

## CONSTRUÇÃO DE UMA FERRAMENTA TECNOLÓGICA PARA DETERMINAR AS MEDIDAS RESUMO DE DADOS DISTRIBUÍDOS POR CLASSE

*Armando Handaya*  
*IFSP-Gru*  
*ahand@ifsp.edu.br*

### **Resumo:**

A construção de uma Tabela de Distribuição de Frequência (TDF) é muito útil para se calcular as medidas-resumo de um conjunto de dados numéricos. Tais medidas como a média, o desvio padrão, a variância, etc. são facilmente obtidas quando essa Tabela corresponde à uma distribuição por pontos. Existem inclusive várias ferramentas tecnológicas que fazem as contas automaticamente. A vasta maioria, senão todos, desses programas de cálculo estatístico apresenta esse recurso. Todavia o quadro se mostra diferente quando os dados não são distribuídos por pontos, mas sim por classes. A Tabela de Distribuição de Frequência por Classes ainda pode ser usada para se calcular as medidas-resumo manualmente. Entretanto as contas são mais complicadas e aqueles mesmos recursos tecnológicos não oferecem recurso para calcular automaticamente essas medidas. Este mini-curso é destinado a todos os interessados em construir uma ferramenta tecnológica usando o software Geogebra para se calcular automaticamente as medidas-resumo neste caso: quando os dados são distribuídos por classes.

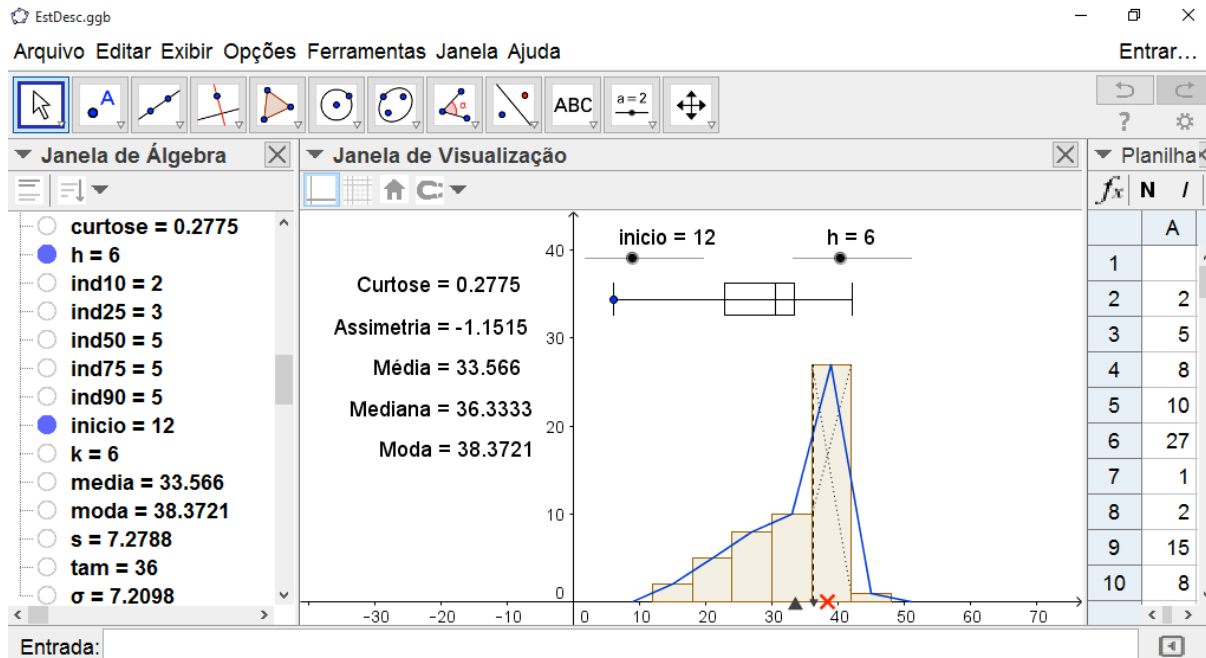
**Palavras-chave:** Geogebra, Estatística Descritiva, Tabela de Distribuição de Frequência.

