

UFT

Mestrado em
Desenvolvimento Regional

Nivelamento em Estatística

Distribuição de Frequências

Prof. Dr. Adriano N. da Paixão

Tópico: Distribuição de frequência

- Referências:

- HOFFMAN, R. **Estatística para economistas**. 4ª Edição Revisada e ampliada. São Paulo: Pioneira Thomson Learning, 2006. – Capítulo 3
- CRESPO, A. A. **Estatística fácil**. 17ª Edição. São Paulo: Saraiva, 2002. – Capítulo 5.

Organização da aula

- 1 – Os níveis de medida
- 2 – Tabela primitiva e tabela Rol
- 3 – A distribuição de freqüências
 - 3.1 – A distribuição de freqüências de uma variável discreta
 - 3.2 – A distribuição de freqüências de uma variável contínua
- 4 – A representação gráfica da distribuição de freqüência
 - 4.1 - A representação gráfica da distribuição de freqüência de uma variável discreta
 - 4.2 - A representação gráfica da distribuição de freqüência de uma variável contínua
- 5 – Exercício

1 – Os níveis de medida

- Toda vez que dispomos de um conjunto de dados podemos classificá-los de acordo com um determinado critério. Dado um conjunto o critério fica na dependência do nível de medida dos dados ou escala que são quatro:
 - escala nominal;
 - escala ordinal;
 - escala de intervalos;
 - escala-razão.

1 – Os níveis de medida

- Escala nominal – quando podem ser distribuídos em um determinado número de categorias mutuamente exclusivas. Ex. gênero (sexo), religião, partido político, etc.
- Escala ordinal – quando podem ser distribuídos em um determinado número de categorias mutuamente exclusivas e que se apresentam ordenadas. A relação entre as categorias pode ser do tipo “maior do que” ou “mais difícil que” ou “superior a”. Ex. hierarquia militar; o *status* socioeconômico (Classes A, B, C, D, E), a dureza dos minerais.

1 – Os níveis de medida

- Escala de intervalos – quando eles podem ser ordenados e é possível quantificar a diferença entre as observações. Ex. temperatura.
- Escala razão – além da propriedade da escala de intervalo, existe uma origem fixa e é possível estabelecer razões entre os dados. Ex. a velocidade, a renda, o capital.

2 – A tabela primitiva e tabela ROL

- Suponhamos que tenhamos feito uma coleta de dados relativos às estaturas de quarenta alunos, que compõem uma amostra de alunos de um colégio A, resultando na seguinte tabela de valores:

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 166 | 160 | 161 | 150 | 162 | 160 | 165 | 167 | 164 | 160 |
| 162 | 161 | 168 | 163 | 156 | 173 | 160 | 155 | 164 | 168 |
| 155 | 152 | 163 | 160 | 155 | 155 | 169 | 151 | 170 | 164 |
| 154 | 161 | 156 | 172 | 153 | 157 | 156 | 158 | 158 | 161 |

- Esse tipo de tabela, cujos dados não foram organizados, denominamos de tabela primitiva.

2 – A tabela primitiva e tabela ROL

- Partindo da tabela acima, é difícil averiguar em torno de qual valor tendem a se concentrar as estaturas dos alunos, qual a menor ou qual a maior altura, ou ainda quanto alunos estão acima da média e quantos estão abaixo da média.
- A maneira mais fácil de organizar um conjunto de dados é em primeiro lugar ordená-los em ordem crescente ou decrescente. A tabela obtida após a ordenação recebe o nome de ROL.

2 – A tabela primitiva e tabela ROL

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 150 | 154 | 155 | 157 | 160 | 161 | 162 | 164 | 166 | 169 |
| 151 | 155 | 156 | 158 | 160 | 161 | 162 | 164 | 167 | 170 |
| 152 | 155 | 156 | 158 | 160 | 161 | 163 | 164 | 168 | 172 |
| 153 | 155 | 156 | 160 | 160 | 161 | 163 | 165 | 168 | 173 |

- Agora podemos saber com facilidade quem qual a menor estatura (150 cm) e qual a maior estatura (173 cm), que a amplitude foi de $173 - 150 = 23$ cm. Podemos verificar ainda que há uma concentração das estaturas em algum valor entre 160 cm e 165 cm.

