

# FAÇA VOCÊ MESMO... PROCEDIMENTO ANACOR. PASSO A PASSO

Ana Paula Couceiro Figueira  
Universidade de Coimbra, Portugal

Use ANACOR to analyze two-way contingency [or correspondence] tables or data that can be expressed as a two-way table, such as brand preference or sociometric choice data. This is also known as correspondence analysis (Manual SPSS, 1990, *Categories*, p. B-23).

Na perspectiva dos autores do Manual do SPSS<sup>1</sup> (cf. SPSS, 1990, *Categories*, p. B-24), tratando-se de dados nominais<sup>2</sup> (categoriais), os procedimentos de associação possíveis, tendo em conta o número de variáveis e os seus níveis e o grau de complexidade de análise, passam pelas técnicas *Optimal scaling*, ou tabelas de contingência ou de correspondência simples e múltipla (*two-way* e *multiway tables*). Os procedimentos *Optimal scaling* permitem associações não lineares, produzindo correlações entre variáveis. Estes procedimentos podem adoptar desde formas mais simples de associação (análise de correspondência para tabela de contingência de dupla entrada – *correspondence analysis for a two-way table*), às mais complexas, isto é, para mais que duas variáveis e para variáveis com número diferente de níveis (cf. SPSS, 1990, Manual, *Categories*, p. A-4).

Neste sentido, são, então, apontados quatro tipos de procedimentos: *ANACOR*, *HOMALS*, *PRINCALS* e *OVERALS* (cf. SPSS, 1990, Manual, *Categories*, p. B-23 e segs.), considerados extensões de técnicas estatísticas clássicas, tipo análise em componentes principais e análise correlacional canónica, possibilitadores da acomodação de variáveis com níveis de medida diferentes (cf. SPSS, 1990, Manual, *Categories*, p. B-24). Somos, ainda, de referir que estas análises permitem revelar, de forma visual, bidimensional, através de gráficos (*plots*), as relações entre as variáveis e entre os seus diferentes níveis.

Embora todos estes procedimentos tenham em comum a análise da homogeneidade<sup>3</sup> dos dados, tendo subjacente como pressuposto a redução das dimensões<sup>4</sup> (análise multivariada), especificamente, a partir do procedimento *ANACOR*, sendo um procedimento para duas variáveis nominais, com vários níveis, analisa os dados a partir de tabelas de correspondência<sup>5</sup> (*two-way tables*), sendo que nas diversas células de cruzamento recaem as frequências observadas dos diferentes níveis das duas categorias (cf. Quadro 1). De

---

<sup>1</sup> SPSS – *Statistical Package for the Social Sciences*- Programa para análise estatística nas Ciências Sociais.

<sup>2</sup> Regra geral, sumariados em tabelas de contingência simples (medida de associação, tendo subjacente o *chi-quadrado*). De facto, embora os resultados a partir dos métodos *ANACOR* e *HOMALS* sejam similares aos obtidos com o método mais analítico - teste - eles são mais facilmente visualizáveis através dos gráficos das suas dimensões, ou seja, permitem a análise das relações entre as variáveis em questão de forma mais sensível e directa (Pan et al., 2001).

<sup>3</sup> «A análise da homogeneidade pode ser, pois, pensada como uma análise de componentes principais, mas para dados nominais. Se as variáveis sobre análise são do tipo numérico (intervalo ou proporcionais) e se é assumida uma associação linear entre elas, então o procedimento correcto será a análise em componentes principais, disponível no *SPSS FACTOR*. (...) os factores encontrados pela análise em componentes principais são, de algum modo, equivalentes às dimensões encontradas pelo procedimento *HOMALS* (...)» (SPSS, 1990, Manual, *Categories*, p. B-26).

<sup>4</sup> Ou seja, as relações entre as variáveis são representadas em poucas dimensões (duas ou três). A sua representação pode adoptar, igualmente, a forma de mapas perceptivos (*perceptual mapping*).

<sup>5</sup> Ou tabela de dupla entrada, contemplando linhas e colunas, correspondentes às duas variáveis em apreço.

resto, a análise de correspondência simples restringe-se às tabelas de dupla entrada, “simple correspondence analysis is limited to two-way tables” (Manual do SPSS, 1990, *Categories*, p. B-25).

