

# Amostragem aleatória simples com reposição (parte 1)

Prof. Caio Azevedo

# Estrutura geral

- Temos uma população de interesse de tamanho  $N$  e desejamos realizar inferências sobre algum parâmetro (média, total, proporção, variância) dessa população, com base em uma amostra de tamanho  $n$ .
- Com algumas adaptações, os resultados a serem vistos poderão ser utilizados mesmo se  $N$  for infinito.
- População: observações univariadas -  $y_1, \dots, y_N$  (variáveis não aleatórias), em que  $y_i$  é a observação relativa ao indivíduo  $i$  (podemos também considerar observações multivariadas). Exemplos: peso, altura, intenção de voto, conhecimento em alguma área.

# Estrutura geral

- Objetivo: estimar  $\mu = \frac{1}{N} \sum_{i=1}^N y_i$  (média, ou proporção se os  $y_i$ 's foram variáveis binárias),  $\tau = \sum_{i=1}^N y_i = N\mu$  (total), variância ( $\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2$ ) (esta, as vezes, tem de ser estimada para se poder fazer inferência para parâmetros de interesse como a média, total, proporção etc), com base na amostra de tamanho  $n$ , com reposição.
- Amostragem aleatória simples com reposição ( $AAS_c \equiv A_1$ ).
- Amostra  $\{y_{k_1}, y_{k_2}, \dots, y_{k_n}\}$ , em que  $k_i \in \{1, 2, \dots, N\}$ . Por exemplo, Se  $N = 5$  e  $n = 3$ , podemos ter  $\{y_2, y_3, y_3\}$ .

# Mecanismo de sorteio da amostra

- 1 Dado que os elementos da população estão numerados de 1 a  $N$ , sorteia-se um elemento, segundo algum procedimento de geração de um número aleatório (função “sample”, no R).
- 2 Repõe-se esse elemento na população.
- 3 Repete-se os procedimentos 1 e 2,  $N-1$  vezes.

# Estimação da média

- Estimador natural (sob duas formas diferentes):

$$\hat{\mu} = \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \quad (1)$$

$$\hat{\mu} = \bar{Y} = \frac{1}{n} \sum_{i=1}^N F_i y_i, \quad (2)$$

em que  $F_i$  é uma v.a. que representa o número de vezes que o  $i$ -ésimo elemento da população apareceu na amostra, e  $s$  representa a amostra sorteada.

# Estimação da média

- Utilizar a forma (1) e considerarmos a distribuição das variáveis  $\mathbf{Y} = (Y_1, \dots, Y_n)^t$  (esses desenvolvimentos dependem da distribuição considerada para a amostra). Neste caso,  $Y_i, i = 1, \dots, n$  são variáveis aleatórias. Esta abordagem é vista nos cursos de Inferência Estatística e pode levar a inferências exatas, desde que as suposições sejam válidas. Note que, neste caso, como há reposição  $Y_i \stackrel{i.i.d}{\sim} D(\theta)$ .

# Estimação da média

- Utilizar a forma (2) e considerarmos a distribuição das variáveis  $\mathbf{F} = (F_1, \dots, F_N)^t$  (mais geral, ou seja, em princípio, os resultados se aplicam, independentemente da distribuição de  $\mathbf{Y}$ ). Neste caso,  $F_i, i = 1, \dots, N$  são variáveis aleatórias ( $y_i, i = 1, 2, \dots, N$  são variáveis não aleatórias). Esta abordagem leva a inferências aproximadas (“n” e “N-n” suficientemente grandes). Uma vantagem é que ela se aplica, em princípio, independentemente da forma da distribuição de  $\mathbf{Y}$ .

# Estimação da média

## ■ Resultados:

1  $F_i \stackrel{i.d.}{\sim}$  binomial( $n, 1/N$ ). Ou seja,

$$P(F_i = f_i) = \binom{n}{f_i} p^{f_i} (1-p)^{n-f_i} \mathbb{1}_{\{0,1,2,\dots,n\}}(f_i)$$

2  $(F_i, F_j) \stackrel{i.d.}{\sim}$  trinomial( $n, 1/N, 1/N$ ), ou seja, os  $F_i$ 's são **dependentes**,

$$\frac{n!}{f_i! f_j! (n-f_i-f_j)!} p_1^{f_i} p_2^{f_j} (1-p_1-p_2)^{n-f_i-f_j} \mathbb{1}_{\{0,1,2,\dots,n-f_j\}}(f_i) \mathbb{1}_{\{0,1,2,\dots,n\}}(f_j)$$

- Exercício: obtenha a distribuição de  $(F_{k_1}, F_{k_2}, \dots, F_{k_r})^t$ , em que  $\{k_1, k_2, \dots, k_r\}$  é um subconjunto de  $\{1, 2, \dots, n\}$ .
- Exercício: obtenha a distribuição de  $(F_1, F_2, \dots, F_N)^t$ .



