Faculdade de Filosofia, Letras e Ciências Humanas
Pós-Graduação em Ciência Política
Graduação em Ciências Sociais

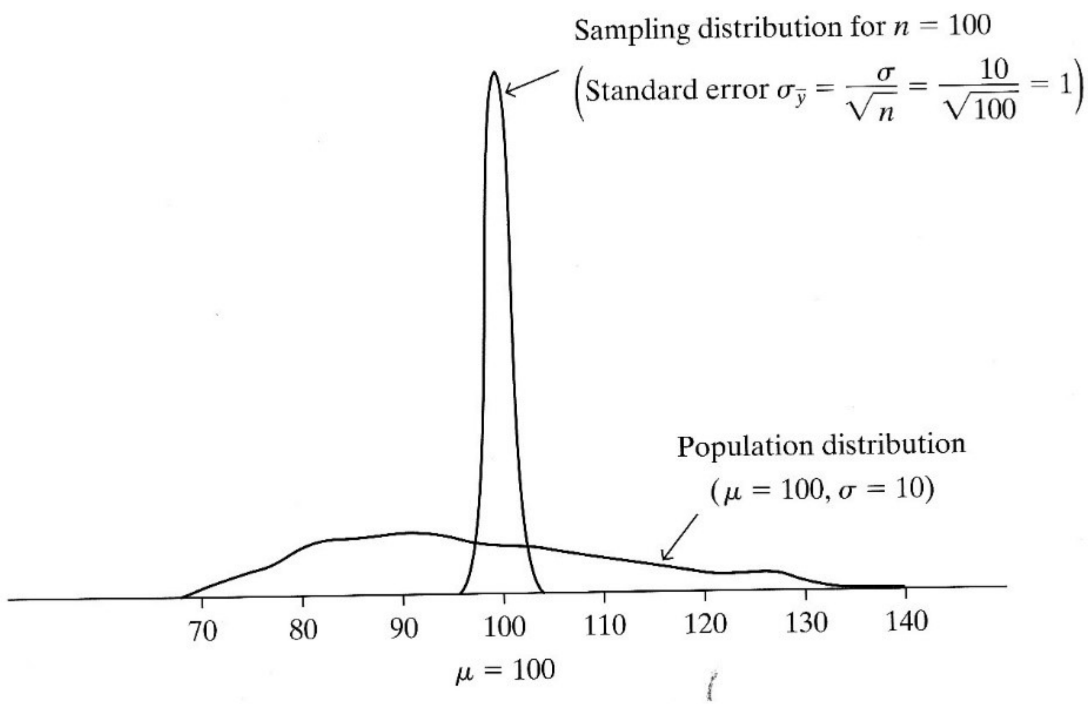
FLS 5028 - Métodos Quantitativos e Técnicas de Pesquisa em Ciência Política
FLP 0406 -Métodos e Técnicas de Pesquisa em Ciência Política

Aula 8 – Inferência Estatística

Glauco Peres da Silva

1º sem / 2021

Inferência Estatística



Inferência Estatística

Compreender um fenômeno não observado a partir da observação de fenômenos observáveis.

Representação da realidade e o papel do acaso. (ex.: os resultados das eleições nos estados ou municípios e a queda de chuvas em certas regiões e o boicote eleitoral em outras).

Os aspectos imprevisíveis irão acarretar que um dado coletado, ainda apresentará variação. (ex.: os resultados das mesmas eleições se realizadas uma semana antes ou depois).

Inferência Estatística

Uma maneira de pensar sobre a inferência é considerar os nossos dados apenas como um dentre vários possíveis.

Exemplos:

1. A idade dos brasileiros coletada pelo PNAD apresentará uma variação que terá componentes sistemáticos (renda, região onde mora, etc.) e não-sistemáticos (referentes à época em que a pesquisa foi feita, domicílios visitados, etc).

2. Pesquisas eleitorais, pois a variação das intenções de voto entre os eleitores terá:

a) um componente estrutural (renda, região, religião, preferência partidária, etc.);

b) e outro circunstancial (relativo as circunstâncias em que a pesquisa foi feita).

Inferência Estatística

Na prática, devemos buscar evidências de que a relação que procuramos explicar é produzida por componentes sistemáticos; a discussão da natureza dos componentes não-sistemáticos **não** é irrelevante.

As pesquisas qualitativas devem também buscar a inferência, tentando separar os componentes sistemáticos e randômicos. Para isso, não são necessárias grandes quantidades de observações.

Inferência Estatística

Modo de encontrar informações sobre a população a partir de amostras.

Dois tipos de inferência:

- Ponto
- Intervalo

Usualmente a inferência se volta para a média ou para a variância.

