

INVESTIGAÇÃO EM ADMINISTRAÇÃO E
GESTÃO EDUCACIONAL

INTRODUÇÃO À
INVESTIGAÇÃO
QUANTITATIVA E
ANÁLISE SPSS

MAGE
Ano Letivo 2020/2021
Maria Álvares



Conteúdo

1. A ESCOLHA DOS MÉTODOS E DAS TÉCNICAS	4
2. DIMENSÕES, INDICADORES E VARIÁVEIS	8
2.1. O Inquérito por questionário: dimensões e indicadores, tipos de variáveis e tipos de resultados	9
3. A ELABORAÇÃO DE UM INQUÉRITO POR QUESTIONÁRIO	12
3.1. Plano de análise	13
3.2. Organização do inquérito	14
3.3. Tipos de perguntas e formas de questionar	15
3.4. Apresentação do questionário	21
4. AMOSTRAGEM	23
5. ORGANIZAR UMA BASE DE DADOS	25
6. O AMBIENTE SPSS	26
6.1. Visão geral	26
6.2. Janelas de dados e de resultados: breve apresentação	28
6.3. Início de uma sessão de trabalho	29
6.3.1. Construção de um ficheiro de dados: definição de Variáveis	31
6.4. Introdução e edição de dados	36
6.5. Importação de bases de dados	36
6.6. Breve apresentação à análise de dados no SPSS	40
BIBLIOGRAFIA	42

Este recurso de aprendizagem, desenvolvido no âmbito da Unidade Curricular e Investigação em Gestão e Administração Educacional do Mestrado de Administração e Gestão Educacional da Universidade Aberta tem como objetivo fornecer alguns conteúdos sobre a condução de uma investigação com recurso a métodos e técnicas de investigação quantitativa. Tem a ambição de incentivar os estudantes incorporarem, no âmbito das suas investigações, uma perspetiva mais abrangente e quantificada, mesmo entre os que pretendem desenvolver uma abordagem sobretudo qualitativa. Refere-se, assim, não apenas à elaboração e análise de inquéritos por questionário, mas também à utilização de outras fontes de dados, como bases de dados pré-existentes ou relatórios e análises estatísticas produzidas por instituições nacionais e internacionais. Organiza-se, sinteticamente em torno de quatro temas:

1. Critérios da opção por um método quantitativo
2. Abordagens quantitativas para a análise de fenómenos sociais: fontes de dados, tipos de informação
2. A construção de um inquérito por questionário
3. Tipos de variáveis e implicações
4. Introdução ao SPSS e suas funcionalidades

Este recurso constitui-se fundamentalmente como um guia: os principais temas são afluídos, mas procurou-se uma visão telescópica, com exemplos adaptados às particularidades da investigação em contextos educacionais. O objetivo é fornecer alguma informação essencial e conselhos práticos aos estudantes no processo de conceção, montagem e implementação dos seus modelos de análise.



Os campos “Saber mais” apontam alguns recursos e referências a explorar para quem deseje aprofundar algumas matérias



São propostos alguns desafios que visam aferir conhecimentos adquiridos

Esperamos que este recurso seja útil. Boas investigações!

1. A escolha dos métodos e das técnicas



Colourful Questions Linda Wood (www.fineartamerica.com)

A escolha dos métodos e técnicas constitui uma etapa relevante do processo de investigação, aquela em que operacionalizamos, em que concretizamos, uma estratégia para responder à questão que desejamos esclarecer. Antes de mergulharmos no exercício de construir e operacionalizar um modelo de análise, convém parar para refletir, de forma a avaliar se:

- a) é claro, mesmo para alguém que não conheça o tema, o que se pretende saber (consegue explicar a sua tese em duas linhas?);
- b) é possível identificar um debate, ou seja, existe um debate científico em que se demarcam posições ou, no contexto de estudo, existe uma questão que suscita perplexidade;
- c) se o investigador consegue, a partir das referências, chegar a uma – ou várias – hipóteses sobre o que pode estar a acontecer e produzir uma teoria explicativa para os fenómenos;
- d) se conhece e manipula as técnicas suficientemente para as utilizar na prossecução dos objetivos de investigação, em particular, se elas são as mais adequadas ao contexto de estudo. É sobretudo sobre este último ponto que se debruça este recurso.

Síntese do processo de investigação

Os diferentes métodos (qualitativos/ quantitativos) e as diversas técnicas (análise de conteúdo, análise estatística, etc.) possuem diferentes vantagens e desvantagens. Em cada pesquisa, é necessário ponderar questões relativas à sua adequabilidade, perguntando se aquilo que se deseja saber pode ser respondido por via da utilização da estratégia escolhida.

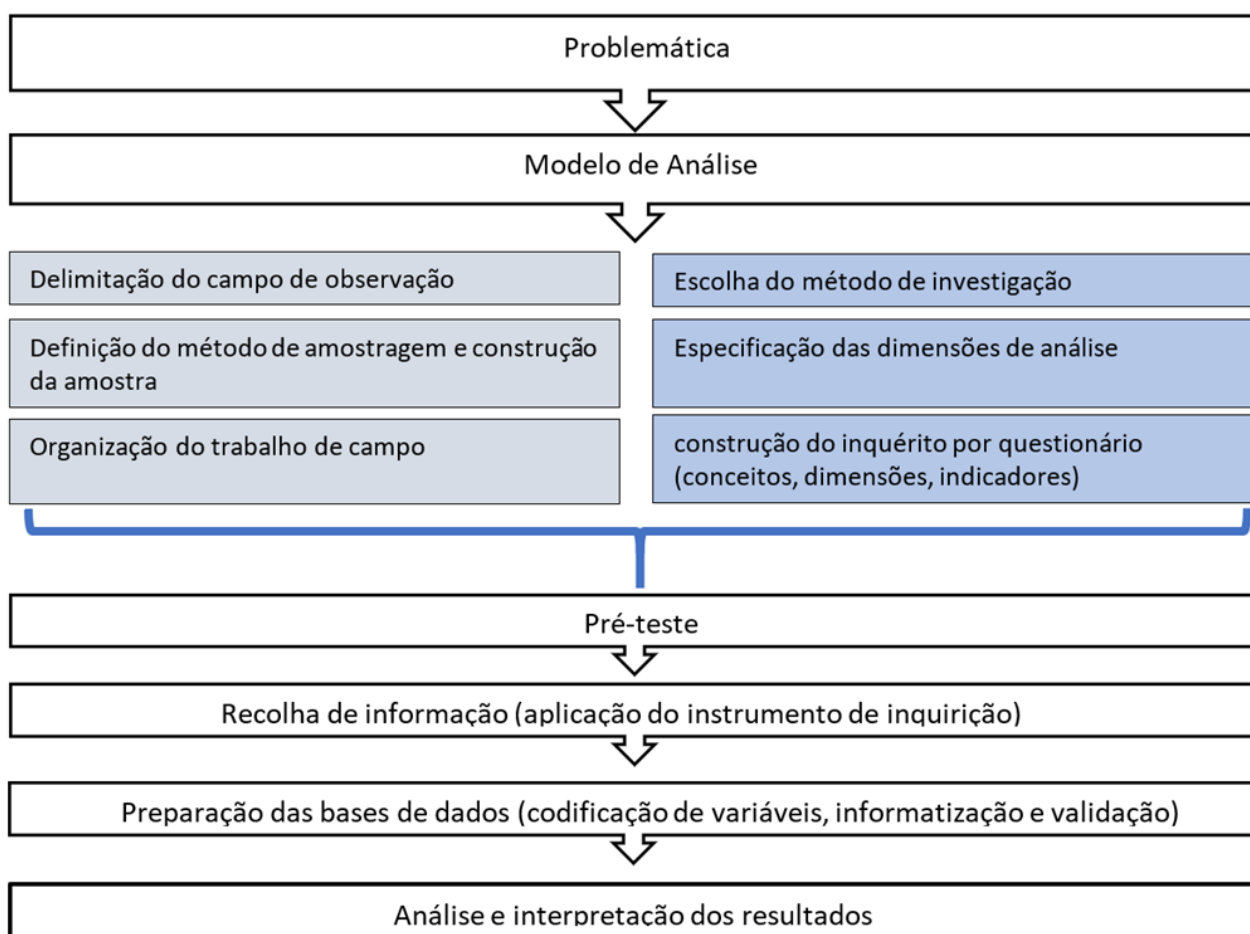
1.1. Abordagem qualitativa e quantitativa

Análise Extensiva / Quantitativa	Análise Intensiva / Qualitativa
Número alargado casos / populações amplas	Número reduzido de casos
Recolha de informação sobre um conjunto limitado e previamente definido de dimensões de análise	Recolha de informação em profundidade e em continuidade; recolha de material informativo heterogéneo
Utilização de instrumentos padronizados de recolha de informação (ex: inquérito por questionário)	Utilização de procedimentos pouco normalizados de recolha e tratamento de informação (entrevistas, grelhas de observação)
Recurso a técnicas de amostragem; análise quantitativa dos dados	Recurso a técnicas de análise de conteúdo (entre outras).

Análise de relações entre variáveis	Exploração de uma multiplicidade de facetas e dimensões, com o objetivo de compreender o fenómeno em estudo na sua totalidade
-------------------------------------	---

A organização de uma investigação a partir de métodos quantitativos – seja com recurso a bases de dados já existentes ou a partir da recolha e tratamento de dados primários exige um trabalho sólido de conceção e planeamento da estratégia. Aplicando uma metáfora culinária à investigação, os métodos qualitativos são como a cozinha tradicional: possui ingredientes e receitas definidas, mas existe alguma margem para a criatividade individual e correção de curso. Os métodos quantitativos, por seu turno, são como a doçaria, que exige grande rigor nos processos e procedimentos e elevada precisão.

Seja qual for a nossa escolha, esta deve ser ponderada em função do objeto de estudo e, principalmente, estar bem suportada numa problematização aprofundada, assentar num modelo de análise claro e deduzido em tarefas bem definidas.



Durante alguns anos, assistiu-se a uma luta paradigmática entre advogados do método qualitativo e do quantitativo. As posições extremadas geraram discussões essencialistas e radicalismos que atrasaram o surgimento do paradigma da ciência moderna, em que o cruzamento de tipos e fontes de informação constitui uma prática desejável. Mas é já hoje boa prática cruzar informação e recolha de dados quantitativos e qualitativos e assegurar a triangulação de informação, ou seja, cruzar fontes. As fontes de informação podem ser secundárias, ou seja, informação já recolhida, ou informação primária, gerada com o propósito de responder aos objetivos da investigação.

1.2. fontes de informação

Fontes de informação secundária:	Fontes de informação primária:
-Fontes bibliográficas	-O inquérito
-Fontes administrativas	-A inquirição
-Fontes estatísticas	Bases de dados Questionários

Existem várias bases de dados que podem ser exploradas. Em educação, a Direção Geral de Estatísticas da Educação e Ciência publica um conjunto de estudos comparativos nacionais e internacionais e disponibiliza os dados recolhidos pelas escolas no âmbito dos sistemas de informação do ME, como InfoEscolas. Existe ainda informação estatística comparativa internacional em fontes oficiais (Eurostat) e a partir de instituições como a [OCDE](#), [Banco Mundial](#) e [IEA](#) (International Association for the Evaluation of Educational Achievement), entre outros. O diretório de informação sobre educação disponível para os alunos do MAGE sintetiza essas fontes.

Uma investigação deve procurar combinar várias fontes de informação (triangulação):

- fontes oficiais
- acesso a bases de dados
- recolha de dados primários



[Reflexões sobre triangulação metodológica](#)

2. Dimensões, Indicadores e variáveis



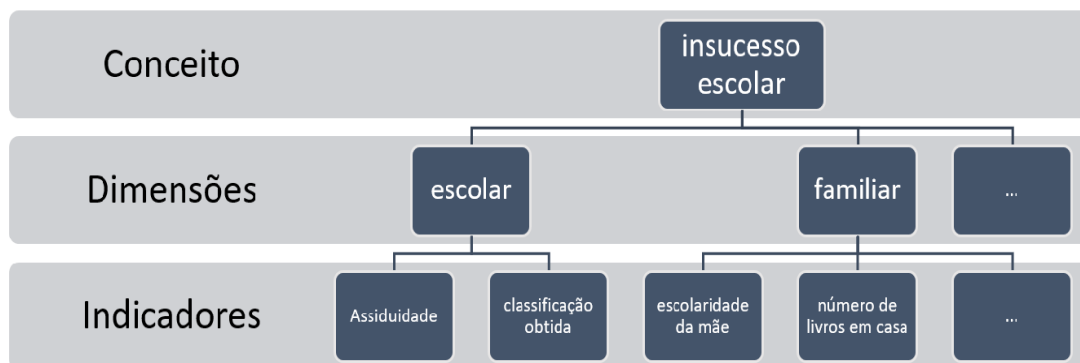
"Multiple Dimensions" Priyadarshini Ohol (www.mojarto.com)

A opção pela recolha autónoma de dados e aplicação de um inquérito por questionário é a mais acertada em situações de pesquisa em que existe já alguma informação exploratória, em que a intenção é descrever um fenómeno, considerando a sua amplitude e é possível equacionar hipóteses explicativas originais, para as quais não existe já informação disponível. A opção pelos quantitativos permite fazer uso de um conjunto de técnicas sofisticadas de análise da informação e a recolha de dados primários, aceder a informação *à medida* das nossas intenções. Esta opção possui, contudo, algumas limitações:

- A recolha de informação pode ser morosa, complexa e dispendiosa;
- Existe uma forte irreversibilidade: a grelha de observação é rígida
- É um processo sequencial e qualquer erro ou ambiguidade tem efeito nas fases seguintes e no resultado;
- A conceção/ redação do questionário é um processo crítico e complexo, orientado pelo quadro teórico de partida (função de comando da teoria) e pela exploração estatística planeada;
- As declarações dos/as inquiridos/as podem ser afetadas por elementos do desenho da investigação: 1) pela formulação das perguntas e categorias de resposta, incluindo aspeto relativos ao texto e à sequência); 2) por variáveis associadas ao contexto, relativas ao inquérito como situação social, como o entrevistador ou local de entrevista;
- Exige forte standardização ao nível do texto das perguntas, modalidades de resposta e sequência, de forma a assegurar a comparabilidade dos resultados e permitir analisar a relação entre variáveis;
- Não é possível compreender casos individuais: a análise do conjunto das respostas sobrepõe-se à análise das respostas de cada indivíduo;

2.1. O Inquérito por questionário: dimensões e indicadores, tipos de variáveis e tipos de resultados

Num inquérito por questionário parte-se da construção teórica – sustentada em reflexões e pesquisas anteriores – de um **conceito**, que define um conjunto de **dimensões** analíticas que se materializam em **indicadores** e se deduzem em **variáveis** (respostas). Veja-se o seguinte exemplo:



[Dimensões e Indicadores de análise IST](#)

Tipos de variáveis

As variáveis são características dos elementos de uma população, quantificáveis ou categorizáveis e que podem assumir valores diferentes em função da unidade de observação (amostra, população). Dividem-se em dois grandes grupos: as qualitativas e as quantitativas.

