



---

# TRABALHO - PARTE 1

## RELATÓRIO - QUESTÃO 3

---

ELIANE RAMOS DE SIQUEIRA RA:155233

GUILHERME PAZIAN RA:160323

HENRIQUE CAPATTO RA:146406

MURILO SALGADO RAZOLI RA:150987

Disciplina: **ME731 - Análise Multivariada**

Professor: **Caio Lucidius Naberezny Azevedo**

## 1. Introdução

Este trabalho visa medir o grau de associação entre o estado da saúde mental dos indivíduos e o status sócio-econômico de seus respectivos pais. Sendo assim, pretendemos encontrar um padrão nestes dados por meio de uma análise de correspondência simples.

O banco de dados que vamos analisar é constituído de 1760 observações, com duas variáveis categorizadas que são: *Saúde Mental*, dividida em quatro categorias (Boa, Presença fraca de sintomas, Presença moderada de sintomas, Debilitado) e *Status Sócio-Econômico dos Pais*, dividido em cinco ordens (A(Boa),B,C,D e E(Baixo)). Os softwares utilizados foram: *R*<sup>1</sup>, versão 3.4.0 e *Rstudio*<sup>2</sup>, versão 1.0.1.

Primeiramente, realizamos um Teste Qui-Quadrado para testarmos a hipótese de independência entre as duas variáveis citadas acima. Depois, caso for indicada a dependência, procuraremos indícios de relações entre as variáveis auxiliados por uma análise descritiva dos perfis médios. Por fim, analisaremos a estrutura deste comportamento utilizando a análise de correspondência simples (ver Azevedo(2017) e Johnson e Wichern (2007)).

## 2. Análise descritiva

Na Tabela 1, apresentamos os dados obtidos. A Lei de Probabilidade geradora da tabela de contingência é a Multinomial, com tamanho  $n=1760$  e 20 categorias.

Tabela 1: Tabela de contingência com os dados obtidos

	A (Alto)	B	C	D	E (Baixo)
Boa	121	57	72	36	21
Presença fraca de Sintomas	188	105	141	97	71
Presença Moderada de Sintomas	112	65	77	54	54
Debilitado	186	60	94	78	71

Obtidos os dados, realizamos o teste Qui-quadrado para identificarmos independência. A hipótese que desejamos testar é:

$H_0 : p_{ij} = p_{i.}p_{.j}$  vs  $H_1 : p_{ij} \neq p_{i.}p_{.j}$ , para pelo menos um par  $(i, j)$ , onde  $p_{ij}$  é a proporção de indivíduos na população total para a  $i$ -ésima categoria da variável *Saúde Mental* e da  $j$ -ésima categoria da variável *Status Sócio-Econômico dos Pais*.  $p_{i.}$  é a proporção de indivíduos na categoria “ $i$ ” da variável *Saúde Mental*.  $p_{.j}$  é a proporção de indivíduos na categoria “ $j$ ” da variável *Status Sócio-Econômico dos Pais*.

Os resultados para o teste qui-quadrado usual são: O quantil teórico de uma distribuição qui - quadrado é:  $\chi^2_{12} = 21,03$ . O

<sup>1</sup><https://cran.r-project.org/>

<sup>2</sup><https://www.rstudio.com/>

valor da estatística qui-quadrado é:  $Q = 31,18$  e seu respectivo p-valor é 0,0529. Pelos resultados anteriores vemos que o teste rejeitou a hipótese de que há independência entre as variáveis.

Como discutimos anteriormente, o fato do teste qui-quadrado indicar dependência necessita de um estudo mais aprofundado dessa estrutura, calculando os perfis médios para cada variável, analisando-as de forma abrangente possível e, posteriormente, aplicando a Análise de Correspondência.

