

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

LCE0211 – Estatística Geral

Taciana Villela Savian
Sala 304, pav. Engenharia, ramal 478913

tvsvavian@usp.br

tacianavillela@gmail.com

Já vimos - Relembrando

- Na **Aula 01** vimos quais os tipos de variáveis: Qualitativas (nominais e ordinais) e Quantitativas (discretas e contínuas);
- Na **Aula 02** vimos como construir as Tabelas de Distribuição de Frequências para variáveis qualitativas (tabelas simples e tabelas de dupla entrada) e vimos como medir a associação entre **duas variáveis qualitativas** (pelo Coeficiente de Contingência de Pearson Corrigido, C^*)
- Na **Aula 03** vimos como medira associação entre **duas variáveis quantitativas** (pelo Coeficiente de Correlação de Pearson, r) e como ajustar um modelo de regressão linear simples para predizer um valor de y para um dado valor de x conhecido.

Estatística Descritiva

- Nessa Aula (Aula 04) vamos ver como construir uma Tabela de Distribuição de Classes de Frequências para **uma variável quantitativa contínua**.

Feita a coleta de dados, os mesmos apresentam-se de maneira desorganizada (sem um padrão crescente/decrescente), ainda sem valor informativo sobre o fenômeno em estudo.

O **primeiro passo** para construção da tabela de distribuição de frequências é a ordenação dos dados (o mais usual é em ordem crescente).

Estatística Descritiva

- Exemplo: **Dados da Lista 1**

Quadro 1. Rendimentos médios, em kg/ha, de 40 híbridos de milho (**dados não ordenados/brutos**).

5291	5357	5480	5387	5373
5384	5330	5519	5337	5366
5318	5285	5306	5444	5337
5323	5313	5425	5417	5313
5386	5448	5310	5397	5271
5298	5297	5378	5373	5443
5368	5438	5417	5479	5307
5268	5498	5410	5342	5455

Estatística Descritiva

- **1º Passo: Ordenar os dados**

Quadro 1. Rendimentos médios, em kg/ha, de 40 híbridos de milho (**dados elaborados**).



5268	5310	5342	5386	5443
5271	5313	5357	5387	5444
5285	5313	5366	5397	5448
5291	5318	5368	5410	5455
5297	5323	5373	5417	5479
5298	5330	5373	5417	5480
5306	5337	5378	5425	5498
5307	5337	5384	5438	5519

Estatística Descritiva

- **2º passo: Determinar o número de classes (k)**

Decidir quantas classes (k) terá a tabela de distribuição de frequências. Embora essa decisão seja arbitrária (o pesquisador pode escolher). Existem algumas fórmulas para auxiliar, por exemplo:

$$k = \sqrt{n} \quad \text{em que } n \text{ é o tamanho da amostra}$$

ou

$$k = 1 + 3,32 \log_{10}(n) \quad \text{Fórmula de Sturges}$$

Usaremos a Fórmula de Sturges!

Estatística Descritiva

- **2º passo: Determinar o número de classes (k)**

Usaremos a Fórmula de Sturges!

$$k = 1 + 3,32 \log_{10}(n)$$

$$k = 1 + 3,32 \log_{10}(40)$$

$$k = 6,3 \text{ classes} \Rightarrow 7 \text{ classes}$$

Cada uma dessas classes representa um intervalo de valores, precisamos definir a amplitude (tamanho) de cada intervalo, ou seja, o tamanho de cada classe.

Classes	Freq.
Classe 1	f1
Classe 2	f2
Classe 3	f3
Classe 4	f4
Classe 5	f5
Classe 6	f6
Classe 7	f7
Total	f.

