

INVESTIGAÇÃO EMPÍRICA EM CIÊNCIAS SOCIAIS:

Um Guia Introdutório

Manuela Magalhães Hill
Andrew Hill

Outubro 1998

WP n.º 1998/10

DOCUMENTO DE TRABALHO

WORKING PAPER





INVESTIGAÇÃO EMPÍRICA EM CIÊNCIAS SOCIAIS: Um Guia Introdutório

Manuela Magalhães Hill & Andrew Hill

WP1998/10
Outubro 1998

Índice

PREFÁCIO.....	1
1. A NATUREZA DA INVESTIGAÇÃO EMPÍRICA.....	2
2. A ESCOLHA DE UM TEMA: PRINCÍPIOS GERAIS	4
3. PASSOS PARA ESCOLHER UM TEMA.....	5
4. A REVISÃO DA LITERATURA.....	6
5. O PLANEAMENTO DA INVESTIGAÇÃO EMPÍRICA.....	9
5.1. AS RELAÇÕES ENTRE AS HIPÓTESES, OS MÉTODOS DE INVESTIGAÇÃO E OS MÉTODOS PARA ANALISAR DADOS.....	9
5.2. A RECOLHA DOS DADOS	15
5.2.1. A fiabilidade das medidas feitas por questionário.....	16
5.2.2. A validade das medidas feitas por questionário.....	21
5.3. AMOSTRAGEM.....	27
5.3.1. Definições: o universo de casos e a amostra de casos	27
5.3.2. A escolha de uma amostra.....	27
6. ALGUNS PROBLEMAS COM A ANÁLISE DE DADOS	32
6.1. O AGRUPAMENTO DE CASOS	32
6.2. PROBLEMAS COM AS CORRELAÇÕES PARAMÉTRICAS (TIPO PEARSON 'R')	35
6.2.1. O problema da não-linearidade	35
6.2.2. O significado estatístico e o significado prático do coeficiente de correlação	35
6.3. PROBLEMAS RELACIONADOS COM O QUI-QUADRADO (χ^2).....	36
6.3.1. Que Qui-quadrado utilizar?	36
6.3.2. O problema das frequências esperadas pequenas.....	37
6.3.3. Qui-quadrado e percentagens.....	37
6.3.4. Um Qui-quadrado fora do usual.....	38
6.4. PROBLEMAS RELACIONADAS COM O TESTE 'T'	38
6.4.1. O Teste 't' para duas amostras independentes: o problema das variâncias desiguais	39
6.4.2. O Teste 't' para duas amostras emparelhadas.....	39
6.4.3. O Teste 't' do "Tolo"	40
6.4.4. Teste 't' e comparações múltiplas	41
6.5. ANÁLISE DE COMPONENTES PRINCIPAIS VERSUS ANÁLISE FACTORIAL.....	43
BIBLIOGRAFIA.....	45
ANEXO.....	46

PREFÁCIO

O presente *working paper* destina-se essencialmente a jovens investigadores que ainda não tenham grande experiência no planeamento de trabalhos empíricos. No entanto, o *working paper* pode também ser útil para alunos de ciências sociais (licenciatura ou mestrado) que pretendam, ou necessitem de, fazer investigação empírica.

A ideia de escrever o presente *working paper* nasceu da experiência dos autores na orientação de teses de mestrado e doutoramento, em particular, na área dos Recursos Humanos. Por este motivo, os exemplos apresentados ao longo deste *working paper* vêm principalmente do contexto dos Recursos Humanos, embora os princípios da investigação empírica indicados possam ser generalizados a qualquer outra área das ciências sociais.

São dois os objectivos deste *working paper*:

- Apresentar duma maneira simples e acessível, os pontos essenciais de uma investigação empírica;
- Referir alguns dos problemas que vulgarmente surgem numa tal investigação e dar conselho prático para os resolver.

Um guia introdutório é exactamente isso -- **introdutório**. Num tal guia não é possível explicar tudo, nem explicar com detalhe exaustivo os assuntos aí tratados. Daí que, de vez em quando, oferecemos referências à literatura especializada de modo que o leitor possa aprofundar o seu conhecimento.

São muitas as investigações em ciências sociais que utilizam questionários para recolher os dados. Embora este *working paper* aborde brevemente os problemas da fiabilidade e da validade de medidas feitas por questionário, não apresenta conselhos sobre a composição de um questionário. Este assunto é tratado separadamente num outro *working paper*, complementar a este, intitulado “**A construção de um questionário**”.

1. A NATUREZA DA INVESTIGAÇÃO EMPÍRICA

> Tipos de investigação empírica

É muito vulgar considerar que há três tipos distintos de investigação empírica em ciências sociais.

- *Investigação pura:*

O objectivo deste tipo de investigação é descobrir factos novos (dados empíricos) para testar deduções feitas de uma teoria que só tem interesse intelectual e que, na altura da investigação, parece não ter aplicação prática. O objectivo deste tipo de investigação é, essencialmente, enriquecer a literatura numa área teórica determinada. Por exemplo, nos anos sessenta, desenvolveram-se muitas investigações puras em psicologia social sobre a teoria de atribuições.

- *Investigação aplicada:*

O objectivo deste tipo de investigação é descobrir factos novos (dados empíricos) para testar deduções (hipóteses) feitas de uma teoria que pode, em princípio, ter aplicações práticas no médio prazo. Por exemplo, nos anos setenta e oitenta, houve muitas investigações que aplicaram aspectos da teoria de atribuição em busca da causa de um tipo de depressão.

- *Investigação aplicável:*

O objectivo deste tipo de investigação é descobrir factos novos (dados empíricos) que sejam capazes de resolver problemas práticos no curto prazo.

Curiosamente, os três tipos de investigação não são tão diferentes como parecem. *O que têm em comum é que todos os tipos de investigação empírica são baseados numa revisão da literatura relevante.*

É óbvio que, numa investigação pura, o tema e as hipóteses de investigação são retirados de uma revisão da literatura. É menos óbvio, mas igualmente certo, que uma investigação aplicada se baseia numa revisão da literatura (como no exemplo anterior). Neste caso, o tema talvez não seja retirado da literatura, mas as hipóteses são deduzidas das teorias aí apresentadas e, muitas vezes, os métodos de investigação vêm da literatura. Em relação à investigação aplicável, o tema não depende na literatura -- foi fornecido na forma de um problema prático que precisa de ser resolvido. Acontece porém que raramente é possível resolver problemas práticos através de uma investigação empírica sem ajuda de uma revisão da literatura relevante. Por exemplo, suponha que a Administração de uma empresa encontra um nível muito elevado de rotação voluntária nos seus trabalhadores. Este é um problema prático porque um nível alto de rotação voluntária faz baixar a produtividade. É possível efectuar uma investigação para resolver o problema, mas o que é que a investigação deve investigar? A literatura na área da sociologia aplicada indica que a rotação voluntária está provavelmente relacionada com vários tipos de atitude, e vários tipos de insatisfação, dos empregados. Portanto, parece útil investigar estes tipos de atitude e insatisfação efectuando entrevistas ou aplicando um questionário.

As duas primeiras secções deste *working paper* tratam mais de investigação pura no sentido que tratam do processo de investigação e do problema de encontrar um tema para investigar. Portanto, estas secções são especialmente relevantes para a investigação académica, em particular para os alunos que têm de fazer um projecto de fim de licenciatura ou para aqueles que tenham de elaborar uma tese de mestrado ou doutoramento.

➤ O processo de investigação empírica

O processo de investigação não é só um processo de aplicação de conhecimentos mas também um processo de planificação e criatividade controlada. Para introduzir a natureza essencial de uma investigação empírica considere-se a seguinte analogia.

Imagine que parte de Lisboa para passar uns dias no Alentejo. A viagem tem um **objectivo** – ir em trabalho ou ir apenas de férias. Para tal, vai ter de fazer **escolhas** – ir de carro, de comboio ou de autocarro. Vai ter de **planear** a viagem -- dia e hora de partida. Vai ter de **pensar adiante** -- que roupas adequadas levar? -- onde vai ficar?

A investigação empírica também pode ser considerada uma viagem. Tal como a viagem ao Alentejo, a “viagem da investigação” tem um **objectivo** e obriga-o a fazer **escolhas**. A viagem da investigação também precisa de **planeamento** e o investigador tem de **pensar adiante**. Tal como a viagem ao Alentejo, a viagem da investigação é uma **viagem de ida e volta**. Começa no “**País Teórico**” chega ao “**País Prático**” e termina mais uma vez no “**País Teórico**”.

Em resumo, a “viagem da investigação” empírica compreende os seguintes aspectos:

- Tem como **objectivo** contribuir para o enriquecimento do conhecimento na área em que escolheu fazer a investigação.
- Precisa de **escolhas** em termos do tema e em termos das hipóteses específicas a testar.
- Obriga-o a **planear** os métodos de recolha de dados.
- Precisa que **pense adiante** para planear as análises de dados antes de começar a parte empírica da investigação.

Vejamos ainda o seguinte: para ir de Lisboa ao Alentejo precisa de usar uma ponte para atravessar o rio Tejo e tem varias opções -- ponte 25 Abril, ponte Vasco da Gama ou ponte de Vila Franca de Xira. Na sua “viagem da investigação”, para passar do “País Teórico” para o “País Prático” também precisa de usar uma ponte porque há um abismo grande entre os dois países (o vale dos investigadores perdidos!). Neste caso, a ponte é a **hipótese da investigação**. Infelizmente esta ponte não existe pelo que tem de ser construída por si -- mas tem várias maneiras de a construir.

O “País Teórico” é a parte teórica da sua investigação e consiste principalmente na revisão de literatura sobre o tema que escolheu. O “País Prático” é a parte empírica da sua investigação e consiste na planeamento e execução do trabalho empírico. Ao efectuar a revisão de literatura vai encontrar teorias, relatórios e artigos sobre investigações empíricas apresentados por diversos autores. A primeira tarefa será utilizar essa literatura para deduzir um hipótese, ou um conjunto de hipóteses, que vai testar na parte empírica. A **hipótese** é uma parte muito importante do processo de investigação porque, como já dissemos, faz uma ponte entre a parte teórica e a parte empírica da investigação. Sendo uma ponte, a hipótese tem um papel fundamental:

A hipótese deve justificar o trabalho da parte empírica da sua investigação.

Para compreender a importância deste ponto suponha que já terminou o trabalho e escreveu o seu relatório ou tese. Quando alguém (por exemplo, o arguente, no caso de ser uma tese) for ler a parte empírica do relatório, ele/ela tem de ficar convencido de que o trabalho **faz sentido e está justificado pela revisão da literatura** apresentada na parte teórica do relatório. Quem for ler a tese, só vai entender o porquê da parte empírica se a hipótese estiver claramente ligada com a revisão da literatura e claramente ligada com o trabalho empírico. Estas ligações são o papel da hipótese, daí a hipótese ter sido associada a uma ponte, na analogia.

Para encontrar uma hipótese claramente ligada com a revisão da literatura é essencial que a revisão da literatura seja **estritamente relevante** para o tema da investigação. Mas, obviamente, antes de fazer uma tal revisão é preciso escolher um tema.

2. A ESCOLHA DE UM TEMA: PRINCÍPIOS GERAIS

Esta secção refere-se a investigação empírica a desenvolver por alunos de mestrado ou doutoramento, e ainda alunos num ano terminal de licenciatura que tenham de elaborar o projecto final.

Vamos usar a palavra “**tema**” para representar um assunto específico, um problema ou uma “linha de investigação”, para os quais existe literatura desenvolvida e sobre os quais há teoria(s) e trabalhos empíricos publicados. Não usamos a palavra “tema” para representar a hipótese da sua investigação. A hipótese virá da sua revisão desta literatura.

A escolha do tema pressupõe três princípios fundamentais:.

➤ O tema deve cumprir as regras da licenciatura ou do mestrado

Isto implica que:

- o aluno deve demonstrar que beneficiou dos conhecimentos adquiridos durante a parte escolar do curso (domínio da literatura e da “cultura” na área que desenvolveu, bem como, conhecimento das técnicas e realidades na área).
- No caso de um Mestrando, deve ter capacidade para desenvolver investigação ao nível de pós-graduação (desenvolver investigação ao nível de pós-graduação é mais exigente que ao nível de licenciatura).

➤ O tema deve ser adequado em escala

Isto significa que o tema deve ser **realístico e maneável**. É muito natural desenvolver um projecto ou uma tese de mestrado demasiado ambiciosos e complicados. Para fazer um projecto ou uma tese de mestrado adequados em termos de escala numa dada área é muito importante não esquecer que tem de enfrentar vários **constrangimentos práticos**, tais como:

- Tempo (o trabalho de investigação precisa **sempre** de mais tempo do que o previsto)
- Acesso à literatura relevante
- Acesso aos dados
- seu conhecimento prévio sobre o tema

Um trabalho de investigação ideal é aquele que apresenta uma investigação relativamente curta em escala, mas com hipóteses simples, claras e **interessantes**.

➤ Deve escolher um tema sobre o qual tem particular interesse

Vai gastar muito tempo e muita energia com o trabalho de investigação que vai desenvolver. É por isso natural que, no curso da investigação, haja períodos em que tudo ande mal - vai ter frustrações e encontrar problemas. Se não tiver muito interesse no tema que escolheu, durante estes períodos difíceis, a sua motivação pode baixar ao ponto de pôr em causa a qualidade do seu trabalho.

3. PASSOS PARA ESCOLHER UM TEMA

Para escolher um tema é útil começar por escolher uma área de interesse antes de passar para o tema propriamente dito. Assim, sugerem-se os seguintes passos:

PASSO 1. Objectivo: Encontrar uma área geral que o interesse
Método: Ler em diagonal alguns livros gerais

Se já decidiu qual a área geral sobre a qual está interessado em desenvolver a sua investigação, é boa ideia começar mesmo assim por ler em diagonal alguns livros e artigos gerais porque é possível que encontre uma área que o venha a interessar mais.

Se ainda não encontrou uma área interessante para a sua investigação, e não tem nenhuma ideia sobre o que é que gostaria de investigar, os livros gerais são úteis porque, normalmente, apresentam teorias e resumos de resultados de investigações empíricas. Neste passo só é importante encontrar uma área geral e não vale a pena entrar nos pormenores dos estudos empíricos que aparecem na literatura.

Encontrada uma área geral que o interesse, passe ao passo seguinte:

PASSO 2. Objectivo: Encontrar um tema dentro da área geral escolhida
Método: Começar a ler artigos nas revistas académicas da especialidade

Paradoxalmente, na busca de um tema, é muitas vezes mais útil usar uma estratégia “do fim para o princípio”, ou seja, começar por ler (em diagonal) os artigos mais recentes na área geral escolhida. Estes artigos podem dar informação sobre os tópicos actualmente considerados importantes e interessantes. Provavelmente não vai compreender tudo nestes artigos, mas os artigos que aparecem nas revistas académicas contêm **referências** sobre as teorias relevantes, livros especializados e outros trabalhos empíricos constantes da literatura, que ajudam à compreensão.

Convém lembrar que a investigação em Ciências Sociais, pode ser ligada com várias disciplinas; por exemplo, a Economia, a Sociologia, a Psicologia, e até a Medicina, pelo que pode ser útil consultar também os Anuários e os “Abstracts” nestas disciplinas. Por exemplo:

Annual Review of Sociology
Annual Review of Psychology

Sociological Abstracts
Psychological Abstracts
Index Medicus

Paralelamente pode usar várias fontes de informação disponíveis em CD-ROM tais como a *Sociofile*. (E sempre há as consultas à “Internet”)

Como resultado do Passo 2 deve ficar com um conjunto de informações, preferivelmente na forma de fotocópias de artigos seleccionados. Esta informação constituirá a literatura relevante para o seu tema. A partir daqui passará para uma nova fase da sua investigação - fazer uma revisão desta literatura em busca de um hipótese (ou hipóteses) para investigar na parte empírica do seu trabalho.

CONSELHO PRÁTICO

É mais fácil e adequado trabalhar primeiro a literatura sobre o tema, e deduzir a partir daí as hipóteses da investigação, do que partir de hipóteses definidas a priori, e ir à procura de literatura que as suporte.

4. A REVISÃO DA LITERATURA

OBJECTIVO: Encontrar uma (ou mais) **Hipóteses Gerais**.

