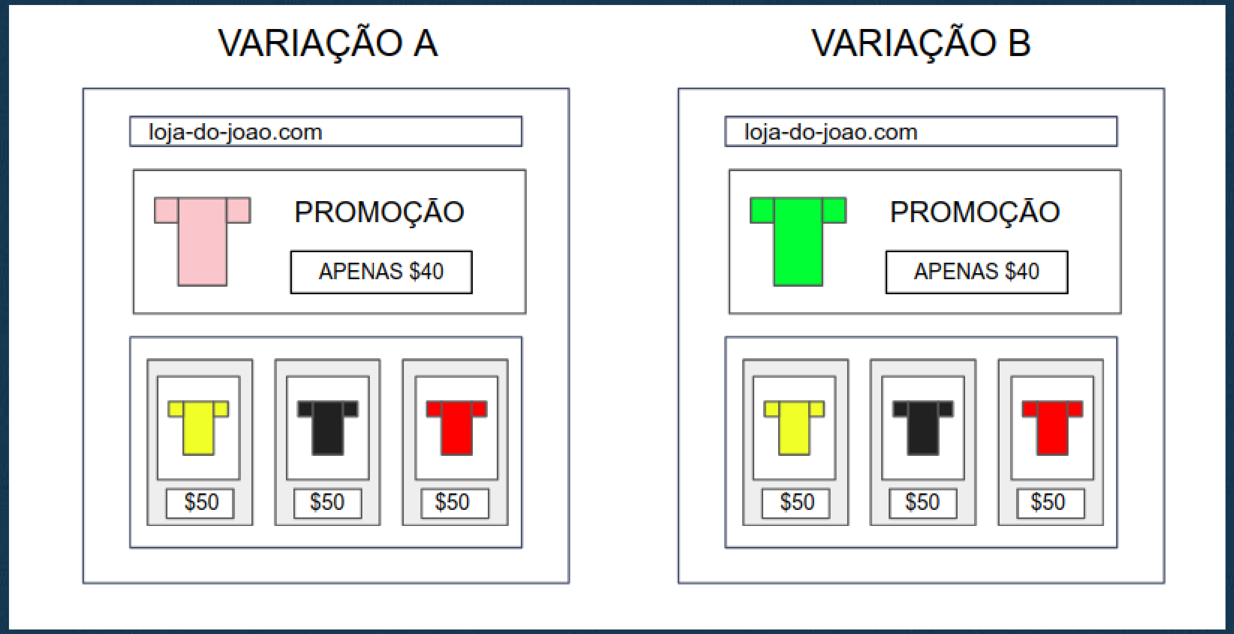
INFERÊNCIA ESTATÍSTICA

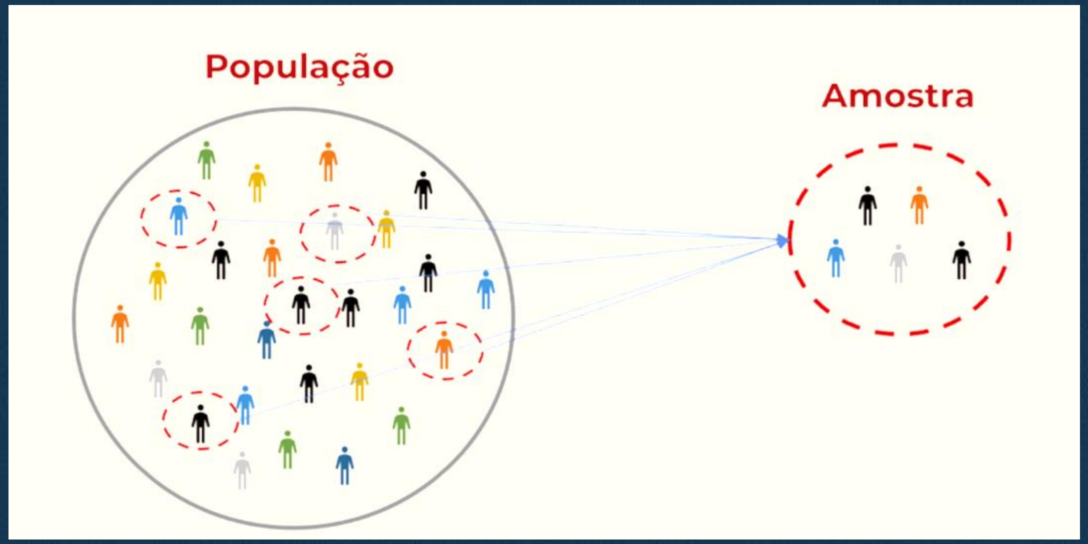
Por que é importante?