# Estimação da média

- Prova (1): para um número  $n$  de repetições, repetimos um procedimento cujo a probabilidade de sucesso (selecionar o elemento  $i$ ) é constante ( $1/N$ ) e contabilizamos o número de sucessos.
- Prova (2): para um número  $n$  de repetições, repetimos um procedimento cujas probabilidades de interesse (selecionar os elementos  $i$  e  $j$ ) são constantes ( $1/n, 1/n$ ) e contabilizamos o número de ocorrências de cada categoria.

## Propriedades dos $F_i$

1  $\mathcal{E}_{A_1}(F_i) = \frac{n}{N}, \mathcal{V}_{A_1}(F_i) = n \frac{1}{N} \left( \frac{N-1}{N} \right) = n \frac{N-1}{N^2}.$

2  $\text{Cov}_{A_1}(F_i, F_j) = -n \frac{1}{N} \frac{1}{N} = -\frac{n}{N^2}.$

3  $\pi_i = 1 - \left(1 - \frac{1}{N}\right)^n$  (probabilidade do  $i$ -ésimo elemento aparecer na amostra). Prova

$$\pi_i = 1 - P(F_i = 0) = 1 - \left(1 - \frac{1}{N}\right)^n.$$

4  $\pi_{ij} = 1 - 2 \left(1 - \frac{1}{N}\right)^n + \left(1 - \frac{2}{N}\right)^n$  (probabilidade do  $i$ -ésimo e  $j$ -ésimo elementos aparecerem na amostra). Exercício: provar esta propriedade.

As provas dos resultados 1, 2 e 3 decorrem das respectivas distribuições de  $F_i$  e  $(F_i, F_j)^t$ .

# Estimação da média : Propriedades do estimador sob $AAS_c$

- Valor esperado

$$\begin{aligned}\mathcal{E}_{A_1}(\hat{\mu}) &= \frac{1}{n} \mathcal{E}_{A_1} \left( \sum_{i=1}^N F_i y_i \right) = \frac{1}{n} \sum_{i=1}^N \mathcal{E}_{A_1}(F_i) y_i = \frac{1}{n} \sum_{i=1}^N \frac{ny_i}{N} \\ &= \sum_{i=1}^N \frac{y_i}{N} = \mu\end{aligned}$$

# Estimação da média : Propriedades do estimador sob $AAS_c$

- Variância do estimador

$$\begin{aligned} \mathcal{V}_{A_1}(\hat{\mu}) &= \frac{1}{n^2} \mathcal{V} \left( \sum_{i=1}^N F_i y_i \right) \\ &= \frac{1}{n^2} \left( \sum_{i=1}^N y_i^2 \mathcal{V}_{A_1}(F_i) + \sum_{\substack{i \neq j, \\ i, j=1, 2, \dots, N}} \text{Cov}_{A_1}(F_i y_i, F_j y_j) \right) \\ &= \frac{1}{n^2} \left( \sum_{i=1}^N y_i^2 \left( \frac{n(N-1)}{N^2} \right) - \sum_{\substack{i \neq j, \\ i, j=1, 2, \dots, N}} y_i y_j \frac{n}{N^2} \right) \\ &= \frac{1}{N^2 n} (N-1) \left( \sum_{i=1}^N y_i^2 - \frac{1}{N-1} \sum_{\substack{i \neq j, \\ i, j=1, 2, \dots, N}} y_i y_j \right) \end{aligned}$$

# Estimação da média : Propriedades do estimador sob $AAS_c$

## ■ Cont.

Lembrando que (Exercício)  $\sum_{i \neq j} y_i y_j = -\sum_{i=1}^N y_i^2 + N^2 \mu^2$  e  $\sum_{i=1}^N (y_i - \mu)^2 = \sum_{i=1}^N y_i^2 - N\mu^2$ , vem que

$$\begin{aligned} \mathcal{V}_{A_1}(\hat{\mu}) &= \frac{1}{N^2 n} (N-1) \left( \sum_{i=1}^N y_i^2 - \frac{1}{N-1} \left( N^2 \mu^2 - \sum_{i=1}^N y_i^2 \right) \right) \\ &= \frac{N-1}{(N-1)N^2 n} \left( N \sum_{i=1}^N y_i^2 - \sum_{i=1}^n y_i^2 - N^2 \mu^2 + \sum_{i=1}^N y_i^2 \right) \\ &= \frac{1}{N^2 n} \left( N \sum_{i=1}^N y_i^2 - N^2 \mu^2 \right) = \frac{1}{n} \left( \frac{1}{N} \sum_{i=1}^N y_i^2 - \mu^2 \right) = \frac{\sigma^2}{n} \end{aligned}$$

# Estimação da média : Propriedades do estimador sob $AAS_c$

- Resumidamente,  $\mathcal{E}_{A_1}(\hat{\mu}) = \mu$  e  $\mathcal{V}_{A_1}(\hat{\mu}) = \frac{\sigma^2}{n}$ . Podemos provar, sob  $AAS_c$ , que  $\hat{\mu}$  é consistente.
- A distribuição exata é bastante complicada de ser obtida (média de uma combinação linear de um vetor aleatório com distribuição multinomial).
- Distribuição assintótica: note que em  $\{F_i\}_{i \geq 1}$  os  $F_i$ 's são identicamente distribuídos mas não independentes. O TLC padrão não se aplica.
- Estimativa  $\tilde{\mu} = \frac{1}{n} \sum_{i \in \mathcal{S}} y_i = \frac{1}{n} \sum_{i=1}^N f_i y_i$ .

