

Universidade Federal de São Carlos  
Centro de Ciências Exatas e de Tecnologia  
Departamento de Estatística

**Análise de Fatores com Dados de Marketing**  
**Estatística Multivariada**

Douglas de Paula Nestlehner  
Paulo Vitor Anselmi

Janeiro, 2023

# Capítulo 1

## Problema Apresentado

Considerando os dados representados na Tabela 1.1:

Tabela 1.1: Primeiras observações da base de dados.

Idade	RndFam	T_Clien	Vinhos	Frutas	Carne	Peixe	Doces	Ouro	V_Web
34	70.95	7.8	239	10	554	254	87	54	1
58	57.09	6.7	464	5	64	7	0	37	5
44	46.10	8.5	57	0	27	0	0	36	8
72	25.36	7.6	19	0	5	0	0	8	6
48	60.49	7.5	637	47	237	12	19	76	5
54	46.89	7.5	43	12	23	29	15	61	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Realizar uma análise de fatores (AF) completa com os dados de marketing.

- Escolher o número de fatores que achar apropriado;
- Apresentar as cargas, comunalidades e variâncias, nomeando adequadamente os fatores;
- Apresentar a análise gráfica;
- Avaliar se é necessária aplicar a rotação dos fatores. Em caso positivo, faça a rotação e apresente os resultados.

### 1.1 Descrição dos dados

A base de dados contém 2216 observações de clientes com dados de marketing de clientes de uma rede de vendas pela web. As variáveis fornecem informações de: Perfil do cliente; Produtos comprados; Sucesso (ou não) das campanhas; Desempenho do canal.

**Obs:** o dataset é maior, contém 28 variáveis, porém, foi reduzido para fins didáticos.

Na Tabela 1.2 temos representado a descrição das variáveis em estudo.

Tabela 1.2: Descrição das variáveis.

Variável	Descrição
Idade	Idade estimada em 19/12/2021 (em anos)
RndFam	Renda familiar anual do cliente
T_Clien	Tempo que é cliente da loja, base 19/12/2019 (em anos)
Vinhos	Total gasto em vinhos nos últimos dois anos
Frutas	Total gasto em frutas nos últimos dois anos
Carne	Total gasto em carne nos últimos dois anos
Peixe	Total gasto em peixe nos últimos dois anos
Doces	Total gasto em doces nos últimos dois anos
Ouro	Total gasto em produtos do tipo “ouro” nos últimos dois anos
VWeb	Número de visitas aos sites de empresa no último mês

## 1.2 Análise Exploratória

Iniciamos realizando uma breve análise exploratória, no intuito de conhecer os dados e identificar algum tipo de problema que possa afetar a análise.

Na Figura 1.1 representamos algumas medidas descritivas, calculadas para cada variável em estudo.

Idade	RndFam	TempClien	Vinhos
Min. : 23.00	Min. : 1.73	Min. : 6.700	Min. : 0.0
1st Qu.: 42.00	1st Qu.: 35.30	1st Qu.: 7.200	1st Qu.: 24.0
Median : 49.00	Median : 51.38	Median : 7.600	Median : 174.5
Mean : 50.18	Mean : 52.25	Mean : 7.641	Mean : 305.1
3rd Qu.: 60.00	3rd Qu.: 68.52	3rd Qu.: 8.100	3rd Qu.: 505.0
Max. : 126.00	Max. : 666.67	Max. : 8.600	Max. : 1493.0
Frutas	Carne	Peixe	Doces
Min. : 0.00	Min. : 0.0	Min. : 0.00	Min. : 0.00
1st Qu.: 2.00	1st Qu.: 16.0	1st Qu.: 3.00	1st Qu.: 1.00
Median : 8.00	Median : 68.0	Median : 12.00	Median : 8.00
Mean : 26.36	Mean : 167.0	Mean : 37.64	Mean : 27.03
3rd Qu.: 33.00	3rd Qu.: 232.2	3rd Qu.: 50.00	3rd Qu.: 33.00
Max. : 199.00	Max. : 1725.0	Max. : 259.00	Max. : 262.00
Ouro	visitweb		
Min. : 0.00	Min. : 0.000		
1st Qu.: 9.00	1st Qu.: 3.000		
Median : 24.50	Median : 6.000		
Mean : 43.97	Mean : 5.319		
3rd Qu.: 56.00	3rd Qu.: 7.000		
Max. : 321.00	Max. : 20.000		

Figura 1.1: Medidas descritivas.