- Qualquer área que trabalhe com amostras
- Pesquisas
- Estudos acadêmicos
- Saúde, farmácia, medicina
- Teste A/B



Inferência Estatística

- População e amostra
- Parâmetro e estatística
- Distribuição amostral
- Teorema do Limite Central
- Valor esperado e erro padrão
- Probabilidade da média amostral
- Probabilidade da proporção amostral



INTRODUÇÃO À INFERÊNCIA ESTATÍSTICA

Inferência Estatística

A inferência estatística é um processo de inferir características de uma população por meio da observação de uma amostra.

A população consiste em todos os ítens de interesse (pessoas, produtos, negócios, etc).

Uma amostra é um subconjunto da população.

Por que precisamos de amostras?

Por que não coletar dados da população inteira?

- Geralmente é impossível conseguir todos esses dados
- Pode ser muito caro coletar informação da população inteira

Na inferência estatística, uma estatística amostral é usada para estimar um parâmetro da população.

Atenção: Se você consegue medir o parâmetro da população, você NÃO precisa fazer uma inferência estatística.

População e amostra

Como estimar?

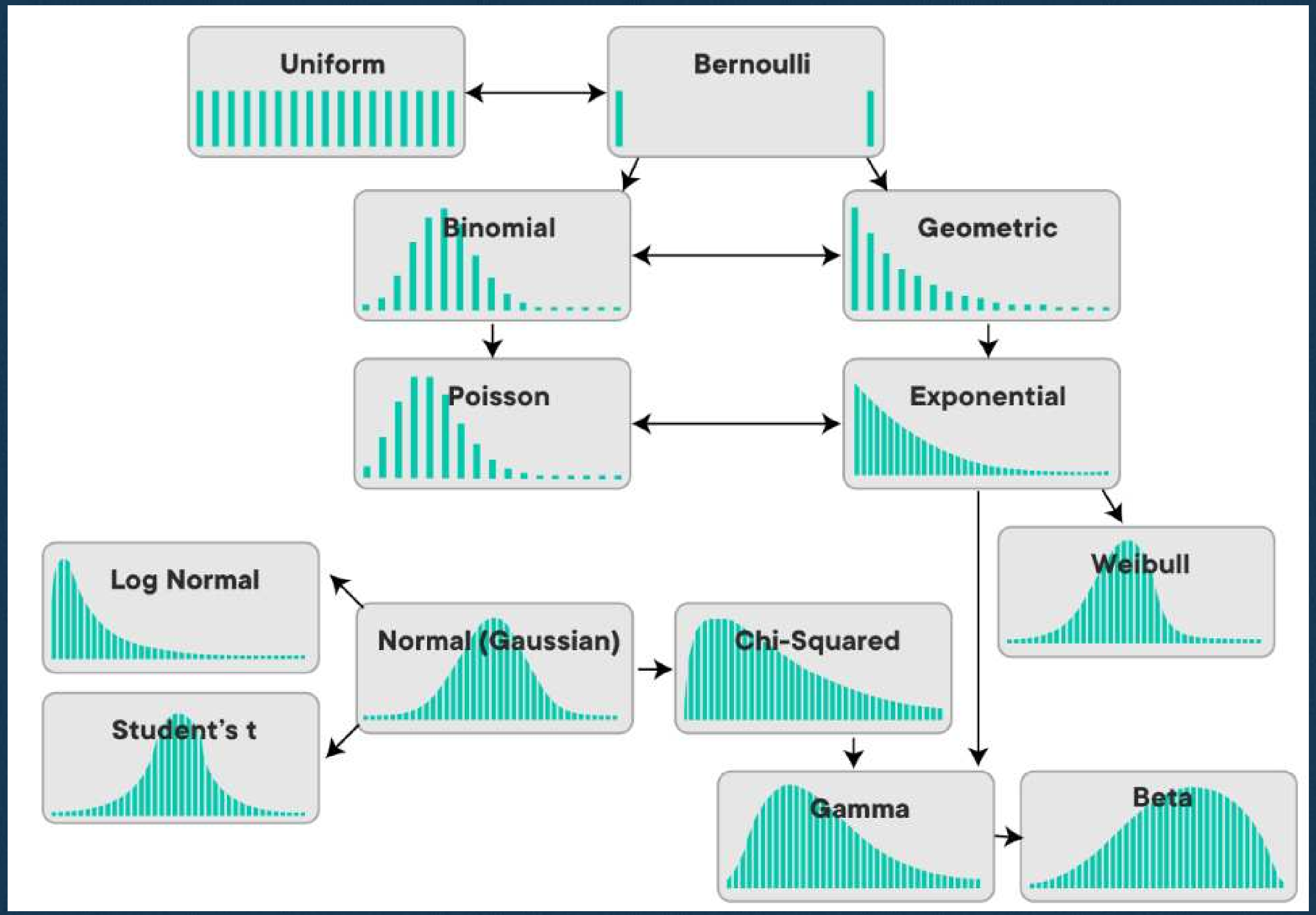
- Intervalo de Confiança
- Teste de Hipóteses