# Estimação da média : Propriedades do estimador sob $AAS_c$

- Discutiremos, com mais detalhes (mais a frente), como se obter os resultados assintóticos mas, por enquanto, sob certas condições, entre elas,  $n$  e  $N-n$  suficientemente grandes, temos que

$$\frac{\hat{\mu} - \mu}{\sqrt{\sigma^2/n}} \xrightarrow[N-n \rightarrow \infty]{n \rightarrow \infty, D} N(0, 1) \quad (4)$$

ou

$\hat{\mu} \approx N(\mu, \sigma^2/n)$ , para  $n$  e  $N-n$  suficientemente grandes.

Problema:  $\sigma^2$ , quase sempre, é desconhecido. Faz-se necessário considerar um estimador consistente (de preferência não viciado), para se poder usar o Teorema de Slutsky .

# Estimação da média : Propriedades do estimador sob $AAS_c$

- Vamos considerar o seguinte estimador

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \hat{\mu})^2 = \frac{1}{n-1} \sum_{i=1}^N F_i (y_i - \hat{\mu})^2.$$

Note que (lembrando que  $\sum_{i=1}^N y_i^2 = N\sigma^2 + N\mu^2$ )

$$\begin{aligned} \mathcal{E}_{A_1}(\hat{\sigma}^2) &= \frac{1}{n-1} \mathcal{E} \left( \sum_{i=1}^n Y_i^2 - n\hat{\mu}^2 \right) = \frac{1}{n-1} \left[ \mathcal{E}_{A_1} \left( \sum_{i=1}^N y_i^2 F_i \right) - n\mathcal{E}(\hat{\mu}^2) \right] \\ &= \frac{1}{n-1} \left[ \sum_{i=1}^N y_i^2 \mathcal{E}_{A_1}(F_i) - n\mathcal{E}_{A_1}(\hat{\mu}^2) \right] \\ &= \frac{1}{n-1} \left[ N(\sigma^2 + \mu^2) \frac{n}{N} - n \left( \mu^2 + \frac{\sigma^2}{n} \right) \right] \\ &= \frac{1}{n-1} [n\sigma^2 + n\mu^2 - n\mu^2 - \sigma^2] = \sigma^2 \end{aligned}$$



# Estimação da média : Propriedades do estimador sob $AAS_c$

- A prova de sua consistência, i.e.,

$$\hat{\sigma}^2 \xrightarrow[n \rightarrow \infty, N-n \rightarrow \infty]{P} \sigma^2 \quad (5)$$

também será discutida mais à frente.

- Portanto, dos resultados (4) e (5), temos que

$$\frac{\hat{\mu} - \mu}{\sqrt{\hat{\sigma}^2/n}} = \frac{\hat{\mu} - \mu}{\sqrt{\sigma^2/n}} \frac{\sigma}{\hat{\sigma}} \xrightarrow[n \rightarrow \infty, N-n \rightarrow \infty]{D} N(0, 1)$$

por Slutsky.

# Intervalo de Confiança

- Estimativa:  $\tilde{\sigma}^2 = \frac{1}{n-1} \sum_{i \in \mathcal{S}} (y_i - \tilde{\mu})^2$ .
- Assim, um intervalo de confiança (assintótico) com coeficiente de confiança de aproximadamente  $\gamma$  é dado por

$$IC(\mu, \gamma) \approx \left[ \hat{\mu} - z_\gamma \sqrt{\frac{\hat{\sigma}^2}{n}}; \hat{\mu} + z_\gamma \sqrt{\frac{\hat{\sigma}^2}{n}} \right]$$

em que  $P(Z \leq z_\gamma) = \frac{1+\gamma}{2}$  e  $Z \sim N(0, 1)$ .

- Erro da estimativa:  $z_\gamma \sqrt{\frac{\hat{\sigma}^2}{n}}$ .

# Testes de Hipótese

- Hipóteses usuais ( $\mu_0$  conhecido)

- 1  $H_0 : \mu = \mu_0$  vs  $H_1 : \mu < \mu_0$ .

- 2  $H_0 : \mu = \mu_0$  vs  $H_1 : \mu > \mu_0$ .

