9
Análise Multivariada de Dados

9.1
Introdução

Uma das dificuldades inerentes em estatística multivariada é a visualização dos dados, principalmente em dimensões maiores que três. No entanto, é importante citar que, muitas vezes, estas medidas apresentam informações redundantes.

A finalidade deste Capítulo é usar a técnica Análise Fatorial para identificar fatores (indicadores). De acordo com Pereira [71], a identificação de novas dimensões na análise fatorial vai recorrer à Análise de Componentes Principais (ACP). Para tal, inicialmente, é preciso calcular as correlações entre as variáveis.

Em concordância com o exposto, a próxima seção reproduz as matrizes de correlação amostral de Pearson e Spearman dos 21 modelos publicados.

9.2
Matriz de correlação de Pearson e Spearman

Muitas vezes é pertinente investigar a existência de uma associação entre dois modelos. A literatura especializada menciona que há diversos métodos para quantificar essa associabilidade. Para fins de análise, aqui são examinados os coeficientes de correlação de Pearson e Spearman.

É importante informar que o coeficiente de correlação de Pearson mede a relação linear dos dados amostrais. Já o coeficiente de correlação de Spearman usa postos de dados amostrais para medir a força da correlação entre dois modelos quaisquer sendo, portanto, desnecessária a suposição de linearidade entre os modelos.

Concomitante a estes conceitos, é interessante realizar um teste de hipótese onde as hipóteses, nula e alternativa, são as seguintes [92]:

- *H0*: $\rho = 0$, isto é, não há correlação entre as duas variáveis.
- *H1:* $\rho \neq 0$, isto é, há correlação entre as duas variáveis.

As Tabelas 9.1 e 9.2 registram os resultados dos coeficientes de correlação de Pearson e Spearman, respectivamente.

Tabela 9.1 – Coeficientes de correlação de Pearson.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21
R1	1,000																				
R2	0,908	1,000																			
R3	0,870	0,760	1,000																		
R4	0,911	0,869	0,823	1,000																	
R5	0,908	0,820	0,849	0,982	1,000																
R6	0,843	0,751	0,762	0,861	0,883	1,000															
R7	0,791	0,649	0,717	0,800	0,834	0,908	1,000														
R8	0,863	0,886	0,820	0,908	0,866	0,751	0,631	1,000													
R9	0,696	0,663	0,642	0,756	0,749	0,595	0,543	0,812	1,000												
R10	0,764	0,697	0,716	0,854	0,865	0,939	0,889	0,775	0,612	1,000											
R11	0,589	0,596	0,600	0,618	0,617	0,496	0,407	0,693	0,864	0,432	1,000										
R12	0,918	0,859	0,820	0,962	0,976	0,861	0,798	0,890	0,808	0,836	0,677	1,000									
R13	0,906	0,850	0,836	0,927	0,938	0,808	0,748	0,909	0,849	0,780	0,720	0,980	1,000								
R14	0,640	0,668	0,392	0,699	0,656	0,634	0,607	0,596	0,443	0,643	0,287	0,706	0,680	1,000							
R15	0,898	0,831	0,786	0,939	0,958	0,834	0,783	0,848	0,777	0,805	0,622	0,981	0,960	0,707	1,000						
R16	0,613	0,659	0,359	0,669	0,611	0,588	0,558	0,593	0,435	0,600	0,277	0,668	0,655	0,994	0,671	1,000					
R17	0,889	0,784	0,801	0,894	0,919	0,788	0,764	0,827	0,781	0,748	0,629	0,951	0,965	0,671	0,971	0,644	1,000				
R18	0,545	0,538	0,599	0,575	0,573	0,401	0,319	0,653	0,843	0,337	0,958	0,611	0,654	0,136	0,571	0,128	0,580	1,000			
R19	0,724	0,691	0,687	0,776	0,772	0,624	0,551	0,828	0,986	0,615	0,918	0,823	0,860	0,437	0,781	0,425	0,783	0,891	1,000		
R20	0,529	0,529	0,607	0,546	0,538	0,348	0,267	0,653	0,827	0,291	0,936	0,576	0,635	0,096	0,540	0,095	0,561	0,992	0,871	1,000	
R21	0,718	0,692	0,711	0,758	0,752	0,581	0,511	0,838	0,976	0,573	0,916	0,803	0,858	0,407	0,759	0,402	0,775	0,896	0,992	0,889	1,000

