

Distribuições e Probabilidade

PRI5003 - Lab 5

Instituto de Relações Internacionais - Universidade de São Paulo

26 de Abril de 2018

Outline

Conceitos Básicos

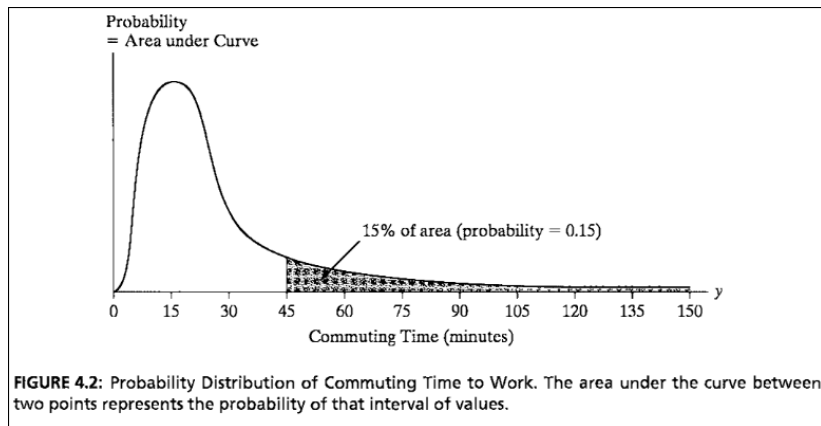
Distribuição Amostral

Erro padrão

Lei dos Grandes Números e Teorema do Limite Central

Revisão

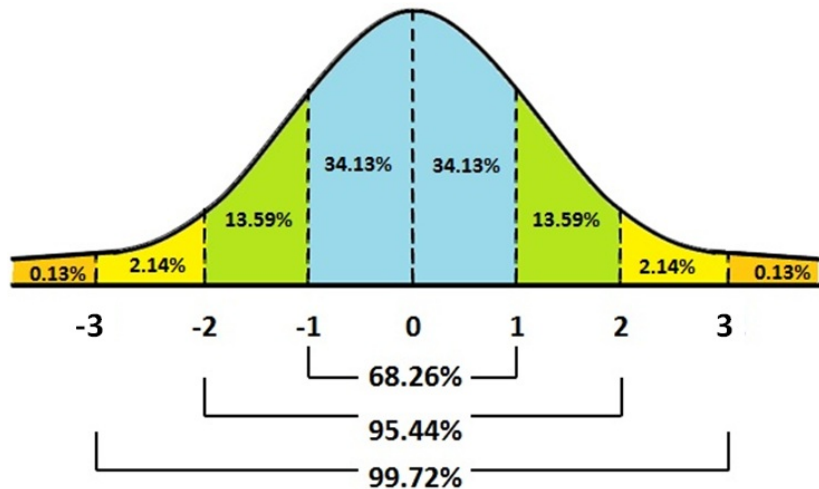
Probabilidade = Área sob a curva



Curva normal

- ▶ $P(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$
- ▶ Simétrica, em forma de sino
- ▶ Caracterizada por dois parâmetros: média μ e desvio-padrão σ
- ▶ Muito comum em fenômenos naturais, e também muito importante em ciências sociais devido ao TLC
- ▶ Normal padronizada: $\mu = 0$ e $\sigma = 1$

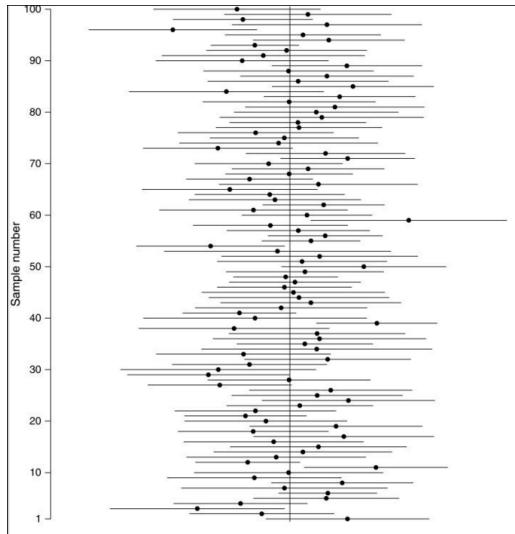
Curva Normal



Distribuição de amostras

- ▶ Em ciências sociais, quase nunca as distribuições são normais.
- ▶ Pior ainda: muitas vezes, não sabemos como se comporta a distribuição real
- ▶ Nós podemos usar amostras para fazer inferências sobre os parâmetros dessa distribuição
- ▶ As estatísticas dessas amostras, quando obtidas repetidamente, seguem uma distribuição conhecida
- ▶ Não estamos mais falando de distribuição de valores de uma variável, mas de estatísticas que emergem de diversas amostras
- ▶ Exemplo: como se comporta a distribuição da intenção de votos em uma eleição?

