

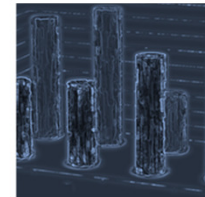
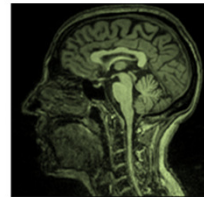
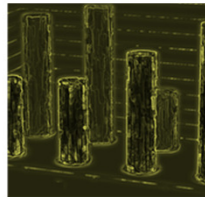
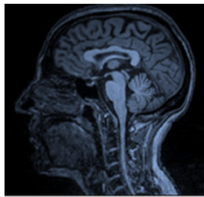


Bioestatística

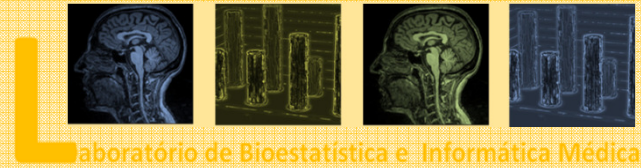
MIEB – 2015/2016

O que é a estatística?

Francisco Caramelo
Miguel Patrício
Bárbara Oliveiros



Disclaimer



- ✖ Nestes slides pretende-se introduzir alguns conceitos de estatística de forma operacional
- ✖ Uma compreensão global e aprofundada dos mesmos conceitos, consultando outras fontes, é recomendada

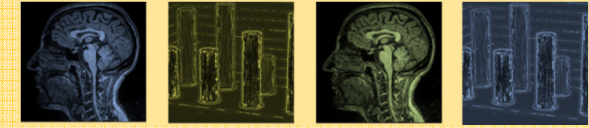
✖ Segundo a Wikipedia:

“Statistics is the study of the collection, organization, analysis, interpretation and presentation of data”



“Data don’t make any sense,
we will have to resort to statistics.”

Definição-exemplo



Laboratório de Bioestatística e Informática Médica

× Estatística

+ Descritiva

+ Inferencial

× Estimação

× Testes de hipóteses

“Como se caracterizam as idades dos doentes de Alzheimer que constam da base de dados dos CHUCs?”

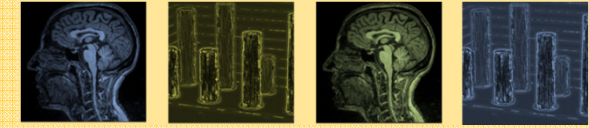
“Como se pode estimar a média da idade de todos os doentes de Alzheimer do sexo feminino, a nível mundial?”

“Em média, os pacientes de Alzheimer do sexo masculino são mais novos que as do sexo feminino?”



- ✗ Podemos dividir a estatística em:
 - + descritiva: recolha, redução, organização e apresentação de dados estatísticos
 - + inferencial: processo de tomada de decisões a partir de dados amostrais:
 - ✗ estimação
 - ✗ Testes de hipóteses

Exemplo [estatística descritiva]



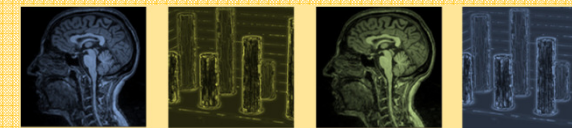
Laboratório de Bioestatística e Informática Médica

- ✗ Se se pretender conhecer o perfil dos doentes que vão a uma determinada consulta, pode-se:
 - + recolher os dados relevantes sobre os mesmos [ex: idade, género, medicação que toma, etc.]
 - + organizar os dados numa base de dados
 - + caracterizar os mesmos, calculando médias, medianas, fazendo gráficos...

Número de processo do paciente	Género	Idade
1	M	32
2	M	41
3	F	18
4	F	23
5	F	15
6	M	54
7	M	20
8	F	22
9	M	34
10	M	15



Exemplo [estatística descritiva]



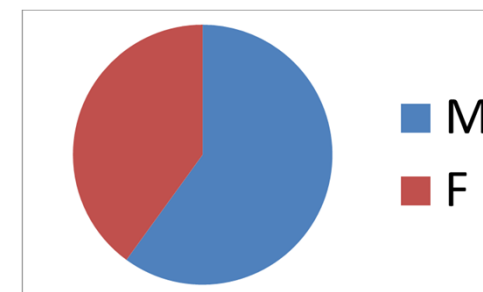
Laboratório de Bioestatística e Informática Médica

- ✗ Se se pretender conhecer o perfil dos doentes que vão a uma determinada consulta, pode-se:
 - + recolher os dados relevantes sobre os mesmos [ex: idade, género, medicação que toma, etc.]
 - + organizar os dados numa base de dados
 - + caracterizar os mesmos, calculando médias, medianas, fazendo gráficos...