- 3  $H_0 : \mu = \mu_0$  vs  $H_1 : \mu \neq \mu_0$ .

- Estatística do teste  $Z_t = \frac{\hat{\mu} - \mu_0}{\hat{\sigma}/\sqrt{n}}$ , em que  $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ .

- Sob  $H_0$ , vimos que  $Z_t \approx N(0, 1)$ , para  $n$  e  $N-n$  suficientemente grandes.

- Defina  $z_t = \frac{\tilde{\mu} - \mu_0}{\tilde{\sigma}/\sqrt{n}}$  o valor calculado da estatística do teste e  $z_c$  o(s) valor(es) crítico(s).

- Defina ainda  $Z \sim N(0, 1)$ .

# Testes de Hipótese

- Procedimento para testar o conjunto de hipóteses [1]
- Valor crítico
  - $P(Z \leq z_c | H_0) = \alpha$ .
  - Se  $z_t \leq z_c$  rejeita-se  $H_0$ , caso contrário, não se rejeita.
- p-valor (nível descritivo)
  - $p - \text{valor} = P(Z \leq z_t | H_0)$

# Testes de Hipótese

- Procedimento para testar o conjunto de hipóteses [2]
- Valor crítico
  - $P(Z \geq z_c | H_0) = \alpha$ .
  - Se  $z_t \geq z_c$  rejeita-se  $H_0$ , caso contrário, não se rejeita.
- p-valor (nível descritivo)
  - $p - \text{valor} = P(Z \geq z_t | H_0)$

# Testes de Hipótese

- Procedimento para testar o conjunto de hipóteses [3]
- Valor crítico
  - $P(Z \leq z_c | H_0) = \frac{1+\alpha}{2}$ .
  - Se  $|z_t| \geq z_c$  rejeita-se  $H_0$ , caso contrário, não se rejeita.
- p-valor (nível descritivo)
  - $p - \text{valor} = 2[1 - P(Z \leq |z_t| | H_0)]$ .

# Estudos de simulação

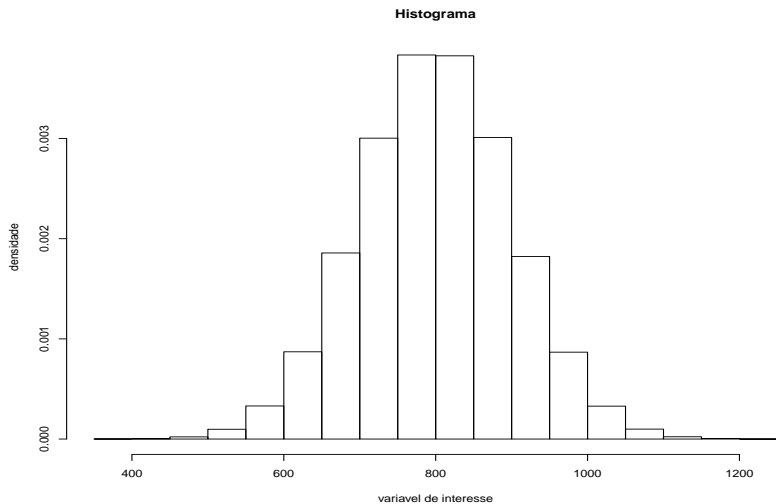
- Distribuição assintótica do estimador para a média. Tamanho da população  $N = 100000$ .
- Cinco cenários, variando em função da variável de interesse na população ( $X$ ).
  - $X \sim N(800, 10000)$
  - $X \sim \text{gama}(5; 0, 00625)$ ,  $E(X) = 800$ ,  $V(X) = 128000$ .
  - $X \sim t_{(7)}(800, 5000)$ ,  $E(X) = 800$ ,  $V(X) = 7000$ .
  - $X \sim U[400; 1200]$ .
  - $X \sim 0.5N(200, 5000) + 0.5N(600, 5000)$

# Estudos de simulação

- Quatro tamanhos amostrais (30, 50, 100, 1000), em termos percentuais, com relação ao tamanho da população (0,03%,0,05%,0,1%,1%).
- Estudar a distribuição amostral (empírica) com base em  $R = 1000$  réplicas (amostras selecionadas da população de interesse).

