

MAE5776 – Análise Multivariada LISTA 3 1º Semestre/2022

Alunos:

Fernando F. Paulos Vieira - n° USP: 13492870

Leandro Alves da Silva - n° USP: 11868023

Thiago Ferreira Miranda - n° USP: 11925711

1 - A partir de uma matriz de dados normalizados $Y_{n \times p}^*$, considere a matriz de covariâncias $nS_{p \times p} = Y^{*'}Y^* = V\Lambda V'$, tal que $V_{p \times p} = (V_1, \dots, V_p)$ e $\Lambda = \text{diag}(\lambda_j)$ são matrizes de autovetores (das colunas de $Y_{n \times p}^*$) e autovalores, respectivamente, e a matriz de distâncias $D_{n \times n}$, tal que seus elementos são função dos elementos de $B_{n \times n} = Y^*Y^{*'} = U\Lambda U'$, com $U_{n \times n} = (U_1, \dots, U_n)$ matriz de autovetores (das linhas de $Y_{n \times p}^*$). Três pesquisadores realizaram análises estatísticas e chegaram à seguinte redução de dimensionalidade de Y^* .

Pesquisador 1: $Y_{n \times p}^* \rightarrow \tilde{Y}_{n \times 2} = Y^*(V_1 \ V_2)$

Pesquisador 2: $Y_{n \times p}^* \rightarrow \tilde{Y}_{n \times 2} = Y^*\left(\frac{V_1}{\sqrt{\lambda_1}} \ \frac{V_2}{\sqrt{\lambda_2}}\right)$

Pesquisador 3: $Y_{n \times p}^* \rightarrow \tilde{Y}_{n \times 2} = Y^*(V_1\sqrt{\lambda_1} \ V_2\sqrt{\lambda_2})$

1.1 - Qual análise estatística cada pesquisador realizou? Que propriedades dos dados estão preservadas em cada caso? Eles partiram do mesmo objetivo? Faça suposições necessárias.

R:

O pesquisador 1 - Realizou uma Análise de Componentes Principais. Buscou preservar a variância total dos dados (ou a maior proporção da variância total que possa ser preservada).

O pesquisador 2 - Realizou uma Análise Fatorial Exploratória. Buscou aproximar a matriz de covariâncias em termos de fatores latentes comuns e específicos, descrevendo as variáveis em função de m fatores.

O pesquisador 3 - Realizou uma Análise de Coordenadas Principais ou Escalonamento Multidimensional. Analogamente ao Pesquisador 1, buscou preservar a variância total dos dados (ou a maior proporção da variância total que possa ser preservada).

Num primeiro momento, acredita-se que estes pesquisadores partiram de um mesmo objetivo de redução de dimensionalidade das variáveis presentes nos dados originais disponíveis.

1.2 - Simule dados e realize as análises dos três pesquisadores. Interprete os resultados.

Amostra	Vetor de Médias	Matriz de Covariâncias
$n_1 = 100$	$\mu_1 = (5, 7, 10, 8, 11)$	$\Sigma_1 = \begin{pmatrix} 1 & 0.3 & 0.5 & 0.2 & 0.7 \\ 0.3 & 1 & 0.2 & 0.4 & 0.6 \\ 0.5 & 0.2 & 2.5 & 1.2 & 1.5 \\ 0.2 & 0.4 & 1.2 & 2 & 1.5 \\ 0.7 & 0.6 & 1.5 & 1.5 & 3 \end{pmatrix}$

1.3 - Para os dados simulados, obtenha uma representação Biplot. Como esse gráfico é construído?

2 - Considere os dados “bodyfat” disponíveis na biblioteca TH.data do R. Neste caso, a matriz de trabalho contém 71 observações avaliadas em 10 variáveis. Gere 5 observações (para tanto, adote um critério) e considere seu novo conjunto de dados “bodyfat_new”. Com base na matriz de trabalho resultante realize as seguintes análises:

2.1 - Componentes Principais.

