Apontamentos

de

Estatística Aplicada à Psicologia II

Ana Paula Amorim

Universidade Minho
Departamento de Matemática
2019-2020

Conteúdo

1 Introdução

Este documento com o título Apontamentos de Estatística Aplicada à Psicologia II serve de suporte à Unidade Curricular (UC) de Estatística Aplicada à Psicologia II do Mestrado Integrado em Psicologia (MIPSI), da Universidade do Minho. Esta UC está inserida no primeiro semestre do terceiro ano do plano estudos do MIPSI e é uma continuação de outra UC, Estatística Aplicada à Psicologia I, lecionada no segundo semestre do segundo ano. Para além dos conteúdos programáticos da UC pretende-se também a familiarização dos alunos com o software IBM SPSS de extrema utilidade na análise de dados fundamental na área da Psicologia.

2 Iniciação ao software da IBM SPSS

2.1 Introdução

A análise estatística de dados realizada de uma forma adequada e válida pressupõe o conhecimento da inferência estatística e a usabilidade de um software estatístico. A nível do software a opção foi o *package* estatístico IBM SPSS. A escolha deste software deve-se a três razões: uma grande adesão na área da Psicologia, uma grande facilidade de utilização e à existência de um amplo conjunto de conteúdos implementados com sistema de ajuda acessível. Nesta secção serão apresentados alguns procedimentos com os respetivos comandos que serão utilizados ao longo desta unidade curricular.

2.2 Criação de ficheiros de dados

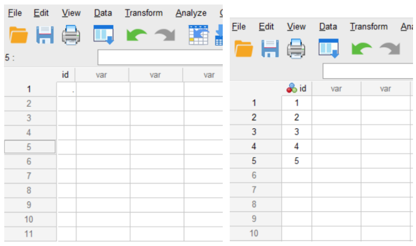
A criação do ficheiro de dados correspondente à tabela seguinte implica o conjunto de procedimentos que se descrevem a seguir.

Nome	Sexo	Idade	Altura	Tratamento
Rita Lima	Feminino	18	1,81	Medicamento
João Sousa	Masculino	17	1,90	Placebo
Daniel Silva	Masculino	19	1,85	Medicamento
Sofia Duarte	Feminino	18	1,68	Placebo
Rui Fernandes	Masculino	18	1,82	Placebo

Os dados são editados na janela do SPSS `Data Editor` e a sequência de comandos a seguir permite abrir um ficheiro novo. No menu principal escolher:

`File` → `New` → `Data`

As linhas do ficheiro representam registos distintos enquanto que as colunas representam as variáveis que se pretendem analisar. Os nomes das variáveis aparecem no início de cada coluna. A primeira variável do ficheiro de dados é chamada `Id` e corresponde ao número identificador (único) de cada caso introduzido no ficheiro.



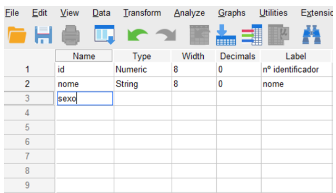
Antes de introduzir os dados da tabela exemplo temos de criar e definir as variáveis. Para criar uma nova variável temos de clicar duas vezes seguidas sobre um dos nomes `var`, que se encontra no cimo das colunas vazias da janela `Data View`, ou em alternativa abrir a janela `Variable View`, onde cada linha representa uma variável.

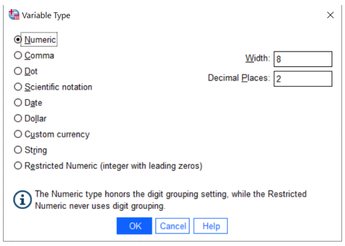
Na coluna `Name` da janela `Variable View` deve escrever-se o nome da variável `sexo` e carregando na tecla `Enter` aparecem todas as definições dessa variável atribuídas automaticamente pelo SPSS.

Para cada variável é necessário definir:

- `Name` - nome da variável deve começar por uma letra;
- `Type` - tipo de variável pode ser *Numeric* (tipo por defeito), *Comma, Dot, Scientific notation, Data, Dollar, Custom currency e String*;
- `Width` - tamanho;
- `Decimals` - número de casas decimais;
- `Label` - descrição da variável;

- `Values` - valores usados para identificar diferentes categorias das variáveis qualitativas;
- `Missing` - valores omissos que podem estar codificados, por exemplo, com o valor 999;
- `Columns` - largura da coluna;
- `Align` - alinhamento dos valores (*Center, Left, Right*);
- `Measure` - tipo de escala (*Scale , Ordinal e Nominal*).





As definições atribuídas de forma automática pelo IBM SPSS devem ser ajustadas ao tipo de valores da variável a introduzir. Assim, por exemplo, na coluna `Type` e na célula correspondente à variável ao clicar no canto direito abre uma caixa de diálogo `Variable Type` onde se deve selecionar a opção adequada ao tipo de dados da variável.

A variável `sexo` que se criou é uma variável tipo `String` mas sempre que possível deve-se codificar estes dados usando o tipo `Numeric` e atribuir um código numérico ao sexo masculino (0) e um código numérico ao sexo feminino (1), como indicado na tabela abaixo. A mesma estratégia de codificação deve ser aplicada à variável `tratamento` com duas categorias: medicamento (1) e placebo (2).

Variável Sexo	*Values*
Masculino	0
Feminino	1
Variável Tratamento	***Values***
Medicamento	1
Placebo	2

Neste tipo de variáveis é recomendado que na coluna `Decimals` seja selecionado o valor 0. Na coluna `Label` deve-se fazer uma breve descrição da variável. A atribuição de valores numéricos às categorias das variáveis `sexo` e `tratamento` é feita na coluna `Values` e `Value Label` como se oberva na figura acima. Para introduzir cada uma destas associações, clicar no botão `Add`.

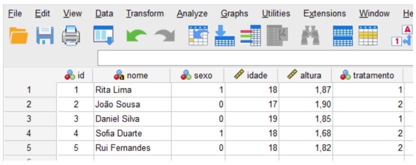
No IBM SPSS existem dois tipos de variáveis: quantitativas contínuas ou discretas (*Scale)* e qualitativas (*Nominal* ou *Ordinal*). As variáveis qualitativas, independentemente do seu tipo (`Nominal` ou `Ordinal`), são tratadas como categóricas na construção de tabelas e gráficos. No exemplo, da construção do ficheiro, as variáveis `sexo` e `tratamento` foram definidas como nominais.

A alteração do tipo de variável realiza-se na coluna `Measure` da janela `Variable View` optando pelo tipo adequado.

As variáveis quantitativas, `idade` e `altura`, foram definidas como `Numeric` (coluna `Type`) e `Scale` (coluna `Measure`).

Após a finalização da definição das variáveis introduzem-se no ficheiro os respetivos valores amostrais. Para as variáveis quantitativas registam-se os respetivos valores e nas variáveis qualitativas registam-se os valores numéricos das respetivas categorias a que pertencem.

No fim da introdução de toda a informação, o ficheiro terá a seguinte apresentação:



2.3 Abertura de ficheiros de dados

Os ficheiros de dados aparecem com uma grande variedade de formatos: EXCEL, tabelas CVS, SAS, SQLServer, $\cdots$, mas o software IBM SPSS está desenvolvido de maneira a permitir a utilização dos diferentes tipos de ficheiros de dados.

Procedimento para abrir um ficheiro de dados:

`Menu File` → `Open` → `Data`

Na caixa de diálogo `Open Data`, selecionar o ficheiro que se deseja abrir

Clicar em `Open`

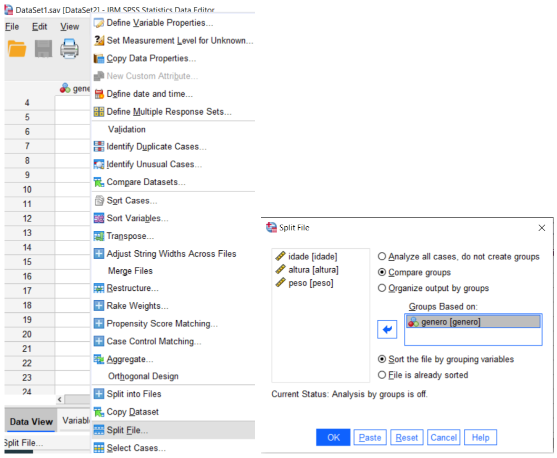
2.4 Manuseamento de ficheiros de dados

Para manusear ficheiros de dados deve-se executar as seguintes opções:

`Menu Data` → `Split File`

Esta opção permite dividir, de uma forma virtual, o ficheiro em diferentes ficheiros com base nas categorias de uma ou mais variáveis. A divisão do ficheiro feita deste modo é fundamental quando se pretende analisar apenas um subconjunto específico de dados como um novo ficheiro.

- `Split File` - janela com:
- lista de variáveis disponíveis no ficheiro;



- `Analyze all cases, do not create groups` - reagrupa os dados num único ficheiro;
- `Compare groups` - apresenta os resultados por grupo mas na mesma tabela;
- `Organize output by groups` - apresenta os resultados por grupo em diferentes tabelas;
- `Groups Based on:` - variável usada como critério de criação dos ficheiros.

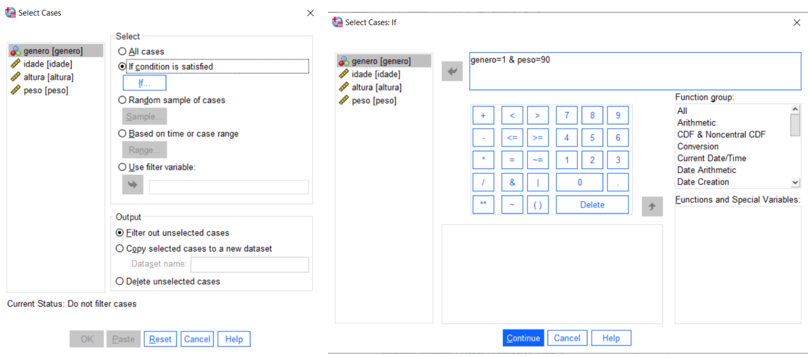
Para se guardar o ficheiro na versão original deve-se fazer os seguintes passos:

`Menu Data` → `Split File`

- Na janela `Split File` realizar as seguintes instruções:
- Em `Groups Based on` - enviar a variável selecionada anteriormente para a lista de variáveis disponíveis;
- Selecionar a opção `Analyze all cases, do not create groups` - que reagrupa os dados novamente num único ficheiro;

- `OK` - a divisão virtual do ficheiro de dados é anulada.

  Menu → `Data` → `Select Cases`



Esta opção permite selecionar os casos de acordo com um determinado critério.

Os principais critérios de seleção são:

- `All cases` - considera todo os casos;
- `If condition is satisfied` - permite definir o critério de seleção;
- `Filter out unselected cases` - filtra os casos não selecionados;
- `Delete unselected cases` - esta opção elimina os casos não selecionados;

Esta caixa de diálogo apresenta outras funcionalidades:

- Permite definir o critério de seleção através de uma condição;
- Grupo de funções preexistentes;
- Teclado que permite definir a condição de seleção;
- Lista de variáveis existentes no ficheiro.

Menu $\rightarrow$ `Transform` $\rightarrow$ `Compute Variable`

Esta opção vai permitir criar uma nova variável através de uma expressão numérica. A expressão numérica pode envolver uma ou mais variáveis existentes.

- `Target variable` - nome da variável a ser criada;
- `Numeric Expression` - expressão numérica a partir da qual será criada a nova variável;

A título de exemplo são apresentados, na tabela abaixo, os símbolos de uso mais frequente.

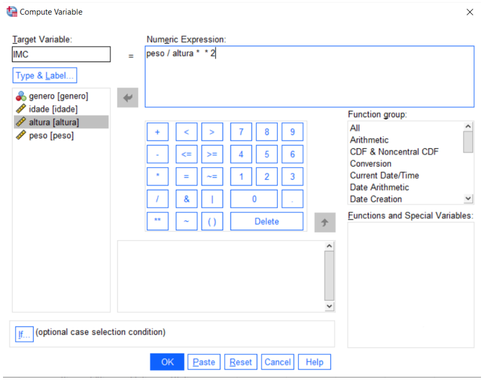
Símbolos	Significado
$\&$	Conjunção ($\wedge$)
$\vert$	Disjunção ($\vee$)
$\sim$	Negação
$\sim=$	Diferente ($\neq$)
$**$	Potência (ex: $4^2 = 4$**2)

A caixa de diálogo também disponibiliza:

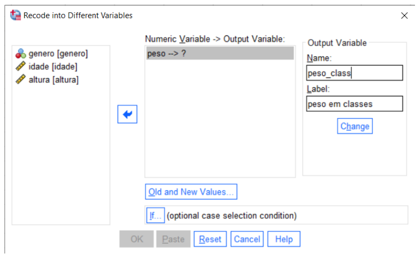
- Teclado que permite construir a expressão numérica;
- Lista de variáveis existentes no ficheiro;
- `Function group`: grupo de funções disponíveis.

Menu $\rightarrow$ `Transform` $\rightarrow$ `Recode into Different Variables`

Esta opção permite criar uma nova variável através da recodificação de uma variável existente no ficheiro de dados.



