

UNIVERSIDADE FEDERAL DE SÃO CARLOS
DEPARTAMENTO DE ESTATÍSTICA

ESTATÍSTICA MULTIVARIADA II

Análise Fatorial de Correspondência Múltipla

Matheus Santello
RA: 818722

Docente: Prof. Dr. Pedro Ferreira Filho

São Carlos - SP
2025

1 Introdução

Muitas marcas, sejam marcas novas ou marcas já consolidadas, fazem o uso de análises estatísticas afim de obter resultados que ajudam em suas tomadas de decisões, auxiliando na gestão das mesmas, no mercado de cerveja não é diferente, resultados de análises multivariadas podem contribuir de diversas formas para compreender a demanda do produto, como por exemplo o gosto pessoal de seus consumidores para um melhor entendimento de suas preferências, portanto a realização de uma análise multivariada pode trazer resultados bastante relevantes para algumas dessas marcas.

2 Objetivo

O objetivo do estudo é a compreensão da preferência dos estudantes com relação a determinadas marcas de cerveja, e compreender a relação conjunta entre elas, afim de fornecer resultados relevantes acerca do mercado de bebidas alcoólicas.

3 Material e Método

Para solução do problema proposto, foi realizado uma análise fatorial de correspondência múltipla a partir de um conjunto de dados referente as notas, de 1 a 10 onde valores altos correspondem a uma maior preferência dos estudantes, em relação a 10 marcas de cerveja, sendo elas:

Sigla	Nome
AC	Anchor Steam
BA	Bass Ale
BE	Beck's
CO	Corona
GB	Gordon-Biersch Märzen
GS	Guinness Stout
HE	Heineken
PW	Pete's Wicked Ale
SA	Samuel Adams Boston Lager
SN	Sierra Nevada Pale Ale

A AFCM se mostrou eficiente para a análise multivariada, por se tratar de um método muito eficaz para a compreensão da estrutura de relação entre as variáveis e suas respectivas categorias.

Para que a AFCM fosse possível, as notas foram categorizadas da seguinte forma: Nota ≤ 4 , classificada como baixa; Nota > 4 e < 8 , classificada como média e Nota ≥ 8 classificada como alta.

4 Resultados

4.1 Tabela percentual de burt

Feita a AFCM, um dos primeiros resultados observados foi a tabela de burt percentual, na qual foi dividida em 2 por conta do grande numero de elementos da tabela.

