

# **Estatística**

**Prof. Diego de Sousa Aguiar**

**Assunto da Aula de Hoje: Introdução**

## **INTRODUÇÃO**

A Estatística pode ser pensada como a ciência de aprendizagem a partir de dados.

Em linhas gerais, a Estatística fornece métodos que auxiliam o processo de tomada de decisão.

Estatística está presente em todas as áreas da ciência que envolvam a coleta e análise de dados.

# INTRODUÇÃO

A Estatística está compreendida em duas partes:

- **Estatística Descritiva:** Reúne um conjunto de técnicas para sumarizar os dados (tabelas, gráficos) e medidas descritivas que permitem tirar muitas informações contidas nos dados.
- **Estatística Indutiva:** Produzir afirmações sobre uma dada característica da população, na qual estamos interessados, a partir de informações colhidas de uma parte dessa população.

## CONCEITO BÁSICOS

- **População (N):** Conjunto de todos os elementos relativos a um determinado fenômeno que possuem pelo menos uma característica em comum, a população é o conjunto Universo, podendo ser finita ou infinita.
  - **Finita:** apresenta um número limitado de observações, que é passível de contagem;
  - **Infinita:** apresenta um número ilimitado de observações que é impossível de contar e geralmente esta associada a processos.
- **Amostra (n):** É um subconjunto da população e deverá ser considerada finita, a amostra deve ser selecionada seguindo certas regras e deve ser representativa, de modo que ela represente todas as características da população como se fosse uma fotografia desta.

## CONCEITO BÁSICOS

- **Pesquisa Estatística:** É qualquer informação retirada de uma população ou amostra, podendo ser através de Censo ou Amostragem.
  - **Censo:** é a coleta exaustiva de informações das "N" unidades populacionais.
  - **Amostragem:** É o processo de retirada de informações dos "n" elementos amostrais, no qual deve seguir um método criterioso e adequado (tipos de amostragem).
- **Dados Estatístico:** é qualquer característica que possa ser observada ou medida de alguma maneira. As matérias-primas da estatística são os dados observáveis, as observações de uma ou mais variáveis.

## CONCEITO BÁSICOS

- **Variável:** É aquilo que se deseja observar para se tirar algum tipo de conclusão, geralmente as variáveis para estudo são selecionadas por processos de amostragem. Os símbolos utilizados para representar as variáveis são as letras maiúsculas do alfabeto, tais como X, Y, Z, ... que pode assumir qualquer valor de um conjunto de dados. As variáveis podem ser classificadas dos seguintes modos:
  - **Qualitativas:** São características de uma população que não pode ser medidas.
  - **Quantitativas:** São características populacionais que podem ser quantificadas, sendo classificadas em discretas e contínuas.

## CONCEITO BÁSICOS

- **Variável:** É aquilo que se deseja observar para se tirar algum tipo de conclusão, geralmente as variáveis para estudo são selecionadas por processos de amostragem. Os símbolos utilizados para representar as variáveis são as letras maiúsculas do alfabeto, tais como X, Y, Z, ... que pode assumir qualquer valor de um conjunto de dados. As variáveis podem ser classificadas dos seguintes modos:
  - **Qualitativas:** São características de uma população que não pode ser medidas.
  - **Quantitativas:** São características populacionais que podem ser quantificadas, sendo classificadas em discretas e contínuas.

## CONCEITO BÁSICOS

- **Qualitativas:** São características de uma população que não pode ser medidas.
  - **Nominal:** são utilizados símbolos, ou números, para representar determinado tipo de dados, mostrando, assim, a qual grupo ou categoria eles pertencem.
  - **Ordinal:** quando uma classificação for dividida em categorias ordenadas em graus convencionados, havendo uma relação entre as categorias do tipo "maior do que", "menor do que", "igual a", os dados por postos consistem de valores relativos atribuídos para denotar a ordem de primeiro, segundo, terceiro e, assim, sucessivamente.



## CONCEITO BÁSICOS

- **Variável:** É aquilo que se deseja observar para se tirar algum tipo de conclusão, geralmente as variáveis para estudo são selecionadas por processos de amostragem. Os símbolos utilizados para representar as variáveis são as letras maiúsculas do alfabeto, tais como X, Y, Z, ... que pode assumir qualquer valor de um conjunto de dados. As variáveis podem ser classificadas dos seguintes modos:
  - **Qualitativas:** São características de uma população que não pode ser medidas.
  - **Quantitativas:** São características populacionais que podem ser quantificadas, sendo classificadas em discretas e contínuas.

