Análise Psicológica (1997), 1 (XV): 19-28

A hipótese de estudo determina a análise estatística: Um exemplo com o modelo ANOVA

TERESA GARCIA-MARQUES (*)

INTRODUÇÃO

Os trabalhos de Fisher relativamente ao modelo estatístico ANOVA, desenvolvidos no campo da agricultura desde 1935, tiveram como linha orientadora a consideração de que o delineamento de um estudo anda de mãos dadas com a sua análise e tratamento estatístico. Mas esta aparente *verdade de Lapalisse* não acompanhou sempre o desenvolvimento de estudos no campo da psicologia. Desde os anos 20 que se podem encontrar referências a estudos com um delineamento experimental multifactorial. No entanto, apenas nos anos 40 é que no campo da psicologia se começam a encontrar referências à modelação destes dados por um modelo ANOVA (Lovie, 1991). Um fenómeno contrário tende agora a verificar-se: o de na literatura experimental recorrer-se indescriminadamente ao uso do modelo ANOVA. Recorre-se ao modelo ANOVA mesmo quando o delineamento do estudo não o permite (por diversas razões que apontaremos adiante) representar os dados. Recorre-se ao modelo ANOVA global para analisar todos os efeitos que lhe estão associados, mesmo quando estes não são contemplados nas nossas hipóteses, isto é, quando análises mais específicas orientaram o estudo.

Neste artigo, pretendemos alertar o leitor para a importante relação que existe entre hipótese de estudo e hipótese estatística. Ao definirmos o *design* ou planeamento do nosso estudo definimos também o modelo estatístico que poderá estar subjacente aos dados recolhidos. Assim, a hipótese empírica associada ao estudo tem uma tradução na linguagem probabilistica que define a hipótese estatistica que queremos testar. Mas, mais do que isso, a especificidade da hipótese de estudo determina o grau de especificidade do efeito que se pretende associado a esse modelo estatístico, pelo que determina o tipo de abordagem estatística a realizar. Com o objectivo de ilustrar os argumentos a apresentar ao longo do artigo, consideremos um hipotético estudo publicitário realizado no campo da publicidade (ver página seguinte).

Como analisar os dados recolhidos sob este design, com vista a testar directamente as hipóteses colocadas?

(*) Instituto Superior de Psicologia Aplicada.

EXEMPLO: **Estudo publicitário**

Objectivo: Verificar se o número de experiências com o produto x apresentado por três anúncios publicitários diferentes tem implicações no grau de satisfação que os sujeitos manifestam com o produto.

A literatura aponta para a possível existência de efeitos da variável sexo sobre o nível de satisfação em geral com produtos. Deste modo, embora não se pretenda estudar a variável sexo *per si*, ela foi incluída no *design*, com o objectivo de controlar (isolar) o seu efeito.

Assim, as hipóteses que determinaram o delineamento do estudo foram:

H1: Existência de diferenças no «grau de satisfação com o produto» manifestado pelos homens e pelas mulheres.

$$(M \neq F)$$

H2: Existência de um impacto diferencial dos três anúncios no «grau de satisfação com o produto» $(A \approx B > C)$

H3: Diminuição do «grau de satisfação com o produto» consoante o número de experiências tidas com ele.

$$(\{1.^a\} > \{2.^a\} > \{3.^a\} > \{4.^a\})$$

Sendo, por tal, o *design* que presidiu à recolha dos dados factorial:

SEXO (2) x ANÚNCIO (3) x N.º DE EXPERIÊNCIAS (4)

A variável dependente (medida) foi o GRAU DE SATISFAÇÃO COM O PRODUTO (operacionalizada por uma questão à qual se associava a possibilidade de resposta numa escala gráfica ancorada em 1- nada satisfeito e 9 - muito satisfeito).