As **variáveis qualitativas** são medidas de atributos que codificamos para poder tratar em modelos estatísticos. O sexo, a nacionalidade, o curso que frequenta ou a religião são alguns exemplos. Este tipo de variáveis reportam simplesmente a informação nominal, referem-se a características que não são ordenáveis. Assumem um número variável de categorias: algumas variáveis são dicotómicas, outras são multicategóricas. Variáveis de sim/não devem ser tratadas como qualitativas dicotómicas (por exemplo num questionário aos hábitos de estudo, perguntas como estuda todos os dias?).

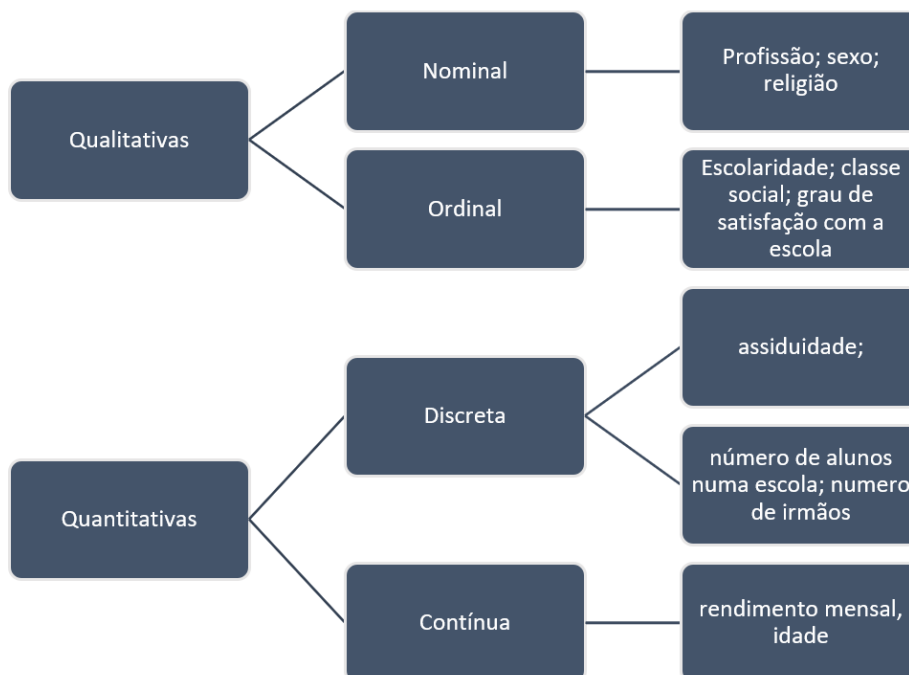
Variáveis como o curso frequentado são politónicas e têm tantas categorias quantas forem necessárias. Algumas politónicas são ordinais, ou seja, são suscetíveis de ter uma organização do menor para o mais pequeno ou vice versa. É caso da escolaridade e da classe social. Estas, pese embora não sejam quantitativas podem, em algumas circunstâncias, ser estatisticamente tratadas como tal.

As variáveis quantitativas são já por si quantidades. São medidas em escalas contínuas ou discretas. Em escalas contínuas temos por exemplo o rendimento familiar ou a idade de um indivíduo. Em escalas discretas temos as que não assumem valores intermédios, como o número de filhos ou de reprovações escolares.

É importante saber qual a tipologia das variáveis usadas, dado que as medidas usadas para tratar a informação são diferentes consoante o tipo e por isso é importante ter presente o que se pretende saber e como (quais as operações estatísticas) o saber. A variável idade, por exemplo, se medida em anos completos, é quantitativa (contínua)

mas, se foi operacionalizada em classes etárias (0 a 5 anos, 6 a 10 anos, etc....), é qualitativa (ordinal). Estas classificações têm implicações nos testes estatísticos passíveis de utilizar para proceder à análise do fenómeno. É recomendado que a categorização seja feita à posteriori, pois uma variável contínua pode ser categorizável, mas o contrário já não é possível.

2.1. tipos de variáveis



DESAFIO FÓRUM 1: O ME pretende lançar um programa de combate ao insucesso escolar. Solicita às escolas que se candidatem ao programa, apresentando um diagnóstico da evolução da situação de insucesso escolar nos últimos anos no agrupamento que dirige. Que variáveis vai utilizar e qual a tipologia dessas variáveis?



DESAFIO FÓRUM 2: Se fosse necessário aprofundar o detalhe e compreender qual a relação entre o insucesso e género e a distribuição entre escolas do agrupamento, que outras variáveis deveria escolher?

2.. Tipos de resultados

Tipos de resultados	
Padrões e perfis de distribuição de variáveis	Análises univariadas
Análise da relação entre variáveis	Análises bivariadas
Análises topológicas, tipologias, modelos de dependência, entre outros	Análises multivariadas

3. A elaboração de um inquérito por questionário



Confusion Painting Emily Maynard

Este capítulo do nosso e-book visa explorar alguma informação básica e conselhos práticos para a elaboração de um inquérito por questionário. Elaborar um bom questionário implica estar bem consciente das etapas e não saltar nenhuma. É frequente e compreensível a tentação de iniciar pela elaboração de perguntas que respondem ao que queremos saber. Mas esse é um caminho que nos pode trazer muitas complicações evitáveis. Pensar cada passo e não seguir atalhos leva-nos mais rapidamente ao objetivo.

3.1. Plano de análise

Para elaborar um bom questionário é fundamental ter explicitados: 1) os objetivos da investigação; 2) as hipóteses; 3) os métodos e técnicas de análise de dados. 4) as escalas de resposta às perguntas do questionário;

Os objetivos da investigação

Habitualmente uma investigação possui um objetivo principal e um, ou vários, objetivos secundários e o principal objetivo numa investigação quantitativa é o de testar a hipótese geral de investigação. Os objetivos secundários, por seu turno, procuram testar hipóteses mais específicas. Por exemplo, no campo da educação, uma investigação sobre sucesso escolar pode testar a relação entre o desempenho escolar e a satisfação face à escola. Os objetivos secundários podem testar outras hipóteses específicas, como:

- o género influencia a relação entre o desempenho escolar e a satisfação face à escola;
- a classe social de origem influencia a relação entre o desempenho escolar e a satisfação face à escola;
- o tipo de curso influencia a relação entre o desempenho escolar e a satisfação face à escola;

O objetivo principal mobiliza as principais variáveis da investigação, mas que podem ter diferentes fontes. Neste caso, a satisfação face à escola tem de ser medida por meio de um questionário aplicado aos alunos, mas a variável “desempenho escolar” pode ser recolhida a partir dos sistemas de informação existentes nas escolas. No exemplo, os objetivos secundários apontam características dos alunos que devem ser incluídas entre as perguntas do questionário.

A formulação da hipótese principal também estabelece uma determinada relação entre as variáveis, classificando de “independente” e “dependente”. Este não é, necessariamente, um atributo da variável mas resulta da relação entre fenómenos que

se estabelece nas hipóteses de trabalho: na nossa hipótese principal podemos considerar que a satisfação face à escola influencia (variável independente) o desempenho escolar (variável dependente) – e explorar se os alunos, quanto mais gostarem da escola, melhor desempenho possuem – mas também é possível equacionar que o desempenho escolar (variável independente) tenha efeitos na satisfação face à escola (variável dependente) – procurando analisar a hipótese de que quanto melhores os resultados escolares, maior a satisfação que possuem com a escola. Como neste caso, na maior parte das situações em investigação social é difícil estabelecer relações de causa-efeito. Podemos – e devemos – identificar correlações, mas muito dificilmente chegaremos a explicações de causa-efeito. Nos fenómenos sociais, as análises baseadas em relações causa-efeito são, de forma geral, recusadas. Os fenómenos sociais são teias complexas de causalidades cruzadas, em que as variáveis interagem entre si, relacionando-se com outras variáveis. Correlação não implica causalidade!

hipóteses da investigação, escalas de resposta e métodos de análise de dados

Cada objetivo liga-se a uma – ou mais- hipóteses de investigação testadas estatisticamente a partir dos dados recolhidos num questionário. As hipóteses de investigação têm de ser deduzidas em hipóteses de trabalho, ou seja, traduzidas em variáveis (e.g. o “género” será operacionalizado a partir da variável “sexo” ou o absentismo usado como “proxy” do risco abandono escolar¹)

3.2. Organização do inquérito por questionário

Os questionários são organizados por secções, cada um correspondendo a um bloco de perguntas que se refere a uma dada dimensão de análise. As secções e perguntas devem ser numeradas sequencialmente e cada questionário deve possuir um número que o identifica.

A planificação das secções do questionário é uma etapa fundamental que evita a introdução de perguntas desnecessárias e/ou a exclusão de perguntas essenciais. Questionários excessivamente longos tornam-se pouco fiáveis, pois os inquiridos rapidamente deixam de atentar às questões. Questionários em que estão em falta questões essenciais, tornam-se inúteis. Cada pergunta deve ser cuidadosamente ponderada e ser claro qual o seu objetivo para esclarecer a questão de partida.

As secções devem espelhar as várias hipóteses que se pretende testar. Assim, no nosso exemplo, serão secções do questionário:

- caracterização pessoal

¹ Uma dedução que implica sempre um enquadramento consistente, teórica e empiricamente suportado num quadro teórico e conceptual.

- trajetória e frequência escolar
- relação/satisfação com a escola

3.3. Tipos de perguntas e formas de questionar

Este é um aspeto essencial a ter em conta na elaboração de uma pesquisa quantitativa. Dele depende, em grande parte, o sucesso da investigação. Um mau questionário inviabiliza uma pesquisa. As questões colocadas podem solicitar informação de diferentes tipos:

3.1. tipos de informação

Tipos de informação	
Factos	“Em que escola andas”?
Opiniões	“Consideras a tua escola uma boa escola?”
Atitudes	Concordas com a frase “esta escola preocupa-se com os seus alunos ”?
Preferências	“esta escola é melhor ou pior que as outras que conheces”?
Valores	“qual a importância que atribuis aos seguintes aspetos da escola:”
Satisfação	“em que medida estás satisfeito/a com a tua escola?”

Razões, crenças, motivos, etc

A formulação das perguntas é um aspeto muito relevante. Por exemplo, a questão “consideras a tua escola uma boa escola” e “em que medida estás satisfeito/a com a tua escola?” não representam a mesma coisa e é possível uma resposta positiva à primeira não ser coincidente com uma evidência de insatisfação na segunda: a escola pode ser boa, mas não se adequar aos projetos do inquirido/a ou não ser boa e, no entanto, satisfazê-lo/a. A situação de inquirição num questionário não permite reconstruir os sentidos e justificações das respostas e por isso o cuidado na formulação das questões deve ser uma preocupação fundamental.

Quando as perguntas se dirigem a factos, é importante ter em conta dois aspetos: a sensibilidade das perguntas (por exemplo, questões sobre o rendimento familiar ou outros aspetos da vida privada tendem a ser mais difíceis de responder) e o conhecimento que o inquirido/a pode efetivamente deter no momento. Um exemplo deste segundo caso é o “número de livros em casa”, frequentemente utilizado como proxy para o capital cultural da família em estudos de educação, mas que trata informação que os alunos, podendo aceder, frequentemente não possuem no

momento em que são questionados, gerando validade baixa do indicador (não mede corretamente o que pretende medir).

