

ME414 - Estatística para Experimentalistas

Amostragem e Distribuições Amostrais

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

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

Aula 11

Amostragem

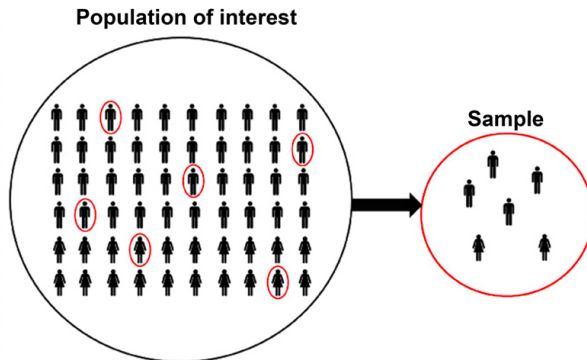
Estimação pontual

Distribuições amostrais

Métodos de amostragem

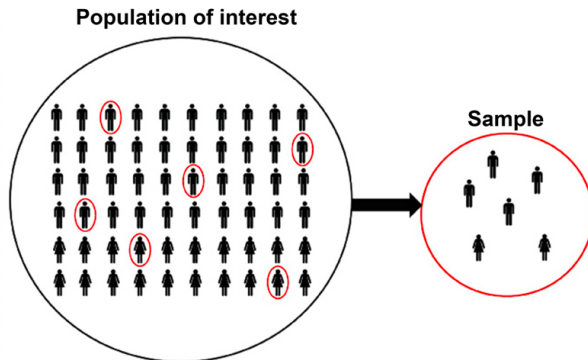
Amostragem

Amostragem



- **População:** conjunto de todos os elementos de interesse no estudo.

Amostragem



- ▶ **População:** conjunto de todos os elementos de interesse no estudo.
- ▶ **Amostra:** subconjunto da população

Amostragem

Características numéricas da população (média, proporção, desvio padrão, etc) são chamados de **parâmetros**

Amostragem

Características numéricas da população (média, proporção, desvio padrão, etc) são chamados de **parâmetros**

Um dos propósitos da inferência estatística é estimar e testar hipóteses a respeito dos parâmetros populacionais utilizando a informação contida na amostra

Amostragem

Características numéricas da população (média, proporção, desvio padrão, etc) são chamados de **parâmetros**

Um dos propósitos da inferência estatística é estimar e testar hipóteses a respeito dos parâmetros populacionais utilizando a informação contida na amostra

Por que utilizar uma amostra?

Amostragem

Características numéricas da população (média, proporção, desvio padrão, etc) são chamados de **parâmetros**

Um dos propósitos da inferência estatística é estimar e testar hipóteses a respeito dos parâmetros populacionais utilizando a informação contida na amostra

Por que utilizar uma amostra?

- ▶ Coletar informação da população toda é caro

Amostragem

Características numéricas da população (média, proporção, desvio padrão, etc) são chamados de **parâmetros**

Um dos propósitos da inferência estatística é estimar e testar hipóteses a respeito dos parâmetros populacionais utilizando a informação contida na amostra

Por que utilizar uma amostra?

- ▶ Coletar informação da população toda é caro
- ▶ Coletar informação da população toda é inviável

Amostragem

Características numéricas da população (média, proporção, desvio padrão, etc) são chamados de **parâmetros**

Um dos propósitos da inferência estatística é estimar e testar hipóteses a respeito dos parâmetros populacionais utilizando a informação contida na amostra

Por que utilizar uma amostra?

- ▶ Coletar informação da população toda é caro
- ▶ Coletar informação da população toda é inviável

Amostragem

Características numéricas da população (média, proporção, desvio padrão, etc) são chamados de **parâmetros**

Um dos propósitos da inferência estatística é estimar e testar hipóteses a respeito dos parâmetros populacionais utilizando a informação contida na amostra

Por que utilizar uma amostra?

- ▶ Coletar informação da população toda é caro
- ▶ Coletar informação da população toda é inviável

Os resultados da amostra apresentam apenas uma estimativa do verdadeiro (e desconhecido) valor do parâmetro. **Com métodos apropriados de amostragem, obteremos "boas" estimativas**

Amostragem aleatória simples

Amostragem aleatória simples é um dos métodos de amostragem mais comuns

¹Existem amostras aleatórias com e sem reposição

Amostragem aleatória simples

Amostragem aleatória simples é um dos métodos de amostragem mais comuns

Definição

Uma amostra aleatória simples (a.a) de tamanho n de uma população de tamanho N é uma amostra selecionada de forma independente e sem reposição¹ de tal maneira que cada elemento na amostra tenha a mesma probabilidade de ser escolhido.

¹Existem amostras aleatórias com e sem reposição

Amostragem aleatória simples

Amostragem aleatória simples é um dos métodos de amostragem mais comuns

Definição

Uma amostra aleatória simples (a.a) de tamanho n de uma população de tamanho N é uma amostra selecionada de forma independente e sem reposição¹ de tal maneira que cada elemento na amostra tenha a mesma probabilidade de ser escolhido.

- ▶ Antigamente se utilizava uma Tabela de número aleatórios

¹Existem amostras aleatórias com e sem reposição

Amostragem aleatória simples

Amostragem aleatória simples é um dos métodos de amostragem mais comuns

Definição

Uma amostra aleatória simples (a.a) de tamanho n de uma população de tamanho N é uma amostra selecionada de forma independente e sem reposição¹ de tal maneira que cada elemento na amostra tenha a mesma probabilidade de ser escolhido.

- ▶ Antigamente se utilizava uma Tabela de número aleatórios
- ▶ Hoje em dia qualquer programa calcula números (pseudo) aleatorios

¹Existem amostras aleatórias com e sem reposição

Amostragem aleatória simples

- ▶ No *R*:

```
sample(x, n)
```

em que x é um vetor com o ID dos elementos da população e n é o tamanho da amostra.

Amostragem aleatória simples

- ▶ No *R*:

```
sample(x, n)
```

em que x é um vetor com o ID dos elementos da população e n é o tamanho da amostra.