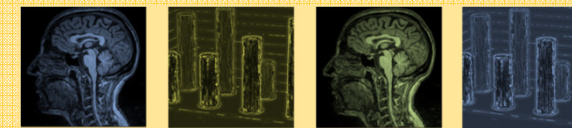
Número de processo do paciente	Género	Idade
1	M	32
2	M	41
3	F	18
4	F	23
5	F	15
6	M	54
7	M	20
8	F	22
9	M	34
10	M	15

✗ Caracterização da variável género:

Sexo	M	F	Total
Número de pacientes	6	4	10
Número de pacientes (%)	60	40	100



Exemplo [estatística descritiva]



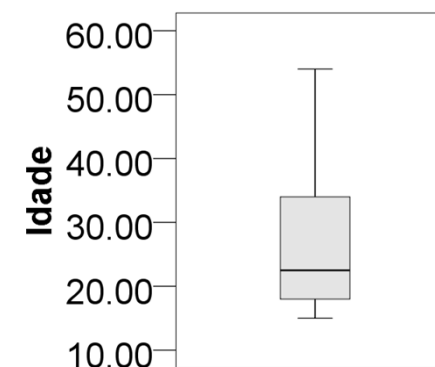
Laboratório de Bioestatística e Informática Médica

- ✗ Se se pretender conhecer o perfil dos doentes que vão a uma determinada consulta, pode-se:
 - + recolher os dados relevantes sobre os mesmos [ex: idade, género, medicação que toma, etc.]
 - + organizar os dados numa base de dados
 - + caracterizar os mesmos, calculando médias, medianas, fazendo gráficos...

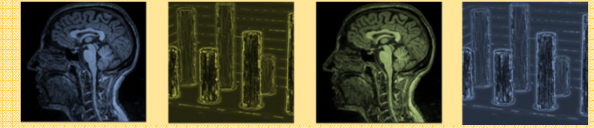
Número de processo do paciente	Género	Idade
1	M	32
2	M	41
3	F	18
4	F	23
5	F	15
6	M	54
7	M	20
8	F	22
9	M	34
10	M	15

✗ Caracterização da variável idade:

	Idade
Média	27.4
Desvio-Padrão	12.7
Mediana	22.5
Mínimo	15
Máximo	54



Estatística inferencial - Estimação



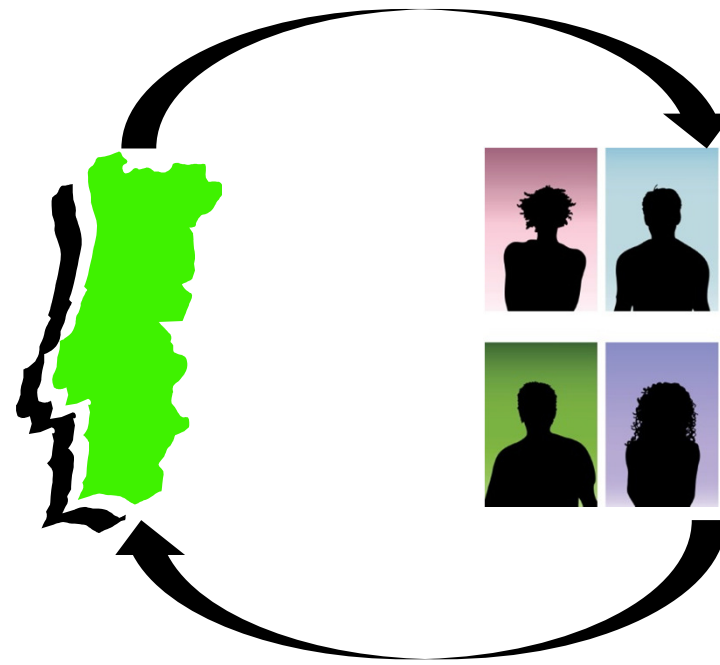
Laboratório de Bioestatística e Informática Médica

- ✗ Pretende-se estudar um conjunto de elementos com uma característica em comum: uma população
- ✗ Pretende-se conhecer medidas relativas a esta população: parâmetros
- ✗ Estuda-se apenas um subgrupo da população, uma amostra
- ✗ Medidas relativas à amostra: estatísticas

Exemplo:

População:
habitantes de
Portugal

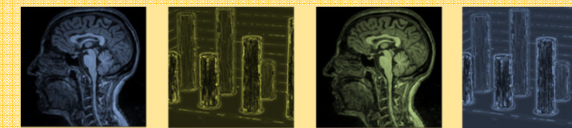
Parâmetro
média
das alturas



Amostra:
estudantes
do curso de
Estatística

Estatística:
média das
alturas

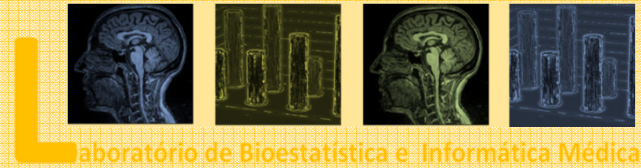
Notação



Laboratório de Bioestatística e Informática Médica

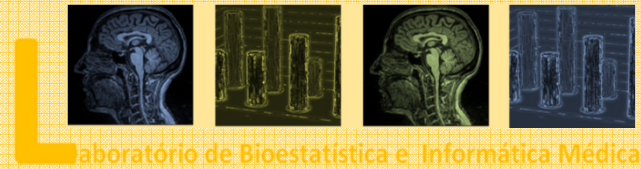
Medida estatística	Parâmetro (relativo à população)	Valor observado (de uma amostra)
Dimensão	N	n
Média	α	\bar{x}
Proporção	π	p
Variância	σ^2	s^2
Desvio padrão	σ	s
Coeficiente de correlação	ρ	r

