Distribuição Amostral e Estimação Pontual de Parâmetros

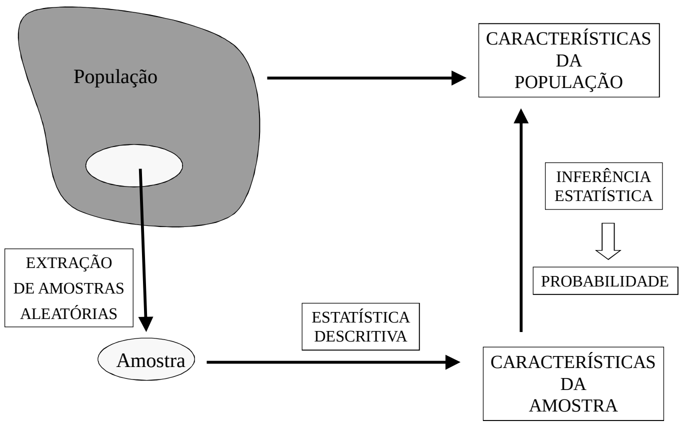
Roteiro

1. Introdução
2. Teorema Central do Limite
3. Conceitos de Estimação Pontual
4. Métodos de Estimação Pontual
5. Referências

Introdução

População e Amostra

- População:
  - √ Conjunto de elementos que apresentam pelo menos uma característica em comum
- População Alvo:
  - √ População de interesse da pesquisa
- Amostra:
  - √ Qualquer subconjunto não vazio da população



Técnicas de Amostragem

- Procedimento a ser adotado na seleção dos elementos da amostra
- O principal objetivo central é obter uma amostra representativa
  - √ Amostra que representa toda a população da melhor maneira possível
- A representatividade depende de:
  - √ Metodologia adotada para seleção da amostra
  - √ Tamanho da amostra

Problema Fundamental da Estatística

A partir da observação de amostras, COMO podemos tirar CONCLUSÕES sobre a POPULAÇÃO ?

Planejando um Experimento

- Identificar seu objetivo
- Coletar dados amostrais
- Usar procedimento aleatório para evitar vício
- Analisar dados e tirar conclusões

Erro Amostral

- Diferença entre um resultado amostral e o verdadeiro resultado populacional
  - √ são resultantes de flutuações amostrais aleatórias.

Erro Não–amostral

- Incorreção na coleta, registro ou análise de dados amostrais
  - √ Ex.
    - – Coleta tendenciosa de amostra
    - – Utilização de instrumento descalibrado
    - – Registro incorreto de dados amostrais

Inferência Estatística

- Definição:
  - √ Procedimentos generalizar características de população a partir da informação contida na amostra.
- Baseia-se na Teoria de Probabilidades
- Áreas:
  - √ Estimação de parâmetros
  - √ Testes de hipóteses.

Estimação de Parâmetros

- Estimação Pontual
- Estimação Intervalar
  - √ Intervalos de Confiança

Teste de Hipóteses

- Hipótese:
  - √ Afirmação (alegação) sobre característica populacional
- Teste de Hipóteses:
  - √ Procedimento padrão para se testar uma afirmativa sobre característica populacional

Conceitos Fundamentais

- Amostra aleatória:
  - √ As variáveis aleatórias X_1, X_2, ..., X_n são uma amostra aleatória de tamanho *n*, se:
    - Forem independentes
    - Cada X_i tiver mesma distribuição de probabilidades

- Parâmetro:
  - √ Quantidades de interesse da população
  - √ Em geral, desconhecidas
    - Média de uma população (μ)
    - Desvio-padrão de uma população (σ)
  - √ Representadas por letras gregas
  - √ Notação para estimador qualquer: θ

- Estatística:
  - √ Qualquer função da amostra que não dependa de parâmetros desconhecidos
  - √ Exemplo : Algumas estatísticas da amostra aleatória X_1, X_2, ..., X_n:

$X_{(1)} = \text{mín}(X_1, X_2, ..., X_n)$

$X_{(n)} = \text{máx}(X_1, X_2, ..., X_n)$

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1}$$

- Distribuição amostral:
  - √ Distribuição de probabilidades de uma estatística
  - √ Exemplo:
    - – Distribuição amostral da média
    - – Parâmetros da distribuição amostral da média

$$\mathrm{E}\left(\overline{X}_n\right) = \mu_X$$

$$\mathrm{Var}\left(\overline{X}_n\right) = \frac{\sigma_X^2}{n}$$

- Espaço paramétrico (Θ)
  - √ Conjunto em que o parâmetro θ toma valores
  - √ Exemplo: Seja a amostra aleatória $X_1, X_2, ..., X_n$ da variável $X \sim N(\mu, \sigma^2)$
    - – Se $\sigma^2=1$, então $\theta = \mu$ é o parâmetro desconhecido e

      $\Theta = \{\mu, -\infty < \mu < \infty\}$
    - – Se $\mu = 0$, então $\theta = \sigma^2$ é o parâmetro desconhecido e

      $\Theta = \{\sigma^2, \sigma^2 > 0\}$

- Estimador de θ
  - √ Qualquer estatística que assuma valores em Θ.
  - √ Notação: $\hat{\theta}$
  - √ Exemplo:

    Alguns estimadores para a média μ de uma população
    - – Média da amostra
    - – Mediana da amostra
    - – X_1
    - – Etc.

- Estimativa de parâmetro populacional:
  - √ é um valor específico, ou um intervalo de valores, usado para estimar parâmetro populacional
- Estimativa pontual:
  - √ é um único valor numérico de uma estatística $\hat{\theta}$

Teorema Central do Limite

Teorema Central do Limite

- Seja X_1, X_2, ..., X_n uma amostra aleatória de tamanho *n* de uma população (finita ou infinita), com média μ e variância finita σ^2. Então

$$Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathrm{N}(0,1)$$

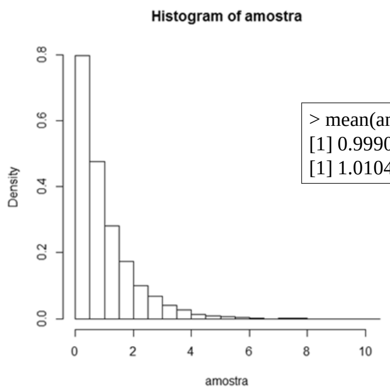
quando $n \to \infty$.

