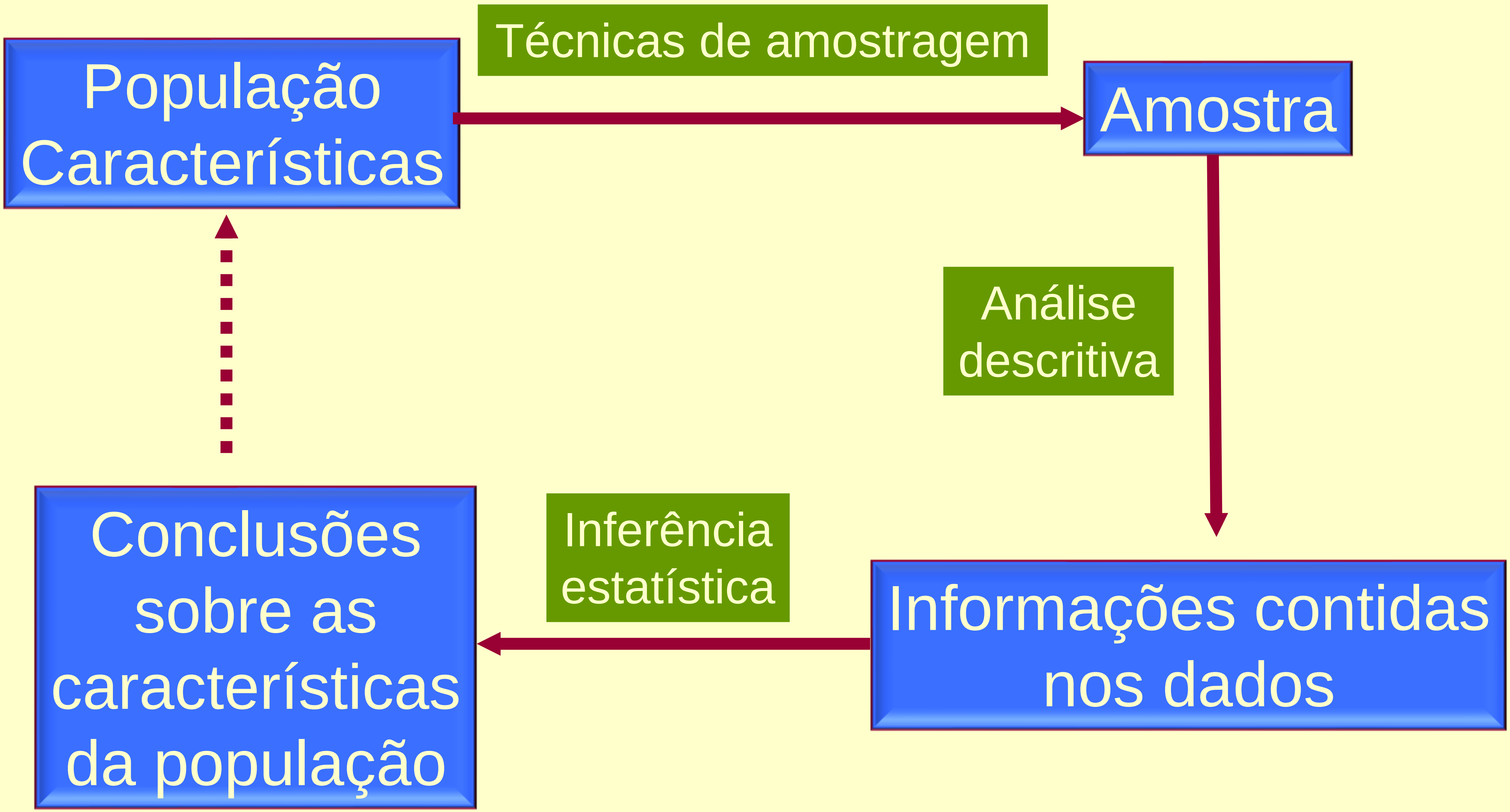
MAE-212: Introdução à Probabilidade e à Estatística II

Aula 01

Inferência Estatística

Estatística

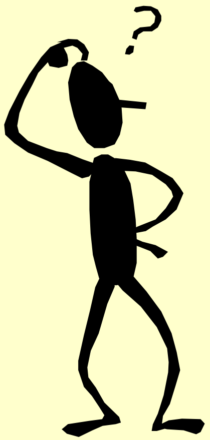


MÉTODO ESTATÍSTICO

Análise Descritiva	Teoria das Probabilidades	Inferência
↓	↓	↓
"Os dados efetivamente observados parecem mostrar que ..."	**"Se a distribuição dos dados seguir uma certa Lei, é esperado ..."**	**Ajuste de um modelo abstrato aos dados de uma amostra**
?	**?**	↓
		Estimação **Testes de hipóteses**

Estatística Descritiva

O que fazer com as observações que coletamos?



