Estimação: (A) Propriedades e Distribuições Amostrais

Wagner H. Bonat
Fernando P. Mayer
Elias T. Krainski

Universidade Federal do Paraná
Departamento de Estatística
Laboratório de Estatística e Geoinformação

01/05/2018

DEST/UFPR

Sumário

Inferência estatística

Seja X uma variável aleatória com função densidade (ou de probabilidade) denotada por $f(x, \theta)$, em que θ é um parâmetro desconhecido. Chamamos de **inferência estatística** o problema que consiste em especificar um ou mais valores para θ, baseado em um conjunto de valores X.

A inferência pode ser feita de duas formas:

- estimativa pontual
- estimativa intervalar

Redução de dados

Um experimentador usa as informações em uma amostra aleatória $X_1, \dots, X_n$ para se fazer inferências sobre θ.

Normalmente n é grande e fica inviável tirar conclusões baseadas em uma longa **lista** de números.

Por isso, um dos objetivos da inferência estatística é **resumir** as informações de uma amostra, da maneira mais **compacta** possível, mas que ao mesmo tempo seja também **informativa**.

Normalmente esse resumo é feito por meio de **estatísticas**, por exemplo, a média amostral e a variância amostral.

População e amostra

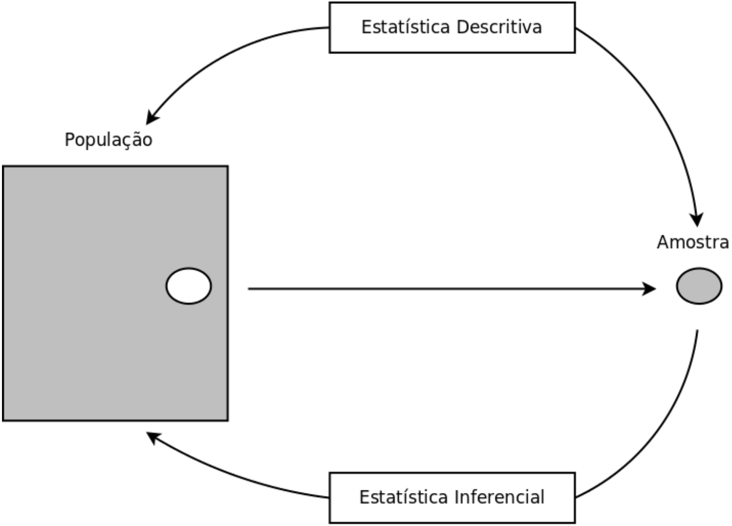
O conjunto de valores de uma característica associada a uma coleção de indivíduos ou objetos de interesse é dito ser uma população.

Uma sequência $X_1, \dots, X_n$ de n variáveis aleatórias independentes e identicamente distribuídas (iid) com função densidade (ou de probabilidade) $f(x, \theta)$ é dita ser uma amostra aleatória de tamanho n da distribuição de X.

Como normalmente $n > 1$, então temos que a fdp ou fp conjunta será

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = f(x_1, \dots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta).$$

População e amostra



Sumário

Parâmetro e Estatística

População $\rightarrow$ censo $\rightarrow$ parâmetro

*Uma medida numérica que descreve alguma característica da **população**, usualmente representada por letras gregas:* θ, μ, σ, ...

Exemplo: média populacional $= \mu$

População $\rightarrow$ amostra $\rightarrow$ estatística

*Uma medida numérica que descreve alguma característica da **amostra**, usualmente denotada pela letra grega do respectivo parâmetro com um acento circunflexo:* $\hat{\theta}$, $\hat{\mu}$, $\hat{\sigma}$, ..., *ou por letras do alfabeto comum:* $\bar{x}$, s, ...