Para já vamos usar a palavra simples **Hipótese** para designar “Hipótese Geral”, mas mais tarde vamos fazer uma distinção entre **Hipóteses Gerais** e **Hipóteses Operacionais**.

MÉTODO: A revisão da literatura envolve três partes:

- **Descrição** das teorias e trabalhos empíricos
- **Avaliação** das teorias e trabalhos empíricos
- **Comparação** das teorias e trabalhos empíricos

As Hipóteses representam conclusões tiradas a partir da **avaliação** e **comparação** dos trabalhos apresentados na literatura, e, portanto, estas duas partes da revisão são muito importantes.

➔ A descrição dos trabalhos empíricos

É conveniente fazer anotações e pequenos resumos de cada um dos trabalhos empíricos que seleccionou. Essas anotações deverão incluir, para além dos resultados e conclusões:

- A situação em que o trabalho foi feito;
(País, local, sector de actividade, tipo de empresa, instituição, organização, etc.)
- A natureza dos casos – famílias, grupos sociais, empresas, sectores de actividade, trabalhadores, etc.;
(Se forem empresas -- dimensão, número de trabalhadores, regime jurídico, etc..)
(Se forem trabalhadores – idade, sexo, profissão, natureza do emprego, etc.)
- A amostra de casos;
(Dimensão da amostra, natureza da amostra - por exemplo, amostra por conveniência, amostra aleatória simples, amostra estratificada – representatividade)
- As hipóteses do trabalho;

- Os métodos usados no trabalho;
(entrevistas, questionários, estudos de casos, etc.)
(Se questionários – a fiabilidade e a validade (se disponíveis))
- Os métodos utilizados na análise dos dados.

➔ A avaliação dos trabalhos empíricos

Efectuada uma descrição conveniente dos textos recolhidos durante a pesquisa bibliográfica, torna-se importante proceder em seguida a uma avaliação sobre os últimos quatro pontos acima referidos (amostra, hipóteses de trabalho e métodos utilizados). A avaliação não é difícil, só precisa de tentar responder às seguintes perguntas:

- A amostra do estudo era adequada? Era representativa do Universo?
(Se não, os resultados só têm valores limitados)
- As hipóteses eram claras e estavam justificadas pela teoria ou pela revisão da literatura?
(Se não, os resultados podem não ser significativos e as conclusões não ser válidas)
- Os métodos usados no trabalho eram adequados?
(Se não, os resultados não são credíveis)
- Os métodos utilizados para analisar os dados foram os mais apropriados? Permitiram testar as hipóteses adequadamente?
(Se não, os resultados e conclusões não são credíveis)

➔ A comparação dos trabalhos empíricos

Esta é a parte mais difícil da revisão da literatura mas é essencial se pretender encontrar uma Hipótese interessante e bem justificada para a parte empírica do seu trabalho.

Uma estratégia que o poderá ajudar a efectuar esta comparação é a seguinte:

1º - Tente agrupar os trabalhos empíricos em dois conjuntos. Um a que chamaremos GRUPO onde as hipóteses testadas são idênticas ou bastante semelhantes entre si, e outro designado por ISOLADOS, formado por trabalhos em que cada trabalho/autor testou uma hipótese diferente. Assim tem-se:

GRUPO

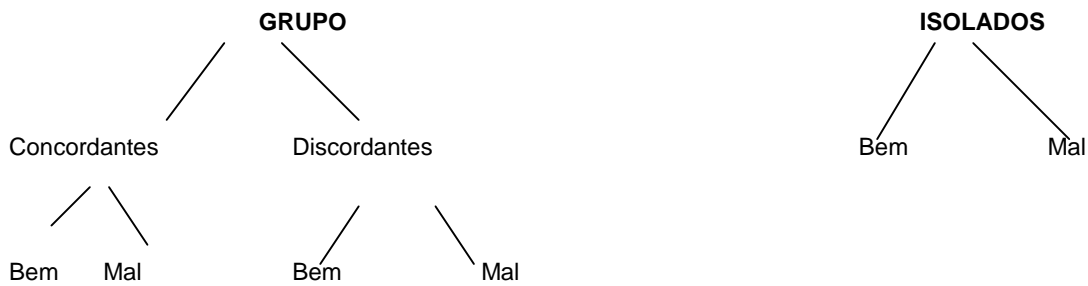
ISOLADOS

2º - Tente agora dividir o GRUPO em dois subgrupos de trabalhos: num deles, as conclusões dos trabalhos são **concordantes**, e no outro, as conclusões são **discordantes**. Assim:



ISOLADOS

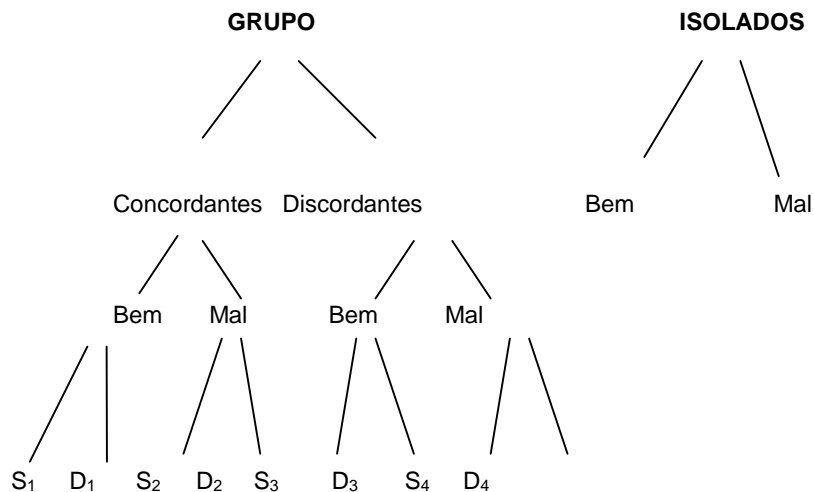
3º - Tente ainda, dentro de cada subgrupo, formar mais dois subgrupos: um em que o trabalho foi tecnicamente **bem feito**, segundo a avaliação que efectuou anteriormente, e outro em que achou que o trabalho foi **mal feito**. Faça o mesmo para os trabalhos ISOLADOS. Tem-se assim:



4º - Finalmente, tente formar mais dois subgrupos dentro de cada um dos subgrupos existentes: um grupo em que os trabalhos **usaram métodos semelhantes (S)**, e outro grupo em que os trabalhos usam **métodos muito diferentes (D)**.

Nota: não é possível fazer isto para os trabalhos isolados

Assim:



Porque é que se torna útil classificar os trabalhos empíricos desta maneira? A resposta é simples. Basicamente, podemos definir quatro tipos diferentes de trabalho empírico que pode desenvolver na sua tese, e os vários subgrupos S e D indicam que tipo(s) de trabalho(s) será razoável efectuar com base na literatura associada ao subgrupo. Os quatro tipos de trabalho empírico são os seguintes:

> - RÉPLICA DE UM TRABALHO ENCONTRADO NA LITERATURA

A réplica utiliza hipóteses, métodos, situação e tipo de amostra iguais aos utilizados no trabalho apresentado na literatura)

REPLICAR um trabalho apresentado na literatura é muito útil quando encontramos um estudo isolado bem feito. Isto porque um tal estudo precisa de ser replicado – especialmente, quando os resultados e conclusões parecem potencialmente importantes.

> - CONFIRMAÇÃO DE UM TRABALHO APRESENTADO NA LITERATURA

Neste tipo do trabalho a hipótese, a situação e o tipo de amostra são iguais às do trabalho encontrado na literatura mas os métodos de investigação, e talvez os métodos para analisar dados são diferentes.

A CONFIRMAÇÃO é útil para S_1 e S_3 porque estes trabalhos utilizaram métodos semelhantes e portanto é importante confirmar as conclusões usando métodos diferentes. A CONFIRMAÇÃO pode também ser útil no caso de um trabalho isolado bem feito.

> MELHORIA DE UM TRABALHO PUBLICADO NA LITERATURA

Neste tipo de trabalho empírico, as hipóteses gerais dum trabalho publicado na literatura são retidas mas as hipóteses operacionais, os métodos de investigação, a amostragem e os métodos para analisar os dados, são mais adequados.

A MELHORIA é útil para S_2 , D_2 , S_4 , D_4 bem como para os trabalhos isolados mal feitos porque estes trabalhos precisam de ser feitos novamente utilizando processos e técnicas mais correctas.

> - EXTENSÃO DE UM TRABALHO APRESENTADO NA LITERATURA

Este é o tipo de trabalho empírico mais criativo. É possível estender um trabalho apresentado na literatura de várias maneiras, por exemplo:

- Inventando uma hipótese alternativa para explicar os resultados de um trabalho na literatura;
- Deduzindo uma hipótese nova, interessante e importante, a partir das conclusões do trabalho apresentado na literatura, se este estiver bem feito.

A EXTENSÃO é útil para S_1 , D_1 e trabalhos isolados bem feitos.

5. O PLANEAMENTO DA INVESTIGAÇÃO EMPÍRICA

5.1. As relações entre as hipóteses, os métodos de investigação e os métodos para analisar dados

Como já referimos no início deste texto, a investigação empírica é uma viagem de ida e volta que começa e termina na literatura. Começa no “País Teórico”, chega ao “País Prático” e volta novamente ao “País Teórico”. As etapas desta viagem podem ser visualizadas esquematicamente na Figura 1.

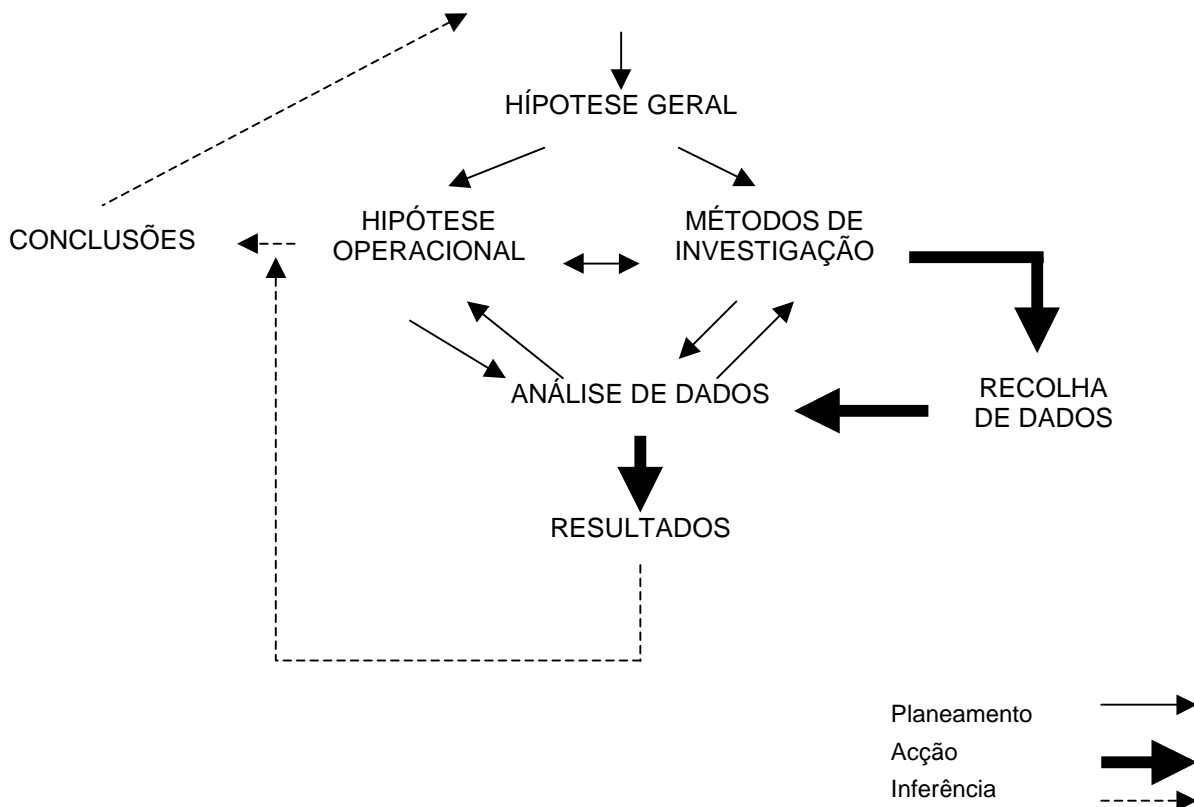
Figura 1

A LITERATURA

DINÂMIA - I

Indagação Socioeconómica

MANUELA MAGALHÃES HILL & ANDREW HILL



Feita a revisão da literatura e estabelecida a hipótese geral do seu trabalho, torna-se necessário operacionalizar a hipótese geral e seleccionar os métodos de investigação. Só depois é que se poderá passar à recolha e análise de dados e à apresentação dos resultados. Estes, vão confirmar ou negar a hipótese operacional estabelecida e é esta confirmação ou negação que vai fornecer as conclusões do seu trabalho empírico. Estas conclusões dão evidência para comentar a literatura (o que é o objectivo principal da investigação).

Suponhamos agora que já efectuou a revisão da literatura e estabeleceu a Hipótese para o trabalho empírico (note que só tem de inventar a sua própria Hipótese se usar o tipo de investigação "EXTENSÃO". Se fizer investigações dos tipos RÉPLICA, CONFIRMAÇÃO ou MELHORIA, a sua Hipótese já existe na literatura).

A próxima fase do planeamento do trabalho empírico começa na tradução da **Hipótese Geral** em **Hipótese Operacional**. A Hipótese Geral é exactamente isso -- geral. A Hipótese Operacional é mais específica e, no caso de os dados serem analisados usando métodos estatísticos, a Hipótese Operacional deverá ser escrita de modo a indicar a natureza das operações estatísticas necessárias para a análise.

Por exemplo:

HIPÓTESE GERAL:

Na empresa X, o absentismo está relacionado com a idade dos trabalhadores.

HIPÓTESE OPERACIONAL:

O número de dias de absentismo observado no ano 1998 na empresa X está significativamente correlacionado com a idade dos trabalhadores.

Quando estamos a planear o trabalho empírico é essencial pensar na Hipótese Operacional, nos métodos da investigação e nas análises de dados, **em conjunto**. É essencial pensar nestes três aspectos em conjunto porque, como a Figura 1 mostra, há interligações entre estas três fases do trabalho, e em cada uma delas, há tipos de constrangimentos ou limitações que automaticamente têm repercussões nas outras.

São muitos os investigadores inexperientes que não pensam nas análises de dados aquando do planeamento do trabalho e, no fim, têm grandes dificuldades em encontrar métodos adequados para analisar os dados.

Pensar nos métodos para analisar dados é tão importante na fase do planeamento que vamos dar um exemplo para ilustrar a natureza dos problemas:

Imagine uma empresa em que a Directora de Recursos Humanos se apercebe de que muitos dos operários numa dada secção andam insatisfeitos. Como a empresa tem um Consultor na área de Recursos Humanos, a Directora falou com ele sobre este assunto e disse que, em sua opinião, a insatisfação é provocada por falta da variedade no trabalho na secção. Pede então ao Consultor que proceda a uma investigação empírica e diz que a Hipótese Geral do trabalho deve ser:

HIPÓTESE GERAL:

A variedade no trabalho está relacionada com a satisfação global no emprego

O Consultor traduz esta hipótese na seguinte Hipótese Operacional:

HIPÓTESE OPERACIONAL:

Os operários que têm variedade no trabalho têm um valor médio de satisfação global superior ao dos operários que não têm variedade.

Para medir a satisfação global e a variedade no trabalho, o Consultor elaborou um pequeno questionário com apenas duas perguntas.

Pergunta 1. SATISFAÇÃO GLOBAL

Em que medida está satisfeito com o seu emprego?

MUITO INSATISFEITO	INSATISFEITO	NEM SATISFEITO NEM INSATISEITO	SATISFEITO	MUITO SATISFEITO
1	2	3	4	5

Pergunta 2. VARIEDADE (Var 1)

Tem variedade no seu trabalho?

NÃO	SIM
0	1

O Consultor aplicou o questionário aos 20 operários da secção. O Quadro 1 mostra as respostas obtidas, onde a variável “Var 1” representa a medida da variedade e “Sat” representa a medida da satisfação global. (Para já ignore a variável “Var 2” que explicaremos mais adiante).

