CLASSIFICAÇÃO E ANÁLISE DE DADOS

Métodos e Aplicações III - CLADMAp III

Editores

Helena Bacelar-Nicolau

Fernanda Sousa

Carlos Marcelo

Ana Sousa Ferreira

Paulo Infante

Adelaide Figueiredo

CLASSIFICAÇÃO E ANÁLISE DE DADOS

MÉTODOS E APLICAÇÕES III - CLADMAp III

Editores
Helena Bacelar-Nicolau
Fernanda Sousa
Carlos Marcelo
Ana Sousa Ferreira
Paulo Infante
Adelaide Figueiredo

Associação Portuguesa de
Classificação e Análise de Dados

Título
Classificação e Análise de Dados – Métodos e Aplicações III

Editores
Helena Bacelar-Nicolau (Universidade de Lisboa)
Fernanda Sousa (Universidade do Porto)
Carlos Marcelo (Instituto Nacional de Estatística)
Ana Sousa Ferreira (Universidade de Lisboa)
Paulo Infante (Universidade de Évora)
Adelaide Figueiredo (Universidade do Porto)

Impressão
Instituto Nacional de Estatística
Av. António José de Almeida
1000-043 LISBOA

1.ª Edição
Lisboa, Abril de 2019
ISSN 2183-8801
Depósito legal 454535/19
Tiragem: 200 exemplares

Todos os direitos reservados. Nenhuma parte desta publicação pode ser reproduzida por processo mecânico, eletrónico ou outro sem autorização escrita dos editores.

Dificuldades na ANOVA com dois fatores não paramétrica com células omissas

Dulce G. Pereira[1] · Anabela Afonso[2]

Resumo Na análise de variância com dois fatores por vezes há combinações de níveis dos fatores que não são observadas. A inexistência destas combinações dificulta a análise dos dados, pois depende do número de células omissas e da sua localização. Já foram propostas algumas abordagens mas apenas para a ANOVA paramétrica. Uma das alternativas não paramétricas à ANOVA consiste em aplicar a abordagem paramétrica às ordens dos valores observados. Neste trabalho ilustraremos que na presença de células omissas este processo pode não ser o mais adequado, sendo necessária mais investigação nesta área.

Palavras-chave: ANOVA, Dados Omissos, Hipóteses de Tipo IV, Modelo de Efeitos Fixos, Modelo de Médias, Somas de Quadrados.

1 Introdução

Muitas experiências envolvem o estudo do efeito de dois ou mais fatores numa variável resposta. Admitindo que existem dois fatores R e C, com r e c níveis respetivamente, diz-se que o delineamento é fatorial quando cada réplica contém todas as rc combinações de níveis destes dois fatores (Montgomery, 2013). No entanto, é usual na prática encontrarem-se situações em que o número de observações por célula não é igual (delineamento desequilibrado) e até podem não ter sido observadas algumas das combinações de níveis (células omissas). Este tipo de situações por vezes ocorre devido a várias razões, que podem ser por conveniência de delineamento devido por exemplo a custos associados, mas também porque o investigador não consegue controlar em absoluto a experiência (por ex., plantas que morrem). Quando o investigador tem controlo sobre a experiência, deve ter muito cuidado para garantir que todas as combinações de níveis dos fatores sejam observadas.

Para a análise dos dados de um delineamento com células omissas foram propostos vários tipos de abordagens paramétricas. Montgomery (2013) propôs

[1] CIMA/IIFA e Departamento de Matemática/ECT, da Universidade de Évora, *dgsp@uevora.pt*
[2] CIMA/IIFA e Departamento de Matemática/ECT, da Universidade de Évora, *aafonso@uevora.pt*

transformar o delineamento com dois fatores num delineamento a um fator com $rc - m$, onde m representa o número de células omissas e realizar uma análise de variância (ANOVA) a um fator. Para testar a interação entre os fatores, propôs a utilização de contrastes de interesse, linearmente independentes. Alguns *software* estatísticos, como sejam o SPSS e SAS, recomendam o uso das somas de quadrados de Tipo IV. As hipóteses de Tipo IV comparam a média de todos os níveis de cada fator com a média obtida a partir de um ou mais níveis comuns do outro fator. Estas somas coincidem com as somas de Tipo III quando não existem células omissas. Mas, vários autores têm salientado que as hipóteses de Tipo IV podem não ser únicas, pelo que o investigador deve definir as suas comparações específicas. Além disso, as conclusões dependem do número de células omissas e da posição em que surgem. Milliken e Johnson (2009) propõem que em vez de se considerar a análise através de um modelo de efeitos fixos se opte por considerar um modelo de médias (*means model*) o qual é muito simples e fácil de interpretar.