	HOMENS			MULHERES		
	ANUNC. A	ANUNC. B	ANUNC. C	ANUNC. A.	ANUNC. B	ANUNC. C
1.ª EXPERIÊNCIA						
2.ª EXPERIÊNCIA						
3.ª EXPERIÊNCIA						
4.ª EXPERIÊNCIA						

A primeira questão que se levanta é «*Que modelo estatístico se associa a este tipo de design?*»

Um modelo ANOVA de efeitos fixos parece ser o que, de uma forma mais simples, poderá representar a estrutura/padrão subjacente aos dados da variável «grau de satisfação com o produto»[1]. Trata-se de um modelo que se define pela amostragem de uma variável aleatória contínua, com distribuição normal, num número fixo de grupos ou estratos da população. Estes grupos podem ser definidos por um só factor ou podem resultar da consideração simultânea de mais dimensões (factores). O modelo ANOVA específico a estes dados poderá ser definido por um número variável de componentes: entre dois [*no caso de se considerar apenas a grande média e o residual*] e nove [se para além de se considerar a *grande média e o residual se incluírem também os três efeitos principais, as três interacções de primeira ordem e a interacção de segunda ordem*]. Testando a significância de cada um destes efeitos, podemos verificar quais os componentes que definem o modelo mais ajustado, ou que melhor representa, os dados.

Um ***efeito principal*** que se verifique estar associado, por exemplo, ao factor **Sexo**, refere que existem diferenças no modo como homens e mulheres manifestam satisfação com o produto. Pelo que o efeito principal desta variável fica definido pelo valor que se adiciona ou se subtrai à média global em cada uma das condições definidas pelos níveis do factor (condição Feminino, Masculino). A existência de ***interacções*** define que a diferença observada entre as condições de uma variável definida num nível de outra variável, não é a mesma que no(s) outro(s) nível(eis) desta última. Um exemplo de um possível efeito de interacção entre Sexo e Tipo de Anúncio verificar-se-ia caso o impacto diferencial dos anúncios não fosse o mesmo em ambos os sexos. A componente de ***Erro*** ou ***Residual***, incorporada em todo e qualquer modelo estatístico, representa o efeito sobre o grau de satisfação de todas as variáveis não incluídas no *design* experimental. Assim, se a variável sexo não tivesse sido controlada, o seu efeito, que se pressupõe significativo, teria sido adiccionado ao termo de Erro, inflaccionando-o. O termo de Erro é por vezes designado como Acaso. Tal reflecte o facto de se considerar *acaso* toda a variação cuja fonte é desconhecida (não explicada).

O modelo estatístico ANOVA pressupõe que os erros (residuais) são variáveis aleatórias independentes (não covariam com as variáveis estudadas), cuja distribuição é assumida como normal, centrada em zero, e com variância constante em todas as condições estudadas. Estes pressupostos têm consequências directas no grau de adequação (validade e precisão) com que o modelo ANOVA representa os dados recolhidos, pelo se torna necessário garantir a sua verificação. Verificando os pressupostos sobre os quais assenta o modelo ANOVA, sabemos não estar a «construir castelos no ar», garantindo a validade das conclusões retiradas do modelo[2].

---

[1] O modelo ANOVA pressupõe que a variável em estudo (variável dependente) seja de natureza intervalar. Alguns autores contestam a natureza intervalar de medidas em *rating-scales* ou outro tipo de escalas gráficas, pelo que colocam igualmente em dúvida a adequação do modelo ANOVA à modelação desses dados, tidos como ordinais. Neste artigo, consideraremos que a variável dependente «grau de satisfação com o produto» é de natureza intervalar. Não cabe no âmbito deste artigo a apresentação da discussão em torno deste assunto, pelo que ao leitor interessado recomendamos a leitura dos trabalhos de: Borgatta e Bohrnstedt (1980), Davidson e Sharma (1988), Gaito (1980) e Michell (1986).

[2] Caso os residuais não tenham média zero, as estimativas dos efeitos estudados são, seguramente, enviezadas. A violação de outros pressupostos não questiona directamente as estimativas mas afecta a integridade da afirmação probabilística, que define um efeito como significativo ou não (inflaccionando os erros Tipo I e Tipo II em porporções desconhecidas que podem atingir valores alarmantes). Os estudos realizados sobre estas violações, estão longe de ser conclusivos. A análise realizada por Box (1954) sugere que a violação simultânea dos pressupostos de normalidade e homocedasticidade, quando esta última é pequena (Fmax<3) não acarreta grandes enviesamentos caso exista igual número de observações por célula. Em caso de violações mais acentuadas ou não existência de igual número de observações por célula, a adequabilidade do modelo ANOVA é posta em causa, pelo que se deve recorrer a outro tipo de modelo estatístico, ou proceder a uma transformação dos dados, com vista a verificar os pressupostos do modelo ANOVA.

