



**MARIELE VILELA BERNARDES PRADO**

**MÉTODOS DE ANÁLISE DE  
CORRESPONDÊNCIA MÚLTIPLA: ESTUDO DE  
CASO APLICADO À AVALIAÇÃO DA  
QUALIDADE DO CAFÉ**

**LAVRAS – MG**

**2012**

**MARIELE VILELA BERNARDES PRADO**

**MÉTODOS DE ANÁLISE DE CORRESPONDÊNCIA MÚLTIPLA:  
ESTUDO DE CASO APLICADO À AVALIAÇÃO DA QUALIDADE DO  
CAFÉ**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Dr. Renato Ribeiro de Lima  
Orientador

Dr. Flávio Meira Borém  
Dr. Ronaldo Rocha Bastos  
Coorientadores

**LAVRAS – MG  
2012**

**Ficha Catalográfica Elaborada pela Divisão de Processos Técnicos da  
Biblioteca da UFLA**

Prado, Mariele Vilela Bernardes.

Métodos de análise de correspondência múltipla : estudo de caso  
aplicado à avaliação da qualidade do café / Mariele Vilela Bernardes  
Prado. – Lavras : UFLA, 2012.

76 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2012.

Orientador: Renato Ribeiro de Lima.

Bibliografia.

1. Análise multivariada. 2. Dados longitudinais. 3. Cafés  
especiais. I. Universidade Federal de Lavras. II. Título.

CDD – 519.535

**MARIELE VILELA BERNARDES PRADO**

**MÉTODOS DE ANÁLISE DE CORRESPONDÊNCIA MÚLTIPLA:  
ESTUDO DE CASO APLICADO À AVALIAÇÃO DA QUALIDADE DO  
CAFÉ**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 22 de junho de 2012.

Dr. Flávio Meira Borém

UFLA

Dr. Marcelo Angelo Cirillo

UFLA

Dr. Ronaldo Rocha Bastos

UFJF

Dra. Sttela Dellyzete Veiga Franco Da Rosa

EMBRAPA

Dr. Renato Ribeiro de Lima  
Orientador

**LAVRAS – MG  
2012**

*A meus pais, José e Silvânia*

*Ao meu esposo, Jair.*

**DEDICO**

## AGRADECIMENTOS

À Universidade Federal de Lavras e ao Programa de Pós-Graduação em Estatística e Experimentação Agropecuária.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão de bolsa de estudos.

Ao professor Renato Ribeiro de Lima, pela orientação, paciência e apoio às minhas escolhas.

Ao professor Ronaldo Rocha Bastos pela efetiva participação em todo o processo de construção deste trabalho. Pela disponibilidade, oportunidade, carinho e principalmente, por apresentar-me à análise de correspondência.

Ao professor Flávio Meira Borém, pela oportunidade de trabalharmos juntos, pela disponibilização dos dados e pelas sugestões e contribuições ao trabalho.

A Mariana e ao Diego pelos conhecimentos compartilhados e pelo auxílio com o banco de dados.

A toda a turma do mestrado em estatística 2010, pela amizade, por todos os bolos de aniversários que comemos juntos, pelas festas e pelas risadas.

As amigas, Elayne (Caronas!), Iábita (Conselhos!) e Tábata (“É nós!”) pela amizade mais que especial.

Ao Sr. Rossoni, por dividir comigo, de forma não igualitária, os períodos de convivência com o Jair.

Aos meus pais, José e Silvânia, por estarem sempre presentes, mesmo tão distantes. Pela dedicação, amor e oportunidades que me proporcionaram chegar até aqui.

Ao meu esposo Jair, principal incentivador desta conquista, pelo apoio, companheirismo e amor.

A Deus.

## LISTA DE TABELAS

Tabela 1	Tabela de contingência com $I$ linhas e $J$ colunas. ....	14
Tabela 2	Tabela de contingência. ....	15
Tabela 3	Escala de classificação de cafés especiais.....	45
Tabela 4	Codificação das categorias das variáveis consideradas na ACM. ....	47
Tabela 5	Principais inércias e porcentagens explicadas e acumuladas da matriz indicadora das médias. ....	52
Tabela 6	Contribuições ( $ctr$ e $cor$ ) das categorias para os 5 primeiros eixos da matriz indicadora das médias. ....	54
Tabela 7	Principais inércias e porcentagens explicadas e acumuladas da matriz $Z^{IK(q)}$ .....	58
Tabela 8	Contribuições ( $ctr$ e $cor$ ) das categorias para os 5 primeiros eixos da matriz $Z^{IK(q)}$ .....	59
Tabela 9	Principais inércias e porcentagens explicadas e acumuladas da matriz $Z^{TK(q)}$ .....	62
Tabela 10	Contribuições ( $ctr$ e $cor$ ) das categorias para os 2 primeiros eixos da matriz $Z^{TK(q)}$ .....	63
Tabela 11	Principais inércias e porcentagens explicadas e acumuladas da matriz BROAD.....	66
Tabela 12	Contribuições das variáveis para a variância do primeiro eixo da matriz BROAD por provador.....	67
Tabela 13	Contribuições das variáveis para a variância do segundo eixo da matriz BROAD por provador.....	67
Tabela 14	Contribuições das variáveis para a variância do terceiro eixo da matriz BROAD por provador.....	67

## LISTA DE FIGURAS

Figura 1	Representação da nuvem de pontos de colunas em $\mathbb{R}^2$ (a) e da nuvem de pontos de linhas em $\mathbb{R}^3$ (b) .....	16
Figura 2	Número de publicações sobre “correspondence analysis” de 2000 a 2011 de acordo com ISI-Web of Science (2012) .....	18
Figura 3	Representação de uma matriz indicadora para variável $q$ .....	20
Figura 4	Representação de uma matriz indicadora para $Q$ variáveis .....	21
Figura 5	Representação de uma matriz indicadora de quatro amostras de café avaliadas em três variáveis.....	22
Figura 6	Representação gráfica do ponto médio $G$ de uma nuvem de pontos. ....	25
Figura 7	Representação do ponto $P'$ , projeção ortogonal do ponto $P$ em $\ell$ .....	30
Figura 8	Representação da projeção ortogonal dos pontos da nuvem sobre a linha $\ell$ .....	31
Figura 9	Representação do ângulo formado pelos vetores $\overline{GM}$ e $\overline{GM}'$ .....	35
Figura 10	Representação da matriz indicadora $Z_i$ .....	39
Figura 11	Representação da construção da matriz super-indicadora $Z^3$ .....	39
Figura 12	Representação da construção da matriz BROAD.....	40
Figura 13	Representação da construção da matriz $Z^{TK(q)}$ .....	42
Figura 14	Gráficos de correspondência das categorias da matriz $Z^m$ para as dimensões 1 e 2 (a), 1 e 3 (b) .....	56
Figura 15	Gráfico de correspondência das categorias da matriz $Z^{IK(q)}$ .....	61
Figura 16	Gráfico de correspondência das categorias da matriz $Z^{TK(q)}$ .....	64
Figura 17	Gráfico de correspondência das categorias para a matriz BROAD .....	69

Figura 18	Gráfico de correspondência da categoria qualidade para a matriz BROAD .....	70
-----------	--	----

## RESUMO

O presente trabalho buscou comparar os resultados obtidos, utilizando diferentes abordagens embasadas na estrutura da análise de correspondência para dados longitudinais. Este trabalho buscou ainda apresentar o método da análise de correspondência de forma intuitiva por meio de conceitos geométricos, descartando a necessidade de conhecimentos mais profundos acerca da álgebra de matrizes, contribuindo assim, para a divulgação do mesmo. Para ilustrar a aplicação das diferentes abordagens foi considerado um conjunto de dados de análise sensorial de amostras de cafês especiais coletadas em diferentes condições e realizada por quatro provadores treinados. A estrutura da análise de correspondência para dados longitudinais foi utilizada considerando os provadores como diferentes momentos de avaliação. A análise de correspondência múltipla foi também aplicada considerando a média entre os provadores. Concluiu-se, ao final do estudo, os resultados indicaram que o uso da estrutura da análise de correspondência para dados longitudinais aplicada à análise sensorial apresentou resultados mais detalhados que a análise de correspondência múltipla aplicada à média dos provadores, podendo, além de relacionar a qualidade aos fatores em estudo, identificar o comportamento dos provadores, tornando possível uma avaliação dos mesmos.

**Palavras-chave:** Análise multivariada. Cafês especiais. Dados longitudinais.

## ABSTRACT

The present study aimed to compare the results obtained from different approaches based on the correspondence analysis for longitudinal data. This study also presents the method of correspondence analysis intuitively by using geometric concepts. Thus, it was eliminated the need for knowledge about matrix algebra. To illustrate the application of different approaches it was considered a data set of sensory analysis of special coffees samples. These samples were collected in different conditions and the sensory evaluations were performed by four trained tasters. The structure of the correspondence analysis for longitudinal data was used by considering the tasters as different moments of evaluations. The multiple correspondence analysis was applied considering the mean of the four tasters. The use of the structure of the correspondence analysis for longitudinal data applied to sensory analysis presented more complete results than multiple correspondence analysis applied to the average of tasters. The correspondence analysis for longitudinal data was adequate to describe the relationship between the coffee samples quality and the factors which are related to that. Moreover, it was possible to study the behavior of the tasters by using this analysis.

**Keywords:** Multivariate analysis. Specialty coffees. Longitudinal data.

## SUMÁRIO

1	INTRODUÇÃO.....	12
2	REFERENCIAL TEÓRICO.....	14
2.1	Tabelas de contingência e conceitos geométricos .....	14
2.2	Análise de correspondência.....	16
2.3	Análise de correspondência múltipla.....	18
2.4	Matriz indicadora.....	19
2.5	Nuvem de pontos .....	23
2.5.1	Ponto médio de uma nuvem de pontos .....	24
2.5.2	Distâncias.....	26
2.5.2.1	Distância entre indivíduos.....	26
2.5.2.2	Distância entre categorias .....	27
2.5.3	Variância .....	28
2.5.4	Eixos principais .....	30
2.5.5	Coordenadas principais .....	32
2.5.6	Contribuições.....	33
2.5.6.1	Contribuição de um ponto a um eixo.....	33
2.5.6.2	Contribuição do eixo para a variância de um ponto .....	34
2.5.7	Fórmulas de transição .....	35
2.5.8	Pontos suplementares.....	36
2.6	Análise de correspondência e dados longitudinais .....	37
2.7	Café.....	42
2.7.1	Classificação do café e análise sensorial .....	44
3	MATERIAL E MÉTODOS .....	46
3.1	Material .....	46
3.2	Métodos .....	47
3.2.1	Construção das matrizes .....	49
3.2.2	Aplicação da AC.....	49
3.2.3	Interpretação dos resultados.....	50
4	RESULTADOS E DISCUSSÃO .....	52
4.1	Análise da matriz indicadora das médias .....	52
4.2	Análise da matriz $Z^{IK(q)}$ .....	57
4.3	Análise da matriz $Z^{TK(q)}$ .....	61
4.4	Análise da matriz BROAD.....	65
5	CONCLUSÃO .....	72
	REFERÊNCIAS.....	73
	APÊNDICE .....	76

## 1 INTRODUÇÃO

Análise sensorial é uma metodologia empregada para avaliar atributos ou características de produtos, através dos sentidos humanos. Na maioria das vezes, o objetivo da análise sensorial é identificar produtos ou amostras que apresentam qualidade superior ou maior aceitação, com base nesses atributos. Porém, os resultados obtidos em uma análise sensorial também podem ser utilizados com o objetivo de identificar quais fatores e como estes interferem na qualidade sensorial de um produto.

Um primeiro passo na identificação dos efeitos de diferentes fatores sobre a qualidade de um produto, ou seja, como esses fatores influenciam os atributos analisados, é estudar a relação entre estes e os atributos. Uma forma de entender essa relação é através da análise de correspondência.

A análise de correspondência é uma técnica de análise multivariada para dados categóricos que possibilita, através de uma redução de dimensionalidade, avaliar graficamente as relações existentes entre variáveis e suas categorias. Essas variáveis estudadas correspondem aos fatores e atributos.

A análise de correspondência pode ser realizada de diferentes formas. No estudo da relação entre duas variáveis, têm-se a análise de correspondência simples. Caso sejam estudadas as relações entre mais de duas variáveis, aplica-se a análise de correspondência múltipla. Além disso, as variáveis podem ser avaliadas em um ou vários momentos. Quando essa avaliação ocorre em mais de um momento, a análise de correspondência é dita longitudinal. De uma forma geral, o uso da análise de correspondência tem aumentado nos últimos anos, porém, a abordagem da análise de correspondência para dados longitudinais tem sido pouco utilizada.

O objetivo deste trabalho é a aplicação de diferentes abordagens da análise de correspondência em dados sensoriais, tendo como embasamento a estrutura da análise de correspondência longitudinal. O presente trabalho busca ainda apresentar a análise de correspondência múltipla de forma intuitiva, por meio de conceitos geométricos básicos, colaborando assim para a divulgação da mesma. Para ilustrar a aplicação das diferentes abordagens, foi avaliado um conjunto de dados de análise sensorial de amostras de cafês especiais coletadas em diferentes condições e realizada por quatro provadores treinados.

## 2 REFERENCIAL TEÓRICO

### 2.1 Tabelas de contingência e conceitos geométricos

Tabelas de contingência são tabelas com classificação cruzada de  $n$  objetos por duas ou mais variáveis, geralmente categóricas, com categorias mutuamente exclusivas e exaustivas. Um modelo clássico para tabela de contingência é apresentado na Tabela 1.

Na Tabela 1,  $n_{ij}$  é a frequência observada (ou contagem) da categoria  $i$  da variável de linha e  $j$  da variável de coluna, sendo  $i = 1, 2, \dots, I$  e  $j = 1, 2, \dots, J$ . Os totais marginais das linhas e das colunas são dados por  $n_{i\bullet}$  e  $n_{\bullet j}$ , respectivamente, e o total geral por  $n_{\bullet\bullet}$ , ou seja,  $n_{i\bullet} = \sum_{j=1}^J n_{ij}$ ,  $n_{\bullet j} = \sum_{i=1}^I n_{ij}$  e  $n_{\bullet\bullet} = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$ .

Tabela 1 Tabela de contingência com  $I$  linhas e  $J$  colunas.

Linhas	Colunas						Totais das linhas
	1	2	...	$j$	...	$J$	
1	$n_{11}$	$n_{12}$	...	$n_{1j}$	...	$n_{1J}$	$n_{1\bullet}$
2	$n_{21}$	$n_{22}$	...	$n_{2j}$	...	$n_{2J}$	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$i$	$n_{i1}$	$n_{i2}$	...	$n_{ik}$	...	$n_{iJ}$	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$I$	$n_{I1}$	$n_{I2}$	...	$n_{Ij}$	...	$n_{IJ}$	$n_{I\bullet}$
Totais das colunas	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet J}$	$n_{\bullet\bullet}$

Tabela 2 Tabela de contingência.

Linhas	Colunas		
	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>
L <sub>1</sub>	2	3	1
L <sub>2</sub>	1	1	2

Cada linha da tabela pode ser interpretada como as coordenadas de um ponto em um espaço  $R^J$ . Da mesma forma, as colunas seriam as coordenadas de pontos em um espaço  $R^I$ . Para exemplificar, considere os dados hipotéticos apresentados na Tabela 2.

Os dados da Tabela 2 podem ser representados por duas nuvens de pontos, conforme apresentado na Figura 1.