Uma pergunta pode também ser geral ou específica. No nosso exemplo, a questão “qual a importância que atribuis aos seguintes aspectos da escola” e “qual a importância que atribuis aos seguintes aspectos da tua escola” não são idênticas. Quem responde a perguntas sobre atitudes, opiniões, satisfação, preferências e gostos interpreta muitas vezes uma pergunta geral em termos da sua situação específica, mas não é necessariamente possível fazer inferências de uma a partir de outra. Recomenda-se que, por secção, as questões gerais precedam questões específicas

Questões abertas e perguntas fechadas

As perguntas podem ser abertas ou fechadas. No primeiro caso, o/a inquirido/a constrói a resposta nas suas próprias palavras. No caso das perguntas fechadas, tem de optar entre as alternativas apresentadas. Em certos casos, a mesma questão pode ser colocada em termos abertos ou fechados. As entrevistas exploratórias e os pré-testes servem, nalguns casos, para identificar as alternativas de resposta mais comuns a questões abertas e “fechar” perguntas. Noutras situações, na fase de análise criam-se categorias a partir de respostas. Um questionário pode ser totalmente composto de perguntas fechadas ou perguntas abertas, mas, na esmagadora maioria dos casos, possui uma combinação de perguntas dos dois tipos.

Tipos de Perguntas	Vantagens	Desvantagens
Perguntas abertas	<ul style="list-style-type: none">-Possibilidade de reconstruir o sentido da resposta e relacionar dados;-Informação mais detalhada, mais completa e mais fiável;-Possibilidade de surgir informação inesperada, maior liberdade exploratória	<ul style="list-style-type: none">- Necessidade de interpretação de respostas;-Codificação mais demorada;-Limitação da análise estatística possível;-Permeabilidade à interpretação pelo codificador
Perguntas fechadas	<ul style="list-style-type: none">-Permite a utilização de técnicas estatísticas;-Parametriza respostas, reduzindo problemas de compreensão;-Reduz o tempo de preparação da base de dados para exploração de resultados	<ul style="list-style-type: none">- A informação, por vezes, não esclarece totalmente;-Resvala, por vezes, numa excessiva simplificação da realidade.-Não permite, por vezes, que o inquirido possa responder o que desejaria.

Um questionário composto por perguntas abertas é usado quando o investigador não possui meios (tempo ou oportunidade) para realizar entrevistas, mas pretende obter informação qualitativa sobre um fenómeno, quando ainda não existe muita literatura sobre o tema em investigação ou quando, existindo, esta não permite identificar as variáveis mais relevantes, recomendando um estudo preliminar. Um questionário composto por perguntas fechadas é adequado quando a natureza das variáveis relevantes é bem conhecida e se pretende recolher informação definida ou quando se pretende criar uma variável a partir de um conjunto de perguntas (dedução das componentes do conceito).

Extensão e clareza das questões

A clareza das questões é um elemento fundamental para se considerar a qualidade de um questionário e, em geral, verifica-se que quanto mais extensa é uma questão, menos clara se torna. A linguagem utilizada deve ser simples e direta, evitando expressões literárias e termos elaborados. As questões devem ser pensadas de forma a serem inteligíveis por pessoas com baixa literacia. A realização de um pré-teste, envolvendo pessoas de perfil diversificado, correspondente à população em estudo é fundamental, entre outros motivos, para avaliar a adequabilidade da formulação das perguntas. Alguns dos lapsos mais frequentes:

Perguntas múltiplas – são perguntas que incluem diferentes questões. Por exemplo: “Em casa estudas num local específico?”. Esta pergunta pressupõe uma resposta positiva a uma pergunta prévia: “estudas em casa?”. A resposta também gerará ambiguidade, pois ficaremos sem saber se a resposta negativa resulta de não estudar em casa, se é porque não possui um local reservado para o estudo ou porque, estudando em casa, tem preferência por variar o local de estudo. A formulação correta da questão implica dedução com a criação de uma primeira pergunta filtro: “Costumas estudar em casa?” (S/N) Se sim: responde às seguintes questões...

Perguntas que misturam conjunções e disjunções – a utilização na mesma pergunta de expressões como “e” e “ou” deve ser evitada por confundir e produzir respostas ambíguas. Por exemplo: “na tua escola existe psicólogo e mediador ou monitor?” Numa situação destas a melhor alternativa seria a apresentação da pergunta: “na tua escola existe:” e uma tabela de contingência para registar a resposta com as opções sim/não em coluna e os três tipos de recursos de apoio em linha.

Perguntas não neutras- são perguntas que induzem uma resposta, seja porque estão formuladas de uma certa maneira, assumem um pressuposto que impele a um dado sentido ou porque a pergunta não admite uma das possibilidades de resposta. No primeiro caso, temos, por exemplo: “O estudo é muito importante para o sucesso

escolar. Costumas estudar em casa?”, perguntas que incluem uma das categorias da resposta (ex: não ou sim, numa pergunta de sim/não) ou expressões como “É normal”, a “maioria” ou “frequentemente”. Um exemplo de má formulação: “quão satisfeito estás com o teu diretor de turma?”. Perguntas que incluam adjetivos quantitativos, como “muito”, “sempre” ou “nunca” são também de evitar, assim como as que sugiram respostas estereotipadas (a não ser em estudos sobre estereótipos).

Perguntas indefinidas- são perguntas em que é deixado ao/à inquirido/a o sentido da resposta, sem que as opções de resposta o esclareçam. Por exemplo: em pergunta aberta, “o que gostas mais na tua escola?” (a questão é sobre o espaço físico? Sobre os professores e outros atores escolares? tudo?).

Fiabilidade e validade das questões

A validade refere-se à capacidade do instrumento de medida – neste caso, um questionário, pergunta ou bateria de perguntas – aferir aquilo que pretende aferir. Ou seja, se é composto de indicadores que refletem efetivamente o que se pretende retratar. A fiabilidade, por seu turno, é a consideração de que a medição feita é constante e não influenciada pelas (variáveis) circunstâncias da aplicação do instrumento de medição (pelo viés de perguntas, pela forma ou contexto de aplicação do questionário). Em alguns casos recorre-se a “perguntas de controlo” para avaliar a consistência da informação recolhida. A mesma questão pode ser colocada em diferentes pontos do questionário, formulada de diferentes formas. Esta estratégia deve ser ponderada pois faz aumentar a dimensão do questionário e potencialmente torná-lo mais desinteressante e repetitivo. No momento de tratamento dos dados existem algumas operações que permitem testar a consistência de um conjunto de questões. Contudo, nessa altura será tarde para recuperar informação. A melhor estratégia é dedicar esforço e atenção à formulação das perguntas, de modo a conseguir respostas válidas e fiáveis.

Perguntas de escala e de escolha múltipla

Nas perguntas fechadas é necessário listar opções de resposta ou definir uma gradação que permite posicionar respostas. Algumas escalas são nominais, outras ordinais e, menos usadas, existem ainda escalas de intervalo e de rácio. As perguntas de escala ou “variáveis gradativas” são frequentemente utilizadas. Usam-se na medição de características físicas (comprimento, tempo de reação) e, nesse caso, resultam da aplicação direta de instrumentos padronizados, ou visam a medição de características associadas ao comportamento humano (“satisfação com um serviço”, “atitude face à escola”, “importância atribuída ao estudo”. No segundo caso, a medição é concretizada pelo/a inquirido/a, que se torna instrumento de medida sobre o objeto de estudo.

Existem dois princípios fundamentais que as escalas devem respeitar: a **exaustividade**, ou seja, uma abrangência que permita representar todas as possibilidades de resposta (o que inclui “não sabe/ não responde” e/ou “Outros”) e a **exclusividade**, ou seja, cada situação particular ser representada numa única opção.

As **escalas de intervalo** são pouco utilizadas no âmbito de questionários das ciências sociais, e caracterizam-se por associar um valor numérico mais elevado na escala a uma quantidade maior da variável medida, assumindo valores negativos e não considerando o valor de zero como ausência como, por exemplo, na medição da temperatura do ar. Já nas **escalas de rácio**, o valor “zero” não é arbitrário - é absoluto ou “real”. É assim possível fazer inferências sobre um rácio de valores na escala. É vulgar tratar as escalas de resposta do tipo “Escala de Avaliação” como escalas de intervalo, mas, na realidade, as escalas de avaliação são somente escalas de ordem, disfarçadas de escalas de intervalo ou rácio.

As **escalas nominais** consistem num conjunto de categorias de resposta qualitativamente diferentes e mutuamente exclusivas. Sendo possível atribuir um código numérico a cada resposta na fase de tratamento dos dados, estes valores não produzem uma realidade ordenável e, portanto, não é legítimo apresentar medidas estatísticas como média ou desvio-padrão sobre medidas de escala ordinal. É, no entanto, possível calcular a moda e produzir análises univariadas e bivariadas com recurso ao SPSS (inferência estatística não paramétrica).

As **escalas ordinais** admitem a ordenação numérica das suas categorias (opções de resposta) mas não a análise da magnitude de diferenças. Algumas questões solicitam uma ordenação de itens, outras a atribuição de um valor numa escala face a cada item. Por exemplo a pergunta “ordena de 1 a 5 a tua satisfação face a estes aspetos da tua escola, atribuindo o valor 5 ao aspeto que consideras mais positivo e 1 ao mais negativo:” e a pergunta “de 1 a 5, sendo 1 o menos satisfatório e 5 o mais satisfatório, indica quão satisfeito estás com os seguintes aspetos relativos à tua escola” são ambas escalas ordinais, mas de diferente tipo, cada uma permitindo operações estatísticas distintas. No geral, são de evitar perguntas do primeiro tipo, pois verificam-se mais limitativas em termos de opções de tratamento de dados. As segundas são

Se apenas se classifica	Se inclui noção de ordem entre categorias	Se inclui informação sobre diferenças entre números	Se permite informação contida nos quocientes
•escala nominal	•escala ordinal	•escala de intervalo	•escala de rácio

habitualmente designadas por escalas de avaliação, ou escalas de Likert/ tipo Likert e permitem a análise de percentis em geral e o coeficiente de correlação ordinal

A distinção entre variáveis nominais, ordinais e numéricas não é, na maioria dos casos, baseada em qualquer propriedade intrínseca das próprias variáveis, mas uma distinção que depende da decisão do investigador. Atualmente a distinção mais frequente é entre variáveis não métricas (nominais e ordinais) e métricas. No caso das variáveis métricas é possível estabelecer comparações quantitativas entre os valores por elas assumidos, o que não acontece com as variáveis não métricas.

As **escalas de avaliação** mais frequentemente utilizadas referem-se a:

Referencial	Escala
Quantidade	1-Nada; 2-Um pouco; 3-Médio; 4-Muito; 5-Bastante
Concordância	1-Discordo; 2- Discordo um pouco; 3-Concordo; 4-Concordo em absoluto
Frequência	1-Nunca; 2-Raramente; 3-Às vezes; 4-Frequentemente; 5-Sempre
Satisfação	1-Nada satisfeito; 2- Insatisfeito; 3-Satisfeito; 4- Muito satisfeito
Probabilidade	1-Nada provável 2- Pouco provável; 3. Provável; 4- Muito provável; 5- Certo

A ordenação dos valores atribuídos a cada categoria (valor que a variável pode assumir) pode ser crescente ou decrescente, sendo importante apenas manter uma dada ordenação e garantir que não se misturam dois tipos de avaliação (por exemplo, frequência e quantidade) na mesma escala.

Nalguns exemplos dados, foi usada uma escala de quatro valores, noutros, uma escala de cinco. A forma das escalas [não é irrelevante](#), mesmo que se sigam todos os critérios de qualidade. Quando a escala tem 5 itens designam-se por “escalas de Likert”, quando tem mais ou menos, é uma “escala tipo Likert”. Likert propôs o método de construção de escalas em 1932, sugerindo a utilização de afirmações e o posicionamento dos inquiridos numa escala de cinco categorias de acordo com a sua concordância, de modo a permitir reservar um ponto médio na escala – o três – para a posição de “indeciso”. A medida de atitude podia ser obtida pela soma dos valores escolhidos, razão pela qual estas escalas são também conhecidas por “escalas adicionadas”. Não existe um número de pontos ideal para a escala. Estas escalas, quando bem construídas, permitem o tratamento como variáveis quantificáveis, mas é relevante para tal garantir igual distância entre os diferentes pontos da escala (natureza gradativa da escala).

A questão do ponto médio é também, por vezes, geradora de algum debate entre os que consideram que as alternativas de resposta dadas devem ser exaustivas e admitir

a posição de “indeciso”, enquanto outros consideram que existe uma tendência para respostas conservadoras e que o ponto médio equivale a uma não reposta, pois o/a inquirido/a acaba não se posicionando, sobretudo se se tratar de um assunto sensível. O ponto neutro pode colocar desafios na fase de análise dos dados, já que, por vezes, é necessário trabalhar uma escala, dicotomizando-a (“concordo”/ “não concordo”), não sendo uma prática aceitável subsumir o ponto médio em nenhuma das duas categorias criadas.



DESAFIO FÓRUM 3

Considera que deve ser utilizado um ponto neutro ou médio numa escala gradativa? No fórum, discuta a questão com os seus colegas e troque argumentos com os que pensam de forma diferente.

As perguntas de Resposta Múltipla (não confundir com perguntas múltiplas) são as que permitem ao/a inquirido/a dar mais do que uma resposta à questão, de acordo com instruções dadas. Podem ser relevantes, mas deve-se ter em conta que os métodos estatísticos para comparar as respostas são restritos a frequências e percentagens. Não é legítimo aplicar estatísticas indutivas na comparação destas respostas (por exemplo o teste do Qui-quadrado para uma amostra) mas é possível considerar cada resposta como uma variável.