Tabela 9.2 – Coeficientes de correlação de Spearman.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21
R1	1,000																				
R2	0,911	1,000																			
R3	0,860	0,735	1,000																		
R4	0,859	0,822	0,822	1,000																	
R5	0,863	0,786	0,855	0,982	1,000																
R6	0,837	0,741	0,756	0,860	0,875	1,000															
R7	0,764	0,615	0,723	0,774	0,808	0,892	1,000														
R8	0,872	0,884	0,810	0,894	0,864	0,761	0,636	1,000													
R9	0,695	0,655	0,647	0,730	0,737	0,591	0,524	0,824	1,000												
R10	0,758	0,681	0,716	0,865	0,870	0,928	0,872	0,784	0,597	1,000											
R11	0,674	0,631	0,633	0,705	0,713	0,564	0,499	0,804	0,997	0,564	1,000										
R12	0,884	0,829	0,812	0,938	0,967	0,857	0,761	0,890	0,803	0,842	0,778	1,000									
R13	0,888	0,835	0,832	0,904	0,927	0,798	0,700	0,921	0,853	0,772	0,830	0,973	1,000								
R14	0,570	0,625	0,287	0,590	0,562	0,603	0,554	0,531	0,344	0,645	0,299	0,625	0,580	1,000							
R15	0,886	0,827	0,818	0,939	0,967	0,858	0,759	0,892	0,809	0,835	0,784	0,999	0,976	0,611	1,000						
R16	0,554	0,617	0,264	0,550	0,519	0,565	0,513	0,520	0,334	0,604	0,288	0,594	0,566	0,994	0,579	1,000					
R17	0,891	0,787	0,845	0,888	0,923	0,791	0,731	0,877	0,824	0,756	0,802	0,950	0,977	0,543	0,957	0,524	1,000				
R18	0,617	0,553	0,659	0,654	0,664	0,510	0,461	0,745	0,933	0,497	0,949	0,690	0,734	0,070	0,701	0,048	0,722	1,000			
R19	0,736	0,690	0,693	0,776	0,784	0,640	0,564	0,851	0,994	0,643	0,989	0,840	0,878	0,374	0,845	0,359	0,851	0,929	1,000		
R20	0,607	0,537	0,687	0,628	0,638	0,486	0,436	0,746	0,914	0,475	0,930	0,660	0,724	0,019	0,673	0,006	0,716	0,985	0,908	1,000	
R21	0,747	0,703	0,721	0,781	0,788	0,630	0,560	0,871	0,990	0,635	0,984	0,841	0,893	0,366	0,847	0,355	0,865	0,922	0,994	0,913	1,000

Ao examinar as Tabelas 9.1 e 9.2, considerando um nível de significância habitualmente adotado de 5%, conclui-se que a maioria dos modelos apresenta coeficientes de correlação significativos, exceto: **R14-R18**, **R14-R20**, **R16-R18** e **R16-R20.**

As Figuras 9.1, 9.2 e 9.3 reproduzem as Tabelas 9.1 e 9.2 do estudo em que os coeficientes de correlação de Pearson e Spearman são descritos.

