

UNIVERSIDADE FEDERAL DO PARANÁ

**HELEN EDUARDA DOS SANTOS LOURENÇO
VITOR NICKHORN KROEFF**

AVALIAÇÃO DA MCA E VARIÁVEIS SUPLEMENTARES

Curitiba
2025

HELEN EDUARDA DOS SANTOS LOURENÇO
VITOR NICKHORN KROEFF

AVALIAÇÃO DA MCA E VARIÁVEIS SUPLEMENTARES

Trabalho apresentado à disciplina CE318 –
Aprendizado não Supervisionado, do Curso
de Graduação em Estatística da Universi-
dade Federal do Paraná, como requisito para
composição parcial da nota.

Prof. Dr. Anderson Luiz Ara Souza

Contents

| | | |
|----------|---|----------|
| 1 | Introdução MCA | 2 |
| 2 | Dados | 2 |
| 3 | Aplicação da MCA | 3 |
| | Inércia | 3 |
| | Visualização | 3 |
| | Qualidade por Dimensão | 4 |
| | Visualização | 4 |
| | Variáveis Suplementares Numéricas | 5 |
| 4 | Conclusão | 6 |
| | REFERÊNCIAS | 7 |

1 Introdução MCA

A Análise de Correspondência Múltipla (MCA) é uma extensão da Análise de Correspondência Simples (CA), técnica utilizada para visualizar relações entre dados categóricos. Seu objetivo é reduzir a dimensionalidade de um conjunto de dados e representar, de forma bidimensional e simplificada, as relações entre as variáveis e os indivíduos. Além disso, na MCA, podemos analisar a quantidade de informação retida em cada dimensão (inércia).

Ainda, na MCA, temos as informações chamadas de suplementares, que são variáveis que não influenciam no cálculo da estrutura geométrica dos dados reduzidos, mas podem ser projetadas nesse espaço, ajudando na interpretação dos resultados.

2 Dados

Decidimos utilizar o conjunto de dados *poison*, disponível no pacote *FactoMineR* do software R, para aplicação da MCA. Os dados são referentes a uma pesquisa realizada com uma amostra de crianças do ensino fundamental que sofreram intoxicação alimentar. Elas foram questionadas sobre os sintomas apresentados e sobre os alimentos que consumiram.

Descrição do conjunto de dados:

```
data(poison)
```

```
# Dimensão
```

```
dim(poison)
```

```
[1] 55 15
```

```
# Variáveis
```

```
summary(poison)
```

| Age | | Time | | Sick | | Sex | | Nausea | | Vomiting | |
|----------|---------|----------|---------|---------|----|-----|----|-----------|----|----------|----|
| Min. | : 4.00 | Min. | : 0.00 | Sick_n: | 17 | F: | 28 | Nausea_n: | 43 | Vomit_n: | 33 |
| 1st Qu.: | 6.00 | 1st Qu.: | 0.00 | Sick_y: | 38 | M: | 27 | Nausea_y: | 12 | Vomit_y: | 22 |
| Median | : 8.00 | Median | : 12.00 | | | | | | | | |
| Mean | : 16.93 | Mean | : 10.16 | | | | | | | | |
| 3rd Qu.: | 10.00 | 3rd Qu.: | 16.50 | | | | | | | | |
| Max. | : 88.00 | Max. | : 22.00 | | | | | | | | |

| Abdominals | | Fever | | Diarrhae | | Potato | | Fish | | Mayo | |
|------------|----|----------|----|-------------|----|-----------|----|---------|----|---------|----|
| Abdo_n: | 18 | Fever_n: | 20 | Diarrhea_n: | 20 | Potato_n: | 3 | Fish_n: | 1 | Mayo_n: | 10 |
| Abdo_y: | 37 | Fever_y: | 35 | Diarrhea_y: | 35 | Potato_y: | 52 | Fish_y: | 54 | Mayo_y: | 45 |

| Courgette | | Cheese | | Icecream | |
|-----------|----|-----------|----|-------------|----|
| Courg_n: | 5 | Cheese_n: | 7 | Icecream_n: | 4 |
| Courg_y: | 50 | Cheese_y: | 48 | Icecream_y: | 51 |

