

## Resolução do exercício: " WHISKY "

Este documento deve ser distribuído aos participantes por os responsáveis do seminário



## Programme PRESTA PROGRAMME DE RECHERCHE ET D'ENSEIGNEMENT EN STATISTIQUE APPLIQUÉE

### Resolução de um exercício de aplicação da Análise Fatorial de Correspondências Simples

#### Objetivo do estudo da Tabela de Dados "Whisky" :

Comparar e seleccionar las marcas de whisky que a presentan, en el mercado francés, a relação «qualidade / preço» más favorável para o consumidor final.

#### III. ETAPAS DE RESOLUÇÃO E PERGUNTAS OPERACIONAIS DO ESTUDO

##### III.1. Primeira parte: Verificação formal dos dados e estudo das distribuições univariadas

1. *Quantas columnas apresenta o arquivo de dados?*

*Resposta :* 6 columnas.

2. *Que contém a primeira columna do arquivo de dados?*

*Resposta:* Os rótulos dos indivíduos, isto é a primeira columna contém o identificador de cada marca de whisky.

3. *Quantos dígitos contém os valores que figuram na primeira columna do arquivo de dados? Estão entre aspas?*

*Resposta:* A primeira columna do arquivo de dados apresenta 2 dígitos. O primeiro whisky está identificado com o valor 01 e o ultimo com o valor 35. O identificador dos indivíduos da tabela não foi gravado entre aspas.

4. *Quais são os valores das cinco variáveis para o whisky N°1 e para o whisky N°35?*

*Resposta:* Whisky N°1: 70,20,1,5,3; Whisky N°35: 95, 100, 3, 12, 0.

5. *O arquivo de dados, tem formato livre, com um caracter branco como separador entre dois valores consecutivos?*

*Resposta:* Sim, o arquivo de dados tem formato livre, com o caracter branco como separador. A verificação formal do conteúdo do arquivo de dados é absolutamente necessária para poder definir os parâmetros de importação dos dados, em SPAD.N, desde um arquivo ASCII.

6. *Quantas variáveis estão definidas no arquivo de rótulos?*

*Resposta:* 5 variáveis.

7. Explique o significado do conteúdo da primeira linha do arquivo de rótulos.

Resposta : Conteúdo da primeira linha: 0001 Preço. O valor 1, em quarta posição, indica que se trata de uma variável contínua. Sendo a primeira linha, o rótulo PREÇO indica que a primeira coluna da Tabela de Dados contém a variável “preços dos whiskys”.

8. Explique o significado do conteúdo da última linha do arquivo de rótulos.

Resposta : Conteúdo da última linha: APRE Apreciación A última coluna do arquivo de dados contém a variável “Apreciação da qualidade dos whiskys pelos especialistas”. Esta linha permite associar a essa variável um rótulo curto (com quatro dígitos alfanuméricos) “APRE” e um rótulo cumprido (em 11 dígitos) “Apreciación” .

9. Quais são as 3 marcas de whisky mais caras e as 3 marcas de whisky mais baratas?

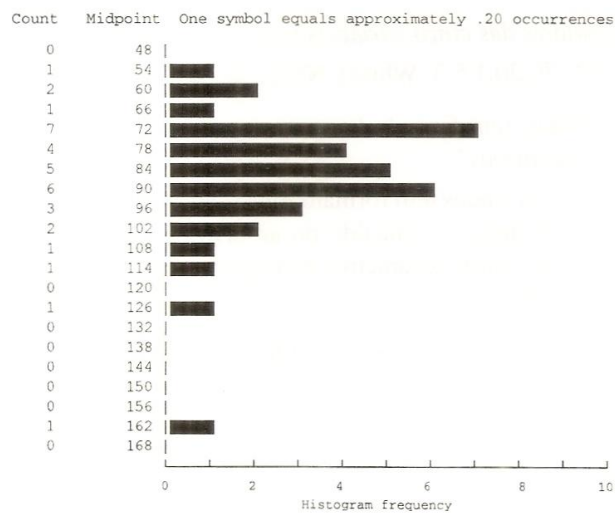
Resposta :

- Os whiskys mais caros são: Whisky N°26 (113 Fr), Whisky N°29 (127 Fr) e Whisky N°30 (160 Fr).
- Os mais baratos são: Whisky N°8 (55 Fr), Whisky N°2 (60 Fr) e Whisky N°13 (62 Fr).

Distribuição dos preços das 35 marcas de whisky

Value	Frequency	Percent	Valid Percent	Cum Percent
55.00	1	2.9	2.9	2.9
60.00	1	2.9	2.9	5.7
62.00	1	2.9	2.9	8.6
65.00	1	2.9	2.9	11.4
70.00	3	8.6	8.6	20.0
73.00	3	8.6	8.6	28.6
74.00	1	2.9	2.9	31.4
77.00	1	2.9	2.9	34.3
78.00	1	2.9	2.9	37.1
80.00	2	5.7	5.7	42.9
82.00	2	5.7	5.7	48.6
83.00	1	2.9	2.9	51.4
85.00	1	2.9	2.9	54.3
86.00	1	2.9	2.9	57.1
87.00	4	11.4	11.4	68.6
90.00	2	5.7	5.7	74.3
93.00	1	2.9	2.9	77.1
95.00	1	2.9	2.9	80.0
96.00	1	2.9	2.9	82.9
100.00	2	5.7	5.7	88.6
110.00	1	2.9	2.9	91.4
113.00	1	2.9	2.9	94.3
127.00	1	2.9	2.9	97.1
160.00	1	2.9	2.9	100.0
Total	35	100.0	100.0	

Variable : Precio



10. Qual é o valor da mediana e a variância da percentagem de malte das 35 marcas de whisky estudadas?

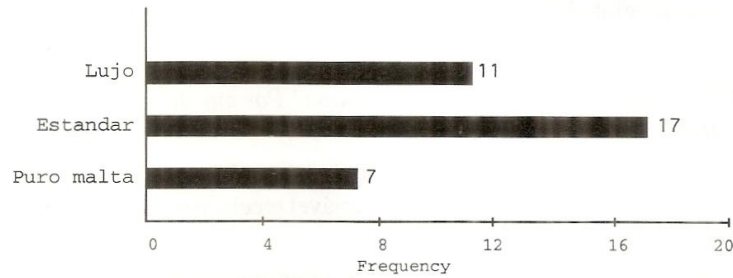
Resposta : A percentagem média de malte das 35 marcas de whisky é de um 40%. A variância da distribuição da porcentagem de malte é a seguinte:  $s^2 = 758,25$ .

