

Modelos Matemáticos e Aplicações (15/16)

Aulas 4 e 5

Manuela Neves

7 e 9 de Março de 2016

1 Intervalos de Confiança

- Distribuições por amostragem
- Intervalos de confiança para μ , σ^2 e p
- Intervalos de confiança - duas médias populacionais
- Intervalos de confiança - duas variâncias populacionais
- Intervalos de confiança – caso geral

2 Testes de Hipóteses

- Exemplo introdutório
- Primeiras noções
- Erros de tipo I e tipo II
- Etapas na construção de um testes de hipóteses
- Testes de Hipóteses– o p – *value*

3 Exercícios

4 Testes de ajustamento

5 Testes de independência

6 Testes não paramétricos

Intervalos de confiança

Falámos até aqui da estimação pontual e métodos de determinar estimadores de um parâmetro desconhecido θ .

Iremos agora tratar a questão da **estimação intervalar**.

Os intervalos são preferíveis quando, em vez de se propôr uma estimativa isolada, $\hat{\theta}$, podemos associar-lhe uma medida de erro $\hat{\theta} \pm \epsilon$, para significar que provavelmente o verdadeiro valor do parâmetro estará em $\hat{\theta} - \epsilon, \hat{\theta} + \epsilon$.

Definição

Considere-se uma amostra aleatória (X_1, X_2, \dots, X_n) de uma população com função de distribuição $F(x|\theta)$. Sejam $\Theta_1^*(X_1, \dots, X_n)$ e $\Theta_2^*(X_1, \dots, X_n)$ duas estatísticas, tais que

$$P(\Theta_1^* < \theta < \Theta_2^*) = 1 - \alpha, \quad 0 < \alpha < 1,$$

onde α é uma constante, não dependente do parâmetro θ .

Diz-se que (Θ_1^*, Θ_2^*) é um intervalo aleatório, que contém θ com probabilidade $1 - \alpha$.

Com a utilização de um intervalo de confiança para estimarmos um parâmetro ficamos a ganhar?

Intervalos de confiança

Efectivamente, pensemos por exemplo, no estimador \bar{X} .

Tem-se $P[\bar{X} = \mu] = 0$, mas já temos uma probabilidade positiva se considerarmos

$$P\{\mu \in]\bar{X} - a, \bar{X} + a[\} \quad \text{com } a > 0$$

ou seja, há uma probabilidade positiva de o intervalo aleatório conter o parâmetro desconhecido.

Definição

A qualquer intervalo (θ_1^*, θ_2^*) , com $\theta_1^* < \theta_2^*$, números reais, que resulta da concretização do intervalo aleatório chama-se **intervalo de confiança** a $(1 - \alpha)100\%$ para θ .

Intervalos de confiança

A determinação de intervalos de confiança para os parâmetros necessita do conhecimento da distribuição dos estimadores envolvidos **distribuições por amostragem**, isto é, são distribuições de funções da amostra aleatória (X_1, X_2, \dots, X_n) , que vamos usar para obter **Intervalos de Confiança**

Relembrando

- Se $X \sim N(\mu, \sigma)$ e σ conhecido $\rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

Distribuições por amostragem

Para obter o Intervalo de Confiança para σ^2

Variável usada	Condições	Distribuição
$\frac{(n-1)S^2}{\sigma^2}$ $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$	$X_i \sim N(\mu, \sigma)$ $i = 1, 2, \dots, n$	$\chi^2_{(n-1)}$

Definição da distribuição χ^2

Se Z_1, Z_2, \dots, Z_n são v.'s a.'s $N(0, 1)$ independentes

\Downarrow
a v.a. $X = Z_1^2 + \dots + Z_n^2$ é tal que $X \sim \chi^2_{(n)}$

Tem-se $E[X] = n$; $Var[X] = 2n$

Distribuições por amostragem

Para obter o I.C. para μ com σ desconhecido

Variável usada	Condições	Distribuição
$\frac{\bar{X} - \mu}{S/\sqrt{n}}$ $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$	$X_i \sim N(\mu, \sigma)$ $i = 1, 2, \dots, n$	$t_{(n-1)}$

Definição da distribuição t – *Student*

Se $Z \sim N(0, 1)$ e $X \sim \chi^2_{(n)}$ são v.a. independentes

\Downarrow

$$T = \frac{Z}{\sqrt{X/n}} \sim t_{(n)}$$

Intervalos de confiança para μ

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para μ quando $X \sim N(\mu, \sigma)$

- Se σ conhecido

$$\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

($z_{\alpha/2} \rightarrow$ valor da v.a. Z tal que $P(Z > z_{\alpha/2}) = \alpha/2$)

- Se σ desconhecido

$$\bar{X} - t_{\alpha/2, (n-1)} \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2, (n-1)} \frac{s}{\sqrt{n}}$$

Observações: Chama-se **precisão da estimativa** à semi-amplitude do intervalo de confiança e **confiança** ou **grau de confiança** a $(1 - \alpha) \times 100\%$. Quanto maior for o intervalo, maior é o grau de confiança, mas menor a precisão da estimativa.

Intervalo de confiança (exemplo)

Exemplo de construção de um I.C. no , para o valor médio de uma normal com variância conhecida (exemplo académico!)

