

Mais sobre amostragem aleatória simples sem reposição

Prof. Caio Azevedo

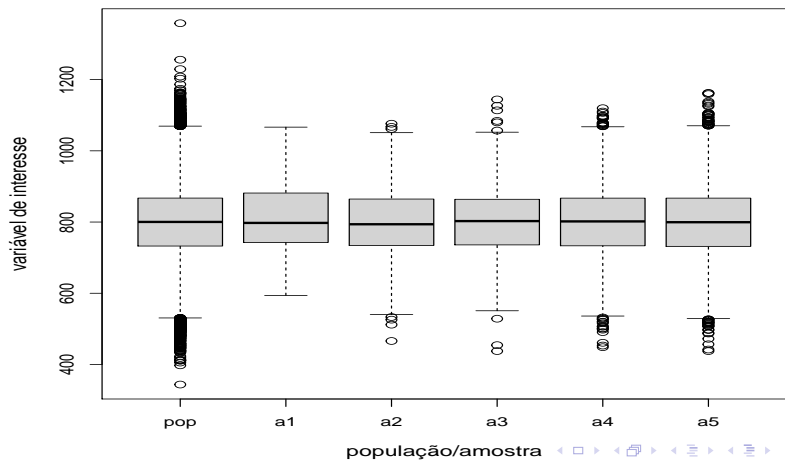
Estudos de simulação: estimação da média

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

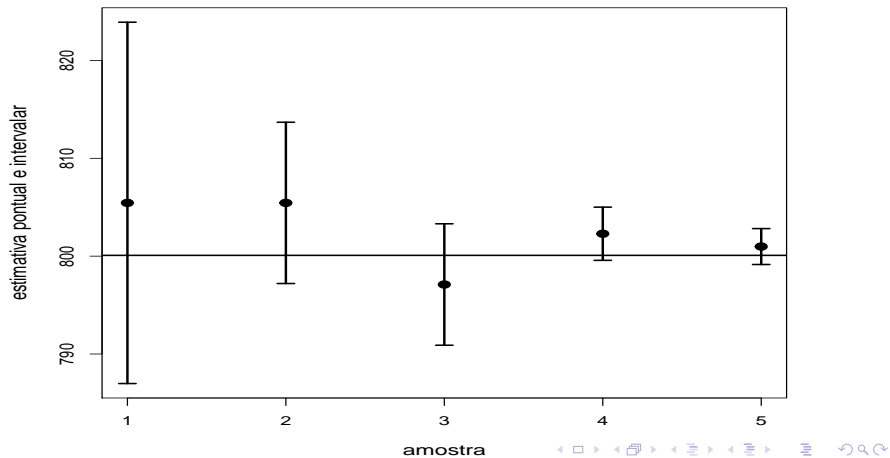
Estudos de simulação

- Quatro tamanhos amostrais (100, 500, 1000, 5000, 10000), em termos percentuais, com relação ao tamanho da população (0,01%,0,05%,0,1%,0,5%,1%).
- Estudar a distribuição da variável resposta na amostra em relação à população bem como o comportamento do estimador para a média.