### **1. Introdução**

Em Estatística Descritiva aprendemos a calcular várias medidas que são os resumos estatísticos como a média, moda, mediana, desvio-padrão, variância, coeficiente de assimetria de Pearson, curtose, .... etc. Aprendemos também que os dados coletados como resultado de um levantamento estatístico podem ser organizados numa Tabela de Distribuição de Frequência (TDF) por pontos ou por classes. Quando eles são apresentados em uma TDF por pontos os cálculos das medidas estatísticas são mais fáceis de se fazer. Quando eles são organizados em uma TDF por classes as contas são mais difíceis. Algumas medidas inclusive requer a adoção de um critério para o seu cálculo. Este é o caso de Moda, que tem vários critérios como Czuber, King e do ponto médio. Cada um desses critérios tem a sua razão de ser ou interpretação geométrica. Conhecer a razão de ser ou, como muitos preferem dizer, a interpretação geométrica, ajuda muito a entender melhor o conceito da Moda. Em muitas escolas do ensino médio, entretanto, as medidas resumo de dados distribuídos por classes muitas vezes são apresentadas pela respectiva fórmula final, desprovida dessa interpretação.

Mesmo nas várias ferramentas tecnológicas existentes que calculam tais medidas, a vasta maioria, senão todas, calculam essas medidas apenas quando os dados são apresentados numa TDF por pontos. Entre elas podemos citar o próprio Excel que possui um módulo de Estatística Descritiva, mas faz as contas com os dados organizados em TDF por pontos. Outros softwares específicos de tratamento estatístico como R, SPSS, Minitab, Statdisk e o próprio Geogebra, ferramenta que será usada neste mini-curso, seguem o mesmo caminho. Existem outros tantos softwares estatísticos, que muito embora não conheçamos cada um deles a fundo, não hesitamos em dizer que em sua maioria, senão todos, funcionam de maneira semelhante. Quanto ao recurso das medidas-resumo também acreditamos funcionar apenas para os dados distribuídos por pontos e não por classes.

Neste mini-curso faremos a construção de uma ferramenta tecnológica quando os dados da Tabela TDF correspondente são organizados por classes, valendo-se para isso a tecnologia do software Geogebra. A ferramenta construída pode ser usada inclusive como apoio ao processo de ensino e aprendizagem de Estatística. Na imagem abaixo pode-se ver a ferramenta já construída e a interpretação geométrica especificamente da Moda de Czuber.



As medidas que serão implementadas nesse mini-curso serão

- Média
- Mediana
- Moda de Czuber
- Moda de King
- Coeficiente de Assimetria
- Curtose
- Desvio Padrão populacional
- Desvio padrão amostral
- Variância populacional
- Variância amostral
- Percentis
- Quartis

## 2. Distribuição de Frequência

Um levantamento de dados resultante de uma pesquisa estatística normalmente resulta em uma tabela onde cada dado pode-se repetir diversas vezes. Sobretudo quando o espaço amostral, ou conjunto de repostas possíveis, é pequeno. Um exemplo disso é uma pesquisa de opinião para a qual as repostas possíveis são apenas duas: sim ou não. Neste caso um levantamento estatístico resultante será muitos *sim's* ou outros *não's*. Um levantamento de dados como esse sempre pode ser organizado em uma Tabela de Distribuição de Frequência onde cada dado é registrado apenas uma única vez. Assim um levantamento estatístico pode registrar 50 *sim's* e 30 *não's* mas a sua Tabela de Distribuição de Frequência terá apenas um *sim* e um *não*. A quantidade de *sim's* e *não's* será registrado em uma outra coluna. Tal quantidade é chamada de Frequência. Assim a Tabela resultante é chamada de Tabela de Distribuição de Frequência, aqui abreviamos por TDF.