## CONCEITO BÁSICOS

- **Quantitativas:** São características populacionais que podem ser quantificadas, sendo classificadas em discretas e contínuas.
  - **Discretas:** são aquelas variáveis que pode assumir somente valores inteiros num conjunto de valores. É gerada pelo processo de contagem, como o número de veículos que passa em um posto de gasolina, o número de estudantes nesta sala de aula.
  - **Contínuas:** são aquelas variáveis que podem assumir um valor dentro de um intervalo de valores. É gerada pelo processo de medição. Neste caso serve como exemplo o volume de água em um reservatório ou o peso de um pacote de cereal.

## CONCEITO BÁSICOS

- **Parâmetros:** são medidas populacionais quando se investiga a população em sua totalidade, neste caso é impossível fazer inferências, pois toda a população foi investigada.
- **Estimadores:** são medidas obtidas da amostra, torna-se possível neste caso utilizarmos as teorias inferências para que possamos fazer conclusões sobre a população.

## CONCEITO BÁSICOS

Parâmetro (valores reais)		Estimador (valores estimados)
$\mu$	Média	$\bar{X}$
$\sigma^2$	Variância	$S^2$
$\sigma$	Desvio Padrão	$S$
$\pi$	Proporção	$\hat{p}$

# REPRESENTAÇÃO EM TABELA

Título

---

---

Cabeçalho
-----------

---

Corpo

---

Rodapé

## SÉRIES ESTATÍSTICA

Uma série estatística é um conjunto de dados ordenados segundo uma característica comum, as quais servirão posteriormente para se fazer análises e inferências.

- **Série Temporal ou Cronológica:** É a série cujos dados estão dispostos em correspondência com o tempo, ou seja, varia o tempo e permanece constante o fato e o local.
- **Série Geográfica ou Territorial:** É a série cujos dados estão dispostos em correspondência com o local, ou seja, varia o local e permanece constante a época e o fato.
- **Série Específica ou Qualitativa:** É a série cujos dados estão dispostos em correspondência com a espécie ou qualidade, ou seja, varia o fato e permanece constante a época e o local.

## SÉRIES ESTATÍSTICA

Uma série estatística é um conjunto de dados ordenados segundo uma característica comum, as quais servirão posteriormente para se fazer análises e inferências.

- **Série Temporal ou Cronológica:** É a série cujos dados estão dispostos em correspondência com o tempo, ou seja, varia o tempo e permanece constante o fato e o local.
- **Série Geográfica ou Territorial:** É a série cujos dados estão dispostos em correspondência com o local, ou seja, varia o local e permanece constante a época e o fato.
- **Série Específica ou Qualitativa:** É a série cujos dados estão dispostos em correspondência com a espécie ou qualidade, ou seja, varia o fato e permanece constante a época e o local.

## Número de Gols da seleção Brasileira em Copas do Mundo

Ano	Número de Gols
1930	5
1934	1
1938	14
1950	22
1954	8
1958	16
1962	14
1966	4
1970	19
1974	6
1978	10
1982	15
1986	10
1990	4
1994	11
1998	14
2002	18
2006	10
2010	9
2014	11
2018	8

Fonte: FIFA



## SÉRIES ESTATÍSTICA

Uma série estatística é um conjunto de dados ordenados segundo uma característica comum, as quais servirão posteriormente para se fazer análises e inferências.

- **Série Temporal ou Cronológica:** É a série cujos dados estão dispostos em correspondência com o tempo, ou seja, varia o tempo e permanece constante o fato e o local.
- **Série Geográfica ou Territorial:** É a série cujos dados estão dispostos em correspondência com o local, ou seja, varia o local e permanece constante a época e o fato.
- **Série Específica ou Qualitativa:** É a série cujos dados estão dispostos em correspondência com a espécie ou qualidade, ou seja, varia o fato e permanece constante a época e o local.

