José Luiz Borin Junior - 10723670
Lucas Barbosa de Figueiredo - 10799150
Lucas Ieks Minicz - 10723972
Luis Guilherme Aparecido de Melo - 10799251
Rodrigo Yuji Ichikawa - 10723930

Métodos Quantitativos Aplicados à Administração de Empresas I
Escalonamento Multidimensional

1. MOTIVAÇÃO

Dado classificações relativas de elementos, inferir suas posições em p dimensões de forma a:
a) manter distâncias relativas inferidas próximas às observadas
b) reduzir p a fim de se ressaltar quantas (ou quais) são as dimensões percebidas como mais relevantes pelos classificadores.

1.1 MAPAS PERCEPTUAIS

Trata-se da representação visual de objetos em um espaço de p dimensões, em que tais objetos são dispostos de acordo com as percepções dos classificadores.

Então, uma aplicação de MDS consiste em, dado avaliações de produtos por clientes (por exemplo), posicionar cada produto em um espaço de 2 ou 3 dimensões para facilitar a visualização, e inferir, com base no conhecimento do estatístico quanto às características de cada produto e de acordo com seus posicionamentos no mapa, o significado de cada eixo e o impacto dos mesmos na percepção dos usuários.

2. INTRODUÇÃO AO MODELO

O Escalonamento Multidimensional é uma análise de interdependência exploratória, utilizada como técnica de inferência sobre objetos dotados de inter-relações métricas ou não-métricas.

Inicialmente, temos as relações percebidas entre cada par de objetos e o resultado desejado é uma projeção em rp de cada ponto de forma a preservar as distâncias.

2.1 ABORDAGEM COMPOSICIONAL OU NÃO-COMPOSICIONAL

COMPOSICIONAL: baseado em atributos definidos explicitamente pelo analista e avaliados pelos respondentes, atribuindo valor semântico às dimensões resultantes.
DECOMPOSICIONAL: livre de atributos, dimensões sem valor semântico, mapas podem ser desenvolvidos para cada respondente.

3. MDS MÉTRICO CLÁSSICO

3.0. DEFINIÇÃO DE MÉTRICA

Uma função d: X x X -> [0, infinito] é uma métrica se, e somente se, satisfaz as seguintes propriedades para qualquer x, y em X:

1 – d(x, y) >= 0;
2 – d(x, y) = 0 <=> x = y
3 – d(x, y) = d(y, x)
4 – d(x, z) <= d(x, y) + d(y, z)

3.1 INTRODUÇÃO

Utilizado quando os dados de entrada e saída (distâncias entre os elementos) são de natureza métrica, representando similaridades/discrepâncias entre estes,onde podemos utilizar metodologias analíticas para encontrar o posicionamento de cada elemento.

A princípio, utilizaria-se um modelo inicialmente arbitrário e melhoraria-lo buscando minimizar o Desajuste (diferença entre as distâncias geradas e as observadas). Porém, para o MDS Métrico, existe uma solução fixa baseada em decomposição de matrizes.

3.2 DECOMPOSIÇÃO

Suponha que os elementos sob análise sejam métricos, ou seja, existe ao menos uma função métrica f sobre os elementos do tipo.

O objetivo do escalonamento multidimensional métrico é, dado uma matriz D(nxn) cujo elemento d (i x j) representa a distância, dada por alguma métrica, entre os elementos i e j, corresponder a cada elemento i um ponto Pi em Rn tal que dist(Pi, Pj) ~= d(i x j) para todos os pares de elementos.

Assumindo uma matriz **D(i x j) das distâncias** entre n elementos e um vetor **m(nx1) de importâncias** de cada elemento (onde **m' * 1(nx1) = 1**),

$$\underset{1\times}{\mathbf{m}^{\mathsf{T}}}\ \underset{I\times 1}{\mathbf{1}} = 1$$

os pontos em alguma dimensão para os elementos podem ser obtidos através do procedimento abaixo:

1 – Calcule **S = -1/2 HDH'**,

$$\underset{I\times I}{\mathbf{S}} = -\tfrac{1}{2}\boldsymbol{\Xi}\mathbf{D}\boldsymbol{\Xi}^{\mathsf{T}}$$

onde **H = I – 1(nx1) * m'**.

$$\underset{I\times I}{\boldsymbol{\Xi}} = \underset{I\times I}{\mathbf{I}} - \underset{I\times 1}{\mathbf{1}}\ \underset{\times I}{\mathbf{m}^{\mathsf{T}}}$$

2 – Encontre os **autovalores A** e **autovetores V** de S.

3 – **F = (1/n)I^(-1/2) * V * A^(1/2)** conterá, na linha i, a representação em Rn do i-ésimo elemento, onde n é o número escolhido de autovalores e autovetores.

$$\mathbf{F} = \mathbf{M}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\ \ (\text{with } \mathbf{M} = \operatorname{diag}\{\mathbf{m}\})$$

4. MDS NÃO-MÉTRICO

4.1 INTRODUÇÃO

Dados de entrada não métricos (de natureza ordinal), representados em relação à preferência ideal de cada respondente.

