P - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}} \stackrel{a}{\sim} N(0, 1)$$

A aproximação pela normal é válida pois $n = 423 \geq 30$, $n\pi_0 = 21.15 > 5$ e $n(1 - \pi_0) = 401.85 > 5$.

- Decisão: Rejeitar H_0 com nível de significância α se $z > N_{1-\alpha/2}(0, 1)$ ou $z < N_{\alpha/2}(0, 1)$

Resposta: Com um nível de significância de 0.05, não se deve rejeitar a hipótese nula portanto não existe evidência estatística suficiente para concluir que a proporção de indivíduos seropositivos de entre os que se injectam com drogas é diferente de 0.05.



		Sucessos	Total
# indivíduos	405	18	423
Proporção	0.958	0.0426	1
$z = \frac{p - 0.05}{\sqrt{0.05(1 - 0.05)/423}}$	$p = 0.0426$	$n = 423$	$z = -0.703$
	≈ 0.95 1.65	≈ 0.975 1.96	≈ 0.99 2.33

O valor-p é portanto

$$p = 2P(Z \leq -0.703) = 0.4821.$$

Um valor muito semelhante é obtido do output do

R:

$$p = 2P(Z \geq \sqrt{0.4938}) = 0.4822$$

ou, equivalentemente,

$$p = P(Z^2 \geq 0.4938) = 1 - pchisq(0.4938, 1) = 0.4822.$$

Nota: não confundir o valor-p com a proporção amostral! ambos usam o mesmo símbolo, p ...

Atente-se nas instruções quanto à existência ou não de correção de continuidade; **é preferível não usar a correção de continuidade.**

```
# usando correção de continuidade
> prop.test(18, 423, 0.05)
    1-sample proportions test with continuity correction

data:  18 out of 423, null probability 0.05
X-squared = 0.3495, df = 1, p-value = 0.5544
alternative hypothesis: true p is not equal to 0.05
95 percent confidence interval:
 0.02616059 0.06767207
sample estimates:
           p 
0.04255319 

# não usando correção de continuidade
> prop.test(18, 423, 0.05, correct=FALSE)
    1-sample proportions test without continuity correction

data:  18 out of 423, null probability 0.05
X-squared = 0.4938, df = 1, p-value = 0.4822
alternative hypothesis: true p is not equal to 0.05
95 percent confidence interval:
 0.02708388 0.06625630
sample estimates:
           p 
0.04255319 

#valor-p
> 1-pchisq(0.4938, 1)
0.4822371
```

[Homepage](#)[Página de Rosto](#)[Índice Geral](#)[Página 43 de 55](#)[Voltar](#)[Full Screen](#)[Fechar](#)[Desistir](#)

1.5.2. Exemplo 2

Uma abordagem que tem sido usada para avaliar a (a)normalidade do número de mortes por cancro verificadas em regiões próximas de centrais nucleares consiste nos estudos de mortalidade proporcional (*proportional-mortality studies*): a proporção de mortes devidas a uma causa específica num grupo de indivíduos expostos é comparada com a correspondente proporção num grupo *grande* de indivíduos não expostos.

Suponhamos, por exemplo, que se verificaram 13 mortes no grupo de todos os trabalhadores de uma central nuclear, do sexo masculino e com idades compreendidas entre 55 e 64 anos, e que 5 dessas mortes foram por cancro. Sabe-se, de outros estudos, que 20% das mortes que ocorrem no país em causa se devem a alguma forma de cancro.

	Sucessos	Total
# indivíduos	5	13
proporção	0.38	1

Problema: Para um nível de significância de 0.05, os dados fornecem evidência estatística suficiente para concluir que existe uma associação significativa entre trabalhar numa central nuclear e morrer por cancro?

- **Dados:** a.a. associada a uma v.a. binária que representa morte por cancro

- $H_0 : \pi = 0.2$

$$H_1 : \pi \neq 0.2$$

- **Estatística de teste:** sob H_0 ,

U : número de mortes por cancro em 13 indivíduos

$$U \sim B(13, 0.2)$$

O tamanho da amostra é pequeno ($n = 13 < 30$).
A estatística de teste toma o valor

$$u_{obs} = 5.$$

- **Decisão:**

O valor- p corresponde à soma das probabilidades pontuais que são menores ou iguais à probabilidade pontual do número observado de sucessos. A instrução em **R**, na página seguinte, dá valor $-p = 0.154$.

