Revisitando 'The Biplot Graphic Display of Matrices with Application to Principal Component Analysis' (Gabriel, 1971).

Luiz Roberto Martins Pinto

Carlos Tadeu dos Santos Dias

Introdução:

Um dos métodos gráficos mais utilizados na análise de dados é o ***Gráfico Biplot*** (***GB***), especialmente aplicado na análise de componentes principais (**PCA**).

O objetivo básico de uma **PCA** é aproximar qualquer matriz Y retangular n x m de posto r por uma matriz n x m, mas de posto r*<r por meio da sua decomposição em valores singulares (**SVD**) de forma que $Y = \sum_{\alpha=1}^{r} \lambda_\alpha p_\alpha q_\alpha$.

Em uma **PCA** espera-se:

1) que (r*<r) expresse ambos a tendência e a variabilidade padrões (não aleatória) do fenômeno em estudo;

2) que $\sum_{\alpha=1}^{r^*<r} \lambda_{\alpha}^{2}$ quantifique esta variabilidade padrão.

Gráfico Biplot (GB)

O **GB** é importante na **PCA** para representar a estrutura da matriz Y, notadamente nos casos em que a variabilidade padrão de Y seja evidenciada para r*=2 ou 3.

Para isto, Gabriel propõe a fatoração dos valores singulares por meio dos valores '*per se*' dos vetores singulares (linhas e colunas) tomados como pesos, estimados na SVD.

E.g. coordenadas $g_{i|\alpha} = \lambda_\alpha \, p_{i|\alpha} \; , \; h_{j|\alpha} = \lambda_\alpha \, q_{j|\alpha}$

Proposição inicial

Devido ao tempo restrito vamos discutir, preferencialmente, a questão dos ‘pesos’ utilizados para a fatoração dos valores singulares λ. Propomos que λ seja fatorado utilizando-se como pesos os quadrados dos vetores singulares.

E.g. coordenadas $g_{i|\alpha} = \lambda_\alpha \, p_{i|\alpha}^2 \; , \; h_{j|\alpha} = \lambda_\alpha \, q_{j|\alpha}^2$

No decorrer da exposição abordaremos outros diferenciais nesta revisitação do artigo de Gabriel, 1971, na medida que o tempo permitir.

Fundamentação para a fatoração

Fatorar se refere a uma decomposição de um objeto matemático em produto de objetos simples.

Para realizar a fatoração dos valores singulares λ os pesos relativos (ω) devem atender aos pressupostos:

a) soma dos pesos = 1, $\Sigma\omega=1$; isto é obrigatório para que

b) $\lambda = \Sigma\, \lambda.\omega$

Neste caso ω é representado por *p* e *q* (vetores singulares associados às linhas e colunas, respectivamente)

Vamos demonstrar que o método para a fatoração de λ proposto por Gabriel não atende a estes pressupostos. $g_{i|\alpha} = \lambda_\alpha \, p_{i|\alpha}$, $h_{j|\alpha} = \lambda_\alpha \, q_{j|\alpha}$

Isto é, demonstraremos que

$\sum abs(p_{i|\alpha}) \neq 1$ portanto, $\lambda_\alpha \neq \sum abs(\lambda_\alpha p_{i|\alpha})$

Por outro lado demonstraremos que o peso apropriado para a a fatoração de λ é o quadrado dos vetores singulares. $g_{i|\alpha} = \lambda_\alpha \, p^2_{i|\alpha}$, $h_{j|\alpha} = \lambda_\alpha \, q^2_{j|\alpha}$

Isto é, demonstraremos que

$\sum p^2_{i|\alpha} = \sum q^2_{j|\alpha} = 1$ portanto, $\lambda_\alpha = \sum abs(\lambda_\alpha p^2_{i|\alpha}) = \sum abs(\lambda_\alpha q^2_{j|\alpha})$.

E isto é derivado de importante propriedade dos vetores singulares $\Sigma p^2 = \Sigma q^2 = 1$ para cada λ.

Dados utilizados

Os dados utilizados neste estudo foram apresentados por Gabriel (1971) na "*Table 1. Facilities and equipment in East Jerusalem in **1967**, by subquarter, from Israel (1968)*".