Exemplo: média amostral $= \bar{x}$

Parâmetros

É importante notar que um parâmetro não é restrito aos modelos de probabilidade. Por exemplo:

- $X \sim \text{N}(\mu, \sigma^2) \Rightarrow$ parâmetros: μ, σ^2
- $Y \sim \text{Poisson}(\lambda) \Rightarrow$ parâmetro: λ
- $Y = \beta_0 + \beta_1 X \Rightarrow$ parâmetros: β_0, β_1
- $L_t = L_\infty[1 - e^{-k(t-t_0)}] \Rightarrow$ parâmetros: L_∞, k, t_0

Estatística

Qualquer função da amostra que não depende de parâmetros desconhecidos é denominada uma estatística, denotada por $T(\mathbf{X}) = T(X_1, X_2, \ldots, X_n)$

Exemplos:

- $T_1(\mathbf{X}) = \sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n$
- $T_2(\mathbf{X}) = \prod_{i=1}^{n} X_i = X_1 \cdot X_2 \cdots X_n$
- $T_3(\mathbf{X}) = X_{(1)}$
- $T_4(\mathbf{X}) = \sum_{i=1}^{n} (X_i - \mu)^2$

Estatística

Qualquer função da amostra que não depende de parâmetros desconhecidos é denominada uma estatística, denotada por $T(\mathbf{X}) = T(X_1, X_2, \ldots, X_n)$

Exemplos:

- $T_1(\mathbf{X}) = \sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n$
- $T_2(\mathbf{X}) = \prod_{i=1}^{n} X_i = X_1 \cdot X_2 \cdots X_n$
- $T_3(\mathbf{X}) = X_{(1)}$
- $T_4(\mathbf{X}) = \sum_{i=1}^{n} (X_i - \mu)^2$

Verificamos que T_1, T_2, T_3 são estatístcas, mas T_4 não.

Como é uma função da amostra, então uma estatística também é uma **variável aleatória** $\rightarrow$ distribuições amostrais

Estimador

Espaço paramétrico

O conjunto Θ em que θ pode assumir seus valores é chamado de **espaço paramétrico**

Estimador

Qualquer estatística que assume valores em Θ é um estimador para θ.

Estimador pontual

Dessa forma, um **estimador pontual** para θ é qualquer estatística que possa ser usada para estimar esse parâmetro, ou seja,

$$\hat{\theta} = T(\mathbf{X})$$

Estimador

Observações:

1. Todo estimador é uma estatística, mas nem toda estatística é um estimador.
2. O valor assumido pelo estimador pontual é chamado de **estimativa pontual**,

$$\hat{\theta} = T(\mathbf{X}) = T(X_1, \ldots, X_n) = t$$

ou seja, o estimador é uma **função** da amostra, e a estimativa é o **valor observado** de um estimador (um número) de uma amostra particular.

Estimação pontual

A ideia geral por trás da estimação pontual é muito simples:

Quando a amostragem é feita a partir de uma população descrita por uma função $f(x, \theta)$, o conhecimento de θ a partir da amostra, gera todo o conhecimento para a população.

Dessa forma, é natural que se procure um **método** para se achar um **bom** estimador para θ.

Existem algumas **propriedades** que definem o que é um bom estimador, ou o "**melhor**" estimador entre uma série de candidatos.

Estimação pontual

Localização do problema: Considere $X_1, \ldots, X_n$ uma amostra aleatóra de uma variável aleatória X com fdp ou fp $f(x, \theta)$, $\theta \in \Theta$. Sejam:

$$\hat{\theta}_1 = T_1(X_1, ..., X_n) \qquad \hat{\theta}_2 = T_2(X_1, ..., X_n)$$

Qual dos dois estimadores pontuais é **melhor** para θ?

Como não conhecemos θ, não podemos afirmar que $\hat{\theta}_1$ é melhor do que $\hat{\theta}_2$ e vice-versa.

O problema da estimação pontual é então escolher um estimador $\hat{\theta}$ que se aproxime de θ segundo algumas **propriedades**.

Estimação pontual

Exemplo 1: Considere uma amostra aleatória $(X_1, \ldots, X_n)$ de uma variável aleatória $X \sim \mathrm{N}(\mu = 3, \sigma^2 = 1)$ e os estimadores pontuais para μ

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{e} \qquad \hat{\theta}_2 = \frac{X_{(1)} + X_{(n)}}{2}$$

Qual dos dois estimadores pode ser considerado como o **melhor** para estimar o verdadeiro valor de μ?

