



Universidade de Brasília

DEPARTAMENTO DE ESTATÍSTICA

16 maio 2023

Entrega 4

Prof. Dr. George von Borries

Análise Multivariada 1

Aluno: Bruno Gondim Toledo | Matrícula: 15/0167636

Lista 5 - Questão 41

a) Indique cada etapa de seu estudo e respectivas conclusões

A primeira etapa numa análise de correspondências seria criar uma tabela de contingência. Depois, as matrizes de perfil de linha e de coluna. Calculamos a inércia total e a estatística Qui-Quadrado. Depois, podemos calcular a matriz dos resíduos, subtraindo do valor observado os valores esperados. Após, é conveniente apresentar gráficos que facilitem a visualização das correspondências, ou seja, as distâncias e agrupamentos entre as variáveis que ajudem a explicar o comportamento dos dados.

Para este subitem, irei apresentar a tabela de contingência para o banco de dados proposto, assim como as matrizes de perfis de linha e coluna; além de um balloonplot que ajuda a visualizar a frequência de ocorrência para cada célula. Demais tópicos serão tratados nos subitens subsequentes.

Table 1: Tabela de contingência: Método por faixa etária

	10-20	25-35	40-50	55-65	70-90
poison	2081	4495	4689	3814	2486
cookgas	45	83	53	38	34
toxicgas	330	913	663	208	40
hang	1736	3326	5417	5595	4303
drown	97	352	601	886	713
gun	537	916	927	506	232
knife	58	180	263	257	179
jump	320	642	571	661	651
other	564	1038	839	590	253

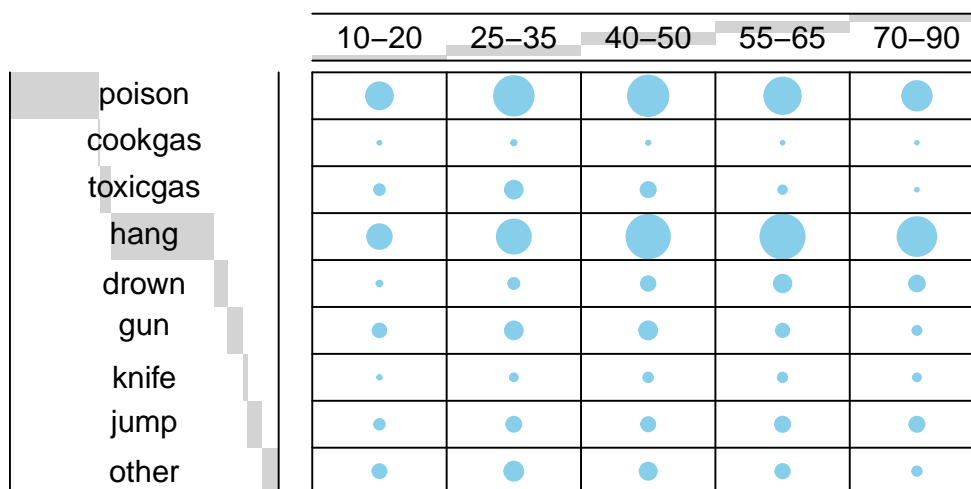
Table 2: Perfis de colunas

	10-20	25-35	40-50	55-65	70-90
poison	0.3607836	0.3763081	0.3343792	0.3037834	0.2796086
cookgas	0.0078017	0.0069485	0.0037795	0.0030267	0.0038241
toxicgas	0.0572122	0.0764337	0.0472795	0.0165671	0.0044989
hang	0.3009709	0.2784429	0.3862939	0.4456392	0.4839726
drown	0.0168169	0.0294684	0.0428582	0.0705695	0.0801935
gun	0.0930999	0.0766848	0.0661057	0.0403027	0.0260938
knife	0.0100555	0.0150691	0.0187549	0.0204699	0.0201327
jump	0.0554785	0.0537463	0.0407188	0.0526483	0.0732201
other	0.0977809	0.0868983	0.0598303	0.0469932	0.0284557

Table 3: Perfis de linha

	10-20	25-35	40-50	55-65	70-90
poison	0.1184742	0.2559066	0.2669513	0.2171364	0.1415315
cookgas	0.1778656	0.3280632	0.2094862	0.1501976	0.1343874
toxicgas	0.1532033	0.4238626	0.3077994	0.0965645	0.0185701
hang	0.0851941	0.1632232	0.2658389	0.2745743	0.2111695
drown	0.0366176	0.1328803	0.2268781	0.3344658	0.2691582
gun	0.1722258	0.2937781	0.2973060	0.1622835	0.0744067
knife	0.0618997	0.1921025	0.2806830	0.2742796	0.1910352
jump	0.1124780	0.2256591	0.2007030	0.2323374	0.2288225
other	0.1717418	0.3160780	0.2554811	0.1796590	0.0770402