Estatística inferencial – testes de hipóteses



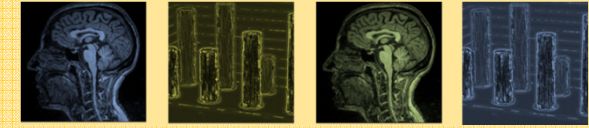
- ✗ Parte-se de uma hipótese de investigação (formula-se a hipótese nula)
- ✗ A partir dos dados amostrais, procura desmentir-se a hipótese nula
- ✗ Resultados possíveis:
 - + mostra-se haver evidência estatística suficiente para rejeitar a hipótese nula
 - + não se mostra haver evidência estatística suficiente para rejeitar a hipótese nula

Estatística inferencial – testes de hipóteses [Exemplos]



- ✗ Exemplos de hipóteses nulas que poderemos tentar desmentir com estatística inferencial:
 - + As mulheres têm o mesmo nível de glicemia que os homens
 - + Fumar não está relacionado com o aumento de probabilidade de sofrer um enfarte de miocárdio
 - + O crescimento do número de bactérias ao longo do tempo, quando colocadas nas soluções A, B ou C, é igual

Exercício



Laboratório de Bioestatística e Informática Médica

✖ As seguintes perguntas podem ser associadas à estatística

- (1) descritiva,
- (2) inferencial (estimação) ou
- (3) inferencial (teste de hipóteses)?

P1: Dos inquiridos do sexo feminino, quantos responderam “sim”?

P2: Qual é o QI médio da população mundial?

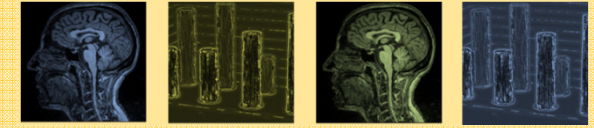
P3: O QI médio da população mundial é 110?

P4: Algum paciente apresentou um valor extremo de nível de glicose?



Resposta: P1 - 1; P2 - 2; P3 - 3; P4 - 1

Como fazer estatística?

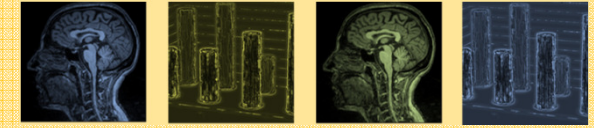


Laboratório de Bioestatística e Informática Médica

✕ Estatística descritiva

- + onde se pode calcular médias, medianas, desvios-padrão, fazer tabelas de frequências, gráficos?

Como fazer estatística?



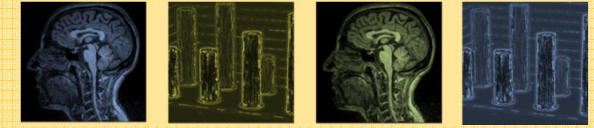
Laboratório de Bioestatística e Informática Médica

✕ Estatística descritiva

- + onde se pode calcular médias, medianas, desvios-padrão, fazer tabelas de frequências, gráficos?

Resposta: calculadora, Excel, SPSS, ...

Como fazer estatística?



Laboratório de Bioestatística e Informática Médica

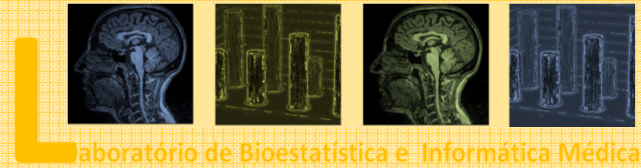
× Estatística descritiva

- + onde se pode calcular médias, medianas, desvios-padrão, fazer tabelas de frequências, gráficos?

Resposta: calculadora, Excel, SPSS, ...

- + Qual a dificuldade?

Como fazer estatística?



✗ Estatística descritiva

- + onde se pode calcular médias, medianas, desvios-padrão, fazer tabelas de frequências, gráficos?

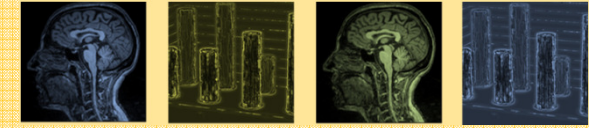
Resposta: calculadora, Excel, SPSS, ...

