

Análise de Correspondência: introdução

Adilson dos Anjos

Objetivo

- ▶ O objetivo dessa aula é apresentar o Teste de χ^2 como introdução à Análise de Correspondência e,
- ▶ Apresentar a Análise de Correspondência (AC);
- ▶ As análises serão realizadas com uso do R.

Pacotes utilizados:

- ▶ ca

Introdução

- ▶ A distribuição χ^2 (pronuncia-se 'qui quadrado') pode ser utilizada em diferentes testes;
- ▶ Em tabelas de frequência do tipo contingência tem-se o interesse em testar a independência entre linhas e colunas;

Exemplos de tabelas de contingência:

Exemplo 1: Foi realizada uma pesquisa com consumidores da capital e do interior sobre a preferência entre duas marcas. Registrou-se a frequência de respostas para cada marca para as diferentes origens de consumidores.

Grupo	Marca A	Marca B
Capital	23	12
Interior	18	13

Exemplo 2: Foi realizada uma pesquisa com 300 consumidores de bolachas recheadas. Registrou-se a preferência entre sabor e recheio.

Grupo	Sabor1	Sabor2	Sabor3
Recheio A	51	15	29
Recheio B	22	27	59
Recheio C	26	44	27

Exemplo 3: Considere uma pesquisa sobre tipos e sabores de sobremesas com 185 consumidores.

- ▶ No R, os dados poderiam ser inseridos da seguinte maneira:

```
sobremesa<-matrix(c(17,33,19,30,15,16,14,15,26),nc=3)
colnames(sobremesa)<-c('flan','pudim','creme')
rownames(sobremesa)<-c("Baunilha", "Chocolate", "Banana")
```

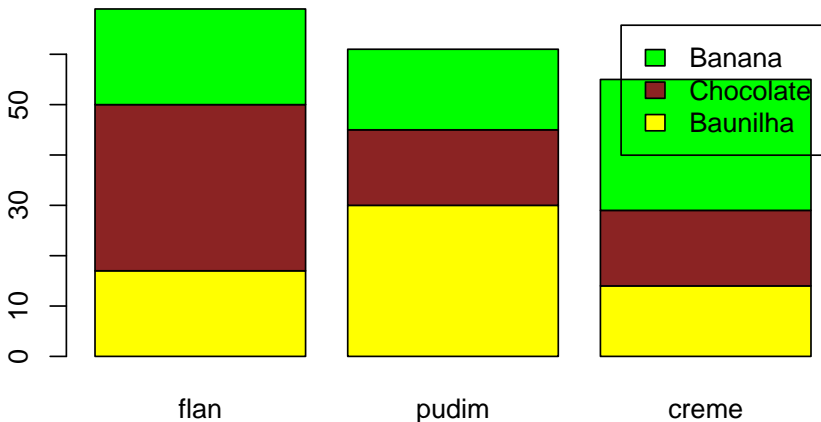

- ▶ A tabela gerada será:

sobremesa

	flan	pudim	creme
Baunilha	17	30	14
Chocolate	33	15	15
Banana	19	16	26

- ▶ Graficamente, os dados poderiam ser visualizados em um gráfico de barras:

```
barplot(sobremesa, legend=T,  
        col=c('yellow', 'brown4', 'green'))
```



- ▶ Na Análise de Correspondência utiliza-se a distribuição χ^2 para compreender a relação entre variáveis;
- ▶ Por esse motivo, vamos entender como utilizar o Teste de χ^2 antes de apresentar a Análise de Correspondência (AC).

- ▶ Na análise de uma tabela de contingência as hipóteses testadas são:
 - ▶ H_0 : as variáveis são independentes,
 - ▶ H_1 : as variáveis não são independentes.
- ▶ ou
 - ▶ H_0 : linhas e colunas são independentes,
 - ▶ H_1 : linhas e colunas não são independentes.