Total de Casos Confirmados de COVID-19 em 21/04/2020

País	Casos Confirmados
Estados Unidos	798742
Espanha	200210
Itália	181228
Alemanha	147139
Reino Unido	124743

Fonte: OMS

## SÉRIES ESTATÍSTICA

Uma série estatística é um conjunto de dados ordenados segundo uma característica comum, as quais servirão posteriormente para se fazer análises e inferências.

- **Série Temporal ou Cronológica:** É a série cujos dados estão dispostos em correspondência com o tempo, ou seja, varia o tempo e permanece constante o fato e o local.
- **Série Geográfica ou Territorial:** É a série cujos dados estão dispostos em correspondência com o local, ou seja, varia o local e permanece constante a época e o fato.
- **Série Específica ou Qualitativa:** É a série cujos dados estão dispostos em correspondência com a espécie ou qualidade, ou seja, varia o fato e permanece constante a época e o local.

Desempenho dos alunos em Estatística em 2019

Avaliação	Número de Alunos
Ótimo/Bom	35
Regular	25
Ruim/Péssimo	20

Fonte: Relatório do professor

## SÉRIES ESTATÍSTICA

Além das séries citadas, podemos ter as **séries mistas** que consiste da combinação de duas séries.

Casos Confirmados de COVID-19 no primeiro trimestre de 2020

País	Meses		
	Janeiro	Fevereiro	Março
Estados Unidos	6	56	186577
Espanha	0	32	102104
Itália	2	1692	104098
Alemanha	4	53	71637
Reino Unido	0	20	25130

Fonte: OMS

## SÉRIES ESTATÍSTICA

Já a distribuição de frequência é o tipo de série estatística na qual permanece constante o fato, o local e a época. Os dados são colocados em classes preestabelecidas, registrando a frequência de ocorrência. Uma distribuição de frequência pode ser classificada em discreta e intervalar.

- **Distribuição de Frequência Discreta ou Pontual:** É uma série de dados agrupados na qual o número de observações está relacionado com um ponto real.
- **Distribuição de Frequências Intervalar:** Na distribuição de frequência, os intervalos parciais deverão ser apresentados de maneira a evitar dúvidas quanto à classe a que permanece determinado elemento.

## SÉRIES ESTATÍSTICA

Já a distribuição de frequência é o tipo de série estatística na qual permanece constante o fato, o local e a época. Os dados são colocados em classes preestabelecidas, registrando a frequência de ocorrência. Uma distribuição de frequência pode ser classificada em discreta e intervalar.

- **Distribuição de Frequência Discreta ou Pontual:** É uma série de dados agrupados na qual o número de observações está relacionado com um ponto real.
- **Distribuição de Frequências Intervalar:** Na distribuição de frequência, os intervalos parciais deverão ser apresentados de maneira a evitar dúvidas quanto à classe a que permanece determinado elemento.



Medição de níveis de ruído em uma unidade fabril

Nível de Ruído (dB)	Frequência
79	2
81	4
82	2
84	3
85	1
86	1

Fonte: Dados Fictícios

## SÉRIES ESTATÍSTICA

Já a distribuição de frequência é o tipo de série estatística na qual permanece constante o fato, o local e a época. Os dados são colocados em classes preestabelecidas, registrando a frequência de ocorrência. Uma distribuição de frequência pode ser classificada em discreta e intervalar.

- **Distribuição de Frequência Discreta ou Pontual:** É uma série de dados agrupados na qual o número de observações está relacionado com um ponto real.
- **Distribuição de Frequências Intervalar:** Na distribuição de frequência, os intervalos parciais deverão ser apresentados de maneira a evitar dúvidas quanto à classe a que permanece determinado elemento.

Notas dos alunos em Estatística do 2019

Notas	Frequência
8  --- 10	25
6  --- 8	27
4  --- 6	10
2  --- 4	12
0  --- 2	6
Total	80

Fonte: Relatório do Professor

## REPRESENTAÇÃO GRÁFICA

Os gráficos são uma forma de apresentação visual dos dados. Normalmente, contém menos informações que as tabelas, mas são de mais fácil leitura. O tipo de gráfico depende da variável em questão.