11. Qual é a distribuição das marcas de whisky segundo a categoria das mesmas?

Resposta: A distribuição da variável Categoria das 35 marcas de whisky é a seguinte

Variable : Categoria

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	Lujo	11	31.4	31.4	31.4
	Estandar	17	48.6	48.6	80.0
	Puro malta	7	20.0	20.0	100.0
	Total	35	100.0	100.0	



12. Qual é o valor médio, o valor da mediana e a variância do tempo de envelhecimento das marcas de whisky estudadas?

Resposta :

Valor médio = 9,44

Valor da mediana = 10

Variância da distribuição = 6,79

Valor mínimo = 5

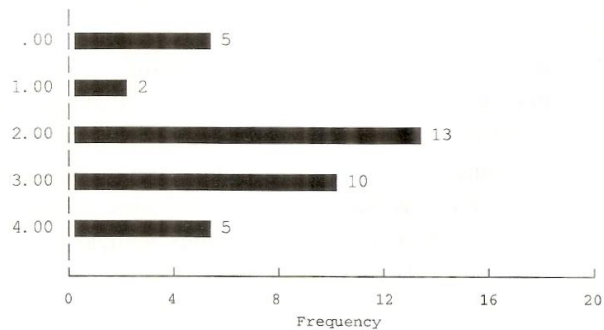
Valor máximo = 12,5

13. Qual é a distribuição da nota de apreciação das marcas de whisky pelos especialistas?

Resposta : A distribuição das notas de apreciação das 35 marcas de whisky é a seguinte :

Variable : Apreciacion de la calidad global

Value	Frequency	Percent	Valid Percent	Cum Percent
.00	5	14.3	14.3	14.3
1.00	2	5.7	5.7	20.0
2.00	13	37.1	37.1	57.1
3.00	10	28.6	28.6	85.7
4.00	5	14.3	14.3	100.0
Total	35	100.0	100.0	





### III.2. Segunda parte : A aplicação do método de A.C.P. à tabela de Dados e interpretação dos resultados

1. *Quais são as variáveis homogêneas que serão eleitas como variáveis ativas da análise como objetivo de comparar as marcas de whisky?*

*Resposta :* Escolhem-se as quatro variáveis ativas seguintes: Preço, Percentagem de Malte, Tempo de envelhecimento e Apreciação da qualidade pelos especialistas (Precio, porcentaje de malta, Tiempo de anejamento e Apreciación de la calidad por los expertos)

A variável Preço é eleita como uma variável ativa porque se trata de estudar a relação qualidade/preço das marcas de whisky . As outras três variáveis constituem três parâmetros diferentes de avaliação da qualidade das marcas de whisky. As duas primeiras medem duas dimensões diferentes, mas em princípio intimamente relacionadas com a qualidade de um whisky. A variável Apreciação representa uma avaliação global desta qualidade.

2. *A variável Categoria, é uma variável contínua? Por que deve ser definida como variável suplementar?*

*Resposta :* Não, a variável Categoria é uma variável categórica que apresenta três modalidades (Luxo, Normal e Puro malte). Esta variável recebe o estatuto de variável suplementar nesta análise por duas razões:

- Em primeiro lugar, porque um dos objetivos da análise é precisamente verificar a relação que existe entre a Categoria atribuída a cada marca de whisky ( por razões de marketing) e a relação qualidade/preço das mesmas. Podemos considerar então a variável Categoria como “variável explicada”(em conseqüência, essa variável é definida como suplementar, para ser projetada no espaço das variáveis ativas, sem ter contribuído à inércia total desse espaço), enquanto que as variáveis Preço, Percentagem de Malte, Envelhecimento e Apreciação constituem as “variáveis explicativas”(variáveis ativas).

- Por outra parte, a variável Categoria não corresponde a nenhuma avaliação (analítica ou global) da qualidade das marcas de whisky. Trata-se de um rótulo que é outorgado a cada marca, com critérios puramente comerciais. De modo que essa variável é conceitualmente não homogênea com o grupo de variáveis ativas. Mas também, a escala de medida desta variável é diferente da escala de medida do grupo de variáveis ativas (todas variáveis contínuas).

3. *Que tipo de Análise em Componentes Principais deve ser utilizado neste caso? Aquele que consiste na decomposição fatorial da inércia da matriz de correlações entre as variáveis ativas? O aquele que consiste na decomposição fatorial da inércia da matriz de variâncias/covariâncias entre as variáveis ativas?*

*Resposta :* Devemos realizar a ACP a partir da matriz de inércia definida pela matriz de correlações ente as variáveis ativas.

As variáveis ativas são conceitualmente homogêneas (cf. &III.2.1.), mas apresentam unidades de medida diferentes, o que influi sobre o valor das variâncias das distribuições que são também muito diferentes.

Para neutralizar o efeito das unidades de medida das variáveis sobre as comparações entre as marcas de whisky, é conveniente realizar essas comparações entre as marcas de whisky, é conveniente realizar essas comparações a partir dos valores das variáveis centradas e reduzidas. Esta transformação dos dados leva a criar uma matriz de inércia (que resume essas comparações) que não é mais que a matriz de correlações entre as variáveis.

4. Quantos eixos fatoriais devem ser conservados na interpretação dos resultados desta análise?

Resposta : Os dois primeiros eixos fatoriais.

O Primeiro Plano Fatorial absorve o 76% da inércia total.

O terceiro eixo fatorial absorve os 16% suplementares da inércia total, mas essa baixa correção complementar corresponde a um eixo de baixo grau de generalidade, como se pode observar nos resultados tabelados dos pontos - marcas de whisky.

5. Quais são as marcas de whisky que são pouco contributivas à inércia projetada sobre o plano I - II e que têm uma má representação nesse plano fatorial?

Resposta : Trata-se das marcas "4", "11" e "23".

INDIVIDUOS			COORDENADAS		CONTRIB.		COS. CUAD.	
P.REL	DISTO		1	2	1	2	1	2
4	2.86	2.01	-0.39	0.31	0.2	0.3	0.08	0.05
11	2.86	1.42	-0.06	0.14	0.0	0.1	0.00	0.01
23	2.86	1.51	0.47	-0.10	0.3	0.0	0.15	0.01

Essas três marcas de whisky apresentam coordenadas muito baixas, contribuição nula ou muito baixa e uma qualidade de representação no plano inferior ou igual a 16%.

6. Quais são as características comuns das marcas "16", "20" e "26", colocadas em evidência pela proximidade desses pontos nesse plano?