De facto, “The ANACOR procedure analyzes correspondence tables. A correspondence table is any two-way table whose cells contain some measurement of correspondence between the rows and the columns” (Manual do SPSS, 1990, *Categories*, p. B-31). Deste modo, a medida de correspondência pode ser uma indicação da similitude, afinidade, confusão, associação ou interação entre as variáveis da linha e da coluna (Manual do SPSS, 1990, *Categories*, p. B-25).

Correspondence analysis is a technique used to summarize the information in a two-way table, such as a contingency table. The row and column variables of the two-way table are assumed to be measured at the nominal level of measurement: therefore the values of the row and column variables represent unordered categories. The SPSS CROSSTABS procedure can also be used to analyze contingency tables – but ANACOR provides a graphic summary in the form of plots which show the relationships between categories of the two variables (Manual SPSS, 1990, *Categories*, p. B-25).

De facto, a análise da relação entre duas variáveis nominais ou categoriais pode ser realizada por via do qui-quadrado. Todavia, quando estas variáveis possuem mais do que dois níveis, é aconselhável a utilização do procedimento ANACOR, ou então, a utilização em concomitância. “For small tables, direct examination of simple statistics such as the row or column percentages might reveal the pattern in the table. The utility of correspondence analysis lies in analysis of large tables, where pattern detection and summarization of the variables’ association is often difficult. If there is an association between the row and column variables – i.e., if the chi-square value is significant – ANACOR may help reveal the nature of the relationship.” (Manual SPSS, 1990, *Categories*, p. B-25).

Neste sentido, para melhor explorar os dados relativos às eventuais relações entre duas variáveis categoriais ou nominais, o procedimento mais aconselhável será a dualidade Qui-Quadrado e ANACOR. Enquanto o Qui-Quadrado evidencia a intensidade da relação entre as duas variáveis, o ANACOR permite a representação gráfica da natureza das relações existentes, distribuindo os dados, fornecendo as suas coordenadas, onde os níveis associados se tendem a aproximar. Ainda, permite uma análise minuciosa entre os diversos níveis das variáveis, de forma isolada, variável a variável, ou de forma associada, tendo em conta as duas variáveis e os seus diferentes níveis.

O ANACOR hierarquiza a informação por ordem decrescente do grau de explicação do fenómeno em estudo, através dos valores singulares ou valores próprios, que medem a contribuição de cada dimensão para a explicação da variação dos dados (cf. Pestana & Gageiro, 2000, p. 360). Os valores próprios são o coeficiente de correlação R de Pearson entre os *scores* das duas variáveis (em linha e em coluna), ou seja, os valores próprios são as correlações entre os resultados em linha ou em coluna para cada dimensão. O quadrado de cada valor próprio designa-se por inércia das dimensões, que mede a importância de cada dimensão. De realçar, ainda, que a inércia, para cada dimensão, corresponde ao quadrado do valor próprio ou simples (*singular value*), ou seja, a correlação entre os resultados das linhas e das colunas, para cada dimensão, isto é, o contributo de cada dimensão para a explicação da variação dos dados. O quociente entre a inércia de cada dimensão e a inércia total dá a proporção de variância explicada pela dimensão (Pestana &

Gageiro, 2000, p. 360), sendo que a inércia total, quando multiplicada pela dimensão da amostra (N), corresponde ao valor do teste qui-quadrado.

## UM EXEMPLO ILUSTRATIVO (DADOS FICTÍCIOS)

Assim, para analisar, rapidamente, a associação ou relação entre Opção de Voto (Política) e Opção Religiosa, ou posição/prática religiosa (cf., igualmente, Alferes, 1997), solicitar os comandos (cf. SPSS) Statistics ou Analyze, Sumarize, Crosstabs [seleccionar as variáveis para as linhas (*rows*) e para as colunas (*columns*)], o que permite analisar o valor do Chi-Square (Pearson)<sup>6</sup>. Esta análise permite, ainda, a obtenção da distribuição dos resultados ao nível das frequências e percentagens.

