

COLEÇÃO AULA ABERTA
SUBCOLEÇÃO CADERNOS PEDAGÓGICOS

INTRODUÇÃO ÀS METODOLOGIAS DA INVESTIGAÇÃO EM MOTRICIDADE HUMANA

MANUAL PRÁTICO DE ANÁLISES DE DADOS COM SPSS

Pablo Tomás Carús
Andreia Cordeiro Fernandes



COLEÇÃO AULA ABERTA
SUBCOLEÇÃO CADERNOS PEDAGÓGICOS

INTRODUÇÃO ÀS METODOLOGIAS DA INVESTIGAÇÃO EM MOTRICIDADE HUMANA

MANUAL PRÁTICO DE ANÁLISES DE DADOS COM SPSS

Pablo Tomás Carús
Andreia Cordeiro Fernandes



Ficha técnica

Título: Introdução às metodologias da investigação em motricidade humana: manual prático de análises de dados com SPSS

Autores: Pablo Tomás Carús, Andreia Cordeiro Fernandes

Edição: Universidade de Évora

Coleção: Aula Aberta

Subcoleção: Cadernos Pedagógicos

Paginação: Divisão de Comunicação da Universidade de Évora

ISBN: 978-972-778-219-2

Évora 2021

AGRADECIMENTOS

Os autores deste livro agradecem aos alunos as sugestões construtivas que muito contribuíram para o seu aperfeiçoamento e qualidade.

ÍNDICE

Preâmbulo	5
1. Preparação da base de dados em SPSS	6
1.1 Abrir o programa SPSS	6
1.2 Preparação do ficheiro de dados	7
2. Transformação de variáveis	12
2.1 Cálculo de novas variáveis a partir das já existentes	12
2.2 Recodificação de variáveis	14
3. Condições de aplicabilidade dos testes paramétricos	19
3.1 Condições de aplicabilidade dos testes paramétricos	19
3.2 Testar a normalidade dos dados das variáveis	19
3.3 Testar a homogeneidade das variâncias	22
4. Estatística descritiva	26
4.1 Medidas de tendência central e de dispersão	26
4.2 Estatística descritiva usando o SPSS	26
4.2.1 Variáveis numéricas contínuas ou “Scale”	27
4.2.2 Variáveis categóricas (“Nominal / Ordinal”)	31
5. Testes paramétricos	33
5.1 Testes paramétricos vs. testes não paramétricos	33
5.2 T-teste para amostras independentes	33
5.3 T-teste para amostras emparelhadas	36
5.4 ANOVA – one way	38
5.5 Teste de Pearson	45
6. Testes não paramétricos	48
6.1 Teste de Chi-quadrado (χ^2) de uma amostra	48
6.2 Teste de Mann – Whitney	50
6.3 Teste de Wilcoxon	54
6.4 Teste de Kruskal – Wallis	56
6.5 Teste de Spearman	59
Bibliografia	63

PREÂMBULO

O *Statistical Package for the Social Sciences* (SPSS) é um programa estatístico informático usado para realizar vários tipos de análises estatísticas, desde as mais simples às mais complexas. As versões 18.0 e anteriores, foram produzidas pela empresa SPSS Inc., com sede em Chicago, USA. Com a aquisição da *International Business Machines Corporation* (IBM), em 2010, as seguintes versões passaram a ter sede em New York, USA. Este é um dos programas estatísticos mais utilizados no mundo dos negócios, para analisar o mercado financeiro e no mundo académico na análise das ciências sociais.

O propósito deste livro visa ensinar como se utiliza o SPSS, de forma aplicada na introdução às metodologias da investigação em motricidade humana. Os destinatários do livro são, em geral, todos aqueles que queiram aprender a realizar análise de dados utilizando o computador, de uma forma simples e sem fórmulas, especialmente estudantes da área da motricidade humana.

O livro está organizado em seis capítulos, sendo que os conteúdos estudados se organizam desde o mais simples até ao mais complexo, começa pela preparação de um ficheiro de dados em SPSS (Capítulo 1). Em seguida explica como se realiza a transformação de variáveis novas, a partir de outras já existentes (Capítulo 2) e as condições de aplicabilidade dos testes paramétricos (Capítulo 3). Finaliza com a aprendizagem do procedimento de computação e interpretação da janela de resultados de alguns testes de estatística descritiva (Capítulo 4), paramétrica (Capítulo 5) e não paramétrica (Capítulo 6) mais usados em motricidade humana, para um nível básico-intermédio.

1. PREPARAÇÃO DA BASE DE DADOS EM SPSS

Depois de completar este capítulo, o aluno deverá ser capaz de:
Preparar um ficheiro de dados em SPSS.

1.1. ABRIR O PROGRAMA SPSS

O presente livro foi escrito para utilizar o IBM SPSS *software*. Assim que o programa SPSS estiver instalado, com os códigos atualizados, deve abrir-se o programa (Figura 1).



Figura 1
Indicação do ícone do programa estatístico SPSS, na barra de ferramentas

Ao abrir o programa SPSS, como indicado anteriormente, irá abrir uma janela que permitirá abrir um novo ficheiro de dados ou abrir um ficheiro de dados existente (Figura 2).

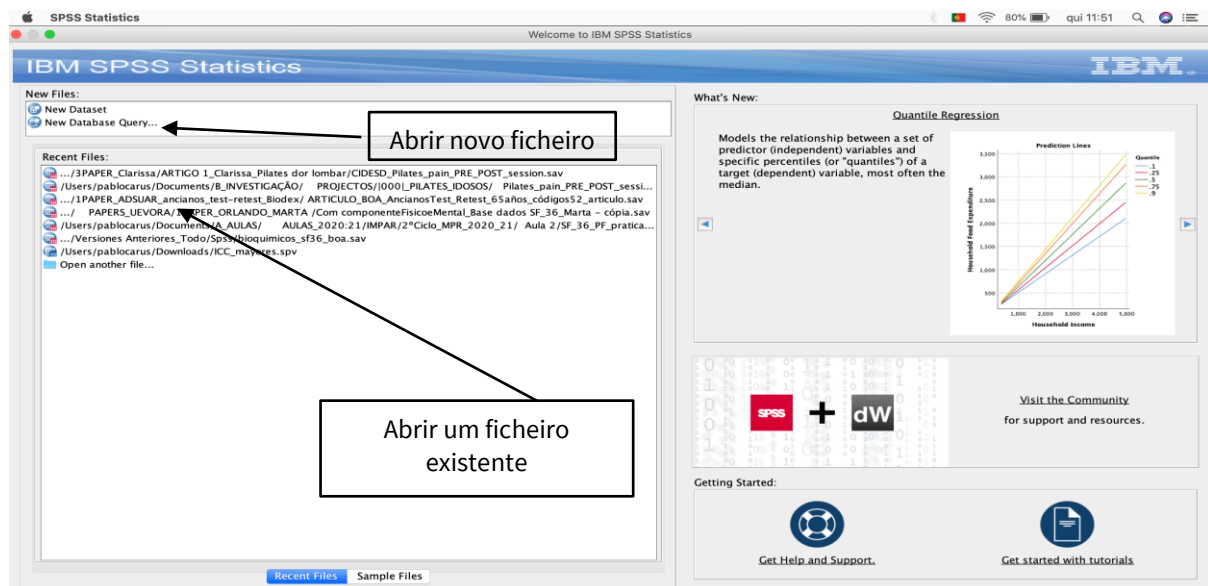


Figura 2
Indicação sobre como abrir um novo ficheiro ou um ficheiro existente

1.2 PREPARAÇÃO DO FICHEIRO DE DADOS

Depois de seleccionarmos a opção “*New Dataset*” na janela apresentada na figura anterior, aparece o Editor de Dados do SPSS, com a opção “*Data View*” pré-seleccionada, que serve para incluir os dados de cada variável, mas antes devemos criar as variáveis, para isso seleccionamos a opção “*Variable View*” (Figura 3).

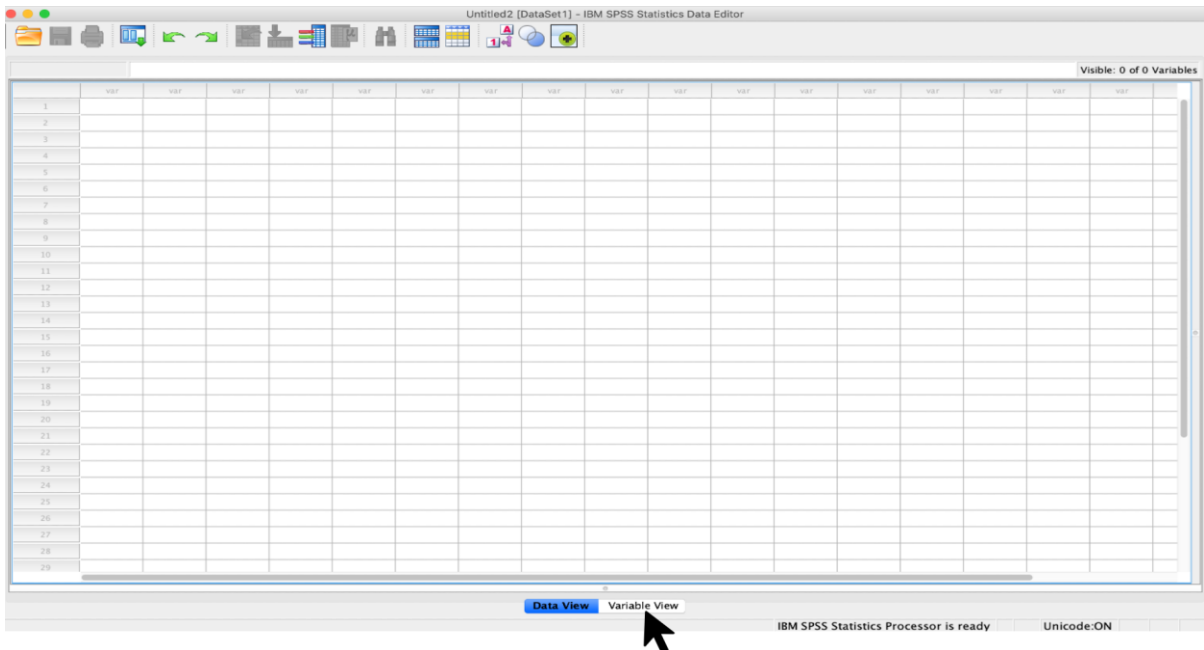


Figura 3
Indicação sobre onde criar as variáveis (“*variable view*”)

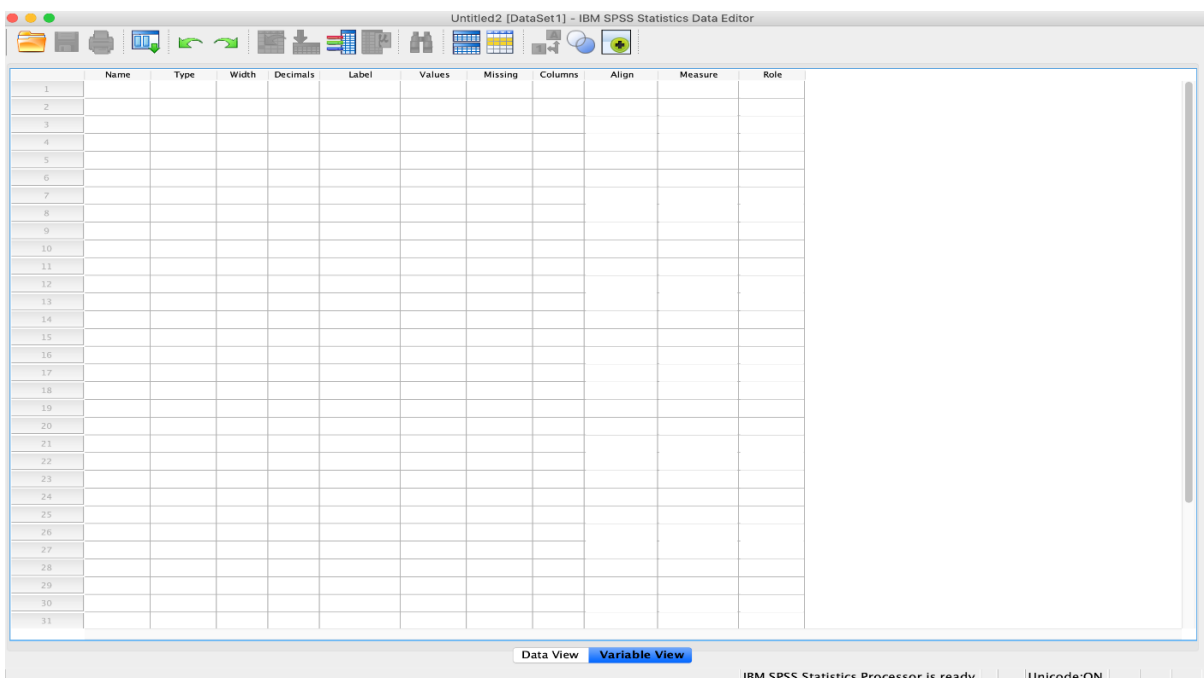


Figura 4
Janela com a tabela de definição das variáveis

Após selecionar a opção “*Variable View*” obter-se-á a janela de definição de variáveis (Figura 4).

No SPSS, cada uma das variáveis precisa de ser definida, ao analisarmos a Figura 5:

	Name	Type	Width	Decimals	Label	Missing	Values	Columns	Align	Measure
1	Participante	Numeric	9	0	Pacientes com ...	None	None	7	Right	Scale
2	Grupo	Numeric	8	0	GRUPO	None	{1, Experimental}...	8	Right	Nominal
3	Idade	Numeric	8	0	IDADE	None	None	8	Right	Scale
4	Peso	Numeric	8	2	PESO	None	None	8	Right	Scale
5	Altura	Numeric	8	2	ALTURA	None	None	8	Right	Scale
6	IMC	Numeric	8	2	Índice de Mass...	None	None	10	Right	Scale
7	anos_diagn...	Numeric	8	0	anos de diagno...	None	None	18	Right	Scale
8	n_medicam...	Numeric	8	0	numero de me...	None	None	16	Right	Scale
9	Género	Numeric	8	0	Género	None	{0, Feminino}...	8	Right	Nominal
10										
11										

Figura 5

Exemplo de definição de variáveis (“*variable view*”)

1. “*Name*”: O nome da variável deve começar com uma letra e pode incluir números, no entanto, sem qualquer intervalo. Se a variável precisar de ser definida em duas palavras, então devem ser unidas, por meio de *underscore*, por exemplo: anos_diagnostico ou n_medicamentos.
2. “*Type*”: É definido o formato das variáveis (numéricas ou não numéricas).
3. “*Width*”: É definido o número de dígitos que uma variável pode ter.
4. “*Decimals*”: É definida a quantidade de números decimais que uma variável pode ter.
5. “*Label*”: É definido o nome completo da variável. O nome da variável pode ser escrito, por extenso, com intervalos entre as palavras, sem abreviaturas.
6. “*Values*”: A codificação da variável é definida por um duplo clique nesta célula, sempre que a variável tiver categorias. Por exemplo, se houver a opção de escolher entre as duas categorias de género – feminino, masculino – então estas categorias podem ser codificadas como 1 = feminino, 2 = masculino. Ao introduzir os dados no computador, estes códigos são introduzidos de acordo com a resposta de um determinado participante (Figura 6).
7. “*Missing*”: Nos estudos baseados em inquéritos, é possível que um participante possa não responder a certas perguntas. Isto cria o problema do valor em falta que pode ser definido nesta coluna.
8. “*Columns*”: É definida a largura do espaço da coluna onde os dados são digitados na vista de dados.
9. “*Align*”: É definido o alinhamento dos dados. Pode ser definido como esquerda, direita ou centro.
10. “*Measure*”: É definido o tipo de variável, que pode ser escala, ordinal ou nominal.

Após definir as variáveis, clicamos na opção “Data View”, na parte inferior do ecrã para abrir o formato de introdução de dados (Figura 7).

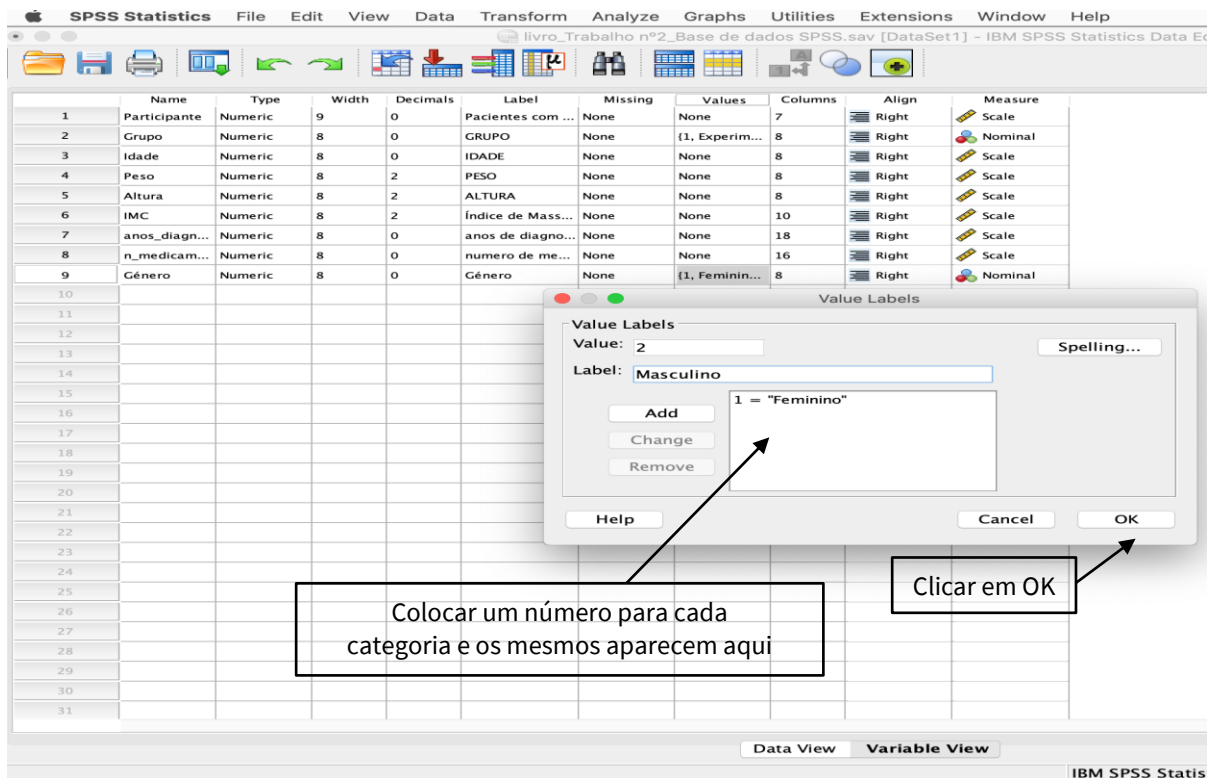


Figura 6
Indicação sobre como inserir "value labels"

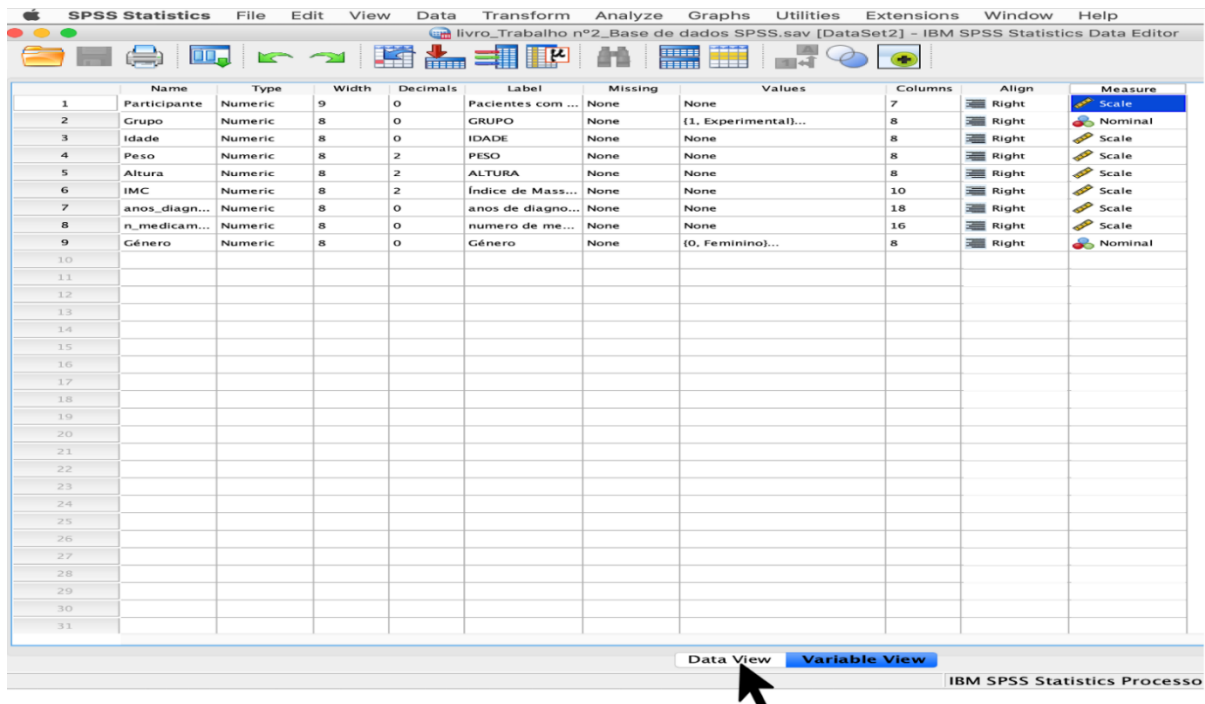


Figura 7
Indicação sobre como voltar à janela de introdução/edição de dados

Para cada variável, os dados podem ser introduzidos em função da coluna. Após introduzir os dados, o ecrã terá o aspeto da Figura 8.

Guardar o ficheiro de dados no local desejado antes do processamento posterior: clicamos em “File” → “Save As...”. Seguidamente, deveremos escolher a pasta de destino e o nome do ficheiro, e, posteriormente, carregar em “Save”(Figura 9).

Após a preparação do ficheiro de dados, o mesmo poderá ser utilizado para diferentes tipos de análises estatísticas disponíveis no SPSS.

Partici pante	Grupo	Idade	Peso	Altura	IMC	anos_diagnostico	n_medicamentos	Género	var	var	var	var
1	2	61	84,00	1,56	34,52	13	3	0				
2	2	65	77,00	1,49	34,68	8	1	0				
3	2	56	72,00	1,57	29,21	15	2	0				
4	2	65	88,00	1,72	29,75	13	2	1				
5	2	51	89,00	1,75	29,06	6	1	1				
6	2	63	73,00	1,56	30,00	5	2	1				
7	2	68	86,00	1,69	30,11	16	1	1				
8	2	63	74,50	1,53	31,83	3	2	1				
9	2	60	86,00	1,64	31,98	3	1	1				
10	2	67	81,50	1,68	28,88	5	1	1				
11	2	63	89,50	1,76	28,89	12	3	1				
12	2	50	71,00	1,70	24,57	18	2	1				
13	2	59	79,00	1,65	29,02	3	1	1				
14	2	48	128,00	1,74	42,28	13	2	1				
15	2	62	84,50	1,67	30,30	4	1	0				
16	2	49	78,00	1,52	33,76	7	2	0				
17	2	65	56,50	1,48	25,79	9	2	0				
18	1	71	73,50	1,56	30,20	13	2	0				
19	1	51	63,00	1,58	25,24	27	1	0				
20	1	61	61,00	1,60	23,63	12	2	0				
21	1	64	85,00	1,57	34,48	23	2	0				
22	1	64	92,50	1,58	37,05	6	1	0				
23	1	70	66,50	1,53	28,41	20	3	0				
24	1	54	100,00	1,65	36,73	4	2	0				
25	1	59	89,00	1,72	30,08	3	1	1				
26	1	51	96,00	1,73	32,08	8	2	1				
27	1	64	90,50	1,65	33,24	6	1	1				
28	1	65	97,50	1,65	35,81	9	1	1				
29	1	69	70,50	1,65	25,90	19	2	1				

Figura 8
Exemplo do “data view” preenchido

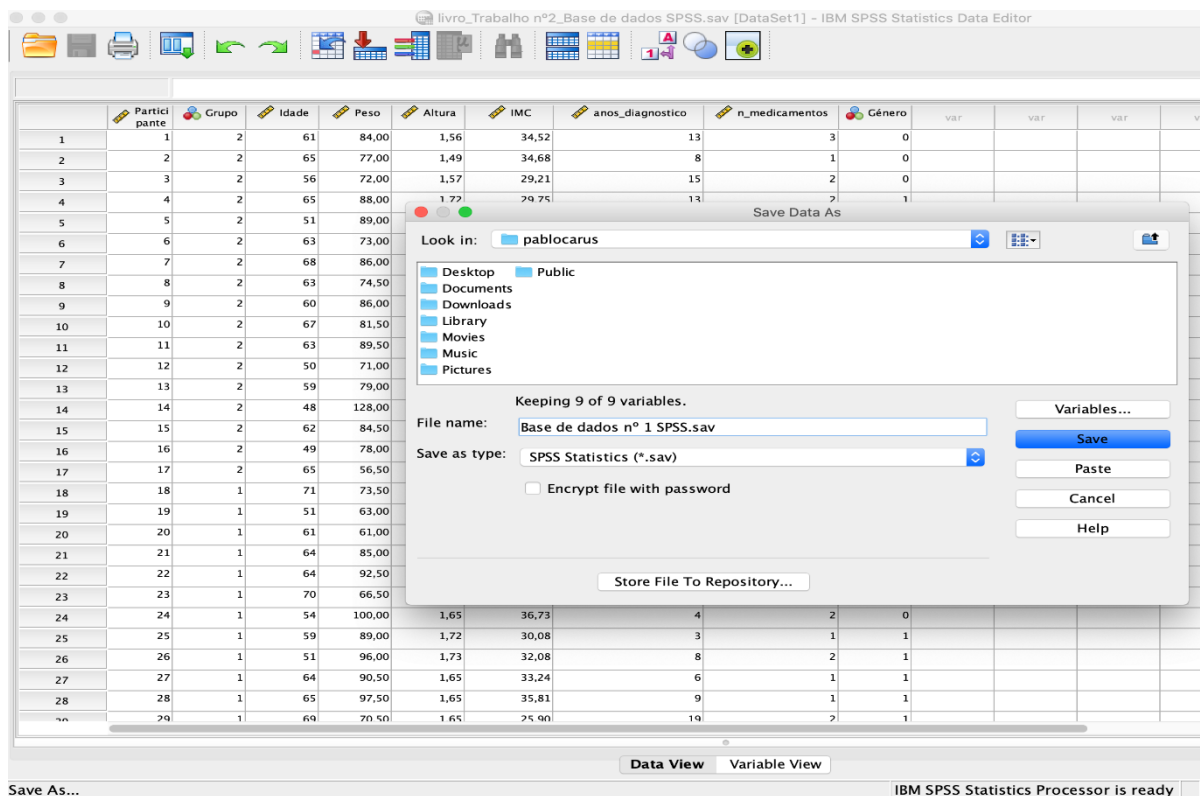


Figura 9
Indicação sobre como guardar um ficheiro de dados

2. TRANSFORMAÇÃO DE VARIÁVEIS

Depois de completar este capítulo, o aluno deverá ser capaz de:
Calcular novas variáveis, a partir das já existentes na base de dados.
Recodificar variáveis já existentes.

2.1 CÁLCULO DE NOVAS VARIÁVEIS A PARTIR DAS JÁ EXISTENTES

Por vezes, os dados do ficheiro necessitam de ser transformados antes de realizar as análises estatísticas. Na opção “Transform” → “Compute Variable” pode calcular uma nova variável, a partir de outras já existentes na base de dados (Figura 10).