- Comentários:
  - √ A aproximação normal para a média amostral depende do tamanho da amostra
  - √ Com população contínua, unimodal e simétrica, na maioria dos casos, o TCL trabalha bem para pequenas amostras (n = 4, 5).
  - √ Em muitos casos de interesse prático, a aproximação normal será satisfatória para $n \geq 30$
  - √ Se $n < 30$, o TCL funcionará se a distribuição da população não for muito diferente da normal

Exemplo – Simulação

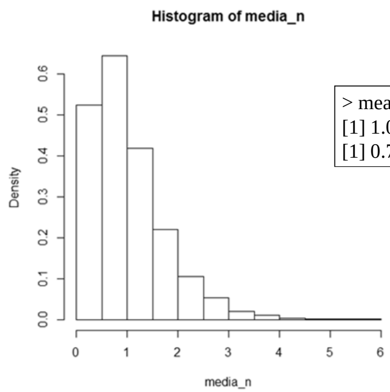
- População exponencial com média 1:
  - √ $\lambda = 1$
  - √ Geração de 10.000 valores dessa população
  - √ Amostra de tamanho 1 (*n* = 1)

- Amostra n =1



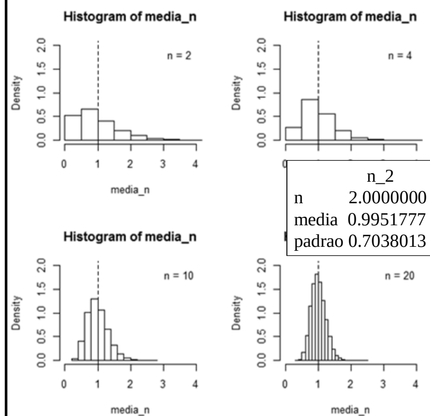
> mean(amostra); sd(amostra)
[1] 0.9990838
[1] 1.010478

- Amostra n = 2



> mean(media_n); sd(media_n)
[1] 1.012711
[1] 0.7129089

- Amostras de tamanhos 2, 4, 10 e 20

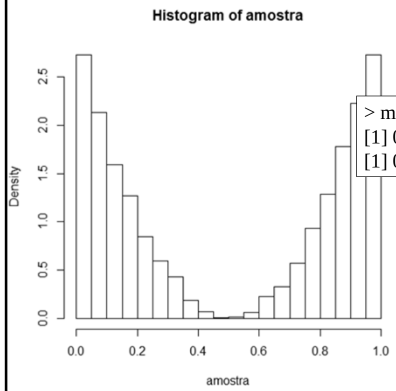


	n_2	n_4	n_10	n_20
n	2.0000000	4.0000000	10.0000000	20.0000000
media	0.9951777	1.0004503	0.9963200	0.9997876
padrao	0.7038013	0.5018562	0.3190797	0.2223573

Exemplo – Simulação

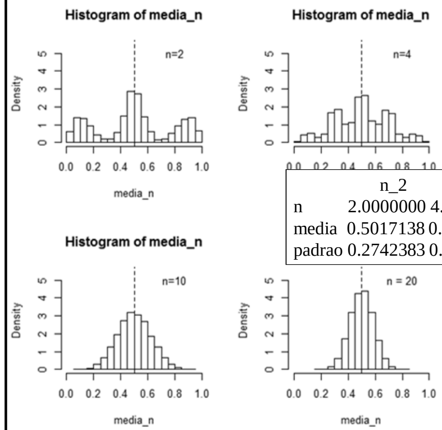
- População com densidade em U:
  - √ f(x) = 12 (x – 0,5)^2
  - √ Geração de 10.000 valores dessa população
  - √ Amostra de tamanho 1 (n = 1)

- Amostra n =1




```
> mean(amostra); sd(amostra)
[1] 0.5061657
[1] 0.3877259
```

- Amostras de tamanhos 2, 4, 10 e 20



	n_2	n_4	n_10	n_20
n	2.0000000	4.0000000	10.0000000	20.00000000
media	0.5017138	0.4986490	0.4994313	0.49993781
padrao	0.2742383	0.1957755	0.1221512	0.08599018

Comparação de Populações

- Considere duas populações:
  - √ População 1: média μ_1 e variância σ_1^2
  - √ População 2: média μ_2 e variância σ_2^2
- Amostras aleatórias das duas populações?
  - √ Amostra da população 1 de tamanho n_1: $\bar{X}_1$
  - √ Amostra da população 2 de tamanho n_2: $\bar{X}_2$

- Caso 1: As duas populações são normais
  - √ Distribuição amostral da diferença

$$\bar{X}_1 - \bar{X}_2 \sim \mathrm{N}(\mu_{\mathrm{dif}}, \sigma^2_{\mathrm{dif}})$$

  - √ Média da diferença de médias amostrais:

$$\mu_{\mathrm{dif}} = \mathrm{E}\left[\bar{X}_1 - \bar{X}_2\right] = \mathrm{E}\left[\bar{X}_1\right] - \mathrm{E}\left[\bar{X}_2\right] = \mu_1 - \mu_2$$

  - √ Variância da diferença de médias amostrais:

$$\sigma^2_{\mathrm{dif}} = \mathrm{Var}\left[\bar{X}_1 - \bar{X}_2\right] = \mathrm{Var}\left[\bar{X}_1\right] + \mathrm{Var}\left[\bar{X}_2\right] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

- Caso 2: populações não normais com tamanhos amostrais maiores que 30
  - √ Pode-se usar o TCL para aproximar a distribuição amostral da diferença:

Distribuição Amostral Aproximada de Diferença de Médias Amostrais

- Suponha:
  - √ Duas populações independentes, com médias μ_1 e μ_2 e variâncias σ_1^2 e σ_2^2
  - √ Amostras aleatórias independentes de tamanhos n_1 e n_2 dessas populações

$$Z = \frac{\bar{X}_{1,n_1} - \bar{X}_{2,n_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1)$$

se as condições do TCL se aplicarem

Exemplo – Vida de Motor

- Motor de turbina de aeronave a jato
  - √ Vida de componente é variável aleatória com média 5000 h e desvio-padrão 40 horas
  - √ Melhoria no componente: média 5050 h e desvio-padrão 30 horas
  - √ Suponha amostra de n_1 = 16 componente do processo antigo e n_2 = 25 do processo aprimorado
- Qual a probabilidade de que a diferença das média amostrais seja no mínimo 25 horas?