Em geral, é possível notar que as variáveis estão em escalas diferentes. Desse modo iremos usar a matriz de correlação para a análise de fatores.

Em alguns casos também é possível notar indícios de outlier (variável “Frutas” por exemplo), entretanto seguimos com a análise, pois não temos controle sobre os dados para verificar os outliers.

Calculamos então a matriz de correlação dos dados, e representamos no grafico da Figura 1.2.

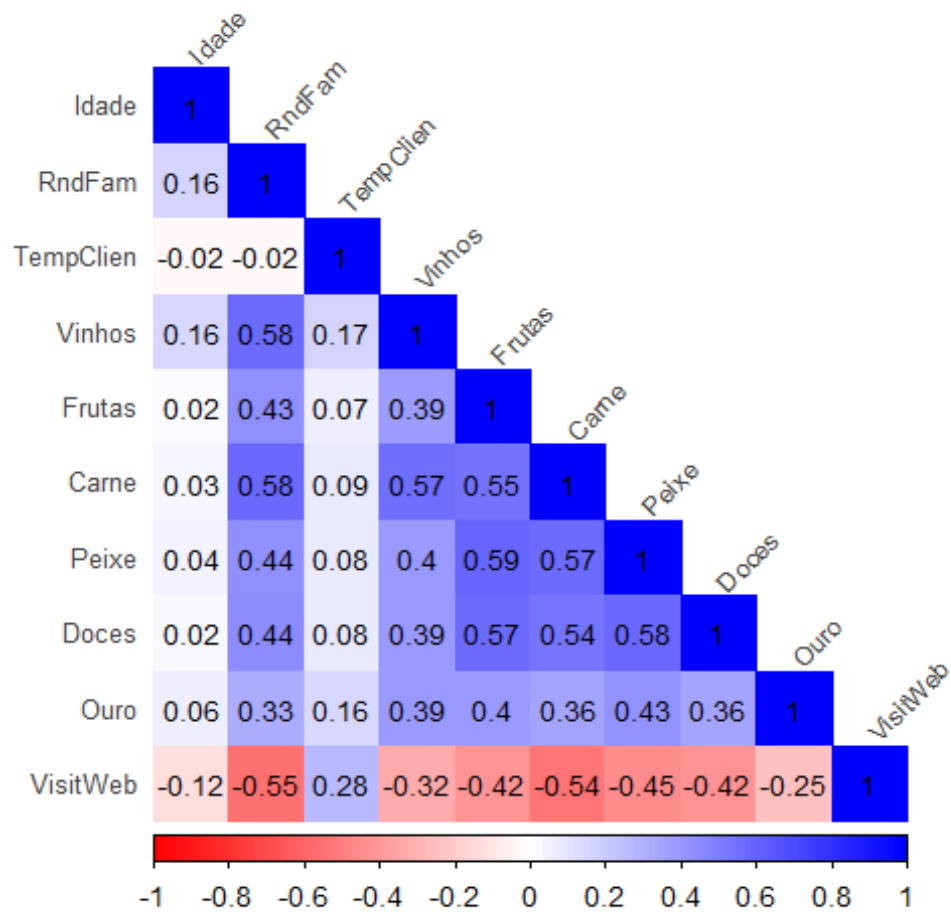


Figura 1.2: Matriz de correlação.