3 – A distribuição de freqüências

- No exemplo que trabalhamos , a variável em questão, estatura, será observada e estudada muito mais facilmente quando dispusermos valores ordenados em uma coluna e colocarmos, ao lado de cada valor, o número de vezes que aparece repetido.
- Denominamos freqüência o número de alunos que fica relacionado a um determinado valor da variável. Obtemos, assim, uma tabela que recebe o nome de distribuição de freqüência. Mas o processo dado ainda é inconveniente, já que exige muito espaço, mesmo quando o número de valores da variável (n) ainda é razoável.

3 – A distribuição de freqüências: tipos de freqüência

- Freqüência simples ou freqüência absoluta (f_i) – são os valores que realmente representam o número de dados em cada classe, como veremos

$$\sum_{i=1}^n f_i = n$$

- Freqüência relativa (fr_i) são os valores das razões entre as freqüências simples e a freqüência total:

$$fr_i = \frac{f_i}{\sum_{i=1}^n f_i}$$

3 – A distribuição de freqüências: tipos de freqüência

- Freqüência acumulada (F_j) – é o total das freqüências de todos os valores inferiores ao limite superior do intervalo de uma classe:

$$F_k = f_1 + f_2 + \dots + f_k$$

ou

$$Fk = \sum_{i=1}^k f_i$$

- Freqüência relativa acumulada (Fr_j) de uma classe é a freqüência acumulada da classe, dividida pela freqüência total da distribuição:

$$Fr_i = \frac{F_i}{\sum_{i=1}^n f_i}$$

3.1 – A distribuição de freqüências de uma variável discreta

- Sejam X_1, X_2, \dots, X_n os valores de uma amostra de n elementos. Admitamos que a variável é medida em escala de intervalos ou em escala razão. Para facilitar a análise é interessante organizar os dados em gráficos e ou tabelas.
- Consideremos por exemplo uma amostra de 200 famílias de certa cidade e seja X_i ($i=1, \dots, n$) o número de filhos dependentes de cada família. Os dados originais seriam: 3, 5, 0, 1, 1, 0, 3, 2, 2, 1, 2 etc.

3.1 – A distribuição de freqüências de uma variável discreta

- A tabela 1 mostra a distribuição de freqüências de X , isto é, dá os valores assumidos por X_i e as freqüências (número de vezes que cada X_i ocorreu) associadas a X_i .
- O número de filhos dependentes é uma variável discreta. Dados de contagem são sempre variáveis discretas. Característica: só podem assumir um número finito de diferentes valores dentro de um intervalo finito. Ex.: sexo, religião, número de pessoas.

Tabela 1 – Distribuição de freqüência de uma variável discreta

Tabela 1 - Distribuição das famílias conforme o número de filhos dependentes

| Número de filhos dependentes (X_i) | Freqüência observada (f_i) | Freqüência acumulada (F_i) | Freqüência relativa (fr_i) | Freqüência relativa acumulada (Fr_i) |
|--|--------------------------------|--------------------------------|--------------------------------|--|
| 0 | 32 | 32 | 0,16 | 0,16 |
| 1 | 46 | 78 | 0,23 | 0,39 |
| 2 | 50 | 128 | 0,25 | 0,64 |
| 3 | 40 | 168 | 0,2 | 0,84 |
| 4 | 16 | 184 | 0,08 | 0,92 |
| 5 | 8 | 192 | 0,04 | 0,96 |
| 6 | 6 | 198 | 0,03 | 0,99 |
| 7 | 0 | 198 | 0 | 0,99 |
| 8 | 2 | 200 | 0,01 | 1 |

Onde: X_i é o valor da variável observada; f_i é a freqüência observada; F_i é a freqüência acumulada; fr_i é a freqüência relativa e Fr_i é a freqüência relativa acumulada.

3.2 – A distribuição de freqüências de uma variável contínua

- A altura, o peso, a área dos estabelecimentos agrícolas, o tempo são exemplos de variáveis contínuas. Normalmente, admite-se que as variáveis econômicas medidas em unidades monetárias, como renda, riqueza, custo e preço são contínuas. Ex.: conjunto das alturas de 200 alunos do sexo masculino em cm.