Inferência Estatística

População	Estimativa de ponto a partir de uma distribuição amostral
μ	$\hat{\mu} = \bar{y} = \frac{1}{n}\sum y_i$
$\sigma_y = \sqrt{\sum \frac{(y_i - \mu)^2}{n}}$	$\hat{\sigma}_y = s_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$
$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{n}}$	$\hat{\sigma}_{\bar{y}} = \mathrm{s}_{\bar{\mathrm{y}}} = \frac{\mathrm{s}_\mathrm{y}}{\sqrt{\mathrm{n}}}$

Inferência Estatística

$\bar{Y}$ é um estimador não viesado de μ.

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$

$E(\bar{Y}) = E(\frac{1}{n}\sum_{i=1}^{n} Y_i)$

$$= \frac{1}{n}\sum_{i=1}^{n} E(Y_i)$$

$= \frac{1}{n} n\mu$

$= \mu$

Inferência Estatística

y_i = variável realizada

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Exemplo. y_i = o número de votos para o Partido dos Trabalhadores no distrito i com n= 650 distritos.

Variância (Y) = A distância de nossos estimadores do parâmetro correto (Y)= σ^2

Variância $(\bar{Y}) = \frac{\sigma^2}{n}$

Inferência Estatística

Exemplo. Variância $(Y_{\text{i}}) = \sigma_i^2$ do número de votos do Partido dos Trabalhadores na Inglaterra no distrito *i*

Variância $(Y_1) = \sigma_1^2$ do número de votos do Partido dos Trabalhadores na Inglaterra no distrito 1

Se n= 650 distritos,
Variância $(\bar{Y}) = \frac{\sigma^2}{n} = \frac{\sigma^2}{650}$ do número de votos do Partido dos Trabalhadores na Inglaterra em 650 distritos

Inferência Estatística

Existem 3 critérios para julgar a qualidade das inferências descritivas:

A **ausência de viés**

Viés = tendência a **erro sistemático** em relação ao parâmetro correto (esperança matemática).

Sua existência depende de nossa teoria e objetivos de pesquisa, não existem dados enviesados a priori.

Exemplo: *survey* eleitoral com entrevistas realizadas na casa das pessoas durante o dia.

Inferência Estatística

A Eficiência: qual a distância de nossos estimadores do parâmetro correto?

Eficiência é um conceito relativo, medido pela variância ($\frac{\sigma^2}{n}$) do estimador através de “n” replicações hipotéticas dos mesmos procedimentos.

Quanto mais reaplicações, maior o número de observações e menor variância. (Quando “n” => infinito, $\frac{\sigma^2}{n}$=> 0).

Para estimadores livres de viés, quanto menor for a variância, maior será sua eficiência.

Inferência Estatística

O dilema entre viés e ineficiência

Muitas situações podem ocorrer de que ganhos de eficiência levam a um pequeno viés. Nestes casos, é preciso pesar um e outro antes de tomar uma decisão.

Embora a formulação deste problema seja mais fácil em análises quantitativas, ele pode ser útil também nas análises qualitativas.

Exemplo

Medida 1: Pesquisa de 650 observações com viés

E(d) $= \mu - 0{,}01$

Var (d) $= \frac{\sigma^2}{650}$

Medida 2: Pesquisa de n =2 com ausência de viés, porém com menos eficiência

E(c) $= \mu$

Var (c) $= \frac{\sigma^2}{2}$

Que medida deveríamos preferir?

Inferência Estatística

O dilema entre viés e ineficiência

Medida **Erro Quadrático Médio** (*Mean Square Error* -MSE) como procedimento na decisão sobre este tipo de problema.

MSE = V(g) + E(g - γ)^2 = variância + (viés)^2

Exemplo

Medida 1: MSE (d) $= \frac{\sigma^2}{650} + \ (0.01)^2$

Medida 2: MSE(c) $= \frac{\sigma^2}{2}$

MSE (d) < MSE (c)

Nossa medida preferida é a Medida 1.

Inferência Estatística

Intervalo de confiança (IC)

O IC para um parâmetro populacional qualquer é um intervalo de números nos quais acredita-se que o verdadeiro parâmetro populacional esteja. A probabilidade de que método produza um intervalo que contenha o parâmetro é chama de **nível de confiança**.

$$\left[\bar{X} \pm Z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$$

Inferência Estatística

Intervalo de Confiança para Populações Finitas

No caso de populações finitas com seleção ***sem*** reposição, deve-se adicionar um termo de correção. Este termos de correção é adicionado ao estimador do erro amostral da seguinte forma:

$$\left[\bar{X} \pm Z\alpha/_2 \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}\right]$$

Em que N é o tamanho total da população.