Primeira Etapa:

Resumo dos dados = Estatística descritiva

Medidas: média, variância, desvio padrão

Gráficos: boxplot, histograma

Inferência Estatística

procura os argumentos estatísticos para fazer afirmações sobre as características de uma população, com base em informações dadas por amostras.

Exemplo 1: observe como uma cozinheira verifica se o prato que ela está preparando está pronto ou não.

Exemplo 2: como você concluiria se um bolo está delicioso ou não.

Essas são decisões baseadas em procedimentos amostrais.

População e Amostra

Estudamos algumas distribuições teóricas de probabilidade que procuram medir a variabilidade de fenômenos casuais de acordo com suas ocorrências: as distribuições de probabilidades de variáveis aleatórias (qualitativas ou quantitativas), por exemplo, distribuição *binomial* e *normal*.

Na prática, frequentemente o pesquisador tem alguma ideia sobre a forma da distribuição, mas não dos valores exatos dos parâmetros que a especificam.

População e Amostra

Por exemplo, parece razoável supor que a distribuição das alturas dos brasileiros adultos possa ser representada por um modelo normal (embora as alturas não possam assumir valores negativos). Mas essa afirmação não é suficiente para determinar qual a distribuição normal correspondente; precisaríamos conhecer os parâmetros (média e variância) dessa normal para que ela ficasse completamente especificada. O propósito do pesquisador seria, então, descobrir (estimar) os parâmetros da distribuição para sua posterior utilização.

População e Amostra

Se pudéssemos medir as alturas de todos os brasileiros adultos, teríamos meios de obter sua distribuição exata e, daí, produzir os correspondentes parâmetros. Mas, nessa situação, não teríamos necessidade de usar a inferência estatística!

Raramente se consegue obter a distribuição exata de alguma variável, ou porque isso é muito dispendioso, ou muito demorado ou, às vezes, porque consiste num processo destrutivo.

Por exemplo, se estivéssemos observando a durabilidade de lâmpadas e testássemos todas até queimarem, não restaria nenhuma para ser vendida. Assim, a solução é selecionar parte dos elementos (amostra), analisá-la e *inferir* propriedades para o todo (população).

Amostragem

Associada a coleta de dados, a tecnologia da amostragem desenvolveu um conjunto de técnicas para obtenção de amostras representativas da população de interesse

Exemplos de utilização:
Pesquisa de mercado, pesquisa de opinião pública, ensaios de medicamentos e em praticamente todo experimento.

Amostragem Aleatória Simples (AAS): é a maneira mais fácil para selecionarmos uma amostra probabilistica de uma população. As unidades são independentemente selecionadas, uma de cada vez, até que o tamanho desejado da amostra seja atingido.

Definição: Uma amostra aleatória simples (AAS) de tamanho n de uma variável aleatória X, com dada distribuição, é o conjunto de n variáveis aleatórias independentes X_1, X_2, ..., X_n, cada uma com a mesma distribuição de X.

Conceitos Básicos

• **População:** é o conjunto de todos os elementos ou resultados sob investigação.

• **Amostra:** é qualquer subconjunto da população.

Exemplos:

1. Consideremos uma pesquisa para estudar os salários dos 500 funcionários da Companhia MB. Seleciona-se uma amostra de 36 indivíduos, e anotam-se os seus salários.
   Variável: salário; População: 500 funcionários; Amostra: 36 indivíduas selecionados

Conceitos Básicos

2. Queremos estudar a proporção de indivíduos na cidade A que são favoráveis a certo projeto governamental. Uma amostra de 200 pessoas é sorteada , e a opinião de cada uma é registrada como sendo a favor ou contra o projeto.

   Variável: resposta de um morador; População: todos os moradores da cidade; Amostra: 200 pessoas selecionadas

3. O interesse é investigar a duração de vida de um novo tipo de lâmpada, pois acreditamos que ela tenha uma duração maior do que as fabricadas atualmente. Então, 100 lâmpadas do novo tipo são deixadas acesas até queimarem. A duração em horas de cada lâmpada é registrada.

   Variável: duração em hora de cada lâmpada; População: todas as lâmpadas fabricadas ou que vem a ser fabricadas por essa empresa com o mesmo processo; Amostra: 100 lâmpadas selecionadas.

   obs: População infinita, modelo teórico para a distribuição da variável populacional

Conceitos Básicos

4. Em alguns casos, fazemos suposições mais precisas sobre a população. Digamos que X represente o peso real de pacotes de café, enchidos automaticamente por uma maquina. Sabe-se que a distribuição de X pode ser representada por uma normal, com parâmetro $\mu(X)$ e $\sigma^2(X)$ desconhecidos. Sorteamos 100 pacotes e medimos seus pesos.