Quando as variáveis são de natureza ordinal, ou são violados os pressupostos de aplicação da ANOVA paramétrica a dois fatores, Hora e Conover (1984) propuseram um teste equivalente em que a estatística de teste F da ANOVA era obtida a partir das ordens das observações. Contudo, as conclusões referentes à possível existência de uma interação significativa não devem ser garantidas como corretas uma vez que os efeitos aditivos dos fatores nas observações originais não são extensíveis às suas ordens e vice-versa. Além disso, o estudo da interação origina probabilidades de erro de Tipo I elevadas (Shah e Madden, 2004). Em alternativa foi sugerido o uso de outras estatísticas de teste com uma distribuição aproximada a um qui-quadrado, que conduzem às mesmas conclusões relativamente aos efeitos principais dos fatores (Shah e Madden, 2004). No entanto, em nenhuma destas alternativas foi tida em conta a possível presença de células omissas.

Neste trabalho vamos aplicar, quer aos valores observados quer às respetivas ordens, as três abordagens paramétricas referidas anteriormente para a ANOVA com dois fatores com células omissas. Recorrendo a alguns exemplos de aplicação, mostraremos que este processo pode não ser o mais adequado como alternativa não paramétrica. Deste modo é necessário mais investigação nesta área.

2 Delineamentos com dois fatores com células omissas

As abordagens paramétricas mais usuais para analisar delineamentos com dois fatores com células omissas, e que estão disponíveis nos *software* estatísticos, são: i) calcular as somas de quadrados de Tipo IV, ii) considerar o modelo das médias, e iii) transformar o delineamento com dois fatores num delineamento com um fator.

De seguida assumiremos que os fatores R e C têm r e c níveis respetivamente e que n_{ij} representa o número de observações no nível i do fator R e nível j do fator C.

2.1 Somas de quadrados de tipo IV

O modelo de efeitos fixos (*effects model*) é definido por (Milliken e Johnson, 2009):

$$Y_{ijk} = \mu + \tau_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \qquad i = 1, \dots, r; j = 1, \dots, c; k = 1, \dots, n_{ij}$$

onde μ representa a média global, τ_i o efeito do i-ésimo nível do fator R, β_j o efeito do j-ésimo nível do fator C, γ_{ij} o efeito devido à interação entre o i-ésimo nível do fator R e o j-ésimo nível do fator C, e $\varepsilon_{ijk} \sim N(0; \sigma^2)$ e independentes. Quando a combinação R_i e C_j não é observada, $n_{ij} = 0$.

Quando existem células omissas Goodnight (1980) sugeriu o uso das somas de quadrados de Tipo IV. Basicamente, uma hipótese é definida como uma hipótese de Tipo IV se compara a média dos níveis de um tratamento com a média de um ou mais níveis comuns do outro tratamento. Assim, espera-se que as hipóteses sejam hipóteses de médias marginais. Quando faltam tratamentos, não se pode considerar a média em todos os níveis do outro tratamento, mas apenas em alguns dos níveis do outro fator.

Esta abordagem tem a vantagem de estar disponível em alguns *software* (por ex.: SPSS e SAS). Contudo, as hipóteses de Tipo IV podem não ser únicas e os *software* testam automaticamente algumas delas (as que são habitualmente interpretáveis), sem as especificar. Para identificar quais as hipóteses testadas é necessário analisar as funções estimadas. Caso as hipóteses não correspondam às de real interesse, recorrendo à programação o utilizador poderá definir as hipóteses a testar. No R desconhecemos a existência de algum pacote com esta abordagem. Recorrendo a alguma programação construímos algumas rotinas, que usam as funções glm e glht do R, para testar as hipóteses de interesse associadas aos contrastes definidos.