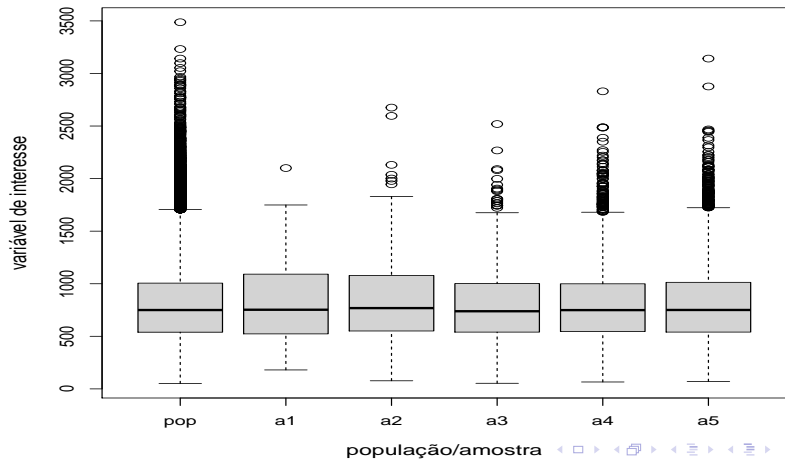
normal



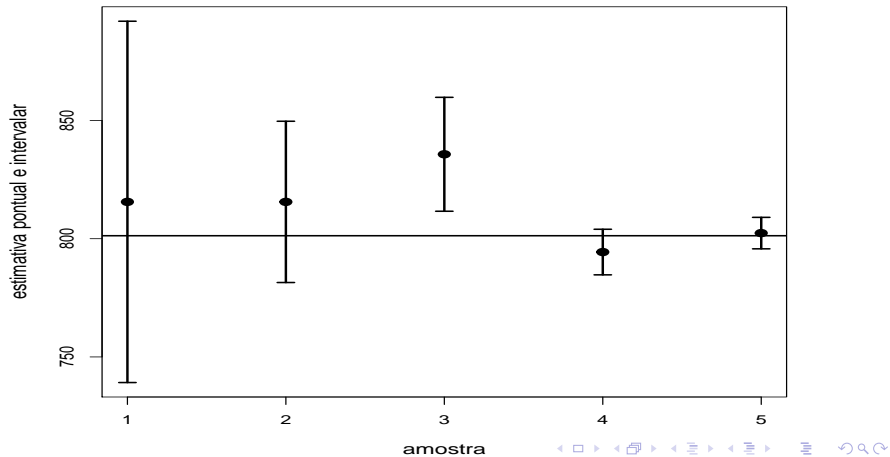
normal



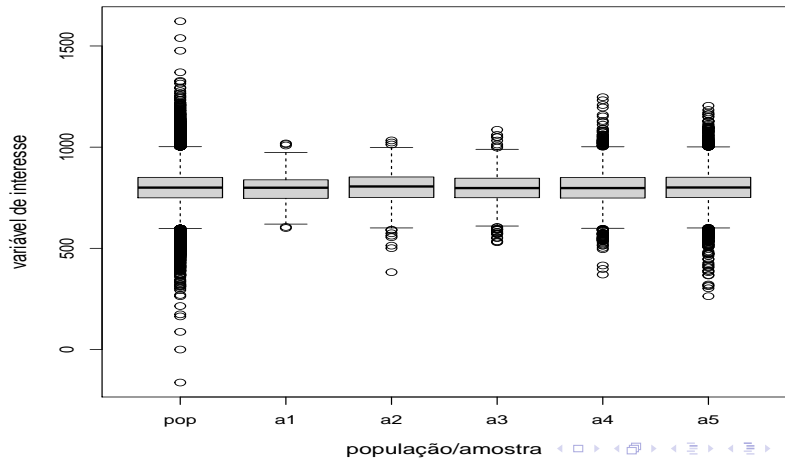
gama



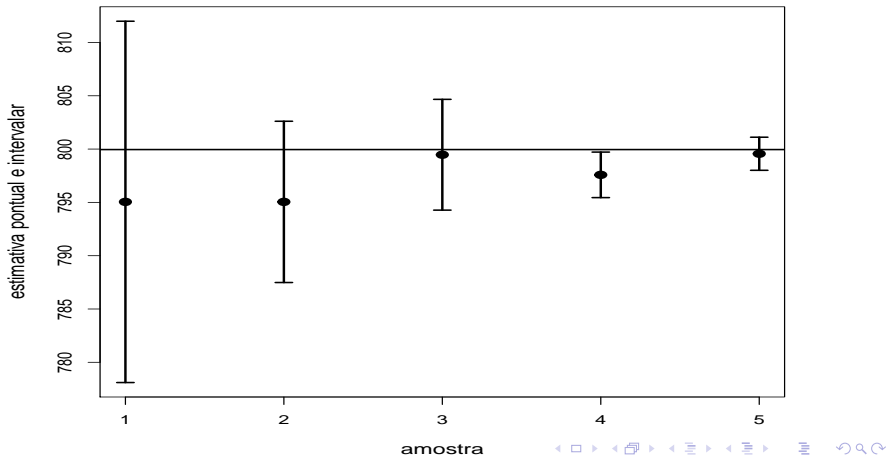
gama



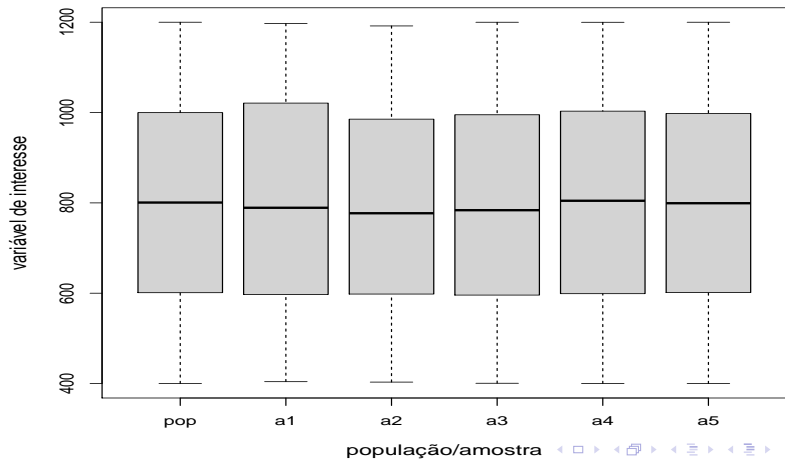
t de Student



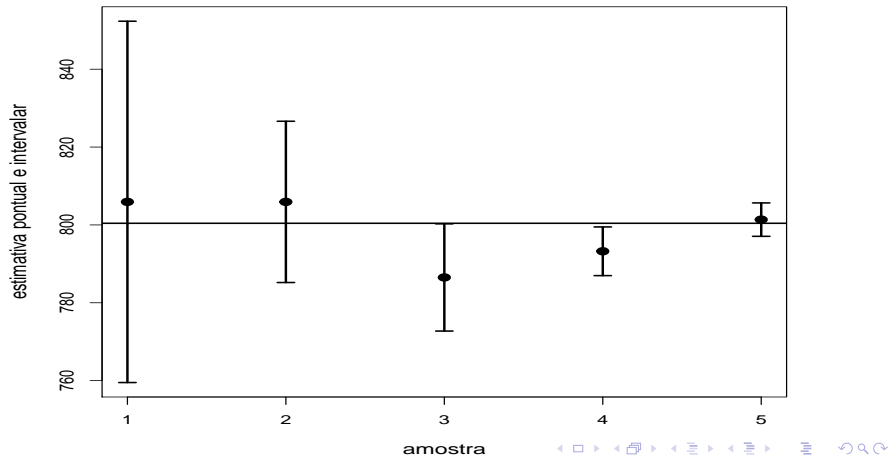
t de Student



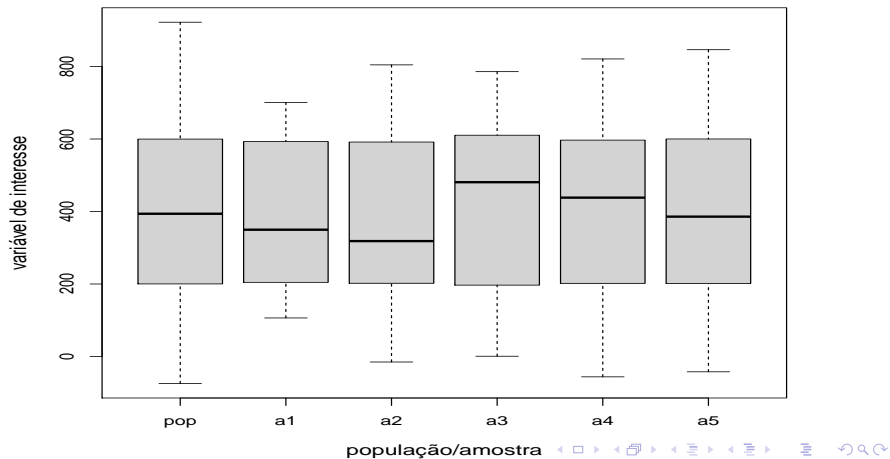
uniforme



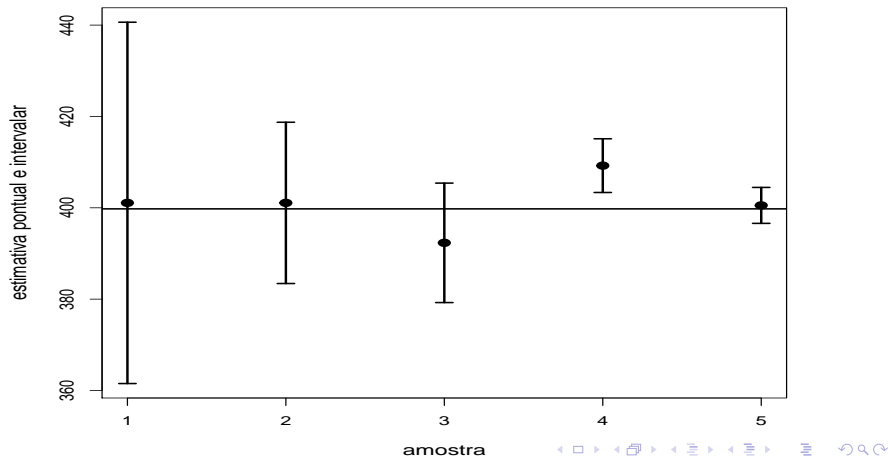
uniforme



mistura de normais



mistura de normais



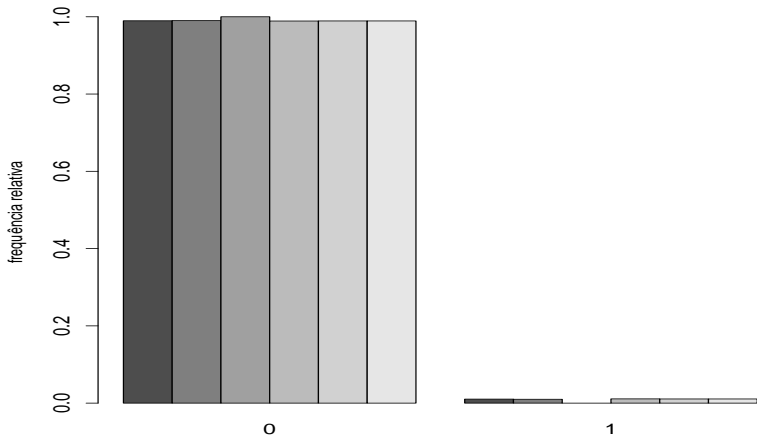
Estudos de simulação: estimação da proporção

- Distribuição da variável de interesse na amostra em relação à distribuição na população sob AAS_s .
- Avaliar a acurácia do estimador.
- Tamanho da população $N = 100000$.

Estudos de simulação: estimação da proporção

- Vários cenários, variando em função do valor verdadeiro da proporção populacional p .
- $p = (0,01; 0,05; 0,10; 0,25; 0,35; 0,50; 0,65; 0,75; 0,9; 0,95; 0,99)^t$.
- A distribuição, (em princípio), da variável de interesse é Bernoulli(p).
- Quatro tamanhos amostrais (100, 500, 1000, 5000, 10000), em termos percentuais, com relação ao tamanho da população (0,01%, 0,05%, 0,1%, 0,5%, 1%).

