

Um agrupamento de turistas – Modelação adequada?

Joel Adilson da Costa Prata

Dissertação de Mestrado
em Prospeção e Análise de Dados

Orientador(a):
Prof^a. Doutora Margarida G.M.S. Cardoso, ISCTE Business School, Departamento de
Métodos Quantitativos

Abril 2011

Um agrupamento de turistas – Modelação adequada?

Um agrupamento de turistas – Modelação adequada?

Joel Adilson da Costa Prata

Dissertação de Mestrado
em Prospeção e Análise de Dados

Orientador(a):
Prof^a. Doutora Margarida G.M.S. Cardoso, ISCTE Business School, Departamento de
Métodos Quantitativos

Abril 2011

Um agrupamento de turistas – Modelação adequada?

“That which does not kill us makes us stronger”

Nietzsche, Friedrich

Agradecimentos

Antes de tudo queria agradecer a Deus por tudo, desde a minha existência até ao fim dos meus dias.

Dedico também o trabalho aos meus pais, João Rodrigues Prata e Hermengarda Abílio da Costa Prata que são o meu suporte emocional, financeiro e o seu amor incondicional.

Dedico de igual modo o trabalho aos meus irmãos, Eng. Márcio Prata e Miguel Prata.

Agradeço a paciência e a sabedoria da minha tutora, Professora Doutora Margarida G.M.S. Cardoso e sem esquecer a Professora Doutora Elisabeth Kastenholz pela disponibilidade dos dados que me foram muito úteis.

Dedico também o trabalho a Dra. Suraya Baptista, ao Mestre Ricardo Gomes, Dr. Luís Carlos Prata, em particular a minha família e amigos pela força e pelo carinho.

Resumo

A zona turística rural no norte de Portugal é uma das áreas escolhidas por alguns turistas que visitam o nosso País. A auto-imagem desses turistas pode aferir-se mediante atributos capazes de caracterizar também o destino visitado.

Neste trabalho aplica-se a metodologia de estimação de modelos de mistura finita para constituição e caracterização dos segmentos de turistas com base em atributos de auto-imagem.

Na modelação atende-se ao tipo de escala – Semântica Diferencial – de dois modos distintos: 1 – as variáveis base são modeladas como métricas e usada uma mistura de distribuições normais; 2 – as variáveis base são modeladas como ordinais e usada uma mistura de multinomiais.

Aplica-se a metodologia de validação cruzada com o auxílio dos índices de concordância para comparar a estabilidade das soluções alternativas de agrupamento, com base na modelação das variáveis como métricas ou ordinais.

No trabalho desenvolvido o agrupamento com base nas variáveis ordinais tem uma maior estabilidade. Este agrupamento é seleccionado e caracterizado.

Palavras-chave: Agrupamento, Modelo de Mistura Finita, Validação Cruzada, Índice de Concordância.

Abstract

The touristic rural zone in the north of Portugal is one of the areas chosen by some tourists who visit our Country. The auto-image of these tourists can be checked by means of attributes able to also characterize the visited destination.

In this paper, the favorite methodology of finite mixture models for constitution and characterization of the segments of tourists on basis of attributes of auto-image is applied.

In the modulation, the type of scale is reached – Semantic Differential – by two different ways: 1 – the base variables are modulated as metrics and a mixture of normal distributions is used; 2 – the base variables are modulated as ordinals and a multinomial mixture is used.

The methodology of cross-validation is applied by using the rand indexes in order to compare the stability of the alternative solutions for Grouping, on basis on the modulation of the variables such as metrics or ordinals.

In the developed work, the Grouping on basis of the ordinal variables has a bigger stability. This Grouping has been selected and characterized.

Key-Words: Grouping, Finite Mixture Model, Cross-validation, Rand index.

Índice

1. Introdução.....	1
1.1. Agrupamento	1
1.2. Soluções de Agrupamento e Propriedades Desejáveis	1
1.3. Objectivos	2
2. Metodologia	2
2.1. Agrupamento usando Modelo de Mistura Finita	2
2.1.1. Consideração de variáveis base métricas.....	3
2.1.2. Consideração de variáveis base ordinais	4
2.1.3. Estimação e Critérios de Informação.....	4
2.2. Validação Cruzada.....	5
2.2.1. Validação Cruzada com amostra ponderada	6
2.3. Índices de Concordância.....	7
3. Análise de Dados.....	8
3.1. Dados sobre Turismo rural entre 1998 até 1999.....	8
3.2. Agrupamento e Análise de Estabilidade.....	9
3.2.1. Variáveis Ordinais (agrupamento sobre amostra global)	9
3.2.2. Variáveis Métricas (agrupamento sobre amostra global).....	11
3.2.3. Resultados da Validação Cruzada	12
3.2.4. Índice de Concordância	13
3.3. Caracterização de Solução Seleccionada.....	13
3.3.1. Alguns Gráficos da Solução Seleccionada	17
4. Limitações	20
5. Conclusões e Perspectivas.....	20
6. Bibliografia.....	21
7. ANEXO	23
7.1. Outros Resultados - Variáveis Ordinais (agrupamento sobre amostra global)	25
7.2. Outros Resultados - Variáveis Métricas (agrupamento sobre amostra global)	27
7.3. Variáveis Ordinais (agrupamento sobre amostra de “treino”).....	29
7.4. Variáveis Ordinais (agrupamento sobre amostra de “teste”).....	33
7.5. Variáveis Métricas (agrupamento sobre amostra de “treino”).....	36
7.6. Variáveis Métricas (agrupamento sobre amostra de “teste”).....	39
7.7. Restantes Quadros da solução Seleccionada.....	42

7.8. Restantes Gráficos da Solução Seleccionada.....	46
--	----

Índice de Tabelas

Tabela 1 - Agrupamento em Validação Cruzada.....	5
Tabela 2 - Formação de amostras TREINO_TESTE e TESTE_TREINO.....	6
Tabela 3 - Agrupamento em Validação Cruzada com amostra Ponderada	7
Tabela 4 - Especificação de nível de medida das variáveis base.....	8
Tabela 5 - Medidas Descritivas para Variáveis Ordinais	9
Tabela 6 - Medidas Descritivas para Variáveis Ordinais (cont.).....	9
Tabela 7 - Análise de resultados das variáveis ordinais	10
Tabela 8 - Medidas Descritivas para Variáveis Métricas	11
Tabela 9 - Medidas Descritivas para Variáveis Métricas (cont.)	11
Tabela 10 - Análise de resultados das variáveis métricas	11
Tabela 11 - Tabela de Contigência para soluções sobre dados ordinais	12
Tabela 12 - Tabela de Contigência para soluções sobre dados métricas.....	12
Tabela 13 - Resultados dos Índices de Concordância	13
Tabela 14 - Resultado da Análise Agrupamento	14
Tabela 15 - Classificação entre agitado e calmo	17
Tabela 16 - Classificação entre organizado e desorganizado	18
Tabela 17 - Classificação entre austero e delicado.....	19
Tabela 18 – Variáveis	24
Tabela 19 - Resultado da Análise do 3ºGrupo	25
Tabela 20 - Resultado da Análise de cada Grupo.....	26
Tabela 21 - Resultado da Análise de 5ºGrupo.....	27
Tabela 22 - Resultado da Análise de cada Grupo.....	28
Tabela 23 - Análise de resultados das variáveis ordinais (agrupamento sobre amostra de “treino”).....	29
Tabela 24 - Resultado da Análise de 3ºGrupo.....	30
Tabela 25 - Resultado da Análise de cada Grupo.....	32
Tabela 26 - Análise de resultados das variáveis ordinais (agrupamento sobre amostra de “teste”).....	33

Tabela 27 - Resultado da Análise de 3ºGrupo.....	34
Tabela 28 - Resultado da Análise de cada Grupo.....	36
Tabela 29 - Análise de resultados das variáveis métricas (agrupamento sobre amostra de “treino”).....	36
Tabela 30 - Resultado da Análise de 5ºGrupo.....	38
Tabela 31 - Resultado da Análise de cada Grupo.....	38
Tabela 32 - Análise de resultados das variáveis métricas (agrupamento sobre amostra de “teste”).....	39
Tabela 33 - Resultado da Análise de 5ºGrupo.....	40
Tabela 34 - Resultado da Análise de cada Grupo.....	41
Tabela 35 - Classificação entre modesto e extravagante.....	42
Tabela 36 - Classificação entre formal e informal	42
Tabela 37 - Classificação entre moderno e tradicional.....	42
Tabela 38 - Classificação entre comum e único	43
Tabela 39 - Classificação entre agradável e desagradável	43
Tabela 40 - Classificação entre racional e emocional	43
Tabela 41 - Classificação entre activo e passivo	44
Tabela 42 - Classificação entre conservador e liberal	44
Tabela 43 - Classificação entre caloroso e frio.....	44
Tabela 44 - Classificação entre juvenil e maduro	45
Tabela 45 - Classificação entre artificial e natural	45
Tabela 46 - Classificação entre complexo e simples.....	45
Tabela 47 - Classificação entre colorido e cinzento	46

Índice de ilustrações

Ilustração 1 - Gráfico com indicador BIC e AIC segundo o estudo das variáveis ordinais (amostra global).....	10
Ilustração 2 - Gráfico com indicador BIC e AIC segundo o estudo das variáveis métricas (amostra global)	12
Ilustração 3 - Gráfico correspondente a variável agitado-calmos	17
Ilustração 4 - Gráfico correspondente a variável organizado-desorganizado.....	18
Ilustração 5 - Gráfico correspondente a variável austero-delicado	19
Ilustração 6 - Gráfico de médias das variáveis métricas	27
Ilustração 7 - Gráfico com indicador BIC segundo o estudo das variáveis ordinais (amostra de treino).....	29
Ilustração 8 - Gráfico com indicador AIC segundo o estudo das variáveis ordinais (amostra de treino).....	30
Ilustração 9 - Gráfico com indicador BIC segundo o estudo das variáveis ordinais (amostra de teste).....	33
Ilustração 10 - Gráfico com indicador AIC segundo o estudo das variáveis ordinais (amostra de teste).....	34
Ilustração 11 - Gráfico com indicador BIC segundo o estudo das variáveis métricas (amostra de treino).....	37
Ilustração 12 - Gráfico com indicador AIC segundo o estudo das variáveis métricas (amostra de treino).....	37
Ilustração 13 - Gráfico com indicador BIC segundo o estudo das variáveis métricas (amostra de teste).....	39
Ilustração 14 - Gráfico com indicador AIC segundo o estudo das variáveis métricas (amostra de teste).....	40
Ilustração 15 - Gráfico correspondente a variável modesto-extravagante	46
Ilustração 16 - Gráfico correspondente a variável formal-informal	47
Ilustração 17 - Gráfico correspondente a variável moderno-tradicional	47
Ilustração 18 - Gráfico correspondente a variável comum-único.....	48
Ilustração 19 - Gráfico correspondente a variável agradável-desagradável	48
Ilustração 20 - Gráfico correspondente a variável racional-emocional	49
Ilustração 21 - Gráfico correspondente a variável activo-passivo.....	49
Ilustração 22 - Gráfico correspondente a variável conservador-liberal.....	50

Ilustração 23 - Gráfico correspondente a variável caloroso-frio	50
Ilustração 24 - Gráfico correspondente a variável juvenil-maduro	51
Ilustração 25 - Gráfico correspondente a variável artificial-natural.....	51
Ilustração 26 - Gráfico correspondente a variável complexo-simples	52
Ilustração 27 - Gráfico correspondente a variável colorido-cinzeno.....	52

1. Introdução

1.1. Agrupamento

O agrupamento é um método de classificação não supervisionado em que se constituem grupos de entidades com base em variáveis que as caracterizam, de modo a que as entidades de um grupo sejam homogéneas entre si e heterogéneas das de grupos distintos. *“Given a collection of n objects individuals, animals, plants etc., each of which is described by a set of p characteristics or variables, derive a useful division into a number of classes. Both the number of classes and the properties of the classes are to be determined.”* (Everitt, B. et al., 2001, p.4)

No presente caso usam-se modelos de mistura finita para obtenção de soluções de agrupamento em que a modelação das variáveis varia considerando-se as variáveis base ora como métricas ora como ordinais.

1.2. Soluções de Agrupamento e Propriedades Desejáveis

Um bom agrupamento pode ser obtido considerando o nível de ajustamento entre a partição obtida (através da análise de agrupamento) e a partição genuína. Quando a partição genuína é desconhecida (classificação não supervisionado), deve-se optar pela identificação de uma boa partição, por meio de propriedades desejáveis, nomeadamente a compacidade-separação e a estabilidade.

“A compacidade que mede a coesão entre os objectos dentro do grupo e a separação das propriedades do agrupamento que mede o isolamento dos grupos, quando comparado com outros grupos, definem a qualidade do agrupamento.

A estabilidade é também conhecida como uma propriedade desejável da solução final do agrupamento. A solução de grupo (estável) deve ficar aproximadamente a mesma quando ocorrem pequenas alterações no procedimento de agrupamento: alternativa parametrização do algoritmo de agrupamento, a introdução de ruído nos dados, diferentes variáveis de agrupamento base, amostras distintas de dados, etc” (Cardoso, M. et al. 2009, p. 2)

1.3. Objectivos

Neste trabalho pretende-se comparar a estabilidade de soluções de agrupamento em que a modelação das variáveis base é feita de forma distinta:

- 1) Considerando que as variáveis são de natureza métrica;
- 2) Considerando que as variáveis são de natureza ordinal.

No âmbito da aplicação dos modelos de mistura finita, utiliza-se geralmente, a modelação com variáveis métricas (usa-se distribuição normal). Mas, neste trabalho, considera-se a adequabilidade de modelação com variáveis ordinais (usando distribuições multinomiais com parametrizações adequadas).

Para comparar a estabilidade das duas soluções irão analisar-se os resultados de validação cruzada de cada uma, nomeadamente os valores dos Índices de Rand e o Rand Ajustado, que associam partições baseadas em duas sub-amostras diferentes.

2. Metodologia

2.1. Agrupamento usando Modelo de Mistura Finita

“No modelo de segmentos latentes, a distribuição das variáveis base Y_q ($q=1,...,Q$) apresenta-se como uma mistura das distribuições, das mesmas variáveis intra-grupos.” (Cardoso, M. 2001, p.4) Frequentemente considera-se o pressuposto que as variáveis Y_q são independentes, seguindo o modelo:

$$f(\underline{y}|\Theta) = \sum_{s=1}^S \lambda_s \prod_{q=1}^Q f_q(y_q | \Theta_{qs}), \quad (1)$$

em que

- λ_s é o parâmetro representando o peso do grupo s (elemento do vector $\underline{\lambda}$);
- f_q representa a f(d)p de um atributo Y_q das entidades a agrupar;

- $\underline{\Theta}_{qs}$ representa o vector de parâmetros modelando a característica Y_q dos indivíduos, intra-grupo s .

“A estimação de um modelo de segmentos latentes é, habitualmente efectuada, via maximização da função de verosimilhança associada a observações independentes de I entidades.” (Cardoso, M. 2001, p.5) Esta função representa-se por:

$$\prod_{i=1}^I \sum_{s=1}^S \lambda_s \prod_{q=1}^Q f(y_{iq} | \underline{\theta}_{qs}), \quad (2)$$

em que

- Y_{iq} é o valor da característica Y_q para a entidade i .
- $\underline{\theta}$ representa o vector de parâmetros modelando a característica Y_{iq} .

2.1.1. Consideração de variáveis base métricas

Nesta perspectiva de agrupamento usando modelo de mistura finita é comum modelar variáveis métricas, usando distribuição normal univariada, ou seja:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \frac{(y-\mu)^2}{\sigma^2} \right\}, \quad (3)$$

em que o μ e σ^2 são média e variância, respectivamente. Sendo a equação mais completa com o uso da independência condicional, i.e. independência de intra-grupos,

$$\sum_{s=1}^S \lambda_s \prod_{q=1}^Q \frac{1}{\sqrt{2\pi\sigma_{qs}^2}} \exp \left\{ -\frac{1}{2} \frac{(y_q - \mu_{qs})^2}{\sigma_{qs}^2} \right\}, \quad (4)$$

2.1.2. Consideração de variáveis base ordinais

Nesta perspectiva de agrupamento usando modelo de mistura finita é comum modelar variáveis ordinais, usando distribuição multinomial com reparametrizações, ou seja:

$$f(\underline{y} | \underline{\pi}) = \prod_{l=1}^L (\pi_l)^{y_l} \quad , \quad (5)$$

sendo $\pi_l = P(l)$, probabilidade da categoria l da resposta ($l=1, \dots, L$). De acordo com o modelo de mistura e considerando independência condicional tem-se, ainda,

$$\sum_{s=1}^S \lambda_s \prod_{q=1}^Q \prod_{l=1}^L (\pi_{ql}^s)^{y_{ql}} \quad , \quad (6)$$

E, de modo a incorporar a natureza ordinal das respostas, utiliza-se a proposta de modelo ordinal de categorias adjacentes (Vermunt, J.K. and Magidson, J. 2005). Nesta proposta tem-se em conta o logaritmo do *odd* de categorias adjacentes – $\log\left(\frac{\pi_{l+1}}{\pi_l}\right)$ – (A. Agresti, 2002)), como soma de uma constante associada ao grupo e de outra dependendo da variável e categoria considerada.

2.1.3. Estimação e Critérios de Informação

As estimativas dos parâmetros obtêm-se usando o algoritmo Latent Gold que implementa uma variante do algoritmo EM – Expectation Maximization (Vermunt, J.K. and Magidson, J. 2005), na tentativa de maximizar a função de probabilidade *a posteriori* (estimativas MAP-*Maximum a posteriori*). Considera-se, assim, o objectivo de maximizar $\log L + \log p(\Theta)$ em que L indica a função de verosimilhança associada ao referido modelo de mistura e $p(\Theta)$ a função de probabilidade *a priori* dos parâmetros.

O número de grupos resultante de um processo de agrupamento pode ser indicado *a priori*, mediante critérios indicados por especialistas do domínio de aplicação. Em

alternativa, procura-se que a estrutura de segmentos se ajuste automaticamente aos dados.

Dois critérios que podem determinar o número de grupos são as seguintes (em que d representa o número de parâmetros livres a estimar):

1. Critério AIC-*Akaike Information Criterion*, (Akaike, 1974);
sua expressão matemática: $AIC = -2\ln L + 2 \cdot d$ (7)

2. Critério BIC-*Bayesian Information Criterion*, (Schwartz, 1978).
sua expressão matemática: $BIC = -2\ln L + \ln(I) \cdot d$ (8)

2.2. Validação Cruzada

A validação cruzada é um método para avaliar e comparar o desempenho de algoritmos de aprendizagem através da partição de dados em subamostras e teste de desempenho nas mesmas subamostras.