2.2 – A distribuição de freqüências de uma variável contínua

- Para construir uma distribuição de freqüência devemos em primeiro lugar, escolher os intervalos de classe. É conveniente termos entre 5 e 20 classes. É conveniente, embora não seja necessário que os intervalos de classe tenham a mesma amplitude.
 - Obs.: se o número de classes é pequeno, os intervalos serão grandes, o que implica em muita perda de informação. Por outro lado, se o número de classes for muito grande, torna a distribuição de freqüências desnecessariamente extensa.

2.2 – A distribuição de freqüências de uma variável contínua

- Ex.: conjunto das alturas de 200 alunos do sexo masculino em cm.
 - Menor altura – 152 cm
 - Maior altura – 194 cm

- Amplitude:

$$A = Valor_{Max} - Valor_{Min} \Rightarrow A = 194 - 152 = 42cm$$

- Se dividirmos $42/9$, a amplitude de cada classe será aproximadamente 5 cm.

2.2 – A distribuição de freqüências de uma variável contínua

- Problema: é necessário que não fique dúvida quanto ao conteúdo; assim se estabelecermos que a primeira classe vai de 152 cm a 157 cm e a classe seguinte vai de 157 a 162, onde deverá ser colocada uma observação de 157 cm?
- Uma forma de superar essa dubiedade é definir intervalos como fechados à esquerda e abertos à direita (ou vice versa); outra maneira é definir os intervalos com precisão.

Tabela 2 – Distribuição de freqüência de uma variável contínua

Tabela 2 - Distribuição de freqüências das alturas de 200 estudantes do sexo masculino

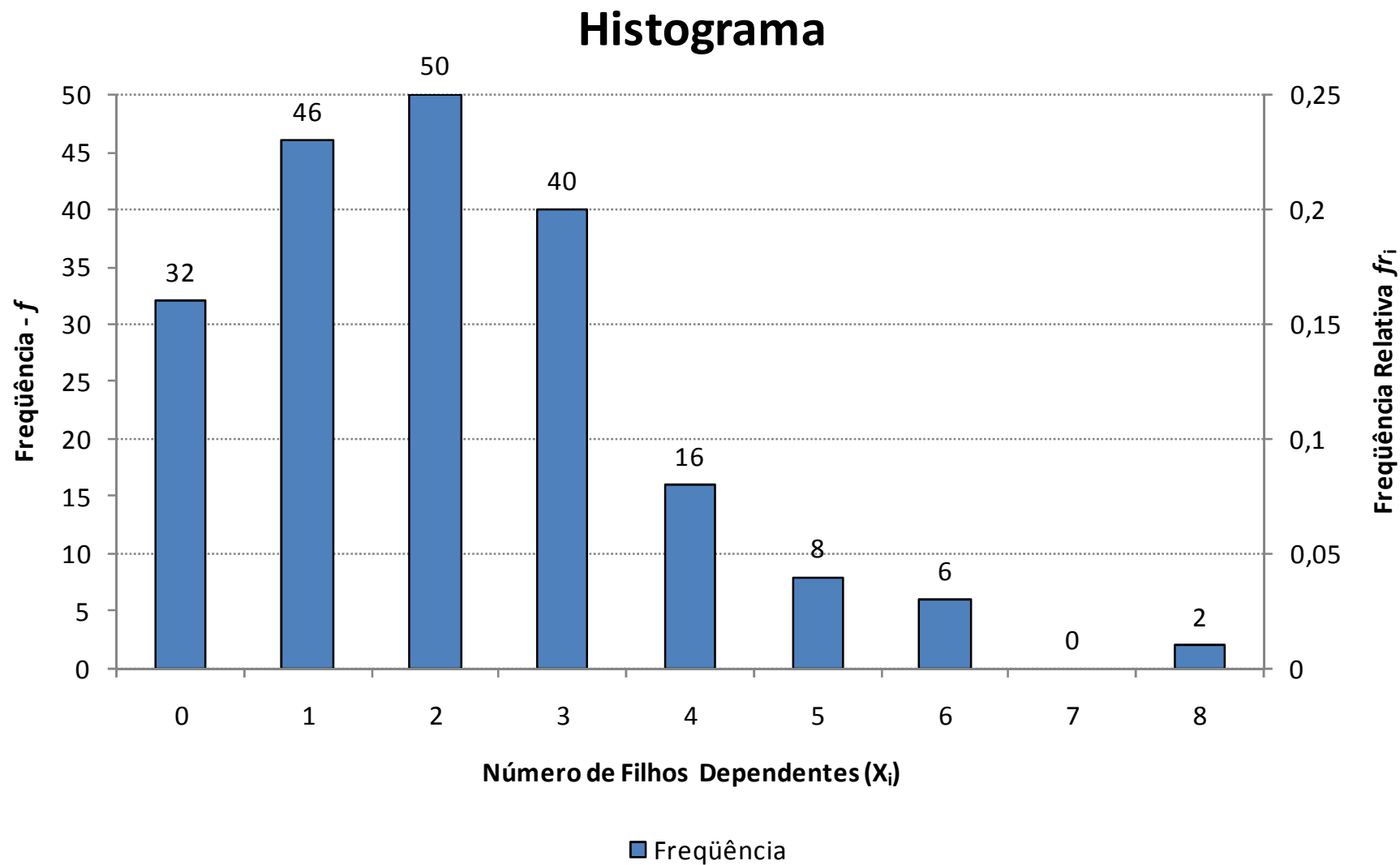
| Limites (cm) | Valor central (X_i) | Freqüência (f_i) | Freqüência acumulada (F_i) | Freqüência relativa (fr_i) | Freqüência relativa acumulada (Fr_i) |
|---------------|-------------------------|----------------------|--------------------------------|--------------------------------|--|
| 151,5 a 156,5 | 154 | 2 | 2 | 0,01 | 0,01 |
| 156,5 a 161,5 | 159 | 6 | 8 | 0,03 | 0,04 |
| 161,5 a 166,5 | 164 | 24 | 32 | 0,12 | 0,16 |
| 166,5 a 171,5 | 169 | 40 | 72 | 0,20 | 0,36 |
| 171,5 a 176,5 | 174 | 46 | 118 | 0,23 | 0,59 |
| 176,5 a 181,5 | 179 | 44 | 162 | 0,22 | 0,81 |
| 181,5 a 186,5 | 184 | 28 | 190 | 0,14 | 0,95 |
| 186,5 a 191,5 | 189 | 8 | 198 | 0,04 | 0,99 |
| 191,5 a 196,5 | 194 | 2 | 200 | 0,01 | 1 |

