Introdução aos Testes de Hipóteses

- Hipóteses estatísticas. Hipótese nula/Hipótese alternativa.
- Passos necessários à elaboração de um teste de hipóteses.
- Testes sobre parâmetros de uma ou de duas populações.
- Teste de normalidade e “teste do qui-quadrado” de ajustamento.

Testes de Hipóteses

Até aqui estivemos a tratar de Intervalos de Confiança, isto é, a construir intervalos de valores que, com uma confiança elevada, contêm o verdadeiro valor do parâmetro desconhecido.

Mas ... vamos agora ver uma outra forma de decidir se uma dada suposição (relativa ao(s) parâmetro(s) desconhecido(s) ou outros aspectos da característica em estudo) é ou não suportada pela informação fornecida pelos dados de uma amostra.
Vamos então **formular Testes de Hipóteses** e indicar como, com base numa amostra observada, se pode tomar decisões relativamente às hipóteses formuladas.

Note-se que os quadros com Intervalos de confiança e os quadros dos Testes de Hipóteses seguem a mesma estrutura, portanto as variáveis usadas na construção dos intervalos de confiança são as que iremos também aqui usar.

Testes de Hipóteses

Uma hipótese estatística é qualquer conjectura sobre aspectos desconhecidos da população (que podem ser parâmetros ou mesmo a forma da distribuição).

Se a hipótese diz respeito a:

- um parâmetro, supondo conhecida a forma da distribuição, a hipótese diz-se paramétrica.
- investigar a forma da distribuição, ou um parâmetro sem admitir o conhecimento da forma da distribuição, a hipótese diz-se não paramétrica.

Um teste de hipóteses é então um procedimento que permite decidir se uma dada hipótese é ou não suportada pela informação fornecida pelos dados de uma amostra.

Testes de Hipóteses

Um teste diz-se **paramétrico** se as hipóteses envolvidas são paramétricas, isto é, dizem respeito ao(s) parâmetros(s), supondo conhecida, pelo menos aproximadamente, a forma da distribuição.

Metodologia a seguir num teste de hipóteses:

- Como uma dada afirmação pode ser verdadeira ou falsa, formulam-se duas hipóteses alternativas:

  Hipótese nula - $\mathbf{H_0}$ - corresponde ao *statu quo*, i.e., é a hipótese que especifica o valor “actual” do(s) parâmetro(s) (é considerada verdadeira até haver evidência estatística para a sua rejeição) e

  Hipótese alternativa - $\mathbf{H_1}$ - onde se especificam o(s) valor(es) a “aceitar” quando se rejeita a hipóteses nula.

Testes de Hipóteses- Exemplos

1. $\mathbf{H}_0 : \mu = 1$ *v.s.* $\mathbf{H}_1 : \mu \neq 1$
2. $\mathbf{H}_0 : \mu = 3$ *v.s.* $\mathbf{H}_1 : \mu < 3$
3. $\mathbf{H}_0 : \mu = 1$ *v.s.* $\mathbf{H}_1 : \mu > 1$
4. $\mathbf{H}_0 : \mu \geq \mu_0$ *v.s.* $\mathbf{H}_1 : \mu < \mu_0$
5. $\mathbf{H}_0 : X \frown$ Normal *v.s.* $\mathbf{H}_1 : X$ segue outra distribuição

O último exemplo é um teste para verificar se uma dada característica segue uma distribuição (neste caso a normal).
Designa-se habitualmente por teste de normalidade.

Exemplo 1

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:

$$\mathbf{H}_0 : \mu = 1 \qquad \textit{v.s.} \qquad \mathbf{H}_1 : \mu \neq 1$$

Testes de Hipóteses

- 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** ou
  - **Não rejeitar H_0** - significa que não há evidência suficiente para rejeitar **H_0**.

Tomar decisões para a população com base numa amostra possui riscos (i.e. podem cometer-se erros). Vejamos o quadro:

	Decisões	
Realidade	**Rejeitar H_0**	**Não rejeitar H_0**
H_0 verdadeira	ERRO de tipo I	não há erro
H_0 falsa	não há erro	ERRO de tipo II

Testes de Hipóteses Paramétricos

P(erro de tipo I ou 1^a espécie)=**P(Rejeitar H_0| H_0 verdadeira)**= α

α - nível de significância

P(erro de tipo II ou 2^a espécie)=**P(Não rejeitar H_0| H_0 falsa)** =β

Costuma atribuir-se um valor muito baixo à probabilidade do erro de 1^a espécie , i.e, habitualmente considera-se

α =P(**Rejeitar H_0| H_0 verdadeira)** = 0.05 ou 0.01

Rejeitar H_0 significa que os dados testemunham fortemente contra H_0

Testes de Hipóteses

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

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** $\longrightarrow$ **rejeita-se H**$_0$
   Se o valor calculado $\notin$ **RC** $\longrightarrow$ **não se rejeita H**$_0$

Testes de Hipóteses

Iremos considerar testes para os parâmetros: valor médio, variância, diferença de valores médios no caso de amostras independentes e emparelhadas, quociente de variâncias e proporções.