Resposta :

Marca	Precio	Graduación	Añejamiento	Apreciación
«16»	73	40	10,5	4
«20»	87	40	8,5	4
«26»	113	45	12	4
<b>Val. Medios</b>	<b>85.7</b>	<b>47.4</b>	<b>9.4</b>	<b>2.2</b>

Trata-se de marcas de whisky que têm recebido dos especialistas, maior nota de apreciação da qualidade. A percentagem de malte dessas marcas é inferior à percentagem de valor médio das 35 marcas. O preço médio dessas três marcas de whisky é superior ao valor médio do preço das 35 marcas, mas a diferença de preços entre essas três marcas é considerável. A marca "26" é quase 50% mais cara do que a marca "16". Igualmente, o tempo médio de envelhecimento dessas três marcas é superior ao valor médio do tempo de envelhecimento das 35 marcas, mas a diferença de tempos de envelhecimento entre elas é considerável.

7. Quais são as características comuns das marcas "30", "33" e "34", colocadas em evidência pela proximidade desses pontos nesse plano?

Resposta :

Marca	Precio	Graduación	Añejamiento	Apreciación
«30»	160	100	12	3
«33»	100	100	10	3
«34»	100	100	11	3
<b>Val. Medios</b>	<b>85.7</b>	<b>47.4</b>	<b>9.4</b>	<b>2.2</b>

Trata-se de três marcas de whisky que apresentam um preço, uma percentagem de malte e um tempo de envelhecimento sistematicamente superiores aos valores médios correspondentes das 35 marcas. A apreciação da qualidade desses whiskys pelos especialistas recebeu um boa pontuação (3), mas não a nota máxima.



8. *Quais são as características comuns das marcas "2", "3" e "8", colocadas em evidência pela proximidade desses pontos nesse plano?*

Resposta :

Marca	Precio	Graduación	Añejamiento	Apreciación
«2»	60	20	5	2
«3»	65	20	7,5	2
«8»	55	30	5	2
<b>Val. Medios</b>	<b>85.7</b>	<b>47.4</b>	<b>9.4</b>	<b>2.2</b>

Trata-se de três marcas de whisky que apresentam um preço, uma porcentagem de malte e um tempo de envelhecimento sistematicamente muito inferiores aos valores médios correspondentes às 35 marcas. A apreciação da qualidade dessas marcas pelos especialistas recebeu uma pontuação regular (2), correspondente à nota média.

9. *Quais são as características comuns das marcas "6", "7" e "9", colocadas em evidência pela proximidade desses pontos nesse plano?*

Resposta:

Marca	Precio	Graduación	Añejamiento	Apreciación
6	73	30	5	0
7	70	30	8	0
9	77	30	5,5	0
<b>Val. Medios</b>	<b>85.7</b>	<b>47.4</b>	<b>9.4</b>	<b>2.2</b>

Trata-se de três marcas que apresentam um preço e uma porcentagem de malte inferiores aos valores médios correspondentes das marcas, mas superiores aos valores correspondentes das marcas "2", "3" e "8". O tempo de envelhecimento, comparável ao das marcas "2", "3" e "8", é também inferior ao tempo de valor médio. Mas essas marcas de whisky receberam a prior notade apreciação da qualidade (0) por parte dos especialistas.

10. *Por que é que o gráfico desse plano fatorial permite ver que as variáveis Preço, Proporção de Malte e Envelhecimento ("Precio", "Proporción de Malta" e "Anejamento") apresentam correlações positivas e relativamente elevadas entre elas?*

Resposta : No primeiro plano fatorial, os vetores - variáveis "Preço" e "Proporção de Malte" apresentam uma norma elevada (o extremo de ambos vetores se coloca relativamente perto do círculo de raio 1) e definem um ângulo agudo muito pequeno. Por outro lado, o vetor - variável "Envelhecimento" apresenta uma norma algo mais reduzida e define com os dois vetores - variáveis anteriores ângulos agudos (correlações positivas), mas de maior importância.

As correlações entre essas variáveis são as seguintes:

	Prec	Grad.	Añej.
Preci	1.00		
Grad.	0.66	1.00	
Añej.	0.48	0.39	1.00

11. *Por que é que o gráfico desse fatorial permite ver que a variável "Apreciação" apresenta uma correlação positiva, mas não muito elevada, com as variáveis "Preço", "Proporção de Malte" e "Envelhecimento"?*

Resposta : Porque os ângulos formados por esse vetor - variável com cada um dos vetores - variáveis "Preço", "Proporção de Malte" e "Envelhecimento" são ângulos agudos (correlações positivas), mas relativamente grandes (baixas correlações).

As correlações entre essas variáveis são as seguintes :

	Prec	Grad.	Añej.
Aprec.	0.32	0.26	0.30

12. Podemos considerar que o primeiro eixo fatorial é uma “boa” combinação linear das quatro variáveis ativas?

Resposta : Não. As variáveis “Preço”, “Proporção de Malte” e “Envelhecimento” apresentam uma boa correlação positiva com o primeiro eixo (ângulo agudo e pequeno) e os extremos dos vetores-variáveis correspondentes estão situados próximos ao círculo de correlações. O vetor-variável “Apreciação” apresenta uma correlação positiva, mas mais baixa com esse eixo (ângulo agudo com o primeiro eixo fatorial, embora maior). Por outro lado esta variável apresenta uma alta correlação positiva com o segundo eixo fatorial.

COORDENADAS DAS VARIÁBLES SOBRE O EIXO I

VARIÁBLES	COORD.	CORRELACION VARIABLE-FACTOR
IDEN - Etiqueta corta	1	1
Prec - Precio	0.86	0.86
Grad - Graduación	0.81	0.81
Añej - Añejamiento	0.73	0.73
Apre - Apreciación	0.57	0.57

13. Quais são as variáveis melhor correlacionadas com o segundo eixo fatorial?

Resposta : Só a variável “Apreciação”. A Tabela de coordenadas-correlações com os eixos das variáveis mostra que a variável “Apreciação” apresenta uma correlação com o segundo eixo de 0,8.

COORDENADAS DAS VARIÁBLES SOBRE O EIXO II

VARIÁBLES	COORD.	CORRELACION VARIABLE-FACTOR
IDEN - Etiqueta corta	2	2
Prec - Precio	-0.23	-0.23
Grad - Graduación	-0.33	-0.33
Añej - Añejamiento	0.02	0.02
Apre - Apreciación	0.80	0.80

14. Que conclusões pode obter de uma “leitura” da estrutura da matriz de correlações das variáveis?

resposta : A avaliação da qualidade das marcas de whisky feita pelos especialistas parece ser relativamente independente do preço das mesmas (em todo caso fracamente relacionada com o...), mas também da percentagem de malte que contém e do tempo de envelhecimento do whisky.

O gráfico do primeiro plano Fatorial no espaço das variáveis, que figura no Anexo, mostra que um direção que poderia ser construída como uma “boa”



combinação linear das variáveis “Preço”, “Envelhecimento” e “percentaje de malte”(designada no gráfico com o nome “Precio/criterios objetivos de la calidad”) apresenta um ângulo quase reto (a) com a direção colinear ao vetor “apreciação”.