Este processo é utilizado para avaliar modelos, principalmente preditivos, baseando-se na divisão de uma amostra de dados em que numa parte da amostra o modelo é estimado e na outra é avaliado.

Neste trabalho, a abordagem da validação cruzada é utilizada no intuito de avaliar a estabilidade de uma solução de agrupamento, com base na comparação de duas estruturas de agrupamento obtidas sobre diferentes amostras.

Eis um quadro exemplificativo sobre o agrupamento em validação cruzada (Martins, M. e Cardoso, M. 2008, p.3):

Etapa	Acção	Saída
1	Divisão da amostra Treino-Teste	Amostras de Treino e Teste
2	Análise de agrupamento na amostra de Treino	Grupos sobre Amostra de Treino
3	Classificação na amostra de Treino e uso de classificador na amostra de Teste	Classes sobre Amostra de Teste
4	Análise de agrupamento na amostra de Teste	Grupos sobre Amostra de Teste
5	Cálculo de índices de concordância entre Grupos e Classes obtidos sobre amostra de teste.	Valor de referência para avaliação de estabilidade

Tabela 1 - Agrupamento em Validação Cruzada

Segundo a Tabela 1 o procedimento de validação cruzada descreve-se em 5 passos, ou etapas.

Na 1ª etapa será efectua-se a divisão da amostra principal em amostras de treino e teste. Na 2ª e 4ª etapa utiliza-se o algoritmo para o agrupamento em ambas amostras. Na 3ª etapa usa-se um classificador na amostra de teste para proceder a uma classificação supervisionada pelos segmentos construídos na amostra de treino. Na 5ª e última etapa usa-se um índice de concordância (índice de Rand, por exemplo) para avaliar os agrupamentos (ver capítulo 2.3).

2.2.1. Validação Cruzada com amostra ponderada

No procedimento da validação cruzada com o uso da amostra ponderada efectua-se uma atribuição de pesos. Neste caso prescinde-se do uso de um classificador (Cardoso, M. et al. 2009). Assim, por exemplo, a primeira partição π^Q refere-se a uma amostra com atribuição do valor 10^{-10} de peso para o Teste (50%) e do valor 1 para o Treino. E a outra partição π^K obtem-se sobre a amostra com atribuição do valor 10^{-10} de peso para o Treino (50%) e do valor 1 para o Teste. O acordo entre as duas partições é, seguidamente, avaliado com um índice de concordância.

Pesos	Frequência	Percentagem
1,00E-10	n/2	50.0
1	n/2	50.0
Total	n	100.0

Tabela 2 - Formação de amostras TREINO_TESTE e TESTE_TREINO

Para validação cruzada com amostra ponderada não existirá nenhum classificador e portanto serão quatro passos para se obter o índice de concordância, conforme a Tabela 3.

Etapa	Acção	Saída
1	Divisão da amostra Treino-Teste e atribuição de pesos de 1 e 1^{-10}	Treino e Teste Ponderados
2	Agrupamento sobre a amostra de Treino Ponderado	Grupos sobre o Treino e Classes sobre o Teste
3	Agrupamento sobre a amostra de Teste Ponderado	Grupos sobre o Teste e Classes sobre o Treino
4	Índice de concordância entre Grupos e Classes de Treino e Teste.	Valor de referência para avaliação de estabilidade

Tabela 3 - Agrupamento em Validação Cruzada com amostra Ponderada

2.3. Índices de Concordância

Os índices de concordância são usualmente utilizados para avaliar se a solução de agrupamento é estável, medindo a associação entre duas partições (classes e grupos), i.e. uma partição é obtida usando a amostra de Treino ponderado para aprendizagem e é designada por classes; A outra partição obtida usando a amostra de Teste ponderada para aprendizagem e é designada por grupos.

“Um exemplo é o Índice de Rand que quantifica a proporção dos pares classificados de uma forma concordante nas duas partições:” (M. Cardoso, 2007, p.2). Segundo (Everitt, B. et al., 2001) sugerem a seguinte expressão do índice de Rand quando o número de classes é o mesmo nas 2 partições (K=Q):

$$Rand = \frac{\binom{n}{2} + \sum_{k=1}^K \sum_{q=1}^Q n_{kq}^2 - \left(\frac{1}{2} \sum_{k=1}^K n_k^2 + \frac{1}{2} \sum_{q=1}^Q n_q^2 \right)}{\binom{n}{2}}. \quad (\text{Rand, 1971}) \quad (9)$$

Em que $\binom{n}{2}$ indica o total de número de pares de observações.

“Hurbert e Arabie estudaram o Índice de Rand com distribuição sob a hipótese de concordância aleatória, baseando-se no modelo hipergeométrico, e sugeriram, um Índice de Rand Ajustado.” (M. Cardoso et al. 2009, p. 2) Vejamos então como é feita a correcção da estatística de Rand:

$$adj - Rand = \frac{\sum_{k=1}^K \sum_{q=1}^Q \binom{n_{kq}}{2} - \sum_{k=1}^K \binom{n_{k.}}{2} \sum_{q=1}^Q \binom{n_{.q}}{2} / \binom{n}{2}}{\frac{1}{2} \left[\sum_{k=1}^K \binom{n_{k.}}{2} + \sum_{q=1}^Q \binom{n_{.q}}{2} \right] - \sum_{k=1}^K \binom{n_{k.}}{2} \sum_{q=1}^Q \binom{n_{.q}}{2} / \binom{n}{2}}. \quad (\text{Hubert \& Arabie, 1985})$$

(10)

3. Análise de Dados

3.1. Dados sobre Turismo rural entre 1998 até 1999

No presente trabalho efectuou-se uma análise sobre dados recolhidos no âmbito de um projecto de investigação sobre o turismo em áreas rurais no Norte de Portugal. A sua recolha foi realizada entre Agosto de 1998 e Julho de 1999, maioritariamente via entrevista directa, procurando representar o mercado nacional e estrangeiro de turistas nas regiões do Minho, Douro e Trás-os-Montes. A amostra considerada integra 2280 respondentes e 16 variáveis estudadas (Kastenholz, 2002)..

O presente trabalho proposto incide numa pergunta do questionário que procura caracterizar a auto-imagem dos turistas usando atributos que se podem aplicar à região.

“Como caracterizaria, aproximadamente, a sua própria pessoa relativamente aos seguintes aspectos?”

Aagitado						Calmo
Organizado						Desorganizado
Austero						Delicado
Modesto						Extravagante
Formal						Informal
Moderno						Tradicional
Comum						Único
Agradável						Desagradável
Racional						Emocional
Activo						Passivo
Conservador						Liberal
Caloroso						Frio/Distante
Juvenil						Maduro
Artificial						Natural
Complexo						Simple
Colorido/Vivo						Cinzento

Tabela 4 - Especificação de nível de medida das variáveis base

A segmentação proposta será, pois realizada sobre o conjunto de variáveis apresentado na Tabela 4 uma estimação de um modelo de mistura finita usando o algoritmo implementado no programa informático Latent Gold. A segmentação foi realizada por etapas, note-se que a análise será realizada sobre 1725 observações que correspondem aos indivíduos que deram respostas completas às questões referidas.

3.2. Agrupamento e Análise de Estabilidade

3.2.1. Variáveis Ordinais (agrupamento sobre amostra global)

De acordo com a natureza das variáveis estudadas foi efectuada uma análise descritiva sobre as variáveis ordinais com intuito de saber os valores dados pelas suas medidas.

	Agitado – Calmo	Organizado – Desorganizado	Austero – Delicado	Modesto – Extravagante	Formal – Informal	Moderno – Tradicional	Comum – Único	Agradável – Desagradável
Mediana	4	2	3	2	3	3	3	2
Mínimo	1	1	1	1	1	1	1	1
Máximo	5	5	5	5	5	5	5	5

Tabela 5 - Medidas Descritivas para Variáveis Ordinais

	Racional – Emocional	Activo – Passivo	Conservador – Liberal	Caloroso – Frio	Juvenil – Maduro	Artificial – Natural	Complexo – Simples	Colorido – Cinzento
Mediana	3	2	3	2	3	4	3	2
Mínimo	1	1	1	1	1	1	1	1
Máximo	5	5	5	5	5	5	5	5

Tabela 6 - Medidas Descritivas para Variáveis Ordinais (cont.)

Ensaando o processo de segmentação sobre as variáveis referidas, modeladas como ordinais obtém-se os resultados na Tabela 7 (BIC e AIC referem-se aos critérios de informação – ver cap. 2.1.3 – e npar ao número de parâmetros livres a estimar):

Número de Grupos	BIC(LL)	AIC(LL)	Npar
Grupo1	77812,12	77463,13	64
Grupo2	76448,67	76006,98	81
Grupo3	75655,97	75121,58	98
Grupo4	75456,96	74829,87	115
Grupo5	75320,93	74601,14	132
Grupo6	75242,28	74429,78	149

Tabela 7 - Análise de resultados das variáveis ordinais

Na Ilustração 1 apresentam-se os resultados:

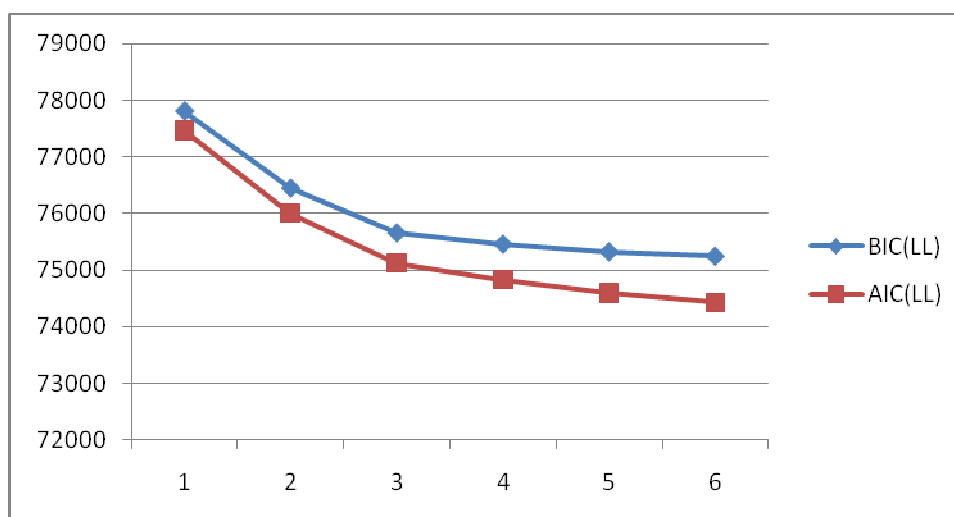


Ilustração 1 - Gráfico com indicador BIC e AIC segundo o estudo das variáveis ordinais (amostra global)

Conforme os indicadores BIC e AIC apresenta-nos um “cotovelo” que permite apontar como adequada solução de três segmentos.

3.2.2. Variáveis Métricas (agrupamento sobre amostra global)

De acordo com a natureza das variáveis estudadas foi efectuada uma análise descritiva sobre as variáveis métricas com intuito de saber os valores dados pelas suas medidas.

	Agitado – Calmo	Organizado – Desorganizado	Austero – Delicado	Modesto – Extravagante	Formal – Informal	Moderno – Tradicional	Comum – Único	Agradável – Desagradável
Média	3,48	2,27	3,42	2,40	3,36	2,75	3,07	1,90
Desvio Padrão	1,219	1,153	,976	,994	1,189	1,118	1,211	,807

Tabela 8 - Medidas Descritivas para Variáveis Métricas

	Racional – Emocional	Activo – Passivo	Conservador – Liberal	Caloroso – Frio	Juvenil – Maduro	Artificial – Natural	Complexo – Simples	Colorido – Cinzento
Média	3,07	2,10	3,30	2,04	3,05	4,18	3,46	2,09
Desvio Padrão	1,241	1,025	1,226	,932	1,224	,816	1,190	,925

Tabela 9 - Medidas Descritivas para Variáveis Métricas (cont.)

Ensaando o processo de segmentação sobre as variáveis referidas, modeladas como ordinais obtém-se os resultados na Tabela 10 (BIC e AIC referem-se aos critérios de informação – ver cap. 2.1.3 – e npar ao número de parâmetros livres a estimar) :

Número de Grupos	BIC(LL)	AIC(LL)	Npar
Grupo1	81602,46	81427,96	32
Grupo2	76098,38	75743,94	65
Grupo3	74629,61	74095,22	98
Grupo4	69312,01	68597,67	131
Grupo5	65981,15	65086,86	164
Grupo6	65182,47	64108,23	197

Tabela 10 - Análise de resultados das variáveis métricas

Na Ilustração 2 apresentam-se os resultados:

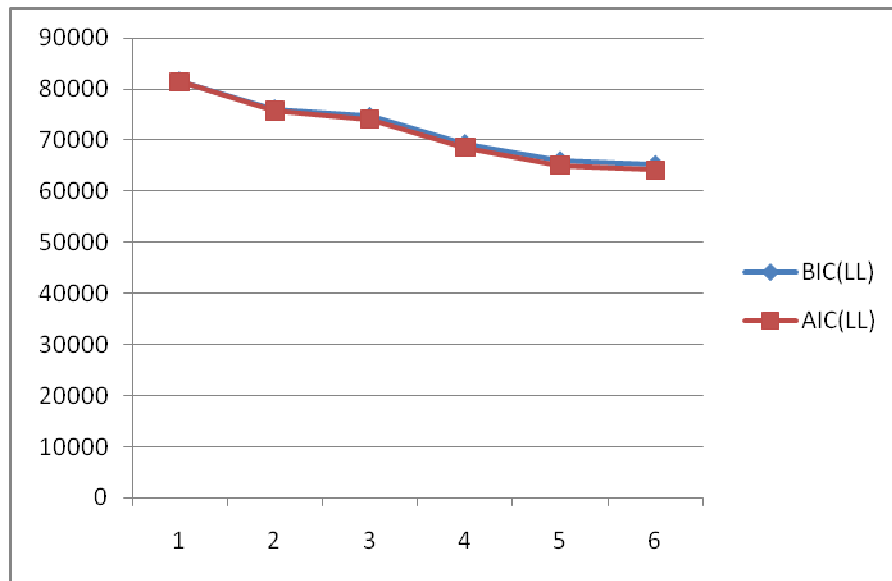


Ilustração 2 - Gráfico com indicador BIC e AIC segundo o estudo das variáveis métricas (amostra global)

Conforme os indicadores BIC e AIC apresenta-nos um “cotovelo” que nos inclina para uma difícil solução de cinco segmentos, uma vez que o referido “cotovelo” não se visualiza facilmente.

3.2.3. Resultados da Validação Cruzada

As Tabela 11 e Tabela 12 são tabelas contigência entre classes e grupos obtidos sobre amostras de teste e treino ponderadas, a partir das quais se irá obter o índice de Rand. São as seguintes:

	Classe 1	Classe 2	Classe 3	Total
Grupo 1	1075	95	31	1201
Grupo 2	4	320	8	332
Grupo 3	4	8	180	192
Total	1083	423	219	1725

Tabela 11 - Tabela de Contigência para soluções sobre dados ordinais

	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Total
Grupo 1	237	0	67	0	5	309
Grupo 2	75	12	235	69	97	488
Grupo 3	43	1	192	21	37	294
Grupo 4	54	0	94	161	39	348
Grupo 5	49	2	74	8	153	286
Total	458	15	662	259	331	1725

Tabela 12 - Tabela de Contigência para soluções sobre dados métricas

3.2.4. Índice de Concordância

Tipo de Variáveis	Número de grupos	Índice de Rand	Índice de Rand Ajustado
Ordinais	3	87%	74%
Métricas	5	68%	13%

Tabela 13 - Resultados dos Índices de Concordância

Segundo os resultados dos índices de concordância, o resultado referente a solução ordinal é o melhor em relação a solução métrica. Quanto maior for o valor de índice de concordância (índice de Rand e Rand Ajustado) melhor será a avaliação a estabilidade, ou seja, há consistência mesmo efectuando ligeiras modificações no processo de agrupamento, considerando as amostras de treino/teste ponderadas.

Note-se que o índice de Rand ajustado tem valores consideravelmente inferiores ao de Rand que neste caso resultam, não só de usar fórmula específica de Rand para $K=Q$ como da correcção própria do Rand ajustado que subtrai a medida de concordância por acaso.

3.3. Caracterização de Solução Seleccionada

Na Tabela 14 a solução de agrupamento seleccionada é característica com base nas probabilidades associadas com a cada categoria de cada variável base, designação usada na base de dados correspondente a (Kastenholtz, 2002), tendo como referência a partição probabilística resultante da estimação do modelo de mistura.

	Grupo1	Grupo2	Grupo3		Grupo1	Grupo2	Grupo3
Dim. Grupo	0,6491	0,2308	0,1202	Grupo Size	0,6491	0,2308	0,1202
Var. base				Indicators			
Calmo				Tradicional			
Muito agitado	0,0621	0,1413	0,0254	Muito moderno	0,0983	0,3182	0,04
Bastante agitado	0,1383	0,2181	0,0787	Bastante moderno	0,2569	0,3668	0,1608
Equilíbrio agitado-calmo	0,27	0,295	0,2134	Equilíbrio moderno-tradicional	0,4204	0,2647	0,4049
Bastante calmo	0,2918	0,2209	0,3202	Bastante tradicional	0,147	0,0408	0,2179
Muito calmo	0,2378	0,1247	0,3624	Muito tradicional	0,0774	0,0095	0,1765
Desorganizado				Único			
Muito organizado	0,2499	0,2711	0,5851	Muito comum	0,1129	0,0399	0,3654
Bastante organizado	0,3262	0,3335	0,3038	Bastante comum	0,1562	0,086	0,2407
Equilíbrio organizado-desorganizado	0,2405	0,2317	0,0891	Equilíbrio único-comum	0,4233	0,3633	0,3107

Um agrupamento de turistas – Modelação adequada?