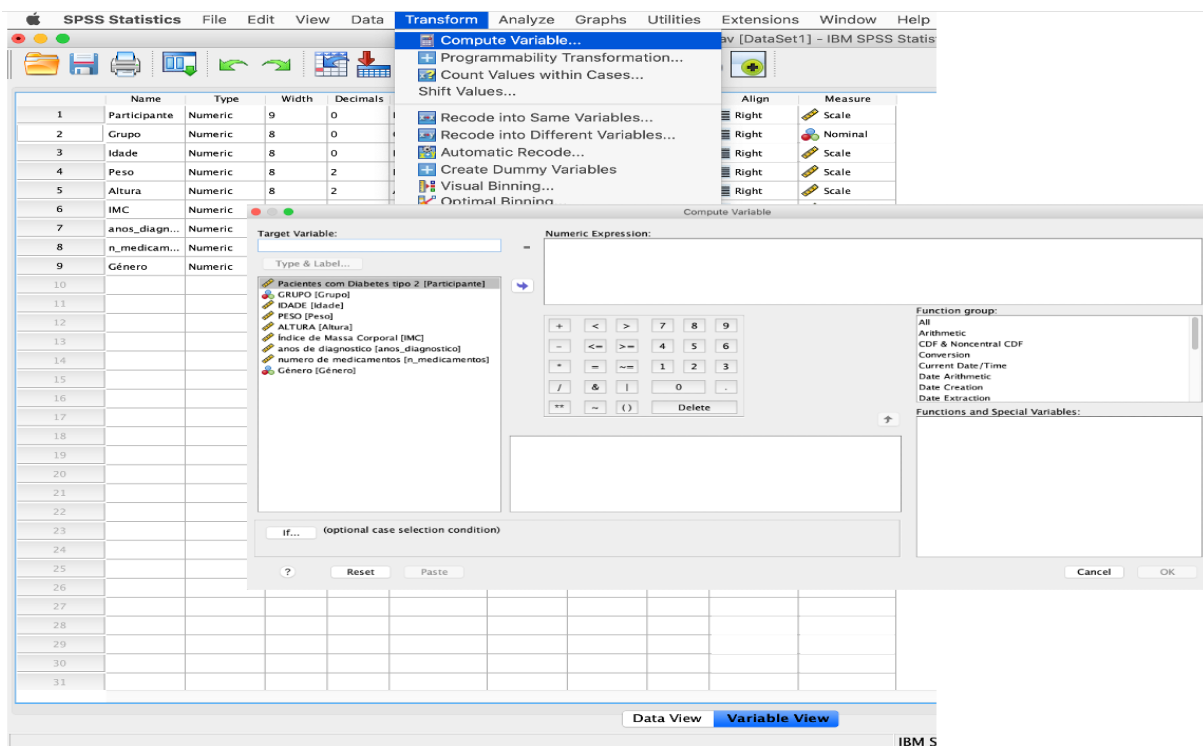


Figura 10

Indicação da opção a selecionar para calcular uma variável, a partir de outras já existentes na base de dados

Por exemplo, vamos determinar a nova variável índice de massa corporal (IMC), através da equação 1, no SPSS (Figura 11).

$$\text{Equação 1: } IMC = m/h^2$$

Onde m , é a massa corporal, em quilogramas e h é altura, em metros.

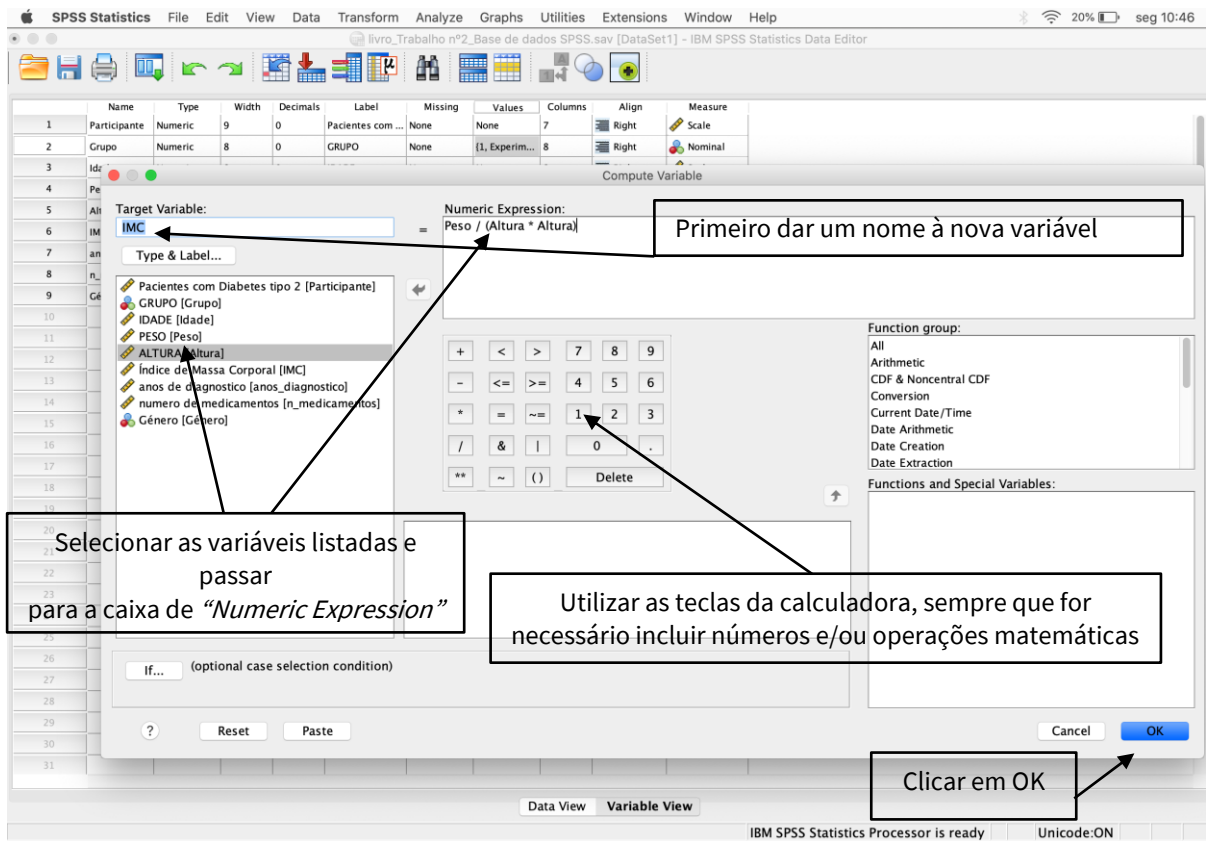


Figura 11

Indicação sobre os passos a seguir para calcular uma variável, a partir de outras já existentes

Após clicarmos em OK, podemos visualizar a nova variável "IMC" no final da lista de variáveis da base de dados na opção "Variable View" e também na última coluna da base de dados na opção "Data View", com os novos dados já calculados para cada participante (Figura 12).

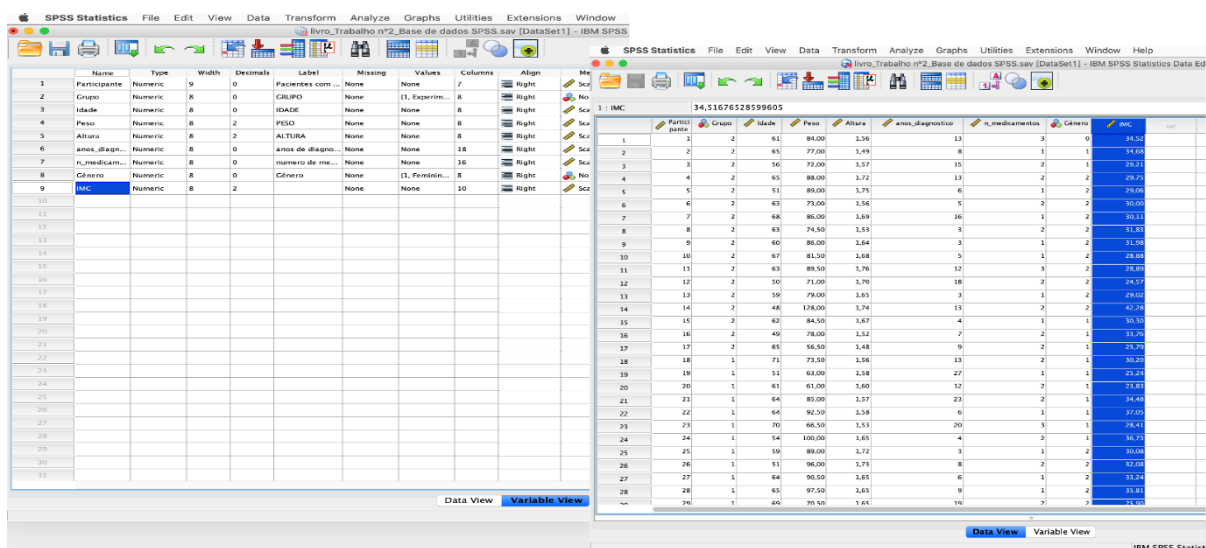


Figura 12

Indicação sobre a localização da coluna respeitante à nova variável

2.2. RECODIFICAÇÃO DE VARIÁVEIS

Para realizarmos a recodificação de variáveis existem duas possibilidades,

- (1) “*Recode into Different Variables...*” que recodifica a variável e cria uma nova variável com a dita recodificação;
- (2) “*Recode into Same Variables...*” que recodifica os dados originais da mesma variável por dados novos.

Vamos exemplificar na variável “região” com 4 categorias: 1 Norte; 2 Centro-norte; 3 Centro-sul; e 4 Sul e pretendemos reduzir para 3 categorias, 1 Norte; 2 Centro; e 3 Sul, em que vamos unificar “Centro-norte” e “Centro-sul” numa categoria “Centro”. Seleccionamos a opção “*Transform*” → “*Recode into Different Variables...*”, como apresentado na Figura 13.

Em seguida, escolhemos a variável que será recodificada, neste exemplo, “Região” e passamo-la para o “*Input Variable*”, assim o “*Output Variable*” ficará ativo para colocar o nome da nova variável, por exemplo: “Região_C”. Posteriormente, clicamos em “*Change*” e o nome da nova variável ficará no quadro como “*Output Variable*” (Figura 14).

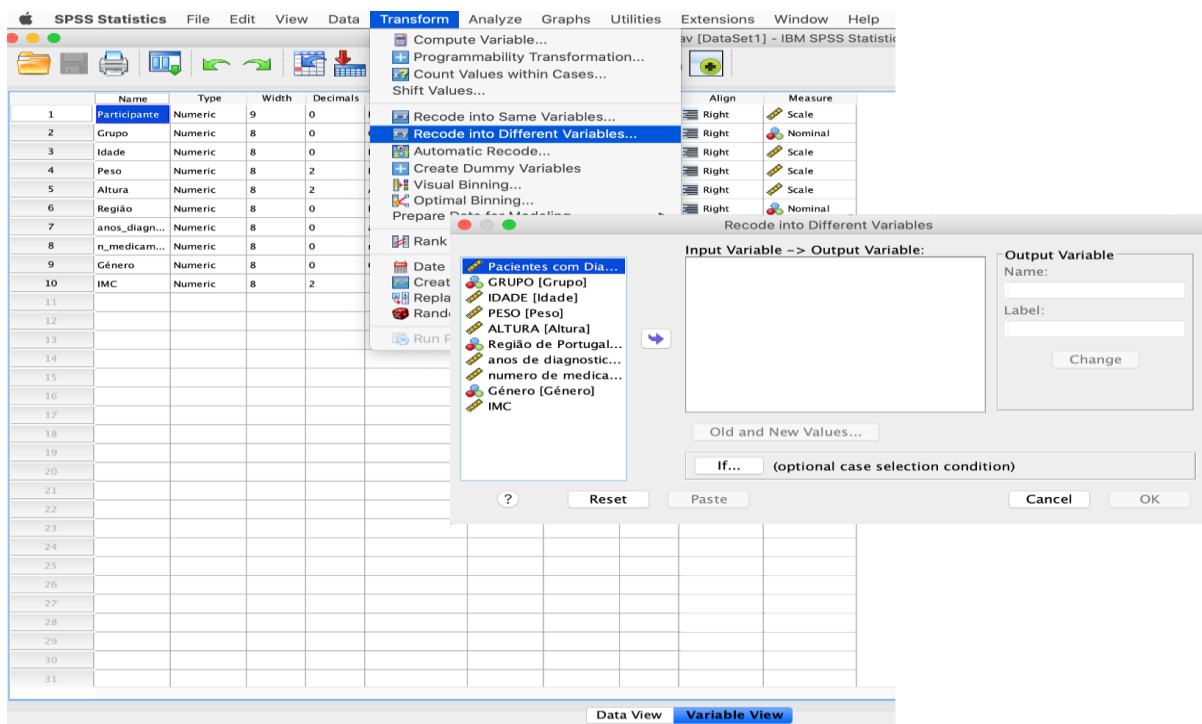


Figura 13
Indicação sobre como recodificar e criar uma nova variável

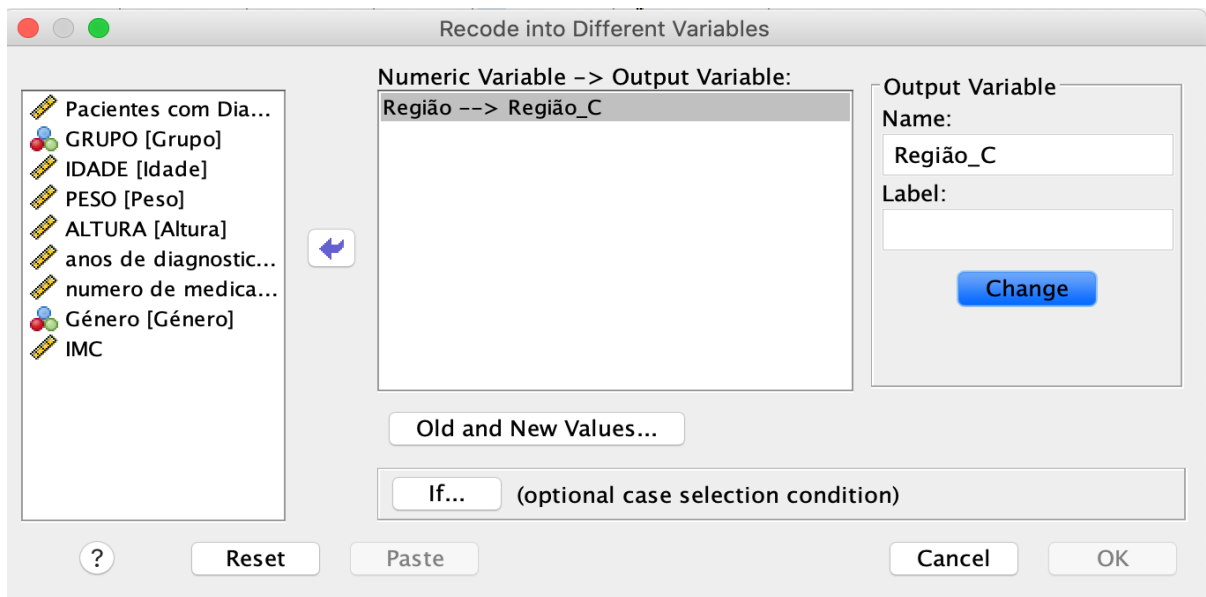


Figura 14
Janela que remete para "recode into different variables"

Para incluir os valores da nova variável, clicamos em “*Old and New Values...*” e colocamos os novos valores da nova variável com 3 categorias: 1 Norte; 2 Centro; e 3 Sul, e clicamos em “*Continue*”, como representado na Figura 15.

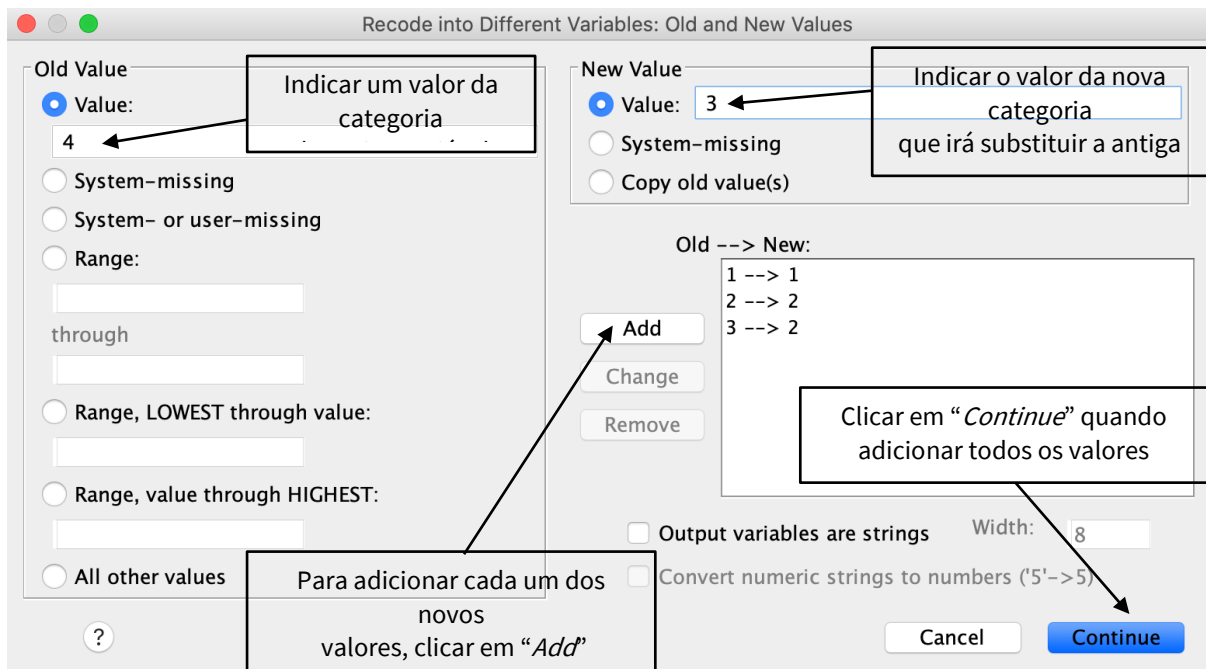


Figura 15
Janela "recode into different variables", na opção "old and new values"

Ao clicar em OK (Figura 16), o SPSS criaremos, na última linha/coluna da base de dados, a nova variável recodificada com o nome “Região_C” com os novos valores.

Na opção 2 pode recodificar uma variável sem ter a necessidade de criar uma nova variável, para isso devemos optar por “Transform” → “Recode into Same Variables...” (Figura 17).

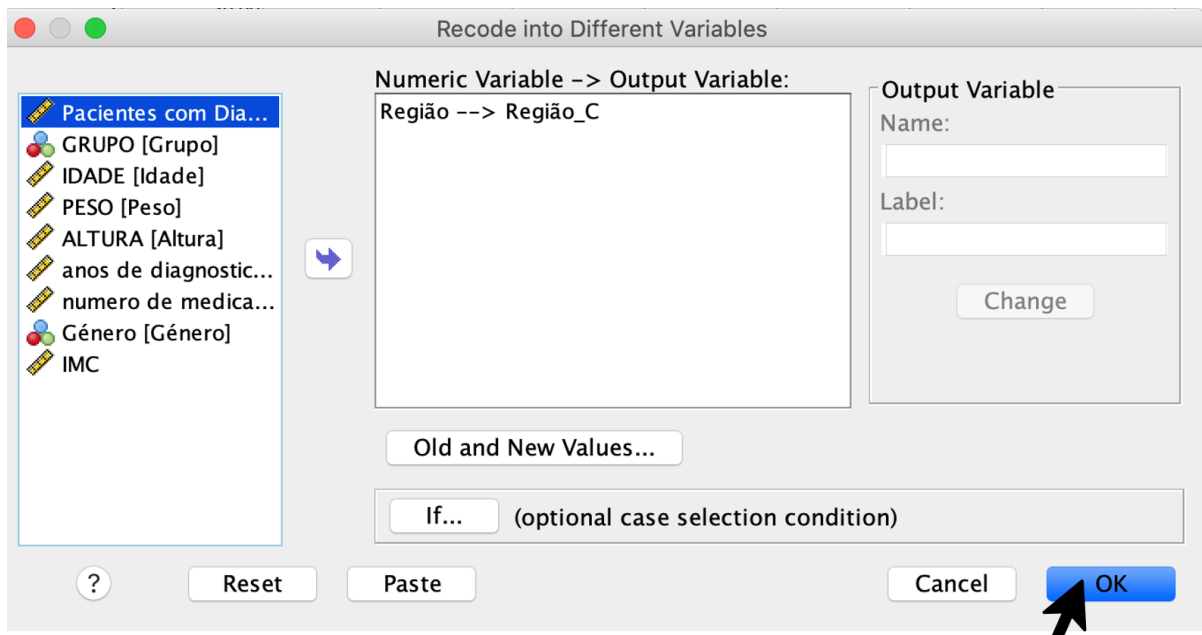


Figura 16
Janela "recode into different variables" preenchida

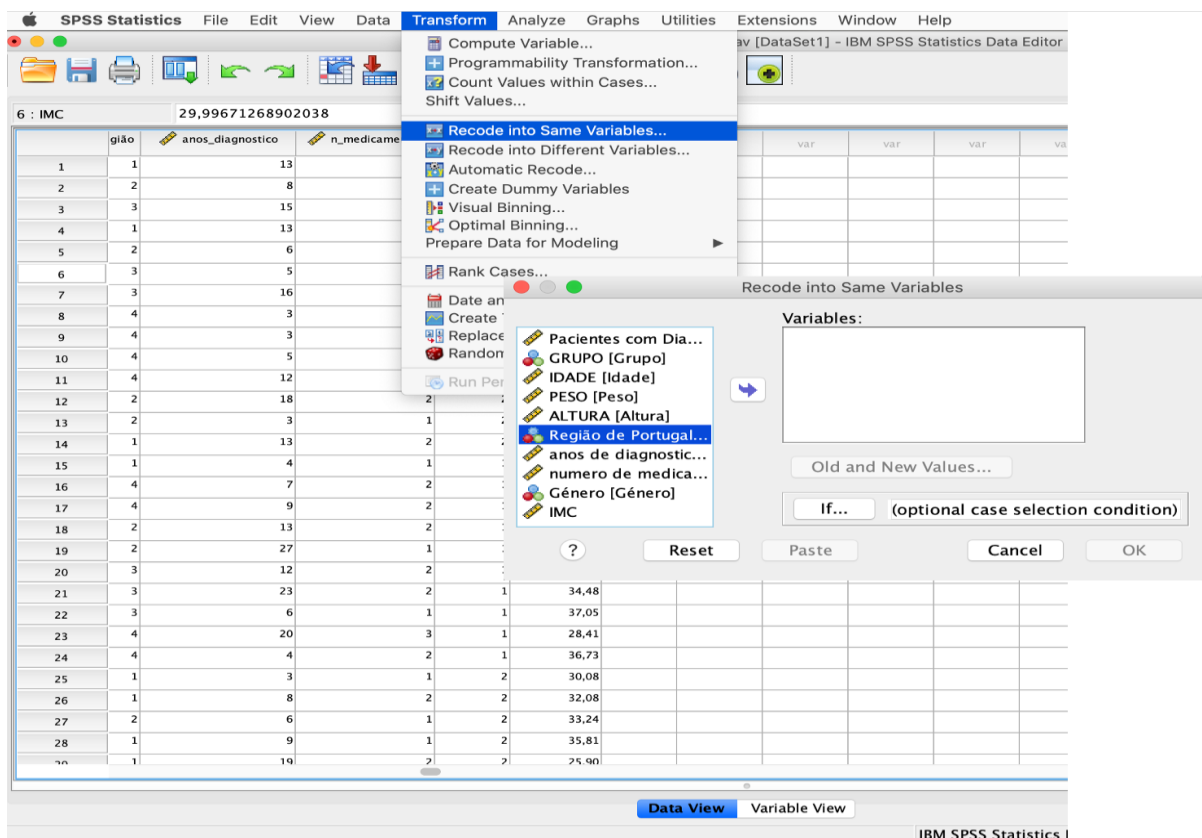


Figura 17
Indicação sobre como recodificar os dados de uma variável, por dados novos

Selecionamos a variável “Região” e passamo-la para a caixa de “Variables” e clicamos em “Old and New Values...”(Figura 18).

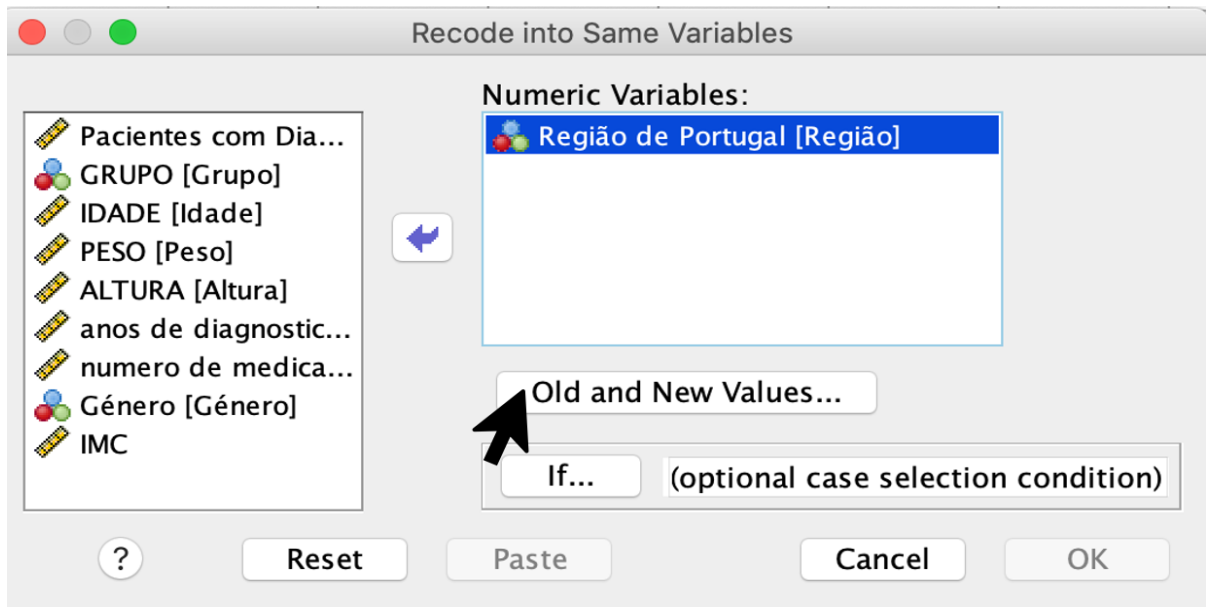


Figura 18
Janela "recode into same variables"

Recodificamos a variável substituindo os valores antigos pelos novos valores e clicamos em “Continue”(Figura 19).

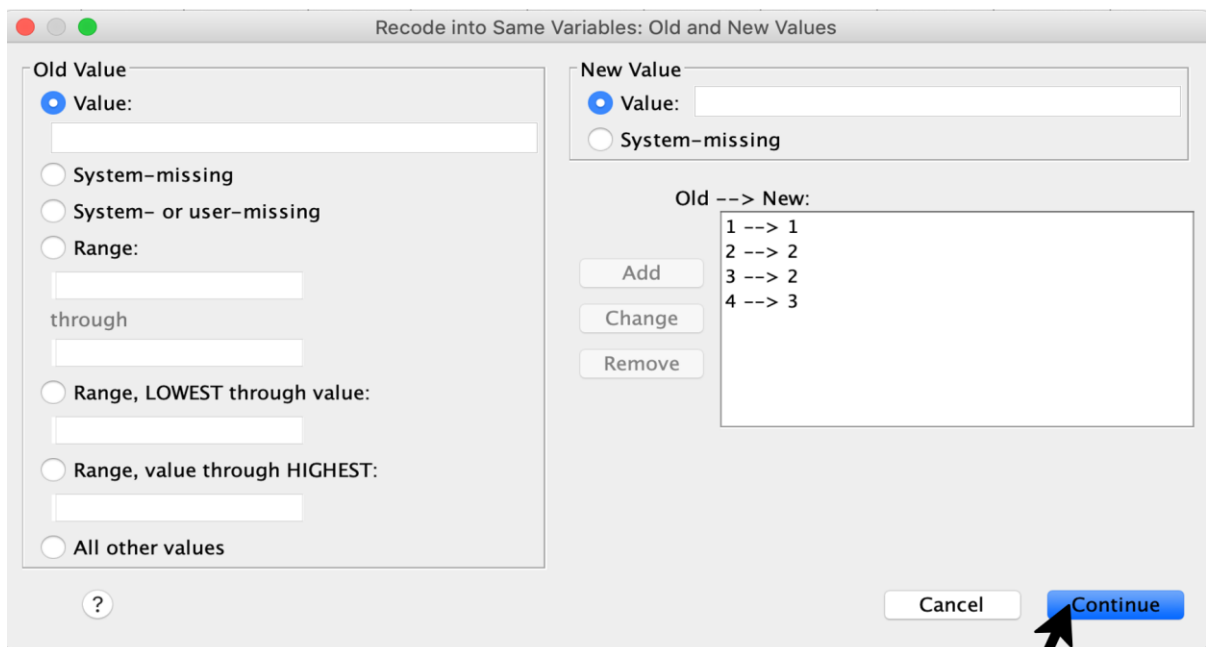


Figura 19
Janela "recode into same variables", na opção "old and new values"

Em seguida, clicamos OK (Figura 20) e o SPSS substituirá os valores indicados na variável já existente.