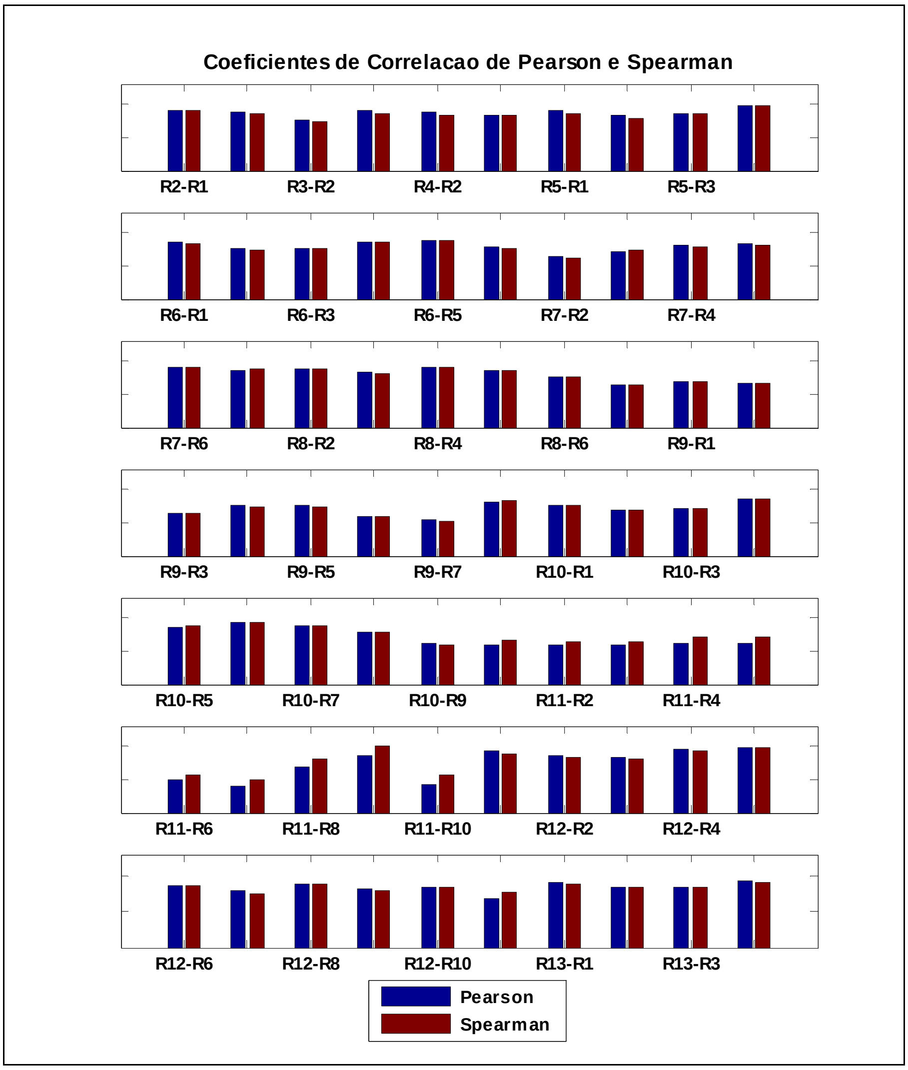

Figura 9.1 Coeficientes de correlação de Pearson e Spearman.