Alguns pontos observados pela matriz de correlação:

- É observado algumas correlações “fortes” (maior que 0.50), entretanto não ultrapassa de 0.60;
- “VisitWeb” tem correlação negativa em quase todas a variáveis;
- Destaque para as variáveis “TempClien” e “Idade” que apresenta correlação baixa para todas a variáveis;

A baixa correlação entre as variáveis pode ser um indicio que o uso do método analise de fatores pode ser inadequado.

Portanto, o calculo de medidas de adequação amostral é uma opção para avaliar se é adequado analisar os dados com a técnica de analise fatorial.

Nesse estudo, verificamos adequação amostral segundo a medida Kaiser-Meyer-Olkin, considerando a medida global (KMO) e a medida para cada varaivel (MSA)

Na Figura 1.3 temos representado as estimativas obtidas, usando função “KMO()” do pacote “psych” no R.

```
> KMO(df)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = df)
Overall MSA = 0.87
MSA for each item =
```

Idade	RndFam	TempClien	Vinhos	Frutas	Carne	Peixe
0.63	0.87	0.44	0.84	0.91	0.88	0.90
Doces	Ouro	Visitweb				
0.91	0.91	0.80				

Figura 1.3: Adequação amostral.

Obtemos um  $KMO = 0.87$  o que é adequado, entretanto observando as medidas para cada variável, temos  $MSA = 0.44$  para a variável “TempClien”, o que é inaceitável.

Portanto, para continuidade do estudo, removemos a variável “TempClien” e observamos novamente as estimativas, Figura 1.4.

```
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = df)
Overall MSA = 0.88
MSA for each item =
```

Idade	RndFam	Vinhos	Frutas	Carne	Peixe	Doces	Ouro
0.62	0.86	0.83	0.91	0.89	0.90	0.91	0.91
Visitweb							
0.86							

Figura 1.4: Adequação amostral sem a variável “TempClien”.

Agora sim temos os dados adequados para a realização da análise de fatores.

Apenas por curiosidade, também realizamos o teste de esfericidade de Bartlett para verificar adequabilidade, obtendo um p-valor  $< 0.05$  indicando que a realização de uma análise fatorial, é adequada para os dados.

## 1.3 Análise Fatorial

Realizamos a análise pelo método dos componentes principais, e considerando a matriz de correlações amostrais. Na Figura 1.5 temos representado as estimativas obtidas para cada fator estimado.

```
> summary(acpcor)
Importance of components:

```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
standard deviation	2.0628	1.0414	0.90680	0.86348	0.76080	0.65614	0.64034	0.60229	0.5572
Proportion of variance	0.4728	0.1205	0.09137	0.08284	0.06431	0.04784	0.04556	0.04031	0.0345
Cumulative Proportion	0.4728	0.5933	0.68464	0.76748	0.83180	0.87963	0.92519	0.96550	1.0000

Figura 1.5: Fatores estimados.

Observamos que o primeiro componente explica cerca 47% da variação dos dados, em conjunto com o segundo 59%, e com o terceiro 68%.

Para auxiliar na escolha do numero de fatores, construímos o gráfico de cotovelo representado na Figura 1.6.