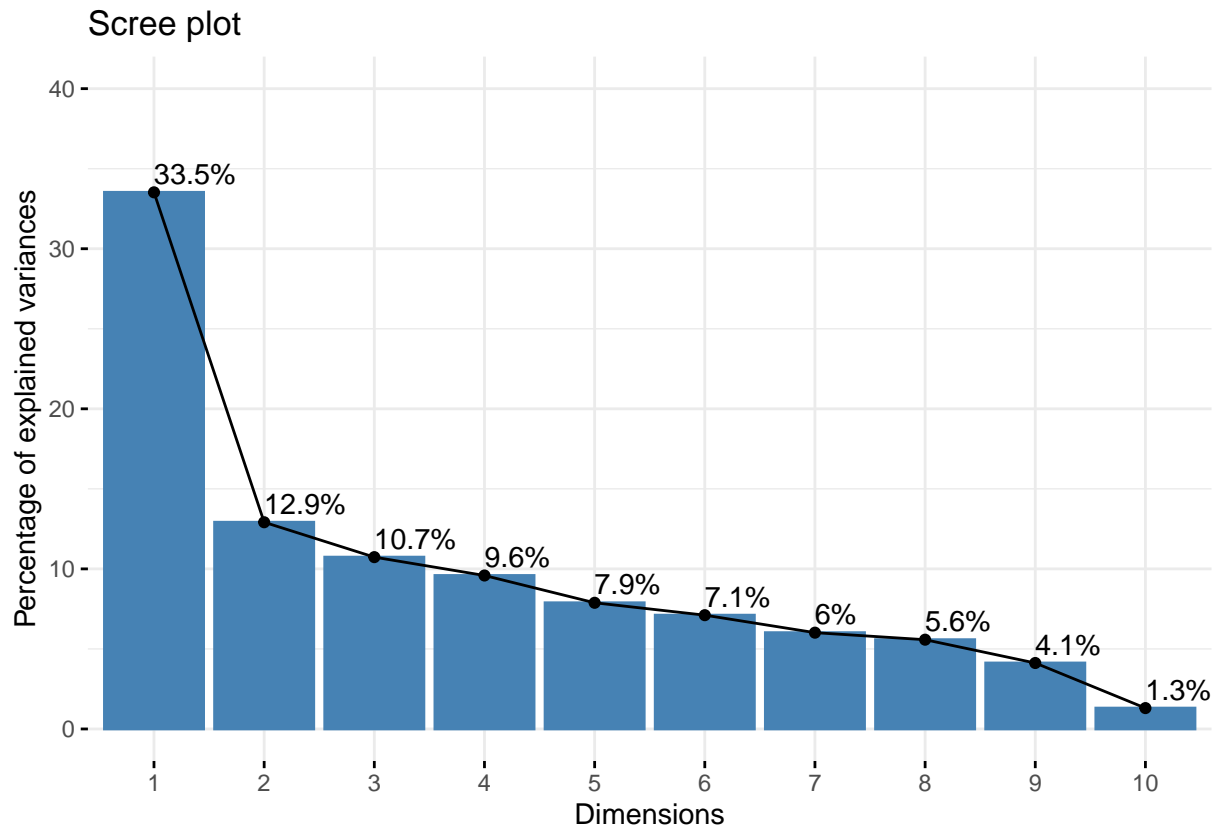
Vamos utilizar como informações suplementares as variáveis *Age* (Idade), *Time* (Tempo), *Sick* (Se está doente ou não), e *Sex* (Gênero). As outras variáveis serão utilizadas para o cálculo da dimensão reduzida.