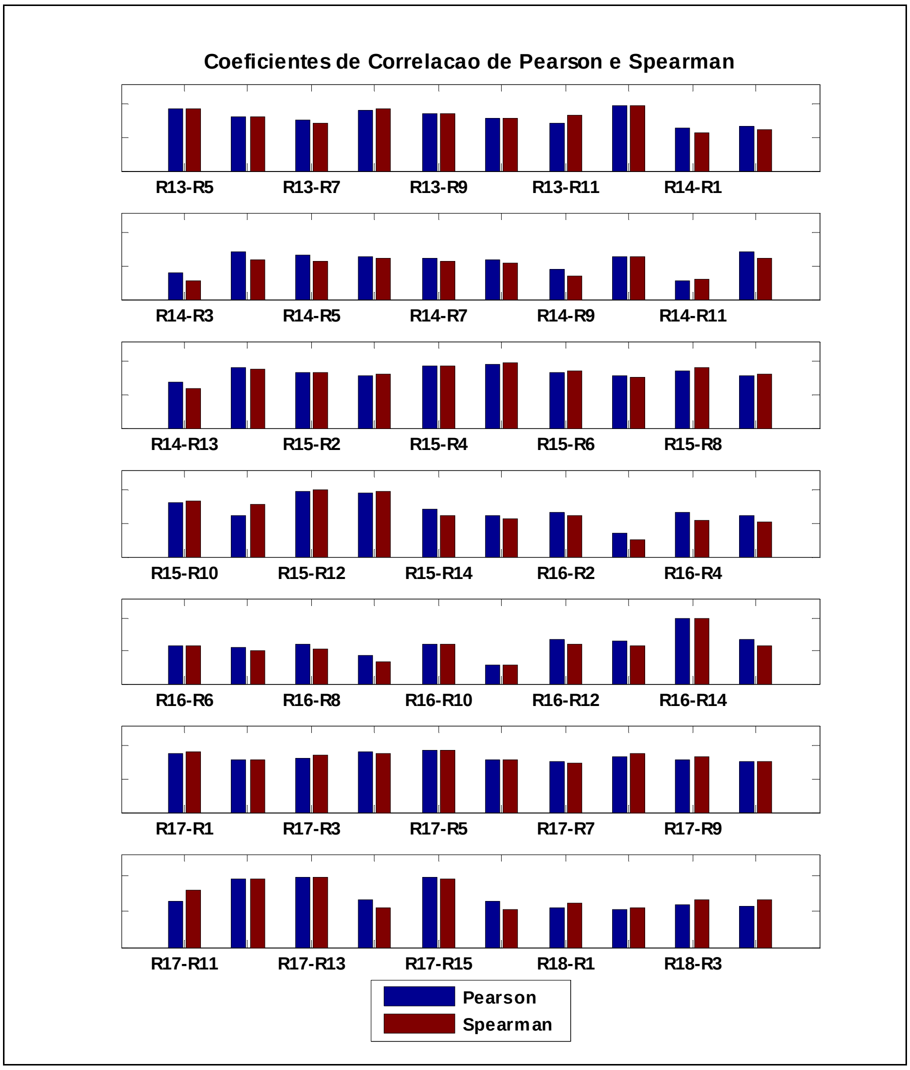

Figura 9.2: Coeficientes de correlação de Pearson e Spearman.