Como analisar a tabela de contingência:

```
chisq.test(sobremesa)
```

Pearson's Chi-squared test

```
data:  sobremesa
```

```
X-squared = 18.317, df = 4, p-value = 0.00107
```

- ▶ Considerando um nível de significância de 5%, como o $p - \text{valor} < 0,05$, rejeita-se H_0 , ou seja, existe uma associação entre linhas e colunas ou,
- ▶ Existe uma associação entre sabor e tipo de sobremesa.
- ▶ Mas como visualizar essa associação?

****Por meio da Análise de Correspondência.****

- ▶ Na Análise de Correspondência, inicialmente, faz-se um teste de χ^2 para testar se existe associação entre linhas e colunas;
- ▶ Caso a associação seja significativa ($p < 0,05$), isso indica que existe uma **CORRESPONDÊNCIA** entre as variáveis;
- ▶ Basicamente, a Análise de Correspondência é uma análise gráfica.

- ▶ Considere os dados do exemplo 1:

```
preferencia1<-matrix(c(23,12,18,13),nc=2)
colnames(preferencia1)<-c('A','B')
rownames(preferencia1)<-c("Capital", "Interior")
```

```
preferencia1
```

	A	B
Capital	23	18
Interior	12	13


```
chisq.test(preferencial)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: preferencial
```

```
X-squared = 0.14836, df = 1, p-value = 0.7001
```

- ▶ Conclusão:
- ▶ Não rejeita-se H_0 , ou seja, **não existe** associação entre sabor e recheio.

- ▶ Considere os dados do exemplo 2:

```
preferencia2<-matrix(c(51,15,29,22,27,59,26,44,27),nc=3)
colnames(preferencia2)<-c('Sabor1','Sabor2','Sabor3')
rownames(preferencia2)<-c("Recheio A", "Recheio B", "Recheio C")
```

```
preferencia2
```

	Sabor1	Sabor2	Sabor3
Recheio A	51	22	26
Recheio B	15	27	44
Recheio C	29	59	27

```
chisq.test(preferencia2)
```

Pearson's Chi-squared test

```
data:  preferencia2
```

```
X-squared = 45.879, df = 4, p-value = 2.61e-09
```

- ▶ Conclusão:
- ▶ Rejeita-se H_0 , ou seja, **existe** associação entre sabor e recheio.

Análise de Correspondência

- ▶ Análise de variáveis categóricas;
- ▶ Tabelas de contingência (frequências de respostas);
- ▶ Avalia a relação entre linhas e colunas;
- ▶ Objetivo é construir uma representação gráfica da associação entre amostras e variáveis;

- ▶ Permite representar linhas (produtos) e colunas (atributos) no mesmo espaço
- ▶ O termo '*análise de correspondência*' vem do fato de que linhas e colunas são combinadas para avaliar a *ligação* entre eles
- ▶ Os dados são obtidos, em geral, por um questionário
- ▶ As respostas são fornecidas para categorias possíveis para uma pergunta.

- ▶ A análise é baseada na decomposição da estatística χ^2 ;
- ▶ A análise é similar a Análise de Componentes Principais;
- ▶ São encontradas as **perfis de linhas** e **perfis de colunas** (Nuvem de pontos);
- ▶ Linhas e colunas são representadas no mesmo espaço por uma função de ligação (não abordado aqui);

- ▶ Análise de correspondência simples acontece quando há apenas duas variáveis categóricas;
- ▶ Quando há mais do que duas, pode-se generalizar a análise;
- ▶ Tem-se então a Análise de Correspondência Múltipla (ACM);
- ▶ Objetivo: estudar indivíduos, estudar as variáveis e suas categorias;

Análise de Correspondência

Exemplos:

- ▶ AC - Dados de Sobremesas (Exemplo livro Silvia)
- ▶ AC - Dados Modelos de automóveis (Mingoti, 2007)

Exemplo da Sobremesa:

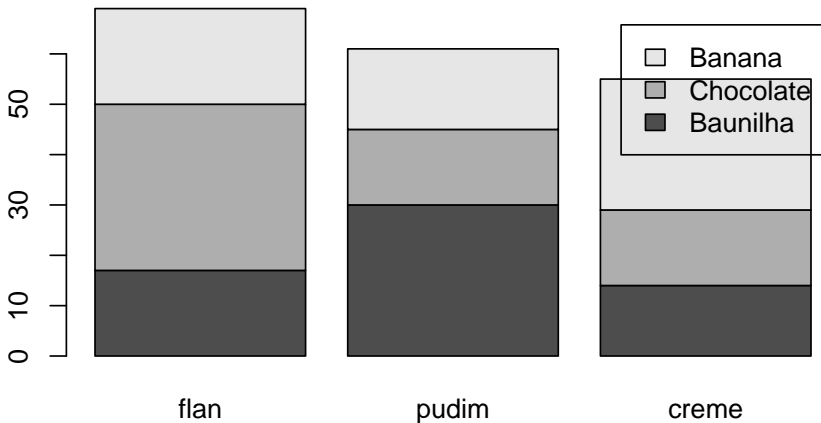
```
sobremesa<-matrix(c(17,33,19,30,15,16,14,15,26),nc=3)
colnames(sobremesa)<-c('flan','pudim','creme')
rownames(sobremesa)<-c("Baunilha", "Chocolate", "Banana")
```

```
sobremesa
```

	flan	pudim	creme
Baunilha	17	30	14
Chocolate	33	15	15
Banana	19	16	26

- Um gráfico de barras:

```
barplot(sobremesa, legend=T)
```



- ▶ Analisando a relação entre linhas e colunas:

```
chisq.test(sobremesa)
```

Pearson's Chi-squared test

```
data:  sobremesa
```

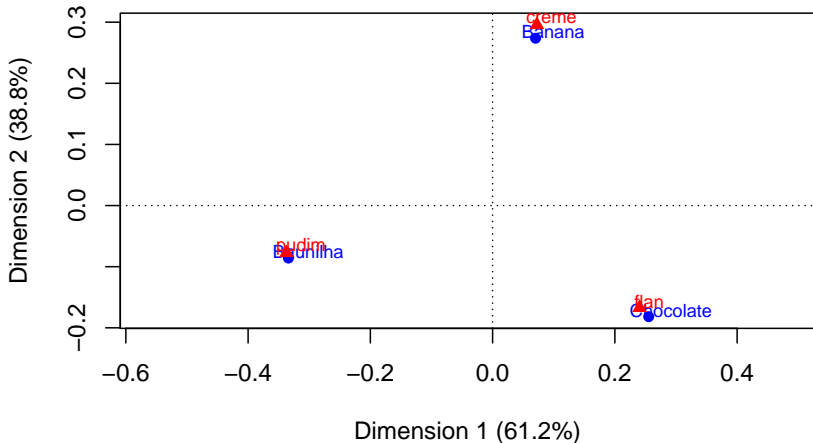
```
X-squared = 18.317, df = 4, p-value = 0.00107
```

- ▶ Como **rejeita-se** a hipótese H_0 , pode-se proceder com a Análise de Correspondência.
- ▶ Utilizando a função `ca` do pacote `ca`:

```
library(ca)  
ca.sobre<-ca(sobremesa)
```

- ▶ Graficamente, pode-se avaliar o resultado da AC:

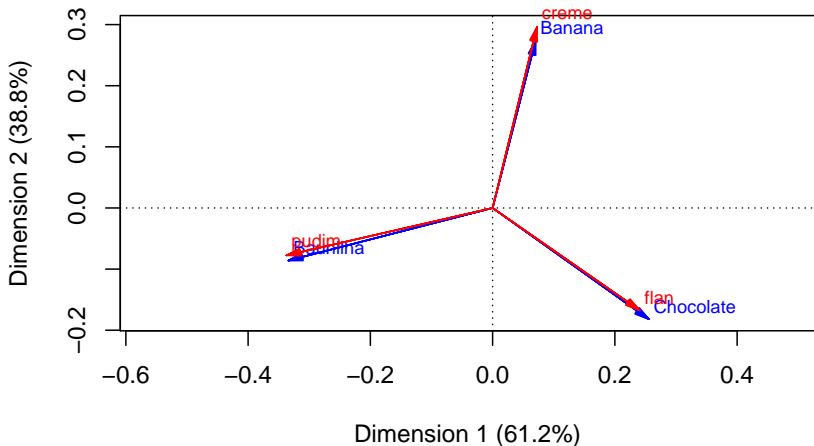
```
X11() # abrir uma nova janela gráfica  
plot(ca.sobre)
```



- ▶ Como interpretar?

- ▶ Uma outra forma de representação Gráfica:

```
plot(ca.sobre, arrows = c(TRUE, TRUE)) # asymmetric map
```



Exemplo Mingoti (2007) sobre automóveis:

- ▶ Nesse exemplo tem-se as seguintes variáveis:
- ▶ F, M -> Feminino, Masculino;
- ▶ 1,2,3-> Faixa etária 20-30/30-40/ ≥ 40 ;
- ▶ T, NT-> Trabalha, Não Trabalha;

Dados:

```
auto<-matrix(c(58,54,36,45,24,45,23,21,15,12,21,14,
               36,33,23,20,12,15,61,59,45,50,48,49,
               24,49,55,25,45,23,15,16,21,32,21,25,
               10,15,24,8,12,12,14,20,13,14,18,15),nc=4)
colnames(auto)<-c('modelo1','modelo2','modelo3','modelo4')
rownames(auto)<-c('F1T','F2T','F3T','F1NT','F2NT','F3NT',
                 'M1T','M2T','M3T','M1NT','M2NT','M3NT')
```

auto

	modelo1	modelo2	modelo3	modelo4
F1T	58	36	24	10
F2T	54	33	49	15
F3T	36	23	55	24
F1NT	45	20	25	8
F2NT	24	12	45	12
F3NT	45	15	23	12
M1T	23	61	15	14
M2T	21	59	16	20
M3T	15	45	21	13
M1NT	12	50	32	14
M2NT	21	48	21	18
M3NT	14	49	25	15

Teste de qui-quadrado:

```
chisq.test(auto)
```

Pearson's Chi-squared test

```
data: auto
```

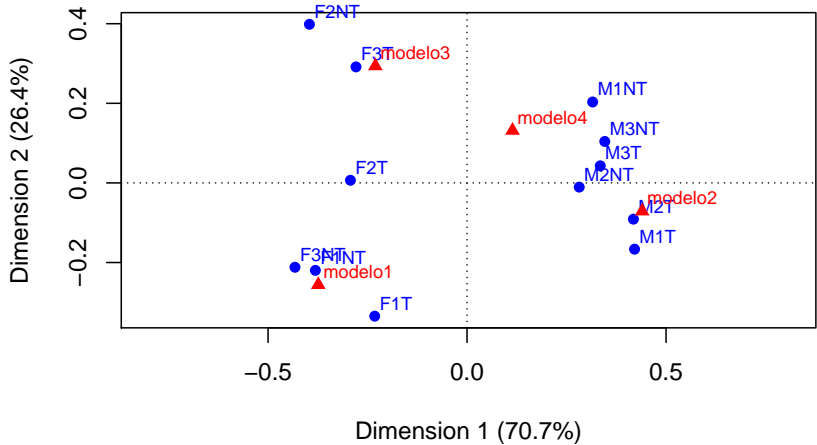
```
X-squared = 225.87, df = 33, p-value < 2.2e-16
```

- ▶ Rejeita-se H_0 .

- ▶ Realizando a AC:

```
ca.auto<-ca(auto)
```

```
plot(ca.auto)
```



- ▶ Como interpretar?