Distribuições amostrais

Prof. Adriano Mendonça Souza, Dr.
Departamento de Estatística
PPGEMQ / PPGEP - UFSM

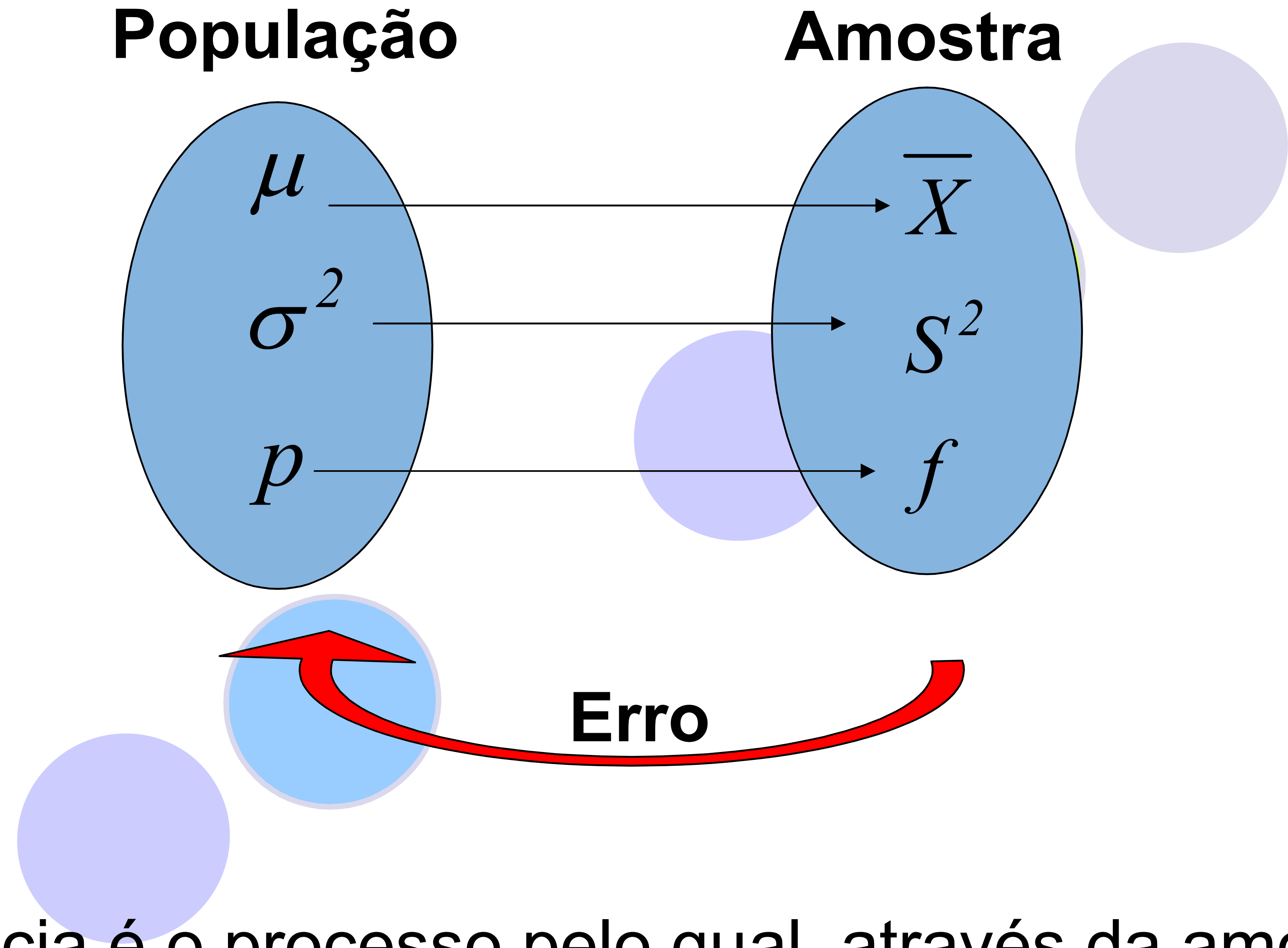
Noções básicas de Inferência

Estatística → **descritiva**

Estatística → **inferencial**

População - Parâmetros desconhecidos (reais)

Amostra - Estatísticas conhecidas (estimativas)



Inferência é o processo pelo qual, através da amostra, são estudadas as características da população.

Distribuições amostrais

População - É todo o conjunto completo de objetos nos quais estamos interessados. É a totalidade das observações possíveis sobre medidas ou ocorrências de determinada variável.

Tipos de população: Finita e Infinita.

Amostra - É a parte da população sendo investigada; é o conjunto das medidas ou sucessos selecionados da população. O princípio de seleção das amostras é importante.

Amostra aleatória de tamanho n - é um conjunto de realizações de "n" variáveis aleatórias independentes. Cada elemento da população tem a mesma probabilidade de ser selecionado.