	AC.alto	AC.baixo	AC.médio	BA.alto	BA.baixo	BA.médio	BE.baixo	BE.médio	CO.alto	CO.baixo	CO.médio	GB.alto	GB.baixo	GB.médio
AC.alto	18.8	0	0	33.3	50	16.7	50	50	16.7	33.3	50	33.3	33.3	33.3
AC.baixo	0	31.2	0	20	40	40	60	40	50	0	50	0	10	90
AC.médio	0	0	50	6.2	12.5	81.2	37.5	62.5	18.8	25	56.2	6.2	0	93.8
BA.alto	40	40	20	15.6	0	0	20	80	40	20	40	20	0	80
BA.baixo	33.3	44.4	22.2	0	28.1	0	66.7	33.3	33.3	11.1	55.6	11.1	22.2	66.7
BA.médio	5.6	22.2	72.2	0	0	56.2	44.4	55.6	22.2	22.2	55.6	5.6	5.6	88.9
BE.baixo	20	40	40	6.7	40	53.3	46.9	0	33.3	20	46.7	6.7	13.3	80
BE.médio	17.6	23.5	58.8	23.5	17.6	58.8	0	53.1	23.5	17.6	58.8	11.8	5.9	82.4
CO.alto	11.1	55.6	33.3	22.2	33.3	44.4	55.6	44.4	28.1	0	0	22.2	0	77.8
CO.baixo	33.3	0	66.7	16.7	16.7	66.7	50	50	0	18.8	0	0	16.7	83.3
CO.médio	17.6	29.4	52.9	11.8	29.4	58.8	41.2	58.8	0	0	53.1	5.9	11.8	82.4
GB.alto	66.7	0	33.3	33.3	33.3	33.3	33.3	66.7	66.7	0	33.3	9.4	0	0
GB.baixo	66.7	33.3	0	0	66.7	33.3	66.7	33.3	0	33.3	66.7	0	9.4	0
GB.médio	7.7	34.6	57.7	15.4	23.1	61.5	46.2	53.8	26.9	19.2	53.8	0	0	81.2
GS.alto	42.9	14.3	42.9	28.6	0	71.4	28.6	71.4	14.3	28.6	57.1	14.3	0	85.7
GS.baixo	16.7	41.7	41.7	8.3	25	66.7	33.3	66.7	33.3	8.3	58.3	16.7	16.7	66.7
GS.médio	7.7	30.8	61.5	15.4	46.2	38.5	69.2	30.8	30.8	23.1	46.2	0	7.7	92.3
HE.alto	0	60	40	10	30	60	40	60	40	0	60	10	0	90
HE.baixo	37.5	25	37.5	0	37.5	62.5	87.5	12.5	12.5	37.5	50	12.5	25	62.5
HE.médio	21.4	14.3	64.3	28.6	21.4	50	28.6	71.4	28.6	21.4	50	7.1	7.1	85.7
PW.alto	66.7	0	33.3	33.3	33.3	33.3	33.3	66.7	66.7	0	33.3	66.7	0	33.3
PW.baixo	9.1	63.6	27.3	18.2	36.4	45.5	27.3	72.7	45.5	9.1	45.5	0	9.1	90.9
PW.médio	16.7	16.7	66.7	11.1	22.2	66.7	61.1	38.9	11.1	27.8	61.1	5.6	11.1	83.3
SA.alto	40	20	40	40	20	40	20	80	20	60	20	0	20	80
SA.baixo	20	20	60	0	20	80	40	60	20	20	60	0	20	80
SA.médio	13.6	36.4	50	13.6	31.8	54.5	54.5	45.5	31.8	9.1	59.1	13.6	4.5	81.8
SN.alto	40	20	40	30	20	50	50	50	30	30	40	20	0	80
SN.baixo	0	75	25	25	25	50	25	75	25	0	75	0	0	100
SN.médio	11.1	27.8	61.1	5.6	33.3	61.1	50	50	27.8	16.7	55.6	5.6	16.7	77.8