Estimação Pontual – Conceitos

Propriedades de um Estimador

- Alguma propriedades importantes:
  - √ Vício
  - √ Consistência
  - √ Eficiência

Vício

- Vício de um estimador:
  - √ Vício($\hat{\theta}$) = E[$\hat{\theta}$] – θ
- Um estimador $\hat{\theta}$ é não viciado (não viesado, não tendencioso) para um parâmetro θ se E[$\hat{\theta}$] = θ
- A esperança de um estimador está relacionada com sua **exatidão**

- Exemplos:
  - √ A média amostral é não viciada para estimar a média verdadeira (populacional):

$$\mathrm{E}\left(\bar{X}_n\right) = \mu_X$$

  - √ X_1 (primeiro item coletado da amostra) é não viciado para estimar a verdadeira média

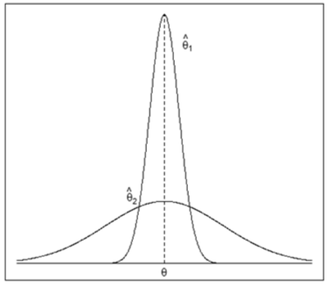
$$\mathrm{E}\left(X_1\right) = \mu_X$$

- A variância amostral é não viciada para estimar a variância populacional (σ^2)?

$$\begin{aligned}
\mathrm{E}\left[S^2\right] &= \frac{1}{n-1}\sum_{i=1}^{n}\mathrm{E}\left[\left(X_i-\bar{X}\right)^2\right] \\
&= \frac{1}{n-1}\mathrm{E}\left[\sum_{i=1}^{n}\left(X_i-\mu+\mu-\bar{X}\right)^2\right] \\
&= \frac{1}{n-1}\mathrm{E}\left[\sum_{i=1}^{n}\left(X_i-\mu\right)^2\right]-\frac{1}{n-1}\mathrm{E}\left[n\left(\bar{X}-\mu\right)^2\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\mathrm{E}\left[\left(X_i-\mu\right)^2\right]-\frac{n}{n-1}\mathrm{E}\left[\left(\bar{X}-\mu\right)^2\right] \\
&= \frac{n}{n-1}\sigma^2-\frac{1}{n-1}\sigma^2 \\
&= \sigma^2
\end{aligned}$$

Variância de Estimador

- $\hat{\theta}_1$ e $\hat{\theta}_2$ estimadores não-viciados de θ
  - √ Variâncias diferentes



  - √ $\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$
    - É mais provável que $\hat{\theta}_1$ produza uma estimativa mais próxima do valor verdadeiro de θ

Estimador de Variância Mínima

- Se considerarmos todos os estimadores não-tendenciosos de θ, aquele com a menor variância será chamado de **estimador não-tendencioso de variância mínima**
  - √ Esse estimador é o mais provável, dentre todos os não-viciados, para produzir uma estimativa que seja próxima do valor verdadeiro

Erro Padrão

- O erro padrão de um estimador $\hat{\theta}$ é o seu desvio-padrão.

$$\sigma_{\hat{\theta}} = \sqrt{\text{Var}\left[\hat{\theta}\right]}$$

  - √ O erro padrão (ou da variância) do estimador está relacionada com sua **precisão**

- √ Se o erro-padrão envolver parâmetros desconhecidos que possam ser estimados, então a substituição daqueles valores produz um erro-padrão estimado
  - Exemplo: O erro padrão da média amostral é: $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$
  - Se não conhecermos s, mas substituímos pelo desvio-padrão amostral, então o erro-padrão amostral estimado da média amostral é: $\hat{\sigma}_{\bar{X}} = \dfrac{S}{\sqrt{n}}$

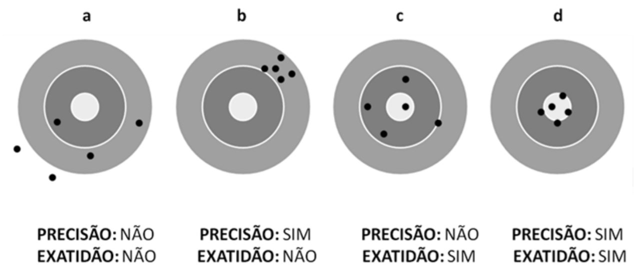
√ Quando o estimador seguir uma distribuição normal, podemos estar confiantes que o valor verdadeiro do parâmetro estará entre dois erros-padrão da estimativa

– Para grandes valores de n este é um resultado útil

√ Nos casos em que o estimador é não-viciado e não normalmente distribuído

– Estimativa do parâmetro, em no máximo 6% das vezes, se desviará do valor verdadeiro tanto quanto 4 erros-padrão

- Quadro comparativo:



Consistência

- Um estimador $\hat{\theta}$ é consistente se à medida em que o tamanho amostral aumenta, seu valor esperado converge para o parâmetro de interesse e sua variância converge para zero.

√ O estimador é consistente se $\lim_{n\to\infty} \mathrm{E}\left[\hat{\theta}\right] = \theta$ e

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

√ Consistência é uma propriedade assintótica (grandes amostras)

- Exemplos:

√ A média amostral é consistente para estimar a média verdadeira

√ O primeiro item coletado da amostra não é consistente para estimar a média populacional.