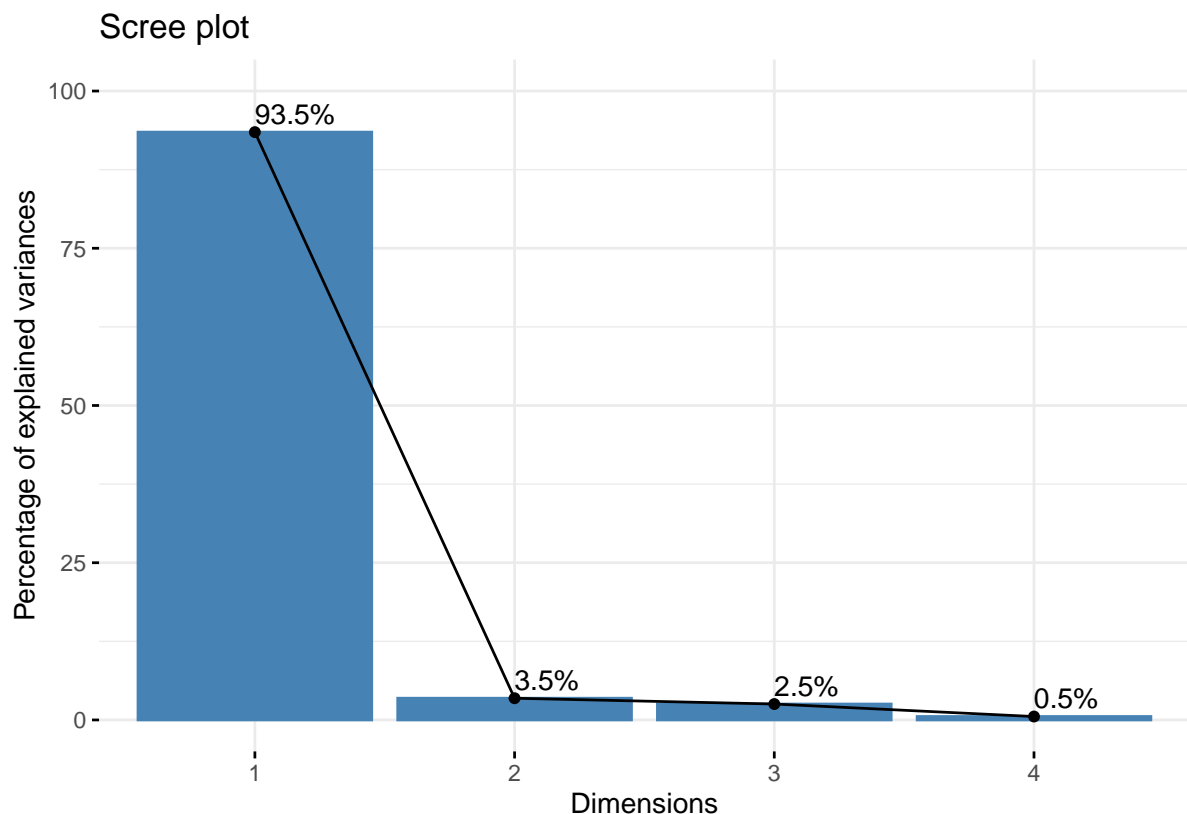
Balloonplot: Frequência de método por faixa etária



b) Calcule e interprete a estatística Qui-Quadrado e a Inércia Total.

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.0608466	93.4646752	93.46468
dim 2	0.0022493	3.4550970	96.91977
dim 3	0.0016490	2.5329913	99.45276
dim 4	0.0003563	0.5472366	100.00000

Pelo cálculo dos autovalores, conseguimos obter a porcentagem da variância explicada por cada dimensão, i.é; com isso, retiramos conclusões acerca da Inércia Total, que neste caso foi calculada em aproximadamente 93.5% logo para a primeira dimensão. Isso indica que com apenas uma dimensão conseguimos explicar mais de 90% da variação total para as categorias propostas nesta análise de correspondência.