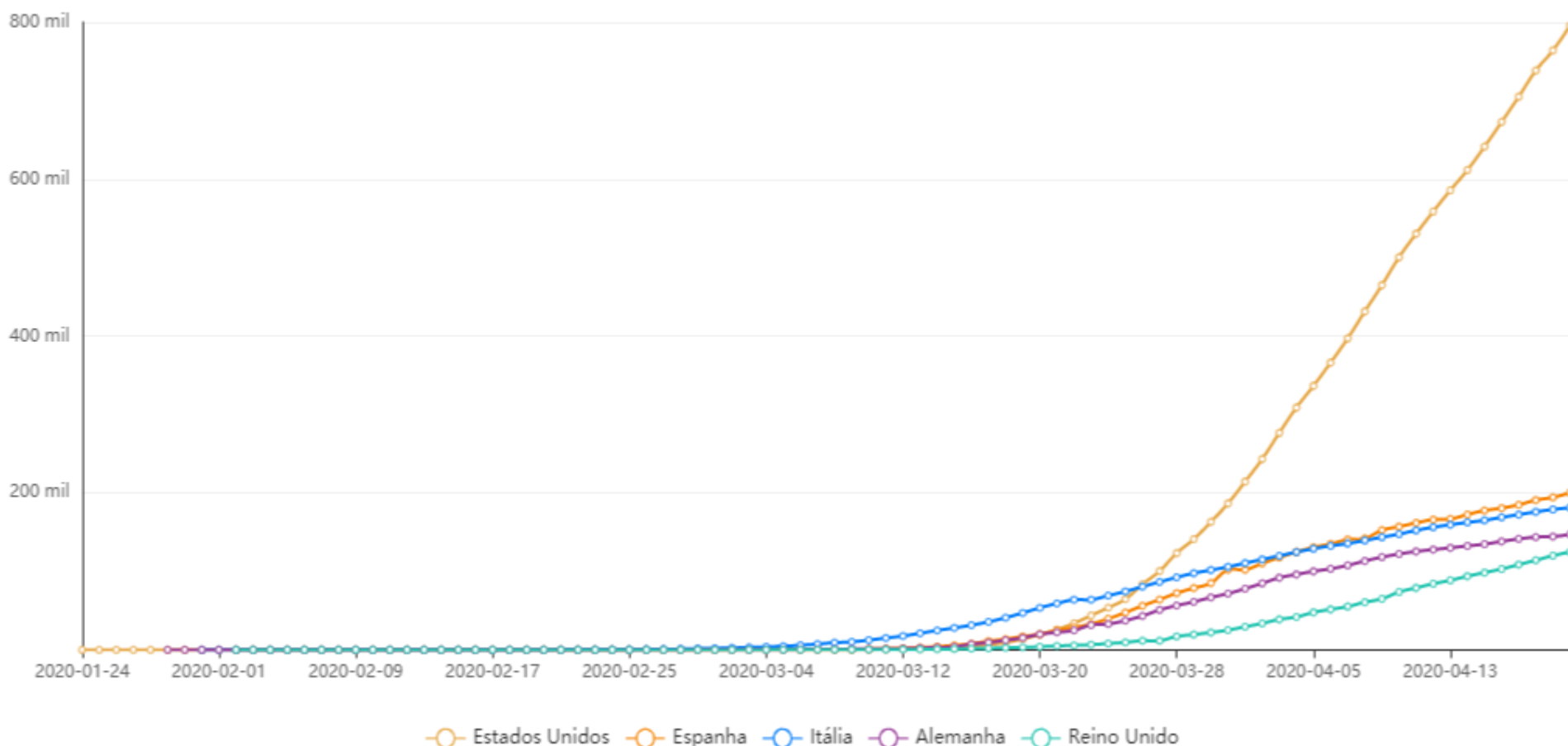
- **Gráfico de Linhas:** é usado para ilustrar uma série temporal.
- **Gráfico de Colunas ou Barras:** Representação gráfica da distribuição de frequências. Este gráfico é utilizado para variáveis nominais e ordinais.
- **Gráficos Circulares ou de Setores:** Representação gráfica da frequência relativa (percentagem) de cada categoria da variável. Este gráfico é utilizado para variáveis nominais e ordinais.

## REPRESENTAÇÃO GRÁFICA

Os gráficos são uma forma de apresentação visual dos dados. Normalmente, contém menos informações que as tabelas, mas são de mais fácil leitura. O tipo de gráfico depende da variável em questão.