3.4. Apresentação do questionário

É boa prática apresentar uma introdução ao questionário, explicando os seus objetivos. Os elementos a ter presentes são:

- Razão da aplicação do questionário;
- Um pedido de cooperação no preenchimento do questionário;
- Tempo médio de preenchimento (de acordo com os pré-testes);
- Descrição curta da natureza geral do questionário (satisfação, opinião, etc.);
- O nome da instituição e/ou do investigador (centro de investigação, empresa);

Existem ainda outros três elementos que é fundamental ter presentes de forma a conduzir uma inquirição com respeito por elementares preocupações éticas:

Garantia de “consentimento informado”, ou seja, deixar claro, após informação sobre os objetivos, características e condições de realização de que o/a entrevistado/a só participa se desejar.

Garantia de anonimato, quer dizer, assegurar explicitamente os/as entrevistados/as de que a sua identidade não será conhecida a partir da informação facultada;

Garantia de confidencialidade, asseverar que a informação que pode comprometer o anonimato não será revelada.

O layout do questionário deve ser apelativo e claro e para isso é necessário um compromisso entre extensão e clareza. Devem ser mantidos espaços adequados entre as perguntas (e dentro das escalas de resposta) e não utilizar caracteres muito pequenos para encurtar o seu tamanho. As questões relativas ao mesmo tema devem estar juntas na mesma secção e as escalas de resposta devem seguir todas o mesmo sentido. É fundamental dar instruções novas quando se muda a forma das perguntas e não assumir que os respondentes sabem responder.



[Síntese sobre elaboração de questionários](#)

4. Amostragem



Maurice Qullinan

Uma amostra é um subgrupo de indivíduos que selecionamos com o objetivo de representar a população do nosso estudo. Esta pode ser representativa ou não representativa. Nem sempre é necessário constituir uma amostra. Podemos ter a possibilidade de trabalhar o universo do fenómeno, o que acontece, por exemplo em projetos de investigação-ação. O seguinte quadro sintetiza as principais formas de amostragem:

	Tipo Amostra	Descrição	Vantagens	Desvantagens
Aleatórias	Aleatória	Grupo aleatório da população	Altamente representativa	Há que conhecer toda a população Custos elevados
	Estratificada	Aleatória dentro de grupos pré-definidos; obedece à % ou quota de representação dos sujeitos em cada estrato	Assegura a representação de todos os grupos da população	Complexa de efectuar Deve contemplar todos os estratos da população
	Clusters	Aleatória dentro dos sucessivos clusters	Possibilidade de escolha aleatória dentro dos clusters	Equivalência dos clusters num mesmo nível
Não aleatórias	Criterion	Escolha de sujeitos de acordo com critério	Amostra adaptada ao estudo	Pouca representatividade
	Conveniência	Grupos intactos	Estudo em contexto real	Pouca generalização de resultados além do grupo
	Quotas	Seleciona sujeitos por quota de representação na população	Assegura a selecção dos sujeitos com as características adequadas	Impossível provar da representatividade
	Acidental	Pedindo voluntários, ou aproveitando um grupo que está disponível no momento	Forma fácil de conseguir sujeitos suficientes para um estudo	Sem representatividade
	Bola de neve	Identifica-se um membro e pede-se que identifique outros membros da população	Alcançar populações difíceis de identificar	Pouca representatividade

Retirado de Coutinho, Clara (2007) Métodos de Investigação em Educação: concebendo o plano de investigação. Policopiado.

Quanto maior for uma amostra, mais reduzido será o “erro amostral”, ou seja, o desvio dos dados recolhidos face às disposições da população de em estudo. Não existem números ideais, pois este valor depende muito da dimensão da população em estudo, das características do fenómeno, entre outros aspetos. Mais importante do que um número, é a garantia de que a amostra é bem selecionada. Uma amostra pequena, contudo, compromete o tipo de tratamento estatístico que é possível realizar e um valor menor que 30 casos presta-se pouco a análises do tipo estatístico.



5. Organizar uma Base de Dados



Maurice Qullinan

A organização da informação de partida é muito importante e pode comprometer todos os resultados a que chegamos. Uma base de dados, qualquer que seja o programa utilizado para análise, tem algumas regras fundamentais que devemos ter sempre em atenção:

Cada linha representa um caso/ indivíduo (os indivíduos não se devem repetir). Cada coluna representa uma variável – nem sempre é imediato definir o que são indivíduos e o que são variáveis. Por exemplo, em dado estudo, os jovens em situação de insucesso são indivíduos e as escolas que frequentam são variáveis. Numa outra investigação, as escolas são os casos e a variável pode ser no número de alunos em situação de insucesso. O nome das variáveis deve estar escrito na primeira linha.

Devemos ter sempre uma variável de ID que nos permite fazer a verificação dos indivíduos, ou acrescentar variáveis se for o caso. Não usar células "unidas"

A codificação das variáveis dicotómicas binárias deve ser de 0 e 1 sendo o valor de 1 o que queremos dar mais relevo no estudo. Por exemplo, se me interessa estudar quem teve uma melhoria de classificação escolar a variável "mais sucesso" deve estar codificada em 1 para "sim" se melhorou as notas e 0 para "não".

As variáveis polinómicas devem estar codificadas por um código que esteja bem definido no documento de suporte à base de dados. As variáveis ordinais devem estar organizadas por ordem crescente, e não ter informação que não tem significado estatístico, como "outros". A categoria "outros" deve ser reservada para variáveis qualitativas.

6. O Ambiente SPSS

A Universidade Aberta disponibiliza aos seus alunos uma licença temporária do "Statistical Package for the Social Sciences" – SPSS, de forma a permitir que os alunos explorem bases de dados existentes ou analisem dados recolhidos por via quantitativa nas suas pesquisas e disponibiliza um [manual de instalação](#). Aqui são brevemente explorados alguns aspetos relacionados com a sua utilização.

6.1. Visão geral

Na janela de dados existem duas formas de visualização – à semelhança do que acontece nas tabelas de ACCESS – ambas em formato matricial: uma permite a visualização dos dados (Data View) e a outra a visualização da sua estrutura (Variable View). É possível aceder a um ou outro modo clicando sobre o respetivo "separador" (Figuras 6.1. e 6.2.).

Figura 6.1: janela de editor de dados (data view)

No modo “data view” as linhas representam os casos ou unidades de observação (geralmente os indivíduos), enquanto as colunas representam as variáveis (figura 6.1.) No modo “variable view”, estão listadas as variáveis e as suas características. Nesse caso, cada linha corresponde a uma variável e as colunas definem as suas propriedades (Figura 6.2.).

Figura 6.2: janela de editor de variáveis (variable view)

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	id	Numeric	3	0	id	None	None	8	Right
2	startdate	String	57	0	startdate	None	None	19	Left
3	idade	Numeric	2	0	Quantos anos tens?	None	None	8	Right
4	Ano	Numeric	3	0	Em que ano andas?	{1, 5ºano}...	None	5	Right
5	Decalage	Numeric	8	0	Diferença da idade real face à idade modal para o ano de escolaridade	None	None	8	Right
6	Sexo	Numeric	3	0	Sexo	{1, Feminino}...	None	7	Right
7	Local	Numeric	3	0	Onde é a tua escola?	{1, Casal de Cambra}...	None	7	Right
8	Turma	String	15	0	Turma	None	None	5	Left
9	Via	Numeric	3	0	Via de ensino:	{1, Regular}...	None	6	Right
10	ASE	Numeric	3	0	Tens algum apoio da Ação Social Escolar (ASE)? (i.e. Recebes da tua...	{1, Sim}...	3	3	Right
11	V8	Numeric	8	0	Quem é o teu/a tua Encarregado/a de Educação?	{1, Pai}...	None	10	Right
12	V8.1	String	36	0	Quem é o teu/a tua Encarregado/a de Educação?[Outro]	None	None	9	Left
13	V9	Numeric	3	0	Até onde é que o teu/ a tua Encarregado/a de Educação estudou?	{1, Não sabe ler nem escrever}...	None	10	Right
14	V10	Numeric	1	0	Já chumbaste alguma vez?	{1, Sim}...	None	8	Right
15	V11.1	Numeric	8	0	Chumbou no 1º Ciclo (1º, 2º, 3º e/ou 4º ano)	{0, Reprovou noutro ciclo}...	4	6	Right
16	V11.2	Numeric	8	0	Chumbou no 2º Ciclo (5º e/ou 6º ano)	{0, Reprovou noutro ciclo}...	4	19	Right
17	V11.3	Numeric	8	0	Chumbou no 3º Ciclo (7º, 8º 3/ou 9º ano)	{0, Reprovou noutro ciclo}...	4	17	Right
18	V12.1	Numeric	1	0	Chumbei porque não estudava	{0, Não Seleccionado}...	None	8	Right
19	V12.2	Numeric	1	0	Chumbei porque a matéria era difícil de perceber	{0, Não Seleccionado}...	None	8	Right
20	V12.3	Numeric	1	0	Chumbei porque os professores não explicavam bem a matéria	{0, Não Seleccionado}...	None	8	Right
21	V12.4	Numeric	1	0	Chumbei porque não gostava dos professores	{0, Não Seleccionado}...	None	8	Right
22	V12.5	Numeric	1	0	Chumbei porque a matéria era pouco interessante	{0, Não Seleccionado}...	None	8	Right

No modo Data View, as células da matriz são preenchidas com os atributos de cada caso em cada variável – por exemplo, as respostas de cada indivíduo a cada pergunta num questionário. As respostas são, em geral, codificadas, pelo que um ficheiro de dados costuma estar preenchido com números (mesmo que representem qualidades e não quantidades), tal como se pode observar na Figura 6.1.

É necessário ter presente que: a) operações entre objetos (i.e. entre linhas) não são, em geral, possíveis; e b) operações entre variáveis (i.e. sobre colunas) afetam, em princípio, todos os casos (i.e., linhas).

Um ficheiro de dados pode ser criado de raiz, ou ser importado de outras fontes (EXCEL, ACCESS). No primeiro caso há que começar por definir a estrutura dos dados, antes de proceder à sua informatização; no segundo caso, as características mais genéricas das variáveis são importadas juntamente com a informação mas devem ser confirmadas.

6.2. Janelas de dados e de resultados: breve apresentação

No SPSS existem várias janelas de trabalho. Habitualmente, numa sessão de trabalho, estão ativas duas janelas: uma referente aos dados (data ou variable view); outra referente aos resultados ou outputs. Na primeira janela visualiza-se a base de dados que está a ser analisada, enquanto na segunda vão sendo acumulados os resultados das análises solicitadas. Para utilizadores mais experimentados, existe ainda a janela de syntax e que permite algumas operações não disponíveis nos comandos.

Editor de Dados (Data Editor) – Janela onde é apresentado o ficheiro de dados (em modo variable ou data view), ou seja, a informação que vai ser trabalhada do ponto de vista estatístico. Esta janela abre automaticamente quando entramos no SPSS, mantendo-se aberta durante toda a sessão de trabalho. Fechar todas as janelas de dados conduz ao encerramento do programa.

Janela de Resultados (Output Navigator) – Janela onde são apresentados os resultados dos procedimentos de análise de dados: quadros de distribuição de frequências, médias, correlações, testes de hipóteses, tabelas, gráficos... Esta janela abre automaticamente na primeira vez que for acionado um comando. Todas as análises feitas ao longo da sessão de trabalho serão apresentadas sequencialmente na janela de resultados. A alternância entre o ficheiro de dados e o de resultados (outputs) faz-se da forma idêntica à de ambiente Windows, para alternar entre vários aplicativos que estejam abertos.

No SPSS, ao contrário do que acontece noutros programas (por exemplo com o EXCEL), o ficheiro de resultados é estático, ou seja, reúne o histórico das análises efetuadas e apresenta o estado do ficheiro de dados quando a análise foi realizada. Assim,

alterações introduzidas ao nível dos dados não provocam alterações nos outputs já obtidos e quando corrigimos o valor de uma variável, por exemplo, se detetarmos que foi mal introduzido, todas as análises já efetuadas sobre essa variável terão de ser refeitas.

A janela de resultados (outputs), está dividida em duas partes: à esquerda situa-se a lista dos itens que fazem parte do output (resultados e outros objetos, por exemplo, caixas de texto), à direita visualizam-se os resultados propriamente ditos, os quais são apresentados pela ordem em que foram pedidos, sendo sempre antecedidos de um título. A estrutura em árvore permite a navegação entre resultados apurados numa sessão de SPSS.

Figura 6.3. janela de resultados (outputs)

The screenshot shows the IBM SPSS Statistics Viewer interface. On the left, a tree view labeled 'Lista de itens' contains a list of objects including 'Crosstabs', 'Title', 'Notes', 'Active Dataset', 'Case Processing Summary', and 'Sexo * Já chur'. A blue arrow points from this list to the main output area. The main area, labeled 'Resultados apurados', displays the following tables:

CROSSTABS
/TABLES=Sexo BY Reprovação
/FORMAT=AVALUE TABLES
/CELLS=COUNT ROW TOTAL
/COUNT ROUND CELL.