Consideremos as hipóteses que guiaram a realização deste estudo. Elas representam os três possíveis efeitos principais associados ao modelo ANOVA. Uma simples leitura das médias obtidas em cada condição experimental poder-nos-ia sugerir que os padrões previstos caracterizam ou não os nossos dados. Ficar-nos-ia, no entanto, uma dúvida: este padrão foi obtido por mero acaso ou reflecte uma realidade passível de ser encontrada na maioria das amostras que recolhêssemos? Esta dúvida pode ser dissipada se considerarmos a probabilidade deste padrão ter surgido por mero acaso na nossa amostra, isto é, se considerarmos a probabilidade dele se verificar na nossa amostra partindo do princípio que não existe na população (Ho é verdadeira). Quanto mais pequena for esta probabilidade, maior a confiança com que se afirma poder encontrar o mesmo padrão de resultados em qualquer outra amostra retirada aleatoriamente da população. Assim, caso a probabilidade seja inferior a um valor pré-estabelecido (α), toma-se o padrão de diferenças encontrado como significativo, isto é, rejeita-se Ho como verdadeira.

Como estudar então a significância dos padrões postulados em cada uma das três hipóteses de estudo? A especificidade com que as hipóteses são colocadas, impelem a uma abordagem estatística confirmatória (por oposição à exploratória) que se centre directamente em cada padrão postulado de efeitos. Tal necessidade, sendo frequentemente compreendida pelos experimentadores, é muitas vezes mal orientada em termos da análise específica que estes levam a cabo, que tende a negligenciar o facto dos dados terem sido recolhidos sob um *design* global. Para compreendermos as consequências, em termos de validade das conclusões retiradas da análise estatística aos dados, do facto de não considerarmos a estreita relação entre o delineamento do estudo e a análise dos dados, atendamos a três abordagens diferentes ao estudo publicitário em questão:

1) Testar cada hipótese por uma ANOVA ONE-WAY (simples)

2) Testar simultaneamente as três hipóteses numa ANOVA 3-WAY (3 factores)

3) Testar cada hipótese em contrastes planeados

Estas três abordagens diferem relativamente às probabilidades de erro (Erro Tipo I e Erro Tipo II) associadas à decisão de significância do efeito. Lembramos que a conclusão de um teste se refere à validade da hipótese nula (Ho), e que a esta conclusão se associam dois tipos de erro. O *erro designado de Tipo I* ocorre quando rejeitamos Ho sendo esta verdadeira, e a sua probabilidade define o *nível de significância* (α) da conclusão. Por outro lado, se Ho for falsa e a tomarmos como verdadeira, cometeremos um *erro designado de Tipo II* e a sua probabilidade é complementar à probabilidade do teste detectar um efeito quando este está presente, que define a *potência do teste* ($1-\beta$). Assim temos,

(α) Erro Tipo I e **($1-\alpha$)** Coeficiente de confiança associado à rejeição de Ho;

(β) Erro Tipo II e **($1-\beta$)** Potência de teste ou coeficiente de confiança associada à aceitação de Ho.

A probabilidade de se rejeitar erroneamente uma hipótese nula numa comparação/teste estatístico, aumenta consoante o número de **comparações independentes** a realizar sob os mesmos (fenómeno que se designa de «*inflacção do alfa*» – ver por exemplo, Garcia-Marques & Azevedo, 1995; Scheffé, 1953, 1959). Se pretendermos ter 95% de confiança na conclusão de que os três efeitos postulados nas hipóteses de estudo estão presentes (ou seja, assumindo a hipótese de errar em 5 de 100 vezes, por decidir pela presença de um efeito quando este, na realidade não existe), então estamos a definir à partida a probabilidade de erro Tipo I como sendo igual a 0.05. Qual a probabilidade de não cometermos nenhum erro Tipo I nas três análises postuladas pela primeira abordagem? A probabilidade da conjunção de acontecimentos independentes é dada pelo produto das probabilidades de cada um desses acontecimentos (ver por exemplo Murteira, 1990) pelo que a probabilidade de não se cometer, nesta análise, nenhum erro Tipo I, caso as comparações sejam independentes, é dada por:

$$(1-0.05)\,(1-0.05)\,(1-0.05) = 0.857$$

Assim sendo, a probabilidade de se concluir erroneamente a validade de pelo menos uma das hipóteses de estudo é de:

$$(1-0.857) = 0.143$$

Verifica-se assim que o valor pretendido de alfa (5%) é inflaccionado para 14,3%.