Quadro 1. Respostas ao questionário aplicado

OPERÁRIO	VARIEDADE VAR 1	SATISFAÇÃO GLOBAL SAT	VARIEDADE VAR 2
1	0	1	0
2	0	2	0
3	0	2	0
4	0	2	0
5	0	3	0
6	1	3	1
7	1	3	1
8	1	4	1
9	1	5	2
10	1	3	3
11	1	2	3
12	1	2	3
13	1	1	3
14	1	1	4
15	1	1	4
16	1	1	4
17	1	1	4
18	1	1	4
19	1	1	4
20	1	1	4

Do Quadro 1 ressalta que 5 operários (números 1, 2, 3, 4 e 5) consideram que não têm variedade no trabalho. O Consultor designou este grupo de operários por “Grupo 1”. Os outros 15 operários (números 6 – 20) que consideram ter variedade no trabalho foram designados por “Grupo 2”.

Para testar a Hipótese Operacional o Consultor calculou a valor médio da satisfação global (variável “Sat”) para cada um dos dois grupos. Estes valores estão apresentados no Quadro 2 .

Quadro 2. satisfação global média por grupos de variedade no trabalho

	VARIEDADE (VAR 1)	
	GRUPO (1) (n = 5) (NÃO)	GRUPO (2) (n = 15) (SIM)
Valor médio de SATISFAÇÃO	2,00	2,00

Os valores médios dos dois grupos são iguais e portanto o Consultor tirou a seguinte conclusão:

CONCLUSÃO:

A Hipótese deve ser rejeitada. Não há relação entre variedade no trabalho e satisfação no emprego.

Mas a Directora de Recursos Humanos não confiou nesta conclusão. Ela apercebeu-se que o Consultor cometeu duas faltas:

Falta 1 - a Hipótese Geral estabelecida pela Directora referia existir uma **“relação”** entre variedade no trabalho e satisfação global. O Consultor foi testar uma hipótese que refere existir **diferença** entre os valores médios da satisfação. A Hipótese Operacional não foi adequada.

Falta 2 - a pergunta feita para medir variedade no trabalho usou uma escala de respostas muito restrita (apenas duas respostas alternativas SIM e NÃO). Para avaliar a relação entre duas variáveis é mais adequado usar uma escala de resposta mais larga porque a avaliação deve ser feita usando um coeficiente de correlação.

A Directora decidiu então fazer ela o seu próprio questionário reescrevendo a Hipótese Operacional da seguinte forma:

HIPÓTESE OPERACIONAL 2:

A variedade no trabalho está correlacionada com a satisfação global

Ela substituiu a pergunta 2 por uma nova:

Pergunta 3 . VARIEDADE (Var2)

Existe variedade no seu trabalho?

NENHUMA	POUCA	MÉDIA	BASTANTE	MUITA
0	1	2	3	4

Esta pergunta foi igualmente aplicada aos 20 trabalhadores da investigação do Consultor e os resultados encontram-se na última coluna do Quadro 1.

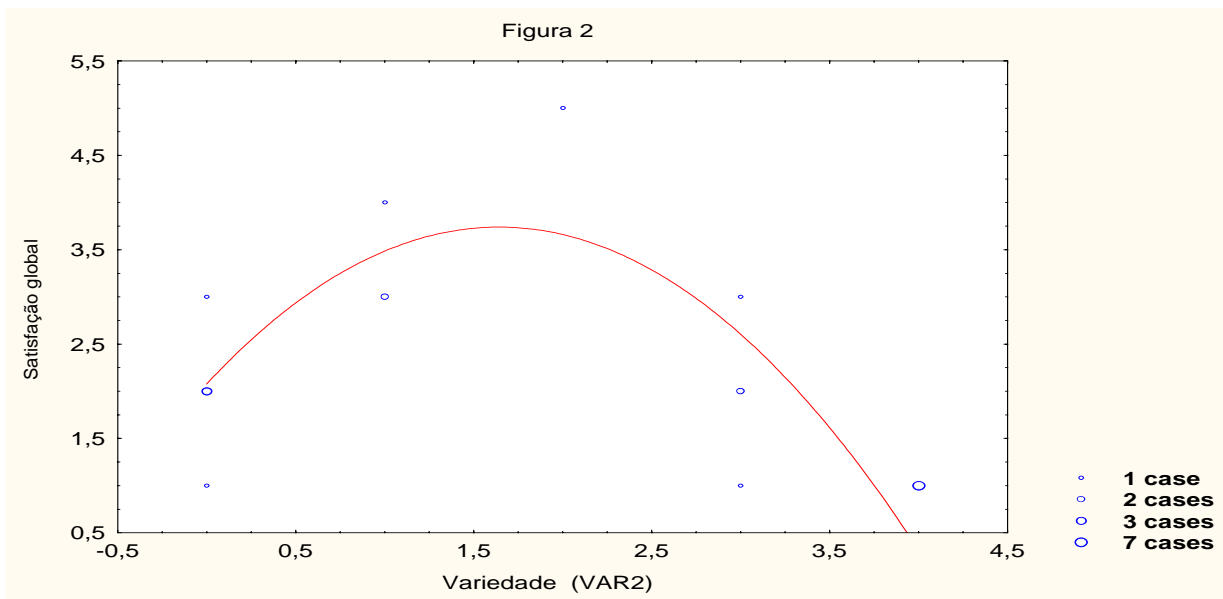
Antes de testar a sua Hipótese Operacional a Directora resolveu, por curiosidade, calcular os valores médios da satisfação global para a sua investigação. Separou então o grupo de trabalhadores que disseram que não tiveram NENHUMA variedade no trabalho (valor 0 na nova medida de Variedade – Var2) do grupo de trabalhadores que disseram ter ALGUM GRAU DE VARIEDADE no trabalho (valores 1, 2, 3 ou 4 na nova medida da Variedade – Var2). O Quadro 3 apresenta estes valores médios.

Quadro 3. Satisfação global média por grupos de variedade no trabalho

	VARIEDADE (VAR 1)	
	GRUPO (1) (n = 5) (Nenhuma)	GRUPO (2) (n = 15) (Alguma)
Valor médio de SATISFAÇÃO	2,00	2,00

Mais uma vez os valores médios são iguais (obviamente !!). Então, para testar a Hipótese Operacional 2, a Directora calculou o coeficiente de correlação de Pearson entre a satisfação global (Sat) e a variedade (VAR 2). O valor encontrado foi -0.482, que é um valor significativo ($p < 0,05$). Este valor suporta a conclusão de que os trabalhadores mais satisfeitos pensam que têm **menos** variedade no trabalho – uma conclusão exactamente ao contrário do que a Directora esperava!! (lembre-se que ela disse ao Consultor que na sua opinião “a insatisfação era provocada por falta de variedade no trabalho”).

Para verificar este resultado inesperado, a Directora calculou ainda o coeficiente de correlação não paramétrica de Spearman. O valor deste coeficiente foi -0,467 (também um valor significativo). Então ela lembrou-se que quando estudou Métodos Quantitativos Ihe foi ensinado que é sempre útil fazer um gráfico dos dados quando se está interessado na relação entre variáveis. E portanto, a Directora resolveu fazer o seguinte gráfico



O gráfico mostra que existe uma **relação não-linear** entre Satisfação e Variedade (Var 2). A Directora calculou então o coeficiente de correlação não-linear entre Satisfação e Variedade (Var 2) ajustando uma função quadrática aos dados. O valor encontrado foi 0.810 ($p < 0,0001$) e portanto ela tirou a seguinte conclusão:

CONCLUSÃO

A Hipótese 2 é aceitável. Há correlação significativa, mas não-linear, entre satisfação e variedade. Os operários mais satisfeitos têm um grau de variedade intermédio e os operários mais insatisfeitos não têm nenhuma ou têm muita variedade no trabalho.

QUE RESULTADOS DEVEMOS TIRAR DESTE EXEMPLO?

CONSELHO PRÁTICO

Cuidado com os Consultores !!

Na fase do planeamento do trabalho devemos verificar sempre se a Hipótese Operacional é uma boa tradução da Hipótese Geral. (a Directora traduziu bem, mas o Consultor não!).

Depois de estabelecer a Hipótese Operacional é necessário seleccionar métodos adequados para recolher os dados (medir as variáveis). Aqui devemos pensar adiante para verificar se, em princípio, as medidas das variáveis serão capazes de testar a Hipótese Operacional duma maneira justa e adequada. Caso contrário, temos que mudar ou a Hipótese ou os métodos para medir as variáveis (a Directora mudou os dois).

Paralelamente, é preciso escolher os métodos para analisar os dados e é necessário verificar se, na prática, estes métodos são adequados para testar a Hipótese Operacional.

Se pretendermos usar métodos estatísticos que assumem pressupostos fortes sobre a natureza dos dados (por exemplo, linearidade, normalidade, homogeneidade), é conveniente pensar em métodos alternativos porque muitas vezes os dados não verificam os pressupostos. Note que, na investigação do Consultor, por causa da natureza da medida de variedade que só tinha duas respostas à pergunta – SIM / NÃO, era impossível encontrar uma relação não-linear entre satisfação global e variedade no trabalho. Portanto, era impossível calcular uma correlação não-linear, pelo que o trabalho do Consultor não encontrou a conclusão certa.

5.2. A recolha dos dados

Os três métodos mais vulgarmente utilizados para efectuar a recolha dos dados, ou seja, para medir as variáveis são:

- Consulta a estatísticas ou a uma base de dados sócio-económicos já existente.
- Entrevista.
- Questionário.

Destes métodos, o mais vulgar é o do questionário – provavelmente porque é mais fácil de aplicar e mais fácil de analisar numa maneira quantitativa. Por causa da popularidade do questionário vamos tratar principalmente deste método embora nos devamos lembrar que entrevistas (bem preparadas) têm algumas vantagens. A construção de um questionário e a comparação entre questionários e entrevistas em termos de vantagens e desvantagens é um assunto longo demais para tratar neste *working paper*, e portanto será tratado separadamente¹. Aqui vamos apenas tratar dois aspectos técnicos das medidas feitas por questionário: A **fiabilidade** das medidas e a **validade** das medidas.

¹ Ver Hill, Manuela e Hill, Andrew (1998), A construção de um Questionário, DINÂMIA, ISCTE

5.2.1.A fiabilidade das medidas feitas por questionário

Um questionário pode ser utilizado tanto para obter factos como para medir opiniões, atitudes, satisfações, etc.. Quando aplicamos um questionário, as respostas obtidas formam um **amostra** tirada do **universo** das respostas possíveis. Quando usamos um questionário que tem perguntas “fechadas” (perguntas que têm um conjunto de respostas alternativas escritas pelo investigador) o universo das respostas alternativas é definido pelo conjunto das respostas alternativas. Por exemplo, considere a seguinte pergunta:

Em que medida está satisfeito com o seu chefe de secção?

Muito pouco	Pouco	Razoavelmente	Bastante	Muito
1	2	3	4	5

Nesta pergunta, o universo de respostas possíveis é definido por cinco respostas e o universo dos valores da satisfação tem cinco valores possíveis (1,2,3,4,5).

Como qualquer outra amostra, a amostra de respostas que obtemos de um questionário contém provavelmente erros. Para entender melhor este ponto imagine que aplicámos esta pergunta 50 vezes (uma vez por semana) ao respondente José. Os resultados são apresentados no Quadro 4.

Quadro 4. frequência das respostas

	RESPOSTAS				
	Muito pouco 1	Pouco 2	Razoavelmente 3	Bastante 4	Muito 5
Frequência	1	3	41	4	1

É evidente que a resposta “típica” é “Razoavelmente”, e o valor mais representativo do grau da satisfação é o valor 3 (o valor médio do grau de satisfação = 3,02). Mas é evidente também que há variação nas respostas de tempo em tempo. Esta variação indica a **falta de estabilidade temporal** na medida da satisfação com o chefe. Os valores 1,2,4 e 5 da satisfação não são **representativos do valor “típico” (valor 3)** e, em termos estatísticos, estes valores são **erros**. Mas de onde vêm estes erros? Principalmente são as variações no estado físico ou mental do José que produzem os erros. Por exemplo, o valor 1 (muito pouco satisfeito) podia ter acontecido quando o José estava deprimido, ou muito cansado, ou zangado depois de uma discussão com o Chefe. O valor 5 pode ter acontecido quando o Chefe o recomendou para promoção.

Uma outra fonte de erros está ligada com a natureza das perguntas e as escalas de resposta usadas no questionário. Por exemplo, perguntas complicadas, perguntas múltiplas e escalas de resposta inadequadas ou ilógicas² trazem dificuldades de interpretação por parte dos respondentes e podem produzir erros.

Normalmente, só se aplica o questionário uma vez e, portanto, só obtemos uma amostra de um só valor (uma resposta) para cada um das perguntas. Deste modo, não sabemos se o **valor observado** (a resposta à pergunta) é ou não representativo do **verdadeiro valor** no universo. Por outras palavras, não sabemos se a resposta é “típica” ou é um **“erro”** causado por um estado não típico do respondente, por uma pergunta mal escrita, ou por outra razão qualquer – por exemplo, o respondente mentiu. Normalmente, aplicamos um questionário a um conjunto de respondentes e, portanto, para cada um das perguntas obtemos um conjunto de valores observados. Naturalmente, há variação nos

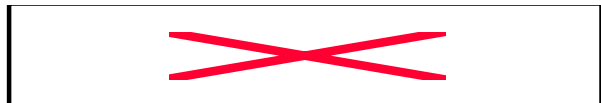
² estes aspectos são discutidos em Hill, Manuela e Hill, Andrew (1998), A construção de um Questionário, DINÂMIA, ISCTE

valores observados porque os respondentes são pessoas diferentes, mas cada um dos valores observados potencialmente contém erro.

A teoria da fiabilidade é baseada no pressuposto de que

$$\text{VALOR OBSERVADO} = \text{VERDADEIRO VALOR} + \text{ERRO} \quad (1)$$

E a fiabilidade de uma medida é definida como



$$\text{Fiabilidade} = \frac{\text{Valor Verdadeiro}}{\text{Valor Observado}} \quad (2)$$

Note que a Fiabilidade é essencialmente uma medida do erro. No caso ideal, quando não há erro, o verdadeiro valor e o valor observado são iguais, como se pode ver pela equação (1), e a Fiabilidade é igual a 1 (o que indica fiabilidade perfeita da medida).

- **Estimativa da fiabilidade (tipo estabilidade temporal)**

Como já dissemos anteriormente, uma fonte de erro é a falta de estabilidade temporal. É possível estimar este erro aplicando duas vezes o questionário ao mesmo grupo de respondentes. Normalmente, o intervalo de tempo entre as aplicações deve ser de, pelo menos, uma semana. A correlação entre os valores resultantes da aplicação 1 e da aplicação 2 do questionário dá uma estimativa do rácio na equação (2) e portanto dá uma estimativa da **Fiabilidade do tipo estabilidade temporal**.

- **Estimativas da fiabilidade de um conjunto de perguntas (fiabilidade interna)**

Podemos estar interessados em usar um conjunto de perguntas para criar uma variável nova. Por exemplo, suponha que queremos criar uma medida da Satisfação Global com o emprego, e suponha que a literatura sobre satisfação no emprego indica que há quatro “componentes” da Satisfação Global: Satisfação com o Chefe, Satisfação com o Ordenado, Satisfação com o Trabalho e Satisfação com os Colegas. Escrevemos as quatro perguntas seguintes para medir estas “componentes”, e vamos designar as perguntas por “itens” no Questionário. (note que a variável “Satisfação Global” é **definida** por quatro itens)

Pergunta:

Em que medida está satisfeito com:

	Muito pouco	Pouco	Razoavel mente	Bastante	Muito
	1	2	3	4	5
O seu ordenado	X				
O seu chefe			X		
O seu trabalho			X		
Os seus colegas					X

(As respostas apresentadas com um X são as respostas de um só respondente, por exemplo, o José)

Neste caso, a nova variável Satisfação Global, é definida por quatro “componentes” e o valor da Satisfação Global pode ser calculado como a soma dos valores das respostas aos quatro itens. Para o José, o valor da Satisfação Global é $(1 + 3 + 3 + 5) = 12$.

Para estimar a Fiabilidade do tipo estabilidade temporal para esta nova variável Satisfação Global, tínhamos de aplicar o questionário pelo menos duas vezes.