15. *Em função da “leitura” da estrutura dessa matriz de correlações das variáveis, como convém interpretar a orientação dos pontos - marcas de whisky sobre o primeiro plano fatorial correspondente a essa nuvem de pontos?*

*Resposta :* Em razão da estrutura das variáveis no Primeiro Plano Fatorial, convém adotar duas orientações principais, tais como está representado no gráfico desse plano no espaço das variáveis:

- Uma orientação segue a direção do vetor - variável “Apreciação”. No gráfico que figura em anexo, está representada essa orientação com o nome “evaluacion global de la calidad” (avaliação global da qualidade). Que isto dizer que os pontos - marcas de whisky que se encontram localizados próximos a essa direção, na extremidade superior do primeiro quadrante do plano, devem ser marcas de whisky que apresentam a melhor apreciação por parte dos especialistas. No outro extremo dessa direção, na parte inferior do terceiro quadrante, os pontos marcas de whisky situados ali devem ser marcas de whisky que apresentam a pior apreciação por parte dos especialistas. Projetando os pontos - marcas whisky sobre essa direção, podemos ordená-los por ordem crescente da apreciação da qualidade das mesmas pelos especialistas.

- A outra orientação segue, praticamente, a direção do vetor - variável preço (“Precio”). No gráfico que figura em anexo, está representada essa orientação com o nome Preço/critério objetivos da qualidade (“Precio/criterios objetivos de la calidad”). Quer isto dizer, que os pontos - marcas de whisky que se encontram localizados próximos a essa direção, no quarto quadrante do plano, devem ser marcas de whisky que apresentam um preço elevado, uma e percentagem de malte e um tempo de envelhecimento superiores também aos valores médios. No outro extremo dessa direção, no segundo quadrante, os pontos marcas de whisky situados ali devem ser marcas de whisky que apresentam um baixo preço, uma percentagem de malte e um tempo de envelhecimento inferiores aos valores médios.

Projetando os pontos - marcas de whisky sobre essa direção, podemos ordená-los no ordem crescente da relação preço/qualidade objetiva (“precio/calidad objetiva”).

### IV.3. Terceira parte: Conclusões da análise e retorno aos dados brutos

1. *Quais são as marcas de whisky que apresentam a relação qualidade/preço mais interessante para o consumidor?*

*Resposta :* O consumidor prefere uma marca de whisky de boa qualidade ao menor preço possível. De modo que tem interesse em escolher as marcas que tenham recebido a maior nota de qualidade global outorgada pelos especialistas e que saiam para o mercado ao menor preço possível.

As marcas que respondem a esses critérios de seleção são: o whisky “16” e o whisky “20”. Devemos ter em conta que as marcas “26”, “31” e “29”, têm também uma boa nota de avaliação global de qualidade, mas seu preço é já superior ao das marcas “16” e “20”.

Na tabela seguinte, apresentamos por ordem crescente de preços, todas as marcas que obtiveram a maior pontuação de qualidade global.



Marca	Precio	Graduación	Categoría	Añejamiento	Apreciación
<<16>>	73	40	2	10.5	4
<<20>>	87	40	2	8.5	4
<<31>>	90	100	3	12	4
<<26>>	113	45	2	12	4
<<29>>	127	100	3	8.5	4
<b>Medias</b>	<b>85.7</b>	<b>47.4</b>	<b>-</b>	<b>9.4</b>	<b>2.2</b>

2. *Quais são as marcas de whisky que apresentam a relação qualidade/preço menos interessante para o consumidor?*

*Resposta:* A marca "35". Trata-se de uma das marcas mais caras, que recebeu a pior pontuação de avaliação global (0).

3. *Compraria você um whisky das marcas "30", "33" o "34"?*

*Resposta:* Não porque se trata de marcas caras (a marca "30" é a mais cara), mas que receberam uma pontuação de avaliação de sua qualidade global (3) que indica que as mesmas são de boa qualidade, mas não as melhores.

Marca	Precio	Graduación	Categoría	Añejamiento	Apreciación
<<30>>	160	100	3	12	3
<<33>>	100	100	3	10	3
<<34>>	100	100	3	11	3
<b>Medias</b>	<b>85.7</b>	<b>47.4</b>	<b>-</b>	<b>9.4</b>	<b>2.2</b>

4. 1. *As marcas de whisky mais baratas, são também as marcas de pior qualidade?*

*Resposta:* Não As marcas de whisky mais baratas ("8", "2" e "13") apresentam uma pontuação de avaliação da qualidade global de valor 2 e incluso, no caso da marca "13", de valor 3, isto é, comparável à marca "30", a mais cara.

Mas também se verifica que essas marcas, as mais baratas, apresentam valores inferiores aos valores médios de graduação de malte e de tempo de envelhecimento.

Marca	Precio	Graduación	Categoría	Añejamiento	Apreciación
<<8>>	55	30	1	5	2
<<2>>	60	20	1	5	2
<<13>>	62	33	2	8	3
<b>Medias</b>	<b>85.7</b>	<b>47.4</b>	<b>-</b>	<b>9.4</b>	<b>2.2</b>

5. 1. *Que relação observa entre a qualidade das marcas de whisky e a Categoria atribuída às mesmas por razões de marketing?*

*Resposta:* As marcas de whisky "Puro Malte" são as mais caras e aquelas que apresentam uma maior graduação em malte. Mas o tempo de envelhecimento dessas marcas é muito variável. A apreciação da qualidade das mesmas por parte dos especialistas varia entre a pior (nota 0) para a marca "35" e a melhor (nota 4) para as marcas "31" e "29".

As marcas de whisky "Normal" apresentam valores próximos dos valores médios de preço e graduação de malte. Mas também apresentam performances muito diferentes no que diz respeito ao tempo de envelhecimento e à avaliação global da qualidade.

As marcas de whisky “Luxo” são aquelas que apresentam valores inferiores aos valores médios de graduação em malte, tempo de envelhecimento e preço.

De modo que a Categoria de uma marca de whisky parece ser um rótulo outorgado pela Câmara de Produtores seguindo um critério diretamente ligado à graduação em malte do produto. Mas, a afetação comercial de uma categoria a uma marca dada, parece determinar o preço da mesma no mercado.

A Categoria que figura no rótulo de um whisky não corresponde à qualidade do produto.

