*Escola Superior de Agricultura*
*"Luiz de Queiroz"*
*Universidade de São Paulo*

# *LCE0211 – Estatística Geral*

Taciana Villela Savian
Sala 304, pav. Engenharia, ramal 478913
tvsavian@usp.br
tacianavillela@gmail.com

# Estatística Descritiva

- **Distribuição de frequências**

  O estudo das distribuições de frequências nos permite conhecer como os valores da variável em estudo se comportam.

Você consegue, a partir da distribuição de frequência, afirmar que duas variáveis (qualitativas) estão associadas?

# Estatística Descritiva

- **Distribuição de frequências – Bidimensionais**

Tabela 4. Distribuição de frequências do tipo de grão de 40 híbridos de milho, segundo as categorias de Aptidão, 2020.

| Tipo de grão | Aptidão | | Total |
|---|---|---|---|
| | Duplo Propósito | Granífero | |
| Dentado | 10 | 7 | 17 |
| Semidentado | 6 | 3 | 9 |
| Semiduro | 7 | 7 | 14 |
| Total | 23 | 17 | 40 |

# Medidas de Associação

- Obter medidas estatísticas que indiquem se **existe ou não relação entre duas variáveis qualitativas**.
- Qual a magnitude dessa relação?

Coeficiente de Contingência de Pearson (C)

- Distribuição de frequência conjunta de duas variáveis qualitativas → Vamos usar Tipo de Grão e Aptidão.

# Estatística Descritiva

- **Distribuição de frequências – Bidimensionais**

Tabela 4. Distribuição de frequências do tipo de grão de 40 híbridos de milho, segundo as categorias de Aptidão, 2020.

| Tipo de grão | Aptidão | | Total |
|---|---|---|---|
| | Duplo Propósito | Granífero | |
| Dentado | 10 | 7 | 17 |
| Semidentado | 6 | 3 | 9 |
| Semiduro | 7 | 7 | 14 |
| Total | 23 | 17 | 40 |

# Medidas de Associação

- Coeficiente de Contingência de Pearson (C)

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Em que: $n$ (ou $f_{..}$) é o tamanho da amostra (40 híbridos); $\chi^2$ é uma estatística chamada quiquadrado.

**Essa estatística ($\chi^2$) baseia-se na diferença entre as frequências absolutas ($\boldsymbol{f_{ij}}$) e as frequências esperadas ($\boldsymbol{fe_{ij}}$) caso não houvesse associação entre as variáveis.**

# Medidas de Associação

## Quem são $f_{ij}$ e $fe_{ij}$?

$f_{ij}$ indica a frequência absoluta (ou observada) na linha i e coluna j da tabela de distribuição de frequência.

$fe_{ij}$ indica a frequência esperada na linha i e coluna j da tabela de distribuição de frequência, caso as variáveis não estejam associadas.

| Tipo de Grão | Aptidão | | Total |
|---|---|---|---|
| | Duplo Propósito | Granífero | |
| Dentado | $f_{11}$=10 | $f_{12}$=7 | $f_{1.}$=17 |
| Semidentado | $f_{21}$=6 | $f_{22}$=3 | $f_{2.}$=9 |
| Semiduro | $f_{31}$=7 | $f_{32}$=7 | $f_{3.}$=14 |
| Total | $f_{.1}$=23 | $f_{.2}$=17 | $f_{..}$=40 |

# Medidas de Associação

**Quem são $\boldsymbol{f_{ij}}$ e $\boldsymbol{fe_{ij}}$ ?**

$fe_{\mathrm{ij}}$ indica a frequência esperada na linha i e coluna j da tabela de distribuição de frequência, caso as variáveis não estejam associadas.

$$\boldsymbol{fe_{ij}} = \frac{\boldsymbol{f_{i.}} \text{ x } \boldsymbol{f_{.j}}}{\boldsymbol{f_{..}}}$$

# Medidas de Associação

$$fe_{11} = \frac{f_{1.} \text{ x } f_{.1}}{f_{..}} = \frac{17\text{x}23}{50} = 9,8$$

**Importante:** $f_{11} = 10$ **e** $fe_{11} = 9,8$.

| Tipo de Grão | Aptidão | | Total |
|---|---|---|---|
| | Duplo Propósito | Granífero | |
| Dentado | $f_{11}$=10 | $f_{12}$=7 | $f_{1.}$=17 |
| Semidentado | $f_{21}$=6 | $f_{22}$=3 | $f_{2.}$=9 |
| Semiduro | $f_{31}$=7 | $f_{32}$=7 | $f_{3.}$=14 |
| Total | $f_{.1}$=23 | $f_{.2}$=17 | $f_{..}$=40 |