Exemplo Dada a amostra referente a 10 alturas, admita-se que os erros de medição são normais de média 0 e desvio padrão 1.5.

```
> x<-c(175,176, 173, 175, 174, 173, 173, 176, 173, 179)
> int.conf.z<-function(x,sigma,conf.level=0.95)
  n <-length(x);xbar<-mean(x)
  alpha <- 1 - conf.level
  zstar <- qnorm(1-alpha/2)
  SE <- sigma/sqrt(n)
  xbar + c(-zstar*SE,zstar*SE)  ## definimos uma função
> int.conf.z(x,1.5)           # basta fazer isto
```

Obteve-se o I.C a 95% para μ]173.7703; 175.6297[

Intervalos de confiança

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para μ

Se X tem **dist. qualquer não normal**

É necessário dispor de uma **amostra de dimensão elevada**, i.e., n grande \rightarrow aplicação do Teorema Limite Central

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1) \quad \text{se } \underline{\sigma \text{ conhecido}}$$

Ou, que é o caso mais frequente,

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim \mathcal{N}(0, 1) \quad \text{se } \underline{\sigma \text{ desconhecido}}$$

Intervalo a $(1 - \alpha) \times 100\%$ de confiança para μ

$$\bar{X} - z_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Intervalos de confiança

Intervalo a $(1 - \alpha) \times 100\%$ de confiança para σ^2 numa população normal

$$\frac{(n-1)s^2}{\chi_{\alpha/2, (n-1)}^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_{1-\alpha/2, (n-1)}^2}$$

($\chi_{\alpha}^2 \rightarrow$ valor da v.a. χ^2 tal que $P[\chi^2 > \chi_{\alpha}^2] = \alpha$)

Intervalo de confiança $(1 - \alpha) \times 100\%$ para p

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

se $X \sim B(n, p)$ e n “grande”

Intervalos de confiança - duas médias populacionais

Intervalos de confiança a $(1 - \alpha) \times 100\%$ para $\mu_1 - \mu_2$ com $X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$ e $X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$ e amostras independentes

- se variâncias conhecidas

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- se variâncias desconhecidas mas se pode admitir variâncias iguais.

$$(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$t_{\alpha/2} \equiv t_{\alpha/2, (n_1+n_2-2)} \quad \text{e} \quad s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

Intervalos de confiança - duas médias populacionais

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para $\mu_1 - \mu_2$ em duas populações quaisquer (não normais), amostras independentes, de dimensões n_1 e n_2 elevadas

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para $\mu_1 - \mu_2$, quando as variâncias σ_1^2 e σ_2^2 são desconhecidas e desiguais

Esta situação foi tratada por Welch-Satterthwaite que considerou uma aproximação t à v.a. usada na construção do intervalo acima, sendo o n^o de graus de liberdade calculados aproximadamente— voltaremos a isto quando tratarmos os Testes de Hipóteses.

Intervalos de confiança - duas variâncias populacionais

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para σ_1^2/σ_2^2 , com $X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$ e $X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$ e amostras independentes

Surge aqui uma nova distribuição, cuja caracterização foi feita por Fisher e Snedecor, baseada no quociente de duas v.a. independentes, com distribuição qui-quadrado.

A distribuição F de Snedecor

Sejam $U \sim \chi_{(m)}^2$ e $V \sim \chi_{(n)}^2$ variáveis aleatórias independentes, então $X = \frac{U/m}{V/n}$ diz-se ter **distribuição F** com **(m, n) graus de liberdade** e representa-se por $X = \frac{U/m}{V/n} \sim F_{(m,n)}$.

Intervalos de confiança - duas variâncias populacionais

Considere-se agora duas amostras aleatórias de dimensão n_1 e n_2 , retiradas de forma independente de $X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$ e $X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$.

Tem-se $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{(n_1-1, n_2-1)}$, variável que permite fazer a estimação de σ_1^2/σ_2^2 . Sob estas condições tem-se

O intervalo de confiança a $(1 - \alpha) \times 100\%$ para $\frac{\sigma_1^2}{\sigma_2^2}$ é

$$\frac{s_1^2}{s_2^2 f_{\alpha/2; (n_1-1, n_2-1)}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2 f_{\alpha/2; (n_2-1, n_1-1)}}{s_2^2}$$

Intervalos de confiança (amostras emparelhadas)

Intervalos de confiança para $\mu_1 - \mu_2$ (amostras emparelhadas)

Se numa dada experiência as observações estão relacionadas, i.e, **emparelhadas** pelo indivíduo - surge aqui o conceito de **bloco**.

