

Índice

Prefácio	5
CAPÍTULO 0	
Guia de consulta rápida	7
CAPÍTULO 1	
Introdução à estatística em perguntas	9
Exercícios	23
CAPÍTULO 2	
Introdução ao SPSS	25
Exercícios	35
CAPÍTULO 3	
Estudo da normalidade	45
Exercícios	49
CAPÍTULO 4	
Testes à associação entre variáveis qualitativas	51
Exercícios	58
CAPÍTULO 5	
Avaliação de correlações	64
Exercícios	70
CAPÍTULO 6	
Testes à média ou mediana para uma amostra	74
Exercícios	79
CAPÍTULO 7	
Testes à média ou mediana para duas amostras independentes	81
Exercícios	86
CAPÍTULO 8	
Testes à média ou mediana para duas amostras emparelhadas	90
Exercícios	95
CAPÍTULO 9	
Testes à média ou mediana para três ou mais amostras independentes	98
Exercícios	105
CAPÍTULO 10	
Testes à média ou mediana para três ou mais amostras emparelhadas	110
Exercícios	119
Bibliografia	123



CAPÍTULO 0

Guia de consulta rápida

O presente livro propõe-se a apresentar uma visão integrada de como a bioestatística poderá começar a ser abordada, de forma prática, com o SPSS. Porém, poderá ser do interesse do leitor consultar o livro tendo em mente uma questão concreta que possa ser respondida recorrendo a um teste de hipóteses, dispensando uma consulta integral de todos os capítulos. Para tanto, será útil ter em conta que o livro consiste no seguinte:

- numa parte introdutória, compreendendo os capítulos 1 e 2. O primeiro destes pretende apresentar as noções fundamentais nas quais se baseia o pensamento estatístico – por exemplo: o que é um teste de hipóteses ou como interpretar o valor-p. O segundo capítulo introduz o *software* de análise estatística SPSS, apresentando os comandos mais relevantes;
- num conjunto de capítulos (do 3 ao 10), em que cada um corresponde, de forma autocontida, a uma técnica estatística diferente. Estando o leitor munido dos conhecimentos da parte introdutória, poderá consultar dos capítulos 3 a 10 o(s) que lhe permite(m) responder à(s) sua(s) questão(ões). Sublinhamos que essa consulta pode ser efetuada por qualquer ordem ou até de forma isolada.

Pensando no leitor que pretenda realizar um teste de hipóteses mas dispense a leitura integral do livro, distinguimos quatro níveis diferentes de conhecimento:

- 1) básicos – estes são conhecimentos prévios mínimos, de carácter teórico, que o leitor deverá obrigatoriamente possuir para conseguir realizar e interpretar um qualquer teste de hipóteses;
- 2) como escolher o teste estatístico adequado – estes são conhecimentos de carácter teórico que o leitor terá de possuir para conseguir escolher o teste estatístico adequado à questão de investigação em estudo;
- 3) como realizar, interpretar e reportar o teste estatístico adequado – estes são conhecimentos de carácter prático que o leitor terá de possuir para conseguir realizar o teste estatístico adequado no SPSS, interpretar os resultados obtidos e reportar os mesmos;
- 4) complementares – estes são conhecimentos prévios, de carácter prático, que o leitor poderá sentir necessidade de possuir para conseguir realizar, interpretar e reportar um teste de hipóteses no SPSS.

Na tabela que se segue, inserimos um resumo do exposto. Nesta, associamos a cada nível de conhecimento (1.ª coluna) os conteúdos correspondentes (2.ª coluna) e indicamos onde os mesmos poderão ser encontrados no livro (3.ª coluna). Sugerimos ao leitor que, para cada um dos níveis de conhecimento, verifique se possui as

competências necessárias relativas aos conteúdos correspondentes. Caso contrário, recomendamos que efetue as leituras sugeridas. Relativamente aos conhecimentos complementares, poderá deixar a consulta das leituras sugeridas para os casos em que, no decorrer do estudo, entenda que a mesma seja estritamente necessária.