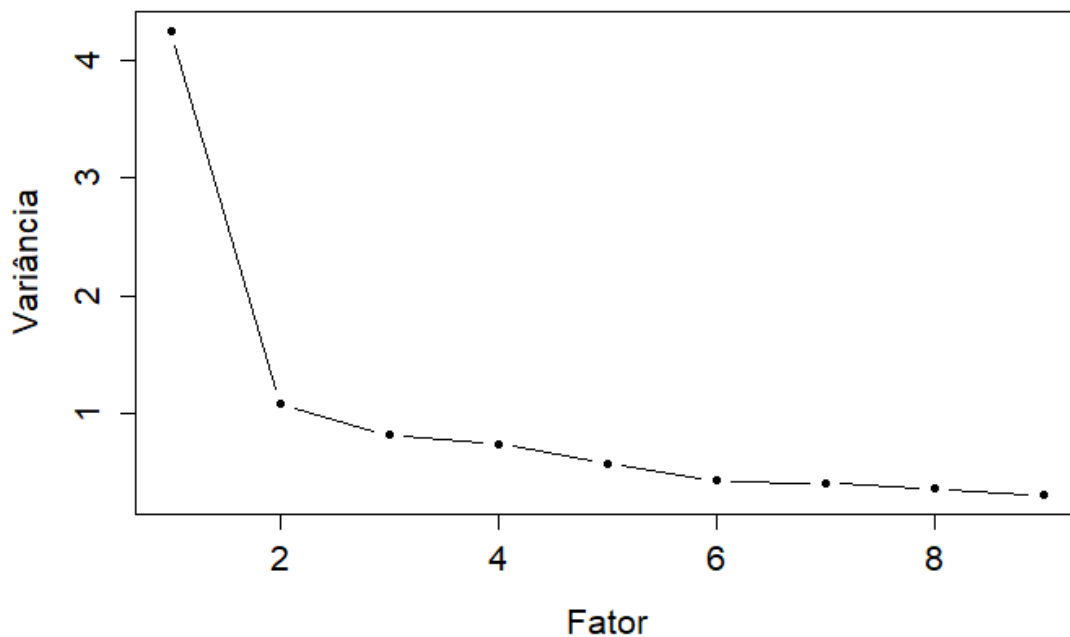


Figura 1.6: Gráfico de cotovelo.

Escolhendo três fatores, teríamos 68% da variância explicada, o que é próximo do critério dos 70%. Pelo critério dos autovalores maiores que 1, escolheríamos dois fatores, entretanto a variância explicada seria de apenas 59%.

Portanto, para continuidade do trabalho, definimos o numero de componente como sendo igual a 3.

Na Figura 1.7 temos representado as cargas fatoriais dos fatores escolhidos.

```
> carfat
```

	Fator 1	Fator 2	Fator 3
Idade	0.1356817	-0.88457164	0.188830687
RndFam	0.7552900	-0.25850565	-0.239611570
Vinhos	0.6942139	-0.25202260	0.126737656
Frutas	0.7507498	0.24028350	0.079011774
Carne	0.8198835	0.03563430	-0.175007413
Peixe	0.7724222	0.20702738	0.095845477
Doces	0.7434847	0.22340377	0.007706999
Ouro	0.5788791	0.06723487	0.671305633
Visitweb	-0.6810943	0.12379332	0.465175050

Figura 1.7: Cargas fatoriais.

Definições de cada fator:

- **Fator 1:** As variáveis “RndFam”, “Vinhos”, “Frutas”, “Carne”, “Peixe”, “Doces” e “Ouro” tem peso significativo positivo no fator, e a variável visitWeb peso significativo negativo. Portanto esse fator traz informações sobre as compras (produtos) e perfil do cliente (RndFam), podemos denota-lo como “Fator de perfil de compra”;
- **Fator 2:** Existe apenas a variável “Idade” como destaque, com contribuição significativa negativa. Com isso, denotamos como “Fator Idade”;
- **Fator 3:** As variáveis “Ouro” e “VisitWeb” são as que se destacam com contribuição significativa positiva. Denotamos esse fator como sendo: “Fator compras Ouro”.