Para ilustrar, sem perda de generalidade, considere-se o delineamento com dois fatores e com duas células omissas apresentado na Tabela 1. É possível comparar R_1 com R_2 através da hipótese de Tipo IV H_0: $(\mu_{11} + \mu_{13})/2 = (\mu_{21} + \mu_{23})/2$, i.e., nos níveis C_1 e C_3, ou da hipótese H_0: $\mu_{11} = \mu_{21}$ que considera apenas o nível C_1.

Tabela 1 – Média das células num delineamento com duas células omissas (μ_{ij} é a média da variável resposta, ou das ordens da variável resposta, dos níveis i e j dos fatores R e C, respetivamente).

	C_1	C_2	C_3
R_1	μ_{11}	μ_{12}	μ_{13}
R_2	μ_{21}		μ_{23}
R_3		μ_{32}	μ_{33}

Todas as possíveis hipóteses de tipo IV, associadas ao exemplo anterior, são apresentadas na Tabela 2.

Tabela 2 – Todas as possíveis hipóteses de Tipo IV para os fatores R e C.

	Para o fator R		Para o fator C
1	$\frac{\mu_{11}+\mu_{13}}{2} = \frac{\mu_{21}+\mu_{23}}{2}$	1	$\frac{\mu_{11}+\mu_{21}}{2} = \frac{\mu_{13}+\mu_{23}}{2}$
2	$\frac{\mu_{12}+\mu_{13}}{2} = \frac{\mu_{32}+\mu_{33}}{2}$	2	$\frac{\mu_{12}+\mu_{32}}{2} = \frac{\mu_{13}+\mu_{33}}{2}$
3	$\mu_{11} = \mu_{21}$	3	$\mu_{11} = \mu_{12}$
4	$\mu_{12} = \mu_{32}$	4	$\mu_{11} = \mu_{13}$
5	$\mu_{13} = \mu_{23}$	5	$\mu_{12} = \mu_{13}$
6	$\mu_{13} = \mu_{33}$	6	$\mu_{21} = \mu_{23}$
7	$\mu_{23} = \mu_{33}$	7	$\mu_{32} = \mu_{33}$

2.2 Modelo das médias

Outra abordagem que muitas vezes se revela muito útil, especialmente quando existem células omissas, é abandonar o modelo de efeitos e empregar, em vez disso, o modelo das médias (*means model*), que é definido por (Milliken e Johnson, 2009):

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \qquad i = 1, \dots, r; j = 1, \dots, c; k = 1, \dots, n_{ij}$$

onde $\mu_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij}$, $\varepsilon_{ijk} \sim N(0; \sigma^2)$ e independentes, e $n_{ij} = 0$ quando a combinação R_i e C_j não é observada.

A aplicação do modelo das médias apresenta várias vantagens: i) permite estimar cada um dos parâmetros do modelo e torna a interpretação mais fácil, ii) permite testar contrastes em simultâneo, e iii) é possível fazer inferência sobre funções com parâmetros que envolvam células omissas, desde que sejam impostas condições (por ex., a inexistência de interação) que devem estar suportadas em alguma evidência experimental (Milliken e Johnson, 2009). Contudo, é necessário recorrer à programação para definir os contrastes de interesse, o que para quem não está familiarizado com as linguagens de programação é uma desvantagem.

2.3 Transformar num delineamento com um fator

Montgomery (2013) propôs transformar o delineamento com dois fatores num delineamento a um fator com $rc - m$, onde m representa o número de células omissas, e realizar uma análise de variância (ANOVA) com um fator. Ou seja, cada nível do fator no modelo com um fator é, na verdade, uma combinação de níveis dos fatores do delineamento original. Para testar a interação entre os fatores, propôs a utilização de contrastes de interesse, linearmente independentes. Este modo de proceder tem a vantagem de ser simples, mas por outro lado a desvantagem de não permitir testar dois contrastes em simultâneo.