	GS.alto	GS.baixo	GS.médio	HE.alto	HE.baixo	HE.médio	PW.alto	PW.baixo	PW.médio	SA.alto	SA.baixo	SA.médio	SN.alto	SN.baixo	SN.médio
AC.alto	50	33.3	16.7	0	50	50	33.3	16.7	50	33.3	16.7	50	66.7	0	33.3
AC.baixo	10	50	40	60	20	20	0	70	30	10	10	80	20	30	50
AC.médio	18.8	31.2	50	25	18.8	56.2	6.2	18.8	75	12.5	18.8	68.8	25	6.2	68.8
BA.alto	40	20	40	20	0	80	20	40	40	40	0	60	60	20	20
BA.baixo	0	33.3	66.7	33.3	33.3	33.3	11.1	44.4	44.4	11.1	11.1	77.8	22.2	11.1	66.7
BA.médio	27.8	44.4	27.8	33.3	27.8	38.9	5.6	27.8	66.7	11.1	22.2	66.7	27.8	11.1	61.1
BE.baixo	13.3	26.7	60	26.7	46.7	26.7	6.7	20	73.3	6.7	13.3	80	33.3	6.7	60
BE.médio	29.4	47.1	23.5	35.3	5.9	58.8	11.8	47.1	41.2	23.5	17.6	58.8	29.4	17.6	52.9
CO.alto	11.1	44.4	44.4	44.4	11.1	44.4	22.2	55.6	22.2	11.1	11.1	77.8	33.3	11.1	55.6
CO.baixo	33.3	16.7	50	0	50	50	0	16.7	83.3	50	16.7	33.3	50	0	50
CO.médio	23.5	41.2	35.3	35.3	23.5	41.2	5.9	29.4	64.7	5.9	17.6	76.5	23.5	17.6	58.8
GB.alto	33.3	66.7	0	33.3	33.3	33.3	66.7	0	33.3	0	0	100	66.7	0	33.3
GB.baixo	0	66.7	33.3	0	66.7	33.3	0	33.3	66.7	33.3	33.3	33.3	0	0	100
GB.médio	23.1	30.8	46.2	34.6	19.2	46.2	3.8	38.5	57.7	15.4	15.4	69.2	30.8	15.4	53.8
GS.alto	21.9	0	0	28.6	28.6	42.9	14.3	14.3	71.4	42.9	0	57.1	85.7	0	14.3
GS.baixo	0	37.5	0	41.7	25	33.3	16.7	58.3	25	8.3	16.7	75	8.3	25	66.7
GS.médio	0	0	40.6	23.1	23.1	53.8	0	23.1	76.9	7.7	23.1	69.2	23.1	7.7	69.2
HE.alto	20	50	30	31.2	0	0	10	60	30	10	10	80	10	30	60
HE.baixo	25	37.5	37.5	0	25	0	12.5	0	87.5	12.5	12.5	75	50	0	50
HE.médio	21.4	28.6	50	0	0	43.8	7.1	35.7	57.1	21.4	21.4	57.1	35.7	7.1	57.1
PW.alto	33.3	66.7	0	33.3	33.3	33.3	9.4	0	0	33.3	0	66.7	66.7	0	33.3
PW.baixo	9.1	63.6	27.3	54.5	0	45.5	0	34.4	0	18.2	18.2	63.6	18.2	36.4	45.5
PW.médio	27.8	16.7	55.6	16.7	38.9	44.4	0	0	56.2	11.1	16.7	72.2	33.3	0	66.7
SA.alto	60	20	20	20	20	60	20	40	40	15.6	0	60	0	40	
SA.baixo	0	40	60	20	20	60	0	40	60	0	15.6	0	20	40	
SA.médio	18.2	40.9	40.9	36.4	27.3	36.4	9.1	31.8	59.1	0	0	68.8	27.3	9.1	63.6
SN.alto	60	10	30	10	40	50	20	20	60	30	10	60	31.2	0	0
SN.baixo	0	75	25	75	0	25	0	100	0	0	50	50	0	12.5	0
SN.médio	5.6	44.4	50	33.3	22.2	44.4	5.6	27.8	66.7	11.1	11.1	77.8	0	0	56.2

Entre os resultados da tabela de burt percentual, pela diagonal principal, se destacam a ausência de notas altas para a variável "BE"(Beck's), entretanto destacam-se as proporções de 31,2% de notas altas para as variáveis "HE"(Heineken) e "SN"(Sierra Nevada Pale Ale), indicando que as duas marcas são as que obtiveram melhor desempenho, destaque para notas baixas de "SN"(Sierra Nevada Pale Ale) que correspondem a 12,5% uma das menores proporções de notas baixas entre todas as variáveis, só não menor do que "GB" com 9,4% de notas baixas.