Consideremos a **amostra emparelhada** (X_i, Y_i) ($i = 1, \dots, n$)

Seja

$$D_1 = X_1 - Y_1; \quad D_2 = X_2 - Y_2; \quad \dots \quad D_n = X_n - Y_n, \text{ isto é,}$$

seja (D_1, D_2, \dots, D_n) a amostra aleatória das diferenças

Intervalos de confiança (amostras emparelhadas)

Se D_1, D_2, \dots, D_n são variáveis aleatórias provenientes de uma lei normal com valor médio $\mu_D = \mu_X - \mu_Y$ e variância σ_D^2 , desconhecida

tem-se
$$\frac{\bar{D} - \mu_D}{S_D/\sqrt{n}} \sim t_{(n-1)}$$

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para μ_D

$$\bar{d} - t_{\alpha/2, (n-1)} \frac{s_D}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2, (n-1)} \frac{s_D}{\sqrt{n}}$$

Se não for possível admitir D_i normais, mas se tenha n 'grande' o intervalo de confiança $(1 - \alpha) \times 100\%$ para μ_D é

$$\bar{d} - z_{\alpha/2} \frac{s_D}{\sqrt{n}} < \mu_D < \bar{d} + z_{\alpha/2} \frac{s_D}{\sqrt{n}}$$

Intervalos de confiança para $p_1 - p_2$

Sejam X_1 e X_2 variáveis aleatórias tais que

$$X_1 \sim \mathcal{B}(n_1, p_1) \quad \text{e} \quad X_2 \sim \mathcal{B}(n_2, p_2);$$

n_1 e n_2 as dimensões de amostras aleatórias independentes

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para $p_1 - p_2$ quando as dimensões das amostras são elevadas

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

Nota: Todos estes intervalos de confiança são calculados no  com uma função que, simultaneamente realiza um teste de hipóteses. Por essa razão vamos fazer exemplos mais tarde.

Os intervalos de confiança apresentados até agora foram construídos supondo conhecida a distribuição dos estimadores (mas havia a distribuição normal, pelo menos aproximadamente, subjacente.)

Se tal não for possível:

- Terá de recorrer-se a distribuições assintóticas, sendo então necessário dispor de amostras ‘grandes’;
- O método geral consiste no seguinte:
Se X é v.a. com densidade $f(x|\theta)$, seja (X_1, \dots, X_n) uma amostra aleatória.
Seja $T(X_1, \dots, X_n)$ um estimador de θ cuja função densidade é $g(t|\theta)$.

Intervalos de confiança – caso geral

Conhecendo $g(t|\theta)$, o objectivo é determinar duas funções $h_1(\theta)$ e $h_2(\theta)$ tais que

$$\int_{-\infty}^{h_1(\theta)} g(t|\theta) dt = \gamma_1 \quad \text{e} \quad \int_{h_2(\theta)}^{+\infty} g(t|\theta) dt = \gamma_2$$

com $\gamma_1 > 0$ e $\gamma_2 > 0$ e $\gamma_1 + \gamma_2 = \alpha$, ou seja tal que

$$P[h_1(\theta) < T < h_2(\theta)] = 1 - \alpha.$$

Mas temos aqui o problema de procurar o melhor dos intervalos relativos ao coeficiente de confiança que se pretende - intervalo de amplitude mínima. (ver Murteira e Casella)

Testes de Hipóteses

Na inferência estatística estudada até aqui tratou-se da obtenção de estimadores e da construção de **Intervalos de Confiança**.

Os **Testes de Hipóteses** são procedimentos de inferência que face a hipóteses estatísticas formuladas, avaliam a sua plausibilidade com recurso a regras que utilizam os dados de uma amostra observada.

Uma hipótese estatística é uma conjectura sobre uma característica desconhecida da população

Um teste de hipóteses é um procedimento estatístico que averigua se os dados sustentam uma hipótese estatística.

A Teoria dos Testes de Hipóteses “põe à prova as nossas suposições”.

Testes de Hipóteses

Seja X uma população com função de distribuição conhecida, dependente de um parâmetro θ , $F(x|\theta)$, ou de um vector de parâmetros $\underline{\theta} = (\theta_1, \dots, \theta_k)$ desconhecidos.

A ideia geral base num teste de hipóteses (teste paramétrico) é a seguinte:

– supor que o parâmetro θ tem um certo valor θ_0 – isto constitui a hipótese a testar - **hipótese estatística** -, que se costuma representar por H_0 , que pode ser verdadeira ou falsa.

Rejeita-se H_0 se houver evidência factual contra ela, apoiando-se outra hipótese.

Testar uma hipótese estatística consiste em aplicar um conjunto explícito de regras para decidir se a conjectura deve ou não ser aceite, com probabilidade mínima de erros.

Exemplo introdutório

Numa linha de engarrafamento de azeite a quantidade deitada em cada garrafa é uma variável aleatória que se admite ter distribuição normal. O processo de enchimento considera-se regulado se $\mu = 1$ *litro*, não sendo de admitir grandes desvios. Para controlar o processo de enchimento escolheram-se ao acaso 20 garrafas da produção diária. Suponha que se obteve uma média de 0.965 litros com um desvio padrão de 0.08 litros.

Poder-se-á dizer que o processo não está regulado? Justifique convenientemente a resposta.

Resolução: Para averiguar se o processo não está regulado, podemos formular um teste das hipóteses:

$$H_0 : \mu = 1$$

v.s.

$$H_1 : \mu \neq 1$$

Testes de Hipóteses—primeiras noções

A Regra de um teste consiste em:

- Dividir o espaço amostra em duas regiões complementares:

RA - região de aceitação – os valores amostrais para os quais a decisão é “aceitar” H_0

RR ou **RC** - região de rejeição ou região crítica – os valores amostrais para os quais H_0 é rejeitada.