3 Exemplos

De seguida ilustra-se a aplicação das abordagens anteriores, aos dados originais e às ordens dos valores observados, em delineamentos com dois fatores e células omissas.

3.1 Exemplo 1

Considere-se que para uma dada experiência foram obtidos os resultados apresentados na Tabela 3.

Tabela 3 – Delineamento com dois fatores com células omissas (adaptado de Kleinbaum *et al.*, 1998).

	C_1	C_2	C_3
R_1	2; 7; 8; 10	2; 5	4; 5; 6
R_2	11; 17; 18		8; 10; 17
R_3		6; 9	3; 5; 8; 16

Na ANOVA paramétrica assume-se que a variável resposta é do tipo contínuo e devem-se validar os pressupostos de normalidade e homogeneidade das variâncias. Neste exemplo verificámos que estes pressupostos eram válidos (Teste Shapiro-Wilk: $W = 0{,}963$, $p = 0{,}581$; Teste Levene: $F = 1{,}098$, $p = 0{,}410$).

3.1.1 Somas de quadrados de tipo IV

Recorrendo ao *software* SPSS, calcularam-se as somas de quadrados de Tipo IV associadas aos dados da Tabela 3, quer dos valores observados quer das ordens das observações, tendo-se obtido os resultados apresentados na Tabela 4.

Tabela 4 – Resultados da ANOVA aplicada aos valores observados e às ordens.

		Valores observados			Ordens das observações		
$F.V.$	$g.l.$	$S.Q.$	F	p	$S.Q.$	F	p
R	2	72,8	2,348	0,132	195,7	4,182	0,038
C	2	31,4	1,013	0,388	45,4	1,023	0,385
$R \times C$	2	3,1	0,100	0,906	10,4	0,234	0,794
Erro	14	217,1			310,8		

Neste caso, as hipóteses de Tipo IV testadas, automaticamente, foram:

- $H_0(R)$: $(\mu_{12} + \mu_{13})/2 = (\mu_{32} + \mu_{33})/2$ e $\mu_{23} = \mu_{33}$
- $H_0(C)$: $(\mu_{11} + \mu_{21})/2 = (\mu_{13} + \mu_{23})/2$ e $(\mu_{12} + \mu_{32})/2 = (\mu_{13} + \mu_{33})/2$
- $H_0(R \times C)$: $\mu_{11} - \mu_{13} - \mu_{21} + \mu_{23} = 0$ e $\mu_{12} - \mu_{13} - \mu_{32} + \mu_{33} = 0$.

Ao nível de significância de 5%, a decisão relativa à existência de efeito do fator R difere com o tipo de abordagem considerada. Com base nos dados originais não

se rejeita a hipótese de ausência de efeito, decidindo-se o contrário com base nas ordens das observações.

Contudo, por ex., quando se troca a ordem dos níveis do fator R (Tabela 5), as somas de Tipo IV associadas a este fator também se alteram bem como a decisão final (Tabela 6). Neste caso, tanto com base nos dados originais como nas ordens das observações conclui-se que existe efeito do fator R na variável resposta. Esta alteração deve-se ao facto de terem sido testadas outras hipóteses para o fator R: $H_0(R)$: $(\mu_{12} + \mu_{13})/2 = (\mu_{32} + \mu_{33})/2$ e $(\mu_{21} + \mu_{23})/2 = (\mu_{11} + \mu_{13})/2$.

Tabela 5 – Troca da ordem das linhas da Tabela 3.

	C_1	C_2	C_3
R_2	11; 17; 18		8; 10; 17
R_3		6; 9	3; 5; 8; 16
R_1	2; 7; 8; 10	2; 5	4; 5; 6

Tabela 6 – Resultados da ANOVA aplicada aos valores observados e às ordens.

		Valores observados			Ordens das observações		
$F.V.$	$g.l.$	$S.Q.$	F	p	$S.Q.$	F	p
R	2	191,8	6,185	0,012	310,7	6,998	0,008

3.1.2 Modelo das médias

Quando se aplica o modelo das médias aos dados originais e às ordens das observações obtêm-se conclusões similares (Figura 1). A registar apenas que o tratamento $R_1 \times C_3$ é posicionado em subgrupos diferentes consoante a análise tenha recaído sobre os dados originais ou sobre as ordens das observações.

