UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

MAT02215 - Estatística Geral 2 - 2024/2

Plano Aula 03 e 04

Markus Stein

Inferência Estatística

Essa semana veremos resultados e extensões de Probabilidade que terão aplicação nas próximas semanas.

- Estatística descritica $\times$ inferência estatística;
  - população e amostra: parâmetros (μ, σ^2, π, ...) $\times$ estatísticas ($\overline{x}$, s^2, p, ...).

Definição **Estatística**: é qualquer valor obtido em função da amostra. Exemplo, $\overline{x}$, s^2, p, ...

Distribuição amostral (Bussab e Morettin - Seção 10.7)

"Toda função de variáveis aleatórias (v.a.s) é uma v.a."

Definição **Distribuição Amostral**: é a distribuição de probabilidade de uma estatística.

Exemplo 1: Seja X a v.a. que denota o número de livros que a população de monitores do curso 'Probabilidade e Estatística' lêem por semestre. Suponha que no último semestre foram lidos 5, 7, 4. Se não soubéssemos essa informação e decidíssemos observar uma amostra de tamanho $n = 2$ para saber a média de livros lidos $\overline{X}$ dessa população.
a. Quais as possíveis amostras? (Cada amostra pode gerar um $\overline{x}$ diferente)
b. Os valores de média calculados com cada amostra formam a distribuição amostral de $\overline{X}$.

Lembrando: Amostra aleatória simples (a.a.s.) = v.a. idependentes e identicamente distribuídas (i.i.d.)

Definição **A.A.S**: Seja $X_1, X_2, \ldots, X_n$ uma a.a.s. de tamanho n de $X \sim f(x;\theta)$, então $X_1 \sim f(x;\theta)$, ..., $X_n \sim f(x;\theta)$ e X_i e X_j são independentes para todo $i \neq j$.

Definição **Erro padrão**: é o desvio padrão de uma estatística. Exemplo, erro padrão da *média amostral* é $\sigma_{\overline{X}} = \sqrt{Var(\overline{X})}$.

Exemplo 2: Proporção amostral $p = \frac{1}{n}\sum_{i=1}^{n} X_i$, para $X_i \in \{0, 1\}$. ...

Teorema central do limite - TCL (Bussab e Morettin - Seção 10.8)

"Garante que uma média amostral se aproxima do seu valor esperado à medida que o tamanho da amostra aumenta (dadas algumas condições...)"

- **Teorema 10.2** e **Corolário 10.1**
- Aplicativo que ilustra o TCL - https://brunamdalmoro.shinyapps.io/TCL_medias/

Estmação Pontual (Bussab e Morettin - Capítulo 11)

- Estatísticas: Estimador *versus* Estimativa.

(cont.) Exemplo 1: Média amostral, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, em que $X_1, \ldots, X_n$ uma amostra aleatória de $X_i \sim Normal(\mu, \sigma^2)$ e σ^2 conhecido:
a. Qual a distribuição amostral de $\overline{X}$?
b. $\overline{X}$ é um bom estimador para a média populacional μ?
c. Como usar $Var(\overline{X})$ para fornecer um grau de certeza sobre usarmos $\overline{X}$ para representar/estimar μ?

- se $X \sim Normal$ então $\overline{X} \sim Normal\left(\mu, \frac{\sigma^2}{n}\right)$! Suposições?

Definição (**Estimador**): Um estimador T do parâmetro θ é qualquer função das observações da amostra, $T = g(X_1, \ldots, X_n)$.

Definição (**Estimativa**): Uma estimativa é um particular valor do estimador. Para uma amostra observada $x_1, \ldots, x_n$ uma estimativa t do parâmetro θ é dada por $t = g(x_1, \ldots, x_n)$.

(cont.) Exemplo 1: E para a média amostral $\widehat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ se σ^2 **desconhecido**?

