

# Distribuição de Frequência em Estatística

---

A **distribuição de frequência** é um conceito fundamental em estatística, sendo uma ferramenta que permite a organização, análise e interpretação de um conjunto de dados. Ela descreve como os dados são distribuídos ao longo de diferentes intervalos ou categorias, oferecendo uma visão clara e concisa das variáveis de interesse.

## O que é uma Distribuição de Frequência?

Uma distribuição de frequência é uma tabela que mostra o número de vezes (frequência) que os dados aparecem em diferentes intervalos ou classes. Essas classes são chamadas de **intervalos de classe** ou **faixas de valores** e são divididas com base no intervalo de valores dos dados.

## Componentes de uma Distribuição de Frequência

Uma distribuição de frequência geralmente inclui os seguintes componentes:

### 1. Classe (ou Intervalo):

Refere-se aos intervalos de valores nos quais os dados são agrupados. As classes são criadas para organizar os dados em grupos com base em suas magnitudes ou valores numéricos.

### 2. Frequência Absoluta (f):

A frequência absoluta indica o número de vezes que um valor ou intervalo de classe aparece no conjunto de dados.

### 3. Frequência Acumulada (F):

A frequência acumulada é a soma das frequências absolutas das classes até o ponto desejado. Ela representa o número total de dados até a classe correspondente.

### 4. Frequência Relativa (fr):

A frequência relativa é a razão entre a frequência absoluta de uma classe e o total de observações no conjunto de dados. Ela é calculada pela fórmula:

$$\text{fr} = \frac{f}{N}$$

Onde \$ f \$ é a frequência absoluta de uma classe e \$ N \$ é o total de dados.

### 5. Frequência Percentual (fp%):

A frequência percentual é a frequência relativa expressa em porcentagem. Ela é calculada multiplicando a frequência relativa por 100.

$$\text{fp\%} = \text{fr} \times 100$$

## Exemplo de Distribuição de Frequência

Imaginemos um conjunto de dados representando as idades de 10 pessoas:

[18, 22, 22, 24, 26, 26, 28, 30, 30, 32]

Se organizarmos esses dados em uma distribuição de frequência com intervalos de 5 anos, teríamos:

Intervalo de Idade	Frequência Absoluta (f)	Frequência Acumulada (F)	Frequência Relativa (fr)	Frequência Percentual (fp%)
18-22	3	3	0.3	30%
23-27	3	6	0.3	30%
28-32	4	10	0.4	40%
Total	10	-	1.0	100%

### Passos para Criar uma Distribuição de Frequência

1. **Organize os dados:**

Coloque os dados em ordem crescente para facilitar a organização e o agrupamento.

2. **Defina o número de classes:**

Determine o número de classes a serem utilizadas. Isso pode ser feito por meio de métodos como a **regra de Sturges** ou **regra de Scott**, que ajudam a calcular a quantidade ideal de intervalos.

3. **Defina os intervalos de classe:**

Defina os intervalos de classe (faixas) com base no intervalo de valores dos dados. Certifique-se de que cada classe tenha a mesma amplitude, ou seja, a diferença entre o limite superior e inferior seja constante.

4. **Conte as frequências absolutas:**

Para cada intervalo, conte o número de dados que se encaixam naquele intervalo.

5. **Calcule as frequências acumuladas, relativas e percentuais:**

A partir das frequências absolutas, calcule as frequências acumuladas, relativas e percentuais.

## Exemplo de Distribuição de Frequência - Didático

Vamos criar um exemplo passo a passo de distribuição de frequência utilizando um conjunto de dados fictício, onde a variável de interesse será a idade de um grupo de pessoas.

### Dados Iniciais

Suponha que temos os seguintes dados representando as idades de 15 pessoas:

[18, 22, 22, 24, 26, 26, 28, 30, 30, 32, 32, 34, 36, 38, 40]

# Passos para Criar a Distribuição de Frequência

## 1. Organize os Dados

O primeiro passo é organizar os dados em ordem crescente:

```
[18, 22, 22, 24, 26, 26, 28, 30, 30, 32, 32, 34, 36, 38, 40]
```

## 2. Defina os Intervalos de Classe

Vamos criar intervalos de classe com uma amplitude de 5 anos. O intervalo de classe será de 18 a 22, 23 a 27, e assim por diante. Aqui estão os intervalos:

- 18-22
- 23-27
- 28-32
- 33-37
- 38-42

## 3. Conte as Frequências Absolutas

Agora, vamos contar quantos dados se encaixam em cada intervalo.

