

CAPÍTULO 8

ANÁLISE FACTORIAL DE CORRESPONDÊNCIAS

Neste capítulo vão analisar-se os procedimentos Anacor, Homals e Princals, usados em tabelas de contingência, isto é, em quadros de distribuição de frequências resultantes do cruzamento de duas ou mais variáveis qualitativas.

- 1) A Anacor ou análise de correspondência simples, é aplicada quando se pretende estudar a relação entre duas variáveis nominais.
- 2) O Homals ou análise da homogeneidade, é aplicado quando se pretende estudar a relação entre mais de duas variáveis nominais.
- 3) O Princals é usado quando se pretende estudar a relação entre variáveis qualitativas de vários níveis (nominais e ordinais).

Todos estes procedimentos são técnicas exploratórias e não confirmatórias, com vista a descobrirem possíveis relações entre variáveis num espaço multidimensional (normalmente a duas dimensões), não sendo portanto testes de hipóteses. O Homals e o Princals permitem ainda analisar a relação entre casos, também designados por objectos pela disposição em linha na base de dados, ou entre variáveis e casos.

A análise factorial de correspondências permite a hierarquização da informação disponível por ordem decrescente do grau de explicação do fenómeno em estudo e permite ainda produzir variáveis compósitas que resumem as relações existentes entre as categorias dos atributos observados (Moreira, M. M. Raúl; Reis, Elizabeth, 1988).

A organização da informação de forma hierarquizada é feita através dos "eixos factoriais" ou "vectores próprios" a que estão associados "valores próprios" que medem a contribuição de cada eixo para a explicação da variação dos dados observados. Uma vez recolhidos os eixos factoriais mais importantes, de acordo com um critério quantitativo, é possível projectar as categorias de cada variável nos seus eixos mais significativos, eliminando aqueles cuja contribuição para a explicação do fenómeno poderá ser considerado pouco importante (Moreira, M. M. Raúl; Reis, Elizabeth, 1988).

1. ANACOR

A Anacor, analisa tabelas de correspondência para duas variáveis nominais, onde cada célula representa as frequências observadas das suas respectivas categorias.

Quando se tem mais de duas variáveis nominais, pode criar-se uma nova variável que seja a combinação de duas iniciais, ou em alternativa recorrer-se ao Homals.

A Anacor pode ser usada como complemento do teste do Qui-Quadrado (capítulo 2). Quando o valor deste teste leva à rejeição da hipótese nula da independência das duas variáveis, a análise das frequências dificilmente revela comportamentos observados nos dados quando existem muitas categorias em linha e em coluna. Nesta situação torna-se vantajoso utilizar a Anacor, que permite ainda representar graficamente a natureza das relações existentes, onde as categorias semelhantes são colocadas mais próximas umas das outras.

Os eixos dos gráficos são as dimensões usadas para representar a variação existente nas categorias, cujo número de dimensões se obtém através da fórmula:

$$\min(l; c) - 1$$

Onde, min significa o número mínimo de categorias; e l e c representam respectivamente o número de categorias em linha e em coluna.

As coordenadas de cada ponto nos gráficos chamam-se *scores* cujos valores são quantificações óptimas atribuídas às linhas e colunas de cada dimensão.

A Anacor hierarquiza a informação por ordem decrescente do grau de explicação do fenómeno em estudo, através dos valores singulares (valores próprios) que medem a contribuição de cada dimensão para a explicação da variação dos dados.

Os valores singulares são o coeficiente de correlação R de Pearson entre os *scores* em linha e em coluna, existindo um valor para cada dimensão, a primeira dimensão é a que mais explica a variação existente nos dados, a segunda dimensão a seguinte e assim sucessivamente.

O quadrado de cada valor singular designa-se por inércia das dimensões, que constitui uma medida da importância de cada dimensão. A proporção de variância explicada por cada dimensão é igual ao quociente entre a respectiva inércia e a inércia total.

Cada categoria contribui para a inércia de cada dimensão, em que os maiores valores indicam as categorias dominantes nessa dimensão, ou seja as que mais se diferenciam das restantes.

A distribuição da inércia entre os scores em linha e em coluna pode ser feita essencialmente por quatro métodos, designados por métodos de normalização:

- Em linha ou *row principal*, que maximiza as distâncias entre as categorias dessa variável, usando-se quando interessa saber como as categorias da variável em linha variam entre si;
- Em coluna ou *column principal*, que maximiza as distâncias entre as categorias dessa variável, usando-se quando interessa saber como as categorias da variável em coluna variam entre si;
- *Symmetrical*, usando-se quando interessa saber as diferenças ou semelhanças entre as duas variáveis;
- *Principal*, usando-se quando interessa saber as semelhanças ou diferenças entre as categorias de cada variável por si só, e não nas diferenças entre as duas variáveis. Assim, não se utilizam neste caso os gráficos conjuntos.

As normalizações *Symmetrical* e *Principal* distribuem a inércia simetricamente entre as categorias em linha e em coluna, não maximizando as distâncias como nas duas primeiras normalizações.

As diferentes normalizações não influenciam os valores singulares, nem a inércia, nem as contribuições de cada categoria de uma variável para uma dada dimensão, mas alteram os scores e as variâncias.

As estatísticas produzidas na Anacor permitem ver a qualidade com que as categorias em linha e em coluna estão representadas nas dimensões.

O quadro 8.1.1 indica as opções de voto e a prática religiosa de uma amostra aleatória de 1008 adultos do concelho X.

Pretende-se ver como se distinguem as preferências religiosas nos partidos políticos, pelo que se vai proceder na Anacor a uma normalização em linha (*row principal*).

Quadro 8.1.1

		PARTIDOS				Total
		1 A	2 B	3 C	4 D	
1 Católica	Count	73	174	136	6	389
	Expected Count	42,5	134,7	174,8	37,0	389,0
	Adjusted Residual	6,3	5,3	-5,0	-6,8	
2 Protestante	Count	30	139	131	20	320
	Expected Count	34,9	110,8	143,8	30,5	320,0
	Adjusted Residual	-1,1	4,0	-1,7	-2,4	
3 Ateia	Count	7	36	186	70	299
	Expected Count	32,6	103,5	134,4	28,5	299,0
	Adjusted Residual	-5,7	-9,8	7,2	9,8	
Total	Count	110	349	453	96	1008
	Expected Count	110,0	349,0	453,0	96,0	1008

As frequências observadas no quadro 8.1.1 podem ser obtidas a partir dos seguintes comandos do syntax do SPSS, seguidos de Run All:

```
Data List Free/ religião partidos Freq.
Weight By Freq.
Value Labels
religião 1 "católica" 2 "protestante" 3 "ateia"
/partidos 1 "A" 2 "B" 3 "C" 4 "D".
Begin Data
1 1 73 1 2 174 1 3 136 1 4 6
2 1 30 2 2 139 2 3 131 2 4 20
3 1 7 3 2 36 3 3 186 3 4 70
END DATA.
ANACOR TABLE = religião (1,3) BY partidos (1,4).
```

Os comandos necessários para a Anacor são:

```
Statistics data Reduction Correspondence Analysis
Row religião Define range Minimum 1 Maximum 3 update
Category Constraints none Continue
Column partidos Define range Minimum 1 Maximum 4 update
Category Constraints none Continue
Model
Dimensions in Solution 2
Distance Measure Chi-Square
Normalization Method Row principal Continue
Statistics
Correspondence table
Overview of row points Overview of column points
Confidence Statistics for Row points Column points
continue
Plot
Biplot Row points Column points Continue OK
```

1.1. RESÍDUOS E TESTE DO QUI-QUADRADO

A relação entre as duas variáveis nominais pode ser analisada através do teste do Qui-Quadrado, representado no quadro 8.1.2, que tem um nível de significância de 0,000, valor inferior a 0,005, denotando a existência de uma relação de dependência entre os partidos políticos e as preferências religiosas.