3 Aplicação da MCA

Inércia

```
# Ajuste da MCA e configuração das informações suplementares
mca_res <- MCA(poison, quanti.sup = 1:2,
               quali.sup = 3:4, graph = FALSE)

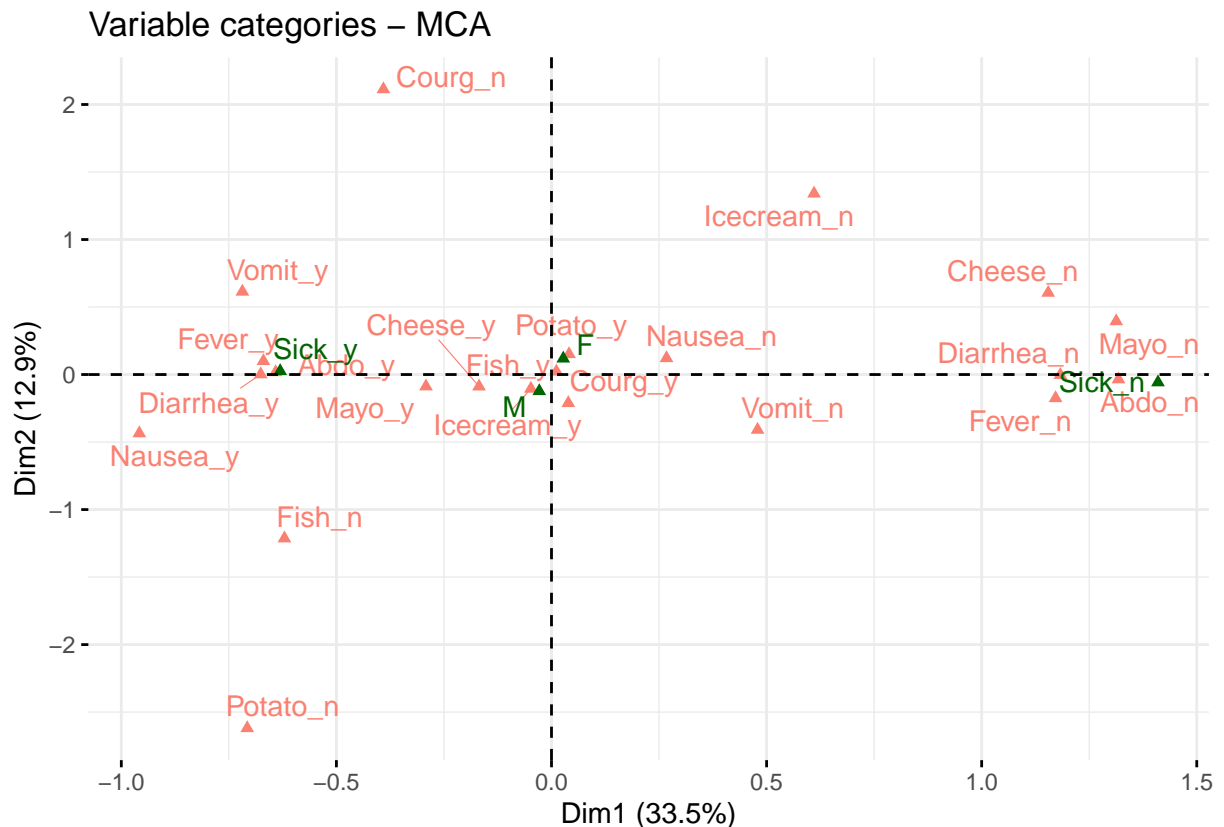
# Quantidade de informação retida em cada dimensão
fviz_screepplot(mca_res,
                addlabels = TRUE,
                ylim = c(0, 40))
```



No gráfico acima, que representa a inércia de cada uma das dimensões, temos que a primeira dimensão retém 33,5% da informação (variabilidade dos dados), a segunda 12,9%, e assim por diante. Juntas, a primeira e a segunda dimensão explicam 46,4% da variabilidade dos dados, com inércia residual de 53,6%.

