

Resumo:

Este artigo tem por objetivo familiarizar o leitor com a utilização do programa informatizado de tratamento estatístico-descritivo de dados lingüísticos (programa Stablex) e esclarecer a finalidade das análises exploratórias, nomeadamente a ACP, (Análise em Componentes Principais), que representa o momento primordial e fundamental da análise exploratória, já que abre o caminho para a Análise Fatorial Discriminante (AFD que é o momento mais fecundo da análise descritiva, paramétrica, objetiva e indutiva, o nec plus ultra da análise exploratória. E, por fim, este artigo pretende desmitificar os ditos “modelos estatísticos e matemáticos” que são à base do método e da metodologia analítica do STABLEX, cuja evocação costuma apavorar os analistas, quando, na verdade, são e fazem a riqueza de uma análise qualitativa, permitindo medir tudo e tudo definir, verificar tudo e tudo controlar.

Palavras-chave: *soft Stablex, análise em componentes principais, análise fatorial.*

Résumé:

Cet article a pour but d'abord de familiariser le lecteur dans l'utilisation de STABLEX, ensuite de l'éclairer sur les finalités des analyses exploratoires, et notamment l'ACP, qui en est le moment primordial et fondamental, puisqu'il ouvre les voies de l'AFD, qui est le moment le plus fécond de l'analyse descriptive, paramétrique, objective et inductive le nec plus ultra de l'analyse exploratoire. Enfin, cet article voudrait démystifier lesdits « modèles statistiques et mathématiques » qui sont à la base de la méthode et de la méthodologie analytique de STABLEX, mais qui font trembler les analystes à la seule évocation, alors que ces modèles sont et font la véritable richesse d'une analyse qualitative, en permettant de tout mesurer, de tout cerner, de tout définir, de tout vérifier et de tout contrôler.

Mots-clés: *soft Stablex, analyse en composantes principales, analyse factorielle.*

Introdução

Esse artigo tem por objetivo esclarecer ou talvez fortalecer a explicação relativa à função e ao funcionamento das estruturas métricas desenvolvidas pelo STABLEX – *programa informatizado de tratamento estatístico-descritivo de dados lingüísticos*, a começar pela Métrica R e a ACP (a Análise em Componentes Principais), dois pontos totalmente dependentes. A ACP tem uma importância capital na análise de qualquer tipo de *corpus*. Mas, para compreender sua importância, é necessário entrar nos arcanos da **Métrica R** para destacar o perfil mais adequado da imagem representativa das “qualidades” inerentes às variáveis do *corpus*.

A motivação deste artigo é simples. Trata-se de disponibilizar um texto, em português, a todos aqueles que têm encontrado dificuldades de leitura e compreensão das publicações em língua francesa que tratam sobre o uso e funcionamento do STABLEX. No que se refere a endereços de publicação, citaríamos nomeadamente, na internet, **Letras & Letras**, *site* que disponibiliza documentos e estudos em *Lexicometria* que podem orientar o leitor na prática do STABLEX, já que possibilitam o acesso aos espaços técnicos e práticos da análise estatística.

Inicialmente iremos concentrar nossa atenção à técnica da ACP que não é nem pode ser considerada uma mera análise “quantitativa”. Ela é uma análise *qualitativa* que serve, desde o início, quando se abre a MACRO dos dados estatísticos, para observar o posicionamento global das variáveis sob forma de uma nuvem de pontos. Os valores que determinam o referido posicionamento são “valores sem dimensão e sem medição”, ou seja, são apenas valores de distanciamento e de posicionamento que *determinam os focos de concentração* das variáveis que constituem o *corpus*.

Para que o leitor possa se familiarizar com o método e a metodologia STABLEX, abordaremos a explicação metódica e metodológica da ACP e da Métrica R a partir de dados concretos, e, sobretudo das MACROS que os grupos de pesquisa Linguagem, Sociedade, Saúde e Trabalho e o Núcleo de Estudos de Hipertexto e Tecnologia Educacional (NEHTE), da Universidade de Pernambuco (sob a responsabilidade dos professores Maria Cristina HENNES SAMPAIO e Antônio Carlos XAVIER, respectivamente) podem colocar a disposição de todos para servir de objeto de estudo e de manipulação.

No entanto, para que a explicação seja completa, é indispensável partir do quadro de correlação para compreender a função e o funcionamento da Métrica R e abordar a ACP com a confecção das imagens realmente representativas da “qualidade” associativa das variáveis que constituem o *corpus*.

É evidente que não vamos nem podemos entrar na explicação matemática das fórmulas utilizadas e na justificativa científica do método estatístico, já que todos esses aspectos são amplamente expostos na obras especializadas cujas referências bibliográficas são conhecidas.

A MACRO, ligada ao STABLEX, fornece, de imediato, as cinco primeiras páginas dos dados estatísticos e da estatística fundamental:

- a. o léxico (classificado por ordem de frequência decrescente, ou por ordem alfabética);
- b. a tabela de contingência de distribuição das frequências (TDF);
- c. a tabela de contingência dos desvios centrados reduzidos (TDR);
- d. os gráficos fundamentais;

e. a régua para determinar o peso e a densidade dos vetores lexicais.

O leitor vai descobrir, nas páginas da TDF, TDR e da Régua, as estruturas disponíveis para a criação de gráficos analíticos que abrem, logo no início, o espaço da ACP. Convém apenas abrir os dados contidos nas planilhas para ver a formação dos gráficos correspondentes, e, além disso, (conforme a perspectiva pedagógica da disponibilização das MACROS), para exercitar-se na confecção de novos gráficos.

A três últimas páginas da MACRO, embora sejam automatizadas, devem ser manipuladas para realizar os cálculos estatísticos descritivos e, portanto, os gráficos correspondentes.

Por enquanto vamos abordar a função e o funcionamento da *Correlação* e da *Métrica R*, automatizados nas páginas 6 e 7 da MACRO.