- vale o resultado $\overline{X} \sim Normal\left(\mu, \frac{\sigma^2}{n}\right)$? (Suposições?)
  - se n for "pequeno" e $X \sim Normal$, então X não segue uma $Normal$;
  - se n for "grande", então pelo TCL teremos que $\overline{X} \sim Normal$.

(cont.) Exemplo 2: E para a proporção amostral $\widehat{\pi} = P = \frac{1}{n}\sum_{i=1}^{n} X_i$?

- quando $P \sim Normal\left(\pi, \frac{\pi(1-\pi)}{n}\right)$? Suposições?

Exemplo 3: Simulação de distribuições de estimadores (*estatísticas*)
a. Para a variância amostral $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - E(X))^2$? E para $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - E(X))^2$?
b. E para outras estatísticas como a mediana, Md, o máximo, $X_{(n)}$, ou o mínimo, $X_{(1)}$, ...?

Propriedades dos estimadores (Bussab e Morettin - Seção 11.2)

(cont.) Exemplo 1: ...

- **Viés** e o Erro Quadrático Médio (EQM)

- **Constistência**

- **Eficiência**

Ler slides das aulas 3 a 4

Fazer exercícios lista 1-2

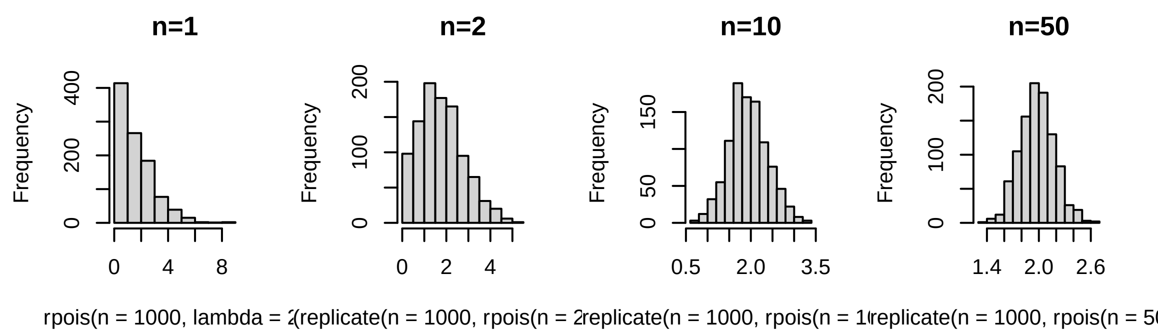
Fazer avaliação pontual 1 da área 1 - vale nota!!!

Ilustração do TCL no R

No R é possível gerar amostras, calcular a média de cada amostra e plotar o histograma:
(*usamos* `replicate` *para gerar 1000 amostras de tamanho* n = *1, 2, 10 e 50*)

- V.a. $X \sim Poisson(2)$. (para $n = 1$ geramos amostra da **distribuição real** de X)

```r
par(mfrow=c(1,4))
hist( rpois( n = 1000, lambda = 2), main = "n=1")
hist( colMeans( replicate( n = 1000, rpois( n = 2, lambda = 2))), main="n=2")
hist( colMeans( replicate( n = 1000, rpois( n = 10, lambda = 2))), main="n=10")
hist( colMeans( replicate( n = 1000, rpois( n = 50, lambda = 2))), main="n=50")
```



- V.a. de $X \sim Uniforme(0,1)$.

```r
par(mfrow=c(1,4))
hist( replicate( n = 1000, runif( n = 1, min = 0, max = 1)), main="n=1")
hist( colMeans( replicate( n = 1000, runif( n = 2, min = 0, max = 1))), main="Uniforme(0,1), n=2")
hist( colMeans( replicate( n = 1000, runif( n = 10, min = 0, max = 1))), main="Uniforme(0,1), n=10")
hist( colMeans( replicate( n = 1000, runif( n = 50, min = 0, max = 1))), main="Uniforme(0,1), n=50")
```