Resposta: Para um nível de significância de 0.05, o teste de hipóteses não é significativo ($p = 0.154$) portanto não se consegue afirmar que existem mais mortes por cancro entre os trabalhadores da central nuclear do que na população em geral.

```

> binom.test(5, 13, p=0.2)

      Exact binomial test

data:  5 and 13
number of successes = 5, number of trials = 13, p-value = 0.1541
alternative hypothesis: true probability of success is not equal to 0.2
95 percent confidence interval:
 0.1385793 0.6842224
sample estimates:
probability of success
      0.3846154

# confirmação do valor-p
# soma das probabilidades pontuais <= ao valor de dbinom(5,13,0.2)
size <- 13
p <- 0.2
x <- 5
all.probs<-dbinom(0:size, prob=p, size=size)
plot(0:size, all.probs)  # probability density function

# probability that the outcome is x=5
prob.x <- dbinom(x=x, prob=p, size=size)

# probability of all outcomes that have a probability less than or equal
  to the probability of the outcome 5 (prob.x)
p.value <- sum(all.probs[all.probs <= prob.x])
p.value  # 0.154

```

[Homepage](#)
[Página de Rosto](#)
[Índice Geral](#)

[Página 45 de 55](#)
[Voltar](#)
[Full Screen](#)
[Fechar](#)
[Desistir](#)

1.5.3. Exemplo 3

Considere-se o problema levantado no Exemplo 2, anterior, mas agora referente aos seguintes dados:

	Sucessos	Total
# indivíduos	8	33
proporção	0.24	1

Problema: Os dados fornecem evidência estatística suficiente para concluir que existe uma associação significativa entre trabalhar numa central nuclear e morrer por cancro?

- **Dados:** a.a. associada a uma v.a. binária que representa morte por cancro

- $H_0 : \pi = 0.2$
 $H_1 : \pi \neq 0.2$

- **Estatística de teste:** sob H_0 ,

$$Z = \frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}} \stackrel{a}{\sim} N(0, 1)$$

(equivalentemente, $Z^2 \stackrel{a}{\sim} \chi^2(1)$)

uma vez que $n = 33 > 30$ e $n\pi_0 = 6.6 > 5$ e $n(1 - \pi_0) = 26.4 > 5$.

- **Decisão:** Rejeitar H_0 com nível de significância α se $z > N_{1-\alpha/2}(0, 1)$ ou $z < N_{\alpha/2}(0, 1)$,
(equivalentemente, $z^2 > \chi_{1-\alpha}^2(1)$).

Para resolver o problema, construiu-se um ficheiro com 33 observações e uma variável aleatória, que se designou por *cancro*, tomando os valores "1" para os indivíduos que morreram por cancro e "0" para os restantes. A localização dos 8 indivíduos com "cancro=1" é irrelevante para a solução do problema.

Resposta: Os resultados das páginas seguintes mostram que, para um nível de significância de 0.05, o teste não é significativo. Isto significa que não se consegue afirmar que o número de mortes observado não está de acordo com a estimativa nacional do país ($z^2=0.371$, $p=0.542$)^a.

^aEsta é uma forma possível de apresentar os resultados: estatística de teste, valor- p .

Mais uma vez, desaconselha-se efetuar uma correção de continuidade.

```
> prop.test(8, 33, 0.2, correct=FALSE)

1-sample proportions test without continuity correction

data: 8 out of 33, null probability 0.2
X-squared = 0.3712, df = 1, p-value = 0.5423
alternative hypothesis: true p is not equal to 0.2
95 percent confidence interval:
 0.1283171 0.4102462
sample estimates:
      p 
0.2424242
```



[Homepage](#)

[Página de Rosto](#)

[Índice Geral](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

[Página 47 de 55](#)

[Voltar](#)

[Full Screen](#)

[Fechar](#)

[Desistir](#)

1.6. Teste de hipóteses sobre a diferença de duas proporções

Problema: Considerem-se duas populações infinitas com indivíduos classificados em duas classes S_1, F_1 na primeira população e S_2, F_2 na segunda população. Suponhamos que se quer testar se a proporção π_1 da classe S_1 na primeira população é igual, menor, ou maior do que a proporção p_2 da classe S_2 na segunda população.