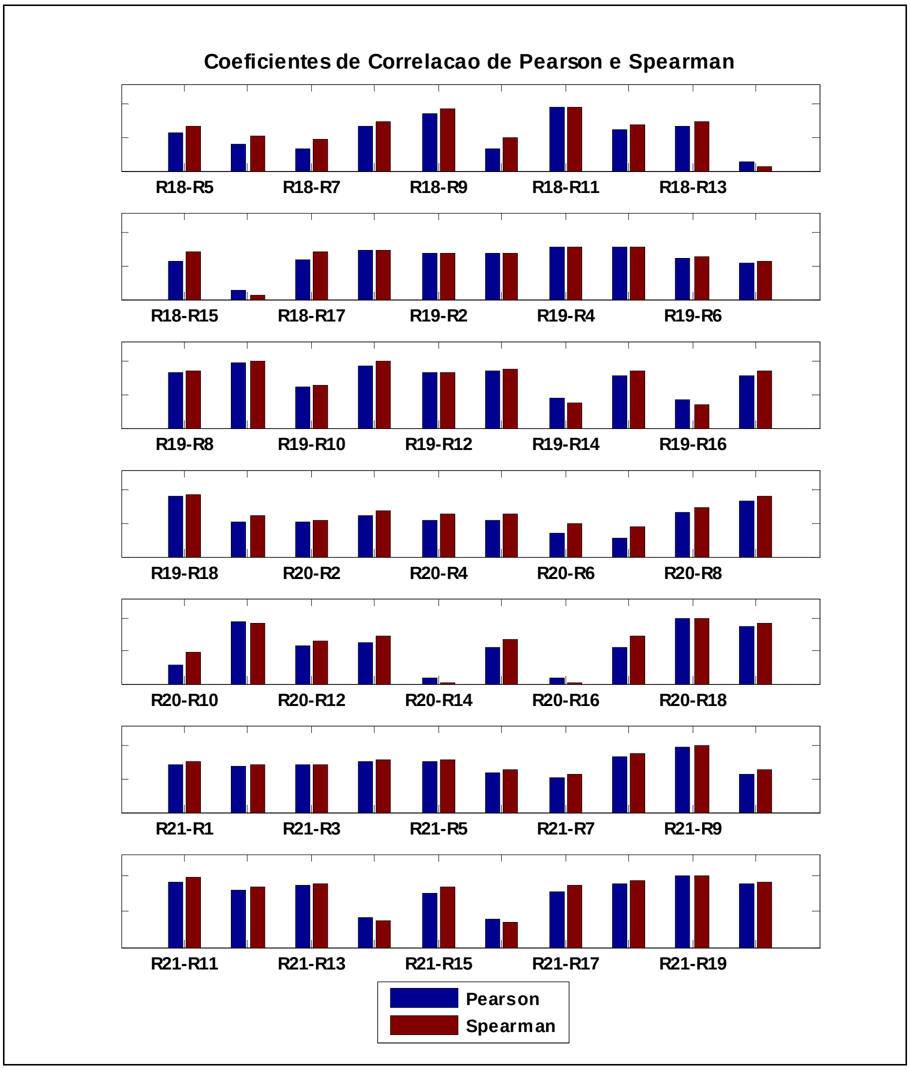

Figura 9.3: Coeficientes de correlação de Pearson e Spearman.

9.3
Análise Fatorial

Segundo Pereira [71], a Análise de Componentes Principais, popularmente chamada de PCA (*Principal Component Analysis*), é aplicada como recurso para a identificação de dimensões abstratas sobre as quais os objetos estudados possam ser projetados satisfatoriamente e, então, estudados em relação a essas dimensões ao invés das dimensões originais das medidas realizadas. Esta metodologia, considerada por muitos autores um dos tipos de análise fatorial, gera um novo conjunto de variáveis, denominado de Componentes Principais. Cada componente principal é uma combinação linear das variáveis originais e são ortogonais entre si, ignorando, então, a redundância de informação. Mingoti [66] cita que uma outra particularidade desta técnica é explicar a variância e covariância de um vetor aleatório, composto de p-variáveis aleatórias.

Porém, antes de aplicar esses conceitos, é importante conceber algumas premissas da análise fatorial. Considere, inicialmente, o problema de analisar a normalidade do vetor aleatório. Há na literatura vários testes com esta finalidade. No entanto, aqui é realizada uma verificação baseada no gráfico de probabilidade qui-quadrado conforme mostrado na Figura 9.4:

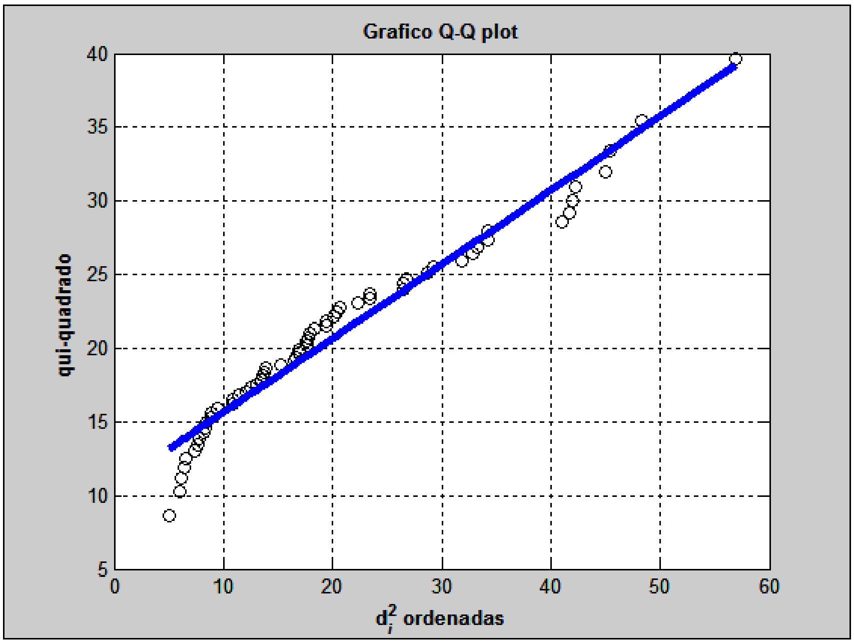

Figura 9.4: Gráfico qui-quadrado (Q-Q *plot*).

Por inspeção visual da Figura 9.4, não é difícil perceber que a normalidade p-variada pode ser coerente com os dados amostrais [66].