A potência de um teste estatístico, estando dependente do próprio modelo estatístico, é função de três parâmetros específicos: (1) Magnitude do efeito estudado (ME); (2) Nível de significância escolhido (α); e (3) Precisão/fidelidade dos dados recolhidos (S/N = Erro Padrão) (ver por exemplo Cohen, 1988 e Azevedo & Garcia-Marques, no prelo).

Nas três abordagens aqui em comparação, a magnitude de efeito estudado é a mesma, visto que as reportamos ao mesmo conjunto de dados. O nível de significância utilizado em cada um pode, caso o desejarmos, ser idêntico. O mesmo não se passa no que diz respeito à variabilidade que se define como Erro. Como já referimos, qualquer efeito incluído num modelo ANOVA é subtraído à variação residual. Qualquer efeito excluído do modelo é adiccionado ao *Residual.* A realização de ANOVAS simples sobre cada um dos efeitos em estudo, postula a construção de três modelos distintos onde se considera apenas um efeito principal. Assim, a proporção da variabilidade associada a cada um dos restantes efeitos principais e associada às interacções é acoplada à variabilidade deixada por explicar no nosso delineamento, pelo que é considerada Erro. Se algum dos efeitos ignorados for de grande magnitude, inflaccionará a quantidade de variabilidade que se atribui «erradamente» ao acaso, pelo que inflaccionará o termo de erro do modelo. Para um efeito observado ser considerado significativo, compara-se a variabilidade que o define com a atribuída ao acaso (calculando-se a estatística de teste), esperando-se que o primeiro se sobreponha ao segundo. Um efeito principal é considerado presente se assumir uma magnitude que o destaque de todo o ruído (acaso). Ora, se o ruído foi inflaccionado, corremos o risco de não detectarmos a presença do dito efeito mesmo quando este está presente (diminuímos a potência do teste, aumentamos a probabilidade de erros Tipo II). Desta forma, os testes a realizar sob esta abordagem têm fraca potência comparativamente às restantes abordagens e, consequentemente, a validade das conclusões extraídas é mais fraca.

Para ilustrar este ponto, comparemos as decisões estatísticas a que nos conduziriam as duas primeiras abordagens. Para o efeito, recorremos ao programa informático de análise de dados STATISTICA, módulo ANOVA/MANOVA, definindo os respectivos designs. O Quadro 1 sumariza os outputs do procedimento efectuado.

1) Testar cada hipótese por uma ANOVA 1-WAY (SIMPLES)

QUADRO 1
Resultados extraídos a partir de uma ANOVA 1-Way (SIMPLES)

1-way ANOVA		DEPENDENT: 1 variable:	SATISFAÇÃO		BETWEEN: 1-SEXO (2): M F	
Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
SEXO	1*	147.4083*	118*	4.510028*	32.68457*	.000000*

1-way ANOVA		DEPENDENT: 1 variable:	SATISFAÇÃO		BETWEEN: 1-ANUNCIO (2): A B C	
Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
ANÚNCIO	2	10.75833	117	5.624573	1.912738	.152265

1-way ANOVA		DEPENDENT: 1 variable:	SATISFAÇÃO		BETWEEN: 1-N.º EXP (4): 1 2 3 4	
Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
N.º EXP	3	12.69722	116	5.530172	2.295990	.081385

Face ao conjunto de resultados das três análises realizadas, seríamos induzidos a concluir que apenas a variável Sexo teve algum efeito sobre o grau de satisfação manifestado pelos sujeitos experimentais. O paradoxo reside essencialmente no facto da variável que se sabia *a priori* ter efeito e por tal ter sido introduzida no *design*, não estar a ser controlada quando se analisam os restantes efeitos. Vejamos, então, em que mudariam as nossas decisões se controlássemos o efeito da variável sexo, em termos estatísticos (Quadro 2).

