

ME414 - Estatística para Experimentalistas

Introdução: Estatística, Aplicações e Conceitos Básicos

Prof. Carlos Trucíos
ctrucios@unicamp.br
ctruciosm.github.io

Instituto de Matemática, Estatística e Computação Científica,
Universidade Estadual de Campinas

Aula 1



UNICAMP

Porque estudar esta matéria?

Estatística e Aplicações

Conceitos Básicos

Estatística Descritiva vs. Estatística Inferencial.

Porque estudar esta matéria?

Porque estudar esta matéria?

"Without data you're just another person with an opinion."

"Data are becoming the new raw material of business."

"The world's most valuable resource is no longer oil, but data."

"Learning from data is virtually universally useful. Master it and you will be welcomed anywhere."

"In God we trust. All others must bring data."

"Numbers have an important story to tell. They rely on you to give them a voice."

"The world is now awash in data and we can see consumers in a lot clearer ways"

"Information is the oil of the 21st century, and analytics is the combustion engine"

Porque estudar esta matéria?

glassdoor

Salários de Analista De Dados em Área de Rio de Janeiro

32 Salários Atualizado em 05/03/2021 10:10:33

Salário médio

R\$ 5.191 / mês



Salários de Cientista De Dados em Área de São Paulo

140 Salários Atualizado em 06/03/2021 08:40:35

Salário médio

R\$ 8.038 / mês



Salários de Analista De Dados em Área de São Paulo

141 Salários Atualizado em 06/03/2021 02:14:25

Salário médio

R\$ 4.995 / mês



Salários de Business Intelligence em Área de São Paulo

5 Salários Atualizado em 04/03/2021 09:27:27

Salário médio

R\$ 6.143 / mês



Salários de Cientista De Dados em Área de Rio de Janeiro

18 Salários Atualizado em 01/03/2021 07:39:59

Salário médio

R\$ 8.838 / mês



Estatística e Aplicações

Estatística e Aplicações

O que é Estatística?

É a ciência de coletar, analisar, modelar e interpretar dados.

Nesta matéria aprenderemos:

- ▶ Como fazer uma análise exploratória de dados (EDA)

Estatística e Aplicações

O que é Estatística?

É a ciência de coletar, analisar, modelar e interpretar dados.

Nesta matéria aprenderemos:

- ▶ Como fazer uma análise exploratória de dados (EDA)
- ▶ Como fazer testes de hipóteses

Estatística e Aplicações

O que é Estatística?

É a ciência de coletar, analisar, modelar e interpretar dados.

Nesta matéria aprenderemos:

- ▶ Como fazer uma análise exploratória de dados (EDA)
- ▶ Como fazer testes de hipóteses
- ▶ Como coletar dados para posteriormente serem analisados

Estatística e Aplicações

O que é Estatística?

É a ciência de coletar, analisar, modelar e interpretar dados.

Nesta matéria aprenderemos:

- ▶ Como fazer uma análise exploratória de dados (EDA)
- ▶ Como fazer testes de hipóteses
- ▶ Como coletar dados para posteriormente serem analisados
- ▶ Como interpretar corretamente os resultados das nossas análises estatísticas.

Estatística e Aplicações

O que é Estatística?

É a ciência de coletar, analisar, modelar e interpretar dados.

Nesta matéria aprenderemos:

- ▶ Como fazer uma análise exploratória de dados (EDA)
- ▶ Como fazer testes de hipóteses
- ▶ Como coletar dados para posteriormente serem analisados
- ▶ Como interpretar corretamente os resultados das nossas análises estatísticas.

Estatística e Aplicações

O que é Estatística?

É a ciência de coletar, analisar, modelar e interpretar dados.

Nesta matéria aprenderemos:

- ▶ Como fazer uma análise exploratória de dados (EDA)
- ▶ Como fazer testes de hipóteses
- ▶ Como coletar dados para posteriormente serem analisados
- ▶ Como interpretar corretamente os resultados das nossas análises estatísticas.

Onde vou aplicar estatística na vida real?

- ▶ Estatística tem campo de atuação em uma infinidade de áreas.

Estatística e Aplicações

O que é Estatística?

É a ciência de coletar, analisar, modelar e interpretar dados.

Nesta matéria aprenderemos:

- ▶ Como fazer uma análise exploratória de dados (EDA)
- ▶ Como fazer testes de hipóteses
- ▶ Como coletar dados para posteriormente serem analisados
- ▶ Como interpretar corretamente os resultados das nossas análises estatísticas.