Conhecimentos	Conteúdos	Leituras sugeridas
Básicos	Distinguir entre variáveis qualitativas nominais, qualitativas ordinais e quantitativas	Capítulo 1: ponto 3
	O que é um teste de hipóteses	Capítulo 1: ponto 17
	Como interpretar o resultado de um teste de hipóteses	Capítulo 1: ponto 21
Como escolher o teste estatístico adequado	Que testes de hipóteses existem	Capítulo 1: ponto 18
	Escolha do teste de hipóteses adequado	Capítulo 1: ponto 19 Capítulo 1: ponto 20
Como realizar, interpretar e reportar o teste estatístico pretendido	Em que situações se deve realizar o teste de hipóteses escolhido	Capítulos 3 a 10
	Como se deverá proceder no SPSS	
	Como deverão ser interpretados os <i>outputs</i> do SPSS	
	Como reportar os resultados	
Complementares	Como formatar uma base de dados de forma adequada	Capítulo 1: ponto 6
	Como são armazenados os dados no SPSS	Capítulo 2: ponto 1
	Onde são apresentados os resultados no SPSS	Capítulo 2: ponto 2
	Como importar dados para o SPSS	Capítulo 2: ponto 3
	Como fazer operações com variáveis no SPSS	Capítulo 2: ponto 4
	Como seleccionar casos no SPSS	Capítulo 2: ponto 5
	Como obter medidas descritivas no SPSS	Capítulo 2: ponto 6
	Como criar tabelas no SPSS	Capítulo 2: ponto 7
	Como criar gráficos no SPSS	Capítulo 2: ponto 8

CAPÍTULO 1

Introdução à estatística em perguntas

Neste capítulo faz-se uma introdução à estatística sob a forma de perguntas e respostas. As questões apresentadas têm por base os fundamentos do pensamento estatístico e pretendem ajudar à compreensão dos conteúdos discutidos nos capítulos subsequentes.

1. Como começa um estudo estatístico? Um estudo começa sempre com uma questão de investigação, à qual se pretende dar resposta. Para se o conseguir, tem de se definir que informação se pretende recolher (variáveis) e qual é a unidade de análise (ver questão 2). Uma vez tendo feito o planeamento de todos os aspetos do estudo estatístico, pode proceder-se à recolha dos dados.

2. O que é uma unidade de análise? A unidade de análise (ou unidade estatística) é a entidade que se examina num estudo estatístico. São exemplos de unidade de análise: pessoas, grupos, objetos (livros, jornais, instrumentos de medida, etc.), unidades geográficas (cidades, países, etc.), partes do corpo humano (pulmão, cérebro, olhos, etc.) ou conceitos sociais (casamentos, divórcios, batizados, etc.). A definição da unidade de análise é geralmente simples e direta, existindo porém situações em que esta se torna mais complexa, como no estudo de órgãos que na maioria das pessoas existem em duplicado (olhos, rins, pulmões, etc.). Neste caso, a unidade de análise pode ser definida como a pessoa em si ou o órgão em causa.

3. Em que diferem as variáveis? As variáveis podem ser classificadas de formas diferentes, sendo dois aspetos importantes a considerar o nível de mensuração da variável e o seu papel no estudo.

– No que respeita ao nível de mensuração, as variáveis podem ser consideradas quantitativas ou qualitativas. Uma variável quantitativa é tal que o atributo que mede se encontra representado por um valor numérico que admite comparações de ordem e cálculos aritméticos. São exemplos de variáveis quantitativas a idade, o peso, o vencimento e o número de visitas num sítio. Uma regra simples de seguir para perceber se uma dada variável é quantitativa é verificar se o conceito de média faz sentido para a variável em questão.

As variáveis qualitativas dizem respeito a atributos que são representados por classes ou categorias, as quais podem ser ordenáveis ou não. No caso destas serem ordenáveis, as variáveis dizem-se ordinais, e são exemplos destas o grau de escolaridade, o estágio de um cancro, o nível socioeconómico e o nível de alerta. Caso as classes ou categorias não admitam ordenação, a variável diz-se nominal. Exemplos de variáveis nominais são o género, o estado civil ou a religião professada.