Erro Quadrático Médio

- O erro quadrático médio de um estimador $\hat{\theta}$ do parâmetro θ é definido como:

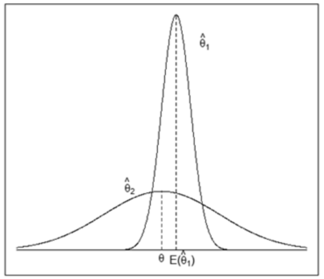
$$\text{EQM}\left[\hat{\theta}\right] = \text{E}\left[\hat{\theta} - \theta\right]^2$$

- EQM – Vício e erro-padrão

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

- O EQM é um critério importante para comparar dois estimadores

√ Estimadores tendenciosos podem ser preferíveis a estimadores não-tendenciosos se tiverem EQM menor



√ Estimativa baseada em $\hat{\theta}_1$ estaria provavelmente mais próxima do valor verdadeiro do que a baseada em $\hat{\theta}_2$

- Estimador ótimo de θ:
  - √ Tem EQM menor ou igual ao EQM de qualquer outro estimador, para todos os valores de θ
  - √ Estimadores ótimos raramente existem

EQM de Estimadores Não-viciados

- No caso em que $\hat{\theta}$ é um estimador não viciado para um parâmetro θ, então:

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

Eficiência

- Dados dois estimadores $\hat{\theta}_1$ e $\hat{\theta}_2$, não viciados para um parâmetro θ, dizemos que $\hat{\theta}_1$ é mais eficiente que θ_2 se $Var[\hat{\theta}_1] < Var[\hat{\theta}_2]$.

- Exemplo:
  - √ No caso de amostra proveniente de distribuição Normal.
    - Média amostral e mediana amostral são não viciadas para estimar a média populacional:

$$\mathrm{E}\left[\bar{X}\right] = \mu \quad \text{e} \quad \mathrm{E}\left[\tilde{X}\right] = \mu$$

    - Média amostral e mediana amostral são consistentes para estimar a média verdadeira

$$\mathrm{Var}\left[\bar{X}\right] = \frac{\sigma^2}{n} \quad \text{e} \quad \mathrm{Var}\left[\tilde{X}\right] = \frac{\pi\sigma^2}{2\,n}$$

    - A média amostral é mais eficiente que a mediana amostral para estimar a média populacional

$$\frac{\mathrm{Var}\left[\bar{X}\right]}{\mathrm{Var}\left[\tilde{X}\right]} = \frac{\frac{\sigma^2}{n}}{\frac{\pi\sigma^2}{2\,n}} = \frac{2}{\pi} = 0,63 < 1$$

Métodos de Estimação Pontual

Métodos para Obtenção de Estimadores

- Obtenção de bons estimadores:
  - √ As propriedades de estimadores não nos orientam sobre como construí-los
- Métodos para obtenção de estimadores pontuais:
  - √ Método dos Momentos
  - √ Método da Máxima Verossimilhança

Método dos Momentos

- Ideia geral:
  - √ Igualar os momentos da população (definidos em termos de esperanças) aos correspondentes momentos da amostra
  - √ Os momentos da população são funções de parâmetros desconhecidos
  - √ Essas equações são resolvidas de modo a se obter estimadores dos parâmetros desconhecidos

Momentos

- Seja $X_1, X_2, ..., X_n$ uma amostra aleatória de população com distribuição de probabilidades expressa por f(x) (função de probabilidade, se X for discreta ou função de densidade de probabilidades, se X for contínua)
  - √ k-ésimo momento populacional $\mathrm{E}[X^k]$
  - √ k-ésimo momento da amostra $\frac{1}{n}\sum_{i=1}^{n} X_i^k$

    com k = 1, 2, ...

Estimadores de Momento

- Seja uma amostra aleatória de função de probabilidade (ou de densidade de probabilidade) com *m* parâmetros desconhecidos $\theta_1, \theta_2, ..., \theta_m$.
  - √ Os estimadores de momento $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m$ são encontrados igualando os *m* primeiros momentos da população aos *m* primeiros momentos da amostra
  - √ Os estimadores serão a solução das equações resultantes

Exemplo

- Estimador de momentos da distribuição exponencial
  - √ $f(x) = \lambda e^{-\lambda x}, x>0$

    $E(X) = 1/\lambda$

    1º momento da amostra: $\overline{X}$

$$\frac{1}{\lambda} = \bar{X}$$

logo

$$\hat{\lambda} = \frac{1}{\bar{X}}$$

Exemplo

- Tempo de falha de módulo eletrônico de motor
  - √ Amostra:
    - n = 8
    - (11,96; 5,03; 67,40; 16,07; 31,50; 7,73; 11,10; 22,38)
    - Média amostral: 21,65
    - Estimativa de momento de λ: $\hat{\lambda} = \frac{1}{21,65} = 0,0462$

Estimador de Momento da Normal

- Amostra aleatória oriunda de população normal, com parâmetros μ e σ².
  - √ Momentos da normal: $\mathrm{E}[X] = \mu$, $\mathrm{E}[X^2] = \mu^2 + \sigma^2$
  - √ Estimador de momentos: $\mu = \bar{X}$, $\mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

$$\hat{\mu} = \bar{X}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}{n} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n}$$

Estimador de Momento da Gama

- Amostra aleatória oriunda de população normal, com parâmetros r e λ.
  - √ Momentos da gama: $\mathrm{E}[X] = \frac{r}{\lambda}$, $\mathrm{E}[X^2] = \frac{r(r+1)}{\lambda^2}$
  - √ Estimador de momentos: $\frac{r}{\lambda} = \bar{X}$, $\frac{r(r+1)}{\lambda^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

$$\hat{r} = \frac{\bar{X}^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2} \qquad \hat{\lambda} = \frac{\bar{X}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}$$

Exemplo

- Continuação exemplo 7.6 – Tempos de falha

$$\bar{X} = 21,65$$

$$\sum_{i=1}^{8} X_i^2 = 6645,43$$

$$\hat{r} = \frac{(21,65)^2}{\frac{1}{8}6645,43 - (21,65)^2} = 1,29$$

$$\hat{\lambda} = \frac{21,65}{\frac{1}{8}6645,43 - (21,65)^2} = 0,0598$$

  - √ r é um pouco maior que 1
  - √ É bem possível que a distribuição gama ou exponencial forneça um modelo razoável para os dados

Método da Máxima Verossimilhança

- Exemplo de Motivação:
  - √ Dados oriundos de população binomial com parâmetros 10 e p_0.
    - p_0: constante e desconhecido
    - Função de probabilidade de X:

$$f(x; p_0) = \mathrm{P}(X = x) = \binom{10}{x} p_0^x (1 - p_0)^{10-x}$$

    - Observa-se X = 3
  - √ Objetivo:
    - – Basear-se no dado disponível para estimar o valor verdadeiro do parâmetro