Mas, quando uma variável é definida por um conjunto de “componentes” há uma fonte de erro não ligada com a estabilidade temporal. Cada um dos itens escritos para medir as componentes é uma parte da nova variável (Satisfação Global), e os quatro itens devem contribuir igualmente para a Satisfação Global. Mais especificamente, a variância da cada um deles deve contribuir igualmente para a variância dos valores da Satisfação Global e a soma das variâncias dos quatro itens deve ser, no caso ideal, um quarto da variância da Satisfação Global. Qualquer diferença nas contribuições pode ser considerada como um erro produzido por falta de consistência entre os itens. Há vários métodos para estimar a fiabilidade ligada com esta fonte de erro e este tipo de fiabilidade chama-se **Fiabilidade interna**. O método mais vulgar para estimar a **Fiabilidade interna** é o coeficiente **alpha (α)** devido ao americano Lee Cronbach. Este coeficiente pode ser interpretado como um coeficiente de correlação.

Para uma medida baseada em k itens o valor de alpha tem a seguinte expressão:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\text{soma das variâncias de cada item}}{\text{variância total dos } k \text{ itens}} \right)$$

Note que há outros métodos para estimar a fiabilidade interna. Destes, o mais vulgar é o chamado método “*split half*” em que os itens são divididos em dois grupos – por exemplo, no nosso exemplo, o Grupo 1 pode conter “satisfação com Ordenado” e “satisfação com o Trabalho” e o Grupo 2 conter “satisfação com Chefe” e “satisfação com os Colegas”. Para cada um dos respondentes o valor total (soma dos valores) dos dois itens em cada um dos grupos é calculado e o coeficiente de correlação entre estes totais dá uma estimativa da fiabilidade interna. Mas este método, embora simples, tem um problema: é que há várias maneiras de dividir os itens (três maneiras no caso de quatro itens) e os coeficientes nem sempre concordam entre si. Qual é o método certo para agrupar os itens? Tradicionalmente, o método par/ímpar era utilizado (Grupo 1 = itens 1 e 3; Grupo 2 = itens 2 e 4) mas é um método arbitrário e portanto, um método muito fraco.

Note-se, contudo, que o método “*split-half*” é um caso especial do método alpha, sendo o valor do alpha igual ao valor médio de **todos** os coeficientes possíveis “*split half*” (no nosso exemplo, três coeficientes). Portanto o coeficiente alpha é preferível.

- **A fiabilidade da uma pergunta aberta (Fiabilidade de concordância)**

Por vezes os investigadores querem usar perguntas abertas, ou seja, perguntas para as quais o respondente tem de construir e escrever as respostas. Quase sempre as respostas dadas às perguntas abertas têm de ser “interpretadas”.

Considere este exemplo em que o investigador quer classificar a resposta de cada um dos respondentes em termos de uma ou outra de duas categorias “Opinião geralmente favorável” e “Opinião geralmente desfavorável”.

Pergunta:

Qual a sua opinião sobre o seu Chefe?

Uma possível resposta pode ser:

“Geralmente é simpático mas tem favoritos na secção e às vezes não é justo. Tem vontade de ouvir os nossos problemas e de vez em quando luta por nós contra a gerência – mas só de vez em quando”.

A opinião deste respondente sobre o Chefe é favorável ou desfavorável?

Obviamente, a resposta é uma mistura de opiniões favoráveis e desfavoráveis, mas o investigador quer classificar a resposta total como “Geralmente Favorável” ou “Geralmente Desfavorável”. A decisão precisa de um julgamento subjectivo porque depende da interpretação do investigador. Mais especificamente, depende do “peso”(importância) que o investigador der às opiniões positivas e às opiniões negativas na resposta. Dois investigadores podem discordar na classificação porque dão “pesos” diferentes (por causa dos seus valores e das suas personalidades) e a possibilidade de variação nas classificações da mesma resposta indica que o próprio investigador é uma fonte de erro na interpretação das respostas.

A este tipo de fiabilidade ligado com o investigador como fonte do erro dá-se o nome de **Fiabilidade da concordância**. Um grau elevado de concordância entre dois ou mais avaliadores indica um grau elevado de fiabilidade.

Quando estamos perante dois avaliadores é possível estimar a fiabilidade da concordância usando várias estatísticas. Por exemplo, imagine que dois avaliadores avaliaram as respostas à pergunta acima para 20 respondentes. O Quadro 5 apresenta os resultados.

Quadro 5. cruzamento das avaliações feitas às respostas à pergunta aberta

		AVALIADOR 2		
		Geralmente favorável	Geralmente desfavorável	TOTAL
AVALIADOR 1	Geralmente Favorável	6	2	8
	Geralmente Desfavorável	5	7	12
	TOTAL	11	9	20

Há dois tipos de métodos para avaliar a fiabilidade da concordância:

Método 1 – Utiliza correlações (Coeficiente Φ ou Cramér's V no caso em que os avaliadores usam mais do que duas categorias). Estas correlações são relacionadas com Qui quadrado (χ^2).

$$\Phi = \sqrt{\chi^2/n} \quad \text{e} \quad \text{Cramer's } V = \sqrt{\chi^2/n(k-1)}$$

Onde n é o número total de respostas e k é o número de categorias usadas pelos avaliadores

Método 2 - Utiliza as percentagens dos casos em que os dois avaliadores concordam na classificação.

No nosso exemplo, o valor do coeficiente Φ é 0,328, pelo que não é significativo ($p > 0,05$) e a percentagem dos casos em que os avaliadores concordam é dada por $[(6 + 7)/20] \times 100 = 65 \%$.

Temos aqui um problema porque o valor 65% revela um grau de concordância razoável (ou pelo menos não mau), mas o valor do coeficiente Φ não é significativo, o que indica que as classificações dos avaliadores não são significativamente concordantes.

A razão deste problema é que a percentagem (Método 2) não corrige as concordâncias que podem acontecer **por acaso**. No exemplo, o avaliador 1 classificou 8 respostas como “Geralmente Favorável” e o avaliador 2 classificou 11 respostas como “Geralmente Favorável”. Portanto, nesta situação, em média, $[(8 \times 11) / 20] = 4,4$ respostas seriam classificadas como “Geralmente Favorável” **só por acaso**. Analogamente, em média, $[(12 \times 9) / 20] = 5,4$ respostas

seriam classificadas como “Geralmente Desfavorável” **só por acaso**. Portanto, o número de respostas onde os avaliadores concordaram, $(6 + 7) = 13$, inclui $(4,4 + 5,4) = 9,8$ casos onde a concordância pode ter acontecido só por acaso.

À luz deste facto, é obvio que a percentagem de concordância deve ser corrigida. Uma estatística que faz esta correcção é o coeficiente **Kappa**. O valor de Kappa indica **a proporção dos casos** em que os avaliadores concordam nas suas classificações depois de fazer a correcção para a concordância que acontece por acaso. Se esse valor for igual a 1,00 existe concordância perfeita entre os avaliadores. É normal multiplicar o valor de **Kappa** por 100 para dar **a percentagem de concordância** entre os avaliadores e este valor apresenta a melhor medida de Fiabilidade da concordância.

Para o exemplo dos dados do Quadro 5 o valor de Kappa é dado por

$$Kappa = \frac{(13 - 9,8)}{(20 - 9,8)} = 0,314$$

Portanto, os avaliadores concordaram em 31,4% das suas classificações (depois da correcção) -- o que é um valor fraco. À luz deste valor, não há grande surpresa que o valor calculado para o coeficiente Φ não seja significativo.

É possível calcular o valor Kappa quando os avaliadores classificam as respostas em mais de duas categorias. Note-se, contudo, que os avaliadores têm de usar o mesmo conjunto de categorias (i.e., no exemplo, os dois avaliadores têm de usar as categorias “Geralmente Favorável” e “Geralmente Desfavorável”. Não é possível um avaliador usar Favorável/Desfavorável e o outro usar Agressivo/Passivo ou qualquer outro conjunto de categorias).

CONSELHO PRÁTICO SOBRE FIABILIDADE

1. *É muito importante dar atenção à fiabilidade das medidas feitas por questionário porque não vale a pena tirar conclusões a partir de medidas que não tenham fiabilidade adequada.*

2. *A escala seguinte dá uma indicação aproximada dos para avaliar o valor da uma medida de fiabilidade.*

<i>Maior que 0,9</i>	<i>.....</i>	<i>Excelente</i>
<i>Entre 0,8 e 0,9</i>	<i>.....</i>	<i>Bom</i>
<i>Entre 0,7 e 0,8</i>	<i>.....</i>	<i>Razoável</i>
<i>Entre 0,6 e 0,7</i>	<i>.....</i>	<i>Fraco</i>
<i>Abaixo de 0,6</i>	<i>.....</i>	<i>Inaceitável</i>

3. *O valor de alpha aumenta com:*

- (a) número de itens (perguntas) no questionário;*
- (b) correlações mais elevadas entre os itens (perguntas);*

Portanto, se quiser criar uma nova variável a partir dum conjunto de perguntas dum questionário, vale a pena usar mais perguntas do que as necessárias e seleccionar depois um conjunto suficientemente grande delas que estejam bem correlacionadas entre si.

4. *Para calcular o valor do coeficiente alpha use a opção SCALE e a sub-opção RELIABILITY ANALYSIS no SPSS (note que é possível usar esta opção para seleccionar um conjunto de componentes óptimas). Use a sub-opção SCALE IF ITEM IS DELETED para avaliar rapidamente o valor do coeficiente alpha para várias combinações de componentes.*

5. *O coeficiente da fiabilidade interna “split half” pode ser calculado usando a opção SCALE no SPSS, mas como já dissemos o coeficiente alpha é preferível.*

6. Como medida da fiabilidade da concordância, o coeficiente Kappa tem a vantagem de ser mais claro e mais simples de interpretar do que o coeficiente Phi (Φ). A vantagem do coeficiente Φ é de, por ser um coeficiente de correlação, o SPSS dar a sua significância. Sugere-se pois o cálculo dos dois coeficientes porque o SPSS não permite a avaliação da significância de Kappa.
7. Para calcular o valor das estatísticas Phi (Φ) e Kappa como estimativas da fiabilidade da concordância use a sequência de opções SUMMARIZE, CROSSTABS e STATISTICS (dentro da opção CROSSTABS).
8. Para aumentar a fiabilidade da concordância na classificação das respostas de perguntas abertas, ou em entrevistas, é muito útil elaborar um sistema para classificar as respostas. Por outras palavras, é conveniente elaborar uma grelha de respostas, ou tipos de respostas, que pertençam a cada uma das categorias usadas para fazer a classificação. Quando os avaliadores usam o mesmo sistema para classificar as respostas a fiabilidade da concordância (normalmente!) aumenta.

5.2.2. A validade das medidas feitas por questionário

Fiabilidade e validade de uma medida não são a mesma coisa, e têm uma relação estranha. Uma medida pode ter boa fiabilidade mas ter pouca validade e, sem fiabilidade adequada, a medida não pode ter validade adequada. Em termos lógicos, podemos afirmar:

A existência de fiabilidade adequada é necessária, mas não suficiente, para garantir validade adequada

Uma medida tem validade se for uma medida da variável que o investigador pretende medir. Mas não é verdade dizer que uma medida é “válida” ou “não válida” -- há graus da validade. Vejamos isto com o seguinte exemplo: suponha que um investigador pretende medir a característica pessoal “generosidade” das pessoas. Para isso, elaborou a seguinte pergunta:

	SIM	NÃO
<i>É uma pessoa generosa?</i>		

Acha que as respostas a esta pergunta vão ser sempre verdadeiras? Provavelmente não! Na nossa sociedade a generosidade é uma característica admirada -- tem grande “valor social” (ao contrário da avareza). Quase todas as pessoas gostam de ser consideradas generosas e portanto muitas pessoas que, na realidade, são pouco generosas, vão mentir nas suas respostas à pergunta. É pois pouco provável que a pergunta dê uma medida razoavelmente válida da generosidade porque a pergunta vai medir, em parte, a tendência em responder numa maneira socialmente aceitável. Falaremos mais sobre esta tendência mais adiante porque isto é um factor que muitas vezes influencia a validade das medidas feitas por questionário.

■ **Tipos de validade**

Há três tipos principais de validade:

- **Validade de conteúdo**
- **Validade teórica**
- **Validade prática**

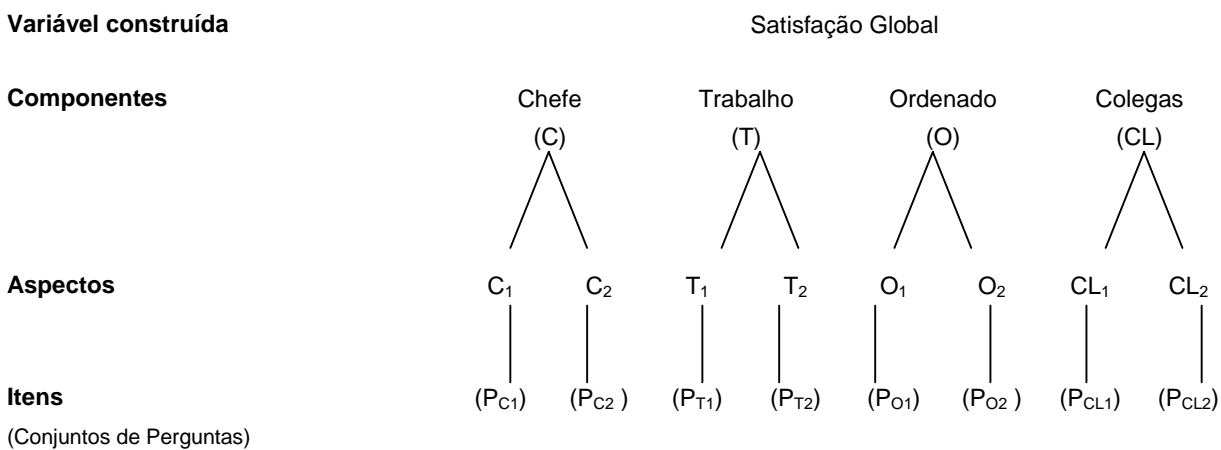
O exemplo acima apresentado, focou a validade teórica de uma só pergunta, mas normalmente só é possível estimar a validade de uma medida obtida através de um conjunto de perguntas (i.e. um conjunto de **itens**). Por exemplo, suponha

que um investigador quer medir a Satisfação Global no emprego usando perguntas do tipo do exemplo da secção 5.2.1. Estas perguntas são itens para medir “componentes” da Satisfação Global (satisfação com o Chefe, satisfação com o Trabalho, satisfação com o Ordenado e satisfação com os Colegas). Uma maneira de obter a medida de Satisfação Global é somar os valores numéricos das respostas aos quatro itens. Neste caso o investigador tem de examinar a validade dessa medida de Satisfação Global. Não é possível examinar a validade dos itens (perguntas individuais) porque é impossível distinguir entre respostas genuínas e respostas falsas.

1. VALIDADE DE CONTEÚDO

No caso de uma variável ser construída a partir de um conjunto de componentes há uma hierarquia como a apresentada na Figura 3. A variável é formada por várias componentes. Normalmente, cada uma das componentes tem, por sua vez, vários aspectos ($C_1, C_2, \dots, CL_1, CL_2$) e sobre cada aspecto serão feitas uma ou várias perguntas (PC_1, \dots, PCL_2).

Figura 3



A validade de conteúdo engloba os dois aspectos seguintes:

↪ A representatividade das componentes

Teoricamente há um **Universo de componentes** a partir do qual a variável Satisfação Global deve ser definida. Se a literatura sobre Satisfação Global referir que o Universo é formado pelas quatro componentes “satisfação com o Chefe”, “satisfação com o Trabalho”, “satisfação com Ordenado” e “satisfação com Colegas”, então a variável Satisfação Global do nosso exemplo terá validade de conteúdo adequada em termos de **representatividade das componentes**. Suponha, contudo, que a literatura diz que o Universo de componentes para a variável Satisfação Global deve incluir, para além destas quatro, as componentes “satisfação com a Gerência” e “satisfação com Condições do Ambiente”. Neste caso, as quatro componentes não são totalmente representativas do Universo, e a validade de conteúdo da medida de Satisfação Global será mais baixa.

↪ A representatividade dos itens como medidas das componentes

Tal como mostra a figura 3, cada uma das componentes tem normalmente um conjunto de aspectos. Para medir a componente por forma a que a medida de Satisfação Global tenha validade de conteúdo adequada, os itens devem ser representativos do **Universo de Aspectos**.