Para recodificar uma variável é necessário definir algumas regras segundo visualização da imagem:

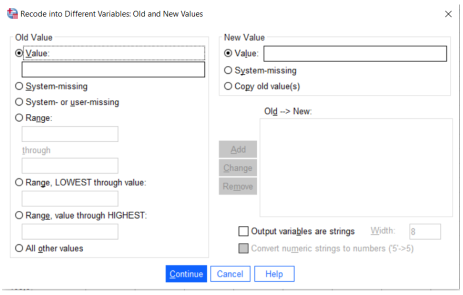


Na caixa de diálogo `Numeric Variable` → `Output Variable` é selecionada a `Variável original` → `Variável` a ser definida

Na caixa de diálogo `Output Variable`:

- `Name`: nome da variável a ser definida;
- `Label`: descrição da nova variável;
- `Change`: clicar após preenchimentos dos campos anteriores;
- `Old and New Values...`: ao clicar abre uma janela onde se definem as regras para a criação da nova variável.

Na caixa de diálogo `Recode into Different Variables: Old and`



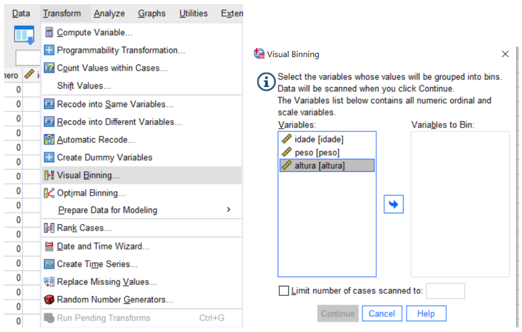
`New Values` o preenchimento das respetivas solicitações deve ser feito da seguinte maneira:

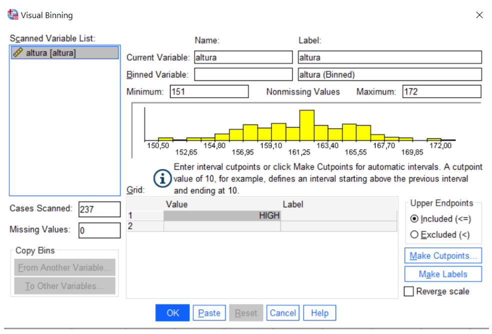
- `Old Value`: valor da variável original;
- `New Value`: valor da nova variável;
- `Value`: permite recodificar os valores um a um;
- `Range`: permite construir intervalos da forma $[a, b]$;
- `Range, LOWEST through value`: permite construir intervalos da forma $[min, b]$;
- `Range, value through HIGHEST`: permite construir intervalos da forma $[a, max]$;
- `Old` $\rightarrow$ `New`: caixa onde se escrevem as regras de codificação da nova variável.

Menu $\rightarrow$ `Transform` $\rightarrow$ `Visual Binning`

Esta opção do menu `Transform` permite agrupar uma variável quantitativa em classes (ou intervalos), criando para isso uma nova variável.

`Visual Binning`$\rightarrow$ `Variables`$\rightarrow$ `Variables to Bin` - da lista de variáveis disponíveis do ficheiro permite a seleção das variáveis quantitativas que se pretendem agrupar em classes.



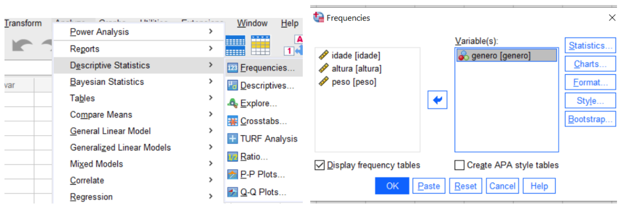


- `Scanned Variable List` - lista de variáveis cujos valores são agrupados em classes;
- `Binned Variable` - nome da nova variável a ser definida em classes;
- `Label` - descrição da nova variável (Label);
- `Minimum` - valor mínimo;
- `Maximum` - valor máximo;
- `Upper Endpoints` - define classes fechadas (`Included`) ou abertas à direita (`Excluded`);
- `Make Cutpoints` - permite definir as classes;
- `Make Labels` - atribuição de labels.

2.5 Cálculo de medidas estatísticas

No menu `Analyze` é possível construir tabelas de frequências, calcular algumas medidas estatísticas e construir alguns gráficos.

Menu → `Analyze` → `Descriptive Statistics` → `Frequencies`



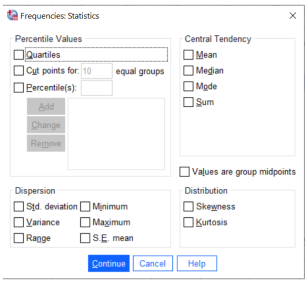
- `Frequency`: frequência absoluta;
- `Percent`: frequência relativa;
- `Valid Percent`: frequência relativa válida;
- `Cumulative Percent`: frequência relativa acumulada.

genero

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Feminino	111	46,8	46,8	46,8
	Masculino	126	53,2	53,2	100,0
	Total	237	100,0	100,0	

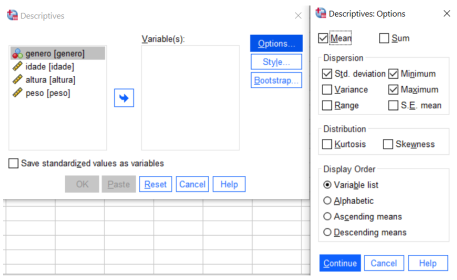
Menu → `Analyze` → `Descriptive Statistics` → `Frequencies` → `Statistics`

- `Percentile Values`: janela de percentis;
- `Central Tendency`: medidas de tendência central;
- `Dispersion`: medidas de dispersão;



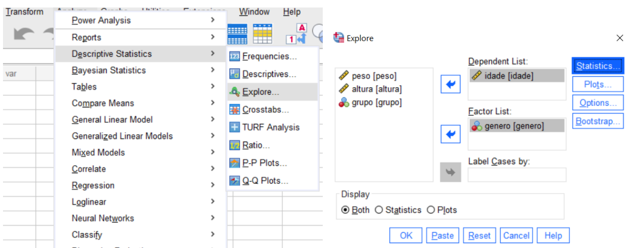
- `Distribution`: medidas de forma da distribuição.

Menu → `Analyze` → `Descriptive Statistics` → `Descriptives`



Menu → `Analyze` → `Descriptive Statistics` → `Explore`

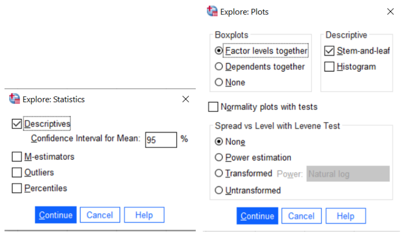
Com este comando é permitido calcular algumas medidas estatísticas e realizar alguns gráficos obtendo-se os resultados em separado para uma variável ou mais variáveis.



- `Dependent List` - conjunto de variáveis a analisar;

- `Factor List` - permite dividir os resultados em grupos;

- `Display` - opções de apresentação de resultados:`Both`- apresenta tudo; `Statistics` - apresenta as medidas estatísticas; `Plots` - apresenta os gráficos.



Na janela `Explore: Statistics`

- `Descriptives` - estatísticas descritivas; intervalo de confiança;

Na janela `Explore:Plots`

- `Boxplots` - caixa-com-bigodes;

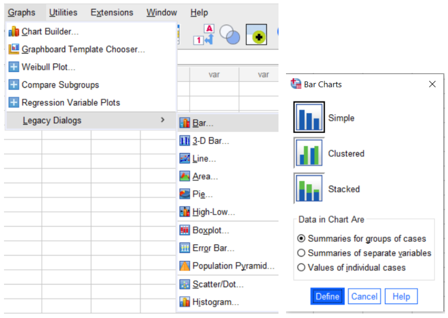
- `Descriptive` - diagrama de caule-e-folhas (`Stem-and-leaf`); histograma (`Histogram`).

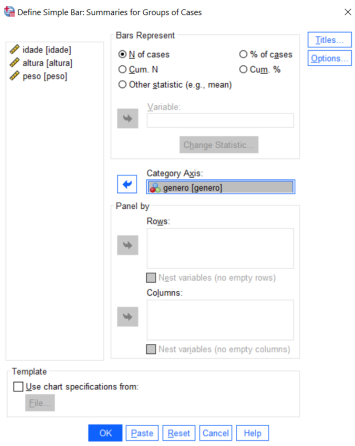
Descriptives

	genero			Statistic	Std. Error
dade	Feminino	Mean		16,44	,172
		95% Confidence Interval for Mean	Lower Bound	16,10	
			Upper Bound	16,78	
		5% Trimmed Mean		16,38	
		Median		16,40	
		Variance		3,286	
		Std. Deviation		1,813	
		Minimum		14	
		Maximum		21	
		Range		7	
		Interquartile Range		3	
		Skewness		,312	,229
		Kurtosis		-,943	,455
	Masculino	Mean		16,45	,167
		95% Confidence Interval for Mean	Lower Bound	16,11	
			Upper Bound	16,78	
		5% Trimmed Mean		16,33	
		Median		16,20	
		Variance		3,518	
		Std. Deviation		1,876	
		Minimum		14	
		Maximum		25	
		Range		11	
		Interquartile Range		3	
		Skewness		1,048	,216
		Kurtosis		2,193	,428

2.6 Construção de gráficos

Menu → `Graphs` → `Legacy Dialogs`→ `Bar`



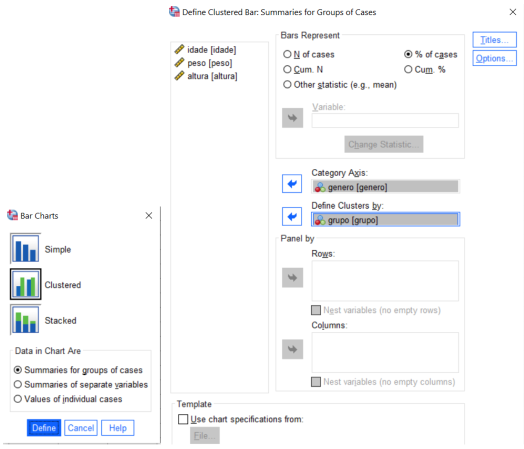


`Bar` → `Bar Charts - Simple` - gráfico para uma variável;

`Define Simple Bar: Summaries for Groups of Cases` - definições do gráfico; ao clicar abre uma janela onde surgem:

lista de variáveis disponíveis do ficheiro:

- `Bars Represent` - apresenta diferentes opções para o eixo dos YY nomeadamente a frequência absoluta (`N of cases`); a frequência relativa (`% of cases`) entre outras;
- `Category Axis` - variável a representar no gráfico.

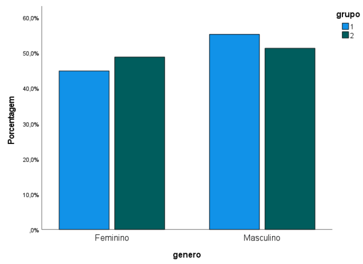


`Bar` → `Bar Charts - Clustered` - gráfico para duas variáveis;

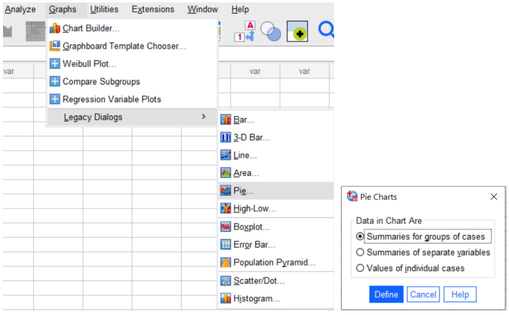
`Define clustered Bar: Summaries for Groups of Cases` - definições do gráfico; abre uma janela com

lista de variáveis disponíveis do ficheiro:

- `Bars Represent` - apresenta diferentes opções para o eixo dos YY nomeadamente a frequência absoluta (`N of cases`); a frequência relativa (`% of cases`) entre outras;
- `Category Axis` - variável a representar no gráfico;
- `Define Clusters by` - variável a representar na legenda do gráfico.

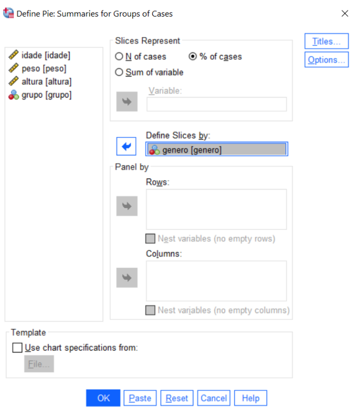


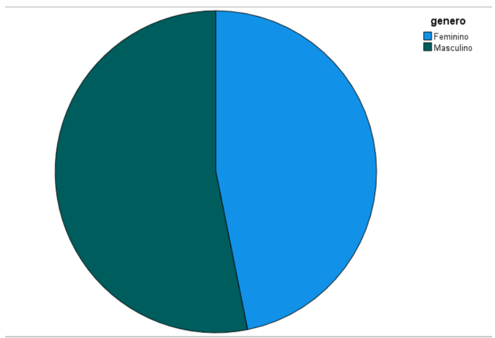
Menu → `Graphs` → `Legacy Dialogs` → `Pie`



- Pie Charts → Summaries for groups of cases
- Define - definições do gráfico na janela Define Pie: Summaries for Groups of Cases:

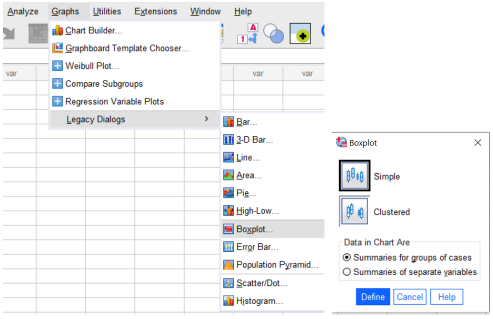
  Lista de variáveis disponíveis no ficheiro:
- Slices Represent → % of cases - frequência relativa;
- Define Slices by: variável a representar no gráfico.