Visualização

```
# Visualização variáveis categóricas dimensões 1 e 2
fviz_mca_var(mca_res,
             repel = TRUE,
             ggtheme = theme_minimal(),
             col.ind = "grey90",
             alpha.ind = 0.95,
             col.var = "salmon"
)
```



No gráfico acima, a dimensão 1 parece separar os indivíduos que não tiveram sintomas e não comeram alguns alimentos (lado direito), daqueles que tiveram sintomas e comeram outros alimentos (lado esquerdo). Fica claro também os indivíduos que não apresentaram os sintomas geralmente não comeram queijo e maionese e comeram abobrinha (*Courg_y*). A variável suplementar *Sex*, que indica o gênero da criança, não parece influenciar nos sintomas.

Qualidade por Dimensão

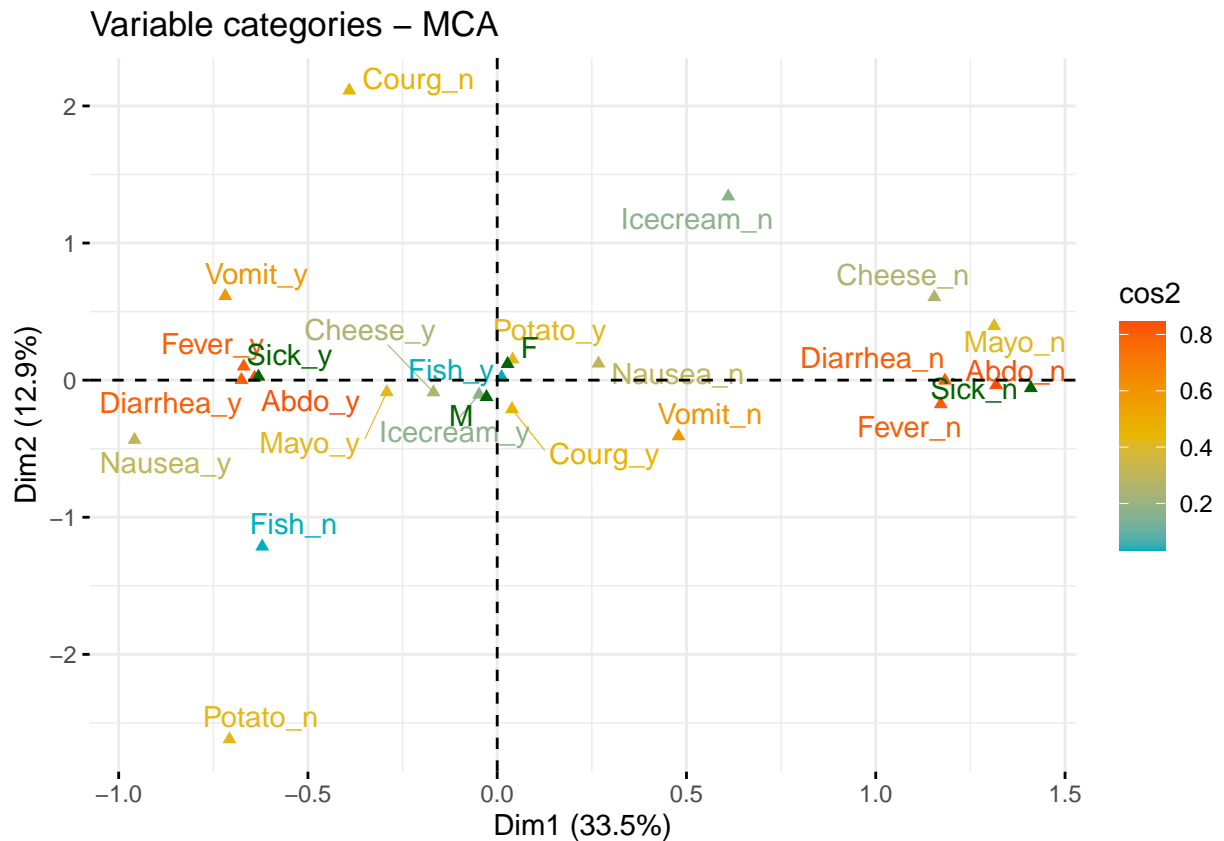
Podemos extrair a correlação das ocorrências com cada dimensão da análise, para verificar se elas representam bem as suas variações. A dimensão 1 explica muito bem a variação das dores abdominais, por exemplo.

```
mca_var <- get_mca_var(mca_res)
round(head(mca_var$cos2, 6), 2)
```