Considere os seguintes pseudo-códigos para um estudo de simulação do comportamento destes dois estimadores:

Estimação pontual

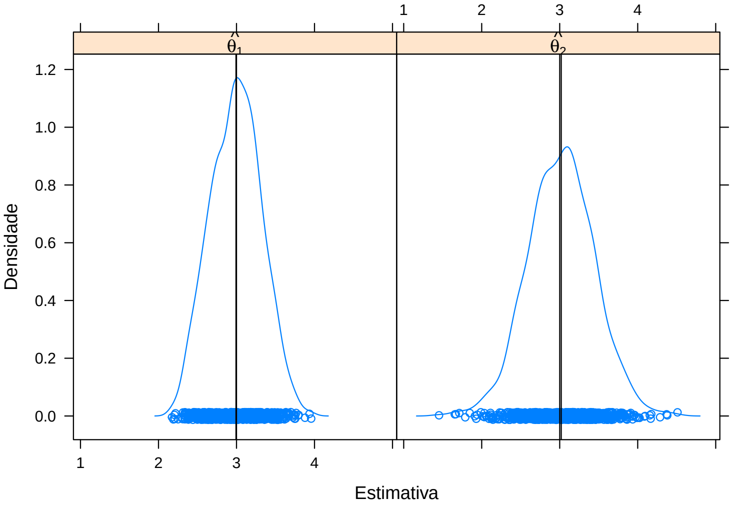
Pseudo-código 1

- Simule uma amostra de tamanho $n = 10$ da distribuição considerada
- Para essa amostra, calcule a média ($\hat{\theta}_1$) e o ponto médio ($\hat{\theta}_2$)
- Repita os passos (1) e (2) acima $m = 1000$ vezes
- Faça um gráfico da densidade das $m = 1000$ estimativas de $\hat{\theta}_1$ e $\hat{\theta}_2$ e verifique seu comportamento

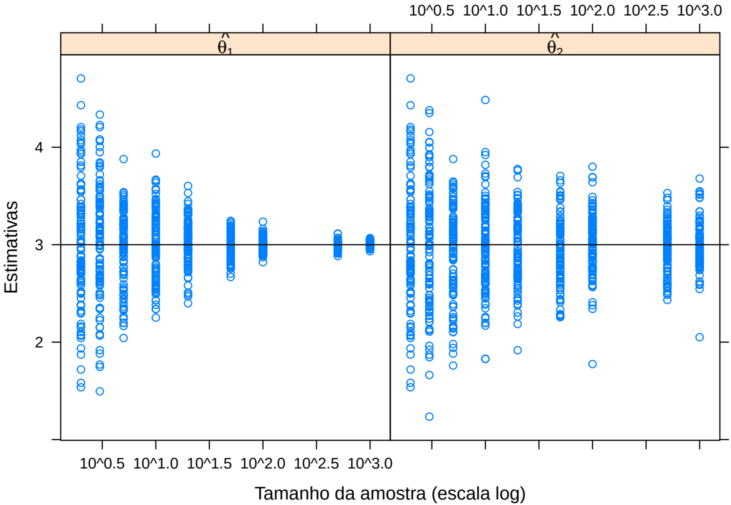
Pseudo-código 2

- Simule amostras de tamanhos (n) 2, 3, 5, 10, 20, 50, 100, 500, 1000 da distribuição considerada
- Para cada amostra de tamanho n, calcule a média ($\hat{\theta}_1$) e o ponto médio ($\hat{\theta}_2$)
- Repita os passos (1) e (2) acima $m = 100$ vezes
- Faça um gráfico das $m = 100$ estimativas de $\hat{\theta}_1$ e $\hat{\theta}_2$ para cada tamanho de amostra n e verifique seu comportamento