- **Gráfico de Linhas:** é usado para ilustrar uma série temporal.
- **Gráfico de Colunas ou Barras:** Representação gráfica da distribuição de frequências. Este gráfico é utilizado para variáveis nominais e ordinais.
- **Gráficos Circulares ou de Setores:** Representação gráfica da frequência relativa (percentagem) de cada categoria da variável. Este gráfico é utilizado para variáveis nominais e ordinais.

Total de Casos Confirmados de COVID-19 até 21/04/2020



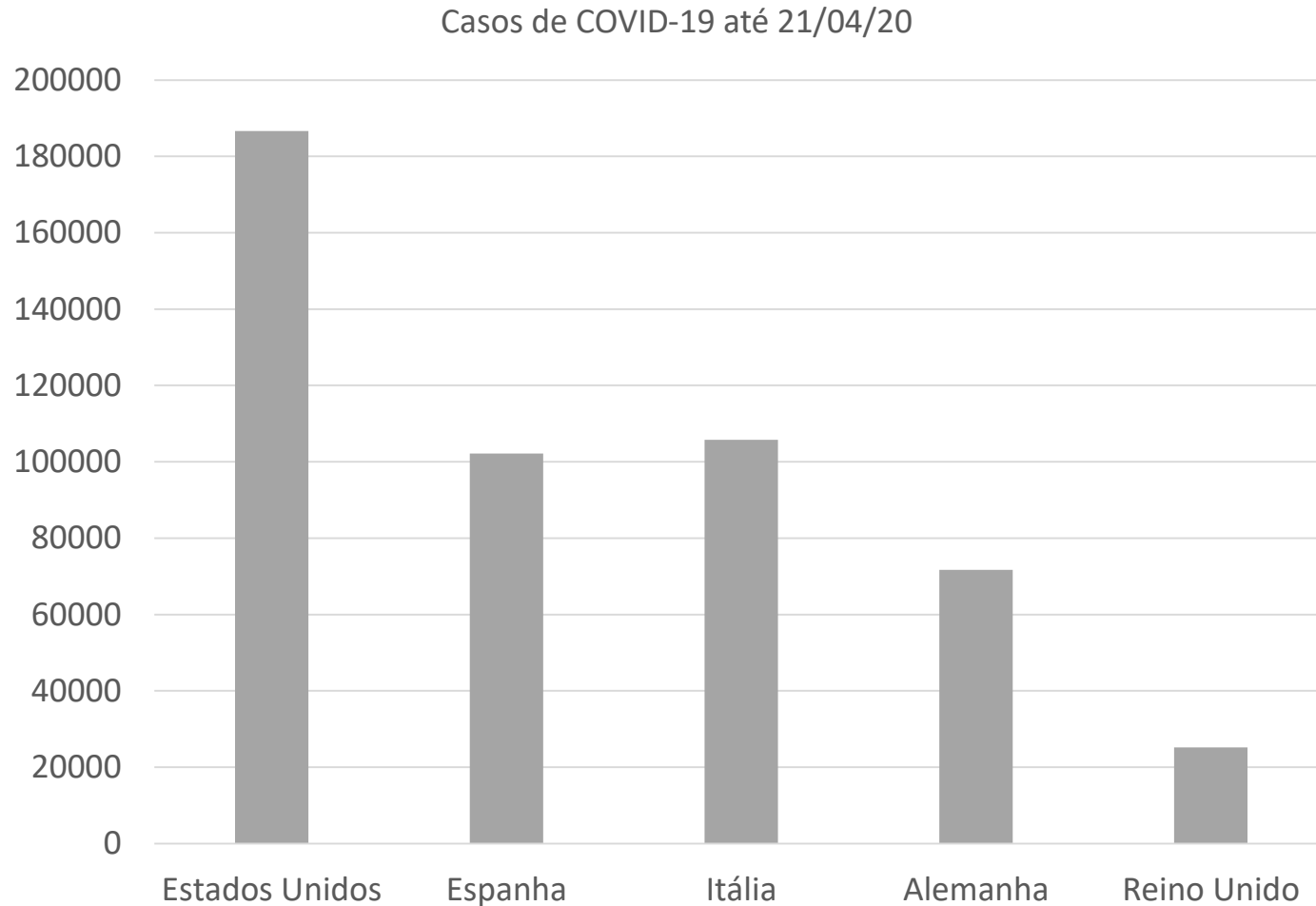
— Estados Unidos — Espanha — Itália — Alemanha — Reino Unido

Fonte: OMS

## REPRESENTAÇÃO GRÁFICA

Os gráficos são uma forma de apresentação visual dos dados. Normalmente, contém menos informações que as tabelas, mas são de mais fácil leitura. O tipo de gráfico depende da variável em questão.