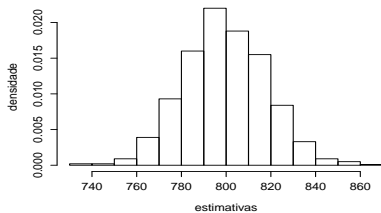


# normal

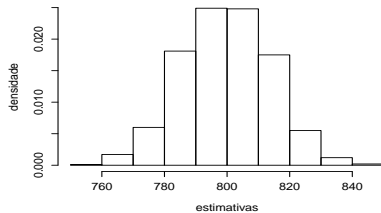


# normal

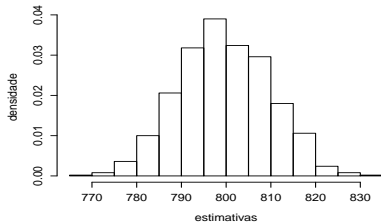
**n = 30 , p-valor (teste-KS) = 0.9536**



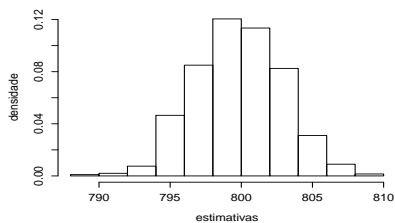
**n = 50 , p-valor (teste-KS) = 0.266**



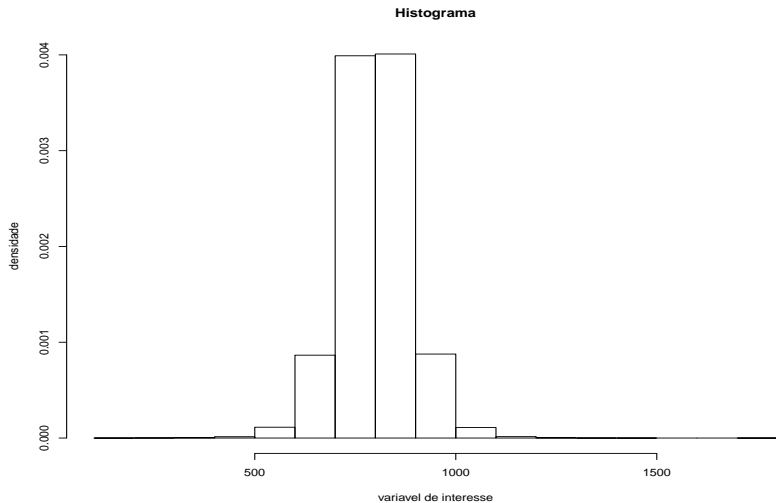
**n = 100 , p-valor (teste-KS) = 0.6521**



**n = 1000 , p-valor (teste-KS) = 0.7906**

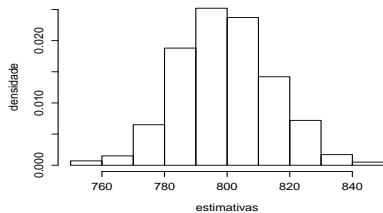


# t de Student

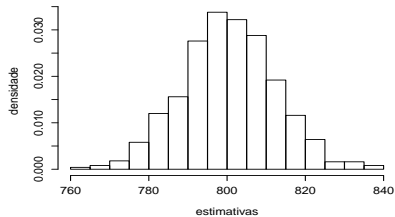


# t de Student

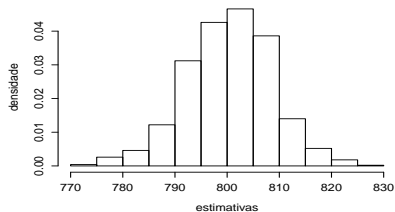
**n = 30 , p-valor (teste-KS) = 0.3804**



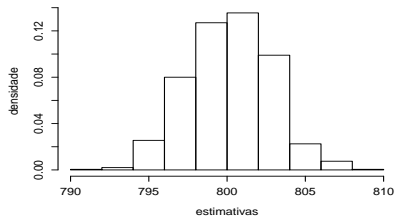
**n = 50 , p-valor (teste-KS) = 0.9877**



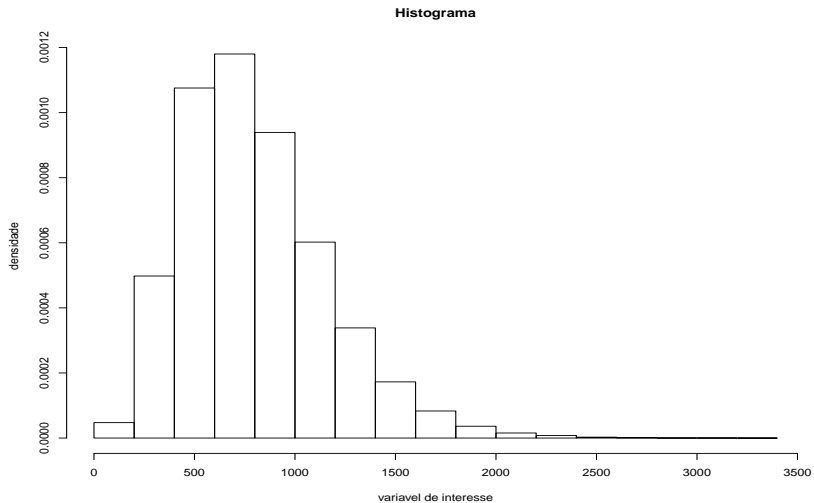
**n = 100 , p-valor (teste-KS) = 0.3352**