Estimação pontual: Pseudo-código 1 - $X \sim N(3, 1)$



Estimação pontual: Pseudo-código 2 - $X \sim N(3, 1)$



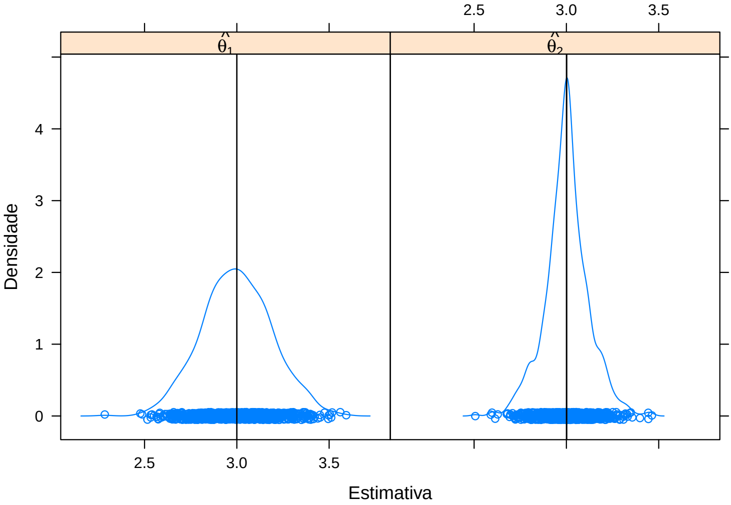
Estimação pontual

Exemplo 2: Considere uma amostra aleatória $(X_1, \ldots, X_n)$ de uma variável aleatória $Y \sim \mathrm{U}(\min = 2, \max = 4)$ (distribuição uniforme no intervalo [2,4]) e os estimadores pontuais para μ

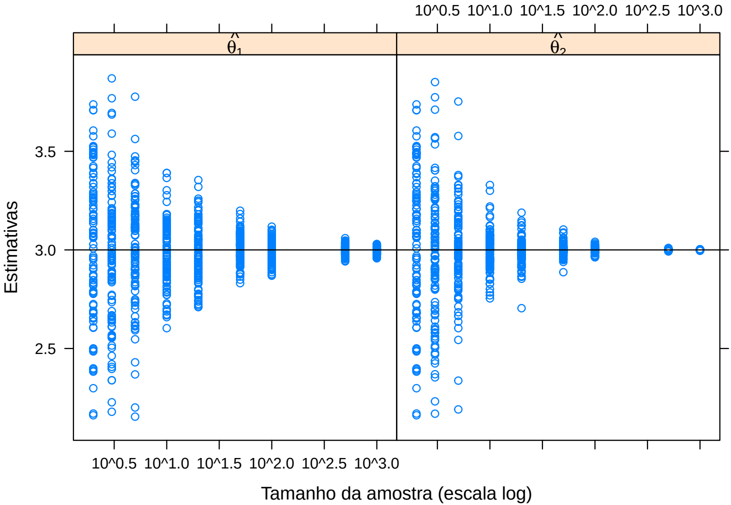
$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{e} \qquad \hat{\theta}_2 = \frac{X_{(1)} + X_{(n)}}{2}$$

Qual dos dois estimadores pode ser considerado como o **melhor** para estimar a média de Y?

Estimação pontual: Pseudo-código 1 - $Y \sim U(2, 4)$



Estimação pontual: Pseudo-código 2 - $Y \sim U(2, 4)$



Propriedades dos estimadores

De modo geral, um "**bom**" estimador deve ser:

1. Não viciado
2. Consistente
3. Eficiente

Propriedades dos estimadores: 1. Vício

Erro quadrático médio (EQM)

O Erro Quadrático Médio (EQM) de um estimador $\hat{\theta}$ de θ é dado por

$$\begin{aligned} \text{EQM}[\hat{\theta}] &= \text{E}[(\hat{\theta} - \theta)^2] \\ &= \text{Var}[\hat{\theta}] + \text{B}[\hat{\theta}]^2 \end{aligned}$$

onde

$$\text{B}[\hat{\theta}] = \text{E}[\hat{\theta}] - \theta$$

é denominado de **vício** do estimador $\hat{\theta}$. Portanto, dizemos que um estimador é **não viciado** para θ quando

$$\text{B}[\hat{\theta}] = 0 \quad \Rightarrow \quad \text{E}[\hat{\theta}] = \theta$$

Propriedades dos estimadores: 1. Vício

Estimador não viciado

Seja $(X_1, \ldots, X_n)$, uma amostra aleatória de uma variável aleatória com fdp ou fp $f(x, \theta)$, $\theta \in \Theta$, dizemos que o estimador $\hat{\theta} = T(\mathbf{X})$ é não viciado para θ se

$$\mathrm{E}[\hat{\theta}] = \mathrm{E}[T(\mathbf{X})] = \theta \qquad \forall\, \theta \in \Theta$$

Um estimador $\hat{\theta}$ é dito **assintoticamente não viciado** se

$$\lim_{n \to \infty} \mathrm{E}[\hat{\theta}] = \theta$$

Ou seja, para grandes amostras, $\hat{\theta}$ passa a ser imparcial.