Crosstabs
[DataSet1] C:\Users\Maria\Desktop\BD Alunos\BD.alunos_18_Jun.sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sexo * Já chumbaste alguma vez?	792	100,0%	0	,0%	792	100,0%

Sexo * Já chumbaste alguma vez?: Crosstabulation

		Já chumbaste alguma vez?		Total
		Sim	Não	
Sexo	Feminino	Count 100	283	383
		% within Sexo 26,1%	73,9%	100,0%
Masculino	Count 138	271	409	
	% within Sexo 33,7%	66,3%	100,0%	
Total		Count 238	554	792
		% within Sexo 30,1%	69,9%	100,0%

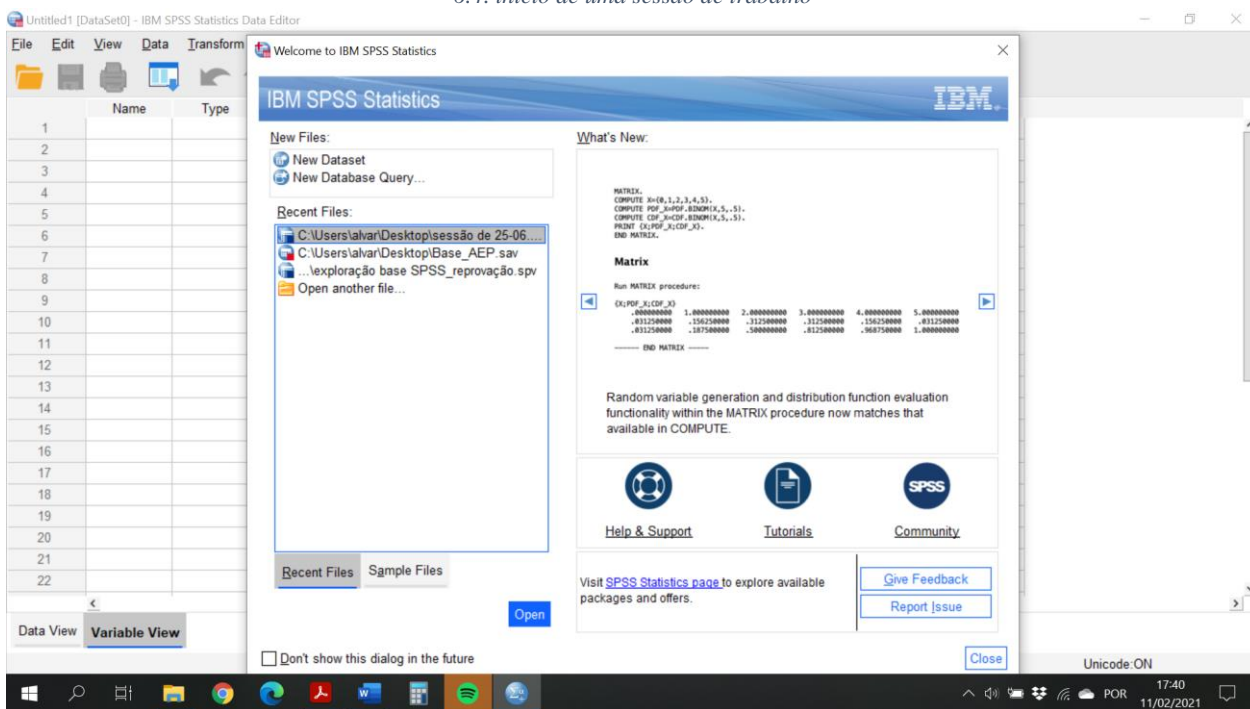
Um duplo clique sobre qualquer item (texto, tabelas ou gráficos) permite editar e modificar as suas características. Consoante a situação, podemos ter acesso a um editor de texto, de tabelas, ou de gráficos, cada um com menus e opções específicas. Entre outras operações, é possível alterar o tipo de letra e clarificar os títulos das figuras, formatar tabelas, por exemplo, alterando a sua aparência ou invertendo linhas e colunas. Pode-se também construir gráficos a partir de valores apresentados em tabelas. Estes valores devem ser previamente selecionados (com a tabela editada) e, com o botão do lado direito do rato, escolhe-se a opção “Create Graph”.

6.3. Início de uma sessão de trabalho

Quando acedemos ao SPSS abre-se automaticamente uma janela de boas vindas. Caso se pretenda construir uma nova base de dados deve escolher-se a opção “new data set”; caso se pretenda analisar os dados de uma base de dados já existente, deve então

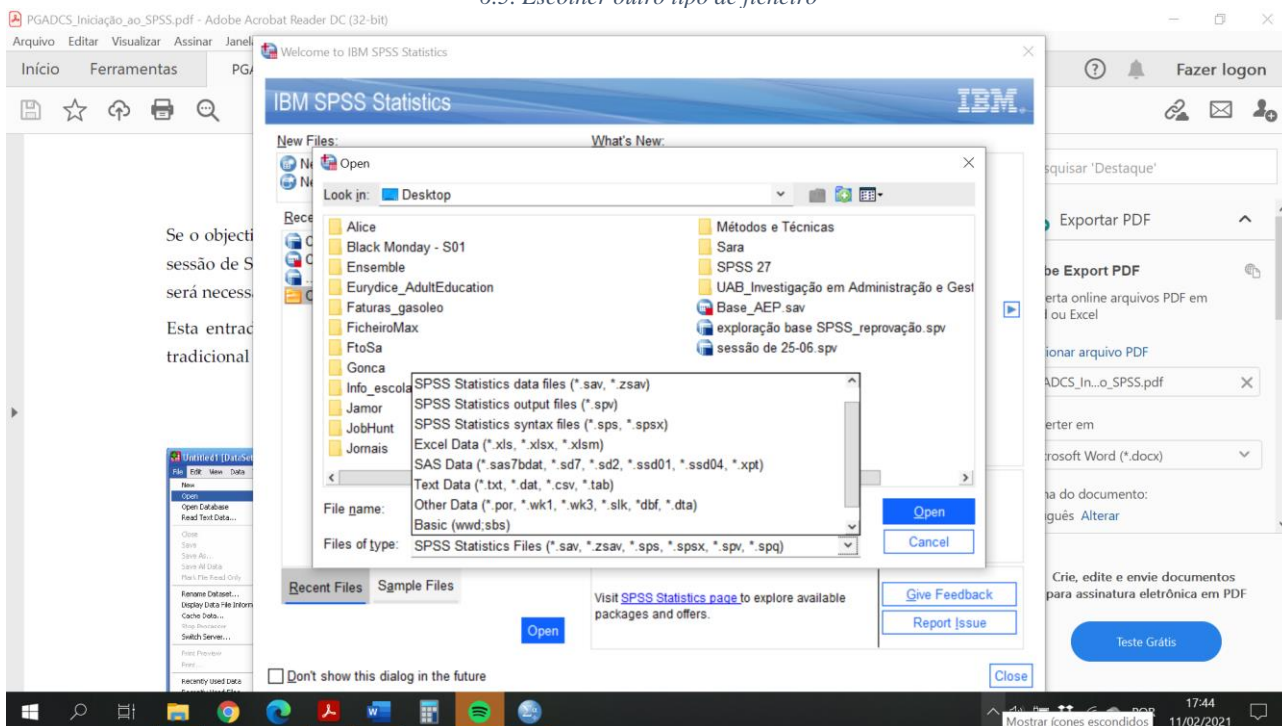
ser escolhida a opção “recent files” ou “open another file”, caso que nos pede para indicar onde está guardada a base de dados).

6.4. início de uma sessão de trabalho



Se o objetivo for abrir um ficheiro que não seja SPSS, então, quando se inicia a sessão, deve escolher-se a opção “Open another type of file”. Também neste caso será necessário dar indicação da localização do ficheiro.

6.5. Escolher outro tipo de ficheiro



6.3.1. Construção de um ficheiro de dados: definição de Variáveis

Como foi referido anteriormente, quando se inicia uma sessão de trabalho no SPSS o Data Editor (janela de Dados) possibilita a construção de raiz de um ficheiro de dados. Essa opção pode também ser escolhida cedendo ao menu “File” New (Data). Surge então uma janela de dados com linhas e colunas em branco. É nessa matriz que serão definidas as variáveis e introduzidos os dados.

As definições de variáveis são realizadas no modo “variable view” (figura 1.2.). Cada linha corresponde a uma variável e as colunas permitem especificar as suas características:

➤ Nome da variável (Variable Name)

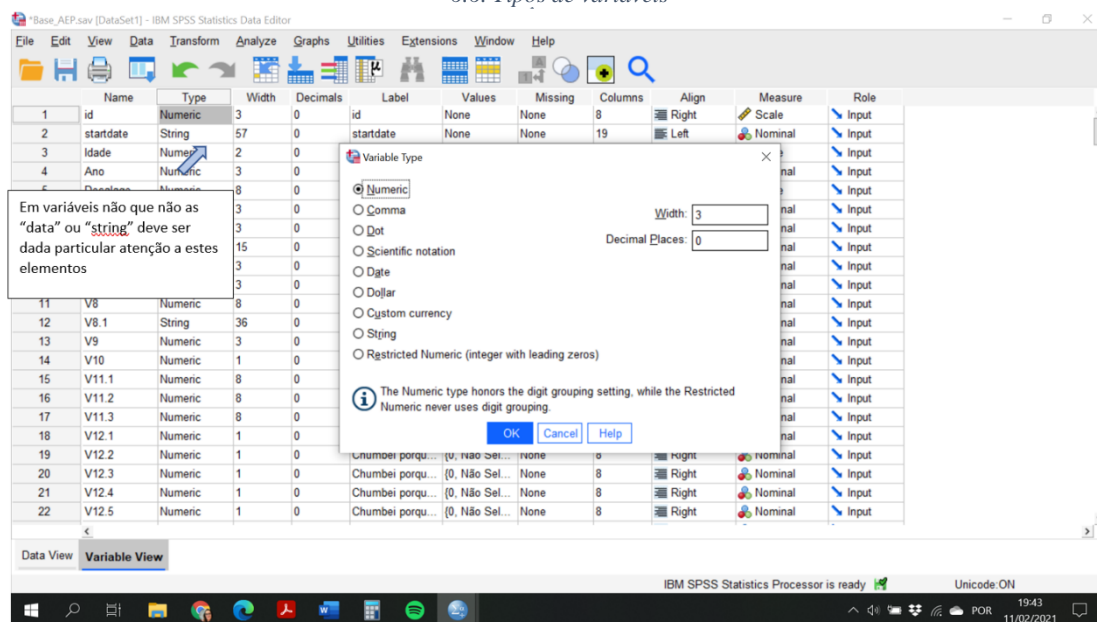
Os nomes das variáveis seguem as seguintes regras:

- Cada nome tem de começar por uma letra. Os restantes caracteres podem ser letras, números, ponto final ou qualquer um dos símbolos @, #, _ or \$;
- O nome de uma variável não pode terminar com um ponto final;
- Deve evitar-se terminar o nome de uma variável com a barra baixa (underscore) pois poderão verificar-se conflitos com nomes de variáveis automaticamente criadas por alguns procedimentos;
- Não podem ser utilizados espaços nem alguns caracteres especiais (por exemplo, !, ?, ', e *);
- Cada nome tem de ser único: não podem coexistir variáveis identificadas pelo mesmo nome. A identificação das variáveis não é sensível à diferenciação maiúsculas/minúsculas: Os nomes VAR1, Var1 e var1 são considerados idênticos;
- Cada nome não pode exceder 64 caracteres. Para atribuir nome a uma variável, basta escrevê-lo na primeira coluna (Name). Em caso de omissão, o SPSS atribui à primeira variável o nome VAR00001, à segunda VAR00002 e assim sucessivamente;

➤ Tipo de variável (Variable Type)

A coluna “variable Type” especifica o tipo de variável. Por defeito, todas as novas variáveis são definidas como numéricas. Posteriormente pode alterar-se o tipo da variável. Para o fazer, e uma vez posicionados na célula correspondente a “type”, basta clicar o botão com as reticências. Aparece, então, a caixa de diálogo que permite seleccionar a definição pretendida para a variável.

6.6. Tipos de variáveis



Numeric – define uma variável cujos valores são números. Os valores são mostrados no formato definido. O Data Editor aceita valores numéricos no formato standard ou em notação científica;

Comma – define uma variável numérica cujos valores são mostrados com vírgulas separando cada três dígitos, com ponto separando a parte inteira da decimal. Os valores podem ser introduzidos com ou sem vírgulas, ou em notação científica;

Dot – define uma variável numérica cujos valores são mostrados com pontos separando cada três dígitos, com vírgula separando a parte inteira da decimal. Os valores podem ser entrados com ou sem pontos ou em notação científica;

Scientific notation – define uma variável numérica cujos valores são mostrados em notação científica, ou seja, com o símbolo E e um expoente potência de 10 (com sinal). O Editor aceita valores numéricos para estas variáveis com ou sem expoente. O expoente pode ser precedido por E ou D, com ou sem sinal, ou apenas pelo sinal. Por exemplo, são representações do mesmo número: 123, 1.23E2, 1.23D2, 1.23E+2, 1.23+2;

Date – define uma variável numérica cujos valores são mostrados em forma de data (dia, mês, ano, ou combinação destes) ou de tempo (horas, minutos, segundos ou combinação destes), de acordo com o formato escolhido de entre a lista fornecida (Figura 7.3). As datas podem ser dadas com barras (/), hifens (-), pontos, vírgulas ou espaços. O intervalo de variação válido para o ano é definido nas opções (no menu Edit, escolher Options e clicar em Data).