Sendo a Inferência Estatística um “caminho” que vai do particular ao geral, pode ter erros associados:

As Possíveis decisões que se podem tomar são :

H_0 verdadeira e “aceitar” $H_0 \rightarrow$ decisão correcta

H_0 verdadeira e rejeitar $H_0 \rightarrow$ decisão incorrecta – erro

H_0 falsa e “aceitar” $H_0 \rightarrow$ decisão incorrecta – erro

H_0 falsa e rejeitar $H_0 \rightarrow$ decisão correcta

Testes de Hipóteses—erros de tipo I e tipo II

Os erros da decisão de **rejeitar ou não rejeitar H_0** são designados, respectivamente por **erro de 1^a espécie** ou **erro de tipo I** e **erro de 2^a espécie** ou **erro de tipo II**, sendo as probabilidades associadas a cada um dos erros habitualmente designadas por

$$\alpha = P(\text{erro de tipo I}) = P(\text{rejeitar } H_0 | H_0 \text{ verdadeiro})$$

$$\beta = P(\text{erro de tipo II}) = P(\text{não rejeitar } H_0 | H_0 \text{ falso}).$$

A α é costume chamar **nível de significância do teste** e a $1 - \beta = P(\text{rejeitar } H_0 | H_0 \text{ falso})$ **potência do teste**.

Testes de Hipóteses

O quadro seguinte resume o que acabamos de expor:

	não rej. H_0	rej. H_0
H_0 verd.	decisão correcta	erro tipo I
H_0 falsa	erro tipo II	decisão correcta

Testes de Hipóteses

Sendo assim, um teste de hipóteses necessita de duas hipóteses:

- 1 uma que é proposta pelo experimentador e
- 2 a outra que é a sua negação.

A primeira, habitualmente representada por H_1 , é chamada a hipótese alternativa (hipótese a pesquisar) , a outra representada por H_0 , é chamada a hipótese nula.

A hipótese a testar é a hipótese nula, que regra geral esperamos rejeitar.

Costuma escrever-se H_0 vs H_1 .

Testes de Hipóteses

A situação desejável seria, portanto, ter α e β tão pequenos quanto possível. Porém, tal não é possível.

De facto, fixada a dimensão da amostra, quando um erro diminui o outro aumenta.

O que é habitual fazer-se – **teoria de Neyman-Pearson** – é fixar o valor de α e tentar minimizar o β . Quando existe um teste nestas condições costuma chamar-se teste mais potente.

A Teoria de Neyman-Pearson estabeleceu que a realização de um teste de hipóteses se baseia na definição de **Estatísticas de Teste** e de **Regiões Críticas** associadas às hipóteses formuladas.

Voltemos ao exemplo

Admita-se então que a amostra observada é a realização da a.a. (X_1, X_2, \dots, X_n) proveniente de uma população normal da qual não se conhece σ . As hipóteses a testar são, como vimos,

$$H_0 : \mu = 1 \quad \text{vs} \quad H_1 : \mu \neq 1;$$

a **Estatística de Teste** é $\frac{\bar{X} - \mu_0}{S/\sqrt{n}}$, que sob a hipótese nula tem distribuição conhecida, que sabemos ser $t_{(n-1)}$. Temos então

$$T = \frac{\bar{X} - 1}{S/\sqrt{n}} \sim t_{(19)},$$

sendo a **região crítica**, correspondente ao nível de significância α dada por $T < -t_{\alpha/2, (19)} \cup T > t_{\alpha/2, (19)}$

A resposta num teste de hipóteses é dada na forma

- **Rejeitar H_0** - significa que os dados observados testemunham fortemente **contra H_0** - neste caso será **adoptada a hipótese H_1**
- **Não rejeitar H_0** - significa que **não há evidência suficiente para rejeitar H_0** .

Etapas na construção de um testes de hipóteses

Passos a seguir na construção de um **teste de hipóteses**:

1. Identificar o(s) parâmetro(s); especificar H_0 e H_1 e o nível de significância α .
2. Escolher uma variável aleatória – **estatística de teste**, que sob H_0 terá distribuição conhecida (pelo menos aproximadamente).
3. Definir a **região de rejeição** ou **região crítica** – **RC** (conjunto de valores da estatística que são menos “plausíveis” caso H_0 seja verdadeira, portanto levam a **rejeitar H_0**).
4. Calcular **o valor** da estatística de teste, para a amostra observada.
5. Se o valor calculado $\in RC \rightarrow$ **rejeita-se H_0**
Se o valor calculado $\notin RC \rightarrow$ **não se rejeita H_0**

Testes de Hipóteses– o *p-value*

A indicação do valor observado da estatística de teste, **por exemplo** z_{cal} , e a indicação de um **valor crítico** z_{α} para decidir, por exemplo,

Rejeitar H_0 se $z_{cal} > z_{\alpha}$ (num teste unilateral à direita)
tem sido recentemente “substituído” pelo cálculo de uma
probabilidade

– **a probabilidade de se observar um valor igual ou mais extremo do que o observado, se a hipótese nula é verdadeira** – a que se chama **valor de prova; valor- p (*p-value*)**

Nota: é esta quantidade que hoje em dia qualquer *software* está preparado para calcular quando se manda realizar um teste. Podemos interpretar o **valor de prova**, **valor-*p*** ou ***p-value*** como a **medida do grau de concordância entre os dados e H₀**

Assim:

- **Quanto menor for o *p-value*, menor é a consistência entre os dados e a hipótese nula**

Habitualmente adopta-se como regra de decisão:

$$\text{rejeitar } H_0 \text{ se } p\text{-value} \leq \alpha$$

Exercício

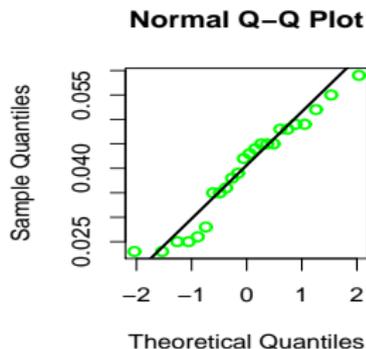
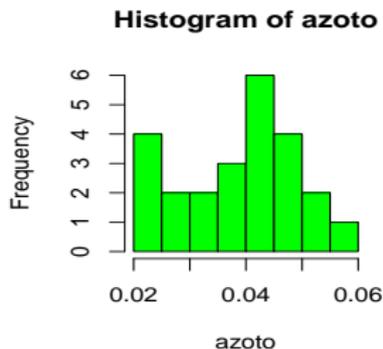
Os dados seguintes referem-se à concentração total de azoto (ppm) na água de um lago que é utilizado como fonte de abastecimento urbano.

0.042	0.023	0.049	0.036	0.045	0.025
0.048	0.035	0.048	0.043	0.044	0.055
0.045	0.052	0.049	0.028	0.025	0.039
0.023	0.045	0.038	0.035	0.026	0.059

- 1 Determine um intervalo de confiança para μ (a 99% de confiança).
- 2 Para ser aceitável como fonte de água potável, o conteúdo médio de azoto deve ser inferior a 0.07 ppm. Acha que os dados são compatíveis com aquele critério?

Exercício–resolução

```
> azoto<-c(0.042,0.048,0.045,0.023,0.023,0.035,0.052,  
+ 0.045,0.049,0.048,0.049,0.038,0.036,0.043, 0.045,  
+ 0.025, 0.044, 0.055, 0.028, 0.025, 0.039,  
+ 0.035, 0.026, 0.059)  
> qqnorm(azoto)# este é um gráfico para uma  
+           #primeira pesquisa da normalidade  
> qqline(azoto)
```



Exercício–resolução

```
> t.test(azoto,mu=0.0,conf.level=0.99)
#alinea a) pode omitir-se mu=0.0
```

One Sample t-test

```
data: azoto
```

```
t = 18.5066, df = 23, p-value = 2.606e-15
```

```
alternative hypothesis: true mean is not equal to 0
```

```
99 percent confidence interval:
```

```
0.03382623 0.04592377
```

```
sample estimates:
```

```
mean of x
```

```
0.039875
```

Exercício–resolução

```
> t.test(azoto, alternative='less', mu=0.07) #alinea b)
```

```
One Sample t-test
```

```
data: azoto
```

```
t = -13.9815, df = 23, p-value = 4.944e-13
```

```
alternative hypothesis: true mean is less than 0.07
```

```
95 percent confidence interval:
```

```
-Inf 0.04356776
```

```
sample estimates:
```

```
mean of x
```

```
0.039875
```

Testes de hipóteses e IC no R

Intervalo de confiança e teste de hipóteses para comparar os valores médios de duas populações

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

Realiza um teste e I.C. para amostras independentes, usando o t de Welch-Satterthwaite para obter uma aproximação ao n^o de graus de liberdade. Por omissão considera `paired = FALSE`, `var.equal = FALSE`.

```
>data(sleep);>sleep ## Uma alternativa a um teste a duas médias  
## forma simples de realizar o teste para comparar 2 grupos  
>t.test(extra ~ group, data = sleep, paired = TRUE)
```

Testes de hipóteses e IC no R

Intervalo de confiança e teste de hipóteses para comparar 2 variâncias de populações que se admitem normais

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

Intervalo de confiança e teste de hipóteses para verificar se a proporção de “sucessos” (em n provas nas quais se observaram x sucessos se pode admitir inferior a 0.4

```
>prop.test(x, n, p = 0.4, alternative = "less",  
          conf.level = 0.99, correct = FALSE)
```

Trata-se de um teste a $p = 0.4$ sem correcção de continuidade. É obtido o intervalo a 99% de confiança.

Exercício (TPC)

Um estudo pretende comparar um tipo de semente melhorada com o tipo de semente tradicional. A semente melhorada passará a ser utilizada se, em média, o crescimento das plantas após 20 dias for superior ao das obtidas das sementes tradicionais. São criadas 15 diferentes situações laboratoriais, variando temperatura e humidade. Em cada situação semeia-se uma semente de cada tipo e obtêm-se os seguintes resultados para o crescimento (em cm) das plantas após 20 dias :

Situação	1	2	3	4	5	6	7	8
Semente melhorada	3.46	3.48	2.74	2.83	4.00	4.95	2.24	6.92
Semente tradicional	3.18	3.67	2.92	3.10	4.10	4.86	2.21	6.91
Situação	9	10	11	12	13	14	15	
Semente melhorada	6.57	6.18	8.30	3.44	4.47	7.59	3.87	
Semente tradicional	6.83	6.19	8.05	3.46	4.18	7.43	3.85	

Deverá passar a usar-se as sementes melhoradas? Responda justificando e explicitando quaisquer hipóteses adicionais que seja necessário impor.