Propriedades dos estimadores: 2. Consistência

Estimador consistente

Seja $(X_1, \ldots, X_n)$, uma amostra aleatória de uma variável aleatória com fdp ou fp $f(x, \theta)$, $\theta \in \Theta$, o estimador $\hat{\theta} = T(\mathbf{X})$ é consistente para θ se satisfaz simultaneamente

$$\lim_{n \to \infty} \mathrm{E}[\hat{\theta}] = \theta$$

e

$$\lim_{n \to \infty} \mathrm{Var}[\hat{\theta}] = 0$$

Propriedades dos estimadores: 3. Eficiência

Eficiência relativa

Sejam $\hat{\theta}_1 = T_1(\mathbf{X})$ e $\hat{\theta}_2 = T_2(\mathbf{X})$ dois estimadores pontuais **não viciados** para θ. A eficiência relativa de $\hat{\theta}_1$ em relação a $\hat{\theta}_2$ é

$$\mathrm{ER}[\hat{\theta}_1, \hat{\theta}_2] = \frac{\mathrm{Var}[\hat{\theta}_1]}{\mathrm{Var}[\hat{\theta}_2]}$$

Se:

- $\mathrm{ER}[\hat{\theta}_1, \hat{\theta}_2] > 1 \Rightarrow \hat{\theta}_2$ é mais eficiente
- $\mathrm{ER}[\hat{\theta}_1, \hat{\theta}_2] < 1 \Rightarrow \hat{\theta}_1$ é mais eficiente

Propriedades dos estimadores

Exemplo: média amostral $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ como estimador da média populacional μ:

$$E(\bar{x}) = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \mu$$

$$\mathrm{Var}(\bar{x}) = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{\sigma^2}{n}$$

Portanto $\bar{x}$ é um estimador **não viciado** e **consistente** para μ.

Propriedades dos estimadores

Exemplo: variância amostral $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ como estimador da variância populacional σ^2:

$$\mathrm{E}(\hat{\sigma}^2) = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \left(\frac{n-1}{n}\right)\sigma^2$$

Portanto $\hat{\sigma}^2$ é um estimador **viciado** para σ^2. (Embora seja um estimador **assintoticamente** não viciado).

Para eliminar esse vício, podemos definir então um novo estimador: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, e

$$\mathrm{E}(S^2) = \mathrm{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \sigma^2$$

que é então um estimador **não viciado** para σ^2.

Exemplo 7.11

Uma amostra $(X_1, \ldots, X_n)$ é retirada de uma população com $X \sim \mathrm{N}(\mu, \sigma^2)$, e dois estimadores são propostos para μ:

$$\hat{\mu}_1 = \bar{X} \quad \text{e} \quad \hat{\mu}_2 = \text{mediana}(X_1, \ldots, X_n)$$

Qual dos dois é melhor para μ?

Exemplo 7.11

Podemos notar que

$$\mathrm{E}(\hat{\mu}_1) = \mathrm{E}(\bar{X}) = \mu$$
$$\mathrm{Var}(\hat{\mu}_1) = \mathrm{Var}(\bar{X}) = \sigma^2/n$$

$$\mathrm{E}(\hat{\mu}_2) = \mathrm{E}(\mathrm{mediana}(X_1, \ldots, X_n)) = \mu$$
$$\mathrm{Var}(\hat{\mu}_2) = \mathrm{Var}(\mathrm{mediana}(X_1, \ldots, X_n)) = (\pi/2)(\sigma^2/n)$$

Portanto, ambos são estimadores não viciados e consistentes. Mas:

$$\mathrm{ER}[\hat{\mu}_1, \hat{\mu}_2] = \frac{\mathrm{Var}[\hat{\mu}_1]}{\mathrm{Var}[\hat{\mu}_2]} = \frac{\sigma^2/n}{(\pi/2)(\sigma^2/n)} = \frac{2}{\pi} = 0,63$$

Como $\mathrm{ER}[\hat{\mu}_1, \hat{\mu}_2] < 1$ então $\hat{\mu}_1 = \bar{X}$ é mais **eficiente**.