Uma outra questão que também merece destaque refere-se à determinação da medida de adequação dos dados, *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* (KMO). Conforme reportado por Pereira [71], para uma boa adequação dos dados à análise fatorial, é necessário que esta medida assuma valores iguais ou superiores a 0,80. Assim sendo, vale informar que o KMO encontrado para este estudo foi de 0,833.

De posse dessas informações, tem-se que as variáveis latentes não observáveis F_1 e F_2 correspondem, respectivamente, a 73,60 % e 15,29% da variância total. Analisando estes fatores (F_1, F_2), pode-se gerar um modelo que explica, aproximadamente, 90% das variações das medidas originais. Neste ponto, é interessante registrar dois conceitos [47]:

1. **Comunalidades** ($\hat{h}_i^2$): quantia total de variância que uma variável original compartilha com todas as outras variáveis incluídas na análise;
2. **Variância específica** ($\hat{\psi}_i$): variância de cada variável, única àquela variável e que não é explicada ou associada com as outras variáveis na análise fatorial.

Para evitar ambigüidade, na interpretação dos resultados, muitas vezes é desejável empregar um método rotacional para simplificar a estrutura fatorial, e, dessa forma, conseguir soluções fatoriais mais simples e teoricamente mais significativas. Neste contexto, é importante esclarecer que foi utilizado, para este estudo, o método de rotação fatorial ortogonal VARIMAX [47].

A titulo de ilustração, a Tabela 9.3 apresenta os resultados finais.

Tabela 9.3 – Comparação entre os *factor loadings* não rotacionados e rotacionados.

Modelo (Ri)	F_1	F_2	F_1	F_2	$\hat{h}_i^2$	$\hat{\psi}_i$
R1	0,9251	0,1561	0,8180	0,4600	0,8802	0,1198
R2	0,8765	0,1277	0,7620	0,4520	0,7845	0,2155
R3	0,8505	-0,0197	0,6490	0,5500	0,7238	0,2762
R4	0,9579	0,1605	0,8460	0,4770	0,9433	0,0567
R5	0,9560	0,1552	0,8410	0,4800	0,9379	0,0621
R6	0,8542	0,3238	0,8680	0,2850	0,8344	0,1656
R7	0,7884	0,3750	0,8490	0,2040	0,7622	0,2378
R8	0,9295	-0,0290	0,7050	0,6070	0,8647	0,1353
R9	0,8706	-0,3683	0,4460	0,8340	0,8936	0,1064
R10	0,8273	0,3562	0,8670	0,2430	0,8114	0,1886
R11	0,7657	-0,5670	0,2390	0,9220	0,9079	0,0921
R12	0,9788	0,1152	0,8340	0,5260	0,9713	0,0287
R13	0,9788	0,0290	0,7800	0,5930	0,9589	0,0411
R14	0,6701	0,5385	0,8600	0,0020	0,7391	0,2609
R15	0,9526	0,1494	0,8350	0,4820	0,9297	0,0703
R16	0,6437	0,5187	0,8270	0,0010	0,6834	0,3166
R17	0,9358	0,1021	0,7920	0,5090	0,8862	0,1138
R18	0,7086	-0,6806	0,1230	0,9750	0,9653	0,0347
R19	0,8910	-0,3969	0,4440	0,8690	0,9515	0,0485
R20	0,6826	-0,7118	0,0840	0,9830	0,9726	0,0274
R21	0,8785	-0,4310	0,4120	0,8870	0,9575	0,0425

Examinando a Tabela 9.3, é possível verificar que o fator 1 reúne melhor as informações relativas às formas funcionais dos modelos (**R6** e **R10**). Por outro lado, o fator 2 refere-se mais à questão dos *clusters* (**R20** e **R18**). Não é difícil perceber que esses fatores são coerentes com a natureza deste estudo. Para auxiliar a interpretação dos fatores, pode-se ainda recorrer a um gráfico de dispersão que examine a localização dos modelos em um sistema de coordenadas criado pelos fatores. (ver Figuras 9.5 e 9.6).