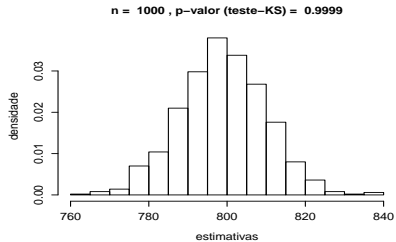
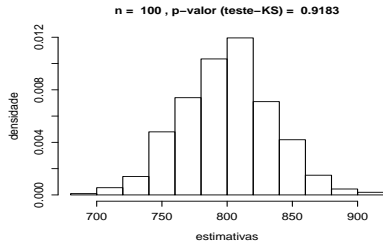
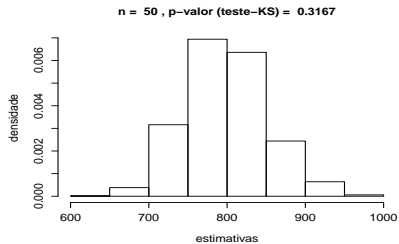
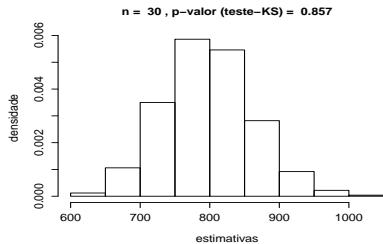
**n = 1000 , p-valor (teste-KS) = 0.802**



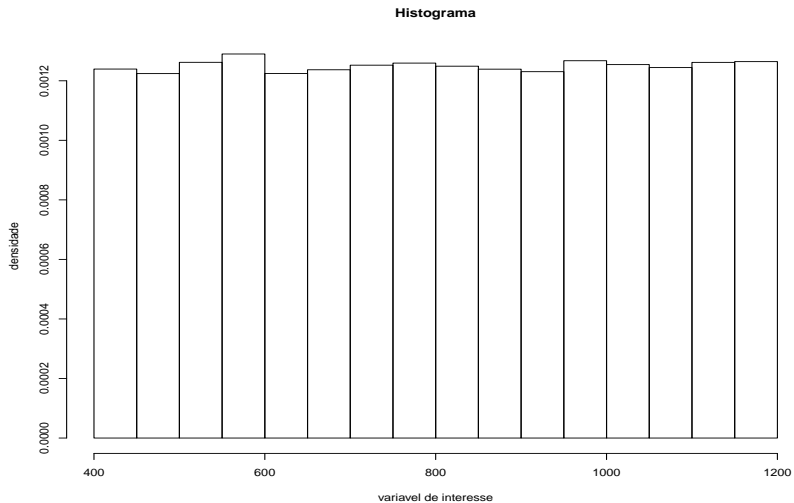
# gama



# gama

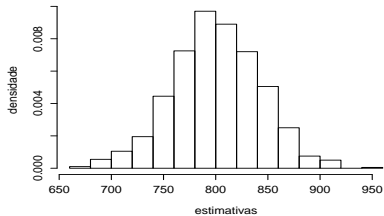


# uniforme

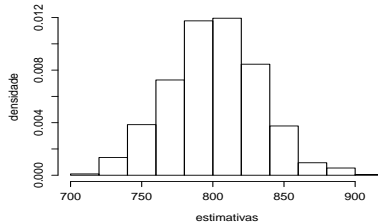


# uniforme

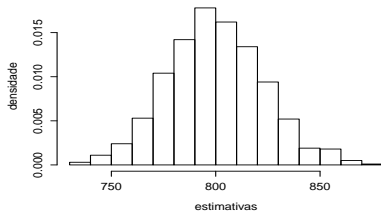
**n = 30 , p-valor (teste-KS) = 0.9761**



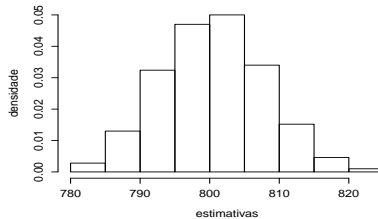
**n = 50 , p-valor (teste-KS) = 0.6903**