4.2 Scree plot variância

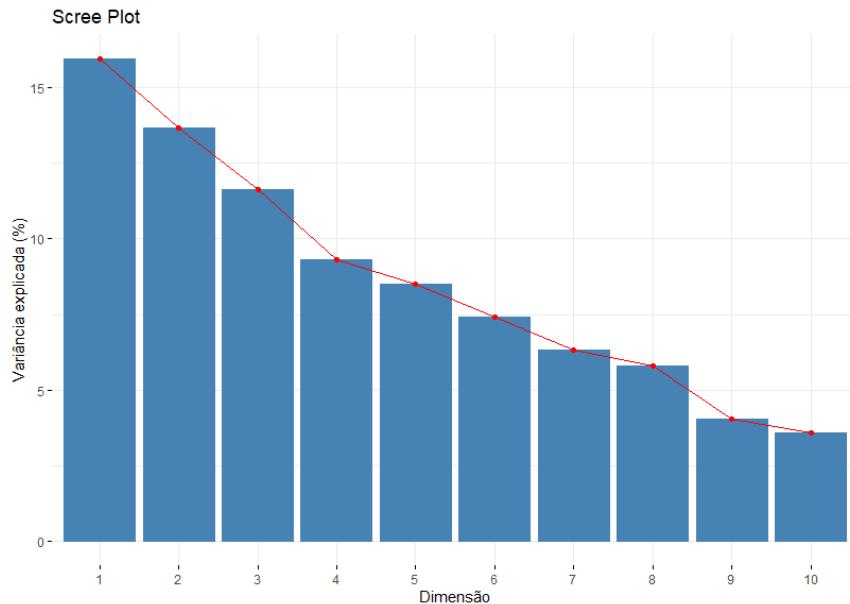


Tabela 1: Autovalores e variâncias explicadas

	eigenvalue	variance %	cumulative %
Dim.1	0.303	15.956	15.956
Dim.2	0.259	13.656	29.612
Dim.3	0.221	11.646	41.259
Dim.4	0.177	9.330	50.589
Dim.5	0.162	8.500	59.089
Dim.6	0.141	7.423	66.512
Dim.7	0.120	6.319	72.830
Dim.8	0.111	5.817	78.647
Dim.9	0.077	4.061	82.708
Dim.10	0.068	3.575	86.283
Dim.11	0.063	3.316	89.599
Dim.12	0.049	2.590	92.189
Dim.13	0.036	1.886	94.075
Dim.14	0.026	1.377	95.451
Dim.15	0.024	1.281	96.733
Dim.16	0.023	1.208	97.941
Dim.17	0.018	0.973	98.913
Dim.18	0.014	0.735	99.649
Dim.19	0.007	0.351	100.000

Apesar do primeiro plano fatorial só explicar 29.612% da variabilidade dos dados, não devemos considerar essa medida como fundamental para redução da dimensionalidade, por conta de que na AFCM a inercia não é função diretamente dos dados, sendo a mesma medida da seguinte forma: $\frac{k}{p} - 1$, portanto, essa medida é utilizada para tem uma noção básica da redução da dimensionalidade e não como fator determinante para tomada de decisão.

4.3 Contribuição das variáveis para 1^a dimensão e 2^a dimensão

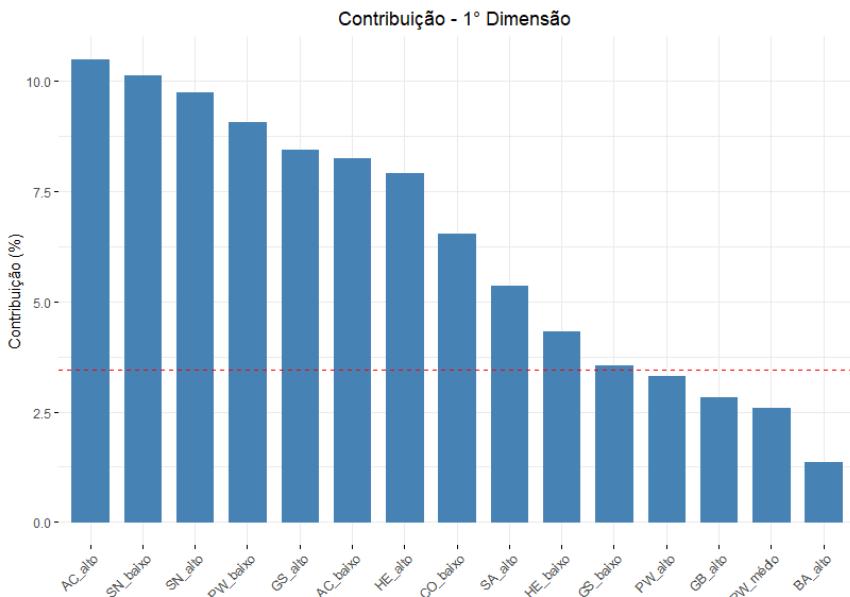


Figura 1: 15 Categorias mais contributivas - Dim 1

A maioria das categorias (15 mais contributivas) ficaram acima da média em contribuição para o primeiro fator, com exceção de "PW_alto", "GB_alto" e "PW_médio", que apesar de abaixo da média, estão muito próximas da mesma, e "BA_alto", com um valor abaixo e distante da média.