Quadro 8.1.2

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	225,276 ^a	6	,000
Likelihood Ratio	240,055	6	,000
Linear-by-Linear Association	190,140	1	,000
N of Valid Cases	1008		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 28,48.

A análise dos resíduos ajustados (quadro 8.1.1) permite observar o seguinte:

- Há mais ateus (resíduos ajustados positivos) nos partidos D e C (resíduos 9,8 e 7,2) e há menos ateus nos partidos A e essencialmente no B (resíduos – 5,7 e – 9,8);
- Há mais católicos nos partidos A e B (resíduos 6,3 e 5,3) e há menos católicos nos partidos C e D (resíduos – 5 e – 6,8);
- Há mais protestantes no partido B (resíduo 4).

1.2. NORMALIZAÇÃO, GRÁFICOS E PERCENTAGENS

A normalização utilizada é a row principal, pois interessa conhecer as diferenças e as semelhanças entre os católicos, protestantes e ateus, no que respeita aos partidos políticos.

Os totais marginais em linha, quadros 8.1.3 ou 8.1.1, mostram que os ateus são menos do que os católicos e os protestantes.

Os totais marginais em coluna mostram que há um maior número de adeptos do partido C seguido dos partidos B, A e D.

Quadro 8.1.3

Correspondence Table

RELIGIÃO	PARTIDOS				Active Margin
	A	B	C	D	
católica	73	174	136	6	389
protestante	30	139	131	20	320
ateia	7	36	186	70	299
Active Margin	110	349	453	96	1008

Em termos de percentagens em linha, quadro 8.1.4, os católicos assemelham-se mais aos protestantes, o que faz com que tenham uma representação próxima nos gráficos.

Quadro 8.1.4

RELIGIÃO	PARTIDOS				Active Margin
	A	B	C	D	
católica	,188	,447	,350	,015	1,000
protestante	,094	,434	,409	,063	1,000
ateia	,023	,120	,622	,234	1,000
Mass	,109	,346	,449	,095	1,000

Estas percentagens são obtidas dividindo a frequência absoluta de cada célula pelo seu total em linha. Por exemplo, $0,188 = 73 / 389$. De igual modo, $0,109$ (designado por massa) é igual a $110 / 1008$.

1.3. DIMENSÕES: INÉRCIA E VALOR SINGULAR

O teste do Qui-Quadrado mostra que existe relação de dependência entre as preferências religiosas e os partidos políticos ($\chi^2 = 225,276$, $\alpha = 0,000$).

O número de dimensões é igual a 2, pois é o valor correspondente ao $[\min(3, 4) - 1]$.

As proporções explicadas nas dimensões 1 e 2 correspondem respectivamente a 96,3% ($0,215 / 0,223$), e a 3,7% da inércia total.

As correlações entre os scores em linha e em coluna para a dimensão 1 e 2 são respectivamente 0,464 e 0,091. O quadrado destes valores dão para cada dimensão a sua inércia, sendo neste caso igual a 0,08% para a dimensão 2.

Assim, a dimensão 1 é mais importante para interpretar os dados do que a dimensão 2.

O pequeno valor da inércia da dimensão 2 conjugado com as elevadas correlações dos scores em coluna (quadro 8.1.5), conduz à decisão de omitir a dimensão 2 para explicar a variação nos dados.

Quadro 8.1.5

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation 2
1	,464	,215			,963	,963	,025	,077
2	,091	,008			,037	1,000	,033	
Total		,223	225,276	,000 ^a	1,000	1,000		

a. 6 degrees of freedom.

1.4. VARIÂNCIA E CORRELAÇÕES

Quando a variância amostral é grande, a localização dos pontos nos gráficos não garante igual comportamento na população, o que já não acontece se a variância for pequena.

Numa solução multidimensional, se a correlação entre as dimensões for grande, pode não ser possível localizar com muita segurança um ponto na dimensão correcta. Tal não é o caso, pois o quadro 8.1.5 mostra que a correlação entre os valores singulares de ambas as dimensões é pequena (0,077).

Este quadro mostra ainda que as variâncias para os valores singulares em ambas as dimensões são muito pequenas ($0,025^2$ e $0,033^2$), garantindo-se que a Anacor produzirá a mesma solução para uma amostra semelhante da mesma população.

Os quadros 8.1.6 apresentam as variâncias e as correlações para os scores em linha e em coluna.

As variâncias (desvios padrões ao quadrado) para cada score em linha e coluna são pequenas, podendo deste modo concluir-se que a solução global dada pelo Anacor é bastante estável.

As correlações entre as duas dimensões, são pequenas para os scores em linha mas elevadas para os scores em coluna (- 0,809 e 0,509). Este resultado deve-se à normalização em linha combinada com o facto da segunda dimensão ter um valor singular muito pequeno, o que indica que a primeira dimensão é suficiente para analisar os dados.

Quadros 8.1.6

Confidence Row Points

RELIGIÃO	Standard Deviation in Dimension		Correlation
	1	2	1-2
católica	,034	,030	-,054
protestante	,051	,047	,050
ateia	,043	,015	,155

Confidence Column Points

PARTIDOS	Standard Deviation in Dimension		Correlation
	1	2	1-2
A	,150	,119	-,809
B	,067	,047	,509
C	,037	,041	-,018
D	,060	,105	,111

1.5. CONTRIBUIÇÃO

Os maiores valores da inércia de cada dimensão, correspondem às categorias dominantes nessa dimensão.

Os quadros 8.1.7 e 8.1.8 mostram as estatísticas que indicam a qualidade com que as categorias representam cada dimensão, ajudando a interpretar os dados.

Assim, em termos das variáveis em linha, os ateus constituem um ponto dominante na dimensão 1, sendo a sua contribuição para a inércia dessa dimensão de 0,665. Os protestantes contribuem mais (0,649) para a inércia da dimensão 2.

Quadro 8.1.7

RELIGIÃO	Mass	Score in Dimension		Inertia	Contribution				
		1	2		Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		Total
					1	2	1	2	
católica	,386	-,410	-,082	,068	,302	,312	,962	,038	1,000
protestante	,317	-,150	,130	,013	,033	,649	,572	,428	1,000
ateia	,297	,695	-,033	,143	,665	,039	,998	,002	1,000
Active Total	1,000			,223	1,000	1,000			

a. Row Principal normalization

A categoria em coluna que mais contribui para a inércia da dimensão 1 é o partido D. Os partidos A e B contribuem mais para a dimensão 2.

Note-se que a inércia da dimensão 2 é muito pequena (0,008), pelo que em termos gráficos os dados só vão ser interpretados em termos de distâncias horizontais, que se referem à dimensão 1.