Parâmetros da população - São os elementos (as quantidades) que caracterizam numericamente a população, isto é, especificam completamente a população. As características da amostra na forma de alguma medida global, chama-se Estatísticas.

Modelos Estatísticos Paramétricos - Para especificar um modelo estatístico precisamos particularizar a classe de modelos P das distribuições subjacentes possíveis, o espaço S dos resultados possíveis para o conjunto de dados disponíveis (observados ou mensurados experimentalmente).

Modelos Paramétricos - São uma classe de distribuições P($\underline{\theta}$), onde é um vetor de parâmetros característicos da distribuição $\underline{\theta}\ e\ \Omega$.

Ω é o espaço paramétrico do modelo, ou os possíveis estados da natureza.

Graus de liberdade

Em física, um ponto que pode mover-se livremente no espaço tri-dimensional têm 3 graus de liberdade. Três variáveis x, y, z podem tomar diferentes valores independentemente.

- Se o ponto se move no plano, ele têm apenas 2 G.L.

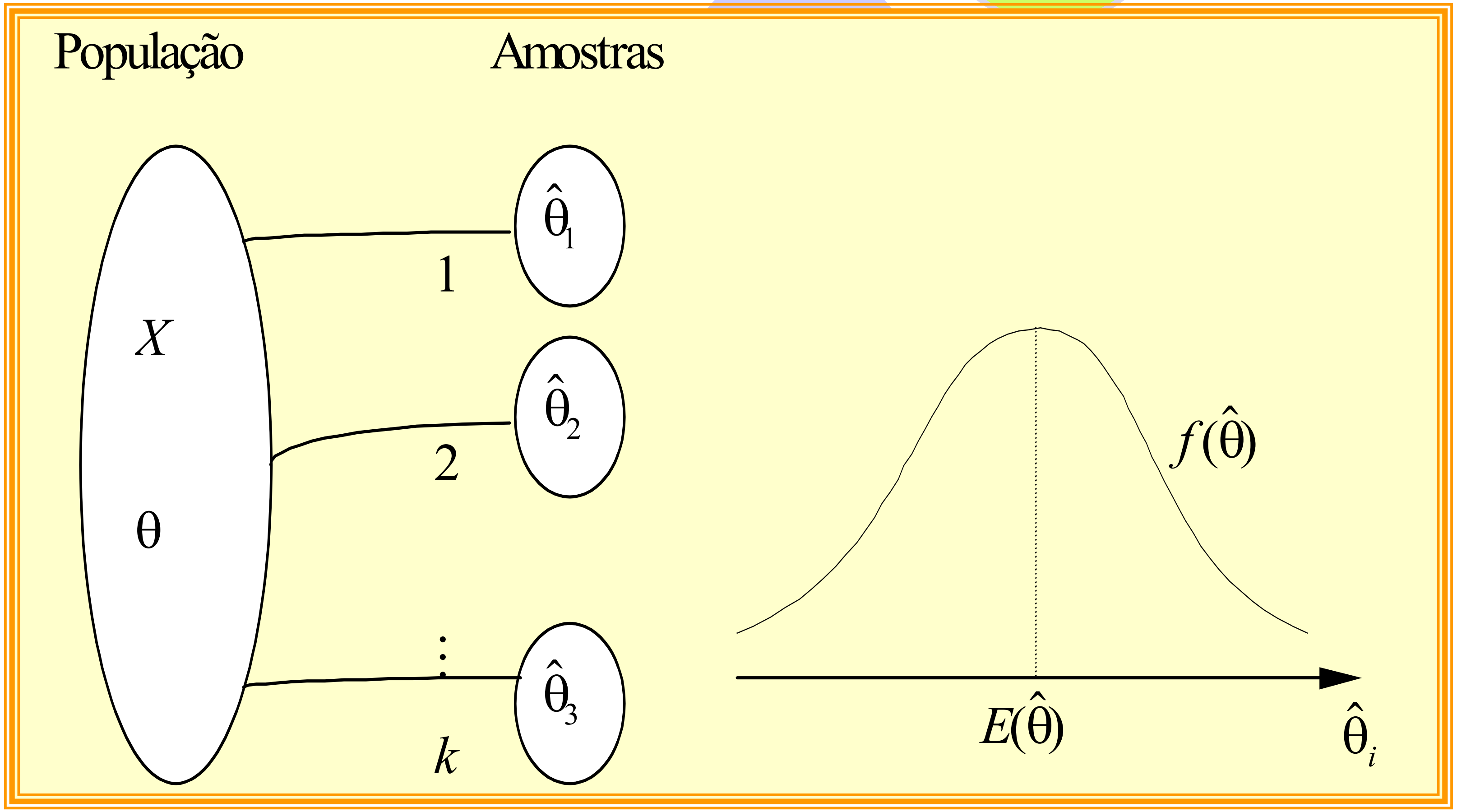
- Em estatística, a soma de *n* quadrados com relação à média amostral tem (n-1) G.L.