Bastante desorganizado	0,1277	0,1159	0,0188
Muito desorganizado	0,0557	0,0477	0,0033
Delicado			
Muito austero	0,0416	0,0198	0,0029
Bastante austero	0,1256	0,0816	0,0236
Equilíbrio austero-delicado	0,4857	0,4314	0,2481
Bastante delicado	0,2444	0,2967	0,3388
Muito delicado	0,1027	0,1705	0,3866
Extravagante			
Muito modesto	0,155	0,0892	0,7473
Bastante modesto	0,3265	0,2637	0,2144
Equilíbrio modesto-extravagante	0,4172	0,4726	0,0373
Bastante extravagante	0,0798	0,1268	0,001
Muito extravagante	0,0214	0,0478	0
Informal			
Muito formal	0,0894	0,038	0,1679
Bastante formal	0,1215	0,0722	0,1698
Equilíbrio formal-informal	0,3344	0,2778	0,3473
Bastante informal	0,2727	0,3166	0,2105
Muito informal	0,182	0,2954	0,1045
Liberal			
Muito conservador	0,0987	0,0216	0,2165
Bastante conservador	0,15	0,0588	0,22
Equilíbrio conservador-liberal	0,3265	0,2289	0,3199
Bastante liberal	0,2655	0,3332	0,1739
Muito liberal	0,1592	0,3575	0,0697
Frio/Distante			
Muito caloroso	0,1236	0,6772	0,6999
Bastante caloroso	0,4343	0,2932	0,275
Equilíbrio caloroso-frio	0,3452	0,0287	0,0244
Bastante frio	0,0809	0,0008	0,0006
Muito frio	0,0161	0	0
Maduro			
Muito juvenil	0,1087	0,2607	0,1279
Bastante juvenil	0,1499	0,2243	0,1634
Equilíbrio juvenil-maduro	0,3211	0,2997	0,3244
Bastante maduro	0,286	0,1666	0,2677
Muito maduro	0,1343	0,0488	0,1165
Bastante único	0,1745	0,2334	0,061
Muito único	0,1331	0,2774	0,0221
Desagradável			
Muito agradável	0,1711	0,5438	0,8635
Bastante agradável	0,4844	0,3902	0,1318
Equilíbrio agradável-desagradável	0,3181	0,0649	0,0047
Bastante desagradável	0,0202	0,001	0
Muito desagradável	0,0062	0,0001	0
Emocional			
Muito racional	0,1454	0,0619	0,0603
Bastante racional	0,2439	0,1542	0,1518
Equilíbrio racional-emocional	0,3148	0,2955	0,294
Bastante emocional	0,1837	0,2561	0,2577
Muito emocional	0,1122	0,2322	0,2363
Passivo			
Muito activo	0,2231	0,5786	0,3807
Bastante activo	0,3773	0,3298	0,3968
Equilíbrio activo-passivo	0,2815	0,0829	0,1824
Bastante passivo	0,0717	0,0071	0,0286
Muito passivo	0,0465	0,0016	0,0114
Natural			
Muito artificial	0,0107	0,0001	0
Bastante artificial	0,0291	0,0009	0
Equilíbrio natural-artificial	0,231	0,033	0,0009
Bastante natural	0,5145	0,3339	0,0678
Muito natural	0,2147	0,6321	0,9312
Simples			
Muito complexo	0,067	0,0601	0
Bastante complexo	0,1913	0,1799	0,0012
Equilíbrio complexo-simples	0,3162	0,3119	0,0184
Bastante simples	0,273	0,2825	0,1535
Muito simples	0,1525	0,1655	0,8269
Cinzento			
Muito vivo	0,1526	0,6955	0,4596
Bastante vivo	0,3718	0,26	0,3871
Equilíbrio vivo-cinzento	0,405	0,0435	0,1458
Bastante cinzento	0,0547	0,0009	0,0068
Muito cinzento	0,0159	0	0,0007

Tabela 14 - Resultado da Análise Agrupamento

Nesta tabela mostra-nos uma visão mais detalhada dos três grupos. Tendo por base a informação da mesma, conclui-se que:

- O Grupo 1, correspondendo a 64,9% da amostra e em termos da variável “calmo” apresenta uma caracterização de bastante calmo (29%); Em termos da variável “desorganizado” apresenta uma caracterização de bastante organizado (33%); Em termos da variável “delicado” apresenta uma caracterização de equilíbrio entre austero e delicado (49%); Em termos da variável “extravagante” apresenta uma caracterização de equilíbrio entre modesto e extravagante (42%); Em termos da variável “informal” apresenta uma caracterização de equilíbrio entre formal e informal (33%); Em termos da variável “tradicional” apresenta uma caracterização de equilíbrio entre moderna e tradicional (42%); Em termos da variável “único” apresenta uma caracterização de equilíbrio entre único e comum (42%); Em termos da variável “desagradável” apresenta uma caracterização de bastante agradável (48%); Em termos da variável “emocional” apresenta uma caracterização de equilíbrio entre racional e emocional (31%); Em termos da variável “passivo” apresenta uma caracterização de bastante activo (38%); Em termos da variável “liberal” apresenta uma caracterização de equilíbrio entre conservador e liberal (33%); Em termos da variável “frio/distante” apresenta uma caracterização de bastante calor (43%); Em termos da variável “maduro” apresenta uma caracterização de equilíbrio entre juvenil e maduro (32%); Em termos da variável “natural” apresenta uma caracterização de bastante natural (51%); Em termos da variável “simples” apresenta uma caracterização de equilíbrio entre complexo e simples (32%); Em termos da variável “cinzento” apresenta uma caracterização de bastante vivo (37%).

- O Grupo 2, correspondendo a 23,1% da amostra e em termos da variável “calmo” apresenta uma caracterização de equilíbrio entre agitado e calmo (30%); Em termos da variável “desorganizado” apresenta uma caracterização de bastante organizado (33%); Em termos da variável “delicado” apresenta uma caracterização de equilíbrio entre austero e delicado (43%); Em termos da variável “extravagante” apresenta uma caracterização de equilíbrio entre modesto e extravagante (47%); Em termos da variável “informal” apresenta uma caracterização de bastante informal (32%); Em termos da variável “tradicional” apresenta uma caracterização de bastante moderno (37%); Em termos da variável “único” apresenta uma caracterização de equilíbrio entre único e comum (36%); Em termos da variável “desagradável” apresenta uma caracterização de muito agradável (54%); Em termos da variável “emocional” apresenta uma caracterização de equilíbrio entre emocional e emocional (30%); Em termos da variável

“passivo” apresenta uma caracterização de muito activo (58%); Em termos da variável “liberal” apresenta uma caracterização de muito liberal (36%); Em termos da variável “frio/distante” apresenta uma caracterização de muito caloroso (68%); Em termos da variável “maduro” apresenta uma caracterização de equilíbrio entre juvenil e maduro (30%); Em termos da variável “natural” apresenta uma caracterização de muito natural (63%); Em termos da variável “simples” apresenta uma caracterização de equilíbrio entre complexo e simples (31%); Em termos da variável “cinzento” apresenta uma caracterização de muito vivo (70%).

- O Grupo 3, correspondendo a 12% da amostra e em termos da variável “calmo” apresenta uma caracterização de muito calmo (36%); Em termos da variável “desorganizado” apresenta uma caracterização de muito organizado (59%); Em termos da variável “delicado” apresenta uma caracterização de muito delicado (39%); Em termos da variável “extravagante” apresenta uma caracterização de muito modesto (75%); Em termos da variável “informal” apresenta uma caracterização de equilíbrio entre formal e informal (35%); Em termos da variável “tradicional” apresenta uma caracterização de equilíbrio entre modesto e tradicional (40%); Em termos da variável “único” apresenta uma caracterização de muito comum (37%); Em termos da variável “desagradável” apresenta uma caracterização de muito agradável (86%); Em termos da variável “emocional” apresenta uma caracterização de equilíbrio entre racional e emocional (29%); Em termos da variável “passivo” apresenta uma caracterização de bastante activo (40%); Em termos da variável “liberal” apresenta uma caracterização de equilíbrio entre conservador e liberal (32%); Em termos da variável “frio/distante” apresenta uma caracterização de muito caloroso (70%); Em termos da variável “maduro” apresenta uma caracterização de equilíbrio entre juvenil e maduro (32%); Em termos da variável “natural” apresenta uma caracterização de muito natural (93%); Em termos da variável “simples” apresenta uma caracterização de muito simples (83%); Em termos da variável “cinzento” apresenta uma caracterização de muito vivo (46%).

3.3.1. Alguns Gráficos da Solução Seleccionada

Nos gráficos seguintes, ilustra-se, a título de exemplo, o comportamento de algumas variáveis base de agrupamento considerando os grupos que resultam de afectação modal i. e. cada indivíduo é afectado ao grupo ao qual associa maior probabilidade de pertença. Assim, através da auto-imagem dos inquiridos é possível visualizar a predominância das suas escolhas e visualizar também a categoria predominante num grupo.

Agitado-Calmo*Grupo Modal. Validação Cruzada: % dentro do Grupo modal

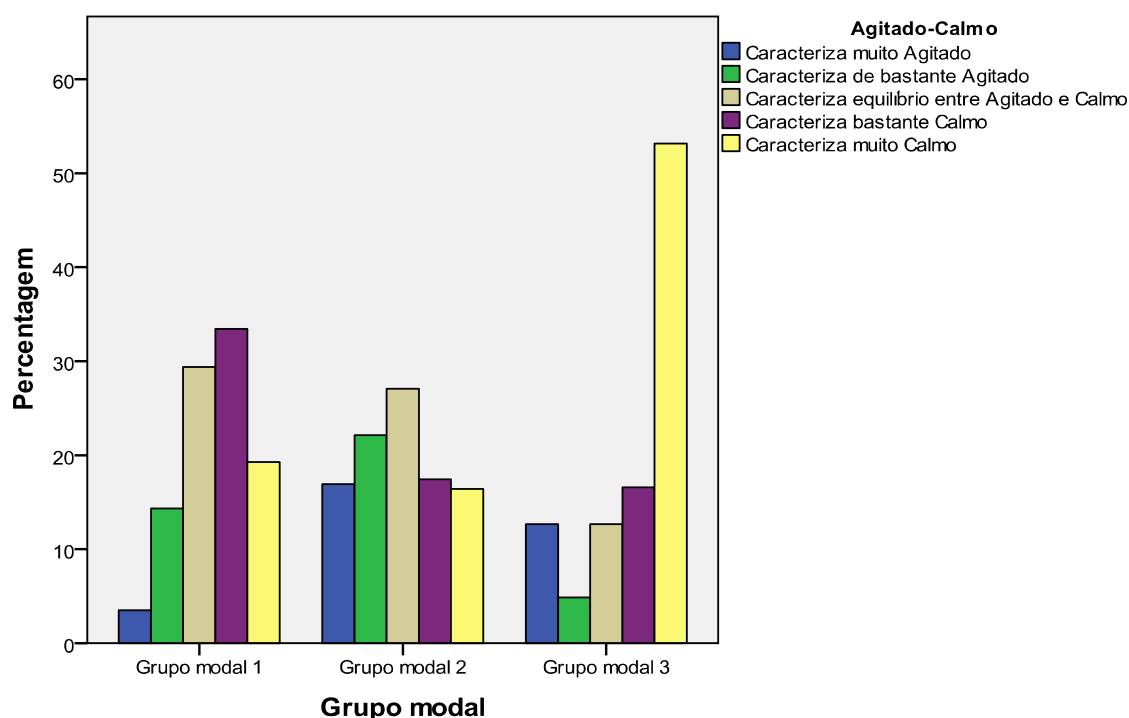


Ilustração 3 - Gráfico correspondente a variável agitado-calmo

Agitado-Calmo*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
agitado-calmo	muito tenso	3,5%	16,9%	12,7%	7,6%
	bastante agitado	14,3%	22,1%	4,9%	15,0%
	equilíbrio agitado-calmo	29,4%	27,1%	12,7%	26,9%
	bastante calmo	33,5%	17,4%	16,6%	27,9%
	muito calmo	19,3%	16,4%	53,2%	22,7%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 15 - Classificação entre agitado e calmo

Organizado-Desorganizado*Grupo Modal. Validação Cruzada: % dentro do Grupo modal

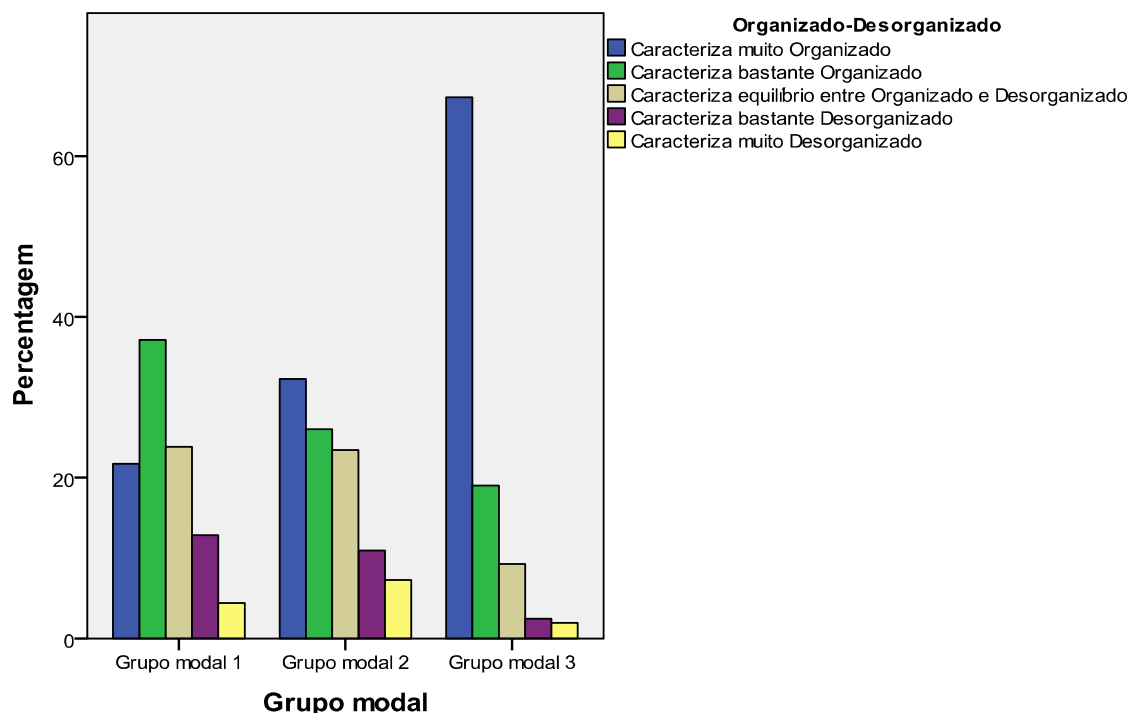


Ilustração 4 - Gráfico correspondente a variável organizado-desorganizado

Organizado-Desorganizado*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
organizado-desorganizado	muito organizado	21,7%	32,3%	67,3%	29,5%
	bastante organizado	37,1%	26,0%	19,0%	32,5%
	equilíbrio organizado-desorganizado	23,9%	23,4%	9,3%	22,0%
	bastante desorganizado	12,9%	10,9%	2,4%	11,2%
	muito desorganizado	4,4%	7,3%	2,0%	4,8%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 16 - Classificação entre organizado e desorganizado

Austero-Delicado*Grupo Modal. Validação Cruzada: % dentro do Grupo modal

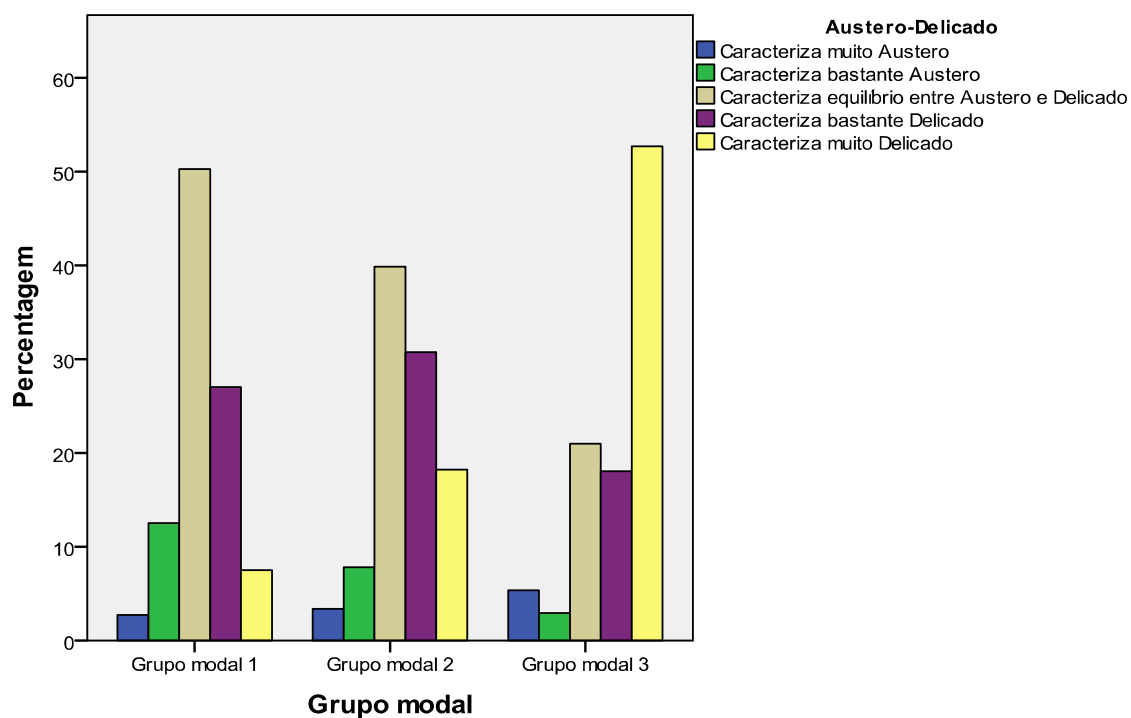


Ilustração 5 - Gráfico correspondente a variável austero-delicado

Austero-Delicado*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
austero-delicado	muito austero	2,7%	3,4%	5,4%	3,2%
	bastante austero	12,5%	7,8%	2,9%	10,3%
	equilíbrio austero-delicado	50,3%	39,8%	21,0%	44,5%
	bastante delicado	27,0%	30,7%	18,0%	26,8%
	muito delicado	7,5%	18,2%	52,7%	15,2%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 17 - Classificação entre austero e delicado

4. Limitações

As limitações encontradas no trabalho foram algumas que a seguir serão enumeradas:

1. Na análise do BIC e do AIC para solução métrica há dificuldade em decidir o número de grupos. A opção tomada foi analisar a solução com 5 grupos.
2. Poderíamos ter considerado outros critérios de informação, para além do BIC e AIC.
3. Só considerámos uma validação cruzada resultante da constituição de duas subamostras ponderadas (uma de treino e outra de teste). Poderíamos ter considerado mais validações cruzadas, evidenciando que a concordância entre as duas partições ocorre, em grande parte, por acaso.