Onde: X_i é o valor da central do intervalo; f_i é a freqüência observada; F_i é a freqüência acumulada; fr_i é a freqüência relativa e Fr_i é a freqüência relativa acumulada.

4.1 – A representação gráfica da distribuição de freqüência de uma variável discreta

- A distribuição de freqüência de uma variável discreta é adequadamente representada por um gráfico de barras. Para traçar o gráfico marcamos sobre um eixo os pontos correspondentes aos valores assumidos pela variável (X_i) Em seguida traçamos as barras verticais cujas bases situam-se sobre o eixo, nos pontos marcados, e cujas alturas são proporcionais às freqüências relativas de cada X_i .

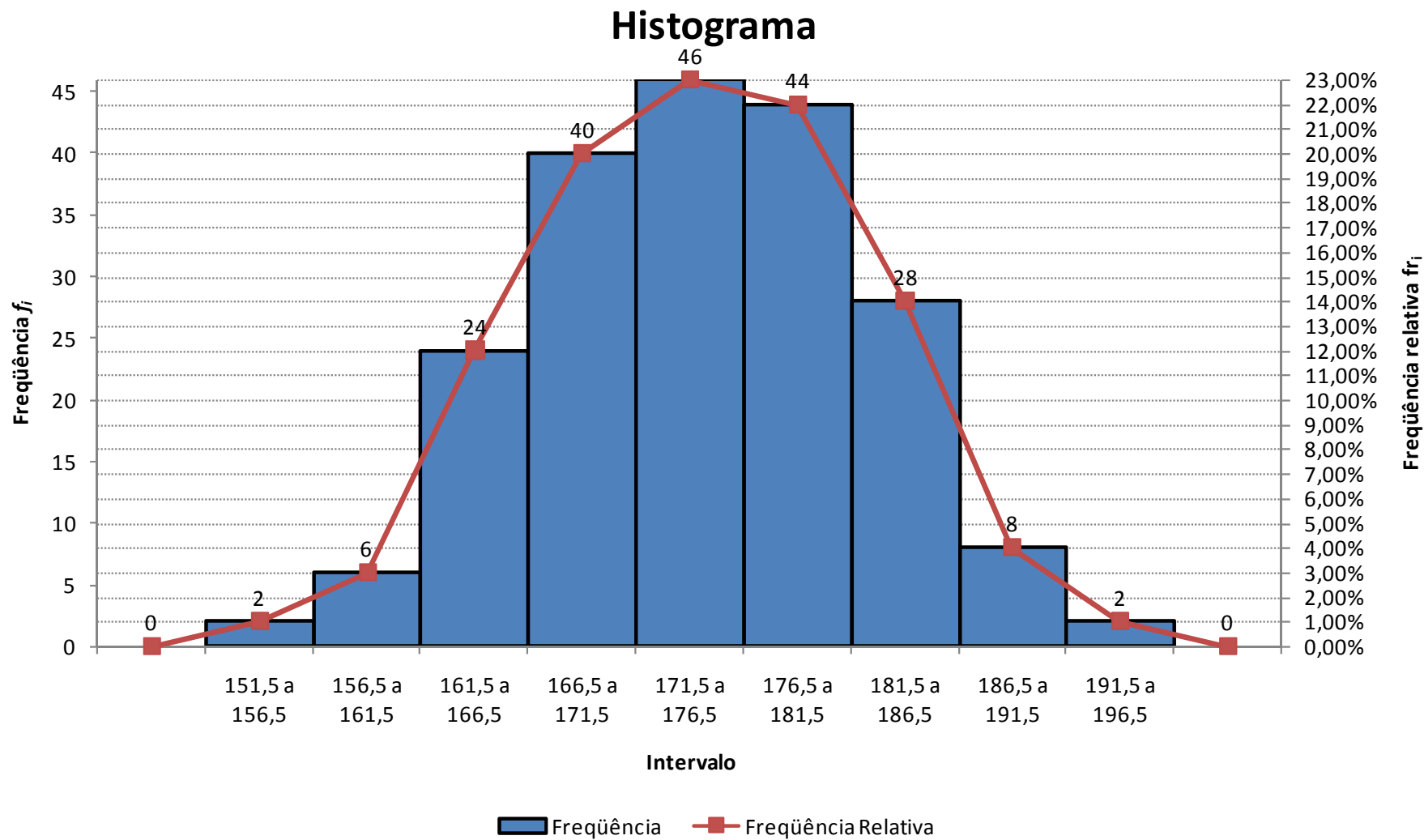
Tabela 1 – Histograma de uma variável discreta



4.2 – A representação gráfica da distribuição de freqüência de uma variável contínua

- Uma distribuição de freqüências contínua pode ser representada por um histograma ou por um polígono de freqüência.
- O histograma é constituído por um conjunto de retângulos cujas bases, situadas sobre um eixo coordenado (X), são os intervalos de classes e cujas áreas são proporcionais às respectivas freqüências ou freqüências relativas como ilustra a figura 2.