Para todos estes testes os passos indicados no slide anterior encontram-se explicitados no quadro "Testes a Médias, Variâncias e Proporções".

Exemplo: de formulação de hipóteses para o valor médio:

$\mathbf{H}_0 : \mu = \mu_0$ *v.s.* $\mathbf{H}_1 : \mu \neq \mu_0$ **teste bilateral**

$\mathbf{H}_0 : \mu \leq \mu_0$ *v.s.* $\mathbf{H}_1 : \mu > \mu_0$ **teste unilateral (à direita)**

$\mathbf{H}_0 : \mu \geq \mu_0$ *v.s.* $\mathbf{H}_1 : \mu < \mu_0$ **teste unilateral (à esquerda)**

Regras de decisão em Testes de Hipóteses

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

– **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.

Regras de decisão em Testes de Hipóteses

Podemos interpretar o **valor de prova** ou **valor-*p*** ou ***p-value*** como a

medida do grau de concordância entre os dados e H_0

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:

rejeitar H_0 se ***p-value*** $\leq \alpha$

Testes de Hipóteses ... outros

Nesta breve introdução à Teoria dos Testes de Hipóteses vimos como formular testes a parâmetros desconhecidos da população, em particular:

- ao valor médio, variância e a uma proporção;
- diferença de dois valores médios em amostras independentes e em amostras emparelhadas;
- quociente de variâncias e diferença de duas proporções.

Mas ... outros testes de hipóteses podem ser necessários nas mais variadas situações.
Note-se, todavia, que os passos a seguir serão sempre os indicados no slide 169.

Testes de Hipóteses ... outros

Como ilustração podemos referir três testes:

- **Teste de normalidade de Shapiro-WilK**;
- **Teste do Qui-quadrado de ajustamento**;
- **Teste à igualdade de três ou mais valores médios em amostras independentes (este é o procedimento estatístico conhecido como Análise de Variância que na sua ideia mais básica corresponde à generalização do problema da comparação de duas médias populacionais.)** Não será tratado neste curso.

Um teste de Normalidade

Consideremos agora o caso de se pretender
testar a forma da distribuição.
Os testes a considerar agora são testes não paramétricos que se designam por testes de ajustamento.

Comecemos com um teste muito importante nas nossas aplicações - **um teste de ajustamento à distribuição normal** - para 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 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).

O teste de Shapiro-Wilk

Valores pequenos de W_{cal} indicam não normalidade, i.e.

RC: $W_{cal} < W_\alpha$

Com W_α $-$ **valor crítico** a consultar numa Tabela.

Alternativamente pode usar-se o valor-$\boldsymbol{p}$,

- Rejeita-se H_0 se valor$-\boldsymbol{p} \leq \boldsymbol{\alpha}$, significando que não se pode admitir que X tem distribuição normal;
- Se valor$-\boldsymbol{p} > \boldsymbol{\alpha}$ não se rejeita $\mathbf{H}_0$, o que significa que a distribuição normal é uma distribuição possível para X.

Exercício 2 (reescrito)

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 fôr 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.

Resolução - Exercício 2

Trata-se de um problema de amostras emparelhadas e como se pretende averiguar se as sementes melhoradas apresentam melhor desempenho, o teste a realizar é o teste unilateral

$$H_0 : \mu_D = 0 \quad vs. \quad H_1 : \mu_D > 0$$

$\boldsymbol{D}$ designa a diferença entre o comprimento das plantas das sementes melhoradas e das tradicionais.

O comando do R para realizar este teste (e da execução do qual se obtém tb o intervalo de confiança unilateral, que não fez parte do programa) é o seguinte:

```
> t.test(nsem,vsem,alternative='greater'),
paired=TRUE)
```

Resolução - Exercício 2 (cont.)

Mas ... para utilizar aquele teste é necessário verificar previamente se pode admitir-se que a diferença entre alturas das plantas provenientes das sementes melhoradas e das tradicionais segue uma distribuição normal.
Comecemos pelo teste de normalidade de Shapiro-Wilk para analisar a hipótese de normalidade:

```
> shapiro.test(nsem-vsem)
Shapiro-Wilk normality test
data:  nsem-vsem
W = 0.9428, p-value = 0.4184
```

Conclusão: Dado que o *p-value* é 0.4184, conclui-se que **não se rejeita a hipótese da normalidade**, pelo que pode realizar-se o teste t.