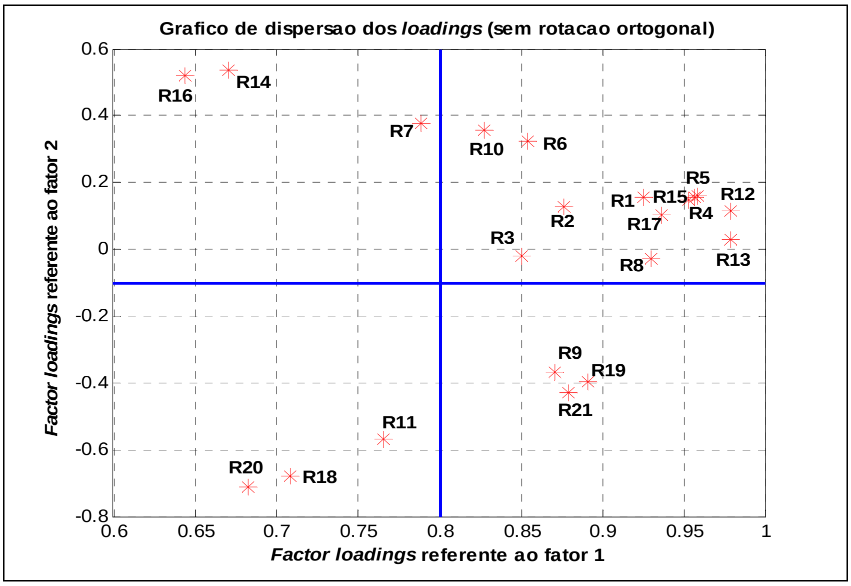

Figura 9.5: Gráfico de dispersão dos *loadings* (sem rotação ortogonal).

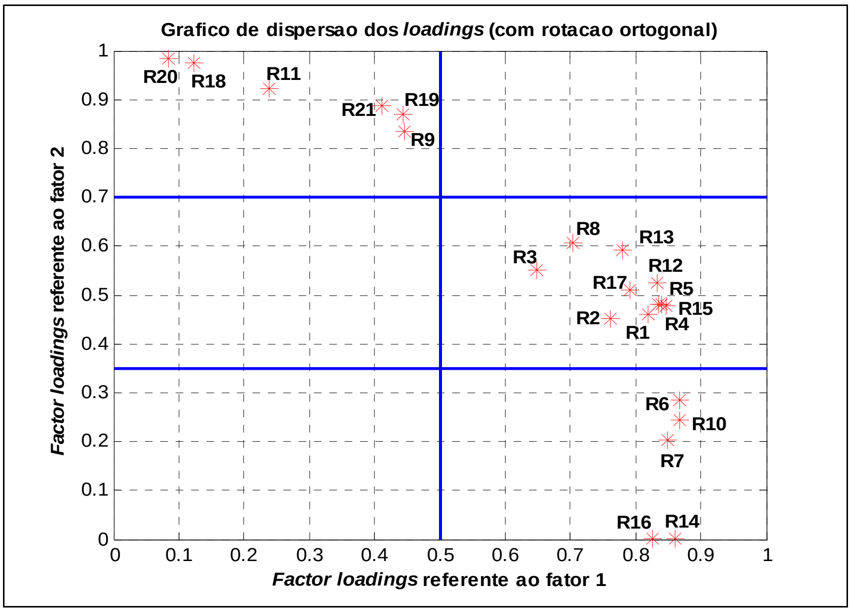

Figura 9.6: Gráfico de dispersão dos *loadings* (com rotação ortogonal).

Após rotacionar os eixos fatoriais, nota-se como os modelos se agrupam:

1. Grupo 1: **R9**, **R11**, **R18**, **R19**, **R20** e **R21**;
2. Grupo 2: **R1**, **R2**, **R3**, **R4**, **R5**, **R8**, **R12**, **R13**, **R15** e **R17**;
3. Grupo 3: **R6**, **R7**, **R10**, **R14** e **R16**.