A interpretação das nuvens de pontos torna-se cada vez mais complexa conforme os números de linhas e colunas aumentam. Uma forma de trabalhar com nuvens multidimensionais é através da redução da dimensionalidade dessas nuvens de pontos em um subespaço ótimo de forma que este assegure a melhor representação possível da nuvem inicial. Essa redução de dimensionalidade é trabalhada através de métodos de análise estatística multivariada como a análise fatorial, componentes principais e análise de correspondência.

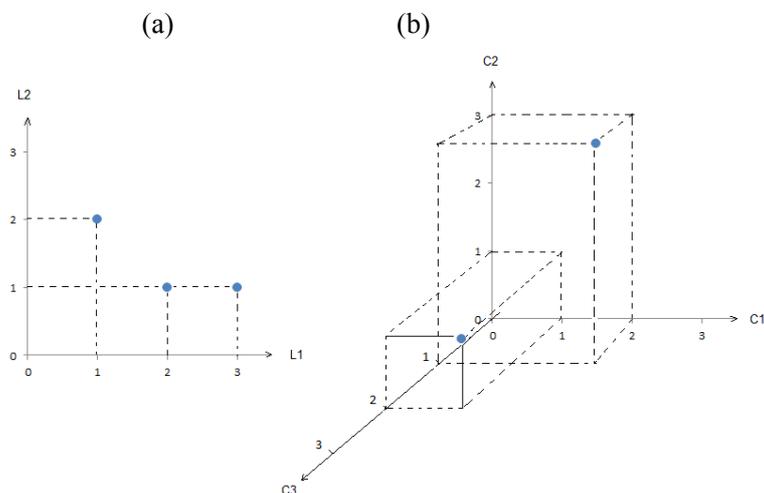


Figura 1 Representação da nuvem de pontos de colunas em  $R^2$  (a) e da nuvem de pontos de linhas em  $R^3$  (b)

## 2.2 Análise de correspondência

A análise de correspondência (AC) é uma das técnicas de análise multivariada para dados categóricos. É usada para explorar geometricamente a relação entre linhas e colunas de uma tabela de contingência em um espaço de baixa dimensão, de modo que a proximidade no espaço indique uma relação ou correspondência entre as mesmas. A AC pode ainda ser entendida, segundo Greenacre e Blasius (2006), como um caso particular da análise de componentes principais (PCA) para dados categóricos, uma vez que PCA trabalha com dados quantitativos.

As primeiras concepções teóricas sobre AC datam do início do século XX. O coeficiente de correlação entre linhas e colunas de uma tabela de

contingência, desenvolvido por Karl Pearson, serviu como base para a técnica da AC. Em 1940, Fisher utilizou a AC ao relacionar cor dos olhos e dos cabelos de crianças de Caithness, Escócia. No ano seguinte, Louis Guttman fundamenta a análise de correspondência múltipla (ACM), na época chamada de *Dual Scaling* (Beh, 2004).

A partir de 1950, amparados nas ideias de Guttman, outros métodos foram desenvolvidos. Porém, foi apenas entre os anos de 1960 e 1970, que o matemático e linguista francês Jean-Paul Benzécri (1992) abordou o método de forma geométrica tornando-o conhecido como análise de correspondência simples e múltipla. Nos anos seguintes, o método foi sendo difundido em diversas escolas de análises de dados. Lebart et al. (1984) e Greenacre (1984) são os responsáveis pela difusão do método nos países de língua inglesa. Ainda na década de 1980, procedimentos da AC foram incluídos em softwares estatísticos, fazendo com que o número de aplicações aumentasse significativamente.

Realizando uma busca pelo tema “*correspondence analysis*”, através da plataforma de pesquisa *ISI –Web of Science*(2012), é possível observar o aumento das publicações sobre o tema na última década, o que está representado na Figura 2.

A análise de correspondência simples (ACS) é aplicável, principalmente na análise de dados apresentados na forma de tabelas de dupla entrada, levando a um mapa que facilita a visualização da associação entre duas variáveis categóricas. A generalização do método para um conjunto de mais de duas variáveis categóricas é chamada de análise de correspondência múltipla (ACM). No presente trabalho, essa generalização é feita através da abordagem geométrica, tomando como base os trabalhos de Le Roux e Rouanet (2004;

2010). Diferentes maneiras de concatenação e agregação de matrizes serão descritas a seguir, utilizando técnicas para dados longitudinais categóricos, propostas por Van Der Heijden (1987). Exemplos práticos serão utilizados para uma melhor compreensão da técnica.

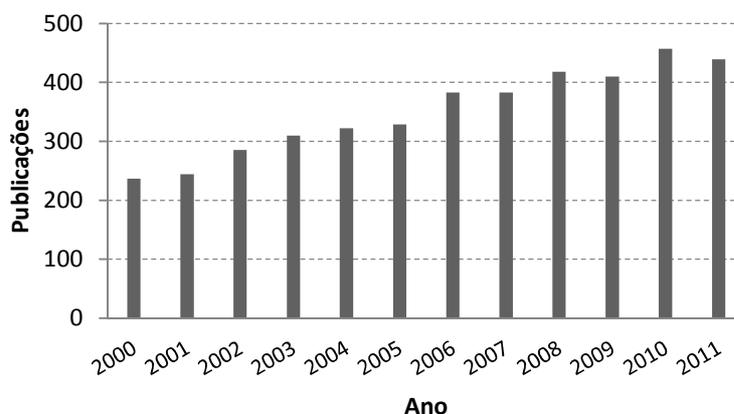


Figura 2 Número de publicações sobre “correspondence analysis” de 2000 a 2011 de acordo com *ISI-Web of Science* (2012)

### 2.3 Análise de correspondência múltipla

Na análise de correspondência múltipla (ACM), os dados são apresentados por meio de duas ou mais variáveis categóricas, onde a dimensionalidade está ligada às categorias de cada variável. Para aplicação da ACM, os dados da tabela de contingência multidimensional são ajustados na

forma de uma matriz retangular em que as linhas representam os objetos de estudo e as colunas representam as categorias de todas as variáveis.

O objetivo da ACM é representar geometricamente as linhas e colunas da matriz resultante da tabela multidimensional em um subespaço ótimo de dimensão reduzida, de modo que, seja possível, observar as relações entre objetos, variáveis e categorias.

As duas formas mais usuais de apresentação da matriz de dados para aplicação da ACM são: matriz indicadora e matriz de Burt. No presente trabalho, destaca-se detalhadamente apenas a matriz indicadora, uma vez que, conforme descrito por Naito (2007), as duas diferentes abordagens da ACM são equivalentes. Naito (2007) conclui que os gráficos resultantes, via matriz indicadora e matriz de Burt, são análogos, divergindo apenas na escala das variáveis.

#### **2.4 Matriz indicadora**

A forma básica de apresentação dos dados para a execução da ACM é na forma de uma matriz indicadora, onde as linhas representam os objetos e as colunas representam as categorias das variáveis em estudo. As categorias devem ser mutuamente exclusivas e exaustivas, isto é, cada indivíduo deve escolher uma, e somente uma categoria para cada variável ou questão.

Na matriz indicadora, os elementos são dispostos na forma de variáveis *dummy*, ou seja, 1 para a categoria escolhida como resposta de uma variável e 0 para as demais categorias da mesma variável.

Seja  $I$  o conjunto de  $n$  indivíduos,  $I = (1, 2, \dots, i, \dots, n)$ ;  $q$  a variável em estudo com  $K_q$  categorias. A matriz indicadora para esta variável  $q$  é apresentada na Figura 3.

	1	2	...	$k$	...	$K_q$
1	1	0	...	0	...	0
2	0	1	...	0	...	0
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$i$	0	0	$\ddots$	1	$\ddots$	0
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$n$	0	0	...	0	...	1
	$n_1$	$n_2$	...	$n_k$	...	$n_{K_q}$

Figura 3 Representação de uma matriz indicadora para variável  $q$

Na Figura 3, tem-se que  $n_k$  é o número total de indivíduos que escolheram a categoria  $k$ .

Considere agora  $Q$  variáveis,  $q = 1, 2, \dots, Q$ . Cada uma com um número  $K_q$  de categorias, tal que o número total de categorias é dado por  $K = \sum_{q=1}^Q K_q$ .

Uma matriz  $Z$ , de ordem  $n \times K$ , é chamada matriz indicadora com elementos  $z_{ik}$ , em que  $z_{ik} = 1$ , se o indivíduo  $i$  escolheu a categoria  $k$  de uma variável  $q$  e  $z_{ik} = 0$ , caso contrário.

Na Figura 4, tem-se a forma geral de uma matriz indicadora  $Z$ .

$$\begin{array}{c}
 \overbrace{\hspace{15em}}^K \\
 \underbrace{\hspace{3em}}^{K_1} \quad \underbrace{\hspace{6em}}^{K_q} \quad \underbrace{\hspace{6em}}^{K_Q} \\
 1 \quad 2 \quad \dots \quad K_1 \quad 1 + \sum_{j=1}^{q-1} K_j \quad \dots \quad \sum_{j=1}^q K_j \quad 1 + \sum_{j=1}^{Q-1} K_j \quad \dots \quad K \\
 \begin{array}{|c|} \hline
 \begin{array}{cccccccccccc}
 1 & 1 & 0 & \dots & 0 & & 0 & 1 & \dots & 0 & & 0 & 1 & \dots & 0 \\
 2 & 0 & 1 & \dots & 0 & & 0 & 1 & \dots & 0 & & 1 & 0 & \dots & 0 \\
 \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
 i & 0 & 0 & \ddots & 1 & & 1 & 0 & \ddots & 0 & & 0 & 1 & \ddots & 0 \\
 \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & \ddots & \vdots \\
 n & 0 & 0 & \dots & 1 & & 1 & 0 & \dots & 0 & & 0 & 1 & \dots & 0 \\
 \hline
 n_1 & n_2 & \dots & & & & \dots & & & \dots & & \dots & & & n_K
 \end{array}
 \end{array}
 \end{array}$$

Figura 4 Representação de uma matriz indicadora para  $Q$  variáveis

A partir da Figura 4, pode-se observar que:

- i. Como o indivíduo  $i$  pode escolher uma única categoria  $k$  da variável  $q$ , tem-se que  $\sum z_{i\cdot}$  para cada variável  $q$  é igual a 1;
- ii.  $\sum z_{i\cdot}$  para todas as categorias é igual ao número de variáveis, dado por  $Q$ . Logo, todos os  $n$  objetos apresentam  $Q$  como valor marginal, ou peso;

- iii.  $\sum z_{\bullet k}$  é igual ao número de indivíduos que escolheram a categoria  $k$ , dado por  $n_k$ ;
- iv. Pelas observações (i) e (iii) tem-se que  $\sum n_k$  para cada categoria é igual ao número de indivíduos, dado por  $n$ . Logo,  $\sum n_k$  para todas as categorias é  $nQ$ .

Para ilustrar a obtenção de uma matriz indicadora, considere um conjunto de dados referentes a quatro amostras de cafês, resultante da avaliação de três variáveis denominadas A, B e C. As variáveis A e B têm duas categorias e a variável C tem três. Um exemplo de matriz indicadora, resultante desse conjunto de dados, é apresentado na Figura 5. Assim, nesse exemplo, tem-se que  $n=4, Q=3, K_1=K_2=2$  e  $K_3=3$ . O total de categorias é  $K=K_1+K_2+K_3=2+2+3=7$ . As categorias referentes ao indivíduo 1 são  $A_1, B_2$  e  $C_1$ .

Amostras	A <sub>1</sub>	A <sub>2</sub>	B <sub>1</sub>	B <sub>2</sub>	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>
1	1	0	0	1	1	0	0
2	0	1	1	0	1	0	0
3	0	1	0	1	0	1	0
4	1	0	0	1	0	0	1
Marginal	2	2	1	3	2	1	1

Figura 5 Representação de uma matriz indicadora de quatro amostras de café avaliadas em três variáveis

## 2.5 Nuvem de pontos

Segundo Le Roux e Rouanet (2010), a ACM, assim como outros métodos de análise geométrica de dados (AGD), é baseada na observação de uma nuvem de pontos. Uma nuvem de pontos é definida como um conjunto finito de pontos em um espaço geométrico. Na AGD, consideram-se tais nuvens como um conjunto de observações dispostas como pontos em um espaço euclidiano com mais de duas dimensões chamado espaço multidimensional. O objetivo da ACM é reduzir esse espaço multidimensional em um subespaço ótimo para que seja possível o estudo das nuvens de pontos.

O procedimento da ACM gera duas nuvens de pontos. Uma referente às linhas, nuvem dos  $I$  pontos dos indivíduos e a outra referente às colunas, nuvem dos  $K$  pontos das categorias.

A dimensão de uma nuvem de pontos é dada pelo número de informações de cada linha, ou coluna, menos um. Como o número de colunas é referente às  $K$  categorias das  $Q$  variáveis, tem-se que a dimensionalidade máxima da nuvem de categorias é dada por:

$$(K_1 - 1) + \dots + (K_q - 1) + \dots + (K_Q - 1) = (K_1 + \dots + K_q + \dots + K_Q) + (-1)Q = K - Q.$$

Para realizar a redução da dimensionalidade de uma nuvem de pontos, é necessário conhecer determinados conceitos geométricos tais como: ponto médio, distância e variância entre os pontos. A redução da dimensionalidade é feita normalmente para  $R^2$ , devido à facilidade de interpretação de uma nuvem de pontos em um plano. Este subespaço ótimo é definido por eixos chamados eixos principais.

### 2.5.1 Ponto médio de uma nuvem de pontos

Le Roux e Rouanet (2010) descrevem o ponto médio de uma nuvem de pontos da seguinte forma: Seja  $P$  um ponto qualquer no espaço e  $(M^i)_{i=1,2,\dots,n}$  pontos da nuvem de indivíduos. O ponto médio da nuvem é o ponto  $G$  do vetor  $\overline{PG}$  obtido a partir da razão entre o somatório dos vetores formados entre o ponto  $P$  e todos os outros pontos da nuvem pelo número total de pontos, ou seja,

$$\overline{PG} = \frac{1}{n} \sum \overline{PM^i}.$$

O ponto  $G$  não depende da escolha do ponto  $P$ , isto é, qualquer que seja o ponto  $P$  escolhido, o ponto  $G$  será sempre o mesmo.

Ao substituir  $P$  pelo ponto  $G$ , tem-se o vetor nulo.

$$\frac{1}{n} \sum \overline{GM^i} = \vec{0},$$

ou seja, a média dos desvios do ponto médio para os pontos da nuvem é igual a zero.

Assim, pode-se definir o ponto  $G$  como a média das coordenadas dos pontos, dado por

$$G = \frac{1}{n} \sum M^i.$$

*Exemplo:* Seja a nuvem de pontos de coluna dada na Figura 1(a), com as coordenadas da Tabela 2. Assumindo o ponto  $P$  como a origem,  $P=O$ , e sendo os vetores  $\overline{OM^i}$  dados pelas coordenadas dos pontos  $M^i$ , tem-se:

$$\overrightarrow{OG} = \frac{1}{3}(\overrightarrow{OM^1} + \overrightarrow{OM^2} + \overrightarrow{OM^3}),$$

$$\overrightarrow{OG} = \frac{1}{3}((2,1) + (3,1) + (1,2)) = \frac{1}{3}(6,4) = (2,1.34).$$

Na Figura 6, apresenta-se a nuvem de pontos, juntamente com o ponto médio  $G$ .