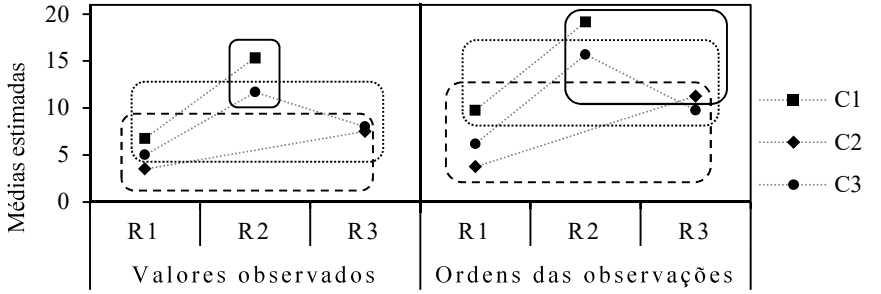

Figura 1 – Subconjuntos homogéneos das médias. (As médias localizadas no mesmo retângulo não são significativamente diferentes a 5%).

3.1.3 ANOVA com um fator

Na ANOVA com um fator a hipótese testada corresponde a

$$H_0: \mu_{11} = \mu_{12} = \mu_{13} = \mu_{21} = \mu_{23} = \mu_{32} = \mu_{33}$$

As conclusões da ANOVA aplicada quer aos dados originais quer às ordens das observações é a mesma, i.e., rejeitar H_0 (Tabela 7).

Tabela 7 – Resultados da ANOVA aplicada aos valores observados e às ordens.

		Valores observados				Ordens das observações			
$F.V.$	$g.l.$	$S.Q.$	$M.Q.$	F	p	$S.Q.$	$M.Q.$	F	p
Fator	6	272,1	45,3	2,924	0,046	453,3	75,5	3,403	0,028
Erro	14	217,1	15,5			310,8	22,2		

O próximo passo é formar as comparações de interesse (contrastes). A título ilustrativo, apresentam-se apenas os contrastes que permitem testar a interação e o efeito do fator R (Tabela 8). Ao nível de significância de 5%, as conclusões não são concordantes apenas na hipótese que compara o nível R_1 com R_2 no nível C_3. Com base nos dados originais não há evidência estatística de que os níveis R_1 e R_2 difiram no nível C_3 concluindo-se o contrário quando se consideram as ordens das observações.

Tabela 8 – Resultados dos contrastes.

		Valores observados			Ordens das observações		
Contrastes	$g.l.$	$\hat{c}$	t	p	$\hat{c}$	t	p
$H_0(R \times C)$:							
$\mu_{11} - \mu_{13} - \mu_{21} + \mu_{23} = 0$	14	-1,92	-0,435	0,670	0,08	0,016	0,988
$\mu_{12} - \mu_{13} - \mu_{32} + \mu_{33} = 0$	14	-1,00	-0,202	0,843	-3,92	-0,661	0,520
$H_0(R)$:							
$(\mu_{11} + \mu_{13})/2 = (\mu_{21} + \mu_{23})/2$	14	-7,6	-3,464	0,004	-9,5	-3,591	0,003
$(\mu_{12} + \mu_{13})/2 = (\mu_{32} + \mu_{33})/2$	14	-3,5	-1,413	0,180	-5,5	-1,870	0,083
$\mu_{11} = \mu_{21}$	14	-8,6	-2,854	0,013	-9,4	-2,617	0,020
$\mu_{12} = \mu_{32}$	14	-4,0	-1,016	0,327	-7,5	-1,592	0,134
$\mu_{13} = \mu_{23}$	14	-6,7	-2,074	0,057	-9,5	-2,470	0,027
$\mu_{13} = \mu_{33}$	14	-3,0	-0,998	0,335	-3,6	-0,996	0,336
$\mu_{23} = \mu_{33}$	14	3,7	1,219	0,243	5,9	1,644	0,122

3.2 Exemplo 2

Como segundo exemplo, considere-se o conjunto de dados apresentado na Tabela 9. À semelhança do exemplo anterior, verificámos que os pressupostos da ANOVA paramétrica eram válidos (Teste Shapiro-Wilk: $W = 0{,}967$, $p = 0{,}676$; Teste Levene: $F = 0{,}547$, $p = 0{,}764$).