Onde vou aplicar estatística na vida real?

- ▶ Estatística tem campo de atuação em uma infinidade de áreas.
- ▶ **Dados são a nova matéria prima dos negócios**

- ▶ **Auditorias:** Procedimentos de amostragem são utilizados para selecionar aleatoriamente as empresas onde serão feitas auditorias.

Estatística e Aplicações

- ▶ **Auditorias:** Procedimentos de amostragem são utilizados para selecionar aleatoriamente as empresas onde serão feitas auditorias.
- ▶ **Economia e Finanças:** mensuração do risco de mercado (volatilidade, Valor em Risco, Expected Shortfall), alocação de carteiras, identificar se alguém será ou não um bom pagador (risco de crédito), previsões macroeconômicas.

- ▶ **Auditorias:** Procedimentos de amostragem são utilizados para selecionar aleatoriamente as empresas onde serão feitas auditorias.
- ▶ **Economia e Finanças:** mensuração do risco de mercado (volatilidade, Valor em Risco, Expected Shortfall), alocação de carteiras, identificar se alguém será ou não um bom pagador (risco de crédito), previsões macroeconômicas.
- ▶ **Marketing:** Segmentação de clientes (CRM), *Pricing*, testes A/B, quais clientes têm mais probabilidade de comprar um determinado produto ou serviço, atributos mais valorizados no momento de escolher um produto, etc.

- ▶ **Auditorias:** Procedimentos de amostragem são utilizados para selecionar aleatoriamente as empresas onde serão feitas auditorias.
- ▶ **Economia e Finanças:** mensuração do risco de mercado (volatilidade, Valor em Risco, Expected Shortfall), alocação de carteiras, identificar se alguém será ou não um bom pagador (risco de crédito), previsões macroeconômicas.
- ▶ **Marketing:** Segmentação de clientes (CRM), *Pricing*, testes A/B, quais clientes têm mais probabilidade de comprar um determinado produto ou serviço, atributos mais valorizados no momento de escolher um produto, etc.
- ▶ Controle de qualidade, reconhecimento de padrões, textmining, criação indicadores de clima laboral, People Analytics, ...

Porque usamos estatística?

- ▶ Resumir a informação contida nos dados
- ▶ Avaliar/Descobrir relações existentes nos dados
- ▶ Ter *insights* baseados nos dados que nos dem alguma vantagem competitiva.

Conceitos Básicos

Definições

- ▶ **Dados:** são a matéria-prima de qualquer análise estatística, são obtidos da medição de determinadas características das unidades de análise.

Definições

- ▶ **Dados:** são a matéria-prima de qualquer análise estatística, são obtidos da medição de determinadas características das unidades de análise.
- ▶ **Elementos:** ou unidade de análise. São as entidades (pessoa, escolas, empresas, países, etc.) a respeito das quais se coletam dados.

Definições

- ▶ **Dados:** são a matéria-prima de qualquer análise estatística, são obtidos da medição de determinadas características das unidades de análise.
- ▶ **Elementos:** ou unidade de análise. São as entidades (pessoa, escolas, empresas, países, etc.) a respeito das quais se coletam dados.
- ▶ **Variável:** Característica de interesse dos elementos.

Definições

- ▶ **Dados:** são a matéria-prima de qualquer análise estatística, são obtidos da medição de determinadas características das unidades de análise.
- ▶ **Elementos:** ou unidade de análise. São as entidades (pessoa, escolas, empresas, países, etc.) a respeito das quais se coletam dados.
- ▶ **Variável:** Característica de interesse dos elementos.
- ▶ **Observação:** Conjunto de medidas obtidas correspondentes a determinado elemento.

Definições

- ▶ **Dados:** são a matéria-prima de qualquer análise estatística, são obtidos da medição de determinadas características das unidades de análise.
- ▶ **Elementos:** ou unidade de análise. São as entidades (pessoa, escolas, empresas, países, etc.) a respeito das quais se coletam dados.
- ▶ **Variável:** Característica de interesse dos elementos.
- ▶ **Observação:** Conjunto de medidas obtidas correspondentes a determinado elemento.

Definições

- ▶ **Dados:** são a matéria-prima de qualquer análise estatística, são obtidos da medição de determinadas características das unidades de análise.
- ▶ **Elementos:** ou unidade de análise. São as entidades (pessoa, escolas, empresas, países, etc.) a respeito das quais se coletam dados.
- ▶ **Variável:** Característica de interesse dos elementos.
- ▶ **Observação:** Conjunto de medidas obtidas correspondentes a determinado elemento.