Amostragem aleatória simples

- No R:

```
sample(x, n)
```

em que x é um vetor com o ID dos elementos da população e n é o tamanho da amostra.

Exemplo

```
IDs <- 1:1000  
n <- 10  
sample(IDs,n)
```

```
## [1] 249 471 845 237 386 592 89 502 245 892
```

Amostragem aleatória simples

Quantas amostras aleatórias simples de tamanho n podem ser selecionadas de uma população com N elementos?

Amostragem aleatória simples

Quantas amostras aleatórias simples de tamanho n podem ser selecionadas de uma população com N elementos?

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Amostragem aleatória simples

A diretoria do CCJE gostaria conhecer a opinião dos alunos da FACC sobre o ensino remoto no 2020.2. Por motivos logísticos e financeiros, saber a opinião de todos os alunos é inviável. Por esse motivo, a diretoria do CCJE pede ajuda aos alunos de MAD211 para selecionar uma amostra de 400 alunos. Como seria feita a amostragem?

Amostragem aleatória simples

A diretoria do CCJE gostaria conhecer a opinião dos alunos da FACC sobre o ensino remoto no 2020.2. Por motivos logísticos e financeiros, saber a opinião de todos os alunos é inviável. Por esse motivo, a diretoria do CCJE pede ajuda aos alunos de MAD211 para selecionar uma amostra de 400 alunos. Como seria feita a amostragem?

- ▶ Usar os ID dos alunos (por exemplo, o número de matrícula)

Amostragem aleatória simples

A diretoria do CCJE gostaria conhecer a opinião dos alunos da FACC sobre o ensino remoto no 2020.2. Por motivos logísticos e financeiros, saber a opinião de todos os alunos é inviável. Por esse motivo, a diretoria do CCJE pede ajuda aos alunos de MAD211 para selecionar uma amostra de 400 alunos. Como seria feita a amostragem?

- ▶ Usar os ID dos alunos (por exemplo, o número de matrícula)
- ▶ Usar `sample(matriculas_alunos, n = 400)`

Amostragem aleatória simples

A diretoria do CCJE gostaria conhecer a opinião dos alunos da FACC sobre o ensino remoto no 2020.2. Por motivos logísticos e financeiros, saber a opinião de todos os alunos é inviável. Por esse motivo, a diretoria do CCJE pede ajuda aos alunos de MAD211 para selecionar uma amostra de 400 alunos. Como seria feita a amostragem?

- ▶ Usar os ID dos alunos (por exemplo, o número de matrícula)
- ▶ Usar `sample(matriculas_alunos, n = 400)`
- ▶ Entrevistar os 400 alunos selecionados.

Estimação pontual

Estimação pontual

- ▶ Frequentemente estamos interessados em conhecer o valor de algumas características (parâmetros) da população

Estimação pontual

- ▶ Frequentemente estamos interessados em conhecer o valor de algumas características (parâmetros) da população
- ▶ Como dificilmente temos informação da população toda, o que é feito é calcular a característica correspondente à amostra, denominada **estatística amostral**.

Estimación pontual

- ▶ Frequentemente estamos interesados em conhecer o valor de algumas características (parâmetros) da população
- ▶ Como dificilmente temos informação da população toda, o que é feito é calcular a característica correspondente à amostra, denominada **estatística amostral**.

Estimação pontual

- ▶ Frequentemente estamos interessados em conhecer o valor de algumas características (parâmetros) da população
- ▶ Como dificilmente temos informação da população toda, o que é feito é calcular a característica correspondente à amostra, denominada **estatística amostral**.

Estatística

Uma estatística é uma característica da amostra, ou seja, uma estatística T é uma função de X_1, X_2, \dots, X_n

Estimação pontual

- ▶ Frequentemente estamos interessados em conhecer o valor de algumas características (parâmetros) da população
- ▶ Como dificilmente temos informação da população toda, o que é feito é calcular a característica correspondente à amostra, denominada **estatística amostral**.

Estatística

Uma estatística é uma característica da amostra, ou seja, uma estatística T é uma função de X_1, X_2, \dots, X_n

Parâmetro

Um parâmetro é uma medida usada para descrever uma característica da população

Estimação pontual

Na prática, estimaremos a característica de interesse (parâmetro) utilizando dados da nossa amostra.

Exemplos

Estimação pontual

Na prática, estimaremos a característica de interesse (parâmetro) utilizando dados da nossa amostra.

Exemplos

- ▶ Saber a opinião dos alunos da FACC sobre o ensino remoto no 2020.2

Estimação pontual

Na prática, estimaremos a característica de interesse (parâmetro) utilizando dados da nossa amostra.

Exemplos

- ▶ Saber a opinião dos alunos da FACC sobre o ensino remoto no 2020.2
- ▶ Saber o salário de todos os professores universitários de instituições publicas no Brasil

Estimação pontual

Na prática, estimaremos a característica de interesse (parâmetro) utilizando dados da nossa amostra.

Exemplos

- ▶ Saber a opinião dos alunos da FACC sobre o ensino remoto no 2020.2
- ▶ Saber o salário de todos os professores universitários de instituições publicas no Brasil
- ▶ Saber a intenção de voto de todos os brasileiros (com idade para votar)

Estimación pontual

Exemplo

Suponha que os salários dos 10000 professores das universidades publicas está no conjunto de dados `salarios`

Estimación pontual

Exemplo

Suponha que os salários dos 10000 professores das universidades publicas está no conjunto de dados `salarios`

Cuja média é 7500.36 e variância é 1493.59

Estimación pontual

Exemplo

Suponha que os salários dos 10000 professores das universidades publicas está no conjunto de dados `salarios`

Cuja média é 7500.36 e variância é 1493.59

Vamos supor que o sistema tem organizado os professores segundo salário (de menor a maior)

Estimación pontual

Exemplo

Suponha que os salários dos 10000 professores das universidades publicas está no conjunto de dados `salarios`