ANEXO

Descomposición da inércia global

EDICION DE LOS VALORES PROPIOS

TRAZA ANTES DE LA DIAGONALIZACION ... 4.0000  
 SUMA DE LOS VALORES PROPIOS .... 4.0000

HISTOGRAMA DE LOS 4 PRIMEROS VALORES PROPIOS

NUMERO	VALOR PROPIO	PORCENT.	PORCENT. ACUMUL.
1	2.2333	55.83	55.83
2	0.8065	20.16	75.99
3	0.6295	15.74	91.73
4	0.3307	8.27	100.00

Coordenadas, contribuições e qualidade de representação das variáveis ativas e dos individuos

COORDENADAS DE LAS VARIABLES SOBRE LOS EJES 1 A 4

VARIABLES	COORDENADAS				CORR. VARIABLE-FACTOR			
	1	2	3	4	1	2	3	4
IDEN - ETIQUETA								
VARIABLES ACTIVAS								
Prec - Precio	0.86	-0.23	-0.14	0.43	0.86	-0.23	-0.14	0.43
Grad - Graduacion	0.81	-0.33	-0.33	-0.36	0.81	-0.33	-0.33	-0.36
Anej - Anejamiento	0.73	0.02	0.68	-0.09	0.73	0.02	0.68	-0.09
Apre - Apreciacion	0.57	0.80	-0.19	-0.02	0.57	0.80	-0.19	-0.02

COORDENADAS, CONTRIBUCIONES Y COSENOS CUADRADOS DE LOS INDIVIDUOS SOBRE LOS EJES 1 A 4

INDIVIDUOS			COORDENADAS				CONTRIBUCIONES				COSENOS CUADR.			
IDENTIFICADOR	P.REL	DISTO	1	2	3	4	1	2	3	4	1	2	3	4
1	2.86	5.22	-1.62	1.12	-1.12	0.30	3.4	4.4	5.7	0.8	0.50	0.24	0.24	0.02
2	2.86	5.92	-2.23	0.50	-0.83	-0.06	6.4	0.9	3.1	0.0	0.84	0.04	0.12	0.00
3	2.86	2.80	-1.61	0.46	-0.03	-0.02	3.3	0.7	0.0	0.0	0.92	0.08	0.00	0.00
4	2.86	2.01	-0.39	0.31	1.32	-0.08	0.2	0.3	8.0	0.1	0.08	0.05	0.87	0.00
5	2.86	2.67	-0.19	1.11	1.16	-0.26	0.0	4.4	6.1	0.6	0.01	0.46	0.50	0.03
6	2.86	7.44	-2.28	-1.29	-0.69	0.27	6.7	5.9	2.1	0.6	0.70	0.23	0.06	0.01
7	2.86	4.88	-1.80	-1.23	0.35	-0.04	4.2	5.3	0.5	0.0	0.67	0.31	0.02	0.00
8	2.86	6.04	-2.18	0.44	-0.93	-0.48	6.1	0.7	4.0	2.0	0.79	0.03	0.14	0.04
9	2.86	6.56	-2.07	-1.34	-0.56	0.39	5.5	6.4	1.4	1.3	0.65	0.28	0.05	0.02
10	2.86	4.95	-0.37	-1.50	1.49	0.60	0.2	8.0	10.0	3.1	0.03	0.45	0.45	0.07
11	2.86	1.42	-0.06	0.14	1.18	0.11	0.0	0.1	6.3	0.1	0.00	0.01	0.98	0.01
12	2.86	3.16	-1.62	-0.58	-0.43	0.07	3.4	1.2	0.9	0.0	0.83	0.11	0.06	0.00
13	2.86	2.52	-1.03	1.07	-0.24	-0.50	1.4	4.1	0.3	2.2	0.42	0.46	0.02	0.10
14	2.86	1.64	0.47	0.77	0.88	0.21	0.3	2.1	3.5	0.4	0.13	0.37	0.48	0.03
15	2.86	0.43	-0.46	0.10	0.46	-0.03	0.3	0.0	1.0	0.0	0.48	0.03	0.50	0.00
16	2.86	2.83	0.23	1.60	0.19	-0.43	0.1	9.0	0.2	1.6	0.02	0.90	0.01	0.07
17	2.86	0.28	-0.38	-0.10	-0.20	0.29	0.2	0.0	0.2	0.7	0.51	0.03	0.14	0.31
18	2.86	0.36	-0.58	0.00	-0.13	0.02	0.4	0.0	0.1	0.0	0.95	0.00	0.05	0.00
19	2.86	0.11	-0.25	-0.06	0.16	0.15	0.1	0.0	0.1	0.2	0.54	0.03	0.22	0.21
20	2.86	2.43	0.26	1.39	-0.61	0.23	0.1	6.9	1.7	0.5	0.03	0.80	0.15	0.02
21	2.86	0.20	-0.39	0.01	0.20	-0.04	0.2	0.0	0.2	0.0	0.79	0.00	0.21	0.01
22	2.86	1.15	-0.62	-0.78	0.38	0.11	0.5	2.2	0.6	0.1	0.34	0.53	0.12	0.01
23	2.86	1.51	0.47	-0.10	1.12	0.16	0.3	0.0	5.7	0.2	0.15	0.01	0.83	0.02
24	2.86	2.96	1.28	0.37	0.57	0.93	2.1	0.5	1.5	7.5	0.55	0.05	0.11	0.29
25	2.86	2.60	-0.95	-0.12	-1.21	0.48	1.1	0.1	6.6	2.0	0.35	0.01	0.56	0.09
26	2.86	5.07	1.78	1.01	0.26	0.90	4.1	3.6	0.3	7.0	0.63	0.20	0.01	0.16
27	2.86	1.63	0.97	0.49	0.62	0.27	1.2	0.9	1.7	0.6	0.57	0.15	0.23	0.05
28	2.86	1.39	0.56	0.68	0.74	-0.27	0.4	1.6	2.5	0.6	0.22	0.33	0.40	0.05
29	2.86	10.54	2.62	0.05	-1.88	0.38	8.8	0.0	16.1	1.2	0.65	0.00	0.34	0.01
30	2.86	19.47	3.94	-1.11	-0.80	1.46	19.8	4.3	2.9	18.3	0.80	0.06	0.03	0.11
31	2.86	6.93	2.20	0.57	-0.37	-1.27	6.2	1.2	0.6	14.0	0.70	0.05	0.02	0.23
32	2.86	4.73	1.45	-0.87	0.07	-1.37	2.7	2.7	0.0	16.1	0.45	0.16	0.00	0.40
33	2.86	4.74	1.80	-0.32	-0.93	-0.73	4.1	0.4	3.9	4.6	0.68	0.02	0.18	0.11
34	2.86	5.03	1.99	-0.32	-0.59	-0.79	5.1	0.4	1.6	5.4	0.78	0.02	0.07	0.13
35	2.86	8.38	1.08	-2.48	0.40	-0.96	1.5	21.7	0.7	8.0	0.14	0.73	0.02	0.11