2) Testar simultaneamente as três hipóteses numa ANOVA 3 -WAY (3 FACTORES)

Construindo o modelo estatístico adequado ao verdadeiro delineamento do estudo, detectamos a presença do efeito esperado de acordo com a hipótese 3, do factor N.° de Experiências. Neste tipo de análise exercemos um verdadeiro controlo da variável sexo, visto que subtraímos o seu efeito ao termo de erro do modelo. Na realidade, nesta análise, o termo de erro é constante para todos os efeitos estudados, e embora os graus de liberdade associados ao erro diminuam, a verdade é que os valores de F associados aos dois efeitos de Anúncio e de Experiência, aumentaram. Trata-se de uma análise mais potente que a da primeira abordagem utilizada. No entanto, contrariamente à primeira abordagem, a Anova *3-Way*, tem as características globais de uma abordagem exploratória, e não faz uso das capacidades de uma abordagem confirmatória que se foca exclusivamente sob os efeitos esperados.

3) Testar cada hipótese em contrastes planeados

A realização de contrastes planeados adicciona à potência de teste da análise global do *design* o ganho em potência de teste que advém da abordagem confirmatória. Para a compreendermos, tenhamos em consideração a noção de contraste. Designa-se por contraste uma comparação muito específica entre as médias de um *design* mais ou menos complexo. Note-se que um conjunto de 4 médias pode estabelecer 3 padrões independentes (ortogonais) de diferenças entre si (o que determina o número de graus de liberdade associados ao factor). Ao definir um contraste, definimos um padrão específico de diferença. Assim ao estudar a significância de um contraste definido ao nível de um efeito principal, contrariamente a testar se existe algum padrão de diferenças entre as médias desse

QUADRO 2
Resultados extraídos a partir de uma ANOVA 3-WAY (3 FACTORES)

3 - way ANOVA	DEPENDENT: 1 variable: SATISFAÇÃO			BETWEEN:	1-SEXO	(2): M F
					2-ANÚNCIO	(3): A B C
					3-N.°EXP	(4): 1 2 3 4
Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
SEXO	**1**	**147.4083**	**96**	**4.712500**	**31.28028**	**.000000**
ANÚNCIO	2	10.7583	96	4.712500	2.28294	.107495
N.°EXP	**3**	**12.6972**	**96**	**4.712500**	**2.69437**	**.050314**
SEX x ANUN	2	2.3583	96	4.712500	.50044	.607835
SEX x N.°EXP	3	.5194	96	4.712500	.11023	.953936
ANUN x N.°EXP	6	1.7472	96	4.712500	.37076	.895883
SEX x ANUNC x N.°EXP	6	.5694	96	4.712500	.12084	.993652

factor, testa-se se aquele padrão específico (ao qual se associa apenas um grau de liberdade) está ou não presente. Tal só é possível se o nosso estudo tiver hipóteses também elas muito específicas.

A relação entre as médias de um mesmo factor é definida por uma combinação linear $Lk = a_iX_i + \ldots + a_mX_m$, onde a_i refere o coeficiente associado ao grupo i e X_i a média do grupo I, onde $\Sigma(a_i)=0$. A ausência de um nível do factor (grupo) coincide com este se se encontrar associado ao coeficiente zero. A significância do contraste é estudada relativamente ao modelo global onde este se define, pelo que, assumindo homogeneidade de variâncias, temos que o erro padrão da comparação é dado por:

$$se_k = \sqrt{MSe \Sigma \frac{a_i^2}{n_i}}$$

em que MSe representa o quadrado médio do erro residual da ANOVA. Assim a estatística de teste associada à hipótese nula, Ho: Lk=0, é definida por $t_k = \frac{L_k}{se_k}$ que tem subjacente uma distribuição T-Student com o número de graus de liberdade associados ao termo de erro (o que é equivalente a uma distribuição $F_{(1,v)}$).