Propriedades dos estimadores

O **erro padrão** de um estimador dá uma ideia da **precisão** da estimativa.

O erro padrão (EP) de um estimador é o seu desvio-padrão (raíz quadrada da variância), ou seja,

$$\mathrm{EP}(\hat{\theta}) = \sqrt{\mathrm{Var}(\hat{\theta})}$$

Exemplo: Sabemos que a distribuição de $\bar{X}$ tem média μ e variância σ^2/n. Então o erro padrão de $\bar{X}$ é

$$\mathrm{EP}(\bar{X}) = \sqrt{\mathrm{Var}(\bar{X})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Sumário

Distribuições amostrais

De maneira geral, uma amostra de tamaho n será descrita pelos valores $x_1, x_2, \ldots, x_n$ das variáveis aleatórias $X_1, X_2, ..., X_n \Rightarrow$ **Amostra Aleatória**

No caso de uma Amostragem Aleatória Simples (AAS) **com reposição**, $X_1, X_2, \ldots, X_n$ serão variáveis aleatórias **independentes e identicamentes distribuídas** (iid) com função de probabilidade (fp) ou função densidade de probabilidade (fdp) conjunta dada por

$$f(x_1, x_2, \ldots, x_n, \boldsymbol{\theta}) = f(x_1, \boldsymbol{\theta}) \cdot f(x_2, \boldsymbol{\theta}) \cdots f(x_n, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i, \boldsymbol{\theta})$$

Onde o mesmo valor do parâmetro $\boldsymbol{\theta}$ é utilizado em cada um dos termos no produto.

Distribuições amostrais

Quando uma amostra $X_1, X_2, \ldots, X_n$ é obtida, geralmente estamos interessados em um resumo destes valores, que pode ser expresso matematicamente pela estatística $T(x_1, x_2, \ldots, x_n)$.

Dessa forma, $Y = T(x_1, x_2, \ldots, x_n)$ **é também uma variável aleatória**. Se Y é uma VA, então ela possui uma **distribuição de probabilidade**.

Uma vez que a distribuição de Y é derivada da amostra $X_1, X_2, \ldots, X_n$, vamos denominá-la de **distribuição amostral** de Y.

Distribuições amostrais

Distribuições amostrais

A distribuição de probabilidade de uma estatística $Y = T(x_1, x_2, \ldots, x_n)$ é denominada de **distribuição amostral** de Y. Assim, uma estatística também é uma variável aleatória, pois seus valores mudam conforme a amostra aleatória.

Exemplo: duas estatísticas comumente utilizadas para o resumo de uma amostra aleatória são a **média amostral** $\bar{x}$, e a **proporção amostral** $\hat{p}$. Cada uma delas também possui uma distribuição amostral.

Distribuição amostral da média

Para estudarmos a distribuição amostral da estatística $\bar{X}$, considere uma população identificada pela VA X, com parâmetros

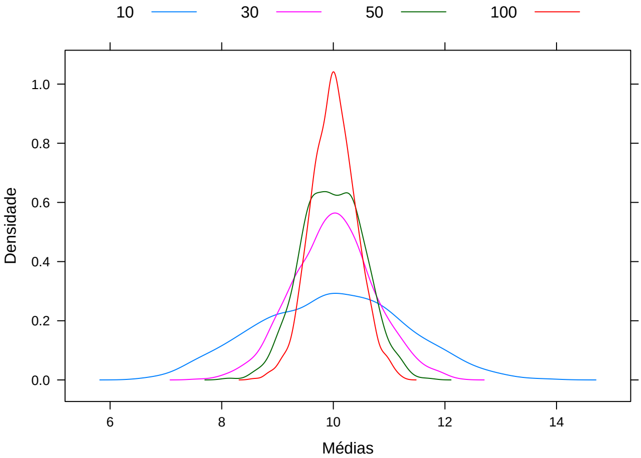
$$\text{E}(X) = \mu = \text{média} \qquad \text{Var}(X) = \sigma^2 = \text{variância}$$

supostamente conhecidos. Em seguida, realizamos os seguintes passos:

1. Retiramos m amostras aleatórias (AAS com reposição) de tamanho n dessa população
2. Para cada uma das m amostras, calculamos a média amostral $\bar{x}$
3. Verificamos a distribuição das m médias amostrais e estudamos suas propriedades

Exemplo 7.13

Seja $X \sim \mathrm{N}(10, 16)$, como se comporta $\bar{X}$ para $n = 10, 30, 50, 100$?