Respostas	Frequência
Sim	50
Não	30

Uma TDF, portanto, resume um levantamento de dados e é uma maneira de organizá-los. Entretanto quando o tamanho do espaço amostral for muito grande uma TDF pode não resumir tanto assim os dados. Por exemplo, considere um levantamento de pesquisa estatística

cujo espaço amostral consiste em idades de pessoas tomadas aleatoriamente a partir de 15 até os 30 anos. Assim o espaço amostral consiste no conjunto de dezesseis números:  $\{15, 16, 17, 18, \dots, 30\}$ . Dependendo do tamanho da amostra tomada, o resultado da pesquisa será poucos de cada: poucos 15, poucos 16, mais uns poucos 17, etc. Assim TDF correspondente não será tão diferente do próprio levantamento estatístico. Neste caso é mais interessante reunir os dados em faixas de mesmo tamanho, essas que são chamadas de classes. Por exemplo de cinco em cinco anos como na tabela abaixo

Faixa de Idade	Frequência
15  — 20	13
20  — 25	25
25  — 30	12

A tabela construída assim resume melhor os dados, embora, há de se dizer aqui, que alguma informação com certeza se perca. Quando os dados são reunidos em faixas como essa, a TDF correspondente será chamada de TDF por classes. Quando não são reunidos em faixas elas são chamadas de TDF por pontos. Os cálculos de medidas resumo são um tanto mais complicado em TDF por classes do que em TDF por pontos. A perda de informações ou a complicação dos cálculos pode ser a razão pela qual os softwares estatísticos só apresentam o recurso de medidas resumo no caso de TDF por pontos.

Neste curso vamos aprender a construir uma ferramenta que a partir de dados distribuídos por frequência, sem faixas, construir um gráfico de barra vertical chamado Histograma e com ele calcular as suas medidas resumo. Esse gráfico Histograma é uma representação gráfica da Tabela de Distribuição de Frequência por classe.

### 3. Comandos e Recursos do Geogebra

O software GeoGebra é um software de matemática dinâmica muito versátil. Criado pelo Prof. Markus Horhenwarter da Universidade de Salzburg tem-se destacado entre os aplicativos educacionais no mundo todo obtendo menção e prêmios internacionais. Com ele pode-se construir ferramentas capazes de atrair atenção dos alunos devido ao dinamismo dos objetos criados nele. Vários ambientes podem ser usados ao mesmo tempo, o que é muito útil. No que tange a este mini-curso pode-se trabalhar com os dados em uma planilha, tratá-los em uma janela de visualização algébrica e apresentar os resultados em outra: janela de visualização geométrica. Podem-se criar também novas ferramentas a partir de outras, e

usando a sua linguagem própria de programação pode-se estender a capacidade de atrair os alunos pelo seu aspecto lúdico.

Os comandos que serão utilizados para a construção da ferramenta em questão serão