Menu → Graphs → Legacy Dialogs→ Boxplot

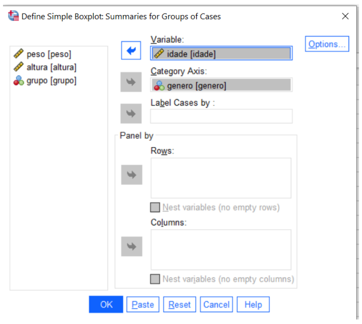
- Boxplot→ Simple - gráfico para uma variável;

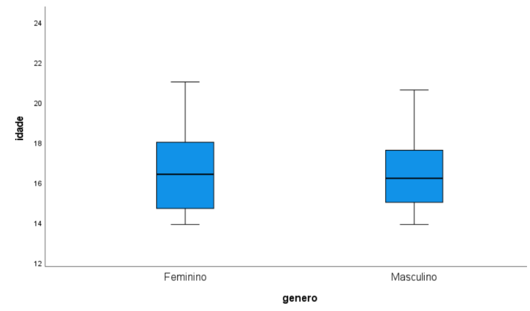


- `Define` - definições do gráfico na janela `Define Simple Boxplot: Summaries for Groups of Cases`

  lista de variáveis disponíveis no ficheiro:

- `Variable:` variável a representar no eixo vertical do gráfico;
- `Category Axis:` variável a representar no eixo horizontal do gráfico.

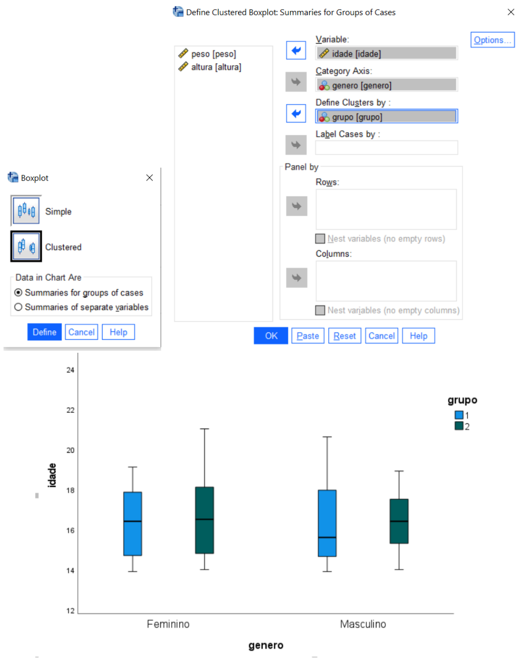




- `Boxplot`$\rightarrow$ `Clustered` - gráfico para duas variáveis;
- `Define` - definições do gráfico na janela `Define Clustered Boxplot: Summaries for Groups of Cases`

  lista de variáveis disponíveis no ficheiro:

- `Variable:` variável a representar no eixo vertical do gráfico;
- `Category Axis:` variável a representar no eixo horizontal do gráfico;



- `Define Clusters by:` variável que aparece na legenda do gráfico.

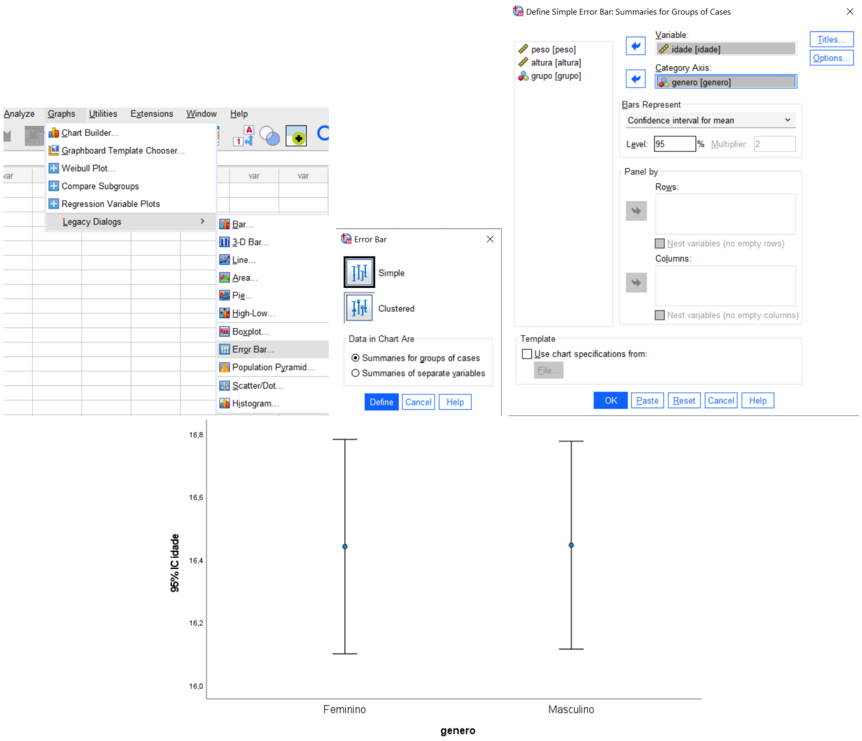
Menu → `Graphs` → `Legacy Dialogs`→ `Error Bar`

- `Error Bar`→ `Simple` - gráfico para uma variável;
- `Define` - definições do gráfico na janela `Define Simple Error Bar: Summaries for Groups of Cases`

  Lista de variáveis disponíveis no ficheiro:

- `Variable:` variável a representar no eixo vertical do gráfico;
- `Category Axis:` variável a representar no eixo horizontal do gráfico.

- `Error Bar`→ `Clustered` - gráfico para duas variáveis;
- `Define` - definições do gráfico na janela-`Define Clustered Error Bar: Summaries for Groups of Cases`



Lista de variáveis disponíveis no ficheiro:

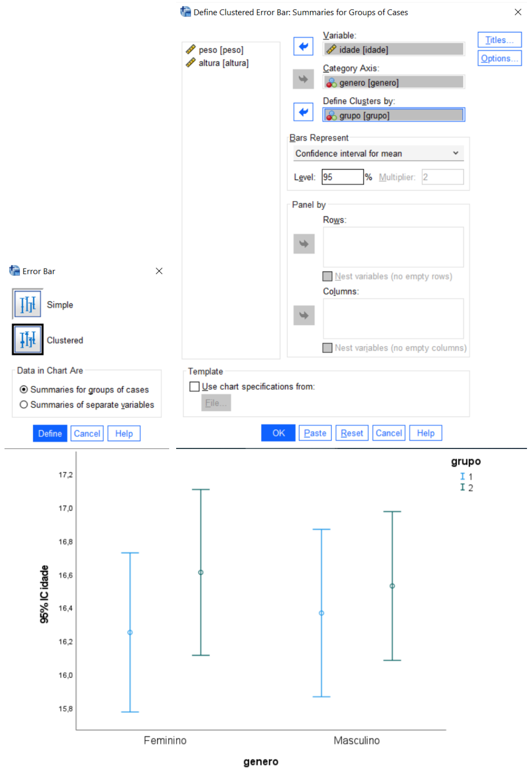
- `Variable:` variável a representar no eixo vertical do gráfico;
- `Category Axis:` variável a representar no eixo horizontal do gráfico;
- `Define Clusters by:` variável que aparece na legenda do gráfico.

Menu → `Graphs` → `Legacy Dialogs`→ `Scatter/Dot`

- `Scatter/Dot`→ `Simple Scatter` - gráfico para duas variáveis;
- `Define` - definições do gráfico na janela `Simple Scatterplot`

Lista de variáveis disponíveis no ficheiro:

- `Y Axis:` variável a representar no eixo vertical do gráfico;



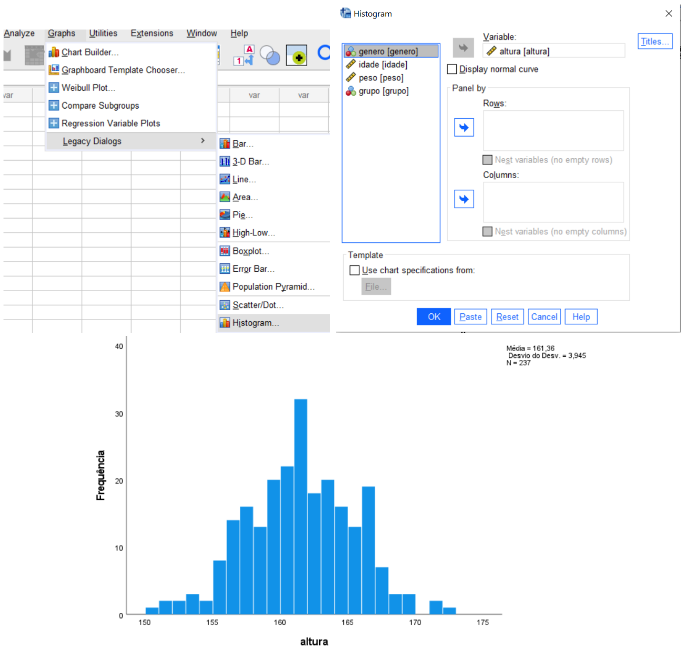
- `X Axis:` variável a representar no eixo horizontal do gráfico.

Menu $\rightarrow$ `Graphs` $\rightarrow$ `Legacy Dialogs`$\rightarrow$ `Histogram`

Lista de variáveis disponíveis no ficheiro:

- `Variable:` variável a representar no gráfico;
- `Titles:` permite escrever títulos no gráfico.

Graphs
Utilities
Extensions
Window
Help
Chart Builder...
Graphboard Template Chooser...
Weibull Plot...
Compare Subgroups
Regression Variable Plots
Legacy Dialogs
Bar...
3-D Bar...
Line...
Area...
Pie...
High-Low...
Boxplot...
Error Bar...
Population Pyramid...
Scatter/Dot...
Histogram...
Scatter/Dot
Simple Scatter
Matrix Scatter
Simple Dot
Overlay Scatter
3-D Scatter
Define
Cancel
Help
Simple Scatterplot
genero [genero]
idade [idade]
grupo [grupo]
Y Axis:
peso [peso]
X Axis:
altura [altura]
Set Markers by:
Label Cases by:
Panel by
Rows:
Nest variables (no empty rows)
Columns:
Nest variables (no empty columns)
Template
Use chart specifications from:
File...
Titles...
Options...
OK
Paste
Reset
Cancel
Help
peso
altura



Analyze
Graphs
Utilities
Extensions
Window
Help
Chart Builder...
Graphboard Template Chooser...
Weibull Plot...
Compare Subgroups
Regression Variable Plots
Legacy Dialogs
Bar...
3-D Bar...
Line...
Area...
Pie...
High-Low...
Boxplot...
Error Bar...
Population Pyramid...
Scatter/Dot...
Histogram...
Histogram
genero [genero]
idade [idade]
peso [peso]
grupo [grupo]
Variable:
altura [altura]
Display normal curve
Panel by
Rows:
Nest variables (no empty rows)
Columns:
Nest variables (no empty columns)
Template
Use chart specifications from:
File...
Titles...
OK
Paste
Reset
Cancel
Help
Média = 161,36
Desvio do Desv. = 3,945
N = 237
Frequência
altura

3 Relembrar alguns testes estatísticos

3.1 Testes de Kolmogorov-Smirnov e de Shapiro-Wilk

O teste de Kolmogorov-Smirnov (K-S) é um teste que se aplica a uma amostra para averiguar se ela provém de uma população contínua (por exemplo, a distribuição normal) com distribuição especificada na hipótese nula.

Este teste baseia-se na comparação da função de distribuição empírica (F_n) com a função de distribuição teórica (F) indicada na hipótese nula. A função de distribuição empírica define-se como a proporção das observações da amostra que são menores ou iguais a x para todos os valores reais x. Graficamente, o teste K-S determina o ponto onde se verifica a maior distância vertical entre as duas funções (empírica e teórica).