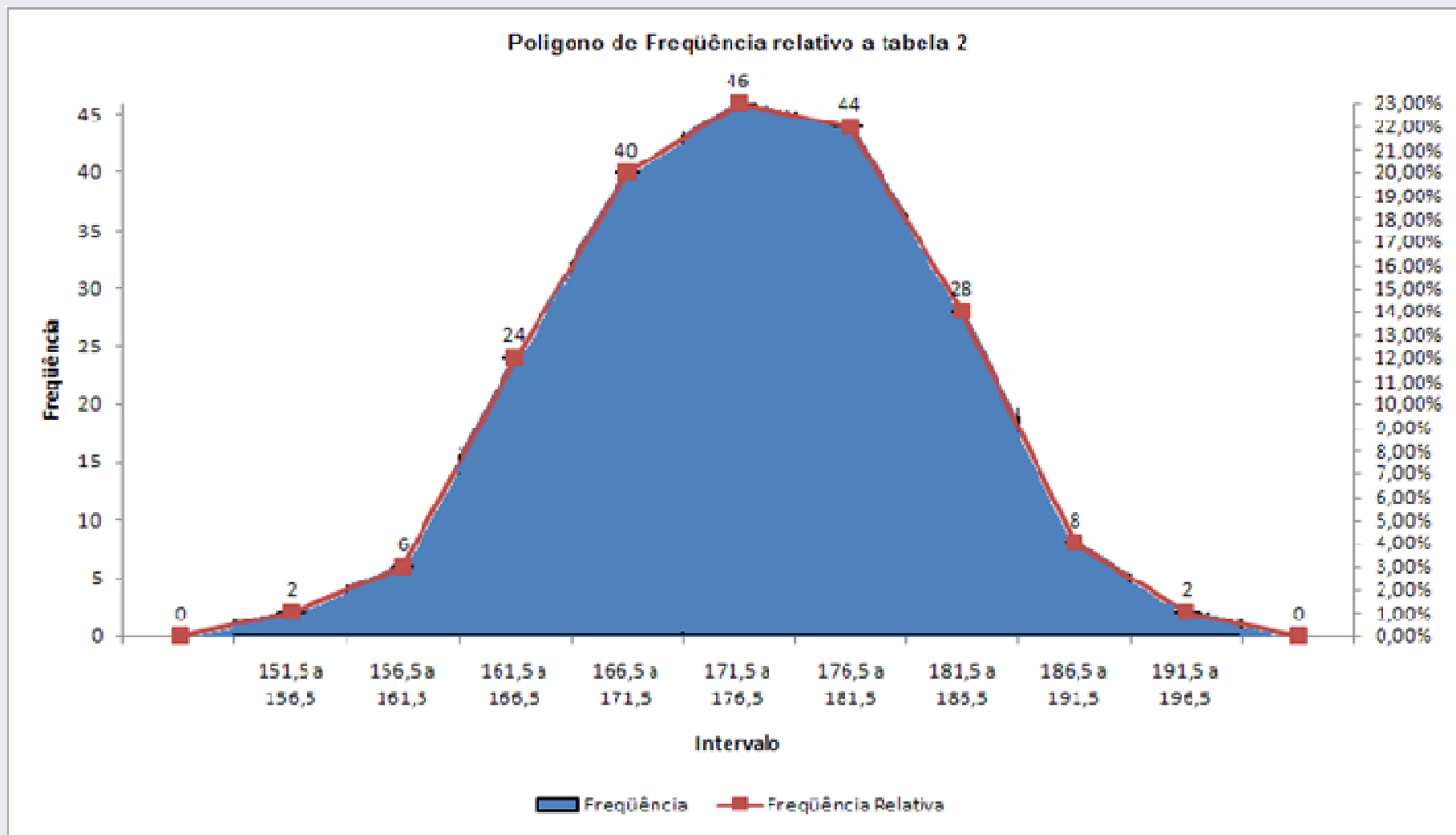
Figura 2 – Histograma de uma variável contínua



4.2 – A representação gráfica da distribuição de freqüência de uma variável contínua

- O polígono de freqüência é obtido ligando, em um sistema de eixos cartesianos ortogonais, os pontos cujas abscissas são os valores centrais das classes e cujas ordenadas são as freqüências correspondentes; ao construir um polígono de freqüências, considere-se, sempre, uma classe inicial e uma final com freqüências iguais a zero.

Figura 3 – Polígono de uma variável contínua



Resumo

- 1 – Os níveis de medida
- 2 – Tabela primitiva e tabela Rol
- 3 – A distribuição de freqüências
 - 3.1 – A distribuição de freqüências de uma variável discreta
 - 3.2 – A distribuição de freqüências de uma variável contínua
- 4 – A representação gráfica da distribuição de freqüência
 - 4.1 - A representação gráfica da distribuição de freqüência de uma variável discreta
 - 4.2 - A representação gráfica da distribuição de freqüência de uma variável contínua

Exercício

- Para as altura de 40 alunos, já classificadas conforme a tabela ROL, abaixo, pede-se:

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 150 | 154 | 155 | 157 | 160 | 161 | 162 | 164 | 166 | 169 |
| 151 | 155 | 156 | 158 | 160 | 161 | 162 | 164 | 167 | 170 |
| 152 | 155 | 156 | 158 | 160 | 161 | 163 | 164 | 168 | 172 |
| 153 | 155 | 156 | 160 | 160 | 161 | 163 | 165 | 168 | 173 |

- Essa é uma variável discreta ou contínua?
- Determine o número de classes dessa distribuição com base na amplitude total.
- Determine os intervalos dessa distribuição.
- Construa a distribuição de freqüências dessa distribuição e calcule a freqüência (f_i), a freqüência relativa (fr_i), a freqüência acumulada (F_i) e a freqüência relativa acumulada (Fr_i)
- Construa o histograma e o polígono de freqüência se for o caso.