Custom currency – variável numérica cujos valores são mostrados num dos formatos de moeda previamente definidos. Os caracteres definidos para a moeda não são incluídos quando se recolhem os dados, mas são mostrados no editor;

String – os valores de uma variável String (alfanumérica) não são numéricos. Consequentemente, não podem ser utilizados em cálculos. Podem conter qualquer cadeia de caracteres até ao comprimento especificado. Neste caso, letras maiúsculas e minúsculas são consideradas diferentes.

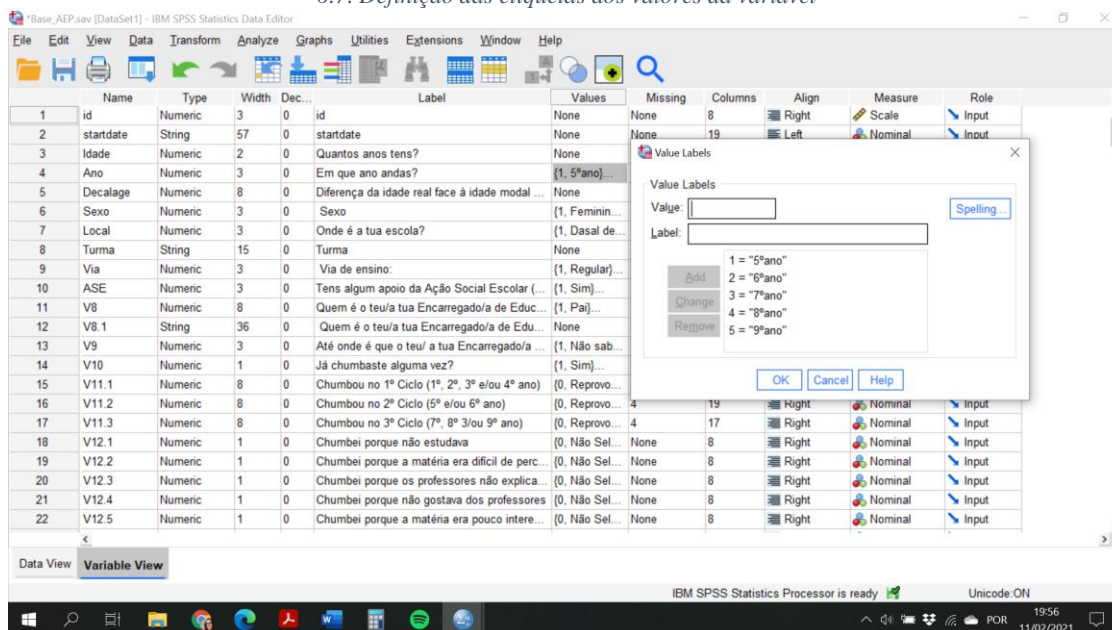
➤ Etiqueta da variável (Variable Label)

Dado que o nome da variável é habitualmente pequeno, é quase sempre conveniente dar-lhe uma etiqueta, a qual poderá ter até 256 caracteres. Essa etiqueta pode aparecer nos outputs, conjuntamente, ou não, com os nomes das variáveis. No entanto, caso a etiqueta seja muito extensa poderá surgir cortada nos quadros de resultados.

➤ Etiquetas dos valores das variáveis (Value Labels)

Aos valores assumidos por uma variável podem ser atribuídas etiquetas específicas, o que é especialmente útil quando se utilizam números para representar categorias não métricas. Por exemplo, se se codificam homens e mulheres com os valores 1 e 2, interessa atribuir a etiqueta “masculino” e “feminino” a cada um dos códigos. Estas etiquetas podem ter até um máximo de 120 caracteres. Não podem ser atribuídas etiquetas a valores de variáveis string (alfanuméricas, definidas com mais de 8 caracteres). Se, para uma dada variável, nos posicionarmos na célula correspondente à coluna “values” e premirmos o botão de reticências, surge a janela para a definição das etiquetas dos valores da variável (value labels). A especificação das etiquetas é feita para cada código de cada vez. A mesma operação é repetida para os restantes códigos e respetivas etiquetas.

6.7. Definição das etiquetas dos valores da variável



➤ Definição dos valores em falta (Missing Values)

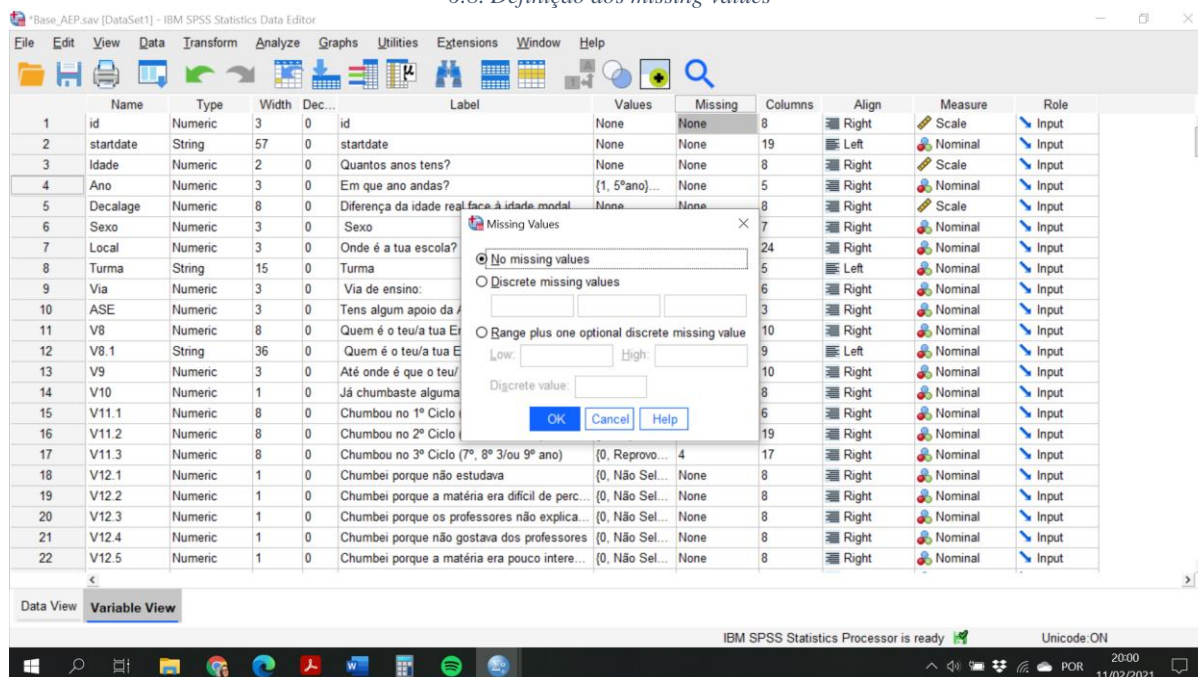
Na coluna “missing Values” podem definir-se valores numéricos específicos que correspondam a respostas não dadas (valores omissos). Esta é uma situação que se justifica sempre que, em vez da ausência de qualquer valor numérico na base de dados (os chamados “system-missing values”, representados por um “.” ou “,”) as não respostas foram codificadas com um, ou mais, códigos numéricos. Estes valores, uma vez definidos pelo utilizador enquanto não-respostas, tomam a designação de “user missing values”.

Note-se que, a recolha exaustiva da informação, incluindo as não respostas, é uma prática aconselhável já que não só permite identificar as razões para a não existência de uma resposta, como ajuda a verificar a integridade da informação recolhida. Poder-se-á pretender distinguir entre respostas omissas porque o respondente não quer responder, porque não sabe responder, ou porque determinada questão não lhe era aplicada. Os valores especificados como omissos podem ser objeto de tratamento especial e excluídos da generalidade dos cálculos.

Para uma dada variável, premindo o botão de reticências, correspondente à coluna “missing” acedemos à janela de definição dos valores em falta (inicialmente a opção que está selecionada é a de “no missing values”

Podem definir-se como omissos um valor composto de três valores individuais distintos, ou um intervalo de valores, ou um intervalo adicionado de mais um valor discreto. Os intervalos de valores omissos apenas podem ser utilizados para variáveis numéricas. Não se podem definir valores omissos para variáveis string com mais de 8 caracteres.

6.8. Definição dos missing values



É recomendável que os códigos utilizados para os valores em falta não sejam confundíveis com os das respostas válidas (por exemplo, podem usar-se valores negativos quando a variável só assume valores positivos, valores elevados quando a variável não assume valores muito altos, ou mesmo o zero sempre que este não faça parte das respostas válidas). No caso das variáveis string (com 8 caracteres ou menos), todos os valores, incluindo cadeias de espaços em branco ou valores não entrados, são considerados válidos, a menos que explicitamente se definam como omissos. Para definir cadeias de espaços, ou valores não entrados, como omissos, é necessário colocar um espaço (um único) num dos campos de “discrete missing values”.

➤ Nível de medida da variável (Variable Measurement Level)

Na coluna “Measure” podemos definir, selecionando a opção desejada, o nível de medida em que as variáveis irão ser trabalhadas: scale (para dados numéricos em escala de rácios ou de intervalos), ordinal, ou nominal. Se a variável estiver definida como sendo do tipo string apenas podem ser escolhidas as opções nominal e ordinal. Note-se que, a especificação do nível de medida não é relevante para a maioria das análises. A sua importância surge apenas nos casos de elaboração de gráficos e tabelas que identificam as variáveis como sendo do tipo escala (scale) ou categoriais (categorical). Neste caso, quer as variáveis definidas como ordinais, quer as definidas como nominais, são tratadas como categoriais. É importante notar que:

SCALE – Variáveis definidas desta forma têm de estar definidas com valores numéricos. Referem-se a variáveis na escala de rácios ou de intervalos, ou tratadas como tal.

ORDINAL – Os valores representam categorias que são ordenáveis (por exemplo, discordo totalmente, discordo, concordo, concordo totalmente). Variáveis definidas como ordinais podem conter quer valores numéricos, quer alfanuméricos. Quando se utilizam valores numéricos, assume-se que cada valor representa uma categoria distinta (por exemplo, 1 = discordo, 2 = concordo, 3 = concordo totalmente). Para variáveis definidas como string, a sua catalogação como ordinais assume que a ordem entre categorias se espelha na ordem alfabética dos valores utilizados. Como princípio geral, é mais fiável utilizar, no caso de variáveis a definir como ordinais, variáveis de tipo numérico, etiquetando convenientemente os valores.

NOMINAL – Os valores representam categorias sem ordem intrínseca – por exemplo curso frequentado, género, escola frequentada. As variáveis nominais podem ser de tipo alfanumérico (string) ou de tipo numérico. Neste último caso, valores diferentes representam distintas categorias (e.g. 1 = Masculino, 2 = Feminino).

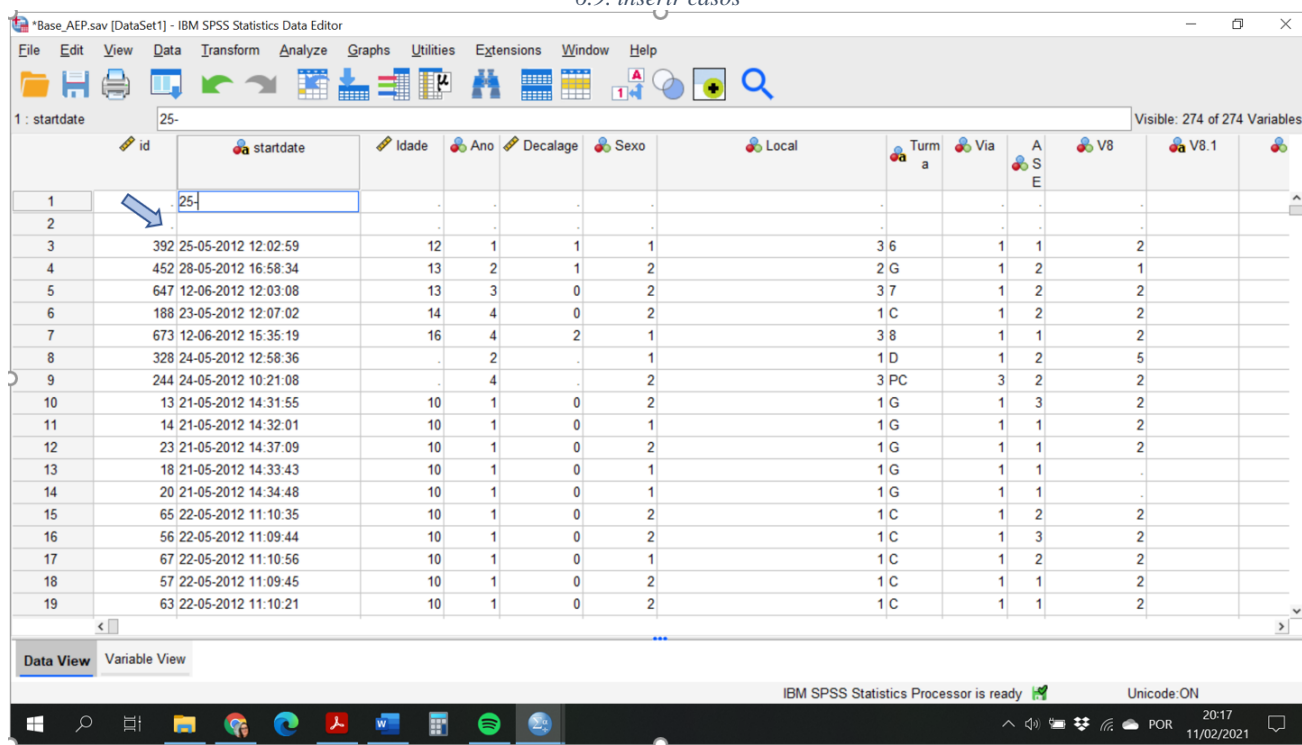
6.4. Introdução e edição de dados

Para introduzir os dados é necessário selecionar “data view” no canto inferior esquerdo da janela de dados (data Editor). Os dados são sempre introduzidos na célula ativa que, quando ativada, fica com um rebordo mais forte.