Dois ou mais contrastes dizem-se ortogonais quando, se referindo a comparações de níveis do mesmo factor, se referem a questões independentes, não correlacionadas. A r condições experimentais (níveis do factor), associam-se r-1 contrastes ortogonais (que coincidem com o número de graus de liberdade associados ao factor). Uma forma rápida de garantir ortogonalidade dos contrastes é garantir que a soma dos produtos dos pesos atribuídos aos i níveis do factor (ai) é zero. Assim, dizemos que j contrastes (j=1,..,r-1) realizados para os valores Xi, aos quais se atribuídos pesos aij, são ortogonais se:

$$\sum_{i=1}^{r} \prod_{j=1}^{r-1} a_{ij} = 0$$

(o somatório da multiplicação dos pesos atribuídos for igual a zero):

EX:

i)	+1	-1/2	-1/2	**ii)**	+1	0	-1
	0	+1	-1		-1/2	+1	-1/2
Πa_{ij} =	**0**	**-1/2**	**+1/2**	Πa_{ij} =	**-1/2**	**0**	**+1/2**
		$\Sigma(\Pi a_{ij}) = 0$				$\Sigma(\Pi a_{ij}) = 0$	

Pode-se definir um tipo específico de contraste quando a natureza da variável independente é ordinal. Isto porque, no contraste, pode-se definir a relação entre esta variável e a variável de medida como linear ou não linear. Nestes casos, o contraste específico define uma *trend* (tendência), qualificada de linear, quadrática, cúbica, etc., positiva ou negativa. Uma *trend* define um padrão específico que se pensa descrever a relação entre as variáveis em estudo, um padrão específico de diferenças entre as médias obtidas nos diferentes níveis da variável ordinal. A terceira hipótese do nosso estudo, define precisamente a existência de uma relação linear negativa entre o número de experiências que os sujeitos tiveram com o produto e o grau de satisfação para com o mesmo.

Mas qual é a vantagem de se realizarem contrastes, relativamente à abordagem global?

A vantagem encontra-se no facto da análise em contraste diminuir os níveis de erro Tipo I e de erro Tipo II que se associam às suas conclusões. Ao diminuir o número de comparações estatísticas que se levam a cabo, exerce-se um maior controlo sobre o erro Tipo I. Diminuem-se, igualmente, os erros do Tipo II por a potência do teste associado a um contraste ser superior à potência associada à abordagem geral, e isto porque nos concentramos na magnitude de um efeito específico em vez da magnitude média de uma classe de efeitos possíveis. Quando no teste global (ANOVA *3-WAY*) consideramos a presença de um efeito associado ao tipo de anúncio, não especificamos que tipo de efeito. Assim, havendo três anúncios, estes podem estabelecer apenas dois padrões de diferença que sejam independentes um do outro (pelo que dois *graus de liberdade*). A variabilidade observada em termos de satisfação com o produto que se associa aos anúncios, é dada pela soma da variabilidade subjacente a qualquer um destes padrões. Assim, a variabilidade média associada aos efeitos do factor (padrões) é calculada dividindo o total dessa variabilidade (soma dos quadrados- SS) pelo seu número de graus de liberdade (resultando no MS). Tal deverá ser a estratégia utilizada caso hipotetizemos a existência de um padrão de diferenças e não o especifiquemos (isto é, caso se pretenda interpretar qualquer padrão que surja nos dados mesmo que diferente do hipotetizado).

Mas, se quisermos fazer uma afirmação precisa relativamente ao tipo de efeito que se associa a um dado factor (uma afirmação que tenha por detrás a ideia de que «ou o efeito é este ou é como se ele não existisse!»), então pretendemos apenas interpretar a variabilidade que se associa a esse padrão específico dos dados. Assim sendo, devemos ter apenas em consideração a magnitude do efeito associado a esse padrão. Se o padrão se ajusta perfeitamente aos dados, esta magnitude atingirá valores que se distanciam dos que seriam de esperar por mero acaso, pelo que será significativa. Paralelamente, para que o argumento seja completo, há que verificar que nenhum outro padrão diferente do que testamos, se associa aos nossos dados. Tal verificação advém de se testar que o residual do contraste não é significativo.