- **18-22:** 3 pessoas (18, 22, 22)
- **23-27:** 3 pessoas (24, 26, 26)
- **28-32:** 4 pessoas (28, 30, 30, 32)
- **33-37:** 3 pessoas (32, 34, 36)
- **38-42:** 2 pessoas (38, 40)

## 4. Calcule a Frequência Acumulada

A frequência acumulada é simplesmente a soma das frequências absolutas à medida que avançamos nas classes.

- **18-22:** 3
- **23-27:** 6 (3 + 3)
- **28-32:** 10 (6 + 4)
- **33-37:** 13 (10 + 3)
- **38-42:** 15 (13 + 2)

## 5. Calcule a Frequência Relativa

A frequência relativa é calculada dividindo a frequência absoluta de cada classe pelo total de dados. Neste caso, temos 15 dados no total.

- **18-22:**  $\frac{3}{15} = 0.2$

- **23-27:**  $\frac{3}{15} = 0.2$
- **28-32:**  $\frac{4}{15} = 0.267$
- **33-37:**  $\frac{3}{15} = 0.2$
- **38-42:**  $\frac{2}{15} = 0.133$

## 6. Calcule a Frequência Percentual

A frequência percentual é simplesmente a frequência relativa multiplicada por 100.

- **18-22:**  $0.2 \times 100 = 20\%$
- **23-27:**  $0.2 \times 100 = 20\%$
- **28-32:**  $0.267 \times 100 = 26.7\%$
- **33-37:**  $0.2 \times 100 = 20\%$
- **38-42:**  $0.133 \times 100 = 13.3\%$

## Tabela de Distribuição de Frequência

Abaixo está a tabela de distribuição de frequência completa:

Intervalo de Idade	Frequência Absoluta (f)	Frequência Acumulada (F)	Frequência Relativa (fr)	Frequência Percentual (fp%)
18-22	3	3	0.2	20%
23-27	3	6	0.2	20%
28-32	4	10	0.267	26.7%
33-37	3	13	0.2	20%
38-42	2	15	0.133	13.3%
<b>Total</b>	<b>15</b>	-	<b>1.0</b>	<b>100%</b>

## Algoritmo para Criar Classes em Distribuição de Frequência

A criação de classes (ou intervalos de classe) em uma distribuição de frequência é um passo fundamental na análise de dados, pois nos permite agrupar os valores de uma variável contínua em intervalos que facilitam a análise e a visualização. Esse processo é especialmente útil quando temos um grande número de dados e queremos resumir a distribuição deles de maneira eficaz.

### Passos para Criar Classes em Distribuição de Frequência

Aqui estão os principais passos para construir as classes de uma distribuição de frequência:

#### 1. Determinação do Número de Classes

O número de classes a ser usado pode ser definido com base no número total de observações ou de acordo com critérios específicos. Um critério comum para determinar o número de classes é a **regra de Sturges**, que pode ser calculada da seguinte maneira:

$$k = 1 + 3.322 \log(n)$$

Onde:

- $k$  é o número de classes
- $n$  é o número de dados na amostra

## 2. Determinação do Intervalo das Classes

O intervalo (ou amplitude) de cada classe é calculado com base na diferença entre o valor máximo e o valor mínimo dos dados, dividida pelo número de classes:

$$\text{Amplitude da Classe} = \frac{\text{Valor Máximo} - \text{Valor Mínimo}}{k}$$

## 3. Criação das Classes

A partir do valor mínimo dos dados, criamos intervalos consecutivos (classes) com a amplitude definida no passo anterior. Cada classe deve cobrir um intervalo de valores dentro do conjunto de dados.

## 4. Distribuição dos Dados nas Classes

Cada valor do conjunto de dados é alocado na classe correspondente com base em seu valor. Ao final, cada classe terá uma frequência absoluta (número de elementos que caem dentro do intervalo da classe).

---

## Exemplo Prático

Vamos usar um conjunto de dados fictício de idades, como já foi mostrado anteriormente:

```
[18, 22, 22, 24, 26, 26, 28, 30, 30, 32, 32, 34, 36, 38, 40]
```

### Passo 1: Determinar o Número de Classes

Para calcular o número de classes  $k$ , usamos a fórmula de Sturges:

$$k = 1 + 3.322 \log(15)$$

Vamos calcular:

\$

$$k = 1 + 3.322 \log(15) \approx 1 + 3.322 \times 1.176 = 1 + 3.91 \approx 4.91$$

\$

Arredondamos para o número inteiro mais próximo, ou seja, 5 classes.