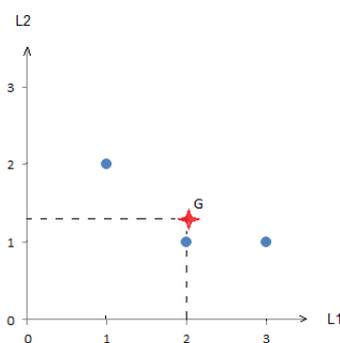


Figura 6 Representação gráfica do ponto médio  $G$  de uma nuvem de pontos

Assim,  $G$  tem coordenadas dadas pela média das coordenadas dos pontos da nuvem. O ponto médio da nuvem de categorias de linha ou de qualquer outra nuvem de pontos é obtido de forma análoga.

## 2.5.2 Distâncias

### 2.5.2.1 Distância entre indivíduos

Sejam os indivíduos  $i$  e  $i'$ , e seja a variável  $q$ , com  $K_q$  categorias, de acordo com Le Roux e Rouanet (2010), a distância entre dois indivíduos é concebida pelas diferentes escolhas de categoria para cada variável  $q$  e é denotada por  $d_q(i, i')$ .

Se dois indivíduos  $i$  e  $i'$ , escolherem a mesma categoria da variável  $q$ , tem-se que a distância entre os dois indivíduos devido à variável  $q$  é nula:

$$d_q(i, i') = 0.$$

Suponha que o indivíduo  $i$  escolha a categoria  $k$ , e o indivíduo  $i'$  escolha a categoria  $k'$  diferente de  $k$ . O quadrado médio da distância entre os indivíduos  $i$  e  $i'$  devido à questão  $q$  é dado por:

$$d_q^2(i, i') = 1/f_k + 1/f_{k'},$$

em que  $f_k$  é a frequência relativa dos indivíduos que escolheram a categoria  $k$ , ou seja,

$$f_k = \frac{n_k}{n},$$

sendo  $n_k$  o número de indivíduos que escolheram a categoria  $k$  e  $n$  o número total de indivíduos.

Sendo  $Q$  o número de variáveis, o quadrado médio da distância total entre  $i$  e  $i'$  é definido por:

$$d^2(i, i') = \frac{1}{Q} \sum_{q=1}^Q d_q^2(i, i').$$

Em uma matriz indicadora, o quadrado médio da distância entre os indivíduos  $i$  e  $i'$  pode ser expresso por:

$$d^2(i, i') = \frac{1}{Q} \sum_{k=1}^K (z_{ik} - z_{i'k})^2 / f_k,$$

em que  $z_{ik} = 1$ , se o indivíduo  $i$  escolheu a categoria  $k$  e  $z_{ik} = 0$  caso contrário,  $z_{i'k}$  é definido da mesma forma.

Quanto menores as frequências das categorias de desacordo, maior será a distância entre os indivíduos. Quanto maior o número de categorias menos frequente escolhidas pelo indivíduo  $i$ , mais afastado do centro estará o ponto  $M^i$  (Le Roux e Rouanet, 2010).

### 2.5.2.2 Distância entre categorias

Seja  $(M^k)_{k=1,2,\dots,K}$  os pontos da nuvem de categorias, o peso do ponto  $M^k$  é dado pelo número de indivíduos que escolheram a categoria  $k$ , e denotado por  $n_k$ . Para cada variável  $q$ , a soma dos pesos dos pontos é  $n$ , e como consequência, a soma para  $K$  é  $nQ$ . O peso relativo  $p_k$  ou massa do ponto  $M^k$  é dado por  $p_k = n_k / nQ = f_k / Q$  e a soma dos pesos relativos para cada variável  $q$  é  $1/Q$ . Logo, a soma dos pesos relativos para todas as variáveis é 1, isto é,

$$p_k = \frac{n_k}{nQ} = \frac{f_k}{Q} \text{ com } \sum_{k \in K_q} p_k = \frac{1}{Q} \text{ e } \sum_{k=1}^K p_k = 1.$$

Seja  $n_{kk'}$  o número de indivíduos que escolheram ambas as categorias  $k$  e  $k'$ , o quadrado da distância entre  $M^k$  e  $M^{k'}$  é dado pela fórmula:

$$d^2(M^k, M^{k'}) = \frac{n_k + n_{k'} - 2n_{kk'}}{n_k n_{k'} / n}.$$

Se  $k$  e  $k'$  são duas categorias diferentes da mesma variável  $q$ , então  $n_{kk'} = 0$ .

Quanto mais categorias  $k$  e  $k'$  escolhidas pelos mesmos indivíduos, menor a distância entre  $M^k$  e  $M^{k'}$ , isto é, quanto mais próximos dois pontos de categorias estiverem, maior a relação de associação entre elas. Quanto menor a frequência da categoria  $k$ , mais distante do centro estará o ponto  $M^k$  (Le Roux e Rouanet, 2010).

### 2.5.3 Variância

A variância de uma nuvem de pontos é definida como a média dos quadrados das distâncias entre os  $M$  pontos da nuvem e um ponto  $P$  qualquer, menos o quadrado da distância entre o ponto médio  $G$  e o ponto  $P$ :

$$V_n = \frac{1}{n} \sum \|\overline{PM}\|^2 - \|\overline{PG}\|^2,$$

em que  $\|\overline{PM}\|$  e  $\|\overline{PG}\|$  são as normas, ou distância entre os pontos, dos respectivos vetores.

Assumindo  $P$  como o ponto médio da nuvem ( $P = G$ ), a variância de uma nuvem de pontos é definida pela média dos quadrados das distâncias entre os pontos da nuvem e o ponto médio  $G$  dada por

$$V_n = \frac{1}{n} \sum \|\overline{GM}\|^2,$$

em que  $\|\overline{GM}\|$  é a norma do vetor  $\overline{GM}$ .

Se o indivíduo  $i$  é representado pelo ponto  $M^i$  e  $G$  é o ponto médio da nuvem, o quadrado da distância de  $M^i$  ao ponto  $G$  é dada por:

$$\|\overline{GM^i}\|^2 = \left( \frac{1}{Q} \sum_{k \in K_i} \frac{1}{f_k} \right) - 1,$$

em que  $K_i$  representa o padrão de respostas do indivíduo  $i$ , isto é, o conjunto de categorias escolhidas pelo indivíduo  $i$ ; e  $f_k$  a frequência relativa dos indivíduos que escolheram a categoria  $k$ .

A variância da nuvem de indivíduos é dada pela razão do número de categorias pelo número de variáveis menos 1, conforme demonstrado em Le Roux e Rouanet (2004), ou seja,

$$V_n^I = \frac{1}{n} \sum_{i=1}^n \|\overline{GM^i}\|^2 = \frac{1}{nQ} \sum_{i=1}^n \sum_{k=1}^K \frac{z_{ik}}{f_k} - \sum_{i=1}^n \frac{1}{n} = \frac{1}{nQ} \sum_{k=1}^K \frac{n_k}{f_k} - 1 = \frac{K}{Q} - 1. \quad (1)$$

Por definição, a variância da nuvem de categorias é dada por:

$$\sum_{k=1}^K p_k \|\overline{GM^k}\|^2.$$

Sabendo que  $\|\overline{GM^k}\|^2 = (1/f_k) - 1$ , tem-se:

$$V_n^K = \sum_{k=1}^K p_k \|\overline{GM^k}\|^2 = \sum_{k=1}^K \frac{f_k}{Q} \left( \frac{1}{f_k} - 1 \right) = \frac{K}{Q} - 1 \quad (2)$$

Pelas equações (1) e (2), tem-se que a variância é a mesma para nuvens de indivíduos e categorias.

Quanto menos frequente uma categoria, mais ela contribui para a variânciageral. Da mesma forma, quanto menos frequente o padrão de respostas de um indivíduo, mais ele contribui para a variância (Le Roux e Rouanet, 2010).

#### 2.5.4 Eixos principais

Seja  $\ell$  uma reta qualquer e  $P$  um ponto não pertencente a  $\ell$ . A projeção ortogonal do ponto  $P$  em  $\ell$  é o ponto  $P'$  tal que  $\overline{PP'}$  é perpendicular a  $\ell$ , ou seja, entre todos os pontos pertencentes a  $\ell$ ,  $P'$  é o ponto relativo à menor distância entre  $\ell$  e  $P$ . Na Figura 7, ilustra-se essa afirmação.

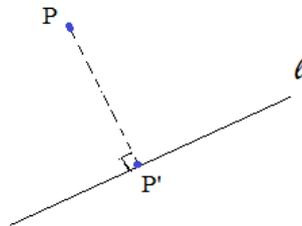


Figura 7 Representação do ponto  $P'$ , projeção ortogonal do ponto  $P$  em  $\ell$

O primeiro eixo principal de uma nuvem é definido como a linha  $\ell$  que passa pelo ponto médio da nuvem, de forma que,  $\ell$  apresente as menores distâncias aos pontos da nuvem e a maior variância da nuvem projetada.

Maximizar a variância da nuvem projetada equivale a minimizar a soma dos quadrados dos desvios residuais. Vale lembrar que desvio residual é o vetor que liga um ponto  $P$  a sua projeção  $P'$ .

*Exemplo:* Considere a nuvem de pontos dada na Figura 1(a). Seja  $\ell$  uma linha que passa pelo ponto médio da nuvem. A projeção ortogonal dos pontos da nuvem na linha  $\ell$  é mostrada na Figura 8.

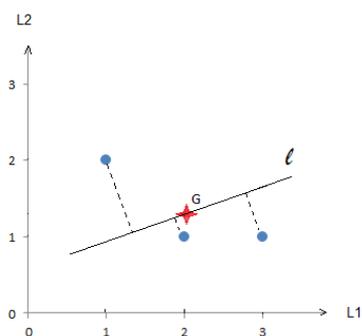


Figura 8 Representação da projeção ortogonal dos pontos da nuvem sobre a linha  $\ell$

O segundo eixo principal é determinado por  $\ell'$ , de forma que  $\ell$  e  $\ell'$  sejam perpendiculares. Os demais eixos principais obtêm-se de forma semelhante tal que  $\ell = 1, 2, \dots, L$  sendo  $L$  a dimensionalidade da nuvem.

A variância da nuvem projetadas obre o eixo  $\ell$  é chamadaa variânciado eixo  $\ell$  ou  $\ell$ -ésimo autovalor da matriz indicadora e é denotada por  $\lambda_\ell$ .

A soma dos autovalores  $\lambda_\ell$  é igual à variância da nuvem de pontos, ou seja,

$$\sum_{\ell=1}^L \lambda_\ell = V_n.$$

Na prática, os  $\lambda_\ell$  são determinados pela fatoração da matriz indicadora através do método de decomposição por valores singulares (DVS). Para mais detalhes sobre o procedimento da DVS consultar Good (1969) e Izenman (2008).

### 2.5.5 Coordenadas principais

A coordenada principal do ponto  $M^i$  da nuvem de indivíduos em relação ao eixo principal é denotada por  $y_\ell^i$ . Para a nuvem de categorias, as coordenadas são definidas da mesma forma, sendo a coordenada principal de  $M^k$  denotada por  $y_\ell^k$ .

Para cada eixo principal, tem-se que a média das coordenadas é nula e a variância é igual ao autovalor. Assim, tem-se

$$\sum \frac{1}{n} y_\ell^i = 0; \sum \frac{1}{n} (y_\ell^i)^2 = \lambda_\ell \text{ e } \sum p_k y_\ell^k = 0; \sum p_k (y_\ell^k)^2 = \lambda_\ell.$$

Assim como os eixos principais, as coordenadas são determinadas pela DVS da matriz indicadora.

### 2.5.6 Contribuições

A interpretação gráfica dos resultados da ACM pode ser confusa, dependendo do número de variáveis estudadas. Devido a essa dificuldade, algumas estatísticas são usadas como apoio à interpretação dos resultados. Essas estatísticas, denominadas de contribuições, colaboram para a interpretação dos eixos obtidos a partir da aplicação da ACM.

Duas contribuições podem ser calculadas: contribuição da categoria ou indivíduo para um determinado eixo e a contribuição do eixo para a variância da categoria ou indivíduo (Le Roux e Rouanet, 2010).

#### 2.5.6.1 Contribuição de um ponto a um eixo

Cada ponto de uma nuvem apresenta uma importância diferente a determinado eixo, isto é, o quanto da variância do eixo é devido ao ponto. Essa importância é chamada de contribuição de um ponto a um eixo e é denotada por  $Ctr$ . Através dos coeficientes dessa contribuição é possível identificar quais os pontos devem ser considerados para a interpretação de cada eixo principal.

Seja  $p$  o peso relativo do ponto  $P$  e  $y$  a coordenada do ponto  $P$  referente ao eixo  $\ell$  de variância  $\lambda$ , a contribuição do ponto  $P$  ao eixo  $\ell$ , segundo Le Roux e Rouanet, (2010), é dado por:

$$Ctr = \frac{(py^2)}{\lambda}. \quad (3)$$

A equação (3) se aplica tanto para pontos relativos a indivíduos como para pontos de categorias.

Sendo  $p_i$  o peso relativo do ponto  $M^i$  dado por  $p_i = Q/nQ = 1/n$ , a contribuição do ponto  $M^i$  da nuvem de indivíduos ao eixo  $\ell$  de variância  $\lambda$  é dada por:

$$Ctr_i = \frac{\left(\frac{1}{n}(y^i)^2\right)}{\lambda}.$$

Da mesma forma, para  $p_k = f_k/Q$ , a contribuição do ponto  $M^k$  da nuvem de categorias ao eixo  $\ell$  de variância  $\lambda$  é dada por:

$$Ctr_k = \frac{\left(\frac{f_k}{Q}(y^k)^2\right)}{\lambda}.$$

### 2.5.6.2 Contribuição do eixo para a variância de um ponto

A contribuição de um eixo para a variância de uma categoria ou indivíduo é denotada por  $Cor$ . A partir da  $Cor$ , é possível identificar quais eixos devem ser considerados para a interpretação de cada ponto.

Seja  $G$  o ponto médio de uma nuvem de pontos e  $M$  um ponto qualquer dessa nuvem.  $M'$  é a projeção ortogonal do ponto  $M$  ao eixo principal  $\ell$ . Segundo Le Roux e Rouanet, (2010), a contribuição relativa é dada por:

$$Cor = \frac{(\overline{GM'})^2}{(\overline{GM})^2} = \cos^2 \theta,$$

sendo  $\theta$  o ângulo formado entre  $\overline{GM}$  e  $\overline{GM'}$  ilustrado na Figura 9.

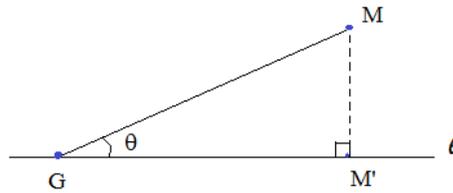


Figura 9 Representação do ângulo formado pelos vetores  $\overline{GM}$  e  $\overline{GM'}$

### 2.5.7 Fórmulas de transição

A visualização das duas nuvens de pontos, nuvem de indivíduos e nuvem de categorias, foram trabalhadas separadamente até aqui. Uma visualização das duas nuvens, em um único espaço, permitiria analisar relações diferentes das estudadas até o momento. Assim, as fórmulas de transição são usadas para a visualização das duas nuvens estudadas em um mesmo plano.

Determinada a nuvem de indivíduos, a partir dos eixos principais, a localização, nesta nuvem, dos pontos relativos às categorias, (ou vice-versa) podem ser definidas pelas fórmulas de transição

$$y^i = \frac{1}{\sqrt{\lambda}} \sum_{k \in K_i} \frac{y^k}{Q} e \quad (4)$$

$$y^k = \frac{1}{\sqrt{\lambda}} \sum_{i \in I_k} \frac{y^i}{n_k}. \quad (5)$$

A partir da equação (4), podem-se calcular as coordenadas principais de um indivíduo através das categorias escolhidas por ele, e com a equação (5), podem-se encontrar as coordenadas principais de uma categoria escolhida por um grupo de indivíduos.