5. Conclusões e Perspectivas

Está análise permitiu propor uma tipologia de agrupamento da amostra com base exclusivamente nas variáveis que se modelaram como ordinais, com a solução de três grupos. Acabou por se aceitar essa solução obtida, porque os resultados do índice de concordância são bons (Índice de Rand: 87% e Índice de Rand Ajustado: 74%) e bastante superiores aos obtidos com modelação de variáveis como métricas. Em relação ao índice de Rand Ajustado para a solução métrica os seus resultados são bastante reduzidos.

Havendo considerado a estabilidade como uma boa propriedade num resultado de agrupamento esta solução com 3 grupos é a mais estável e aquela que se opta por interpretar.

No futuro seria interessante analisar mais bases de dados de modo semelhante – comparando a modelação de variáveis ordinais vs métricas – de modo a consolidar a ideia de que são melhores os resultados obtidos com a modelação ordinal. Esta avaliação de desempenho deveria ser realizada como no caso presente, através de validação cruzada ponderada.

6. Bibliografia

Monografia:

Maroco, J. (2007). *Análise Estatística com Utilização do SPSS 3ª Edição*. Edições Silabo.

Cardoso, M. (2009). *Data Mining*. Textos de apoio a UC de Data Mining-Mestrado de Prospecção e Análise de Dados.

Vermunt, J.K. Magidson, J. (2005). *Technical Guide for Latent Gold 4.0: Basic and Advanced*. Belmont Massachusetts: Statistical Innovations Inc.

Agresti, A. (2002). *Categorical Data Analysis*. Hoboken, NJ: John Wiley and Sons.

Everitt, B. Landau, S. Leese, M. (2001). *Cluster Analysis Fourth Edition*. London: John Wiley and Sons.

Tese:

Kastenholz, E. (2002). *The Role and Marketing Implications of Destination Images on Tourist Behavior: The case of Northern Portugal*. Aveiro

Periódicos científicos:

Cardoso, Margarida G. M. S. (2007) *Clustering and Cross-Validation*. IASC 07 - Statistics for Data Mining, Learning and Knowledge Extraction, Proceedings. Ed. by C. Ferreira, C. Lauro, G. Saporta and M. Souto de Miranda. Aveiro, Portugal, p.32 (CD-ROM). ISBN 978-90-73592-26-1.

Cardoso, Margarida G. M. S.; Carvalho, André Ponce de Leon F. de; Faceli, Katti (2009). *Evaluation of clustering results: the trade-off bias-variability*. In *Classification as a Tool for Research*. Proceedings of the 11th IFCS Biennial Conference. Dresden, March 13-18, 2009 *Studies in Classification, Data Analysis, and Knowledge Organization* Springer, Berlin-Heidelberg-New York. Hermann Locarek-Junge, Claus Weihs (editors). P. 201-208.

Cardoso, Margarida G.M.S. (2001).1 Modelos de Segmentos Latentes: Aplicações em Marketing. Temas em Métodos em Métodos Quantitativos II. Ed. Por Manuel Alberto M. Ferreira, Rui Menezes e Margarida G. M. S. Cardoso. Ed. Sílabo. p. 206-230.

Martins, Cristina; Cardoso, Margarida G.M.S. (2009) Evaluation of clusters of credit cards holders. Revista de Ciências da Computação (Universidade Aberta), Vol. III, nº3, p.1-11.

Jesus, Neuza Brito; Cardoso, Margarida G.M.S. (2008) Análise de Agrupamento Incremental - Segmentação de Pontos de Retalho. Revista de Ciências da Computação (Universidade Aberta), Vol. II, nº2, p. 26-38.

Serra, P.; Cardoso, Margarida G. M. S.; Salgueiro, F. (2009) Satisfação laboral percebida: agrupamento adequado? Livro de Resumos. XVII Congresso Anual da Sociedade Portuguesa de Estatística.

W. M. Rand. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association, 66: 846-850, 1971.

L. Hubert and P. Arabie. Comparing partitions. Journal of Classification, 2: 193-218, 1985.

7. ANEXO

Variáveis Base de Segmentação

Variáveis		Único	Níveis	Maduro	Níveis
Calmo	Níveis	Muito comum	1	Muito juvenil	1
Muito agitado	1	Bastante comum	2	Bastante juvenil	2
Bastante agitado	2	Equilíbrio único-comum	3	Equilíbrio juvenil-maduro	3
Equilíbrio agitado-calmo	3	Bastante único	4	Bastante maduro	4
Bastante calmo	4	Muito único	5	Muito maduro	5
Muito calmo	5	Desagradável	Níveis	Natural	Níveis
Desorganizado	Níveis	Muito agradável	1	Muito artificial	1
Muito organizado	1	Bastante agradável	2	Bastante artificial	2
Bastante organizado	2	Equilíbrio agradável-desagradável	3	Equilíbrio natural-artificial	3
Equilíbrio organizado-desorganizado	3	Bastante desagradável	4	Bastante natural	4
Bastante desorganizado	4	Muito desagradável	5	Muito natural	5
Muito desorganizado	5	Emocional	Níveis	Simples	Níveis
Delicado	Níveis	Muito racional	1	Muito complexo	1
Muito austero	1	Bastante racional	2	Bastante complexo	2
Bastante austero	2	Equilíbrio racional-emocional	3	Equilíbrio complexo-simples	3
Equilíbrio austero-delicado	3	Bastante emocional	4	Bastante simples	4
Bastante delicado	4	Muito emocional	5	Muito simples	5
Muito delicado	5	Passivo	Níveis	Cinzento	Níveis
Extravagante	Níveis	Muito activo	1	Muito vivo	1
Muito modesto	1	Bastante activo	2	Bastante vivo	2
Bastante modesto	2	Equilíbrio activo-passivo	3	Equilíbrio vivo-cinzento	3
Equilíbrio modesto-extravagante	3	Bastante passivo	4	Bastante cinzento	4
Bastante extravagante	4	Muito passivo	5	Muito cinzento	5
Muito extravagante	5	Liberal	Níveis	-	-
Informal	Níveis	Muito conservador	1	-	-
Muito formal	1	Bastante conservador	2	-	-
Bastante formal	2	Equilíbrio conservador-liberal	3	-	-
Equilíbrio formal-informal	3	Bastante liberal	4	-	-
Bastante informal	4	Muito liberal	5	-	-
Muito informal	5	Frio/Distante	Níveis	-	-
Tradicional	Níveis	Muito caloroso	1	-	-
Muito moderno	1	Bastante caloroso	2	-	-
Bastante moderno	2	Equilíbrio caloroso-frio	3	-	-
Equilíbrio moderno-tradicional	3	Bastante frio	4	-	-
Bastante tradicional	4	Muito frio	5	-	-
Muito tradicional	5	-	-	-	-

Tabela 18 – Variáveis

7.1. Outros Resultados - Variáveis Ordinais (agrupamento sobre amostra global)

Solução com 3 grupos	
Número de casos	1725
Número de Parâmetros (Npar)	98
Variação de semente	10
Melhor sementação	422332
Log-verosimilhança (LL)	-37462,8
BIC (baseada em LL)	75655,97
AIC (baseada em LL)	75121,58

Tabela 19 - Resultado da Análise do 3ºGrupo

	Grupo1	Grupo2	Grupo3		Grupo1	Grupo2	Grupo3
Overall Probability	0,6491	0,2308	0,1202	-	-	-	-
Indicators				Indicators			
Calmo				Tradicional			
Muito agitado	0,31	0,4895	0,2006	Muito moderno	0,2776	0,5699	0,1525
Bastante agitado	0,6221	0,3364	0,0415	Bastante moderno	0,6999	0,2752	0,0248
Equilíbrio agitado-calm	0,707	0,2337	0,0593	Equilíbrio moderno-tradicional	0,7413	0,1592	0,0994
Bastante calmo	0,7774	0,1553	0,0674	Bastante tradicional	0,7876	0,078	0,1344
Muito calmo	0,554	0,1638	0,2822	Muito tradicional	0,4531	0,0565	0,4904
Desorganizado				Único			
Muito organizado	0,4783	0,2488	0,2729	Muito comum	0,3659	0,1538	0,4803
Bastante organizado	0,739	0,1947	0,0663	Bastante comum	0,7855	0,1142	0,1004
Equilíbrio organizado-desorganizado	0,703	0,2423	0,0547	Equilíbrio único-comum	0,7604	0,1746	0,065
Bastante desorganizado	0,7492	0,2191	0,0317	Bastante único	0,6828	0,2883	0,0288
Muito desorganizado	0,6087	0,3398	0,0515	Muito único	0,4227	0,4883	0,089
Delicado				Desagradável			
Muito austero	0,5134	0,2597	0,2269	Muito agradável	0,2943	0,3929	0,3128
Bastante austero	0,7889	0,1778	0,0332	Bastante agradável	0,7926	0,1823	0,0251
Equilíbrio austero-delicado	0,7391	0,2056	0,0553	Equilíbrio agradável-desagradável	0,9057	0,0802	0,0141
Bastante delicado	0,6527	0,2664	0,0809	Bastante desagradável	0,8635	0,1365	0
Muito delicado	0,314	0,2715	0,4145	Muito desagradável	0,8127	0,1873	0
Extravagante				Emocional			
Muito modesto	0,3983	0,156	0,4457	Muito racional	0,5669	0,2217	0,2114
Bastante modesto	0,776	0,1644	0,0596	Bastante racional	0,8882	0,0852	0,0267
Equilíbrio modesto-	0,7192	0,2599	0,0209	Equilíbrio racional-	0,6719	0,2377	0,0904

Um agrupamento de turistas – Modelação adequada?

extravagante				emocional			
Bastante extravagante	0,5874	0,4118	0,0007	Bastante emocional	0,6629	0,2675	0,0695
Muito extravagante	0,372	0,6192	0,0089	Muito emocional	0,319	0,3735	0,3075
Informal				Passivo			
Muito formal	0,3788	0,1785	0,4427	Muito activo	0,348	0,4333	0,2187
Bastante formal	0,7594	0,1319	0,1087	Bastante activo	0,7679	0,1781	0,054
Equilíbrio formal-informal	0,7481	0,1811	0,0709	Equilíbrio activo-passivo	0,8519	0,092	0,0561
Bastante informal	0,7294	0,2228	0,0478	Bastante passivo	0,8555	0,048	0,0965
Muito informal	0,4311	0,4029	0,166	Muito passivo	0,5789	0,0499	0,3712
Liberal				Natural			
Muito conservador	0,4336	0,1511	0,4153	Muito artificial	0,8625	0,1375	0
Bastante conservador	0,8245	0,0684	0,1071	Bastante artificial	0,9575	0,0374	0,0051
Equilíbrio conservador-liberal	0,725	0,1616	0,1133	Equilíbrio natural-artificial	0,9096	0,087	0,0034
Bastante liberal	0,7415	0,22	0,0385	Bastante natural	0,8416	0,1421	0,0164
Muito liberal	0,3834	0,5076	0,109	Muito natural	0,324	0,3924	0,2836
Frio/Distante				Simples			
Muito caloroso	0,2107	0,4985	0,2908	Muito complexo	0,5052	0,4838	0,011
Bastante caloroso	0,7951	0,1587	0,0462	Bastante complexo	0,7914	0,2029	0,0057
Equilíbrio caloroso-frio	0,9275	0,0423	0,0301	Equilíbrio complexo-simples	0,746	0,2405	0,0135
Bastante frio	0,9674	0,0055	0,0271	Bastante simples	0,7966	0,1643	0,039
Muito frio	0,9281	0,0008	0,0711	Muito simples	0,3072	0,2506	0,4422
Maduro				Cinzento			
Muito juvenil	0,292	0,5013	0,2067	Muito vivo	0,246	0,5213	0,2327
Bastante juvenil	0,6842	0,2497	0,0661	Bastante vivo	0,7899	0,1541	0,056
Equilíbrio juvenil-maduro	0,7237	0,1857	0,0906	Equilíbrio vivo-cinzento	0,8883	0,0416	0,07
Bastante maduro	0,7865	0,1397	0,0738	Bastante cinzento	0,8618	0,0257	0,1124
Muito maduro	0,5372	0,1852	0,2776	Muito cinzento	0,7174	0,0003	0,2823

Tabela 20 - Resultado da Análise de cada Grupo

7.2. Outros Resultados - Variáveis Métricas (agrupamento sobre amostra global)

Solução com 5 grupos	
Número de casos	1725
Número de Parâmetros (Npar)	164
Variação de sementação	10
Melhor sementação	543226
Log-verossimilhança (LL)	-32379,43
BIC (baseada em LL)	65981,15
AIC (baseada em LL)	65086,86

Tabela 21 - Resultado da Análise de 5ºGrupo

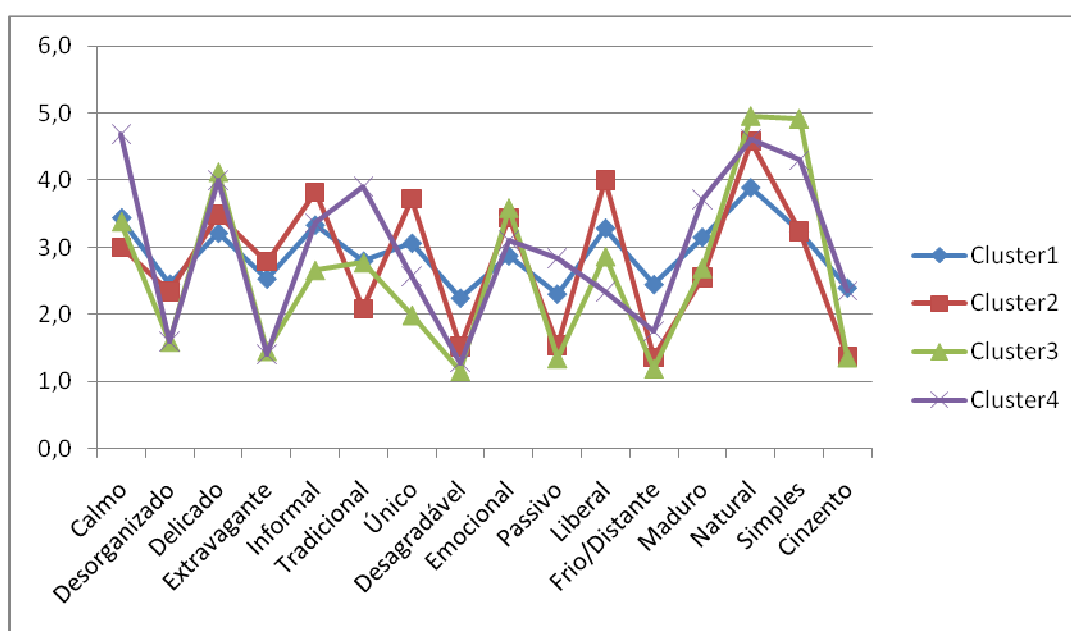


Ilustração 6 - Gráfico de médias das variáveis métricas

Um agrupamento de turistas – Modelação adequada?

	Grupo1	Grupo2	Grupo3	Grupo4	Grupo5
Cluster Size	0,419	0,2546	0,1576	0,1425	0,0263
Indicators					
Calmo					
Média	3,426	3,4514	3,3746	3,4384	3,6009
Desorganizado					
Média	2,4108	2,1939	2,3602	1,9911	2,5552
Delicado					
Média	3,3485	3,4213	3,2129	3,8259	3,0474
Extravagante					
Média	2,527	2,2869	2,5996	2,0609	2,5098
Informal					
Média	3,4246	3,4329	3,1881	3,443	3,0861
Tradicional					
Média	2,7483	2,8066	2,6473	2,6091	2,5986
Único					
Média	3,1065	3,0388	3,1506	2,9673	3,1565
Desagradável					
Média	2,1162	1,9356	2,1696	1	2,1529
Emocional					
Média	2,9876	3,1866	2,8827	3,4356	2,5776
Passivo					
Média	2,2476	1,9229	2,3125	1,6984	2,3113
Liberal					
Média	3,3734	3,3099	3,151	3,4512	3,2866
Frio/Distante					
Média	2,2503	2,0244	2,445	1	2,4218
Maduro					
Média	3,1272	3,0413	2,9516	2,7878	2,7825
Natural					
Média	4	5	3	5	1,7403
Simples					
Média	3,278	3,8085	2,9972	3,9639	2,7738
Cinzento					
Média	2,2586	2,1455	2,3389	1,1212	2,2408

Tabela 22 - Resultado da Análise de cada Grupo

7.3. Variáveis Ordinais (agrupamento sobre amostra de “treino”)

Foram colocadas as variáveis ordinais e usamos um filtro como peso que será a variável TREINO_TESTE.