Quando definimos as hipóteses de estudo publicitário sob análise, associámos um padrão específico a cada uma delas e definimos a análise como confirmatória. Ao definir a análise como confirmatória, afirmamos querer apenas saber se os dados davam ou não suporte às nossas hipóteses, em vez de nos prepararmos para interpretar todo e qualquer padrão que pudesse emergir dos dados. Sendo assim, podemos definir concretamente os padrões que as nossas hipóteses associam a cada efeito principal em termos de contrastes ou comparações planeadas, e testar a sua significância. No Quadro 3, sumarizamos os resultados da análise da significância dos contrastes definidos.

Quando procedemos a uma análise tão específica, qualquer das hipóteses de estudo parece ter suporte empírico (a um nível de a de 5%). Pelo que podemos concluir pela não rejeição das hipóteses em estudo. Os anúncios A e B, promovendo igual satisfação com o produto, parecem ser mais eficazes que o anúncio C. Porém, tal como se previa, o número de experiências com o produto leva a uma progressiva diminuição dessa satisfação. A análise global realizada sugere interpretações de carácter hipotético relativas à ausência de efeitos interactivos. Assim, ela sugere que o efeito de consumo progressivo se verifica independentemente do anúncio que apresentou o produto ao sujeito e que todos estes efeitos são constantes para ambos os sexos (apesar de homens e mulheres diferirem na satisfação com o produto).

Os objectivos da análise estatística de um conjunto de dados são os mesmos que conduzi-

QUADRO 3

Resultados extraídos a partir da análise de contrastes

ANOVA	DEPENDENT: 1 variable:	SATISFAÇÃO			BETWEEN: 1-ANÚNCIO: A B C			
	Contraste:	A	B	C				
		1	1	-2				
	df Effect	SS Effect	MS Effect	df Error	SS Error	MS Error	F	p-level
ANÚNCIO	1	18.70	18.70	96	452.40	4.71	3.97	.049184
RESIDUAL	1	2.81	2.81	96	452.40	4.71	0.59	n.s.

ANOVA	DEPENDENT: 1 variable:	SATISFAÇÃO	BETWEEN: 1-N.ºEXP (4): 1 2 3 4					
	Contraste: 1	2	3	4				
	3	1	-1	-3	(TREND LINEAR negativo)			
	df Effect	SS Effect	MS Effect	df Error	SS Error	MS Error	F	p-level
N.ºEXP	1	38.00	38.00	96	452.40	4.71	8.06	.005512
RESIDUAL	2	.09	0.45	96	452.4	4.71	.00	n.s.

ram à recolha desses mesmos dados. Quanto maior o carácter exploratório do estudo, maior o carácter exploratório da análise estatística. Quanto mais precisas e específicas as hipóteses em estudo, mais precisas devem ser as análises estatísticas que as testam. Em condições ideais, as hipóteses de um estudo traduzem-se em hipóteses que se associam a um modelo estatístico. No entanto, nem sempre se consegue desenvolver o modelo estatístico que se associa ao delineamento do estudo. Para certos delineamentos, tem sido extremamente difícil quer encontrar algoritmos que estimem eficientemente os parâmetros do modelo que lhes pode estar subjacente, quer estudar as suas características distribucionais. Como exemplo, temos o caso de modelos de análise de variância em heteroscedasticidade e os modelos log-lineares para medidas repetidas. A outros planos de estudo podem associar-se modelos probabilísticos com características computacionais mais fáceis mas não directamente disponíveis em programas estatísticos. Por todas estas razões é de extrema importância que o investigador considere a análise estatística dos seus dados ao conceber o plano do seu estudo. Quanto mais perfeita for a relação entre os objectivos do estudo, o seu delineamento experimental e a abordagem estatística, mais a função desta última é cumprida.

Há que deixar de pensar nas decisões sobre a análise estatística dos dados como uma tarefa independente da do planeamento do estudo a ser realizada após a recolha dos dados. Ao conceber um estudo, há que conceber um estudo passível de ser modelado em termos de partição de variabilidade e estudo da sua probabilidade, por algum dos modelos já desenvolvidos em estatística. Ao deduzir da teoria as hipóteses que orientam a construção de *design*, há que procurar verificar o seu nível de especificidade e perceber se pretendemos apenas perceber a presença ou ausência de um efeito específico ou a teoria nos dá liberdade de interpretar outro tipo de padrões nos resultados. Deste modo situar-nos-emos no contínuo confirmatório-exploratório que definirá a análise dos nossos dados.