Estatística Descritiva

- Terceiro passo: Amplitude de classe (h)**

Tabela 1. Distribuição de Classes de Frequências

Classes	Freq.
Classe 1	f1
Classe 2	f2
Classe 3	f3
Classe 4	f4
Classe 5	f5
Classe 6	f6
Classe 7	f7
Total	f.

$$h = \frac{A}{k} = \frac{251}{7} = 35,9$$

Qual o tamanho de cada classe? Qual a amplitude de cada classe (h)?

$$h = \frac{A}{k}$$

“A” representa a amplitude dos dados →

A = Maior valor – Menor valor

A= 5.519 – 5268 = 251

e...k (é número de classes) =7

Ou seja, cada uma das sete classes terá uma amplitude de 35,9 unidades.

Estatística Descritiva

- Quarto passo: Construção da 1ª classe**

Tabela 1. Distribuição de frequência.

Classes	Freq.
Classe 1	f1
Classe 2	f2
Classe 3	f3
Classe 4	f4
Classe 5	f5
Classe 6	f6
Classe 7	f7
Total	f.

Primeira Classe: o menor valor dos dados ($x_{\min}=5.268$) deverá ser o limite inferior da primeira classe.

Classe 1: $[x_{\min} ; x_{\min} + h)$

Lembrando: h é a amplitude de classe

Desta forma, a classe 1 é representada pelo seguinte intervalo:

Classe 1: $[5.268 ; 5.268 + 35,9)$
 $[5.268 ; 5.303,9)$

Estatística Descritiva

- **Quarto passo: Construção da 1ª classe**

Estabelecer uma notação para representação das classes na tabela.

- Parêntesis: indica que o valor não pertence à classe em questão
- Colchetes: indica que o valor pertence ao intervalo.

$(1 ; 3)$ é a notação de uma classe em que nem o valor 1, nem o valor 3 pertencem ao intervalo;

$(1 ; 3]$ é a notação de uma classe em que o valor 1 não pertence ao intervalo mas o valor 3 pertence;

$[1 ; 3)$ é a notação de uma classe em que o valor 1 pertence ao intervalo mas o valor 3 não pertence; (usaremos essa)

$[1 ; 3]$ é a notação de uma classe em que ambos os valores, 1 e 3, pertencem ao intervalo;

Estatística Descritiva

- Quarto passo: Construção da 1ª classe**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[5.268 ; 5.303,9)	f1
Classe 2	f2
Classe 3	f3
Classe 4	f4
Classe 5	f5
Classe 6	f6
Classe 7	f7
Total	f.

Primeira Classe: o menor valor dos dados ($x_{\min}=5.268$) deverá ser o limite inferior da primeira classe.

Classe 1: $[x_{\min} ; x_{\min} + h)$

Lembrando: h é a amplitude de classe

Desta forma, a classe 1 é representada pelo seguinte intervalo:

Classe 1: $[5.268 ; 5.268 + 35,9)$
 $[5.268 ; 5.303,9)$

Estatística Descritiva

- Quinto passo: Construção das demais classes**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[5.268 ; 5.303,9)	f1
[5.303,9 ; 5.339,8)	f2
[5.339,8 ; 5.375,7)	f3
[5.375,7 ; 5.411,6)	f4
[5.411,6 ; 5.447,5)	f5
[5.447,5 ; 5.483,4)	f6
[5.483,4 ; 5.519,3)	f7
Total	f.

Para construção das demais ($k = 7$) classes da tabela basta transpor o limite superior de cada classe para o limite inferior da classe subsequente e somar a esse valor a amplitude de classe ($h = 35,9$).

1ª classe: (5.268 ; 5.303,9]

2ª classe: (5.303,9 ; 5.339,8]



$$5.303,9 + 35,9$$

Estatística Descritiva

- **Sexto passo: Determinar as frequências em cada classe (contar quantos valores tem em cada classe construída)**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[5.268 ; 5.303,9)	f1
[5.303,9 ; 5.339,8)	f2
[5.339,8 ; 5.375,7)	f3
[5.375,7 ; 5.411,6)	f4
[5.411,6 ; 5.447,5)	f5
[5.447,5 ; 5.483,4)	f6
[5.483,4 ; 5.519,3)	f7
Total	f.