- + Qual a dificuldade?

Dominar um programa de computador é fácil. As partes difíceis são

- 1) ter uma base de dados consolidada
- 2) escolher os resultados que melhor informam ou melhor ilustram o que se pretende

Como fazer estatística?



Laboratório de Bioestatística e Informática Médica

× Estatística descritiva

- + onde se pode calcular médias, medianas, desvios-padrão, fazer tabelas de frequências, gráficos?

Resposta: calculadora, Excel, SPSS, ...

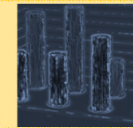
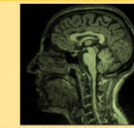
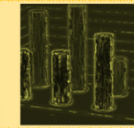
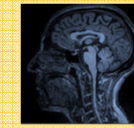
- + Qual a dificuldade?

Dominar um programa de computador é fácil. As partes difíceis são

- 1) ter uma base de dados consolidada
- 2) escolher os resultados que melhor informam ou melhor ilustram o que se pretende

- + Como fazer no SPSS? R: ver [passos_SPSS_descritiva.pdf](#)

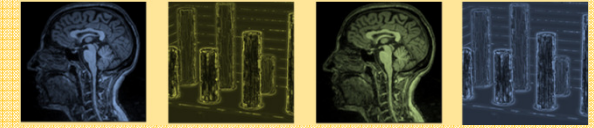
Como fazer estatística?



Laboratório de Bioestatística e Informática Médica

- × Estatística inferencial – estimação
 - + O que queremos?

Como fazer estatística?



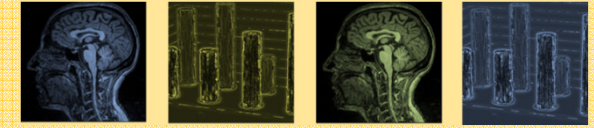
Laboratório de Bioestatística e Informática Médica

✕ Estatística inferencial – estimação

+ O que queremos?

Resposta: estimar um parâmetro (exemplo: média das alturas dos Portugueses)

Como fazer estatística?



Laboratório de Bioestatística e Informática Médica

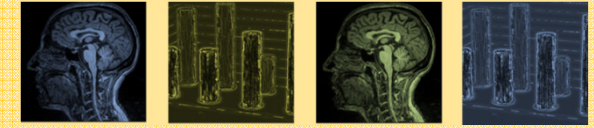
✕ Estatística inferencial – estimação

+ O que queremos?

Resposta: estimar um parâmetro (exemplo: média das alturas dos Portugueses)

+ Como o fazemos?

Como fazer estatística?



Laboratório de Bioestatística e Informática Médica

✕ Estatística inferencial – estimação

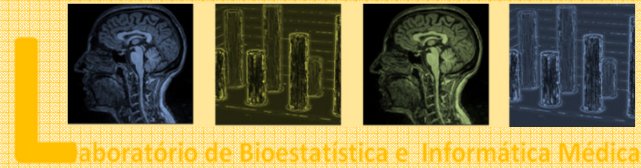
+ O que queremos?

Resposta: estimar um parâmetro (exemplo: média das alturas dos Portugueses)

+ Como o fazemos?

Resposta: determinamos uma estatística (ou seja, média das alturas de uma amostra que se pretende representativa da população)

Como fazer estatística?



✗ Estatística inferencial – estimação

+ O que queremos?

Resposta: estimar um parâmetro (exemplo: média das alturas dos Portugueses)

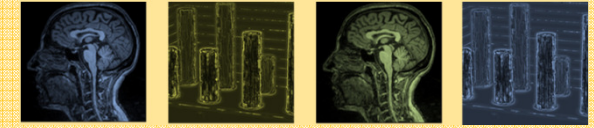
+ Como o fazemos?

Resposta: determinamos uma estatística (ou seja, média das alturas de uma amostra que se pretende representativa da população)

+ Pontos a ponderar:

- ✗ Qual é a população?
- ✗ Como escolher a amostra? É representativa?
- ✗ ...

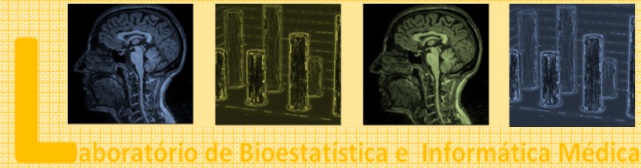
Como fazer estatística?



Laboratório de Bioestatística e Informática Médica

- ✗ Estatística inferencial – estimação: Que amostra escolheria se pretender saber a média das alturas dos habitantes em Portugal?

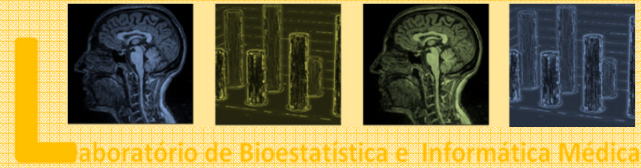
Como fazer estatística?