Assim, para construir uma medida de Satisfação Global que tenha validade de conteúdo adequada, é preciso encontrar o Universo de Aspectos para cada uma das componentes. Por exemplo, no caso da componente “satisfação com o

Trabalho”, há vários Aspectos do trabalho com os quais um respondente pode estar satisfeito ou não, tais como: o grau de responsabilidade que possui, o grau de variedade do seu trabalho, o grau de interesse pelo trabalho que desempenha, o grau de controle sobre os métodos usados, o grau de fadiga etc.. Usando a literatura sobre a área (e bom senso!) é possível construir uma lista com todos os Aspectos **importantes** do trabalho com os quais o respondente pode estar ou não satisfeito. Estes Aspectos formam o **Universo de Aspectos**. Os itens (perguntas) escolhidos para medir a componente devem ser representativos deste universo. Normalmente, a pergunta associada a cada um dos Aspectos deve contribuir para que a medida (no exemplo, Satisfação Global) tenha validade de conteúdo adequada.

Note que não é possível calcular um valor numérico para a validade de conteúdo

2. A VALIDADE TEÓRICA

Uma medida tem uma boa validade teórica se for uma medida da variável que o investigador pretender medir. Há três tipos de validade teórica que tratam de aspectos ligeiramente diferente do conceito.

- **Validade convergente**
- **Validade discriminante**
- **Validade factorial**

Validade convergente

Uma medida (M_{1X}) da variável ‘X’ tem validade convergente se a medida concordar bem com outras medidas da mesma variável, por exemplo, M_{2X} , M_{3X} , M_{4X} que já sejam aceites como medidas adequadas da variável ‘X’. O grau da concordância é estimado por meio do coeficiente de correlação. Para ter validade convergente adequada as correlações entre M_{1X} e M_{2X} , entre M_{1X} e M_{3X} e entre M_{1X} e M_{4X} devem ser significativas. (nota: é normal usar duas ou três outras medidas para avaliar a validade convergente de uma medida M_{1X}).

Validade discriminante

Uma medida da variável ‘X’ tem validade discriminante se não estiver correlacionada significativamente com medidas de variáveis que teoricamente não estão relacionadas com a variável ‘X’. Por exemplo M_{1X} não deve ter correlações significativas com as medidas M_{1Y} e M_{1Z} das variáveis Y e Z .

Campbell e Fiske (1959) propuseram uma técnica sofisticada chamada “*multi método – multi traço*” para avaliar as validades convergente e discriminante simultaneamente. Mas esta técnica é complicada e é principalmente usada no contexto de psicometria.

Validade factorial

Este tipo de validade teórica é útil quando o investigador quer definir e medir uma nova variável usando um conjunto de perguntas num questionário. É possível definir a nova variável aplicando a técnica de Análise Factorial – técnica semelhante à técnica de Análise de Componentes Principais. O peso ou saturação de cada um dos itens (perguntas que, em conjunto, vão definir a nova variável) no factor representa a validade factorial do item (pergunta).

Voltemos ao exemplo apresentado para construir a variável “Satisfação Global”. Suponha que a literatura refere que a variável Satisfação Global tem seis componentes -- satisfação com o Trabalho, Ordenado, Chefe, Colegas,

Administração e Ambiente de trabalho. Para cada uma destas seis componentes foi elaborada uma pergunta, o que não dará grande validade de conteúdo, mas permite um exemplo mais simples. (Normalmente, como é evidente na Figura 3, cada uma das componentes tem vários Aspectos e cada um deles está ligado com pelo menos um item pergunta no questionário). O questionário foi preenchido por 60 respondentes como parte de um inquérito sobre absentismo e os dados são apresentados no Anexo. Os resultados da Análise Factorial encontram-se no Quadro 6.

Quadro 6. Resultados da Análise Factorial

VARIÁVEL (PERGUNTA)	FACTOR 1
TRABALHO	0,712
ORDENADO	0,595
CHEFE	0,688
COLEGAS	0,651
ADMINISTRAÇÃO	0,782
AMBIENTE	0,440

O factor obtido é uma combinação linear dos itens e representa a nova variável “Satisfação Global”. É importante notar duas coisas: primeiro, o peso de um item representa a correlação entre o item e o factor; segundo, os itens não têm pesos iguais. Por exemplo o item relacionado com o ambiente de trabalho tem menor peso (0,440) que o item relacionado com o próprio trabalho (0,712) . Este facto implica que a satisfação com o trabalho é mais importante do que a satisfação com o ambiente de trabalho na definição e medida da Satisfação Global.

O **coeficiente de validade factorial** de cada um dos itens é o peso do item no factor. Por exemplo, a validade factorial do item “Em que medida está satisfeito com o seu trabalho?” é 0,712.

3. A VALIDADE PRÁTICA

São muitos os métodos para avaliar a validade prática de uma medida. Os dois métodos clássicos são:

- A validade preditiva (“*Predictive validity*”)
- A validade simultânea (“*Concurrent validity*”)

↳ A validade preditiva

A validade preditiva de uma medida é a validade que essa medida tem para prever valores noutra variável. Por exemplo, suponha que o Director(a) de Recursos Humanos queria usar um teste para seleccionar empregados. Obviamente, não vale a pena usar o teste se o valor que um candidato obteve no teste não conseguir prever bem o seu desempenho no emprego. O coeficiente de validade preditiva do teste é a correlação entre os valores do teste e os valores de uma medida de desempenho no emprego. Para estimar este coeficiente é preciso recolher valores do teste, num dado momento, para uma amostra de candidatos seleccionados **sem utilização do teste como parte do processo de selecção**. Mais tarde, é preciso recolher os valores de desempenho para os candidatos seleccionados. A correlação entre os valores do teste e os valores de desempenho representa o coeficiente de validade preditiva.

↪ A validade simultânea

Neste tipo de validade, os valores da variável a prever são recolhidos ao mesmo tempo (aproximadamente) que os valores do teste. Por exemplo, o Director(a) de Recursos Humanos pode obter os valores do teste e os valores de desempenho no emprego para uma amostra dos empregados da empresa. A correlação entre os valores do teste e os valores de desempenho é uma estimativa da validade simultânea do teste.

A diferença entre os coeficientes de validade preditiva e simultânea parece não ser grande --- mas pode sê-lo. Isto porque na validade simultânea o coeficiente da validade é calculado a partir dos dados dos empregados enquanto que no caso de validade preditiva o coeficiente de validade é calculado a partir dos dados dos candidatos. É natural que os valores de avaliação de desempenho dos empregados abranja uma gama restrita dos valores de desempenho, e portanto, a diferença entre a validade preditiva e a validade simultânea pode ser grande. É, contudo, possível corrigir o valor do coeficiente de validade simultânea, tornando mais fácil e mais barato calcular este coeficiente de validade prática. (**vide** Murphy & Davidshofer: 1988).

CONSELHO PRÁTICO SOBRE VALIDADE

☛ **A influência da tendência em dar respostas socialmente desejáveis**

Suponha que pretende usar perguntas dum questionário para medir atitudes, opiniões ou satisfações, mas não quer usar as perguntas para criar uma nova variável (como Satisfação Global no nosso exemplo). Por outras palavras, pretende analisar as respostas a cada pergunta individualmente. Neste caso, vale a pena ter em conta a natureza das perguntas. Se considerar que as respostas podem ser muito influenciadas por “valores sociais” (lembre-se do exemplo da questão “É uma pessoa generosa?”) é essencial utilizar um questionário **anónimo**. Provavelmente o anonimato não vai eliminar totalmente o problema, mas ajuda.

É possível aplicar uma medida da tendência de usar respostas “socialmente desejáveis” (por exemplo a medida do Edwards (1957) para corrigir as respostas do questionário ou para eliminar da amostra os respondentes que mostram fortemente esta tendência). É possível também construir a sua própria medida desta tendência (vide Hill: 1986), mas este processo é complicado e requer muito tempo. Vale mais a pena tentar escrever as perguntas de modo mais neutro, se for possível, e testá-las numa amostra pequena.

☛ **A validade de conteúdo**

Se quiser usar um conjunto de perguntas para definir e medir uma nova variável (como Satisfação Global no exemplo), a validade de conteúdo é importante. Portanto, antes de escrever as perguntas, use a literatura para encontrar todas as componentes relevantes e todos os Aspectos das componentes da variável que quer definir.

☛ **A validade factorial**

Se quiser criar uma variável nova a partir de um conjunto de perguntas:

- Use a técnica de Análise factorial. Não use a técnica de Componentes Principais (explicaremos a razão mais adiante na secção 6.5). A Análise factorial não é uma técnica – é uma família de técnicas. Para criar uma variável nova é melhor usar o tipo de Análise factorial designada por “Análise factorial Alpha” porque esta técnica cria uma nova variável (um factor) em que as medidas da variável têm o valor do coeficiente alpha (a fiabilidade interna) máximo. Para fazer uma Análise factorial (tipo alpha) use as seguintes opções do SPSS:

DATA REDUCTION
FACTOR
EXTRACTION

Dentro da opção Extraction use

METHOD = ALPHA
CORRELATION MATRIX
EIGENVALUES GREATER THAN 1
MAXIMUM ITERATIONS = 50

Se a análise conduzir a duas ou mais factores, e só quer um factor, deve repetir a análise usando a opção

Number of factors = 1 (em vez da opção “Eigenvalues greater than 1)

- Para calcular os valores dos casos da nova variável (i.e. o factor) use as opções
SCORES
SAVE AS VARIABLES
METHOD = REGRESSION

Estas opções criam uma **nova variável** (FAC_1) no seu ficheiro de dados.

Não se esqueça de o GRAVAR !!

Notas:

1. Os valores (“scores”) da nova variável tomam em linha de conta os pesos dos itens no factor obtido através da análise factorial, e não são simplesmente a soma dos valores das respostas aos itens (perguntas) do questionário. É melhor usar os “scores” do que a soma simples dos valores das respostas -- especialmente no caso em que os pesos das variáveis no factor não são aproximadamente iguais.
2. É essencial verificar se o valor do coeficiente Kaiser–Mayer–Olkin (KMO) é adequado para fazer a análise factorial. Este valor deve ser, pelo menos, igual a 0,6 e preferivelmente igual a 0,8 ou mais. Use a opção DESCRIPTIVES dentro da opção FACTOR no SPSS para verificar o valor de KMO.

5.3. Amostragem

5.3.1. Definições: o universo de casos e a amostra de casos

Uma investigação lida normalmente com um conjunto de casos. Os casos podem ser pessoas, empresas, sectores da indústria, países ou qualquer outra entidade que o investigador queira considerar.

Ao conjunto **total** dos casos dá-se o nome de **População** ou **Universo** de casos. [Vamos usar a palavra “Universo” porque muitas pessoas pensam (incorrectamente) que, neste contexto, a palavra “População” só se refere à situação em que os casos são pessoas]. Suponha então que um investigador quer estudar as políticas dos hotéis Portugueses, e suponha que há um total de 8500 hotéis no país. Neste caso o Universo é o conjunto dos 8500 hotéis. Note que é o investigador que define o Universo. Um investigador que só queira analisar a actividade hoteleira em Lisboa definiria o Universo como sendo o conjunto de todos os hotéis em Lisboa.

Acontece, porém, que muitas vezes o investigador não tem tempo, nem recursos suficientes, para analisar cada um dos casos do Universo, pelo que, nesta situação, só é possível considerar uma parte dos casos que constituem o Universo. Esta parte de casos é uma **amostra** do Universo. Assim, o que o investigador pretende ou pode fazer, na maioria das situações, é analisar os dados da amostra, tirar conclusões, e extrapolar as conclusões para o Universo. Mas este processo de **extrapolação das conclusões** pode correr bem ou pode correr mal. Se a amostra dos casos for retirada do Universo de modo que seja **representativa** desse Universo, é possível aceitar, com razoável confiança, que as conclusões obtidas utilizando a amostra possam ser extrapoladas para o Universo. Mas, se a amostra for retirada sem ter sido tida em consideração a sua representatividade, não é possível extrapolar as conclusões para o Universo com confiança. Daí que se torna importante escolher uma amostra representativa.

5.3.2. A escolha de uma amostra

Há duas “famílias” de métodos para escolher uma amostra:

- Os métodos de amostragem casual
- Os métodos de amostragem não – casual (também chamada “dirigida”)

➤ OS MÉTODOS DE AMOSTRAGEM CASUAL

Os métodos de amostragem casual são preferíveis quando pretendemos extrapolar para o Universo os resultados obtidos a partir da amostra porque têm duas grandes vantagens:

- É possível demonstrar a representatividade da amostra;
- É possível estimar (estatisticamente) o grau de confiança com o qual as conclusões tiradas da amostra se aplicam ao Universo.

Os métodos mais vulgares de amostragem casual podem ser de vários tipos:

- Amostragem aleatória simples
- Amostragem sistemática
- Amostragem estratificada
- Amostragem por clusters
- Amostragem multi-etapas
- Amostragem multi-fásica

Explicações claras e simples destes métodos (e dos métodos de amostragem não casual) podem ser encontradas em Reis, Melo, Andrade e Calapez (1996) e portanto não vamos aqui explicar todos eles.

De entre os métodos de amostragem casual, o método de amostragem estratificada e o método de amostragem por clusters são talvez os mais úteis e utilizados em Ciências Sociais.

➔ Método de amostragem estratificada

Este método é especialmente útil quando se pretende obter uma amostra representativa de um Universo grande segundo várias variáveis pré-identificadas. Por exemplo, suponha que pretendíamos estudar as políticas dos hotéis nacionais e queríamos analisar os resultados em termos de dimensão e local geográfico dos hotéis. Para isso devemos seleccionar uma amostra de hotéis a partir do conjunto de hotéis nacionais que seja representativa em termos de dimensão (número de trabalhadores) e local. Para termos uma amostra estratificada precisamos de efectuar os seguintes passos:

- ✓ Decidir quanto ao número e natureza dos estratos.

Por exemplo:

Dimensão: 3 estratos --- (menos de 50 trabalhadores, 51 – 100, mais de 100)

Local: 3 estratos --- (Norte, Centro, Sul)

- ✓ Encontrar uma lista de todos os hotéis em Portugal que inclua; nome, local e dimensão)
- ✓ A partir desta lista construir um Quadro para caracterizar o Universo. (Quadro 7 em baixo)
- ✓ Decidir qual o tamanho da amostra -- qual a fracção de amostragem (por exemplo, 10% do Universo)
- ✓ Seleccionar uma amostra **aleatória** de 10% dos hotéis em cada um dos estratos (Quadro 8)

Quadro 7. Caracterização do Universo

REGIÃO	DIMENSÃO				
		Menos de 50	51 - 100	Mais de 100	TOTAL
NORTE		1100	740	160	2000
CENTRO		900	500	100	1500
SUL		3500	1200	300	5000
TOTAL		5500	2440	560	8500

Quadro 8. Caracterização da amostra = 10% do Universo

REGIÃO	DIMENSÃO				
		Menos de 50	51 - 100	Mais de 100	TOTAL
NORTE		110	74	16	200
CENTRO		90	50	10	150
SUL		350	120	30	500
TOTAL		550	244	56	850

Note que em cada uma das células do Quadro 8 o número de casos (hotéis) é igual a 10% do número de casos na célula correspondente no Universo. A amostra é representativa do Universo **se** a amostra de casos em cada um das células for escolhida **por amostragem aleatória**.

➡ Método de amostragem por clusters

Este método é útil quando temos um Universo grande mas os casos são agrupados em unidades. Por exemplo, suponha que queríamos estudar a relação entre o absentismo voluntário no emprego e a satisfação dos empregados dos hotéis de Lisboa. Neste caso o Universo é o conjunto de todos os empregados dos hotéis de Lisboa. Mas neste Universo há grupos ou “clusters” de casos -- o conjunto de empregados em cada um dos hotéis forma um “cluster”. Para escolher uma amostra por clusters devemos seguir os seguintes passos:

- ✓ Encontrar uma lista dos hotéis de Lisboa;
- ✓ Decidir qual o número de hotéis sobre os quais a investigação vai ser aplicada (Por exemplo 10%);
- ✓ Escolher 10% dos hotéis por amostragem aleatória;
- ✓ Efectuar a investigação sobre todos os empregados dos hotéis que constituem a amostra de 10% .