Classe 1

Quadro 1. Rendimentos médios, em kg/ha, de 40 híbridos de milho (dados elaborados).

5268	5310	5342	5386	5443
5271	5313	5357	5387	5444
5285	5313	5366	5397	5448
5291	5318	5368	5410	5455
5297	5323	5373	5417	5479
5298	5330	5373	5417	5480
5306	5337	5378	5425	5498
5307	5337	5384	5438	5519

Estatística Descritiva

- Sexto passo: Determinar as frequências em cada classe (contar quantos valores tem em cada classe construída)**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[5.268 ; 5.303,9)	6
[5.303,9 ; 5.339,8)	10
[5.339,8 ; 5.375,7)	6
[5.375,7 ; 5.411,6)	6
[5.411,6 ; 5.447,5)	6
[5.447,5 ; 5.483,4)	4
[5.483,4 ; 5.519,3)	2
Total	40

Quadro 1. Rendimentos médios, em kg/ha, de 40 híbridos de milho **(dados elaborados)**.

5268	5310	5342	5386	5443
5271	5313	5357	5387	5444
5285	5313	5366	5397	5448
5291	5318	5368	5410	5455
5297	5323	5373	5417	5479
5298	5330	5373	5417	5480
5306	5337	5378	5425	5498
5307	5337	5384	5438	5519

Estatística Descritiva

- **Completando a tabela com os demais tipos de frequências (relativa e percentual)**

Tabela 1. Distribuição de frequência do Rendimentos médios, em kg/ha, de 40 híbridos de milho.

Classes	Frequência absoluta (f_i)	Frequência relativa ($f r_i$)	Frequência percentual ($f \%_i$)
[5.268 ; 5.303,9)	6	$6/40 = 0,15$	$0,15 \times 100 = 15\%$
[5.303,9 ; 5.339,8)	10	$10/40 = 0,25$	$0,25 \times 100 = 25\%$
[5.339,8 ; 5.375,7)	6	$6/40 = 0,15$	$0,15 \times 100 = 15\%$
[5.375,7 ; 5.411,6)	6	$6/40 = 0,15$	$0,15 \times 100 = 15\%$
[5.411,6 ; 5.447,5)	6	$6/40 = 0,15$	$0,15 \times 100 = 15\%$
[5.447,5 ; 5.483,4)	4	$4/40 = 0,10$	$0,10 \times 100 = 10\%$
[5.483,4 ; 5.519,3)	2	$2/40 = 0,05$	$0,05 \times 100 = 5\%$
Total	40	1,0000	100%

Estatística Descritiva

- **Completando ainda mais tabela: ponto médio da classe e frequência acumulada.**

Tabela 1. Distribuição de frequência do Rendimentos médios, em kg/ha, de 40 híbridos de milho.

Classes	Ponto médio (x_i)	Freq. Absoluta (f_i)	Freq. Acumulada (fAc_i)	Freq. Acumulada percentual ($fAc\%$)
[5.268 ; 5.303,9)	5.285,9	6	6	$(6 / 40) \times 100 = 15\%$
[5.303,9 ; 5.339,8)	5.321,8	10	$6 + 10 = 16$	$(16/40) \times 100 = 40\%$
[5.339,8 ; 5.375,7)	5.357,7	6	$6 + 10 + 6 = 22$	$(22/40) \times 100 = 55\%$
[5.375,7 ; 5.411,6)	5.393,6	6	$6 + 10 + 6 + 6 = 28$	$(28/40) \times 100 = 70\%$
[5.411,6 ; 5.447,5)	5.429,5	6	$6 + 10 + 6 + 6 + 6 = 34$	$(34/40) \times 100 = 85\%$
[5.447,5 ; 5.483,4)	5.465,4	4	$6 + 10 + 6 + 6 + 6 + 4 = 38$	$(38/40) \times 100 = 95\%$
[5.483,4 ; 5.519,3)	5.501,3	2	$6 + 10 + 6 + 6 + 6 + 4 + 2 = 40$	$(40/40) \times 100 = 100\%$

Que tipo de gráfico usar?