A última página 8 trata da análise da *Regressão*, especializada na Análise Fatorial Discriminante (AFD), o *ne plus ultra* da análise estatística descritiva, objetiva e indutiva.

1. A Correlação

Pouco importa a origem dos dados. O importante é observar os quadros e compreender a finalidade da análise estatística.

Pouco importa igualmente a extensão do *corpus*, e, portanto o número de colunas do quadro de correlação.

Todavia, para facilitar a observação e a manipulação, vamos abrir aqui quadros e tabelas de dimensão limitada para ver as diferenças estruturais e melhor compreender a função e o funcionamento das referidas tabelas e quadros de correlação.

1.1 Correlação de 8 variáveis

Vejamos aqui um quadro de correlação de 8 variáveis (no caso presente tirado das *Três Cidras do Amor*, publicado na internet em *Letras & Letras*)

	T1	T2	T3	T4	T5	T6	T7	T8
T1	1,000	0,895	0,751	0,864	0,883	0,804	0,828	0,826
T2	0,895	1,000	0,777	0,815	0,809	0,766	0,752	0,799
T3	0,751	0,777	1,000	0,649	0,640	0,609	0,578	0,688
T4	0,864	0,815	0,649	1,000	0,859	0,937	0,931	0,882
T5	0,883	0,809	0,640	0,859	1,000	0,878	0,918	0,824
T6	0,804	0,766	0,609	0,937	0,878	1,000	0,958	0,811
T7	0,828	0,752	0,578	0,931	0,918	0,958	1,000	0,841

T8	0,826	0,799	0,688	0,882	0,824	0,811	0,841	1,000
-----------	-------	-------	-------	--------------	-------	-------	-------	--------------

Como ler esse quadro de correlação?

O quadro fundamental, de forma triangular, é transformado em quadro “quadrado”, cuja leitura é evidente, tomando conta do ponto de convergência referencial da coluna e da linha.

1.2 Como proceder à leitura dos valores de correlação

Devemos saber que:

- o valor de correlação entre duas variáveis é igual ao valor do co-seno do ângulo formado por duas variáveis (entre duas linhas de dados da TDR).
- o grau de significação do co-seno depende do valor do ângulo formado pelas duas rectas de estimação.
- o grau de correlação entre duas variáveis depende do ângulo formado pelas duas rectas de estimação.
- o grau de correlação é, portanto, um valor trigonométrico, como se pode observar na tabela seguinte:

x	0	30°	45°	60°	90°
co-seno x = r	1	$\sqrt{3}/2$	$\sqrt{2}/2$	1/2	0
<i>valores</i>	1	0,866	0,707	0,500	0
correlação	1	+	=	-	0

- o grau de correlação é predeterminado e fixo:
 - com $0,866 \leq r \leq 1$, a correlação é significativa positiva
 - com $0,707 \leq r \leq 0,866$, a correlação é positiva
 - com $0,500 \leq r \leq 0,707$, a correlação é fraca
 - com $0 \leq r \leq 0,500$, a correlação é inexistente.

Para fixar a dependência da correlação das variáveis convém fazer uma leitura dos valores em função dos valores trigonométricos dados pela tabela trigonométrica.

Por facilitar a leitura do quadro de correlação, basta colocar em negrito os valores superiores ou iguais a 0,866 representativos de uma ligação significativa positiva entre duas variáveis.

É preciso lembrar que a correlação não é transitiva. Ela refere-se apenas às variáveis emparelhadas (duas a duas). Daí a leitura do quadro linha/coluna ou coluna/linha:

1. T1 tem uma ligação significativa com T2 e T5 ($r = 0,895$ e $0,883$)
2. T2 tem ligação significativa apenas com T1 ($r = 0,895$)
3. T3 não tem ligação significava positiva com os outros textos
4. T4 tem ligação significativa com T6, T7 e T8 ($r = 0,937$; $0,931$ e $0,882$)
5. T5 com T1 e T7 ($r = 0,883$ e $0,918$)
6. T6 com T4 e T7 ($r = 0,937$ e $0,958$)
7. T7 com T4, T5 e T6 ($r = 0,931$; $0,918$ e $0,958$)
8. T8 com T4 ($r = 0,882$)

A leitura do quadro de correlação faz-se à luz da tabela dos valores trigonométricos.

Então vemos claramente que o valor limite $r = 0,866$ serve de referência significativa positiva para o co-seno, e, portanto para a estimação do coeficiente de correlação.

De fato, os valores do co-seno são significativamente superiores aos valores do seno quando o ângulo varia de 0° até 45° , momento em que se opera a inversão da correlação em função da complementaridade do seno e do co-seno.

Essa tabela poderia ser completada com os valores complementares do seno e do co-seno para melhor observar a inversão da correlação das variáveis emparelhadas:

x	0	30°	45°	60°	90°
co-seno x	1	$\sqrt{3}/2$	$\sqrt{2}/2$	$1/2$	0
<i>valores</i>	1	0,866	0,707	0,500	0
seno x	0	$1/2$	$\sqrt{2}/2$	$\sqrt{3}/2$	1
<i>valores</i>	0	0,500	0,707	0,866	1

Talvez a significação dos valores, apesar da explicação trigonométrica, não seja tão evidente e esclarecedora da realidade descritiva quanto desejado ou esperado. Por isso é preciso recorrer à Métrica R para poder analisar a correlação não do ponto de vista do emparelhamento das variáveis, mas do ponto de vista do conjunto das variáveis, objeto da ACP.

É por isso, precisamente, que a MACRO calcula automaticamente a matriz de passagem, que figura no cabeçalho da página de correlação.

1.3 A matriz de passagem da correlação à Métrica R

A matriz de passagem permite, conforme a significação da palavra “passagem”, abordar a análise “qualitativa” do conjunto das variáveis, e detectar a “qualidade” da correlação que une ou opõe as variáveis em função dos grupos constitutivos formados pela projeção geométrica nos gráficos apropriados à comparação “qualitativa” dos objetos.

