



## Introdução à análise estatística com SPSS

### Guião nº5: Comparação de médias de duas amostras emparelhadas

#### Investigação sobre a especialização do hemisfério cerebral

Numa certa experiência, a facilidade com que as palavras apresentadas no campo visual esquerdo e direito são reconhecidas é investigada. Foi dito aos participantes que, após um pequeno intervalo, uma palavra iria aparecer num dos campos visuais e tempo de que leva a reconhecê-la é guardado. As palavras apresentadas foram sorteadas e mostradas em ambos os campos visuais e os tempos medianos (em milissegundos) no reconhecimento das palavras estão na tabela 1. Será que existem diferenças nos tempos de reconhecimento médios de palavras entre o campo visual direito e o campo visual esquerdo?

**Tabela 1-** Resultados experimentais obtidos sobre os tempos medianos (mseg) no reconhecimento de palavras dos participantes.

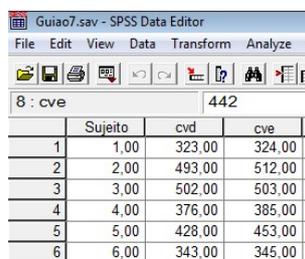
<b>Sujeitos</b>	<b>Campo visual direito</b>	<b>Campo visual esquerdo</b>	<b>Sujeitos</b>	<b>Campo visual direito</b>	<b>Campo visual esquerdo</b>
1	323	324	8	439	442
2	493	512	9	682	683
3	502	503	10	703	998
4	376	385	11	598	600
5	428	453	12	456	462
6	343	345	13	653	704
7	523	543	14	652	653

Identifique as variáveis e as hipóteses estatísticas em estudo.



## Abrir o SPSS e construir o ficheiro com os dados

Abrir o SPSS e depois a opção **type in data** e clicar na janela do **variable view** para criar as variáveis. Como o problema é de amostras emparelhadas é necessário criar três variáveis: a primeira variável designada por *sujeito*, na coluna do **label** indicará *elemento da amostra*; a segunda variável designada por *cvd* e na coluna do **label** indicará *campo visual direito*; a terceira variável designada por *cve* e na coluna do **label** indicará *campo visual esquerdo*. Na coluna **decimals**, colocar o valor 0 já que os dados não têm casas decimais. Depois da inserção dos dados na janela do **data view**, o resultado final deve ser como o apresentado na figura 1.



	Sujeito	cvd	cve
1	1,00	323,00	324,00
2	2,00	493,00	512,00
3	3,00	502,00	503,00
4	4,00	376,00	385,00
5	5,00	428,00	453,00
6	6,00	343,00	345,00

**Figura 1-** Representação dos 6 primeiros elementos e dos tempos no campo visual direito e no campo visual esquerdo, respectivamente.

## Explorar os dados

A primeira etapa é sempre examinar os dados e verificar se existe algum comportamento estranho.

O objectivo é criar uma tabela com médias e desvios-padrão, com a indicação sobre a distribuição dos dados através do histograma e da caixa de bigodes. Para o cálculo das médias e desvios-padrão, utiliza o comando **analyse/descriptive statistics /explore**. As variáveis dependentes (**dependent**) são o *cvd* e o *cve* e o **factor list** (variável independente que contém os grupos participantes) fica vazio. No quadro **label cases by** colocar a variável *sujeito*. Na opção **statistics**, clicar na opção **descriptives**. Na opção **plots**, desactivar o **stem-and leaf** e activar a opção **normality plots with tests**. Alterar a opção do boxplot de **factors levels together** para **dependents together**. Depois clicar no **continue** e no **ok**



Preencha a seguinte tabela:

**Tabela 2** - Resultados da estatística descritiva para variável tempos medianos (mseg) no reconhecimento de palavras.

Grupos	n	Média	Desvio- padrão	Mediana	Mínimo	Máximo
Campo visual direito						
Campo visual esquerdo						

Responda às seguintes questões:

- Acha que os resultados descritivos apoiam a hipótese científica?
- Da análise da caixa de bigodes, existem outliers e extremos? No caso afirmativo, que valores têm os outliers/extremos.
- Da análise dos testes de ajustamento, que pode concluir?