- Não conhecemos p_0, mas podemos considerar o cenário em p_0 = ½.
  - √ Sob esta particular condição, a probabilidade de gerar o dado que realmente observamos (X = 3) é:

$$\mathrm{P}(X = 3) = f(3; 0,5) = \binom{10}{3} (0,5)^3 (0,5)^7 \approx 0,117$$

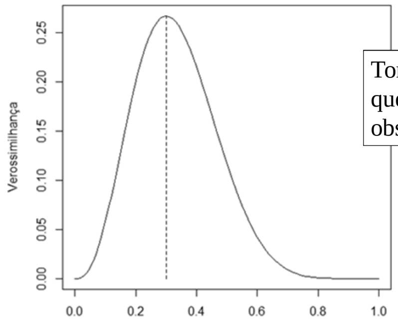
- Podemos calcular esta probabilidade sob a condição que p_0 = p

$$L(p; 3) = \binom{10}{3} (p)^3 (1 - p)^7, \ p \in [0, 1]$$

  - √ Essa função é denominada função de verossimilhança e denotamos por $L(p; 3)$

Princípio da Máxima Verossimilhança

- Devemos usar como nossa estimativa de p0 o valor de p que faz $L(p; 3)$ o maior possível



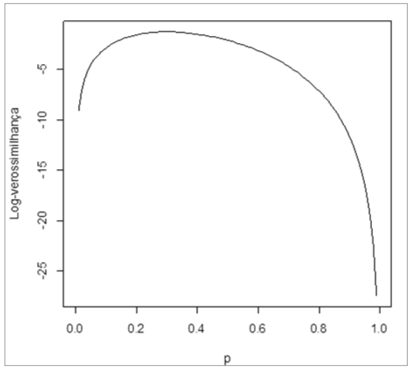
Toma-se o valor do parâmetro que torna mais provável o dado observado

$$\mathrm{P}(X = 3) = f(3; 0,3) = \binom{10}{3} (0,3)^3 (0,7)^7 \approx 0,267$$

Função de Log-Verossimilhança

- Como o log é uma função crescente, o valor de p que maximiza $L(p; 3)$ é o mesmo que maximiza log $L(p; 3)$
  - √ Em geral, é conveniente maximizar log $L(p; 3)$ ao invés de $L(p; 3)$
  - √ Assim, em nosso exemplo, a função de log-verossimilhança é definida como:

$$\begin{aligned} l(p; 3) &\stackrel{\text{def}}{=} \ln L(p; 3) \\ &= 3 \ln p + 7 \log(1 - p) + \ln \binom{10}{3} \end{aligned}$$

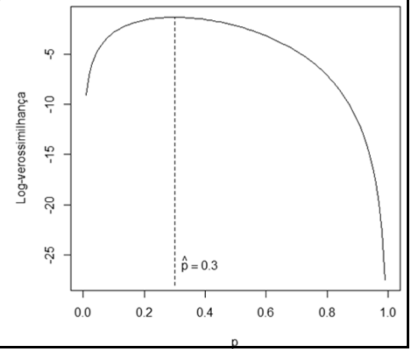


$$l(p;3)3\ln p + 7\ln(1-p) + \ln\binom{10}{3}$$

√ O ponto crítico da função é um ponto no domínio de em que a derivada é zero

$$0 = \frac{\partial}{\partial p} l(p;3)$$
$$= \frac{3}{p} - \frac{7}{1-p} = 3(1-p) - 7p$$

$$\hat{p} = \frac{3}{10}$$



- A estimativa é determinada pelo valor de X
  - √ Se tivéssemos observado X = k, teríamos a estimativa $\hat{p} = \frac{k}{10}$
- Em nosso exemplo, o estimador de máxima verossimilhança é?

$$\hat{p} = \frac{X}{10}$$

Função de Verossimilhança – Definição

- Suponha X uma variável aleatória com distribuição de probabilidades f(x; θ), em que θ é um único parâmetro desconhecido.
  - √ Sejam x_1, x_2, ..., x_n os valores observados de amostra aleatória de tamanho n.
  - √ A função de verossimilhança da amostra é:

$$L(\theta; x_1, x_2, \ldots x_n) = f(x_1;\theta) f(x_2;\theta) \ldots f(x_n;\theta)$$

Estimador de Máxima Verossimilhança

- O estimador de máxima verossimilhança (EMV) de θ é o valor $\hat{\theta}$ que maximiza a função de verossimilhança $L(\theta; x)$

$$L(\hat{\theta}; \mathbf{x}) = \max_{\theta \in \Theta} L(\theta; \mathbf{x})$$

√ No caso discreto, o EMV é um estimador que maximiza a probabilidade de ocorrência dos valores da amostra

Exemplo – Distribuição de Bernoulli

- Seja X uma variável aleatória de Bernoulli

√ Função de probabilidade:

$$f(x;p) = \begin{cases} p^x(1-p)^{1-x} & , x = 0, 1 \\ 0 & , \text{caso contrário} \end{cases}$$

p é o parâmetro desconhecido a ser estimado

- Função de verossimilhança da amostra

$$\begin{aligned} L(p;\mathbf{x}) &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \ldots p^{x_n}(1-p)^{1-x_n} \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^n x_i}(1-p)^{\sum_{i=1}^n (1-x_i)} \\ &= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} \end{aligned}$$

- Função de log-verossimilhança

$$l(p;\mathbf{x}) = \sum_{i=1}^{n} x_i \ln(p) + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

- Derivada da função de log-verossimilhança:

$$l'(p;\mathbf{x}) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p}$$

- Estimativa de máxima verossimilhança da amostra **x**:

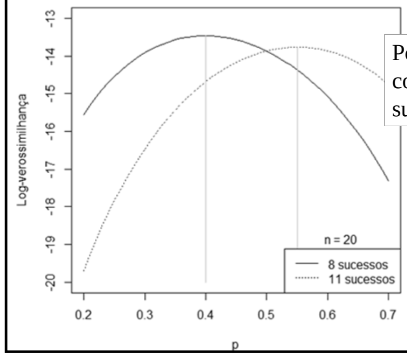
$$l'(\hat{p};\mathbf{x}) = \frac{\sum_{i=1}^n x_i}{\hat{p}} - \frac{n - \sum_{i=1}^n x_i}{1-\hat{p}} = 0 \qquad \hat{p} = \frac{\sum_{i=1}^n x_i}{n}$$

- Estimador de máxima verossimilhança para amostras de Bernoulli $\hat{p} = \bar{X}$

√ Proporção de sucessos na amostra

Exemplo

- Bernoulli com parâmetro p desconhecido
  - √ Duas amostras de n= 20 com $\sum_{i=1}^{20} x_i = 8$ e $\sum_{i=1}^{20} x_i = 11$



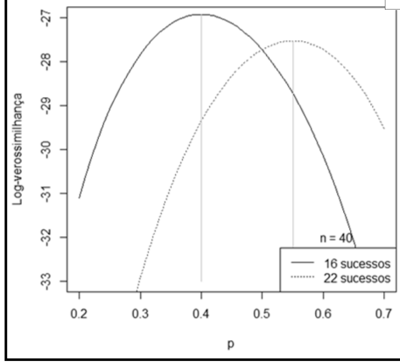
Pontos de máxima verossimilhança correspondentes à proporção de sucessos de cada amostra

$$\hat{p}_1 = \frac{8}{20} = 0,40$$

$$\hat{p}_2 = \frac{11}{20} = 0,55$$

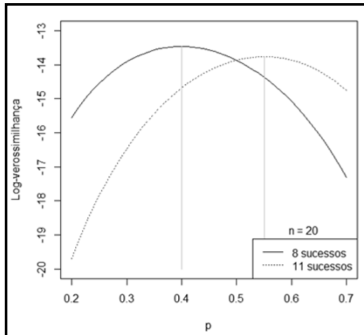
√ Duas amostras de n= 40 com $\sum_{i=1}^{20} x_i = 16$ e $\sum_{i=1}^{20} x_i = 22$

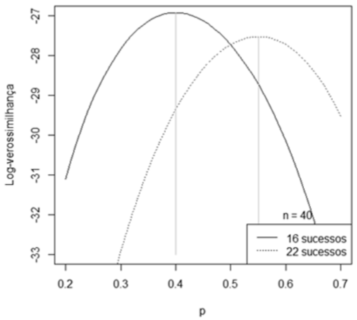
Pontos de máxima verossimilhança correspondentes à proporção de sucessos de cada amostra



$$\hat{p}_1 = \frac{16}{40} = 0,40$$

$$\hat{p}_2 = \frac{22}{40} = 0,55$$





√ As curvas crescem (decrescem) mais acentuadamente nas proximidades dos pontos de máximo na amostra maior (n = 40)

– Variância do EMV é menor para amostras maiores)

- Amostragem industrial pode ser modelada como amostras de variável aleatória de Bernoulli:
  - √ n itens selecionados ao acaso de linha de produção
  - √ Variável Aleatória:

$$X_i = \begin{cases} 1 & \text{, item classificado como não conforme} \\ 0 & \text{, caso contrário} \end{cases}$$

  - √ Estatística de teste:

$\sum_{i=1}^{n} x_i$: quantidade de itens não conformes na amostra

$\hat{p}$: proporção de itens não conformes na amostra

Exemplo – Distribuição Exponencial

- Seja X uma variável aleatória exponencial
  - √ Função de densidade de probabilidade:

$$f(x;\lambda) = \begin{cases} \lambda\, e^{-\lambda x} & , x > 0 \\ 0 & , \text{caso contrário} \end{cases}$$

λ é o parâmetro desconhecido a ser estimado

- Função de verossimilhança da amostra

$$\begin{aligned} L(\lambda;\mathbf{x}) &= \lambda\, e^{-\lambda x_1} \lambda\, e^{-\lambda x_2} \ldots \lambda\, e^{-\lambda x_n} \\ &= \prod_{i=1}^{n} \lambda\, e^{-\lambda x_i} \\ &= \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} \end{aligned}$$

- Função de log-verossimilhança

$$l(\lambda;\mathbf{x}) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

- Derivada da função de log-verossimilhança:

$$l'(\lambda;\mathbf{x}) = \frac{\partial}{\partial\lambda} l(\lambda;\mathbf{x}) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

- Estimativa de máxima verossimilhança da amostra **x**:

$$l'(\hat{\lambda};\mathbf{x}) = \frac{n}{\hat{\lambda}} - \sum_{i=1}^{n} x_i = 0 \qquad \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$$

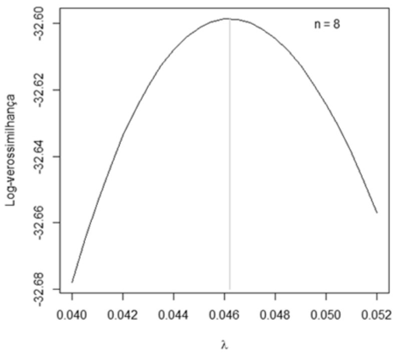
- Estimador de máxima verossimilhança para amostras exponenciais $\hat{\lambda} = \frac{1}{\bar{X}}$
  - √ O EMV da exponencial é igual a seu estimador de momentos

Exemplo

- (Continuação Ex. 7.6)
  - √ Tempo de falha de módulo eletrônico de motor
  - √ Amostra:
    - n = 8
    - (11,96; 5,03; 67,40; 16,07; 31,50; 7,73; 11,10; 22,38)
    - Média amostral: 21,65 $\bar{X} = 21,65$
    - Estimativa de máxima verossimilhança de λ:

$$\hat{\lambda} = \frac{1}{21,65} = 0,0462$$

- Função de log-verossimilhança da amostra

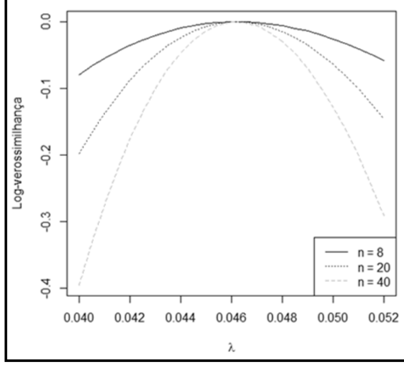


Função atinge máximo para λ=0,0462

√ Função é relativamente plana na região do máximo

– Parâmetro não é estimado muito precisamente

- Comparação exponenciais:
  - √ Curvas das diferenças de log-verossimilhança com o máximo em cada tamanho amostral
  - √ Mantida mesma média amostral de 21,65



n	$l(\lambda_{máx})$
8	–32,599
20	–81,497
40	–162,994

Inclinação da curva de log-verossimilhança acentua-se com aumento da amostra.