– No que respeita ao papel num estudo, as variáveis podem ser dependentes ou independentes. Uma variável independente é uma variável que não é condicionada pelas restantes variáveis. Uma variável dependente é tal que a sua variação é influenciada por outras variáveis. Por exemplo, considere-se um estudo da influência do desporto na função pulmonar, em que foram consideradas três variáveis: o fluxo expiratório forçado médio (FEF), a prática de desporto (expressa por duas categorias: «pratica desporto» e «não pratica desporto») e os hábitos tabágicos (expressos por duas categorias: «fuma» e «não fuma»). Neste caso, é de esperar que a prática de desporto e os hábitos tabágicos tenham influência no fluxo expiratório forçado médio. Esta última será uma variável dependente, sendo a prática desportiva e os hábitos tabágicos variáveis independentes.

Em determinadas circunstâncias, deve também ter-se presente que pode haver variáveis relacionadas tanto com as variáveis independentes como com as dependentes que é necessário tomar em conta. Estas variáveis são designadas por confundentes. Um exemplo deste aspeto: é sabido que a sobrevivência de doadores de rim é superior à da população em geral; apenas tendo em conta este conhecimento, concluir-se-ia que para viver mais seria aconselhável ser dador de rim. No entanto, existe uma variável que afeta a sobrevivência e está relacionada com os doadores de órgãos: o controlo clínico apertado a que os doadores ficam sujeitos. Neste exemplo, tem-se uma variável independente (ser ou não dador de rim), uma variável dependente (sobrevivência) e variáveis confundentes (que exprimirão o controlo e acompanhamento clínico, por exemplo: a frequência com que se faz um *check-up* ou análises ao sangue). O tipo de relação que uma variável confundente estabelece com as variáveis independentes e as dependentes pode ser diferente de caso para caso, mas o resultado da sua omissão numa análise estatística é sempre o da deturpação da relação entre as variáveis independentes e dependentes. Por conseguinte, a observação de variáveis confundentes é importante para deduzir a verdadeira relação que possa existir entre a variável dependente e as variáveis independentes.

4. Uma variável representada por um número é uma variável quantitativa? Há variáveis que são registadas numericamente sem que os números representem uma quantidade, mas apenas uma classe ou categoria. Por vezes isto é óbvio, como no caso em que se codifica o género atribuindo um número a pessoas do sexo masculino e outro a pessoas do sexo feminino. Há também exemplos menos intuitivos: o tamanho da roupa na Europa é frequentemente representado por números. Apesar do tamanho ser representado numericamente, tal não significa que a variável seja quantitativa; apenas aceita comparações de ordem. De facto, umas calças de tamanho 44 são maiores do que umas de tamanho 42, pelo que podemos admitir que se trata de uma variável qualitativa ordinal, pois os números 44 e 42 não representam quantidades de algo. Outro exemplo dramático de uma variável numérica que não é quantitativa é o código postal, e, neste caso, considerá-la ordinal também seria injustificado: o que significa, neste caso, dizer que 3000 é superior a 1000?

5. A distinção das variáveis pelo nível de mensuração é importante ou é apenas um detalhe? É muito importante porque as representações possíveis e as análises estatísticas que se podem fazer são diferentes consoante o tipo de variável. Note-se

que, para uma variável nominal, apenas se podem fazer contagens, por exemplo podemos dizer quantos indivíduos do género masculino e do género feminino existem numa amostra, não fazendo sentido, por exemplo, calcular a média para o género. O nível de mensuração também se repercute no tipo de gráficos que se pode fazer. Para variáveis qualitativas nominais, é usual recorrer a gráficos de barras ou outros que representem frequências (absolutas ou relativas), não havendo outras medidas sumárias a considerar. Relativamente a uma variável ordinal, é possível calcular outros tipos de estatísticas, para além das frequências, tais como a mediana, os percentis ou a amplitude interquartil (diferença entre percentil 75 e o percentil 25). Há ainda mais opções de escolha para representar graficamente uma variável ordinal, como, por exemplo, os diagramas de extremos e quartis.