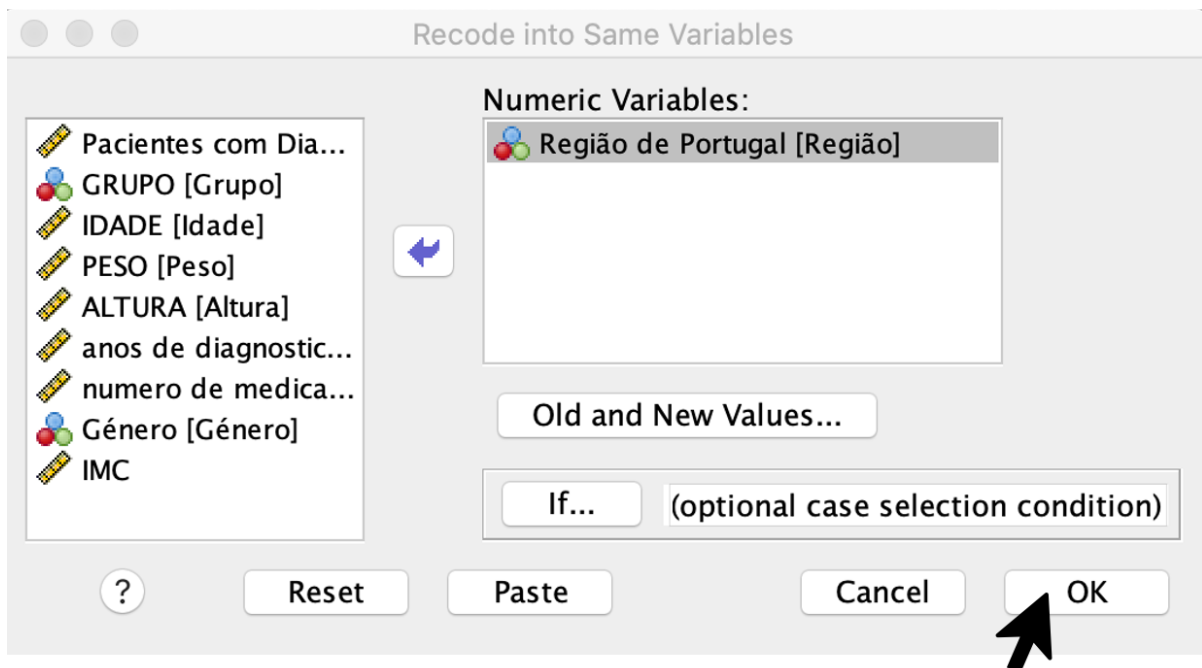


Figura 20
Janela "recode into same variables" preenchida

3. CONDIÇÕES DE APLICABILIDADE DOS TESTES PARAMÉTRICOS

Depois de completar este capítulo, o aluno deverá ser capaz de:

Conhecer as condições de aplicabilidade dos testes paramétricos.

Aprender o procedimento de computação para testar a distribuição normal dos dados e as variâncias populacionais homogêneas usando SPSS.

Interpretar os testes de normalidade e homogeneidade no visor de resultados.

3.1 CONDIÇÕES DE APLICABILIDADE DOS TESTES PARAMÉTRICOS

A comparação de parâmetros populacionais (média, mediana, variância, ...) das amostras é uma das necessidades mais frequentes nas análises estatísticas em Motricidade Humana. Para realizar este tipo de análises, existem os testes paramétricos e os testes não-paramétricos. O uso dos testes paramétricos implica o cumprimento de dois pressupostos: 1) que a variável dependente possua uma distribuição normal, e 2) que as variâncias populacionais sejam homogêneas caso se esteja a comparar duas ou mais populações (Maroco, 2018).

Com a utilização do SPSS podem resolver-se estas questões. Os testes Shapiro-Wilk e Kolmogorov-Smirnov são utilizados para testar a normalidade dos dados. O teste Shapiro-Wilk é mais adequado para testar a normalidade em amostras de pequena dimensão ($n < 30$). Relativamente, ao segundo pressuposto, este refere que as variâncias populacionais devem ser homogêneas, para realizar essa verificação utiliza-se o teste de Levene (Maroco, 2018).

3.2 TESTAR A NORMALIDADE DOS DADOS DAS VARIÁVEIS

Para verificar a distribuição normal dos dados de cada variável devemos selecionar a opção: **Analyze** → **Descriptive Statistics** → **Explore** (Figura 21).

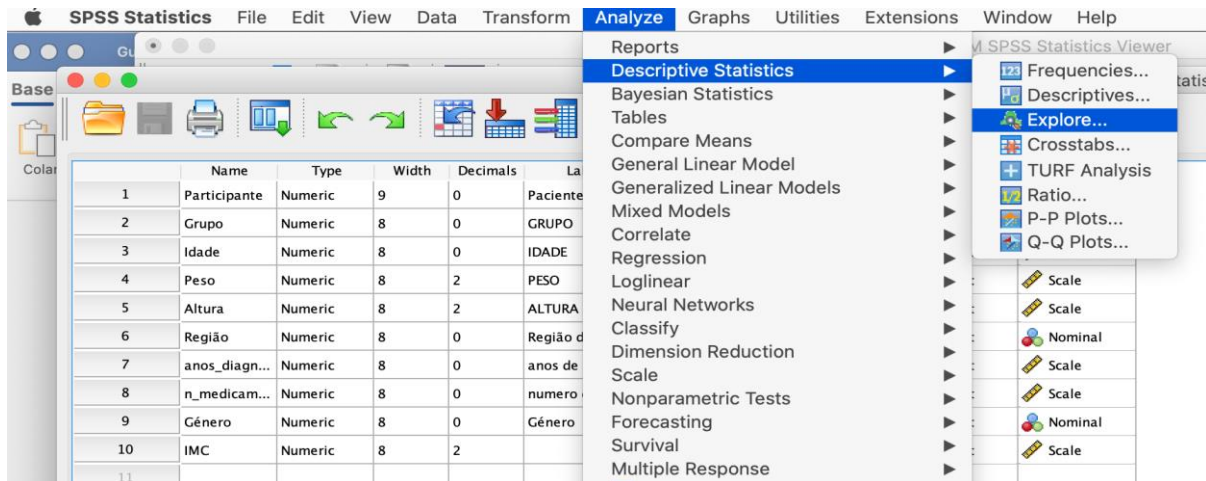


Figura 21

Indicação da opção a seleccionar para verificar a distribuição normal dos dados de cada variável

Em seguida colocamos a(s) variável(eis), na caixa de “*Dependente List*”, e clicamos em “*Plots...*”, na parte superior direita da janela (Figura 22).

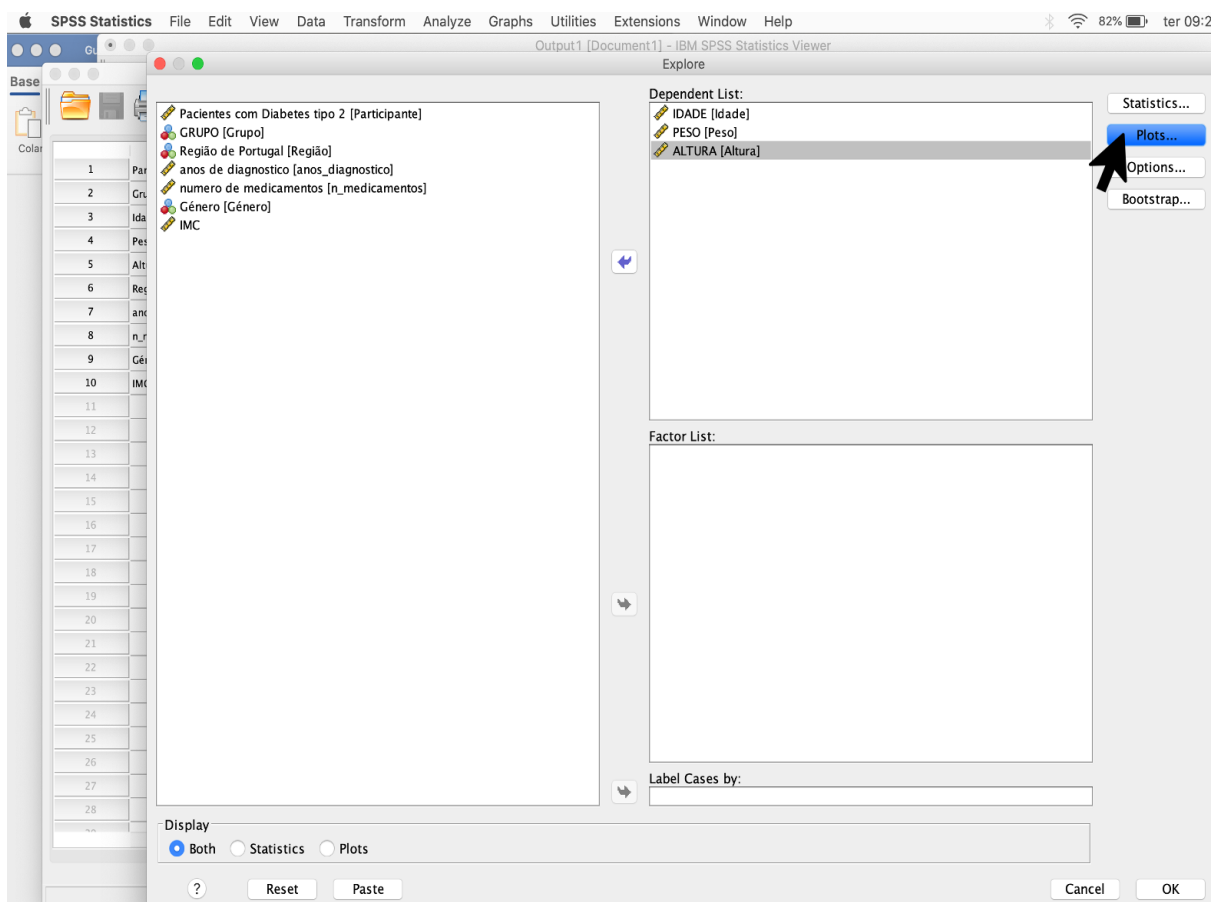


Figura 22

Janela "explore", com indicação para seleccionar a opção "plots"

Na janela que surge ao selecionar a opção “Explore: Plots”, devemos selecionar a opção de “Normality plots with tests”, sendo que as outras opções podem ficar tal como aparecem. A seguir clicamos em “Continue” e OK (Figura 23).

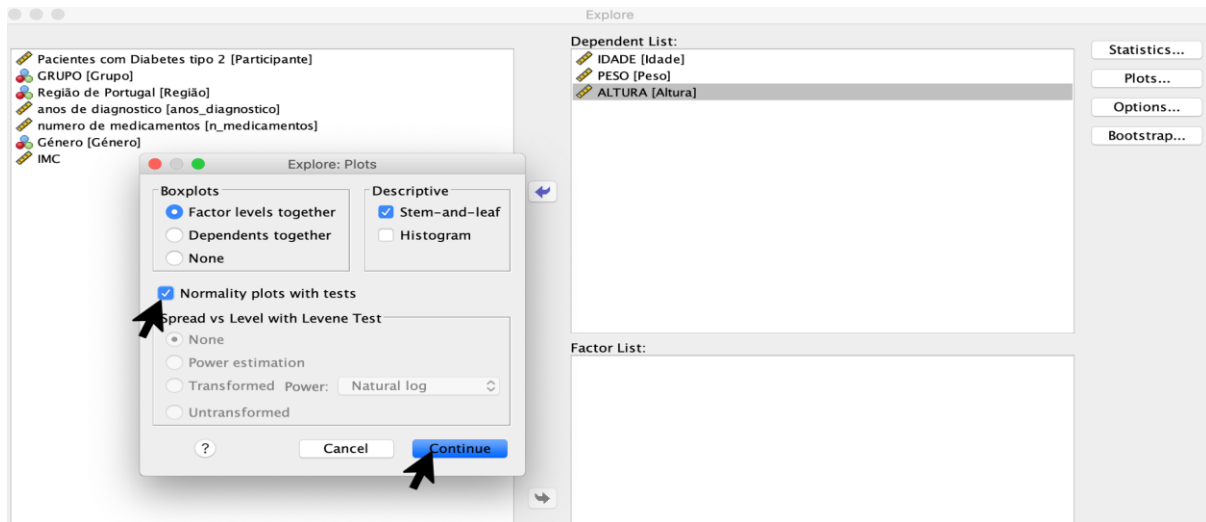
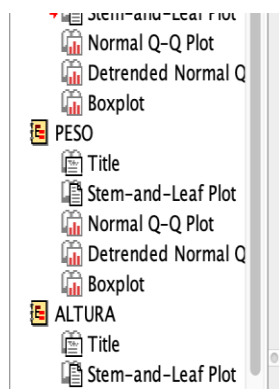


Figura 23

Indicação das opções a selecionar na janela " explore: plots", para obter o teste de normalidade

A janela de resultados mostra os mesmos resultados obtidos para os testes de Kolmogorov-Smirnov e Shapiro-Wilk. Se estes testes forem significativos, então os dados apresentam uma distribuição não-normal. Assim, para que os dados apresentem uma distribuição normal, estes testes não devem ser significativos, ou seja, devem ter um p -value superior a 0,05. No exemplo, uma vez que a amostra é menor que 50 ($N=32$), devemos dar atenção ao teste de Shapiro-Wilk, sendo que podemos observar que o resultado é significativo para a variável “IDADE” ($p = 0,023$), mas não é significativo para o “PESO” ($p = 0,160$), nem para a “ALTURA” ($p = 0,160$), pelo que se pode concluir que os dados “PESO” e “ALTURA” apresentam uma distribuição normal, enquanto que a variável “IDADE” apresenta uma distribuição não-normal dos dados (Figura 24).



Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
IDADE	,169	32	,021	,922	32	,023
PESO	,112	32	,200*	,952	32	,160
ALTURA	,139	32	,122	,952	32	,160

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

a. Lilliefors Significance Correction

Figura 24

Janela de resultados referente ao teste de normalidade

3.3 TESTAR A HOMOGENEIDADE DAS VARIÂNCIAS

Na análise das condições de testes paramétricos é necessário também analisar se as variâncias são homogêneas. Na Figura 24, verificamos que o SPSS não obteve resultados para o teste de Levene, pois o mesmo aparece desativado. Isto sucede porque não estamos a comparar duas ou mais populações e não colocando uma variável na caixa de “*Factor List*”, como por exemplo: a variável “GRUPO”, o SPSS não irá apresentar na janela de resultados a homogeneidade das variâncias, ou seja, os resultados para o teste de Levene. Para isso, temos que seleccionar a opção: **Analyze** → **Descriptive Statistics** → **Explore**, a seguir colocamos as variáveis dependentes na caixa de “*Dependente List*” e na caixa de “*Factor List*” colocamos a variável que desejamos comparar, por exemplo a variável GRUPO (No capítulo 4 será explicado como dividir uma variável em grupos), para comparar os grupos experimental e de controlo da nossa base de dados. Depois clicamos em “*Plots...*” (Figura 25).

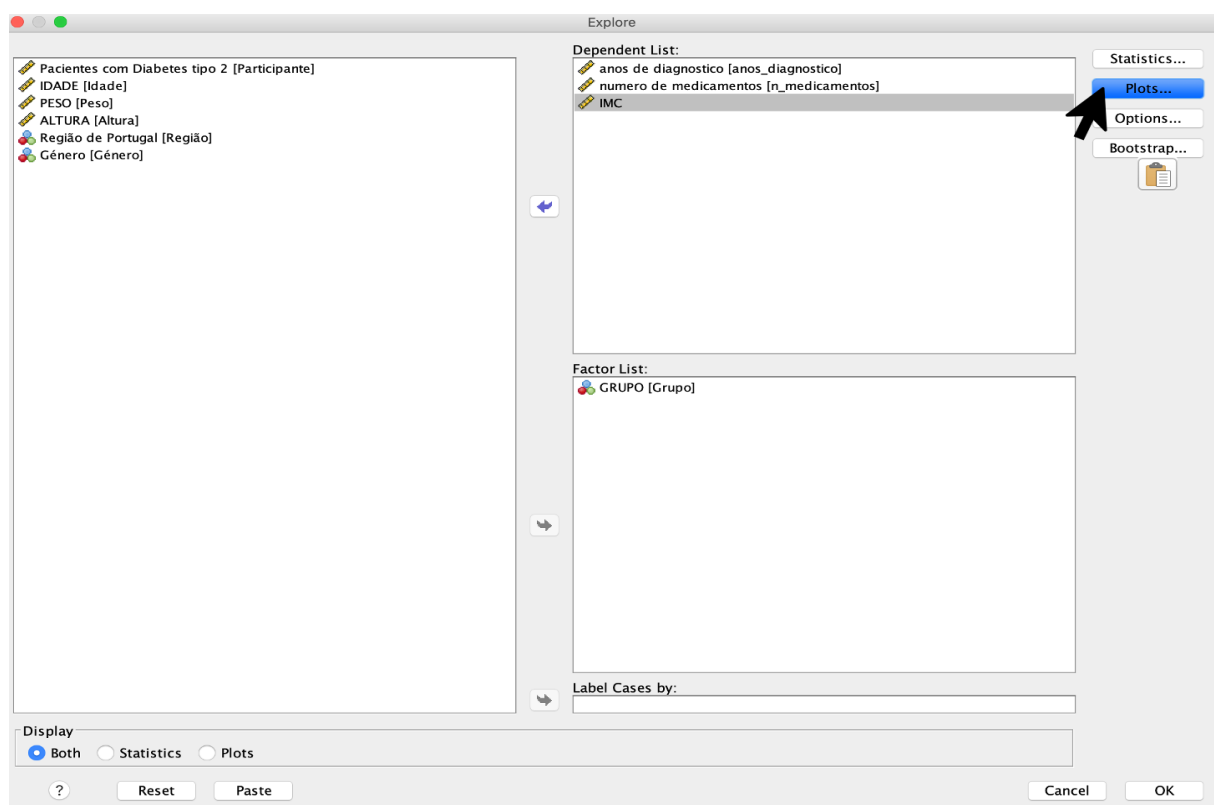


Figura 25
Janela "explore", com indicação para seleccionar a opção "plots"

No quadro que surge em “*Explore: Plots*”, devemos seleccionar a opção “*Untransformed*” no “*Spread vs Level with Levene Test*” para obter a homogeneidade das variâncias. Também podemos seleccionar, ao mesmo tempo, a normalidade dos dados com a opção de “*Normality plots with tests*” (como explicado no ponto 4.2) e, na mesma análise também será apresentada a normalidade dos dados. As outras opções podem ficar tal como surgem. A seguir clicamos em “*Continue*” e OK (Figura 26).

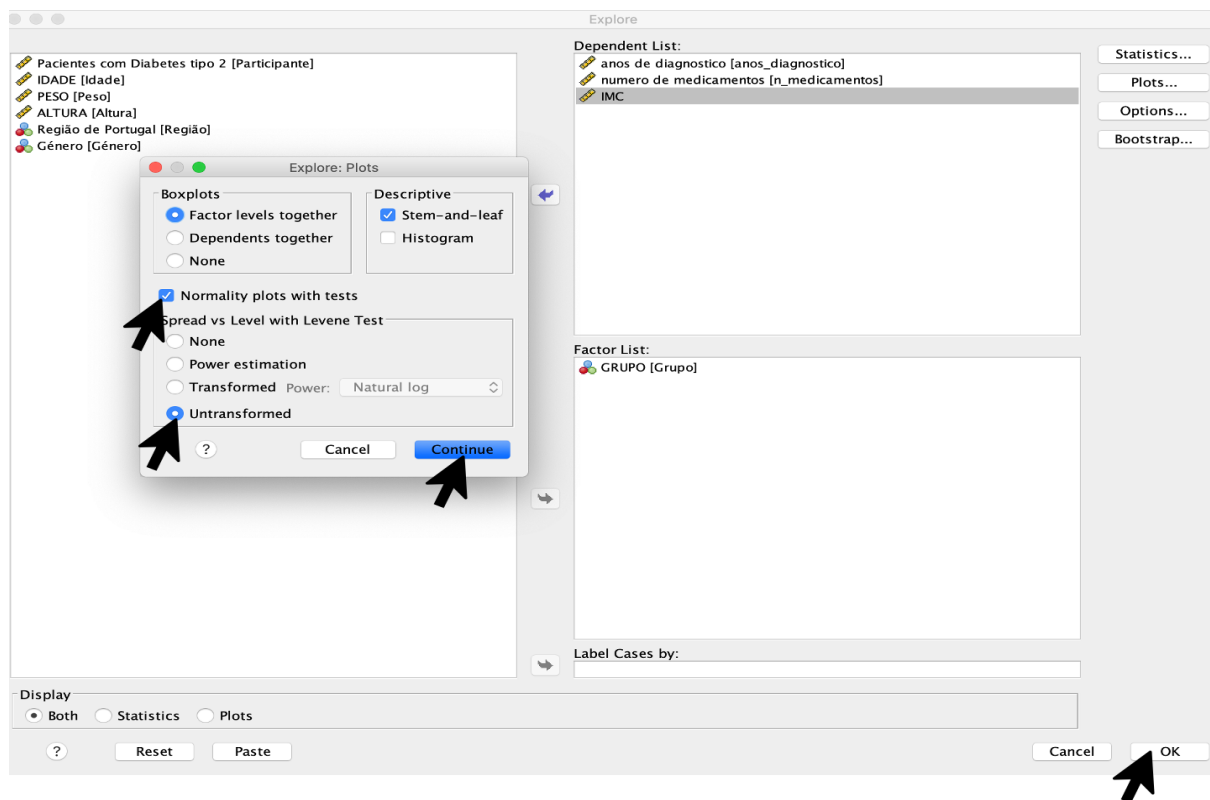


Figura 26

Indicação das opções a selecionar na janela " *explore: plots*", para obter o teste de homogeneidade

A janela de resultados apresentada na Figura 27 mostra uma tabela com os testes de Normalidade (Kolmogorov-Smirnov e Shapiro-Wilk), que já foram explicados no ponto 3.2, mas adicionámos resultados para os grupos experimental e de controlo, pois adicionámos a variável GRUPO no "Factor List". Na segunda tabela, a janela de resultados mostra o teste de homogeneidade das variâncias (teste de Levene). Como estamos a comparar médias, devemos dar atenção aos valores de p "Based on Mean". Se o teste de Levene é significativo, então as variâncias não são homogéneas. As variâncias são homogéneas, se o teste não apresentar um valor com diferença estatisticamente significativa. No exemplo (Figura 27), observamos que não é significativo para nenhuma das três variáveis estudadas: "anos de diagnóstico" (0,099), "número de medicamentos" (0,980), e "IMC" (0,257). Assim, podemos concluir que as três variáveis apresentam variâncias homogéneas.

Em suma, para aplicar testes paramétricos numa variável, esta deve ser homogénea e apresentar uma distribuição normal dos dados. No exemplo apresentado na Figura 27 verificámos um $N < 50$ ($N=32$), e assim analisámos o valor de p do teste de Shapiro-Wilk. Verificámos que na variável "anos de diagnóstico" os valores de p não são significativos ($p = 0,248$ para o grupo experimental e $p = 0,112$ para o grupo de controlo), o que indica que existe uma distribuição normal dos dados nesta variável. No entanto, para a variável "número de medicamentos" os valores de p são estatisticamente significativos para o grupo experimental e para o grupo de controlo ($p = 0,006$ e $p = 0,001$; respetivamente) concluindo que os dados desta variável apresentam uma distribuição não normal. Por último, a variável "IMC" o valor de p para o grupo experimental não é estatisticamente significativo ($p = 0,277$), contudo o valor de p para o grupo de controlo é estatisticamente

significativo ($p=0,032$). Nesta última variável podemos concluir que a mesma, não cumpre uma distribuição normal dos dados, ainda que seja cumprido num dos grupos, no outro grupo o mesmo não se verifica, pois, para considerar que uma variável apresenta uma distribuição normal dos dados, os valores de p de todos os grupos não devem ser estatisticamente significativos.

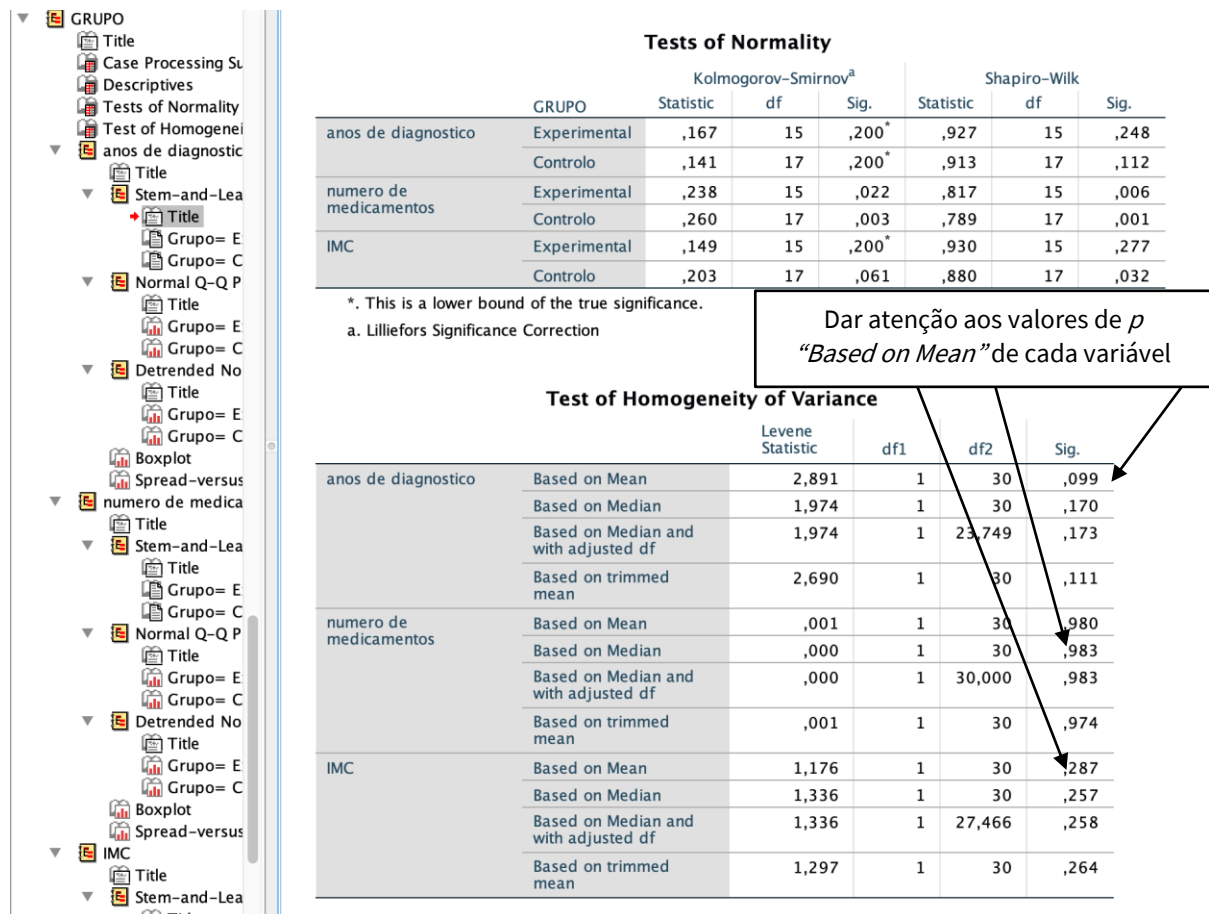


Figura 27
Janela de resultados referente ao teste de homogeneidade

De uma forma mais prática, e para saber se deve aplicar testes paramétricos ou a sua alternativa não paramétrica, basta ter como referência a Tabela 1, que resume o exemplo apresentado.

Como pode verificar, só quando cumpre a normalidade e homogeneidade das variâncias é que podemos aplicar testes paramétricos. Pelo contrário, quando a variável não cumpre alguma das duas condições (normalidade e/ou homogeneidade), temos que aplicar um teste não paramétrico, como alternativa.

Variável	Normalidade	Homogeneidade	Teste paramétrico ou teste não paramétrico
anos de diagnostico	Grupo Experimental: V Grupo de Controlo: V	V	Teste paramétrico
número de medicamentos	Grupo Experimental: X Grupo de Controlo: X	V	Teste não paramétrico
IMC	Grupo Experimental: V Grupo de Controlo: X	V	Teste não paramétrico

V: cumpre; X: não cumpre

Tabela 1
Síntese sobre quando aplicar testes paramétricos ou não paramétricos

4. ESTATÍSTICA DESCRITIVA

Depois de completar este capítulo, o aluno deverá ser capaz de:

Compreender a aplicação da estatística descritiva na investigação em motricidade humana.

Aprender o procedimento de computação de várias estatísticas descritivas usando o SPSS.

Interpretar as estatísticas descritivas no visor de resultados.

4.1 MEDIDAS DE TENDÊNCIA CENTRAL E DE DISPERSÃO

Quando se descreve uma variável através da estatística descritiva, o que se pretende é reduzir um conjunto de números a índices numéricos, que representem adequadamente esse conjunto de números. Determinados índices representam uma descrição reducionista, pois reduzem muito o conjunto de números, contudo é precisamente o que se pretende, realizar uma descrição sintética dos resultados. As principais estatísticas descritivas são: as medidas de tendência central e as medidas de dispersão ou variabilidade.

Nas medidas de tendência central, a média é a estatística mais utilizada com as variáveis numéricas contínuas (no SPSS “*Scale*”). Esta representa o valor médio de todos os valores e, consiste na soma dos valores dividido pelo número total de valores. Com variáveis ordinais, (no SPSS “*Ordinal*”) a medida de tendência central mais adequada é a mediana. A mediana é o valor de um grupo de dados que ordenados por ordem crescente, 50% dos dados devem ser superiores e os restantes 50% devem ser inferiores. Com variáveis nominais (“*Nominal*” no SPSS), a medida de tendência central mais utilizada é a moda, esta, por sua vez, corresponde ao valor mais frequente.