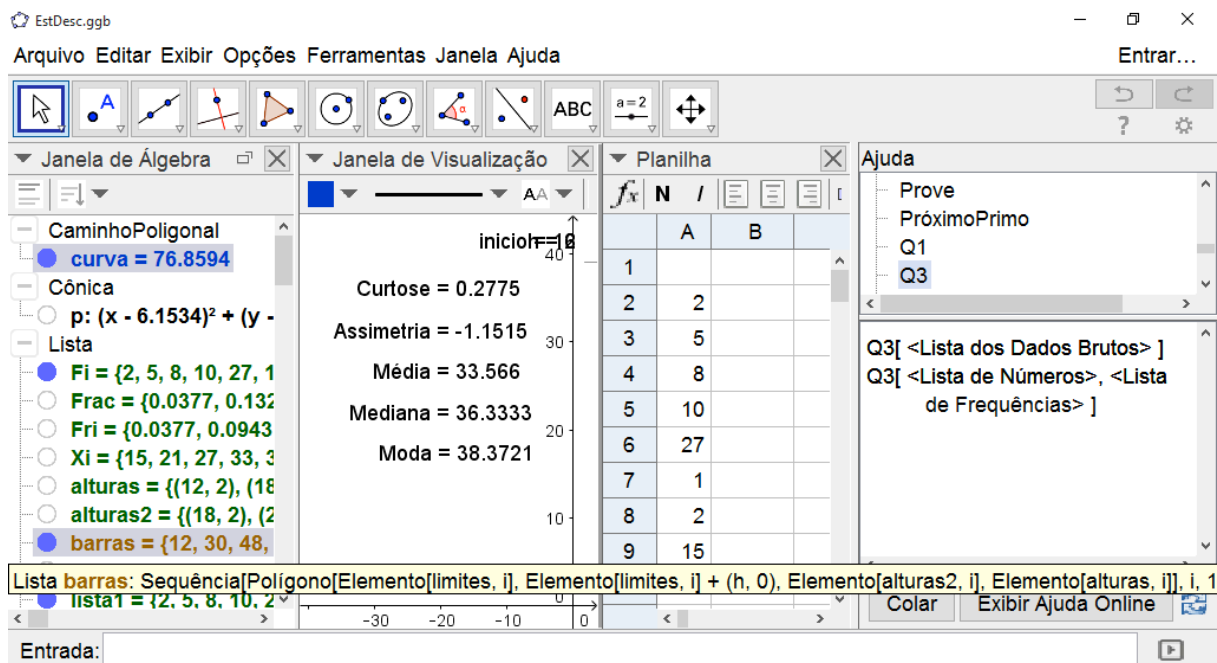
- CaminhoPoligonal
- Comprimento
- Concatenar
- Elemento
- Histograma
- ÍndiceDe
- Máximo
- ParteDaLista
- Poligono
- Se
- Segmento
- Sequência
- Soma
- Sqrt
- x

Os objetos e recursos que serão utilizados serão

- Caminho Poligonal
- Cônica
- Lista
- Número
- Ponto
- Ponto de interseção
- Ponto sobre círculo
- Reta
- Reta perpendicular
- Segmento
- Texto
- Translação de pontos

- Valor Booleano

Na imagem abaixo apresentamos algumas das várias janelas possíveis e exemplo de um comando que será utilizado na proposta do curso.



The screenshot shows the GeoGebra interface with three main panels: Algebra, Visualization, and Spreadsheet. The Algebra panel shows a list of objects including a polygonal path, a conic section, and a list of data. The Visualization panel shows a histogram with various statistical measures calculated. The Spreadsheet panel shows a table with columns A and B, containing numerical data.

	A	B
1		
2	2	
3	5	
4	8	
5	10	
6	27	
7	1	
8	2	
9	15	

Statistical measures shown in the Visualization panel:

- Curiose = 0.2775
- Assimetria = -1.1515
- Média = 33.566
- Mediana = 36.3333
- Moda = 38.3721

Algebra panel list:

- Lista barras: Sequência[Polígono[Elemento[limites, i], Elemento[limites, i] + (h, 0), Elemento[alturas2, i], Elemento[alturas, i]], i, 1
- lista1 = {2, 5, 8, 10, 27, 1, 1, 2}

#### 4. Considerações Finais

Os cálculos de medidas-resumo dados tabelados por uma Tabela de Distribuição de Frequência por classe são feitos aqui através de uma ferramenta tecnológica. Tal ferramenta será construída com ajuda do software Geogebra e pode ser usada para ajudar no processo de ensino-aprendizagem de Estatística Descritiva, uma vez que as ferramentas existentes são voltadas apenas para uma Tabela de Distribuição sem classes.

#### 5. Agradecimentos

Agradecemos ao Instituto Federal de Educação, Ciência e Tecnologia que proporciona momentos de atuação profissional na área de Educação Matemática e fomento à participação nesse evento como uma Capacitação Externa de seus servidores.

#### 6. Referências

- MILONE, Giuseppe. **Estatística Geral e Aplicada**. São Paulo: Pioneira Thomson Learning, 2004.
- GEOGEBRA.org. **Manual do Geogebra**. Disponível em <https://www.geogebra.org/manual/pt/Manual> Acesso: 16/02/2016