| | Dim 1 | Dim 2 | Dim 3 | Dim 4 | Dim 5 |
|----------|-------|-------|-------|-------|-------|
| Nausea_n | 0.26 | 0.05 | 0.25 | 0.00 | 0.02 |
| Nausea_y | 0.26 | 0.05 | 0.25 | 0.00 | 0.02 |
| Vomit_n | 0.34 | 0.25 | 0.01 | 0.11 | 0.00 |
| Vomit_y | 0.34 | 0.25 | 0.01 | 0.11 | 0.00 |
| Abdo_n | 0.85 | 0.00 | 0.00 | 0.01 | 0.00 |
| Abdo_y | 0.85 | 0.00 | 0.00 | 0.01 | 0.00 |

Visualização

```
fviz_mca_var(mca_res, col.var = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE,
  ggtheme = theme_minimal())
```



Com as variáveis representadas em um mapa de cores de acordo com a qualidade por dimensão, fica claro que a variável *Fish*, que indica se a criança comeu peixe ou não, não está sendo bem representada por essas duas dimensões.

Abaixo, puxando os valores apenas dessa variável, vemos que ela está muito mais correlacionada com as dimensões 3 e 4.

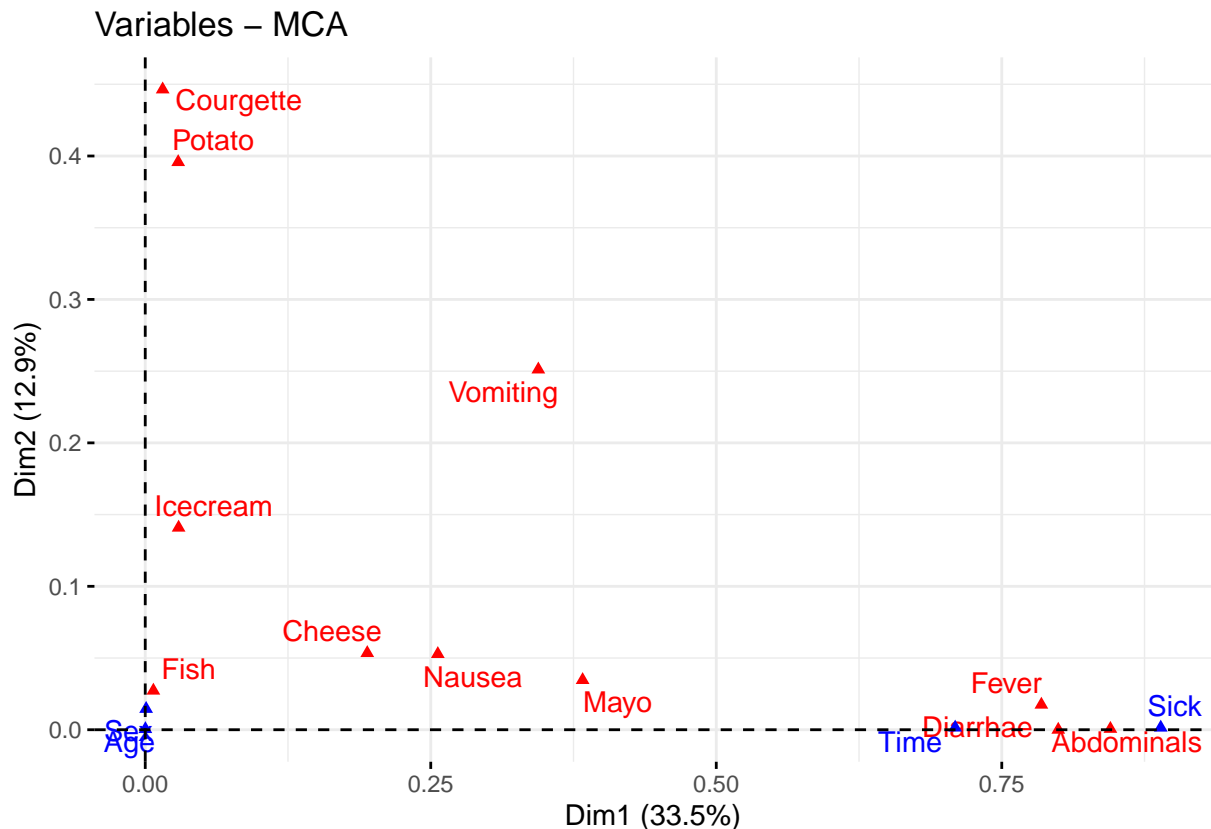
```
round(mca_var$cos2[13:14,],2)
```

| | Dim 1 | Dim 2 | Dim 3 | Dim 4 | Dim 5 |
|--------|-------|-------|-------|-------|-------|
| Fish_n | 0.01 | 0.03 | 0.25 | 0.57 | 0 |
| Fish_y | 0.01 | 0.03 | 0.25 | 0.57 | 0 |

Variáveis Suplementares Numéricas

Escolhemos *Age* e *Time* como informações suplementares anteriormente, e ambas são variáveis numéricas. Uma forma de visualizar a correlação das variáveis com as dimensões e incluir essas duas informações, seria o gráfico abaixo.

```
fviz_mca_var(mca_res, choice = "mca.cor",
             repel = TRUE)
```