Após a estimação do modelo houve o surgimento de 3 segmentos segundo o critério BIC e também para o critério AIC. Observe a Tabela 23 (BIC e AIC referem-se aos critérios de informação e npar ao número de parâmetros livres a estimar):

Número de Grupos	BIC(LL)	AIC(LL)	Npar
Grupo1	39284,25	38979,21	64
Grupo2	38703,44	38317,38	81
Grupo3	38381,43	37914,34	98
Grupo4	38264,88	37716,76	115
Grupo5	38239,96	37610,83	132
Grupo6	38235,1	37524,94	149

Tabela 23 - Análise de resultados das variáveis ordinais (agrupamento sobre amostra de “treino”)

Observe em seguida o gráfico segundo os critérios BIC e AIC :

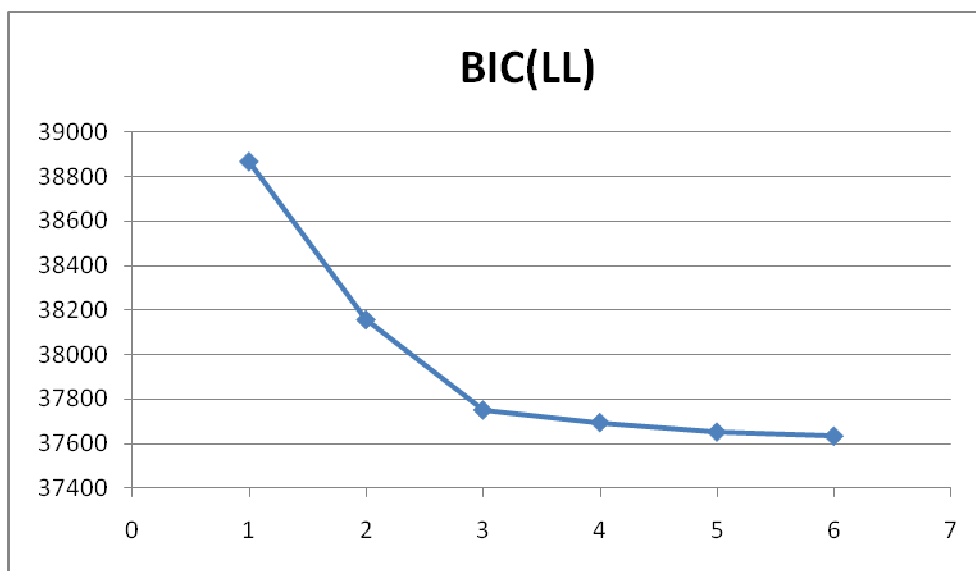


Ilustração 7 - Gráfico com indicador BIC segundo o estudo das variáveis ordinais (amostra de treino)

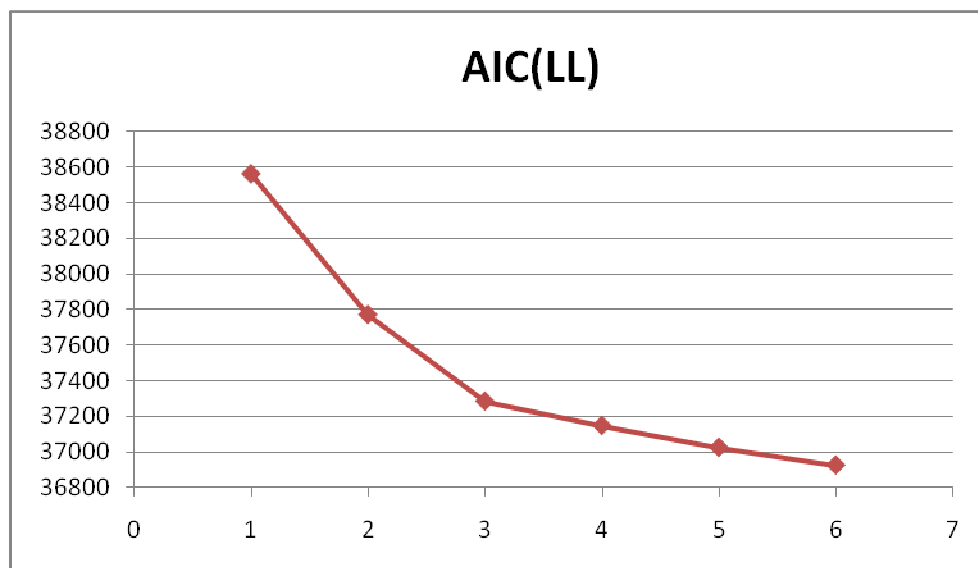


Ilustração 8 - Gráfico com indicador AIC segundo o estudo das variáveis ordinais (amostra de treino)

De acordo com o indicador BIC verifica-se um *cotovelo* para uma solução de 3 grupos o que justifica a consideração deste número, indicando uma mudança no gráfico correspondente a este indicador precisamente na solução com 3 grupos.

Ilustrando um resumo do Latent Gold em relação ao 3º grupo:

Solução com 3 grupos	
Número de casos	857
Número de Parâmetros (Npar)	98
Variação Sementação	348718
Melhor Sementação	669881
Log-verosimilhança (LL)	-18544,0713
BIC (based on LL)	37749,9795
AIC (based on LL)	37284,1426

Tabela 24 - Resultado da Análise de 3º Grupo

A caracterização dos grupos nesta solução será:

	Grupo1	Grupo2	Grupo3		Grupo1	Grupo2	Grupo3
Grupo Size	0,6796	0,1982	0,1222	Grupo Size	0,6796	0,1982	0,1222
Indicators				Indicators			
Calmo				Tradicional			
Muito agitado	0,0654	0,1181	0,0274	Muito moderno	0,0998	0,3667	0,0344
Bastante agitado	0,1522	0,2127	0,088	Bastante moderno	0,2523	0,3587	0,1446
Equilíbrio agitado-calmo	0,2519	0,2722	0,2008	Equilíbrio moderno-tradicional	0,4293	0,2362	0,4089
Bastante calmo	0,2859	0,239	0,3139	Bastante tradicional	0,1556	0,0331	0,2462
Muito calmo	0,2445	0,158	0,3699	Muito tradicional	0,0631	0,0052	0,1659
Desorganizado				Único			
Muito organizado	0,2585	0,2267	0,582	Muito comum	0,1201	0,0361	0,3716
Bastante organizado	0,33	0,317	0,3054	Bastante comum	0,1564	0,0784	0,235
Equilíbrio organizado-desorganizado	0,2359	0,2483	0,0898	Equilíbrio único-comum	0,4259	0,3557	0,3107
Bastante desorganizado	0,1258	0,1451	0,0197	Bastante único	0,1729	0,2406	0,0613
Muito desorganizado	0,0498	0,0629	0,0032	Muito único	0,1247	0,2891	0,0215
Delicado				Desagradável			
Muito austero	0,0475	0,022	0,006	Muito agradável	0,1837	0,6251	0,8293
Bastante austero	0,1264	0,0806	0,0351	Bastante agradável	0,485	0,3323	0,1631
Equilíbrio austero-delicado	0,4585	0,4029	0,2819	Equilíbrio agradável-desagradável	0,304	0,0419	0,0076
Bastante delicado	0,2601	0,315	0,3536	Bastante desagradável	0,0238	0,0007	0
Muito delicado	0,1075	0,1796	0,3234	Muito desagradável	0,0034	0	0
Extravagante				Emocional			
Muito modesto	0,1513	0,0823	0,7727	Muito racional	0,1452	0,0539	0,0695
Bastante modesto	0,3393	0,2681	0,199	Bastante racional	0,2395	0,139	0,1615
Equilíbrio modesto-extravagante	0,4105	0,4711	0,0276	Equilíbrio racional-emocional	0,3063	0,2775	0,2912
Bastante extravagante	0,0808	0,1347	0,0006	Bastante emocional	0,1931	0,2733	0,2588
Muito extravagante	0,0181	0,0438	0	Muito emocional	0,116	0,2563	0,219
Informal				Passivo			
Muito formal	0,0853	0,0262	0,1227	Muito activo	0,2381	0,5485	0,3647
Bastante formal	0,1111	0,0533	0,136	Bastante activo	0,3623	0,3356	0,382
Equilíbrio formal-informal	0,3407	0,2554	0,3551	Equilíbrio activo-passivo	0,2716	0,1012	0,1971
Bastante informal	0,2768	0,3243	0,2456	Bastante passivo	0,0788	0,0118	0,0393
Muito informal	0,1861	0,3408	0,1406	Muito passivo	0,0493	0,003	0,0169

Um agrupamento de turistas – Modelação adequada?

Liberal					Natural			
Muito conservador	0,0964	0,0085	0,2715		Muito artificial	0,007	0	0
Bastante conservador	0,1489	0,0321	0,237		Bastante artificial	0,027	0,0002	0
Equilíbrio conservador-liberal	0,3523	0,1862	0,3168		Equilíbrio natural-artificial	0,219	0,0131	0,0001
Bastante liberal	0,2677	0,3464	0,136		Bastante natural	0,535	0,2448	0,0273
Muito liberal	0,1347	0,4268	0,0387		Muito natural	0,212	0,7419	0,9725
Frio/Distante					Simples			
Muito caloroso	0,1292	0,7052	0,8117		Muito complexo	0,066	0,0627	0
Bastante caloroso	0,4497	0,2721	0,1798		Bastante complexo	0,193	0,1873	0,0006
Equilíbrio caloroso-frio	0,3286	0,022	0,0084		Equilíbrio complexo-simples	0,317	0,3146	0,0115
Bastante frio	0,0805	0,0006	0,0001		Bastante simples	0,266	0,2699	0,1184
Muito frio	0,012	0	0		Muito simples	0,159	0,1655	0,8695
Maduro					Cinzento			
Muito juvenil	0,109	0,2629	0,1516		Muito vivo	0,166	0,7742	0,502
Bastante juvenil	0,1614	0,2412	0,1914		Bastante vivo	0,372	0,2007	0,3656
Equilíbrio juvenil-maduro	0,3051	0,2825	0,3085		Equilíbrio vivo-cinzento	0,396	0,0248	0,1268
Bastante maduro	0,287	0,1646	0,2474		Bastante cinzento	0,047	0,0003	0,0049
Muito maduro	0,1375	0,0489	0,1011		Muito cinzento	0,019	0	0,0006

Tabela 25 - Resultado da Análise de cada Grupo

7.4. Variáveis Ordinais (agrupamento sobre amostra de “teste”)

Foram colocadas as variáveis como ordinais e usamos um filtro como peso que será a variável TESTE_TREINO.

Após a estimação do modelo houve o surgimento de 3 segmentos segundo o critério BIC e também para o critério AIC. Observe a Tabela 26 (BIC e AIC referem-se aos critérios de informação e npar ao número de parâmetros livres a estimar):

Número de Grupos	BIC(LL)	AIC(LL)	Npar
Grupo1	39284,25	38979,21	64
Grupo2	38703,44	38317,38	81
Grupo3	38381,43	37914,34	98
Grupo4	38264,88	37716,76	115
Grupo5	38239,96	37610,83	132
Grupo6	38235,1	37524,94	149

Tabela 26 - Análise de resultados das variáveis ordinais (agrupamento sobre amostra de “teste”)

Observe em seguida o gráfico segundo os critérios BIC e AIC:

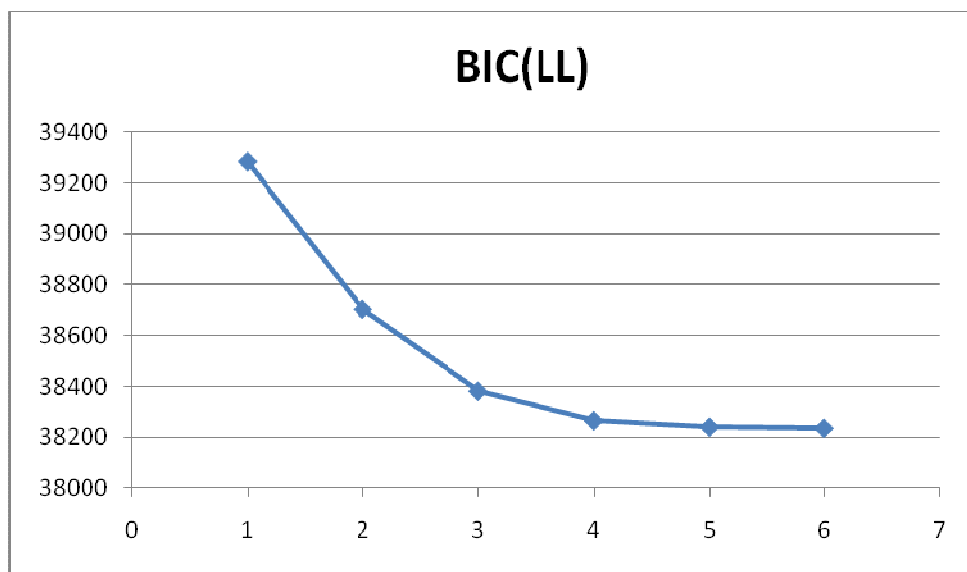


Ilustração 9 - Gráfico com indicador BIC segundo o estudo das variáveis ordinais (amostra de teste)

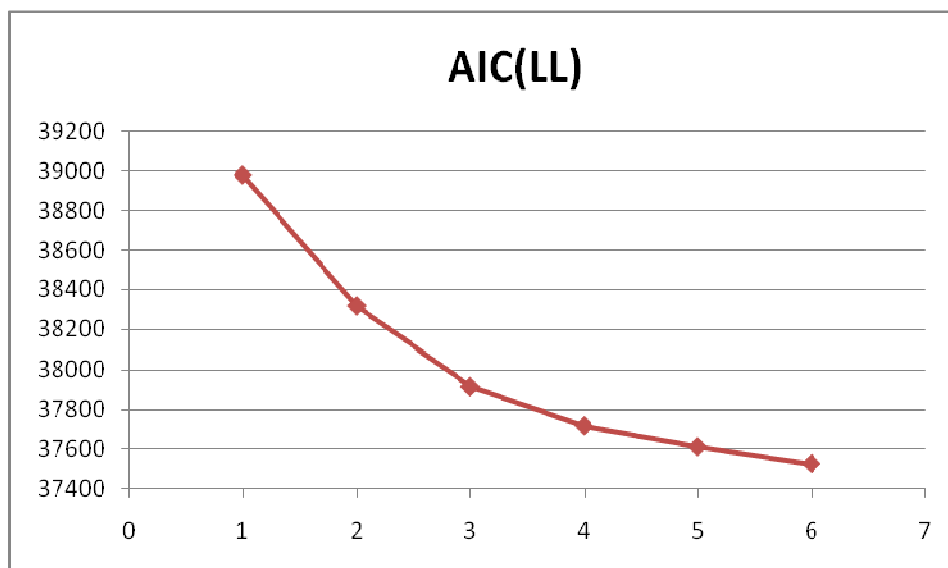


Ilustração 10 - Gráfico com indicador AIC segundo o estudo das variáveis ordinais (amostra de teste)

De acordo com o indicador BIC verifica-se um *cotovelo* para uma solução de 3 grupos o que justifica a consideração deste número, indicando uma mudança no gráfico correspondente a este indicador precisamente na solução com 3 grupos.

Ilustrando um resumo do Latent Gold em relação ao 3º grupo :

Solução com 3 grupos	
Número de casos	868
Número de Parâmetros (Npar)	98
Variação Sementação	553187
Melhor Sementação	1882850
Log-verosimilhança (LL)	-18859
BIC (based on LL)	38381,4
AIC (based on LL)	37914,3

Tabela 27 - Resultado da Análise de 3ºGrupo

A caracterização dos grupos nesta iteração será:

	Grupo1	Grupo2	Grupo3			Grupo1	Grupo2	Grupo3
Grupo Size	0,6213	0,2552	0,1235	Grupo Size	0,6213	0,2552	0,1235	
Indicators				Indicators				
Calmo				Tradicional				
Muito agitado	0,06	0,1597	0,0212	Muito moderno	0,101	0,2759	0,0505	
Bastante agitado	0,1276	0,2172	0,0659	Bastante moderno	0,2643	0,3683	0,1854	
Equilíbrio agitado-calmo	0,2902	0,316	0,2193	Equilíbrio moderno-tradicional	0,4077	0,2898	0,4007	
Bastante calmo	0,2967	0,2067	0,3283	Bastante tradicional	0,1355	0,0492	0,1868	
Muito calmo	0,2255	0,1005	0,3652	Muito tradicional	0,0914	0,0169	0,1766	
Desorganizado				Único				
Muito organizado	0,2371	0,3079	0,5873	Muito comum	0,1045	0,0457	0,3408	
Bastante organizado	0,3229	0,3445	0,3036	Bastante comum	0,1534	0,0956	0,2434	
Equilíbrio organizado-desorganizado	0,2475	0,2169	0,0883	Equilíbrio único-comum	0,4178	0,3707	0,3224	
Bastante desorganizado	0,1304	0,0939	0,0176	Bastante único	0,178	0,2249	0,0668	
Muito desorganizado	0,0621	0,0368	0,0032	Muito único	0,1462	0,2631	0,0267	
Delicado				Desagradável				
Muito austero	0,0353	0,0174	0,0009	Muito agradável	0,1566	0,4715	0,9153	
Bastante austero	0,1248	0,083	0,0124	Bastante agradável	0,4877	0,4383	0,0831	
Equilíbrio austero-delicado	0,5165	0,4651	0,1907	Equilíbrio agradável-desagradável	0,3304	0,0887	0,0016	
Bastante delicado	0,2299	0,2803	0,3164	Bastante desagradável	0,0162	0,0013	0	
Muito delicado	0,0935	0,1543	0,4796	Muito desagradável	0,0092	0,0002	0	
Extravagante				Emocional				
Muito modesto	0,1579	0,1057	0,6757	Muito racional	0,1449	0,066	0,0581	
Bastante modesto	0,3069	0,2631	0,2536	Bastante racional	0,2476	0,1634	0,1517	
Equilíbrio modesto-extravagante	0,4265	0,4681	0,068	Equilíbrio racional-emocional	0,323	0,3088	0,3021	
Bastante extravagante	0,082	0,1152	0,0025	Bastante emocional	0,1752	0,2425	0,25	
Muito extravagante	0,0266	0,0479	0,0002	Muito emocional	0,1093	0,2191	0,2381	
Informal				Passivo				
Muito formal	0,0892	0,0425	0,2377	Muito activo	0,2063	0,6145	0,3884	
Bastante formal	0,1311	0,0839	0,2118	Bastante activo	0,3955	0,3177	0,4136	
Equilíbrio formal-informal	0,3329	0,2866	0,3262	Equilíbrio activo-passivo	0,2925	0,0634	0,17	
Bastante informal	0,2712	0,3138	0,161	Bastante passivo	0,063	0,0037	0,0204	
Muito informal	0,1756	0,2732	0,0632	Muito passivo	0,0427	0,0007	0,0077	

Liberal				Natural			
Muito conservador	0,0974	0,0329	0,1694	Muito artificial	0,0147	0,0003	0
Bastante conservador	0,1507	0,0777	0,2009	Bastante artificial	0,0306	0,0022	0
Equilíbrio conservador-liberal	0,3073	0,242	0,314	Equilíbrio natural-artificial	0,2389	0,057	0,0048
Bastante liberal	0,2664	0,3203	0,2087	Bastante natural	0,4951	0,3841	0,1389
Muito liberal	0,1782	0,3271	0,107	Muito natural	0,2207	0,5562	0,8562
Frio/Distante				Simples			
Muito caloroso	0,117	0,6736	0,5666	Muito complexo	0,0693	0,057	0,0002
Bastante caloroso	0,4221	0,2951	0,373	Bastante complexo	0,1919	0,172	0,0039
Equilíbrio caloroso-frio	0,3598	0,0305	0,058	Equilíbrio complexo-simples	0,316	0,3082	0,0401
Bastante frio	0,0808	0,0008	0,0024	Bastante simples	0,2756	0,2927	0,2195
Muito frio	0,0204	0	0,0001	Muito simples	0,1472	0,1701	0,7364
Maduro				Cinzento			
Muito juvenil	0,1075	0,2664	0,1026	Muito vivo	0,1392	0,6377	0,426
Bastante juvenil	0,1388	0,2107	0,1352	Bastante vivo	0,3723	0,3015	0,405
Equilíbrio juvenil-maduro	0,338	0,3145	0,3365	Equilíbrio vivo-cinzento	0,4131	0,0591	0,1597
Bastante maduro	0,2853	0,1627	0,2901	Bastante cinzento	0,0625	0,0016	0,0086
Muito maduro	0,1305	0,0456	0,1356	Muito cinzento	0,0128	0,0001	0,0006

Tabela 28 - Resultado da Análise de cada Grupo

7.5. Variáveis Métricas (agrupamento sobre amostra de “treino”)

Foram colocadas as variáveis métricas e usamos um filtro como peso que será a variável TREINO_TESTE.