# Medidas de Associação

**Importante:** $\boldsymbol{f_{11} = 10}$ **e** $\boldsymbol{fe_{11} = 9,8.}$

| Tipo de Grão | Aptidão | | Total |
|---|---|---|---|
| | Duplo Propósito | Granífero | |
| Dentado | $f_{11}\ (fe_{11})$ | $f_{12}(fe_{12})$ | $f_{1.}$=17 |
| Semidentado | $f_{21}\ (fe_{21})$ | $f_{22}(fe_{22})$ | $f_{2.}$=9 |
| Semiduro | $f_{31}\ (fe_{31})$ | $f_{32}(fe_{32})$ | $f_{3.}$=14 |
| Total | $f_{.1}$=23 | $f_{.2}$=17 | $f_{..}$=40 |

# Medidas de Associação

**Importante:** $\boldsymbol{f_{11} = 10}$ **e** $\boldsymbol{fe_{11} = 9,8}$.

| Tipo de Grão | Aptidão | | Total |
|---|---|---|---|
| | Duplo Propósito | Granífero | |
| Dentado | $\mathbf{10\ (9,8)}$ | $\mathbf{7}$ $(\boldsymbol{fe_{12}})$ | $f_{1.}$=17 |
| Semidentado | $\mathbf{6}$ $(\boldsymbol{fe_{21}})$ | $\mathbf{3}$ $(\boldsymbol{fe_{22}})$ | $f_{2.}$=9 |
| Semiduro | $\mathbf{7}$ $(\boldsymbol{fe_{31}})$ | $\mathbf{7}$ $(\boldsymbol{fe_{32}})$ | $f_{3.}$=14 |
| Total | $f_{.1}$=23 | $f_{.2}$=17 | $f_{..}$=40 |

# Medidas de Associação

**Importante:** $\boldsymbol{f_{11} = 10}$ **e** $\boldsymbol{fe_{11} = 9,8}$.

| Tipo de Grão | Aptidão | | Total |
|---|---|---|---|
| | Duplo Propósito | Granífero | |
| Dentado | $\mathbf{10\ (9,8)}$ | $\mathbf{7\ (7,2)}$ | $f_{1.}$=17 |
| Semidentado | $\mathbf{6\ (5,2)}$ | $\mathbf{3\ (3,8)}$ | $f_{2.}$=9 |
| Semiduro | $\mathbf{7\ (8,1)}$ | $\mathbf{7\ (5,9)}$ | $f_{3.}$=14 |
| Total | $f_{.1}$=23 | $f_{.2}$=17 | $f_{..}$=40 |

# Medidas de Associação

- Coeficiente de Contingência de Pearson (C)

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Em que: $n$ (ou $f_{..}$) é o tamanho da amostra (40 híbridos); $\chi^2$ é uma estatística chamada quiquadrado.

**Essa estatística ($\chi^2$) baseia-se na diferença entre as frequências efetivamente observadas ($\boldsymbol{f_{ij}}$) e as frequências esperadas ($\boldsymbol{fe_{ij}}$) caso não houvesse associação entre as variáveis.**

# Medidas de Associação

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(f_{ij} - fe_{ij})^2}{fe_{ij}}, \boldsymbol{r} \text{ é } \boldsymbol{n^{o}\ linhas\ e\ s} \text{ é } \boldsymbol{n^{o}\ colunas}$$

$$\chi^2 = \frac{(f_{11} - fe_{11})^2}{fe_{11}} + \ldots + \frac{(f_{32} - fe_{32})^2}{fe_{32}}$$

| Tipo de Grão | Aptidão | | Total |
|---|---|---|---|
| | Duplo Propósito | Granífero | |
| Dentado | $\mathbf{10\ (9,8)}$ | $\mathbf{7\ (7,2)}$ | $f_{1.}$=17 |
| Semidentado | $\mathbf{6\ (5,2)}$ | $\mathbf{3\ (3,8)}$ | $f_{2.}$=9 |
| Semiduro | $\mathbf{7\ (8,1)}$ | $\mathbf{7\ (5,9)}$ | $f_{3.}$=14 |
| Total | $f_{.1}$=23 | $f_{.2}$=17 | $f_{..}$=40 |