Quadro 8.1.8

PARTIDOS	Mass	Score in Dimension		Inertia	Contribution				
		1	2		Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		Total
					1	2	1	2	
A	,109	-1,250	-2,529	,042	,171	,698	,864	,136	1,000
B	,346	-,895	,925	,062	,278	,296	,961	,039	1,000
C	,449	,551	-,049	,029	,136	,001	1,000	,000	1,000
D	,095	2,089	-,232	,089	,415	,005	1,000	,000	1,000
Active Total	1,000			,223	1,00	1,000			

a. Row Principal normalization

1.6. PERMUTAÇÕES

Por vezes é útil ordenar as categorias das linhas e das colunas, quando há razões para acreditar que nelas existe uma ordem, sem no entanto a conhecer. Esta ordem pode ser alcançada considerando os scores das linhas e das colunas como variáveis.

O quadro 8.1.9 apresenta as permutações, que neste caso mantiveram a ordem inicial.

Quadro 8.1.4

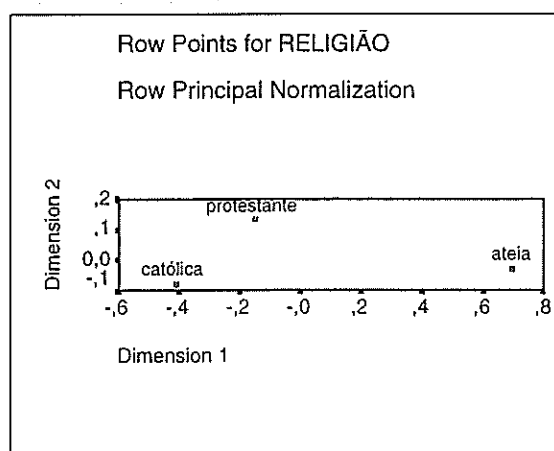
Permuted Correspondence Table According to Dimension 1

RELIGIÃO	PARTIDOS				Active Margin
	A	B	C	D	
católica	73	174	136	6	389
protestante	30	139	131	20	320
ateia	7	36	186	70	299
Active Margin	110	349	453	96	1008

1.7. GRÁFICOS

A interpretação dos quadros de frequências na dimensão 1, mostra que as categorias em linha mais próximas são mais semelhantes entre si do que as mais afastadas. A distância entre as categorias da variável em linha, vista no eixo horizontal, mostra que as categorias que mais se afastam entre si são os católicos e os ateus, indicando diferentes preferências partidárias (gráfico 8.1.1).

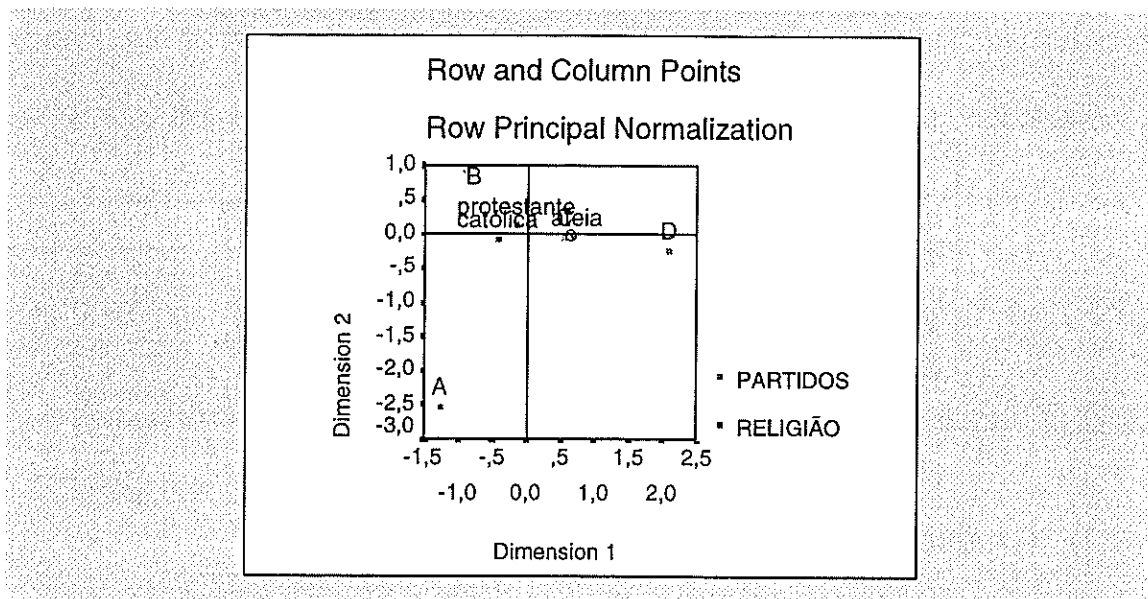
Gráfico 8.1.1



Em termos da representação gráfica das duas variáveis, gráfico 8.1.2, desenha-se um segmento de recta a partir da origem (0,0) até cada categoria em coluna (partidos), fazendo depois uma projecção ortogonal das categorias em linha (religião) para estes segmentos, revelando-se assim a natureza das relações observadas nos dados.

Verifica-se que a maioria das pessoas dos partidos C e D são ateus, e que a maioria das pessoas do partido B são protestantes, sendo católicos a maioria das pessoas do partido A.

Gráfico 8.1.2



2. HOMALS

O Homals ou análise de correspondência múltipla, é uma técnica de redução dos dados, que permite representar a relação entre três ou mais variáveis nominais em poucas dimensões, normalmente em duas ou três.

Se todas as variáveis fossem quantitativas e se existisse uma associação linear entre elas, tornava-se apropriada a análise factorial (capítulo 9).

A Homals considera-se uma análise das componentes principais para dados nominais, procedendo a uma partição dos objectos ou casos em grupos homogéneos. Quantifica as variáveis atribuindo a cada dimensão scores óptimos que permitem a maior separação entre as categorias.

Os scores associados aos casos e às categorias, são designados respectivamente por object scores e category quantifications.

Admita que se pretende analisar a relação entre o destino habitual de férias, férias, a motivação principal de férias, motivo, e as licenciaturas, curso, dos quadros superiores da empresa Y. Os resultados observados numa amostra aleatória de 25 trabalhadores constam do quadro 8.2.1.

Os destinos habituais de férias considerados são: 1 = Costa de Prata, 2 = Costa Verde, 3 = Montanhas, 4 = Algarve.

Os cursos superiores considerados são: 1 = Medicina, 2 = Psicologia, 3 = Economia e Gestão.

Os motivos principais de férias são: 1 = Fugir ao calor, 2 = Contactar com a natureza, 3 = Relaxar, 4 = Procurar aventura, 5 = Conviver.

O Homals assume que os labels das categorias são números inteiros positivos e sequenciais, começando no 1.