Ainda, calculando a estatística Qui-Quadrado, sob as seguintes hipóteses:

$$\begin{cases} H_0 : \nexists \text{ associação entre as variáveis} \\ H_1 : \exists \text{ associação entre as variáveis} \end{cases}$$

```
##
## Pearson's Chi-squared test
##
## data:  tabela
## X-squared = 3462.2, df = 32, p-value < 2.2e-16
```

sob um nível de significância $\alpha = 0,05$, observamos um p-valor $< 0,001$. Logo, a decisão do teste é rejeitar a hipótese nula H_0 . Portanto, aparenta haver associação entre a faixa etária com o método de suicídio empregado.

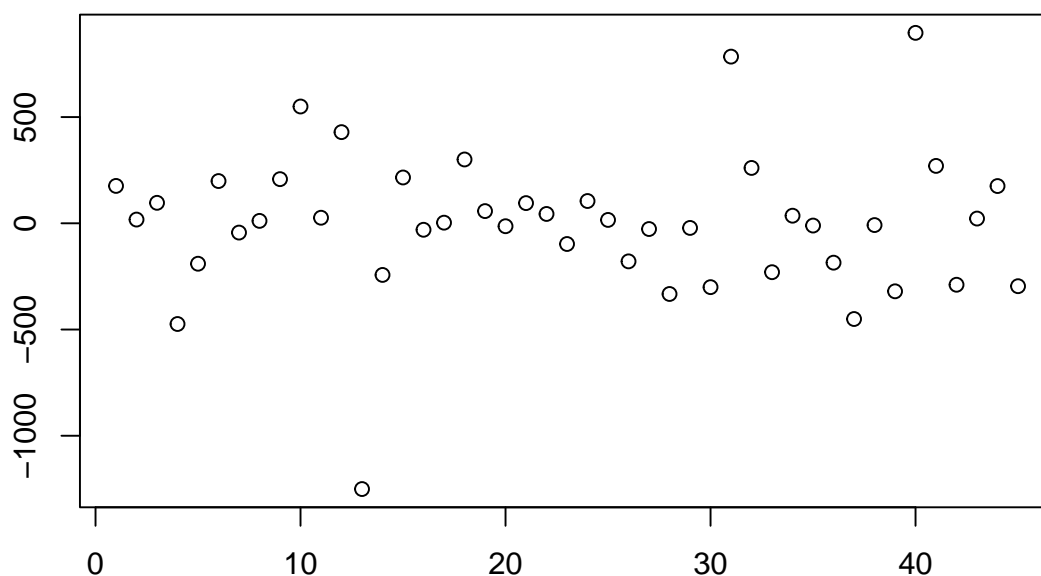
c) Análise os resíduos em cada célula da tabela e interprete os resultados

	10-20	25-35	40-50	55-65	70-90
poison	175.93964	549.794385	57.47063	-332.67698	-450.527678
cookgas	17.56019	26.174664	-13.71090	-21.72726	-8.296698
toxicgas	96.38201	429.198526	95.03449	-300.50795	-320.107066
hang	-474.04355	-1250.797883	44.00404	784.47698	896.360404
drown	-190.30458	-242.981479	-97.48684	260.63437	270.138524
gun	198.82874	215.678275	104.84750	-230.08533	-289.269189
knife	-43.62491	-30.455887	15.93236	35.79668	22.351754
jump	11.43770	2.995732	-179.16801	-10.63655	175.371122
other	207.82475	300.393667	-26.92328	-185.27396	-296.021173