Distribuição Populacional

<u>Exemplo Eleição para o governo da California</u>
y =1 (voto para o Partido Republicano) ou y=0 (voto para o Partido Democrata)

Se 50% da população apoia o candidato Republicano, o restante apoia o Democrata,

E(y) = $\bar{y} = \sum p(y)y = 0 * 0.50 + 0 * 1 = 0.5$

$s_y = 0.50$

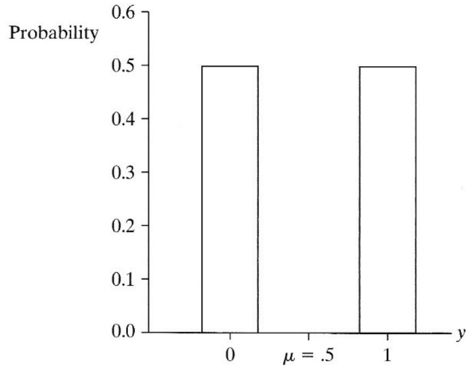
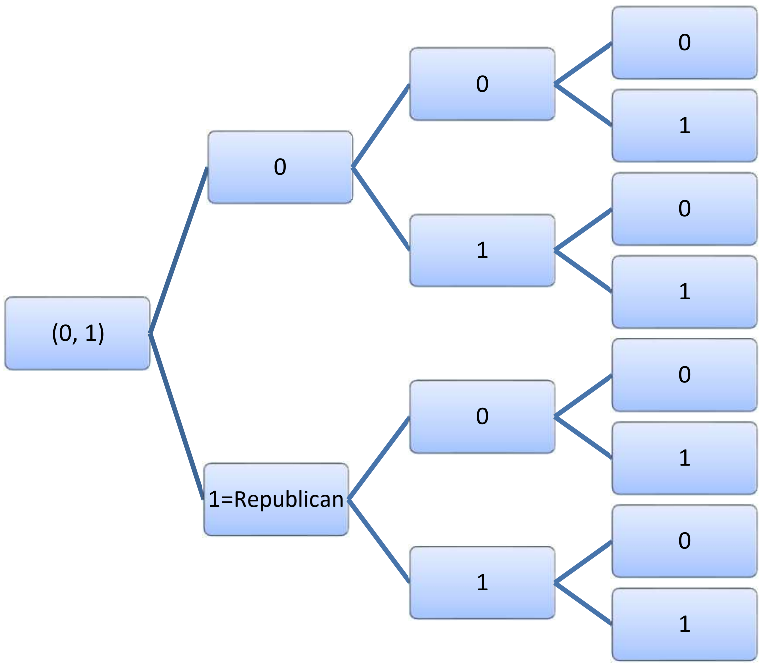


FIGURE 4.12: The Population Distribution when $y = 0$ or 1, with Probability 0.50 Each. This is the distribution for a vote, with 1 = vote for Republican candidate and 0 = vote for Democratic candidate.

Distribuição Amostral

Exemplo Eleição para o governo da California

Podemos construir uma distribuição amostral para uma amostra de n=3 eleitores de uma população na qual 50% dos eleitores apoiem o candidato Republicano:

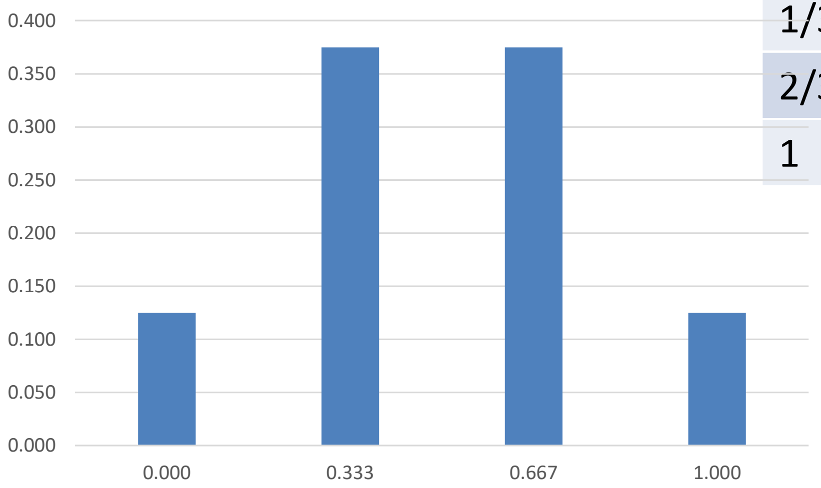


Distribuição Amostral

Exemplo Eleição para o governo da California

Podemos construir uma distribuição amostral para uma amostra de n=3 eleitores de uma população na qual 50% dos eleitores apoiem o candidato Republicano:

Distribuição amostral da média populacional
quando n=3 eleitora (P=0 or P=1)