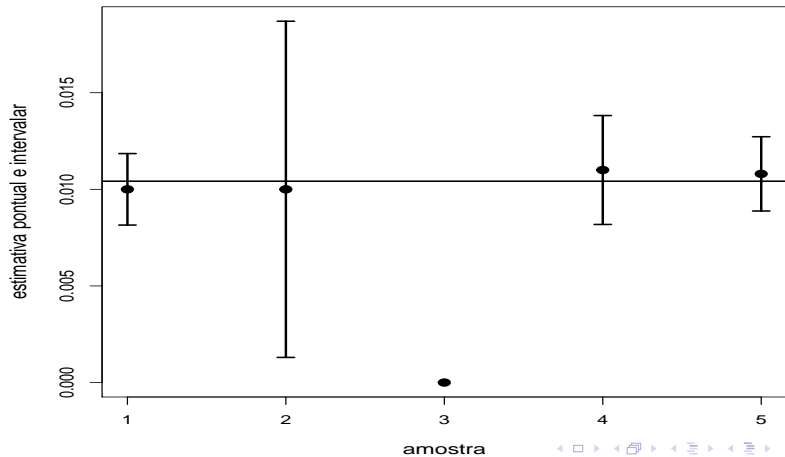
$p=0,01$



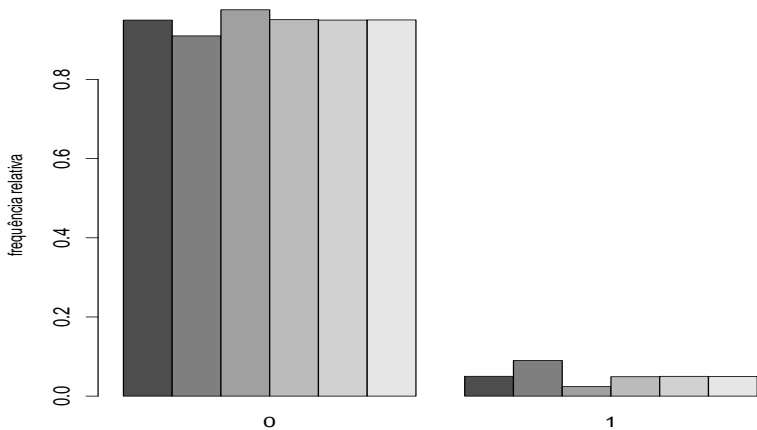
valores



$p=0,01$



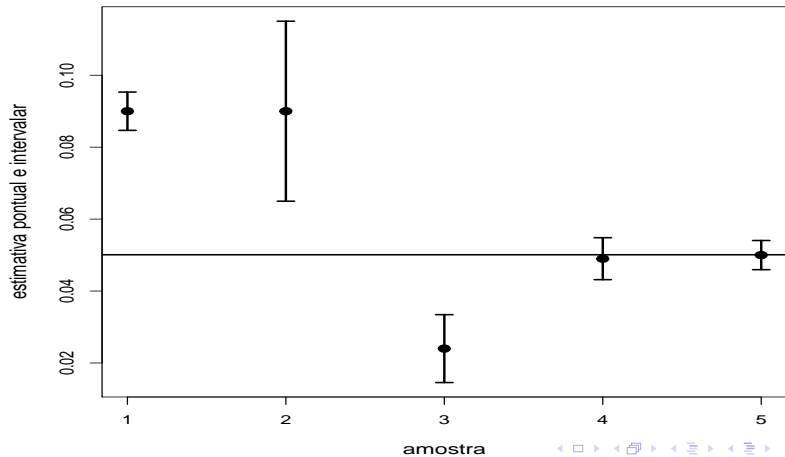
$p=0,05$



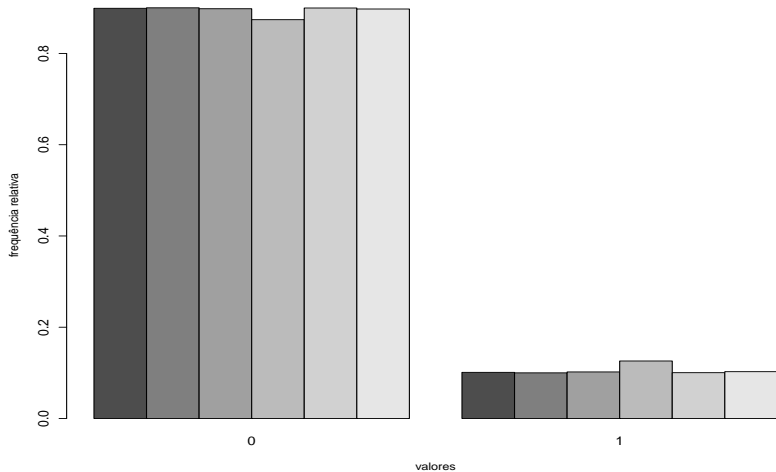
valores



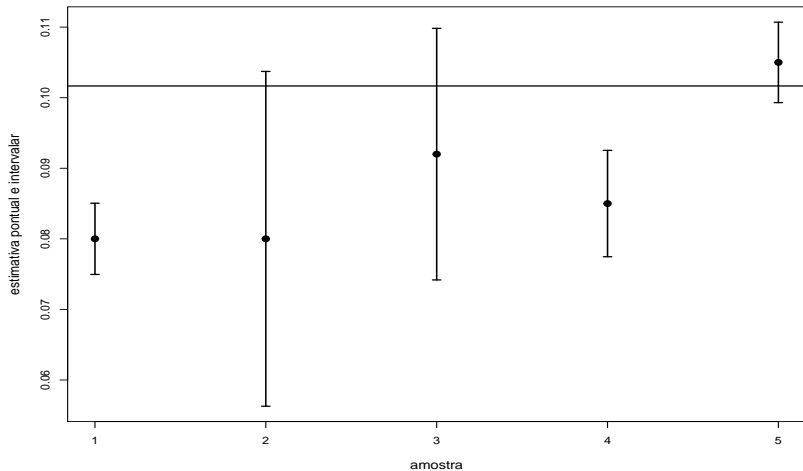
$p=0,05$



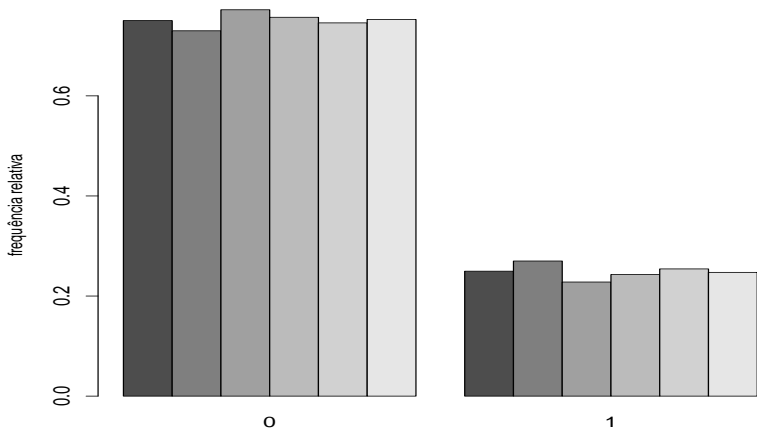
$$p=0,1$$



$p=0,1$



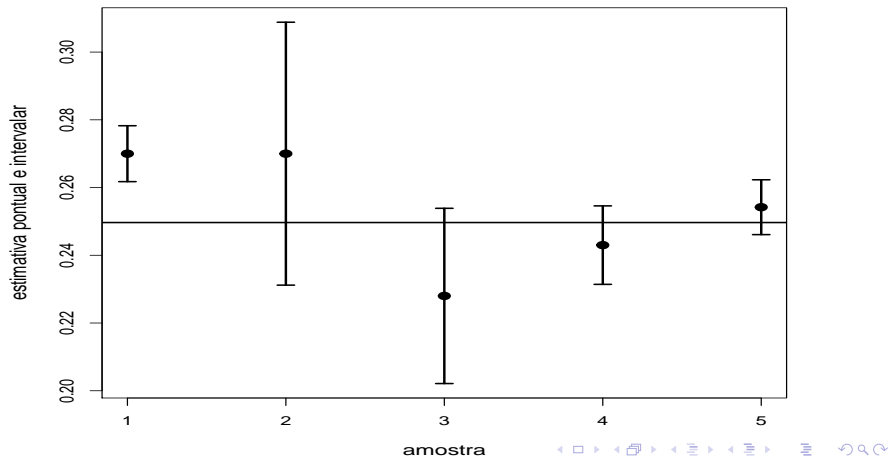
$p=0,25$



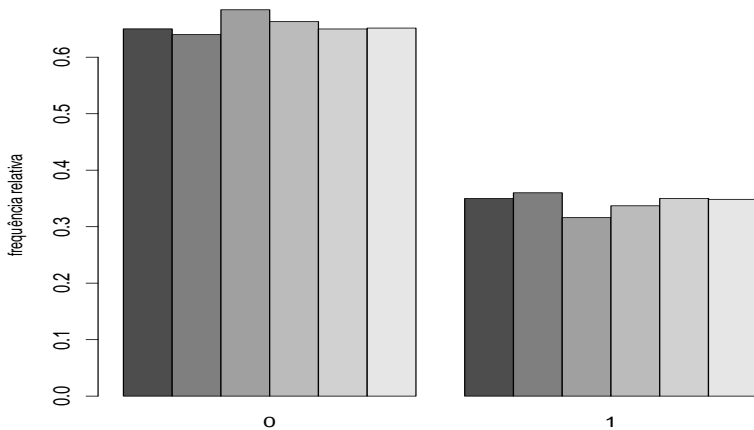