Os efeitos nas Linhas referem-se aos I=8 níveis de '*Facilities and equipment*' (F) e os efeitos nas Colunas referem-se aos J=9 níveis de '*subquarter*' (Q) .

Modelo utilizado

Considere o modelo geral

$$Y = F + Q + FxQ$$

ou

$$Y_{ij} = \mu + f_i + q_j + (fq)_{ij}$$

Vamos demonstrar o que afirmamos avaliando a interação *FxG por meio da PCA*:

$$fq_{ij} = Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}$$

Função R utilizada

Uma função em código R foi criada para a realização da PCA e do Gráfico Biplot, e atende aos modelos:

a) Interação

b) AMMI

c) GGE

d) EGE

e) Ajuste média geral

f) Gráfico Biplot para todos os efeitos da matriz.

A função r ‘bcpa.m’ é uma ampliação da função ‘bpca’ criada por Faria et. al. 2013.

Resultados

Destaques, pressupostos:

Gabriel meth: $\sum abs(p_{i|\alpha}) \neq 1$ (e.g. PC_1: $\sum abs(p_{i|\alpha=1}) = 2.52$)

Gabriel mod. meth: $\sum abs(p^2_{i|\alpha}) = \sum p^2_{i|\alpha} = 1$

As coordenadas GB para ambos os métodos de fatoração são apresentados na Tabela 2.

Tabela 1: $\lambda_{\alpha=1} = 59.69$

Tabela 2:
Gabriel meth: $\sum abs(\lambda_{\alpha=1} p_{i|\alpha=1}) = 150.72 \neq \lambda_{\alpha=1}$ e $\sum abs(\lambda_{\alpha=1} q_{j|\alpha=1}) = 151.86 \neq \lambda_{\alpha=1}$

Gabriel mod. meth: $\sum abs(\lambda_{\alpha=1} p^2_{i|\alpha=1}) = \sum abs(\lambda_{\alpha=1} q^2_{j|\alpha=1}) = 59.69 = \lambda_{\alpha=1}$

Table 1: PCA for interaction matrix FxQ. Weights for factoring singular values for the to estimation of the coordinates Biplot graphic for the F and Q effects described in Table 1 in Gabriel (1971) by 'Gabriel' and 'Gabriel modified' methods. Principal Components 1 to 4 (5 to 7 not shown).

Weights for factoring singular values								
	Gabriel method *				Gabriel modified method **			
F effect	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
To	-0,49	0,11	-0,35	0,47	-0,24	0,01	-0,12	0,22
K	-0,17	0,03	-0,35	-0,63	-0,03	0,00	-0,12	-0,40
B	0,55	0,23	-0,07	0,12	0,31	0,05	-0,01	0,01
E	-0,28	-0,72	0,10	0,08	-0,08	-0,52	0,01	0,01
W	0,49	-0,22	-0,45	0,02	0,24	-0,05	-0,20	0,00
Ra	-0,15	0,23	0,38	-0,51	-0,02	0,05	0,14	-0,26
Tv	-0,17	0,53	0,11	0,27	-0,03	0,29	0,01	0,07
Re	0,22	-0,18	0,63	0,18	0,05	-0,03	0,39	0,03
sum(abs)	2,52	2,24	2,43	2,28	1,00	1,00	1,00	1,00
Q effect	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
O1	-0,28	-0,36	0,48	-0,05	-0,08	-0,13	0,23	0,00
O2	-0,25	-0,23	0,35	-0,11	-0,06	-0,05	0,12	-0,01
O3	-0,29	0,21	-0,33	0,63	-0,08	0,04	-0,11	0,40
O4	-0,32	-0,14	-0,15	0,24	-0,10	-0,02	-0,02	0,06
M1	0,51	-0,35	-0,01	-0,04	0,26	-0,12	0,00	0,00
M2	0,62	0,03	0,04	0,34	0,38	0,00	0,00	0,12
D1	0,12	0,00	-0,24	-0,31	0,01	0,00	-0,06	-0,10
D2	-0,13	0,05	-0,54	-0,53	-0,02	0,00	-0,30	-0,28
R	0,03	0,79	0,41	-0,17	0,00	0,63	0,17	-0,03
sum(abs)	2,54	2,16	2,55	2,43	1,00	1,00	1,00	1,00
Singular values (λ)								
λ	59,69	37,80	19,51	13,06	59,69	37,80	19,51	13,06
λ^2	3563,31	1429,03	380,77	170,69	3563,31	1429,03	380,77	170,69
$\lambda^2/\Sigma\lambda^2$	0,63	0,25	0,07	0,03	0,63	0,25	0,07	0,03
Acumul.	0,63	0,88	0,95	0,98	0,63	0,88	0,95	0,98