### Passo 2: Determinar o Intervalo das Classes

Agora, calculamos a amplitude das classes. O valor mínimo é 18 e o valor máximo é 40.

\$

$$\text{Amplitude da Classe} = \frac{40 - 18}{5} = \frac{22}{5} = 4.4$$

\$

Arredondamos para o número inteiro mais próximo, ou seja, 4. Agora sabemos que cada classe terá uma amplitude de 4.

### Passo 3: Criar as Classes

Com a amplitude de 4, criamos as classes a partir do valor mínimo (18). As classes serão:

- 18-22
- 23-27
- 28-32
- 33-37
- 38-42

### Passo 4: Distribuição dos Dados nas Classes

Agora vamos distribuir os dados nas classes:

- **18-22:** 18, 22, 22 (3 dados)
- **23-27:** 24, 26, 26 (3 dados)
- **28-32:** 28, 30, 30, 32 (4 dados)
- **33-37:** 32, 34, 36 (3 dados)
- **38-42:** 38, 40 (2 dados)

---

### Tabela Completa com os Cálculos

Intervalo de Idade	Frequência Absoluta (f)	Frequência Acumulada (F)	Frequência Relativa (fr)	Frequência Percentual (fp%)
18-22	3	3	0.2	20%
23-27	3	6	0.2	20%
28-32	4	10	0.267	26.7%
33-37	3	13	0.2	20%

Intervalo de Idade	Frequência Absoluta (f)	Frequência Acumulada (F)	Frequência Relativa (fr)	Frequência Percentual (fp%)
38-42	2	15	0.133	13.3%
<b>Total</b>	<b>15</b>	<b>-</b>	<b>1.0</b>	<b>100%</b>

## Resumo do Algoritmo para Criar Classes

1. **Determine o número de classes** (usando a fórmula de Sturges ou outra metodologia).
2. **Calcule a amplitude das classes** com base no intervalo total dos dados.
3. **Crie as classes** a partir do valor mínimo dos dados, utilizando a amplitude calculada.
4. **Distribua os dados nas classes** e calcule a frequência de cada classe.
5. **Calcule as frequências acumuladas, relativas e percentuais.**

Esse processo permite agrupar os dados de maneira eficiente, ajudando a visualizar e entender melhor a distribuição dos valores na amostra.

## Exemplo de Distribuição de Frequência para Dados de Ponto Flutuante

Quando lidamos com dados de ponto flutuante (decimais), o processo de criação de classes e a construção da tabela de distribuição de frequência segue a mesma lógica, mas devemos estar atentos às casas decimais para definir corretamente os intervalos e garantir uma análise precisa.

### Passos para Criar Classes em Distribuição de Frequência para Dados de Ponto Flutuante

1. **Determine o número de classes**, utilizando a fórmula de Sturges ou outra metodologia.
2. **Calcule a amplitude das classes**, considerando o intervalo dos dados e a precisão necessária para os números decimais.
3. **Crie as classes**, com intervalos baseados na amplitude calculada, ajustando as casas decimais conforme necessário.
4. **Distribua os dados nas classes**, contando quantos valores caem dentro de cada intervalo.
5. **Calcule as frequências acumuladas, relativas e percentuais**, como na tabela de dados inteiros.

### Exemplo Prático com Dados de Ponto Flutuante

Vamos considerar um conjunto de dados fictício de idades com pontos flutuantes:

```
[18.2, 22.5, 22.1, 24.3, 26.7, 26.4, 28.9, 30.0, 30.5, 32.8, 32.2, 34.1, 36.3, 38.5, 40.2]
```

### Passo 1: Determinar o Número de Classes

Utilizando a fórmula de Sturges para o número de classes  $k$ :

$$k = 1 + 3.322 \sqrt{\log(n)}$$

Onde  $n = 15$ , portanto:

$$k = 1 + 3.322 \sqrt{\log(15)} \approx 1 + 3.322 \times 1.176 = 1 + 3.91 \approx 5.91$$

Arredondamos para 6 classes.

### Passo 2: Determinar o Intervalo das Classes

O valor mínimo é 18.2 e o valor máximo é 40.2. A amplitude da classe é calculada da seguinte forma:

$$\text{Amplitude da Classe} = \frac{40.2 - 18.2}{6} = \frac{22}{6} \approx 3.67$$

A amplitude será de aproximadamente 3.7. Vamos arredondar para 3.7.