Cuja média é 7500.36 e variância é 1493.59

Vamos supor que o sistema tem organizado os professores segundo salário (de menor a maior)

```
salarios <- sort(salarios)
```

Estimación pontual

Exemplo

Suponha que os salários dos 10000 professores das universidades publicas está no conjunto de dados `salarios`

Cuja média é 7500.36 e variância é 1493.59

Vamos supor que o sistema tem organizado os professores segundo salário (de menor a maior)

```
salarios <- sort(salarios)
```

vamos a seleccionar uma amostra de tamanho 100 e calcular o salário médio e a variância

Estimação pontual

```
amostra1 <- head(salarios,100) #100 primeiros registros na base  
amostra2 <- tail(salarios,100) #100 últimos registros na base  
amostra3 <- sample(salarios,100) # a.a de tamanho 100
```

Estimação pontual

```
amostra1 <- head(salarios,100) #100 primeiros registros na base  
amostra2 <- tail(salarios,100) #100 últimos registros na base  
amostra3 <- sample(salarios,100) # a.a de tamanho 100
```

Qual amostra você acha nos dará “boas” estimativas? por quê?

Estimação pontual

```
amostra1 <- head(salarios,100) #100 primeiros registros na base  
amostra2 <- tail(salarios,100) #100 últimos registros na base  
amostra3 <- sample(salarios,100) # a.a de tamanho 100
```

Qual amostra você acha nos dará “boas” estimativas? por quê?

$\mu = 7500.36$ e $\sigma^2 = 1493.59$

```
c(mean(amostra1), mean(amostra2), mean(amostra3))
```

```
## [1] 7396.064 7603.682 7499.218
```

```
c(var(amostra1), var(amostra2), var(amostra3))
```

```
## [1] 123.0519 141.2240 1376.7597
```

Estimação pontual

E se quisermos saber a proporção de professores que ganham mais de 7500 reais?

```
sum(salarios > 7500)/length(salarios) #prop. populacional
```

```
## [1] 0.4991
```

Estimación puntual

E se quisermos saber a proporção de professores que ganham mais de 7500 reais?

```
sum(salarios > 7500)/length(salarios) #prop. populacional
```

```
## [1] 0.4991
```

```
sum(amostra1 > 7500)/length(amostra1)
```

```
## [1] 0
```

```
sum(amostra2 > 7500)/length(amostra2)
```

```
## [1] 1
```

```
sum(amostra3 > 7500)/length(amostra3)
```

```
## [1] 0.46
```

Estimação pontual

- ▶ A média amostral é um estimador da média populacional μ

Estimação pontual

- ▶ A média amostral é um estimador da média populacional μ
- ▶ A variância amostral é um estimador da variância populacional σ^2

Estimação pontual

- ▶ A média amostral é um estimador da média populacional μ
- ▶ A variância amostral é um estimador da variância populacional σ^2
- ▶ A proporção da amostra é um estimador da proporção populacional p

Distribuições amostrais

Distribuições amostrais

- ▶ No exemplo anterior, temos que $\mu = 7500.36$ e (para a amostra3) $\bar{x} = 7499.22$ mas. . . .

Distribuições amostrais

- ▶ No exemplo anterior, temos que $\mu = 7500.36$ e (para a amostra3) $\bar{x} = 7499.22$ mas. . . .
- ▶ E se tivéssemos escolhido outra amostra?

Distribuições amostrais

- ▶ No exemplo anterior, temos que $\mu = 7500.36$ e (para a amostra3) $\bar{x} = 7499.22$ mas. . . .
- ▶ E se tivéssemos escolhido outra amostra?

Distribuições amostrais

- ▶ No exemplo anterior, temos que $\mu = 7500.36$ e (para a amostra3) $\bar{x} = 7499.22$ mas....
- ▶ E se tivéssemos escolhido outra amostra?

```
c(mean(sample(salarios,100)), mean(sample(salarios,100)),  
  mean(sample(salarios,100)), mean(sample(salarios,100)))
```

```
## [1] 7511.493 7507.844 7503.407 7497.451
```

Distribuições amostrais

- ▶ No exemplo anterior, temos que $\mu = 7500.36$ e (para a amostra3) $\bar{x} = 7499.22$ mas....
- ▶ E se tivéssemos escolhido outra amostra?

```
c(mean(sample(salarios,100)), mean(sample(salarios,100)),  
  mean(sample(salarios,100)), mean(sample(salarios,100)))
```

```
## [1] 7511.493 7507.844 7503.407 7497.451
```

- ▶ Cada amostra levará a valores diferentes de \bar{x}

Distribuições amostrais

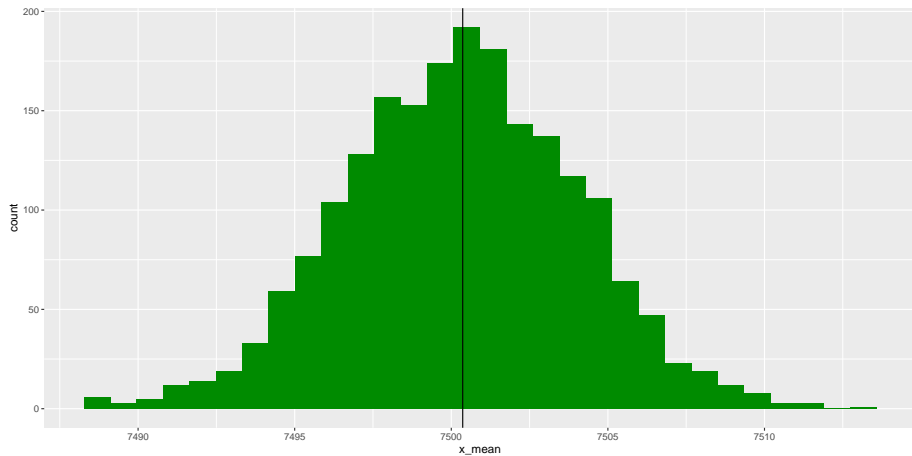
- ▶ No exemplo anterior, temos que $\mu = 7500.36$ e (para a amostra3) $\bar{x} = 7499.22$ mas....
- ▶ E se tivéssemos escolhido outra amostra?

```
c(mean(sample(salarios,100)), mean(sample(salarios,100)),  
  mean(sample(salarios,100)), mean(sample(salarios,100)))
```

```
## [1] 7511.493 7507.844 7503.407 7497.451
```

- ▶ Cada amostra levará a valores diferentes de \bar{x}
- ▶ Se considerarmos, digamos 2000 amostras, qual é a distribuição amostral de \bar{x} ?