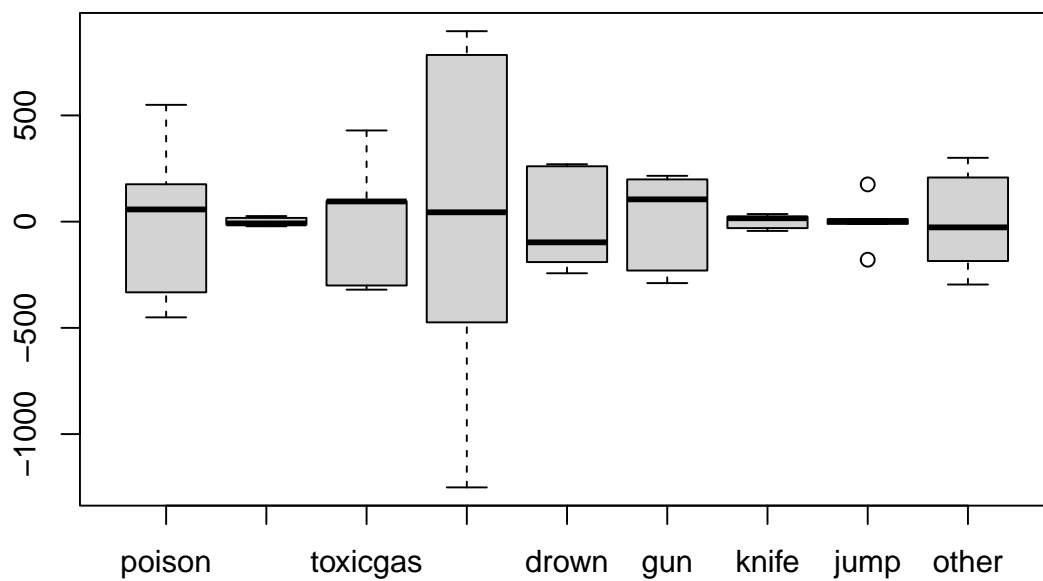
Aqui, observamos o resíduo para cada célula. Este resultado foi obtido calculando os valores totais de cada linha, e dos valores totais de cada coluna. Posteriormente, para cada i,j ; multiplicou-se esses dois valores, e posteriormente dividiu-se pelo N total de frequência observada. Com isso, obtemos os valores esperados para cada célula. Depois, subtraiu-se os valores observados dos valores esperados, obtendo assim os resíduos, que representam o quanto os valores esperados divergiram dos valores observados. Claramente, se somarmos todos os resíduos, iremos obter 0.

Podemos ainda graficar os resíduos, primeiramente com um clássico gráfico de dispersão. Podemos também avaliar a dispersão dos resíduos com boxplots, um agrupando pelo método e o outro pela faixa etária.