### **Aplicar o teste paramétrico t-student para diferença de médias, amostras emparelhadas**

Vamos aplicar o teste t-student para diferença de médias, amostras emparelhadas, sem eliminar os outliers e partir do princípio que os dados seguem uma distribuição normal.

Através do comando **analyse/compare means/paired-samples t test**, clicar nas variáveis *cvd* e *cve* de forma que o quadro designado por **current selections** fique preenchido. Depois clicar na seta de maneira que as variáveis aparecem no quadro **paired variables**. Clicar no **OK**.

Responda às seguintes questões:



- Escreva o valor do  $t$  (t), os graus de liberdade (**df**) e o valor do p-value (**sig 2-tailed**).
- Considerando os níveis de significância usuais, será que os investigadores têm razão? Justifique analisando os valores amostrais obtidos na tabela 2.
- Verifique que as variáveis *cvd* e *cve* estão correlacionadas.

### **Aplicar os testes não paramétricos: Wilcoxon e Sinais para diferença de medianas, amostras emparelhadas**

Os testes t-Student são influenciados pela existência de outliers e extremos e pela não normalidade das variáveis. Se não o removermos então deve-se utilizar um teste não paramétrico e analisar a influência dos outliers e extremos.

Através do comando **analyse/non parametric tests/ 2-related samples**, clicar nas variáveis *cvd* e *cve* de forma que o quadro designado por **current selections** fique preenchido. Depois clicar na seta de maneira que as variáveis aparecem no quadro **test pair list**. Clicar no teste dos Sinais (Sign, o Wilcoxon por default já está activado). Na opção **exact** escolher o método *exact*. Clicar no **Continue** e depois no **OK**.

A diferença entre o Wilcoxon e o Sinais está relacionada com a simetria ou não da variável que resulta do emparelhamento. Através do comando **transform/compute** iremos criar uma variável designada por *diff* que resulta da diferença entre *cvd* e *cve*. A nova variável é colocada no quadro **target variable** e no quadro **numeric expression** a condição *cvd-cve*. Clicar no **OK**.

Construir a caixa de bigodes através do comando **analyse/descriptive statistics /explore**. Colocar a variável *diff* no quadro das variáveis dependentes (**dependent**) e clicar no **ok**.

Responda às seguintes questões:



- A caixa de bigodes para a variável *diff* é simétrica? A variável *diff* segue uma distribuição Normal?
- O resultado do teste de Wilcoxon é idêntico ao do teste t?
- O resultado dos Sinais é idêntico ao teste do Wilcoxon? Justifique qual dos testes não paramétrico deve utilizar.

### Eliminação dos outliers/extremos

A selecção de dados é vista como a criação de um filtro e é realizada através do comando: **data/selected cases**. A seguir temos que clicar em **if condition is satisfied** e depois no **if**. No novo quadro, a condição que vamos impor é:  $diff > -50$  (ou seja, todos em que a diferença entre as duas variáveis é superior a -50) no painel direito. Depois é clicar no **Continue** e no **OK**. Sempre que se cria um filtro é necessário verificar se ficou bem feito. Verifique.

Repita as etapas da exploração dos dados e da aplicação dos testes e responda às seguintes questões:

- Da análise dos testes de ajustamento, que pode concluir?
- A conclusão relativa ao teste t continua válida (considere que a variável *diff* seguia uma distribuição normal)?
- As conclusões relativas aos testes do Wilcoxon e dos Sinais continuam válidas?

Num relatório ou artigo os resultados devem ser apresentados e escritos da seguinte forma:



**Tabela 3** - Resultados da estatística descritiva para variável tempos medianos (mseg) no reconhecimento de palavras.

Grupos	n	Média (M)	Desvio-padrão (DP)	Mediana (Med)	Amplitude inter- quartil (AIQ)
Campo visual direito (CVD)					
Campo visual esquerdo (CVE)					
CVD-CVE*					

\* teste dos sinais;  $p=$  (2-tailed)

Complete os espaços:

O tempo mediano no reconhecimento de palavras pelo campo visual direito (Med= ; AIQ= ) foi significativamente diferente do tempo médios do campo visual esquerdo (Med= ; AIQ= )  $p=$  (2-tailed);  $\alpha = 0.05$ . Este resultado confirma que o tempo para o reconhecimento de palavras é diferente consoante o campo visual.