Distribuição amostral da média

Através do estudo da distribuição da média amostral chegamos em um dos resultados mais importantes da inferência estatística.

Distribuição amostral da média

- $\text{E}(\bar{X}) = \mu_{\bar{X}} = \mu$
- $\text{Var}(\bar{X}) = \sigma^2_{\bar{X}} = \sigma^2/n$

Portanto, se

$$X \sim \text{N}(\mu, \sigma^2) \quad \text{então} \quad \bar{X} \sim \text{N}(\mu_{\bar{X}}, \sigma^2_{\bar{x}})$$

mas, como

$$\mu_{\bar{X}} = \mu \quad \text{e} \quad \sigma^2_{\bar{X}} = \sigma^2/n$$

então, a **distribuição amostral** da média amostral $\bar{X}$ é

$$\bar{X} \sim \text{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Distribuição amostral da média

Pode-se mostrar que, para amostras suficientemente grandes, **a média amostral $\bar{X}$ converge para o verdadeiro valor da média populacional** μ (é um **estimador não viesado** de μ).

Além disso, a variância das médias amostrais $\sigma^2_{\bar{X}}$ tende a diminuir conforme $n \to \infty$ (é um estimador **consistente**).

Estes resultados sugerem que, quando o tamanho da amostra aumenta,

<u>independente do formato da distribuição da população original</u>,

a distribuição amostral de $\bar{X}$ aproxima-se cada vez mais de uma distribuição Normal, um resultado fundamental na teoria de probabilidade conhecido como **Teorema Central do Limite**.

Distribuição amostral da média

Teorema Central do Limite (TCL)

Para amostras aleatórias simples $(X_1, X_2, \dots, X_n)$, retiradas de uma população com média μ e variância σ^2, a distribuição amostral da média $\bar{X}$, terá forma dada por

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

no limite quando $n \to \infty$, onde $Z \sim \mathrm{N}(0, 1)$.

- Se a população for normal, então $\bar{X}$ terá distribuição *exata* normal.
- A rapidez da convergência para a normal depende da distribuição da população da qual as amostras foram geradas.

Distribuição amostral da média

Ver figura dist_amostrais.pdf

Distribuição amostral da média

Em palavras, o teorema garante que que para n grande, a distribuição da média amostral, devidamente padronizada, **se comporta segundo um modelo normal** com média 0 e variância 1.

Pelo teorema, temos que quanto maior o tamanho da amostra, **melhor é a aproximação**.

Estudos envolvendo simulações mostram que, em muitos casos, **valores de n ao redor de 30** fornecem aproximações bastante boas para as aplicações práticas.

Distribuição amostral da média

Quando calculamos a probabilidade de um valor estar em um determinado intervalo de valores, podemos usar o modelo Normal, como vimos anteriormente.

No entanto, quando temos uma **amostra**, e queremos calcular probabilidades associadas à **média amostral** (a probabilidade da média amostral estar em um determinado intervalo de valores), precisamos necessariamente usar os resultados do TCL.

Distribuição amostral da média

Exemplo: Uma máquina de empacotamento que abastece pacotes de feijão apresenta distribuição normal com média de 500 g e desvio-padrão de 22 g. De acordo com as normas de defesa do consumidor, os pacotes de feijão não podem ter peso inferior a 2% do estabelecido na embalagem.

a. Determine a probabilidade de **um pacote** selecionado aleatoriamente ter a peso inferior a 490 g.
b. Determine a proabilidade de **20 pacotes** selecionados aleatoriamente terem peso médio inferior a 490 g.
c. Como podemos interpretar os resultados dos itens anteriores? O que é mais indicado para se tomar uma decisão sobre o funcionamento da máquina: selecionar um pacote ou uma amostra de pacotes?