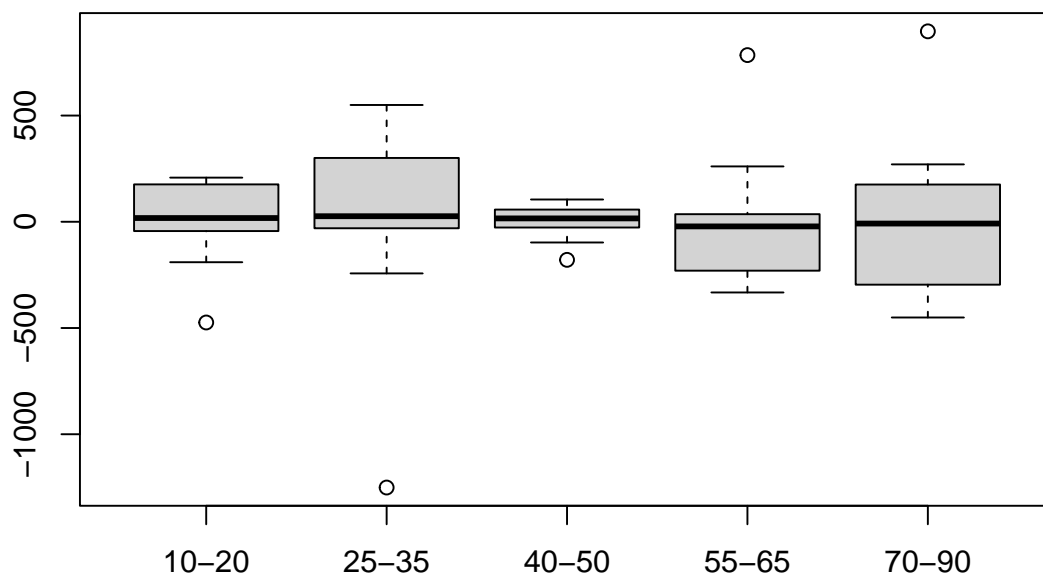
Gráfico de dispersão dos resíduos



Boxplot: Distribuição dos resíduos por método

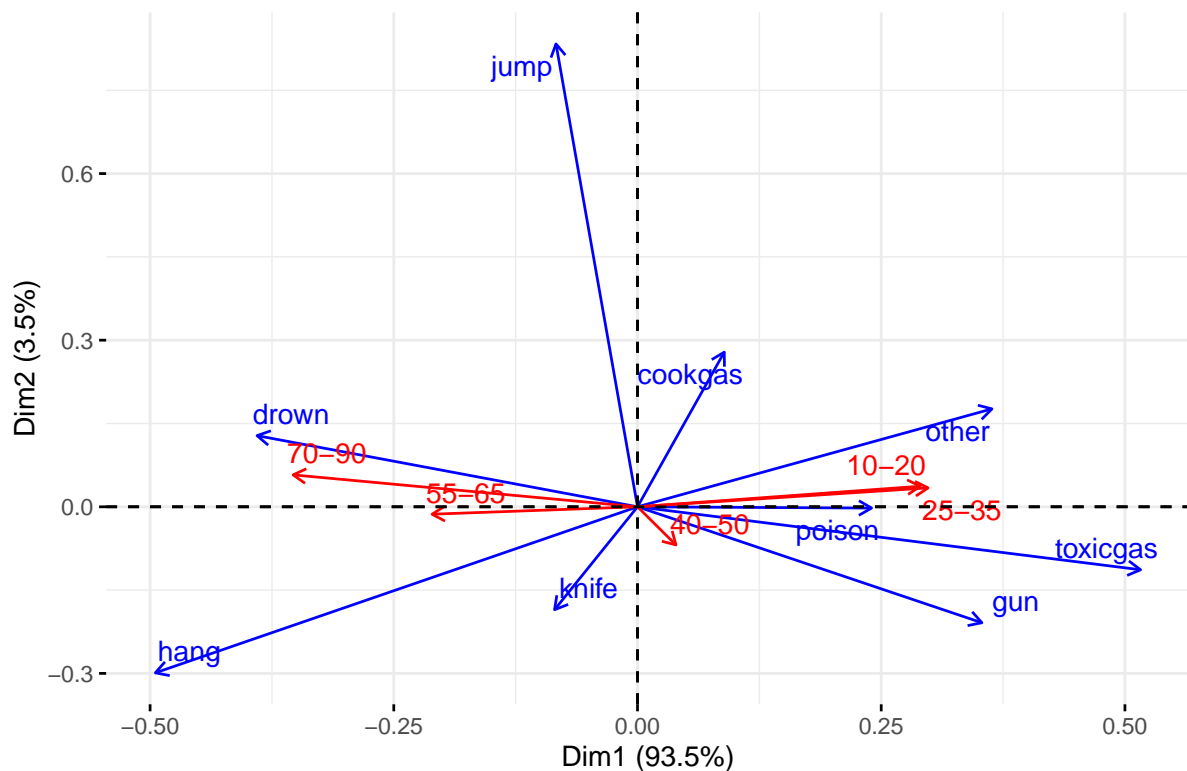


Boxplot: Distribuição dos resíduos por faixa etária



d) Apresente um mapa de correspondências e faça um breve resumo dos resultados gerais da análise.

CA – Biplot



```
## $`Dim 1`
## $`Dim 1`$row
##          coord
## drown    -0.4313269
## hang     -0.1970848
## knife    -0.1581071
## jump     -0.0891699
## poison    0.1030459
## cookgas   0.3183065
## gun       0.3598742
## other     0.3609461
## toxicgas  0.6325202
##
## $`Dim 1`$col
##          coord
## 70-90    -0.35329839
## 55-65    -0.21075164
## 40-50     0.03946632
## 10-20     0.29001828
## 25-35     0.29810810
```

Observando o *biplot* de *Gabriel* e observando a descrição da primeira dimensão das colunas e das linhas, notamos que aparenta haver um agrupamento dos métodos *drown*, *hang*, *knife* e *jump*; e por outro lado, os métodos *poison*, *cookgas*, *gun*, *other*, *toxicgas*. No caso das faixas etárias, existe um agrupamento nas idades avançadas 55-65 e 70-90; por outro lado as idades mais jovens 10-20 e 25-35. A faixa etária 40-50 está pouco explicada neste caso, não se associando a nenhum dos agrupamentos anteriores, o que faz sentido por ser a idade intermediária entre elas.

Pelo biplot, notamos que o primeiro agrupamento descrito dos métodos (*drown*, *hang*...) se associa

ao primeiro agrupamento das idades descritas (idades avançadas); enquanto o segundo agrupamento de métodos descritos (*poison*, *cookgas*, ...) se associa mais com o segundo agrupamento etário descrito (jovens).

Com isso, podemos concluir que aparenta existir uma diferença na preferência de método de suicídio de acordo com o agrupamento etário da pessoa; mais jovens tendendo a utilizar métodos como arma, gás tóxico, veneno e outros; enquanto idosos recorrendo ao enforcamento, afogamento e outros.