Table 2: Coordinates and effects for the Biplot graphic for I and J levels for FxQ interaction described in Table 1 in Gabriel (1971). Principal Components (PC) 1 to 4 (5 to 7 not shown) for fatoração de λ_α by 'Gabriel' and 'Gabriel modified' methods.

$g_{i\alpha}$	Gabriel					Gabriel modified				
	PC1	PC2	Effect**	PC3	PC4	PC1	PC2	Effect**	PC3	PC4
To	-29,53	-4,15	29,82	-6,75	6,10	-14,61	-0,46	14,62	-2,34	2,85
K	-10,06	-0,95	10,11	-6,74	-8,24	-1,70	-0,02	1,70	-2,33	-5,19
B	33,09	-8,53	34,17	-1,43	1,52	18,34	-1,92	18,44	-0,10	0,18
E	-16,89	27,25	32,06	1,91	0,99	-4,78	19,65	20,22	0,19	0,07
W	29,43	8,40	30,60	-8,74	0,31	14,51	1,87	14,63	-3,91	0,01
Ra	-8,91	-8,52	12,33	7,32	-6,63	-1,33	-1,92	2,34	2,75	-3,36
Tv	-9,96	-20,19	22,51	2,22	3,55	-1,66	-10,78	10,91	0,25	0,96
Re	12,85	6,70	14,49	12,21	2,40	2,77	1,19	3,01	7,64	0,44
Sum(abs)	150,72	84,69	-	47,32	29,73	59,69	37,80	-	19,51	13,06
$h_{i\alpha}$	PC1	PC2	Effect**	PC3	PC4	PC1	PC2	Effect**	PC3	PC4
O1	-16,52	13,58	21,38	9,30	-0,71	-4,57	4,88	6,68	4,43	-0,04
O2	-15,21	8,70	17,52	6,78	-1,40	-3,88	2,00	4,36	2,36	-0,15
O3	-17,20	-7,81	18,89	-6,48	8,26	-4,96	-1,62	5,21	-2,15	5,23
O4	-19,27	5,47	20,03	-2,97	3,16	-6,22	0,79	6,27	-0,45	0,76
M1	30,32	13,11	33,03	-0,19	-0,51	15,40	4,55	16,05	0,00	-0,02
M2	36,88	-0,99	36,90	0,84	4,45	22,79	-0,03	22,79	0,04	1,52
D1	7,08	0,02	7,08	-4,60	-4,10	0,84	0,00	0,84	-1,09	-1,29
D2	-7,73	-2,06	8,00	-10,62	-6,95	-1,00	-0,11	1,01	-5,78	-3,70
R	1,65	-30,01	30,06	7,93	-2,20	0,05	-23,83	23,83	3,23	-0,37
Sum(abs)	151,86	81,76	-	49,71	31,74	59,69	37,80	-	19,51	13,06
*dist***	17,80	9,79	11,15	5,71	3,62	7,02	4,45	5,09	2,30	1,54

*Effect: e.g. $i.effect = sqrt((g_{i|\alpha=1})^2 + (g_{i|\alpha=2})^2)$; ** $dist = mean(\{i.effect, j.effect\})*0.5, \quad \forall\, i, j, a = 1,2$