### **2.5.8 Pontos suplementares**

Os pontos, que representam linhas e colunas na AC, são chamados pontos ativos. Pontos ativos são aqueles responsáveis por determinar a orientação dos eixos principais, fornecendo assim, as informações necessárias para a construção das nuvens de baixa dimensionalidade.

Na AC, é possível incluir novas informações em uma análise previamente realizada. Tais informações são representadas pelos pontos suplementares. Os pontos suplementares são plotados no gráfico de correspondência, mas não contribuem para a construção do mesmo. Tais pontos possuem massa igual a zero, não apresentando contribuição a variância dos eixos.

Pontos suplementares podem ser usados para representar informações adicionais sobre os objetos em estudo, para informações invariáveis ao longo do tempo, como sexo e raça, ou ainda para variáveis pouco frequentes, pois estas apresentam contribuições exorbitantes para a variância da nuvem, podendo levar a um resultado falso.

As coordenadas dos pontos suplementares são encontradas, utilizando as fórmulas de transição apresentadas no item anterior. A partir da equação (4), é possível calcular as coordenadas principais de um indivíduo suplementar através das categorias das variáveis ativas escolhidas por ele. Da mesma forma, a

equação (5) permite encontrar as coordenadas principais de qualquer categoria de uma variável suplementar que tenha sido a escolha de um grupo de indivíduos ativos.

## **2.6 Análise de correspondência e dados longitudinais**

Segundo Menard (1991), pesquisa longitudinal diz respeito à coleta e análise de dados ao longo do tempo. Nesse tipo de pesquisa, os dados são coletados para cada item ou variável, para dois ou mais períodos de tempo distintos; os sujeitos ou casos analisados são os mesmos, ou pelo menos comparáveis entre um período e outro. Através de estudos longitudinais, é possível descrever padrões de mudanças, estabelecer direções e proporções às relações causais, como por exemplo, o efeito do tempo sobre as variáveis e indivíduos.

De acordo com a definição de Menard (1991), vários tipos de pesquisa podem ser considerados como longitudinais. O estudo em que as variáveis e objetos analisados permanecem os mesmos entre um período e outro, recebe o nome de estudos em painel.

Como em outras análises estatísticas, uma análise descritiva dos dados longitudinais pode ser de grande ajuda. A avaliação gráfica dos dados longitudinais é extremamente útil para reconhecer as características mais marcantes e fornecer informações sobre padrões de mudanças nos objetos de estudo ao longo do tempo.

Entre as técnicas de análise exploratória de dados longitudinais, a análise de correspondência (AC) tem apresentado bons resultados para estudos qualitativos. Conforme discutido por Van Der Heijden (2005), o uso da

AC para dados longitudinais, apesar de não ter sido citada na literatura muitas vezes, permanece ainda hoje como excelente maneira de analisar dados multivariados categóricos em painel devido a sua flexibilidade de uso. A AC não faz nenhuma suposição de distribuição e não apresenta problemas com observações faltantes, muito comuns em dados em painel, além de permitir a análise de variáveis categóricas (Menard, 2008; Vieira et al., 2011).

Seja  $Z$  a matriz indicadora apresentada na Figura 4 (Página 21) com  $n$  indivíduos e  $Q$  variáveis, cada uma com um número  $K_q$  de categorias, somando um total  $K$  de categorias. Considere agora  $T$  tempos, e que em cada tempo  $t$  uma matriz  $Z_t$  é formada. A matriz super-indicadora  $Z^q$  é uma matriz cúbica de ordem  $n \times K \times T$  com elementos  $z_{ikt}^q$  em que  $q$  indica a variável,  $i$  os objetos,  $k$  as categorias da variável  $q$  e  $t$  o momento da observação.

A construção de uma matriz super-indicadora é exemplificado a partir da matriz indicadora apresentada na Figura 5.

Seja  $Z$  a matriz indicadora para quatro amostras de café apresentada na Figura 5, em que as 4 amostras de café são representadas nas linhas e as 3 variáveis analisadas, com  $K_1 = 2; K_2 = 2; K_3 = 3$ , somando um total  $K = 7$  de categorias, representadas nas colunas. Os elementos da matriz são dados por  $z_{ik}^q$ , sendo o elemento  $z_{42}^3 = 0$ , representado como a casela hachurada. Considere agora que as amostras de café foram avaliadas em 3 tempos diferentes, e que em cada tempo  $t$  uma matriz  $Z_t$  é formada. A matriz super-indicadora  $Z^3$  é uma matriz cúbica de ordem  $4 \times 7 \times 3$ , resultante da junção das três matrizes  $Z_t$  como indicado na Figura 11.

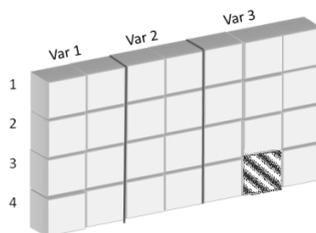


Figura 10 Representação da matriz indicadora  $Z_t$

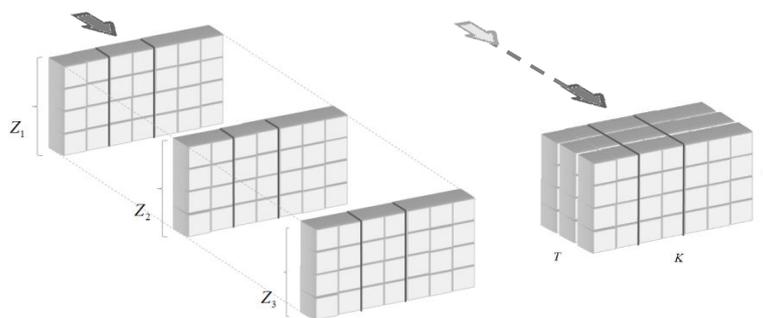


Figura 11 Representação da construção da matriz super-indicadora  $Z^3$

A matriz  $Z^q$  foi mencionada pela primeira vez por Saporta(1981). Como a AC é uma técnica aplicável apenas a matrizes bidimensionais, a matriz super-indicadora deve ser reduzida a uma matriz de duas dimensões.

Em princípio, existem três formas para a redução de dimensionalidade de uma matriz super-indicadora, são elas: analisar cada “fatia” da matriz separadamente; concatenar as fatias da matriz; analisar as matrizes marginais da

matriz super-indicadora. Essas diferentes maneiras de realizar a redução da dimensionalidade da matriz  $Z^q$  são apresentadas por Van Der Heijden (1987). Trataremos aqui da concatenação das fatias da matriz super-indicadora e do estudo das matrizes marginais.

Uma forma possível de concatenação de uma matriz super-indicadora, é a chamada matriz BROAD, proposta por Visser (1985). A matriz BROAD é obtida através da concatenação horizontal das  $Z_t$  matrizes, formando uma matriz de ordem  $n \times KT$ , com os indivíduos nas linhas e as categorias das variáveis sendo repetidas em cada tempo nas colunas. A construção da matriz BROAD a partir da matriz super-indicadora  $Z^3$  (Figura 11) é ilustrada na Figura 12.

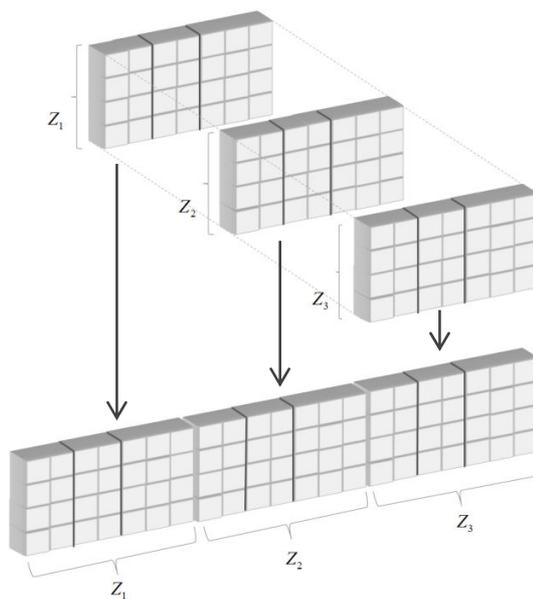


Figura 12 Representação da construção da matriz BROAD

A ACM da matriz BROAD permite identificar as diferentes relações entre as variáveis em cada tempo, isto é, como essa relação muda do tempo  $t$  para o tempo  $t'$ .

A matriz super-indicadora pode resultar em três matrizes marginais diferentes. A primeira alternativa é agregar a matriz  $Z^q$  no tempo, obtendo a matriz marginal  $Z^{IK(q)}$  com elementos  $z_{ik\bullet}^q$ . Agregando pelos objetos tem-se a matriz marginal  $Z^{TK(q)}$  com elementos  $z_{\bullet kt}^q$ . Agregando pelas categorias, tem-se a matriz  $Z^{IT(q)}$ , com elementos  $z_{i\bullet t}^q$ , a qual não apresenta resultados interessantes uma vez que  $z_{i\bullet t}^q = Q$ .

A construção da matriz marginal  $Z^{TK(q)}$  a partir da matriz super-indicadora  $Z^3$  (Figura 11) é apresentada na figura 13.  $Z^{37(3)}$  é obtida a partir do “achatamento” da super-indicadora de forma que cada um de seus elementos é dado por  $z_{\bullet kt}^q$ . Conforme a matriz indicadora da Figura 5, o elemento  $z_{\bullet 21}^1$ , identificado em negrito na Figura 13 recebe o valor 2. As demais matrizes marginais são construídas de forma semelhante, levando em conta a agregação a ser feita.

As matrizes marginais,  $Z^{IK(q)}$  e  $Z^{TK(q)}$  são consideradas, no estudo, por contribuírem para uma melhor compreensão dos resultados da AC, como mostrado por Van Der Heijden (1987). A análise da matriz marginal  $Z^{IK(q)}$  fornece informações sobre a estrutura média no tempo considerado, enquanto que a matriz marginal  $Z^{TK(q)}$  fornece informações de frequência das categorias ao longo do tempo.

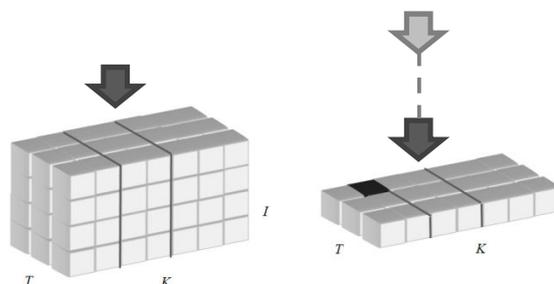


Figura 13 Representação da construção da matriz  $Z^{TK(q)}$

A matriz  $Z^{IK(q)}$  pode ser considerada uma matriz restrita da matriz BROAD, uma vez que os escores de uma dimensão qualquer, da categoria  $k$  da variável  $q$  para todos os tempos devem ser os mesmos para categorias iguais, isto é, categorias iguais devem apresentar os mesmos escores para todos os tempos em cada dimensão. Caso não exista uma mudança considerável na estrutura dos dados, entre os tempos, a solução de cada dimensão (ou os escores) para a matriz restrita  $Z^{IK(q)}$  será parecida com a solução das respectivas dimensões da matriz BROAD. Uma análise de correlação entre os escores pode ser usada para verificar a existência ou não dessa mudança (Van Der Heijden, 1987).

## 2.7 Café

O café, uma das bebidas mais populares do mundo, é cultivado em mais de 80 países, nas regiões tropicais e subtropicais do planeta. De acordo com

Souza (2006), o consumo mundial está relacionado ao prazer, ao relaxamento e ao convívio social, podendo ainda, representar um hábito requintado quando apreciado de acordo com seu cultivo e origem.

No Brasil, a área plantada ultrapassa dois milhões de hectares, colocando o país como o maior produtor mundial de café. Com uma produção aproximada de 43 milhões de sacas em 2011, o Brasil foi responsável por mais de 30% da produção mundial, segundo dados da Organização Internacional do Café (2012).

Para o ano 2012 essa porcentagem pode crescer ainda mais, uma vez que a produção brasileira estimada para o ano é de aproximadamente 50 milhões de sacas. Entre os produtores nacionais, o estado de Minas Gerais aparece como o maior produtor nacional de café, com uma produção estimada de 52% para safra 2012 (CONAB, 2012).

O Brasil, apesar de grande fornecedor mundial, é reconhecido no mercado internacional como produtor de cafés comuns e de baixo preço. Buscando mudar essa visão e aumentar sua participação no mercado de cafés especiais, alguns produtores têm investido na melhoria da qualidade, já que o Brasil apresenta nível tecnológico e parque cafeeiro favoráveis a isso (Giomo e Borém, 2011).

O segmento dos cafés especiais surgiu na década de 70, nos Estados Unidos, com a criação da Specialty Coffee Association of America (SCAA) por produtores interessados no incentivo à produção e ao consumo de cafés especiais. Esses cafés diferenciam-se dos demais por apresentarem características específicas relacionadas ao aspecto físico dos grãos, local de origem, forma de cultivo e processamento, produção limitada, dentre outras (Giomo e Borém, 2011). De acordo com a Associação Brasileira de Cafés

Especiais (BSCA), o segmento representa cerca de 12% do mercado internacional.

De acordo com Uejo Neto (2007), o conceito de cafés especiais é bastante amplo e pode ser percebido de formas diferentes por produtores, importadores, distribuidores e consumidores. Segundo a percepção da qualidade sensorial, um café especial deve corresponder a uma bebida adocicada, sem ocorrência de asperezas ou adstringência, apresentando notas superiores a 80 pontos.

Segundo Souza (2006), a diferenciação do café, com base em parâmetros de qualidade, é feita através de fatores genéticos, ambientais e daqueles relacionados à condução e manejo da lavoura cafeeira, às alterações físico-químicas, fisiológicas e bioquímicas que ocorrem nos grãos ou sementes do café durante o processamento. Saber como esses fatores influenciam na qualidade do café, e de que forma ocorre essa influência é fundamental para a produção de cafés especiais.

### **2.7.1 Classificação do café e análise sensorial**

No Brasil, as primeiras normas para classificação do café foram elaboradas em 1917. Atualmente, a classificação dos cafés comercializados nas bolsas de mercadorias, denominados cafés commodities, é conduzida pela Instrução Normativa nº 8, de 11 de junho de 2003 do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) (Brasil, 2003). Para os cafés especiais, a classificação leva em conta, além da nota global, as pontuações dos atributos sensoriais que constituem a qualidade da bebida. Entre as diferentes

metodologias disponíveis para a análise sensorial dos cafés especiais, o método da SCAA tem sido bastante utilizado (Malta, 2011).

Conforme o protocolo para análise sensorial de café da SCAA (2009), a análise sensorial busca descrever sabores e aromas, determinar diferentes características sensoriais entre as diferentes amostras e determinar uma preferência.

A análise sensorial deve ser realizada por provadores devidamente treinados e habilitados pela SCAA, sendo avaliados onze atributos para o café: fragrância/aroma, uniformidade, ausência de defeitos (xícara limpa), doçura, sabor, acidez, corpo, finalização, equilíbrio, defeitos e avaliação global/final. Nessa avaliação, são atribuídas notas entre 6 e 10 com intervalos de 0,25 pontos. Notas abaixo de 6 pontos são aplicáveis apenas aos cafés comerciais.