Na Tabela 2, observamos a proporção estimada dividindo a quantidade de indivíduos em cada casela pelo total de entrevistados. Pode-se verificar que, na primeira variável, a categoria que possui maior quantidade de indivíduos, com 34,49, têm pais com alto status sócio-econômico e, para a segunda variável, a categoria “presença fraca de sintomas” é a que tem maior proporção de entrevistados, por volta de 34,2%. Podemos ver que 10,68% dos indivíduos tem presença fraca de sintomas e tem pais com alto status sócio-econômico e 10,57 dos indivíduos é debilitados e os pais tem alto status sócio-econômico.

Tabela 2: Proporções estimadas (pelo total de observações) x 100

	A (Alto)	B	C	D	E (Baixo)	Total
Boa	6,88	3,24	4,09	2,05	1,19	17,44
Presença fraca de Sintomas	10,68	5,97	8,01	5,51	4,03	34,20
Presença Moderada de Sintomas	6,36	3,69	4,38	3,07	3,07	20,57
Debilitado	10,57	3,41	5,34	4,43	4,03	27,78
Total	34,49	16,31	21,82	15,06	12,33	100,00

Na Tabela 3, há o ranqueamento para cada variável. Como podemos observar, as análises feitas anteriormente, para Tabela 2, são observadas nesta Tabela.

Tabela 3: Ranqueamento da saúde mental e do status socioeconômico dos pais

	A (Alto)	B	C	D	E (Baixo)	Rank
Boa	6,88	3,24	4,09	2,05	1,19	4
Presença fraca de Sintomas	10,68	5,97	8,01	5,51	4,03	1
Presença Moderada de Sintomas	6,36	3,69	4,38	3,07	3,07	3
Debilitado	10,57	3,41	5,34	4,43	4,03	2
Rank	1	3	2	4	5	

Na Tabela 4, de perfil das linhas, vemos que, para uma boa saúde mental, os perfis-medios relacionados aos indivíduos cujos pais pertencem aos maiores status socio-econômicos, denominados (**A,B,C**, em ordem decrescente), se destacam em relação a proporção vista na população com respectivamente 39,41%, 18,57%, 23,45%. Ou seja, há indícios de que uma boa saúde mental pode estar relacionada a melhores condições sócio-econômicas. Podemos observar que para quem está debilitado, os perfis-medios que se destacam em relação a proporção populacional são de valores 38,04%, 15,95%, 14,52% e correspondentes aos status economicos **A,D** e **E**, ou seja, há indícios que quem está debilitado pode ter uma pior situação econômica ou tem pais com o status socio-econômico mais alto.

Tabela 4: Perfil das linhas (x100)

	A (Alto)	B	C	D	E (Baixo)	Total
Boa	<b>39,41</b>	<b>18,57</b>	<b>23,45</b>	11,73	6,84	1
Presença fraca de Sintomas	31,23	<b>17,44</b>	<b>23,42</b>	<b>16,11</b>	11,79	1
Presença Moderada de Sintomas	30,94	<b>17,96</b>	21,27	14,92	<b>14,92</b>	1
Debilitado	<b>38,04</b>	12,27	19,22	<b>15,95</b>	<b>14,52</b>	1
Total	34,49	16,31	21,82	15,06	12,33	1

Na Tabela 5, vemos o perfil de coluna. Podemos ver que 34,20% dos indivíduos possuem presença fraca de sintomas, 17,44% apresentam boa saúde mental, 20,57% Presença moderada de sintomas e 27,78% estavam debilitados. Outrossim, os possuintes de pais no mais baixo status socioeconômico, **E**, 24,88% apresentam presença moderada de sintomas e 32,72% são classificados com saúde mental debilitada, destacando-se em relação as proporções encontradas na população (20,57% e 27,78%, respectivamente). Para mais, para os possuintes de pais no status **C**, 36,72% apresentam presença fraca de sintomas, porcentagem maior do que a porcentagem encontrada na população, 34,20%, ocorrendo o mesmo para indivíduos cujos pais são classificados no status socioeconômico **D**. Vê-se concomitantemente que dos entrevistados possuintes de pais no mais alto status, **A**, 19,93% delas apresentam boa saúde mental, número maior do que os 17,44% da proporção populacional e podemos ver que 30,64% dos indivíduos que tem pais no status **A** estão debilitados. Por último, dos que possuem os pais no status **B**, 19,86% possuem boa saúde mental e 36,59% têm presença fraca de sintomas, índices superiores as proporções encontradas na população. 17,44% e 34,20%, na devida ordem).