Quadro 8.2.1

Férias	Curso	Motivo
2	1	5
4	3	4
4	3	5
4	3	1
4	3	4
3	2	2
2	1	5
1	2	1
2	1	5
3	2	3
1	2	2
3	2	3
1	2	1
3	2	3
2	1	5
2	3	4
2	3	5
4	1	4
1	3	5
4	3	4
4	3	4
2	3	3
1	2	1
3	3	2
2	2	5

Pretende-se analisar a relação entre as características das três variáveis nominais, pelo que se podem usar os seguintes comandos:

```

Statistics Data Reduction Optimal Scaling
All variables multiple nominal
One set Homogeneity(Homals) Define
Variables férias (1 4) curso (1 3) motivação (1 5)
Dimensions in solution 2 Options
Display
Frequencies Eigenvalues Discrimination measures
Category quantification Object scores
Plots
Category quantification Object scores
Discrimination measures Continue OK

```

O tratamento dos dados pode ser obtido, igualmente, usando a syntax do SPSS onde se inscrevem as seguintes instruções, prosseguindo com Run All:

```

DATA LIST FREE /férias curso motivação.
VARIABLE LABELS
    férias "férias habituais"
    curso "Curso superior"
    motivação "motivação principal".
VALUE LABELS
    férias 1 "C.Prata" 2 "C.Verde" 3 "Montanha" 4 "Algarve"
    /curso 1 "Medicina" 2 "Psicologia" 3 "Economia e
    Gestão"
    /motivação "fugir ao calor" 2 "contacto com a natureza"
    3 "relaxar" 4 "procurar aventura" 5 "conviver".
BEGIN DATA
4 3 1
2 1 5 4 3 4 1 2 1 1 2 2 3 2 3 2 3 5 4 3 4 1 2 1
4 3 4 3 2 2 2 1 5 3 2 3 2 1 5 4 1 4 4 3 4 3 3 2
4 3 5 2 1 5 3 2 3 1 2 1 2 3 4 1 3 5 2 3 3 2 2 5
END DATA.
HOMALS VARIABLES=férias(4) curso(3) motivação(5)
/DIMENSION=2
/PRINT=EIGEN QUANT
/PLOT=DISCRIM (15) OBJECT (férias curso motivação) (10)
QUANT(15).

```

2.1. DIMENSÕES

Os valores próprios ou eigenvalues medem a informação dada por cada dimensão. O seu valor máximo é 1 e ocorre quando as categorias que caracterizam essa dimensão são por ela explicadas na totalidade.

Neste exemplo, o quadro 8.2.2 mostra que as duas dimensões explicam respectivamente 0,8312 e 0,6596 da variação dos dados. Estes valores elevados permitem a

agregação nítida das diferentes categorias, discriminando bem cada variável. Levam portanto à formação de grupos diferenciados de categorias de variáveis.

Quadro 8.2.2

Dimension	Eigenvalue
1	,8312
2	,6596

2.2. GRÁFICOS DAS QUANTIFICAÇÕES

Depois de examinar os valores próprios deve olhar-se para a quantificação dos casos ou *object scores*, que são úteis para identificar outliers e comportamentos existentes nos dados.

Os casos ou *objectos* estão próximos quando os *scores* são semelhantes para todas as variáveis.

Os gráficos 8.2.1, 8.2.2 e 8.2.3 mostram que não existem outliers e que as pessoas da Empresa Y procuram a Costa Verde para convívio, o Algarve para a procura de aventura e as montanhas pelo contacto com a natureza.

Gráfico 8.2.1

Pontos identificados pelos destinos de férias e pelas motivações

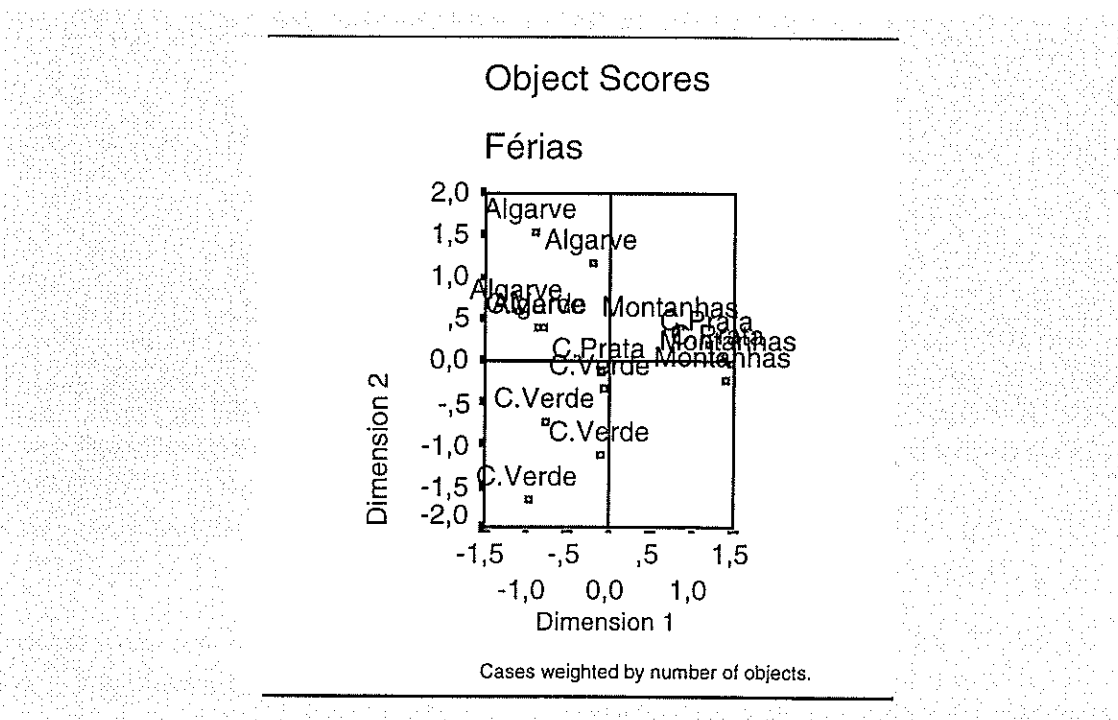


Gráfico 8.2.2

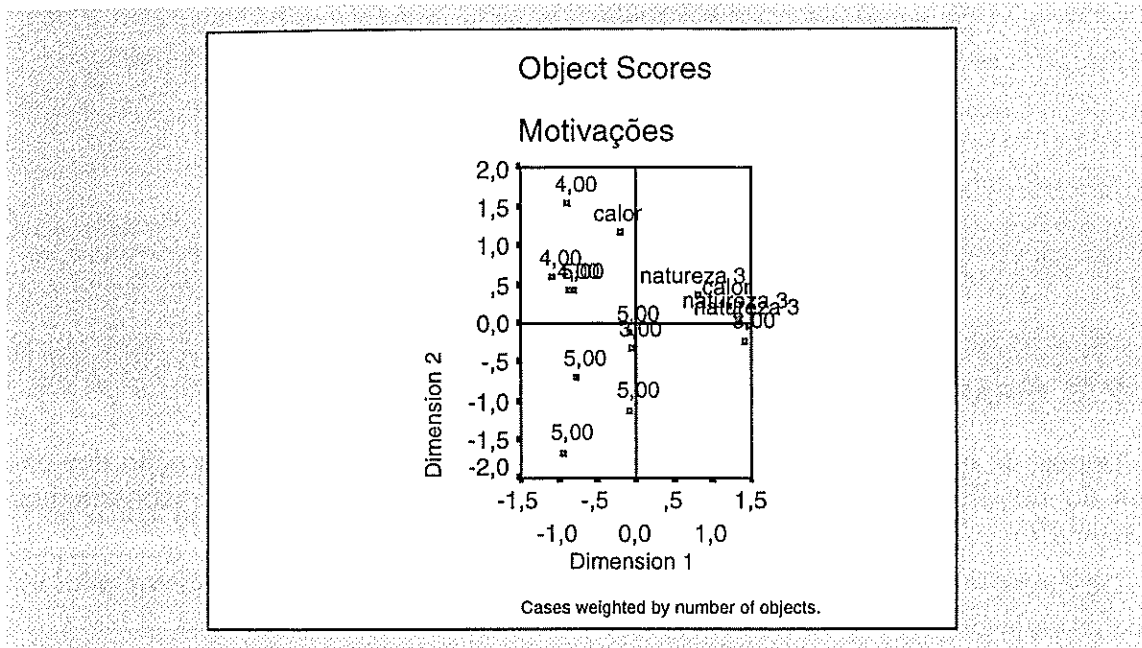
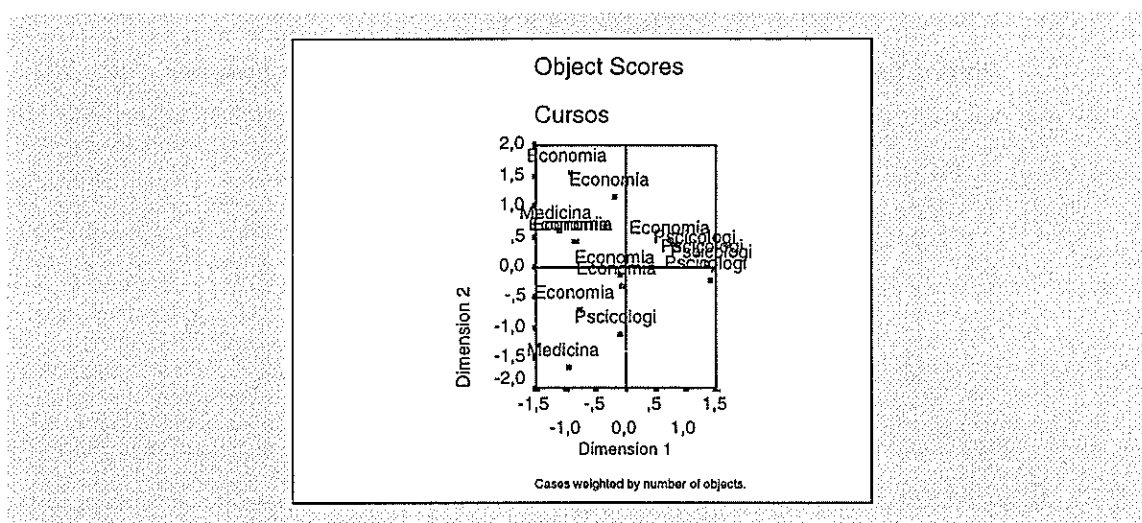