- **Dados:** a.a. x_1, \dots, x_{n_1} da primeira população e y_1, \dots, y_{n_2} da segunda população da v.a. binária associada à classe em estudo

- (i) $H_0 : \pi_1 = \pi_2$ ^a, $H_1 : \pi_1 \neq \pi_2$
- (ii) $H_0 : \pi_1 = \pi_2$, $H_1 : \pi_1 > \pi_2$
- (iii) $H_0 : \pi_1 = \pi_2$, $H_1 : \pi_1 < \pi_2$

- **Estatística de teste:** sob H_0 ,

Se $n_i \geq 30$, $n_i p_i > 5$ e $n_i p_i (1 - p_i) > 5$, para $i = 1, 2$, então

$$Z = \frac{P_1 - P_2}{\sqrt{P(1-P) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \approx N(0, 1)$$

ou, equivalentemente,

$$Z^2 \approx \chi^2(1).$$

^aNotar que só estamos a testar a **igualdade** de proporções, e não $\pi_1 - \pi_2 = \pi_0$ para um valor π_0 qualquer. A situação tratada aqui é a mais comum.

Na expressão anterior, e para as amostras consideradas:

- $p_1 = \frac{1}{n_1}$ (#sucessos observados na amostra 1) é a proporção (amostral) na amostra 1
 - $p_2 = \frac{1}{n_2}$ (#sucessos observados na amostra 2) é a proporção (amostral) na amostra 2
 - $\bar{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ é uma proporção amostral combinada de p_1 e p_2
- **Decisão:** Rejeitar H_0 com **nível de significância α** se:
 - (i) $z \geq N_{1-\alpha/2}(0, 1)$ ou $z \leq N_{\alpha/2}(0, 1)$ (equivalentemente, $z^2 \geq \chi_{\alpha}^2(1)$)
 - (ii) $z \geq N_{1-\alpha}(0, 1)$
 - (iii) $z \leq N_{\alpha}(0, 1)$

Notas:

1. Efectuar um teste bilateral sobre a diferença de duas proporções ou um teste χ^2 para a homogeneidade de proporções é equivalente; ambas as abordagens fornecem o mesmo valor- p .
2. Caso uma das condições de aplicação não seja satisfeita, deve-se usar o **teste exacto de Fisher** que utiliza a distribuição hipergeométrica. A instrução em **R** é **fisher.test(...)**; em particular, esta instrução fornece um intervalo de confiança para o **odds ratio** $\frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$.

Observação: Da teoria sobre a proporção amostral, sabe-se que (distribuições assintóticas)

$$P_1 \stackrel{a}{\sim} N\left(\pi_1, \frac{\pi_1(1-\pi_1)}{n_1}\right)$$

$$P_2 \stackrel{a}{\sim} N\left(\pi_2, \frac{\pi_2(1-\pi_2)}{n_2}\right).$$

Dado que as populações são independentes,

$$\text{Var}(P_1 - P_2) = \text{Var}(P_1) + \text{Var}(P_2)$$

e portanto

$$P_1 - P_2 \stackrel{a}{\sim} N\left(\pi_1 - \pi_2, \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}\right),$$

isto é,

$$\frac{(P_1 - P_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}} \stackrel{a}{\sim} N(0, 1).$$

Sob H_0 tem-se $\pi_1 = \pi_2 (= \pi, \text{digamos})$ e portanto

$$\frac{P_1 - P_2}{\sqrt{\pi(1-\pi)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \stackrel{a}{\sim} N(0, 1).$$

Pode agora estimar-se π pela média dos sucessos obtidos nas duas populações,

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}.$$

Substituindo π por p na expressão anterior obtém-se a fórmula especificada atrás na estatística de teste.

Instruções em R:

```
prop.test(x, n, p = NULL,
          alternative = c("two.sided",
                        "less", "greater"),
          conf.level = 0.95,
          correct = TRUE)
```

onde

- x é um vector de sucessos (em cada grupo) **ou** uma matriz com 2 colunas onde se indica o número de sucessos e insucessos (por esta ordem).
- n é um vector com o número de experiências realizadas (em cada grupo); só se indica n na situação em que x não é uma matriz.
- p é um vector de probabilidades de sucesso (com o mesmo comprimento que o número de grupos).
- `correct=TRUE` corresponde a considerar uma correcção de continuidade.

Notar que a instrução pode ser usada para comparar proporções em mais de 2 grupos.

1.6.1. Exemplo 1

Realizou-se um inquérito a 204 pessoas residentes num lar de idosos. De entre 55 indivíduos que se identificaram como hipertensos, 24 estavam a seguir uma dieta com restrições em sódio. Dos restantes 149 indivíduos sem hipertensão, 36 seguiam também uma dieta com quantidades de sódio limitadas.

	Não Hipert.	Hipert.
C/ Restr. Na ⁺	36	24
Total	149	55

Problema: Com um nível de significância de 0.01, pode-se concluir que a dieta com restrições em sódio é mais seguida pelos indivíduos hipertensos?