Parâmetros e estatísticas

	Parâmetros (população)	**Estatísticas (amostra)**
Média	μ	$\bar{x}$
Desvio Padrão	σ	s
Variância	σ^2	s^2
Proporção	ρ	$\bar{p}$
	Letras gregas	Letras do latim

DISTRIBUIÇÃO AMOSTRAL

Distribuições



Teoria da Distribuição Amostral

A teoria de distribuição amostral assume que as amostras são aleatórias e representativas da população.

Parâmetros e estatísticas

- Só existe uma população. O parâmetro é uma constante cujo valor pode ser desconhecido.
- Uma estatística é uma variável aleatória cujo valor depende da amostra aleatória.

Distribuição Amostral

A média da amostra é uma variável aleatória já que o valor da média varia de amostra para amostra.

Pergunta: Se a média da amostra varia de amostra para amostra, como podemos usá-la para estimar o parâmetro da população?

Resposta: Se as amostras são aleatórias, suas médias formam um padrão no longo prazo.

Exemplo

	Variável Aleatória			
	X_1	X_2	X_3	X_4
	6	10	8	4
	5	10	4	3
	1	8	4	3
	4	1	6	2
	6	6	8	4
	7	7	8	6
	1	5	10	5
	5	5	9	1
	4	6	4	2
	7	4	9	5
	8	5	8	6
	9	2	7	7
	9	1	2	3
	6	10	2	6
Médias	**5.57**	**5.71**	**6.36**	**4.07**

Uma amostra aleatória coletada da população – uma única distribuição de valores de X

Distribuição de médias de cada coleta da população – distribuição amostral

NOTAÇÕES

Distribuição Amostral da Média

Suponha que você queira estimar a média da população μ (um parâmetro).

Pegamos uma amostra aleatória da população e computamos a média amostral $\bar{x}$ (uma estatística).

Coletando muitas amostras resultam em diferentes médias amostrais, uma para cada amostra.

Notação

A variável aleatória $\overline{X}$ representa os valores possíveis de uma amostra aleatória.

A distribuição amostral da média é a distribuição de probabilidade associada à $\overline{X}$.

Notação

População: A variável aleatória X é uma característica da população. Ela tem média (valor esperado) e desvio padrão:

$$E(X) = \mu$$

$$DP(X) = \sqrt{\sigma^2} = \sigma$$

Notação

Amostras: Para amostras de tamanho n, $\overline{X}$ é a média amostral. A média e o **erro padrão** de $\overline{X}$ são:

$$E(\bar{X}) = \mu$$

$$EP(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Nota: o **erro padrão** de $\bar{X}$ é o desvio padrão da distribuição das médias amostrais.