Centróides dos componentes:

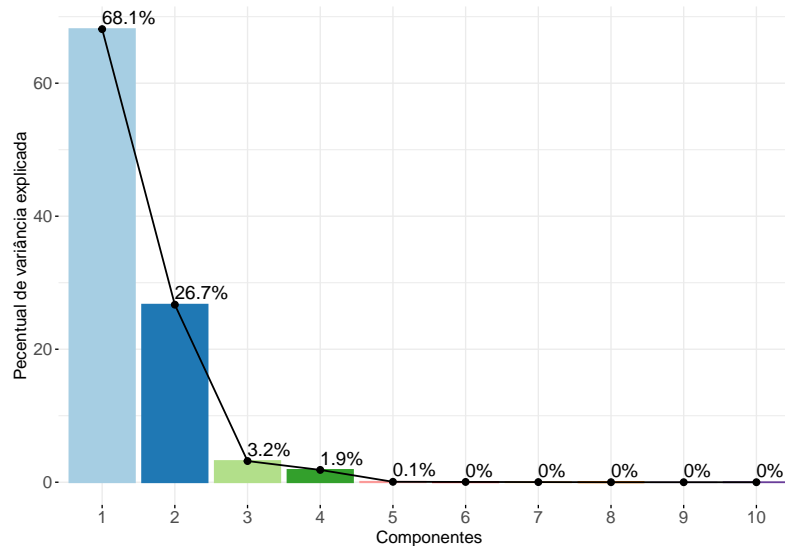
PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
50.75	30.6439	87.2789	105.0844	6.5048	9.3218	3.8605	4.2793	3.8764	5.383

Autovalores:

PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
415.484	162.7966	19.4313	11.2885	0.3823	0.2692	0.1336	0.0345	0.0077	0.0018

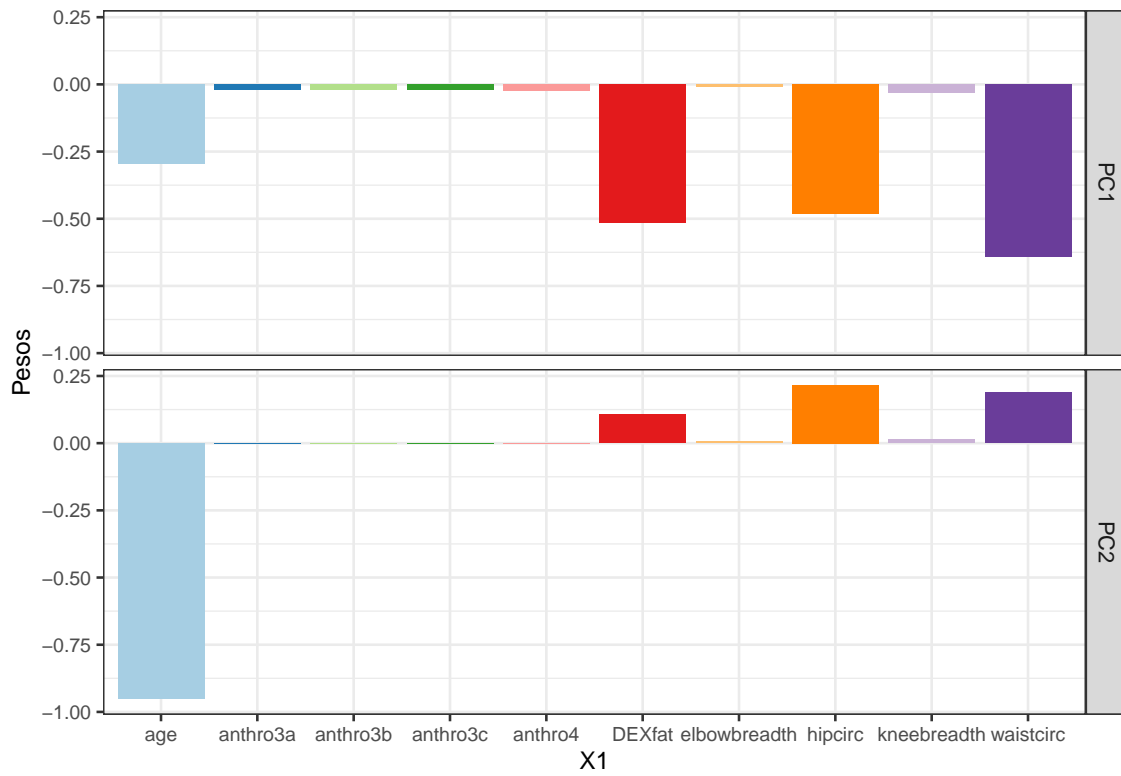
Importância dos componentes:

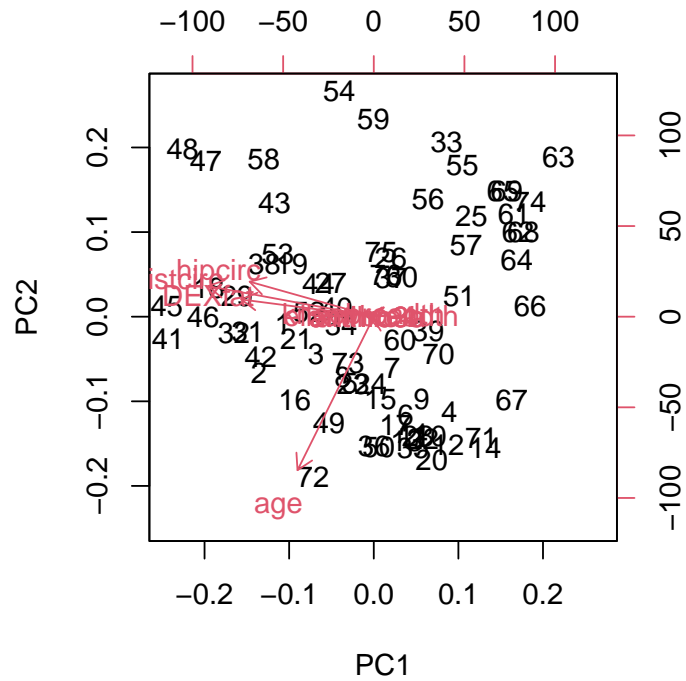
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	20.3834	12.7592	4.4081	3.3598	0.6183	0.5189	0.3656	0.1859	0.088	0.0421
Proportion of Variance	0.6813	0.2670	0.0319	0.0185	0.0006	0.0004	0.0002	0.0001	0.000	0.0000
Cumulative Proportion	0.6813	0.9483	0.9801	0.9986	0.9993	0.9997	0.9999	1.0000	1.000	1.0000



Autovetores:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
age	-0.2959	-0.9515	0.0355	-0.0753	0.0094	0.0023	0.0044	0.0023	-0.0003	0.0002
DEXfat	-0.5140	0.1068	0.2621	0.8026	0.0261	0.1014	0.0228	0.0014	-0.0014	-0.0024
waistcirc	-0.6428	0.1895	-0.7129	-0.2054	-0.0080	0.0151	-0.0094	-0.0039	0.0025	0.0016
hipcirc	-0.4817	0.2169	0.6491	-0.5460	-0.0312	-0.0203	0.0105	0.0016	0.0007	0.0001
elbowbreadth	-0.0078	0.0071	-0.0156	0.0082	0.2307	-0.3360	0.9058	0.1138	0.0083	0.0051
kneebreadth	-0.0327	0.0143	0.0146	0.0168	0.7956	-0.4714	-0.3769	-0.0086	0.0289	-0.0087
anthro3a	-0.0191	-0.0009	0.0016	0.0406	-0.1997	-0.3557	-0.0229	-0.4422	-0.6344	0.4827
anthro3b	-0.0187	-0.0020	0.0026	0.0479	-0.2752	-0.3689	-0.0712	-0.0442	0.7192	0.5111
anthro3c	-0.0220	0.0005	-0.0017	0.0473	-0.2914	-0.3843	-0.1690	0.8156	-0.2657	-0.0187
anthro4	-0.0254	-0.0024	0.0003	0.0573	-0.3340	-0.4936	-0.0513	-0.3525	0.0930	-0.7109