Que é a matriz de passagem?

É o vetor representativo do valor médio da correlação de cada variável.

O cálculo é automático.

Variável	T1	T2	T3	T4	T5	T6	T7	T8
Média	0,856	0,827	0,711	0,867	0,851	0,845	0,851	0,834

A finalidade do vetor da correlação média é a passagem “obrigatória” para a Métrica R, e, portanto para abordar a ACP.

2. A Métrica R e a ACP

Vamos diretamente aos resultados que figuram na página 7 da MACRO para comentar o conteúdo e explicar a finalidade.

	Média	1	2	3	4	5	6	7	8
r	0,830	0,856	0,827	0,711	0,867	0,851	0,845	0,851	0,834
ρ	0,552	0,517	0,563	0,703	0,499	0,525	0,534	0,526	0,552
r²		0,733	-0,683	-0,506	0,751	0,725	0,715	0,724	0,695
ρ²		-0,267	0,317	0,494	-0,249	-0,275	-0,285	-0,276	0,305
x	0,915	0,928	0,913	0,856	0,933	0,926	0,923	0,925	0,917
y	0,776	0,758	0,781	0,851	0,749	0,762	0,767	0,763	0,776
r¹²	0,445	0,508	0,427	0,179	0,538	0,494	0,477	0,492	0,446
		0,010	0,010	0,010	0,010	0,010	0,010	0,010	0,010
r'	0,661	0,712	0,653	0,423	0,734	0,703	0,691	0,701	0,668
ρ'	0,104	0,033	0,125	0,406	-0,003	0,049	0,068	0,051	0,104
ρ¹²	0,025	0,001	0,016	0,164	0,000	0,002	0,005	0,003	0,011
r''	0,000	0,026	-0,004	-0,119	0,037	0,021	0,015	0,020	0,003
ρ''	0,000	-0,036	0,011	0,151	-0,054	-0,028	-0,018	-0,026	0,000

Apresentação da matriz em função dos pares formados pelo co-seno r e o seno ρ.

2.1 O co-seno e o seno à base dos cálculos

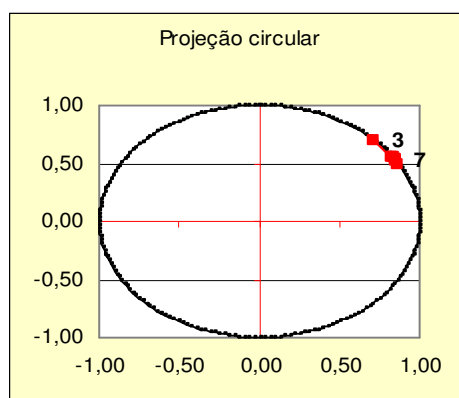
Habitualmente o seno é dado antes do co-seno. Mas como, de fato, o co-seno é um “dato fundamental”, vamos observar a ordem “básica” co-seno / seno como coordenadas trigonométricas para posicionar cada ponto no círculo trigonométrico.

Assim, na primeira linha figuram os valores médios da correlação de cada variável, quer dizer o valor “real” do co-seno de cada variável. Trata-se de um dato fundamental.

Na segunda linha figura o valor complementar do seno: um valor calculado.

Não iremos reproduzir aqui as fórmulas que permitem calcular os diferentes valores da matriz, já que as mesmas constam em vários documentos que acompanham o STABLEX.

O círculo de projeção é dado inicialmente sob forma de quadrado. Para transformar o “círculo quadrado” em “círculo redondo”, convém colar o “quadrado” no “círculo” que figura na última página 7 da MACRO, e configurar a apresentação do gráfico.



A projeção circular mostra a posição exata de cada ponto (variável) sobre o círculo.

Vê-se imediatamente que os pontos formam dois grupos distintos, onde a variável T3 (o 3) figura à parte, por causa do seno elevado que o projeta para cima.

Mas, além disso, nota-se também que as variáveis formam um corpus “homogêneo”, já que a dispersão é a mínima, ou insignificante à primeira vista.

2.2 A projeção ortogonal do co-seno e do seno no círculo trigonométrico

Como calcular os valores de projeção ortogonal?

Quem diz “projeção ortogonal”, diz necessariamente “valores quadrados”: (r^2) para o co-seno e (p^2) para o seno.

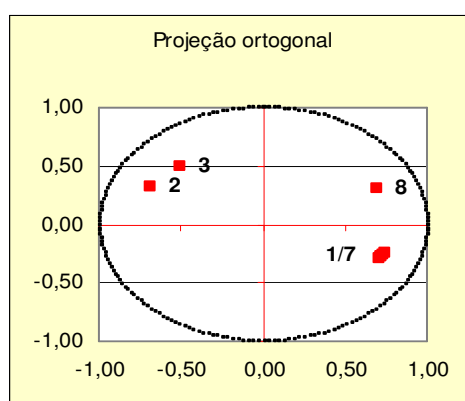
Mas não diz apenas “valores quadrados”, diz também “valores quadrados reduzidos à inércia mínima”, ou “com o mínimo de inércia”.

Para reduzir os quadrados à inércia mínima convém afetar os cálculos do signo algébrico dos pares correspondentes da “inércia mínima” que figuram na última posição da matriz.

Como fazer os cálculos com o signo algébrico correspondente?

Basta aplicar a formula seguinte: $\{=(C2^2)*(C14/ABS(C14))\}$, que permite afetar o cálculo quadrado do signo algébrico correspondente na mesma operação.

Convém notar os espaços reservados para efetuar esses cálculos.



O interesse da projeção ortogonal é evidente, visto que os pontos se posicionam em quatro grupos distintos:

1. À esquerda, o 2 e o 3, com o seno positivo e o co-seno negativo, mas separados.
2. À direita o 8, no quarto superior, com o seno e o co-seno positivos, à parte.
3. À direita, no quarto inferior, o resto dos textos praticamente confundidos ou agregados, com o seno negativo e o co-seno positivo.