Gráfico 8.2.3

Pontos identificados pelos cursos



2.3. MEDIDAS DE DISCRIMINAÇÃO

As medidas de discriminação informam sobre as variáveis mais importantes em cada dimensão, que são as que permitem identificar o significado da dimensão.

Estas medidas assumem o valor máximo 1, quando a discriminação é perfeita e os object scores caem em grupos mutuamente exclusivos.

Só quando existem não respostas é que o valor máximo pode ser maior do que 1.

O quadro 8.2.3, mostra que a dimensão 1, representada no eixo horizontal, discrimina os cursos (score elevado na dimensão 1 e muito mais baixo na dimensão 2 das férias. Por sua vez, a dimensão 2 separa os destinos e a motivação dos cursos.

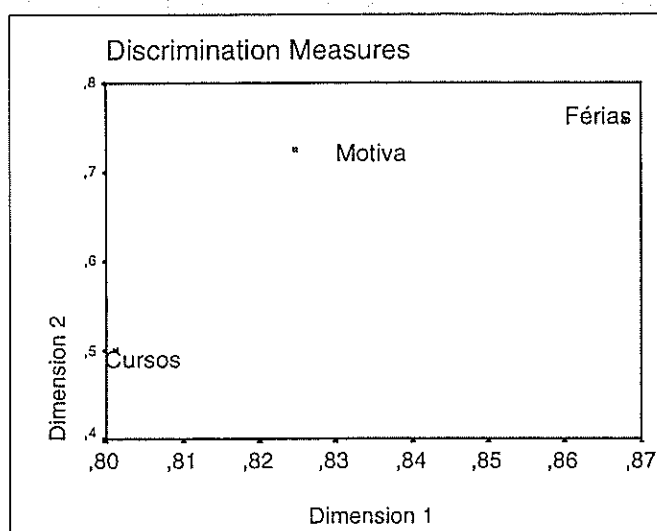
Quadro 8.2.3

Discrimination measures per variable per dimension

Variable	Dimension	
	1	2
CURSOS	,801	,499
FÉRIAS	,868	,755
MOTIVAÇÃO	,825	,724

Igual conclusão sobre as medidas de discriminação pode ser obtida no gráfico 8.2.4.

Gráfico 8.2.4



2.4. QUANTIFICAÇÕES DAS CATEGORIAS

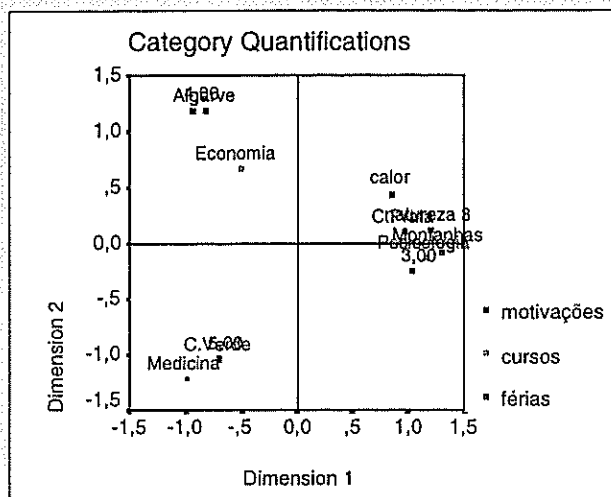
As quantificações das categorias representam as médias dos object scores dessas categorias, e são as coordenadas de cada ponto no gráfico *Category Quantification*. As categorias de diferentes variáveis estão próximas umas das outras se tiverem scores semelhantes.

Dado que os valores próprios são elevados, existe uma boa discriminação entre as variáveis, resultando num gráfico em que as observações se distanciam do (0,0) e formam grupos bem identificados.

O gráfico 8.2.5 da *Category Quantification* permite ver que dos licenciados da empresa Y, os médicos vão essencialmente para a Costa Verde para conviver; os

economistas vão para o Algarve para a aventura; e os psicólogos vão para as Montanhas e para a Costa de Prata, à procura da natureza, de relaxar e de fugir ao calor.

Gráfico 8.2.5



3. PRINCALS

O Princals é um método não linear das componentes principais, que permite usar os níveis de medida nominal e ordinal, explicando a maior variação dos dados através dum número reduzido de dimensões sem perda de informação. Este método revela relações entre as variáveis, entre os casos e entre as variáveis e os casos.

O Princals possibilita a investigação de possíveis relações não lineares através da atribuição de quantificações óptimas ou scores às variáveis.

As variáveis podem ser quantificadas como: numéricas (NUM); ordinais (ORDI); nominais com igual quantificação em ambas as dimensões (single nominal ou SNOM); e nominais com diferentes quantificações em ambas as dimensões (multiple nominal ou MNOM).

- A quantificação NUM aplica-se aos valores numéricos atribuídos às categorias, dando-lhes as propriedades de variáveis de intervalo/rácio.
- A quantificação ORDI atribui à variável as propriedades de variáveis ordinais.
- A quantificação SNOM atribui à variável as propriedades de variáveis nominais, com a restrição dos objectos da mesma categoria terem a mesma quantificação para todas as dimensões. Quando todas as variáveis são SNOM e há apenas uma dimensão o Princals é igual ao Homals.
- A quantificação MNOM, atribui à variável as propriedades de variáveis nominais, podendo os objectos na mesma categoria ter quantificações diferentes para todas as dimensões. Se todas as variáveis forem MNOM o Princals equivale ao Homals.