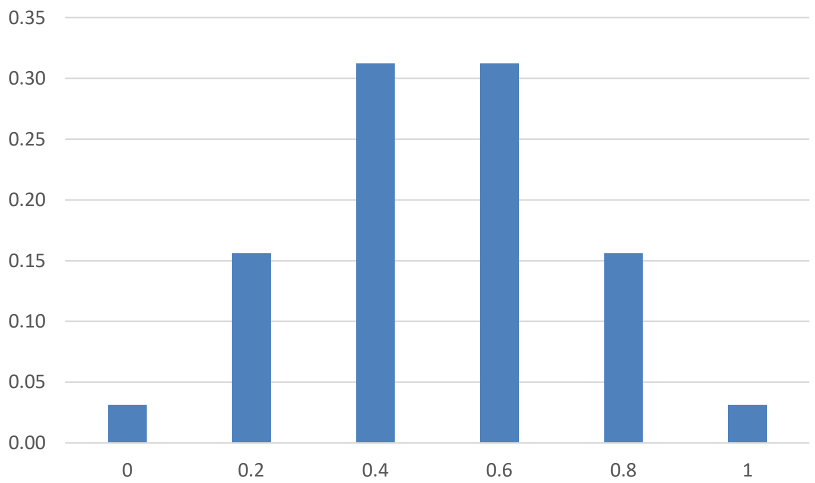


Proporção amostral	Probabilidade
0	1/8=0.125
1/3	3/8=0.375
2/3	3/8= 0.375
1	1/8=0.125

Distribuição Amostral

Exemplo Eleição para o governo da California

Podemos construir uma distribuição amostral para uma amostra de n=5 eleitores de uma população na qual 50% dos eleitores apoiem o candidato Republicano:



$\mu = 0.5$

Proporção amostral	Proabilidade
0/5= 0 cara em 5 lançamentos	0.03
1/5= 1 cara em 5 lançamentos	0.16
2/5= 2 caras em 5 lançamentos	0.31
3/5= 3 caras em 5 lançamentos	0.31
4/5= 4 caras em 5 lançamentos	0.16
5/5= 5 caras em 5 lançamentos	0.03

Conforme *n* aumenta, a distribuição amostral da proporção amostral parece se tornar mais simétrica ("mais normal", com as probabilidades concentrando-se mais em torno de 0,50 (a média da proporção amostral).

Distribuição Amostral

Exemplo Eleição para o governo da California

n	Média Amostral	Erro Amostral
3	0.50	0.29
100	0.50	0.05
2705	0.50	0.01

A Distribuição Amostral da Proporção Amostral

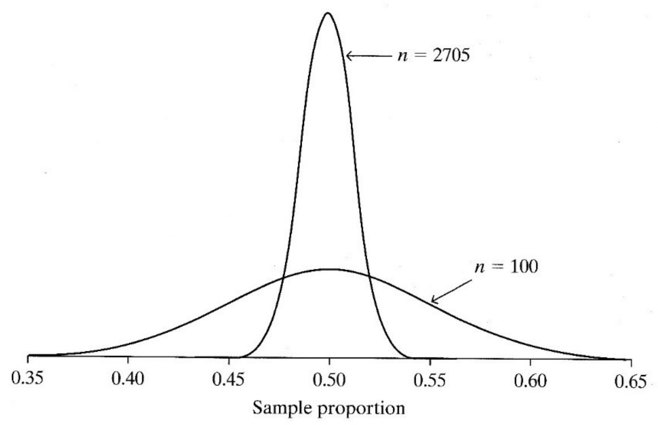


FIGURE 4.13: The Sampling Distributions of the Sample Proportion, when $n = 100$ and when $n = 2705$. These refer to sampling from the population distribution in Figure 4.12.

Teorema do Limite Central

O **Teorema do Limite Central (TLC)** nos diz que um número grande de variáveis independentes e identicamente distribuídas terão distribuição aproximadamente normal, independentemente da distribuição original.

$$\overline{X_n} \sim N\left(\mu_X, \frac{\sigma^2}{n}\right), \text{ quando } n \to \infty$$

Leis dos Grandes Números

A **Lei dos Grandes Números (LGN)** é um teorema que diz que conforme aumenta o número de amostras em um processo aleatório, a diferença percentual entre o valor esperado e o valor medido tende a zero.

$$\overline{X_n} \xrightarrow[n \to \infty]{} E(X)$$

Ex.: Falácia do jogador

TLC vs LGN

LGN nos garante que quando o número de amostras tende ao infinito, a média dessas observações é igual ao valor esperado.

TLC nos garante que quando o tamanho da amostra aumenta, a distribuição dos valores amostrais tende a uma distribuição normal.

TLC nos garante precisão!
LGN nos garante a estimativa!