O resultado final da análise sensorial é calculado através da soma dos 11 atributos pontuados subtraindo o valor referente ao atributo defeitos. Os resultados finais podem ser expressos segundo a escala de classificação apresentada na Tabela 3.

Para o êxito da análise, é importante seguir todas as orientações e normas estabelecidas pela SCAA referentes aos equipamentos, preparação das amostras e avaliação descritas no protocolo.

Tabela 3 Escala de classificação de cafés especiais.

Pontuação	Descrição	Classificação
90-100	Exemplar	Cafés Especiais
85-89,99	Excelente	Cafés Especiais
80-84,99	Muito bom	Cafés Especiais
<80	Bom	Café não especial

FONTE: Specialty Coffee Association of America (2009).

### **3 MATERIAL E MÉTODOS**

#### **3.1 Material**

A base de dados utilizada neste trabalho para ilustrar a aplicação da ACM foi obtida junto aos responsáveis por um projeto multi-institucional, que tem o objetivo de coletar informações técnico-científicas com o intuito de monitorar a qualidade do café produzido na microrregião da Serra da Mantiqueira em Minas Gerais.

Os dados estudados são referentes aos resultados finais da análise sensorial de 120 amostras de café coletadas no ano de 2011. As amostras foram coletadas de acordo com as características de altitude, cor do fruto, processamento e vertente, sendo que a vertente refere-se à encosta de uma montanha e é classificada conforme a exposição ao sol.

A análise sensorial foi realizada por um grupo de quatro provadores treinados utilizando a metodologia da SCAA (2009). O resultado final foi categorizado segundo a classificação da SCAA apresentada na Tabela 3 e recebeu o nome de variável qualidade.

As categorias das variáveis utilizadas na ACM e suas respectivas codificações são apresentadas na Tabela 4.

Tabela 4 Codificação das categorias das variáveis consideradas na ACM.

Variável	Categoria	Codificação
Altitude	1	Alt1
	2	Alt2
	3	Alt3
Processamento	1	Proc1
	2	Proc2
Cor do Fruto	1	Cr1
	2	Cr2
Vertente	1	Vert1
	2	Vert2
Qualidade	Bom	Qld_Bom
	Muito Bom	Qld_MB
	Excelente	Qld_Exc
	Exemplar	Qld_Exp
Provedor	1	P1
	2	P2
	3	P3
	4	P4

### 3.2 Métodos

Conforme os objetivos propostos, a aplicação das diferentes abordagens da análise de correspondência foi considerada com o intuito de comparar os resultados obtidos, identificando as vantagens no uso de cada uma delas. Para a aplicação dessas diferentes abordagens, utilizou-se um conjunto de dados de análise sensorial de amostras de café realizada por quatro provedores treinados.

Para uma primeira análise, foi considerada a média das notas finais entre quatro provadores. Na análise, considerando a estrutura para dados longitudinais, os quatro provadores foram analisados como diferentes tempos observados. Apesar de não serem estritamente longitudinais, os dados da análise sensorial apresentaram uma estrutura de correlação das observações ao longo dos diferentes provadores, tornando possível a aplicação do método proposto em 2.6.

- i. A aplicação das diferentes abordagens seguiram os seguintes passos:
- ii. Construção da matriz indicadora considerando a média entre os provadores para avaliar a associação entre a qualidade das amostras de café e as variáveis: altitude, processamento, cor do fruto e vertente; Construção das matrizes  $Z^{IK(q)}$ ,  $Z^{TK(q)}$  e BROAD buscando verificar se tais matrizes apresentaram contribuições diferentes para a interpretação dos resultados;
- iii. Aplicação da AC nas matrizes encontradas.
- iv. Interpretação dos resultados;

O número de variáveis é  $Q = 5$  com  $K = 3 + 2 + 2 + 2 + 4 = 13$ , sendo 3 altitudes, 2 processamentos, 2 cores de frutos, 2 vertentes e 4 classificações diferentes de qualidade, com  $n = 120$  amostras e número de provadores  $T = 4$ .

As análises foram realizadas no software R (R Development Core Team, 2011), utilizando-se os pacotes *homals* (Leeuw e Mair, 2009), *ca* (Greenacre e Nenadic, 2010).

### 3.2.1 Construção das matrizes

Para a obtenção da matriz indicadora das médias, uma nova variável foi criada, antes da categorização dos resultados finais, considerando a média entre os quatro provadores. Estas médias foram categorizadas segundo a classificação da SCAA. A matriz indicadora de ordem  $n \times K = 120 \times 13$ , foi definida a partir da matriz de frequências de ordem  $n \times Q = 120 \times 5$ , das variáveis: altitude, cor do fruto, processamento, vertente e média dos resultados finais da análise sensorial.

A matriz de frequências com provadores nas linhas e categorias nas colunas resulta na matriz marginal  $Z^{TK(q)}$  de ordem  $T \times K = 4 \times 13$ .

A matriz marginal  $Z^{IK(q)}$  de ordem  $n \times K = 120 \times 13$  é resultado da soma das frequências de cada categoria para todos os provadores.

A matriz BROAD de ordem  $n \times TK = 120 \times 52$  foi construída a partir da concatenação horizontal das matrizes indicadoras  $Z_t$  de ordem  $n \times K$  para cada provador.

As matrizes indicadoras foram obtidas utilizando a função *homals* do pacote de nome análogo.

### 3.2.2 Aplicação da AC

Para aplicação da AC foi utilizado o pacote *ca*. A função *ca* apresenta as inércias principais, além dos escores (coordenadas dos pontos) e contribuições para linhas e colunas da matriz de dados. Utilizando os escores foram construídos os gráficos de correspondência para as categorias.

### 3.2.3 Interpretação dos resultados.

O primeiro critério usado para interpretação dos resultados da ACM é definir o número de dimensões que devem ser consideradas. Para isso, foi calculada a média dos autovalores ( $\bar{\lambda}$ ) e considerados todos os autovalores superiores a essa média. Essa média é dada pela divisão da variância total pela dimensionalidade da nuvem de categorias (Van Der Heijden, 1987).

Para interpretação de cada eixo, consideraram-se as categorias que apresentaram contribuições aos eixos superiores ou próximos à contribuição média ( $\overline{ctr}$ ). Quanto mais afastadas da origem do eixo, mais as categorias contribuem para a variância do mesmo.

A matriz BROAD, de ordem  $n \times KT$ , apresenta normalmente um número elevado de categorias, o que pode dificultar a interpretação dos resultados. Uma forma de auxiliar as interpretações é utilizar restrições, tanto para a contribuição média das variáveis, dada por  $1/Q$ , como para contribuições médias das variáveis em momentos específicos,  $1/QT$  e contribuições médias para um determinado provador  $1/T$ .

A contribuição do eixo para a variância de cada categoria (*cor*) é usada para avaliar quanto de cada categoria foi explicada pelos eixos considerados. Uma vez que a variável foi satisfatoriamente explicada pelos primeiros eixos, incluir novos eixos para a interpretação torna-se desnecessário, pois um novo eixo acrescentaria informações não significativas para a interpretação daquela variável podendo ainda, comprometer a visualização gráfica.

A associação entre as categorias é observada graficamente através da localização de cada uma delas em relação aos eixos. Pontos de categorias,

localizados próximos a origem apresentam associações pequenas com as demais. Enquanto que, quanto mais afastadas da origem, e próximas umas das outras, maior a associação entre as categorias. De forma geral, categorias presentes em um mesmo quadrante, apresentam características comuns.

Para verificar a existência, ou não, de uma possível diferença na estrutura dos dados entre os provedores, foi utilizado o coeficiente de correlação, através do comando *cor* para os escores das matrizes BROAD e  $Z^{K(q)}$ .

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Análise da matriz indicadora das médias

A nuvem de pontos resultante da análise da matriz indicadora das médias,  $Z^m$ , de ordem  $n \times K = 120 \times 13$ , apresenta uma variância de  $(K/Q) - 1 = (13/5) - 1 = 1,6$ . Sua dimensionalidade é dada por  $K - Q = 13 - 5 = 8$ , isto é, o número máximo de eixos necessários para explicar a variância total da nuvem de pontos é 8.

A partir da variância total e da dimensionalidade da nuvem de categorias, é possível obter a variância média da nuvem, dada por  $\bar{\lambda} = \left( \frac{K}{Q} - 1 \right) / (K - Q) = 1,6/8 = 0,2$ . Conhecendo a variância média, foram considerados os cinco primeiros eixos que apresentaram autovalores superiores a  $\bar{\lambda}$  (Tabela 5). Os cinco primeiros eixos, explicam 76,9% da variância da nuvem de categorias.

Tabela 5 Principais inércias e porcentagens explicadas e acumuladas da matriz indicadora das médias.

Dimensão	Inércia	%	% Acumulada
1	0,326933	20,4	20,4
2	0,279708	17,5	37,9
3	0,223916	14,0	51,9
4	0,200000	12,5	64,4
5	0,200000	12,5	76,9
Total	1,230557		

Na Tabela 6, apresentam-se as contribuições das categorias para a inércias dos eixos (*ctr*) e as contribuições dos eixos para a inércia das categorias (*cor*), considerando os cinco primeiros eixos. As variáveis altitude e qualidade estão relacionadas com o primeiro eixo, sendo a contribuição das mesmas para o primeiro eixo 35,9% e 50%, respectivamente. A variável qualidade também está associada ao segundo eixo, juntamente com a variável cor do fruto, responsáveis por explicar respectivamente 50% e 23,4% da variância do eixo dois. As variáveis altitude, vertente e qualidade estão relacionadas ao terceiro eixo, contribuindo com 20,5%, 20,2% e 50% da variância deste eixo respectivamente. O quarto eixo é explicado principalmente pela variável processamento, sendo esta responsável por 67,2% da variância do eixo. As variáveis altitude e vertente estão relacionadas também com o quinto eixo, respondendo por 40,8% e 46% da variância do mesmo.

Como o quarto e quinto eixos não contribuem para a variância da variável qualidade ( $cor = 0$ ), e sendo esta a principal variável de interesse, a interpretação dos eixos ficará restrita aos três primeiros eixos. Os três primeiros eixos explicam 68,4% ( $57,7 + 9,4 + 1,3$ ) da variância da categoria Alt\_1, 42,1% da Alt\_2 e 45,5% da Alt\_3. Para a variável cor do fruto, a contribuição dos três primeiros eixos é de 57% para ambas as categorias. As categorias Qld\_B, Qld\_MB, Qld\_Exc e Qld\_Exp da variável qualidade receberam, respectivamente, uma *cor* de 62,6%, 69,6%, 67,3% e 76,8% referente aos três primeiros eixos. A contribuição dos três primeiros eixos para as categorias das variáveis processamento e vertente foram relativamente baixos, sendo que para as categorias Proc\_1 e Proc\_2 a *cor* foi de 19,6% e para Vert\_1 e Vert\_2 de

26,9%. Desta forma, os três primeiros eixos explicam satisfatoriamente a relação entre as variáveis altitude, cor do fruto e qualidade.

Tabela 6 Contribuições (*ctr* e *cor*) das categorias para os 5 primeiros eixos da matriz indicadora das médias.

Categorias	ctr (%)					cor (%)				
	Eixos					Eixos				
	1	2	3	4	5	1	2	3	4	5
Alt1	<b>23,5</b>	4,5	0,8	<b>9,1</b>	0,0	57,7	9,4	1,3	13,7	0,0
Alt2	3,5	<b>10,3</b>	7,1	4,4	<b>20,4</b>	8,5	21,6	12	6,6	30,6
Alt3	<b>8,9</b>	1,2	<b>12,6</b>	0,9	<b>20,4</b>	21,8	2,5	21,2	1,3	30,6
Total	35,9	16	20,5	14,4	40,8	88	33,5	34,5	21,6	61,2
Proc1	0,3	4,4	2,8	<b>33,6</b>	1,4	1,1	12,3	6,2	67,1	2,8
Proc2	0,3	4,4	2,8	<b>33,6</b>	1,4	1,1	12,3	6,2	67,1	2,8
Total	0,6	8,8	5,6	67,2	2,8	2,2	24,6	12,4	134,2	5,6
Cor1	6,1	<b>11,7</b>	1,9	5,4	5,2	19,9	32,8	4,3	10,7	10,3
Cor2	6,1	<b>11,7</b>	1,9	5,4	5,2	19,9	32,8	4,3	10,7	10,3
Total	12,2	23,4	3,8	10,8	10,4	39,8	65,6	8,6	21,4	20,6
Vert1	0,6	0,9	<b>10,1</b>	3,9	<b>23,0</b>	2,0	2,4	22,5	7,8	46,1
Vert2	0,6	0,9	<b>10,1</b>	3,9	<b>23,0</b>	2,0	2,4	22,5	7,8	46,1
Total	1,2	1,8	20,2	7,8	46,0	4,0	4,8	45,0	15,6	92,2
Qld_B	0,1	<b>20,3</b>	<b>22,9</b>	0,0	0,0	0,2	32,8	29,6	0,0	0,0
Qld_MB	<b>14,3</b>	2,1	<b>12,8</b>	0,0	0,0	40,1	5,0	24,5	0,0	0,0
Qld_Exc	0,4	<b>26,1</b>	6,4	0,0	0,0	1,0	55,4	10,9	0,0	0,0
Qld_Exp	<b>35,2</b>	1,5	7,9	0,0	0,0	64,5	2,4	9,9	0,0	0,0
Total	50,0	50,0	50,0	0,0	0,0	105,8	95,6	74,9	0,0	0,0

Para a interpretação de cada eixo, foram consideradas as categorias que apresentaram contribuições superiores ou próximas a  $\overline{ctr} = 1/K = 1/13 = 0,0769 = 7,69\%$ . Estas contribuições (*ctr*) estão apresentadas em negrito na Tabela 6. As categorias consideradas para o primeiro eixo são responsáveis por 81,9% da variância do mesmo, enquanto que as categorias selecionadas para o segundo e terceiro eixo respondem por 80,1% e 76,4%, respectivamente.

Na Figura 14, têm-se os gráficos de correspondência das categorias para as três primeiras dimensões. Considerando a Figura 14 e as categorias selecionadas na Tabela 6, pode-se observar que para o primeiro eixo, as categorias Alt\_1 e Qld\_Exp estão localizadas à direita enquanto que Alt\_3 e Qld\_MB estão à esquerda do eixo. Dessa forma, pode-se concluir que Alt\_1 está relacionado a cafés de qualidade superior, enquanto que Alt\_3 a cafés com qualidade Muito Bom. Para o segundo eixo, as variáveis Alt\_2, Cor\_1 e Qld\_Exc opõem-se a Cor\_2 e Qld\_B, demonstrando a relação de cafés classificados por Excelente com as categorias 1, para cor do fruto; e 2 para altitude, enquanto que cafés classificados como Bom estão relacionados a categoria 2, da variável cor do fruto. O terceiro eixo (Figura 14-b) separa as categorias Alt\_3, Vert\_2, Qld\_MB das categorias Vert\_1 e Qld\_B, confirmando a relação da categoria 3, da variável altitude, com cafés de qualidade Muito Bom. A terceira dimensão apresenta também uma relativa associação entre as variáveis vertente e qualidade.