Eleição



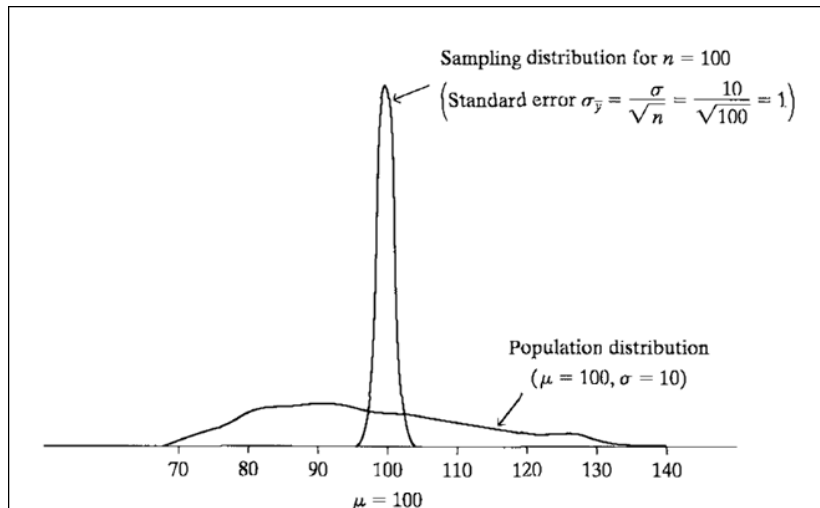
Distribuição de amostras

- ▶ Se fizermos diversas amostras de uma mesma população, podemos esperar variabilidade nos resultados.
- ▶ Esse erro é inerente ao fato de que estamos utilizando apenas uma parte da população: a amostra
- ▶ Por isso institutos de pesquisas têm números diferentes para uma eleição. Mas essa variação segue um padrão conhecido
- ▶ Para o caso da média, podemos calcular a variação esperada de diferentes amostras por meio do desvio-padrão
- ▶ Na prática, não precisamos tirar diversas amostras: podemos calcular o erro padrão pela fórmula $\frac{\sigma}{\sqrt{n}}$

Erro padrão

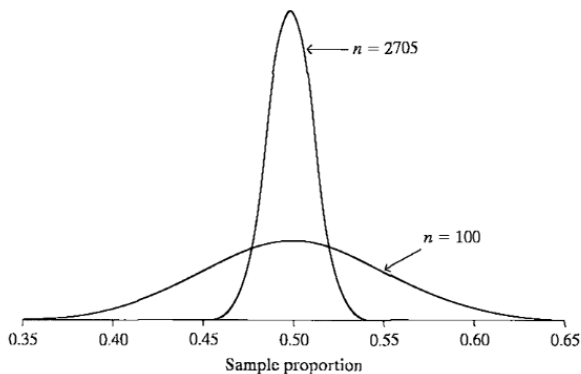
É o erro que esperamos obter se fizermos diversas amostras de uma mesma população

Erro padrão



Maior amostra - menor erro

In summary, error results from estimating μ by \bar{y} , because we sampled only part of the population. This error, which is the **sampling error**, tends to decrease as the sample size n increases. The standard error is fundamental to inferential procedures that predict the sampling error in using \bar{y} to estimate μ .



Lei dos Grandes Números e Teorema do Limite Central

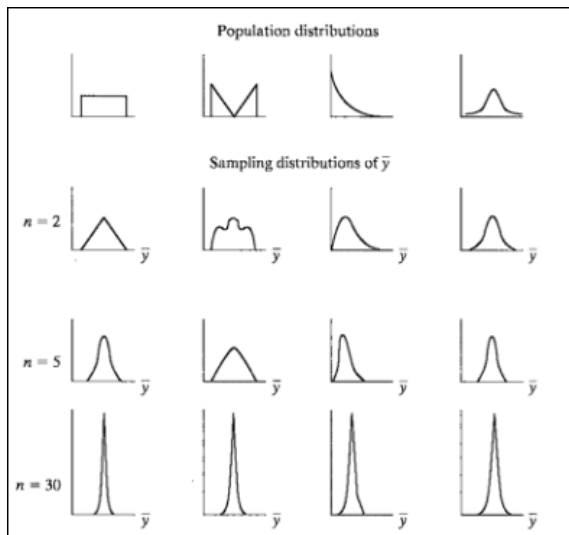
Lei dos Grandes Números

A média amostral converge para o valor esperado da distribuição à medida em que o n cresce. Exemplo: jogar dados repetidas vezes

Teorema do Limite Central

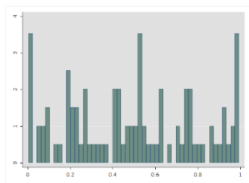
A distribuição amostral da média se aproxima de uma normal, independentemente da distribuição original da variável. Uma aproximação

Teorema do Limite Central



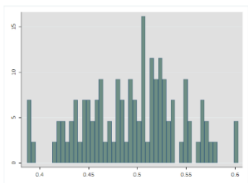
TLC vs Lei dos Grandes Números

Caso 1



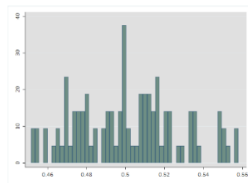
n de cada amostra = 1

Quantidade de amostras = 100



n de cada amostra = 30

Quantidade de amostras = 100

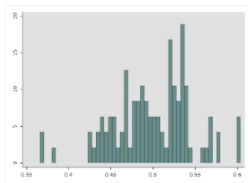


n de cada amostra = 200

Quantidade de amostras = 100

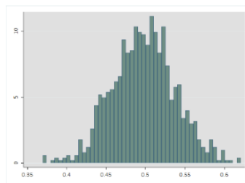
TLC vs Lei dos Grandes Números

Caso 2



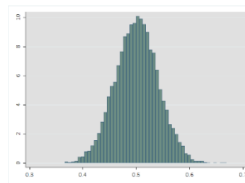
n de cada amostra = 50

Quantidade de amostras = 100



n de cada amostra = 50

Quantidade de amostras = 1000



n de cada amostra = 50

Quantidade de amostras = 10000

Questões

1. Por que alterar o n de cada amostra (mantendo-se constante a quantidade de amostras) é diferente de alterar a quantidade de amostras (mantendo-se constante o n de cada amostra)?
2. Quais são os nomes técnicos dos fenômenos observados para cada caso? Quais são as definições técnicas de cada um deles? Como se distinguem?

Revisão

- ▶ Distribuição da população **vs** distribuição dos dados amostrais **vs** distribuição amostral
- ▶ Desvio-padrão **vs** erro padrão
- ▶ Número de amostras **vs** número de observações