- ✗ Estatística inferencial – estimação: Que amostra escolheria se pretender saber a média das alturas dos habitantes em Portugal?

Resposta: podemos consultar o Census. Nesse caso, amostra=população

Como fazer estatística?

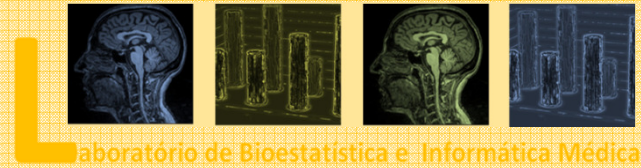


- ✗ Estatística inferencial – estimação: Que amostra escolheria se pretender saber a média das alturas dos habitantes em Portugal?

Resposta: podemos consultar o Census. Nesse caso, amostra=população

- + E se não tiver acesso ao Census?

Como fazer estatística?



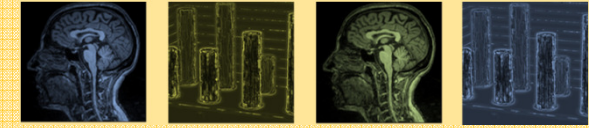
- ✗ Estatística inferencial – estimação: Que amostra escolheria se pretender saber a média das alturas dos habitantes em Portugal?

Resposta: podemos consultar o Census. Nesse caso, amostra=população

- + E se não tiver acesso ao Census?

Resposta: pergunta-se a um conjunto de pessoas (amostra) qual a sua altura? Este conjunto de pessoas deve ser tão representativo da população de Portugal quanto possível. Por exemplo, poderá ser boa ideia estratificar por géneros (porquê?).

Como fazer estatística?



Laboratório de Bioestatística e Informática Médica

- ✗ Estatística inferencial – estimação: Que amostra escolheria se pretender saber a média das alturas dos habitantes em Portugal?

Resposta: podemos consultar o Census. Nesse caso, amostra=população

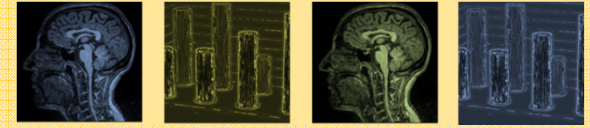
- + E se não tiver acesso ao Census?

Resposta: pergunta-se a um conjunto de pessoas (amostra) qual a sua altura? Este conjunto de pessoas deve ser tão representativo da população de Portugal quanto possível. Por exemplo, poderá ser boa ideia estratificar por géneros (porquê?).

- + E se um médico especialista numa doença rara pretender estimar um parâmetro relativamente aos pacientes com essa doença? R: poderá ter de usar uma amostra por conveniência...

- ✗ Estatística inferencial: testes de hipóteses
 - + Existem diferentes testes de hipóteses, mas todos têm em comum:
 - ✗ partir de uma hipótese (hipótese nula)
 - ✗ determinar um valor p . Se ele for mais baixo que o nível de significância α , rejeita-se a hipótese nula.

Testes de hipóteses - exemplo

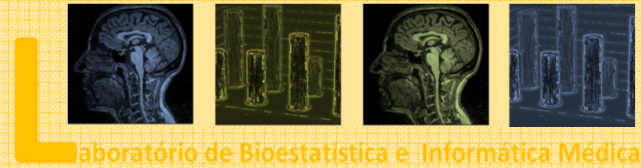


Laboratório de Bioestatística e Informática Médica

✗ Ponto de partida

- + escolha do nível de significância α
- + formulação da questão de investigação: definição da hipótese nula H_0

Testes de hipóteses - exemplo

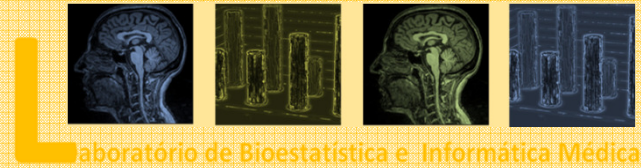


✗ Ponto de partida

- + escolha do nível de significância α
- + formulação da questão de investigação: definição da hipótese nula H_0

Exemplo: $\alpha=0.05$; H_0 : As mulheres têm o mesmo nível de glicemia que os homens

Testes de hipóteses - exemplo



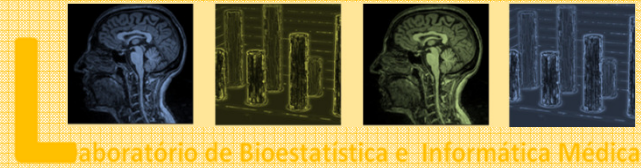
✗ Ponto de partida

- + escolha do nível de significância α
- + formulação da questão de investigação: definição da hipótese nula H_0

Exemplo: $\alpha=0.05$; H_0 : As mulheres têm o mesmo nível de glicemia que os homens

- + (já com base de dados consolidada – estatística descritiva) Escolha do teste estatístico