Assim, variável **Opção Política**, cujos níveis poderão ser o **Partido A** o **Partido B**, **Partido C**, **Partido D** e **Sem Opção**;

A variável **Opção Religiosa**, cujos níveis poderão ser **Católico Praticante**, **Católico Não Praticante**, **Muçulmano**, **Judeu**, **Sem Opção**;

Após a construção da base de dados, em que

**variable name:** designação do nome da variável (exemplo: Opção Política)

**type:** numeric

**labels** (exemplo):

- 1= "Partido A"
- 2= "Partido B"
- 3= "Partido C"
- 4= "Partido D"
- 5= "Sem Opção"

**Nota:** proceder do mesmo modo para a situação da opção religiosa.

N=30 sujeitos

2 variáveis nominais, cada uma com 5 níveis

**Variáveis**  
Opção Política  
Opção Religiosa

### Níveis da variável Opção Política (coluna)

- 5 níveis
- 1= "Partido A"
- 2= "Partido B"
- 3= "Partido C"
- 4= "Partido D"
- 5= "Sem Opção Partidária"

<sup>6</sup> Neste sentido, caso o valor do qui-quadrado seja [ (12 dp, N=193)=16.4416, p=.172] [na situação de 4 níveis para a opção política e 5 níveis para a opção religiosa (dp=12), com 193 sujeitos, o valor de Pearson é de 16.44164, com um nível de significância de .172, i.é., não significativo], significa que as duas variáveis são independentes (não estão relacionadas), isto é, não existe uma relação entre, por exemplo, a opção política e a opção religiosa.

*Grosso modo*, somos de referir que o teste Qui-Quadrado estuda a relação entre duas variáveis nominais. Porém, quando as duas variáveis nominais têm muitos níveis é mais vantajoso utilizar o procedimento ANACOR.

Comparando os resultados por esta via com os obtidos através do ANACOR, podemos dizer que o valor de Pearson (mede a intensidade da associação entre variáveis intervalo/rácio é igual ao valor da inércia total x N . Assim, por exemplo, caso o total da inércia seja = .08519 e se o N=193, então, temos um valor de Pearson de 16.4416.

**Níveis da variável Opção Religiosa (linha)**

5 níveis

- 1= "Católico Praticante"
- 2= "Católico Não Praticante"
- 3= "Muçulmano"
- 4= "Judeu"
- 5= "Sem Opção Religiosa "

Deste modo,

valor do **CHI-SQUARE (qui-quadrado)**

[ (16 dp, N=30)=18.72103, p=.28339]

O que significa que as duas variáveis são independentes, i.é., não estão relacionadas.

Assim, podemos dizer que não se regista uma relação entre a opção religiosa e a opção política.

**Quadro 1**

Resultados obtidos a partir do *crosstabs*: Freqüências registadas em cada célula

//////////	Católico P.	Católico N. P.	Muçulmano		Judeu
Sem Opção					
Partido A	2	1	0	0	2
Partido B	0	4	1	1	1
Partido C	0	0	3	0	2
Partido D	2	1	0	1	2
Sem Opção	1	2	1	1	2

**ANACOR**

comandos a utilizar

opta-se pela *Statistics* (6.1 para Macintosh) ou *Analyze* (11.0 para Windows) → *data reduction*  
*correspondence analysis* → *row (opção religiosa), define range (mínimo 1 e máximo 5) e*  
*column (opção política), define range (mínimo 1 e máximo 5) → 2 dimensões → opção:*  
*normalização canónica ou principal*<sup>7</sup>

<sup>7</sup> Os comandos para a **Normalização** especificam os métodos possíveis: normalização canónica, principal, rprincipal e cprincipal, para os resultados das linhas e das colunas. Somente os resultados e as variâncias são afectados. Os contributos e os perfis não sofrem alterações.

Assim, na **Normalização Canónica**, para cada dimensão, as linhas são a ponderação média das colunas dividida pelo seu valor próprio, e as colunas são a ponderação média das linhas dividida pelo seu valor próprio. Esta é a opção, por defeito, tomada pelo procedimento, caso o investigador não tome qualquer opção específica. Este tipo de normalização deve ser utilizado quando se está interessado, fundamentalmente, na análise das diferenças ou similitudes entre as variáveis.