Distribuições amostrais



Distribuições amostrais

- A estimador $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ é uma variável aleatória, e como variável aleatória tem um valor médio (valor esperado), um desvio padrão e uma distribuição de probabilidade.

Distribuições amostrais

$$\sum_{i=1}^n X_i$$

- ▶ A estimador $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ é uma variável aleatória, e como variável aleatória tem um valor médio (valor esperado), um desvio padrão e uma distribuição de probabilidade.
- ▶ A distribuição de probabilidade da média amostral é chamada de **distribuição amostral** da média.

Distribuições amostrais

$$\sum_{i=1}^n X_i$$

- ▶ A estimador $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ é uma variável aleatória, e como variável aleatória tem um valor médio (valor esperado), um desvio padrão e uma distribuição de probabilidade.
- ▶ A distribuição de probabilidade da média amostral é chamada de **distribuição amostral** da média.
- ▶ Conhecer a distribuição amostral de alguma **estatística** (média, proporção, variância, etc), bem como suas propriedades, nos possibilitará fazer afirmações a respeito de quão próximas de, por exemplo, quão próximas a média da amostra está da média populacional (ou a variância amostral da variância populacional ou a proporção amostral da proporção populacional, etc)

Distribuição da média amostral

Sejam X_1, \dots, X_n uma amostra aleatória de uma distribuição com média μ e variância σ^2

Distribuição da média amostral

Sejam X_1, \dots, X_n uma amostra aleatória de uma distribuição com média μ e variância σ^2

$$\blacktriangleright \mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n} \underbrace{\mathbb{E}(X_1 + \dots + X_n)}_{\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)} = \frac{1}{n} n\mu = \mu$$

Distribuição da média amostral

Sejam X_1, \dots, X_n uma amostra aleatória de uma distribuição com média μ e variância σ^2

- ▶ $\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n} \underbrace{\mathbb{E}(X_1 + \dots + X_n)}_{\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)} = \frac{1}{n} n\mu = \mu$
- ▶ Quando o valor esperado de um estimador pontual é igual ao parâmetro, dizemos que o estimador é não viesado.

Distribuição da média amostral

Sejam X_1, \dots, X_n uma amostra aleatória de uma distribuição com média μ e variância σ^2

- ▶ $\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n} \underbrace{\mathbb{E}(X_1 + \dots + X_n)}_{\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)} = \frac{1}{n} n\mu = \mu$
- ▶ Quando o valor esperado de um estimador pontual é igual ao parâmetro, dizemos que o estimador é não viesado.
- ▶ A média amostral é um estimador não viesado da média populacional.

Distribuição da média amostral

Sejam X_1, \dots, X_n uma amostra aleatória de uma distribuição com média μ e variância σ^2

- ▶ $\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n} \underbrace{\mathbb{E}(X_1 + \dots + X_n)}_{\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)} = \frac{1}{n} n\mu = \mu$
- ▶ Quando o valor esperado de um estimador pontual é igual ao parâmetro, dizemos que o estimador é não viesado.
- ▶ A média amostral é um estimador não viesado da média populacional.
- ▶ $\mathbb{V}(\bar{X}) = \mathbb{V}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2} \underbrace{\mathbb{V}(X_1 + \dots + X_n)}_{\mathbb{V}(X_1) + \dots + \mathbb{V}(X_n)} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

Distribuição da média amostral

- ▶ No exemplo dos salários, vimos que a distribuição da média amostral tinha a forma de uma distribuição normal.

Distribuição da média amostral

- ▶ No exemplo dos salários, vimos que a distribuição da média amostral tinha a forma de uma distribuição normal.
- ▶ Será que isso é uma coincidência?

Distribuição da média amostral

- ▶ No exemplo dos salários, vimos que a distribuição da média amostral tinha a forma de uma distribuição normal.
- ▶ Será que isso é uma coincidência?

Distribuição da média amostral

- ▶ No exemplo dos salários, vimos que a distribuição da média amostral tinha a forma de uma distribuição normal.
- ▶ Será que isso é uma coincidência?

Resultado

Se $X_1, X_2, \dots, X_n \sim N(\mu, \sigma)$, então

$$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$$

Distribuição da média amostral

Teorema Central do Limite

Sejam X_1, X_2, \dots, X_n (para n grande) v.as **independentes** e **identicamente distribuídas** com $E(X_1) = \mu$ e $V(X_1) = \sigma^2 < \infty$. Então,

$$\frac{(\bar{X}_n - \mu)}{\sigma/\sqrt{n}} \sim_{aprox} N(0, 1) \quad \equiv \quad \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \sim_{aprox} N(0, 1)$$

Distribuição da média amostral

Teorema Central do Limite

Sejam X_1, X_2, \dots, X_n (para n grande) v.as **independentes e identicamente distribuídas** com $E(X_1) = \mu$ e $V(X_1) = \sigma^2 < \infty$. Então,

$$\frac{(\bar{X}_n - \mu)}{\sigma/\sqrt{n}} \sim_{\text{aprox}} N(0, 1) \quad \equiv \quad \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \sim_{\text{aprox}} N(0, 1)$$

Vamos ver o TCL (Teorema Central do Limite) através de simulações?

Distribuição da média amostral

Por que precisamos da distribuição da média amostral?