1. $F(x)$ representa a função distribuição assumida pelos dados na hipótese H_0;
2. $F_n(x)$ representa a função distribuição empírica dos dados;
3. As observações são ordenadas de forma crescente $(x_{(1)}, x_{(2)}, \cdots, x_{(n)})$;
4. $F(x_{(i)}) = P(X \leq x_{(i)})$;
5. $F_n(x) = \dfrac{1}{n}\sum_{i=1}^{n} I_{\{(-\infty,x]\}}(x_{(i)}) = \begin{cases} 0, \text{se } x < x_{(1)} \\ \dfrac{k}{n}, \text{se } x_{(k)} \leq x < x_{(k+1)} \\ 1, \text{se } x > x_{(n)} \end{cases}$

   A função indicatriz I_A, é definida da seguinte forma: $I_A = \begin{cases} 1, & \text{se} & x \in A \\ 0, & \text{se} & \text{c.c.} \end{cases}$

Como a função de distribuição empírica é descontínua e a função de distribuição hipotética é contínua, vamos considerar duas estatísticas:

$$D^+ = \sup_{x_{(i)} \in \mathbb{R}} |F(x_{(i)}) - F_n(x_{(i)})|$$

$$D^- = \sup_{x_{(i)} \in \mathbb{R}} |F(x_{(i)}) - F_n(x_{(i-1)})|$$

A estatística de teste é dada por:

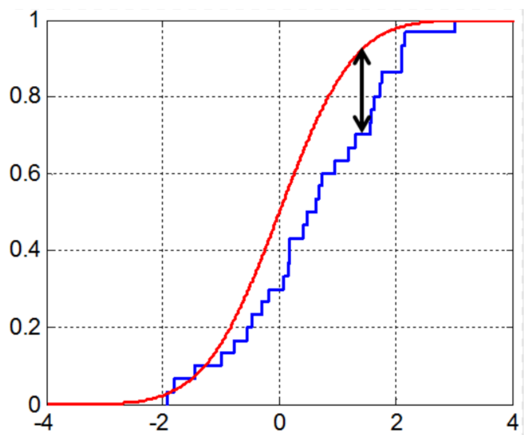
$$D_n = \max(D^+, D^-)$$

As hipóteses a testar são:

$$H_0 : F(x) = F_0(x), \quad \text{para todo o x} \quad \text{vs} \quad H_1 : F(x) \neq F_0(x) \quad \text{para algum x}$$

Rejeita-se a hipótese de ajustamento para valores muito grandes da estatística de teste, pois significam um grande afastamento entre F e F_n, ou seja os dados amostrais não apresentam uma distribuição semelhante à da função $F(x)$. Para um dado n e nível de significância α, rejeita-se a hipótese nula quando p-value $< \alpha$.

Na figura abaixo encontra-se a representação gráfica das funções de distribuição $F(x)$ (linha vermelho) e $F_n(x)$ (linha azul), para os valores amostrais x.



O teste K-S, no software IBM SPSS, encontra-se implementado para as distribuições Normal, Exponencial e Uniforme. Apesar de ser um teste direcionado às distribuições contínuas também se encontra implementado para a distribuição discreta de Poisson. Este teste encontra-se implementado na versão teste exato, para amostras de pequena dimensão, e na versão teste assintótico para grandes amostras.

Para uma amostra o teste de K-S, no software IBM SPSS, tem a seguinte sequência de passos:

`Analyze` → `Nonparametric Tests` → `Legacy Dialogs` → `1-Sample K-S`

Abre uma caixa de diálogo: `One-Sample Kolmogorov-Smirnov Test`

- `Test Variable List` - seleção das variáveis a testar
- `Test Distribution` - distribuição de teste: `Normal`, `Uniforme`, `Exponencial` e `Poisson`.
- `OK`

Importa referir que o teste de K-S pressupõe que os parâmetros, μ e σ, da distribuição Normal sejam conhecidos. No caso dos parâmetros não serem conhecidos, devem ser estimados a partir dos dados da amostra, o que conduz a uma tendência para não rejeitar a hipótese nula um maior número de vezes do que o desejável. Para contornar este problema foi implementada uma correção ao teste de K-S, designada por correção de Lillefors.

Esta correção está disponível no SPSS mas no menu `Analyze` → `Descriptive Statistics` → `Explore`, selecionar o botão `Plots` e ativar a opção `Normality plots with tests`. Com esta opção o SPSS fornece o `QQ-gráfico Normal` para a amostra e uma tabela resumo dos testes de ajustamento. Na tabela, à esquerda encontra-se o teste de `Kolmogorov-Smirnov` com a correção de Lillefors e à direita encontra-se o teste de `Shapiro-Wilk`.

Dos dois testes, o de `Shapiro-Wilk` é mais indicado para amostras de pequena dimensão e o teste de `Kolmogorov-Smirnov` com a correção de Lillefors para amostras de grande dimensão, $(n \geq 30)$. A regra de decisão para estes testes de ajustamento consiste em não rejeitar H_0 quando p-value $\geq \alpha$, sendo α o nível de significância do teste.

3.2 Teste de aleatoriedade de sequências

O teste de aleatoriedade de sequências é um teste usado para verificar se uma amostra foi obtida através de um processo aleatório. Dada uma amostra $(x_1, x_2, \cdots, x_n)$ pretende-se saber se estes valores foram obtidos aleatoriamente ou se apresentam algum tipo de tendência.

As hipóteses de teste são definidas como:

$$H_0 : \text{a amostra é aleatória} \quad \text{vs} \quad H_1 : \text{a amostra não é aleatória}$$

O teste aplica-se a variáveis dicotómicas (0 e 1) ou a variáveis dicotomizadas e faz a análise de sequências dos dois tipos de símbolos (0 e 1) na amostra.

Quando a amostra resulta da observação de uma variável quantitativa, é necessário fixar um ponto de corte (média, mediana, moda ou personalizado) e atribuir um dos símbolos (0 ou 1) a todas as observações, consoante o valor da amostra fique abaixo ou acima do valor do ponto de corte.

O teste considera as diferentes sequências de símbolos iguais. Estas sequências podem ser unitárias, ou seja constituídas por um único símbolo, ou serem sequências com um número de símbolos iguais superior a um.

Situações em que na amostra aparecem muito poucas ou demasiadas sequências iguais, podem indicar a presença de uma tendência ao longo do tempo ou de uma agregação de acontecimentos semelhantes, indicando que a amostra não é aleatória.

Como exemplificação considere-se um teste de avaliação de imagens cujo resultado é uma sequência de símbolos 0 e 1 (0 - não identificou a imagem; 1 -

identificou a imagem) apresentado a seguir:

1 0 1 0 0 1 0 1 1 1 0 1 1 0 1 0 0 1 1 1 0

As subamostras para os símbolos 0 e 1 têm dimensões respetivamente $n_0 = 10$ e $n_1 = 12$.

O número total de sequências da amostra, separadas pelo hífen, é de catorze.

$$\mathbf{1 - 0 - 1 - 00 - 1 - 0 - 111 - 0 - 11 - 0 - 1 - 00 - 111 - 00}$$

A estatística de teste não é mais do que o total de sequências, R (runs), que se observa na amostra.

A hipótese nula é rejeitada para valores grandes ou valores pequenos de R. O teste encontra-se implementado no software IBM SPSS nas versões de teste exato (com a estatística R) e teste assintótico (com a estatística Z).

A hipótese nula é rejeitada para valores grandes ou pequenos de R, para a estatística assintótica Z, a região de rejeição ou região crítica (RC) associada ao nível de significância α é dada por:

$RC = \{z(x_1, \cdots, x_n) : z < z_{\alpha/2} \quad \text{ou} \quad z > z_{1-\alpha/2}\}$

O output deste teste no IBM SPSS apresenta uma tabela com a seguinte informação:

- `Test Value` - ponto de corte utilizado na realização do teste;
- `Cases < Test Value` - número de observações inferiores ($<$);
- `Cases >= Test Value` - número de observações superiores ou iguais ($>=$) ao ponto de corte;
- `Total Cases` - número total de casos;
- `Number of Runs` - número de sequências (R);

- `Z` - valor da estatística de teste R com aproximação à Normal;
- `Asymp. Sig. (2-tailed)`: p-value (Sig.) assintótico (2 extremidades);
- `Exact Sig. (2-tailed)`: p-value (Sig.) exato (2 extremidades);
- `Point Probability`: probabilidade pontual de se observar o valor $R = r_{obs}$.

3.3 Teste de Levene

O teste de Levene é um teste usado para avaliar se k populações apresentam variâncias iguais ($\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$), i.e., se as populações são homocedásticas. As hipóteses a testar são:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

vs

$$H_1 : \exists i, j : \sigma_i^2 \neq \sigma_j^2 \;\; (i \neq j; i, j = 1, \cdots, k)$$

Sendo o teste de Levene um dos testes mais potentes para fazer a verificação da homogeneidade de variâncias o software IBM SPSS optou pela sua única implementação.

Para k populações o teste Levene surge no output do teste paramétrico de comparação de k médias populacionais.

No menu: `Analyze` $\rightarrow$ `Compare Means` $\rightarrow$ `Independent-Samples T Test` para duas populações;

`Analyze` $\rightarrow$ `Compare Means` $\rightarrow$ `One-Way ANOVA` para mais de duas populações;

4 Testes paramétricos

4.1 Introdução

Neste tópico são abordados os testes de hipóteses paramétricos para a diferença de valores médios em duas ou mais populações normais, nos casos seguintes:

- Teste t para duas amostras independentes (homogeneidade de variâncias);
- Teste t para duas amostras independentes (heterogeneidade de variâncias);
- Teste t para duas amostras emparelhadas;
- Análise de Variância (ANOVA) um fator para três ou mais amostras independentes;
- Análise de Variância (ANOVA) para medidas repetidas.

4.2 Recordar alguns conceitos

Ao conjunto de valores amostrais que levam à rejeição da hipótese nula, chama-se região crítica (RC) do teste. Este conjunto é geralmente definido através de uma estatística designada como estatística de teste, T. A estatística de teste quantifica a "distância"dos dados em relação à hipótese H_0.

$$\text{"Rejeita-se} \quad H_0 \quad \text{se e só se} \quad (x_1, \ldots, x_n) \in RC\text{"}$$

$$\Updownarrow$$

$$\text{"Rejeita-se} \quad H_0 \quad \text{se e só se} \quad T \in RC\text{"}$$

Na formulação das hipóteses pretende-se que a hipótese nula contenha sempre uma igualdade e a hipótese alternativa contenha sempre uma desigualdade expressa pela utilização de um dos seguintes símbolos $>$; $<$ ou $\neq$. O símbolo usado na hipótese alternativa está relacionado com a designação do teste: teste unilateral à direita ($>$), teste unilateral à esquerda ($<$) e teste bilateral ($\neq$).

A hipótese nula é considerada verdadeira ao longo do procedimento de teste até ao momento em que haja evidência clara apontando em sentido contrário e nesse caso rejeita-se então a hipótese nula.

Uma decisão baseada num teste tem associado sempre um risco de erro, que pode acontecer de dois modos distintos: (i) se o teste levar à rejeição de H_0 e H_0 for verdadeira. Este erro diz-se erro tipo I ou de primeira espécie, e a probabilidade de o cometer denota-se por α e chama-se nível de significância do teste, isto é:

$$\alpha = P((X_1, \ldots, X_n) \in RC | H_0 \text{ é verdadeira})$$

(ii) o teste pode levar à não rejeição de H_0 e H_0 ser falsa. Este erro diz-se erro tipo II ou de segunda espécie, e a probabilidade de o cometer denota-se por β.

Regras de decisão de um teste de hipóteses:

- com a região crítica: consiste em fixar o nível de significância α do teste e determinar a região crítica associada ao teste. Se o valor observado da estatística de teste T pertencer à região crítica então rejeitamos H_0, caso contrário não rejeitamos.

- com valor-p (*p-value*): o valor-p representa a probabilidade de se observar um valor da estatística de teste T, tão ou mais afastado do que o que foi observado, supondo H_0 verdadeira. Se o valor-p for menor ou igual ao nível de significância α do teste rejeitamos H_0.

4.3 Teste t para duas amostras independentes

Nos testes de diferença de valores médios em populações normais, consideram-se duas amostras aleatórias e independentes de dimensão n e m:

$(X_1, \ldots, X_n)$ de dimensão n, tal que $X \frown N(\mu_X, \sigma_X)$,

$(Y_1, \ldots, Y_m)$ de dimensão m, tal que $Y \frown N(\mu_Y, \sigma_Y)$

as variâncias, σ_X^2 e σ_Y^2, são desconhecidas mas consideradas iguais ou desconhecidas mas consideradas diferentes.

A aplicação destes testes exige que os seguintes pressupostos sejam verificados: as duas amostras são aleatórias e independentes provenientes de duas populações normais $N(\mu_X, \sigma_X)$ e $N(\mu_Y, \sigma_Y)$. Como as variâncias populacionais são desconhecidas mas consideradas iguais ou desconhecidas e consideradas diferentes, originam testes distintos para cada um dos casos.

A escolha do par de hipóteses a testar será um dos seguintes casos:

$$H_0 : \mu_X - \mu_Y = 0 \quad vs \quad H_1 : \mu_X - \mu_Y \neq 0$$

$$H_0 : \mu_X - \mu_Y = 0 \quad vs \quad H_1 : \mu_X - \mu_Y < 0$$

$$H_0 : \mu_X - \mu_Y = 0 \quad vs \quad H_1 : \mu_X - \mu_Y > 0$$

Para as amostras de X e de Y as médias amostrais são designadas por $\overline{X}$ e $\overline{Y}$ e as variâncias amostrais representadas por S_X^2 e S_Y^2, respetivamente.

Caso A - Teste para a diferença de valores médios, μ_X e μ_Y, em populações normais, considerando variâncias desconhecidas, mas supostamente iguais (homogeneidade de variâncias).

A estatística de teste que pressupõe que as variâncias σ_X^2 e σ_Y^2 são desconhecidas mas iguais, é dada por:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{(n+m-2)}}} \frown t_{n+m-2} \quad (1.1)$$

Ao nível de significância $\alpha \in [0, 1]$, as regiões críticas e *p-value* estão resumidas no quadro seguinte:

H_1	região crítica	p-value
$\mu_X - \mu_Y \neq 0$	$\lvert t\rvert > t_{n+m-2,1-\alpha/2}$	$2P(T \geq \lvert t_{\text{obs}}\rvert \mid H_0)$
$\mu_X - \mu_Y < 0$	$t < t_{n+m-2,\alpha}$	$P(T \leq t_{\text{obs}} \mid H_0)$
$\mu_X - \mu_Y > 0$	$t > t_{n+m-2,1-\alpha}$	$P(T \geq t_{\text{obs}} \mid H_0)$

o valor $t_{n-1,1-\alpha}$ representa o quantil $1-\alpha$ da distribuição t-Student com (n-1) graus de liberdade.