➤ OS MÉTODOS DE AMOSTRAGEM NÃO – CASUAL

Estes métodos não são aconselháveis quando se pretende extrapolar para o Universo os resultados ou conclusões obtidos com a amostra, mas podem ser úteis no início de um investigação – por exemplo, para testar as primeiras versões dos questionários.

Dentro deste tipo de métodos os mais vulgares são:

- **Amostragem por conveniência**
- **Amostragem por quotas**

➔ **método de amostragem por conveniência**

Neste método os casos escolhidos são os casos que são facilmente disponíveis (muitas vezes, os amigos dos amigos!). O método tem vantagens por ser rápido, barato e fácil. As desvantagens são que os resultados e conclusões só se aplicam à amostra, não podendo ser extrapolados para o Universo. Isto porque não há garantia de que a amostra seja razoavelmente representativa do Universo (muitas vezes não é óbvio o que é o Universo!).

➔ **método de amostragem por quotas**

Este método é análogo ao método de amostragem estratificada mas com uma diferença importante. Em vez de se escolher uma amostra aleatória dentro de cada uma dos estratos na etapa final (Passo 5 no exemplo acima), escolhe-se uma amostra não-aleatória (de tamanho determinado pela fracção de amostragem). Às vezes, as amostras dentro dos estratos são pouco diferentes das amostras por conveniência, e quase sempre há um enviesamento na selecção dos casos. Portanto não é necessariamente razoável extrapolar para o Universo os resultados e conclusões tirados da amostra. Contudo, e em geral, este método é melhor do que o da amostragem simplesmente por conveniência.

CONSELHO PRÁTICO SOBRE AMOSTRAGEM

1. Restrição da investigação

No contexto de um projecto de licenciatura ou de uma tese de Mestrado o tempo para fazer investigação é limitado. Portanto é de bom senso restringir a natureza e tamanho da investigação. É possível fazer isto por meio da escolha de um tema em que a investigação pode ser feita num Universo pequeno de casos -- por exemplo, todos os empregados de uma empresa. Neste caso não é preciso escolher uma amostra, e portanto não há problemas de amostragem. É verdade que num tal estudo não é possível extrapolar as conclusões a outras empresas no mesmo sector, mas isto não importa; o que importa é que o trabalho seja bem feito.

2. A estruturação de uma amostra estratificada

Numa investigação relativamente simples é provavelmente melhor usar uma amostra por clusters ou uma amostra estratificada. Mas como decidir quanto aos estratos de uma amostra estratificada? Não é difícil. Só precisa de pensar nos objectivos da investigação. No exemplo do estudo sobre as políticas dos hotéis nacionais o objectivo é identificar as políticas. É razoável pensar que os hotéis de maior dimensão tenham talvez políticas diferentes dos hotéis de menor dimensão (muitas vezes os grandes hotéis fazem parte de uma grande empresa multinacional e os pequenos hotéis pertencem a uma família) e, portanto, "dimensão" deva ser utilizada para estratificar a amostra. É também razoável considerar que os hotéis têm vários tipos de clientes, por exemplo, turistas, pessoas de negócio, viajantes comerciais etc.. Provavelmente, a grande maioria dos turistas vão para os hotéis do Algarve e uma parte menor para os hotéis de Lisboa. A maioria dos clientes dos hotéis de Lisboa e do Porto são provavelmente pessoas de negócios. É assim provável que os hotéis que têm tipos de clientes diferentes tenham políticas diferentes, portanto pode ser útil utilizar também "Local" para estratificar a amostra. Convém notar que, em muitas das investigações, a revisão da literatura que deu luz aos objectivos do estudo é muito útil na escolha das variáveis para estratificar a amostra.

3. O problema da representatividade de uma amostra reduzida

Na maioria das situações, o número de questionários recebidos não coincide com o número de casos da amostra – há sempre um conjunto de pessoas ou instituições que não respondem ao questionário ou não dão a informação solicitada. Vamos usar o termo amostra reduzida quando a amostra obtida é apenas uma parte da amostra alvo (a amostra completa), ou seja, da amostra inicialmente planeada. Isto é um problema muito vulgar no caso de questionários enviados por correio. Se a amostra alvo foi escolhida por meio de métodos aleatórios, a existência de uma amostra reduzida traz problemas relativos à sua representatividade. Se a amostra alvo for bem planeada, ela é representativa do Universo, mas não há garantia de que a amostra reduzida também o seja. Se, por exemplo, a amostra alvo foi escolhido utilizando o método por quotas o problema parece mais fácil de resolver pois é relativamente fácil encontrar mais casos num estrato onde haja falta de casos – quaisquer casos que tenham as características do estrato servem. (Mas lembre-se que se utilizar o método por quotas, a amostra alvo não é necessariamente representativa do Universo). O problema é mais difícil resolver no caso de uma amostra estratificada. Neste caso há quatro possibilidades:

- *Persuadir os inquiridos que não deram resposta a preencher o questionário. (É muito provável que este processo aumente o tamanho da amostra, mas é menos provável que resolva inteiramente o problema).*
- *Escolher casos alternativos para obter a amostra alvo (A vantagem é óbvia, mas as desvantagens são que precisa de tempo e não é certo que seja possível obter suficientes casos adicionais).*
- *Analisar as características dos “não–respondentes” para decidir se será provável que a falta de informação destes casos tenha introduzido um enviesamento na amostra. (A desvantagem aqui é que a decisão não é totalmente objectiva, e é um processo difícil. Mas há literatura nesta área e, por vezes, é possível “corrigir” os dados (vide: Moser e Kalton:1986).*
- *Provavelmente a solução mais prática (mesmo que não seja a melhor solução) é comparar a amostra reduzida com uma amostra do mesmo tamanho retirada por quotas do Universo. Se a diferença for pequena em todos os estratos, a amostra reduzido pode ser tratada como sendo uma amostra por quotas. Se a diferença for grande em alguns dos estratos é relativamente fácil aumentar o número de casos tal como se faz numa amostra por quotas (ou eliminar casos usando um método aleatório). Para ilustrar este processo usaremos o exemplo da investigação das políticas dos hotéis nacionais. O quadro 7 mostra que o Universo tem 8500 casos no total. A amostra estratificada com a fracção amostral de 10% é formada por 850 casos e apresentada no quadro 8. Imagine que só obtivemos os 600 casos do Quadro 9, o que representam uma fracção amostral do Universo igual a $(600 \times 100)/8500 = 7,059\%$. O quadro 10 apresenta o número de casos para cada um dos estratos para uma amostra de 7,059% dos casos do Universo. Para calcular o número de casos em cada estrato do Quadro 10 calcula-se a percentagem dos casos no estrato no Universo e multiplica-se esta percentagem por 600. Por exemplo, no estrato ‘Dimensão (51 – 100) no Norte’ há 740 casos do Universo de 8500 casos. Portanto este estrato contém $(740 \times 100)/ 8500 = 8,7\%$ dos casos no Universo. Numa amostra por quota que tem 600 casos, o estrato ‘Dimensão (51 – 100) no Norte’ deve ter $(8,7 \times 600)/100 = 52$ casos.*

Quadro 9. A amostra reduzida de 600 casos

REGIÃO	DIMENSÃO				TOTAL
	Menos de 50	51 - 100	Mais de 100		
NORTE	80	50	10	140	
CENTRO	60	40	10	110	
SUL	240	90	20	350	
TOTAL	380	180	40	600	

Quadro 10. A amostra por quotas de 600 casos

REGIÃO	DIMENSÃO				
		Menos de 50	51 - 100	Mais de 100	TOTAL
NORTE		78	52	11	141
CENTRO		64	35	7	106
SUL		247	85	21	354
TOTAL		389	172	39	600

É evidente que a amostra reduzida e a amostra por quotas de 600 casos são relativamente semelhantes. Para obter uma **amostra por quotas** perfeita só precisamos de eliminar 15 casos e obter mais 15 casos. Neste exemplo podemos dizer que a amostra reduzida é razoavelmente representativa do Universo **como amostra estratificada**. Mas o exemplo não é típico, e geralmente é mais razoável concluir que a amostra é **adequada como uma amostra por quotas**.

Há outras maneiras de resolver o problema da representatividade da amostra, como sejam, utilizando ponderadores. Neste caso é aconselhável consultar um especialista em estatística.

6. ALGUNS PROBLEMAS COM A ANÁLISE DE DADOS

Nesta secção discutiremos várias questões práticas (problemas e soluções) relacionadas com a análise estatística de dados. Para ilustrar alguns pontos vamos usar o conjunto de dados apresentado em Anexo. Estes dados representam os resultados de um inquérito realizado numa fábrica onde a ausência ao emprego era um problema preocupante. Antes de ler esta secção convém ler a descrição (em Anexo) da natureza e objectivos deste inquérito.

6.1. O agrupamento de casos

Por vezes torna-se útil (ou necessário) usar uma variável contínua para agrupar os casos em dois grupos. Por exemplo, suponha que queríamos usar os dados do Anexo para comparar os casos que têm um nível elevado de absentismo (AUSÊNCIA) com os casos que têm um nível baixo de absentismo. Para fazer isto temos de recodificar a variável AUSÊNCIA noutra variável (por exemplo, AUSEN_2) que só vai ter dois valores (Baixa Ausência = 1 e Alta Ausência = 2). Os casos que têm valor 1 na variável AUSEN_2 constituem o Grupo 1 (Baixa Ausência) e os casos que têm valor 2 constituem o Grupo 2 (Alta Ausência). Os três métodos mais vulgares para fazer este tipo de recodificação são:

- **Partição pelo valor médio da variável que se pretende agrupar**
- **Partição pela mediana.**
- **O método de “Grupos Extremos”**

➔ **Partição pelo valor médio da variável que se pretende agrupar**

O valor médio da variável AUSÊNCIA é 10,43. Portanto, os casos com valores de AUSÊNCIA inferior a 10,43 têm o valor 1 na variável AUSEN_2 e formam o Grupo 1. Os casos com valores iguais ou superiores a 10,43 vão ter o valor 2 na variável AUSEN_2 e formam Grupo 2.

Este método só é aceitável se a distribuição da variável contínua (AUSÊNCIA) for aparentemente simétrica.

Para efectuar este tipo de agrupamento de casos use os seguintes comandos do SPSS:

```

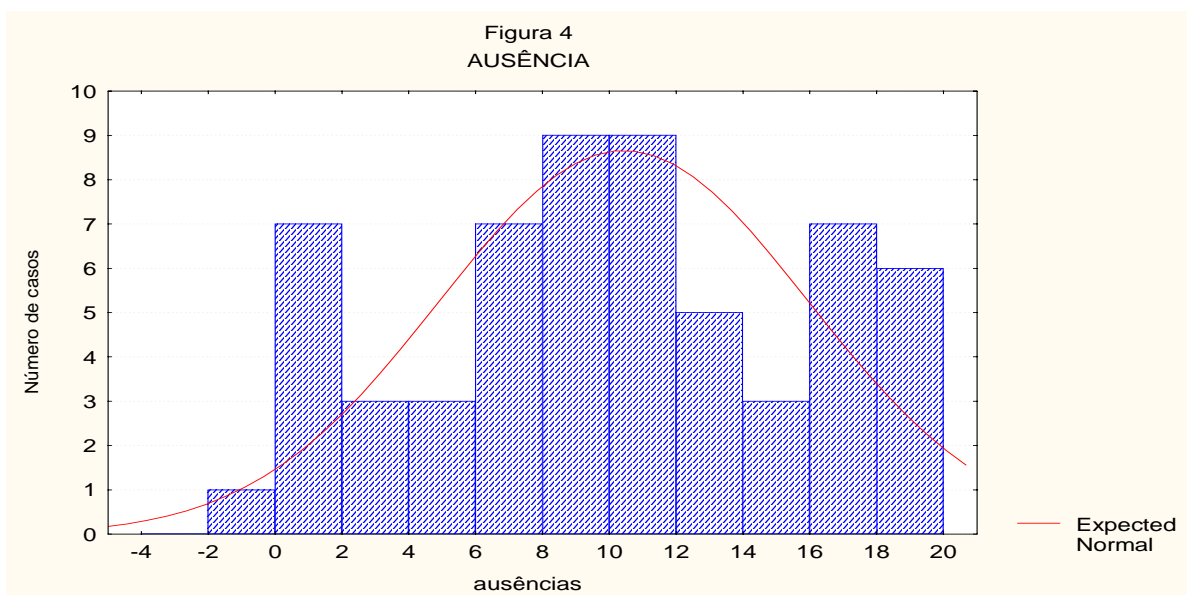
TRANSFORM
  RECODE
    INTO DIFFERENT VARIABLE
      input variable = (neste exemplo, AUSÊNCIA)
      output variable = (neste exemplo, AUSEN_2)
    OLD AND NEW VALUES
    
```

➔ **Partição pela mediana.**

Este método é análogo ao método (a) mas utiliza a mediana em vez do valor médio da variável contínua. A mediana da variável AUSÊNCIA é 10,16. Portanto, os casos no Grupo 1 vão ter valores de AUSÊNCIA inferiores a 10,16 e os casos no Grupo 2 vão ter valores iguais ou superiores a 10,16.

Este método é mais útil do que o anterior porque pode ser utilizado quando a distribuição da variável contínua (AUSÊNCIA) é simétrica ou assimétrica.

Infelizmente tanto o primeiro método como o segundo trazem problema. Olhemos para a figura 4 que apresenta a distribuição da AUSÊNCIA, e recordemos que o valor da mediana igual a 10,16. A matriz de dados apresentada em Anexo revela que o caso 3, que pertence ao Grupo 1, tem o valor 10 na variável AUSÊNCIA e que, por exemplo, o caso 14, que pertence ao Grupo 2, tem o valor 11 na variável AUSÊNCIA. A diferença de AUSÊNCIA entre os dois casos é pequena (11 – 10), ou seja, 1 dia de ausência. **Mas os dois casos pertencem a grupos diferentes!** Ainda pior, o valor médio da variável AUSÊNCIA para os casos do Grupo 1 é 5,96, e o valor médio para os casos do Grupo 2 é 14,92, portanto a diferença entre o caso 3 e o caso “típico” do Grupo 1 (o caso médio do grupo1) é (10 – 5,96) = 4,04, e esta diferença é maior que a diferença entre os casos 3 e 14 que pertencem a grupos diferentes!. O caso 3 é pouco representativo dos casos do Grupo 1 em termos de AUSÊNCIA (e o caso 14 é pouco representativo dos casos do Grupo 2). Infelizmente casos como o 3 e 14 não são casos isolados, há vários outros casos que mostram este problema da falta de representatividade.



Para resolver esta problema é possível usar outro método para efectuar a partição dos dados.

➔ **O método de “Grupos Extremos”**

Neste método vamos retirar um conjunto de casos no meio de distribuição. Para isso é preciso estabelecer dois valores da variável contínua (AUSÊNCIA) como valores de *cut-off*. Isto pode ser feito de duas maneiras:

Método 1. Quando a distribuição for relativamente simétrica

Valor inferior de *cut-off* = Valor médio (ou mediana) - 0,25 desvios padrão (d.p.)

Valor superior de *cut-off* = Valor médio (ou mediana) + 0,25 desvios padrão (d.p)

O Grupo 1 é formado por todos os casos com valores abaixo do valor inferior de *cut-off*.

O Grupo 2 é formado por todos os casos com valores acima do valor superior de *cut-off*.

No nosso exemplo, o desvio padrão da variável AUSÊNCIA é 5,53, portanto tem-se:

▪ *Usando o valor médio (10,43)*

$$\text{O valor inferior de } cut-off = 10,43 - (5,53 \times 0,25) = 9,048$$

$$\text{O valor superior de } cut-off = 10,43 + (5,53 \times 0,25) = 11,812$$

▪ *Usando a mediana (10,16)*

$$\text{os } cut-offs \text{ são } 10,16 - (5,53 \times 0,25) = 8,778$$

$$10,16 + (5,53 \times 0,25) = 11,543$$

Método 2. Quando a distribuição é assimétrica

$$\text{O valor inferior de } cut-off = \text{Mediana} - (0,25 \times \text{intervalo interquartil}) = 8,165$$

$$\text{O valor superior de } cut-off = \text{Mediana} + (0,25 \times \text{intervalo interquartil}) = 12,155$$

São duas as vantagens do método de partição pelos Grupos Extremos. Por um lado, este método evita situações semelhantes às dos casos 3 e 14 no exemplo, e por outro, pode aumentar o poder de uma análise onde se pretenda comparar os grupos de casos. (Se existir uma diferença “genuína” entre os grupos de casos, é mais provável que a análise a consiga encontrar). Mas também há desvantagens: por um lado a amostra fica mais reduzida do que a amostra inicial, e por outro, após a eliminação dos casos intermédios, pode acontecer que a amostra não seja representativa do Universo.