**n = 100 , p-valor (teste-KS) = 0.6922**

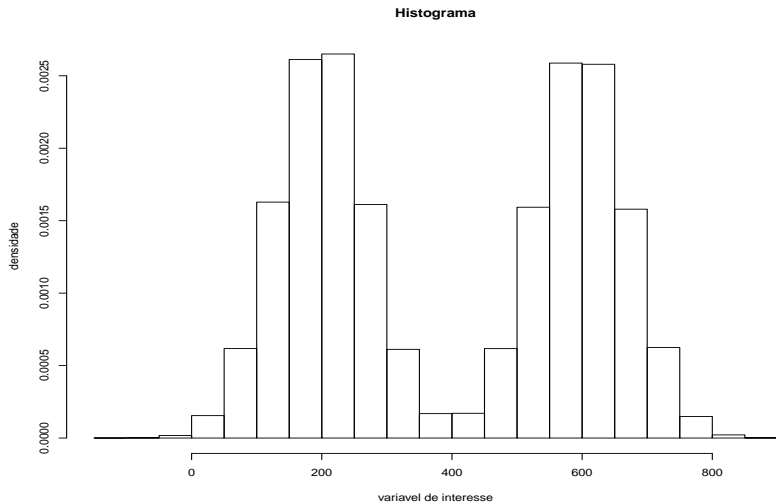


**n = 1000 , p-valor (teste-KS) = 0.8906**



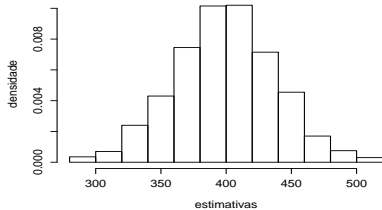


# mistura de duas normais

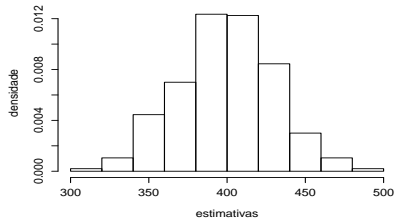


# mistura de duas normais

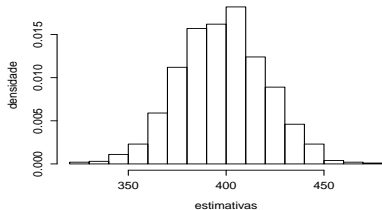
**n = 30 , p-valor (teste-KS) = 0.9695**



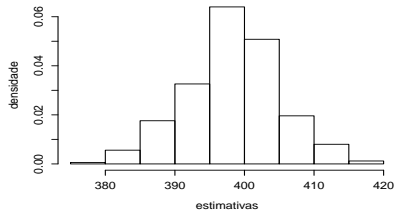
**n = 50 , p-valor (teste-KS) = 0.7368**



**n = 100 , p-valor (teste-KS) = 0.779**



**n = 1000 , p-valor (teste-KS) = 0.2342**



# Determinação do tamanho amostral

- Estabelece-se algum critério de interesse acerca da acurácia/precisão na estimativa da média populacional.
- Sob o estimador proposto, calcula-se o tamanho da amostra, com base em sua distribuição assintótica obtida e critério estabelecido.
- Erro de estimativa:  $z_\gamma \sqrt{\frac{\hat{\sigma}^2}{n}}$ . Fixa-se um erro de estimativa de interesse.
- Probabilidade do módulo da diferença  $P(|\hat{\mu} - \mu| < \delta) > \epsilon$ ,  $\delta > 0$ ,  $\epsilon \in (0, 1)$ .

# Determinação do tamanho amostral: erro da estimativa

$$\delta = z_{\gamma} \sqrt{\frac{\sigma^2}{n}} \rightarrow n = \frac{z_{\gamma}^2 \sigma^2}{\delta^2}$$

Em geral, o (um) valor de  $\sigma^2$  é obtido através de pesquisas anteriores ou de uma amostra piloto, de tamanho apropriado.

## Determinação do tamanho amostral: precisão

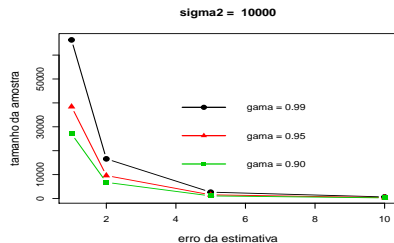
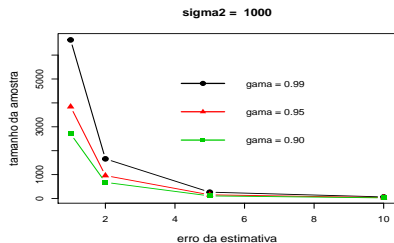
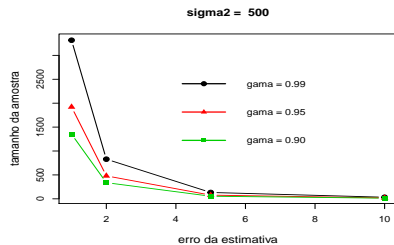
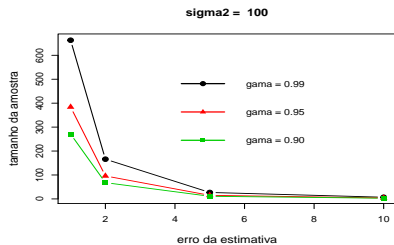
$$\begin{aligned} P_{A_1} (|\hat{\mu} - \mu| < \delta) > \epsilon &\Leftrightarrow P_{A_1} \left( \left| \frac{\hat{\mu} - \mu}{\sqrt{\hat{\sigma}^2/n}} \right| < \frac{\sqrt{n}\delta}{\sigma} \right) > \epsilon \\ &\Leftrightarrow P_{A_1} \left( |Z| < \frac{\sqrt{n}\delta}{\sigma} \right) > \epsilon \Leftrightarrow \frac{\sqrt{n}\delta}{\sigma} = z_\gamma \end{aligned}$$

em que  $Z \approx N(0, 1)$ . O que leva ao mesmo procedimento oriundo de se fixar o erro da estimativa.