Caso B-Teste para a diferença de valores médios, $\mu_X - \mu_Y$, em populações normais independentes com variâncias desconhecidas mas diferentes.

A estatística de teste que pressupõe que as variâncias σ_X^2 e σ_Y^2 são desconhecidas mas diferentes, é dada por:

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{n\,m}{n+m}}} \approx t_\nu \quad (1.2)$$

Nota: os graus de liberdade ν da distribuição t-Student (t_ν) são obtidos por aproximação pela fórmula de Welch.

Ao nível de significância $\alpha \in [0,1]$, as regiões críticas e *p-value* estão resumidas no quadro seguinte:

H_1	região crítica	p-value
$\mu_X - \mu_Y \neq 0$	$\lvert t\rvert > t_{\nu,1-\alpha/2}$	$2P(T \geq \lvert t_{\text{obs}}\rvert \mid H_0)$
$\mu_X - \mu_Y < 0$	$t < t_{\nu,\alpha}$	$P(T \leq t_{\text{obs}} \mid H_0)$
$\mu_X - \mu_Y > 0$	$t > t_{\nu,1-\alpha}$	$P(T \geq t_{\text{obs}} \mid H_0)$

No software da IBM SPSS estes testes encontram-se no menu: *Analyze* $\rightarrow$ *Compare Means* $\rightarrow$ *Independent$-$Samples T Test* .

No output destes teste aparece o teste de Levene para avaliar a igualdade de variâncias das populações, considerando as seguintes hipóteses:

$$H_0 : \sigma_X^2 = \sigma_Y^2 \text{ vs } H_1 : \sigma_X^2 \neq \sigma_Y^2$$

A interpretação do output do teste de Levene foi realizada na secção 3.

Os dados das duas amostras a considerar devem estar numa só coluna devendo existir no ficheiro de dados uma outra coluna que identifique a amostra de origem de cada observação. No caso de não existir a variável identificadora da amostra de origem, as amostras podem ser discriminadas por uma variável dicotómica ou por um valor de corte (*cut point*) que divide os dados como menores ou iguais e superiores ao valor de corte, se este for o caso indicado para o estudo a realizar.

No output do software aparecem dois tipos de resultados, um assumindo a igualdade de variâncias e outro que não admite a igualdade de variâncias, para cada um dos casos são utilizadas as estatísticas de teste adequadas e apresentadas anteriormente.

Exemplo: Uma auditoria externa a uma empresa gerou uma onda de stress entre os colaboradores da mesma. Pretende-se avaliar se o stress médio sentido 7 dias antes da auditoria é igual para os dois géneros?

O output do teste de normalidade (K-S) da variável stress 7 dias antes da auditoria para ambas as amostras de género, permite concluir que as amostras são provenientes de populações normais (Homens - estatística=0.088 e *p-value*=0.200; Mulheres - estatística=0.097 e *p-value*=0.200).

Na tabela da estatística descritiva observa-se que:

- `N` - dimensões das amostras de X e Y respetivamente $n = 38$ e $m = 42$;

Tests of Normality

	Sexo	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Stress - 7 dias antes da auditoria (maior valor, maior stress)	Masculino	,088	38	,200*	,976	38	,569
	Feminino	,097	42	,200*	,974	42	,458

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Group Statistics

	Sexo	N	Mean	Std. Deviation	Std. Error Mean
Stress - 7 dias antes da auditoria (maior valor, maior stress)	Masculino	38	24,92	8,079	1,311
	Feminino	42	25,86	8,739	1,348

- `Mean` - médias de stress 7 dias antes da auditoria $\overline{x} = 24.92$ e $\overline{y} = 25.86$;
- `Std. Deviation` - desvios padrões corrigidos $s_X = 8.079$ e $s_Y = 8.739$;
- `Std. Error Mean` - estimativa dos desvios padrões das médias amostrais $s_X/\sqrt{n} = 1.311$ e $s_Y/\sqrt{m} = 1.348$.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Stress - 7 dias antes da auditoria (maior valor, maior stress)	Equal variances assumed	,624	,432	-,496	78	,621	-,936	1,888	-4,695	2,822
	Equal variances not assumed			-,498	77,960	,620	-,936	1,880	-4,680	2,807

Na tabela podemos observar: dois conjuntos de resultados, um assumindo a igualdade de variâncias (primeira linha da tabela) e baseado na estatística (1.1) e outro (segunda linha da tabela) que admite a não igualdade de variâncias e cujos resultados foram obtidos através da estatística (1.2).

Nas duas primeiras colunas da primeira linha da tabela, apresenta-se o resultado do teste de Levene, que permite assumir a igualdade de variâncias para o stress nos dois géneros (F(0.624), *p-value*=0.432).

Assumindo a igualdade de variâncias: o campo **t** contém o valor observado da estatística de teste T definida em (1.1); o campo **df** (*degrees of freedom*) contém o número de graus de liberdade; o campo **Sig. (2-tailed)** contém o *p-value* do teste bilateral $2 * P(T > |t_{obs}|)$; o campo **Mean Difference** contém a diferença entre as médias; **Std. Error Difference** é o denominador da estatística de teste (1.1); os campos **95% Confidence Interval of the Difference; Lower and Upper** contêm os extremos do intervalo de confiança 95 % para a diferença de valores médios, $\mu_X - \mu_Y$.

Em relação aos resultados da última tabela conclui-se que a média de stress 7 dias antes da auditoria é idêntica para os dois géneros (t(-0.496; 78), *p-value*=0.621). Esta conclusão é confirmada pelo intervalo de confiança de 95% para a diferença de valores médios de stress 7 dias antes da auditoria, com o valor zero a pertencer ao intervalo ((-4.695,2.822)).

Muitas vezes surgem estudos de investigação que envolvem dependência amostral, isto é, as observações de uma amostra estão relacionadas com as de outra amostra (correlação amostral não nula). Este caso será analisado no teste seguinte.

4.4 Teste t para duas amostras emparelhadas

O teste t para duas amostras emparelhadas ou dependentes é aplicado quando o conjunto de indivíduos que constitui a amostra é avaliado em dois momentos temporais ou é submetido a duas condições experimentais distintas com o intuito de avaliar a evolução no tempo da característica em estudo ou avaliar o impacto na variável resposta em diferentes condições experimentais.

Os seguintes pressupostos devem ser verificados:

As duas amostras $X_1, \cdots, X_n$ e $Y_1, \cdots, Y_n$ são emparelhada, formando pares de observações $(X_i, Y_i)\ \ i = 1, 2, ..., n$;

Cada amostra deve ser formada por observações independentes e retiradas da mesma população (amostra aleatória);

As amostras são provenientes de populações normais.

Seja $(X_1, Y_1), \cdots (X_n, Y_n)$ uma amostra aleatória de pares e considere-se a variável aleatória diferença definida como $D_i = X_i - Y_i$ com $i = 1, \cdots, n$. A estatística de teste a considerar neste caso é a seguinte:

$$T = \frac{\overline{D}}{S_D/\sqrt{n}} \frown t_{n-1} \quad (1.3)$$

com $\overline{D}$ a média e S_D^2 a variância da amostra $(D_1, \cdots, D_n)$. Pretende-se testar um dos seguintes pares de hipóteses:

$$H_0 : \mu_D = 0 \quad vs \quad H_1 : \mu_D \neq 0$$

$$H_0 : \mu_D = 0 \quad vs \quad H_1 : \mu_D < 0$$

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

O intervalo de $100(1-\alpha)\%$ de confiança para $\mu_X - \mu_Y$, (σ_X e σ_Y desconhecidos):

$$\left]\overline{D} - c\frac{S_D}{\sqrt{n}}, \overline{D} + c\frac{S_D}{\sqrt{n}}\right[\qquad \text{com} \quad c = t_{n-1;1-\alpha/2}$$

onde c é o quantil $1 - \alpha/2$ da dsitribuição t-Student com (n-1) graus de liberdade.

No IBM SPSS o teste e o intervalo encontram-se no menu:

Para as duas amostras originais

Analisar $\rightarrow$ *Comparar Médias* $\rightarrow$ *Amostras emparelhadas Teste t*

Para a amostra das diferenças

Analisar $\rightarrow$ *Comparar Médias* $\rightarrow$ *Uma amostra Teste t*$\rightarrow$ Valor de teste=0

A região crítica a considerar quando se testa a hipótese nula contra cada uma das alternativas bilateral ou unilateral, ao nível de significância $\alpha \in]0, 1[$, é respetivamente:

$$\text{Se} \quad H_1 : \mu_D \neq 0 \quad RC = \{t : |t| > t_{n-1;1-\alpha/2}\}$$

$$RC = \{t : t < t_{n-1;\alpha/2} \quad \text{ou} \quad t > t_{n-1;1-\alpha/2}\}$$

$$\text{Se} \quad H_1 : \mu_D < 0 \quad RC = \{t : t < t_{n-1;\alpha}\}$$

$$\text{Se} \quad H_1 : \mu_D > 0 \quad RC = \{t : t > t_{n-1;1-\alpha}\}$$

o teste para amostras emparelhadas compara as médias de duas variáveis medidas em dois momentos temporais para um mesmo grupo. A regra de decisão consite em rejeitar H_0 se *p-value*$\leq \alpha$ com α o nível de significância do teste.

Exemplo: Pretende-se avaliar se o aproximar da auditoria tem efeito sobre os colaboradores da empresa, considerando as variáveis stress 30 dias e stress 7 dias antes da auditoria. A normalidade dos dados é assumida.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	Stress - 30 dias antes da auditoria (maior valor, maior stress) - Stress - 7 dias antes da auditoria (maior valor, maior stress)	-17,775	8,913	,997	-19,759	-15,791	-17,837	79	,000

Na tabela o intervalo de confiança 95% para a diferença de valores médios, $]-19.76, -15.79[$ indica que o valor médio do stress 30 dias antes da auditoria é inferior ao valor médio do stress 7 dias antes da auditoria.

O *p-value* fornecido na tabela corresponde ao teste bilateral, $H_0 : \mu_D = 0$ vs $H_1 : \mu_D \neq 0$ onde μ_D representa a diferença entre os valores médios do stress 30 dias e o stress 7 dias antes da auditoria. Como (t(-17.84,79), *p-value*$<$ 0.001), concluimos que os valores médios do stress nos dois momentos temporais de 30 e de 7 dias antes da auditoria são diferentes e que o valor médio do stress 30 dias antes da auditoria representa o menor dos dois valores.

4.5 ANOVA um fator com três ou mais amostras independentes

A técnica estatística denominada ANOVA ($\underline{an}alysis\ \underline{o}f\ \underline{va}riance$) foi desenvolvida po Ronald Fisher (1890-1962) para avaliar a igualdade de valores médios em mais de duas populações normais.

A ANOVA é um conjunto de modelos estatísticos nos quais a variância amostral é decomposta em diferentes componentes permitindo estudar a influência dos fatores na característica de interesse (variável dependente ou variável resposta). Considerando uma variável dependente quantitativa em escala intervalar ou de razão e uma variável independente qualitativa em escala nominal, denominada fator, a análise da variância de um fator com k níveis, consiste em testar em simultâneo a igualdade dos k valores médios das populações sobre as quais a variável dependente foi mensurada.

Amostras	Observações	Dimensão	Média amostral
amostra 1	$X_{11}, \cdots, X_{1n_1}$	n_1	$\overline{X}_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
amostra i	$X_{i1}, \cdots, X_{in_i}$	n_i	$\overline{X}_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
amostra k	$X_{k1}, \cdots, X_{kn_k}$	n_k	$\overline{X}_k$

Análise da variância:

Tem-se k amostras de observações independentes (k amostras aleatórias), sendo as amostras independentes entre si;

Cada uma das k amostras $X_{i1}, \cdots, X_{in_i}$ é proveniente de uma distribuição Normal, $N(\mu_i, \sigma_i)$ com $i = 1, \cdots, k$.

A variância das k populações é a mesma (caso homocedástico):

$$\sigma_1^2 = \sigma_2^2 = \cdots \sigma_k^2 = \sigma^2$$

Pretende-se testar as seguintes hipóteses:

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$

vs

H_1 : pelo menos um dos μ_i é diferente $\quad i = 1, 2, ..., k$

A aplicação deste método baseia-se no seguinte:

Estimar a variância σ^2 por dois métodos distintos, um que não depende da veracidade de H_0 e outro que depende.

Em seguida comparam-se as duas estimativas obtidas. Se H_0 for verdadeira as duas estimativas deverão ser próximas, caso contrário diferem significativamente.

Notação:

- Dimensão total das k amostras: $n = n_1 + \cdots + n_k$

- $X_{ij} \quad j$-ésima observação da amostra do nível i

- Média da amostra i:
$$\overline{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$$

- Média global:
$$\overline{\overline{X}} = \frac{1}{n} \sum_{i=1}^{k} n_i \overline{X}_i$$

- A variância global:
$$\frac{1}{n} \sum_{i} \sum_{j} (X_{ij} - \overline{\overline{X}})^2$$

No teste da hipótese da igualdade dos valores médios é essencial ter em atenção a variabilidade dos dados observados relativos a cada população. Se a variabilidade dos dados (em cada um dos grupos) em torno de cada uma das médias amostrais dos respetivos grupos (variabilidade dentro de cada amostra) é grande em comparação com a variabilidade entre as médias amostrais (variabilidade entre amostras), então não se rejeitará a H_0.
Se, pelo contrário a variabilidade dos dados dentro de cada amostra é pequena em comparação com a variabilidade entre amostras, então rejeita-se H_0.