Portanto, se usar o método dos “Grupos Extremos”, e tiver uma amostra aleatória estratificada escolhida para ser representativa de um Universo, é essencial verificar que, depois de eliminar os casos, a amostra ainda fica representativa do Universo. (Se não o for, elimine menos casos e verifique mais uma vez a representatividade).

6.2. Problemas com as correlações paramétricas (Tipo Pearson 'r')

6.2.1.O problema da não-linearidade

Quando estamos a analisar dados deparamo-nos por vezes com uma correlação pequena entre duas variáveis quando a teoria prevê a existência de uma correlação maior. Nesta situação é importante testar a relação entre as variáveis (isto faz-se com a opção "Graphs" no SPSS). Será que a relação entre as variáveis é não-linear?.

Se a relação for não-linear tem duas opções:

- Transformar uma, ou as duas variáveis, de modo que a relação entre elas passe a ser linear;

(Note que às vezes esta transformação pode complicar a interpretação da correlação...).

- Se a relação não-linear tiver uma forma relativamente regular, ajuste uma curva. (Para fazer isto use as opções STATISTICS / REGRESSION / CURVE ESTIMATION no SPSS. O comando CURVE ESTIMATION calcula automaticamente uma "correlação" não-linear).

Note que teoricamente a coeficiente de correlação do Pearson precisa de uma distribuição bivariada normal que mostre homogeneidade e um comportamento linear. Mas o mais importante destes pressupostos é a linearidade do comportamento.

6.2.2.O significado estatístico e o significado prático do coeficiente de correlação

Numa amostra grande, o coeficiente de correlação pode ter significado estatístico aceitável mas pouco significado prático. Afirmar isto é o mesmo que dizer que se o investigador recolher uma segunda amostra de dados (do mesmo Universo), é muito provável que o novo coeficiente de correlação tenha um valor semelhante ao valor da primeira amostra. Podemos assim dizer que o coeficiente de correlação da primeira amostra **é fiável**. Mas nem uma nem outra das correlações permitiria ao investigador predizer, com exactidão, o valor de uma das variáveis para um dado caso (observação) com base no valor da outra variável. Por exemplo, para uma amostra de 400 casos, um coeficiente de correlação (r) entre as variáveis A e B igual a 0,1 é **estatisticamente significativo** ($p < 0,05$), mas a percentagem da variância dos valores de A explicada pela variância dos valores de B só é $r^2 \times 100 = [(0,1)^2 \times 100] = 1\%$. Este valor não tem **significado prático**.

Geralmente é mais importante avaliar o significado estatístico de um coeficiente de correlação, mas no caso em que o investigador quer usar uma das variáveis para prever valores noutra variável, é importante avaliar o significado prático também.

6.3. Problemas relacionados com o Qui-quadrado (χ^2)

6.3.1. Que Qui-quadrado utilizar?

Nos dados apresentados em Anexo há duas variáveis dicotomizadas --- SEXO e AUSEN_2. Imagine que usamos o programa SPSS para calcular o valor do Qui-quadrado para estas variáveis. O output obtido encontra-se nos quadros 11 e 11 (a).

Quadro 11.

Cruzamento Sexo x AUSEN2

			Ausen2		Total
			AUSÊNCIA ABAIXO DA MEDIANA	AUSÊNCIA ACIMA DA MEDIANA	
SEXO	MASCULINO	Count	16	18	34
		Expected Count	17,0	17,0	34,0
	FEMININO	Count	14	12	26
		Expected Count	13,0	13,0	26,0
TOTAL		Count	30	30	60
		Expected Count	30,0	30,0	60,0

Quadro 11(a)

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,271 ^b	1	,602		
Continuity Correction ^a	,068	1	,794		
Likelihood Ratio	,272	1	,602		
Fisher's Exact Test				,795	,397
N of Valid Cases	60				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 13,00.

O Quadro 11 (a) apresenta três tipos (e três valores) de Qui-quadrado -- *Pearson* (que é o Qui-quadrado "normal"), o *Likelihood ratio* e o *Yates continuity correction*. Qual deles usar? Neste exemplo, os três tipos conduzem à mesma conclusão, mas por vezes podem conduzir a conclusões diferentes.

Normalmente usamos o valor do Qui-quadrado de *Pearson*, mas este valor e o do "Likelihood ratio" são quase idênticos, e quando a dimensão da amostra aumenta ficam cada vez mais idênticos. A vantagem do *Likelihood ratio* é que é

menos influenciado pelo tamanho da amostra, e portanto no caso de amostras pequenas (por exemplo, menos que 25 casos) é melhor usar o *Likelihood ratio*.

A maioria dos livros de estatística mais antigos referem que, no caso de o valor do Qui-quadrado ser calculado sobre uma tabela de contingência 2 x 2, é importante usar o *continuity correction*. Mas isto não é bem verdade! Só é preciso usar o *continuity correction* quando os totais marginais na tabela são fixos e não variam de uma amostra para outra. No âmbito da investigações em Ciências Sociais isso é pouco provável e portanto, geralmente, não é preciso usar o *continuity correction*.

6.3.2. O problema das frequências esperadas pequenas

A nota (b) colocada por baixo do Quadro 11(a) indica “0 cells (0%) have expected count less than 5”. Isto pode ser comprovado no Quadro 11 que apresenta as frequências esperadas 17, 17, 13. Aqui não há problema pois todos os valores são superiores a 5. Mas por vezes uma ou mais células têm frequências esperadas inferiores a 5, e quando 20% (ou mais) das células têm frequências esperadas inferiores a 5, o valor do Qui-quadrado não é valido. Para corrigir este problema, é possível usar o teste exacto de Fisher, mas infelizmente este teste só se aplica aos cruzamentos do tipo 2 x 2. No caso de termos cruzamentos com mais de duas categorias, pelo menos numa das variáveis, é possível juntar níveis (categorias) de uma ou das duas variáveis. Por exemplo, se tivermos 5 categorias para a variável “IDADE” (menos que 21, 21 –30, 31-40, 41 – 50, e mais que 50) é possível reduzir o número de categorias até 3 (por exemplo, menos que 21, 21 – 40. mais que 40). Esta redução aumentará os valores das frequências esperadas em cada célula e portanto pode ajudar a resolver o problema. Mas se quiser fazer uma tal redução -- **cuidado!** É importante reduzir o número de categorias de modo a ficar com um conjunto de categorias que **façam sentido**.

6.3.3. Qui-quadrado e percentagens

Nunca aplique o teste do Qui-quadrado a percentagens! Para visualizar a situação considere o seguinte exemplo:

Um estudo sobre a satisfação dos 65 clientes dos hotéis com duas (**) ou quatro (****) estrelas produziu os dados do quadro 12.

O valor do Qui-quadrado para este quadro é igual a 2,229, $P > 0,10$. Obviamente não existe relação significativa entre satisfação dos clientes e tipo de hotel. Se transformamos as frequências do Quadro 12 em percentagens (em coluna, em linha e total) temos o quadro 12 (a).

Quadro 12

	TIPO DE HOTEL		TOTAL
	****	**	
Número de clientes Satisfeitos	12	18	30
Número de clientes Insatisfeitos	8	27	35
TOTAL	20	45	65

Quadro 12.a

	(% EM COLUNA)		(% EM LINHA)		(% TOTAL)	
	Tipo de hotel		Tipo de hotel		Tipo de hotel	
	****	**	****	**	****	**
% Sat.	60	40	40	60	19	28
% Insat	40	60	23	77	12	41
Qui-quadrado=	8,0		6,78		3,44	
	(p < ,01)		(p < ,01)		(p < .06)	

Segundo os valores do Qui-quadrado calculados a partir das percentagens encontramos relação significativa entre a satisfação dos clientes e o tipo de hotel, mesmo no caso da (% total) onde a significância da relação é fraca (6%). É evidente que as conclusões para os três tipos de percentagem são muito diferentes da conclusão tirada da análise das frequências do Quadro 12. Repetimos -- **Nunca aplique o teste do Qui-quadrado a percentagens!**

6.3.4. Um Qui-quadrado fora do usual

O Qui-quadrado é um teste de hipótese nula. A técnica compara um conjunto de frequências observadas com um conjunto de frequências esperadas com base na hipótese nula. Normalmente as frequências esperadas são as frequências esperadas ao acaso, mas é possível usar frequências esperadas com base numa teoria.

Por exemplo, tomemos os dados do Anexo e suponhamos que tínhamos uma teoria que nos diz que 25% dos empregados estariam pouco satisfeitos com o Chefe, 50% estariam razoavelmente satisfeitos e 25% estariam muito satisfeitos. Sendo a amostra de 60 casos, 25% corresponde a 15 empregados e 50% a 30 empregados. Estes valores podem ser usados como sendo as frequências esperadas. As frequências observadas são calculadas a partir da matriz de dados no Anexo. Estas frequências são apresentadas no quadro seguinte:

Quadro 13.

	Pouco satisfeito	Razoavelmente satisfeito	Muito satisfeito
Frequências observadas	11	20	29
Frequências esperadas	15	30	15

Qui-quadrado = 17,467, g.l. = 2, p < 0,001

Neste caso verifica-se que a distribuição observada é significativamente diferente da distribuição esperada com base na teoria – há muito mais empregados satisfeitos com o Chefe do que o investigador esperava com base na teoria. Parece pois que o investigador precisa de outra teoria!

6.4. Problemas relacionadas com o Teste 't'

Há vários tipos da Teste t

- O teste t para uma só amostra
- O teste t para duas amostras independentes
- O teste t para duas amostras emparelhadas

Nesta Secção só vamos tratar de problemas dos últimos dois tipos do Teste t porque estes são os tipos mais vulgarmente utilizados em ciências sociais aplicadas.

6.4.1. O Teste 't' para duas amostras independentes: o problema das variâncias desiguais

O objectivo do teste t para duas amostras independentes é calcular uma estatística t para testar a hipótese nula que a diferença entre os dois valores médios de uma variável dependente obtidos a partir de duas amostras independentes (ou seja, dois grupos de casos independentes) seja zero. Por exemplo, no caso dos dados do anexo, é possível testar se a AUSENÇA média difere significativamente entre trabalhadores do sexo masculino e trabalhadores do sexo feminino.

O problema do Teste t é que exige que a variância da variável dependente não difira significativamente entre as duas amostras (no exemplo, homens e mulheres). No caso em que este pressuposto não é válido é preciso calcular o valor da estatística t de outra maneira.

O SPSS aplica automaticamente um teste de igualdade das variâncias (o teste F de Levene), e também calcula automaticamente dois valores da estatística t – o primeiro para utilizar quando o pressuposto de variância é válido, e o outro para utilizar quando o pressuposto não é válido. Para escolher o valor da estatística t adequado, o investigador precisa de olhar para a significância da estatística F apresentada pelo SPSS como parte do “output”. Se o valor de F for significativo ($p < 0,05$), é preciso utilizar o valor de t associado com a expressão “*Equality of variances not assumed*” no SPSS “output”. Se o valor de F não for significativo, utilize o valor de t associado com a frase “*Equality of variance assumed*”.

6.4.2. O Teste 't' para duas amostras emparelhadas

Neste contexto “amostras emparelhadas” são amostras **relacionadas**, ou seja, não independentes. Por exemplo, imagine que em Janeiro 1998 aplicámos um questionário a uma amostra de 100 pessoas para medir a Satisfação Global no emprego. Em Junho 1998 aplicámos a mesma medida às mesmas pessoas mais uma vez. Nesta investigação temos **uma variável dependente** (Satisfação Global) e **uma variável independente** (momento de aplicação do questionário). A variável independente só tem dois níveis (valores) -- Janeiro e Junho. No fim temos duas amostras de dados (valores de Satisfação Global) -- uma amostra retirada em Janeiro e outra retirada em Junho. Mas as duas amostras de dados não são independentes -- são **relacionadas** porque foram fornecidas pelas mesmas pessoas, o que implica que é provável que os dados retirados em Janeiro estejam significativamente **correlacionados** com os dados retirados em Junho. Por causa de não serem independentes, as duas amostras de dados não podem ser analisadas por meio de um Teste t para amostras independentes. Temos de aplicar o **Teste t para amostras emparelhadas** (também chamada “amostras relacionadas”, “amostras correlacionadas” ou “amostras dependentes”).

Na aplicação do Teste t para amostras emparelhadas o objectivo é testar a hipótese nula que a diferença entre os dois valores médios da variável dependente (Satisfação global) é igual a zero. No nosso exemplo, seria testar que a satisfação média global não diferiu entre Janeiro e Junho de 1998.

O teste t para amostras emparelhadas é aplicável quando:

- Um só grupo de casos dá valores da variável dependente duas vezes (isto é, em dois momentos ou em duas situações diferentes).

ou

- Dois grupos de casos, relacionados de certa forma, dão valores da variável dependente. Nesta situação é essencial que cada um dos casos no Grupo 1 esteja relacionado com um só caso no Grupo 2. Por exemplo, o Grupo 1 é um grupo de homens e o Grupo 2 é formado pelas mulheres dos homens. Mas a relação entre os casos dos dois Grupos pode ser da outra forma. Por exemplo, os casos podem estar seleccionados de modo que o nível de inteligência de cada um dos casos do Grupo 1 seja igual ao nível de inteligência de um só caso do Grupo 2. Nesta situação os casos dos dois grupos estão relacionados em termos de inteligência.

O problema com o Teste t para amostras emparelhadas é que precisa de uma codificação dos dados fora do usual no SPSS. Para utilizar o SPSS correctamente é essencial que os níveis da variável independente sejam representados como variáveis diferentes no ficheiro de dados. Por exemplo, a variável independente do nosso exemplo é “Momento de aplicação do questionário”, mas no ficheiro de dados temos de colocar os níveis desta variável (Janeiro e Junho) como duas variáveis. Os valores destas “variáveis” para cada caso são os valores da Satisfação Global. Não é adequado criar uma variável “Momento” no ficheiro de dados porque é impossível codificar os dois valores da Satisfação Global numa tal variável.

É fácil esquecer que o Teste t para amostras emparelhadas precisa de uma codificação fora do normal, e é muito fácil abusar deste Teste -- como vamos demonstrar na próxima Secção.

6.4.3. O Teste ‘t’ do “Tolo”

As opções para usar a teste t para amostras emparelhadas no SPSS são:

STATISTICS
COMPARE MEANS
PAIRED-SAMPLE T TEST

Quando são seleccionadas estas opções, o computador mostra um painel para seleccionar as variáveis, e o investigador tem de fazer entrar duas variáveis. Não há problemas se o ficheiro de dados estiver bem construído para fazer o Teste t para amostras emparelhadas e se o investigador perceber o significado do teste .

Mas considere as variáveis do Anexo “IDADE” e “ANTIGUIDADE NA EMPRESA”. Um investigador pode usar estas variáveis para fazer uma Teste t para amostras emparelhadas pensando que um só grupo de casos deu valores para estas variáveis e portanto é preciso um Teste t para amostras emparelhadas.

Este Teste t é aquilo que designamos por um Teste t do “Tolo” porque as variáveis IDADE e ANTIGUIDADE NA EMPRESA são duas variáveis dependentes totalmente diferentes. É pouco interessante saber que o valor médio da IDADE é significativamente maior que o valor médio da ANTIGUIDADE NA EMPRESA. (Seria impossível encontrar o contrário!).

Se não compreender isto imagine que as variáveis eram PESO e ALTURA. Provavelmente são poucas as pessoas que querem investigar se, num amostra de empregados, os empregados são, em média, mais pesados do que altos. Seria uma investigação ridícula. Daí que é preciso muita cautela ao escolher o tipo de Teste t para analisar os seus dados.

6.4.4. Teste ‘t’ e comparações múltiplas

Por vezes o investigador pretende comparar dois grupos de casos em termos de um conjunto de variáveis usando o teste t. Por exemplo, no caso dos dados apresentados em Anexo suponha que queríamos comparar dois grupos de empregados -- Grupo 1 (Ausência baixa) e Grupo 2 (Ausência alta) em termos de um conjunto de oito variáveis dependentes: Antiguidade na empresa, Idade e seis medidas de satisfação (satisfação com Trabalho, Ordenado, Chefe, Colegas, Administração e Ambiente no trabalho). Para agrupar os casos vamos usar o método (b) da secção 6.1, portanto os casos do Grupo 1 têm valores de AUSÊNCIA abaixo da mediana e os casos do Grupo 2 têm valores de AUSÊNCIA acima da mediana. O quadro 14 apresenta os valores médios das variáveis dependentes.