Tabela 5: Perfil das colunas (x100)

	A (Alto)	B	C	D	E (Baixo)	Total
Boa	<b>19,93</b>	<b>19,86</b>	<b>18,75</b>	13,58	9,68	17,44
Presença fraca de Sintomas	30,97	<b>36,59</b>	<b>36,72</b>	<b>36,60</b>	32,72	34,20
Presença Moderada de Sintomas	18,45	<b>22,65</b>	20,05	20,38	<b>24,88</b>	20,57
Debilitado	<b>30,64</b>	20,91	24,48	<b>29,43</b>	<b>32,72</b>	27,78
	100,00	100,00	100,00	100,00	100,00	100,00

### 3. Análise Inferencial

Na Tabela 6, apresentamos os resultados obtidos via Análise de Correspondência Simples. Observamos que 94,96 % da variabilidade é explicada pelas duas primeiras componentes, uma quantidade razoável que nos leva a considerarmos apenas elas para as próximas análises.

Tabela 6: Inércia

	Valor Singular	Inércia	Porcentagem	Porcentagem Acumulada
1	0,10	0,01	59,09	59,09
2	0,08	0,01	35,88	94,96
3	0,03	0,00	5,04	100,00

Na Figura 1, vemos o biplot. Por meio dele, observamos que os status socio-economico dos pais **B** e **C** estão relacionados à presença fraca de sintomas, ou seja, pessoas que possuem esse sintoma possuem pais naquele status econômico. Podemos ver também que quem apresenta uma presença moderada de sintomas dispõe de pais com status econômico mais baixo, nas classes **D** e **E**. Para o status econômico mais alto, nível **A**, vemos que temos dois comportamentos interessantes devido à distinção dos níveis de sintomas: o primeiro é a relação do mais alto status, **A** com o nível de saúde mental Debilitado, uma relação mais forte do que a segunda, a de **A** com Boa saúde mental.