O nome desta técnica estatística resulta de analisar a decomposição da variância total da variável resposta em duas partes: uma parte corresponde à variabilidade entre os diferentes grupos e a outra à variabilidade dentro dos grupos. A decomposição da variabilidade é representada de forma similar à variância como uma soma de quadrados (as variâncias são obtidas das somas de quadrados divididas pelos respetivos graus de liberdade).

$$\underbrace{(X_{ij} - \overline{\overline{X}})}_{\substack{\text{desvio da observação} \\ X_{ij} \text{ em relação à} \\ \text{média global}}} = \underbrace{(\overline{X}_i - \overline{\overline{X}})}_{\substack{\text{desvio da média do} \\ \text{grupo em relação à} \\ \text{média global}}} + \underbrace{(X_{ij} - \overline{X}_i)}_{\substack{\text{desvio da observação} \\ X_{ij} \text{ em relação à média} \\ \text{do grupo}}}$$

$$\underbrace{\sum_i \sum_j \left(X_{ij} - \overline{\overline{X}}\right)^2}_{\substack{\text{SST} \\ \text{total} \\ n-1 \;\; g.l.}} = \underbrace{\sum_i n_i \left(\overline{X}_i - \overline{\overline{X}}\right)^2}_{\substack{\text{SSA} \\ \text{entre amostras} \\ k-1 \;\; g.l.}} + \underbrace{\sum_i \sum_j \left(X_{ij} - \overline{X}_i\right)^2}_{\substack{\text{SSE} \\ \text{dentro das amostras} \\ n-k \;\; g.l.}}$$

Significando que a soma dos quadrados total é igual à soma dos quadrados entre amostras (devida aos grupos) com a soma dos quadrados dentro das amostras (devida ao erro).

As duas fontes de variabilidada são devidas a à variação que resulta da diferença entre indivíduos dentro de cada grupo (variação *within*, residual ou variância dentro do grupo) e à variação que resulta das diferenças introduzidas pelos grupos (variação *between*, ou variância entre grupos).

Sob a validade de $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ tem-se que

$$\frac{SSA}{\sigma^2} \frown \chi^2_{k-1} \quad \text{e} \quad \frac{SSE}{\sigma^2} \frown \chi^2_{n-k} \qquad \text{são independentes.}$$

$\dfrac{SSE}{n-k}$ é um estimador centrado de σ^2 mesmo quando H_0 não é verdadeira e $\dfrac{SSA}{k-1}$ é um estimador centrado de σ^2 quando H_0 é verdadeira.

Sob a validade de $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$:

$$F = \frac{MSA}{MSE} \frown F_{k-1;n-k}$$

com $MSA = \dfrac{SSA}{k-1}$ e $MSE = \dfrac{SSE}{n-k}$.

O *p-value* é dado por $P(F \geq F_{obs}|H_0)$ e a regra de decisão consiste em

rejeitar H_0 se valor-p$\leq \alpha$ sendo α o nível de significância do teste.

A tabela seguinte representa a forma habitual de apresentar os resultados da ANOVA:

Fonte de variação	Graus de liberdade	Soma de quadrados	Média de quadrados	Estatística F
Entre amostras	$k-1$	SSA	MSA	$\dfrac{MSA}{MSE}$
Dentro das amostras	$n-k$	SSE	MSE	
Total	$n-1$	SST		

Exemplo: Num estudo sobre depressão os pacientes são avaliados por um diagnóstico baseado no BDI (depressão mínima, depressão média, depressão moderada e depressão severa) sendo-lhe atribuído um score do inventário de Beck (BDI - *Beck Depression Inventory*)) como quantificador do grau de depressão. Os pacientes foram distribuídos aleatoriamente por 4 tratamentos (1-nenhum, 2-placebo, 3-homeopático e 4-farmacêutico). Petende-se analisar se os valores médios dos scores do inventário de Beck são iguais para os diferentes tratamentos.

Variável dependente - score do inventário de Beck;

Variável independente - tratamento (fator com quatro níveis): 1-nenhum, 2-placebo, 3-homeopático e 4-farmacêutico.

Vamos analisar primeiro uma tabela com as principais estatísticas descritivas, de modo a obtermos o máximo de informação sobre os diferentes grupos.

Na barra de menu escolher:

Analyze → Compare Means → One Way-ANOVA

Dependent List: score do inventário de Beck

Factor: tratamento

Options: Statistics

Descriptive

Homogeneity of variance test

Continue → OK

Descriptives

score do inventário de Beck

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
nenhum	25	41,72	6,328	1,266	39,11	44,33	27	51
placebo	25	35,04	7,056	1,411	32,13	37,95	25	53
homeopático	25	35,52	6,634	1,327	32,78	38,26	19	47
farmacêutico	25	26,88	6,704	1,341	24,11	29,65	15	43
Total	100	34,79	8,451	,845	33,11	36,47	15	53

A média total do score do inventário de Beck é de 34.79. Por tratamento as médias variam entre 26.88 e 41.72. As amplitudes dos intervalos de confiança são nomeadamente: Tratamento nenhum 5.22; tratamento placebo 5.82; tratamento homeopático 5.48 e tratamento farmacêutico 5.54.

- Uma análise alternativa pode ser realizada escolhendo no menu:

  Graphs→ Legacy Dialogs → Error Bar

  Escolher *Simple* e premir *Define*

  Escolher as variáveis dependente e independente (nominal)

Para analisar a normalidade dos dados (com os testes de Kolmogorov-Smirnov e de Shapiro-Wilk) escolher no menu:

  Analyze → Descriptive Statistics → Explore

  Dependent List: score do inventário de Beck; Factor List: tratamento

  Plots: Normality plots with tests; Continue → OK

Tests of Normality

	tipo de tratamento administrado ao paciente	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
score do inventário de Beck	nenhum	,138	25	,200*	,955	25	,323
	placebo	,103	25	,200*	,948	25	,223
	homeopático	,149	25	,159	,959	25	,393
	farmacêutico	,106	25	,200*	,966	25	,538

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

As amostras dos diferentes tratamentos têm uma dimensão pequena (< 30) pelo que consideramos o teste de Shapiro-Wilk. Para todos os tratamentos decide-se pela não rejeição da normalidade dos dados.

Para validar o pressuposto de homogeneidade de variâncias escolher no menu:

  Analyze → Compare Means → One-Way ANOVA

  Dependent List: score do inventário de Beck; Factor List: tratamento

  Options→ Statistics: Homogeneity of variance test; Continue→ OK

Em relação ao teste de Levene, o pressuposto de homogeneidade de variâncias nas amostras dos 4 tratamentos é verificada (Levene=0.119, *p-value*=0.949), pelo

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
score do inventário de Beck	Based on Mean	,119	3	96	,949
	Based on Median	,115	3	96	,951
	Based on Median and with adjusted df	,115	3	93,637	,951
	Based on trimmed mean	,123	3	96	,947

ANOVA

score do inventário de Beck

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2779,710	3	926,570	20,730	,000
Within Groups	4290,880	96	44,697		
Total	7070,590	99			

que não se rejeita a hipótese de igualdade de variâncias populacionais entre os tratamentos.

Em relação à tabela de ANOVA os resultados sugerem que pelos menos um dos tratamentos tem um score médio de inventário de Beck diferente dos restantes, dado que o (F(20.730),*p-value*$<$0.0001), rejeitamos a hipótese de igualdade de médias populacionais para os 4 tratamentos.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu \quad \text{vs} \quad H_1 : \mu_i \neq \mu \quad i \in \{1, 2, 3, 4\}$$

No teste de ANOVA, rejeitar H_0 significa que existem diferenças significativas entre pelo menos um par de valores médios considerados. Interessa, neste caso, identificar que valores médios diferem entre si. Os métodos para realizar comparações múltiplas (comparando todos os possíveis pares de médias) podem ser selecionados no menu:

Analyze $\rightarrow$ Compare Means $\rightarrow$ One-way ANOVA $\rightarrow$ Post Hoc

Quando se assume a igualdade de variâncias dos diferentes grupos, o IBM SPSS disponibiliza um conjunto alargado de testes, dos quais destacam-se três: o teste de Tukey HSD (H*onestly Significant Difference*); o teste de Scheffé e o teste de Bonferroni.

Na tabela destes testes são realizadas todas as comparações múltiplas entre os diferentes grupos, sendo calculadas as diferenças entre as médias de dois grupos (**Mean Difference**); a estimativa do desvio padrão dessa diferença (**Std. Error**); os valores-p (**Sig.**) dos testes bilaterais e um intervalo de confiança para cada diferença de valores médios correspondente a cada comparação. Na coluna **Mean Difference** são assinalados com um asterisco sempre que o teste conduz à rejeição da hipótese nula.

O teste de Tukey HSD é um destes métodos, podendo ser usado para decidir os responsáveis pela rejeição (ou seja, que populações são distintas). Para além da tabela de comparações múltiplas o teste de Tukey HSD fornece também uma tabela de subconjuntos homogéneos de médias (grupos em que não se rejeita que as médias sejam iguais). No exemplo a análise de comparações de pares com o teste de Tukey HSD observa-se que o score médio do tratamento farmacêutico (valor mais baixo igual a 26.88) e do tratamento nenhum (alor mais mais alto igual a 41.72) encontram-se isolados. Os scores médios dos tratamentos placebo (35.04) e homeopático (35.52) não apresentam diferenças estatisticamente significativas.

Nas comparações de dois a dois dos tratamentos, os testes de Tukey HSD, Bonferroni e Scheffe produzem intervalos de confiança para a diferença de médias.

4.6 ANOVA para medidas repetidas para três ou mais amostras emparelhadas

Tal como a análise de variância para amostras independentes a ANOVA de medições repetidas testa a existência de diferenças entre médias populacionais. Neste caso as medições da variável dependente são repetidas no mesmo indivíduo experimental, fazendo com que as amostras não sejam independentes, mas são relacionadas pelo mesmo indivíduo. Na técnica da ANOVA para medidas repetidas temos uma soma de quadrados adicional destinada a quantificar a variabilidade intra-sujeitos.

Com a técnica da ANOVA queremos comparar o efeito de um fator (com k níveis) sobre uma variável resposta X, comparando as médias de k amostras.

	Amostra 1	$\cdots$	Amostra k
Sujeito 1	X_{11}	$\cdots$	X_{1n}
Sujeito 2	X_{21}	$\cdots$	X_{2n}
$\cdots$	$\cdots$	$\cdots$	
Sujeito n	X_{n1}	$\cdots$	X_{nk}

As hipóteses a testar numa ANOVA de medidas repetidas são:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu$$

$$vs$$

$$H_1 : \text{pelo menos um dos } \mu_i \text{ é diferente}$$

Nota: k representa o número de níveis (tratamentos) - número de momentos temporais ou condições experimentais.

A tabela seguinte representa a forma habitual de apresentar os resultados:

Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrados médios	Estatística F
Fatorial (Entre amostras)	$k-1$	SSA	$MSA = \frac{SSA}{k-1}$	$\frac{MSA}{MSE}$
Sujeitos (Intrassujeitos)	$n-1$	SSS	$MSS = \frac{SSS}{n-1}$	
Erro (Intra-amostras)	$(n-1)(k-1)$	SSE=SST-SSA-SSS	$MSE = \frac{SSS}{(n-1)(k-1)}$	
Total	$kn-1$	SST		

A fonte de variação dos erros pode ser designada por interação *SujeitosxTratamentos*. Este termo será uma medida do grau em que o efeito do tratamento depende dos sujeitos da amostra (e vice-versa). Um erro elevado (interação elevada) significa que diferentes sujeitos respondem de forma diferente ao tratamento.

Sob a validade de $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu$:

$$F = \frac{MSA}{MSE} \frown F_{(k-1);(k-1)(n-1)}$$

com $MSA = \dfrac{SSA}{k-1}$ e $MSE = \dfrac{SSE}{(n-1)(k-1)}$.

O valor-p é dado por $P(F \geq F_{obs}|H_0)$ rejeitando-se H_0 para valores pequenos do valor-p. Regra de decisão: rejeito H_0 se valor-p$\leq \alpha$ sendo α o nível de significância de teste.

Exemplo: Uma empresa realizou um estudo para avaliar se o aproximar da hora da auditoria à empresa aumentava o stress dos seus colaboradores.

Neste planeamento experimental o conjunto dos colaboradores da empresa, formam a amostra e foram sujeitos à avaliação do stress em três momentos temporais: 30 dias, 15 dias e sete dias antes da auditoria. Consideremos o factor: 30 dias, 15 dias e sete dias (factor within subject) e a variável dependente: stress (variável quantitativa em escala absoluta). Neste estudo as fontes de variação dos dados são: os colaboradores que são diferentes (efeito residual) e apresentam graus de stress diferentes (efeito factor - aproximar da hora da auditoria).

> H_0 : Não há diferenças entre o stress sentido 30 dias, 15 dias e 7 dias antes da auditoria.
>
> vs
>
> H_1 : Há diferenças entre o stress sentido 30 dias, 15 dias e 7 dias antes da auditoria.