Para auxiliar a visualização da definição dos fatores, plotamos os graficos das cargas dois a dois, representado na Figura 1.8

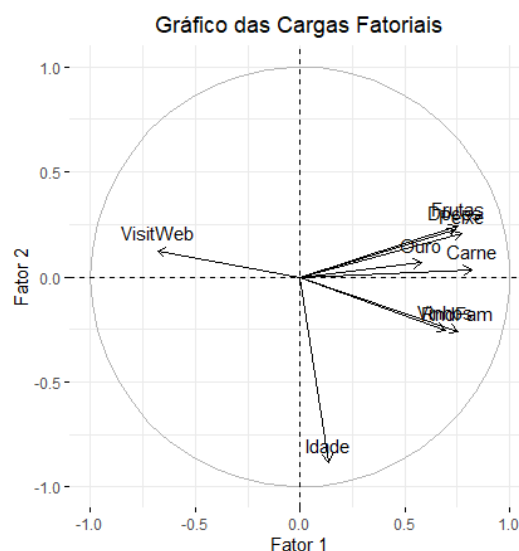


Figura 1.8: Gráficos das cargas por fator.

Obs: Não incluímos os gráficos: Fator 1 x Fator 3; e Fator 2 x Fator 3, pois as imagens ocupariam muito espaço, e a interpretação é a mesma apresentada nas definições de cada fator.

No intuito de verificar se obtivemos um bom ajuste aos dados, calculamos a comunalidade e variâncias específicas, representado na Figura 1.9.

```
> AF.mad
      comunalidades var.especificas
Idade      0.8365335      0.1634665
RndFam     0.6947019      0.3052981
Vinhos     0.5615107      0.4384893
Frutas     0.6276043      0.3723957
Carne      0.7041064      0.2958936
Peixe      0.6486827      0.3513173
Doces      0.6027381      0.3972619
Ouro       0.7902728      0.2097272
Visitweb   0.6956021      0.3043979
```

Figura 1.9: Estimativas das comunalidades e variâncias específicas.

Notamos que, a menor comunalidade é de 0.56, valor a baixo do esperado, mas aceitável.

Calculamos também a matriz de resíduos, representado na tabela abaixo:

	Idade	RndFam	Vinhos	Frutas	Carne	Peixe	Doces	Ouro	VisitWeb
Idade	0.00	-0.12	-0.18	0.11	-0.01	0.10	0.12	-0.08	-0.01
RndFam	-0.12	0.00	0.02	-0.06	-0.07	-0.07	-0.06	0.07	0.10
Vinhos	-0.18	0.02	0.00	-0.08	0.03	-0.10	-0.07	-0.08	0.12
Frutas	0.11	-0.06	-0.08	0.00	-0.06	-0.04	-0.04	-0.11	0.03
Carne	-0.01	-0.07	0.03	-0.06	0.00	-0.05	-0.08	-0.00	0.10
Peixe	0.10	-0.07	-0.10	-0.04	-0.05	0.00	-0.04	-0.10	0.01
Doces	0.12	-0.06	-0.07	-0.04	-0.08	-0.04	0.00	-0.09	0.05
Ouro	-0.08	0.07	-0.08	-0.11	-0.00	-0.10	-0.09	0.00	-0.17
VisitWeb	-0.01	0.10	0.12	0.03	0.10	0.01	0.05	-0.17	0.00

A partir disso calculamos os resíduos, vimos que o lado esquerdo da expressão vale 52% e o lado direito 3,7% com isso temos 56% mostrando que temos um ajuste satisfatório do modelo.

Seguindo para a análise se é necessário a rotação temos em primeira observação a matriz de rotação dos resíduos:

Diante disso é possível realizar a análise das cargas fatoriais com e sem rotação que será representado na Figura 1.10 :



	Idade	RndFam	Vinhos	Frutas	Carne	Peixe	Doces	Ouro	VisitWeb
Idade	0.00	-0.12	-0.18	0.11	-0.01	0.10	0.12	-0.08	-0.01
RndFam	-0.12	0.00	0.02	-0.06	-0.07	-0.07	-0.06	0.07	0.10
Vinhos	-0.18	0.02	0.00	-0.08	0.03	-0.10	-0.07	-0.08	0.12
Frutas	0.11	-0.06	-0.08	0.00	-0.06	-0.04	-0.04	-0.11	0.03
Carne	-0.01	-0.07	0.03	-0.06	0.00	-0.05	-0.08	-0.00	0.10
Peixe	0.10	-0.07	-0.10	-0.04	-0.05	0.00	-0.04	-0.10	0.01
Doces	0.12	-0.06	-0.07	-0.04	-0.08	-0.04	0.00	-0.09	0.05
Ouro	-0.08	0.07	-0.08	-0.11	-0.00	-0.10	-0.09	0.00	-0.17
VisitWeb	-0.01	0.10	0.12	0.03	0.10	0.01	0.05	-0.17	0.00