Os quatro níveis de quantificação conduzem a diferentes interpretações dos resultados, pelo que se deve explorar aquelas que se mostrem adequadas à variável e facilitem a análise.

Por exemplo, dado não existir um comportamento monótono crescente ou decrescente no número de acidentes de viação nas diferentes faixas etárias, uma vez que os mais jovens e os mais velhos tendem a ter mais acidentes do que os de meia idade, pode preferir-se a quantificação nominal à ordinal para as idades.

Do mesmo modo, embora os partidos políticos sejam uma variável nominal, pode-se querer quantificá-los de modo a manterem a ordem dos mais de esquerda aos mais de direita, usando-se o nível ordinal.

A aplicação prática do modelo vai ser feita a uma amostra aleatória de 10 pessoas.

As variáveis são:

- Gosto pelo convívio, *conv*, com 4 categorias: 1 = não gosta, 2 = gosta pouco, 3 = gosta moderadamente, 4 = gosta muito;
- Cuida da aparência, *apare*, com 4 categorias: 1 = nada, 2 = pouco, 3 = ocasionalmente, 4 = sempre;
- Frequência de leitura, *ler*, com 4 categorias: 1 = não lê, 2 = lê irregularmente, 3 = lê todas as semanas, 4 = lê todos os dias;
- Frequência de Tv, *tv*, com 2 categorias: 1 = muita, 2 = pouca;
- Actividade sexual, *sexual*, com 3 categorias: 1 = rara, 2 = regular, 3 = intensa.

Todas as variáveis são quantificadas como ORDI.

O quadro 8.3.1 apresenta a base de dados.

Quadro 8.3.1

Casos	<i>conv</i>	<i>apare</i>	<i>ler</i>	<i>tv</i>	<i>sexual</i>
1	1	1	1	2	2
2	2	2	2	2	2
3	1	2	1	1	1
4	4	4	2	2	3
5	4	4	4	1	3
6	3	3	3	1	2
7	2	3	3	2	1
8	3	4	1	1	1
9	4	3	1	2	2
10	4	1	3	2	2

Pretende-se analisar a relação existente entre as casos (pessoas), entre as cinco variáveis e entre as suas categorias.

O quadro 8.3.2 apresenta a tabela de frequências do Princals. Pode ver-se, por exemplo, que dos inquiridos 3 têm uma actividade sexual rara, 5 regular e 2 intensa.

Quadro 8.3.2

List of Variables			
Variable	Variable Label	Number of Categories	Measurement Level
CONV	gosta de conviver	4	Ordinal
APARE	1 "nada" 2 "pouco" 3 "algumas vezes" 4	4	Ordinal
LER	1 "não lê" 2 "lê irregularmente" 3 "lê semp4	4	Ordinal
TV	1 "muito" 2 "pouco"	2	Ordinal
SEXUAL	1 "rara" 2 "regular" 3 "intensa"	3	Ordinal
Marginal Frequencies			
=====			
Variable	Missing	Categories	
		1	2
CONV	0	2	2
APARE	0	4	2
LER	0	2	2
TV	0	4	6
SEXUAL	0	3	5

A base de dados e o procedimento Princals pode ser conseguida através dos seguintes comandos inscritos na syntax do SPSS, seguidos da instrução Run All:

```

DATA LIST / conv 1 apare 2 ler 3 Tv 4 sexual 5.
VARIABLE LABELS conv "gosta de conviver"
    apare "cuida da aparência"
    ler "frequência de leitura"
    Tv "frequência de tv"
    sexual "actividade sexual"
VALUE LABELS conv 1 "não gosta" 2 "gosta pouco" 3 "gosta
moderadamente" 4 "gosta muito"
    /apare 1 "nada" 2 "pouco" 3 "ocasionalmente" 4 "sempre"
    /ler 1 "não lê" 2 "lê irregularmente" 3 "lê todas as
semanas" 4 "lê todos os dias"
    /Tv 1 "muita" 2 "pouca"
    /sexual 1 "rara" 2 "regular" 3 "intensa".
BEGIN DATA
11122
22222
11211
42423
44423
33312
23321
31411

```

```

41322
43122
END DATA.
PRINCALS VARIABLES=conv TO apare (4)
                    sexo (2) estado (3)
/ANALYSIS=conv TO apare (ORDI) sexo (MNOM) estado (ORDI)
/PRINT=ALL
/PLOT=OBJECT (8) QUANT (conv ler apare sexo estado) (8)
LOADINGS (8).

```

3.1. DIMENSÕES E QUALIDADE DO MODELO

Os valores próprios são utilizados como indicadores do número de dimensões necessárias para explicar os dados, medindo a variância atribuída a cada dimensão.

Quando as variáveis não são MNOM, como é o caso, o número de dimensões a reter é dado pelo número dos valores próprios maiores do que $1/(n^{\circ} \text{ variáveis})$.

Quando as variáveis são MNOM o número máximo de dimensões é dado pelos valores próprios maiores do que $[1 / (n^{\circ} \text{ categorias} - n^{\circ} \text{ variáveis})]$. No entanto, o ideal é manter um pequeno número de dimensões que facilite a interpretação.

O quadro 8.3.3 mostra que os valores próprios são ambos maiores que $1/5$ (0,5471 e 0,2661), pelo que se utilizam as duas dimensões para interpretar a associação entre as variáveis.

Quadro 8.3.3

Dimension	Eigenvalue
1	,5471
2	,2661

Os valores do ajustamento, *fit*, e da perda, *loss*, apresentados no quadro 8.3.4, mostram como o Princals se ajusta aos dados quantificados.

$$\text{total fit} + \text{total loss} = n^{\circ} \text{ dimensões, isto é, } 0,8132 + 1,1868 = 2.$$

O máximo valor que o *Fit* pode assumir iguala o número de dimensões, indicando que a relação é perfeita. Neste caso a relação não é perfeita pois 0,8132 é menor do que 2.

A *Loss*, 1,1868, dá a diferença entre o ajustamento óptimo e o ajustamento feito.

A *total loss* é decomposta na *multiple loss* e na *single loss*.

Quadro 8.3.4

Iteration Number	Total Fit	Total Loss	Multiple Loss	Single Loss
17	,8132	1,1868	1,0049	,1819

3.2. VARIÁVEIS MAIS IMPORTANTES POR DIMENSÃO

Os quadros 8.3.5 e 8.3.6 apresentam, respectivamente o multiple fit e o single fit, onde são identificadas as variáveis mais importantes por dimensão, cujos valores são superiores à média.

As variáveis mais explicadas na dimensão 1 são a actividade sexual, o cuidar da aparência e o gosto pela leitura, enquanto que a dimensão 2 parece separar a frequência de televisão do gosto pelo convívio.

Quadro 8.3.5

Summary of Analysis			
Multiple Fit			
Variable	Row Sums	Dimension	
-----		1	2
CONV	1,100	,456	,644
APARE	,948	,755	,192
LER	,872	,695	,177
TV	,908	,102	,805
SEXUAL	1,148	,857	,291
Mean:	,995	,573	,422

Quadro 8.3.6

Single Fit			
Variable	Row Sums	Dimension	
-----		1	2
CONV	,830	,424	,406
APARE	,713	,679	,034
LER	,728	,687	,041
TV	,908	,102	,805
SEXUAL	,888	,844	,044
Mean:	,813	,547	,266

O quadro 8.3.7 mostra que as variáveis mais explicativas da primeira dimensão têm loadings do mesmo sinal, neste caso negativo, o que indica uma associação no mesmo sentido entre as categorias dessas variáveis. Deste modo, as baixas categorias de sexual estão associadas às baixas categorias de apare e de ler, assim como, as categorias elevadas de sexual estão associadas às categorias elevadas de apare e ler.