Análise de Componentes Principais com dados padronizados

Centróides dos componentes:

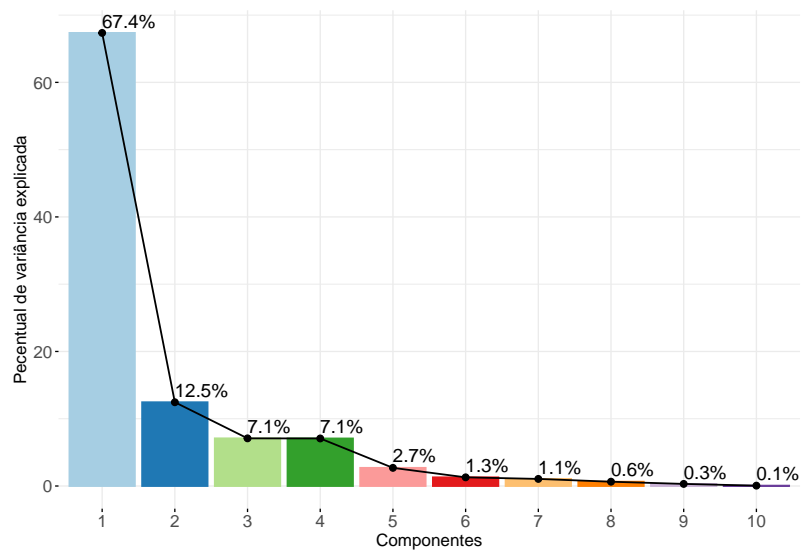
PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
50.75	30.6439	87.2789	105.0844	6.5048	9.3218	3.8605	4.2793	3.8764	5.383

Autovalores:

PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
6.7355	1.2452	0.7084	0.7067	0.2697	0.1289	0.1054	0.0638	0.0311	0.0054

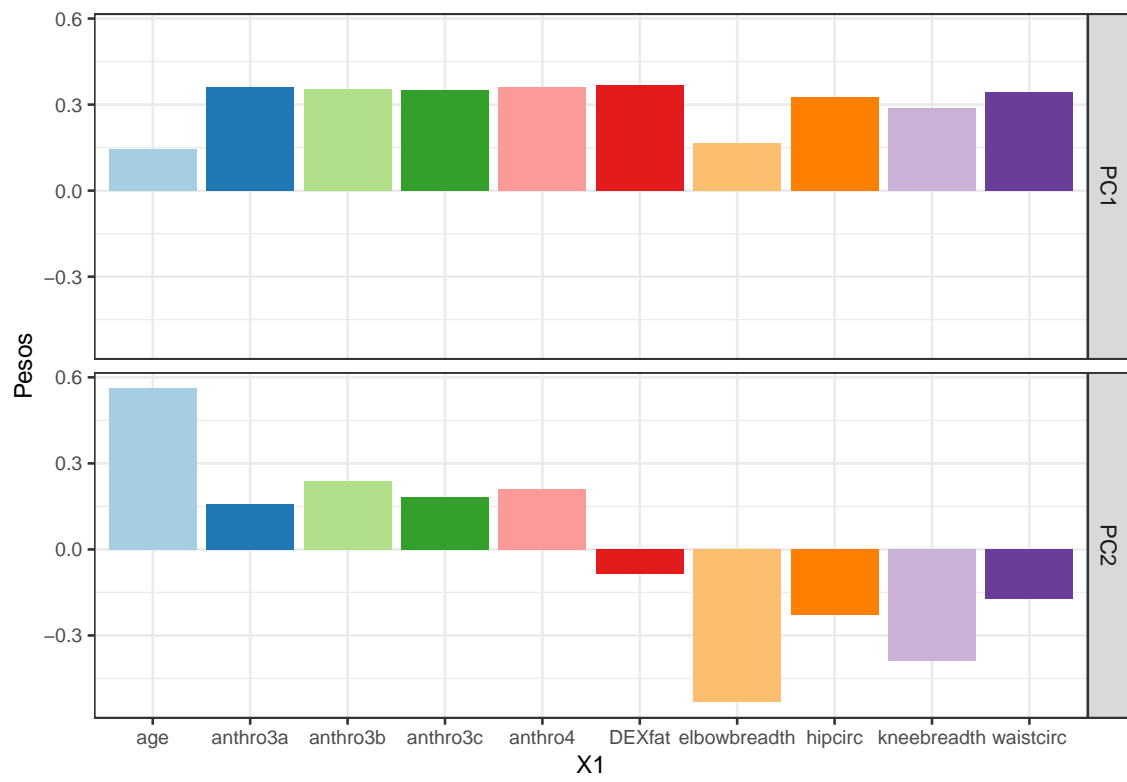
Importância dos componentes:

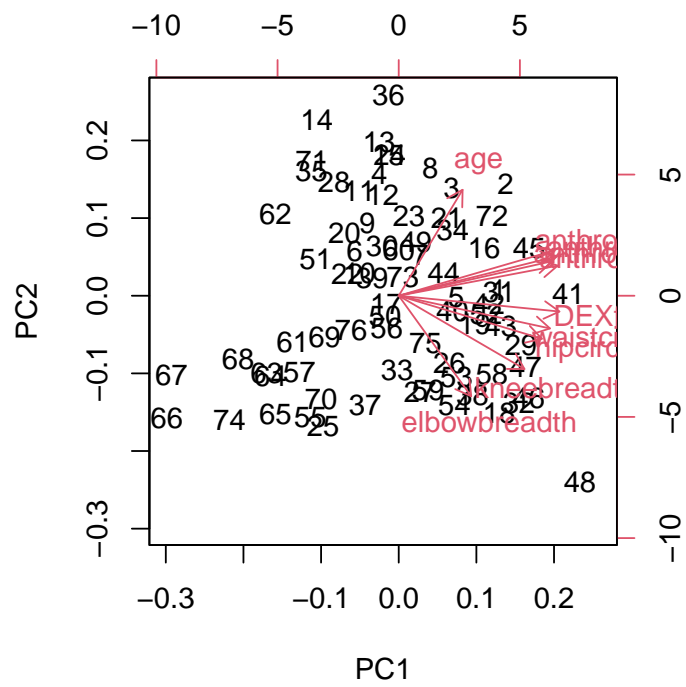
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	2.5953	1.1159	0.8416	0.8407	0.5193	0.3590	0.3247	0.2525	0.1763	0.0733
Proportion of Variance	0.6736	0.1245	0.0708	0.0707	0.0270	0.0129	0.0105	0.0064	0.0031	0.0005
Cumulative Proportion	0.6736	0.7981	0.8689	0.9396	0.9666	0.9794	0.9900	0.9964	0.9995	1.0000