Picos acentuados = maior precisão na estimação

Exemplo – Distribuição Normal

- Seja X uma variável aleatória normal
  - √ Função de probabilidade:

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

  - √ Situações de estimação dos parâmetros:
    - – μ é desconhecido e σ^2 é conhecido
    - – μ e σ^2 são desconhecidos

- Função de verossimilhança da amostra
  - √ μ desconhecida e σ^2 conhecido

$$L(\mu;\sigma^2,\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

- Função de log-verossimilhança

$$l(\mu;\sigma^2,\mathbf{x}) = -\frac{n}{2}\ln(2\pi\sigma^2) - (2\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\mu)^2$$

- Derivada da função de log-verossimilhança:

$$l'(\mu;\sigma^2,\mathbf{x}) = \frac{\partial}{\partial\mu} l(\mu;\sigma^2,\mathbf{x}) = -(2\sigma^2)^{-1}(-2)\sum_{i=1}^{n}(x_i-\mu)$$

$$= -(\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\mu)$$

- Estimativa de máxima verossimilhança da amostra **x**:

$$l'(\hat\mu;\sigma^2,\mathbf{x}) = (\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\hat\mu) = 0 \qquad \hat\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

- Estimador de máxima verossimilhança da média verdadeira para amostras normais

$$\hat\mu = \bar X$$

- Função de verossimilhança da amostra
  - √ μ e σ^2 desconhecidos

$$L(\mu,\sigma^2;\mathbf{x}) = \prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

- Função de log-verossimilhança

$$l(\mu,\sigma^2;\mathbf{x}) = -\frac{n}{2}\ln(2\pi\sigma^2) - (2\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\mu)^2$$

- Derivadas parciais da log-verossimilhança:

$$\frac{\partial}{\partial\mu} l(\mu,\sigma^2;\mathbf{x}) = -(2\sigma^2)^{-1}(-2)\sum_{i=1}^{n}(x_i-\mu)$$

$$= (\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\mu)$$

$$\frac{\partial}{\partial\sigma^2} l(\mu,\sigma^2;\mathbf{x}) = -\left(\frac{n}{2}\right)\frac{(2\pi)}{2\pi\sigma^2} + \frac{2}{(2\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= \frac{1}{2\sigma^2}\left[-n + \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

- Estimativas de máxima verossimilhança da amostra **x**:

$$\frac{\partial}{\partial\mu} l(\hat\mu,\hat\sigma^2;\mathbf{x}) = (\hat\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\hat\mu) = 0 \qquad \hat\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\frac{\partial}{\partial\sigma^2} l(\hat\mu,\hat\sigma^2;\mathbf{x}) = \frac{1}{2\hat\sigma^2}\left[-n + \frac{1}{\hat\sigma^2}\sum_{i=1}^{n}(x_i-\hat\mu)^2\right] = 0$$

$$\hat\sigma = \frac{\sum_{i=1}^{n}(x_i-\hat\mu)^2}{n}$$

- Estimador de máxima verossimilhança da média e variâncias verdadeiras para amostras normais

$$\hat{\mu} = \bar{X}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$$

  - √ O EMV da variância amostral é viciado para a variância verdadeira

Comentários

- Para usar a estimação de máxima verossimilhança, a distribuição da população tem de ser conhecida ou suposta
- Em geral, o método da máxima verossimilhança produz estimadores com boas propriedades estatísticas
  - √ Têm boas propriedades assintóticas dos estimadores
  - √ (são consistentes e assintoticamente eficientes)

Estimador de Máxima Verossimilhança - Propriedades

- Sob condições gerais e não-restritivas, quando uma amostra de tamanho *n* for grande e se $\hat{\theta}$ for um EMV do parâmetro θ, então:
  - √ $\hat{\theta}$ é um estimador consistente
    - – Assintoticamente não viciado
  - √ A variância de $\hat{\theta}$ é assintoticamente eficiente
  - √ θ tem distribuição assintoticamente normal

Propriedade da Invariância

- Seja $\hat{\theta}$ um estimador de máxima verossimilhança de θ.
  - √ Então o EMV de qualquer função $h(\theta)$ desse parâmetro é a mesma função $h(\hat{\theta})$ do estimador $\hat{\theta}$
- No caso da distribuição normal
  - √ O EMV de σ é $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}\right]^{1/2}$
  - √ O EMV de σ não é o desvio-padrão amostral S

Complicações no Uso da Estimação de Máxima Verossimilhança

- Nem sempre é simples maximizar a função de verossimilhança
- Pode não ser possível utilizar diretamente métodos de cálculo para determinar o máximo de $L(\theta; \mathbf{x})$

Exemplo – Distribuição Gama

- Seja X uma variável aleatória exponencial
  - √ Função de densidade de probabilidade:

$$f(x; r, \lambda) = \begin{cases} \frac{\lambda^r x^{r-1}}{\Gamma(r)} e^{-\lambda x} & , x > 0 \\ 0 & , \text{caso contrário} \end{cases}$$

r e λ são os parâmetros desconhecidos a serem estimados

  - √ Função gama: $\Gamma(r) = \int_0^\infty y^{r-1} \mathrm{e}^{-y} \, \mathrm{d}y$