# Tamanhos amostrais

- Situações hipotéticas: cruzamento entre os níveis de diferentes fatores de interesse
  - $\delta \in \{1, 2, 5, 10\}$ .
  - $\gamma \in \{0, 9; 0, 95; 0, 99\}$ .
  - $\sigma^2 \in \{100, 500, 1000, 10000\}$ .

# Tamanhos amostrais



# Estimação do total populacional

- $\tau = \sum_{i=1}^N y_i = N\mu.$
- Estimador “natural”:  $\hat{\tau}_u = \sum_{i=1}^n Y_i.$  Problema: se os  $y_i$ 's foram positivos,  $\hat{\tau}_u$  sempre subestimar $\hat{a}$   $\tau.$
- Alternativa  $\hat{\tau} = N\hat{\mu}.$
- Estimativa  $\tilde{\tau} = N\tilde{\mu}$



# Propriedades do estimador

- $\mathcal{E}_{A_1}(\hat{\tau}) = \mathcal{E}_{A_1}(N\hat{\mu}) = N\mathcal{E}(\hat{\mu}) = N\mu = \tau$  (não viciado).
- $\mathcal{V}_{A_1}(\hat{\tau}) = N^2\mathcal{V}_{A_1}(\hat{\mu}) = N^2\frac{\sigma^2}{n}$  (a imprecisão associada à estimação do total é maior do que aquela associada à média).

# Propriedades do estimador

- Normalidade assintótica, como

$$\frac{\hat{\mu} - \mu}{\sqrt{\hat{\sigma}^2/n}} \xrightarrow[n \rightarrow \infty, N-n \rightarrow \infty]{D} N(0, 1),$$

lembrando que  $N$  é fixo, temos que

$$\frac{N\hat{\mu} - N\mu}{\sqrt{N^2\hat{\sigma}^2/n}} \xrightarrow[n \rightarrow \infty, N-n \rightarrow \infty]{D} N(0, 1) \rightarrow \frac{\hat{\tau} - \tau}{\sqrt{N^2\hat{\sigma}^2/n}} \xrightarrow[n \rightarrow \infty, N-n \rightarrow \infty]{D} N(0, 1)$$

# Intervalo de Confiança

- Estimativa:  $\tilde{\sigma}^2 = \frac{1}{n-1} \sum_{i \in \mathcal{S}} (y_i - \tilde{\mu})^2$ .
- Assim, um intervalo de confiança (assintótico) com coeficiente de confiança de aproximadamente  $\gamma$  é dado por

$$IC(\mu, \gamma) \approx \left[ \hat{\mu} - z_\gamma N \sqrt{\frac{\hat{\sigma}^2}{n}}; \hat{\mu} + z_\gamma N \sqrt{\frac{\hat{\sigma}^2}{n}} \right]$$

em que  $P(Z \leq z_\gamma) = \frac{1+\gamma}{2}$  e  $Z \sim N(0, 1)$ .

- Erro da estimativa:  $z_\gamma N \sqrt{\frac{\hat{\sigma}^2}{n}}$ .

# Testes de Hipótese

- Hipóteses usuais ( $\tau_0$  conhecido)
  - 1  $H_0 : \tau = \tau_0$  vs  $H_1 : \tau < \tau_0$ .
  - 2  $H_0 : \tau = \tau_0$  vs  $H_0 : \tau > \tau_0$ .
  - 3  $H_0 : \tau = \tau_0$  vs  $H_0 : \tau \neq \tau_0$ .
- Estatística do teste  $Z_t = \frac{\hat{\tau} - \tau_0}{N\hat{\sigma}/\sqrt{n}}$ , em que  $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ .
- Sob  $H_0$ , vimos que  $Z_t \approx N(0, 1)$ , para  $n$  e  $N-n$  suficientemente grandes.
- Defina  $z_t = \frac{\hat{\tau} - \tau_0}{N\hat{\sigma}/\sqrt{n}}$  o valor calculado da estatística do teste e  $z_c$  o(s) valor(es) crítico(s).
- Defina ainda  $Z \sim N(0, 1)$ . Os procedimentos são análogos ao caso da média, com as devidas adaptações.

# Determinação do tamanho amostral: erro da estimativa

Analogamente ao caso da média populacional, temos que

$$\delta = z_{\gamma} N \sqrt{\frac{\sigma^2}{n}} \rightarrow n = \frac{z_{\gamma}^2 N^2 \sigma^2}{\delta^2}$$

Isto vale para qualquer um dos dois critérios: erro da estimativa e precisão.