Exercício (TPC)

Pretende-se testar se a proporção de ulmeiros afectados pela grafiose é idêntica em duas zonas A e B. Na zona A foi recolhida uma amostra aleatória de 30 ulmeiros e verificou-se que 20 estavam afectados pela grafiose. Na zona B recolheu-se uma amostra de 35 ulmeiros e verificou-se que 27 estavam afectados pela grafiose.

Que conclusão se pode tirar ao nível de significância de 0.05?

Métodos não paramétricos

Os procedimentos de ICs e Testes de Hipóteses tratados até agora costumam designar-se por **métodos paramétricos**.

Sempre que os procedimentos usados se preocupam com o comportamento da população (e não com os seus parâmetros) ou então com os parâmetros mas sem a exigência do conhecimento de hipóteses distribucionais dizemos, de uma forma geral, que estamos perante **métodos não paramétricos**.

Nestes procedimentos estão incluídos os **testes de ajustamento** ('goodness-of-fit tests') e os testes também ditos **testes não paramétricos** ('nonparametric tests' ou 'distribution-free tests')

Testes de ajustamento

Iremos começar por considerar os Testes de Ajustamento ('*goodness-of-fit tests*') dos quais referiremos:

- o teste de Shapiro Wilk;
- o teste de Kolmogoroff-Smirnov e
- testes do **Qui-quadrado**

Começemos com um teste muito importante nas nossas aplicações - **um teste de ajustamento à distribuição normal** - i.e. permite averiguar se um dado conjunto de observações se pode considerar proveniente de uma população com distribuição normal – é um **teste de normalidade**,

O Teste de Shapiro Wilk

que se tem revelado ser um dos mais potentes. Vejamos em síntese como se processa.

O teste de normalidade de Shapiro-Wilk

Seja X a característica em estudo na população.

Formulam-se as hipóteses:

H_0 : X tem distribuição Normal

H_1 : X não tem distribuição Normal

Calcula-se o valor da estatística de teste $W_{cal} = \frac{b^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$

com b constante a determinar a partir dos dados e com recurso a uma tabela (que não será aqui dada).

Não usaremos o Teste de Shapiro-Wilk com recurso a tabelas, vamos executá-lo apenas no .

O teste de normalidade de Shapiro-Wilk no R

```
> shapiro.test(azoto)

      Shapiro-Wilk normality test

data:  azoto
W = 0.944, p-value = 0.2001
```

Um outro teste de ajustamento, que além da normalidade permite considerar o ajustamento a outras distribuições é o **teste de Kolmogorov-Smirnov**.

Este teste, introduzido por Kolmogorov e por Smirnov em 1933, baseia-se na diferença (distância) entre a **função de distribuição postulada na hipótese nula** e a **função de distribuição empírica**, construída a partir da amostra.

O teste de Kolmogorov Smirnov

Formulam-se as hipóteses:

H_0 : X tem distribuição F_0

H_1 : X tem distribuição diferente de F_0

A estatística de teste é $D = \sup_x |F_0(x) - F_n^*(x)|$, onde F_0 designa a função de distribuição considerada e F_n^* a função de distribuição empírica.

O comportamento daquela variável D foi completamente especificado por aqueles autores no caso de X ter uma lei normal com parâmetros conhecidos.

Lilliefors estudou a mesma estatística quando X era exponencial.

Actualmente os softwares estão preparados para realizar o teste para outras distribuições.

O teste de Kolmogorov Smirnov no R

Exemplo: Consideremos $n = 200$ valores gerados a partir de uma distribuição de Weibull, com os parâmetros abaixo especificados:

```
>x.wei<-rweibull(n=200,shape=2.1,scale=1.1)
+           # gera valores de uma v.a. com
+           # distribuição de Weibull, com
+           # parâmetros de forma=2.1 e escala=1.1
>ks.test(x.wei,"pweibull",shape=2,scale=1)
```

One-sample Kolmogorov-Smirnov test

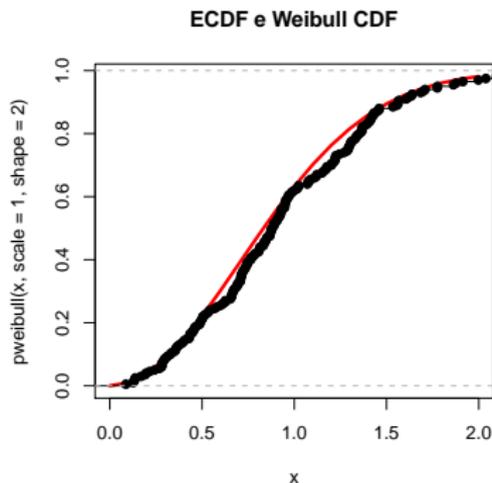
data: x.wei

D = 0.0765, p-value = 0.1918

alternative hypothesis: two-sided

O teste de Kolmogorov Smirnov no R

```
> x<-seq(0,2,0.1)
> plot(x,pweibull(x,scale=1,shape=2),type="l",col="red",
+      lwd=3, main="ECDF e Weibull CDF")
>plot(ecdf(x.wei),add=TRUE)
```



O Teste de ajustamento do Qui-quadrado

Um teste muito usado **baseado em contagens (frequências)** - é o teste do **Qui-quadrado** (K. Pearson).