Esse posicionamento é representativo da afinidade dos pontos, e, portanto das variáveis do corpus analisado.

Então a ACP não mostra apenas uma mera afinidade de posicionamento ou de distanciamento. Mostra mais do que isso: revela e salienta que o fenômeno é intrínseco e inerente à essência das variáveis. A referida distribuição é devida às “características qualitativas”, que deverão ser determinadas pela AFD.

2.3 A projeção cartesiana

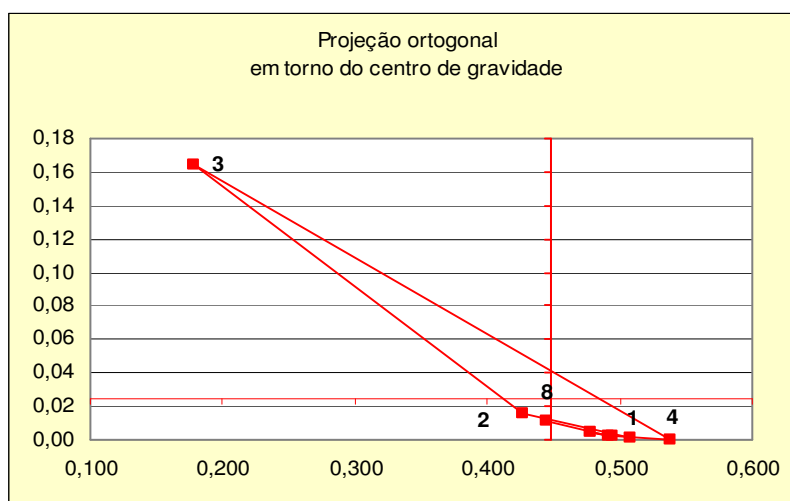
Essa projeção não tem um interesse imediato, a não ser o de formar um gráfico dentro de uma ampla perspectiva, ou seja, “afastada” da origem das coordenadas. Como é fácil de realizar, deixamos isso como exercício prático.

2.4 A projeção ortogonal cartesiana em torno do centro de gravidade

Essa projeção é feita a partir dos valores “quadrados” de r' e ρ' para cada ponto.

Os valores quadrados do co-seno e seno figuram em vermelho na matriz.

Convém adaptar a apresentação do gráfico para destacar o “centro de gravidade” da projeção em função da escala marginal.



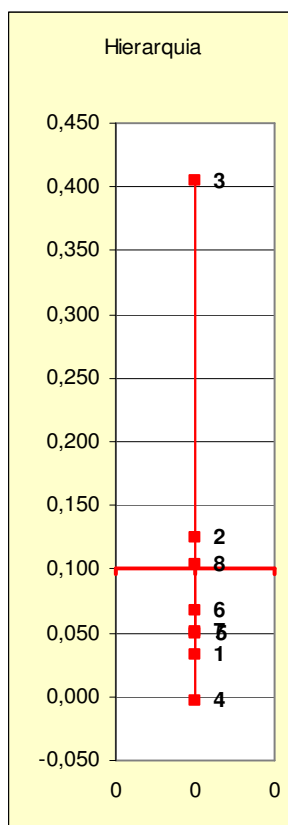
As coordenadas do centro de gravidade (o baricentro) $O - O (0,445; 0,025)$ – são as médias do co-seno (r'^2) e do seno (ρ'^2) calculadas nas linhas correspondentes da *Métrica R*.

Qualquer comentário é supérfluo. Basto observar para “ver”, para “saber e compreender”. A explicação e a justificação virão com a AFD.

2.5 A projeção hierárquica dos pontos

A visão da hierarquia dos pontos reforça consideravelmente o posicionamento das variáveis, como para melhor focalizar a força intrínseca através do distanciamento devido às características essenciais as quais convém definir, identificar e analisar.

As coordenadas de cada ponto têm um co-seno idêntico (aqui 0,010) e o seno ρ' que permite escalonar verticalmente o posicionamento.



Figuram acima da média o 8, o 2 e o 3; e, abaixo da média, o 6; logo o 7 e o 5 praticamente confundidos; depois o 1 e afinal o 4.

A AFD vai ser orientada objetivamente pela ACP que determina o posicionamento “qualitativo” das variáveis.

2.6 A projeção oblíqua da nuvem de pontos

A projeção oblíqua é feita a partir do último par de coordenadas, r'' e ρ'' , com a inércia máxima em torno do centro $O(0, 0)$.

Esse gráfico confirma de maneira evidente a qualidade da correlação entre as 8 variáveis.

De fato, essa estrutura mostra claramente a importância do peso do T3, que faz contrapeso às outras variáveis, minimizando o desvio de T2 e T8. O desvio de T2 e T8 anunciado pela projeção circular, é “qualitativamente” evidenciado pela projeção oblíqua.

A imagem vale todos os discursos do mundo, e até é bem melhor do que qualquer comentário.

Basta olhar e observar para ver e saber qual é realmente a qualidade da distribuição. O posicionamento e o distanciamento dos pontos são representativos das afinidades lexicais, textuais e narrativas:

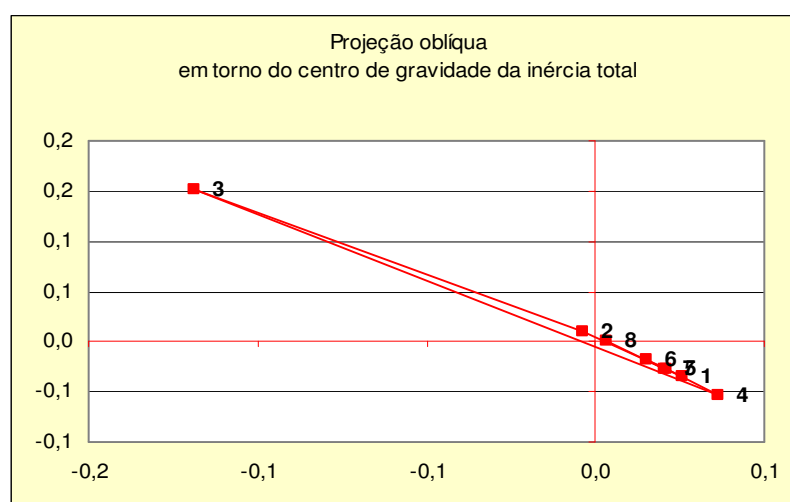
T3 destaca-se nitidamente como se fosse um caso a parte.