- Função de verossimilhança da amostra

$$\begin{aligned} L(\lambda, r; \mathbf{x}) &= \frac{\lambda^r x_1^{r-1}}{\Gamma(r)} e^{-\lambda x_1} \frac{\lambda^r x_2^{r-1}}{\Gamma(r)} e^{-\lambda x_2} \dots \frac{\lambda^r x_n^{r-1}}{\Gamma(r)} e^{-\lambda x_n} \\ &= \prod_{i=1}^n \frac{\lambda^r x_i^{r-1}}{\Gamma(r)} e^{-\lambda x_i} \\ &= \left(\frac{\lambda^r}{\Gamma(r)} \right)^n \mathrm{e}^{-\lambda \sum_{i=1}^n x_i} \left(\prod_{i=1}^n x_i \right)^{r-1} \end{aligned}$$

- Função de log-verossimilhança

$$l(\lambda, r; \mathbf{x}) = nr \ln(\lambda) - n \ln(\Gamma(r)) + (r-1) \sum_{i=1}^n \ln(x_i) - \lambda \sum_{i=1}^n x_i$$

- Derivadas da função de log-verossimilhança:

$$\frac{\partial}{\partial \lambda} l(\lambda, r; \mathbf{x}) = \frac{nr}{\lambda} - \sum_{i=1}^n x_i$$

$$\frac{\partial}{\partial r} l(\lambda, r; \mathbf{x}) = n \ln(\lambda) + \sum_{i=1}^n \ln(x_i) - \frac{n\Gamma'(r)}{\Gamma(r)}$$

- Estimativa de máxima verossimilhança da amostra **x**:

$$\frac{n\hat{r}}{\hat{\lambda}} - \sum_{i=1}^n x_i = 0 \qquad \hat{\lambda} = \frac{n\hat{r}}{\sum_{i=1}^n x_i} = \frac{\hat{r}}{\bar{x}}$$

$$n \ln(\hat{\lambda}) + \sum_{i=1}^n \ln(x_i) = \frac{n\Gamma'(\hat{r})}{\Gamma(\hat{r})}$$

  - √ Não há solução exata para essas equações

Aplicação

- Método Lincoln-Peterson de Marcação e Recaptura
  - √ Objetivo: estimar tamanho de população de animais
  - √ Variáveis:
    - t: número de animais capturados e marcados
    - k: número de animais recapturados
    - r: número de animais marcados que são recapturados
    - N: população total (desconhecida)
  - √ Os valores t e k são fixos (não são aleatórios)

    São determinados no planejamento do estudo
  - √ r: observação amostral (pode variar)

- Função de verossimilhança da amostra

$$L(N;r) = \frac{\binom{t}{r}\binom{N-t}{k-r}}{\binom{N}{k}}$$

- O espaço paramétrico são os inteiros positivos
  - √ Não se pode usar Cálculo
- Razão de verossimilhança para sucessivos valores da população total

$$\frac{L(N;r)}{L(N-1;r)}$$

  - √ N é mais provável que N–1 quando a razão > 1

- Desenvolvimento razão de verossimilhanças:

$$\frac{L(N;r)}{L(N-1;r)} = \frac{\frac{\cancel{\binom{t}{r}}\binom{N-t}{k-r}}{\binom{N}{k}}}{\frac{\cancel{\binom{t}{r}}\binom{N-t-1}{k-r}}{\binom{N-1}{k}}} = \frac{\binom{N-t}{k-r}\binom{N-1}{k}}{\binom{N-t-1}{k-r}\binom{N}{k}}$$

$$= \frac{\frac{\cancelto{N-t}{(N-t)!}}{\cancelto{N-t-k+r}{(N-t-k+r))!}\ \cancel{(k-r)!}}\ \frac{\cancel{(N-1)!}}{\cancel{k!}\cancel{(N-1-k))!}}}{\frac{\cancel{(N-t-1)!}}{\cancel{(k-r)!}\cancel{(N-t-1-k+r))!}}\ \frac{\cancelto{N}{N!}}{\cancel{k!}\cancelto{N-k}{(N-k))!}}}$$

$$= \frac{(N-t)(N-k)}{(N(N-t-k+r)}$$

- Determinação máximo:

$$\frac{L(N;r)}{L(N-1;r)} > 1 \Leftrightarrow (N-t)(N-k) > N(N-t-k+r)$$

$$\cancel{N^2} - \cancel{tN} - \cancel{kN} + tk > \cancel{N^2} - \cancel{tN} - \cancel{kN} + rN$$

$$tk > rN \Rightarrow \frac{tk}{r} > N$$

  - √ Seja [x]: parte inteira de x

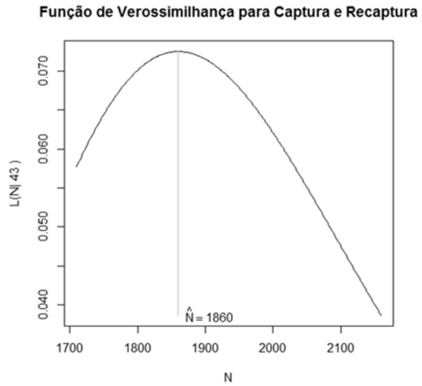
$$\begin{cases} L(N;r) > L(N-1;r) & \text{, para } N < \left[\frac{tk}{r}\right] \\ L(N;r) \le L(N-1;r) & \text{, para } N \ge \left[\frac{tk}{r}\right] \end{cases}$$

  - √ Isso nos dá o EMV de N:

$$\hat{N} = \left[\frac{tk}{r}\right]$$

Exemplo

- Objetivo:
  - √ Estimar população de peixes
- Procedimento:
  - √ São marcados t = 200 peixes
  - √ Peixes são devolvidos ao habitat
  - √ Aguarda-se que os peixes devolvidos misturem-se à população
  - √ São recapturados k = 400 peixes
  - √ Dos peixes recapturados, há r = 43 peixes marcados
- Obs. Dados obtidos em simulação com N=2000



$$\hat{N} = \left[\frac{(200)(400)}{43}\right] = 1860$$

- √ População estimada: 1.860 peixes
- √ População real: 2000 (na prática, desconhecido)

Referências

Bibliografia Recomendada

- Montgomery, D. C. (LTC)

  Estatística Aplicada e Probabilidade para Engenheiros

- Pinheiro, J. I. D et al. (Campus)

  Probabilidade e Estatística: Quantificando a Incerteza