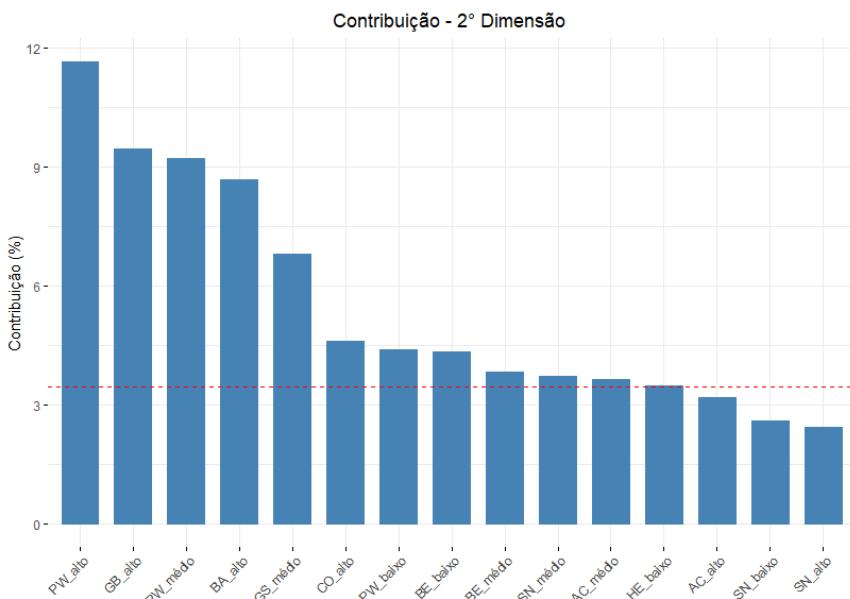


Figura 2: 15 Categorias mais contributivas - Dim 2

Já na segunda dimensão, as categorias (entre as 15 mais contributivas) que antes se encontravam abaixo da média, para esse fator sua contribuição é mais significativa, isso é um bom indicativo para redução da dimensionalidade para os dois primeiros fatores.

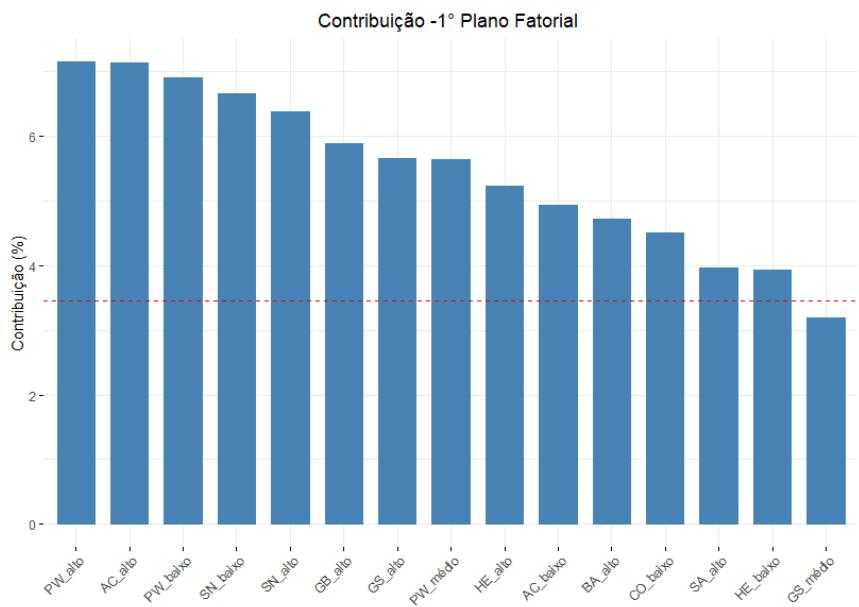


Figura 3: 15 Categorias mais contributivas - 1º plano fatorial

Na contribuição das categorias (entre as 15 mais contributivas) em relação ao primeiro plano fatorial, temos que apenas uma categoria abaixo da média, porém a mesma não se distanciou tanto do valor esperado, o restante das categorias tem contribuições acima da média, sendo um ótimo indicativo de que os dois primeiros fatores são suficientes para representar as categorias.