T2 e o T8 colam ao centro de gravidade como para representar a norma do grupo.

T1, T5, T6 e T7 formam um grupo concentrado.

T4 encerra a disposição.

Qualquer que seja o *corpus* analisado, convém sempre tomar conta do conjunto dos gráficos produzidos pela *Métrica R*, para evitar de formular uma conclusão errada ou confusa.

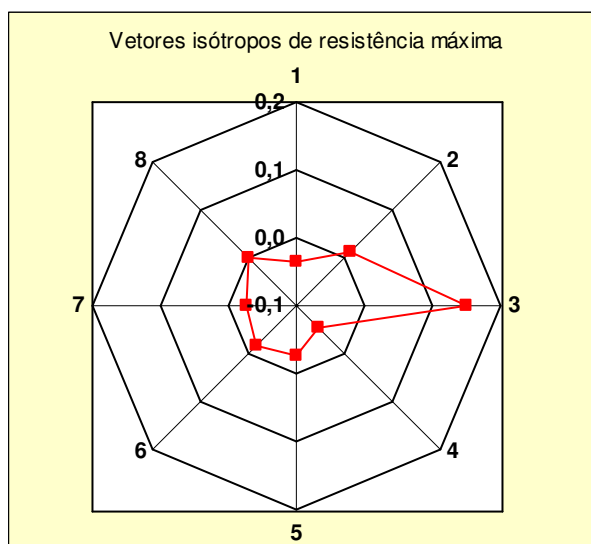


Os três últimos gráficos são representativos da “qualidade” de correlação lexical, textual e discursiva dos contos que entram na composição do *corpus*. Essas “qualidades” descritivas não são uma utopia fenomenal, já

que os elementos responsáveis pela “qualidade” analítica da ACP são ancorados nas estruturas profundas dos textos e dos discursos produzidos pelos textos. Essas meras aparências descritivas têm um fundamento “essencial”. Não vamos dizer que a ACP tenha como objetivo uma descrição essencial dos componentes, mas tem como função essencial a de evidenciar os fenômenos inerentes a esses componentes relevantes das qualidades essenciais intrínsecas. Cabe então, à AFD, identificar os elementos para poder defini-los e analisa-los concretamente.

2.7 Os vetores isotrópicos de inércia total

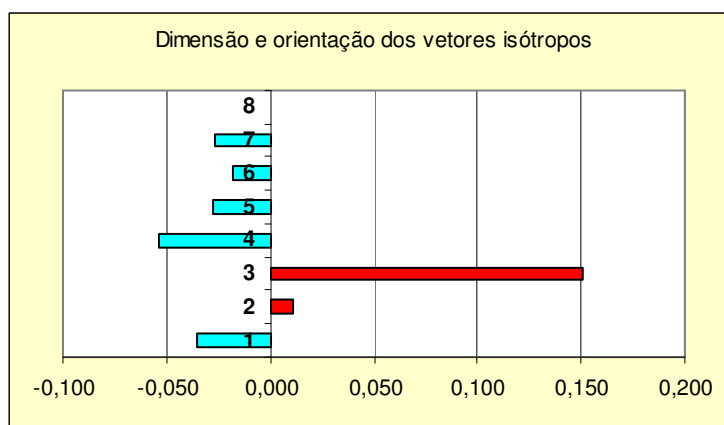
Para poder medir e observar a resistência da correlação convém destacar os vetores isotrópicos de inércia total, representados pelos valores do seno ρ :



Os vetores são concêntricos. O centro (aqui negativo: -0,1) é o ponto de referência, que permite observar a posição positiva ou negativa dos pontos em relação ao círculo central (0,0) e o comprimento de cada vetor.

2.8 A dimensão e a orientação dos vetores isotrópicos

Mas a dimensão e a orientação dos vetores são determinadas pelo diagrama das barras que coloca sistematicamente os vetores de resistência máxima positiva à direita e os vetores de resistência máxima negativa à esquerda.



Esse gráfico termina a ACP e completa “a descrição qualitativa do *corpus*”.

Não se trata aqui de definir o posicionamento dos componentes, mas a força de integração das variáveis no *corpus* através do comprimento e da orientação dos vetores.

A T3 é um caso particular, visto a resistência de integração que a caracteriza.

A T2, embora seja orientada para a direita, não se afasta muito do “eixo normativo”.

A T8 é centrada, como se fosse o elemento representativo da “norma” com a qual se confunde.

Quanto aos outros cinco componentes, estes são, por excelência, o núcleo fundamental do *corpus*, o grupo de maior influência “qualitativa” na temática comum.

3. Recapitulação das fórmulas da Métrica R

Reproduzimos aqui as fórmulas da **Métrica R** que figuram no CD de instalação do soft para facilitar a manipulação da MACRO.

A análise da **Correlação** e a análise fatorial, através da **Métrica R**, são feitas automaticamente nas duas folhas seguintes a dos Gráficos:

3.1 A Correlação. Apenas é preciso ativar o comando **Corrélation** (Correlação) na célula A1 para que apareça uma matriz quadrada correspondente à TDF com o cálculo do coeficiente marginal \bar{r} de cada variável, afixado nas duas primeiras linhas.