   Variável: peso de cada pacote; População: todos os pacotes enchidos ou que virão a ser enchidos pela máquina; Amostra: 100 medidas dos pacotes selecionados.

5. Para investigar a "honesticidade" de uma moeda, nós a lançamos 50 vezes e contamos o número de caras observadas. A população pode ser considerada como tendo a distribuição da variável X, assumindo o valor 1, com probabilidade p, se ocorrer cara, e assumindo o valor 0, com probabilidade 1-p, se ocorrer coroa.

   X: distribuição Bernoulli.

Conceitos Básicos

- **Parâmetro:** é uma medida (constante) usada para descrever uma característica da população. Por exemplo, média e variância populacional da variável de interesse X, $\mu(X)$ e $\sigma^2(X)$, respectivamente.

- **Estatística:** uma característica da amostra, ou seja, uma estatística T é uma função da amostra (X_1, X_2, …, X_n) da variável de interesse X. Por exemplo, média amostral de X, que também é uma variável aleatória. $\bar{X} = \frac{1}{n}\{X_1 + X_2 + \ldots + X_n\}.$

- **Variável:** qualquer característica associada a uma população, por exemplo, peso, altura da população brasileira.

As estatísticas mais comuns são:

$$\bar{X} = 1/n \sum_{i=1}^{n} X_i \text{ : média da amostra,}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2 \text{ : variância da amostra,}$$

$$X_{(1)} = \min\left(X_1, X_2, \ldots, X_n \right)\text{: o menor valor da amostra,}$$

$$X_{(n)} = \max\left(X_1, X_2, \ldots, X_n \right)\text{: o maior valor da amostra,}$$

$$W = X_{(n)} - X_{(1)} \text{ : amplitude amostral,}$$

$$X_{(i)} = \text{a } \; i - \text{ésima maior observação da amostra}$$

Os símbolos mais comuns são dados na tabela a seguir.

Denominação	População	Amostra
Média	$\mu = E(X)$	$\bar{X} = \sum X_i / n$
Mediana	$\mathrm{Md} = Q_2$	$\mathrm{md} = q_2$
Variância	$\sigma^2 = \mathrm{Var}(X)$	$S^2 = \sum (X_i - \bar{X})^2/(n-1)$
Nº de elementos	N	n
Proporção	p	$\hat{p}$
Quantil	$Q(p)$	$q(p)$
Quartis	Q_1, Q_2, Q_3	q_1, q_2, q_3
Intervalo inter-quartil	$d_Q = Q_3 - Q_1$	$d_q = q_3 - q_1$
Função densidade	$f(x)$	histograma
Função de distribuição	$F(x)$	$F_e(x)$

Conceitos Básicos

Problemas de Inferência

Exemplo: Indicamos por X o número de caras obtidos depois de lançar a moeda 50 vezes, sabemos que, se tomados alguns cuidados quando do lançamento, X segue uma distribuição binomial com $n=50$ e p.

P=1/2: moeda honesta.

Lançada a moeda, vamos supor que tenham ocorridos 36 caras. Esse resultado traz evidência de que a moeda seja honesta?

Para tomarmos uma decisão, podemos partir do princípio de que a moeda não favorece nem cara nem coroa, isto é, $p = 1/2$. Com essa informação e com o modelo binimial, podemos encontrar qual a probabilidade de se obterem 36 caras ou mais, e esse resultado nos ajudaria a tomar uma decisão. Suponha que a decisão foi rejeitar a "honestidade" da moeda: qual é a melhor estimativa para ***p***, baseando-se no resultado observado?

Dois problemas básicos da Inferência Estatística: teste de hipóteses e estimação.

Conceitos Básicos

Problemas de Inferência

Exemplo: Às vezes, o modelo teórico associado ao problema não é tão evidente. Por exemplo, a máquina de encher pacote de café automaticamente, digamos que ela esteja regulada para enchê-los segundo uma distribuição normal com média 500g e desvio padrão de 100g. Sabemos também que, às vezes, a máquina desregula-se e, quando isso acontece, o único parâmetro que se altera é a média, permanecendo a mesma variância. Para manter a produção sob controle, iremos colher uma amostra de 100 pacotes e pesá-los. Como essa amostra nos ajudará a tomar uma decisão?

Usarmos a média da amostra como informação pertinente para uma decisão. Mesmo que a máquina esteja regulada, dificilmente a média amostral será igual a 500g, dado que os pacotes apresentam certa variabilidade no peso. Se a média amostral não se afastar muito de 500g, não existirão razões para suspeitarmos da qualidade do procedimento de produção.

Problema: decidir o que é próximo ou distante de 500g.