Uma **observação** contém **dados** referentes a várias características (*i.e.*, **variáveis**) que foram medidas dos **elementos**.

Definições

Tabela 1: Exemplo de um conjunto de dados.

	Milhas/galão	Cilindros	Potencia	Cambio
Mazda RX4	21.0	6	110	1
Datsun 710	22.8	4	93	1
Hornet Sportabout	18.7	8	175	0

Identifique: observação, variáveis, elementos e dados.

De onde vêm os dados?

Existem diversas fontes de dados, mas elas podem ser classificadas em dois tipos.

- ▶ **Fontes de dados internas:** dados gerados pela própria empresa/organização.
- ▶ **Fontes de dados externas:** dados vindos de fora da empresa/organização. Por sua vez, dados externos podem ser classificados em:
 - ▶ fontes primárias: quando são criados/coledados e publicados pela organização externa (dados do IBGE, Bureaus)
 - ▶ fontes secundárias: quando os dados publicados pela instituição, vem de fato de outra (dados publicados em um journal, mas que são indicadores da FGV)

Fontes primárias são preferidas, pois temos menor chance de erro.

Definições

Tipos de variável

- ▶ Categórica ou Qualitativa
- ▶ Numérica ou Quantitativa
 - ▶ Discretas: se o conjunto de valores que pode assumir é finito ou enumerável (se podemos contá-los)
 - ▶ Contínuas: se pode assumir qualquer valor em um intervalo.

Escalas de Medida

- ▶ **Escala Nominal:** os valores são apenas um rótulo (unicamente para diferenciá-los).
 - ▶ Sexo
 - ▶ Religião
 - ▶ Raça
 - ▶ Nacionalidade

Escalas de Medida

- ▶ **Escala Nominal:** os valores são apenas um rótulo (unicamente para diferenciá-los).
 - ▶ Sexo
 - ▶ Religião
 - ▶ Raça
 - ▶ Nacionalidade
- ▶ **Escala Ordinal:** existe uma ordem, mas comparações de diferença e proporção não fazem sentido
 - ▶ Grau de instrução
 - ▶ Nível socioeconômico
 - ▶ Qualidade de um produto (escalas tipo *Liker*)

Variáveis categóricas estão em escala nominal ou ordinal

Escalas de Medida

- ▶ **Escala Intervalar:** comparações de diferença fazem sentido, mas não de proporção. O zero **não** indica ausência da característica.
 - ▶ Graus Fahrenheit
 - ▶ Pontuações em um teste psicológico

Escalas de Medida

- ▶ **Escala Intervalar:** comparações de diferença fazem sentido, mas não de proporção. O zero **não** indica ausência da característica.
 - ▶ Graus Fahrenheit
 - ▶ Pontuações em um teste psicológico
- ▶ **Escala de Proporção:** Comparações de diferença e proporção fazem sentido. O zero indica que não existe nada para a variável.
 - ▶ Peso
 - ▶ Altura
 - ▶ Distância
 - ▶ Preço

Variáveis numéricas estão em escala intervalar ou de proporção

Escalas de Medida

Conhecer o tipo de variável e sua escala de medida é importante por vários motivos:

- ▶ Nos ajudará a interpretar corretamente os dados
- ▶ Fazer análises apropriadas dependendo do tipo de variáveis e escala de medição. (exemplo: se tivermos a variável cor de olhos onde 0 significa cor preto, 1 cor marrom, 2 cor verde e 3 cor azul; não faz sentido nenhum calcular a média desses valores, ou faz?)
- ▶ Veremos na próxima aula que existem gráficos para cada tipo de variável e que o tratamento que damos aos dados muda dependendo do seu tipo/escala.

Estrutura dos Dados

Corte transversal

Os dados são coletados no mesmo intervalo de tempo. Podemos pensar neles como observar uma foto.

Estrutura dos Dados

Corte transversal

Os dados são coletados no mesmo intervalo de tempo. Podemos pensar neles como observar uma foto.

Séries temporais

Dados coletados de forma cronológica (a ordem como são coletados é importante)

Estrutura dos Dados

Corte transversal

Os dados são coletados no mesmo intervalo de tempo. Podemos pensar neles como observar uma foto.

Séries temporais

Dados coletados de forma cronológica (a ordem como são coletados é importante)

Outras estruturas

- ▶ Corte transversal agrupados
- ▶ Dados de painel (longitudinais)

Estatística Descritiva vs. Estatística Inferencial.

Estatística Descritiva vs. Estatística Inferencial.