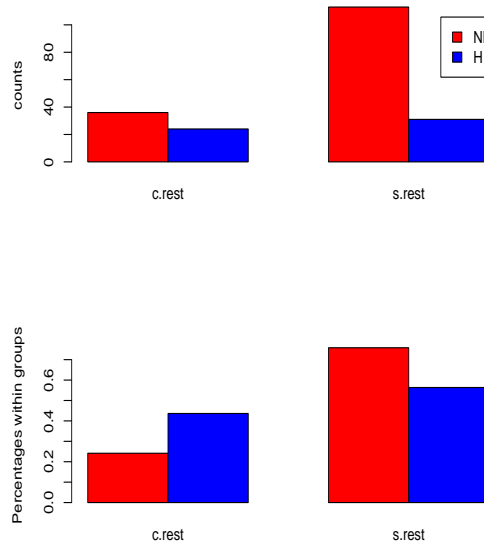
- Dados: a.a. de duas populações independentes da v.a. que representa a adesão ou não à dieta;
 $n_i \geq 30$, $n_i p_i > 5$ e $n_i(1 - p_i) > 5$ para $i = 1, 2$.
- $H_0 : \pi_1 = \pi_2$
 $H_1 : \pi_1 \neq \pi_2$
- Estatística de teste: sob H_0 ,

$$Z = \frac{(P_1 - P_2) - (\pi_1 - \pi_2)}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \stackrel{a}{\sim} N(0, 1),$$

$$U = Z^2 \stackrel{a}{\sim} \chi^2(1) \quad \left(P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}\right)$$
- Decisão: Rejeitar H_0 com nível de significância α se $z < N_{\frac{\alpha}{2}}(0, 1)$ ou $z > N_{1-\frac{\alpha}{2}}(0, 1)$; equivalentemente, $z^2 > \chi^2_{1-\alpha}(1)$.

Observação: problemas com uma estrutura semelhante a esta, em que não se tem acesso ao ficheiro de dados mas apenas aos números finais da experiência, podem ser resolvidos à mão ou criando um ficheiro de dados que reproduza os resultados obtidos.

	Hipert.		Não Hipert.	
	# suc.	Total	# suc.	Total
# ind.	24	55	36	149
p	p_1	n_1	p_2	n_2
0.2941	0.4363	55	0.2416	149
$z = 2.7084$	$z_{0.90}$ 1.29	$z_{0.95}$ 1.65	$z_{0.99}$ 2.33	$z_{0.999}$ 3.07



Resposta: Com um nível de significância de 0.01, pode-se concluir que as proporções são diferentes ($p = 2 * P(Z \geq 2.7084) \approx 0.007$). Os resultados amostrais sugerem que as restrições em sódio são maiores entre os hipertensos.

1.7. Teste de hipóteses sobre a razão de variâncias de duas distribuições normais

Problema: testar se a variância σ_X^2 desconhecida de uma variável aleatória X que segue uma distribuição normal é igual, maior ou menor do que a variância σ_Y^2 desconhecida de uma variável aleatória Y que segue uma distribuição normal. Supõe-se que as variáveis aleatórias X e Y são independentes, e que para além das variâncias também não se conhecem as médias populacionais.

- **Dados:** a.a. x_1, \dots, x_{n_X} e y_1, \dots, y_{n_Y} de v.a. contínuas X e Y , resp., definidas em cada uma das populações
- (i) $H_0 : \sigma_X^2 = \sigma_Y^2$, $H_1 : \sigma_X^2 \neq \sigma_Y^2$
- (ii) $H_0 : \sigma_X^2 = \sigma_Y^2$, $H_1 : \sigma_X^2 > \sigma_Y^2$
- **Estatística de teste:**

$$F = \frac{S_X^2}{S_Y^2} \sim F(n_X - 1, n_Y - 1)$$

onde

- $\bar{X} = \frac{1}{n_X} \sum_{k=1}^{n_X} X_k$ é a média amostral de X
- $S_X^2 = \frac{1}{n_X - 1} \sum_{k=1}^{n_X} (X_k - \bar{X})^2$ é a variância amostral de X
- $\bar{Y} = \frac{1}{n_Y} \sum_{k=1}^{n_Y} Y_k$ é a média amostral de Y
- $S_Y^2 = \frac{1}{n_Y - 1} \sum_{k=1}^{n_Y} (Y_k - \bar{Y})^2$ é a variância amostral de Y .