### Passo 3: Criar as Classes

As classes serão definidas a partir do valor mínimo, com a amplitude de 3.7:

- 18.2 - 21.9
- 22.0 - 25.7
- 25.8 - 29.5
- 29.6 - 33.3
- 33.4 - 37.1
- 37.2 - 40.9

### Passo 4: Distribuição dos Dados nas Classes

Agora vamos distribuir os dados nas classes:

- **18.2 - 21.9:** 18.2, 22.1 (2 dados)
- **22.0 - 25.7:** 22.5, 24.3, 26.4 (3 dados)
- **25.8 - 29.5:** 26.7, 28.9, 30.0 (3 dados)
- **29.6 - 33.3:** 30.5, 32.8, 32.2 (3 dados)
- **33.4 - 37.1:** 34.1, 36.3 (2 dados)
- **37.2 - 40.9:** 38.5, 40.2 (2 dados)

---

## Tabela Completa com os Cálculos



Intervalo de Idade (Classe)	Frequência Absoluta (f)	Frequência Acumulada (F)	Frequência Relativa (fr)	Frequência Percentual (fp%)
18.2 - 21.9	2	2	0.133	13.3%
22.0 - 25.7	3	5	0.2	20%
25.8 - 29.5	3	8	0.2	20%
29.6 - 33.3	3	11	0.2	20%
33.4 - 37.1	2	13	0.133	13.3%
37.2 - 40.9	2	15	0.133	13.3%
<b>Total</b>	<b>15</b>	<b>-</b>	<b>1.0</b>	<b>100%</b>

## Resumo do Algoritmo para Criar Classes em Dados de Ponto Flutuante

1. **Determine o número de classes** (com base na fórmula de Sturges ou outras).
2. **Calcule a amplitude das classes**, considerando a precisão decimal dos dados.
3. **Crie as classes** a partir do valor mínimo e usando a amplitude calculada.
4. **Distribua os dados nas classes** e calcule a frequência absoluta de cada classe.
5. **Calcule as frequências acumuladas, relativas e percentuais.**

Essa abordagem permite a construção de uma distribuição de frequência precisa para dados de ponto flutuante, ajudando na visualização e compreensão da distribuição dos dados.

## Tipos de Distribuição de Frequência

Existem diferentes formas de representar e organizar distribuições de frequência, dependendo do tipo de dados e da análise desejada:

- **Distribuição de Frequência Simples:**  
Quando os dados são organizados em uma única variável, sem a necessidade de subdividir em mais categorias ou características.
- **Distribuição de Frequência Agrupada:**  
Usada quando os dados são contínuos ou têm muitos valores diferentes. Neste caso, os dados são agrupados em intervalos.
- **Distribuição de Frequência Cumulativa:**  
A distribuição de frequência acumulada mostra como os dados se acumulam à medida que você avança pelas classes. Essa distribuição é útil para visualizar a quantidade total de dados até um determinado ponto.

## Por que Utilizar Distribuição de Frequência?

1. **Simplificação dos Dados:**  
Organizar os dados em uma distribuição de frequência torna a análise mais simples, permitindo

identificar padrões e tendências de forma rápida.

## 2. Visualização das Características dos Dados:

As distribuições de frequência ajudam a visualizar a dispersão, a concentração de valores e a simetria ou assimetria dos dados. Isso é importante para entender a distribuição dos dados e decidir sobre a melhor análise estatística a ser feita.

## 3. Comparação de Dados:

Permite a comparação entre diferentes conjuntos de dados ao examinar suas distribuições de frequência. Você pode comparar distribuições de diferentes variáveis ou até de diferentes grupos de indivíduos.

## 4. Fundamento para Cálculos Estatísticos:

A distribuição de frequência é a base para muitos cálculos estatísticos, como a média, mediana, moda, desvio padrão, entre outros.

# Uso de Ferramentas na Construção de Distribuições de Frequência

---

No contexto de análise de dados, o uso de ferramentas de software, como **Python**, é essencial para automatizar, validar e realizar cálculos precisos em grandes volumes de dados. Ao criar distribuições de frequência, a análise manual pode ser complexa e propensa a erros, enquanto ferramentas como **pandas**, **numpy** e **matplotlib** permitem realizar esses cálculos e gerar visualizações de maneira rápida e eficiente. Abaixo, discutimos o uso dessas ferramentas e por que elas são importantes no processo de criação de distribuições de frequência.