Neste caso aplica-se a técnica de ANOVA com medições repetidas dado que é um planeamento experimental intra-sujeitos e o mesmo grupo de indivíduos vai ser avaliado quanto ao *stress* em três momentos temporais (30 dias, 15 dias e 7 dias).

Na barra de menus escolher:

Analyze → General Linear Model → Repeated Measures

Executar os seguintes passos: Within-Suject Factor Name: escrever stress (variável em estudo)

Number of levels - 3 (medições repetidas aos: 30 dias, 15 dias e 7 dias)

Add → Define

Options: Descriptive statistics

EM means: Display Means for: stress

Compare main effects: Bonferroni

Continue

Plots: Horizontal Axis- stress → Add → Continue → OK

O pressuposto da esfericidade: quando o factor *within* tem mais do que dois

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
stress	,988	,917	2	,632	,988	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: stress

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

níveis, é necessário que se verifique a esfericidade da matriz de covariâncias. Trata-se de um pressuposto semelhante à homogeneidade de variâncias, mas neste caso da ANOVA com medidas repetidas. No exemplo, não rejeito a hipótese nula de ser possível assumir a esfericidade da matriz de covariâncias ($\chi^2_{(2)} = 0.917$, valor-p=0.632). Caso não seja possível assumir a esfericidade, deve-se considerar para a estatística F da tabela da ANOVA os graus de liberdade obtidos pelo critério de Greenhouse-Geisser.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
stress	Sphericity Assumed	15966,558	2	7983,279	209,269	,000
	Greenhouse-Geisser	15966,558	1,977	8076,624	209,269	,000
	Huynh-Feldt	15966,558	2,000	7983,279	209,269	,000
	Lower-bound	15966,558	1,000	15966,558	209,269	,000
Error(stress)	Sphericity Assumed	6027,442	158	38,148		
	Greenhouse-Geisser	6027,442	156,174	38,594		
	Huynh-Feldt	6027,442	158,000	38,148		
	Lower-bound	6027,442	79,000	76,297		

A significância do efeito do ”aproximar da hora da auditoria”lê-se na linha onde está escrito que a esfericidade é assumida (*Sphericity Assumed*). Rejeita-se H_0, F(2,158)=209,27 e valor-p=0.000, ou seja, o nível de stress foi influenciado pelo aproximar da hora da auditoria.

O efeito do aproximar da hora da auditoria - o stress parece ser mais baixo 30 dias antes da auditoria do que o stress 15 dias e 7 dias antes da auditoria.

No teste de ANOVA, rejeitar H_0 significa que existem diferenças significativas entre pelo menos um par de valores médios considerados. Interessa, neste caso, ver que valores médios diferem entre si. Métodos para realizar comparações múltiplas (comparando todos os possíveis pares de médias) estão disponibilizados no SPSS, dos quais se destaca o teste de Bonferroni.

As hipóteses a testar são:

$H_0 : \mu_i = \mu_j$ v.s. $H_1 : \mu_i \neq \mu_j$

para todos os pares de medidas repetidas.

Estimates

Measure: MEASURE_1

stress	Mean	Std. Error	95% Confidence Interval Lower Bound	Upper Bound
1	7,638	,384	6,874	8,401
2	24,425	,840	22,753	26,097
3	25,413	,938	23,545	27,280

Pairwise Comparisons

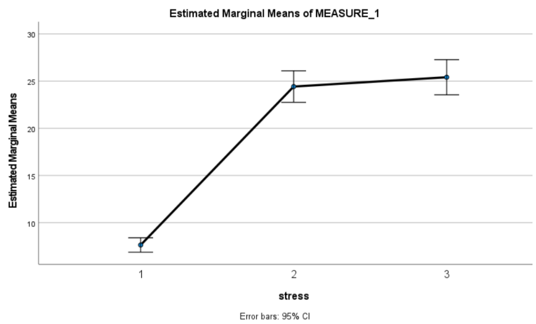
Measure: MEASURE_1

(I) stress	(J) stress	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-16,787*	,923	,000	-19,045	-14,530
	3	-17,775*	,997	,000	-20,212	-15,338
2	1	16,788*	,923	,000	14,530	19,045
	3	-,988	1,008	,991	-3,454	1,479
3	1	17,775*	,997	,000	15,338	20,212
	2	,988	1,008	,991	-1,479	3,454

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.



O stress ocorrido 30 dias antes é estatisticamente diferente do stress ocorrido 15 dias e 7 dias antes da auditoria (Mean Difference (2-1)= 16.79 e valor-p=0.923; Mean Difference (3-1)= 17.78 e valor-p=0.997). O teste de comparação entre o stress ocorrido aos 15 dias e 7 dias antes da auditoria, indica que não são estatisticamente diferentes (Mean Difference (3-2)=0.988 e valor-p=0.991).

5 Testes não paramétricos

5.1 Introdução

Neste tópico do programa serão apresentados os testes de diferenças não-paramétricos, para as seguintes situações:

- Teste U de Mann-Whitney para duas amostras independentes;
- Teste de Kruskal-Wallis para três ou mais amostras independentes;
- Teste de Wilcoxon para duas amostras emparelhadas;
- Teste de Friedman para três ou mais amostras emparelhadas.

5.2 Teste U de Mann-Whitney

O teste U de Mann-Whitney também designado por teste de Wilcoxon-Mann-Whitney é um teste não paramétrico, usado para verificar se duas amostras independentes provêm da mesma população ou de populações que diferem apenas da localização. É um dos testes não paramétricos mais potentes e constitui uma boa alternativa ao teste-t de comparação de valores médios (médias populacionais) quando não se pode assumir a normalidade dos dados ou quando as amostras são pequenas ou ainda quando as variáveis são de escala pelo menos ordinal.

Pressupostos do teste: (i) a variável de estudo é medida numa escala (no mínimo ordinal) que permite a ordenação das observações e subsequente atribuição de ordens ou *ranks*; (ii) as amostras aleatórias, $(X_1, \cdots, X_{n_1})$ e $(Y_1, \cdots, Y_{n_2})$, são provenientes de populações independentes, X e Y, respetivamente e (iii) as funções de distribuição das populações de X e de Y são contínuas.

Quando as populações apresentam formas idênticas, as medidas de localização das duas populações são iguais. No entanto se as medidas de localização forem iguais não implica que as distribuições sejam idênticas.

As hipóteses de interesse consistem num dos pares seguintes:

$H_0 : F_X = F_Y$ vs $H_1 : F_X \neq F_Y$;

$H_0 : F_X = F_Y$ vs $H_1 : F_X < F_Y$;

$H_0 : F_X = F_Y$ vs $H_1 : F_X > F_Y$;

onde F_X e F_Y representam as funções de distribuições das variáveis X e Y nas respetivas populações.

O teste U de Mann-Whitney, no IBM SPSS, aparece nas versões de teste exato e de teste assintótico. Na prática a distribuição assintótica é razoavelmente precisa para amostras de dimensão superior ou igual a 6, pelo que será a opção usual.

O teste de Mann-Whitney é acessível no IBM SPSS através das opções: Analyze $\rightarrow$ Nonparametric Tests $\rightarrow$ Legacy Dialogs $\rightarrow$ 2 Independent Samples

Para os diferentes testes o correspondente *p-value*:

H_1	p-value
$F_X - F_Y \neq 0$	$2P(Z \geq \lvert z_{\text{obs}}\rvert \lvert H_0)$
$F_X - F_Y < 0$	$P(Z \leq z_{\text{obs}} \lvert H_0)$
$F_X - F_Y > 0$	$P(Z \geq z_{\text{obs}} \lvert H_0)$

Regra de decisão: rejeito H_0 quando o *p-value* for inferior ou igual ao nível de significância α.

Exemplo: As diferenças entre homens e mulheres podem ser consideradas significativas para a variável idade (anos)? As hipóteses a testar são:

H_0 : Os grupos feminino e masculino são idênticos na idade. vs H_1 : Os grupos feminino e masculino não são idênticos na idade.

As variáveis envolvidas nas hipóteses do teste são a idade (anos) variável quantitativa em escala absoluta e a variável sexo variável qualitativa nominal (como *string*) que precisa de recodificação (como numérica) para a execução do teste.

Tests of Normality

	Sexo do paciente	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Idade do paciente	Masculino	,155	38	,022	,892	38	,002
	Feminino	,182	42	,001	,901	42	,001

a. Lilliefors Significance Correction

Mann-Whitney Test

Ranks

	Sexo do paciente	N	Mean Rank	Sum of Ranks
Idade do paciente	Masculino	38	43,67	1659,50
	Feminino	42	37,63	1580,50
	Total	80		

Test Statistics[a]

	Idade do paciente
Mann-Whitney U	677,500
Wilcoxon W	1580,500
Z	-1,170
Asymp. Sig. (2-tailed)	,242

a. Grouping Variable: Sexo do paciente

Variável Sexo (alterar de sequência de caracteres para numérica)

- Transformar→ Recodificar nas mesmas variáveis
- Variáveis de sequência de caracteres : Sexo
- Valores antigo e novo
- Valor antigo:h Novo valor: 0 → Adicionar
- Valor antigo:m Novo valor: 1 → Adicionar
- Continuar
- OK

Nas definições da variável alterar os valores e rótulos:

- Valor:0 Rótulo: homem → Adicionar
- Valor:1 Rótulo: mulher → Adicionar

A não normalidade para a variável idade presente nos dois grupos de indivíduos (feminino e masculino) conduz à aplicação do teste de Mann-Whitney para duas amostras independentes com o seguinte resultado para o teste assintótico (Z=-1.170 ; p-value=0.242). Donde se conclui queas distribuições para a idade por sexo são idênticas, logo os dois grupos são idênticos para as idades.

5.3 Teste de Kruskal-Wallis

O teste de Kruskal-Wallis é usado para comparar mais de duas distribuições de uma variável definida no mínimo em escala ordinal, e cujas observações constituem mais de duas amostras independentes. De uma forma um pouco abusiva no lugar de distribuições fala-se em medidas de localizações centrais (medianas) de k populações.

Caso a normalidade das k amostras não seja verificada, o teste não paramétrico de Kruskal - Wallis pode ser usado como alternativa à ANOVA um factor. Neste teste as observações são substituídas pela sua ordem na amostra conjunta, testando-se a igualdade das distribuições das k populações. Este teste pode também ser usado no caso heterocedástico. O número de grupos definido pela variável independente são 3 ou mais e a escala de medida da variável dependente é no mínimo ordinal, como já referido.

O teste de Kruskal-Wallis é acessível no SPSS através das opções:

Analyze $\rightarrow$ Nonparametric Tests $\rightarrow$ k Independent Samples

Regra de decisão do teste: rejeito H_0 se *p-value*$\leq \alpha$.

Quando se rejeita H_0 deve-se investigar quais os pares de amostras que diferem entre si. Assim, realizam-se tantos testes não paramétricos do tipo teste de Mann-Whitney quantas as comparações duas a duas das amostras. Estas comparações irão permitir identificar os grupos com diferenças significativas.

A realização de vários testes de Mann-Whitney para fazer comparações múltiplas, com o objetivo de identificar grupos de semelhança entre as amostras conduz a um aumento do erro tipo I, ou seja, podemos encontrar resultados significativos que não o sejam na realidade. Assim para ultrapassar esta desvantagem devemos efetuar a correção de Bonferroni que consiste em comparar o *p-value* do teste de Mann-Whitney não com o nível de significância α, mas com α dividido pelo número de combinações de pares a comparar.

Exemplo : Averiguar se as dificuldades relacionadas com a profissão contribuem para o grau de stress sentido diariamente. As hipóteses a testar são: H_0 : Não há diferença entre os indivíduos que apresentam diferentes níveis de dificuldades profissionais ao nível do stress sentido diariamente e H_1 : Há diferença entre os indivíduos que apresentam diferentes níveis de dificuldades profissionais ao nível do stress sentido diariamente. Neste estudo pretende-se comparar quatro grupos de indivíduos que avaliaram as dificuldades sentidas com a profissão como sem dificuldade, pouca dificuldade, muita dificuldade e muitíssima dificuldade, representa a variável independente nominal que será colocada no quadro *Test Variable List*. A variável dependente é o grau de stress sentido diariamente é ordinal e será inserida no quadro *Grouping Variable*. Por baixo do quadro *Grouping Variable* aparece um botão *Define Groups* onde se deve colocar os valores mínimo e máximo correspondentes à codificação da variável dependente.