Testes de hipóteses - exemplo



✗ Ponto de partida

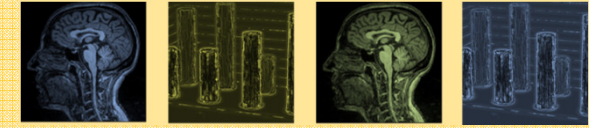
- + escolha do nível de significância α
- + formulação da questão de investigação: definição da hipótese nula H_0

Exemplo: $\alpha=0.05$; H_0 : As mulheres têm o mesmo nível de glicemia que os homens

- + (já com base de dados consolidada – estatística descritiva) Escolha do teste estatístico

A escolha depende de vários factores. Neste caso, assumindo que “n” é pequeno, o teste escolhido é o Mann-Whitney U

Testes de hipóteses - exemplo



Laboratório de Bioestatística e Informática Médica

✗ Ponto de partida

- + escolha do nível de significância α
- + formulação da questão de investigação: definição da hipótese nula H_0

Exemplo: $\alpha=0.05$; H_0 : As mulheres têm o mesmo nível de glicémia que os homens

- + (já com base de dados consolidada – estatística descritiva) Escolha do teste estatístico

A escolha depende de vários factores. Neste caso, assumindo que “n” é pequeno, o teste escolhido é o Mann-Whitney U

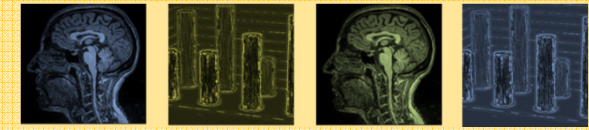
- + Cálculo do p-valor



Test Statistics ^a	
	VAR00002
Mann-Whitney U	5.000
Wilcoxon W	15.000
Z	-.929
Asymp. Sig. (2-tailed)	.353
Exact Sig. [2*(1-tailed Sig.)]	.486 ^b

a. Grouping Variable: VAR00001
b. Not corrected for ties.

Testes de hipóteses - exemplo



Laboratório de Bioestatística e Informática Médica

✗ Ponto de partida

- + escolha do nível de significância α
- + formulação da questão de investigação: definição da hipótese nula H_0

Exemplo: $\alpha=0.05$; H_0 : As mulheres têm o mesmo nível de glicémia que os homens

- + (já com base de dados consolidada – estatística descritiva) Escolha do teste estatístico

A escolha depende de vários factores. Neste caso, assumindo que “n” é pequeno, o teste escolhido é o Mann-Whitney U

- + Cálculo do p-valor
- + Comparação do p-valor com α :

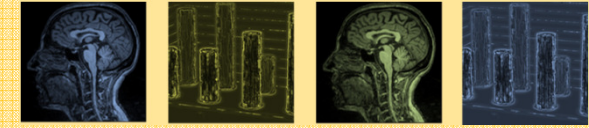
Como $p > \alpha$, não podemos rejeitar a hipótese nula, ou seja, não conseguimos mostrar que as mulheres têm nível de glicémia diferente dos homens

Test Statistics ^a	
	VAR00002
Mann-Whitney U	5.000
Wilcoxon W	15.000
Z	-.929
Asymp. Sig. (2-tailed)	.353
Exact Sig. [2*(1-tailed Sig.)]	.486 ^b

a. Grouping Variable: VAR00001
b. Not corrected for ties.



Testes de hipóteses - exercício



Laboratório de Bioestatística e Informática Médica

- ✖ Cada uma das questões de investigação seguintes foi alvo de um estudo estatístico. Conhecendo os níveis de significância de cada estudo e os p-valores que foram obtidos, para cada estudo:
 - + identifique a hipótese nula
 - + indique se se deve ou não rejeitar a hipótese nula
 - + responda à questão de investigação

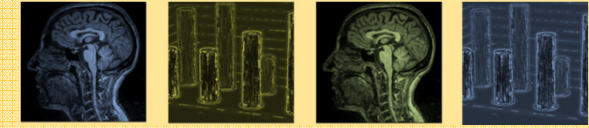
Q1: Os doentes de Alzheimer têm scores cognitivos diferentes da restante população? ($p=0.001$, $\alpha=0.05$)

Q2: À nascença, os bebés albinos têm um peso maior que os outros? ($p=0.362$, $\alpha=0.05$)

Q3: Há associação entre obesidade e risco cardíaco? ($p=0.09$, $\alpha=0.1$)

Q4: O batimento cardíaco é alterado durante um jogo de futebol? ($p=0.11$, $\alpha=0.1$)

Testes de hipóteses - exercício



Laboratório de Bioestatística e Informática Médica

- ✖ Cada uma das questões de investigação seguintes foi alvo de um estudo estatístico. Conhecendo os níveis de significância de cada estudo e os p-valores que foram obtidos, para cada estudo:
 - + identifique a hipótese nula
 - + indique se se deve ou não rejeitar a hipótese nula
 - + responda à questão de investigação