Autovetores:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
age	0.1456	0.5618	0.7324	0.3430	0.0407	-0.0149	0.0778	-0.0334	0.0094	-0.0046
DEXfat	0.3663	-0.0838	-0.0746	0.2012	0.1138	-0.0947	0.0292	0.8841	0.0780	0.0471
waistcirc	0.3447	-0.1733	-0.0219	0.2481	0.4187	0.6043	-0.4527	-0.1811	-0.0962	-0.0397
hipcirc	0.3263	-0.2279	-0.1581	0.3918	0.3615	-0.5205	0.3510	-0.3711	-0.0228	-0.0021
elbowbreadth	0.1653	-0.5331	0.5998	-0.5266	0.1871	-0.0455	0.1178	0.0189	-0.0016	-0.0044
kneebreadth	0.2866	-0.3884	0.1281	0.3342	-0.7856	0.0795	-0.0228	-0.1060	-0.0589	0.0140
anthro3a	0.3624	0.1596	-0.0591	-0.2252	-0.1031	-0.2918	-0.4290	-0.1220	0.5764	-0.4026
anthro3b	0.3541	0.2394	-0.1185	-0.2759	-0.0899	-0.0663	0.0453	0.0187	-0.7239	-0.4338
anthro3c	0.3505	0.1835	-0.1789	-0.2029	-0.0611	0.4818	0.6445	-0.0886	0.3333	0.0117
anthro4	0.3608	0.2094	-0.0829	-0.2641	-0.0695	-0.1567	-0.2215	-0.1097	-0.1152	0.8035