O sinal contrário nos loadings indica que as baixas categorias de uma variável estão associadas a elevadas categorias da outra.

As duas variáveis explicativas da segunda dimensão variam no mesmo sentido, pois os loadings são do mesmo sinal, neste caso positivo.

As variáveis apare e ler têm scores elevados e semelhantes ($-0,824$ e $-0,829$), o que significa terem uma elevada correlação (valores próximos de -1).

Quadro 8.3.7

Component Loadings		
Variable	Dimension	
	1	2
CONV	-,651	,637
APARE	-,824	-,185
LER	-,829	-,203
TV	,320	,897
SEXUAL	-,919	,209

A matriz das correlações, apresentada no quadro 8.3.8, mostra que uma maior actividade sexual está associada a uma maior frequência de leitura, a maiores cuidados com a aparência, a um maior gosto pelo convívio e a uma menor frequência televisiva. Todas estas correlações são elevadas e positivas (0,731; 0,67; 0,635), excepto a ligada à tv que é negativa e muito fraca ($-0,071$).

Igualmente se conclui que uma maior frequência de tv está associada a maior gosto pelo convívio, a menores cuidados com a aparência e a uma menor frequência de leitura. No entanto as estas correlações são fracas (0,249; $-0,403$; $-0,387$).

A maior frequência de leitura está associada a um maior cuidado com a aparência (correlação positiva média 0,505) e a um maior gosto pelo convívio (correlação positiva fraca 0,411).

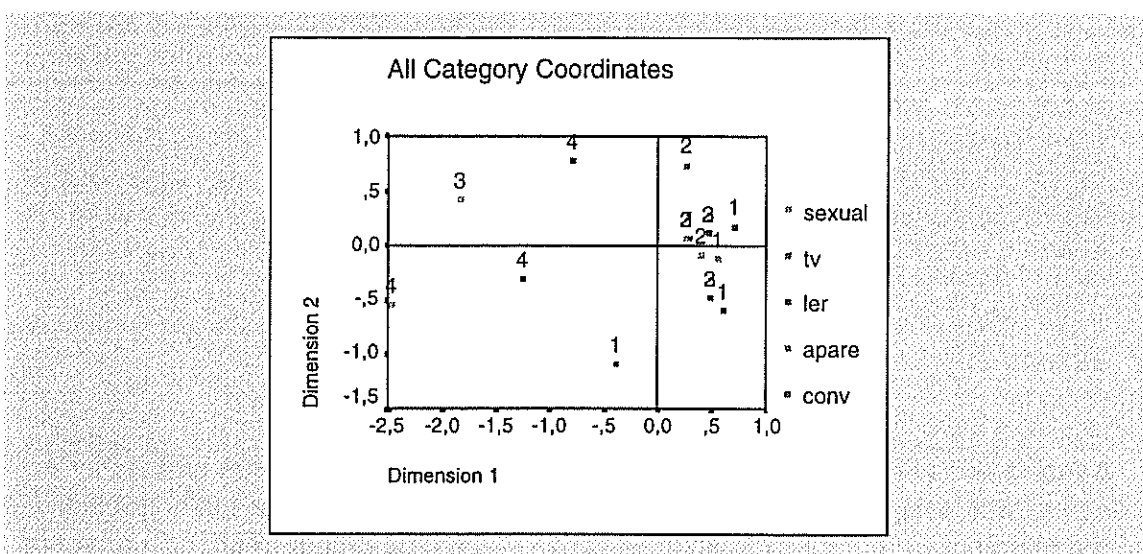
Quadro 8.3.8

* Correlations between Optimally Scaled Variables *

	CONV	APARE	LER	TV	SEXUAL
CONV	*				
APARE	,411	*			
LER	,347	,505	*		
TV	,249	-,403	-,387	*	
SEXUAL	,635	,670	,731	-,071	*

O gráfico 8.3.1 mostra a associação comentada entre as categorias das variáveis.

Gráfico 8.3.1



3.3. SEPARAR CATEGORIAS

Os quadros 8.3.9 a 8.3.11 apresentam as quantificações das categorias, onde as que têm valores semelhantes são aquelas que a análise Princals não consegue diferenciar, sugerindo-se ou o seu agrupamento ou outra quantificação diferente da usada.

Observando os quadros seguintes, pode ver-se que existem quantificações semelhantes para todas as variáveis, exceptuando a *tv* e a *sexual*.

A análise Princals não consegue separar as categorias 2 e 3 de *convive* (gosta pouco ou moderadamente), de *ler* (irregularmente ou todas as semanas) e de *apare* (liga pouco ou ocasionalmente).

Quadro 8.3.9

Variable: CONV gosta de conviver
 Type: Ordinal Missing: 0
 Category: Marginal Frequency
 Quantification

Category	Frequency	Coordinate
1	2	-,94
2	2	-,75
3	2	-,75
4	4	1,22

Single Category Coordinates

Category	Dimension	
	1	2
1	,61	-,60
2	,49	-,48
3	,49	-,48
4	-,80	,78

Multiple Category Coordinates

Category	Dimension	
	1	2
1	,70	-,51
2	,71	,24
3	,17	-1,29
4	-,79	,78

=====

Variable: APARE 1 "nada" 2 "pouco" 3 " algumas vezes" 4
 Type: Ordinal Missing: 0
 Category: Marginal Frequency
 Quantification

Category	Frequency	Coordinate
1	4	-,36
2	2	-,32
3	3	-,32
4	1	3,00

Single Category Coordinates

Category	Dimension	
	1	2
1	,29	,07
2	,26	,06
3	,26	,06
4	-2,47	-,55

Quadro 8.3.10

Apare

Multiple Category Coordinates

Category	Dimension	
	1	2
1	,38	-,32
2	-,27	,76
3	,50	,12
4	-2,47	-,57

=====

Variable: LER 1 "não lê" 2 "lê irregularmente" 3 "lê s

Type: Ordinal Missing: 0

Category:	Marginal Frequency	Quantification
1	2	-,85
2	2	-,58
3	3	-,57
4	3	1,52

Single Category Coordinates

Category	Dimension	
	1	2
1	,70	,17
2	,48	,12
3	,47	,12
4	-1,26	-,31

Multiple Category Coordinates

Category	Dimension	
	1	2
1	,57	,73
2	,63	-,49
3	,47	,10
4	-1,27	-,26

=====

Variable: TV 1 "muito" 2 "pouco"