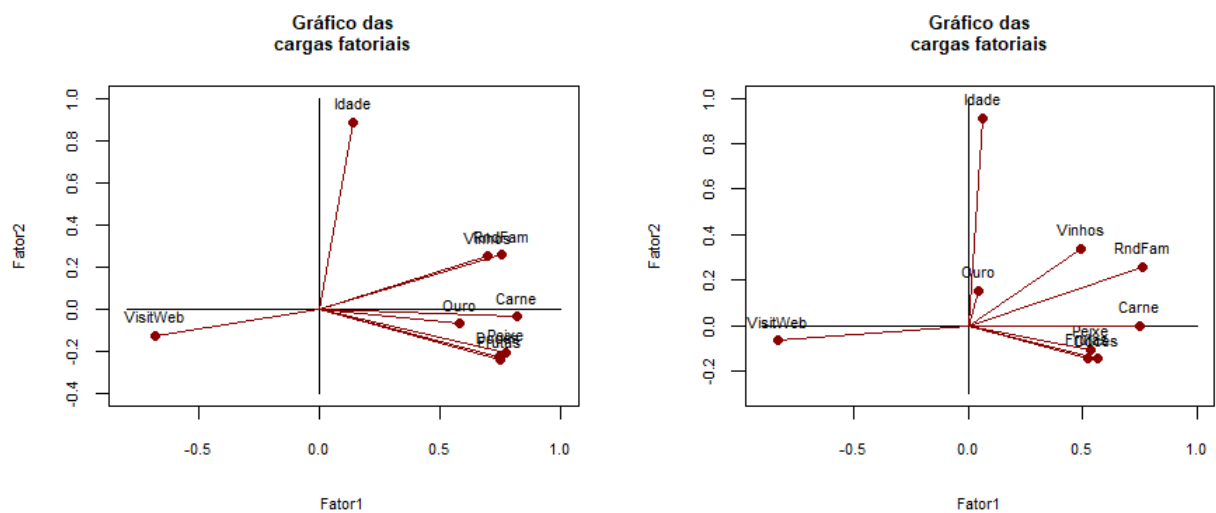


Figura 1.10: Cargas fatoriais dos dois primeiros fatores. (à esquerda: sem rotação; à direita: com rotação)

Comparando os gráficos da Figura acima, percebemos que com a rotação não há uma diferença nítida das variáveis em relação aos fatores. tendo idade e VisitWeb cargas mais elevadas, ao passo que olhando para o gráfico direito isso se mantém. Não sendo necessário a rotação. Isso pode ser visto pela tabela abaixo.

	Idade	RndFam	Vinhos	Frutas	Carne	Peixe	Doces	Ouro	VisitWeb
Fator.1	0.14	0.76	0.69	0.75	0.82	0.77	0.74	0.58	-0.68
Fator.2	0.88	0.26	0.25	-0.24	-0.04	-0.21	-0.22	-0.07	-0.12
Fator.3	0.19	-0.24	0.13	0.08	-0.17	0.10	0.01	0.67	0.46
Fator.1.1	0.06	0.76	0.49	0.52	0.75	0.53	0.56	0.04	-0.83
Fator.2.1	0.91	0.26	0.34	-0.14	-0.00	-0.10	-0.14	0.15	-0.07
Fator.3.1	0.00	0.21	0.45	0.58	0.38	0.59	0.51	0.87	-0.03