- Gráficos: Visualização mais sugestiva do que tabelas;
- Forma alternativa de apresentar as distribuições de frequências;
- Tipo de gráfico também irá depender do tipo e número de variáveis em estudo.
 - Qualitativas: gráfico de setores/barras/colunas (Já vimos em aulas anteriores);
 - Quantitativas: Histogramas, Polígonos de frequência, Ogivas (para frequências acumuladas).

Que tipo de gráfico usar?

- **Histogramas:** é uma sequência de retângulos postos lado a lado em que cada retângulo tem como base a amplitude da classe ($h=35,9$) e como altura a frequência (absoluta/relativa);

Que tipo de gráfico usar?

- HISTOGRAMA

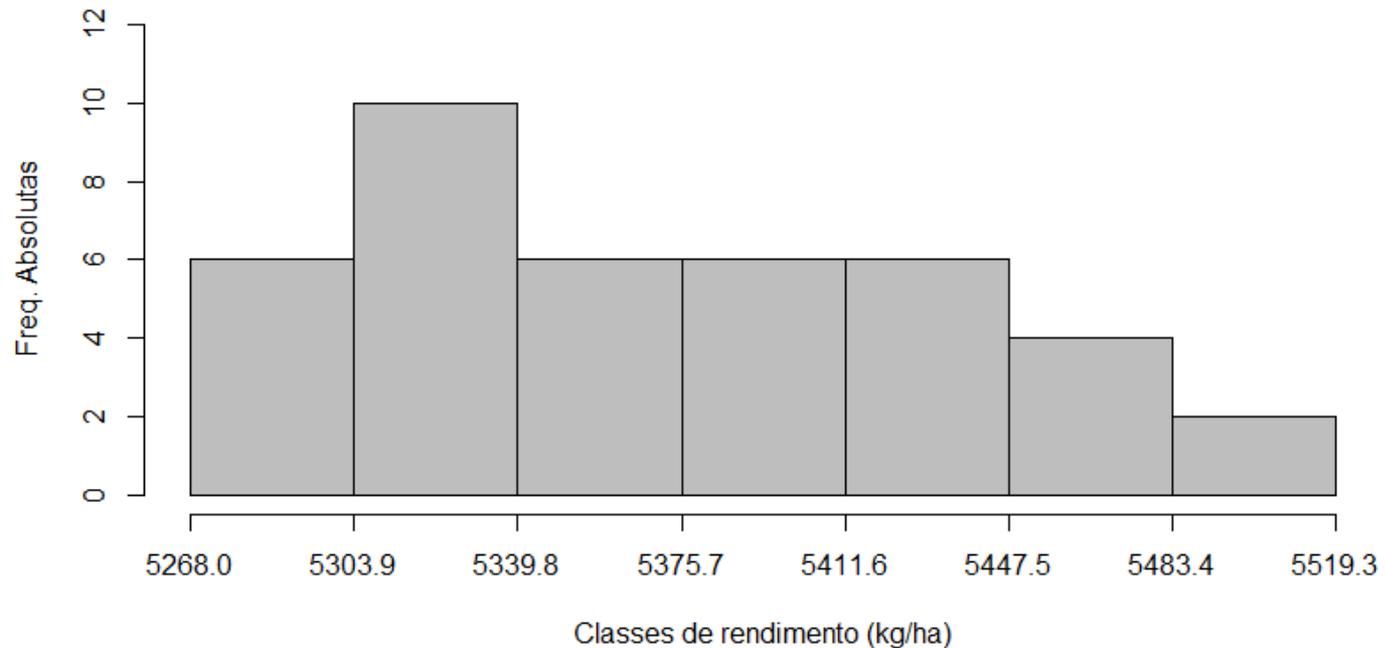


Figura 1. **Histograma** da distribuição de frequências dos rendimentos médios, em kg/ha, de 40 híbridos de milho.