No canto superior esquerdo da janela “data editor” pode visualizar-se o nome da variável e o número de linha da célula ativa. O valor da célula ativa aparece no campo editor de célula, acima do data editor. Os valores introduzidos só são registados quando se pressiona “enter”, ou quando se seleciona outra célula. Para introduzir dados não numéricos, a variável deverá ter sido previamente definida como variável alfanumérica (string).

É sempre possível inserir novos casos ou novas variáveis.

6.9. inserir casos



	id	startdate	Idade	Ano	Decalage	Sexo	Local	Turma	Via	A S E	V8	V8.1
1	25	25-
2
3	392	25-05-2012 12:02:59	12	1	1	1	.	3 6	1	1	2	.
4	452	28-05-2012 16:58:34	13	2	1	2	.	2 G	1	2	1	.
5	647	12-06-2012 12:03:08	13	3	0	2	.	3 7	1	2	2	.
6	188	23-05-2012 12:07:02	14	4	0	2	.	1 C	1	2	2	.
7	673	12-06-2012 15:35:19	16	4	2	1	.	3 8	1	1	2	.
8	328	24-05-2012 12:58:36	.	2	.	1	.	1 D	1	2	5	.
9	244	24-05-2012 10:21:08	.	4	.	2	.	3 PC	3	2	2	.
10	13	21-05-2012 14:31:55	10	1	0	2	.	1 G	1	3	2	.
11	14	21-05-2012 14:32:01	10	1	0	1	.	1 G	1	1	2	.
12	23	21-05-2012 14:37:09	10	1	0	2	.	1 G	1	1	2	.
13	18	21-05-2012 14:33:43	10	1	0	1	.	1 G	1	1	.	.
14	20	21-05-2012 14:34:48	10	1	0	1	.	1 G	1	1	.	.
15	65	22-05-2012 11:10:35	10	1	0	2	.	1 C	1	2	2	.
16	56	22-05-2012 11:09:44	10	1	0	2	.	1 C	1	3	2	.
17	67	22-05-2012 11:10:56	10	1	0	1	.	1 C	1	2	2	.
18	57	22-05-2012 11:09:45	10	1	0	2	.	1 C	1	1	2	.
19	63	22-05-2012 11:10:21	10	1	0	2	.	1 C	1	1	2	.

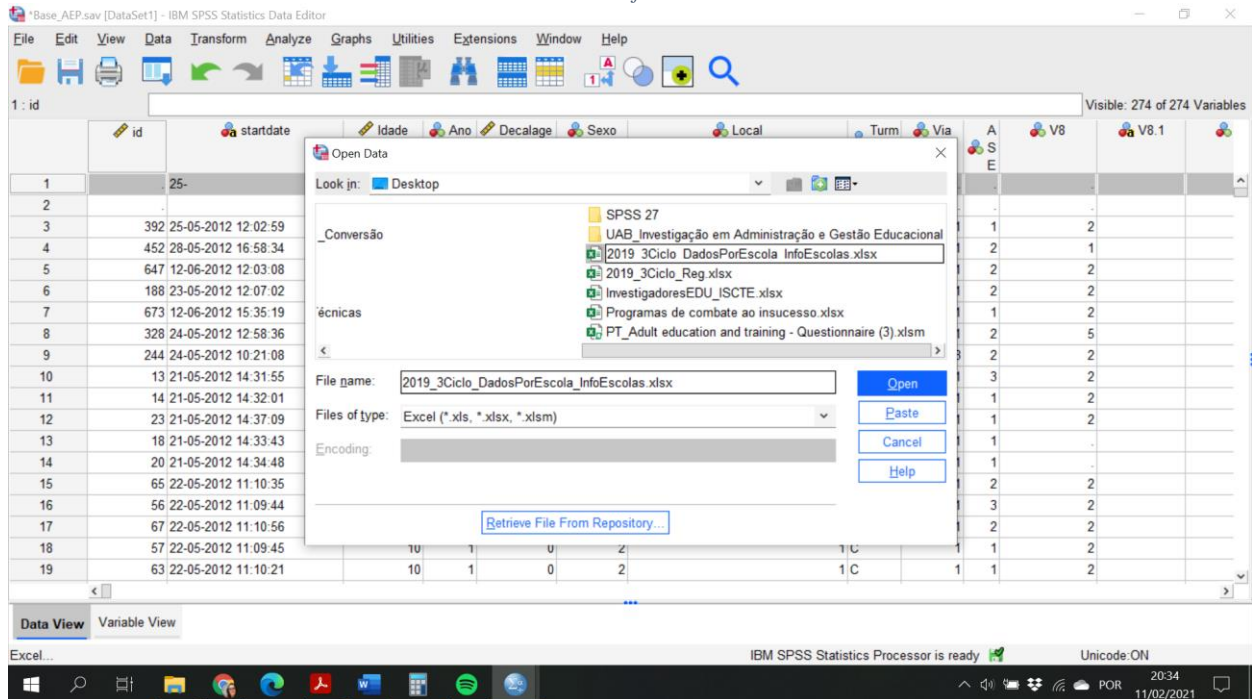
6.5. Importação de bases de dados

A importação de bases de dados no SPSS pode ser feita a partir do menu: File/open/data.

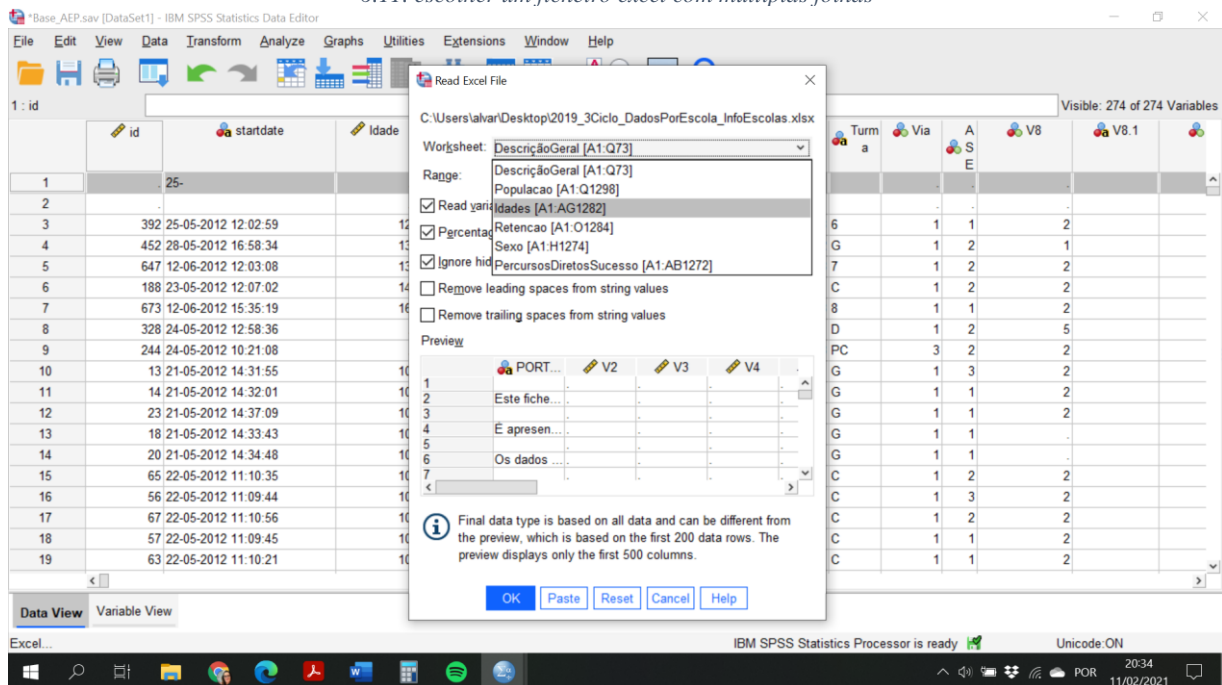
Para importar uma base de dados em excel com várias folhas (como a do Infoescolas), uma vez escolhida a sequência, o processo é idêntico à abertura de um ficheiro SPSS, indicando-se a base de dados que se pretende importar e especificando o tipo de ficheiro de origem (neste caso, Excel). Esta sequência de procedimentos tem de ser

repetida para cada uma das folhas do ficheiro de Excel que se pretende trabalhar em SPSS, o que significa que, no final, se irá dispor de tantos ficheiros SPSS (*.sav) quantas as folhas excel que se deseja importar.

6.10. Escolher ficheiros excel



6.11. escolher um ficheiro excel com múltiplas folhas

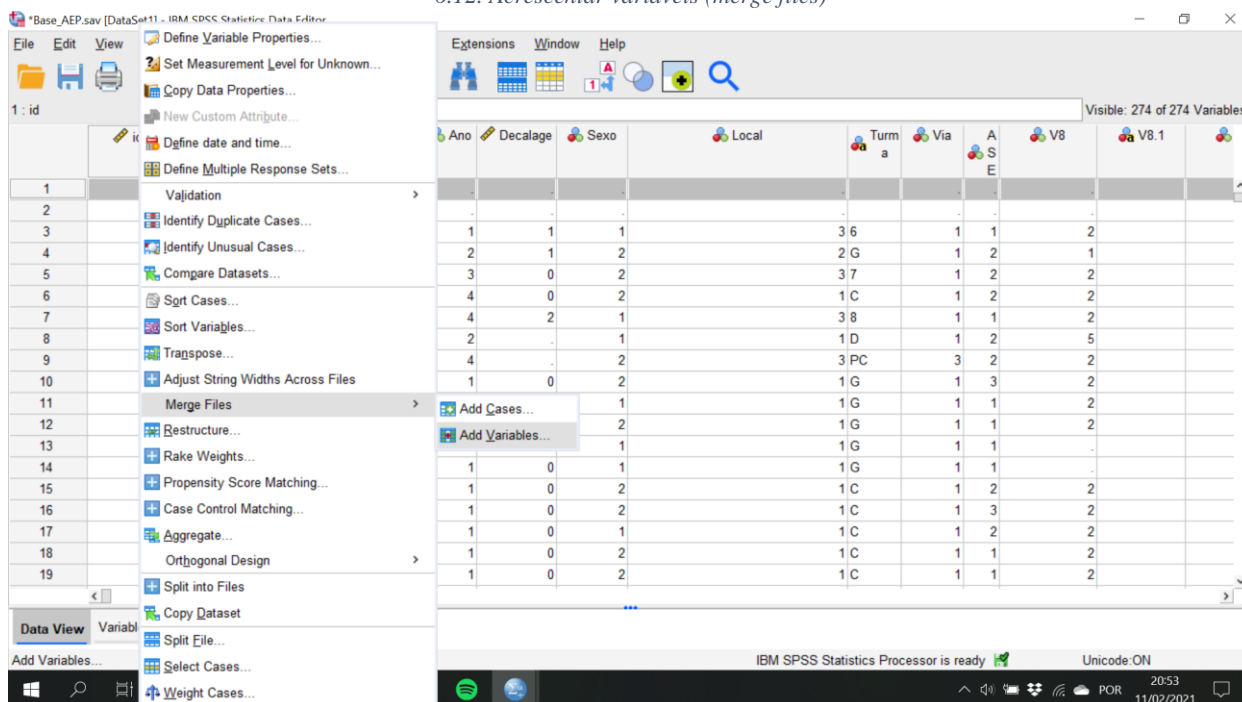


Se o ficheiro excel que queremos trabalhar em SPSS tiver múltiplas folhas, criamos vários ficheiros SPSS e podemos desejar juntá-los numa única base de dados. Também pode ocorrer a necessidade de juntar, não novas variáveis, mas casos.