3.2 O cálculo da **Métrica R** é feito na folha seguinte a partir do coeficiente marginal \bar{r} anteriormente calculado. Para efetuar o cálculo na folha inteira basta ativar o comando **Métrique** (Métrica) na célula A1. Assim são automaticamente calculados os valores relativos a cada variável:

a) primeiro par, r e ρ , o co-seno e o seno da projeção normal no círculo $C(0, 1)$;

b) segundo par, x e y , as coordenadas da projeção cartesiana;

c) terceiro par, r' e ρ' , as coordenadas da rotação angular mediante as quais são feitas as nuvens de pontos, relacionando o co-seno r com o seno ρ' . Por exemplo, para observar uma configuração concentrada ou relacionando o co-seno (0) com o seno ρ' para realizar uma projeção lateral no eixo do seno (OY);

d) quarto par, r'' e ρ'' , as coordenadas da dispersão em torno de centro de gravidade. Os gráficos que estão abaixo das matrizes de passagem, aqui apresentadas, servirão de guia para a melhor observação possível. Aconselhamos, ao leitor, seguir toda a indicação e explicação dadas nos artigos de *Lexicometria*, (referenciados na bibliografia indicada no CD do Soft STABLEX, como dito no paragrafo introdutório, ou ainda no, como, por exemplo, em *Letras & Letras*).

3.3 Os gráficos da Métrica R. Abaixo dos cálculos figuram 4 gráficos automaticamente ativados:

a) o gráfico de nuvem de pontos das variáveis em torno do centro de gravidade ou de inércia CG (r'' , ρ'');

b) o gráfico sob forma de círculo quadrado, localizando as variáveis no espaço circular de centro $C(0, 0)$ (tanto pode ser o círculo como o quadrado com os quadrantes);

c) a projeção hierárquica no eixo do seno ρ' das variáveis de coordenadas $(0, \rho')$. Para realizar o tal gráfico é preciso atribuir o valor 0 ao co-seno r , preenchendo com 0 a linha 9 (a partir da célula C9);

d) o gráfico serial dado apenas pelo seno ρ'' , representando a distribuição circular em torno do centro de inércia. Enquanto amostra imediata das características da ACP esses gráficos podem ser multiplicados, já que se trata de perceber, sob o melhor ângulo, a qualidade da ligação entre as variáveis consideradas globalmente.

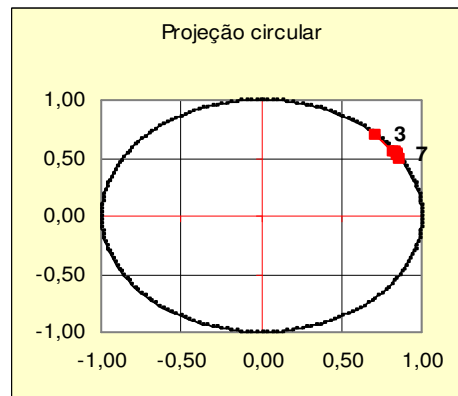
4. Os ângulos

O valor absoluto do ângulo de projeção de cada ponto no círculo de projeção é calculado a partir do valor de correlação r , que corresponde ao co-seno do ângulo formado com o raio de origem \overline{OA} do círculo trigonométrico.

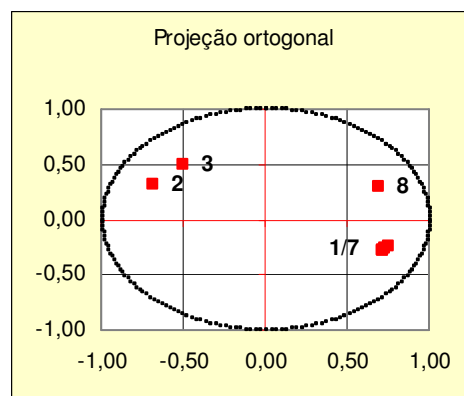
Var	T1	T2	T3	T4	T5	T6	T7	T8
Co-seno	0,856	0,827	0,711	0,867	0,851	0,845	0,851	0,834
Ângulo	32°36'	35°49'	46°47'	31°18'	33°10'	33°51'	33°10'	35°04'

Todos os ângulos são ligeiramente superiores a 30°, ao valor limite de correlação significativa.

Com um co-seno de 0,771 – ($r = 0,771 < 0,886$) –, a variável T3 forma um ângulo de 46°47' com o raio \overline{OA} , origem do círculo trigonométrico. Esse ângulo corresponde, de fato à maior resistência do grupo.



Quanto à projeção ortogonal, as coordenadas são corrigidas em função dos quadrados do co-seno (-0,506) e do seno (0,494).



O cálculo do ângulo se faz automaticamente a partir do valor algébrico de referência.

Com um co-seno de -0,500 já sabemos que a correlação de T3 não é “significativa”.

Por isso indicamos apenas como calcular os ângulos de projeção dos pontos seja no círculo de correlação, seja no círculo de projeção ortogonal.

Fazemos esta observação para indicar que o método funciona com a precisão de um relógio. Não há nenhum espaço para qualquer aproximação e muito menos para qualquer “elucubração”.

5. Os gráficos

A MACRO confecciona automaticamente os “gráficos básicos” para se ter uma “visão imediata”, que corresponde à “realidade geométrica” da configuração a partir do momento em que o cálculo algébrico projeta-se no cálculo geométrico.

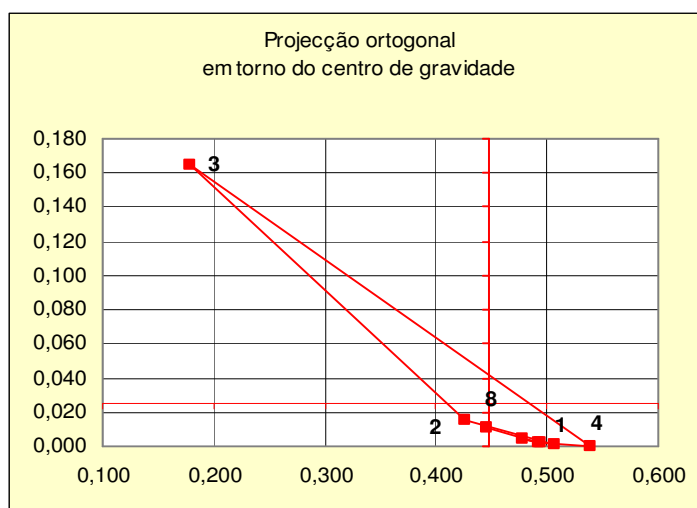
Qual é o gráfico que deve ser escolhido para servir de modelo crítico para fundamentar a “análise objetiva e indutiva”?