Tabela 9 – Delineamento com dois fatores com células omissas (adaptado de Kleinbaum *et al.*, 1998).

	C_1	C_2	C_3
R_1	20; 21; 23		8; 10; 12
R_2		6; 9	3; 5; 8; 8
R_3	2; 7; 8; 11	1; 5	2; 4; 9

3.2.1 Somas de quadrados de Tipo IV

Na Tabela 10 apresentam-se os resultados relativos às somas de quadrados de Tipo IV associadas aos dados da Tabela 9, quer com base nos valores observados quer nas ordens dos valores observados. Estes resultados foram obtidos com auxílio do *software* SPSS.

Tabela 10 – Resultados da ANOVA aplicada aos valores observados e às ordens.

		Valores observados			Ordens das observações		
$F.V.$	$g.l.$	$S.Q.$	F	p	$S.Q.$	F	p
R	2	299,1	18,834	<0,001	264,0	6,345	0,011
C	2	153,2	9,649	0,002	49,9	1,199	0,331
$R \times C$	2	70,4	4,431	0,032	26,9	0,646	0,539
Erro	14	299,1			291,3		

Neste exemplo, as hipóteses de Tipo IV testadas, automaticamente, foram:

- $H_0(R)$: $(\mu_{11} + \mu_{13})/2 = (\mu_{31} + \mu_{33})/2$ e $(\mu_{22} + \mu_{23})/2 = (\mu_{32} + \mu_{33})/2$
- $H_0(C)$: $(\mu_{11} + \mu_{31})/2 = (\mu_{13} + \mu_{33})/2$ e $(\mu_{22} + \mu_{32})/2 = (\mu_{23} + \mu_{33})/2$
- $H_0(R \times C)$: $\mu_{11} - \mu_{13} - \mu_{31} + \mu_{33} = 0$ e $\mu_{22} - \mu_{23} - \mu_{32} + \mu_{33} = 0$.

Relativamente à interação entre os dois fatores obtêm-se conclusões contraditórias, consoante a abordagem considerada. Com base nos dados originais, conclui-se que existe interação significativa ao passo que com base nas ordens das observações não há evidência estatística dessa interação. Este resultado evidencia que existência de interação nos dados originais não implica a existência de interação nas ordens das observações, tal como já foi salientado por vários autores (e.g., Hora & Conover, 1984). Além disso, ambas as estatísticas de teste dos efeitos principais são significativas quando se consideram os dados originais, ao passo que com base nas ordens das observações a estatística associada ao efeito C não é significativa.

3.2.2 Modelo das médias

Considerando o modelo das médias, constituem-se três subconjuntos com os dados originais e apenas dois com base nas ordens das observações (Figura 2). A principal diferença reside no facto de com base nos dados originais se concluir que a média do tratamento $R_1 \times C_1$ difere das restantes médias, mas a média das ordens

deste tratamento não difere da média das ordens dos tratamentos $R_1 \times C_3$ e $R_2 \times C_2$.

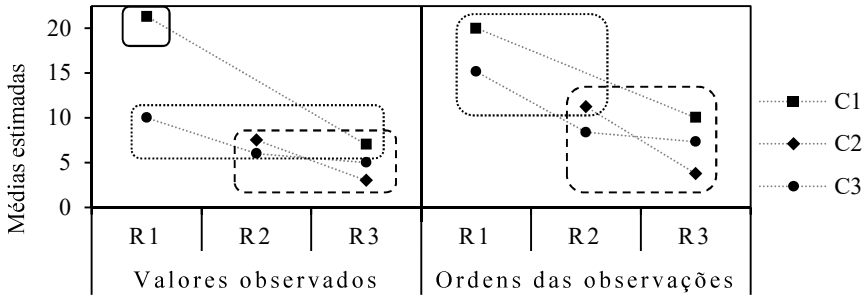

Figura 2 – Subconjuntos homogéneos das médias.