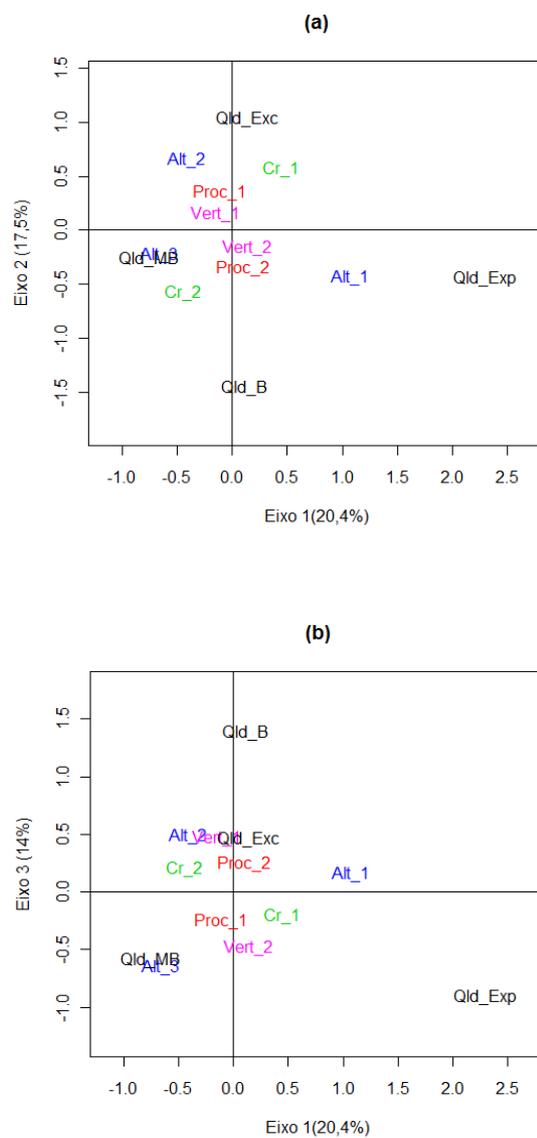


Figura 14 Gráficos de correspondência das categorias da matriz  $Z^m$  para as dimensões 1 e 2 (a), 1 e 3 (b)

#### 4.2 Análise da matriz $Z^{IK(q)}$

A matriz marginal da matriz super-indicadora  $Z^q$ , agregada pelos provadores, é a matriz  $Z^{IK(q)}$  de ordem  $n \times K = 120 \times 13$ , ordem esta, igual à apresentada pela matriz indicadora das médias,  $Z^m$ . Apesar de apresentarem mesma ordem, as matrizes  $Z^{IK(q)}$  e  $Z^m$  são obtidas de formas diferentes. Enquanto  $Z^{IK(q)}$  é uma matriz de frequências das categorias pelos provadores, ou seja, cada elemento é referente ao número de vezes que cada categoria foi relacionada a uma mesma amostra  $i$  considerando os quatro provadores, a matriz  $Z^m$  é uma matriz indicadora, considerando a média entre as notas dadas pelos quatro provadores como valor para categorizar a variável qualidade.

A AC da matriz  $Z^{IK(q)}$  resulta em uma nuvem de pontos com variância igual a 1,26 e dimensionalidade 12. A variância média da nuvem de categorias é dada por  $\bar{\lambda} = 1,26/12 = 0,105$ , sendo assim, foram considerados os cinco primeiros eixos que apresentaram autovalores superiores a  $\bar{\lambda}$  (Tabela 7). Os cinco eixos juntos, explicam 84,9% da variância da nuvem de categorias.

Na Tabela 8, apresentam-se as contribuições *ctr* e *cor* das categorias para os cinco primeiros eixos. É possível identificar quais foram as variáveis que mais contribuíram para a variância dos cinco primeiros eixos. As variáveis altitude, cor do fruto e qualidade estão relacionadas com o primeiro eixo, sendo suas contribuições 28%, 37% e 29,7%, respectivamente. A variável que mais contribui para a variância do segundo eixo é da altitude com uma *ctr* de 68,7%. As variáveis cor do fruto e qualidade apresentam contribuições de 13% e 11,7%. As variáveis altitude e processamento são responsáveis por 23,6% e 49,6% da

Tabela 7 Principais inércias e porcentagens explicadas e acumuladas da matriz  $Z^{K(q)}$ .

Dimensão	Inércia	%	% Acumulada
1	0,256210	20,3	20,3
2	0,214807	17,0	37,3
3	0,201649	16,0	53,3
4	0,200000	15,8	69,1
5	0,200000	15,8	84,9
Total	1,072666		

variância do terceiro eixo e de 23,6% e 49,6% da variância do quarto eixo. A variável vertente explica 75,6% da variância do quinto eixo.

A interpretação gráfica ficará restrita aos dois primeiros eixos, uma vez que os demais eixos apresentam contribuições muito baixas ou nulas para a variância das categorias da variável qualidade (Tabela 8), sendo esta a variável de maior interesse para o estudo. Os dois primeiros eixos explicam 37,3% da variância total da nuvem de categorias.

Na Tabela 8, identifica-se também, quanto da variância de cada categoria é devido aos dois primeiros eixos. A partir da *cor*, é possível calcular que a categoria Alt\_1 tem 81,6% (32,7+48,9) de sua variância devido aos dois primeiros eixos, enquanto que para as categorias Alt\_2 e Alt\_3 a contribuição é de 62,9% e 20,1%. A contribuição dos dois primeiros eixos para cada uma das categorias das variáveis processamento, cor do fruto e vertente são respectivamente, 8,2%, 61,5% e 5,5%. Para a variável qualidade, a *cor* dos três primeiros eixos para suas categorias são: 29,1% para Qld\_B, 42,2% para Qld\_MB, 21,2% para Qld\_Exc e 57,1% para Qld\_Exp. Note a baixa

contribuição dos dois primeiros eixos para as categorias das variáveis processamento e vertente. Desta forma, os primeiros eixos explicam satisfatoriamente a relação entre as variáveis altitude, cor do fruto e qualidade, mesmo resultado apresentado para a matriz indicadora das médias.

Tabela 8 Contribuições (*ctr* e *cor*) das categorias para os 5 primeiros eixos da matriz  $Z^{IK(q)}$ .

Categoria	ctr (%)					cor (%)				
	Eixo 1	Eixo 2	Eixo 3	Eixo 4	Eixo 5	Eixo 1	Eixo 2	Eixo 3	Eixo 4	Eixo 5
Al_1	<b>17,0</b>	<b>30,3</b>	2,3	<b>7,6</b>	0,0	32,7	48,9	3,5	11,4	0,0
Al_2	0,9	<b>38,0</b>	<b>17,7</b>	1,2	3,9	1,7	61,2	26,7	1,8	5,9
Al_3	<b>10,1</b>	0,4	<b>32,8</b>	<b>14,8</b>	3,9	19,4	0,7	49,6	22,2	5,9
Total	28,0	68,7	52,8	23,6	7,8	53,8	110,8	79,8	35,4	11,8
Pr_1	0,5	3,2	<b>19,7</b>	<b>24,8</b>	0,8	1,3	6,9	39,7	49,5	1,6
Pr_2	0,5	3,2	<b>19,7</b>	<b>24,8</b>	0,8	1,3	6,9	39,7	49,5	1,6
Total	1,0	6,4	39,4	49,6	1,6	2,6	13,8	79,4	99,0	3,2
Cr_1	<b>18,5</b>	6,5	1,3	6,5	7,5	47,5	14,0	2,5	13,0	14,9
Cr_2	<b>18,5</b>	6,5	1,3	6,5	7,5	47,5	14,0	2,5	13,0	14,9
Total	37,0	13	2,6	13,0	15,0	95,0	28,0	5,0	26,0	29,8
Vr_1	2,1	0,0	2,1	7,0	<b>37,8</b>	5,4	0,1	4,3	13,9	75,7
Vr_2	2,1	0,0	2,1	7,0	<b>37,8</b>	5,4	0,1	4,3	13,9	75,7
Total	4,2	0,0	4,2	14,0	75,6	10,8	0,2	8,6	27,8	151,4
Qld_B	4,4	<b>8,2</b>	0,1	0,0	0,0	11,4	17,7	0,3	0,0	0,0
Qld_MB	6,1	0,9	0,3	0,0	0,0	37,6	4,6	1,4	0,0	0,0
Qld_Exc	1,7	2,1	0,5	0,0	0,0	10,2	11,0	2,2	0,0	0,0
Qld_Exp	<b>17,5</b>	0,5	0,2	0,0	0,0	55,8	1,3	0,4	0,0	0,0
Total	29,7	11,7	1,1	0,0	0,0	115,0	34,6	4,3	0,0	0,0

Para a interpretação de cada eixo, foram consideradas as categorias que apresentaram contribuições superiores ou próximas a  $\overline{ctr} = 1/K = 1/13 = 0,0769 = 7,69\%$ . Tais contribuições estão apresentadas em negrito na Tabela 8. As categorias consideradas para o primeiro eixo são Alt\_1, Alt\_3, Cr\_1, Cr\_2 e Qld\_Exp, que juntas são responsáveis por 81,6% da variância do referido eixo. Para o segundo eixo foram selecionadas as variáveis Alt\_1, Alt\_2 e Qld\_B representando 76,5% da variância do mesmo.

Na Figura 15, apresenta-se o gráfico de correspondência das categorias para as duas primeiras dimensões. Considerando-se a Figura 15 e as categorias selecionadas na Tabela 8, segundo o critério  $ctr \geq \overline{ctr}$ , pode-se observar que para o primeiro eixo, as categorias Alt\_1, Cr\_1 e Qld\_Exp encontram-se em um mesmo lado do eixo, indicando associação entre tais categorias de variáveis. O segundo eixo opõe as altitudes 1 e 2.

Assim, como nos resultados exibidos para a matriz das médias, os resultados apresentados pela AC da matriz  $Z^{K(q)}$  associam as categorias Alt\_1 e Qld\_Exp, ao mesmo tempo que opõem Qld\_Exp a categoria Alt\_3. Outra semelhança a ser observada entre os resultados apresentados pela AC das matrizes  $Z^{K(q)}$  e  $Z^m$ , é variável Qld\_B estar relacionada a Cr\_2. Tais observações são definidas observando apenas o primeiro eixo do gráfico de correspondência da AC da matriz  $Z^{K(q)}$ , enquanto que para matriz  $Z^m$ , é necessário observar os dois primeiros. De forma geral, os resultados das duas matrizes apontam altitude e cor do fruto como as variáveis que apresentam maior associação com a qualidade do café. Cafés de qualidade superior estão diretamente relacionados à categoria 1 da variável altitude, enquanto categoria 2 da variável cor do fruto está relacionado a cafés de qualidade inferiores.

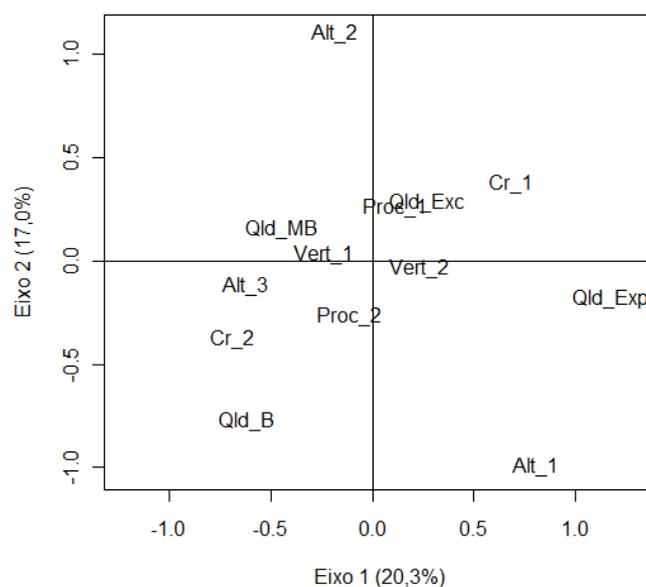


Figura 15 Gráfico de correspondência das categorias da matriz  $Z^{TK(q)}$

#### 4.3 Análise da matriz $Z^{TK(q)}$

A matriz  $Z^{TK(q)}$  é resultado da agregação da matriz super-indicadora  $Z^q$  pelas amostras, ou seja, é a matriz de frequências das categorias considerando as 120 amostras. Cada elemento  $z_{\bullet kt}^q$  de  $Z^{TK(q)}$  é referente ao número de vezes que cada categorial foi relacionada por um dos provedores.

AC da matriz  $Z^{TK(q)}$  de ordem  $T \times K = 4 \times 13$ , resulta em uma nuvem de categorias com variância 0,02 e dimensionalidade 3. Conforme a Tabela 9, os dois primeiros eixos explicam 99,3% da variância total da nuvem de categorias. Dessa forma, a interpretação dos resultados será referente a eles.

Para a interpretação dos eixos, foram consideradas as categorias que apresentaram contribuições superiores a  $\overline{ctr} = 1/K = 1/13 = 0,0769 = 7,69\%$  para as categorias e  $\overline{ctr} = 1/T = 1/4 = 0,25 = 25\%$  para os provadores. Essas contribuições ( $ctr$ ) estão apresentadas em negrito na Tabela 10. As variáveis altitude, cor do fruto, processamento e vertente não são apresentadas na Tabela 10, pois apresentam contribuições nulas, o que era esperado, uma vez que suas frequências foram as mesmas para cada um dos provadores.

As categorias consideradas para o primeiro eixo são Qld\_MB, Qld\_Exc e Qld\_Exp que juntas contribuem com 99,6% da variância do eixo. Para o segundo eixo, 99,8% de sua variância é explicada pelas categorias Qld\_B, Qld\_Exc e Qld\_Exp. O provador P4 é responsável por 65,5% da variância do primeiro eixo, enquanto que P1 e P2 explicam 52,6% e 47% da variância do

Tabela 9 Principais inércias e porcentagens explicadas e acumuladas da matriz  $Z^{TK(q)}$ .

Dimensão	Inércia	%	% Acumulada
1	0,014360	74,1	74,1
2	0,004874	25,2	99,3
3	0,000135	0,7	100,0
Total	0,019369		

segundo eixo. Desta forma, é possível verificar a relação de associação do provador P4 com as qualidades Qld\_MB, Qld\_Exc e Qld\_Exp e os provadores P1 e P3 com as qualidades Qld\_B, Qld\_Exc e Qld\_Exp. O provador P3 apresenta contribuições nulas ( $ctr = 0$ ), o que significa que o perfil de P3 é o perfil médio.

Por meio do gráfico de correspondências para as categorias da matriz  $Z^{TK(q)}$ , apresentado na Figura 16, e das categorias selecionadas na Tabela 10, é possível identificar as associações entre as categorias da variável qualidade e os provadores.

O primeiro eixo opõe a categoria Qld\_MB e o provador P4 às categorias Qld\_Exc e Qld\_Exp. O segundo eixo, separa as categorias Qld\_B, Qld\_Exp e provador P2 da categoria Qld\_Exc e provador P1.

Tabela 10 Contribuições ( $ctr$  e  $cor$ ) das categorias para os 2 primeiros eixos da matriz  $Z^{TK(q)}$ .