Por último, para variáveis quantitativas, além das frequências ou percentis, também é possível apresentar médias, medianas ou outras estatísticas como o desvio-padrão. Há um grande leque de opções gráficas para representar variáveis quantitativas.

6. Como formatar a base de dados de forma adequada? As variáveis medidas para cada unidade de análise devem ser registadas e organizadas de forma correta para serem posteriormente analisadas. Apesar de existirem várias formas de organizar uma base de dados, a maior parte dos programas informáticos de análise estatística impõe que a base de dados seja organizada numa tabela em que as variáveis estejam em colunas e em que cada linha represente uma unidade de análise. No caso do SPSS, esta é a norma. Por exemplo, no estudo da função pulmonar referido na questão 3, cada linha da tabela representa uma pessoa diferente e existem três colunas: numa regista-se o valor do fluxo expiratório forçado médio (FEF) e nas outras assinalam-se os grupos a que pertencem as pessoas, como representado na tabela seguinte. Nesta última, na coluna «Desporto», distingue-se quem «pratica desporto» (codificado com 1) de quem «não pratica desporto» (codificado com 0) e, na coluna «Hábitos tabágicos», distingue-se quem «fuma» (codificado com 1) de quem «não fuma» (codificado com 0).

Sujeito	FEF	Desporto	Hábitos tabágicos
1	3.85	1	1
2	3.72	0	0
3	3.78	0	1
4	3.93	1	1

Ao registar os dados, é importante ter muito claro quais são as variáveis e quais as unidades de análise. Na questão 2, considerou-se um exemplo em que seria necessário discutir a unidade de análise a considerar, a pessoa ou o olho. Esta decisão determina a organização da base de dados. No caso de se decidir que a unidade de análise é o olho, cada pessoa contribuirá com até duas unidades de análise, ficando cada uma destas unidades registada numa linha. Todavia, se a decisão for no sentido de a unidade de análise ser a pessoa, então todos os dados da pessoa ficarão na mesma linha. Porém, as observações relativas ao olho direito e ao olho esquerdo de uma mesma pessoa deverão ficar registadas em colunas distintas.

Existem algumas regras que devem ser tidas em atenção ao criar uma base de dados, que se resumem seguidamente:

- a base de dados deve ser anónima. Assim, a identificação de cada indivíduo deve ser protegida por códigos que apenas o responsável pelos dados deve conhecer;
- as variáveis nominais e ordinais devem ser codificadas numericamente. Por exemplo, «sim» pode ser substituído pelo valor 1 e «não» pelo valor 0;
- cada variável quantitativa deve ser preenchida sempre na mesma unidade métrica. Por exemplo, no caso de se registar a idade de crianças, estas devem estar todas em meses ou todas em anos;
- os valores em falta devem ser registados usando um código numérico que seja suficientemente diferenciador de tudo o resto (por exemplo, 9999) ou deixando simplesmente a célula correspondente em branco.

Finalmente, um aspeto importante na criação da base de dados é o facto de, simultaneamente, ser indispensável produzir um documento com a legenda das variáveis. Geralmente, este documento denomina-se plano de operacionalização de variáveis (POV).

7. O que é o POV? O plano de operacionalização de variáveis (ou POV) é um documento que tem por objetivo explicar o que é cada uma das variáveis e que coexiste com a base de dados. Pode ser construído de diversas formas, mas uma das mais fáceis e eficazes é por intermédio de uma tabela. As linhas desta tabela correspondem a cada uma das variáveis. Quanto às colunas, estas poderão incluir o nome da variável, a sua descrição, os códigos associados à mesma e o nível de mensuração. A tabela que segue mostra o POV para o exemplo da função pulmonar abordado na questão anterior.