Um outro exercício que pode ser feito é combinar os *factor scores* das empresas em uma representação gráfica como a da Figura 9.7:

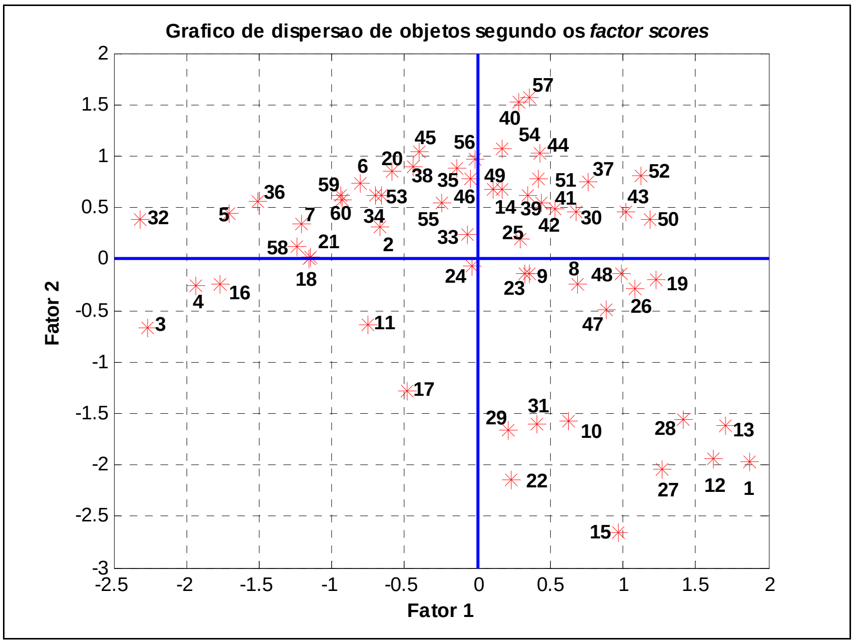

Figura 9.7: Gráfico de dispersão de objetos segundo os *factor scores.*

Inspecionando a Figura 9.4, percebe-se que a maioria das concessionárias (vide primeiro e segundo quadrantes) se destaca com relação à análise *cluster*. Conclusão análoga pode ser feita com relação à forma funcional (vide primeiro e quarto quadrantes). Já as empresas 3, 4, 11, 16, 17 e 24 (vide terceiro quadrante) dão uma contribuição menor a estes 2 fatores.

Por último, pode-se utilizar o primeiro fator para uma simples ordenação (*ranking*) das concessionárias em estudo. Mingoti [66] discorre que existem alguns questionamentos em relação a esta aplicação. Entretanto, vale informar que este procedimento tem sido amplamente utilizado para este fim.

Tabela 9.4 – *Ranking* das concessionárias.

	Concessionária	Posição
CLUSTER 1	AES-SUL	1
	CEAL	8
	CEEE	14
	CELPA	13
	CELTINS	12
	CEPISA	10
	CERON	11
	COSERN	4
	ENERGIPE	6
	ESCELSA	5
	MANAUS	9
	PIRATININGA	3
	RGE	2
	SAELPA	7
CLUSTER 2	BANDEIRANTES	5
	CEB	17
	CELESC	13
	CELG	16
	CELPE	3
	CEMAR	14
	CEMAT	15
	CEMIG	10
	CERJ	8
	COELBA	12
	COELCE	9
	COPEL	4
	CPFL	2
	ELEKTRO	1
	ELETROPAULO	11
	ENERSUL	6
	LIGHT	7
CLUSTER 3	BOA VISTA	29
	BRAGANTINA	18
	CAUIÁ	24
	CAT-LEO	19
	CEA	28
	CELB	6
	CENF	22
	CFLO	12
	CHESP	13
	COCEL	7
	CPEE	8
	CSPE	3
	DEMEI	9
	ELETROACRE	21
	ELETROCAR	17
	JAGUARI	5
	JOÃO CESA	4
	MOCOCA	15
	MUXFELDT	1
	NACIONAL	10
	NOVA PALMA	2
	PANAMBI	23
	POÇOS DE CALDAS	14
	SANTA CRUZ	20
	SANTA MARIA	16
	SULGIPE	11
	URUSSANGA	27
	V. PARANAPANEMA	26
	XANXERÊ	25

De posse das informações apresentadas na Tabelas 9.4, conclui-se o seguinte:

	***CLUSTER* 1**	***CLUSTER* 2**	***CLUSTER* 3**
Melhor Empresa	AES-SUL	ELEKTRO	MUXFELDT
Pior Empresa	CEEE	CEB	BOA VISTA