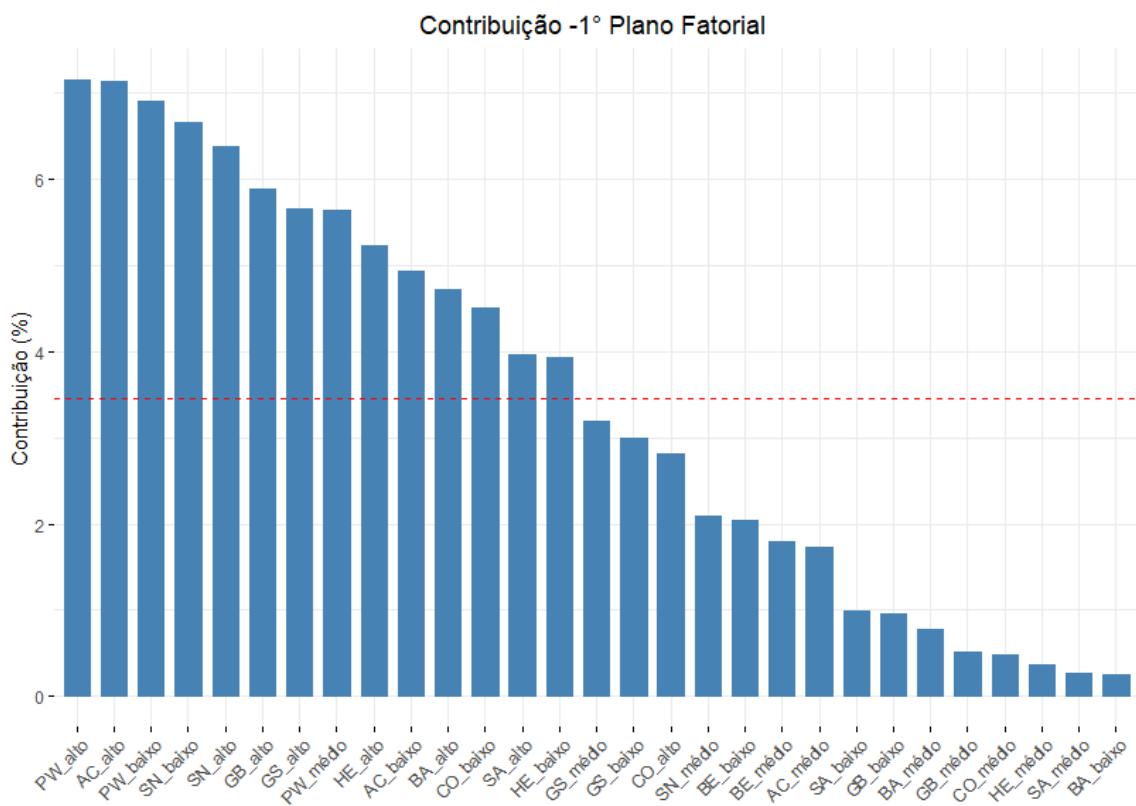


Figura 4: Contribuição de todas as categorias - 1º plano fatorial

Quando observamos as contribuições de todas as categorias em relação ao 1º plano fatorial, percebemos que a maioria das categorias muito abaixo da média, são as categorias referentes as notas médias, o que faz sentido visto que na projeção das categorias no plano fatorial, aquelas classificadas como "médio" se concentram ao redor do ponto (0,0) para os dois primeiros fatores, portanto, uma baixa contribuição das notas médias não atrapalharam no entendimento das preferências dos alunos com relação as marcas de cerveja, sendo assim seguiremos o estudo com os dois primeiros fatores.