Uso da função r

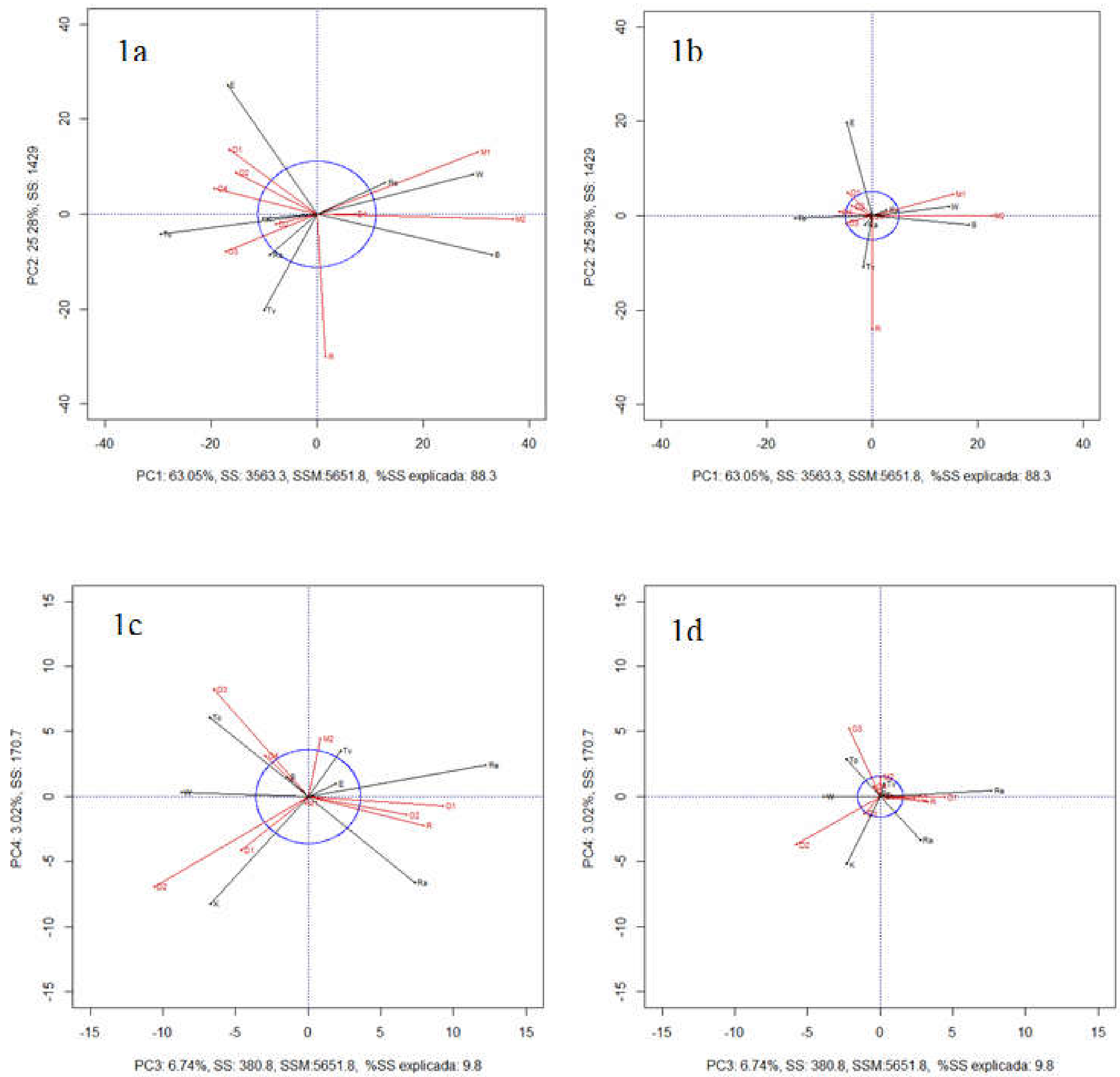

Figura 1: Gráficos Biplot. Fatoração de λ: (1a) em função de $p_{i|\alpha}$ e $q_{j|\alpha}$, (1b) em função de $p_{i|\alpha}^2$ e $q_{j|\alpha}^2$.

O Código R para executar a função bpca.m está disponível sob requisição:

Luiz Roberto Martins Pinto

luizroberto.uesc@gmail.com

Agradeço a atenção!