Que tipo de gráfico usar?

- **Polígonos de frequência:** é um gráfico que se obtém unindo por uma poligonal os pontos correspondentes as frequências das diversas classes, centradas nos respectivos pontos médios. Para se obterem as interseções dos polígonos com o eixo - X, cria-se, em cada extremo do diagrama uma classe de frequência nula;

Que tipo de gráfico usar?

- **POLÍGONO DE FREQUÊNCIAS**

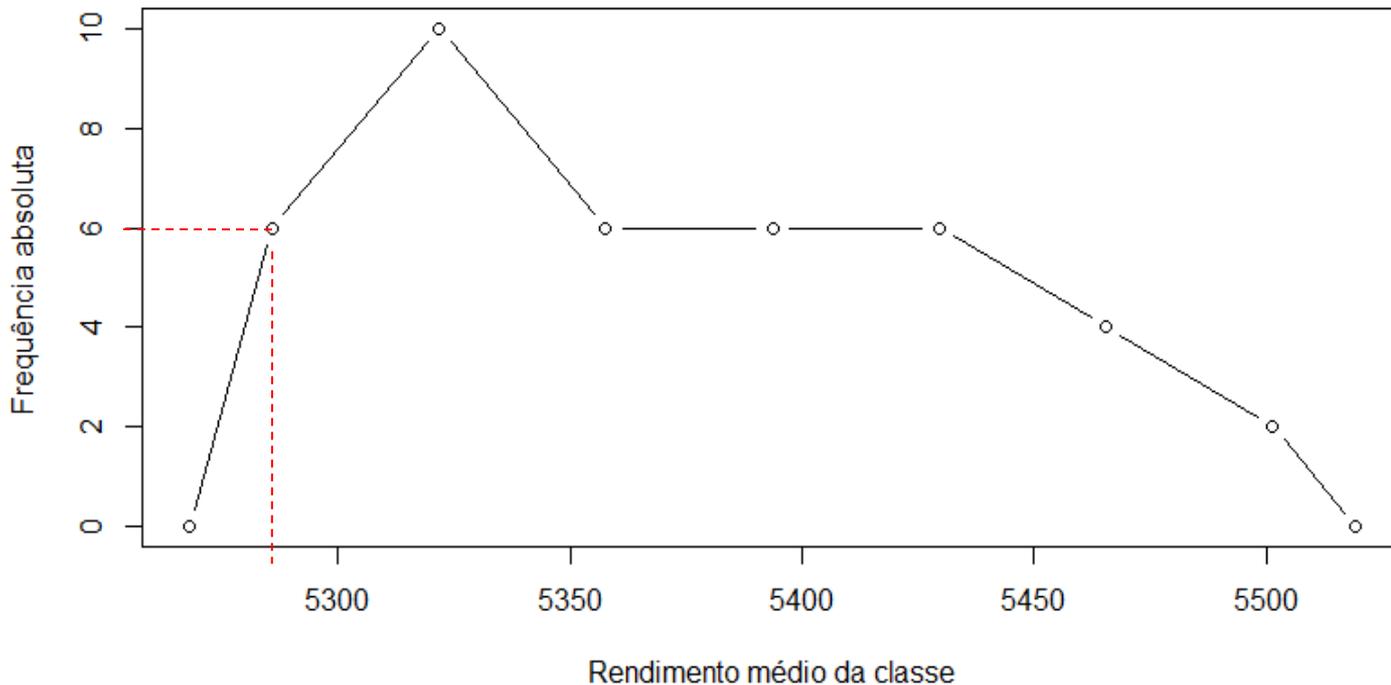


Figura 2. **Polígono de frequência** da distribuição dos rendimentos médios, em kg/ha, de 40 híbridos de milho.

Que tipo de gráfico usar?

