

PLN para a Ciência Política e Políticas Públicas Públicas

Professora: Lorena Barberia

Semana 2



Tópicos da Aula

1 Desenho da Pesquisa

Desenho de Pesquisa

Metodologia *Project-Based Learning* ancorada no projeto *Mapping Political Elites COVID-19 Vaccine Tweets in Brazil*.

Desenho de Pesquisa

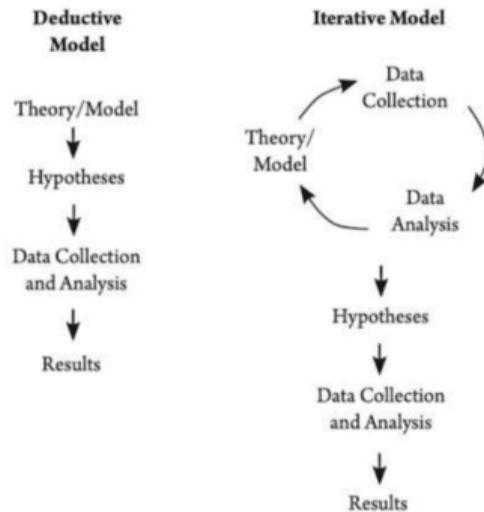


Figure: Métodos dedutivos vs. indutivos

Métodos de Inferência e Amostragem

Table: Classificação de desenhos de amostras de acordo com Hansen et al. (1983)

Plano Amostral	Método de Inferência dos Dados Gerados pelo Plano Amostral	
	Baseado no Desenho	Baseado no Modelo
Amostra Probabilística	A	B
Amostra dependente do modelo	C	D

Desenho de Pesquisa

- Podemos pensar em 2 diferentes tipos de planos de amostragem:
 - Os planos de **amostragem probabilística** atribuem a cada membro da população uma probabilidade conhecida diferente de zero de inclusão na amostra. Seleção aleatória
 - Os planos de **amostragem dependentes no modelo** assumem que as variáveis de interesse na pesquisa seguem uma distribuição de probabilidade conhecida e que a escolha dos elementos da amostra maximiza a precisão da estimativa para as estatísticas de interesse.

Desenho de Pesquisa

- *Desenhos Prospectivos* (Prospective Study Designs)
Uma amostra probabilística de indivíduos em algum ponto inicial e, em seguida, medida repetidamente ao longo do tempo em ondas.

Desenho de Pesquisa

- 2 diferentes tipos de métodos de inferência:
 - Modelos **não paramétricos**, e.g. $y \sim N(xB + \epsilon, \sigma^2)$ não há suposições relativas à natureza da população da qual uma amostra é extraída ou métodos baseados na distribuição de probabilidade para a seleção da amostra.
 - Modelos **paramétricos**, suposições relativas à distribuição específica da população e parâmetros ou métodos que consideram a distribuição de probabilidade para a variável aleatória de interesse.

Desenho de Pesquisa

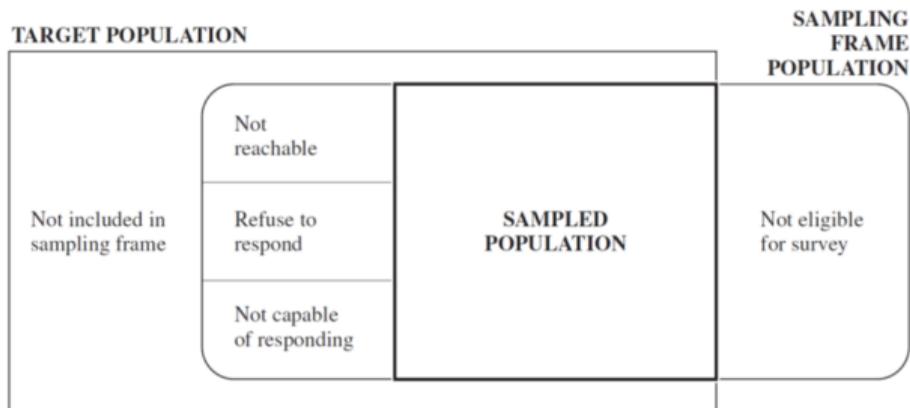
- Perguntas importantes sobre a seleção da amostra:
 - A amostra é selecionada em um único estágio ou em vários estágios?
 - O agrupamento (*clusters*) é usado em um ou mais estágios de amostra?
 - A estratificação é empregada em um ou mais estágios da amostra?
 - Os elementos são selecionados com probabilidades iguais?

Características de uma "boa" estratégia de seleção da amostra

- As características da amostra são representativas da população de interesse do estudo
 - População alvo (o total de indivíduos que queremos estudar)
 - Sampling frame (quem é possível amostrar)
 - Sampling unit (como está organizada a amostra, e.g. famílias)

População alvo e sampling frame

Figure: Lohr(2019)

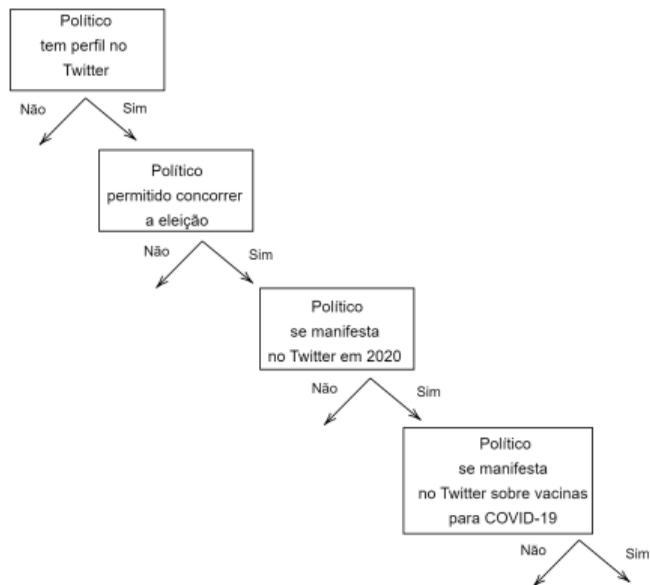


Corpus

A construção de um corpus envolve uma estratégia de amostragem e de desenho de pesquisa.

População alvo e sampling frame

Figure: COVID-19 Vaccine Tweets de Elites Políticas



O Problema em Aprendizado com Supervisão

- Y = classificação (e.g. posicionamento, sentimento, etc.)
- X = vetor de características preditoras
- *Training Data Set* = $(x_1, y_1, \dots, x_N, y_N)$
- Coleta e Avaliação de outras Amostras para avaliar se o *training data set* corretamente classifica outras amostras
- Statistical Framework: $y = f(x) + \epsilon$

Dúvidas?