R1: H0: Os doentes de Alzheimer têm scores cognitivos iguais aos da restante população. A hipótese nula deve ser rejeitada: conclui-se que os doentes de Alzheimer têm scores cognitivos mais baixos que a restante população

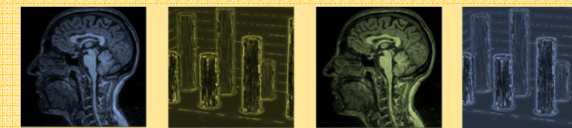
R2: H0: À nascença, os bebés albinos têm um peso igual aos outros. A hipótese nula não pode ser rejeitada: não se pode afirmar que os bebés albinos têm um peso maior que os outros

R3: H0: Não há associação entre obesidade e risco cardíaco. A hipótese nula deve ser rejeitada: conclui-se que há associação entre obesidade e risco cardíaco

R4: H0: O batimento cardíaco não é alterado durante um jogo de futebol. A hipótese nula não pode ser rejeitada (apesar de haver uma tendência de rejeição, por p estar próximo de α): não se pode concluir que o batimento cardíaco é alterado durante um jogo de futebol

- ✗ Estatística inferencial: testes de hipóteses
 - + Existem diferentes testes de hipóteses, mas todos têm em comum:
 - ✗ partir de uma hipótese (hipótese nula)
 - ✗ determinar um valor p. Se ele for mais baixo que o nível de significância α , rejeita-se a hipótese nula.
 - + Algumas questões:
 - ✗ Que testes de hipóteses há? Quais as hipóteses subjacentes?
 - ✗ Como escolher o teste correcto?
 - ✗ Como interpretar o resultado de um teste?
 - ✗ Qual deve ser o valor de α ?

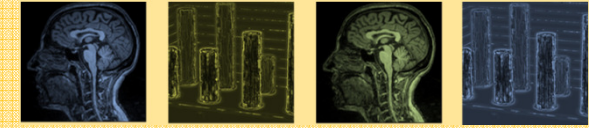
Testes de hipóteses (alguns...)



Laboratório de Bioestatística e Informática Médica

Número de grupos	Dependência	Testes paramétricos (média)	Testes não paramétricos (mediana)
1		T-student	Wilcoxon signed rank
2	Emparelhadas	T-student (emparelhadas)	Wilcoxon
	Independentes	T-student (independentes)	Mann-Whitney
3 ou mais	Emparelhadas	ANOVA de medidas repetidas	Friedman
	Independentes	ANOVA	Kruskal-Wallis

Testes de hipóteses (alguns...)



Laboratório de Bioestatística e Informática Médica

✕ Pontos importantes para escolha do teste de hipóteses:

+ Quantos grupos amostrais estão envolvidos? →

Queremos comparar a média de uma variável de um grupo com um determinado valor? Ou a média entre n grupos?

+ Há dependência entre os grupos? →

Existe informação mútua partilhada entre os grupos que possa ser reflectida na variável?

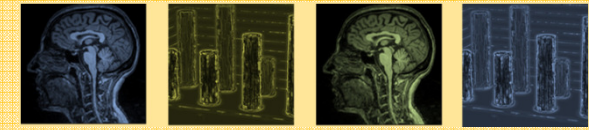
+ A variável que pretendemos comparar é qualitativa ou quantitativa? →

Os valores que a variável toma dividem os dados em categorias que se distinguem por características não numéricas ou são contagens/medições?

+ Se for quantitativa, a sua distribuição é normal (para cada um dos grupos envolvidos)? →

O histograma da variável, para cada grupo, tem uma forma geométrica aproximada a uma distribuição normal?

Testes de hipóteses (alguns...)



Laboratório de Bioestatística e Informática Médica

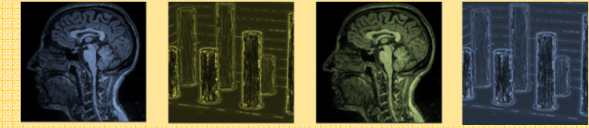
- ✗ Pretende-se comparar o valor médio das alturas dos portugueses e dos japoneses

Quantos grupos amostrais estão envolvidos?

2 grupos amostrais

Número de grupos	Dependência	Testes paramétricos (média)	Testes não paramétricos (mediana)
1		T-student	Wilcoxon signed rank
2	Emparelhadas	T-student (emparelhadas)	Wilcoxon
	Independentes	T-student (independentes)	Mann-Whitney
3 ou mais	Emparelhadas	ANOVA de medidas repetidas	Friedman
	Independentes	ANOVA	Kruskal-Wallis

Testes de hipóteses (alguns...)