Resolução - Exercício 2 (cont.)

```
> t.test(nsem,vsem,alternative='greater',
paired=TRUE)
```

Paired t-test
data: nsem and vsem
t = 0.1396, df = 14, p-value = 0.4455
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
-0.07744645 Inf
sample estimates: mean of the differences
0.006666667

Como o *p-value*= 0.4455 é superior a 0.05, não se rejeita H_0.

Conclusão: Não se pode afirmar que μ_D seja superior a 0, portanto não se pode aconselhar o uso das sementes melhoradas face às tradicionais.

Os slides que se apresentam a seguir contêm matéria que, este ano lectivo, não será considerada para avaliação.

Teste de ajustamento do qui-quadrado

Falámos já num teste de ajustamento - o teste Shapiro - Wilk - **teste de normalidade**.

Mas, ... a necessidade de considerar o ajustamento a outras distribuições (contínuas ou discretas) pode surgir nas mais variadas aplicações.

O “teste do qui-quadrado” foi desenvolvido por Pearson (1900) como teste de ajustamento para dados nominais ou dados classificados em classes, i.e., é um teste que usa a contagem do número de valores que se observam em classes ou categorias.

Vamos introduzir o **teste do qui-quadrado** com um Exemplo, relembrando as etapas necessárias à realização de um teste de hipóteses - slide 165.

Exemplo

A descendência originada pelo cruzamento de dois dados tipos de plantas pode ser qualquer um dos três genótipos que representaremos por A, B e C. Um modelo teórico de sucessão genética indica que os tipos A, B e C devem aparecer na razão $1:2:1$. Efectuou-se o cruzamento daqueles dois tipos tendo-se classificado 90 plantas. A sua classificação genética foi registada na tabela:

Genótipos	A	B	C
	18	44	28

Estão estes dados de acordo com o modelo genético?

O que se pretende aqui averiguar é se X segue uma dada distribuição ou não

... no caso do exemplo, aquela averiguação inicia-se pela formulação das seguintes hipóteses:

Teste de ajustamento do qui-quadrado

H$_0$ **:** $p_1 = 0.25, \quad p_2 = 0.5, \quad p_3 = 0.25$
H$_{1:}$ pelo menos uma das probabilidades é diferente do formulado.

O teste sugerido por Karl Pearson (para o caso de as observações se encontrarem classificadas em k classes) consiste em considerar a estatística de teste:

$$\boldsymbol{X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}}$$

$k-$ nº de classes; O_i- frequência observada, é a frequência absoluta em cada classe; E_i- frequência esperada, dada por $E_i = n\,p_i$ com p_i a probabilidade da classe i, se a hipótese **H**$_0$ verdadeira; n é o número (total) de observações independentes.

O valor de X^2 em cada classe é uma medida da proximidade entre os dados e a Hipótese nula. Quanto menor for o valor de X^2 mais plausível é a hipótese **H**$_0$.

Teste de ajustamento do qui-quadrado

Pearson mostrou que X^2 tem distribuição assintótica
$\boldsymbol{X^2 \sim \chi^2_{(k-1)}}$ onde k é o número de classes em que as observações foram agrupadas.

A região crítica, para um nível de significância α, é definida como:
RC: $\boldsymbol{X^2_{cal} > \chi^2_{\alpha,(k-1)}}$

Resolução:

	A	B	C
O_i	18	44	28
p_i	0.25	0.5	0.25
$E_i = np_i$	22.5	45	22.5

$X^2_{cal} = \frac{(18-22.5)^2}{22.5} + \frac{(44-45)^2}{45} + \frac{(28-22.5)^2}{22.5} = 2.27$ e $\chi^2_{0.05,2} = 5.99$.

Então $X^2_{cal} < \chi^2_{0.05,2}$, logo não se rejeita H_0 a um nível de significância de 5%, portanto não há razões para duvidar que os dados estejam de acordo com o modelo genético.

Teste do qui-quadrado - Observações

- A variável X^2, tem distribuição assintótica. Qual a dimensão que a amostra deverá ter para que a aproximação à distribuição $\chi^2_{(k-1)}$ seja válida?
  - Alguns autores consideram que a aproximação só é válida se a frequência esperada, E_i, verifica $E_i \geq 5$;
  - Cochran (1954) sugeriu que se podia considerar classes com $E_i = 1$, desde que 80% das classes apresente $E_i > 5$. Sempre que as frequências esperadas de algumas classes forem inferiores a 1 essas classes devem agrupar-se com as adjacentes por forma a atingir a frequência mínima desejada.

- Sempre que para determinar a probabilidade p_i for necessário estimar parâmetros, a distribuição χ^2 virá

$$\chi^2_{(k-1-\text{n}^{\underline{o}}\ \text{de parâmetros estimados})}$$

Votos de que os resultados correspondam às vossas expectativa e trabalho realizado ao longo do semestre.
BOA SORTE

BOAS FESTAS
Dez/2011