Estatística Descritiva

- ▶ Utilizada para descrever os dados
 - ▶ Medidas de tendência central
 - ▶ Medidas de variabilidade
 - ▶ Tabelas de frequência
 - ▶ Gráficos

Estatística Descritiva vs. Estatística Inferencial.

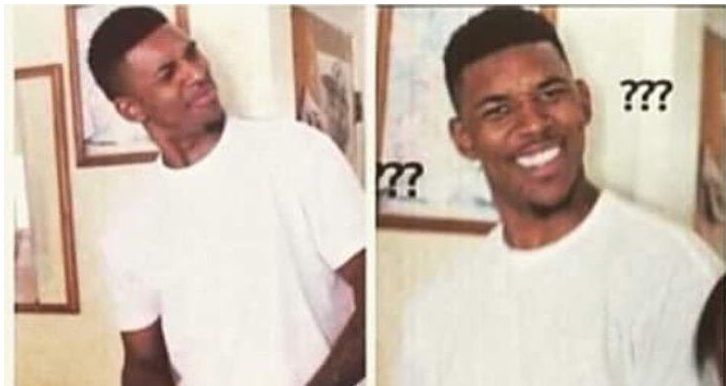
Estatística Descritiva

- ▶ Utilizada para descrever os dados
 - ▶ Medidas de tendência central
 - ▶ Medidas de variabilidade
 - ▶ Tabelas de frequência
 - ▶ Gráficos

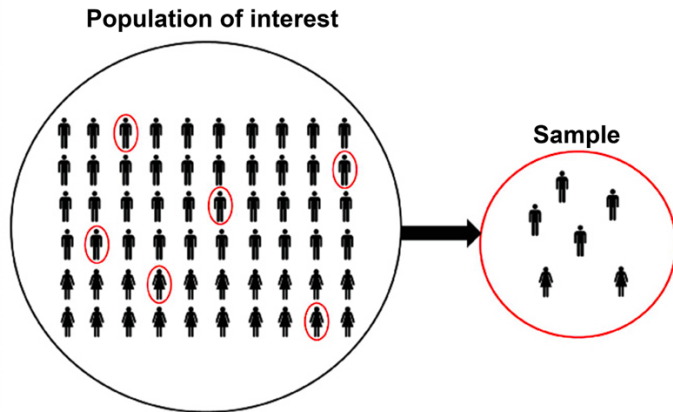
Estatística Inferencial

- ▶ Utiliza dados de uma **amostra** para inferir conclusões para a **população**
 - ▶ Estimativas pontuais
 - ▶ Intervalos de Confiança
 - ▶ Testes de Hipóteses

Estatística Descritiva vs. Estatística Inferencial.



Estatística Descritiva vs. Estatística Inferencial.



Estatística Descritiva vs. Estatística Inferencial.

- ▶ **População** é o conjunto de todos os *elementos*¹ que são objeto de estudo (elementos dos quais desejamos obter informação)
- ▶ Uma **amostra** é um subconjunto da população.
- ▶ Utilizando os dados da **amostra** vamos **inferir** determinadas características da **população**.

¹Usamos o termo **elementos**, pois uma população não é necessariamente formada por indivíduos.

Exemplo

Queremos saber a opinião dos moradores do RJ sobre as medidas adotadas para conter o avanço da COVID-19, para o estudo são selecionados 5000 moradores do RJ e se lhes pergunta quão satisfeitos/insatisfeitos estão com as medidas adotadas para conter o avanço da COVID-19 na cidade (se está totalmente insatisfeito, insatisfeito, nem satisfeito/nem insatisfeito, satisfeito, muito satisfeito). Além disso, são também registrados o sexo, idade, e bairro dos moradores selecionados.

- Identifique: população, amostra, variáveis, elementos, etc.

Estatística no século XXI



Figura 1: Softwares livres

- ▶ Tudo é feito no computador
- ▶ Excel é sempre uma opção razoável
- ▶ Existem vários *softwares* livres, nós utilizaremos *R*

Porque não usar apenas Excel?

A tendência é utilizar software livre para analisar dados.

Leituras recomendadas

- ▶ Anderson, D. R; Sweeney, D. J.; e Williams, T. A. (2008). *Estatística Aplicada à Administração e Economia*. 2ed. Cengage Learning. **Cap 1**
- ▶ Freund, J. E.; Perles, B. M. (2014). *Modern elementary statistics*. 12ed. Pearson College Division. **Chapter 1**

Filmes/Séries recomendados