Type: Ordinal Missing: 0

Category:	Marginal Frequency	Quantification
1	4	-1,22
2	6	,82

Single Category Coordinates

Category	Dimension	
	1	2
1	-,39	-1,10
2	,26	,73

Quadro 8.3.11

```

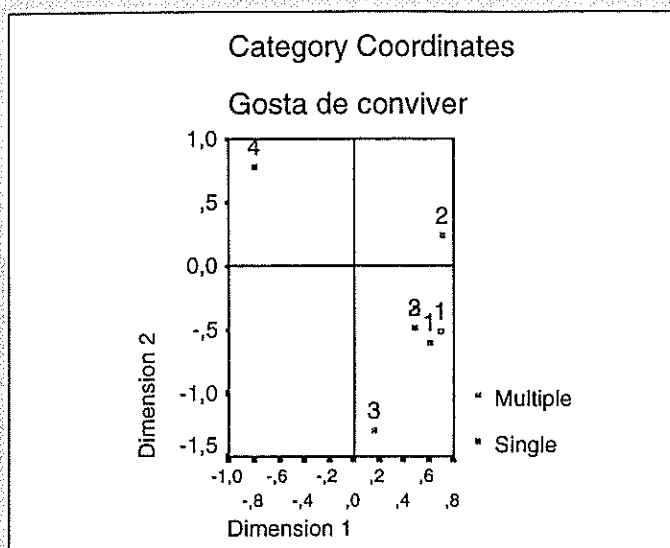
tv
Multiple Category Coordinates
-----
Category          Dimension
                   1          2
    1          -,39      -1,10
    2           ,26         ,73
=====
Variable: SEXUAL      1 "raro" 2 "regular" 3 "intensa"
Type: Ordinal          Missing:          0
Category:          Marginal Frequency          Quantification
-----
    1              3              -,61
    2              5              -,43
    3              2              1,99
Single Category Coordinates
Category          Dimension
                   1          2
    1           ,56         -,13
    2           ,40         -,09
    3          -1,83         ,42
Multiple Category Coordinates
Category          Dimension
                   1          2
    1           ,40         -,82
    2           ,50         ,36
    3          -1,85         ,34

```

Em termos dos gráficos Quant, quando as categorias têm semelhantes coordenadas single e multiple significa que são bem discriminadas pela análise Princals. Caso contrário, os pontos apresentam-se afastados no gráfico significando que não são bem discriminadas pela análise Princals.

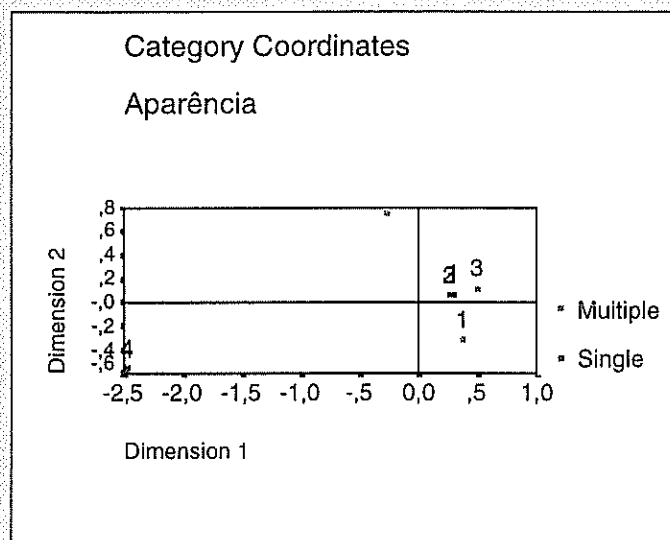
O gráfico 8.3.2 mostra que as coordenadas single e multiple da variável gosto pelo convívio encontram-se afastadas nas categorias 2 e 3, significando, como já se tinha dito, que não são diferenciadas.

Gráfico 8.3.2



O gráfico 8.3.3, mostra que existe uma diferença nas coordenadas *single* e *multiple* das categorias 2 e 3 da variável cuida da aparência, ou seja, não são separadas estas categorias pelo Princals.

Gráfico 8.3.3



O gráfico 8.3.4 mostra que existe uma diferença nas coordenadas *single* e *multiple* das categorias 2 e 3 da variável frequência de leitura.

Os gráficos 8.3.5 e 8.3.6, referentes à tv e à actividade sexual apresentam todas a categorias bem discriminadas.

Gráfico 8.3.4

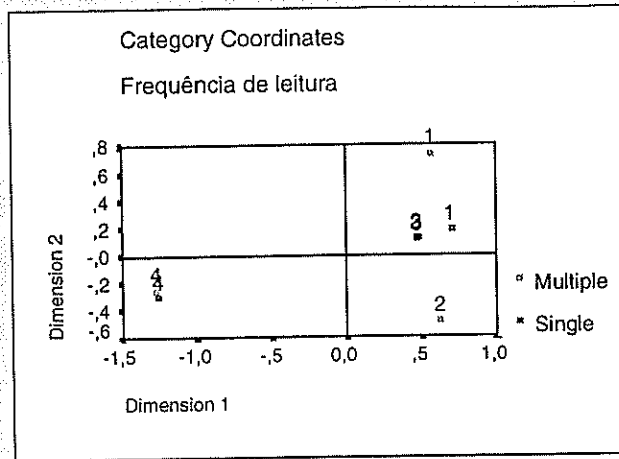


Gráfico 8.3.5

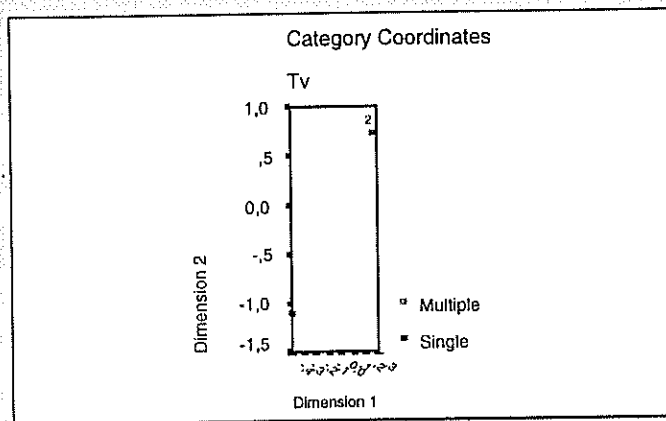
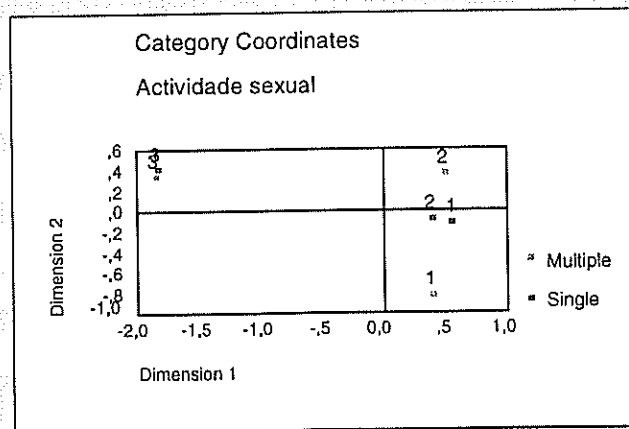


Gráfico 8.3.6



3.4. SEPARAR CASOS

O procedimento Princals permite igualmente ver se existem casos (pessoas) outliers e casos semelhantes, usando para tal a quantificação óptima (scores) atribuída aos casos, cujos valores constam do quadro 8.3.12.

Quadro 8.3.12

The Object Scores are:
=====

Object *	Dimension	
	1	2
1 *	,83	,21
2 *	,69	,25
3 *	,57	-1,24
4 *	-1,23	1,26
5 *	-2,47	-,57
6 *	,45	-1,12
7 *	,74	,23
8 *	-,11	-1,46
9 *	,23	1,20
10 *	,30	1,24

O gráfico 8.3.7 dos object scores, mostra têm comportamentos semelhantes (próximos no gráfico) as pessoas com os códigos: 9 e 10; a 1 com a 2 e a 7, e a 3 com a 6. No quadro 8.3.1 podem ver-se as categorias correspondentes a cada um destes casos.

Gráfico 8.3.7