valores



$p=0,25$



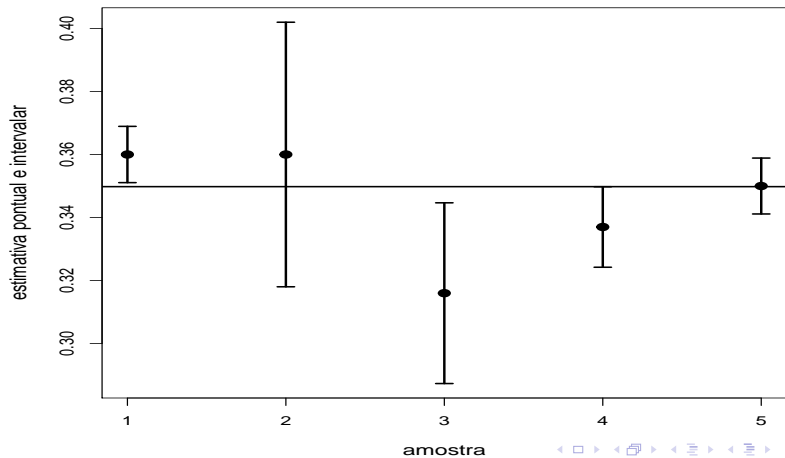
$p=0,35$



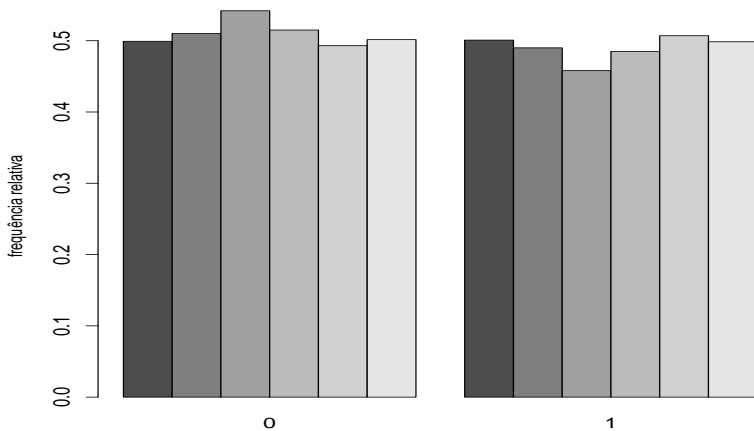
valores



$p=0,35$



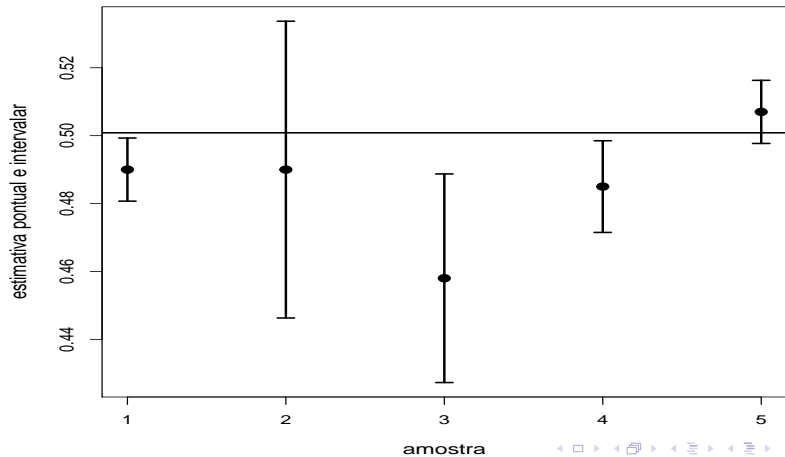
$p=0,50$



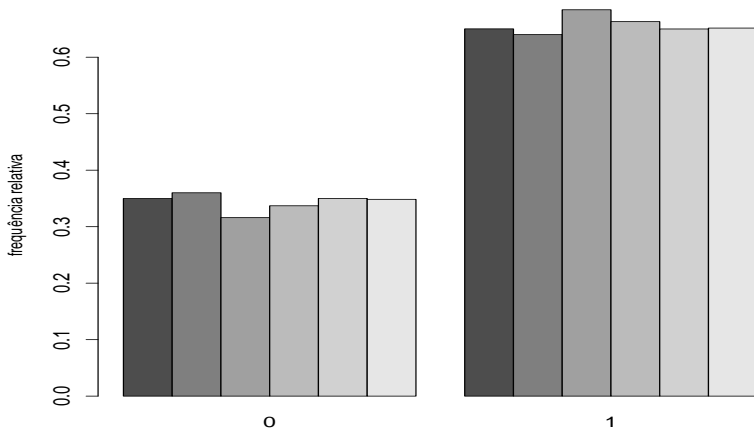
valores



$p=0,50$



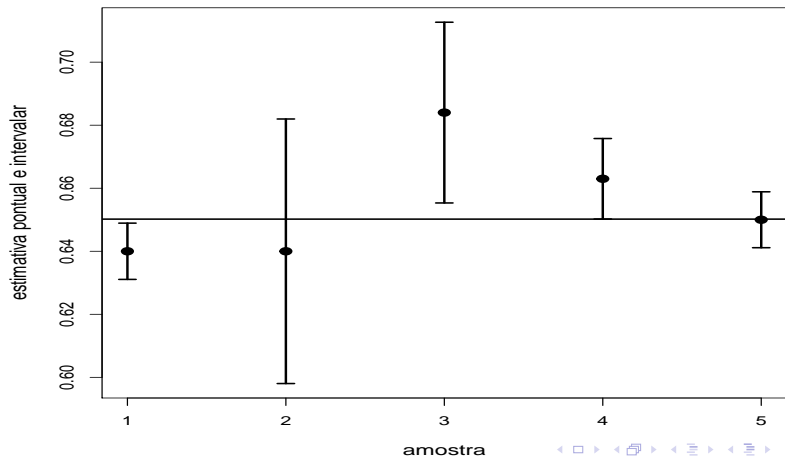
$p=0,65$



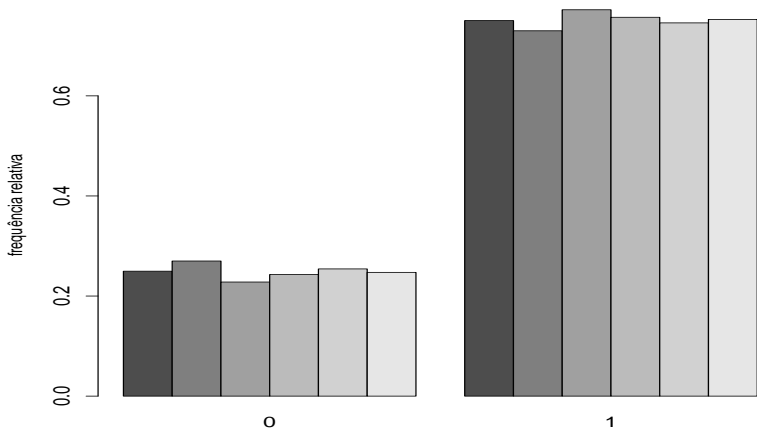
valores



$p=0,65$