Atenção aos graus de liberdade de F : têm de ser definidos de acordo com o numerador e o denominador da fracção.

Para **testes unilaterais** escolhe-se $F > 1$ de acordo com H_1 (para obter a região de rejeição definida abaixo).

Para **testes bilaterais** não é necessária nenhuma escolha especial de F .

- **Decisão:** Rejeitar H_0 com **nível de significância** α se
 - (i) $F > F_{1-\alpha/2}(n_X - 1, n_Y - 1)$
ou $F < F_{\alpha/2}(n_X - 1, n_Y - 1)$
 - (ii) $F \geq F_{1-\alpha}(n_X - 1, n_Y - 1)$

Notas:

- o teste da razão de variâncias não é robusto a desvios da normalidade.
- o teste não pode ser aplicado a amostras que não sejam independentes.

Instruções em R:

```
## Instrução 1
var.test(x, y,
         ratio = 1,
         alternative = c("two.sided",
                        "less", "greater"),
         conf.level = 0.95,
         ...)

OU

## Instrução 2
var.test(formula, data, subset, na.action,
         ...)
```

onde

- x e y são vectores numéricos
- formula refere-se a uma fórmula do tipo

$$var_1 \sim var_2$$

onde var_1 é a variável contínua em causa e var_2 é o factor com duas categorias que separa os indivíduos da amostra em duas populações independentes

- data é a matriz ou dataframe de onde provêm os dados
- subset é um vector (opcional) que especifica as observações que se pretendem usar.

1.7.1. Exemplo 1

Duas amostras aleatórias independentes de duas espécies diferentes de ratinhos forneceram os resultados apresentados ao lado para os níveis de glucose no plasma depois de uma experiência traumática.

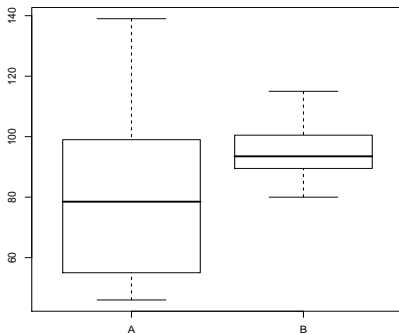
Problema: Os dados apresentados são suficientes para concluir que existem diferenças ao nível da dispersão entre as duas populações? Em particular, efetue um teste de hipóteses que tenha como hipótese alternativa a afirmação de que a variância é maior na espécie A.

- Dados: a.a. de duas pop. independentes e aprox. normais
- $H_0 : \sigma_1^2 = \sigma_2^2$
 $H_1 : \sigma_1^2 > \sigma_2^2$
- Estatística de teste: sob H_0 ,
 $F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$
Para testes unilaterais escolhe-se $F > 1$ de acordo com H_1 .
- Decisão: Rejeitar H_0 com nível de significância α se $f > F_{1-\alpha}(n_1 - 1, n_2 - 1)$

Observ.	Glucose (X_1) Espécie A	Glucose (X_2) Espécie B
1	54	93
2	99	91
3	105	93
4	46	115
5	70	80
6	87	104
7	55	108
8	58	83
9	139	88
10	91	95
11		94
12		97
Total	$n_1 = 10$	$n_2 = 12$
Média	$\bar{x}_1 = 80.4$	$\bar{x}_2 = 95.1$
D. padrão	$s_1 = 29.21$	$s_2 = 9.99$
Variância	$s_1^2 = 852.93$	$s_2^2 = 99.72$
$f = 8.55$	$\nu = (9, 11)$	
$F_{0.95}(\nu)$	$F_{0.99}(\nu)$	$F_{0.999}(\nu)$
≈ 2.90	≈ 4.64	≈ 8.14

Resposta: Assumindo normalidade nas duas populações e com um nível de significância de .05, pode-se rejeitar H_0 e concluir que a dispersão é maior nos ratinhos da raça A do que nos da raça B.
($p = P(F \geq 8.55) \approx 0.0008$).