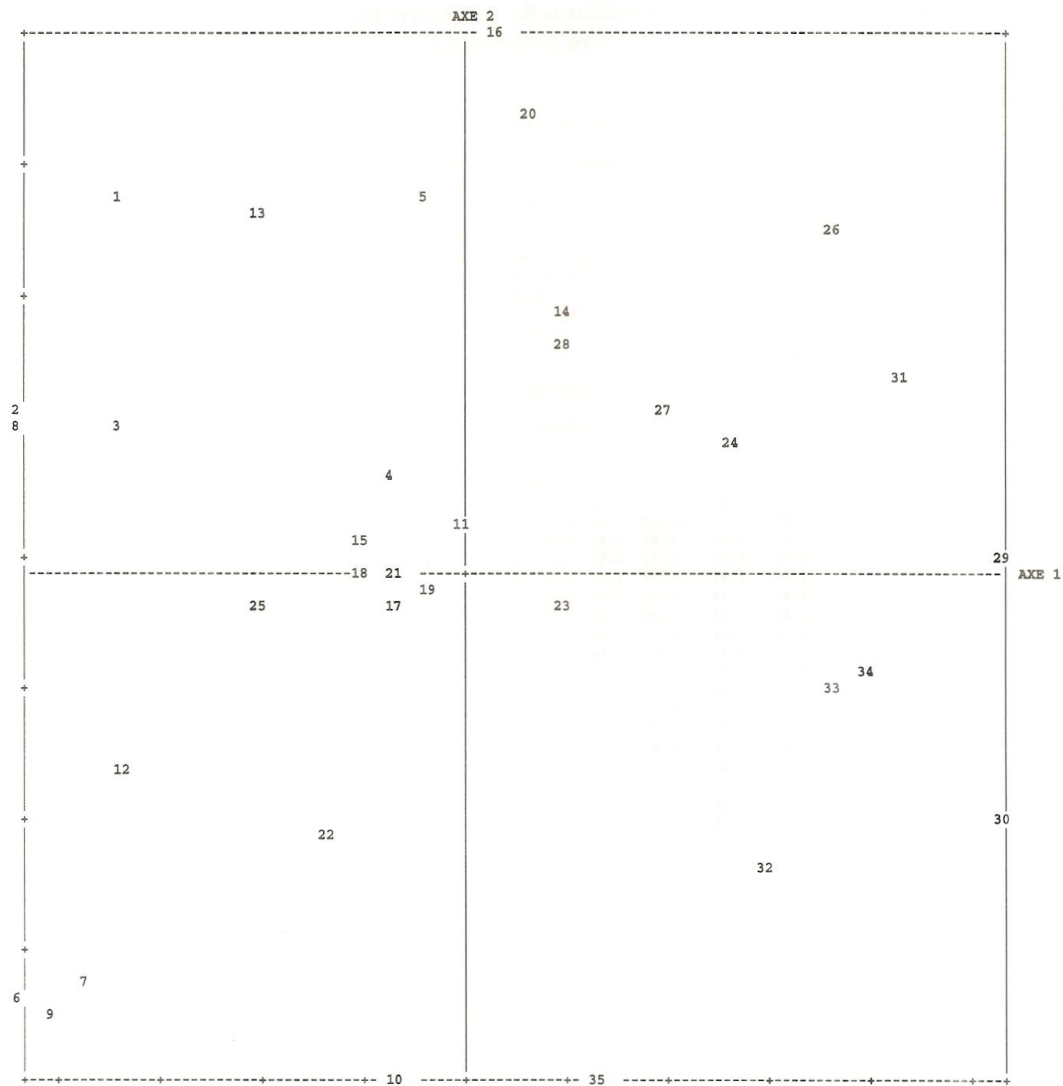


## Coordenadas e valores das modalidades das variáveis suplementares

COORDENADAS Y VALORES-TEST DE LAS MODALIDADES SOBRE LOS EJES 1 A 4

MODALIDADES				VALORES-TEST					COORDENADAS					DISTO.
IDEN	ETIQUETA	EPEC.	P.ABS	1	2	3	4	0	1	2	3	4	0	
3 . Categor a														
Cat1	- Lujo	11	11.00	-3.6	-0.5	0.6	0.5	0.0	-1.35	-0.12	0.12	0.07	0.00	1.84
Cat2	- Estandar	17	17.00	-0.1	2.1	1.2	1.5	0.0	-0.02	0.34	0.16	0.15	0.00	0.16
Cat3	- Puro malta	7	7.00	4.2	-2.1	-2.2	-2.4	0.0	2.15	-0.64	-0.59	-0.47	0.00	5.61

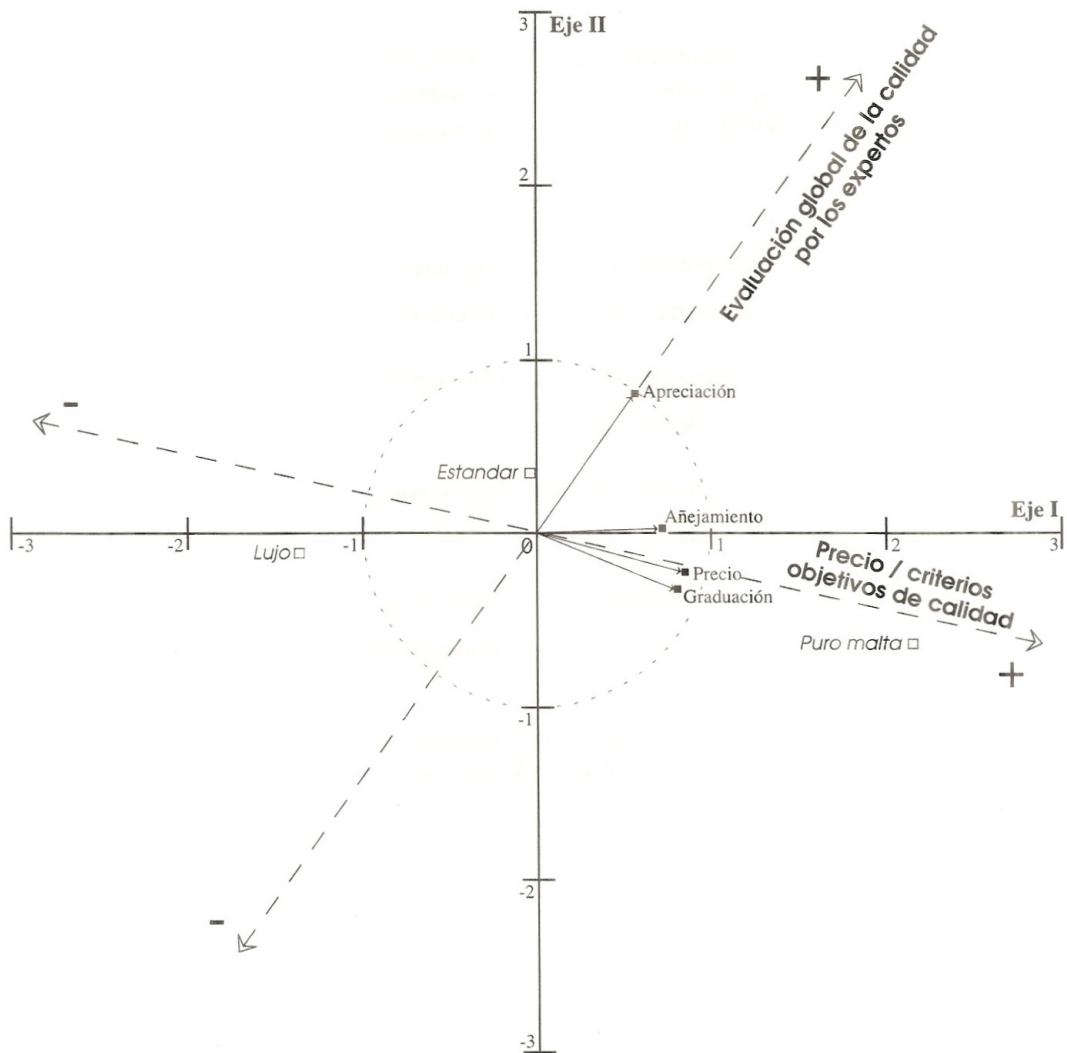
### Primeiro Plano Fatorial no espaço dos indivíduos