- **Gráfico de Linhas:** é usado para ilustrar uma série temporal.
- **Gráfico de Colunas ou Barras:** Representação gráfica da distribuição de frequências. Este gráfico é utilizado para variáveis nominais e ordinais.
- **Gráficos Circulares ou de Setores:** Representação gráfica da frequência relativa (percentagem) de cada categoria da variável. Este gráfico é utilizado para variáveis nominais e ordinais.



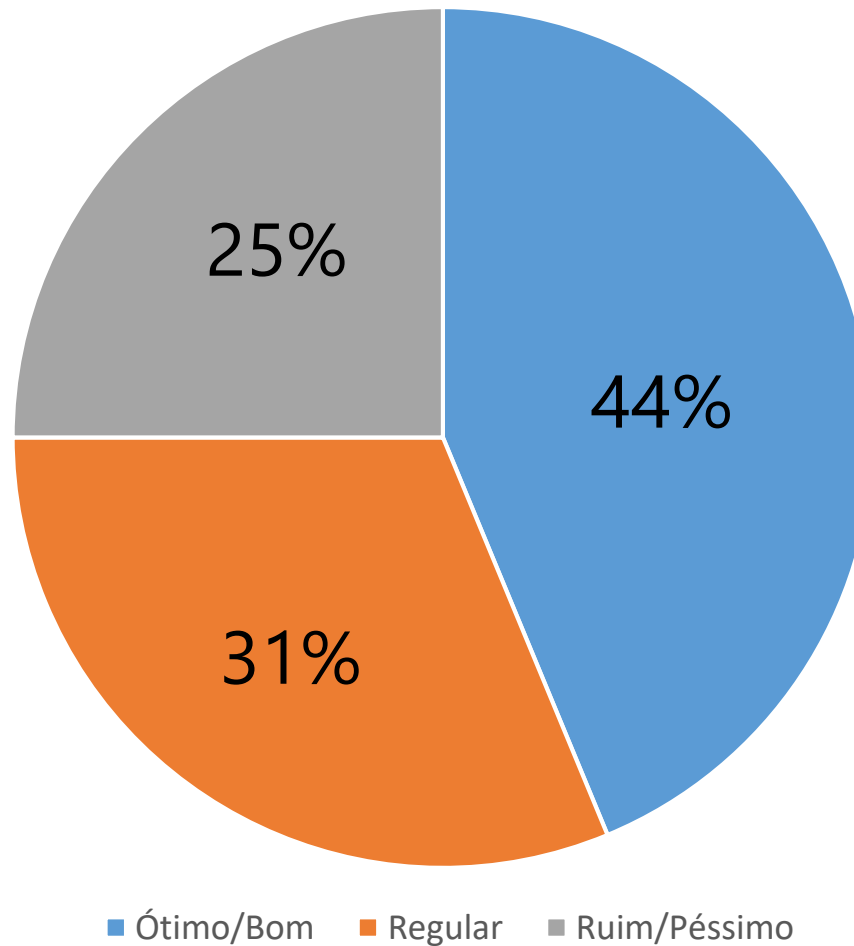


## REPRESENTAÇÃO GRÁFICA

Os gráficos são uma forma de apresentação visual dos dados. Normalmente, contém menos informações que as tabelas, mas são de mais fácil leitura. O tipo de gráfico depende da variável em questão.

- **Gráfico de Linhas:** é usado para ilustrar uma série temporal.
- **Gráfico de Colunas ou Barras:** Representação gráfica da distribuição de frequências. Este gráfico é utilizado para variáveis nominais e ordinais.
- **Gráficos Circulares ou de Setores:** Representação gráfica da frequência relativa (percentagem) de cada categoria da variável. Este gráfico é utilizado para variáveis nominais e ordinais.