4.4 Categorias projetadas - 1º plano fatorial

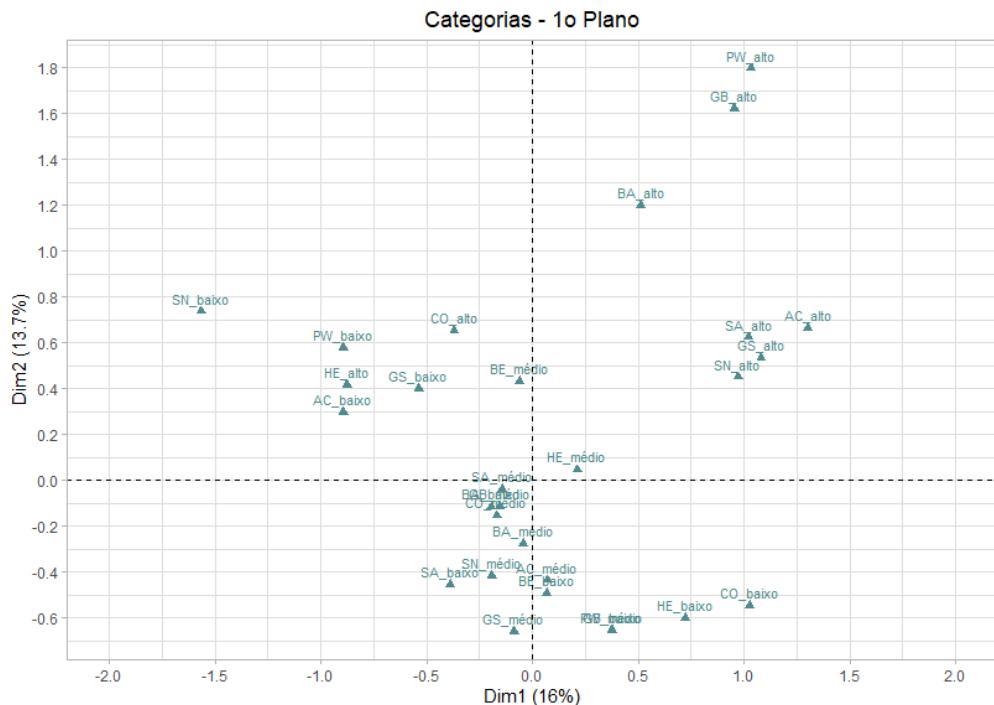


Figura 5: Categorias primeiro plano

Ao analisar a projeção das categorias no plano factorial, de cara da para destacar 3 concentrações de categorias, uma no primeiro quadrante do gráfico, outra no segundo, e por último, no terceiro quadrante, a concentração de categorias do primeiro quadrante, indica uma preferência comum entre as marcas SN, GS, SA, AC, PW, GB e BA, na concentração do segundo quadrante, temos que aqueles que preferem as marcas HE e CO, tendem a não gostar de SN, AC, GS e PW, para as categorias do quarto quadrante, aqueles que não gostam de HE, tendem a não gostar de CO de GB e de BE.

Essas relações são explicitadas pela tabela de burt percentual:

Linha	Coluna	Valor (%)
SN.baixo	PW.baixo	100.0
HE.baixo	BE.baixo	87.5
GS.alto	SN.alto	85.7
SN.baixo	GS.baixo	75.0
SN.baixo	AC.baixo	75.0
SN.baixo	HE.alto	75.0
AC.baixo	PW.baixo	70.0
PW.alto	AC.alto	66.7
GB.alto	PW.alto	66.7
PW.alto	SN.alto	66.7

Tabela 2: Maiores valores da tabela percentual.

4.5 Conclusão

Concluimos então, que a AFCM se mostrou muito eficiente para o problema, explicitando relações entre as variáveis categóricas, nos permitindo tirar conclusões acerca das preferências dos alunos em relação as marcas de cerveja.

A AFCM nos revelou que há uma preferência em comum para as marcas SN, GS, SA, AC, PW, GB e BA, e outra preferência para as marcas HE e CO, onde aqueles que às preferem, temdem a não gostar de SN, AC, GS e PW, já para aqueles que não gostam de HE e CO, também não gostam de CO, GB e BE.