Matriz de correlações

	Prec	Grad	A ej	Apre
Prec	1.00			
Grad	0.66	1.00		
A ej	0.48	0.39	1.00	
Apre	0.32	0.26	0.30	1.00

Primeiro Plano Fatorial no espaço dos indivíduos



## COMENTÁRIOS

### a) Significado dos valores - teste

Para caracterizar a relação existente entre o  $x$ -ésimo eixo fatorial de uma ACP e a  $k$ -ésima modalidade de uma variável nominal suplementar de um estudo, podemos proceder de uma maneira muito similar ao caso de um clássico teste de hipótese.

Se formulamos as seguintes hipóteses:

$H_0$  : Os  $n_k$  valores  $F_\alpha(i)$ , coordenadas dos indivíduos sobre o eixo  $x$ , foram extraídos ao acaso sobre os  $n$  indivíduos. É dizer que  $H_0 : \mu = \bar{x}_k = \bar{x}$ ,

$H_1$  : Os  $n_k$  valores  $F_\alpha(i)$ , coordenadas dos indivíduos sobre o eixo  $x$ , não foram extraídos ao acaso sobre os  $n$  indivíduos. É dizer que  $H_1 : \mu \neq \bar{x}_k \neq \bar{x}$ .

Se consideramos que esses valores constituem uma “amostra” aleatória simples, sem reposição, de  $n_k$  observações entre os  $n$  indivíduos, podemos então calcular a probabilidade de observar um valor “médio da amostra” que seja tão extremo como o observado na classe  $k$  de indivíduos.

Chamamos:

- $\bar{x}$  : ao valor médio das coordenadas sobre o eixo  $x$ , dos  $n$  indivíduos,  $\bar{x} = \frac{1}{n} \sum_n F_\alpha(i) = 0$ , já que os eixos fatoriais são variáveis centradas por construção.
- $\sigma^2$  : a variância das coordenadas sobre o eixo  $\alpha$ , dos  $n$  indivíduos, isto é, a variância da  $\alpha$ -ésima componente principal. De modo que  $\sigma^2 = \lambda_\alpha$ .
- $n_k$  : o número de indivíduos que apresentaram a  $k$ -ésima modalidade da variável nominal suplementar considerada.
- $\bar{x}_k$  : o valor médio das coordenadas sobre o eixo  $\alpha$  dos  $n_k$  indivíduos que apresentaram a  $k$ -ésima modalidade da variável nominal suplementar considerada,  $\bar{x}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} F_\alpha(i)$ .

Sob a hipótese de que os valores das coordenadas dos indivíduos da classe que nos interessa foram selecionados ao acaso, a variável aleatória  $M_k$  se ajusta a uma variável normal de parâmetros  $M_k \equiv N(\bar{x}, \sigma^2)$ .

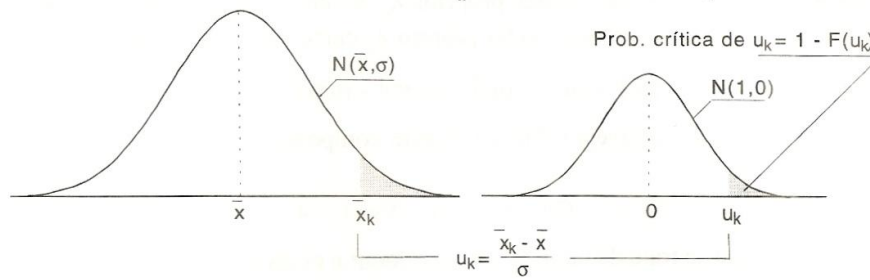
Sendo :

$$E_{H_0}(M_k) = \bar{x} \quad \text{e} \quad \text{Var}_{H_0}(M_k) = \frac{\lambda}{n_k} \cdot \frac{n - n_k}{n - 1} = \sigma^2.$$

Centrando e reduzindo a variável aleatória  $M_k$ , podemos definir uma variável  $U_k$ , que será uma variável aleatória normal reduzida  $u_k = \frac{m_k - \bar{x}}{\sigma}$ ,  $U_k \equiv N(0,1)$ .



Se  $n$  e  $n_k$  são suficientemente grandes, se pode demonstrar que a probabilidade associada à ocorrência de um valor dado de  $M_k$  é igual à probabilidade que a lei normal supere o valor  $u_k$  calculado sobre o “valor médio da amostra” dos  $n_k$  indivíduos da classe que nos interessa.



O valor  $u_k = \frac{\bar{x}_k - \bar{x}}{\sigma}$  é chamado “valor de prova” ou “valor - teste”.

É dizer que:  $G_\alpha(k) = u_k \times \sigma$ , valor da prova se pode interpretar como a distância - em número de desvios padrão - entre o ponto representativo da  $k$ -ésima modalidade da variável nominal que nos interessa sobre o eixo  $\alpha$  e a origem dos eixos fatoriais.

De modo que, se a uma modalidade de uma variável nominal suplementar lhe corresponde um “valor - teste” superior (inferior) ou igual a 1,98 (ou -1,98) sobre um eixo  $\alpha$ , deve-se rejeitar  $H_0$  e conservar  $H_1$ . Concluimos então que essa modalidade está “bem representada” sobre o eixo  $\alpha$ .