Não há nem “modelo normativo” nem lei normativa para responder a tal pergunta.

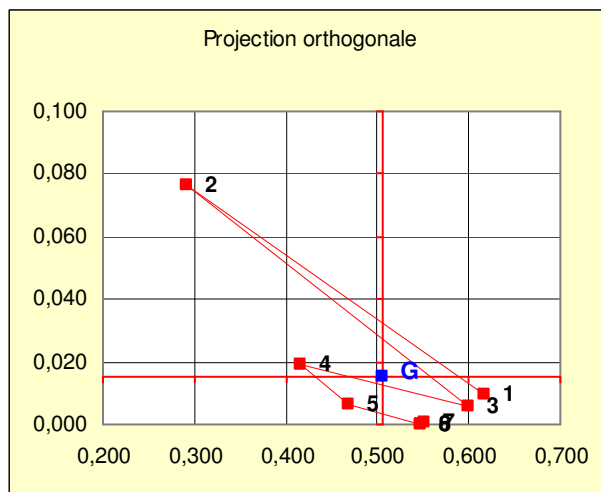
Pode-se dizer apenas que será ou serão escolhidos o gráfico ou os gráficos mais idôneos ou apropriados à “visualização” direta do fenômeno estudado, tendo como finalidade a realização de uma análise paramétrica objetiva, verificada e verificável.

Para justificar a necessária adaptação do modelo à função demonstrativa e pedagógica da análise, é preciso dizer que não há “caso geral” para fixar a priori uma determinada solução. Por isso, vale a pena colocar lado a lado ambos os gráficos de projeção ortogonal das *Três Cidras* e dos contos de Perrault, para observar as diferenças e evitar todo comentário supérfluo.

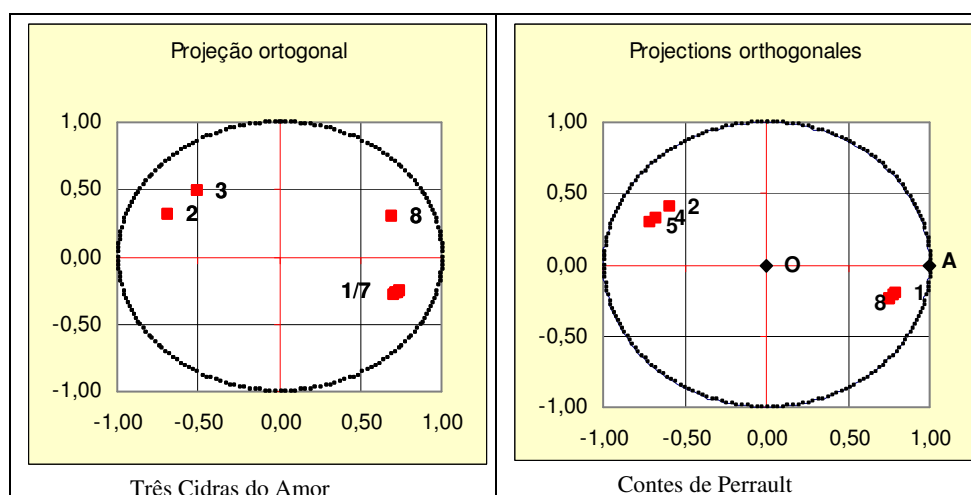
5.1 A projeção ortogonal dos oito contos das Três Cidras do Amor



5.2 A projeção ortogonal dos oito contos de Perrault



Quaisquer que sejam os modelos de gráficos, não há dois gráficos análogos, como se pode observar com os oito contos das *Três Cidras* e os oito contos de Perrault.



Convém então escolher o gráfico ou os gráficos mais relevantes e apropriados à ACP.

6. Os gráficos da Régua

Já que os gráficos realizados, quando se utiliza a Régua para calcular o peso e a densidade do vetor criado por lematização ou por discriminação, são gráficos análogos aos da ACP produzidos pela Métrica R, vamos indicar aqui como confeccionar os gráficos adequados da parábola ou da curva de Laplace-Gauss (o sino) para facilitar a observação do dito vetor.

Não percamos nunca de vista que o cálculo algébrico é um cálculo de medição e de comparação; que o cálculo geométrico é um cálculo de representação e de visualização; e que o cálculo aritmético é um cálculo de verificação e de confirmação.

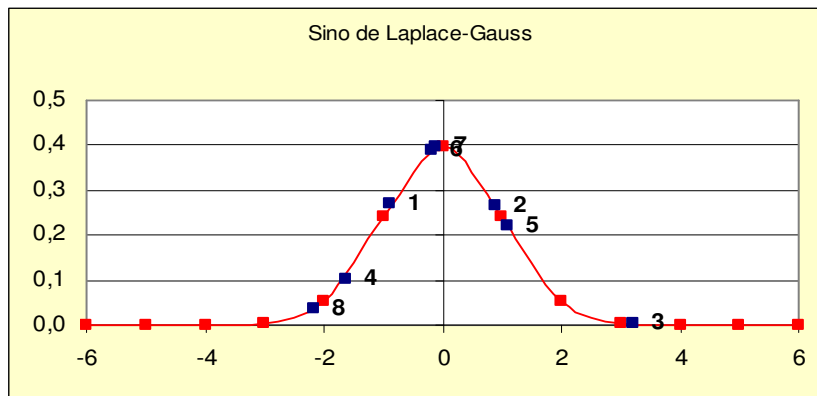
Daí o interesse em acompanhar a determinação do peso e da densidade do vetor pela visualização do fenômeno que o caracteriza.