3.2.3 ANOVA com um fator

Na ANOVA com um fator a hipótese testada corresponde a

$$H_0: \mu_{11} = \mu_{13} = \mu_{22} = \mu_{23} = \mu_{31} = \mu_{32} = \mu_{33}.$$

Quer com base nos dados originais quer com base nas ordens das observações, conclui-se que os resultados das ANOVAS são concordantes e conduzem à mesma decisão (dados originais: $F_{6;14} = 13{,}297,\ p < 0{,}001$; ordens das observações: $F_{6;14} = 3{,}783, p = 0{,}019$).

Tendo em conta os resultados obtidos quando se consideraram as somas de Tipo IV, focamo-nos apenas nos contrastes que permitem testar a existência de interação.

Só concluímos que existe interação significativa ($\alpha = 5\%$) se considerarmos os dados originais e usarmos a hipótese definida no primeiro contraste (Tabela 11).

Portanto, a decisão varia com o contraste considerado.

Tabela 11 – Resultados dos contrastes.

		Valores observados			Ordens das observações		
Contrastes	$g.l.$	$\hat{c}$	t	p	$\hat{c}$	t	p
$H_0(R \times C)$:							
$\mu_{11} - \mu_{13} - \mu_{31} + \mu_{33} = 0$	14	9,33	2,963	0,010	2,17	0,425	0,677
$\mu_{22} - \mu_{23} - \mu_{32} + \mu_{33} = 0$	14	3,50	0,987	0,340	6,46	-1,125	0,279

4 Conclusão

As hipóteses de Tipo IV testadas automaticamente pelos *software* dependem da posição em que surgem as células omissas e do seu número. Estas hipóteses não são únicas e não comparam exatamente os mesmos valores médios, o que pode dar

origem a decisões discordantes. Nos *software* as hipóteses a testar são dadas automaticamente sendo necessário pedir informação adicional para se saber quais foram as hipóteses testadas. Recorrendo à programação, o utilizador pode escolher as hipóteses de interesse.

Com o modelo das médias e com a ANOVA com dois fatores considerando as somas de quadrados de Tipo IV podemos testar exatamente as mesmas hipóteses. No entanto, no modelo das médias é mais simples e fácil de controlar as hipóteses de real interesse.

Quando se opta pela realização dos contrastes na ANOVA com um fator, como nem todos os contrastes geram a mesma conclusão, à semelhança do que acontece quando se usam as hipóteses de Tipo IV, é preciso saber quais as hipóteses que se pretendem testar.

Com este estudo salientámos o quão importante é saber o que se pretende testar, de forma a evitar conclusões erróneas. Como as conclusões obtidas com base nos dados originais e com base nas ordens das observações nem sempre coincidem, se os dados forem quantitativos, recomendamos que primeiro se deve tentar transformar os dados originais de forma a verificarem os pressupostos da ANOVA paramétrica.

As alternativas não paramétricas à ANOVA, de um modo geral, não permitem a existência de células omissas, pelo que é necessária mais investigação nesta área.

Agradecimentos

Este trabalho é financiado por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia no âmbito do projeto «UID/MAT/04674/2019(CIMA)».

Referências

GOODNIGHT, J. H. (1980). Test of hypotheses in fixed effects linear models. *Communications in Statistics*, 9: 167-180.

HORA, S. & CONOVER, W. J. (1984). The F-statistic in the two-way layout with rank-score transformed data. *Journal of the American Statistical Association*, 79, 668–673.

KLEINBAUM, D. G., KUPPER, L. L., MULLER, K. E., & NIZAM, A. (1998) *Applied Regression Analysis and Other Multivariable Methods*. Duxbury.

MILLIKEN, G. A. & JOHNSON, D. E. (2009). *Analysis of Messy Data. Volume 1: Designed Experiments*, Second Edition, Chapman and Hall/CRC.

MONTGOMERY, D. G. (2013). *Design and Analysis of Experiments*, 8ª Edição. John Wiley & Sons.

SHAH, D. A. & MADDEN, L. V. (2004). Nonparametric analysis of ordinal data in designed factorial experiments. *Phytopathology*, 94, 33-43.