- ▶ Quando calculamos \bar{x} não podemos esperar que esse valor seja exatamente igual a μ

Distribuição da média amostral

Por que precisamos da distribuição da média amostral?

- ▶ Quando calculamos \bar{x} não podemos esperar que esse valor seja exatamente igual a μ
- ▶ Mas conhecer a distribuição da média amostral nos fornece informações probabilísticas da diferença entre a média amostral e a média da população.

Distribuição da média amostral

1. Seja X o número ligações semanais de telemarketing recebidas por clientes de uma operadores de celular (X tem média de 7 e desvio padrão de 3). Os clientes, cansados de receberem tantas ligações, resolvem processar a empresa. A empresa nega este fato mas aceita escolher uma amostra de 100 pessoas e se, em média, o número de ligações semanais for maior do que 9, eles darão aos clientes uma indenização pelos transtornos. Qual é a probabilidade de, em média, os 100 usuários receberem mais do que 9 ligações semanais?

Distribuição da média amostral

1. Seja X o número ligações semanais de telemarketing recebidas por clientes de uma operadores de celular (X tem média de 7 e desvio padrão de 3). Os clientes, cansados de receberem tantas ligações, resolvem processar a empresa. A empresa nega este fato mas aceita escolher uma amostra de 100 pessoas e se, em média, o número de ligações semanais for maior do que 9, eles darão aos clientes uma indenização pelos transtornos. Qual é a probabilidade de, em média, os 100 usuários receberem mais do que 9 ligações semanais?
 - ▶ X tem $\mu = 7$ e $\sigma = 3$

Distribuição da média amostral

1. Seja X o número ligações semanais de telemarketing recebidas por clientes de uma operadores de celular (X tem média de 7 e desvio padrão de 3). Os clientes, cansados de receberem tantas ligações, resolvem processar a empresa. A empresa nega este fato mas aceita escolher uma amostra de 100 pessoas e se, em média, o número de ligações semanais for maior do que 9, eles darão aos clientes uma indenização pelos transtornos. Qual é a probabilidade de, em média, os 100 usuários receberem mais do que 9 ligações semanais?

- ▶ X tem $\mu = 7$ e $\sigma = 3$
- ▶ Pelo TCL

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{100}} \approx N(0, 1)$$

Distribuição da média amostral

1. Seja X o número ligações semanais de telemarketing recebidas por clientes de uma operadores de celular (X tem média de 7 e desvio padrão de 3). Os clientes, cansados de receberem tantas ligações, resolvem processar a empresa. A empresa nega este fato mas aceita escolher uma amostra de 100 pessoas e se, em média, o número de ligações semanais for maior do que 9, eles darão aos clientes uma indenização pelos transtornos. Qual é a probabilidade de, em média, os 100 usuários receberem mais do que 9 ligações semanais?

- ▶ X tem $\mu = 7$ e $\sigma = 3$
- ▶ Pelo TCL

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{100}} \approx N(0, 1)$$

- ▶ Queremos: $P(\bar{X} > 9)$

Distribuição da média amostral

temos que: X tem $\mu = 7$ e $\sigma = 3$ e $\frac{\bar{X} - \mu}{\sigma/\sqrt{100}} \sim N(0, 1)$

Distribuição da média amostral

temos que: X tem $\mu = 7$ e $\sigma = 3$ e $\frac{\bar{X} - \mu}{\sigma/\sqrt{100}} \sim N(0, 1)$

$$P(\bar{X} > 9) = P(\bar{X} - 7 > 9 - 7) = P\left(\underbrace{\frac{\bar{X} - 7}{3/\sqrt{100}}}_{\sim N(0,1)} > \underbrace{\frac{9 - 7}{3/\sqrt{100}}}_{6.666667}\right)$$

Distribuição da média amostral

temos que: X tem $\mu = 7$ e $\sigma = 3$ e $\frac{\bar{X} - \mu}{\sigma/\sqrt{100}} \sim N(0, 1)$

$$P(\bar{X} > 9) = P(\bar{X} - 7 > 9 - 7) = P\left(\underbrace{\frac{\bar{X} - 7}{3/\sqrt{100}}}_{\sim N(0,1)} > \underbrace{\frac{9 - 7}{3/\sqrt{100}}}_{6.666667}\right)$$

$$P(Z > 6.666667) = 1 - P(Z \leq 6.666667)$$

Distribuição da média amostral

temos que: X tem $\mu = 7$ e $\sigma = 3$ e $\frac{\bar{X} - \mu}{\sigma/\sqrt{100}} \sim N(0, 1)$

$$P(\bar{X} > 9) = P(\bar{X} - 7 > 9 - 7) = P\left(\underbrace{\frac{\bar{X} - 7}{3/\sqrt{100}}}_{\sim N(0,1)} > \underbrace{\frac{9 - 7}{3/\sqrt{100}}}_{6.666667}\right)$$

$$P(Z > 6.666667) = 1 - P(Z \leq 6.666667)$$

```
1-pnorm(6.666667)
```

```
## [1] 1.308387e-11
```


Distribuição da média amostral

2. Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?

Distribuição da média amostral

2. Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?
 - ▶ X_i : lado superior da moeda no i -ésimo lançamento (1: cara, 0:coroa)

Distribuição da média amostral

2. Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?
- ▶ X_i : lado superior da moeda no i -ésimo lançamento (1: cara, 0:coroa)
 - ▶ $X_i \sim \text{bernoulli}(p = 0.5)$

Distribuição da média amostral

2. Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?
- ▶ X_i : lado superior da moeda no i -ésimo lançamento (1: cara, 0:coroa)
 - ▶ $X_i \sim \text{bernoulli}(p = 0.5)$
 - ▶ $E(X_1) = p = 0.5$

Distribuição da média amostral

2. Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?
- ▶ X_i : lado superior da moeda no i -ésimo lançamento (1: cara, 0:coroa)
 - ▶ $X_i \sim \text{bernoulli}(p = 0.5)$
 - ▶ $E(X_1) = p = 0.5$
 - ▶ $V(X_1) = pq = 0.25$

Distribuição da média amostral

2. Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?