Apesar da idade não estar correlacionada com essas duas dimensões, o tempo tem uma correlação forte com a dimensão 1 e fica próximo dos sintomas mais fortes, como febre, diarreia e dores abdominais, além da variável *Sick*, que indica se a criança passou mal.

Coordenadas variáveis numéricas:

```
mca_res$quanti
```

```
$coord
```

| | Dim 1 | Dim 2 | Dim 3 | Dim 4 | Dim 5 |
|------|-------------|-------------|-------------|-------------|------------|
| Age | -0.01198562 | -0.02141403 | -0.25866390 | 0.20587533 | 0.01725215 |
| Time | -0.84214754 | -0.03894016 | -0.05226732 | -0.04636355 | 0.02743571 |

4 Conclusão

A ausência de certos alimentos, como queijo e maionese, está associada à ausência de sintomas, enquanto o consumo de outros alimentos aparece vinculado aos sintomas. A variável suplementar tempo foi importante para a análise, uma vez que os indivíduos com maior tempo de exposição ou reação estão localizados na direção oposta ao centro da dimensão 1, o que indica a associação entre o tempo e sintomas mais graves. A MCA facilitou a visualização e interpretação dos padrões nos dados, explicando, em apenas duas dimensões, uma parcela significativa da variabilidade dos dados.

REFERÊNCIAS

ARA, A. Análise de Correspondência e MCA (Notas de aula). 2025. Curso de Estatística, Universidade Federal do Paraná, Curitiba, 2025.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, 25(1), 1–18. <https://doi.org/10.18637/jss.v025.i01>

MCA - Multiple Correspondence Analysis in R: Essentials - Articles - STHDA. 24 de setembro de 2017, <https://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/114-mca-multiple-correspondence-analysis-in-r-essentials/>.