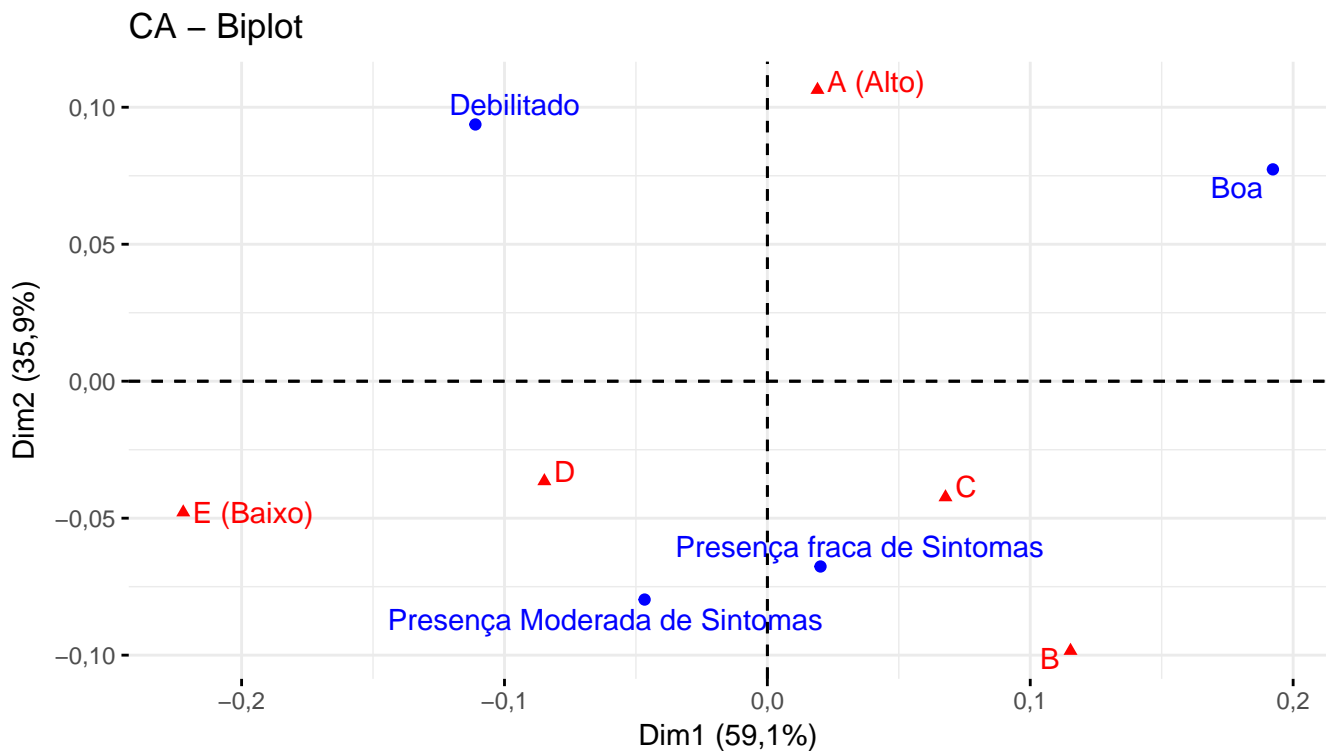


Figura 1: Biplot das componentes

#### 4. Conclusão

Podemos ver que primeramente há uma relação de dependência entre as duas variáveis analisadas, *Saúde Mental* e *Status Sócio-Econômico dos Pais*, verificadas através do Teste Qui-Quadrado. Com o perfil das linhas e colunas, conjecturamos alguns padrões que poderiam ser identificadas através da Análise de Correspondência, podemos ver pelo perfil das colunas que para o status sócio-econômico **A** os indivíduos estão mais relacionados ou a uma boa saúde ou a está debilitado. Podemos ver nesse perfil que tem menor status sócio-econômico está mais relacionado a presença moderada de sintomas ou está debilitado. Os indivíduos cujos pais estão nos status sócio-econômicos **C** e **D** estão mais relacionados a presença fraca de sintomas. Para entrevistados que cujos pais tem status sócio\_econômicos mais baixo, **E**, ou possuem presença moderada de sintomas ou está debilitado. Para indivíduos no qual os pais tem status **B**, ou possuem uma boa saúde mental ou presença fraca de sintomas.

Depois, ao ajustarmos um modelo de Análise de Correspondência Simples, 94,96% da variabilidade dos dados é explicada pelas duas primeiras componentes deste modelo, o que nos levou a considerarmos apenas estas duas para o restante das análises.

Vemos, através do Biplot, que os status **B** e **C** estão mais relacionadas à presença fraca de sintomas, **D** e **E** mais relacionadas a presença moderada e **A**, alusiva tanto ao status Debilitado, quanto ao Bom, porém a primeira relação é mais forte que a segunda. De modo geral, as relações não são contundentes, dificultando a identificação de padrões definidos.

Portanto, podemos ver que tanto a análise dos perfis médios e quanto a do biplot concordam nas conclusões obtidas.

## 5. Bibliografia

- Azevedo, C. L. N (2017). Notas de aula sobre análise multivariada de dados
- Johson, R. A. and Wichern, D. W. (2007). Applied Multivariate Statistical Analysis, 7a edição, Upper Saddle River, NJ : Prentice-Hall.