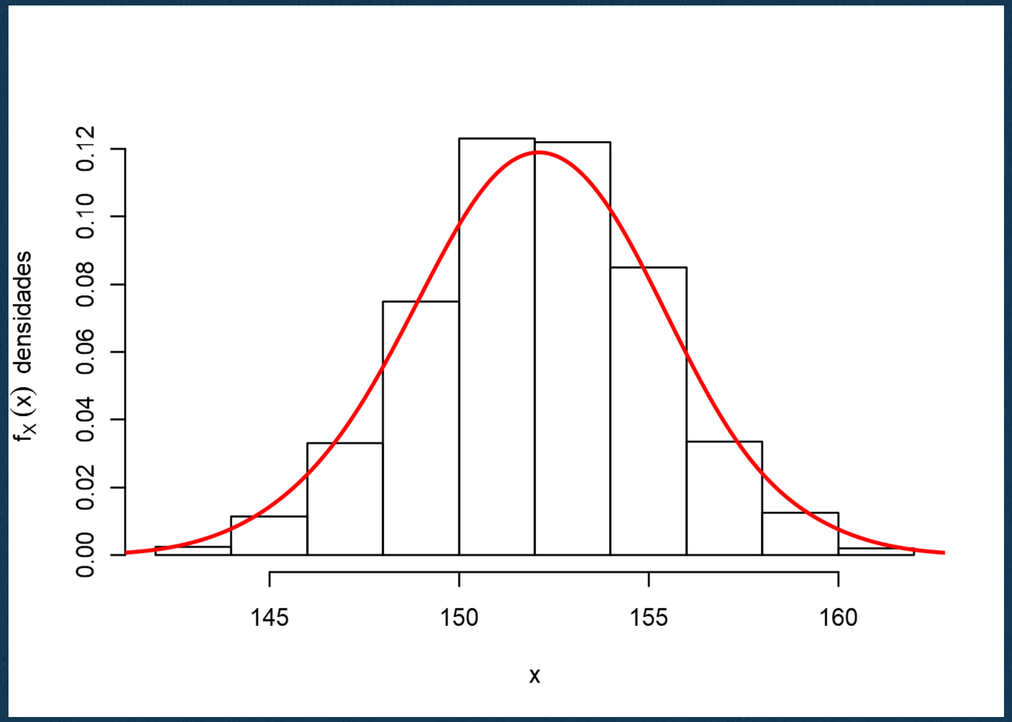
Simulador

http://onlinestatbook.com/stat_sim/sampling_dist/

TEOREMA DO LIMITE CENTRAL

Teorema do Limite Central

O TLC afirma que quando o tamanho da amostra aumenta, a distribuição amostral da sua média aproxima-se cada vez mais de uma distribuição normal.



Importante!

(1) “Qualquer distribuição vira uma normal se eu aumento o tamanho do n?”
Não, o TLC não diz que qualquer distribuição de X “vira” normal.
O que acontece é que a <u>distribuição das médias das amostras</u> se aproxima da normalidade à medida que o tamanho da amostra aumenta.

(2) “Eu preciso coletar várias amostras para usar o TLC?”
Não, você só precisa de uma amostra.
Você pode usar essa amostra para fazer inferências sobre a população.

TEOREMA DO LIMITE CENTRAL PARA A MÉDIA

Pressuposto

Para qualquer população X com valor esperado μ e desvio padrão σ, a distribuição amostral de X vai ser aproximadamente normal se o tamanho da amostra for maior ou igual a 30.

A distribuição amostral é normal se:

- As amostras foram retiradas de uma população com uma distribuição normal, ou
- Se o tamanho da amostra é $n \geq 30$ (Teorema do Limite Central).

Exemplo: QI

(a) As pontuações de QI para adultos são normalmente distribuídas com média 100 e desvio padrão 15. Se um adulto for selecionado aleatoriamente, qual é a probabilidade de seu QI ser maior que 110?

$$E(X) = 100$$
$$\mathrm{DP}(X) = \sigma = 15$$

$$P(X > 110) = 1 - P(X \leq 110) = 0{,}2525$$

Exemplo: QI

(b) Suponha que várias amostras de 10 adultos sejam coletadas e o QI médio dos 10 adultos seja calculado. Qual é a probabilidade de que a média de uma amostra seja superior a 110?

$$E(\bar{X}) = 100$$

$$EP(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{10}} = 4{,}74$$

$$P(\bar{X} > 110) = 1 - P(\bar{X} \leq 110) = 0{,}0175$$

Exemplo

Garrafas de cerveja são enchidas de modo que contenham em média 330 ml de cerveja em cada garrafa. Suponha que a quantidade de cerveja em uma garrafa seja normalmente distribuída com um desvio padrão de 4 ml.