Categorias	ctr (%)		cor (%)	
	Eixo 1	Eixo 2	Eixo 1	Eixo 2
Qld_B	0,4	<b>20,5</b>	5,2	87,2
Qld_MB	<b>45,1</b>	0,2	99,6	0,1
Qld_Exc	<b>12,4</b>	<b>55,2</b>	39,9	60,0
Qld_Exp	<b>42,1</b>	<b>24,2</b>	83,4	16,3
P1	13,6	<b>52,6</b>	43,1	56,7
P2	20,9	<b>47,0</b>	56,6	43,2
P3	0,0	0,0	3,8	1,3
P4	<b>65,5</b>	0,3	99,7	0,2

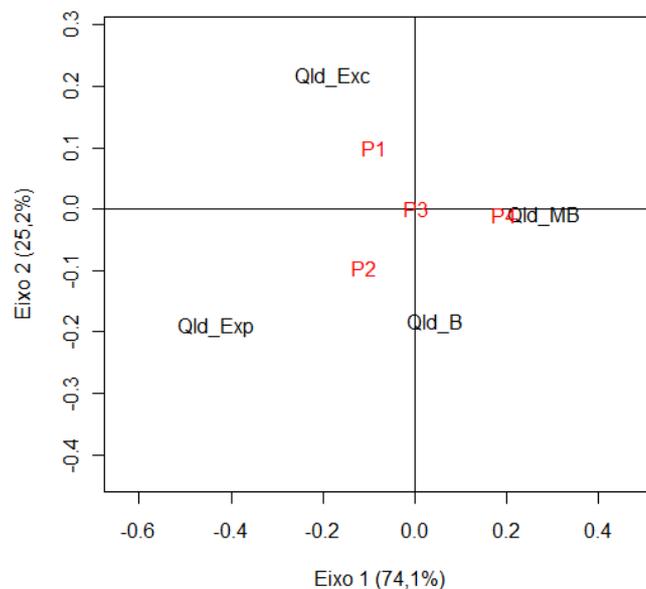


Figura 16 Gráfico de correspondência das categorias da matriz  $Z^{TK(q)}$

Analisando os dois eixos é possível observar as tendências de cada provador. Enquanto o provador P1 apresenta uma preferência em classificar as amostras de café como Excelente, o provador P2 apresenta uma preferência pelas qualidades Bom e Exemplar. O provador P4 apresenta uma predisposição para a utilização da categoria Muito Bom e, ao mesmo tempo, uma tendência em preterir a categoria Exemplar. Como era esperado, devido à sua contribuição nula ( $ctr = 0$ ), o provador P3 encontra-se no centro do gráfico, representando a média dos provadores. P3 pode ser usado para representar todo o grupo de forma

satisfatória. Por estarem distribuídos próximos a origem dos eixos, ou seja, em torno de um perfil médio, os perfis dos provadores P1, P2 e P4 apresentam-se de forma similar.

#### 4.4 Análise da matriz BROAD

Aplicando a AC na matriz BROAD de ordem  $n \times KT = 120 \times 52$ , obtém-se a nuvem de categorias com variância  $(KT/QT) - 1 = (52/20) - 1 = 1,6$  e dimensionalidade  $KT - QT = 52 - 20 = 32$ .

A partir do cálculo da variância média da nuvem de categorias dada por  $\bar{\lambda} = 1,6/32 = 0,05$ , foram considerados, a princípio, os oito primeiros eixos que apresentaram autovalores superiores a  $\bar{\lambda}$  (Tabela 11), que juntos apresentam uma contribuição de 82,3% da variância da nuvem de categorias.

As contribuições das categorias para os oito primeiros eixos e a contribuição dos eixos para cada categoria são apresentadas no Apêndice A. Para fins práticos, foram considerados para a análise final apenas os três primeiros eixos, que contribuem em 43,3% da variância total da nuvem de categorias. Considerar mais que três eixos, torna a interpretação um tanto complexa, podendo comprometer a análise.

Como a matriz BROAD apresenta um número muito grande de informações, a interpretação foi restrita às variáveis que contribuem acima da contribuição média das variáveis, dada por  $1/Q = 1/5 = 0,2$ . Outras restrições utilizadas foram: considerar as médias das variáveis para cada provador, dada por  $1/QT = 1/20 = 0,05$  e a contribuição média de cada provador,  $1/T = 1/4 = 0,25$ . Essas contribuições são apresentadas nas Tabelas 12, 13 e 14,

Tabela 11 Principais inércias e porcentagens explicadas e acumuladas da matriz BROAD.

Dimensão	Inércia	%	% Acumulada
1	0,265148	16,6	16,6
2	0,219713	13,7	30,3
3	0,207207	13,0	43,3
4	0,204670	12,8	56,0
5	0,202553	12,7	68,7
6	0,104502	6,5	75,2
7	0,060419	3,8	79,0
8	0,052654	3,3	82,3
Total	1,316866		

sendo que as contribuições que satisfazem as restrições mencionadas anteriormente encontram-se em negrito.

As variáveis altitude, cor do fruto e qualidade apresentam maiores contribuições para a variância do primeiro eixo, e juntas são responsáveis por 95,1% desta variância. Para o segundo eixo, apenas a variável altitude apresenta contribuição considerável ao eixo, contribuindo com 50% da variância do mesmo. A variável altitude aparece novamente com uma contribuição relevante ao terceiro eixo, e juntamente com a variável processamento, são responsáveis por 81,2% da variância do terceiro eixo.

Tabela 12 Contribuições das variáveis para a variância do primeiro eixo da matriz BROAD por provador.

Variável	Provador				Total
	1	2	3	4	
Altitude	<b>6,8</b>	<b>6,8</b>	<b>6,8</b>	<b>6,8</b>	<b>27,2</b>
Cor	<b>8,8</b>	<b>8,8</b>	<b>8,8</b>	<b>8,8</b>	<b>35,2</b>
Processamento	0,2	0,2	0,2	0,2	0,8
Vertente	1,0	1,0	1,0	1,0	4,0
Qualidade	<b>7,9</b>	<b>7,4</b>	<b>10,2</b>	<b>7,2</b>	<b>32,7</b>
Total	24,7	24,2	<b>27,0</b>	24,0	100,0

Tabela 13 Contribuições das variáveis para a variância do segundo eixo da matriz BROAD por provador.

Variável	Provador				Total
	1	2	3	4	
Altitude	<b>12,5</b>	<b>12,5</b>	<b>12,5</b>	<b>12,5</b>	<b>50,0</b>
Cor	<b>4,0</b>	<b>4,0</b>	<b>4,0</b>	<b>4,0</b>	16,0
Processamento	<b>4,2</b>	<b>4,2</b>	<b>4,2</b>	<b>4,2</b>	16,8
Vertente	0,6	0,6	0,6	0,6	2,4
Qualidade	3,2	2,0	<b>5,5</b>	3,8	14,5
Total	24,5	23,3	<b>26,8</b>	<b>25,1</b>	100,0

Tabela 14 Contribuições das variáveis para a variância do terceiro eixo da matriz BROAD por provador.

Variável	Provador				Total
	1	2	3	4	
Altitude	<b>6,9</b>	<b>6,9</b>	<b>6,9</b>	<b>6,9</b>	<b>27,6</b>
Cor	0,2	0,2	0,2	0,2	0,8
Processamento	<b>13,4</b>	<b>13,4</b>	<b>13,4</b>	<b>13,4</b>	<b>53,6</b>
Vertente	3,4	3,4	3,4	3,4	13,6
Qualidade	1,4	1,7	0,4	0,5	4,0
Total	<b>25,3</b>	<b>25,6</b>	24,3	24,4	100,0

O equilíbrio entre as contribuições para cada provador, expressas no total das colunas das Tabelas 12, 13 e 14, indicam como os provadores estão concordantes. Isto é, apresentando contribuições próximas, pode-se dizer que os provadores discriminam as variáveis de forma similar, sendo o provador P3 aquele que mais discrimina nas duas primeiras dimensões.

Os valores no interior das Tabelas 12, 13 e 14 são referentes às contribuições das variáveis em relação a cada provador. As variáveis selecionadas, considerando contribuições superiores as contribuições médias das variáveis para cada provador, coincidem, em sua maioria, com as variáveis selecionadas através da média das variáveis, demonstrando, mais uma vez, uma concordância entre os provadores.

As variáveis altitude, cor do fruto, processamento e vertente apresentaram sempre as mesmas contribuições para cada um dos provadores. Fato esperado, uma vez que essas variáveis são fixadas, isto é, cada amostra apresenta sempre a mesma categoria para cada uma dessas variáveis, independente do provador. Essas informações podem ainda ser confirmadas no gráfico de correspondência apresentado na Figura 17. A codificação utilizada para a variável qualidade faz referência a três informações: provador, variável e categoria. Sendo que o ponto referente ao provador P4 para a categoria exemplar da variável qualidade é dado por P4\_Qld\_Exp.

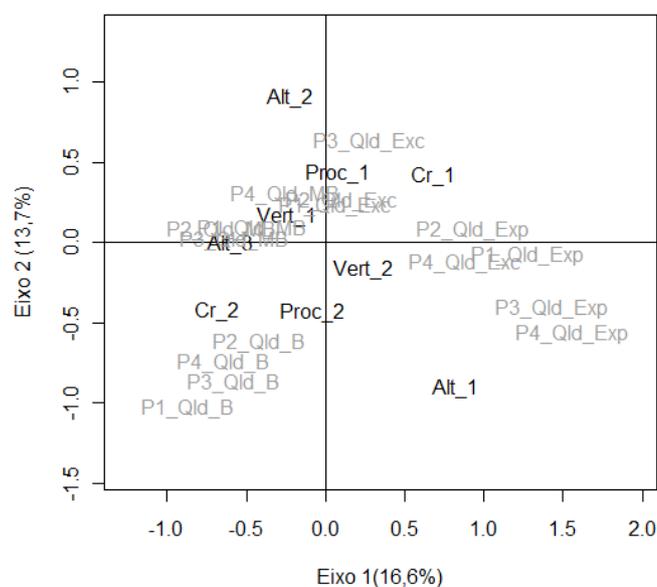


Figura 17 Gráfico de correspondência das categorias para a matriz BROAD

Para uma melhor observação do resultado final da análise sensorial em relação aos provedores, uma vez que na Figura 17 tem-se uma sobreposição muito grande, um segundo gráfico de correspondência é apresentado na Figura 18, exibindo apenas a variável qualidade. A partir desse gráfico, é possível observar que as categorias da variável qualidade formam grupos distintos. A disposição em forma de ferradura ressalta a característica ordinal na variável qualidade, sendo que as classificações seguem de Bom até Exemplar. Fica claro também, uma tendência do provedor P4 em pontuar as amostras abaixo dos

demais provadores, principalmente em relação às amostras de café de melhor qualidade. Enquanto que para os demais provadores algumas amostras são consideradas de qualidade Exemplar, para P4 tais amostras são ainda referentes à qualidade Excelente. Uma correção quanto a essa subestimação pode ser necessária.

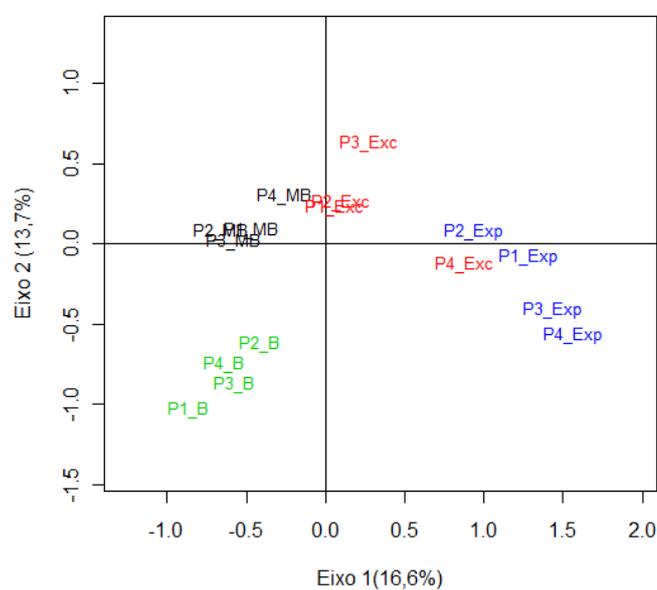


Figura 18 Gráfico de correspondência da categoria qualidade para a matriz BROAD

Agregando a matriz super-indicadora  $Z^q$  pelos provadores obtêm-se a matriz  $Z^{IK(q)}$ .  $Z^{IK(q)}$  é a matriz de frequências das categorias pelos quatro

providores, isto é, a matriz  $Z^{IK(q)}$  é o resultado da redução de dimensionalidade da matriz super-indicadora  $Z^q$  de forma que cada elemento  $z_{ik}^q$  indique o número de vezes em que cada categoria  $k$  da variável  $q$  foi relacionada à amostra  $i$ , considerando os quatro provedores.  $Z^q$  pode também ser considerada como uma matriz restrita da matriz BROAD, uma vez que os escores de categorias iguais devem ser os mesmos para todos os tempos em cada dimensão. Dessa forma, não havendo diferença entre os provedores, espera-se que os escores das amostras da matriz restrita  $Z^{IK(q)}$  sejam próximos aos escores das amostras da matriz BROAD, ou seja, que os resultados correspondentes à AC das duas matrizes não apresentem diferenças significativas. A relação entre os resultados das matrizes BROAD e  $Z^{IK(q)}$  é verificada por meio do coeficiente de correlação, aplicado aos escores das amostras das duas matrizes. A correlação entre os escores da primeira dimensão foi  $r_1 = 0,9979$  e da segunda foi  $r_2 = 0,9628$ . O resultado apresenta uma alta relação entre os escores resultantes da AC das duas matrizes, indicando que a diferença de estrutura nos dados, entre os provedores, não é considerável, isto é, os provedores classificam as amostras de café de formas bem parecidas. Em suma, essa informação confirma a ideia de que os resultados obtidos no estudo das relações das categorias seriam análogos considerando as informações dos quatro provedores (BROAD) ou uma restrição destas informações ( $Z^{IK(q)}$ ).

## 5 CONCLUSÃO

O uso da estrutura da ACM para dados longitudinais, considerando como tempo, cada provador, foi satisfatório. Os resultados apresentados foram aproximados aos apresentados pela ACM da matriz indicadora, considerando as médias entre os quatro provadores. A vantagem da utilização da matriz BROAD e suas matrizes restritas  $Z^{IK(q)}$  e  $Z^{TK(q)}$ , é que a partir delas, novas informações podem ser observadas. Utilizando a matriz BROAD, além da associação entre a qualidade das amostras e as variáveis altitude, processamento, cor do fruto e vertente, é possível verificar a mudança na estrutura dos dados para cada provador. Pela matriz  $Z^{IK(q)}$ , verifica-se a estrutura média entre os dados e é possível comparar se esta estrutura difere da apresentada pela matriz BROAD, podendo assim confirmar o equilíbrio entre os provadores. A matriz  $Z^{TK(q)}$  apresenta a mudança de frequência das categorias para os provadores, ou seja, é possível analisar a relação dos provadores com cada uma das categorias das variáveis.

Como resultado final, o presente trabalho constatou que o uso da ACM para matrizes do tipo BROAD são úteis, quando o objetivo do estudo é analisar o equilíbrio entre provadores. Tal análise é ainda complementada pelo estudo das matrizes restritas  $Z^{IK(q)}$  e  $Z^{TK(q)}$ .