A dispersão ou variabilidade da amostra indica o quão separados ou agrupados estão os valores. A medida adequada para variáveis numéricas contínuas ou “*Scale*” é a variância. A variância é o quadrado do desvio padrão, e é calculada através da média quadrática das pontuações diferenciais. No entanto, quando se apresenta uma média, é habitual usar o desvio padrão como medida de dispersão, sendo que este é calculado como a raiz quadrada da variância.

Neste sentido, é importante que os documentos científicos apresentem as estatísticas descritivas de todas as variáveis, preferencialmente, em tabelas, na seção de resultados. Assim propicia-se a comparação com os resultados de outras investigações. Apesar deste capítulo ter como objetivo abordar a estatística descritiva, nos capítulos seguintes será explicado como realizar uma tabela científica através da inclusão do nome das variáveis, unidades de medida, *p-values*, *eta square*, etc.

4.2 ESTATÍSTICA DESCRITIVA USANDO O SPSS

O SPSS permite realizar várias análises de estatística descritiva. Nesta seção será mostrado de que forma podemos aplicar a estatística descritiva diretamente.

4.2.1 VARIÁVEIS NUMÉRICAS CONTÍNUAS OU “SCALE”

Um exemplo de analisar a estatística descritiva de uma variável numérica contínua ou “Scale” é através do comando “Descriptives”: **Analyze** → **Descriptive Statistics** → **Descriptives** (Figura 28).

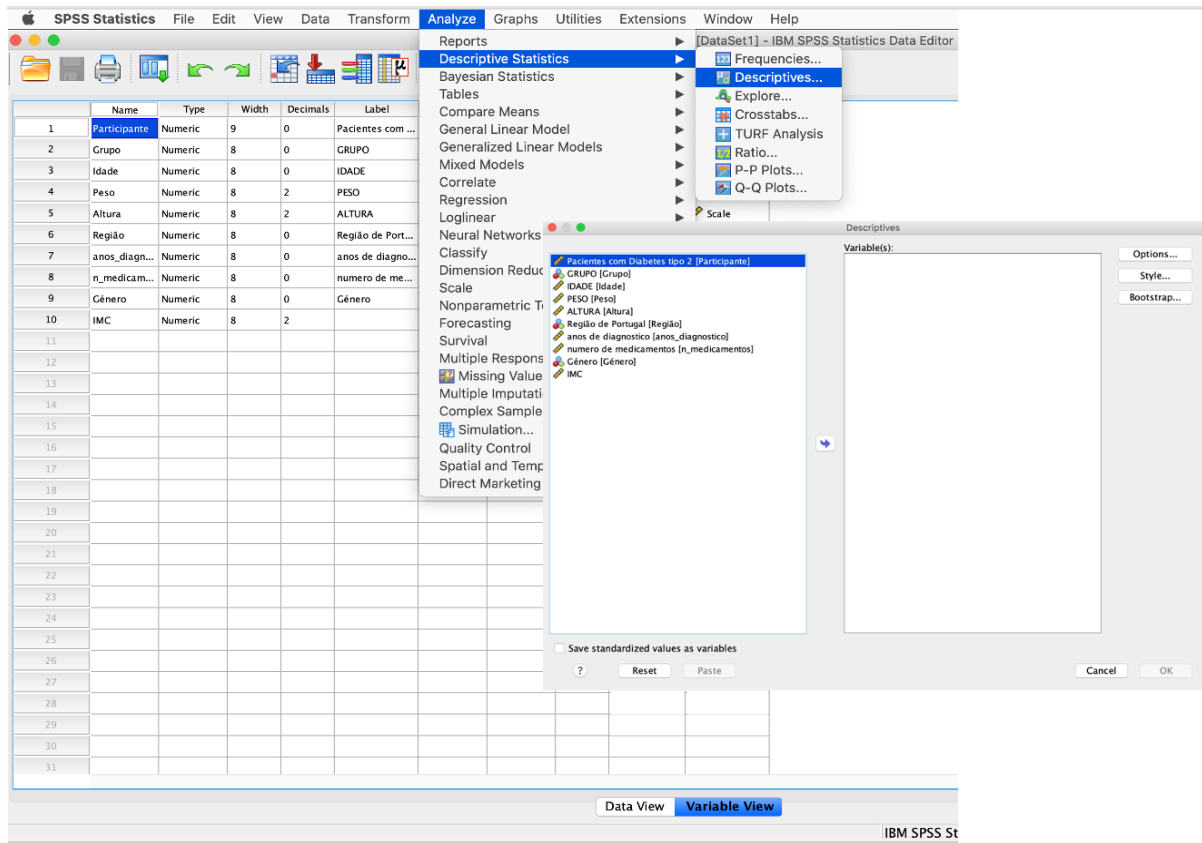


Figura 28
Indicação sobre as opções a selecionar para realizar a análise de estatística descritiva de uma variável numérica contínua

Em seguida, selecionamos a(s) variável(eis) (só do tipo “Scale”) da lista de variáveis, passamo-la(s) para a caixa de “Variable(s)” e clicamos OK (Figura 29).

No SPSS surgirá a janela com os resultados selecionados e a Figura 30 exemplifica a informação obtida.

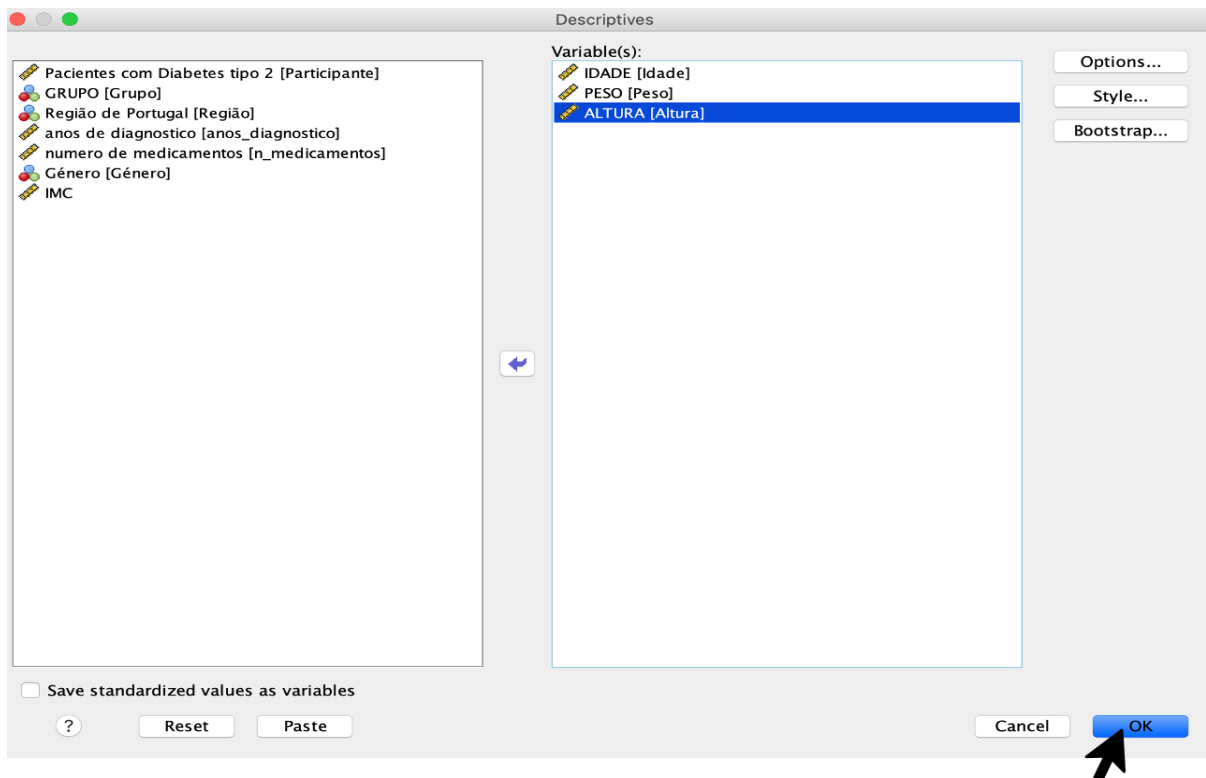


Figura 29
Exemplo de variáveis numéricas contínuas a selecionar para realizar a análise de estatística descritiva

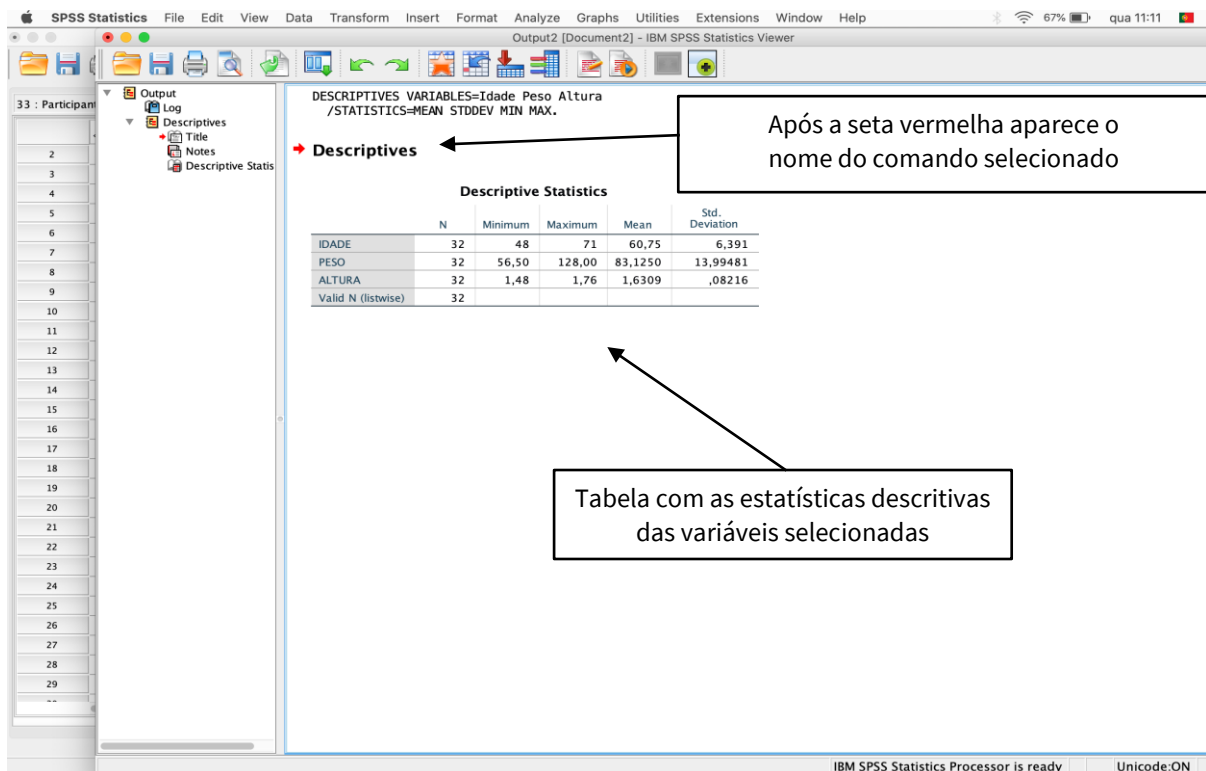
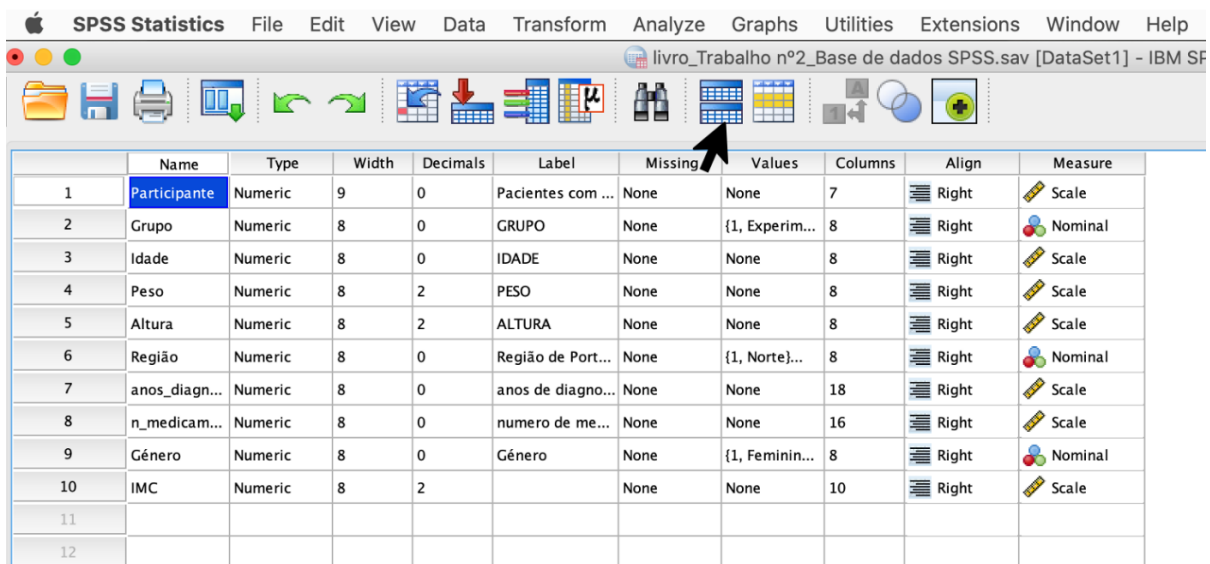


Figura 30
Janela de resultados da análise de estatística descritiva de variáveis numéricas contínuas

Na tabela apresentada podemos ver a lista das variáveis selecionadas com o número de participantes (“N”), o valor mínimo (“Minimum”) e o valor máximo (“Maximum”), a média (“Mean”) e o desvio padrão (“Std. Deviation”) (Figura 31). No exemplo, a variável idade: contém 32 participantes, sendo que a pessoa mais nova tem 48 anos e a mais velha tem 71 anos, a média de idade dos participantes é de 60,75 anos e o desvio padrão é 6,39 anos (o mais correto será apresentar da seguinte forma: 60,75 ± 6,39 anos).

Por vezes, pode ser necessário expressar os resultados das estatísticas descritivas em subgrupos. Por exemplo se pretendermos determinar a média ± desvio padrão da idade para o género feminino e masculino, realizamos um simples passo, que deve acontecer antes da análise descritiva, este consiste na divisão do ficheiro de dados, isto é, dividir o ficheiro de dados, considerando a variável “Género”. Para isso, clicamos no ícone “Split File”, na barra de opções da parte superior da base de dados.



	Name	Type	Width	Decimals	Label	Missing	Values	Columns	Align	Measure
1	Participante	Numeric	9	0	Pacientes com ...	None	None	7	Right	Scale
2	Grupo	Numeric	8	0	GRUPO	None	{1, Experim...	8	Right	Nominal
3	Idade	Numeric	8	0	IDADE	None	None	8	Right	Scale
4	Peso	Numeric	8	2	PESO	None	None	8	Right	Scale
5	Altura	Numeric	8	2	ALTURA	None	None	8	Right	Scale
6	Região	Numeric	8	0	Região de Port...	None	{1, Norte}...	8	Right	Nominal
7	anos_diagn...	Numeric	8	0	anos de diagno...	None	None	18	Right	Scale
8	n_medicam...	Numeric	8	0	numero de me...	None	None	16	Right	Scale
9	Género	Numeric	8	0	Género	None	{1, Feminin...	8	Right	Nominal
10	IMC	Numeric	8	2		None	None	10	Right	Scale
11										
12										

Figura 31
Indicação da localização do ícone de “split file”

Em seguida, devemos selecionar a opção “Compare groups”, que permitirá passar a variável “Género” para a caixa de “Groups Based on:”, e clicamos OK (Figura 32).

Na janela de resultados surge a indicação de que o ficheiro está dividido por género, através da mensagem representada na Figura 33.

A partir deste momento podemos voltar a aplicar as análises de estatísticas descritivas (lembre: **Analyze → Descriptive Statistics → Descriptives**). Colocamos a variável “Idade” (pode qualquer uma das várias variáveis tipo “Scale”) na caixa “Variable(s)” e clicamos OK (Figura 28 e 29). O utilizador volta a visualizar a janela de resultados do SPSS com a tabela das análises de estatística descritiva para o género feminino e masculino (Figura 34).

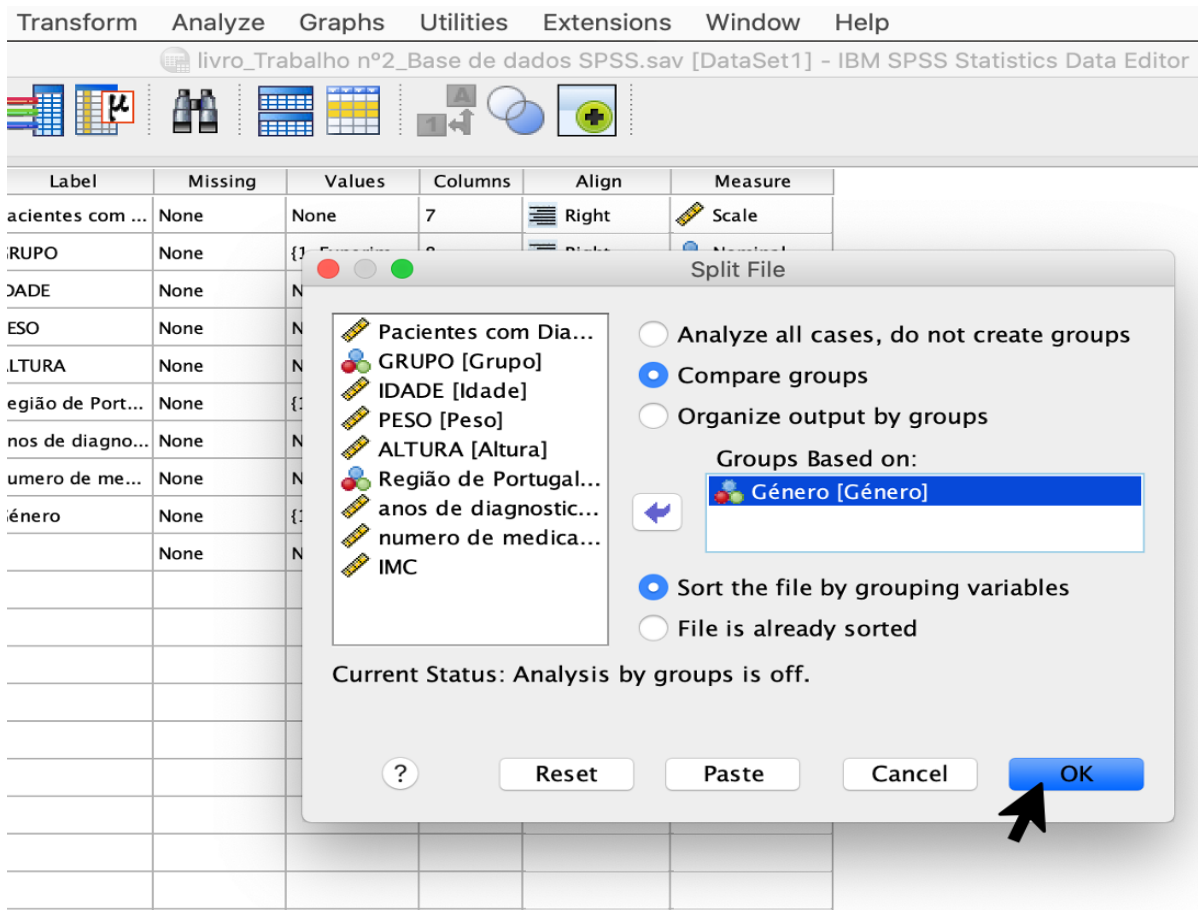


Figura 32
Indicação das opções a seleccionar para realizar "split file"

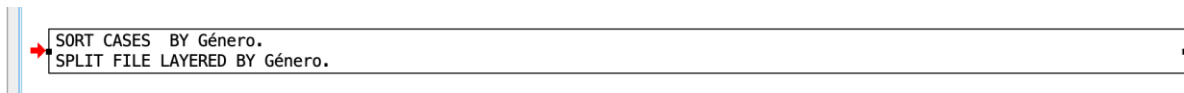


Figura 33
Indicação de que o ficheiro está dividido por género

→ **Descriptives**

		Descriptive Statistics				
Género		N	Minimum	Maximum	Mean	Std. Deviation
0	IDADE	1	61	61	61,00	.
	Valid N (listwise)	1				
Feminino	IDADE	13	49	71	61,23	6,821
	Valid N (listwise)	13				
Masculino	IDADE	18	48	69	60,39	6,427
	Valid N (listwise)	18				

Figura 34
Janela de resultados da análise de estatística descritiva para o género feminino e masculino

4.2.2 VARIÁVEIS CATEGÓRICAS (“NOMINAL / ORDINAL”)

Analisar a estatística descritiva de uma variável categórica ou “Nominal/Ordinal” do, forma direta, pode ser realizada através do comando frequências: **Analyze** → **Descriptive Statistics** → **Frequencies**. Em seguida, colocamos a variável(eis) categórica(s), que pretende analisar na caixa de “Variable(s)” e clicamos OK (Figura 35).

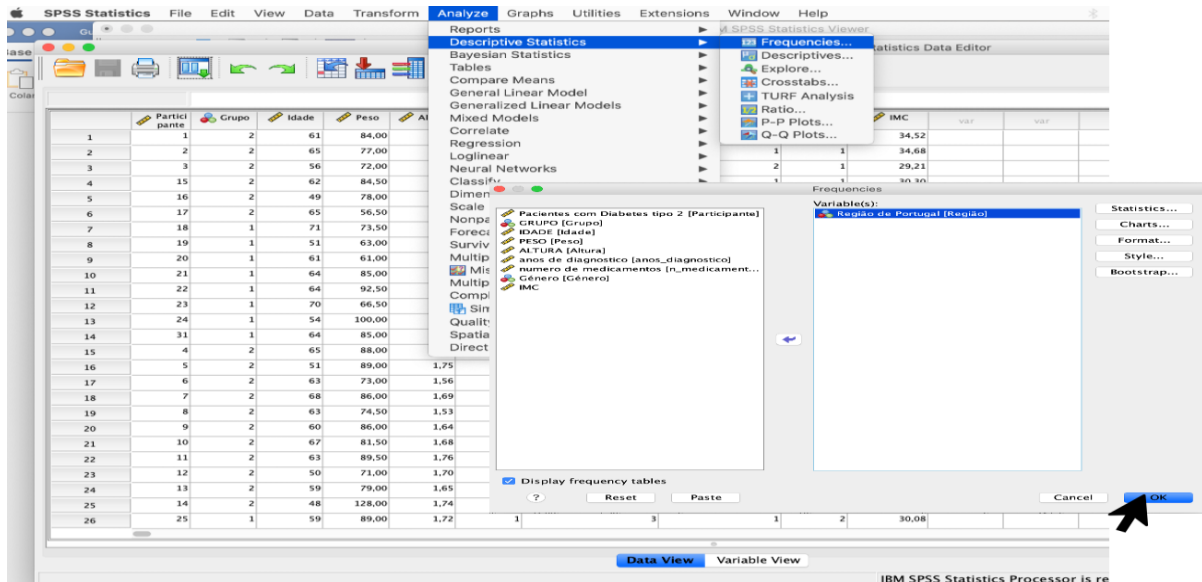


Figura 35

Indicação sobre as opções a selecionar para realizar a análise de estatística descritiva de uma variável categórica

Após clicarmos em OK, será apresentado o visor de resultados.

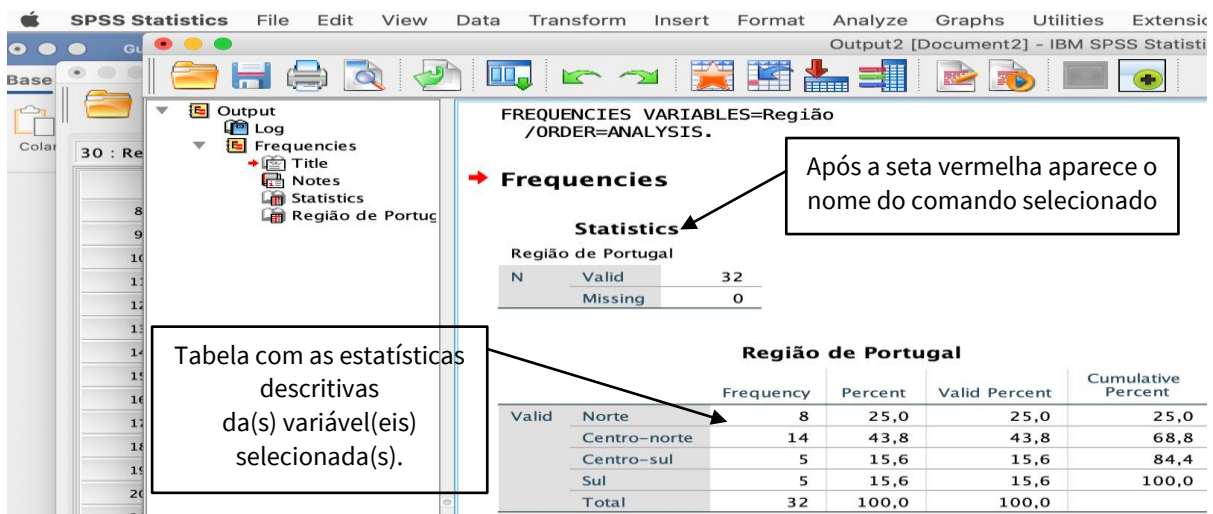


Figura 36

Janela de resultados com a análise de estatística descritiva das variáveis categóricas

A janela de resultados mostra a tabela da variável “Região de Portugal”. Ao interpretarmos a tabela da Figura 36 podemos verificar as quatro categorias da variável listadas na parte esquerda da tabela (Norte, Centro-norte, Centro-sul e Sul). A coluna seguinte “*Frequency*” representa o número de participantes que pertencem a cada categoria (neste caso Região). Na coluna seguinte “*Percent*”, analisamos a percentagem dos participantes por categoria. Neste caso, podemos descrever que 8 indivíduos pertencem à região Norte, representando 25% de um total de 32 indivíduos; 14 indivíduos pertencem à região Centro-norte (representa 43,8% do total do grupo); 5 indivíduos pertencem à região Centro-sul (representa 15,6% do total do grupo); e 5 indivíduos pertencem à região Sul (representa 15,6% do total do grupo).

Outro exemplo, pode ser obter os resultados das análises de estatística descritiva da variável “Região de Portugal”, mas divididos pela variável “Género” (Lembro o ponto 4.2.1, devemos dividir, previamente, o ficheiro de dados usando ícone “*Split File*”, na barra de opções da parte superior da base de dados). Em seguida, devemos selecionar a opção “*Compare groups*”, que permitirá passar a variável “Género” para a caixa de “*Groups Based on:*” e, para finalizar, clicamos OK.

Uma vez que o ficheiro se encontra dividido pela variável “Género”, as análises da estatística descritiva realizam-se, através do comando frequências: **Analyze → Descriptive Statistics → Frequencies**. Em seguida, colocamos a variável “Região de Portugal” na caixa de “*Variable(s)*” e clicamos OK. A interpretação dos resultados é igual ao explicado anteriormente, contudo, neste caso os resultados aparecem divididos por género (Figura 37).

→ **Frequencies**

Statistics

Região de Portugal

Região de Portugal	N	Valid	Missing
0		1	0
Feminino		13	0
Masculino		18	0

Região de Portugal

Género	Valid	Região de Portugal	Frequency	Percent	Valid Percent	Cumulative Percent	
0	Valid	Norte	1	100,0	100,0	100,0	
		Feminino	Valid	Norte	1	7,7	7,7
		Centro-norte		7	53,8	53,8	
		Centro-sul		4	30,8	30,8	
		Sul		1	7,7	7,7	
Total	13	100,0	100,0				
Masculino	Valid	Norte	6	33,3	33,3	33,3	
		Centro-norte	7	38,9	38,9	72,2	
		Centro-sul	1	5,6	5,6	77,8	
		Sul	4	22,2	22,2	100,0	
		Total	18	100,0	100,0		

Figura 37

Janela de resultados com a análise de estatística descritiva das variáveis categóricas, divididos por género feminino e masculino

5. TESTES PARAMÉTRICOS

Depois de completar este capítulo, o aluno deverá ser capaz de:

Conhecer alguns dos testes paramétricos mais usados para comparar médias e realizar correlações.

Aprender o procedimento de computação de alguns dos testes paramétricos usando SPSS.

Interpretar os testes paramétricos estudados no visor de resultados.

5.1 TESTES PARAMÉTRICOS VS. TESTES NÃO PARAMÉTRICOS

Considerando as condições de aplicabilidade dos testes paramétricos analisadas no Capítulo 3, há a necessidade de compreender alguns dos testes paramétricos e da sua alternativa não paramétrica. A Figura 38 permite analisar alguns dos testes paramétricos aplicados em motricidade humana e a sua alternativa de testes não paramétricos.