Após a estimação do modelo houve o surgimento de 5 segmentos segundo o critério BIC e também para o critério AIC. Observe a Tabela 29 (BIC e AIC referem-se aos critérios de informação e npar ao número de parâmetros livres a estimar):

Número de Grupos	BIC(LL)	AIC(LL)	Npar
Grupo1	40663,47	40511,36	32
Grupo2	37829,36	37520,39	65
Grupo3	37433,69	36967,85	98
Grupo4	35577,99	34955,29	131
Grupo5	33233,41	32453,84	164
Grupo6	33495,33	32558,9	197

Tabela 29 - Análise de resultados das variáveis métricas (agrupamento sobre amostra de “treino”)

Observe em seguida o gráfico segundo os critérios BIC e AIC:

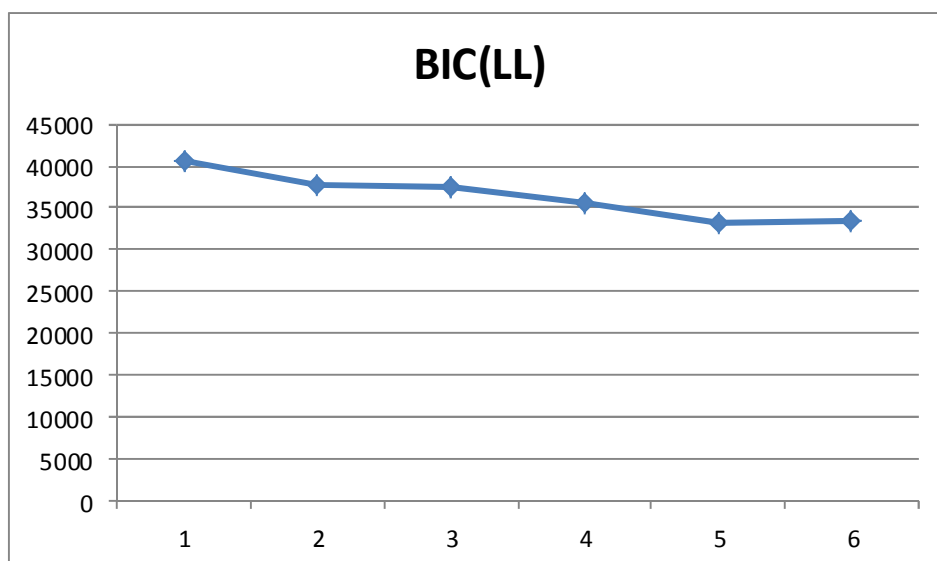


Ilustração 11 - Gráfico com indicador BIC segundo o estudo das variáveis métricas (amostra de treino)

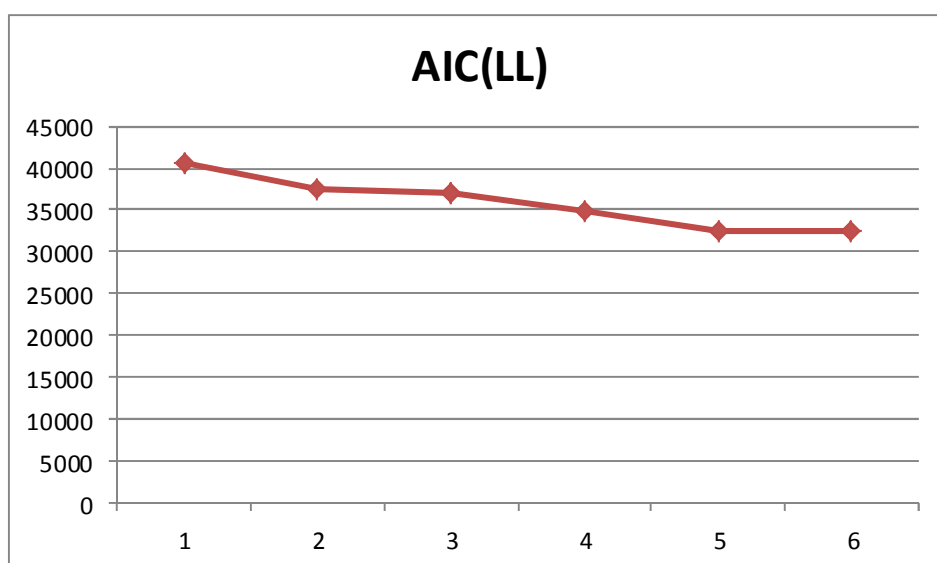


Ilustração 12 - Gráfico com indicador AIC segundo o estudo das variáveis métricas (amostra de treino)

De acordo com o indicador BIC verifica-se um *cotovelo* para uma solução de 5 grupos o que justifica a consideração deste número, indicando uma mudança no gráfico correspondente a este indicador precisamente na solução com 5 grupos.

Ilustrando um resumo do Latent Gold em relação a solução com 5 grupos :

Solução com 5 grupos	
Número de casos	857
Número de Parâmetros (Npar)	164
Variação Sementação	10
Melhor Sementação	543226
Log-verossimilhança (LL)	-16062,92
BIC (based on LL)	33233,41
AIC (based on LL)	32453,84

Tabela 30 - Resultado da Análise de 5ºGrupo

A caracterização dos grupos nesta solução será:

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Grupo Size	0,4115	0,2625	0,1471	0,1109	0,068
Indicators					
Calmo					
Média	3,422	3,5778	3,4683	3,291	3,5925
Desorganizado					
Média	2,4423	2,2267	2,0238	2,4387	2,1251
Delicado					
Média	3,3306	3,4	3,7937	3,0842	3,4526
Extravagante					
Média	2,5194	2,2756	2,0555	2,6565	2,3157
Informal					
Média	3,523	3,4889	3,5556	3,2569	2,8804
Tradicional					
Média	2,7548	2,8	2,5476	2,7569	2,5288
Único					
Média	3,0845	2,9111	3,0873	3,1494	3,0344
Desagradável					
Média	2,1316	1,9156	1	2,3947	1,6573
Emocional					
Média	2,9449	3,1422	3,4921	2,8343	3,2268
Passivo					
Média	2,3001	1,9378	1,754	2,3887	2,4332
Liberal					
Média	3,3767	3,2845	3,3968	3,1923	2,7388
Frio/Distante					
Média	2,2896	1,96	1	2,652	2,0325
Maduro					
Média	3,1204	3,0622	2,7063	3,0618	2,8391
Natural					
Média	4	5	5	3	2,6412
Simples					
Média	3,2351	3,7778	4,0159	2,9486	3,2395
Cinzento					
Média	2,2715	2,1867	1	2,5476	1,8655

Tabela 31 - Resultado da Análise de cada Grupo

7.6. Variáveis Métricas (agrupamento sobre amostra de “teste”)

Foram colocadas as variáveis métricas e usamos um filtro como peso que será a variável TESTE_TREINO.

Após a estimação do modelo houve o surgimento de 5 segmentos segundo o critério BIC e também para o critério AIC. Observe a Tabela 32 (BIC e AIC referem-se aos critérios de informação e npar ao número de parâmetros livres a estimar):

Número de Grupos	BIC(LL)	AIC(LL)	Npar
Grupo1	41107,6	40955,08	32
Grupo2	38796,85	38487,05	65
Grupo3	38285,58	37818,49	98
Grupo4	37321,43	36697,06	131
Grupo5	35133,9	34352,24	164
Grupo6	34058,13	33119,19	197

Tabela 32 - Análise de resultados das variáveis métricas (agrupamento sobre amostra de “teste”)

Observe em seguida o gráfico segundo o critério BIC:

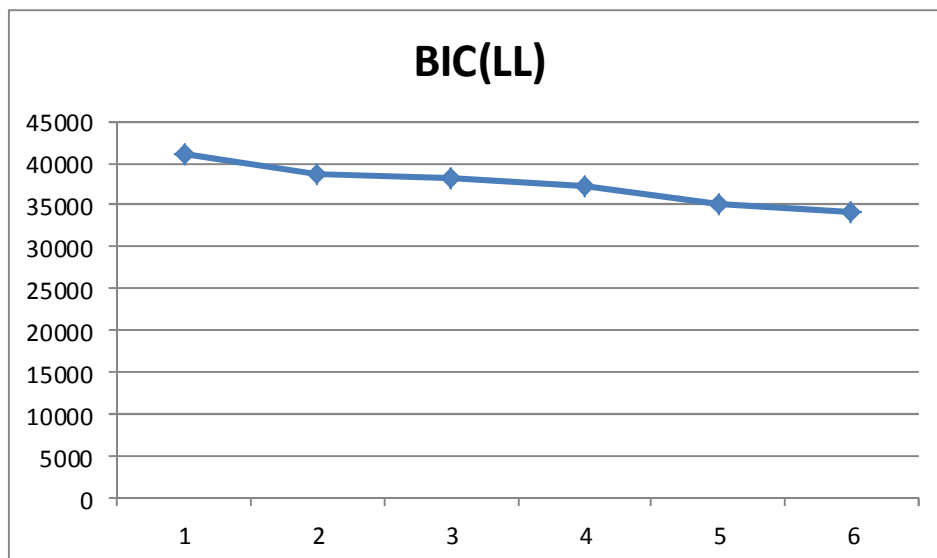


Ilustração 13 - Gráfico com indicador BIC segundo o estudo das variáveis métricas (amostra de teste)

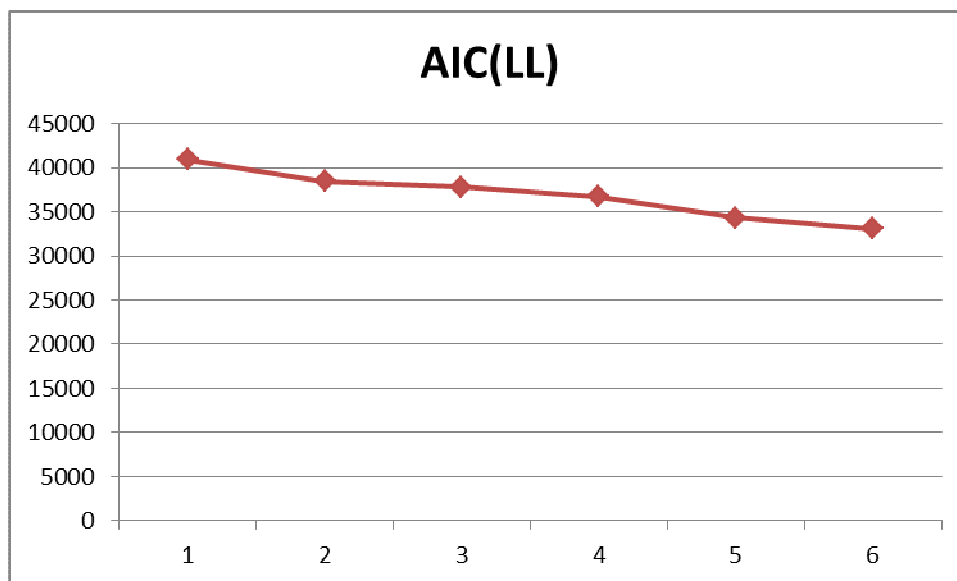


Ilustração 14 - Gráfico com indicador AIC segundo o estudo das variáveis métricas (amostra de teste)

De acordo com o indicador BIC verifica-se um *cotovelo* para uma solução de 5 grupos o que justifica a consideração deste número, indicando uma mudança no gráfico correspondente a este indicador precisamente na solução com 5 grupos.

Ilustrando um resumo do Latent Gold em relação a solução com 5 grupos :

Solução com 5 grupos	
Número de casos	868
Número de Parâmetros (Npar)	164
Variação Sementação	10
Melhor Sementação	1729797
Log-verossimilhança (LL)	-16544,3
BIC (based on LL)	35133,9
AIC (based on LL)	34352,24

Tabela 33 - Resultado da Análise de 5ºGrupo

A caracterização dos grupos nesta solução será:

	Grupo1	Grupo2	Grupo3	Grupo4	Grupo5
Cluster Size	0,4094	0,291	0,1492	0,1222	0,0281
Indicators					
Calmo					
Média	3,4149	3,2064	3,4161	3,8314	3,0537
Desorganizado					
Média	2,4044	2,2824	2,447	1,5736	2,6672
Delicado					
Média	3,3262	3,4418	3,1452	4,05	3,0515
Extravagante					
Média	2,5833	2,6067	2,7576	1	2,9178
Informal					
Média	3,3476	3,4815	3,3289	2,8941	3,0989
Tradicional					
Média	2,7328	2,6099	2,6528	3,1863	2,2067
Único					
Média	3,1432	3,2436	3,2539	2,4579	3,3773
Desagradável					
Média	2,1389	1,8112	2,2809	1	2,4931
Emocional					
Média	3,0158	3,2654	2,8456	3,1636	2,4627
Passivo					
Média	2,2047	1,8464	2,2615	1,7458	2,2532
Liberal					
Média	3,3773	3,5342	3,2767	2,9326	3,9551
Frio/Distante					
Média	2,2474	1,7807	2,5081	1,4772	2,4983
Maduro					
Média	3,1197	2,8614	2,9313	3,2649	2,6813
Natural					
Média	4	5	3	4,6129	1,763
Simples					
Média	3,2966	3,6362	2,9438	4,323	2,4948
Cinzento					
Média	2,2588	1,7993	2,3859	1,8002	2,4918

Tabela 34 - Resultado da Análise de cada Grupo

7.7. Restantes Quadros da solução Seleccionada

Modesto-Extravagante*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
modesto-extravagante	muito modesto	12,8%	14,6%	79,5%	21,1%
	bastante modesto	36,0%	20,1%	14,1%	29,9%
	equilíbrio modesto-extravagante	42,4%	43,8%	6,3%	38,4%
	bastante extravagante	7,2%	15,1%		8,1%
	muito extravagante	1,6%	6,5%		2,5%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 35 - Classificação entre modesto e extravagante

Formal-Informal*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
formal-informal	muito formal	5,1%	6,3%	33,2%	8,7%
	bastante formal	13,2%	7,0%	11,2%	11,6%
	equilíbrio formal-informal	37,4%	24,7%	18,0%	32,3%
	bastante informal	31,2%	25,8%	10,7%	27,5%
	muito informal	13,1%	36,2%	26,8%	19,9%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 36 - Classificação entre formal e informal

Moderno-Tradicional*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
moderno-tradicional	muito moderno	5,7%	37,5%	17,6%	14,2%
	bastante moderno	29,7%	31,5%	4,4%	27,1%
	equilíbrio moderno-tradicional	43,8%	25,3%	31,7%	38,3%
	bastante tradicional	15,7%	4,2%	15,6%	13,1%
	muito tradicional	5,1%	1,6%	30,7%	7,4%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 37 - Classificação entre moderno e tradicional

Um agrupamento de turistas – Modelação adequada?

Comum-Único*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
comum-único	muito comum	7,0%	8,3%	51,7%	12,6%
	bastante comum	18,2%	6,8%	12,7%	15,0%
	equilíbrio comum-único	46,3%	29,4%	21,5%	39,6%
	bastante único	18,6%	21,4%	3,9%	17,4%
	muito único	9,9%	34,1%	10,2%	15,3%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 38 - Classificação entre comum e único

Agradável-Desagradável*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
agradável-desagradável	muito agradável	15,0%	60,7%	89,8%	34,0%
	bastante agradável	51,8%	31,5%	7,8%	42,0%
	equilíbrio agradável-desagradável	31,0%	6,8%	2,4%	22,2%
	bastante desagradável	1,8%	,8%		1,3%
	muito desagradável	,5%	,3%		,4%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 39 - Classificação entre agradável e desagradável

Racional-Emocional*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
racional-emocional	muito racional	10,1%	10,9%	21,0%	11,6%
	bastante racional	29,2%	6,8%	3,9%	21,2%
	equilíbrio racional-emocional	31,9%	31,0%	24,4%	30,8%
	bastante emocional	21,3%	25,0%	11,2%	20,9%
	muito emocional	7,5%	26,3%	39,5%	15,5%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 40 - Classificação entre racional e emocional

Um agrupamento de turistas – Modelação adequada?

Activo-Passivo*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
activo-passivo	muito activo	16,6%	64,6%	59,5%	32,4%
	bastante activo	44,4%	25,3%	17,1%	36,9%
	equilíbrio activo-passivo	29,4%	8,3%	9,8%	22,4%
	bastante passivo	6,8%	1,0%	3,9%	5,2%
	muito passivo	2,8%	,8%	9,8%	3,2%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 41 - Classificação entre activo e passivo

Conservador-Liberal*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
conservador-liberal	muito conservador	6,3%	6,5%	32,7%	9,5%
	bastante conservador	17,6%	3,4%	11,7%	13,7%
	equilíbrio conservador-liberal	34,0%	19,8%	29,8%	30,3%
	bastante liberal	31,1%	25,0%	8,3%	27,0%
	muito liberal	11,0%	45,3%	17,6%	19,4%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 42 - Classificação entre conservador e liberal

Caloroso-Frio*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
caloroso-frio	muito caloroso	9,8%	73,2%	78,5%	32,1%
	bastante caloroso	48,0%	22,9%	13,2%	38,3%
	equilíbrio caloroso-frio	33,0%	3,9%	6,3%	23,4%
	bastante frio	7,7%		1,5%	5,3%
	muito frio	1,5%		,5%	1,0%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 43 - Classificação entre caloroso e frio

Um agrupamento de turistas – Modelação adequada?