- ▶ X_i : lado superior da moeda no i -ésimo lançamento (1: cara, 0:coroa)
- ▶ $X_i \sim \text{bernoulli}(p = 0.5)$
- ▶ $E(X_1) = p = 0.5$
- ▶ $V(X_1) = pq = 0.25$
- ▶ TCL: $\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \sim_{\text{aprox}} N(0, 1)$

Distribuição da média amostral

2. Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?

- ▶ X_i : lado superior da moeda no i -ésimo lançamento (1: cara, 0:coroa)
- ▶ $X_i \sim \text{bernoulli}(p = 0.5)$
- ▶ $E(X_1) = p = 0.5$
- ▶ $V(X_1) = pq = 0.25$
- ▶ TCL: $\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \sim_{\text{aprox}} N(0, 1)$
- ▶ $P\left(\sum_{i=1}^n X_i > 495\right) = P\left(\underbrace{\frac{\sum_{i=1}^n X_i - 900 \times 0.5}{\sqrt{0.25} \times \sqrt{900}}}_Z > \underbrace{\frac{495 - 900 \times 0.5}{\sqrt{0.25} \times \sqrt{900}}}_3\right)$

Distribuição da média amostral

```
1-pnorm(3)
```

```
## [1] 0.001349898
```


Distribuição da média amostral

```
1-pnorm(3)
```

```
## [1] 0.001349898
```

Mas poderíamos também ter obtido a resposta utilizando diretamente a distribuição binomial ($Y \sim \text{Binom}(n = 900, p = 0.5)$)

Distribuição da média amostral

```
1-pnorm(3)
```

```
## [1] 0.001349898
```

Mas poderíamos também ter obtido a resposta utilizando diretamente a distribuição binomial ($Y \sim \text{Binom}(n = 900, p = 0.5)$)

$$P(Y > 495) = 1 - P(Y \leq 495)$$

Distribuição da média amostral

```
1-pnorm(3)
```

```
## [1] 0.001349898
```

Mas poderíamos também ter obtido a resposta utilizando diretamente a distribuição binomial ($Y \sim \text{Binom}(n = 900, p = 0.5)$)

$$P(Y > 495) = 1 - P(Y \leq 495)$$

```
1-pbinom(495, 900, prob = 0.5)
```

```
## [1] 0.001200108
```

Distribuição da média amostral

```
1-pnorm(3)
```

```
## [1] 0.001349898
```

Mas poderíamos também ter obtido a resposta utilizando diretamente a distribuição binomial ($Y \sim \text{Binom}(n = 900, p = 0.5)$)

$$P(Y > 495) = 1 - P(Y \leq 495)$$

```
1-pbinom(495, 900, prob = 0.5)
```

```
## [1] 0.001200108
```

A diferença nas respostas é devido a que o TCL é uma aproximação

Distribuição da média amostral

3. Uma amostra aleatório de tamanho $n = 120$ é extraída de uma distribuição uniforme no intervalo $[0,1]$ ($U_{[0,1]}$). Qual é a probabilidade de que $P(|\bar{X} - 1/2| \leq 0.1)$?

Distribuição da média amostral

3. Uma amostra aleatório de tamanho $n = 120$ é extraída de uma distribuição uniforme no intervalo $[0,1]$ ($U_{[0,1]}$). Qual é a probabilidade de que $P(|\bar{X} - 1/2| \leq 0.1)$?
- A média da distribuição $U_{[a,b]}$ é $\frac{a+b}{2}$, no nosso caso: $\frac{0+1}{2} = 1/2$

Distribuição da média amostral

3. Uma amostra aleatório de tamanho $n = 120$ é extraída de uma distribuição uniforme no intervalo $[0,1]$ ($U_{[0,1]}$). Qual é a probabilidade de que $P(|\bar{X} - 1/2| \leq 0.1)$?
- ▶ A média da distribuição $U_{[a,b]}$ é $\frac{a+b}{2}$, no nosso caso: $\frac{0+1}{2} = 1/2$
 - ▶ A variância da distribuição $U_{[a,b]}$ é $\frac{(b-a)^2}{12}$, no nosso caso: $1/12$

Distribuição da média amostral

3. Uma amostra aleatório de tamanho $n = 120$ é extraída de uma distribuição uniforme no intervalo $[0,1]$ ($U_{[0,1]}$). Qual é a probabilidade de que $P(|\bar{X} - 1/2| \leq 0.1)$?

- ▶ A média da distribuição $U_{[a,b]}$ é $\frac{a+b}{2}$, no nosso caso: $\frac{0+1}{2} = 1/2$
- ▶ A variância da distribuição $U_{[a,b]}$ é $\frac{(b-a)^2}{12}$, no nosso caso: $1/12$
- ▶ Aplicando o TCL:

$$\frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} = \frac{\sqrt{120}(\bar{X} - 1/2)}{1/\sqrt{12}} \sim_{approx} N(0, 1)$$

Distribuição da média amostral

3. Uma amostra aleatório de tamanho $n = 120$ é extraída de uma distribuição uniforme no intervalo $[0,1]$ ($U_{[0,1]}$). Qual é a probabilidade de que $P(|\bar{X} - 1/2| \leq 0.1)$?