Laboratório de Bioestatística e Informática Médica

- ✗ Pretende-se comparar o valor médio das alturas dos portugueses e dos japoneses

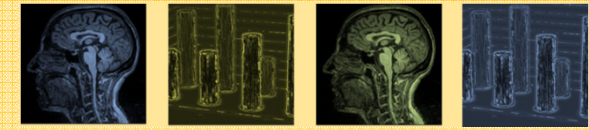
Há dependência entre os grupos?

2 grupos amostrais

Número de grupos	Dependência	Testes paramétricos (média)	Testes não paramétricos (mediana)
1		T-student	Wilcoxon signed rank
2	Emparelhadas	T-student (emparelhadas)	Wilcoxon
	Independentes	T-student (independentes)	Mann-Whitney
3 ou mais	Emparelhadas	ANOVA de medidas repetidas	Friedman
	Independentes	ANOVA	Kruskal-Wallis

independentes

Testes de hipóteses (alguns...)



Laboratório de Bioestatística e Informática Médica

- ✗ Pretende-se comparar o valor médio das alturas dos portugueses e dos japoneses

A variável é quantitativa e a sua distribuição é normal?

sim ↓

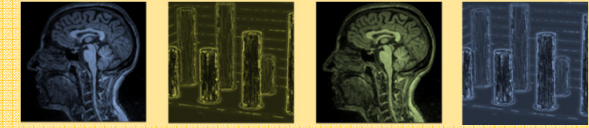
↓ não

2 grupos amostrais

Número de grupos	Dependência	Testes paramétricos (média)	Testes não paramétricos (mediana)
1		T-student	Wilcoxon signed rank
2	Emparelhadas	T-student (emparelhadas)	Wilcoxon
	Independentes	T-student (independentes)	Mann-Whitney
3 ou mais	Emparelhadas	ANOVA de medidas repetidas	Friedman
	Independentes	ANOVA	Kruskal-Wallis

independentes

Testes de hipóteses - exercício

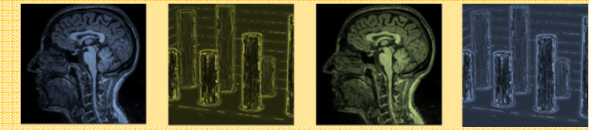


Laboratório de Bioestatística e Informática Médica

- ✗ Que teste de hipótese usaria para verificar se à nascença os bebês albinos têm um peso igual aos outros? Sabe-se que as distribuições de pesos dos bebês, albinos ou outros, são normais.

Número de grupos	Dependência	Testes paramétricos (média)	Testes não paramétricos (mediana)
1		T-student	Wilcoxon signed rank
2	Emparelhadas	T-student (emparelhadas)	Wilcoxon
	Independentes	T-student (independentes)	Mann-Whitney
3 ou mais	Emparelhadas	ANOVA de medidas repetidas	Friedman
	Independentes	ANOVA	Kruskal-Wallis

Testes de hipóteses - exercício

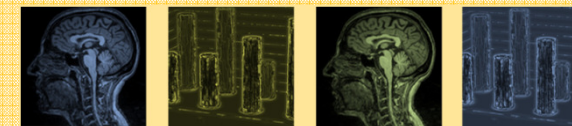


Laboratório de Bioestatística e Informática Médica

- ✗ O peso de pessoas obesas a tomar diariamente um suplemento comercial para redução de peso foi medido semanalmente durante 5 semanas, começando no dia da primeira toma. Sabendo que apenas na medição da segunda semana a distribuição dos pesos não foi normal, que teste estatístico usaria para averiguar se o peso dos pacientes foi alterado?

Número de grupos	Dependência	Testes paramétricos (média)	Testes não paramétricos (mediana)
1		T-student	Wilcoxon signed rank
2	Emparelhadas	T-student (emparelhadas)	Wilcoxon
	Independentes	T-student (independentes)	Mann-Whitney
3 ou mais	Emparelhadas	ANOVA de medidas repetidas	Friedman
	Independentes	ANOVA	Kruskal-Wallis

Tabela de testes



Laboratório de Bioestatística e Informática Médica

Variável Dependente	2 Grupos de Indivíduos diferentes	Mais de 2 Grupos de indivíduos diferentes	Antes - Depois (só 2 medições)	Antes - Depois (mais de 2 medidas repetidas)	Associação entre 2 variáveis	Associação com várias variáveis (contínuas ou nominais)
Quantitativa	Teste T de Student para a diferença entre duas médias	Anova de 1 factor	Teste t de Student para pares emparelhados	Anova medidas repetidas	Regressão/Correlação Linear simples (Pearson)	Regressão Linear Múltipla Anova 2 ou mais factores Ancova
Nominal	Regras de Cochran: - χ^2 - χ^2 corrigido - T.E.Fischer	χ^2	McNemar K index	Q de Cochran	Odds Ratio	Regressão Binária Logística
Ordinal	U de Mann-Whitney	Kruskal-Wallis	Wilcoxon matched-pairs	Friedman	Spearman	
Tempo até ao evento	Log Rank	Log-Rank			Cox-Regression	Cox-Regression