Por exemplo, considerando os valores da tabela de “valores - teste” das modalidades da variável “Categoria”...

COORDENADAS Y VALORES-TEST DE LAS MODALIDADES SOBRE LOS EJES 1 A 3

MODALIDADES		VALORES-TEST			COORDENADAS			DISTO.
IDEN - ETIQUETA	EFF. P.ABS	1	2	3	1	2	3	
3 . Categoria								
Lujo - Lujo	11 11.00	-3.6	-0.5	0.6	-1.35	-0.12	0.12	1.84
Esta - Estandar	17 17.00	-0.1	2.1	1.2	-0.02	0.34	0.16	0.16
PuMa - Puro malta	7 7.00	4.2	-2.1	-2.2	2.15	-0.64	-0.59	5.61

$n$  : numero de whiskys do estudo.

$n_k$  : numero de whiskys que apresentam a modalidade «Luxo» = 11

$x_k$  : «valor médio da coordenada» dos  $n_k$  whiskys sobre o eixo I = -1.35

$$\sigma^2 = \frac{\lambda_1}{n_k} \cdot \frac{n - n_k}{n - 1} = \frac{2.2333}{11} \cdot \frac{35 - 11}{35 - 1} = 0.1433$$

$$\sigma = 0.3786$$

$$\text{valor - test}_1(\text{Lujo}) = \frac{x_k}{\sigma} = \frac{-1.35}{0.3786} = -3.6$$

A modalidade ‘Luxo’ da variável “Categoria” está bem representada sobre o eixo I.

### b) Significado das coordenadas sobre os eixos unitários

Na tabela seguinte, as colunas chamadas “antigos eixos unitários” são os vetores - próprios unitários associados aos valores próprios  $\lambda_k$ . Se chamamos  $u_1$  - por exemplo - ao vetor próprio unitário associado ao primeiro vetor próprio  $\lambda_1$ , então verifica-se que:

$$\|u\| = (u_1^1)^2 + (u_2^1)^2 + (u_3^1)^2 + (u_4^1)^2 = (0.57)^2 + (0.54)^2 + (0.49)^2 + (0.38)^2 = 1$$

As coordenadas das variáveis sobre a primeira componente principal se obtém da seguinte maneira:

$$G_1(j) = \sqrt{\lambda_1} \cdot u_j^1 \quad \forall j = 1, \dots, 4$$

Nesta caso, como se trata de uma Análise de Componentes principais normada e reduzido (a matriz de inércia é a matriz de correlação das quatro variáveis ativas”), as coordenadas das variáveis sobre cada componente principal são iguais à correlação das variáveis com cada componente principal.

Por exemplo, a coordenada da variável ‘preço’, sobre a primeira componente principal (equivalente à correlação da variável ‘preço’ com a primeira componente”), é igual a:

$$G_1(\text{Precio}) = \sqrt{2.2333} \cdot (0.57) = 0.86$$

COORDENADAS DE LAS VARIABLES SOBRE LOS EJES 1 A 3

VARIABLES	COORDENADAS			CORR. VAR-FACTOR			ANTIGUOS EJES UNITARIOS		
	1	2	3	1	2	3	1	2	3
VARIABLES ACTIVES									
Prec - Precio	0.86	-0.23	-0.14	0.86	-0.23	-0.14	0.57	-0.26	-0.18
Grad - Graduaci n	0.81	-0.33	-0.33	0.81	-0.33	-0.33	0.54	-0.37	-0.41
A ej - A ejamiento	0.73	0.02	0.68	0.73	0.02	0.68	0.49	0.02	0.86
Apri - Apresiasi n	0.57	0.80	-0.19	0.57	0.80	-0.19	0.38	0.89	-0.24

### c) Significado dos intervalos de Anderson

Seja uma amostra de  $n$  indivíduos, de uma variável aleatória normal  $p$ -dimensional  $N_p(u, \Delta)$ , e seja  $V^* = \frac{n}{n-1} \cdot V$  a matriz corrigida de variâncias dessa amostra.

Se  $\gamma_i$  e  $\lambda_i$  designam os  $i$ -ésimos valores próprio de  $\Delta$  e de  $V^*$  respectivamente, T. ANDERSON demonstron que  $\sqrt{n-1}(\lambda_i - \gamma_i)$  converge para uma lei normal  $LG(0, \gamma_i \sqrt{2})$ . De modo que  $\ln \gamma_i$  tem uma distribuição aproximada com  $LG\left(\ln \gamma_i, \sqrt{\frac{2}{n-1}}\right)$ , o que permite afirmar que a probabilidade associada ao seguinte intervalo é de 0.95 :

$$P\left[\ln \gamma_i - 1.96 \sqrt{\frac{2}{n-1}} < \ln \lambda_i < \ln \gamma_i + 1.96 \sqrt{\frac{2}{n-1}}\right] = 0.95 .$$

(\*) in : SAPORT *Analyse des données et statistique*», Ed. Technip, Paris, 1990, pp.272-273.

Em conseqüência:

$$\lambda_i \cdot \exp\left(-1.96\sqrt{\frac{2}{n-1}}\right) < \ln \gamma_i < \lambda_i \cdot \exp\left(1.96\sqrt{\frac{2}{n-1}}\right) \text{ ao n nível de confiança } \alpha = 0.95.$$

Vemos então que é possível construir um intervalo de confiança, ao nível  $\alpha = 0,95$ , para os diferentes valores próprios, empregando a fórmula de Anderson sempre que  $n$  seja suficientemente grande.

Mas vemos também que essas propriedades só podem ser aplicadas para o caso de matrizes de variância que resultem do estudo de uma variável gaussiana p-dimensional. Essas propriedades NÃO SE APLICAM no caso mais frequente de utilização da ACP, na qual a matriz de inércia é a matriz de correlações.

No entanto, o procedimento CORBI do programa SPAD.N edita por defeito os intervalos de Anderson correspondentes à sequência de valores próprios resultantes da decomposição da matriz de inércia, ainda quando se trata - como neste caso - de uma ACP normalizada

O arquivo de resultados produzido por SPAD.N contém os resultados seguintes, os quais — neste caso - não devem ser utilizados.

INTERVALOS LAPLACIANOS DE ANDERSON (UMBRAL 0.95)

NUMERO	LIMITE INFERIOR	VALOR PROPIO	LIMITE SUPERIOR
1	1.3883	2.2333	3.5925
2	0.5014	0.8065	1.2973
3	0.3913	0.6295	1.0126
4	0.2056	0.3307	0.5320

EXTENSION Y POSICION RELATIVA DE LOS INTERVALOS

1	.....*
2	.....*
3	.....*
4	*.....*