Juvenil-Maduro*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
juvenil-maduro	muito juvenil	6,6%	32,6%	25,4%	14,6%
	bastante juvenil	18,4%	16,7%	8,8%	16,9%
	equilíbrio juvenil-maduro	34,5%	26,6%	25,4%	31,7%
	bastante maduro	31,0%	15,1%	15,6%	25,6%
	muito maduro	9,5%	9,1%	24,9%	11,2%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 44 - Classificação entre juvenil e maduro

Artificial-Natural*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
artificial-natural	muito artificial	,9%	,5%		,7%
	bastante artificial	2,8%	,3%		1,9%
	equilíbrio artificial-natural	21,9%	6,0%		15,8%
	bastante natural	55,0%	22,7%	5,4%	41,9%
	muito natural	19,4%	70,6%	94,6%	39,7%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 45 - Classificação entre artificial e natural

Complexo-Simples*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
complexo-simples	muito complexo	4,2%	13,0%	,5%	5,7%
	bastante complexo	20,1%	14,6%	1,0%	16,6%
	equilíbrio complexo-simples	32,4%	28,4%	2,4%	27,9%
	bastante simples	32,3%	17,4%	7,8%	26,1%
	muito simples	11,0%	26,6%	88,3%	23,7%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 46 - Classificação entre complexo e simples

Colorido-Cinzeno*Grupo modal – Validação Cruzada

% dentro do Grupo modal

		Grupo modal			Total
		1	2	3	
colorido-cinzeno	muito colorido	11,5%	74,7%	61,0%	31,5%
	bastante colorido	43,0%	20,6%	15,6%	34,8%
	equilíbrio colorido-cinzeno	39,4%	4,4%	17,6%	29,0%
	bastante cinzeno	4,8%	,3%	3,4%	3,7%
	muito cinzeno	1,1%		2,4%	1,0%
Total		100,0%	100,0%	100,0%	100,0%