Considere-se os valores possíveis da característica X , repartidos em k classes, A_1, A_2, \dots, A_k , mutuamente exclusivas. Seja

- n_i – frequência absoluta observada da classe A_i ; $\sum_{i=1}^k n_i = n$
- p_i – a probabilidade desconhecida de obter uma observação na classe $A_i, \forall i$;
- p_{0i} – a probabilidade de obter uma observação na classe A_i supondo que a observação foi extraída de uma população com a distribuição especificada em H_0 , i.e. $p_{0i} = P(A_i|H_0)$.

O Testes de ajustamento do Qui-quadrado

$H_0 : p_i = p_{i0} \quad i = 1, 2, \dots, k \quad \text{v.s.}$

$H_1 : \text{pelo menos um dos } p_i \neq p_{i0}$

A Estatística do teste é

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_{i0})^2}{np_{i0}}$$

Se H_0 verdadeira $\chi^2 \sim \chi_{(k-1)}^2$ i.e., a distribuição é assintótica, i.e., válida para dimensões de amostra elevada.

Região Crítica–RC: é unilateral direita

Rejeita-se H_0 se $\chi_{calc}^2 > \chi_{\alpha, (k-1)}^2$

Se **houver necessidade de estimar parâmetros** a estatística passa a ter assintoticamente distribuição $\chi_{(k-\nu-1)}^2$, onde ν é o número de parâmetros estimados.

O Testes de ajustamento do Qui-quadrado no R

Na tabela seguinte estão representados os resultados de um estudo experimental sobre o efeito do gorgulho Azuki do feijão. Introduziram-se larvas desse orgulho nos feijões que as alimentaram. As crisálidas saíram através de um buraco feito no feijão e, como tal, o n. de buracos por feijão indica-nos o n. de adultos que saíram. Observados 100 feijões obtiveram-se os seguintes resultados:

n. de gorgulhos saídos de 1 feijão	0	1	2	3	4
frequência observada	60	22	10	5	3

Poderá considerar-se o n^o. de gorgulhos por feijão uma v.a. com distribuição de Poisson?

```
>num<-c(0,1,2,3,4); freq<-c(60,22,10,5,3)
>lambda_est<-sum(num*freq)/100;lambda_est
>probs<-c(c(dpois(num[-5],lambda_est)),
+ ppois(3,lambda_est,lower.tail=F))
>chisq.test(freq, p = probs)
```

O Testes de ajustamento do Qui-quadrado no R

Chi-squared test for given probabilities

```
data: freq
```

```
X-squared = 19.678, df = 4, p-value = 0.000578
```

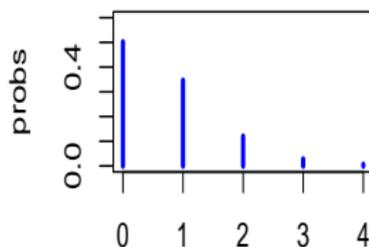
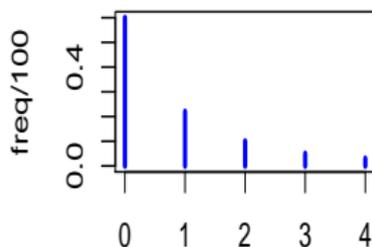
```
Warning message:
```

```
In chisq.test(freq, p = probs): Chi-squared approximation  
may be incorrect
```

```
>par(mfrow=c(1,2)) #vamos só visualizar
```

```
>plot(num,freq/100,type="h", ylim=c(0,.6),lwd=3)
```

```
>plot(num,probs,type="h",ylim=c(0,0.6),lwd=3)
```



Outros Testes de ajustamento

- `library(nortest)` é necessário carregar este package
 - `ad.test()` teste de normalidade de Anderson-Darling
 - `cvm.test()` teste de normalidade de Cramer-Von Mises
 - `lillie.test()` teste de Lilliefors
 - `pearson.test(x.norm)` teste de normalidade do qui-quadrado de Pearson
- `library(vcd)` carregar o package para o próximo teste
 - `goodfit()` ajusta uma distribuição discreta

Tabelas de contingência

Suponhamos que os indivíduos de uma amostra são classificados de acordo com dois critérios (factores) A e B (qualitativos ou quantitativos).

Consideremos r níveis do critério A e c níveis do critério B . Portanto os n valores observados são classificados de acordo com 2 diferentes factores (critérios).

É costume apresentar as frequências observadas O_{ij} na célula (i, j) de uma tabela a que se chama **tabela de contingência**

	B_1	\dots	B_j	\dots	B_c	
A_1	O_{11}	\dots	O_{1j}	\dots	O_{1c}	$O_{1.}$
A_2	O_{21}	\dots	O_{2j}	\dots	O_{2c}	$O_{2.}$
\vdots						
A_r	O_{r1}	\dots	O_{rj}	\dots	O_{rc}	$O_{r.}$
	$O_{.1}$	\dots	$O_{.j}$	\dots	$O_{.c}$	

$\sum_{i=1}^r \sum_{j=1}^c O_{ij} = n$ e O_{ij} representa o número de elementos da amostra classificados nas categorias A_i e B_j .