O output do teste de Kruskal-Wallis apresenta duas tabelas. A primeira desiganada como *Ranks* apresenta duas colunas: N, número de indivíduos por grupo e *Mean Rank*, onde surge a ordem média de cada grupo. A segunda tabela *Test Statistics* apresenta o valor da estatística do teste de Kruskal-Wallis (χ^2); os correspondentes graus de liberdade (*df*); o valor-pdo teste assintótico (*Asymp. Sig.*); o valor-p do teste exato (*Exact Sig.*) e na última linha da tabela o (*Point Probability)*. Para o exemplo, o valor de $\chi^2 = 10.723$ com 3 graus de liberdade e um valor-p=0.013<0.05, pelo que a hipótese H_0 será rejeitada, i.e., existem diferenças estatisticamente significativas entre os grupos. Dado o resultado signi-

Kruskal-Wallis Test

Ranks

	Dificuldades sentidas com a profissão	N	Mean Rank
Grau de stress sentido	sem dificuldade	14	27,86
	pouca dificuldade	28	37,04
	muita dificuldade	26	45,38
	muitíssima dificuldade	12	52,75
	Total	80	

Test Statistics[a,b]

	Grau de stress sentido
Kruskal-Wallis H	10,723
df	3
Asymp. Sig.	,013
Exact Sig.	,011
Point Probability	,000

a. Kruskal Wallis Test

b. Grouping Variable: Dificuldades sentidas com a profissão

ficativo do teste de Kruskal-Wallis temos de realizar testes não paramétricos (teste de Mann-Whitney) para comparar dois a dois os quatro grupos em comparação. As comparações de dois a dois permitem identicar entre que grupos existem diferenças aplicando a estes testes a Correção de Bonferroni, que consiste em dividir o nível de significãncia do teste ($\alpha = 0.05$) pelo número de testes dois a dois a realizar. Para o exemplo temos de realizar seis testes de Mann-Whitney pelo que se obtém uma Correção de Bonferroni $\alpha = 0.008$ que será comparado com cada valor-p dos testes de Mann-Whitney a realizar. Da realização do teste de Mann-Whitney para o

Mann-Whitney Test

Ranks

	Dificuldades sentidas com a profissão	N	Mean Rank	Sum of Ranks
Grau de stress sentido	sem dificuldade	14	17,71	248,00
	pouca dificuldade	28	23,39	655,00
	Total	42		

Test Statistics[a]

	Grau de stress sentido
Mann-Whitney U	143,000
Wilcoxon W	248,000
Z	-1,529
Asymp. Sig. (2-tailed)	,126
Exact Sig. [2*(1-tailed Sig.)]	,163[b]

a. Grouping Variable: Dificuldades sentidas com a profissão

b. Not corrected for ties.

par (”sem dificuldade”; ”pouca dificuldade”) o output indica o valor da estatística de teste U=143 e um valor-p=0.019$>$ 0.008 pelo que indica não haver diferenças significativas entre os grupos compardos.

Na tabela a seguir são apresentados os resultados dos seis testes de comparação dois a dois:

Par	Estatística U	valor-p	Decisão ($\alpha = 0.008$)
sem dif. vs pouca dif.	143,00	0,163	não rejeitar
sem dif. vs muita dif.	104,50	0,027	não rejeitar
sem dif. vs muitíssima dif.	37,50	0,015	não rejeitar
pouca dif. vs muita dif.	283,50	0,126	não rejeitar
pouca dif. vs muitíssima dif.	98,50	0,039	não rejeitar
muita dif. vs muitíssima dif.	125,00	0,343	não rejeitar

Os testes de Mann-Whitney realizados realçam que com a aplicação da Correção de Bonferroni nenhum dos seis testes apresenta diferenças significativas.

5.4 Teste de Wilcoxon

O teste de Wilcoxon é usado num contexto de duas amostras emparelhadas quando os dados não são provenientes de populações normais e se pretende analisar as diferenças entre duas condições para a mesma amostra de indivíduos. O teste permite comparar a mesma variável dependente em dois momentos temporais ou em duas condições experimentais diferentes. As amostras obtidas dizem-se emparelhadas quando se pretende comparar a resposta dada a dois tratamentos ou estímulos X e Y aplicados ao mesmo indíviduo ou a indivíduos tão semelhantes quanto possível.

A variável dependente é medida no mínimo em escala ordinal. Calculam-se as diferenças de cada par de observações, considerando o sinal e ordenam-se as diferenças da menor para a maior, prescindindo dos sinais; por fim somam-se as ordens das diferenças positivas e negativas para provar que são iguais.

O teste de Wilcoxon é acessível no IBM SPSS através das opções:

Analyze $\rightarrow$ Nonparametric Tests $\rightarrow$ Legacy Dialogs $\rightarrow$ 2 Related Samples

Pretendemos testar as seguintes hipóteses:

$H_0 : F_X = F_Y$ vs $H_1 : F_X \neq F_Y$;

$H_0 : F_X = F_Y$ vs $H_1 : F_X < F_Y$;

$H_0 : F_X = F_Y$ vs $H_1 : F_X > F_Y$;

O software IBM SPSS para o teste de Wilcoxon calcula dois valores-p, um para o texte exato (*Exact tests*) e outro para o teste assintótico este último a partir da distribuição normal. A aproximação à distribuição normal é adequada para amostras de dimensões $n > 20$, contudo para amostra de pequena dimensão são apresentados os dois valores-p (exato e assintótico) o que pode criar alguma confusão com a diferença de valores que podem apresentar.

Cálculo do *p-value* para os diferentes testes:

- teste bilateral: *p-value*=$2 * P(Z \geq |Z_{obs}||H_0)$
- teste unilateral à direita: *p-value*=$P(Z \geq Z_{obs}|H_0)$
- teste unilateral à esquerda: *p-value*=$P(Z \leq Z_{obs}|H_0)$

Regra de decisão: rejeitar H_0 quando *p-value*$\leq \alpha$.

Exemplo: Os níveis de stress relacionados com a saúde e a profissão são idênticos, a um nível de significância,$\alpha = 5\%$? A variável dependente - grau de stress (com 4 níveis: ”sem stress”; ”pouco stress”; ”muito stress”e ”muitíssimo stress”) - é avaliada em duas situações distintas (a nível da saúde e a nível profissional) para os mesmos 80 indivíduos que formam a amostra.

As hipóteses a testar são:

H_0 : Não há diferença entre os níveis de stress relacionados com a saúde e a profissão.

H_1 : Há diferença entre os níveis de stress relacionados com a saúde e a profissão.

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
Nível de stress relacionado com a profissão - Nível de stress relacionado com a saúde	Negative Ranks	38[a]	28,22	1072,50
	Positive Ranks	22[b]	34,43	757,50
	Ties	20[c]		
	Total	80		

a. Nível de stress relacionado com a profissão < Nível de stress relacionado com a saúde

b. Nível de stress relacionado com a profissão > Nível de stress relacionado com a saúde

c. Nível de stress relacionado com a profissão = Nível de stress relacionado com a saúde

Test Statistics[a]

	Nível de stress relacionado com a profissão - Nível de stress relacionado com a saúde
Z	-1,196[b]
Asymp. Sig. (2-tailed)	,232

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

Neste estudo temos o mesmo grupo de indivíduos avaliados ao nível do grau de stress em dois níveis saúde e profissional. A variável dependente - nível de stress - é medida numa escala ordinal para a saúde e para a profissão. Neste estudo estamos a avaliar a mesma variável dependente em duas condições diferentes (stress a nível da saúde e stress a nível profissional). No SPSS as opções no menu são:

Analyze $\rightarrow$ Nonparametric Tests $\rightarrow$ Legacy Dialogs $\rightarrow$ 2 Related Samples

Como as amostras são emparelhadas, necessitamos de escolher o par de variáveis que pretendemos comparar:

Variável 1: nível de stress relacionado com a saúde;

Variável 2: nível de stress relacionado com a profissão.

Do output do SPSS são apresentadas duas tabelas: Ordens e Estatística de Teste. Na tabela das ordens aparecem três colunas: N (dimensões dos grupos); Ordem Média (positivas e negativas) e a Soma das Ordens. Na tabela da Estatística de Teste aparece o valor do Teste de Wilcoxon, representado por Z e o respetivo valor-p. Para o exemplo obtém-se Z=-1.196 e p-value=0.232.

Conclusão: Como o *p-value*=$0.232 > 0.05$, não se rejeitar H_0, pelo que se deve concluir que não há diferença entre os graus de stress para as duas situações em avaliação (nível de saúde e nível profissional). Na tabela das ordens verifica-se contudo uma tendência para os indivíduos considerarem a profissão com menos stress que o stress devido à saúde (38 pares com ordens negativas contra 22 pares com ordens positivos).

5.5 Teste de Friedman

O teste de Friedman é a versão não-paramétrica alternativa da ANOVA com medidas repetidas. Este teste aplica-se a comparações de mais de duas populações (k populações) de onde foram extraídas as amostras dependentes em que a variável dependente em análise é definida em escala no mínimo ordinal. Com o teste de Friedman pretendemos averiguar se há diferenças na variável dependente que foi avaliada em três ou mais momentos temporais ou condições experimentais distintas.

A hipótese a testar pode ser escrita como: H_0 : Não há diferença na avaliação dos diferentes momentos temporais ou H_0 : Não há diferença na avaliação das diferentes condições experimentais.

A hipótese alternativa, H_1, é a negação da respetiva hipótese nula.

Regra de decisão do teste: rejeitar H_0 se *p-value* $\leq \alpha$, nível de significância.

Exemplo: Uma auditoria externa a uma empresa gerou uma onda de stress entre os colaboradores da mesma. Averigue se o aproximar da auditoria tem efeito significativo sobre o stress sentido pelos colaboradores. Considere as medições, do grau de stress, realizadas em três momentos temporais: 30 dias antes, 15 dias antes e 7 dias antes da auditoria.

H_0 : Não há diferença entre o grau de stress nos três momentos temporais. H_1 : Há diferença entre o grau de stress nos três momentos temporais. No IBM SPSS as opções no menu são:

Analyze $\rightarrow$ Nonparametric Tests $\rightarrow$ Legacy Dialogs $\rightarrow$ K Related Samples

O output do teste apresenta duas tabelas: a das ordens (Ranks) e a Estatística de Teste (Test Statistics).

Friedman Test

Ranks

	Mean Rank
Nível de stress relacionado com a saúde	2,39
Nível de stress relacionado com a profissão	2,09
Nível de stress relacionado com o social	1,52

Test Statistics[a]

N	80
Chi-Square	45,445
df	2
Asymp. Sig.	,000

a. Friedman Test

Na tabela temos a média das ordens (Mean Rank) para as variáveis nível de stress (relacionado com a saúde, relacionado com a profissão e relacionado com o social).

Na tabela observamos o valor da estatística do Teste de Friedman (Chi-Square); os graus de liberdade (df=degrees of freedom); o nível de significância (Asymp. Sig).

Conclusão: como o p-value=0.000 < 0.05, rejeitamos H_0, pelo que concluímos que há diferença entre os níveis de stress para situações em contexto de saúde, profissão e social.

Para identificar as possíveis diferenças entre os níveis de stress, vamos fazer comparações de pares recorrendo ao teste de Wilcoxon.

Com o objetivo de minimizar o Erro Tipo I associado ao elevado número de testes de comparação vamos aplicar a correção de Bonferroni, para o nível de significância α, vamos comparar o valor-p dos testes de Wilcoxon com

$$\alpha^* = \alpha/\text{número de comparações.}$$

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
Nível de stress relacionado com o social - Nível de stress relacionado com a profissão	Negative Ranks	32[a]	17,31	554,00
	Positive Ranks	1[b]	7,00	7,00
	Ties	47[c]		
	Total	80		
Nível de stress relacionado com a profissão - Nível de stress relacionado com a saúde	Negative Ranks	38[d]	28,22	1072,50
	Positive Ranks	22[e]	34,43	757,50
	Ties	20[f]		
	Total	80		
Nível de stress relacionado com o social - Nível de stress relacionado com a saúde	Negative Ranks	52[g]	30,83	1603,00
	Positive Ranks	6[h]	18,00	108,00
	Ties	22[i]		
	Total	80		

Test Statistics[a]

	Nível de stress relacionado com o social - Nível de stress relacionado com a profissão	Nível de stress relacionado com a profissão - Nível de stress relacionado com a saúde	Nível de stress relacionado com o social - Nível de stress relacionado com a saúde
Z	-4,972[b]	-1,196[b]	-5,980[b]
Asymp. Sig. (2-tailed)	,000	,232	,000

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

Neste exemplo há três pares de combinações de variáveis para aplicar o teste de Wilcoxon. Considerando ”nível de stress relacionado com a saúde”=1, ”nível de stress relacionado com a profissão”=2 e ”nível de stress relacionado com o social”=3, obtemos os seguintes pares: (1,2), (1,3) e (2,3).

Da aplicação do teste de Wilcoxon resultaram os seguintes valores:

(Para $\alpha = 0.05$ obtemos $\alpha^* = 0.05/3 = 0.017$)

Pares	Est. de Wilcoxon	p-value	$\alpha^* = 0.017$
(1,2)	-1.196	0.232	Não rejeitar H_0
(1,3)	-5.980	< 0.001	Rejeitar H_0
(2,3)	-4.972	< 0.001	Rejeitar H_0

Concluí-se que o nível de stress relacionado com o social é considerado diferente e com uma ordem média de nível de stress inferior às ordens médias das outras situações analisadas (tabela dos Ranks do Teste de Friedman).

Bibliografia

Field, A.; *Discovering Statistics Using SPSS*. Sage, London, 3 edition, 2009.

Howell, D.C.; *Statistical Methods for Psychology*. Belmont, USA: Cengage Wadsworth, 7 edition, 2010.

Marôco, J.; *Análise Estatística com o SPSS Statistics*. Reporter Number, 7.^{a} Edição, 2018.

Martins, C.; *Manual de Análise de Dados Quantitativos com Recurso ao IBM SPSS: Saber decidir, fazer, interpretar e redigir*. Psiquilíbrios, Braga, 2011.

Ribeiro C.; Silva J.; Pimenta C.; Murteira, B.; *Introdução à Estatística*. MacGraw-Hill, 3.^{a} Edição, 2010.

Velosa S.F.; Pestana, D.D.; *Introdução à Probabilidade e à Estatística*. Fundação Caloute Gulbenkian, 4.^{a} Edição, 2010.