6.1 O “sino” de Laplace-Gauss

Primeiro, é preciso criar a curva de Laplace-Gauss mediante a função predeterminada de EXCEL: Lei normal, média 0, desvio padrão 1, cumulativa “falso”.

Segundo, realizar o sino com as densidades da Régua.

Terceiro, colar o segundo gráfico no primeiro.



Como as variáveis são identificadas, a localização é evidenciada pelo gráfico.

Convém coordenar as cores para reforçar o impacto da imagem.

Como as estruturas são “ativas”, os resultados são atualizados com a medição de um novo segmento vetorial.

6.2 A parábola dá uma imagem mais eficiente

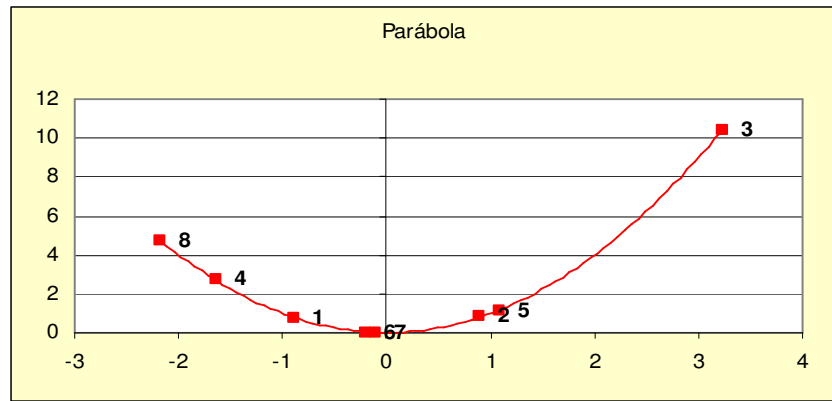
A parábola dá uma imagem mais eficiente das densidades do que o “sino” de Laplace-Gauss, embora as medidas sejam exatamente as mesmas para os dois gráficos.

Qualquer que seja a imagem escolhida, a preferência sempre deve ir para a mais representativa da “qualidade” da distribuição.

A imagem deve ter um impacto visual forte e preciso.

É evidente que as características da parábola que segue são idênticas às do “sino” precedente. Os pontos da nuvem distribuem-se de modo idêntico nos dois gráficos.

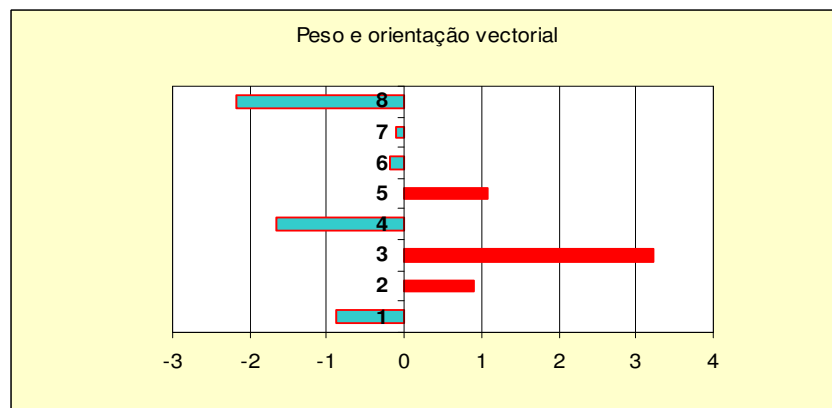
Qual é então a diferença?



A diferença resulta da própria construção do gráfico. No caso do “sino”, o gráfico é sobrecarregado por causa da contingência da elaboração que requer uma dupla estrutura. Enquanto no caso da parábola a implantação dos pontos é uma implantação direta, que dá mais relevo às densidades, a todas as densidades, tanto às positivas como às negativas ou às “neutras”.

6.3 Os vetores paramétricos

Como para a ACP, os vetores paramétricos dão uma visão exata do peso e da densidade dos elementos analisados através da dimensão e da orientação do segmento, salientando as características fundamentais do elemento analisado e do impacto da variável no corpus inteiro.



Esses três tipos de gráficos são idênticos apesar da diferença estrutural, mas cada um revela uma sensibilidade particular já que apresenta a mesma imagem sob um ângulo particular.

A filosofia do método é sempre a mesma: projetar o cálculo algébrico dentro do próprio cálculo geométrico para favorecer a melhor visualização do fenômeno observado, já que esse fenômeno não é visto sob um ponto de vista quantitativo, mas sob um ponto de vista “QUALITATIVO”.

A análise paramétrica é uma análise objetiva e indutiva. É feita de precisão e de verificação. É constante. É evidente. É “científica” (*stricto sensu*).

7. Conclusão

Conclusão não há nem pode haver, porque a ACP não tem uma finalidade em si. A ACP é uma análise “qualitativa” que desemboca necessariamente na AFD, a Análise Fatorial Discriminante, que pode ser definida como uma “análise integral”, já que todos os elementos e todos os componentes, qualquer que seja a “dimensão” que se considere, passam pelo crivo da análise paramétrica, uma “análise exploratória”.

O leitor pode referir-se à própria experiência da análise de dados que realizou com o STABLEX, pode referir-se igualmente aos estudos e ensaios que foram publicados na internet, pode consultar os documentos disponibilizados pelo Grupo de Pesquisa Linguagem, Sociedade, Saúde e Trabalho e o Núcleo de Estudos de Hipertexto e Tecnologia Educacional (NEHTE), sob a responsabilidade dos professores Cristina SAMPAIO e Antônio Carlos XAVIER.

Depois da ACP, vocês terão em breve a possibilidade de se familiarizar com a AFD, que constituiu a segunda parte do presente estudo.

¹ André **CAMLONG**, Professor Dr.
Universidade de Toulouse II – Le Mirail
camlong.andre@orange.fr