Tabelas de contingência–testes de independência

Se a tabela de contingência resultou da classificação dos indivíduos da amostra segundo os níveis de cada um dos critérios, regra geral pretende-se com este estudo inferior da eventual existência de alguma relação ou associação entre os dois critérios de classificação. As hipóteses a testar são:

H_0 : A e B são independentes

H_1 : A e B não são independentes

A estatística do teste é

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - e_{ij})^2}{e_{ij}},$$

onde e_{ij} representa a estimativa da frequência esperada, se a hipótese H_0 fosse verdadeira, i.e. $e_{ij} = \frac{o_{i.} \cdot o_{.j}}{n}$

Se H_0 verdadeira, $\chi^2 \sim \chi^2_{(r-1)(s-1)}$.

Rejeita-se a hipótese H_0 se $\chi^2_{cal} > \chi^2_{\alpha, (r-1)(s-1)}$

Tabelas de contingência

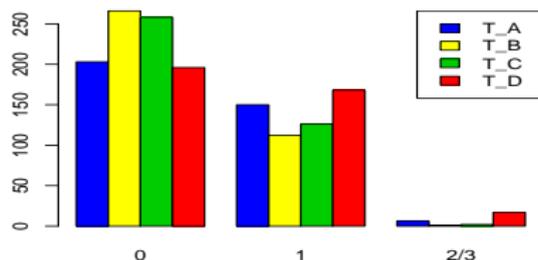
Submeteram-se ramos florais de macieiras “Golden Delicious”, a quatro tratamentos. Contou-se o número de frutos produzidos a fim de verificar se existe ou não uma relação entre os diferentes tratamentos e a frutificação. Vejamos os resultados no seguinte quadro:

Tratamentos	N. de frutos			Totais
	0	1	2 ou 3	
A	203	150	6	359
B	266	112	1	379
C	258	126	2	386
D	196	168	17	381
Totais	923	556	26	1505

Pretendemos testar a hipótese nula , de que não há relação entre os tratamentos e a frutificação, **ou seja, que são independentes.**

Resolução do Exercício no R

```
>frutos<-matrix(c(203,150,6,266,112,1,258,126,2,196,168,17),
+ nc=3,byrow=T,
+ dimnames=list(c("T_A", "T_B", "T_C","T_D"),c("0", "1","2/3")))
>frutos
>chisq.test(frutos)
>chisq.test(frutos)$expected
>(chisq.test(frutos)$residuals)^2
>barplot(frutos,names=c("0","1","2/3"),col=c(4,7,3,2),
+ cex.names=1,beside=T)
>legend("topright",c("T_A","T_B","T_C","T_D"),fill=c(4,7,3,2))
```



Pressupostos a verificar:

- as frequências esperadas em cada classe não devem ser inferiores a 5, quando o número total de observações é ≤ 20 ;
- se $n > 20$ não deverá existir mais do que 20% das células com frequências esperadas inferiores a 5, nem deverá existir nenhuma com frequência esperada inferior a 1.
- se nos casos anteriores as condições não se verificarem deve-se juntar linhas ou colunas (desde que tal junção tenha significado).
- a realização de um teste de independência não deve terminar com a rejeição da hipótese nula. Deve analisar-se a contribuição de cada célula para o valor de X^2 .

Testes Não Paramétricos

Consideremos agora que pretendemos realizar testes a parâmetros mas as hipóteses da normalidade ou da aproximação à normal não são verificadas. Como já dissemos tais testes têm, habitualmente, a designação “*distribution-free*” mas muitos autores designam-nos por testes não paramétricos.

Vamos apenas deixar aqui uma breve nota sobre dois testes muito usuais que se podem aplicar quando o teste $t_{student}$ não é válido. Estes testes são baseados nas ordens das observações, ou seja, na posição de cada observação na amostra ordenada.

Enquanto os testes paramétricos exigem que as variáveis em causa sejam quantitativas, os testes não paramétricos que vamos usar podem aplicar-se também a variáveis qualitativas, desde que as elas sejam ordinais.

Testes de Wilcoxon

Teste de Wilcoxon - teste não paramétrico para o estudo da mediana de uma população ou para comparar as medianas em duas amostras emparelhadas

Teste de Wilcoxon-Mann-Whitney - teste não paramétrico adequado à comparação de duas amostras independentes

`wilcox.test(A,B)` realiza o teste que atrás chamámos de Wilcoxon-Mann-Whitney para as duas amostras independentes A e B .

`wilcox.test(A,B,paired=T)` realiza o teste que atrás chamámos de Wilcoxon para as duas amostras A e B , mas agora consideradas emparelhadas.

Os testes de Wilcoxon no R

A tabela seguinte dá a percentagem de concentração de zinco, determinada por dois métodos diferentes, em 9 amostras de comida:

Amostra	EDTA titration	Espectrometria atómica
1	7.2	7.6
2	6.1	6.8
3	4.9	4.8
4	5.9	5.7
5	9.0	9.7
6	8.5	9.1
7	6.6	7.0
8	4	4.7
9	5.2	4.9

Poder-se-á afirmar que existe uma diferença significativa entre os resultados dos dois métodos?