4.6 Apêndice

```
1 install.packages(c("dplyr", "FactoMineR", "factoextra", "ggplot2", "gridExtra", "GDAtools"))
2 install.packages("plotly")
3 library(plotly)
4 library(dplyr)
5 library(FactoMineR)
6 library(factoextra)
7 library(ggplot2)
8 library(gridExtra)
9 library(GDAtools)

10
11 #Categorizando os dados
12
13 Dados1_cat <- Dados1_[, !(colnames(Dados1_) %in% c("Col1"))]
14
15
16 Dados1_cat <- Dados1_cat %>%
17   mutate(across(1:10, ~ case_when(
18     . <= 4 ~ "baixo",
19     . > 4 & . < 8 ~ "m_dio",
20     . >= 8 ~ "alto"
21   )))
22
23 View(Dados1_cat)
24
25
26 dados=Dados1_cat[,1:10]
27 dados
28
29 #Tabela Lógica
30
31 dados.logica=dichotom(dados, out = "numeric")
32 dados.logica
33
34 #Tabela Burt
35
36 dados.burt = burt(dados)
37 View(dados.burt)
38
39 #Tabela de burt percentual
40
41 diagonal_princ = diag(dados.burt)
42 diagonal_princ
43
44 tabela_percentual <- matrix(0, nrow = nrow(dados.burt), ncol = ncol(dados.burt))
45
46 total_individuos <- 32
47
48 for (i in 1:nrow(dados.burt)) {
49   for (j in 1:ncol(dados.burt)) {
50     if (i == j) {
51       # Diagonal principal: divide pelo total (32)
52       tabela_percentual[i, j] <- (dados.burt[i, j] / total_individuos) * 100
53     } else {
```

```

54     # Fora da diagonal: divide pelo valor da diagonal da linha (i)
55     tabela_percentual[i, j] <- (dados.burt[i, j] / diagonal_princ[i]) * 100
56   }
57 }
58 }

59 rownames(tabela_percentual) <- rownames(dados.burt)
60 colnames(tabela_percentual) <- colnames(dados.burt)

61
62
63
64

65 tabela_percentual <- round(tabela_percentual, 1)
66 View(tabela_percentual)
67 write.csv2(tabela_percentual, "tabela_percentual")
68
69
70 #AFC
71
72 res.mca <- MCA(Dados1_cat, graph = F)
73
74
75 #Autovalores, inercia e inercia total
76
77 eig.val <- get_eigenvalue(res.mca);
78 round(eig.val,3)
79
80 #Scree Plot
81
82 fviz_eig(res.mca, xlab = "Dimenso", main= "Scree Plot", linecolor="red",
83           ylab = "Varancia explicada (%)")
84
85
86 #Resultados para as variaveis
87 var <- get_mca_var(res.mca)
88
89 #Coordenadas
90 var$coord[,1:2]
91
92 #Contribucoes
93 var$contrib[,1:2]
94
95 #Cosseno 2 (Correlacao: Categoria x Dimenso)
96
97 round(var$cos2,1)
98 var$cos2
99
100 #Contribuicao das variaveis para o dimensao 1 - 15 primeiras
101
102 fviz_contrib(res.mca, choice = "var", axes = 1, top = 15) +
103   theme(plot.title=element_text(hjust=0.5))+
104   labs(title = "Contribuicao - 1 Dimenso", y = "Contribuicao (%)")
105
106 #Dim 2
107
108 fviz_contrib(res.mca, choice = "var", axes = 2, top = 15) +
109   theme(plot.title=element_text(hjust=0.5))+
110   labs(title = "Contribuicao - 2 Dimenso", y = "Contribuicao (%)")
111
112 #1o Plano

```

```

113
114 fviz_contrib(res.mca, choice = "var", axes = 1:2, top = 15) +
115   theme(plot.title=element_text(hjust=0.5))+
116   labs(title = "Contribui o -1 Plano Fatorial ", y = "Contribui o (%)")
117
118 #Contribuicao das vari veis
119
120 fviz_contrib(res.mca, choice = "var", axes = 1:2) +
121   theme(plot.title=element_text(hjust=0.5))+
122   labs(title = "Contribui o -1 Plano Fatorial ", y = "Contribui o (%)")
123
124 #Categorias no 1 Plano Fatorial
125
126 fviz_mca_var(res.mca, repel = F, labelsize = 3, geom = c("point", "text"), pointsize = 2,
127   col.var = "darkslategray4") + theme_light()+
128   labs(title = "Categorias - 1o Plano")+
129   scale_x_continuous(breaks = scales::pretty_breaks(n = 15), limits = c(-2,2)) +
130   scale_y_continuous(breaks = scales::pretty_breaks(n = 15)) +
131   theme(plot.title = element_text(hjust = 0.5))
132
133
134 #Grafico interativo para ajudar na visualizacao das categorias
135
136 gr = fviz_mca_var(res.mca, repel = F, labelsize = 3, geom = c("point", "text"), pointsize = 2,
137   col.var = "darkslategray4") + theme_light()+
138   labs(title = "Categorias - 1o Plano")+
139   scale_x_continuous(breaks = scales::pretty_breaks(n = 15), limits = c(-2,2)) +
140   scale_y_continuous(breaks = scales::pretty_breaks(n = 15)) +
141   theme(plot.title = element_text(hjust = 0.5))
142 grafico_interativo <- ggplotly(gr)
143 grafico_interativo

```