## REFERÊNCIAS

- ASSOCIAÇÃO BRASILEIRA DE CAFÉS ESPECIAIS (BSCA - Brazil Specialty Coffee Association). Disponível em: <<http://bsca.com.br/index.php>>. Acesso em: 25 de março de 2012.
- BEH, E. J. Simple correspondence analysis: a bibliographic review. **International Statistical . Review**, v. 72, n.2, p. 257-284, ago. 2004.
- BENZÉCRI, J.-P. **Correspondence Analysis Handbook**. New York: Dekker. 1992.
- BRASIL.Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº8, de 11 de junho de 2003. [Aprova o Regulamento Técnico de Identidade e de Qualidade para a classificação do Café Beneficiado Grão Cru]. **Diário Oficial [da] República Federativa do Brasil**, Brasília, 13 de junho de 2003. Seção 1, p.22-29.
- CHARNET, R., FREIRE, C. A. L.; CHARNET, E. M. R.; BONVINO, H. **Análise de modelos de regressão linear**. Editora Unicamp, 2008.
- CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da Safra Brasileira: café, Safra 2012**. Disponível em: <<http://www.conab.gov.br>>. Acesso em: 25 de jan. de 2012.
- GIOMO, G. S., BORÉM, F. M. Cafés especiais no Brasil: opção pela qualidade. **Informe Agropecuário**, Belo Horizonte, v. 32, n. 261, p. 7-16, mar./abr. 2011.
- GOOD, I.J. Some applications of the singular value decomposition of a matrix. **Technometrics**, 11, p. 823-831, 1969.
- GREENACRE, M. J. **Theory and Applications of Correspondence Analysis**. London: Academic Press. 1984.
- GREENACRE, M. J.; BLASIUS, J.. **Multiple correspondence analysis and related methods**. Boca-Raton: Chapman-Hall. 2006.

GREENACRE, M. J.; NENADIC, O. **ca**: Simple, Multiple and Joint Correspondence Analysis. R package version 0.33, 2010.

HUSSON, F.; JOSSE, J.; LE, S.; MAZET, J. **FactoMineR**: Multivariate Exploratory Data Analysis and Data Mining with R. R package version 1.16, 2011.

*ISI – Web of Science*. Disponível em: < <http://isiknowledge.com>>. Acesso em: 23 de março de 2012.

IZENMAN, A.J., **Modern Multivariate Statistical Techniques**: Regression, Classification, and Manifold Learning. Springer, Philadelphia, 2008.

LE ROUX, B., ROUANET, H. **Geometric data analysis**. From correspondence analysis to structured data analysis. Dordrecht, the Netherlands: Kluwer-Springer. 2004.

LE ROUX, B., ROUANET, H. **Multiple Correspondence Analysis**. Londres: SAGE, 2010.

LEBART, L., MORINEAU, A., WARWICK, K. Multivariate Descriptive Statistical Analysis. New York: Wiley. 1984.

MALTA, M.R. Critérios utilizados na avaliação da qualidade do café. Informe Agropecuário, Belo Horizonte, v. 32, n. 261, p. 114-126, mar./abr. 2011.

LEEuw, J.; MAIR, P. **Gifi Methods for Optimal Scaling in R**: The Package homals. Journal of Statistical Software, 31(4), 1-20, 2009.

MENARD, S. **Longitudinal Research**. Newbury Park: SAGE, 1991.

MENARD, S. **Handbook of Longitudinal Research Design**: Measurement and Analysis across the Social Sciences. Elsevier. 2008.

NAITO, S.D.N.P. **Análise de correspondências generalizada**. 2007. 156p. Dissertação de mestrado em Bioestatística - Universidade de Lisboa, Lisboa.

ORGANIZAÇÃO INTERNACIONAL DE CAFÉ. **Produção total dos países exportadores**. Disponível em: < <http://www.ico.org/index.asp>>. Acesso em: 25 de março de 2012.

R Development Core Team (2011). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

SAPORTA, G. **Méthodes Exploratoires d'Analyse de Donnés Temporelles**, Unpublished doctoral dissertation, l'Université P. et M. Curie, Paris, 1981.

SOUZA, M. C. M. de. **Cafés sustentáveis e denominação de origem: a certificação de qualidade na diferenciação de cafés orgânicos, sombreados e solidários**. 2006. 177 p. Tese de Doutorado em Ciência Ambiental – Universidade de São Paulo, São Paulo.

SPECIALTY COFFEE ASSOCIATION OF AMERICA. **SCAA Protocols - Cupping Specialty Coffee**. Long Beach: SCAA. 2009. 7p.

VAN DER HEIJDEN, P. G. M. **Correspondence Analysis of Longitudinal Categorical Data**. Leiden: D.S.W.O. Press, Holland, 1987.

VAN DER HEIJDEN, P. G. M. Correspondence Analysis of Longitudinal Data. **Encyclopedia of Biostatistics**. 2nd ed., Peter Armitage and Theodore Colton, eds., Vol. 2, pp. 1230-1234, John Wiley and Sons Ltd., Chichester, 2005.

VIEIRA, M. D. T., BASTOS, R. R., SOUZA, A. C., HIPPERT, H. S. On the use of exploratory and confirmatory longitudinal data analysis. **Advances and Applications in Statistics**, v. 22, n. 2, p. 129-156, 2011.

VISSER, R. A. **The Analysis of Longitudinal Data in Behavioural and Social Research**. Leiden: D.S.W.O. Press, 1985.

UERJO NETO, E. **Compreendendo os cafés especiais**: Coffee Traveler, 2007. Disponível em: <<http://coffeetraveler.net/educacao-servicos-by-enseineto/compreendendo-os-cafes-especiais/>>. Acesso em: 23 de jan. de 2012.

## APÊNDICE

Tabela 1 Principais contribuições (*ctr*) das categorias para os 8 primeiros eixos da matriz BROAD.

Categoria	ctr (%)							
	Eixo 1	Eixo 2	Eixo 3	Eixo 4	Eixo 5	Eixo 6	Eixo 7	Eixo 8
A1_1	4,2	6,2	1,2	1,7	0,5	0,0	1,8	0,0
A1_2	0,3	6,3	4,6	1,7	2,3	0,6	0,5	0,0
A1_3	2,3	0,0	1,1	6,9	4,9	0,5	0,5	0,0
A2_1	4,2	6,2	1,2	1,7	0,5	0,0	1,8	0,0
A2_2	0,3	6,3	4,6	1,7	2,3	0,6	0,5	0,0
A2_3	2,3	0,0	1,1	6,9	4,9	0,5	0,5	0,0
A3_1	4,2	6,2	1,2	1,7	0,5	0,0	1,8	0,0
A3_2	0,3	6,3	4,6	1,7	2,3	0,6	0,5	0,0
A3_3	2,3	0,0	1,1	6,9	4,9	0,5	0,5	0,0
A4_1	4,2	6,2	1,2	1,7	0,5	0,0	1,8	0,0
A4_2	0,3	6,3	4,6	1,7	2,3	0,6	0,5	0,0
A4_3	2,3	0,0	1,1	6,9	4,9	0,5	0,5	0,0
Total	27,2	50,0	27,6	41,2	30,8	4,4	11,2	0,0
P1_1	0,1	2,1	6,7	0,3	2,6	0,2	0,0	0,0
P1_2	0,1	2,1	6,7	0,3	2,6	0,2	0,0	0,0
P2_1	0,1	2,1	6,7	0,3	2,6	0,2	0,0	0,0
P2_2	0,1	2,1	6,7	0,3	2,6	0,2	0,0	0,0
P3_1	0,1	2,1	6,7	0,3	2,6	0,2	0,0	0,0
P3_2	0,1	2,1	6,7	0,3	2,6	0,2	0,0	0,0
P4_1	0,1	2,1	6,7	0,3	2,6	0,2	0,0	0,0
P4_2	0,1	2,1	6,7	0,3	2,6	0,2	0,0	0,0
Total	0,8	16,8	53,6	2,4	20,8	1,6	0,0	0,0
C1_1	4,4	2,0	0,1	0,2	3,0	1,6	0,4	0,0
C1_2	4,4	2,0	0,1	0,2	3,0	1,6	0,4	0,0
C2_1	4,4	2,0	0,1	0,2	3,0	1,6	0,4	0,0
C2_2	4,4	2,0	0,1	0,2	3,0	1,6	0,4	0,0
C3_1	4,4	2,0	0,1	0,2	3,0	1,6	0,4	0,0
C3_2	4,4	2,0	0,1	0,2	3,0	1,6	0,4	0,0
C4_1	4,4	2,0	0,1	0,2	3,0	1,6	0,4	0,0
C4_2	4,4	2,0	0,1	0,2	3,0	1,6	0,4	0,0
Total	35,2	16,0	0,8	1,6	24,0	12,8	3,2	0,0
VI_1	0,5	0,3	1,7	6,5	2,8	0,1	0,1	0,0

Tabela 1, conclusão

V1_2	0,5	0,3	1,7	6,5	2,8	0,1	0,1	0,0
V2_1	0,5	0,3	1,7	6,5	2,8	0,1	0,1	0,0
V2_2	0,5	0,3	1,7	6,5	2,8	0,1	0,1	0,0
V3_1	0,5	0,3	1,7	6,5	2,8	0,1	0,1	0,0
V3_2	0,5	0,3	1,7	6,5	2,8	0,1	0,1	0,0
V4_1	0,5	0,3	1,7	6,5	2,8	0,1	0,1	0,0
V4_2	0,5	0,3	1,7	6,5	2,8	0,1	0,1	0,0
Total	4,0	2,4	13,6	52	22,4	0,8	0,8	0,0
Qld1_B	1,5	2,6	0,1	0,8	0,0	7,3	1,1	0,0
Qld1_MB	1,3	0,1	0,4	0,2	0,0	1,5	0,1	0,2
Qld1_Exc	0,0	0,5	0,8	0,1	0,1	3,5	5,5	1,9
Qld1_Exp	5,1	0,0	0,1	0,2	0,1	0,8	10,8	2,8
Qld2_B	0,6	1,6	0,0	0,0	0,7	20,3	0,3	1,2
Qld2_MB	2,4	0,0	0,8	0,0	0,0	0,6	0,4	10,4
Qld2_Exc	0,0	0,4	0,0	0,1	0,4	2,5	14,5	7,5
Qld2_Exp	4,4	0,0	0,9	0,1	0,0	2,0	12,0	0,0
Qld3_B	0,8	2,3	0,1	0	0,0	13,6	0,1	0,2
Qld3_MB	2,6	0,0	0,0	0,2	0,0	0,2	0,5	0,3
Qld3_Exc	0,4	2,6	0,1	0,6	0,0	1,6	2,4	0,2
Qld3_Exp	6,4	0,6	0,2	0,2	0,0	0,8	11,1	1,1
Qld4_B	1,3	2,1	0,0	0,0	0,0	21,9	1,3	0,1
Qld4_MB	0,7	1,2	0,1	0,0	0,0	2,8	9,1	0,2
Qld4_Exc	2,9	0,1	0,3	0,1	0,0	1,3	15,7	21,3
Qld4_Exp	2,3	0,4	0,1	0,1	0,0	0,2	0,1	52,7
Total	32,7	14,5	4,0	2,7	1,3	80,9	85	100,0

Tabela 2 Principais contribuições (*cor*) das categorias para os 8 primeiros eixos da matriz BROAD.

Categoria	cor (%)							
	Eixo 1	Eixo 2	Eixo 3	Eixo 4	Eixo 5	Eixo 6	Eixo 7	Eixo 8
A1_1	33,7	40,7	7,3	10,6	3,0	0,0	3,3	0,0
A1_2	2,4	41,3	28,8	10,4	13,8	2,0	0,8	0,0
A1_3	18,0	0,0	7,1	42,2	29,6	1,6	0,8	0,0
A2_1	0,7	18,7	55,5	2,4	21,3	0,8	0,0	0,0
A2_2	0,7	18,7	55,5	2,4	21,3	0,8	0,0	0,0
A2_3	46,5	17,8	0,6	1,8	24,7	6,5	1,0	0,0
A3_1	46,5	17,8	0,6	1,8	24,7	6,5	1,0	0,0
A3_2	5,7	2,8	14,4	53,1	22,8	0,2	0,3	0,0
A3_3	5,7	2,8	14,4	53,1	22,8	0,2	0,3	0,0
A4_1	9,1	12,8	0,6	3,6	0,2	17,1	1,5	0,0
A4_2	9,6	0,4	2,6	1,3	0,0	4,5	0,2	0,3
A4_3	0,2	3,7	6,0	0,5	0,5	12,4	11,3	3,4
P1_1	32,5	0,1	0,3	1,2	0,3	2,0	15,6	3,5
P1_2	33,7	40,7	7,3	10,6	3,0	0,0	3,3	0,0
P2_1	2,4	41,3	28,8	10,4	13,8	2,0	0,8	0,0
P2_2	18,0	0,0	7,1	42,2	29,6	1,6	0,8	0,0
P3_1	0,7	18,7	55,5	2,4	21,3	0,8	0,0	0,0
P3_2	0,7	18,7	55,5	2,4	21,3	0,8	0,0	0,0
P4_1	46,5	17,8	0,6	1,8	24,7	6,5	1,0	0,0
P4_2	46,5	17,8	0,6	1,8	24,7	6,5	1,0	0,0
C1_1	5,7	2,8	14,4	53,1	22,8	0,2	0,3	0,0
C1_2	5,7	2,8	14,4	53,1	22,8	0,2	0,3	0,0
C2_1	3,8	8,5	0,2	0,1	3,6	52,0	0,5	1,6
C2_2	18,5	0,3	4,6	0,1	0,1	1,6	0,7	15,6
C3_1	0,3	2,3	0,2	0,8	2,2	6,9	23,3	10,5
C3_2	31,6	0,2	5,2	0,6	0,0	5,8	19,8	0,0
C4_1	33,7	40,7	7,3	10,6	3,0	0,0	3,3	0,0
C4_2	2,4	41,3	28,8	10,4	13,8	2,0	0,8	0,0
V1_1	18,0	0,0	7,1	42,2	29,6	1,6	0,8	0,0
V1_2	0,7	18,7	55,5	2,4	21,3	0,8	0,0	0,0
V2_1	0,7	18,7	55,5	2,4	21,3	0,8	0,0	0,0
V2_2	46,5	17,8	0,6	1,8	24,7	6,5	1,0	0,0
V3_1	46,5	17,8	0,6	1,8	24,7	6,5	1,0	0,0
V3_2	5,7	2,8	14,4	53,1	22,8	0,2	0,3	0,0
V4_1	5,7	2,8	14,4	53,1	22,8	0,2	0,3	0,0

Tabela 2, conclusão

V4_2	5,1	11,6	0,5	0,0	0,1	32,8	0,1	0,2
Qld1_B	23,8	0,0	0,2	1,1	0,0	0,8	1,0	0,5
Qld1_MB	3,1	15,7	0,5	3,2	0,0	4,6	4,0	0,3
Qld1_Exc	40,8	3,3	0,9	0,9	0,1	2,1	16,1	1,3
Qld1_Exp	33,7	40,7	7,3	10,6	3,0	0,0	3,3	0,0
Qld2_B	2,4	41,3	28,8	10,4	13,8	2,0	0,8	0,0
Qld2_MB	18,0	0,0	7,1	42,2	29,6	1,6	0,8	0,0
Qld2_Exc	0,7	18,7	55,5	2,4	21,3	0,8	0,0	0,0
Qld2_Exp	0,7	18,7	55,5	2,4	21,3	0,8	0,0	0,0
Qld3_B	46,5	17,8	0,6	1,8	24,7	6,5	1,0	0,0
Qld3_MB	46,5	17,8	0,6	1,8	24,7	6,5	1,0	0,0
Qld3_Exc	5,7	2,8	14,4	53,1	22,8	0,2	0,3	0,0
Qld3_Exp	5,7	2,8	14,4	53,1	22,8	0,2	0,3	0,0
Qld4_B	8,2	11,1	0,2	0,2	0,2	54,8	1,9	0,1
Qld4_MB	8,9	12,8	1,0	0,3	0,3	14,3	26,3	0,5
Qld4_Exc	19,2	0,4	1,7	0,7	0,1	3,5	23,7	28,0
Qld4_Exp	12,9	1,7	0,4	0,2	0,0	0,5	0,1	58,4