Nome	Descrição	Código	Mensuração
FEF	Fluxo expiratório forçado (l/s)		quantitativa
Desporto	Pratica desporto?	0 = não; 1 = sim	nominal
Hábitos tabágicos	Fuma?	0 = não; 1 = sim	nominal

O POV deve acompanhar a tabela de valores que constitui a base de dados. Uma prática usual é registar os valores da base de dados numa folha de cálculo (por exemplo, Microsoft® Excel®) e numa outra folha escrever o POV.

Caso a introdução dos valores seja feita diretamente no SPSS, existem já duas folhas apropriadas para o efeito (*Data View* e *Variable View*). Neste caso, a componente *Variable View* permite registar o POV, apresentando porém mais campos do que a tabela referida.

8. Qual é a diferença entre estatística descritiva e inferencial? São vários os autores que dividem a estatística em duas partes: estatística descritiva e estatística inferencial. Estas correspondem a finalidades distintas da estatística, que se complementam. Por um lado, tem-se a sumarização de dados e a sua representação e, por

outro, a capacidade de extrapolação das conclusões para um grupo mais abrangente que o observado (a amostra).

A parte da estatística que tem por objetivo a descrição dos dados designa-se por estatística descritiva. A descrição dos dados pode ser realizada por três vias diferentes: a tabular, a gráfica e a textual. Na descrição tabular usam-se tabelas nas quais se apresentam estatísticas dos dados, como, por exemplo, a frequência absoluta e relativa, a média, a mediana, o desvio-padrão e a amplitude interquartil. Na representação gráfica pretende traduzir-se a distribuição dos valores obtidos e as suas eventuais relações. Já a forma textual refere-se à possibilidade de em texto corrido serem elencadas estatísticas apropriadas, por exemplo, «a média amostral do fluxo expiratório forçado médio em desportistas é de 3.91 l/s».

A estatística inferencial tem por objetivo avaliar em que medida observações feitas relativamente a uma amostra são extensíveis para a população de onde a amostra é retirada. De facto, em muitos estudos, os valores que se encontram na base de dados referem-se apenas a uma parte da população que se pretende analisar. As relações que se observam são sem dúvida válidas na amostra, no entanto, podem não ser verificáveis na população. A estatística inferencial fornece as ferramentas necessárias para inferir, com uma determinada confiança, conclusões para a população a partir do observado na amostra.

9. Qual é a relação entre população e amostra? É comum distinguir-se entre população-alvo, população acessível e amostra. A população-alvo, muitas vezes referida simplesmente como população, é o conjunto das unidades de análise que se encontra sob estudo e para o qual se pretende retirar conclusões. A população acessível é o subconjunto da população-alvo que pode ser acedida pelo investigador. A amostra é o subconjunto da população acessível relativamente ao qual se observam características ou fazem medições.

Exemplificando, para o estudo do FEF em desportistas abordado na questão 3, poderá eventualmente definir-se a população (mais especificamente a população-alvo) como sendo o conjunto de todos os praticantes de atletismo com idade entre os 16 e os 18 anos que estiveram presentes nos últimos 6 meses em pelo menos duas competições. Nem todos os elementos desta população serão contactáveis ou estarão disponíveis para a realização do teste espirométrico, indispensável para o estudo, numa clínica específica. A população acessível é constituída pelos atletas disponíveis para participar no estudo. Como a população acessível poderá ter uma dimensão considerável, não será sempre prático incluir todos os seus elementos num estudo. A amostra será assim constituída pelo subconjunto da população acessível que de facto seja incluída no estudo.

10. Qual é a diferença entre estatísticas e parâmetros? As estatísticas são medidas relativas à amostra, por exemplo média e desvio-padrão de uma das variáveis observadas. Os parâmetros são medidas relativas à população. Naturalmente, apenas se poderá conhecer com exatidão um parâmetro analisando toda a população, isto é, realizando um censo. Porém, um dos procedimentos mais comuns em estatística é a obtenção de estimativas dos parâmetros a partir das estatísticas.