TESTES PARAMÉTRICOS		TESTES NÃO PARAMÉTRICOS
T–teste para amostras independentes	—————→	Mann – Whitney teste
T–teste para amostras emparelhadas	—————→	Teste de Wilcoxon
ANOVA – one way	—————→	Kruskal – Wallis teste
Teste de Pearson	—————→	Teste de Spearman

Figura 38

Testes paramétricos vs. testes não paramétricos

Os testes paramétricos serão estudados neste capítulo 5 e, a seguir, no capítulo 6 serão estudados os testes não paramétricos.

5.2 T-TESTE PARA AMOSTRAS INDEPENDENTES

O T-teste para amostras independentes utiliza-se para estudar a diferença entre duas médias de amostras independentes (ou grupos diferentes). Por exemplo, quando queremos comparar as médias entre o grupo experimental e o grupo de controlo das variáveis “altura” e “anos de diagnóstico” antes de uma intervenção de exercício físico (pré-intervenção ou *baseline*), devemos selecionar a opção: **Analyze → Compare Means → Independent – Samples T Teste...** (Figura 39).

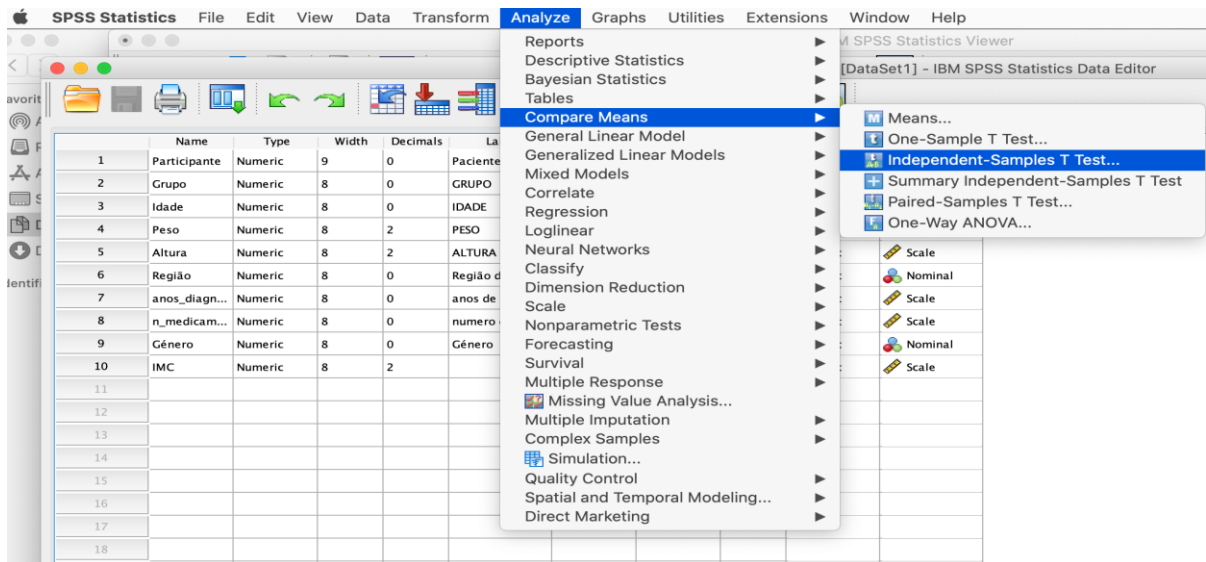


Figura 39

Indicação sobre as opções a selecionar para realizar o T-Teste para Amostras Independentes

Em seguida, colocamos as variáveis dependentes na caixa de “*Test Variable(s):*” e a variável de agrupação na caixa “*Grouping Variable:*”. Verificamos que a nossa variável “Grupo” aparece com dois pontos de interrogação “Grupo (? ?)”. Para definir os grupos a comparar e considerando que o T-teste para amostras independentes só compara dois grupos ao mesmo tempo, devemos clicar em “*Define Groups...*” (Figura 40).

No exemplo mostrado na Figura 40 após colocar os grupos a comparar: 1 grupo experimental e 2 grupo de controlo, devemos colocar na caixa de “*Define Groups*”, os números dos grupos que o teste irá comparar (colocando os mesmos números que aparecem na base de dados), posteriormente, clicamos em “*Continue*” (Figura 41).

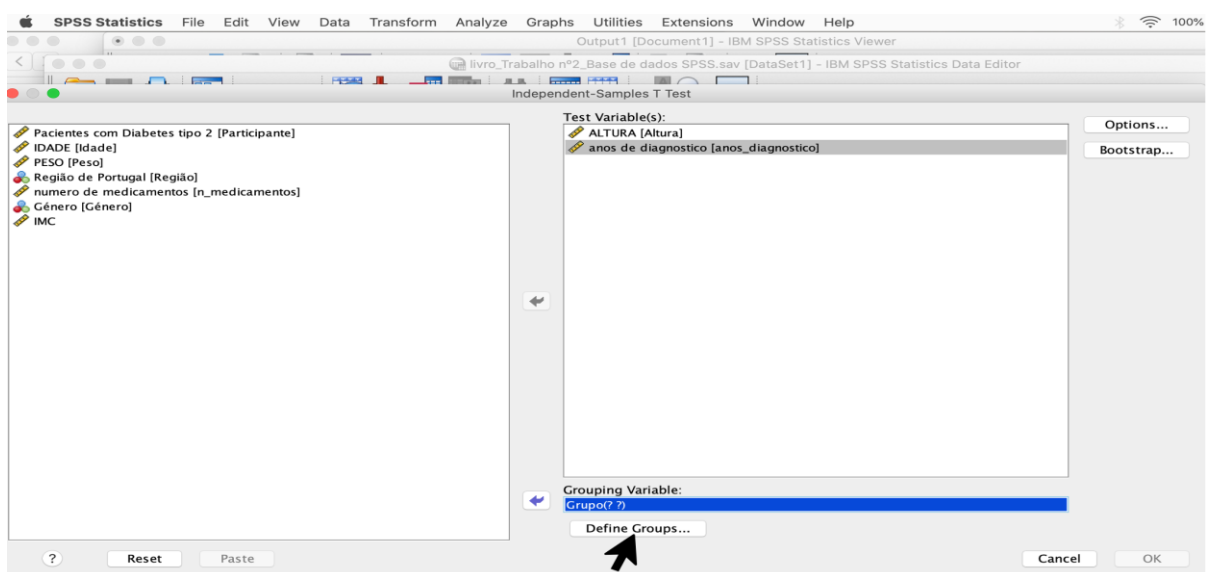


Figura 40

Indicação sobre onde definir os grupos que queremos comparar

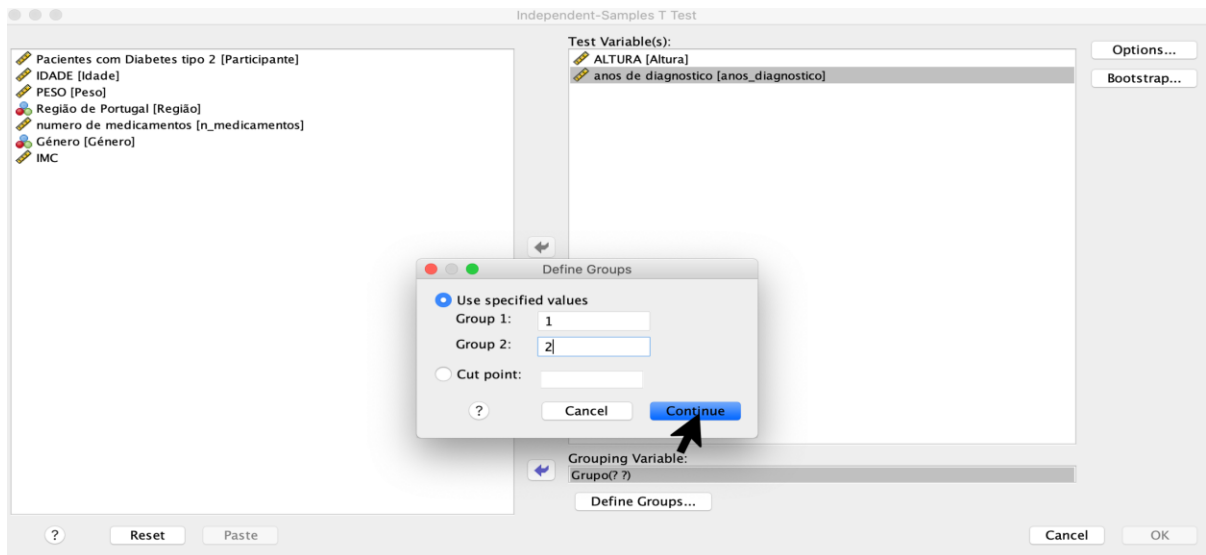


Figura 41
Indicação sobre o passo a seguir à definição dos grupos

Na caixa de “*Grouping Variable*” aparecem agora os grupos que o teste irá comparar “Grupo (1 2)”. Seguidamente, clicamos OK (Figura 42).

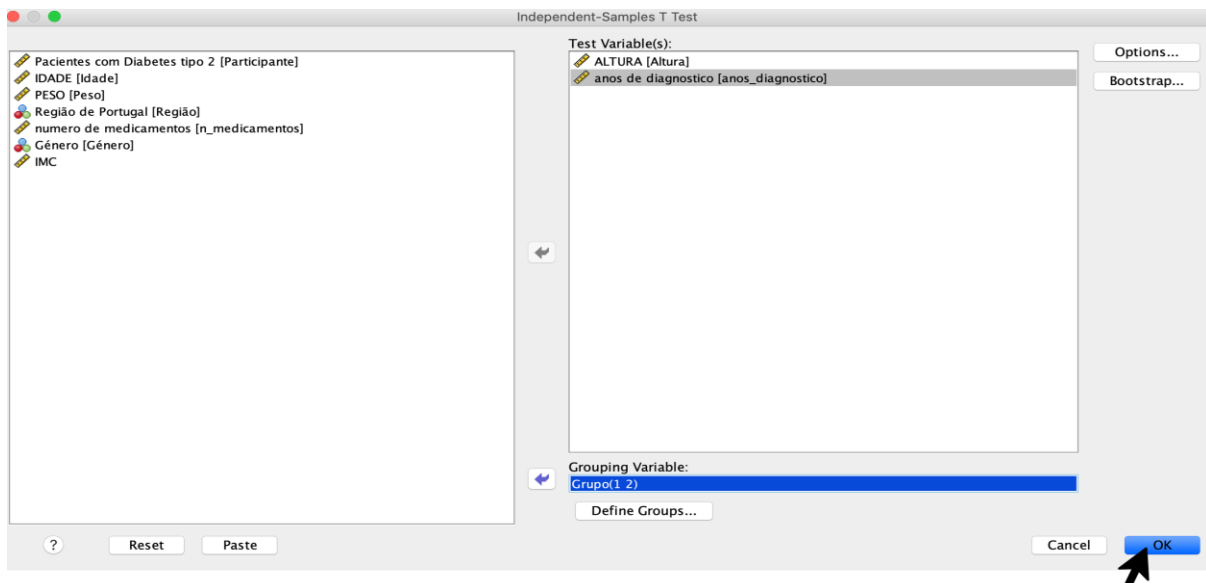


Figura 42
Exemplo da janela “*Independent-Samples T-Test*” preenchida

O SPSS mostrará na janela de resultados (Figura 43) uma primeira tabela de estatística descritiva e uma segunda tabela com o T-teste para amostras independentes.

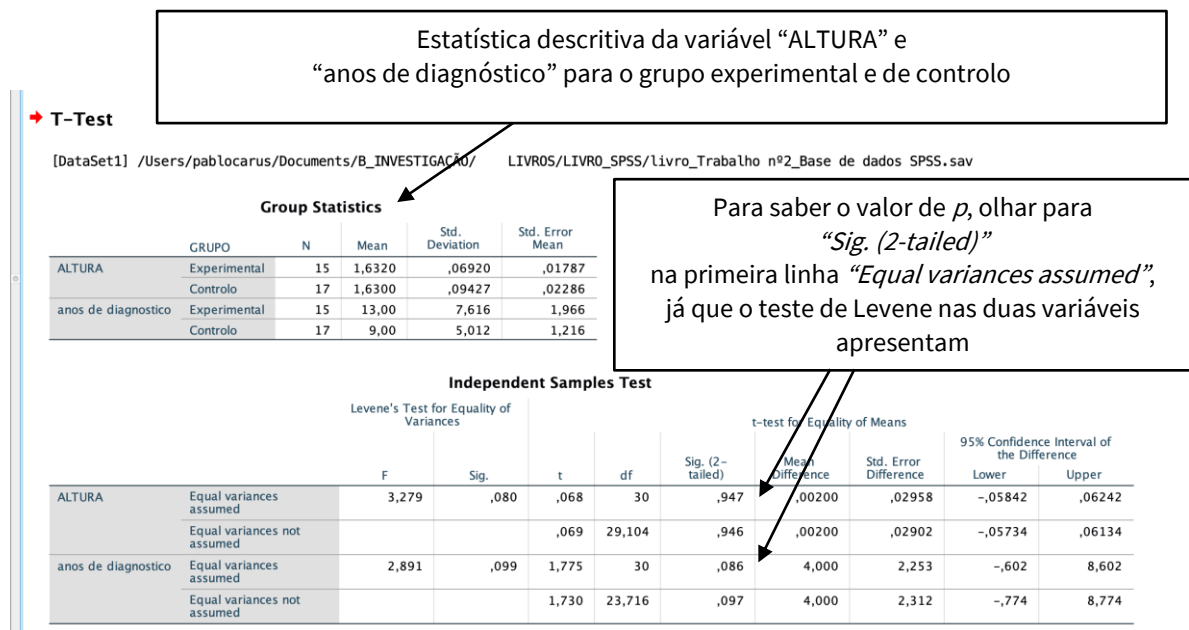


Figura 43

Indicação sobre onde verificar os resultados do T-Teste para Amostras Independentes

Mediante as tabelas, é possível observar que para a variável “ALTURA” o valor de $p = 0.947$ (maior de 0,050) o que indica que não existem diferenças significativas entre a média do grupo experimental $1,63 \pm 0,06$ e a média do grupo de controlo $1,63 \pm 0,09$. Para a variável “anos de diagnóstico” o valor de $p = 0,086$ (também maior de 0,050) indica igualmente que não existem diferenças significativas entre a média dos dois grupos ($13,00 \pm 7,61$ vs. $9,00 \pm 5,01$).

5.3 T-TESTE PARA AMOSTRAS EMPARELHADAS

O T-teste para amostras emparelhadas é o teste a utilizar quando se pretende comparar duas populações de onde foram extraídas duas amostras emparelhadas relativamente a uma variável dependente quantitativa (Maroco, 2018). Por exemplo, 32 mulheres com obesidade participaram num programa de exercício de 16 semanas e, que se avaliou o peso antes e após o programa de exercício, para realizar esta análise estatística aplica-se o T-teste para amostras emparelhadas, que servirá para averiguar se o programa de exercício foi eficaz. Para realizar este teste selecionamos a opção: **Analyze → Compare Means → Paired – Samples T Teste...** (Figura 44).

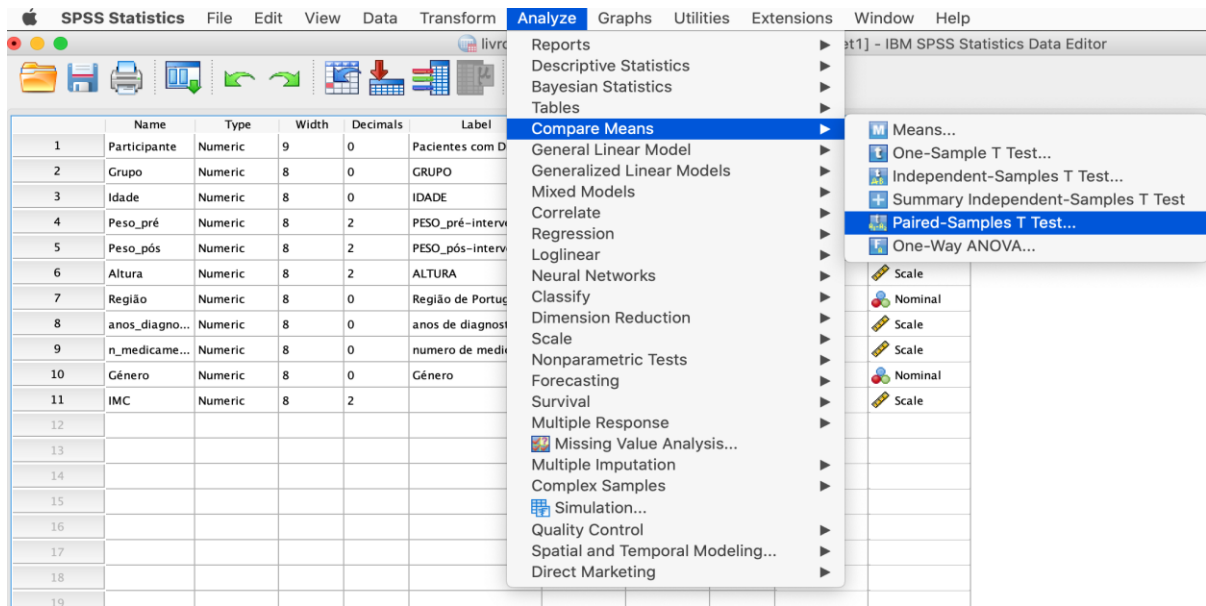


Figura 44

Indicação sobre as opções a selecionar para realizar o T-Teste para Amostras Emparelhadas

Em seguida, colocamos a variável “PESO_pré-intervenção” no “*Pair – Variable1*” e a variável “PESO_pós-intervenção” no “*Pair – Variable2*” na caixa de “*Paired Variables*” e clicamos OK (Figura 45).

O SPSS mostrará na janela de resultados (Figura 46), três tabelas em que a análise deve forçar-se na primeira tabela, que contém a estatística descritiva e na terceira tabela que indicará o valor de p do T-teste para amostras emparelhadas. Neste exemplo, podemos observar que existiram diferenças estatisticamente significativas, uma vez que o valor de p é igual a 0,001 ($p < 0,050$), e que é notória uma diferença entre a média do peso avaliada antes do programa de exercício (pré-intervenção: $83,1 \pm 13,9$), e a média do peso avaliada após 16 semanas de exercício físico (pós-intervenção: $81,4 \pm 12,8$). Logo, podemos concluir que o programa de exercício permitiu melhorar significativamente o peso das mulheres obesas.

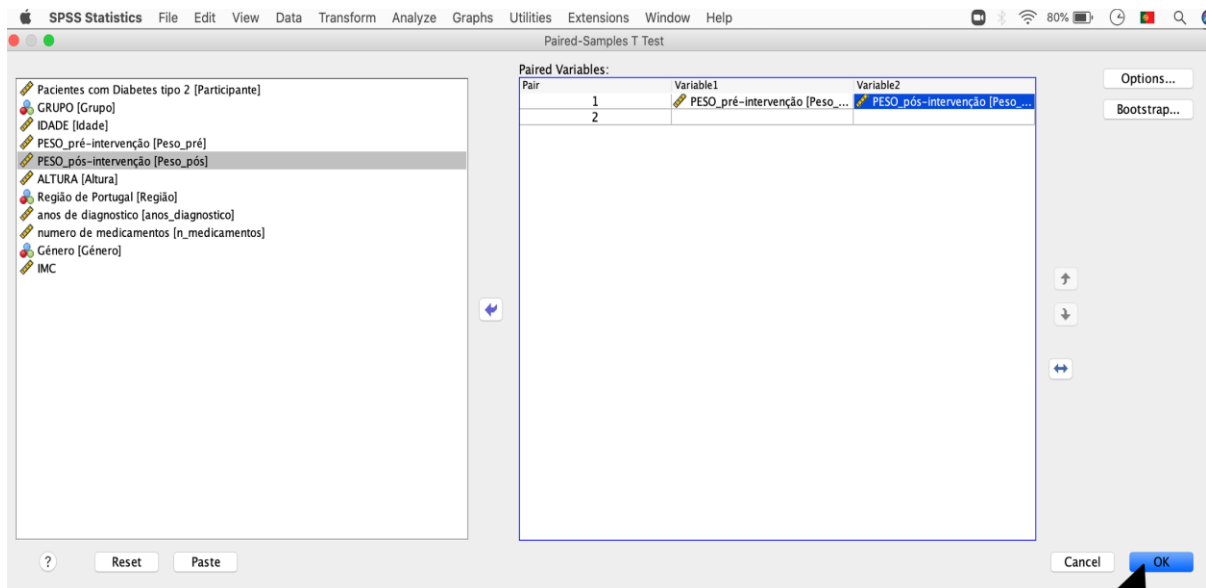


Figura 45
Exemplo da janela "Paired-samples T-Test" preenchida

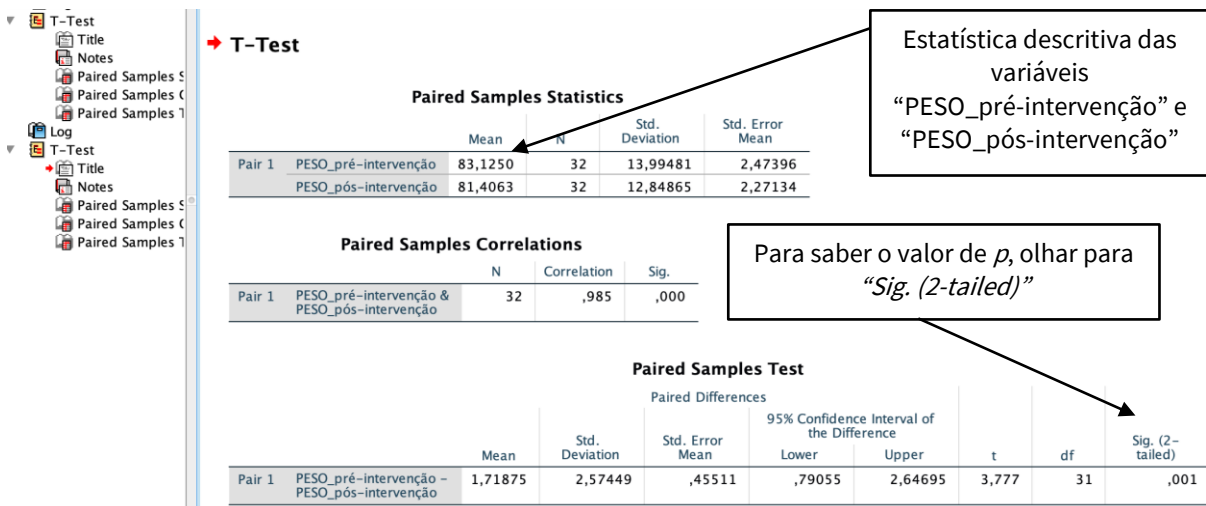


Figura 46
Indicação sobre onde verificar os resultados do T-Teste para Amostras Emparelhadas

5.4 ANOVA – ONE WAY

Se o utilizador quer comparar três médias de três grupos diferentes não é recomendável utilizar três T-testes para amostras independentes, porque a utilização de múltiplos T-testes inflaciona também a taxa de erro Tipo I. Assim, se uma hipótese for testada ao nível da significância 0,05; o erro real seria muito superior a este, e a conclusão não será fiável. Para solucionar este problema podemos utilizar a ANOVA – one way, um teste que permite verificar se existem diferenças entre as médias de mais de dois grupos (não emparelhados) (Verma, 2016). Quando a ANOVA – one way mostra uma diferença significativa ($p < 0.050$) entre as médias de vários grupos significa que há diferença entre pelo menos duas das

médias, mas não indica entre quais existem diferenças. Então é necessário realizar uma análise adicional, designada de análise de contrastes, que mostra a diferença entre todos os pares de médias, no contexto da amostra total (Camacho, 2002).

Por exemplo, para verificar se há diferenças para a “ALTURA” entre as três regiões de Portugal da variável “Região”, temos que selecionar a opção: **Analyze → Compare Means → One-Way ANOVA...** (Figura 47).

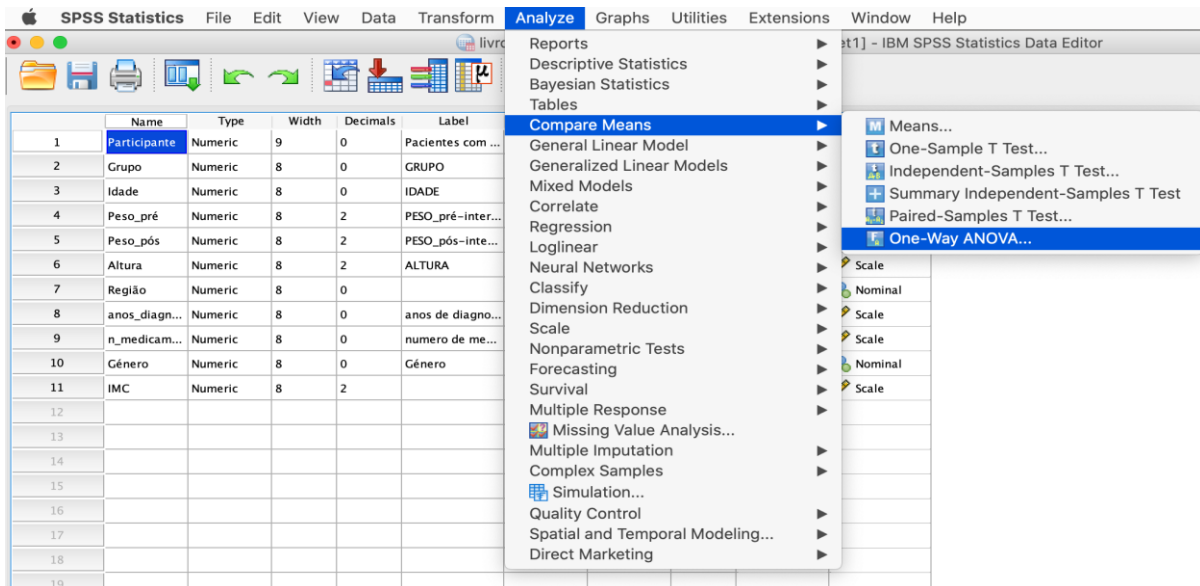


Figura 47
Indicação sobre as opções a selecionar para realizar a ANOVA - One Way

Seguidamente, selecionamos a variável “ALTURA” e passamo-la para a caixa “*Dependent List.*”, colocamos a variável “Região” na caixa “*Factor*” e clicamos OK (Figura 48).

Isto seria suficiente para realizar uma análise de variância, mas como serão realizados contrastes de médias dois a dois de *Tukey*, antes é necessário selecionarmos a opção “*Post Hoc...*”, na parte superior direita (Figura 49).