Quadro 14. Aplicação do Teste “t” para comparações múltiplas

Variável	Grupo		t	Significância do t
	ausência baixa	ausência alta		
Antiguidade na empresa	14,50	9,47	2,083	p = 0,042
Idade	46,29	35,20	3,764	p < 0,0001
Trabalho	2,37	1,70	3,912	p < 0,0001
Ordenado	2,27	1,77	3,060	p = 0,003
Chefe	2,63	1,97	3,721	p < 0,0001
Colegas	2,43	1,70	4,269	p < 0,0001
Administração	2,13	1,57	3,113	p = 0,003
Ambiente	2,17	1,80	2,257	p = 0,028

Quando pretendemos aplicar um conjunto de testes t há três passos a seguir:

- 1º Comparar os dois grupos usando o teste T^2 de **Hotelling**. Este é um teste multivariado para comparar os dois grupos em termos do **conjunto inteiro** das variáveis dependentes. Um valor significativo de T^2 indica que pelo menos uma, e talvez mais, das variáveis dependentes é significativa, e neste caso podemos passar ao passo 2. Se o valor de T^2 não for significativo não há nada mais a fazer porque nenhuma das variáveis dependentes pode ser considerada significativa.
No caso dos dados do exemplo, $T^2 = 47,974$ ($p = 0,00009$), portanto continuamos para o passo 2
- 2º Fazer um conjunto de testes t **com correcção Bonferroni**.
Porque é que precisamos de fazer um correcção, e o que é a correcção Bonferroni?
O teste t é um teste de hipótese nula (os valores médios dos dois grupos são iguais) e normalmente queremos rejeitar esta hipótese ao nível de $p = 0,05$. Com base no quadro 14, que também apresenta os valores do teste t para todas as variáveis, parece que podemos rejeitar a hipótese nula para todas as variáveis. Mas não é bem assim. O valor 0,05 é a probabilidade de rejeitar a hipótese nula quando esta hipótese é verdadeira, e portanto o valor 0,05 (o nível de significância) **é um nível (probabilidade) de erro**. Tecnicamente chama-se a este erro, erro do Tipo 1. Este erro é a probabilidade de concluirmos, incorrectamente, que os valores médios dos dois grupos não são iguais. Além disso, é a probabilidade de erro quando fazemos **só um** teste t. Quando aplicamos uma “**família**” (conjunto) de testes, a probabilidade de erro aumenta porque há uma probabilidade de erro associada com cada um dos testes. Se usarmos ‘k’ testes, e quisermos utilizar o nível de significância

$\alpha = 0,05$, a probabilidade de pelo menos um dos testes dar conclusões erradas é α' (o que é chamada “a taxa do erro da família”)

$$\text{A taxa de erro da família} = \alpha' = 1 - (1 - \alpha)^k$$

No caso do nosso exemplo em que há oito variáveis tem-se:

$$\alpha' = 1 - (1 - 0,05)^8 = 0,337$$

Esta probabilidade de erro não é aceitável e para **manter a taxa de erro da família** ao nível de $p = 0,05$ é preciso fazer uma correcção.

A correcção mais vulgarmente utilizada é a correcção de Bonferroni . Esta correcção é muito simples, basta dividir a taxa de erro da família que queremos manter pelo número de testes. No nosso exemplo, queremos aplicar o teste t a oito variáveis e pretendemos usar uma taxa de erro da família $p = 0,05$. Portanto calculamos $0,05 / 8 = 0,00625$. Para ser considerado significativo ao nível $p = 0,05$ o valor do t em cada um dos testes deve ser significativo ao nível $p = 0,00625$, e isto precisa de um valor $t = 2,84$ (**vide** as Tabelas de Dunn³).

Repare-se que pelo quadro 14, todas as variáveis dependentes têm valores de t significativos antes de fazer a correcção. Efectuada a correcção, **duas das variáveis dependentes têm valores de t não significativos** (Antiguidade na empresa e Satisfação com o ambiente no trabalho).

3º Verifique as correlações entre as variáveis dependentes.

No caso ideal estas variáveis não devem estar significativamente correlacionadas. Para o nosso exemplo, o quadro 15 revela que todas as variáveis ligadas com satisfação estão correlacionadas entre si, mas nenhuma delas tem correlação elevada com a variável Idade. É então razoável pensar que os vários tipos de satisfação representam aspectos diferentes da **uma variável** -- “Satisfação Global no emprego” (**vide** Quadro 6 na secção sobre “validade factorial”). Portanto é possível que os resultados dos testes t, mesmo com correcções de Bonferroni, sejam artificiais. Talvez os grupos “Ausência Alta” e “Ausência Baixa” só difiram em termos de Satisfação Global (e Idade por não estar correlacionada com as vários tipos de satisfação).

Quadro 15.

Correlations

	TRABAL	ORDENADO	CHEFE	COLEGAS	ADMIN	AMBIENTE
Idade	,135	,212	,159	,273	,130	,109
TRABAL	1,000	,476	,403	,453	,650	,284
ORDENADO		1,000	,383	,462	,470	,193
CHEFE			1,000	,375	,489	,452
COLEGAS				1,000	,522	,278
ADMIN					1,000	,305
AMBIENTE						1,000

Numa situação como esta em que muitas das variáveis dependentes estão correlacionadas entre si seria melhor realizar uma análise factorial com estas variáveis e aplicar o teste t aos valores dos factores. A análise factorial (tipo alpha)

³ Estas tabelas são as tabelas dos valores críticos para o teste de comparações múltiplas Bonferroni. Ver por exemplo em Howell, D.C. (1992). *Statistical Methods for Psychology*, Duxbury Press, An Imprint of Wadsworth Publishing Company, Belmont, California.

aplicada às seis medidas de satisfação produziu os valores (*scores*) do factor “Satisfação Global” para todos os casos. O valor médio dos *scores* para cada grupo de trabalhadores é apresentado no Quadro 16.

(Note que a natureza do factor “Satisfação Global” é apresentada no Quadro 6)

Quadro 16.

Variável	Grupo		t	Significância de t
	ausência baixa	ausência alta		
Satisfação Global	0,508	-0,508	5,142	p < 0,0001
Idade	46,29	35,20	3,764	p < 0,0001

O teste T^2 de Hotelling aplicado ao conjunto das variáveis Satisfação Global e Idade dá um valor $T^2 = 42,364$ ($p < 0,000001$). Portanto, os grupos de “Ausência Baixa” e “Ausência Alta” diferem significativamente em termos **do conjunto** destas variáveis. Para fazer uma correcção de Bonferroni (com taxa da família $p = 0,05$) calcula-se $0,05 / 2 = 0,025$. Ressalta do quadro 16 que as duas variáveis, Idade e Satisfação Global, são significativas após a correcção (ambos os valores da significância são inferiores a 0,025).

6.5. Análise de componentes principais versus Análise factorial

Estas técnicas parecem semelhantes mas têm objectivos diferentes. A **análise de Componentes Principais** tem como objectivo principal a análise da **variância total** de cada um das variáveis num conjunto de variáveis. Teoricamente a variância total de cada variável pode ser decomposta em três partes:

- **Variância comum** (variância partilhada com as outras variáveis do conjunto com as quais a variável está correlacionada).
- **Variância específica** (variância que só pertence à variável e que não é partilhada com as outras variáveis do conjunto).
- **Variância associada com o erro**

As três partes são aditivas, ou seja:

$$\text{Variância Total} = \text{Variância comum} + \text{Variância específica} + \text{Variância erro}$$

Sendo uma análise de variância total, a técnica de Componentes Principais é uma análise que inclui variância associada com o erro.

Teoricamente a correlação entre duas variáveis só depende da variância comum, e o objectivo principal da **Análise Factorial** é analisar a **variância comum** num conjunto de variáveis para entender, ou “explicar”, as correlações entre essas variáveis.

É possível usar as duas técnicas para agrupar variáveis em termos de Componentes ou Factores, e é possível usar as duas técnicas para reduzir a complexidade de um conjunto de dados. Mas quando as medidas das variáveis não têm fiabilidade perfeita (e portanto as medidas contêm “erro” -- o que é normal no caso de variáveis tais como opiniões,

atitudes e satisfações) é melhor **não analisar a variância associada com o erro**. Portanto nestes casos é preferível usar a técnica de Análise Factorial.

Não vamos dar conselho prático sobre as opções no SPSS para fazer uma Análise Factorial porque há vários tipos de Análise Factorial e muitas opções. Se quiser fazer uma tal análise é melhor pedir conselho a um profissional de estatística.

BIBLIOGRAFIA

- BABBIE, E. & HALLEY, F. (1995) *Adventures in Social Research: Data analysis using SPSS for Windows*. Thousand Oaks California: Pine Forge Press.
- BRYMAN, A. & CRAMER, D. (1993) *Análise de Dados em Ciências Sociais. Introdução às Técnicas Utilizando o SPSS*. Oeiras: Celta.
- CAMPBELL, D.T. & FISKE, D.W. (1959) Convergent and discriminant validation by the multitrait – multimethod matrix. *Psychological Bulletin*, 56, 81 – 105.
- EDWARDS, A.L. (1957) *The social desirability variable in personality assessment and research*. New York: Dryden
- HILL, A.B. (1986) The influence of social desirability on the Ai3 measure of anal character. *British Journal of Medical Psychology*, 59, 105 – 109.
- HILL, M. M. & HILL, A.B. (1998) *A construção de um questionário*. Lisboa: Dinâmia.
- HOWELL, D.C. (1992) *Statistical methods for Psychology* 3rd edition. Belmont, Ca.: Duxbury press.
- MOSER, C.A. & KALTON, G. (1986) *Survey Methods in Social Investigation*. Aldershot, U.K.: Gower Publishing Company.
- MURPHY, K.R. & DAVIDSHOFER, C.O. (1988) *Psychological Testing: Principles and Applications*. Englewood Cliffs, NJ: Prentice – Hall International Inc.
- REIS, E., MELO, P., ANDRADE, R. & CALAPEZ, T. (1996) *Estatística Aplicada*. Lisboa: Edições Sílabo, LDA.

ANEXO

O Director do Departamento de Produção queixa-se que os limites de produção não estão a ser atingidos por causa do elevado absentismo observado na empresa e pede ao Director de Recursos Humanos para estudar o problema e analisar em que medida este absentismo pode ser reduzido.

A partir de uma lista de possíveis factores relacionados com o absentismo, o Director de Recursos Humanos aplicou um pequeno questionário aos 60 trabalhadores do departamento de Produção.

As primeiras perguntas desse questionário diziam respeito ao sexo, idade do trabalhador, antiguidade na empresa e secção onde trabalha e as seis últimas perguntas procuravam medir o grau de satisfação sobre vários aspectos do emprego. Recolheu ainda o número de dias no último ano que o trabalhador se ausentou do serviço sem justificação. A partir destes dados construiu um ficheiro com as seguintes variáveis:

SECÇÃO:	(1, 2 ou 3)
SEXO:	(1 - Masculino; 2 – Feminino)
ANTIGUIDADE:	Nº de anos de trabalho na empresa
IDADE	Idade do trabalhador
TRABALHO:	Grau de satisfação com o trabalho que desempenha
ORDENADO:	Grau de satisfação com o ordenado
CHEFE:	Grau de satisfação com o chefe
COLEGAS:	Grau de satisfação com os colegas
ADMINIST:	Grau de satisfação com a administração
AMBIENTE:	Grau de satisfação com o ambiente no trabalho
AUSÊNCIA:	Nº de dias/ano de ausência ao trabalho sem justificação

As variáveis TRABALHO, ORDENADO, CHEFE e COLEGAS ADMINIST e AMBIENTE foram codificadas numa escala de 1 a 3, onde 1 - pouco satisfeito, 2 – razoavelmente satisfeito e 3 – muito satisfeito.

A variável AUSEN2 apresentada na última coluna do ficheiro de dados é uma recodificação da variável AUSÊNCIA em duas categorias: 1 - ausência baixa (abaixo da mediana); 2 - ausência alta (acima da mediana).

FICHEIRO DE DADOS

Caso	Secção	Sexo	Antiguid	Idade	Trabalho	Ordenad o	Chefe	Colegas	Admin	Ambiente	Ausência	Ausen2
1	1	2	3	32	3	3	3	3	2	2	4	1
2	3	1	10	38	2	3	2	1	1	2	17	2
3	2	2	8	45	2	2	3	2	1	2	10	1
4	3	1	5	59	2	2	1	2	1	2	7	1
5	3	2	32	57	2	2	3	2	1	2	3	1
6	2	1	8	28	3	3	2	2	2	1	19	2
7	1	1	11	40	2	2	3	2	2	3	4	1
8	1	2	4	46	3	3	3	3	2	2	8	1
9	3	2	1	19	3	2	2	2	2	3	12	2
10	2	1	5	52	2	2	2	1	1	2	11	2
11	1	2	3	43	3	2	3	1	3	1	14	2
12	3	2	5	37	2	2	3	3	2	2	8	1
13	2	1	8	22	2	3	3	2	1	2	15	2
14	1	1	17	49	2	2	3	3	2	2	11	2
15	3	2	8	60	3	3	3	3	3	2	0	1
16	1	2	2	21	1	2	2	1	1	2	18	2
17	2	1	14	46	2	2	1	2	1	2	8	1
18	2	2	9	36	2	1	1	2	1	1	17	2
19	2	2	14	42	3	3	3	2	3	2	5	1
20	3	1	15	56	2	3	3	3	3	2	10	1
21	1	2	-1	-1	3	2	1	3	2	1	7	1
22	2	2	30	54	1	1	2	1	1	2	11	2
23	1	2	7	39	2	2	3	2	2	3	2	1
24	1	1	50	63	2	2	3	3	2	3	10	1
25	3	1	18	38	2	2	2	2	2	3	11	2
26	1	2	-1	-1	2	2	2	3	2	2	9	1
27	3	2	25	44	2	1	2	1	1	2	8	1
28	2	1	1	20	1	2	2	2	1	1	20	2
29	2	1	15	41	1	2	3	2	1	2	9	1
30	3	1	3	30	3	3	3	3	3	3	18	2
31	3	1	20	33	3	3	3	2	3	3	8	1
32	3	2	15	45	1	2	3	2	2	1	10	2
33	1	1	5	24	1	1	2	1	2	3	19	2
34	2	1	15	54	1	2	1	1	1	1	15	2
35	1	2	8	25	3	2	3	2	3	2	9	1
36	3	2	1	18	2	1	1	1	1	1	19	2
37	3	1	7	41	2	2	3	3	3	2	5	1
38	2	1	4	51	3	1	3	2	2	3	2	1
39	3	2	30	69	2	2	3	2	2	2	9	1
40	3	1	14	33	1	2	1	2	1	2	14	2
41	2	1	15	45	2	2	3	3	2	3	6	1
42	3	1	24	46	3	2	3	3	3	2	2	1
43	1	2	8	34	1	1	2	1	1	2	17	2
44	3	2	3	22	1	1	1	1	1	1	16	2
45	1	1	8	31	2	1	3	1	2	2	11	2
46	2	1	13	30	1	1	2	1	1	1	20	2
47	2	1	9	34	2	2	2	2	2	1	9	1
48	3	1	15	44	3	3	3	3	3	2	2	1
49	1	1	9	26	2	2	3	2	2	3	16	2
50	3	1	10	42	2	2	3	2	1	2	2	1
51	1	2	7	30	1	2	2	3	2	2	11	2
52	2	1	8	57	2	3	2	3	3	2	2	1
53	3	1	4	22	1	1	1	1	2	1	19	2
54	3	2	25	49	3	3	2	3	3	3	2	1
55	1	1	4	35	2	1	2	3	2	2	13	2
56	2	2	30	50	1	2	2	2	1	2	12	2
57	1	1	15	53	3	3	3	2	3	2	9	1
58	2	1	6	52	2	2	1	1	2	2	17	2
59	2	1	7	45	3	1	2	3	2	2	11	2
60	2	2	20	55	1	2	1	2	1	1	13	2