Com o desenvolvimento de programas estatísticos, a análise dos dados de um estudo foi apenas facilitada em termos de cálculos e rapidez de implementação de algoritmos. Os programas substituem o conhecimento das fórmulas e da sua aplicação, e **nada mais**. Os programas não decidem as análises a serem levadas a cabo, nem nos elucidam sobre a validade das conclusões a que nos induzem. O centro de decisão e controlo de todo o processo ainda está no utilizador e não na máquina. Lá porque uma batedeira bate ovos e massa, não quer dizer que saiba fazer bolos. Ela não distinguirá, por exemplo farinha de cimento... e nós que partamos os dentes ao digerir as conclusões de alguns estudos publicados.

REFERÊNCIAS

Azevedo, M. & Garcia-Marques, T. (no prelo). Que confiança podemos ter nas conclusões estatísticas que apresentamos?. *Psicologia.*

Borgatta, E. F., & Bohrnstedt, G. W. (1980). Level of measurement: Once over again. *Sociological Methods and Research, 9*, 147-160.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effects of inequality of variance in the one-way classification. *Annual of Mathematics Statistics, 25*, 290-302.

Davidson, M. L., & Sharma, A. R. (1988). Parametric statistics and levels of measurement. *Psychological Bulletin, 140*, 137-144.

Gaito, J. (1980). Measurement scales and statistic: Resurgence of an old misconception. *Psychological Bulletin, 87*, 564-567.

Garcia-Marques, T., & Azevedo, M. (1995). A inferência estatística múltipla e o problema da inflação do nível de alfa: a ANOVA como exemplo. *Psicologia, 10* (1), 195-220.

Lovie, A. D. (1991). *New developments in statistics for psychology and the social sciences*. London: BPS and Routledge.

Michell, J. (1986). Measurement scales and statistic: a clash of paradigms. *Psychological Bulletin, 100*, 398-407.

Murteira, B. J. F. (1990). *Probabilidades e estatística* (Vol. I). Lisboa: McGraw-Hill.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika, 40*, 87-104.

Scheffé, H. (1959). *Analysis of variance*. New York: Wiley.

RESUMO

As decisões sobre a análise estatística dos dados de uma investigação, estando intimamente relacionadas com o delineamento do estudo e com as hipóteses que o geram, não podem ser encaradas como uma tarefa independente, a ser realizada após a recolha de dados.

Para ilustrar a estreita relação entre hipótese de estudo e hipótese estatística, dados de um estudo publicitário são abordados de três modos diferentes, tendo por subjacente o modelo ANOVA. A primeira abordagem negligencia o delineamento factorial do estudo e analisa cada efeito principal por uma ANOVA One-way. A segunda tem em consideração a natureza factorial do delineamento experimental mas não a especificidade das hipóteses em estudo. A terceira procura a total adequação entre hipótese de estudo e abordagem estatística, realizando análises específicas definidas em contrastes ou comparações planeadas. Uma leitura comparativa das conclusões retiradas de cada abordagem permite perceber o argumento de estreita relação entre a planificação de um estudo e seu tratamento estatístico.

Palavras-chave: Análise Estatítica, ANOVA, Contrastes.

ABSTRACT

There is a close relation between experimental hypothesis, experimental design and statistical data analysis. In order to understand the way statistical data analysis is dependent of design and hypothesis this article discusses three different approaches to the same set of data, all them within an ANOVA model. The first approach neglects the factorial nature of the design and analyses the data with an One-way ANOVA. The second approach has in consideration the nature of the design but not the specificity of the hypothesis that orientated the study. The third approach tries to achieve a total fit between statistical and study hypothesis, suggesting a couple of planned comparisons. The conclusions associated with these three different approaches allow us to understand the argument of a close relation between the way we plan the study and the way we analyze its data.

Key words: Statistical Analysis, ANOVA, Contrast.