Quanto à **Normalização Principal**, é uma aproximação das distâncias do qui-quadrado, i.é., as distâncias entre os pontos da linha e os pontos da coluna. As distâncias representam a distância entre a linha ou a coluna, correspondendo ao seu perfil médio. Este tipo de normalização deve ser utilizado quando se está interessado quer nas diferenças entre os níveis da variável da linha quer nas diferenças entre os níveis da variável da coluna, mas não nas diferenças entre as variáveis.

Na **RPrincipal**, as distâncias entre os pontos da linha são aproximações das distâncias do qui-quadrado. Este método maximiza as distâncias entre os pontos da linha. Deve ser utilizado quando se está interessado, fundamentalmente, nas diferenças ou similitudes entre os níveis ou categorias da variável da linha.

Na **CPrincipal**, as distâncias entre os pontos da coluna são aproximações das distâncias do qui-quadrado. Este método maximiza as distâncias entre os pontos da coluna. Deve ser utilizado quando se está interessado, fundamentalmente, nas diferenças ou similitudes entre os níveis ou categorias da variável da coluna.

A opção depende, pois, do objectivo, ou objectivos, do estudo.

Variáveis da ANACOR = Opção Política (5) Opção Religiosa (5) / 2 dimensões / Print = Freq Eigen Discrim Quant / Plot = Quant.

ANACOR table=opção política (1,5) por opção religiosa (1,5). A análise irá utilizar uma tabela de contingências simples, isto é, uma tabela simples de dupla entrada, que reproduz a distribuição de frequências.

Os resultados nas dimensões significam as coordenadas de localização dos níveis das variáveis nas dimensões.

Para além da representação gráfica parcelar (*plots*), tendo em conta os agrupamentos encontrados para os níveis de cada variável sob análise, este procedimento permite, ainda, a visualização conjunta, dos diversos níveis das duas variáveis.

Neste sentido, as informações mais relevantes a reter e a analisar serão:

N=30

Dp=16

Inércia total=.62403

Pearson=18.72 [.62403 x 30]

Resultados obtidos:

**\* Variável Opção Política:**

Nível da variável	Frequência	Dimensão 1	Dimensão 2
Partido A	5	-.675	.765
Partido B	7	-.019	-1.128
Partido C	5	1.613	.511
Partido D	6	-.680	.433
Sem Opção	7	-.068	-.155

**\* Variável Opção Religiosa:**

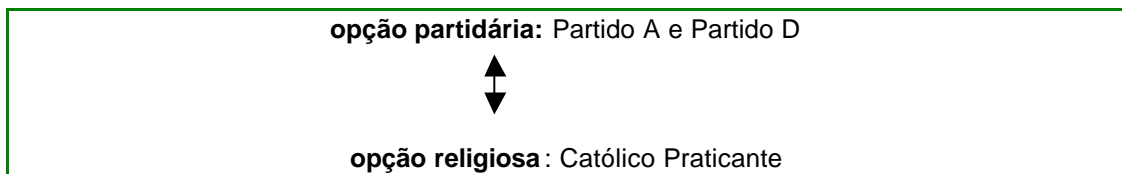
Nível da variável	Frequência	Dimensão 1	Dimensão 2
Católico Praticante	5	-.921	.932
Católico Não Praticante	8	-.325	-.942
Muçulmano	5	1.576	.104
Judeu	3	-.424	-.589
Sem Opção	9	.067	.458

Pese embora, a partir destes dados, com esta amostra, se registre que as duas variáveis não estão associadas, de forma significativa, não significa que não possamos analisar as relações ou agrupamentos existentes.