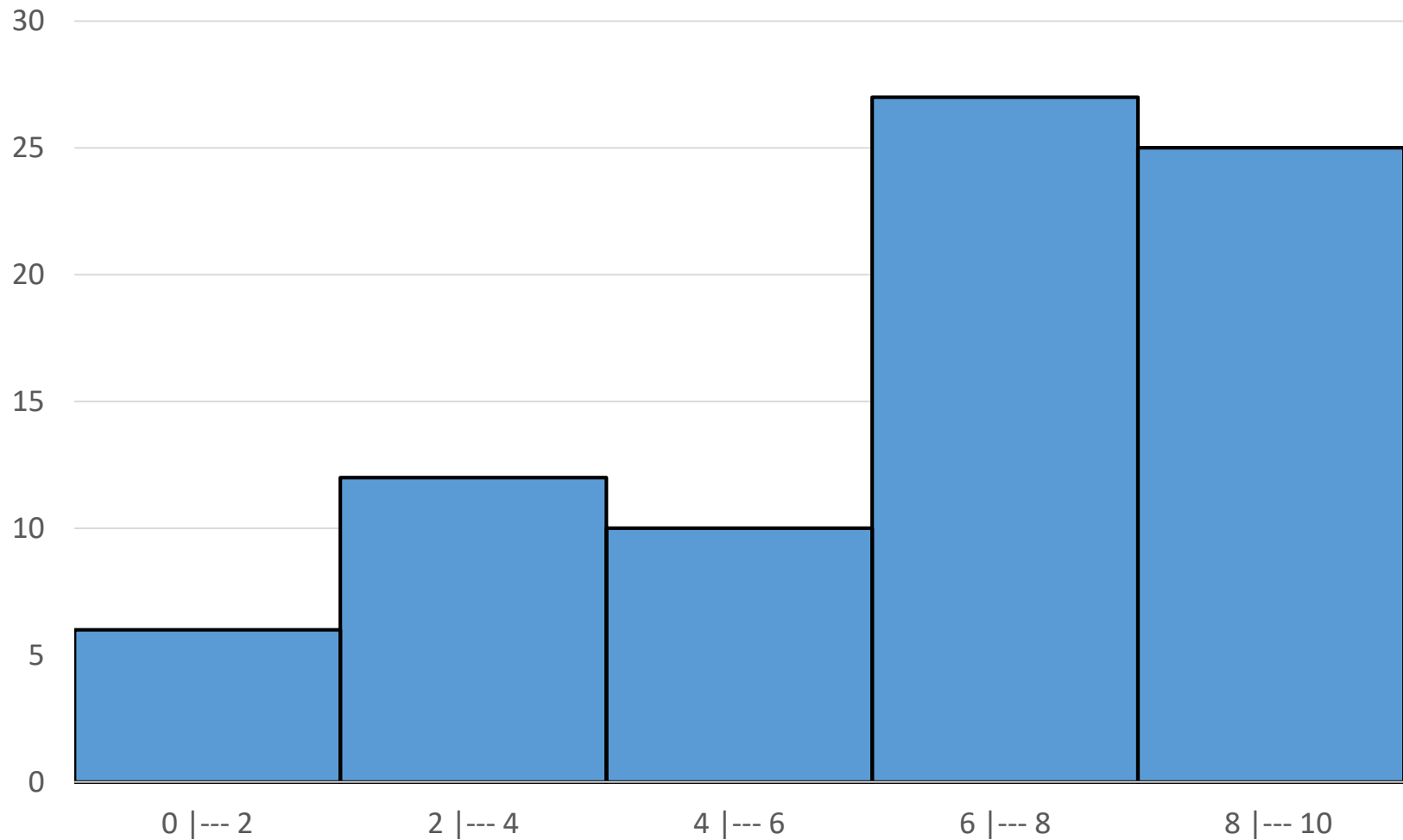
Desempenho dos alunos em Estatística em 2019



## REPRESENTAÇÃO GRÁFICA

Exemplos de gráficos para a análise de uma distribuição de frequência temos os **Histogramas**.

Notas dos alunos em Estatística em 2019



# **Estatística**

**Prof. Diego de Sousa Aguiar**

**Assunto da Aula de Hoje: Introdução**