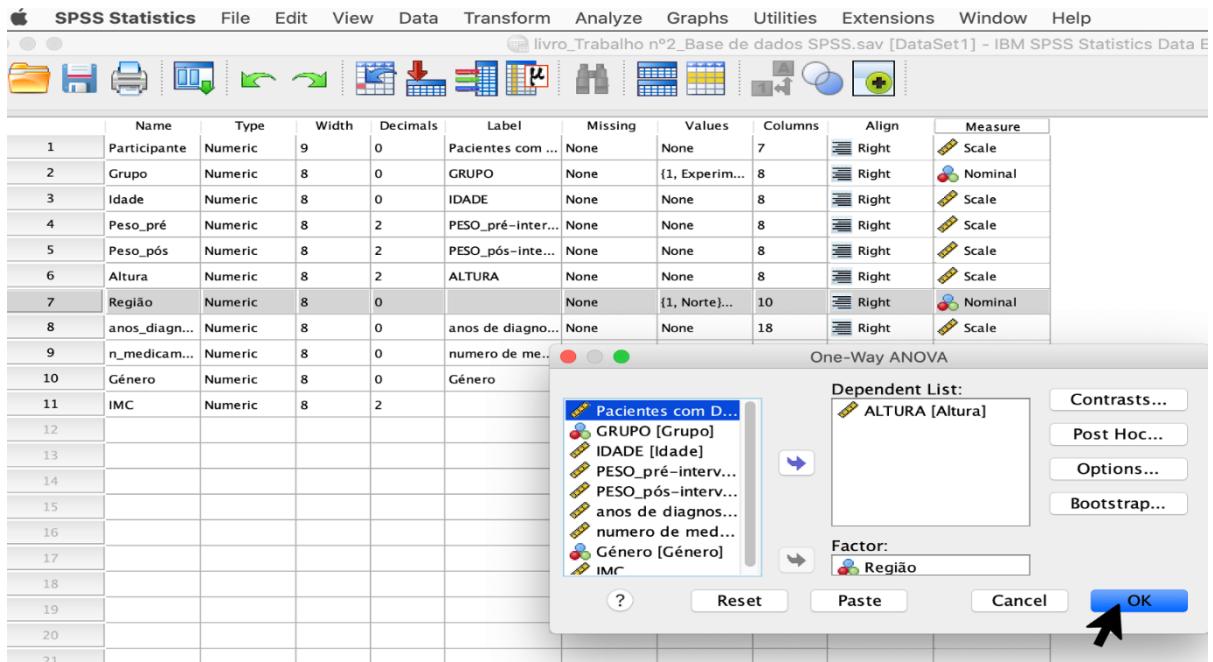


Figura 48
Janela "One-Way ANOVA" preenchida

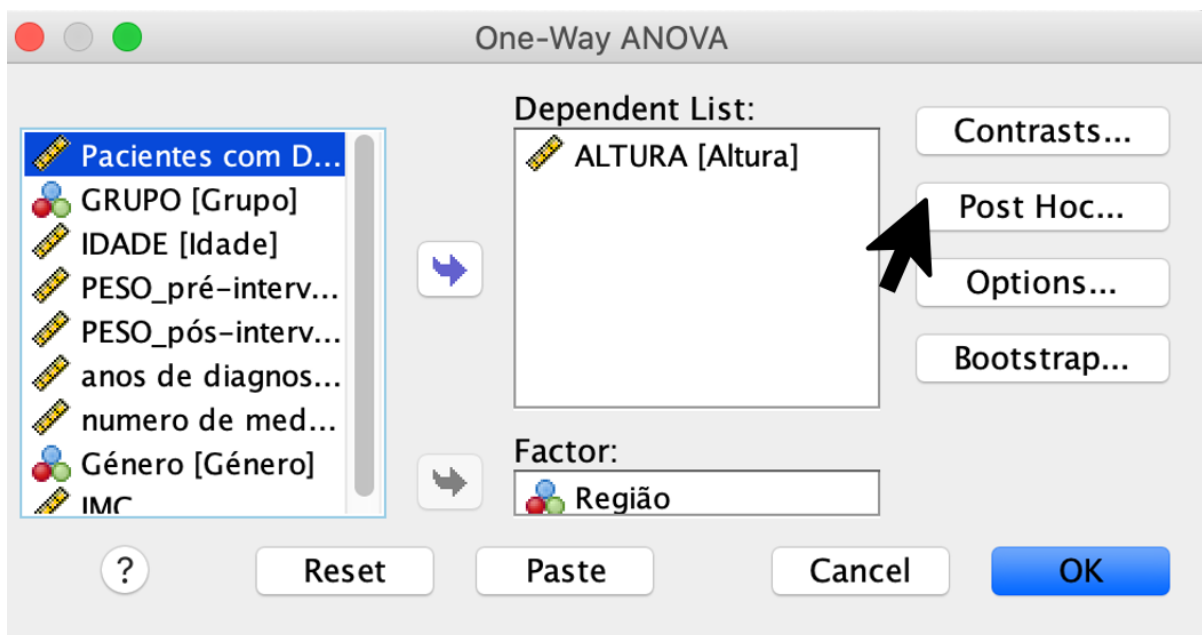


Figura 49
Janela "One-Way ANOVA", na opção "Post Hoc"

Em seguida selecionamos a opção “*Tukey*” em “*Equal Variances Assumed*” e clicamos em “*Continue*” (Figura 50).

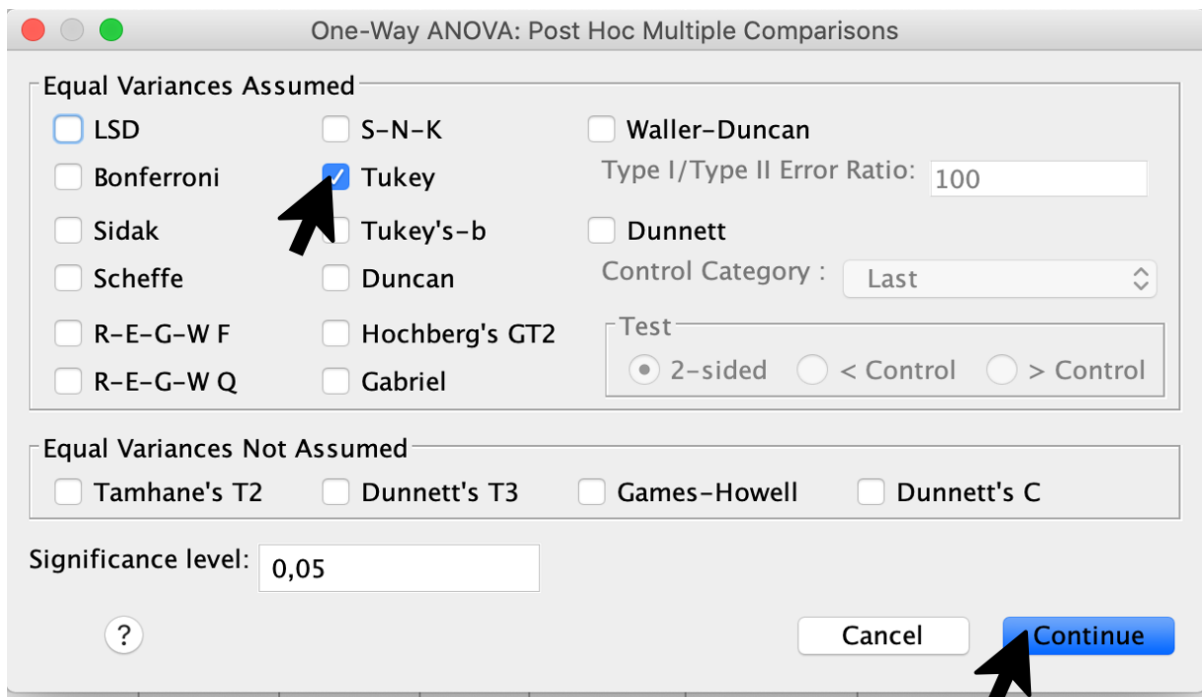


Figura 50
Indicação das opções a selecionar na janela "One-Way ANOVA: Post Hoc Multiple Comparisons"

Para que o visor de resultados mostre também a estatística descritiva, é necessário selecionarmos a opção “*Options...*”, na parte superior direita (Figura 51).

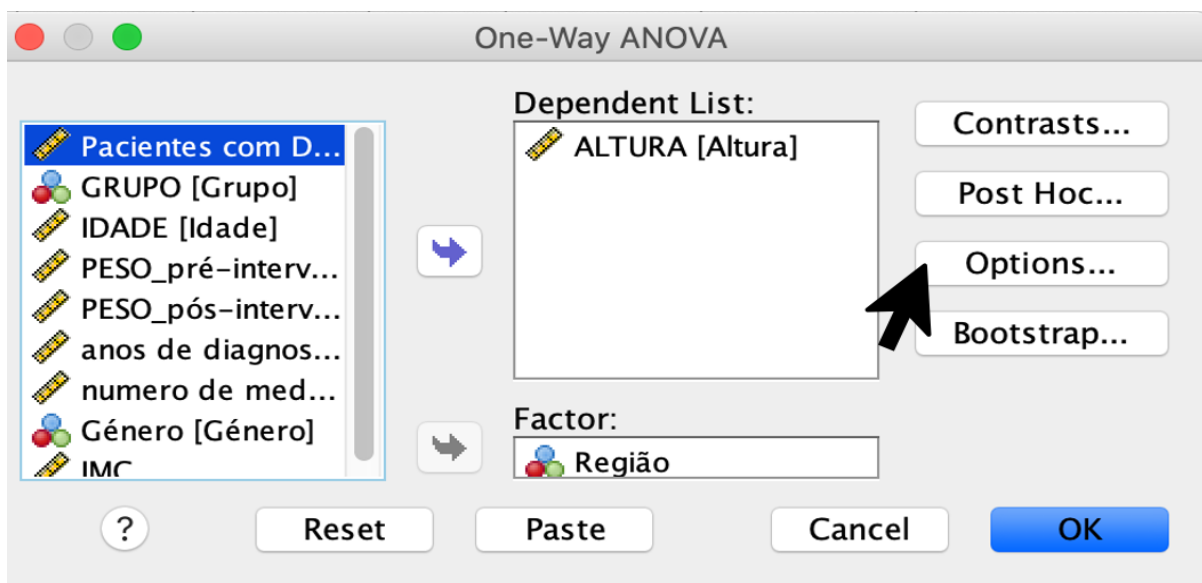


Figura 51
Indicação para selecionar "Options", na janela "One-Way ANOVA"

Posteriormente, selecionamos a opção “*Descriptive*” em “*Statistics*”, também deverão ser selecionadas as opções “*Homogeneity of variance test*” (Teste de Levene) e “*Welch*”, para finalizar, clicamos em “*Continue*” (Figura 52).

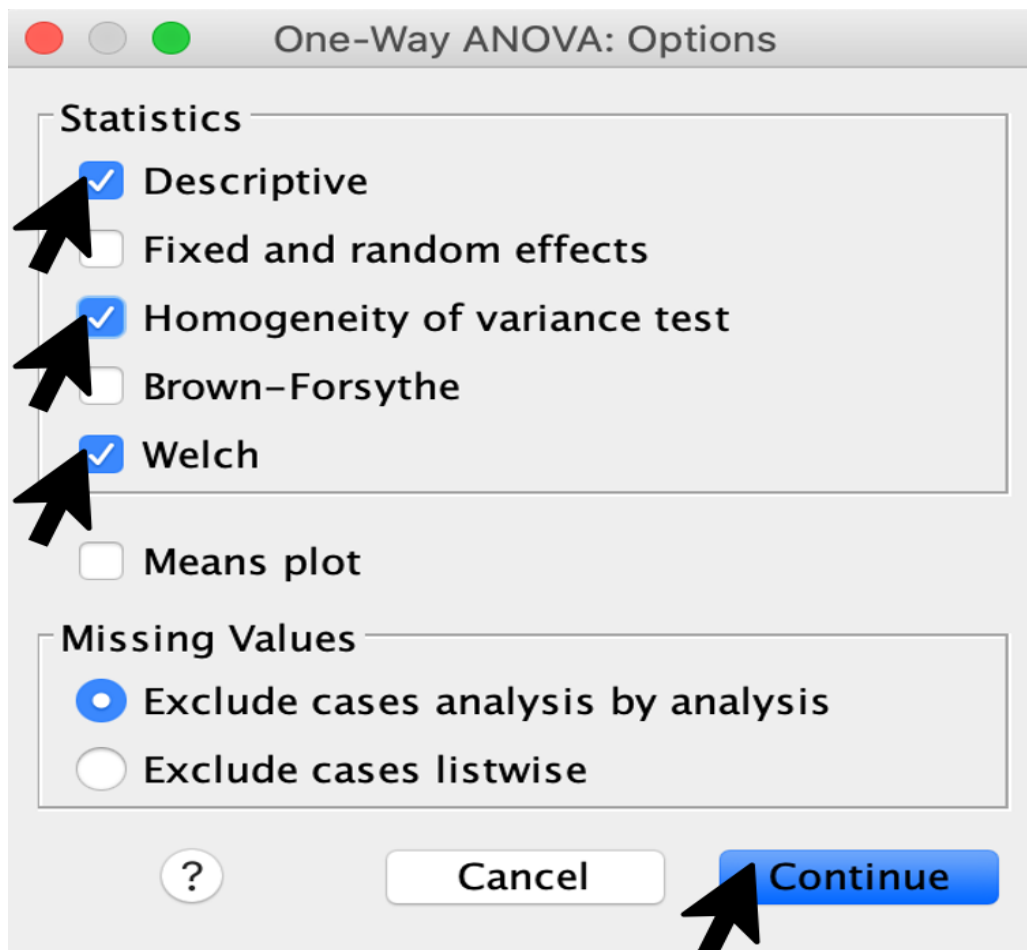


Figura 52
Indicação das opções a selecionar na janela "One-Way ANOVA: Options"

E, em seguida, clicamos em OK, para que o SPSS mostre o visor de resultados com todas as análises solicitadas (Figura 53).

A janela de resultados mostrará uma primeira tabela de estatísticas descritivas, uma segunda tabela com a homogeneidade das variâncias (teste de Levene), uma terceira tabela com as análises de variância (ANOVA) e uma quarta tabela com o teste robusto de análises de variância de Welch (que será considerado quando o teste de Levene é significativo: $p < 0,050$). Podemos observar que a média da altura (em metros) para a região “Norte” é $1,68 \pm 0,06$; para a região “Centro” é $1,63 \pm 0,09$; e para a região “Sul” é $1,58 \pm 0,06$. A tabela de homogeneidade das variâncias mostra que, o teste de Levene não é significativo ($p = 0,106$, indicando que as variâncias dos grupos são iguais), assim devemos olhar para o teste de ANOVA, para saber se existem diferenças entre grupos. O valor de $p = 0,039$; pelo que é possível afirmar que existem diferenças significativas entre as médias dos três grupos na altura (analisar Figura 54).

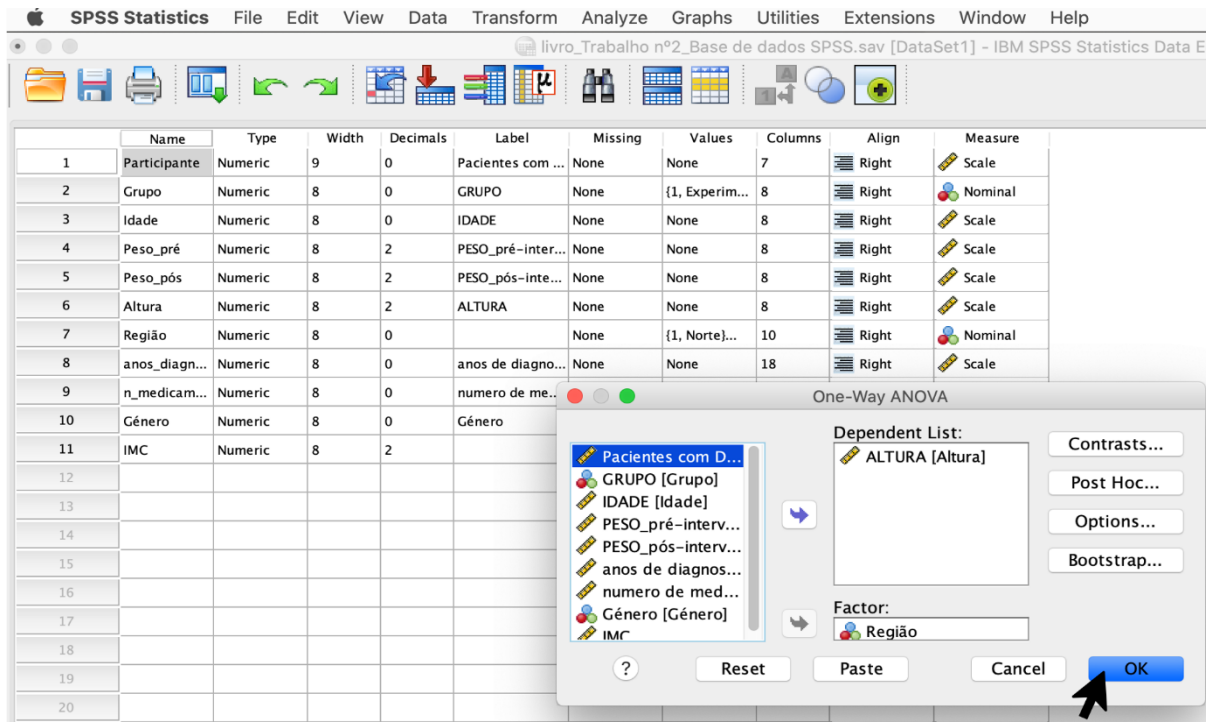


Figura 53
Janela "One-Way ANOVA" preenchida

Se pretendemos verificar se os grupos apresentam diferenças estatisticamente significativas devemos observar a tabela de "Multiple Comparisons", onde podemos observar que existem diferenças estatisticamente significativas (Figura 55), na altura entre a região do Norte e a região do Sul ($p = 0,031$), sendo que não há diferenças significativas entre a região Norte e a região Centro ($p = 0,373$) e, entre a região Centro e a região Sul ($p = 0,233$) também não existem diferenças estatisticamente significativas.

→ **Oneway**

Descriptives

Estatística descritiva da variável "ALTURA" para as regiões: "Norte", "Centro" e "Sul"

ALTURA	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Norte	8	1,6800	,06047	,02138	1,6294	1,7306	1,56	1,74
Centro	15	1,6347	,08975	,02317	1,5850	1,6844	1,49	1,76
Sul	9	1,5811	,06030	,02010	1,5348	1,6275	1,48	1,69
Total	32	1,6309	,08216	,01452	1,6013	1,6606	1,48	1,76

Test of Homogeneity of Variances

Valor de p do teste de Levene

ALTURA		Levene Statistic	df1	df2	Sig.
Based on Mean		2,428	2	29	,106
Based on Median		1,603	2	29	,219
Based on Median and with adjusted df		1,603	2	26,987	,220
Based on trimmed mean		2,375	2	29	,111

ANOVA

Valor de p do teste de ANOVA

ALTURA		Sum of Squares	df	Mean Square	F	Sig.
Between Groups		,042	2	,021	3,620	,039
Within Groups		,167	29	,006		
Total		,209	31			

Robust Tests of Equality of Means

Valor de p do teste de Welch (só será considerado para comparar as médias dos grupos quando o teste de Levene é significativo ($p < 0,050$))

ALTURA		Statistic ^a	df1	df2	Sig.
Welch		5,508	2	18,048	,014

a. Asymptotically F distributed.

Figura 54
Indicação dos resultados obtidos após a aplicação da ANOVA - One Way

Post Hoc Tests

Multiple Comparisons

Dependent Variable: ALTURA

Tukey HSD

(I) Região	(J) Região	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Norte	Centro	,04533	,03327	,373	-,0368	,1275
	Sul	,09889*	,03692	,031	,0077	,1901
Centro	Norte	-,04533	,03327	,373	-,1275	,0368
	Sul	,05356	,03204	,233	-,0256	,1327
Sul	Norte	-,09889*	,03692	,031	-,1901	-,0077
	Centro	-,05356	,03204	,233	-,1327	,0256

*. The mean difference is significant at the 0.05 level.

Figura 55

Janela de resultados referente ao "Post Hoc Testes: Multiple Comparisons"

5.5 TESTE DE PEARSON

O coeficiente de correlação r de Pearson é usado para medir a magnitude e o sinal da relação linear entre duas variáveis quantitativas. O coeficiente r varia entre -1 e 1. Em valor absoluto, quanto mais perto de 0 menor a relação entre variáveis, e quanto mais perto de 1 maior é a relação entre as duas variáveis. Não existe nenhuma relação linear entre as variáveis quando $r=0$, e a relação perfeita será quando $r=1$ ou $r=-1$. O sinal positivo de r indica que quanto maior for o valor de uma variável, também o valor da outra variável será maior, e se o sinal for negativo, então isso indica que quando uma variável tiver um valor maior, a outra irá ter um valor menor (Camacho, 2002). Por exemplo, uma relação de sinal positivo pode dar-se entre a altura e o peso, em geral, quanto mais altura maior é o peso da pessoa. Uma relação de sinal negativo pode dar-se entre o número de horas de atividade sedentária e a esperança de vida, em geral, quantas mais horas de atividade sedentária, menor será a esperança de vida.

Imaginemos que queremos saber as relações bivariadas entre as variáveis "tender points_/18", "duração dos sintomas anos", "SF-36 Função Física_pre" e "SF-36 Dor_pre", então devemos selecionar a opção: **Analyze** → **Correlate** → **Bivariate** (Figura 56).

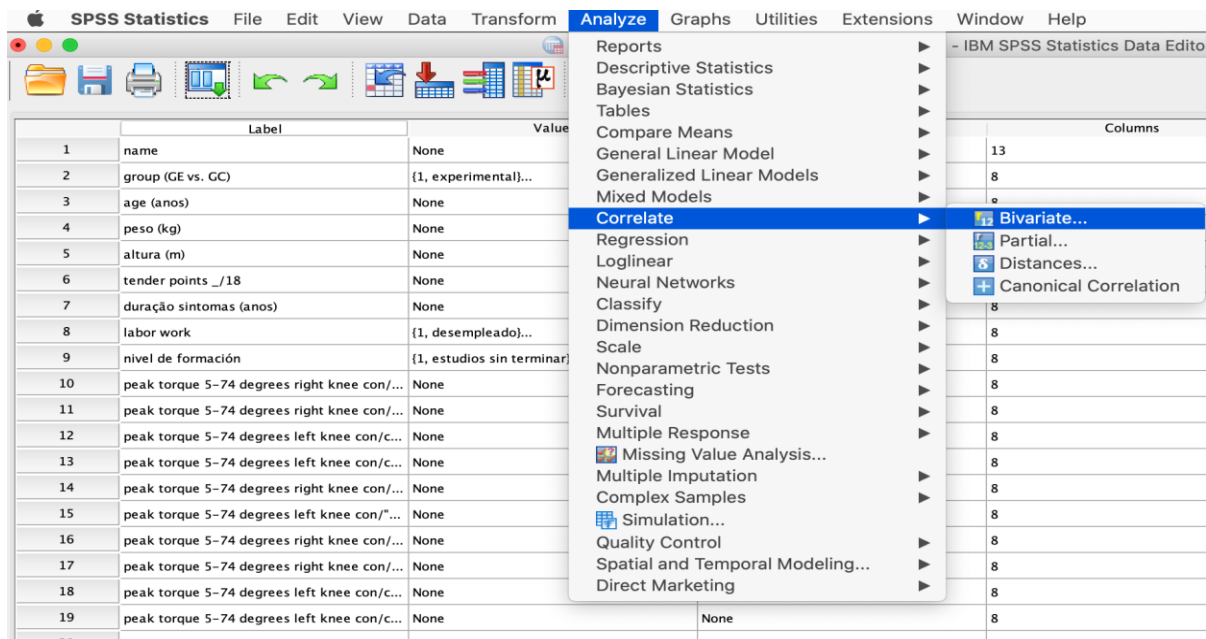


Figura 56

Indicação das opções a selecionar para realizar a Correlação de *Pearson*

Em seguida selecionamos as variáveis desejadas e transferimo-las para a caixa de “*Variables*”, clicando, posteriormente, em OK, sabendo que se encontra pré-selecionada a opção de correlação de *Pearson* (Figura 57).

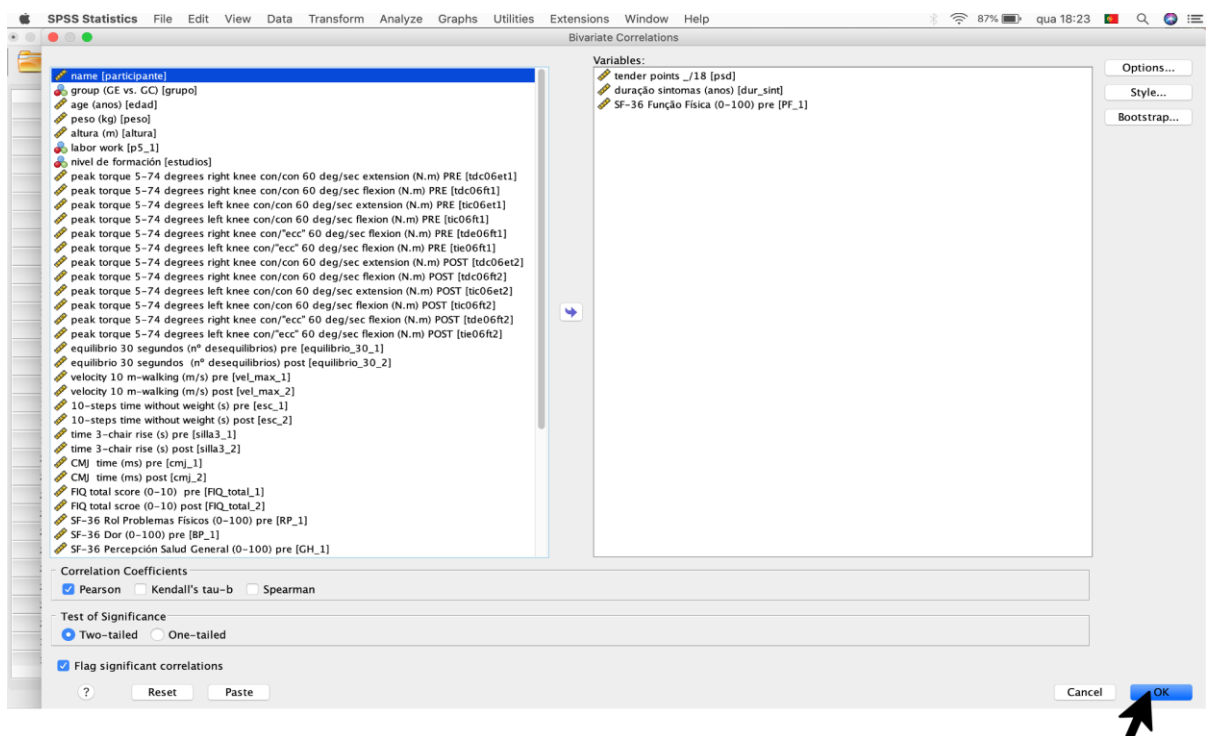


Figura 57

Janela da Correlação de *Pearson* preenchida

Na janela de resultados (Figura 58), aparece uma tabela com os coeficientes r de Pearson e os valores de p (para determinar se existe ou não uma correlação significativa). A tabela mostra as correlações significativas com um ou dois asteriscos (* quando o valor de $p < 0.050$; ** quando o valor de $p < 0.010$) ao lado do coeficiente r , mas se queremos saber o valor de p , devemos observar na tabela a coluna “Sig. (2-tailed)”. Por exemplo, o único coeficiente de correlação significativo, que podemos observar na tabela é entre “duração sintomas (anos)” e “SF-36 Função Física (0-100) pre”, pois verifica-se que o valor de $p = 0,030$ (menor que 0,050) e que o $r = -0,384$ é inverso (r com sinal negativo), isto é, quando os anos de duração de sintomas aumentam a função física diminui. Não é possível observar mais coeficientes de correlação significativos na tabela (o resto apresenta um $p > 0,050$).

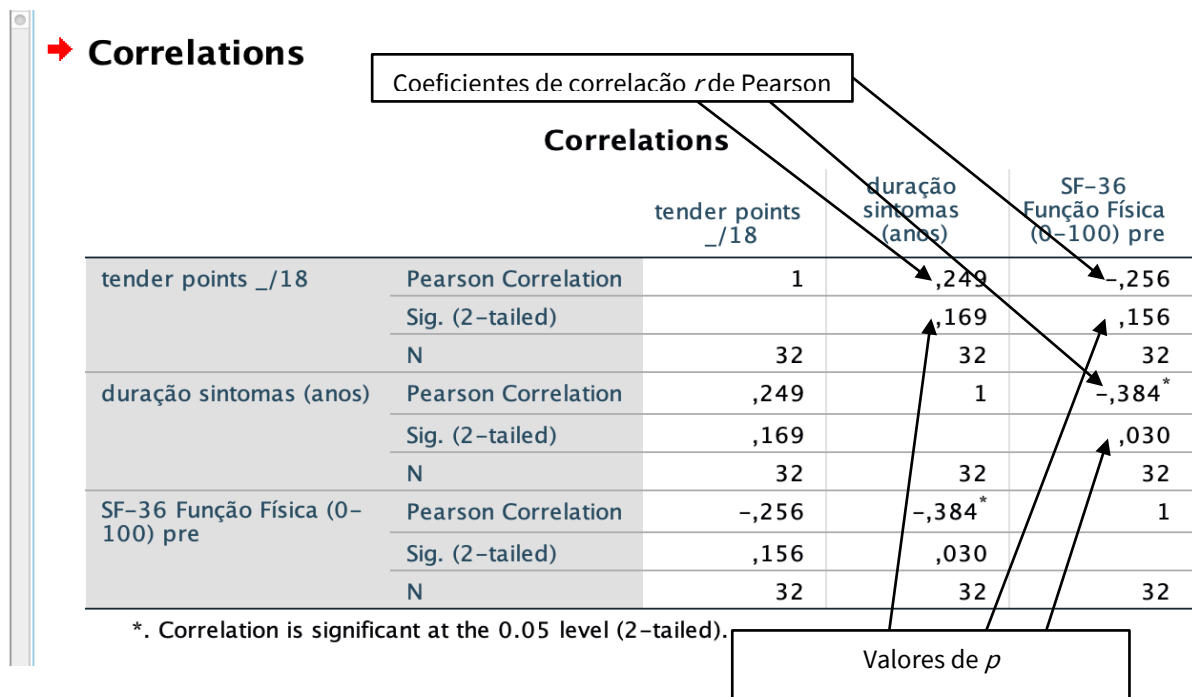


Figura 58
Indicação dos resultados obtidos após a aplicação da Correlação de *Pearson*

6. TESTES NÃO PARAMÉTRICOS

Depois de completar este capítulo, o aluno deverá ser capaz de:

Conhecer alguns dos testes não paramétricos mais usados para comparar médias, frequências e correlações.

Aprender o procedimento de computação de alguns dos testes não paramétricos usando SPSS.

Interpretar os testes não paramétricos estudados no visor de resultados.

Os testes não paramétricos utilizam-se quando as variáveis não cumprem os pressupostos de distribuição normal dos dados e/ou homogeneidade das variâncias, o que, geralmente, acontece com amostras pequenas e com variáveis categóricas. Também se usam testes não paramétricos quando algum dos grupos tem variância zero, ou seja, quando são comparados dois grupos de sujeitos independentes, e num dos grupos todos os sujeitos obtêm a mesma pontuação (Maroco, 2018).