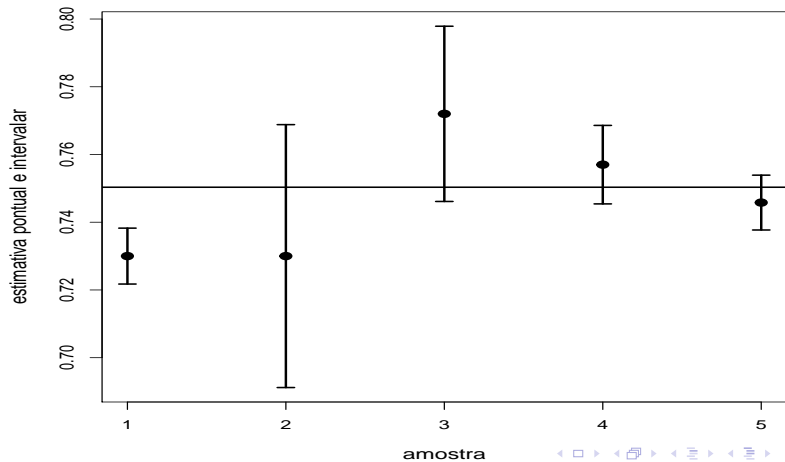
$p=0,75$



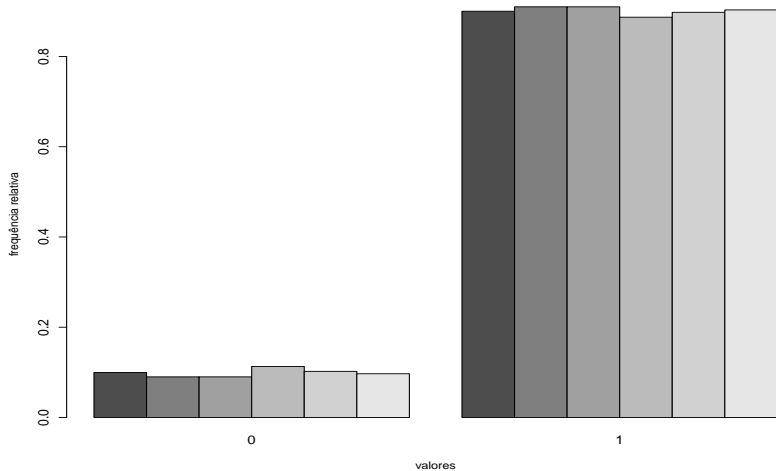
valores



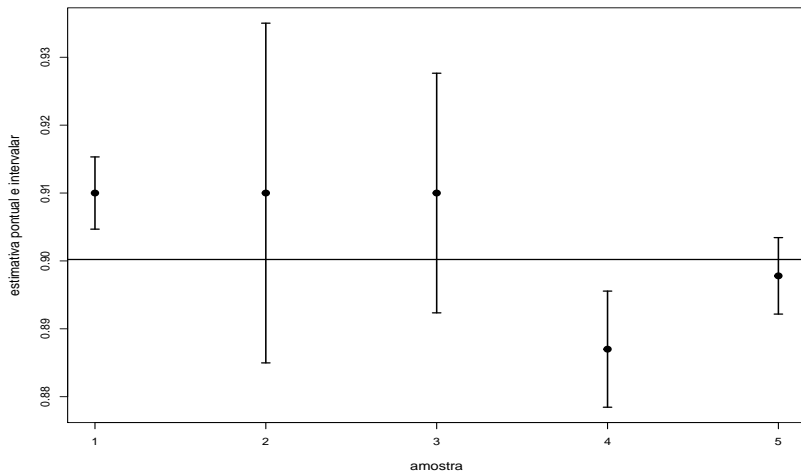
$p=0,75$



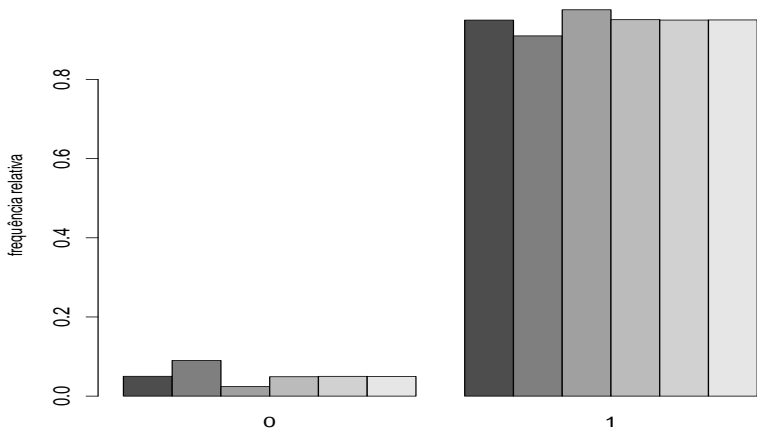
$p=0,90$



$p=0,90$



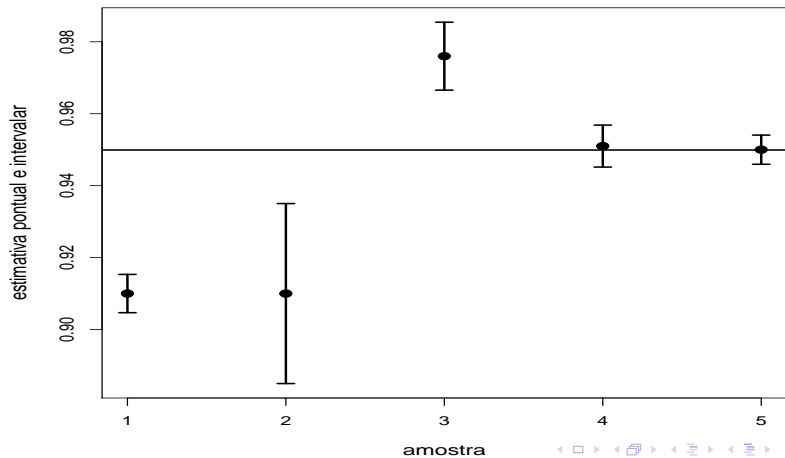
$p=0,95$



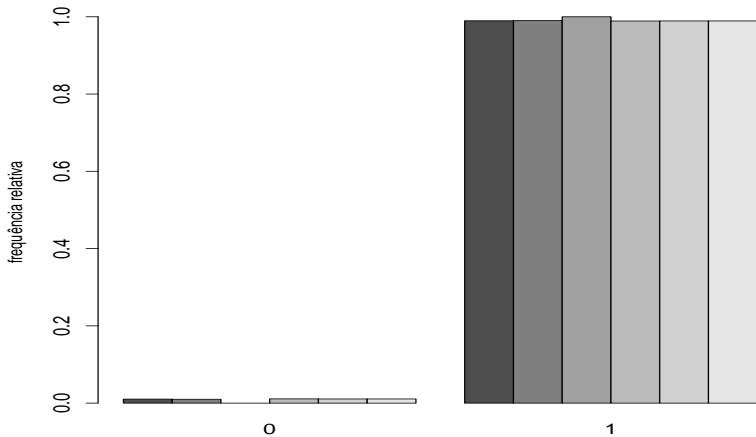
valores



$p=0,95$



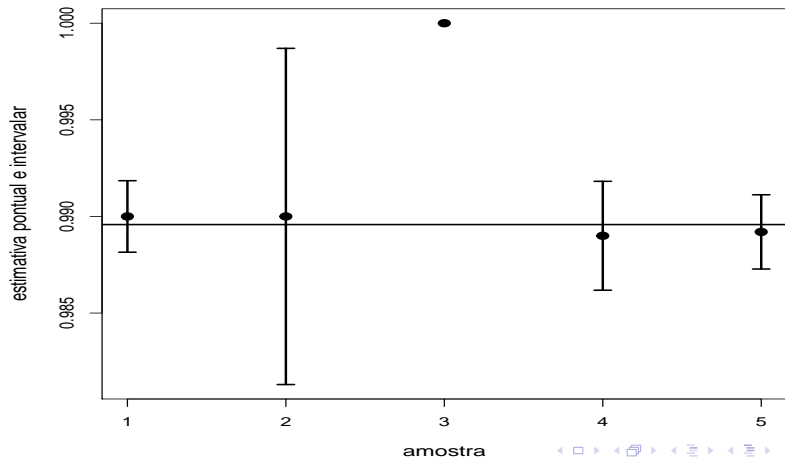
$p=0,99$



valores



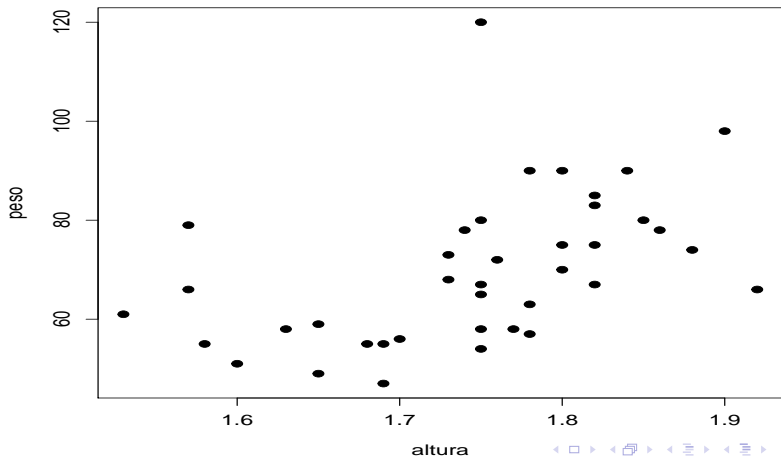