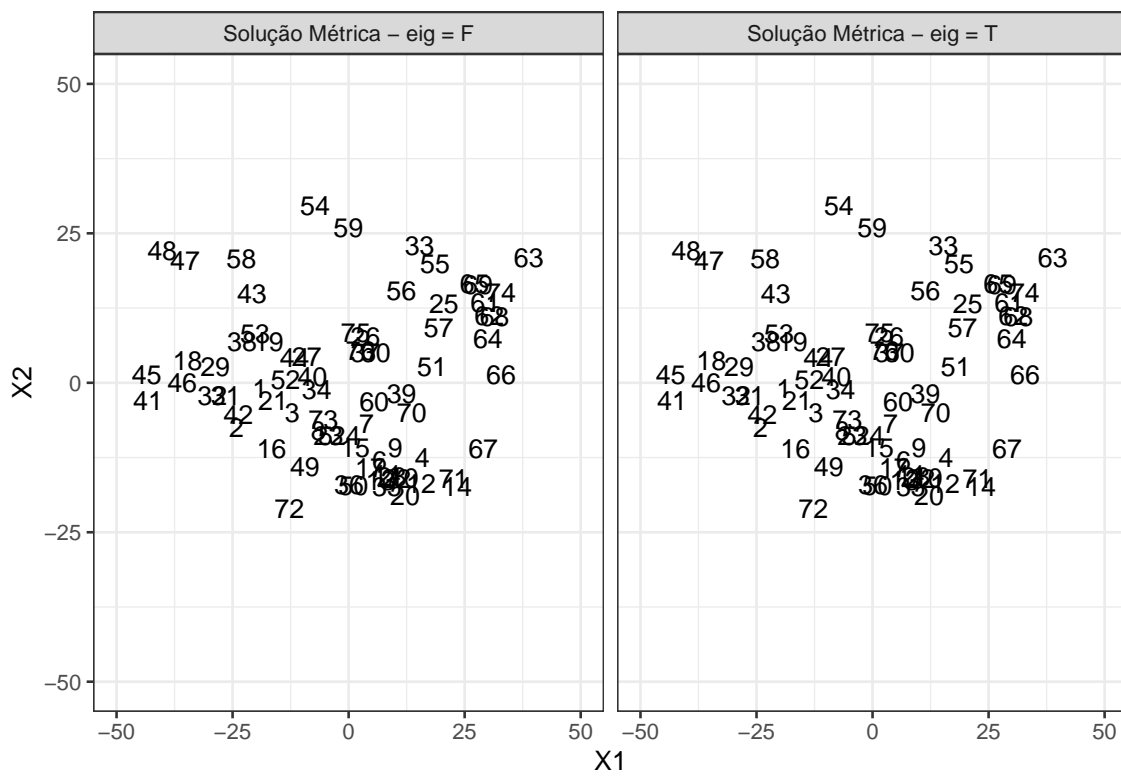


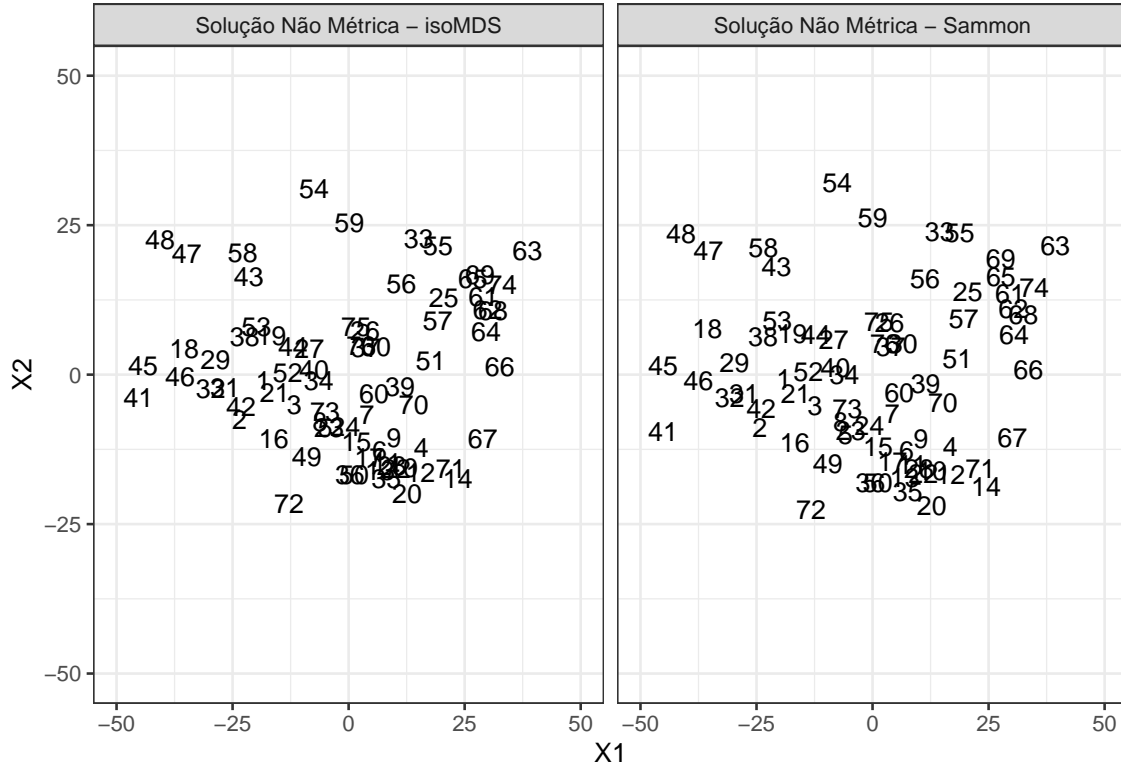


2.2 - Escalonamento Multidimensional (ou Coordenadas Principais) – Compare as soluções métricas e não-métricas.

```
> Initial stress      : 0.00731
> stress after 10 iters: 0.00486, magic = 0.461
> stress after 20 iters: 0.00468, magic = 0.500
> stress after 30 iters: 0.00468, magic = 0.500

> initial value 4.018740
> final value 3.493559
> converged
```





2.3 - Análise Fatorial Exploratória – Solução de MVS (rotacionar, se for de interesse). Em cada caso, que proporção da variância total dos dados pode ser explicada por 2 componentes? Quais variáveis mais influenciaram na redução de dimensionalidade? Represente os dados em eixos bidimensionais, identifique as observações e compare os resultados das três análises.

2.4 - Escolha uma das variáveis do banco de dados e obtenha uma tabela de contingência categorizando esta variável de acordo com faixas etárias das observações. Realize uma Análise de Correspondência e comente sobre o padrão de associação presente nesses dados.