- **Ogivas de Galton** (para frequências acumuladas): gráfico representativo de uma distribuição acumulada de frequências. Trata-se de uma poligonal ascendente formada ligando-se os pontos de coordenadas $(LS_i ; fAc_i)$ em que LS_i é o limite superior da classe i e fAc_i é a frequência acumulada até a classe i .

Que tipo de gráfico usar?

- **OGIVA DE GALTON CRESCENTE**

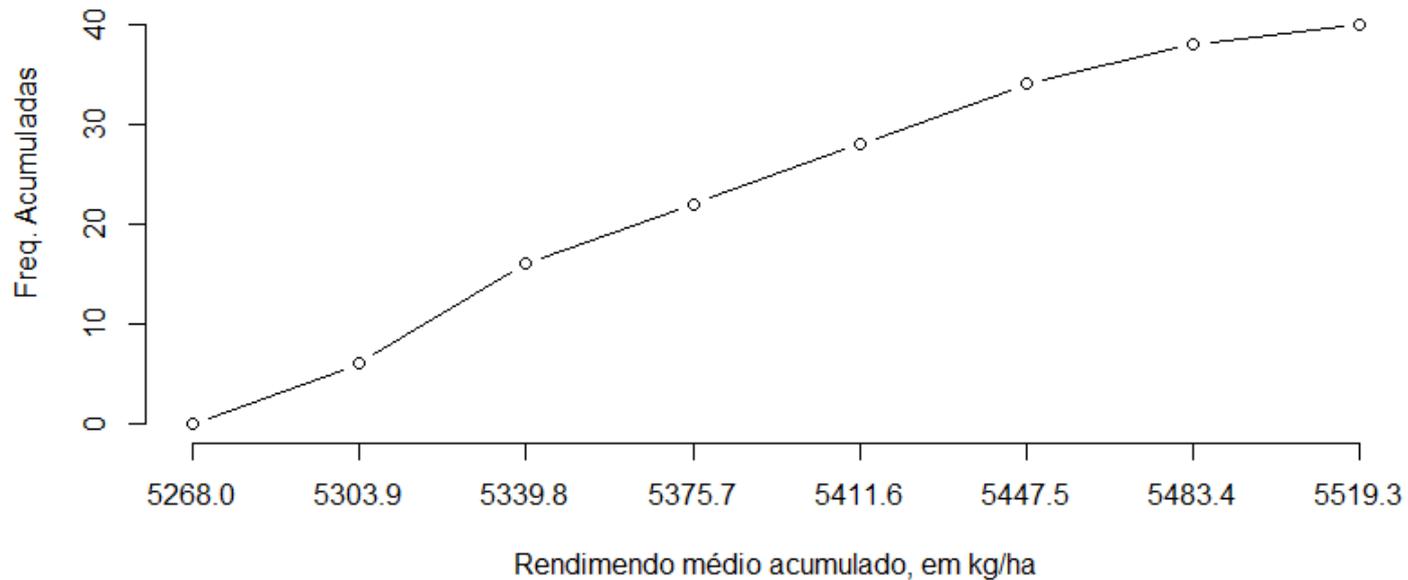


Figura 3. Ogiva crescente de Galton da distribuição acumulada dos rendimentos médios, em kg/ha, de 40 híbridos de milho.

Exercício 1

Quadro1. Densidade da madeira (g/cm^3) de 30 clones de *Eucalyptus grandis* medidos em um plantio comercial, em Piracicaba, em 2015.

0,347	0,373	0,297	0,360	0,338	0,357	0,343	0,345	0,392	0,330
0,405	0,364	0,294	0,267	0,413	0,295	0,427	0,333	0,324	0,230
0,445	0,327	0,359	0,446	0,256	0,329	0,471	0,306	0,328	0,230

i) Construir a tabela de distribuição de frequências para os dados de densidade da madeira (Absoluta, Relativa, Percentual, Acumulada e Acumulada percentual);

ii) Escolha uma das frequências obtidas para ser representada graficamente e faça uma interpretação do gráfico.