$p=0,99$



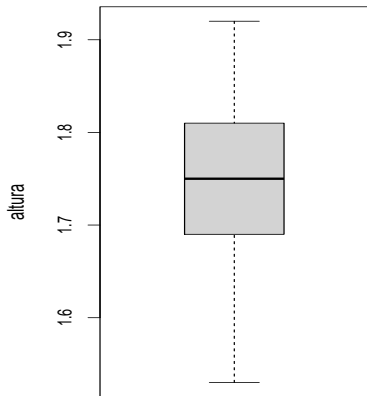
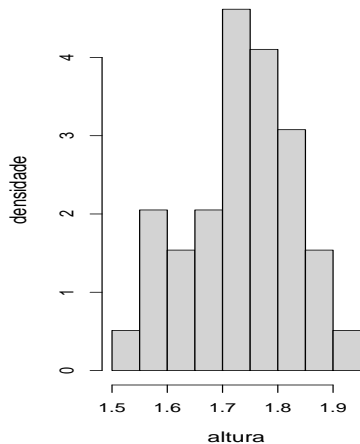
Análise dos dados de uma dada turma

- Este exemplo não é realístico pois o tamanho da população é muito pequeno $N = 39$.
- Apenas para ilustrar algumas das técnicas vistas em sala.
- Estimar a média da altura e do peso bem como a proporção de alunos procedentes da RMC (região metropolitana de Campinas).