6.1 TESTE DE CHI-QUADRADO (χ^2) DE UMA AMOSTRA

O teste de Chi-quadrado de uma amostra é utilizado para comprovar se a distribuição de frequências, de uma variável nominal, se ajusta a uma distribuição modelo, por exemplo, em que todas as frequências sejam iguais. Neste caso, para sabermos se a variável “Região” segue uma distribuição uniforme, devemos selecionar a opção: **Analyze → Nonparametric Tests → Legacy Dialogs → Chi-square**. (Figura 59).

Em seguida, selecionamos a variável “Região” e transferimo-la para “*Test Variable List*.” e clicamos OK (Figura 60).

A janela de resultados mostra que, o valor de χ^2 é igual a 2,688 e que não é significativo, pois o valor de $p = 0,261$ (ou seja, superior a 0,050). Assim, pode ser interpretado que as três categorias da variável “Região” estão distribuídas de forma igual (igualdade de frequências em todas as categorias da variável) (Figura 61).

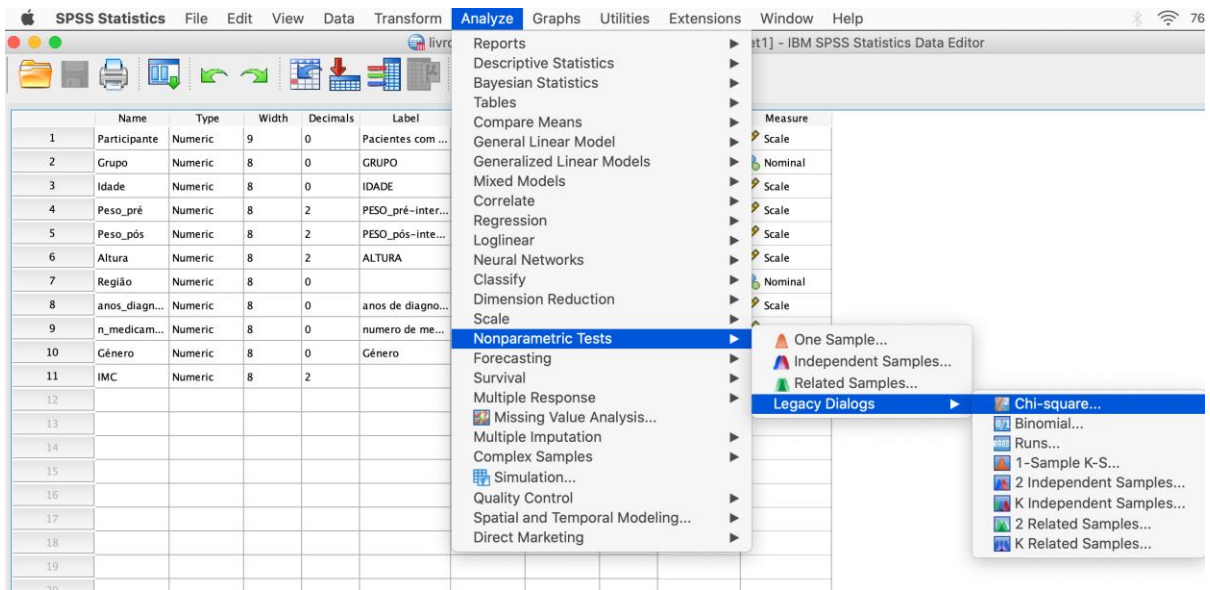


Figura 59
Indicação sobre as opções a selecionar para realizar o Teste de Chi-Quadrado

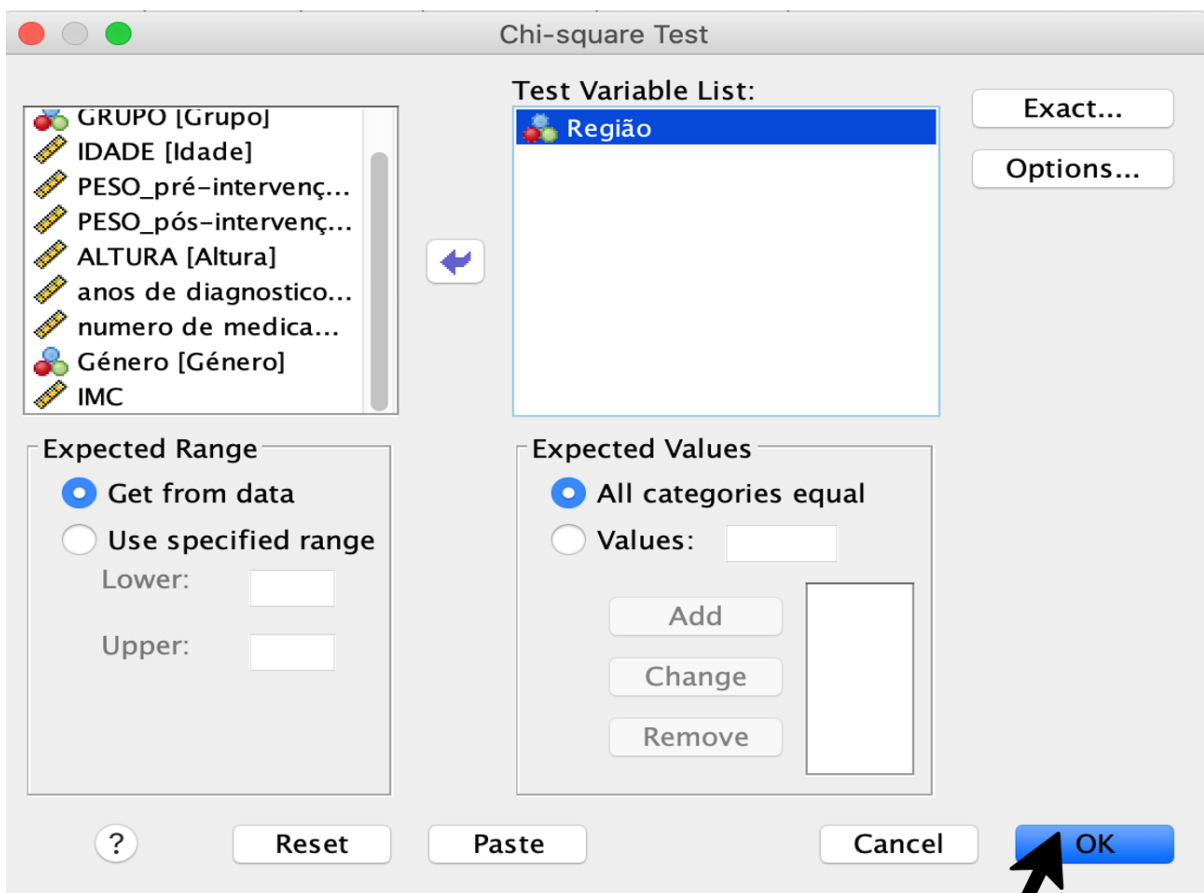


Figura 60
Janela "Chi-Square Test" preenchida

Chi-Square Test

Frequencies

Frequências de cada categoria da variável e total

	Região		
	Observed N	Expected N	Residual
Norte	8	10,7	-2,7
Centro	15	10,7	4,3
Sul	9	10,7	-1,7
Total	32		

Test Statistics

	Região
Chi-Square	2,688 ^a
df	2
Asymp. Sig.	,261

Valor de χ^2

a. 0 cells (0,0%) have expected frequencies less than 5. The minimum expected cell frequency is 10,7.

Valor de p

Figura 61

Indicação dos resultados obtidos após a aplicação do Teste de Chi-Quadrado

6.2 TESTE DE MANN - WHITNEY

O teste de Mann - Whitney é usado para comparar duas amostras não relacionadas. É o teste não paramétrico alternativo ao T-teste para amostras independentes, quando os dados da variável não cumprem com a normalidade e/ou homogeneidade das variâncias. Neste sentido, se queremos comparar as médias de dois grupos (grupo experimental e grupo de controlo) para a variável "IMC" num momento determinado, que pode ser antes de um programa de exercício físico (*baseline*). Para realizar esta análise, selecionamos a opção: **Analyze** → **Nonparametric Tests** → **Legacy Dialogs** → **2 Independent Samples...** (Figura 62).

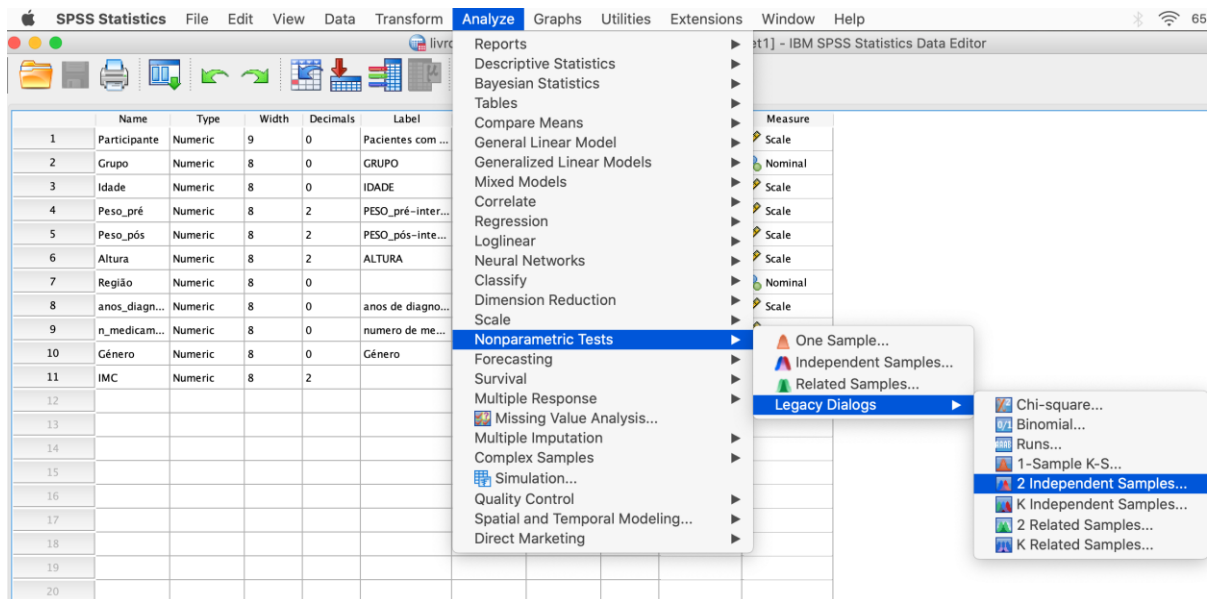


Figura 62
Indicação sobre as opções a selecionar para realizar o Teste de *Mann-Whitney*

Por definição, o SPSS indica que o tipo de teste que estamos a solicitar é a U de Mann – Whitney (na parte inferior esquerda) (Figura 63). Em seguida, devemos colocar a variável dependente, neste exemplo “IMC”, na caixa de “*Test Variable List:*” e a variável “Grupo” na caixa “*Grouping Variable:*”. Devemos definir os grupos que queremos comparar, já que o teste de Mann – Whitney (tal como, no caso do seu homólogo paramétrico, o T-teste para amostras independentes) só compara dois grupos ao mesmo tempo, pelo que devemos clicar em “*Define Groups...*” (Figura 63).

A seguir colocamos os grupos que iremos comparar. No presente exemplo, só temos dois grupos para comparar: 1 grupo experimental e 2 grupo de controlo, logo devemos colocar na caixa de “*Two Independent Samples: Define Groups*” os números dos grupos que o teste irá comparar (lembrar que devemos colocar os mesmos números que aparecem na base de dados) e clicamos em “*Continue*” (Figura 64).

Na caixa de “*Grouping Variable*” surge os grupos que o teste irá comparar “Grupo (1 2)”, depois é só clicarmos em OK (Figura 65).

O SPSS abre a janela de resultados com uma primeira tabela de *ranks* e, uma segunda tabela com as estatísticas de contraste (Figura 66).

A estatística de Mann – Whitney é igual a 109,5. O *rank* médio do grupo experimental é 17,70 e o *rank* médio do grupo de controlo é 15,44. Para comprovar se o teste de Mann – Whitney é estatisticamente significativo, analisamos a probabilidade exata (sempre que não existam pontuações iguais). Neste exemplo, o valor de *p* é superior a 0,050 ($p=0,502$) logo não é estatisticamente significativo, assim podemos afirmar que não existem diferenças significativas entre o grupo experimental e o grupo de controlo para a variável “IMC”.

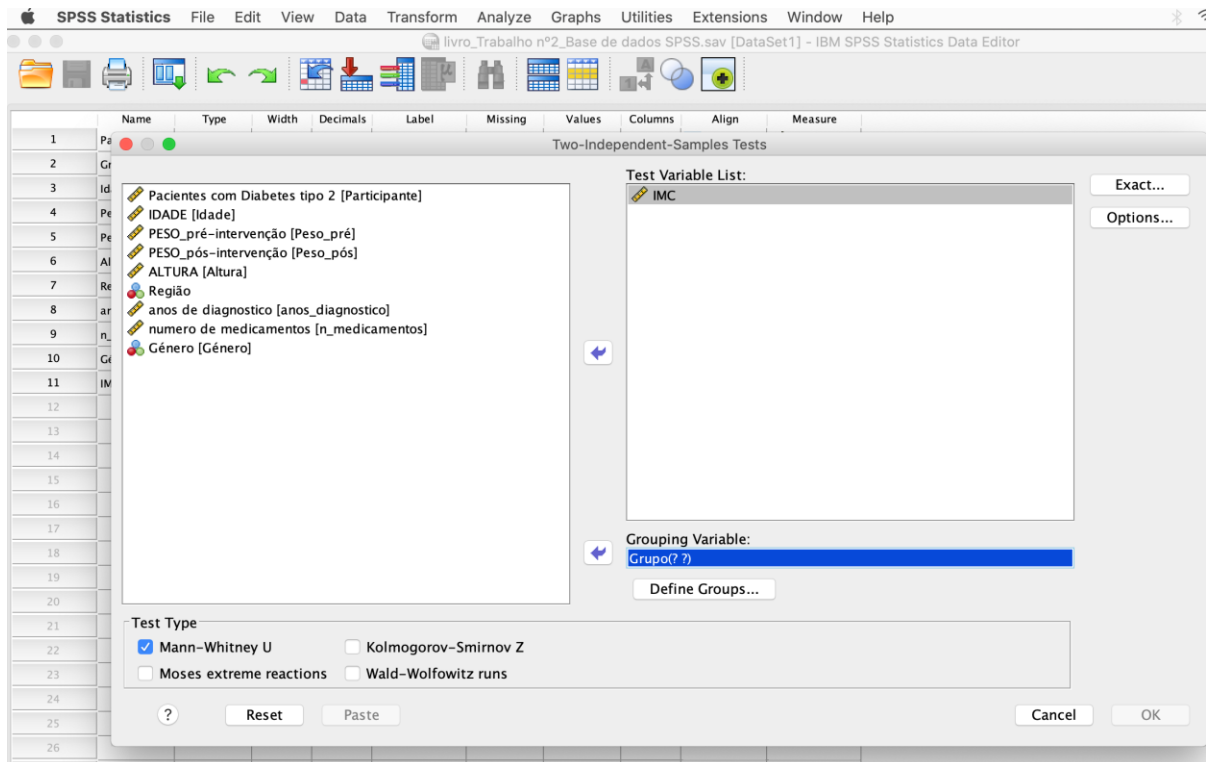


Figura 63
 Janela "Two Independent Samples-Tests", com indicação da seleção de "Mann-Whitney U" e a solicitação a definição dos grupos a comparar

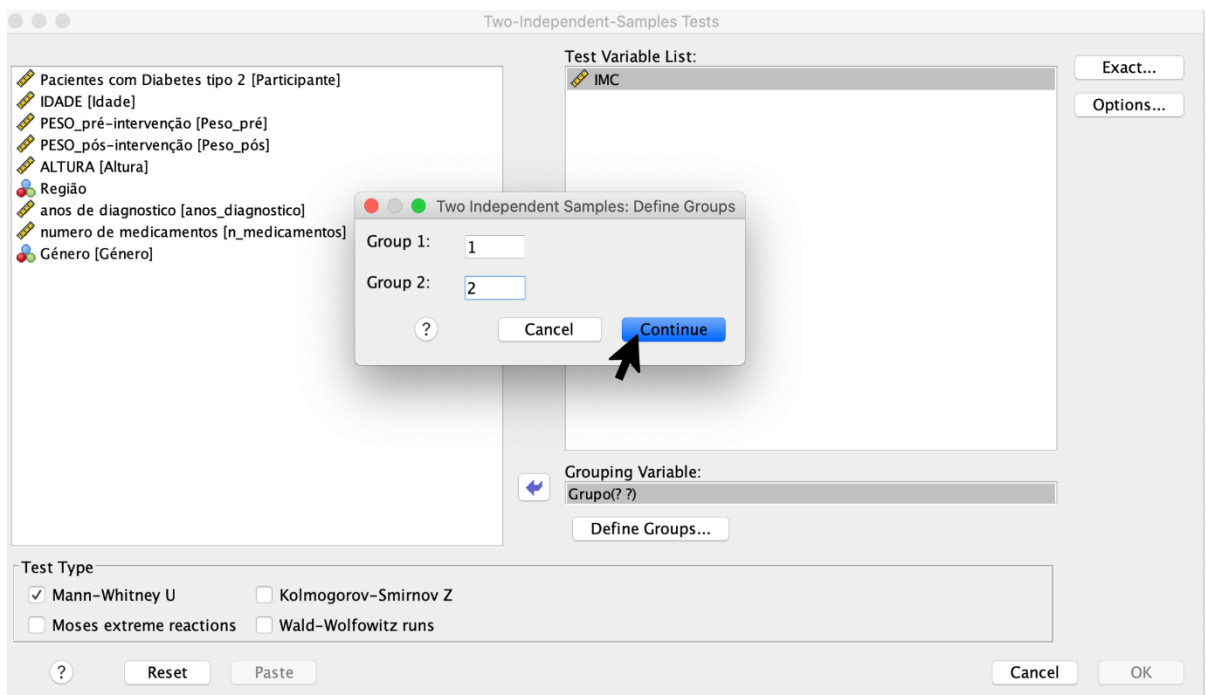


Figura 64
 Indicação sobre o passo a seguir à definição dos grupos

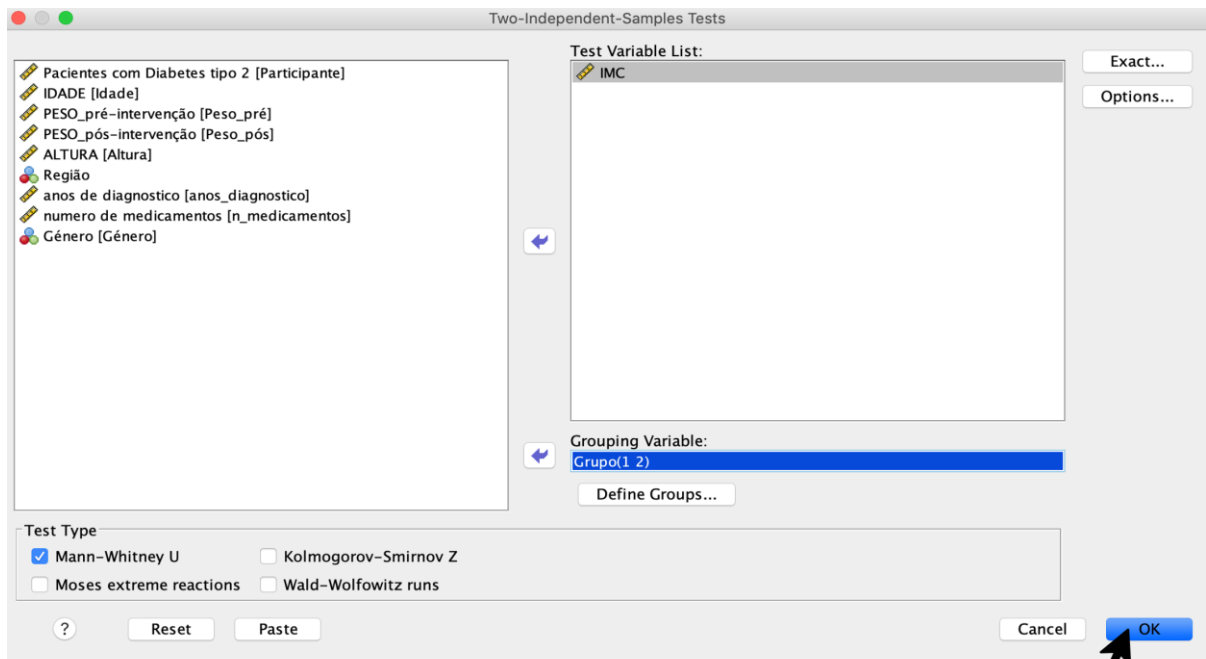


Figura 65
Janela "Two-Independent - Samples Tests" preenchida

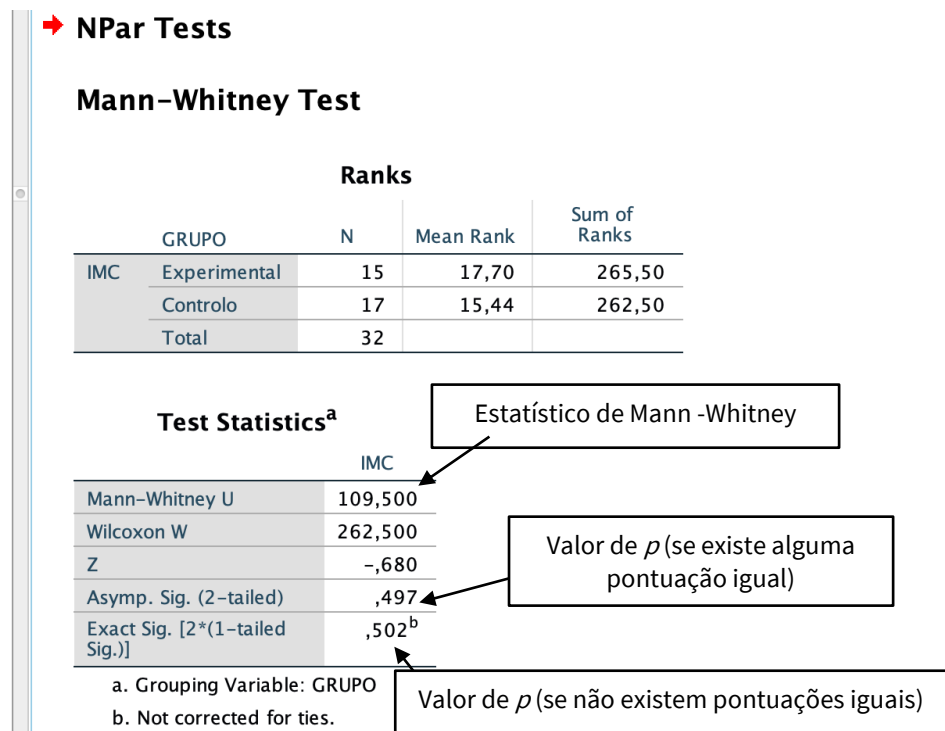


Figura 66
Indicação dos resultados obtidos com o Teste de *Mann-Whitney*

6.3 TESTE DE WILCOXON

O teste de Wilcoxon é usado para comparar duas amostras relacionadas. É o teste não paramétrico alternativo ao T-teste para amostras emparelhadas, quando as variáveis dependentes não cumprem a normalidade e/ou a homogeneidade das variâncias. Imagine que 32 sujeitos participam num programa de exercício de fortalecimento e, que queremos saber se existem diferenças estatisticamente significativas no “Counter Movement Jump (CMJ)” avaliado antes e após o programa de exercício, para isso temos que selecionar a opção: **Analyze** → **Nonparametric Tests** → **Legacy Dialogs** → **2 Related Samples...** (Figura 67).

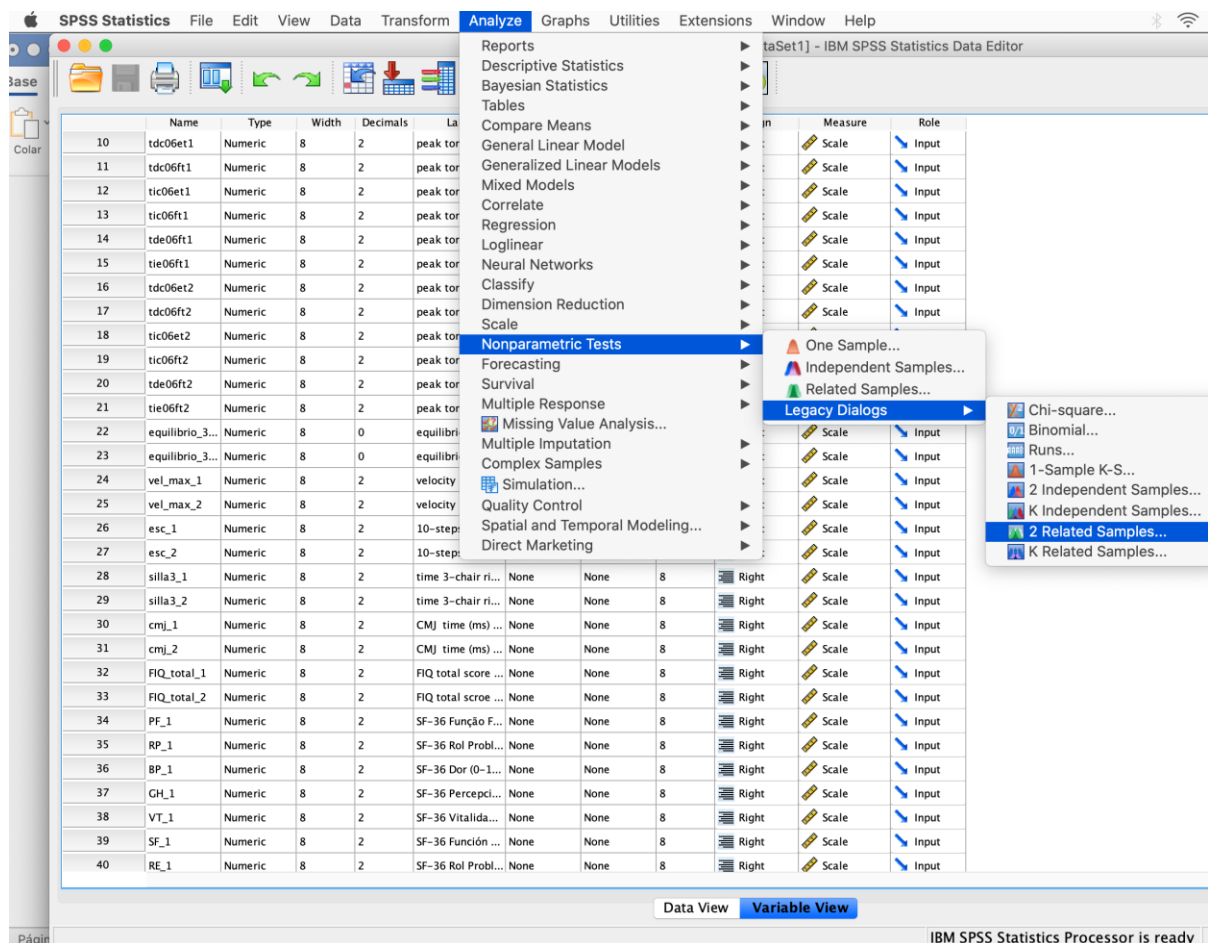


Figura 67
Indicação das opções a selecionar para realizar o Teste de *Wilcoxon*

Observe que o SPSS indica o teste de Wilcoxon em “*Test Type*”. Na caixa de “*Test Pairs:*” colocamos a variável “CMJ time (ms) pre” no “*Pair – Variable1*”, e a variável “CMJ time (ms) post” no “*Pair – Variable2*”. Em “*Options...*”, podemos solicitar as estatísticas descritivas, selecionando “*Descriptive*” e, a seguir clicando em “*Continue*”. Para terminar clicamos em OK (Figura 68).