- ▶ A média da distribuição $U_{[a,b]}$ é $\frac{a+b}{2}$, no nosso caso: $\frac{0+1}{2} = 1/2$
- ▶ A variância da distribuição $U_{[a,b]}$ é $\frac{(b-a)^2}{12}$, no nosso caso: $1/12$
- ▶ Aplicando o TCL:

$$\frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} = \frac{\sqrt{120}(\bar{X} - 1/2)}{1/\sqrt{12}} \sim_{approx} N(0, 1)$$

- ▶ Queremos: $P(|\bar{X} - 1/2| \leq 0.1)$

Distribuição da média amostral

$$\frac{\sqrt{120}(\bar{X} - 1/2)}{1/\sqrt{12}} \sim_{approx} N(0, 1)$$

Distribuição da média amostral

$$\frac{\sqrt{120}(\bar{X} - 1/2)}{1/\sqrt{12}} \sim_{approx} N(0, 1)$$

► $P(|\bar{X} - 1/2| \leq 0.1) = P(-0.1 \leq \bar{X} - 1/2 \leq 0.1)$

Distribuição da média amostral

$$\frac{\sqrt{120}(\bar{X} - 1/2)}{1/\sqrt{12}} \sim_{approx} N(0, 1)$$

- ▶ $P(|\bar{X} - 1/2| \leq 0.1) = P(-0.1 \leq \bar{X} - 1/2 \leq 0.1)$
- ▶ $P(|\bar{X} - 1/2| \leq 0.1) = P\left(\frac{-0.1}{1/\sqrt{12}} \leq \frac{\bar{X} - 1/2}{1/\sqrt{12}} \leq \frac{0.1}{1/\sqrt{12}}\right)$

Distribuição da média amostral

$$\frac{\sqrt{120}(\bar{X} - 1/2)}{1/\sqrt{12}} \sim_{approx} N(0, 1)$$

- ▶ $P(|\bar{X} - 1/2| \leq 0.1) = P(-0.1 \leq \bar{X} - 1/2 \leq 0.1)$
- ▶ $P(|\bar{X} - 1/2| \leq 0.1) = P\left(\frac{-0.1}{1/\sqrt{12}} \leq \frac{\bar{X} - 1/2}{1/\sqrt{12}} \leq \frac{0.1}{1/\sqrt{12}}\right)$
- ▶ $P(|\bar{X} - 1/2| \leq 0.1) = P\left(\underbrace{\sqrt{120} \frac{-0.1}{1/\sqrt{12}}}_{-3.794733} \leq Z \leq \underbrace{\sqrt{120} \frac{0.1}{1/\sqrt{12}}}_{3.794733}\right)$

Distribuição da média amostral

$$\frac{\sqrt{120}(\bar{X} - 1/2)}{1/\sqrt{12}} \sim_{approx} N(0, 1)$$

- ▶ $P(|\bar{X} - 1/2| \leq 0.1) = P(-0.1 \leq \bar{X} - 1/2 \leq 0.1)$
- ▶ $P(|\bar{X} - 1/2| \leq 0.1) = P\left(\frac{-0.1}{1/\sqrt{12}} \leq \frac{\bar{X} - 1/2}{1/\sqrt{12}} \leq \frac{0.1}{1/\sqrt{12}}\right)$
- ▶ $P(|\bar{X} - 1/2| \leq 0.1) = P\left(\underbrace{\sqrt{120} \frac{-0.1}{1/\sqrt{12}}}_{-3.794733} \leq Z \leq \underbrace{\sqrt{120} \frac{0.1}{1/\sqrt{12}}}_{3.794733}\right)$

Distribuição da média amostral

$$\frac{\sqrt{120}(\bar{X} - 1/2)}{1/\sqrt{12}} \sim_{approx} N(0, 1)$$

- ▶ $P(|\bar{X} - 1/2| \leq 0.1) = P(-0.1 \leq \bar{X} - 1/2 \leq 0.1)$
- ▶ $P(|\bar{X} - 1/2| \leq 0.1) = P\left(\frac{-0.1}{1/\sqrt{12}} \leq \frac{\bar{X} - 1/2}{1/\sqrt{12}} \leq \frac{0.1}{1/\sqrt{12}}\right)$
- ▶ $P(|\bar{X} - 1/2| \leq 0.1) = P\left(\underbrace{\sqrt{120} \frac{-0.1}{1/\sqrt{12}}}_{-3.794733} \leq Z \leq \underbrace{\sqrt{120} \frac{0.1}{1/\sqrt{12}}}_{3.794733}\right)$

```
pnorm(3.794733) - pnorm(-3.794733)
```

```
## [1] 0.9998522
```

Métodos de amostragem

Métodos de amostragem

- ▶ Até agora discutimos a amostragem aleatória simples e discutimos propriedades de \bar{X} e \bar{p} quando se usa esta forma de amostragem.

Métodos de amostragem

- ▶ Até agora discutimos a amostragem aleatória simples e discutimos propriedades de \bar{X} e \bar{p} quando se usa esta forma de amostragem.
- ▶ Contudo, não é o único método de amostragem

Métodos de amostragem

- ▶ Até agora discutimos a amostragem aleatória simples e discutimos propriedades de \bar{X} e \bar{p} quando se usa esta forma de amostragem.
- ▶ Contudo, não é o único método de amostragem
- ▶ Discutiremos brevemente outros métodos de amostragem.

Métodos de amostragem

Amostragem estratificada

- ▶ Os elementos da população são divididos em grupos chamados de *estratos* de forma que cada elemento pertence unicamente a um estrato

Métodos de amostragem

Amostragem estratificada

- ▶ Os elementos da população são divididos em grupos chamados de *estratos* de forma que cada elemento pertence unicamente a um estrato
- ▶ Selecionamos uma amostra aleatoria de cada um dos estratos

Métodos de amostragem

Amostragem estratificada

- ▶ Os elementos da população são divididos em grupos chamados de *estratos* de forma que cada elemento pertence unicamente a um estrato
- ▶ Selecionamos uma amostra aleatoria de cada um dos estratos
- ▶ Quanto mais homogêneos os estratos, melhor a nossa amostra

Métodos de amostragem

Amostragem por conglomerados

- ▶ Dividimos a população em conglomerados (grupos) em que cada elemento pertence unicamente a um conglomerado.