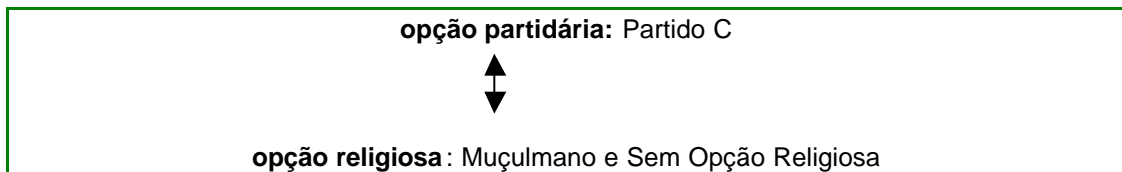
Assim, e segundo uma leitura mais ou menos grosseira, obtiveram-se 3 grupos distintos e interpretáveis,

**Leitura dos resultados, nos 3 subgrupos encontrados:**

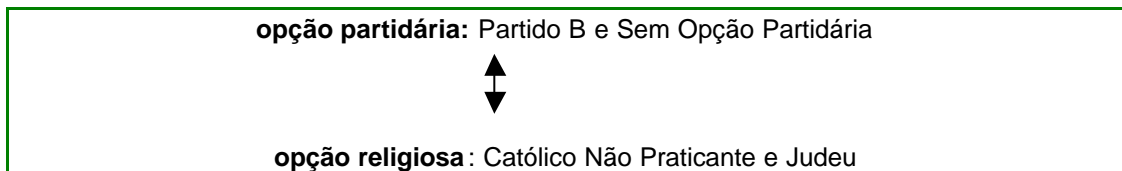
- 1) Os sujeitos que assumem uma opção partidária do tipo A ou do tipo D (Partidos A e D) referem, como opção religiosa, o católico praticante, ou, dito de outro modo, os católicos praticantes predominam nos partidos A e D,



- 2) Os sujeitos que assumem uma opção partidária do tipo C (Partido C) referem, como opção religiosa, o ser Muçulmano ou Sem Opção Religiosa,



- 3) Os sujeitos que assumem uma opção partidária do tipo B (Partido B) e os que referem não ter opção partidária estão associados às opções religiosas Judeu e Católico Não Praticante,



## BIBLIOGRAFIA RECOMENDADA

- ALFERES, V. R. (1997). *Investigação científica em Psicologia*. Teoria & Prática (pp. 123-124 e pp. 186-190). Coimbra: Almedina.
- BIDARRA, M. G. (1994). *Estudo dos processos sociocognitivos e análise das práticas de formação*. Dissertação de Doutoramento em Psicologia Pedagógica não publicada. Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra.
- FIGUEIRA, A. P. M. C. C. (2001). *Das epistemologias pessoais à epistemologia das práticas educativas. Estudo das vivências metodológicas numa amostra de professores dos 3º Ciclo do Ensino Básico e do Ensino Secundário, das disciplinas de Matemática, Português e Inglês*. Dissertação de Doutoramento em Psicologia Pedagógica não publicada. Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra.
- <http://forrest.psych.unc.edu/research/vista-frames/winvista64plugins.html>
- PAN, B., XU, F., WU, Z., AOYAGI, K., et al. (2001). Study on the relationship of indoor air quality indexes by Correspondence Analysis and Homogeneity Analysis. *Acta Med. Nagasaki*, 46, 33-37.
- PEREIRA, A. (1999). *SPSS - Guia prático de utilização. Análise de dados para Ciências Sociais e Psicologia*. Lisboa: Edições Sílabo.
- PESTANA, M. H., & Gageiro, J. N. (1998). *Análise de dados para Ciências Sociais. SPSS e complementariedades* (1ª ed.). Lisboa: Edições Sílabo.
- PESTANA, M. H., & Gageiro, J. N. (2000). *Análise de dados para Ciências Sociais. A complementariedade do SPSS* (2ª ed., pp. 360-368). Lisboa: Edições Sílabo Lda.
- SPSS (1990). *SPSS for the Macintosh* [Computer program - Version 6.1]. Chicago: IL Author.
- SPSS (1990). *SPSS: Advanced statistics user's guide* [Computer manual]. Chicago: IL Author.
- SPSS (1990). *SPSS: Base system user's guide* [Computer manual]. Chicago: IL Author.
- SPSS (1990). *SPSS: Categories* [Computer manual] (pp. A-2-A-4, pp. B-23-B-29 e pp. B-31-B-47). Chicago: IL Author.
- SPSS (1990). *SPSS: Tables* [Computer manual]. Chicago: IL Author.

**Contactar**

**Revista Iberoamericana de Educación**

**Principal OEI**