- Se está fixado, apenas *(n-1)* dos X's podem ser escolhidos independentemente, o n-ésimo estará automaticamente fixado.

- Usualmente, dadas "*n*" observações, perde-se tantos G.L. quantos são os parâmetros desconhecidos que são estimados.

A estatística que estima θ é comumente $\hat{\theta}$. E como diferentes amostras originam valores distintos de $\hat{\theta}$, $\hat{\theta}$ é, ele próprio, uma variável aleatória. Nesse último caso temos uma ***Distribuição Amostral*** da estatística $\hat{\theta}$.

O gráfico seguinte ilustra tal distribuição onde para cada amostra $\hat{\theta}_i$, calculamos o valor da estatística $\hat{\theta}$. Os valores $\hat{\theta}_i$ formam uma $\hat{\theta}$ nova população, a Distribuição Amostral da estatística. .



Estatística

É uma função de uma ou mais variáveis aleatórias que não dependem de parâmetros desconhecidos. É uma função das observações ou dados contidos na amostra.

- Média amostral ($\overline{X}$);
- Máximo da amostra (y_n ou $x_{máx}$);
- Amplitude da amostra (r);
- Variância amostral (S^2).

Se nós repetirmos o experimento e calcularmos, a partir de várias amostras, o valor da estatística $\mu(x_1,\ldots,x_n)$, teremos a distribuição amostral da estatística $\mu(x_1,\ldots,x_n)$.

Distribuições por amostragem

Considerar todas as possíveis amostras de tamanho *n* retiradas da população. Para cada amostra calcular a estatística de interesse, obtendo-se desta maneira uma distribuição desses resultados originando uma distribuição por amostragem. Desta maneira pode-se obter as distribuições por amostragem da média, da variância, da proporção e de outras estatísticas.

Amostragem com ou sem reposição

- Se o processo de retirada for **com reposição**, N^n = número de amostras de tamanho *n* que poderão ser extraídas com reposição da população de tamanho N;

- Se o processo de retirada for **sem reposição**, o número combinatório C_N^n =número de amostras de tamanho *n* extraídas, sem reposição, da população de tamanho N.

Distribuição amostral das médias

Admita-se que todas as amostras possíveis de tamanho *n* são retiradas, **sem reposição**, de uma população finita de tamanho N.

Distribuição amostral das médias

Se a média e o desvio padrão da distribuição amostral das médias forem designados por

$$\boxed{\mu(\bar{x}) \ e \ \sigma(\bar{x})}$$

e os valores correspondentes da população o forem por μ e σ, respectivamente, então:

$$\boxed{\mu(\bar{x}) = \mu}$$

$$\boxed{\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

Distribuição amostral das médias

Se a população for infinita, ou se a amostragem for tomada **com reposição**, os resultados acima reduzir-se-ão a:

$$\boxed{\mu(\bar{x}) = \mu}$$

$$\boxed{\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}}}$$

Exemplos:

1) Consideremos uma população formada por 5 números: 2, 3, 6, 8, 11.

Considerar todas as amostras possíveis com 2 elementos cada uma, que possam ser retiradas desta população:

a) com reposição;

b) sem reposição.

Calcular a média da distribuição amostral das médias e o erro padrão.

2) Admite-se que as alturas de 3.000 estudantes do sexo masculino de uma universidade são normalmente distribuídas com média $\mu = 172{,}72$ cm e desvio padrão $\sigma = 7{,}62$ cm. Se forem obtidas 80 amostras de 25 estudantes cada uma, quais serão a média e o desvio padrão esperado da distribuição amostral das médias resultantes, se a amostragem for:

1º) Com reposição;

2º) Sem reposição.

Distribuição amostral da proporção

Seja X uma população infinita, onde:

p - representa o sucesso e;

q=1 - p - é o fracasso.

Como teremos "n" a.a dessa população, logo trabalharemos com uma v.a. binomial, com parâmetros:

$$\mu(x) = n.p \qquad v(x) = n.p.q$$

Logo a distribuição de freqüência relativa será: $f = \frac{x}{n} = \frac{n^o\, de\, elementos\;\, do\, evento}{tamanho\, da\, amostra}$, logo:

$$\boxed{\mu(f) = E[f] = \left[\frac{x}{n}\right] = \frac{np}{n} = p}$$

$$\boxed{\sigma^2(f) = \sigma^2\left[\frac{x}{n}\right] = \frac{1}{n^2}\sigma^2[x] = \frac{npq}{n^2} = \frac{pq}{n}}$$

Exemplo:

Seja uma população formada por 2 mulheres e 1 homem. Considerando-se sucesso a ocorrência de mulheres, retirar todas as amostras possíveis de tamanho 2 (n = 2), **com reposição**.