# Medidas de Associação

$$\chi^2 = \frac{(f_{11} - fe_{11})^2}{fe_{11}} + \ldots + \frac{(f_{32} - fe_{32})^2}{fe_{32}}$$

$$\chi^2 = \frac{(10-9{,}8)^2}{9{,}8} + \frac{(7-7{,}2)^2}{7{,}2} + \frac{(6-5{,}2)^2}{5{,}2} + \ldots + \frac{(7-5{,}9)^2}{5{,}9} = 0{,}6552$$

| Tipo de Grão | Aptidão | | Total |
|---|---|---|---|
| | Duplo Propósito | Granífero | |
| Dentado | **10 (9, 8)** | **7 (7, 2)** | $f_{1.}$=17 |
| Semidentado | **6 (5, 2)** | **3 (3, 8)** | $f_{2.}$=9 |
| Semiduro | **7 (8, 1)** | **7 (5, 9)** | $f_{3.}$=14 |
| Total | $f_{.1}$=23 | $f_{.2}$=17 | $f_{..}$=40 |

# Medidas de Associação

$$\chi^2 = \frac{(10-9{,}8)^2}{9{,}8} + \frac{(7-7{,}2)^2}{7{,}2} + \frac{(6-5{,}2)^2}{5{,}2} + \ldots + \frac{(7-5{,}9)^2}{5{,}9} = 0{,}6552$$

- **Percebemos que:**

$\chi^2 = 0$ **somente quando** $f_{ij} = fe_{ij}$**, indicando a não associação entre as variáveis;**

$\chi^2 > 0$**significa que** $f_{ij} \neq fe_{ij}$ **e temos indicação de associação entre as variáveis;**

**Problema:** $\chi^2$**é um valor que pode variar de zero a infinito.**

# Medidas de Associação

- Coeficiente de Contingência de Pearson Corrigido (C*)

$$C^* = \frac{C}{\sqrt{(t-1)/t}},$$

em que: *t* é o mínimo entre o número de linhas (r) e o número de colunas (s) da tabela de distribuição de frequência conjunta (tabela de contingência), ou seja,

***t = min(r,s).***

***Em uma tabela com 3 linhas e 2 colunas, o valor de t = min(3,2) = 2***

***Em uma tabela com 3 linhas e 4 colunas, o valor de t = min(3,4) = 3***

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

$$\chi^2 = \sum_{i=1}^{s}\sum_{j=1}^{r}\frac{\left(f_{ij} - fe_{ij}\right)^2}{fe_{ij}}$$

**O Coeficiente C* assume valores entre 0 e 1 e quanto mais próximo de um mais forte é a associação entre as variáveis qualitativas estudadas.**

# Medidas de Associação

- Para o exemplo, temos:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(f_{ij} - fe_{ij}\right)^2}{fe_{ij}} = 0{,}6552$$

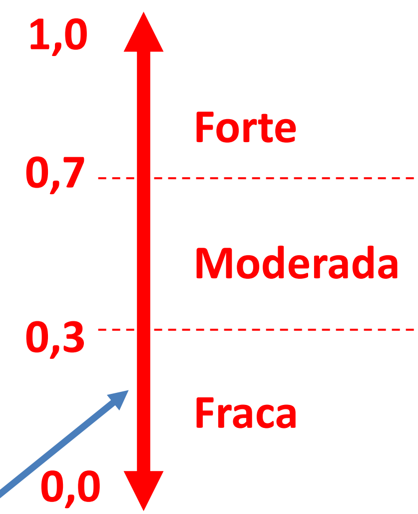
$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{0{,}6552}{0{,}6552 + 40}} = 0{,}1269$$

$$C^* = \frac{C}{\sqrt{(t-1)/t}} = \frac{0{,}1269}{\sqrt{(2-1)/2}} = 0{,}18$$

O coeficiente de contingência de Pearson corrigido indica uma associação bastante fraca entre as variáveis Tipo de grão e aptidão do híbrido.

# Medidas de Associação

$$0 < C^* < 1$$




- Para o exemplo, temos:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(f_{ij} - fe_{ij}\right)^2}{fe_{ij}} = 0{,}6552$$

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{0{,}6552}{0{,}6552 + 40}} = 0{,}1269$$

$$C^* = \frac{C}{\sqrt{(t-1)/t}} = \frac{0{,}1269}{\sqrt{(2-1)/2}} = 0{,}18$$

O coeficiente de contingência de Pearson corrigido indica uma associação bastante fraca entre as variáveis Tipo de grão e aptidão do híbrido.

# Medidas de Associação

- Exercício:

  Calcular (e interpretar) o Coeficiente de Contingência de Pearson Corrigido para:

  - Tipo de Grão versus Resistência à Ferrugem;
  - Aptidão versus Resistência à Ferrugem;