Não há solução fixa, portanto busca-se melhorar o modelo minimizando a estatística.

4.2 STRESS (DEFINIÇÃO E MINIMIZAÇÃO)

Diferentemente do MDS métrico, onde é possível utilizar uma fórmula fechada para se obter pontos que aproximem as distâncias inicialmente dadas, para dados não métricos é necessário definir uma função de custo e tentar obter pontos que a minimizem.

STRESS é uma função do custo de tentar representar os n elementos com seus respectivos pontos em relação às suas distâncias predefinidas. O STRESS não normalizado é definido pela seguinte fórmula:

$$\sigma(\mathbf{X}) = \sum_{i<j} (\hat{d}_{ij} - d_{ij}(\mathbf{X}))^2$$

e, normalizado, na seguinte fórmula:

$$\sigma_1(\mathbf{X}) = \sqrt{\frac{\sum_{i<j} (\hat{d}_{ij} - d_{ij}(\mathbf{X}))^2}{\sum_{i<j} d_{ij}^2(\mathbf{X})}}$$

Note que a distância inicial não é representada por ‘delta’, mas sim por ‘d chapéu’: isso acontece pois ‘d chapéu’, na verdade, trata-se de uma função monotônica aplicada sobre ‘delta’, o que é feito pois, em uma análise não-métrica, o estatístico se preocupa mais em manter a ordem das distâncias dos pontos, e não necessariamente a mesma distância inicialmente observada. Assim, utiliza-se um algoritmo específico (Método do Gradiente ou SMACOF) para minimizar o STRESS.

5. VALIDAÇÃO

A forma tradicional - recomendada por Kruskal em seu artigo de 1964 - de se realizar testes de qualidade de ajuste de MDS consiste em comparar o STRESS final obtido com a seguinte tabela, que associa níveis de STRESS com qualidade de ajuste:

STRESS	Qualidade de ajuste
0.200	pobre
0.100	justo
0.050	bom
0.025	excelente
0.000	perfeito

Outra alternativa é a validação de STRESS-por-ponto, na qual trava-se um ponto, calcula-se a média de seus resíduos e, caso essa média exceda um valor predefinido, o estatístico considera a sua remoção do modelo a fim de melhorar o nível de STRESS médio - tal remoção é justificada pelo pressuposto de que o ponto em questão resulta em um problema na coleta dos dados.

6. APLICAÇÃO

In [1]:
```
from sklearn.manifold import MDS
from sklearn.metrics import euclidean_distances

import pandas
```

In [2]:
```
data = pandas.read_csv("c_0000.csv").drop(["vx", "vy", "vz", "m", "id"], axis=1)[:1000].apply(p
andas.to_numeric)
data -= data.mean()
```

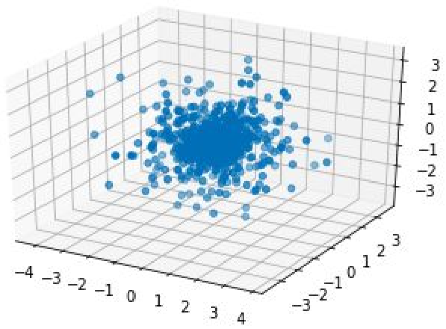
In [3]:
```
mds = MDS(n_components=3, max_iter=300000, eps=1e-9).fit(data)
pos = mds.embedding_
stress = mds.stress_
```

```
In [3]: mds = MDS(n_components=3, max_iter=300000, eps=1e-9).fit(data)
        pos = mds.embedding_
        stress = mds.stress_
```

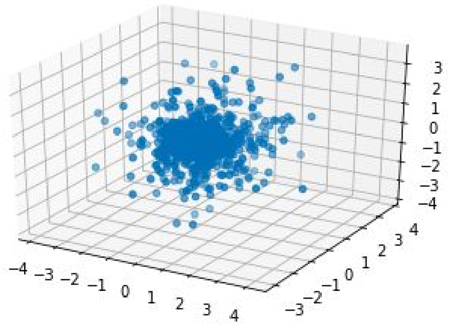
```
In [4]: stress
Out[4]: 1.6544098579173896e-05
```

```
In [5]: from matplotlib import pyplot as ppl
        from mpl_toolkits import mplot3d
```

```
In [6]: ppl.axes(projection = '3d').scatter(data["x"], data["y"], data["z"])
        ppl.show()
```




```
In [7]: ppl.axes(projection = '3d').scatter(-pos[:, 2], -pos[:, 0], -pos[:, 1])
        ppl.show()
```



7. REFERÊNCIAS

Análise de Dados Multivariados - James Lattin, J. Douglas Caroll e Paul E. Green
Análise Multivariada de Dados - Hair, Black, Babin, Anderson e Tatham
Goodness-of-Fit Assessment in Multidimensional Scaling and Unfolding - Patrick Mair e Ingwer Borg
Metric (mathematics) - Wikipedia
Metric Multidimensional Scaling (MDS): Analyzing Distance Matrices - Hervé Abdi
Multidimensional Scaling - NCSS Statistical Software