Observação: os tamanhos amostrais em causa são demasiado pequenos para o resultado sobre a normalidade poder ser assumido com segurança.



```
> glucose <- c(54, 99, 105, 46, 70, 87, 55,
58, 139, 91, 93, 91, 93, 115, 80, 104,
108, 83, 88, 95, 94, 97)
> raca <- c(rep("A", 10), rep("B", 12))

> glucose.A <- glucose[especie=="A"]
> glucose.B <- glucose[especie=="B"]

> var.test(glucose ~ especie,
alternative="greater")
```

F test to compare two variances

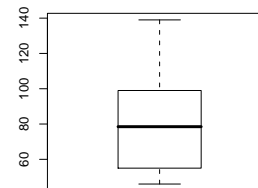
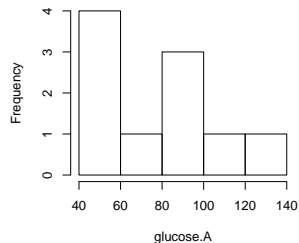
```
data: glucose by especie
F = 8.5533, num df = 9, denom df = 11,
p-value = 0.0007928
alternative hypothesis: true ratio of variances
is greater than 1
95 percent confidence interval:
2.953263      Inf
sample estimates:
ratio of variances
8.553309
```

```
# instrução equivalente
> var.test(glucose.A, glucose.B,
alternative="greater")
```

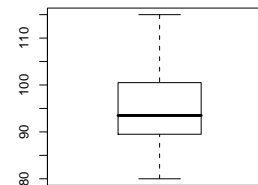
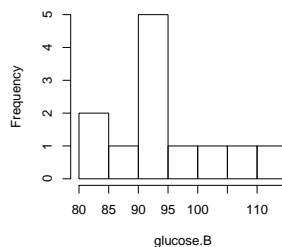
F test to compare two variances

```
data: glucose.A and glucose.B
F = 8.5533, num df = 9, denom df = 11,
p-value = 0.0007928
alternative hypothesis: true ratio of variances
is greater than 1
95 percent confidence interval:
2.953263      Inf
sample estimates:
ratio of variances
8.553309
```

glucose: raça A



glucose: raça B

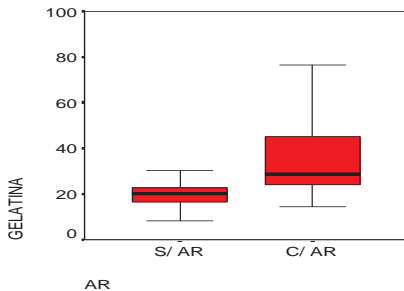


1.7.2. Exemplo 2

Gruber *et al.* realizaram um estudo com o objectivo de encontrar um marcador sérico para a artrite reumatóide (AR) suficientemente poderoso para reflectir mecanismos patológicos subjacentes à doença. Foram medidos, entre outras variáveis, os níveis de circulação de gelatinase B no sêrum, em pacientes com AR e em pacientes sem AR (*controles*). Os resultados obtidos são apresentados na tabela ao lado.

Problema: Os dados são suficientes para, com um nível de significância de 0.005, concluir que a variância do nível de gelatinase B no sêrum é diferente em indivíduos com AR e em indivíduos sem AR?

- **Dados:** a.a. de duas pop. independentes e aprox. normais
- $H_0 : \sigma_1^2 = \sigma_2^2$
 $H_1 : \sigma_1^2 \neq \sigma_2^2$
- **Estatística de teste:** sob H_0 ,
 $F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$
- **Decisão:** Rejeitar H_0 com nível de sig. α se $f > F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$ ou $f < F_{\alpha/2}(n_1 - 1, n_2 - 1)$



Observ.	C/ AR (X_1)	S/ AR (X_2)
1	26.8	23.4
2	19.1	30.5
3	52.6	10.1
4	14.5	17.3
5	22.7	24.4
6	43.5	19.7
7	25.4	8.4
8	29.8	20.4
9	27.6	16.3
10	16.1	16.5
11	76.5	22.2
12	36.7	20.4
13	57.2	
14	71.3	
15	25.3	
16	46.7	
17	30.9	
18	27.5	
19	17.2	
20	31.6	
Total	$n_1 = 20$	$n_2 = 12$
Média	$\bar{x}_1 = 34.95$	$\bar{x}_2 = 19.13$
Variância	$s_1^2 = 314.36$	$s_2^2 = 36.65$
$f = 8.58$	$\nu = (19, 11)$	
$F_{0.95}(\nu)$	$F_{0.99}(\nu)$	$F_{0.999}(\nu)$
≈ 2.65	≈ 4.10	≈ 7.01

Resposta: Assumindo normalidade e independência nas duas populações, com um nível de significância de .005 pode-se rejeitar H_0 e concluir que as variâncias são diferentes. Contudo, a amostra 2 tem um tamanho amostral pequeno e a sua distribuição empírica apresenta alguma assimetria. Assim, o teste para a diferença de variâncias fica mais frágil e aconselha-se a realização do teste de Levéne.