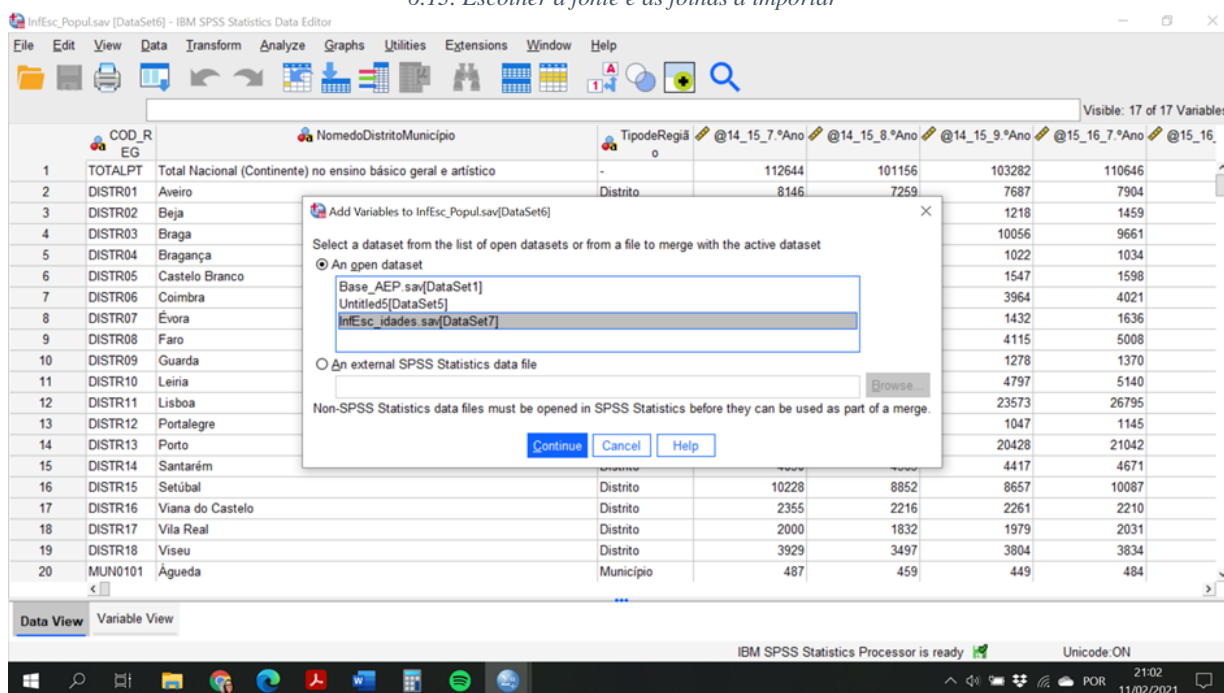
➤ Junção de ficheiros de dados: acrescentar variáveis (ADD VARIABLES)

No caso de se pretender juntar duas bases que possuem os mesmos casos, mas variáveis distintas, é fundamental existir uma variável comum que os identifique e que possua exatamente o mesmo nome, escrito da mesma forma na coluna “name”. Verificada essa condição é possível juntar as variáveis (todas ou parte delas) num só ficheiro. Para realizar a operação, depois de abrir uma das bases de dados que vai integrar o ficheiro final, selecionar no Menu Data a opção “merge files” e, dentro deste “add variables”.

6.12. Acrescentar variáveis (merge files)

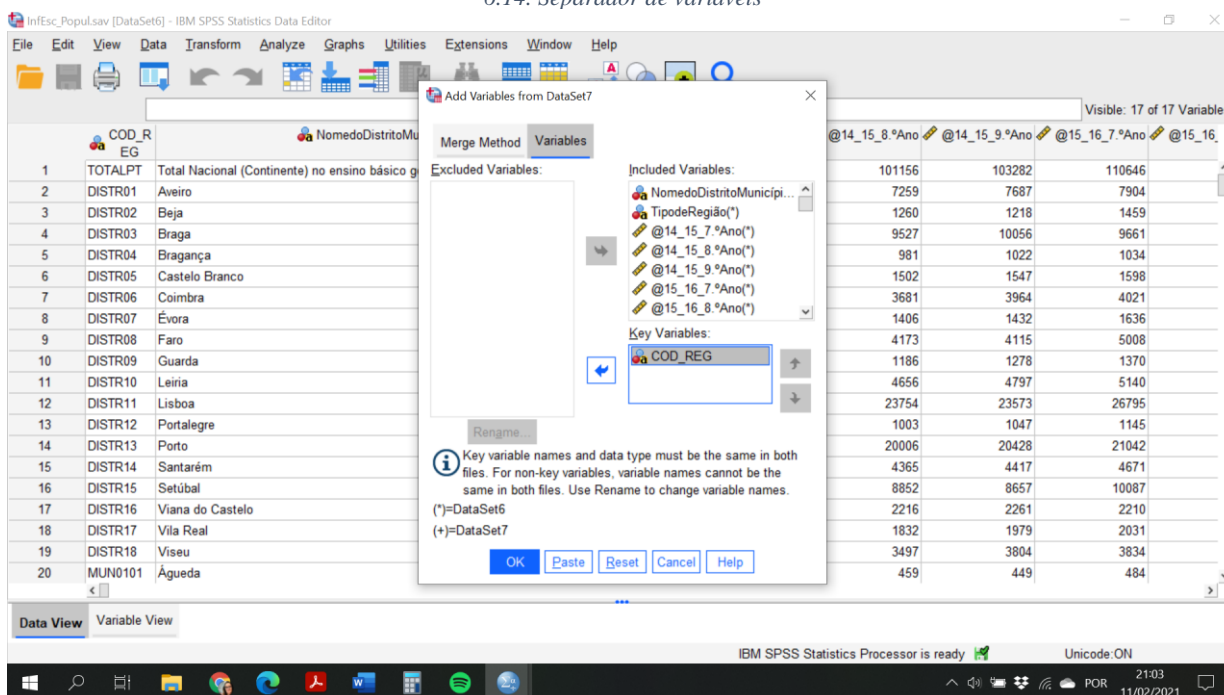


6.13. Escolher a fonte e as folhas a importar

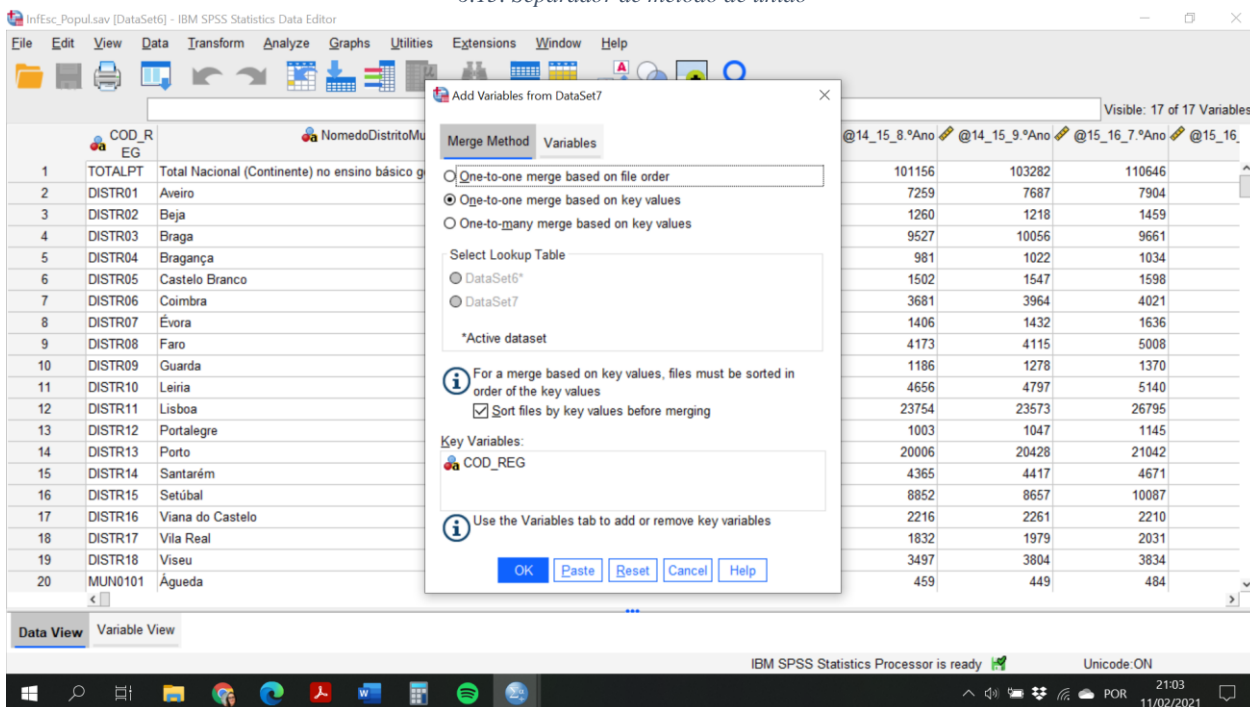


Após a especificação do ficheiro de dados, surge a janela que completa a junção dos dois ficheiros. No separador “variables” são listadas todas as variáveis, tanto as do ficheiro de partida (ativo), como as que se pretende acrescentar. As variáveis que não se pretende reter podem ser seleccionadas e incluídas no quadro “excluded variables”. Nesse mesmo quadro estão listadas, por defeito, todas as variáveis que sejam comuns a ambos os ficheiros de dados, por exemplo as variáveis de identificação.

6.14. Separador de variáveis



6.15. Separador de método de união



Na junção de bases de dados é recomendável a existência de uma variável-chave que identifique os casos e assegure a correta colagem dos ficheiros e que deve ser identificada como **Key Variable**. É também necessário que ambos os ficheiros de dados estejam ordenados por ordem ascendente da variável-chave (clique na coluna da variável, selecionando-a toda e, com o botão direito do rato seleccionar “sort ascending”). No caso de se pretender juntar mais do que duas bases de dados, o procedimento que acabou de ser descrito deverá ser repetido até que todas as variáveis estejam no mesmo ficheiro. A base deve então ser gravada, de preferência com outro nome.

6.6. Breve apresentação à análise de dados no SPSS

A exploração de dados baseia-se no teste de hipóteses concebidas pelo investigador. Esta testagem resulta da aplicação de testes estatísticos inferenciais, ou seja, que concluem sobre a probabilidade de se poder (ou não) extrapolar (inferir) os resultados obtidos da amostra para a população de onde esta foi extraída. A possibilidade de inferir (validade experimental) é condicionada pela existência de validade interna (garantia de que os dados – amostra, forma das perguntas e conceptualização das variáveis – não interferem no resultado, ou seja, que se está a medir o que se pretende medir) e validade externa (que existe uma fidedigna representação da realidade). O SPSS permite analisar:

- Padrões e perfis de distribuição de variáveis (análises univariadas)
- Relação entre variáveis (análises bivariadas)
- Topologias, tipologias, modelos de dependência, entre outros (análises multivariadas)

Os testes realizados pelo SPSS podem ser paramétricos ou não paramétricos. Os testes paramétricos – também designados de teste t – exigem variáveis quantitativas e que a amostra tenha uma distribuição normal, ou seja, que se conheçam os parâmetros da população, que estes sejam seguidos na amostra e que esta assuma uma curva em forma de sino. Os testes não paramétricos podem ser de aderência ou ajustamento – como de Kolmogorov-Smirnov ou do Qui-Quadrado - quando se pretende testar se uma amostra foi retirada de uma população com uma determinada distribuição ou testes de independência – como o teste de Mann Whitney, Kruskal-Wallis – que pretende testar se duas variáveis são independentes ou se existe relação entre elas. Alguns dos testes disponíveis:

Tipo de teste/ variáveis	Nominais	Ordinais	Quantitativas
Independência de 2 grupos		Teste Mann-Whitney	Teste T
Independência de K grupos		Kruskal-Wallis	One-Way Anova
Comparação de pares de variáveis	Teste de independência de Qui-Quadrado	Komolgorov-Smirnoff	
Validação da relação amostras emparelhadas		WILCOXON	Teste T para amostras emparelhadas
Medidas de associação / correlação	V de Cramer ETA	RHO de Spearman TAU-B de Kendall	ETA R de Pearson

(amostras emparelhadas: os mesmos indivíduos – por exemplo, inquiridos em dois momentos distintos - ou a assunção lógica da relação entre duas respostas)

Existem ainda no SPSS um conjunto de funcionalidades que permitem apresentar e relacionar variáveis de natureza distinta, a partir de tabelas de contingência (crosstabs).

Representação gráfica

Estão ainda disponíveis no SPSS um conjunto de funcionalidades que permitem apresentar e relacionar variáveis a partir da sua representação gráfica. Cada forma de representação deve respeitar o tipo de variáveis em consideração

Tipo de variáveis	Tipo de representação Gráfica
nominais, ordinais e quantitativas	Tabelas de frequência
nominais e ordinais	Gráficos de barras e circulares
ordinais e quantitativas	Caixa de bigodes
quantitativas	Gráficos de dispersão; gráficos de médias e desvio-padrões; histograma; Scatterplot” de X e Y
catóricas (qualitativas)	Gráfico de colunas por estratos da segunda variável
quantitativa e qualitativa	Box-Plot”, para cada estrato ou categoria da variável qualitativa após categorizar a quantitativa



[Estatística com o SPSS](#)

BIBLIOGRAFIA

Carvalho, Helena (2017) *Análise Multivariada de Dados Qualitativos: Utilização de Análise de Correspondências Múltiplas com o SPSS*. Sílabo. 12ª Edição

Coutinho, Clara Pereira (2018) *Metodologia de Investigação em Ciências Sociais e Humanas - Teoria e Prática*. Almedina

Hill, Manuela Magalhães e Andrew Hill (2008) *Investigação por questionário*. Edições Sílabo. 2ª Edição

Laureano, Raul e Maria do Carmo Botelho (2017) *SPSS Statistics: O Meu Manual de Consulta Rápida*. Sílabo. 2ª Edição

Maroco, João (2003). *Análise Estatística com utilização do SPSS*. Sílabo. 2ª Edição

Pereira, Alexandre e Teresa Patrício (2013). *SPSS: guia prático de utilização: análise de dados para Ciências Sociais e Psicologia*. Sílabo. 8ª Edição