Métodos de amostragem

Amostragem por conglomerados

- ▶ Dividimos a população em conglomerados (grupos) em que cada elemento pertence unicamente a um conglomerado.
- ▶ Selecionamos uma amostra aleatoria dos conglomerados

Métodos de amostragem

Amostragem por conglomerados

- ▶ Dividimos a população em conglomerados (grupos) em que cada elemento pertence unicamente a um conglomerado.
- ▶ Seleccionamos uma amostra aleatoria dos conglomerados
- ▶ Todos os elementos contidos nos conglomerados seleccionados formam parte da nossa amostra

Métodos de amostragem

Amostragem por conglomerados

- ▶ Dividimos a população em conglomerados (grupos) em que cada elemento pertence unicamente a um conglomerado.
- ▶ Selecionamos uma amostra aleatoria dos conglomerados
- ▶ Todos os elementos contidos nos conglomerados selecionados formam parte da nossa amostra
- ▶ Quanto mais heterogeneo sejam os conglomerados, melhor a nossa amostra

Métodos de amostragem

Amostragem por conglomerados

- ▶ Dividimos a população em conglomerados (grupos) em que cada elemento pertence unicamente a um conglomerado.
- ▶ Selecionamos uma amostra aleatoria dos conglomerados
- ▶ Todos os elementos contidos nos conglomerados selecionados formam parte da nossa amostra
- ▶ Quanto mais heterogeneo sejam os conglomerados, melhor a nossa amostra
- ▶ Ex. amostrar bairros

Métodos de amostragem

Amostragem sistemática

- ▶ Muitas vezes, selecionar uma amostra aleatória simples pode não ser viável ou consumir muito tempo

Métodos de amostragem

Amostragem sistemática

- ▶ Muitas vezes, selecionar uma amostra aleatória simples pode não ser viável ou consumir muito tempo
- ▶ Ex: clientes de uma determinada loja física

Métodos de amostragem

Amostragem sistemática

- ▶ Muitas vezes, selecionar uma amostra aleatória simples pode não ser viável ou consumir muito tempo
- ▶ Ex: clientes de uma determinada loja física
- ▶ Uma alternativa é utilizar amostragem sistemática

Métodos de amostragem

Amostragem sistemática

- ▶ Muitas vezes, selecionar uma amostra aleatória simples pode não ser viável ou consumir muito tempo
- ▶ Ex: clientes de uma determinada loja física
- ▶ Uma alternativa é utilizar amostragem sistemática
- ▶ Consiste em amostrar elementos a cada k elementos

Métodos de amostragem

Amostragem sistemática

- ▶ Muitas vezes, selecionar uma amostra aleatória simples pode não ser viável ou consumir muito tempo
- ▶ Ex: clientes de uma determinada loja física
- ▶ Uma alternativa é utilizar amostragem sistemática
- ▶ Consiste em amostrar elementos a cada k elementos
- ▶ Ex: amostrar o primeiro cliente que entra na loja, o cliente $1 + k$, $1 + 2k$, \dots $1 + nk$ até completarmos nossa amostra.

Métodos de amostragem

Amostragem sistemática

- ▶ Muitas vezes, selecionar uma amostra aleatória simples pode não ser viável ou consumir muito tempo
- ▶ Ex: clientes de uma determinada loja física
- ▶ Uma alternativa é utilizar amostragem sistemática
- ▶ Consiste em amostrar elementos a cada k elementos
- ▶ Ex: amostrar o primeiro cliente que entra na loja, o cliente $1 + k$, $1 + 2k$, $\dots 1 + nk$ até completarmos nossa amostra.
- ▶ Se tivermos o tamanho total de população N e quisermos uma amostra de tamanho n , escolhemos elemento de $k = N/n$ em k . Ex: k , $2k$, $\dots nk$

Métodos de amostragem

Amostragem por conveniência

- ▶ Amostragem não-probabilística

Métodos de amostragem

Amostragem por conveniência

- ▶ Amostragem não-probabilística
- ▶ A amostra é identificada por conveniência

Métodos de amostragem

Amostragem por conveniência

- ▶ Amostragem não-probabilística
- ▶ A amostra é identificada por conveniência
- ▶ Ex: ao se fazer uma pesquisa na universidade, podemos selecionar estudantes voluntários (simplesmente por eles estarem disponíveis)

Métodos de amostragem

Amostragem por conveniência

- ▶ Amostragem não-probabilística
- ▶ A amostra é identificada por conveniência
- ▶ Ex: ao se fazer uma pesquisa na universidade, podemos selecionar estudantes voluntários (simplesmente por eles estarem disponíveis)

Métodos de amostragem

Amostragem por conveniência

- ▶ Amostragem não-probabilística
- ▶ A amostra é identificada por conveniência
- ▶ Ex: ao se fazer uma pesquisa na universidade, podemos selecionar estudantes voluntários (simplesmente por eles estarem disponíveis)

Amostragem de julgamento

- ▶ Amostragem não-probabilística

Métodos de amostragem

Amostragem por conveniência

- ▶ Amostragem não-probabilística
- ▶ A amostra é identificada por conveniência
- ▶ Ex: ao se fazer uma pesquisa na universidade, podemos selecionar estudantes voluntários (simplesmente por eles estarem disponíveis)

Amostragem de julgamento

- ▶ Amostragem não-probabilística
- ▶ A pessoa que conhece profundamente o tema do estudo escolhe os elementos que julga serem mais importantes da população

Métodos de amostragem

Amostragem por conveniência

- ▶ Amostragem não-probabilística
- ▶ A amostra é identificada por conveniência
- ▶ Ex: ao se fazer uma pesquisa na universidade, podemos selecionar estudantes voluntários (simplesmente por eles estarem disponíveis)

Amostragem de julgamento

- ▶ Amostragem não-probabilística
- ▶ A pessoa que conhece profundamente o tema do estudo escolhe os elementos que julga serem mais importantes da população
- ▶ Ex: um reporter escolher alguns deputados para darem sua opinião sobre algum fato.

Leituras recomendadas

- ▶ Anderson, D. R; Sweeney, D. J.; e Williams, T. A. (2008). *Estatística Aplicada à Administração e Economia*. 2ed. Cengage Learning. **Cap 7**
- ▶ Morettin, P.A; e Bussab, W. de O. (2004). *Estatística Básica*. 5ed, Saraiva. **Cap 10**