Qual é a probabilidade de que uma garrafa selecionada aleatoriamente tenha menos de 325 ml de cerveja?

$$E(X) = 330$$
$$\mathrm{DP}(X) = \sigma = 4$$

$$P(X < 325) = 0{,}1056$$

Exemplo

(b) Qual é a probabilidade de que um pack de 6 cervejas selecionado aleatoriamente tenha uma quantidade média inferior a 325 ml?

$$E(\bar{X}) = 330$$

$$EP(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{6}} = 1{,}63$$

$$P(\bar{X} < 325) = 0{,}0011$$

Exemplo

(c) Qual é a probabilidade de que um pack de 12 cervejas selecionado aleatoriamente tenha uma quantidade média inferior a 325 ml?

$$E(\bar{X}) = 330$$

$$EP(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{12}} = 1{,}15$$

$$P(\bar{X} < 325) = 0{,}00001$$

TEOREMA DO LIMITE CENTRAL PARA A PROPORÇÃO

Proporção

Em alguns problemas na inferência, estimamos a proporção da população (percentual), geralmente para uma variável qualitativa.

O parâmetro da população é desconhecido. Estimamos esse parâmetro pegando uma amostra e computando a proporção amostral:

$$\bar{p} = \frac{x}{n}$$

Onde x é o número de sucessos e
n é o tamanho da amostra.

Exemplo

Em um grupo de 45 estudantes, 15 declaram que preferem aulas presenciais. Ao selecionar um aluno do grupo, qual a probabilidade que ele prefira aulas presenciais?

x = 15

n = 45

$\bar{p} = \frac{15}{45}$ = 0,333 = 33,3%

Proporção

A distribuição amostral para a proporção da amostra é denotada por $\bar{P}$, que representa todos os valores possíveis da proporção da amostra.

Para uma amostra de tamanho n retirada de uma população com proporção populacional ρ, o valor esperado e o erro padrão de $\bar{P}$ são:

$$E(\bar{P}) = \rho$$

$$\mathrm{EP}(\bar{P}) = \sqrt{\frac{\rho(1-\rho)}{n}}$$

Pressuposto

Para qualquer proporção da população ρ, a distribuição amostral de $\bar{P}$ é aproximadamente normal se o tamanho da amostra n for suficientemente grande.

A distribuição amostral é normal se:

- As amostras foram retiradas de uma população com uma distribuição normal, ou
- Se o tamanho da amostra é $n\rho > 5$ e $n(1 - \rho) > 5$.

Exemplo

Muitas pessoas se candidatam a empregos para servirem como paramédicos ou bombeiros, mas não conseguem cumprir os padrões básicos de aptidão física. Um estudo recente descobriu que 77% de todos os candidatos a paramédicos e bombeiros nos EUA estavam com sobrepeso ou obesos.

Exemplo

(a) Quais são o valor esperado e o erro padrão da proporção da amostra derivada de uma amostra aleatória de 100 candidatos para posições de paramédico ou bombeiro?

Como 77% de todos os candidatos a paramédicos e bombeiros tem sobrepeso ou obesidade, a proporção da população é ρ = 0,77

$$E(\bar{P}) = 0{,}77,$$

$$\mathrm{EP}(\bar{P}) = \sqrt{\frac{0{,}77(1-0{,}77)}{100}} = 0{,}042$$

Exemplo

(b) Suponha que um corpo de bombeiros tenha 100 candidatos para as posições de paramédico e bombeiro. Qual é a probabilidade de que menos de 70 estejam com sobrepeso ou obesidade?

A proporção da amostra é $\bar{p} = \frac{70}{100} = 0{,}7$

$$P(\bar{P} < 0{,}7) = 0{,}048$$