Tabela 47 - Classificação entre colorido e cinzeno

7.8. Restantes Gráficos da Solução Seleccionada

Modesto-Extravagante*Grupo Modal. Validação Cruzada: % dentro do Grupo modal

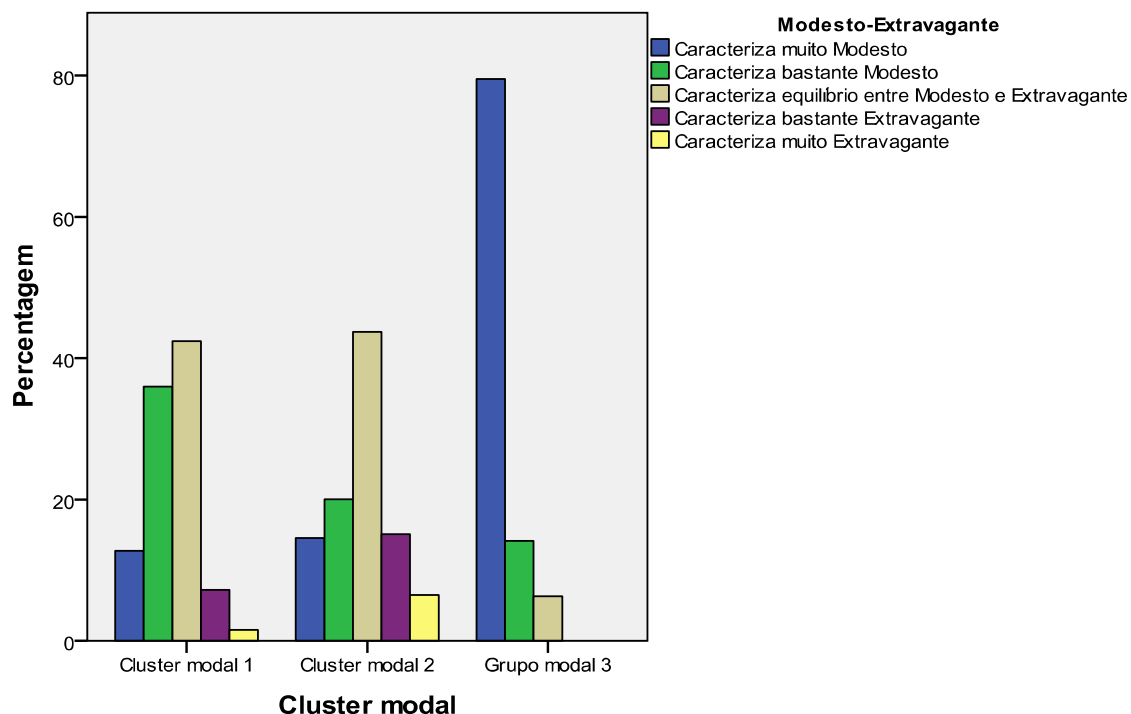


Ilustração 15 - Gráfico correspondente a variável modesto-extravagante

Formal-Informal*Grupo Modal. Validação Cruzada= % dentro do Grupo modal

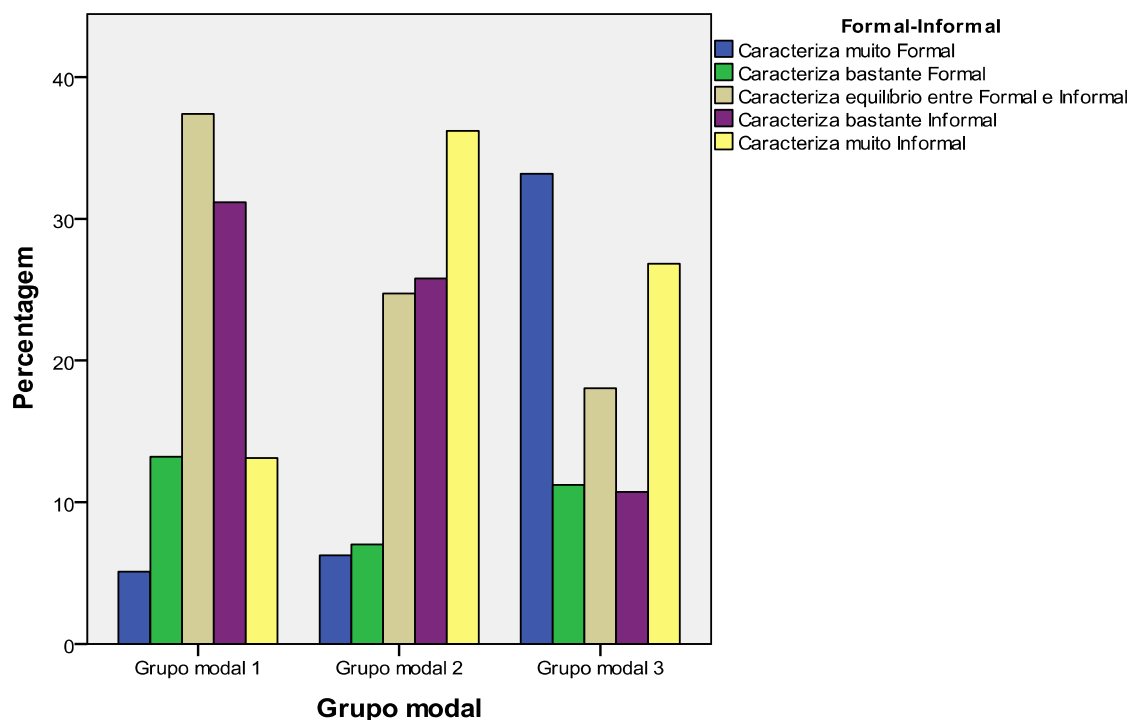


Ilustração 16 - Gráfico correspondente a variável formal-informal

Moderno-Tradicional*Grupo Modal. Validação Cruzada= % dentro do Grupo modal

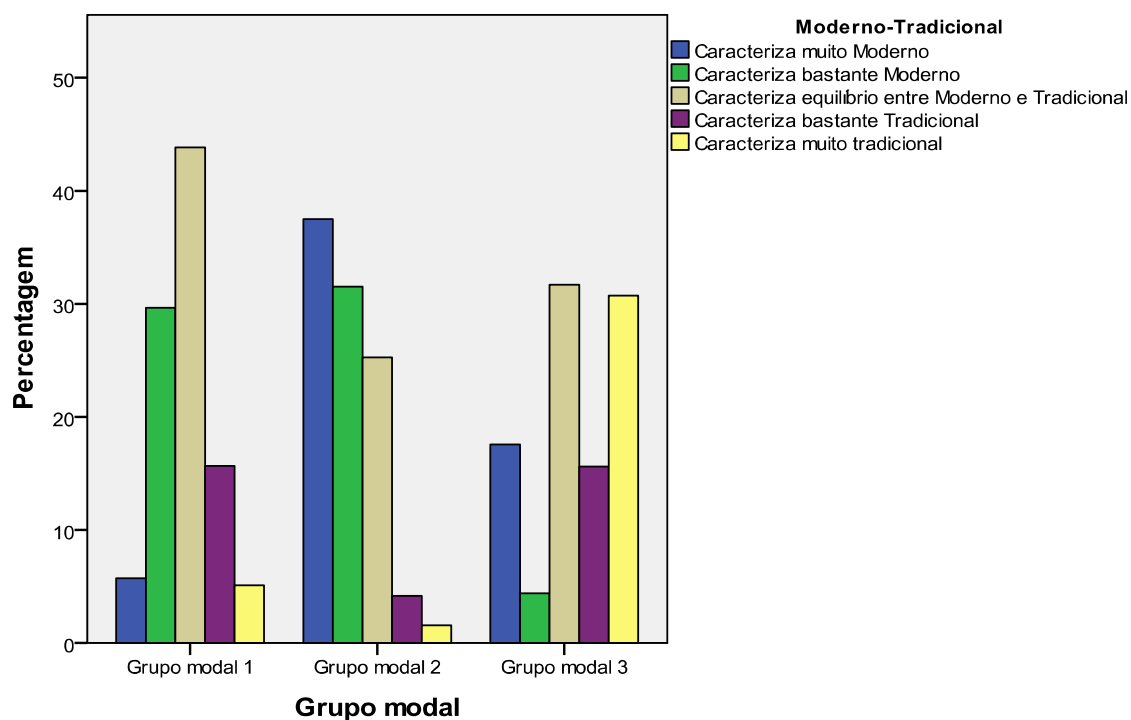


Ilustração 17 - Gráfico correspondente a variável moderno-tradicional

Comum-Único*Grupo Modal. Validação Cruzada= % dentro do Grupo modal

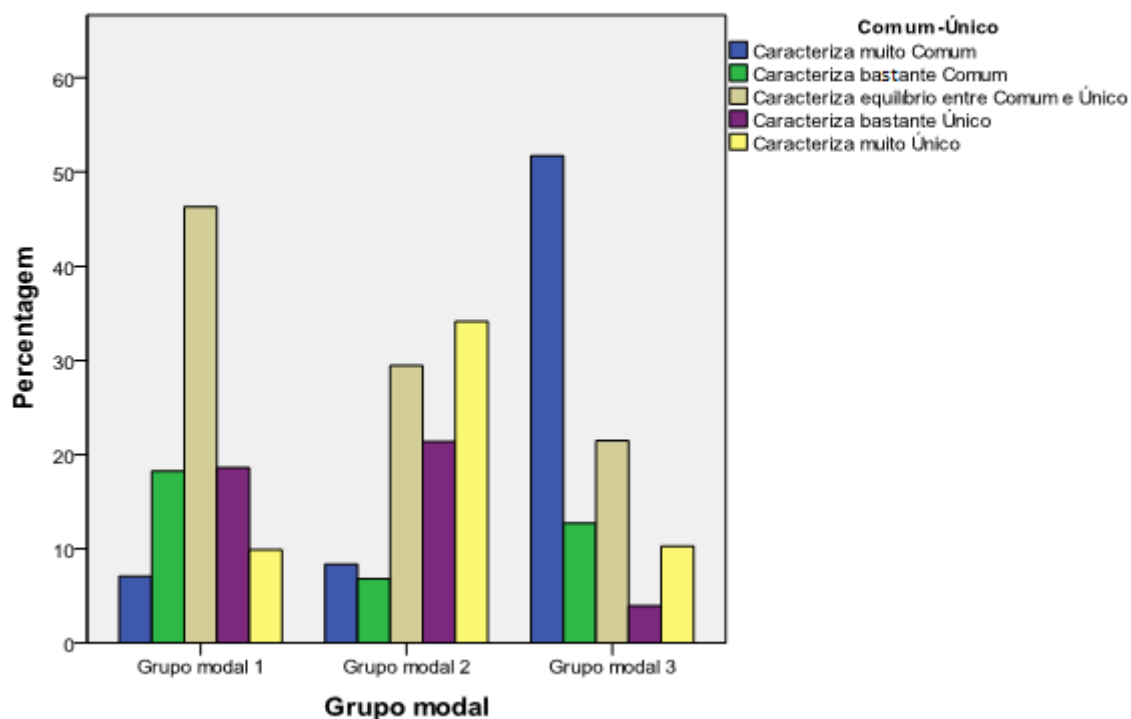


Ilustração 18 - Gráfico correspondente a variável comum-único

Agradável-Desagradável*Grupo Modal. Validação Cruzada= % dentro do Grupo modal

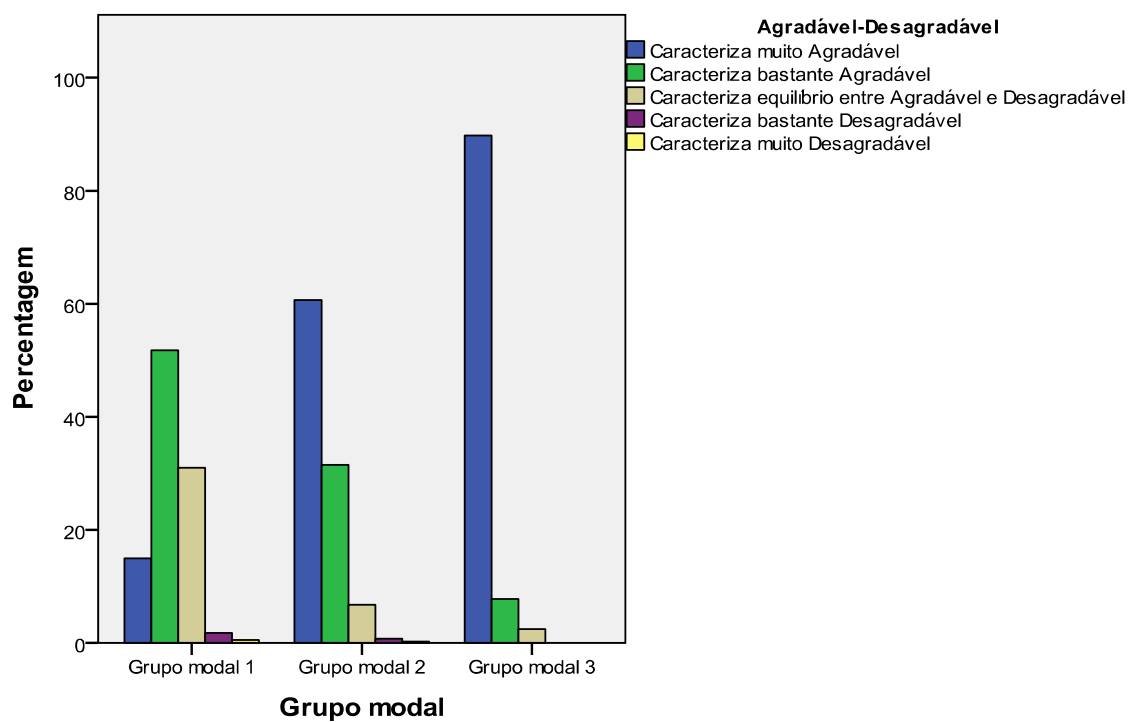


Ilustração 19 - Gráfico correspondente a variável agradável-desagradável

Racional-Emocional*Grupo Modal. Validação Cruzada= % dentro do Grupo modal

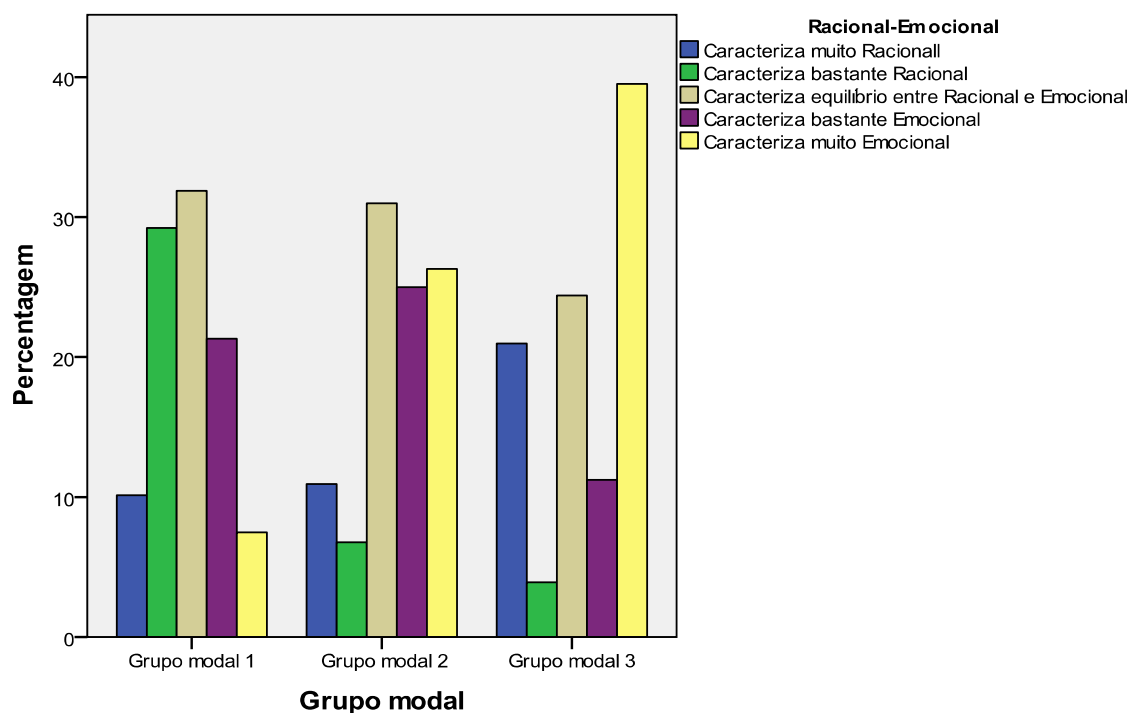


Ilustração 20 - Gráfico correspondente a variável racional-emocional

Activo-Passivo*Grupo Modal. Validação Cruzada= % dentro do Grupo modal

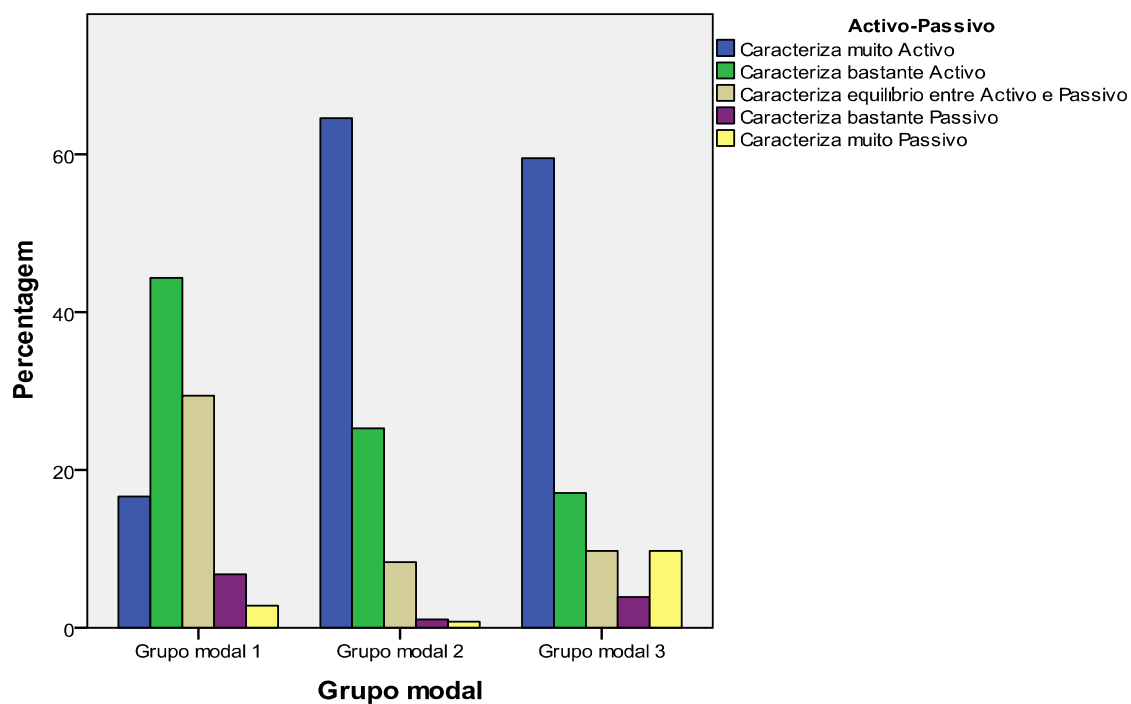


Ilustração 21 - Gráfico correspondente a variável activo-passivo

Conservador-Liberal*Grupo Modal. Validação Cruzada= % dentro do Grupo modal

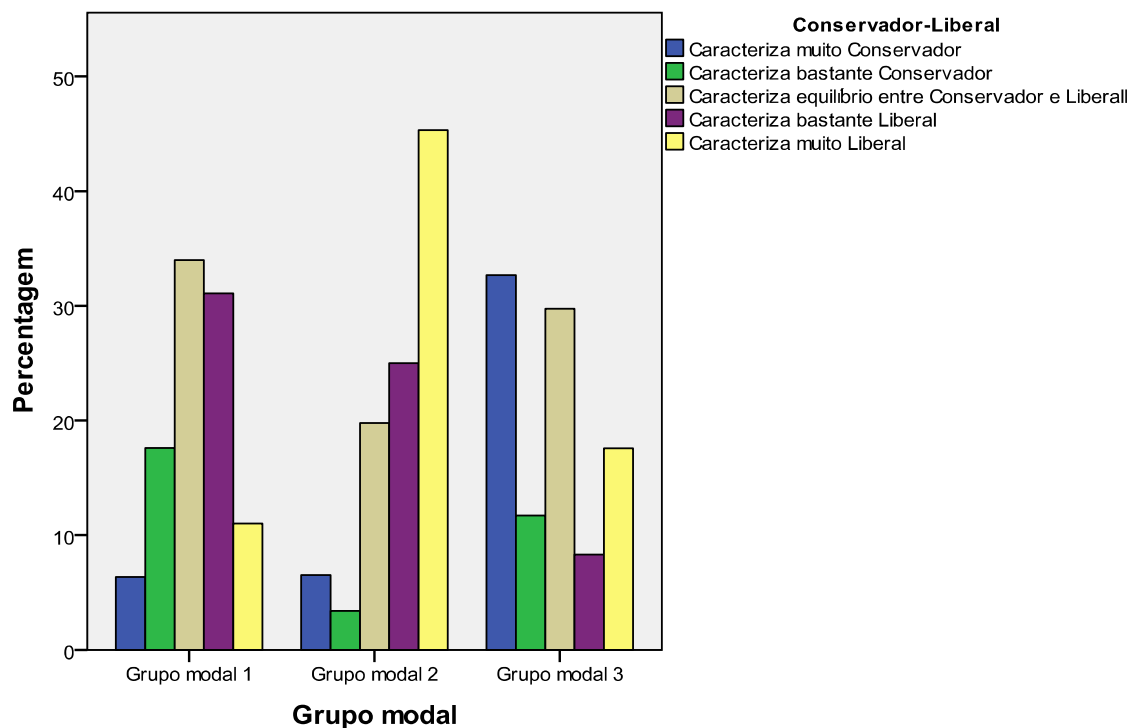


Ilustração 22 - Gráfico correspondente a variável conservador-liberal

Caloroso-Frio*Grupo Modal. Validação Cruzada= % dentro do Grupo modal

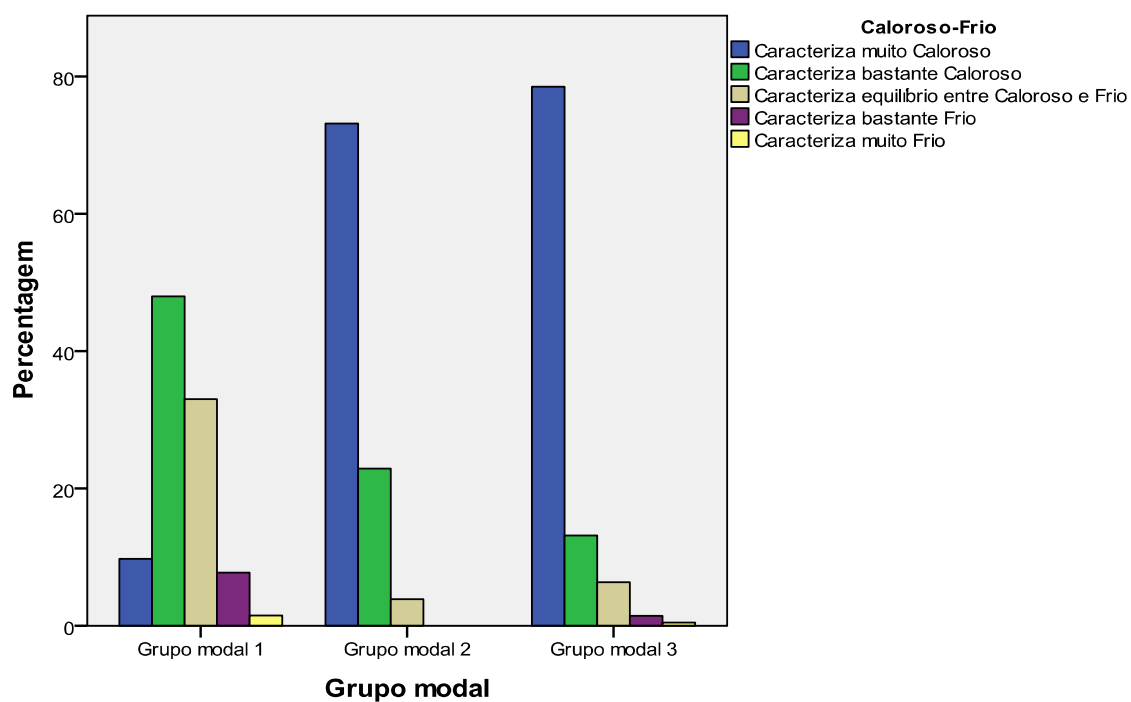


Ilustração 23 - Gráfico correspondente a variável caloroso-frio

Juvenil-Maduro*Grupo Modal. Validação Cruzada= % dentro do Grupo modal

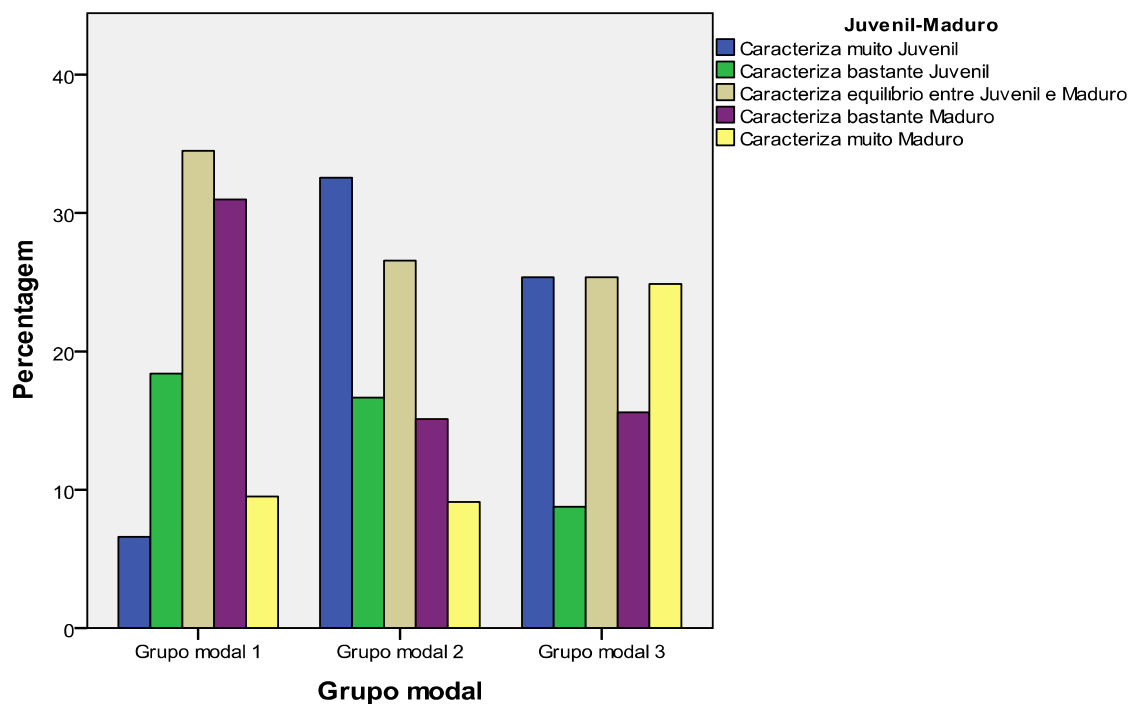


Ilustração 24 - Gráfico correspondente a variável juvenil-maduro

Artificial-Natural*Grupo Modal. Validação Cruzada= % dentro do Grupo modal

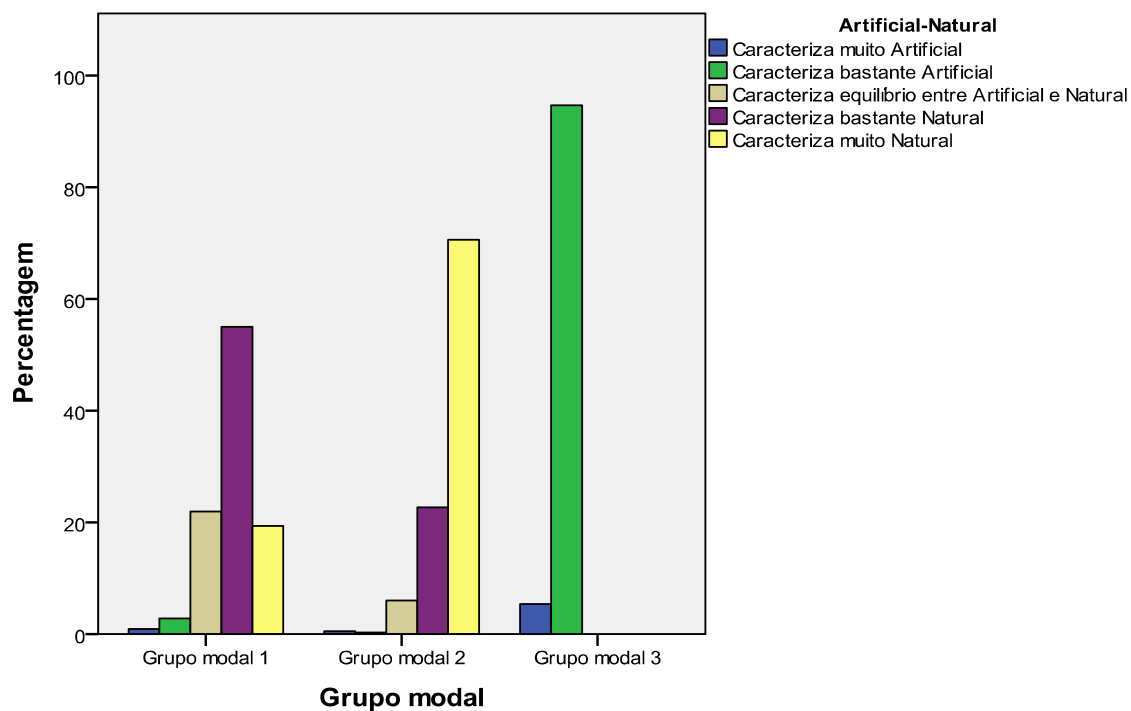


Ilustração 25 - Gráfico correspondente a variável artificial-natural

Complexo-Simples*Grupo Modal. Validação Cruzada= % dentro do Grupo modal

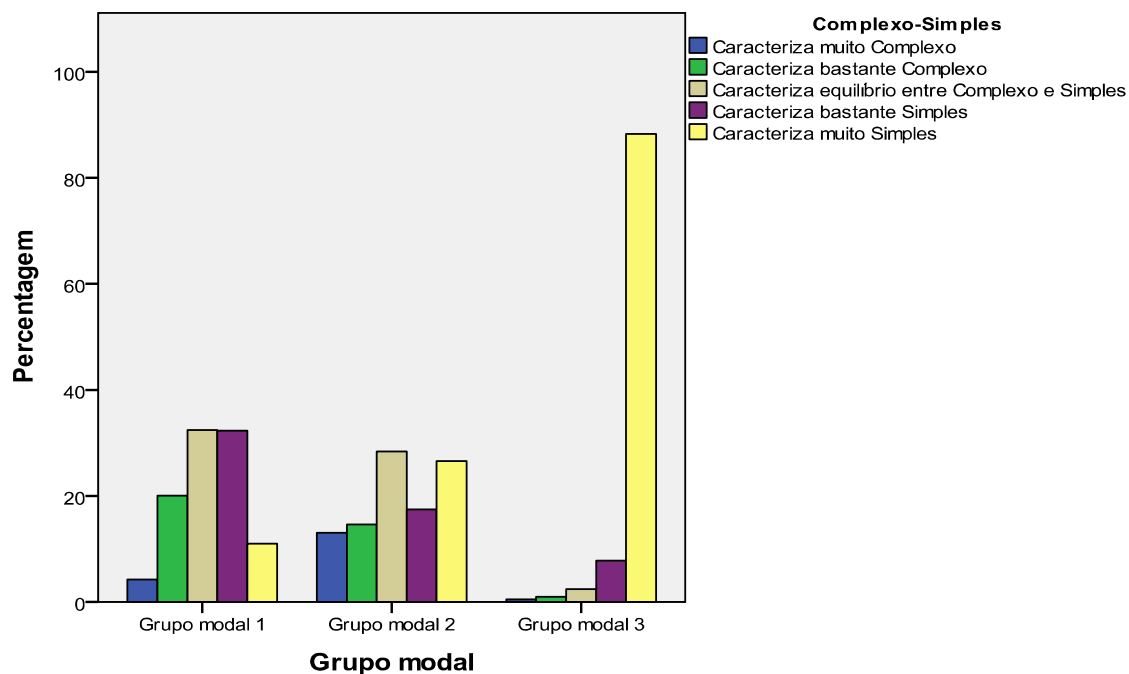


Ilustração 26 - Gráfico correspondente a variável complexo-simples

Colorido-Cinzeno*Grupo Modal. Validação Cruzada= % dentro do Grupo modal

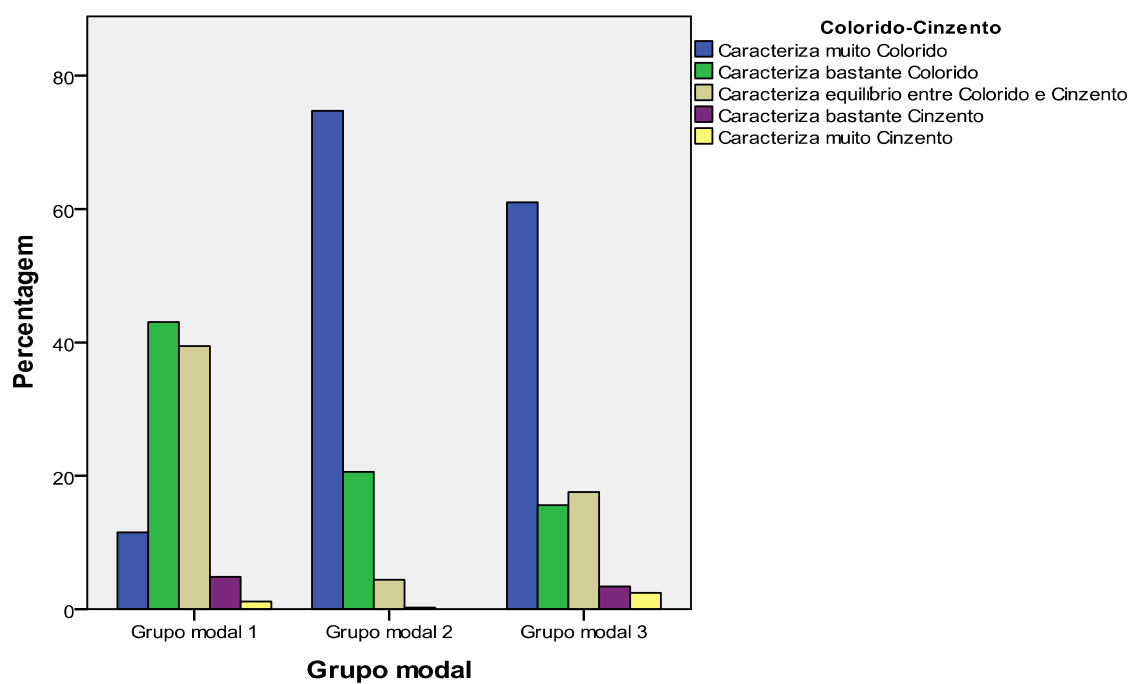


Ilustração 27 - Gráfico correspondente a variável colorido-cinzeno