## Exemplo python

- No seguinte link [Distribuição de frequência COLAB Python](#) há um exemplo completo em python de distribuição de frequência
- No seguinte link [Gerar dataset COLAB](#)

Para executar os arquivo acima abra o site [COLAB REsearch]

# Por Que Usar Ferramentas no Processo de Distribuição de Frequência?

## 1. Automatização do Processo:

- A criação de uma tabela de distribuição de frequência envolve várias etapas: cálculo de classes, contagem de elementos em cada classe, cálculo de frequências acumuladas e relativas, entre outros. Ao usar ferramentas, podemos automatizar todo esse processo, o que não só economiza tempo, mas também reduz as chances de erro humano.

## 2. Escalabilidade:

- Para grandes volumes de dados, a análise manual se torna inviável. Ferramentas como **pandas** são projetadas para lidar com datasets grandes, permitindo que você trabalhe com

milhões de dados de forma eficiente, sem perder desempenho.

### 3. Precisão nos Cálculos:

- Calculando distribuições de frequência manualmente, há sempre o risco de cometer erros ao contar ou arredondar valores. Ferramentas de software garantem que os cálculos sejam feitos com precisão, seguindo fórmulas matemáticas bem definidas.

### 4. Visualização:

- Uma das principais vantagens de usar ferramentas como **matplotlib** ou **seaborn** em Python é a capacidade de gerar visualizações, como histogramas, que permitem uma interpretação rápida e intuitiva dos dados. As visualizações são essenciais para comunicar os resultados de maneira clara e eficaz para diferentes públicos.

### 5. Facilidade de Reprodutibilidade:

- Ao utilizar ferramentas e escrever código para análise de dados, você cria um processo que pode ser facilmente reproduzido. Isso é especialmente importante quando você precisa realizar a mesma análise em diferentes conjuntos de dados ou em atualizações futuras dos dados.

## Exemplo de Implementação com Python

Aqui está um exemplo de como Python pode ser utilizado para calcular e visualizar uma distribuição de frequência com dados de ponto flutuante. Utilizamos as bibliotecas **pandas** para manipulação dos dados e **matplotlib** para visualização.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# Dados fictícios de ponto flutuante
dados = [18.2, 22.5, 22.1, 24.3, 26.7, 26.4, 28.9, 30.0, 30.5, 32.8, 32.2,
34.1, 36.3, 38.5, 40.2]

# Criação do DataFrame
df = pd.DataFrame(dados, columns=['Idade'])

# Número de classes a partir da fórmula de Sturges
num_classes = 6

# Definindo os intervalos com a amplitude dos dados
bins = np.histogram_bin_edges(df['Idade'], bins=num_classes)
labels = [f"{int(bins[i])}-{int(bins[i+1])}" for i in range(len(bins)-1)]

# Criando a tabela de distribuição de frequência
df['Classe'] = pd.cut(df['Idade'], bins=bins, labels=labels, right=False)

# Calculando a frequência absoluta (f), acumulada (F), relativa (fr) e
```

```

percentual (fp%)
tabela_freq = df['Classe'].value_counts().sort_index().reset_index()
tabela_freq.columns = ['Classe', 'f']
tabela_freq['F'] = tabela_freq['f'].cumsum()
tabela_freq['fr'] = tabela_freq['f'] / tabela_freq['f'].sum()
tabela_freq['fp (%)'] = tabela_freq['fr'] * 100

# Exibindo a tabela
print(tabela_freq)

# Gerando um histograma para visualização
plt.hist(df['Idade'], bins=bins, edgecolor='black', alpha=0.7)
plt.title('Histograma de Idade')
plt.xlabel('Faixa Etária')
plt.ylabel('Frequência')
plt.show()

```

## Resultado Esperado

A execução desse código resultará em:

1. **Tabela de Frequência:** Mostrando as classes, frequência absoluta, acumulada, relativa e percentual, como já discutido anteriormente.
2. **Histograma:** Uma representação visual da distribuição dos dados, facilitando a compreensão de sua dispersão e concentração.

---

O uso de ferramentas de análise de dados, como Python e suas bibliotecas (pandas, numpy, matplotlib), oferece vantagens significativas quando estamos lidando com grandes volumes de dados e processos de análise complexos. Além de garantir a precisão dos cálculos, essas ferramentas permitem gerar visualizações claras, reprodutibilidade e, o mais importante, facilitam a tomada de decisões informadas com base nos dados.

## Conclusão

A **distribuição de frequência** é uma das ferramentas mais poderosas e versáteis na estatística. Ela ajuda a organizar, descrever e analisar dados, tornando-os mais compreensíveis e acessíveis para tomada de decisões. Seja em uma análise exploratória de dados, em um estudo de amostra ou na modelagem de dados para aprendizado de máquina, entender como construir e interpretar distribuições de frequência é um passo fundamental para qualquer análise estatística.