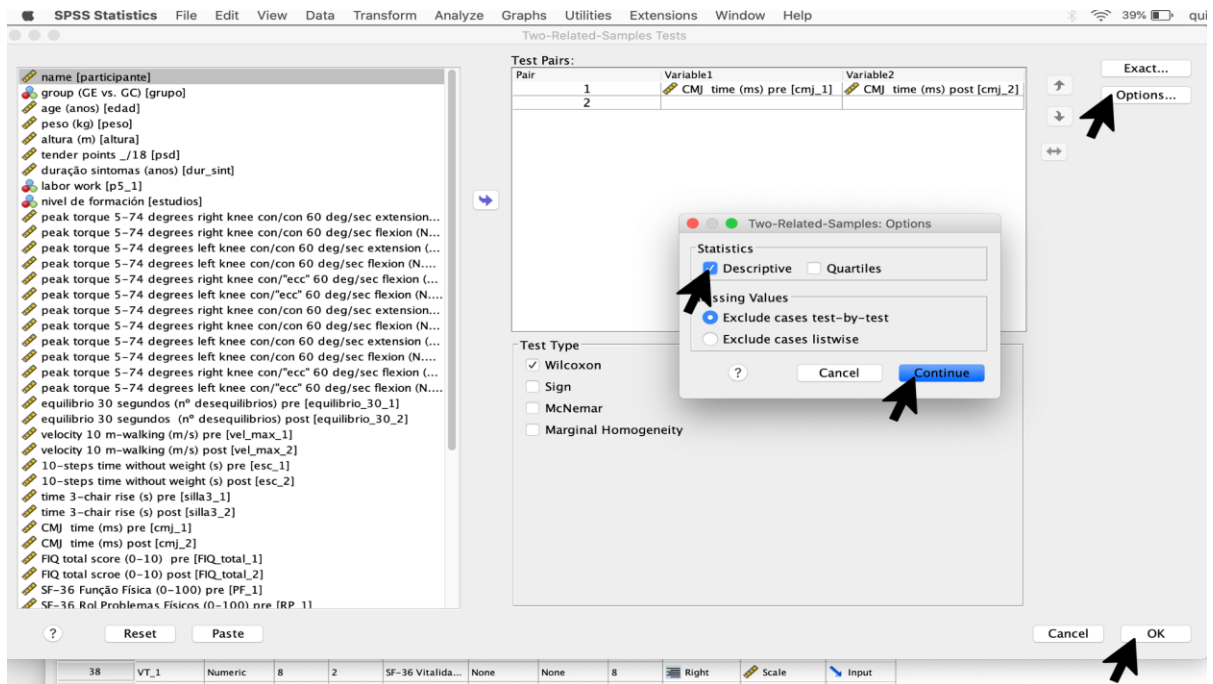


Figura 68

Indicação das opções a selecionar na janela "Two-Related-Samples: Options"

Na janela de resultados do SPSS observamos duas tabelas, a primeira mostra as estatísticas descritivas (número de sujeitos avaliados, média, desvio padrão, mínimo e máximo) para cada uma das variáveis selecionadas. A segunda apresenta os *ranks* positivos, negativos, médios e a soma dos mesmos. Para saber se há diferenças estatisticamente significativas devemos dar atenção à terceira tabela, onde observamos que não existem diferenças estatisticamente significativas ($p = 0,804$ superior a $0,050$) entre a média do CMJ avaliada antes do programa de exercício ($288,28 \pm 70,28\text{mm}$) e a média do CMJ avaliada após o programa ($288,06 \pm 55,90\text{mm}$), logo podemos concluir que o programa de exercício não teve efeitos significativos no CMJ dos participantes do estudo (Figura 69).

➔ NPar Tests

Tabela de estatísticas descritivas

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
CMJ time (ms) pre	32	288,2812	70,28478	174,00	408,00
CMJ time (ms) post	32	288,0625	55,90426	200,00	402,00

Wilcoxon Signed Ranks Test

Tabela de ranks

Ranks

		N	Mean Rank	Sum of Ranks
CMJ time (ms) post – CMJ time (ms) pre	Negative Ranks	16 ^a	14,31	229,00
	Positive Ranks	13 ^b	15,85	206,00
	Ties	3 ^c		
	Total	32		

- a. CMJ time (ms) post < CMJ time (ms) pre
- b. CMJ time (ms) post > CMJ time (ms) pre
- c. CMJ time (ms) post = CMJ time (ms) pre

Test Statistics^a

	CMJ time (ms) post – CMJ time (ms) pre
Z	–,249 ^b
Asymp. Sig. (2-tailed)	,804

Para saber o valor de p , olhar para “Asymp. Sig. (2-tailed)”

- a. Wilcoxon Signed Ranks Test
- b. Based on positive ranks.

Figura 69

Indicação dos resultados obtidos após realizar o Teste de *Wilcoxon*

6.4 TESTE DE KRUSKAL – WALLIS

O teste de Kruskal – Wallis é usado para verificar se existem diferenças entre várias amostras não relacionadas, quando as variáveis dependentes não cumprem a normalidade e/ou a homogeneidade das variâncias, sendo a alternativa não paramétrica ao teste ANOVA – one way.

Imagine que queremos saber se há diferenças no “IMC” (variável que não cumpre com uma distribuição normal dos dados) entre as três regiões de Portugal da variável “Região”. Lembre-se que a variável “Região” tem três categorias: 1 – Norte; 2 – Centro; e 3 – Sul, para efetuar a análise devemos selecionar a opção: **Analyze** → **Nonparametric Tests** → **Legacy Dialogs** → **K Independent Samples...** (Figura 70).

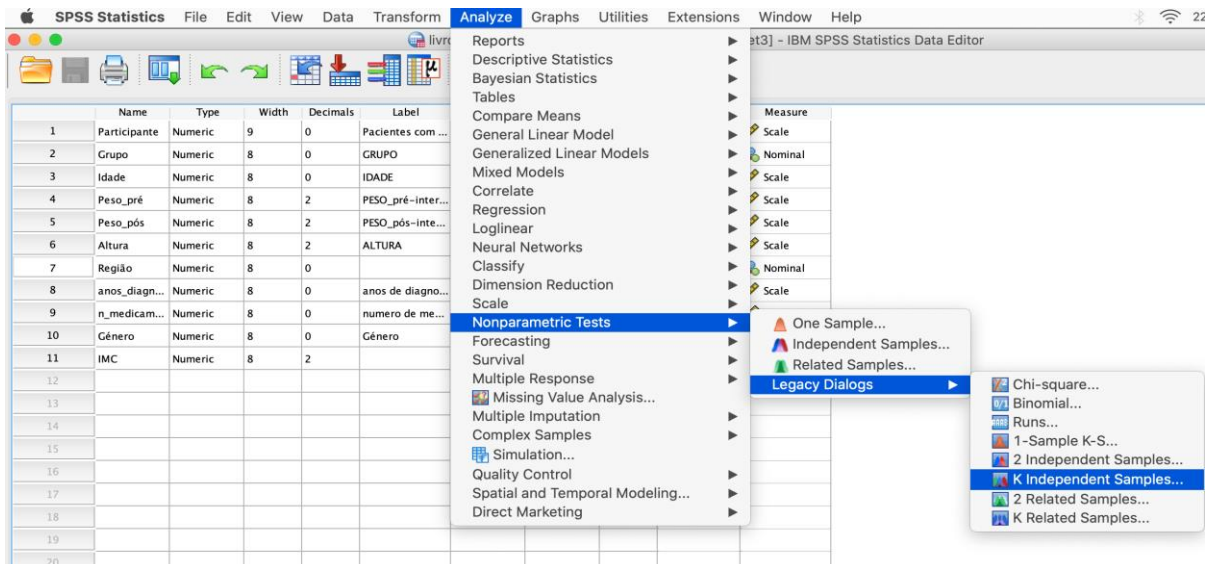


Figura 70
Indicação das opções a selecionar para realizar o Teste de *Kruskal-Wallis*

Observe que o SPSS marca o teste de Kruskal – Wallis como selecionado em “*Test Type*” (parte inferior esquerda). Selecionamos a variável “IMC” e passamo-la para a caixa “*Test Variable List:*” e a variável “Região” para a caixa “*Grouping Variable:*”. Devemos clicar em “*Define Range...*” para definir os grupos (Figura 71).

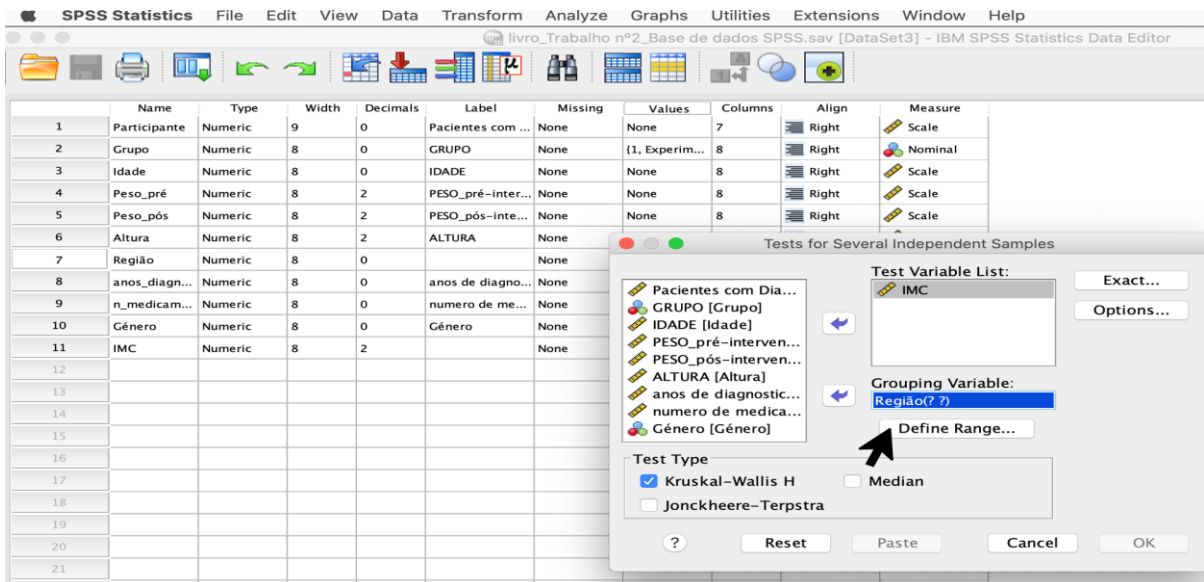


Figura 71
Janela " *Tests for Several Independent Samples*", com indicação para selecionar " *Define range*"

Em seguida, devemos escrever os valores 1 e 3 como mínimo e máximo, respetivamente (que são os grupos da variável “Região” na base de dados), e clicamos em “*Continue*” (Figura 72).

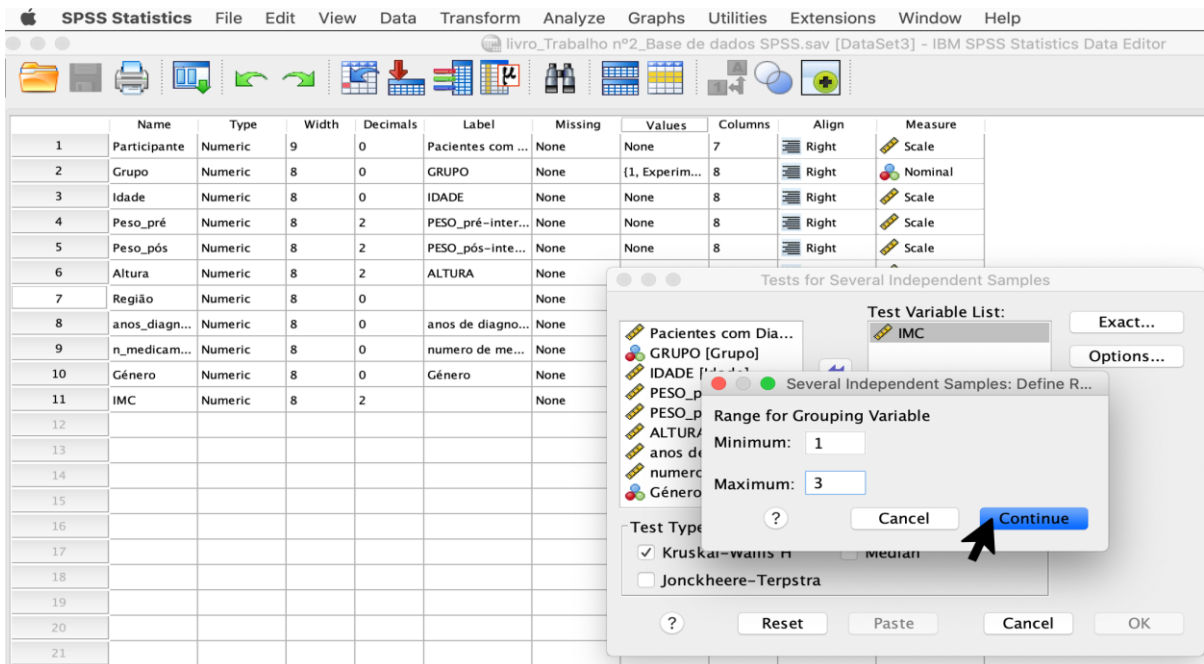


Figura 72
Indicação dos valores a colocar na janela "Several Independent Samples: Define Range"

Em "Grouping Variable", podemos verificar que os grupos aparecem definidos corretamente como "Região (1 3)". Finalmente, clicamos em OK para obter os resultados (Figura 73).

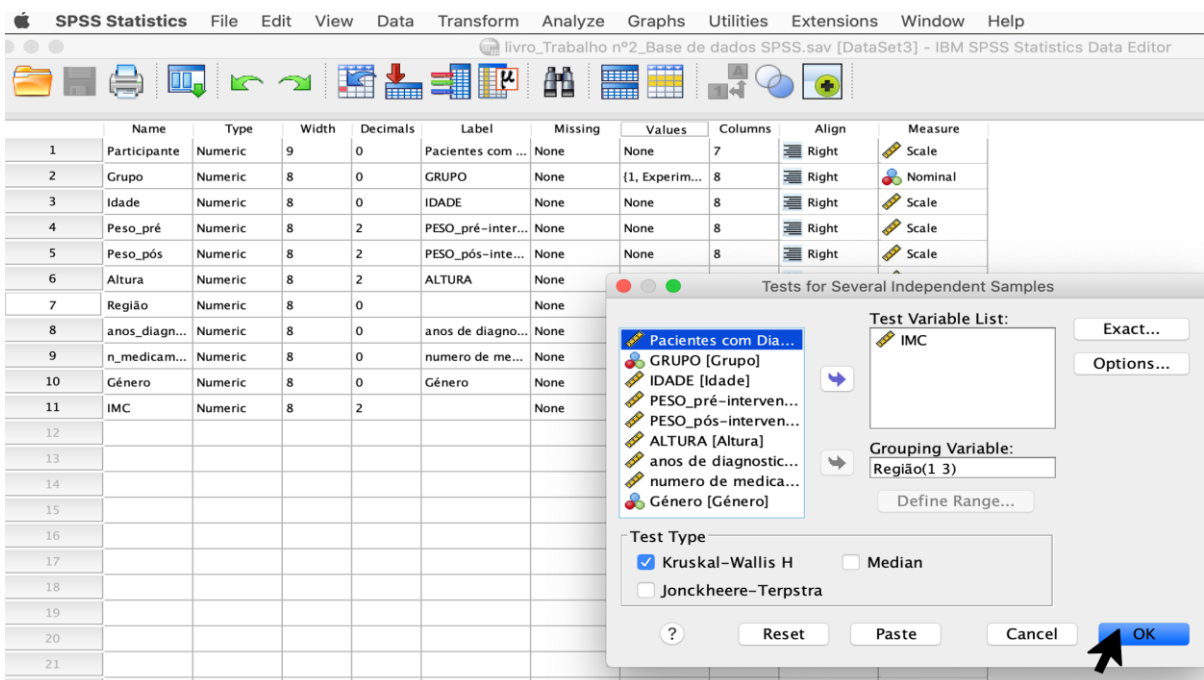


Figura 73
Janela "Tests for Several Independent Samples" preenchida

A janela de resultados mostra uma primeira tabela de *ranks*, mas para sabermos se existem diferenças significativas no IMC entre as três regiões de Portugal, devemos observar na segunda tabela o valor de p em “Asymp. Sig.”. Neste exemplo, os resultados mostram que $p = 0,521$; o que indica que não há diferenças significativas no IMC entre as três regiões de Portugal. Para comparar as regiões duas a duas, é necessário realizar o teste de Mann – Whitney, já que o teste de Kruskal – Wallis não realiza comparações dois a dois (analisar Figura 74).

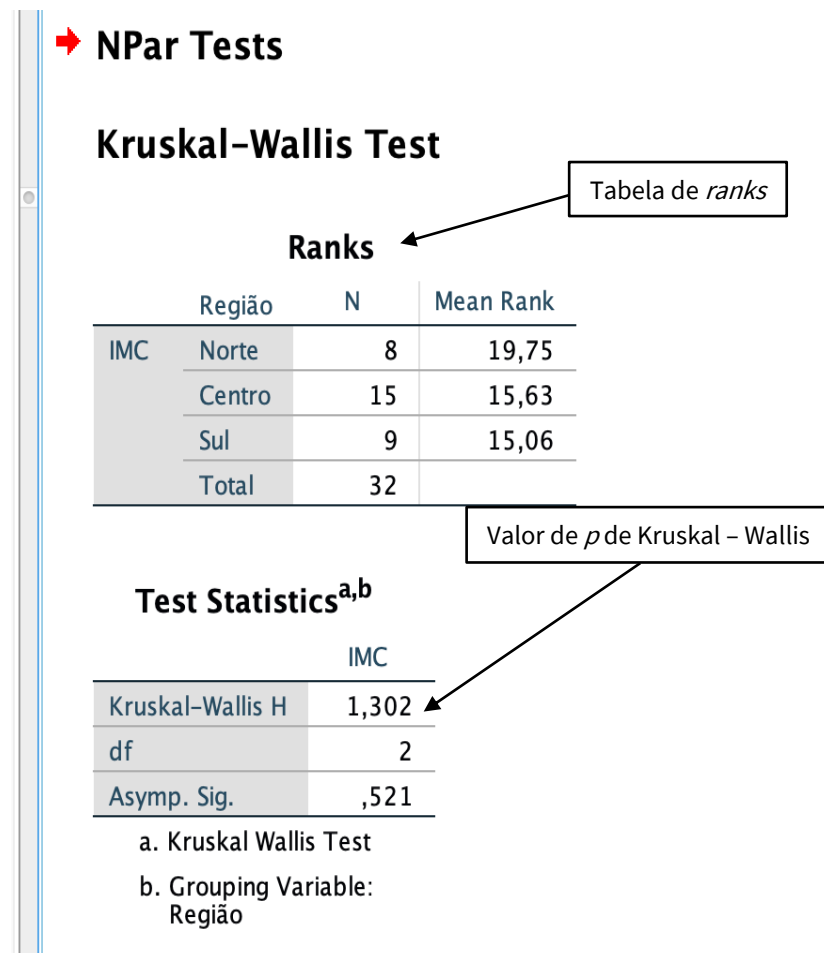


Figura 74
Indicação dos resultados obtidos após a realização do Teste de *Kruskal-Wallis*

6.5 TESTE DE SPEARMAN

O coeficiente de Spearman (ρ) é usado para medir a correlação entre duas variáveis quando não apresentam uma distribuição normal dos dados ou quando a variável tem uma escala ordinal. Tal como, o coeficiente r de Pearson, o coeficiente ρ de Spearman, segue as mesmas normas de interpretação, variando entre -1 e 1. Quanto mais perto de 0, menor é a relação entre as duas variáveis, e quanto mais perto de 1 ou -1, maior é a relação entre variáveis. O zero indica que não existe nenhuma relação linear entre as duas variáveis, sendo a relação perfeita quando $\rho = 1$ ou $\rho = -1$. O sinal indica a direção da correlação:

o sinal positivo de ρ indica que quanto maior o valor de uma variável, maior será o valor da outra variável, o que também pode indicar que quanto menor é o valor de uma variável menor será o valor da outra variável. No caso do sinal negativo, quanto maior o valor de uma variável, menor será o valor da outra variável (Camacho, 2002).

Em seguida, queremos verificar as correlações das variáveis “velocity 10 m – walking”, “10 – steps time without weight”, “CMJ” e “SF-36 physical function”. Como estas não cumprem com uma distribuição normal dos dados, devemos usar o teste de Spearman, para isso selecionamos a opção: **Analyze** → **Correlate** → **Bivariate** (Figura 75).

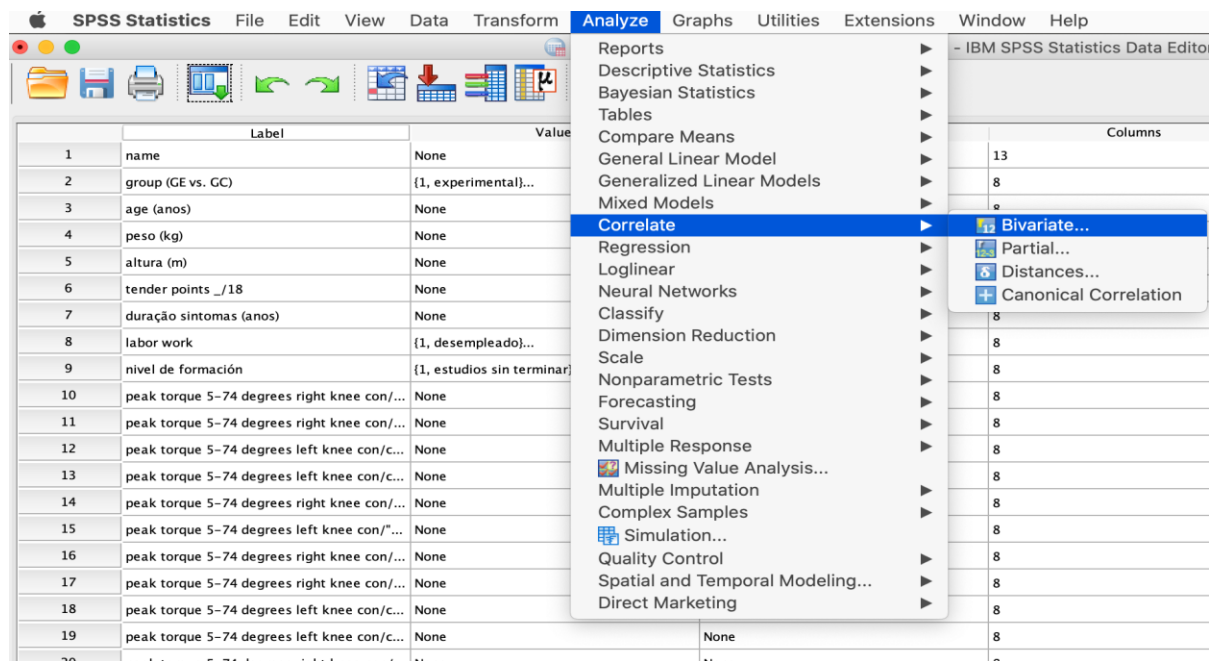


Figura 75
Indicação das opções a selecionar para realizar a Correlação de *Spearman*

Atenção que aparece pré-selecionado a correlação de Pearson, então devemos de anular essa opção e selecionamos a opção “Spearman”, no canto inferior esquerdo da Figura 76. Seguidamente, selecionamos as variáveis desejadas e transferimo-las para a caixa de “Variables”, clicando em OK, para finalizar.

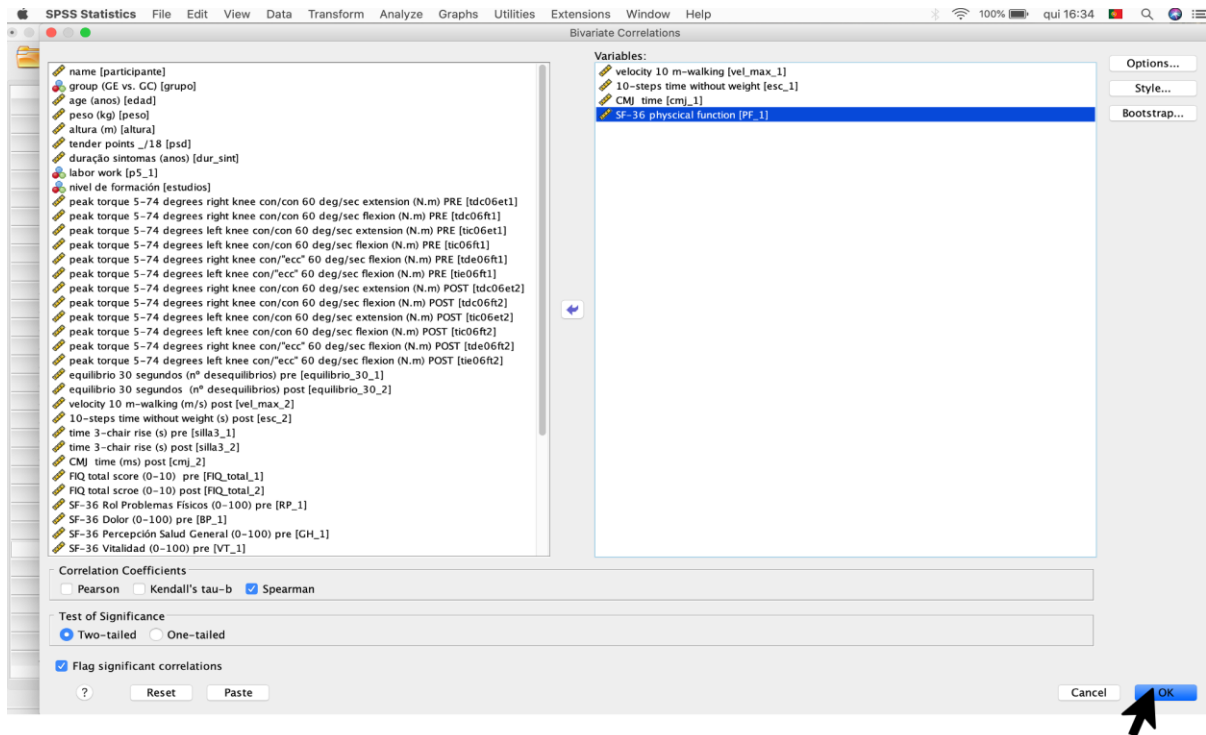


Figura 76
Janela "Bivariate Correlations" preenchida

Na janela de resultados do SPSS, é possível observar uma tabela com os coeficientes ρ de Spearman e os valores de p . Tal como acontece, na correlação de Pearson, a tabela mostra as correlações significativas com um ou dois asteriscos (* quando o valor de $p < 0.050$; ** quando o valor de $p < 0.010$) ao lado do coeficiente ρ de Spearman. Pode ainda obter o valor exato de p em "Sig. (2-tailed)" da tabela. Se observarmos os resultados obtidos, a variável "velocity 10 m – walking", evidencia uma correlação significativa inversa ($p < 0,000$; $\rho = -0,654$) com "10 – steps time without weight"; e uma correlação significativa direta ($p = 0,009$; $\rho = 0,454$) com "CMJ". É ainda de salientar, a variável "10 – steps time without weight" que mostra uma correlação significativa inversa ($p < 0,000$; $\rho = -0,684$) com "CMJ (analisar Figura 77).

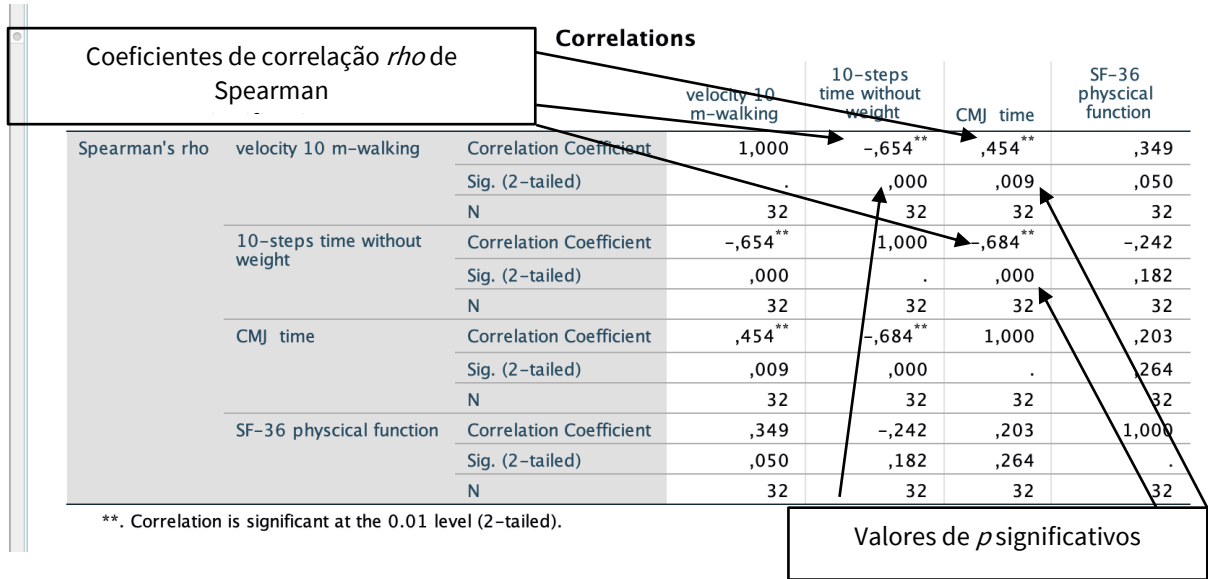


Figura 77
Indicação dos resultados obtidos após realizar a Correlação de *Spearman*

BIBLIOGRAFIA

Camacho, J. (2002). Estadística con SPSS para Windows, versión 11. Madrid: Rama.

Maroco, J. (2018). Análise estatística com o SPSS statistics. (7ª ed). Pêro Pinheiro: ReportNumber.

Verma, J.P. (2016). Sports Research with Analytical Solution using SPSS. New York: John Wiley & Sons, Inc.