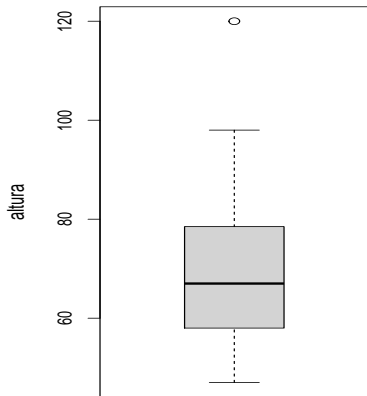
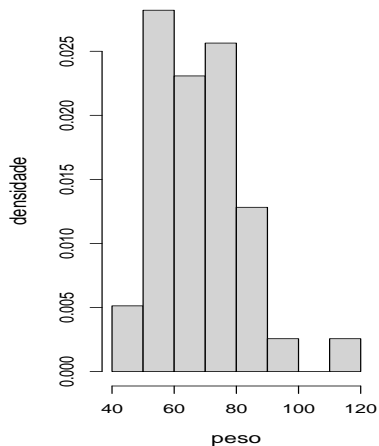
Gráfico de dispersão: altura \times peso (cor = 0,473)



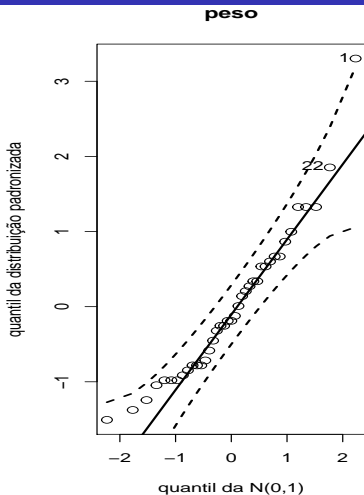
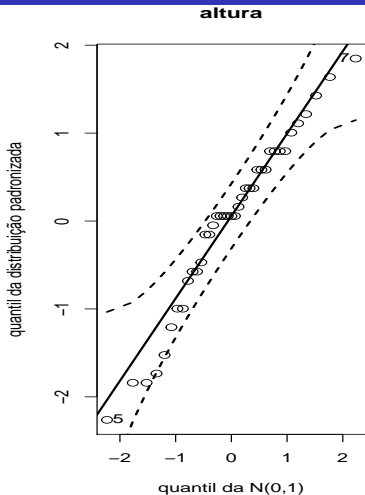
Histograma e boxplot: altura



Histograma e boxplot: peso



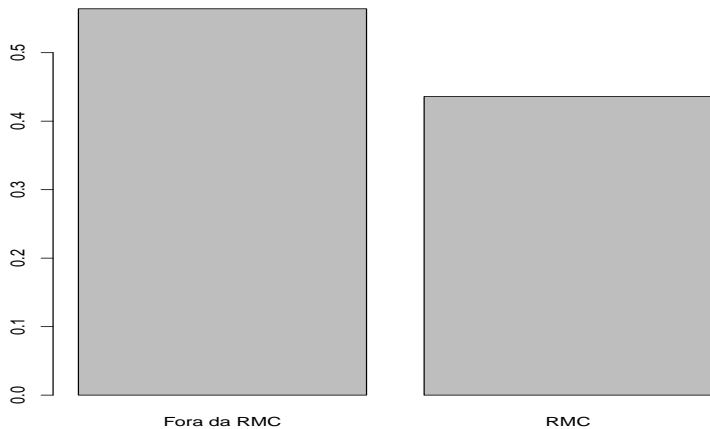
Gráficos de quantil quantil (normalidade)



Medidas resumo

Variável	Média	Var.	DP	CV(%)	Min.	Med.	Max.	CA
altura	1,74	0,01	0,09	5.44	1,53	1,75	1,92	-0,43
peso	69,87	230,38	15,18	21.72	47,00	67,00	120,00	0,94

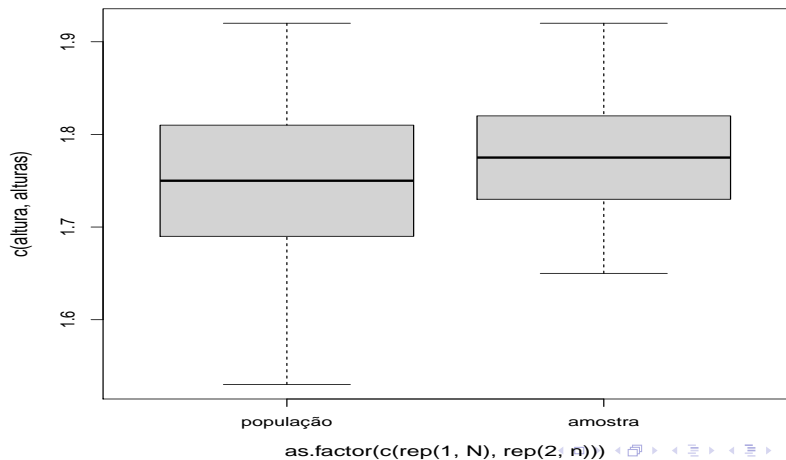
Distribuição da procedência ($p(\text{RMC}) = 43,59\%$)



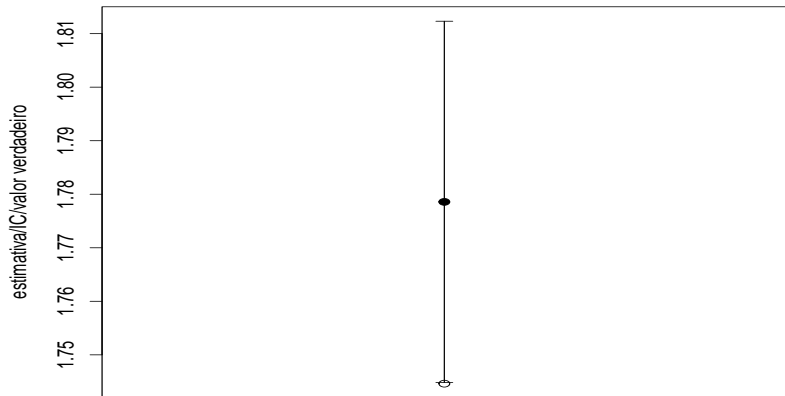
Estimação da média da altura

- $\delta = 0,04$ m (precisão), $\tilde{s}^2 = 0,01m$ (populacional), $\gamma = 0,95$ (nível de confiança), $N = 39$.
- $n = \frac{1}{\frac{\delta^2}{s^2 z_\gamma^2} + \frac{1}{N}} = 14$.
- Resultados: $\tilde{\mu} = 1,74m$; $IC(95\%) = [1,70m; 1,78m]$

Box-plot das alturas: população \times amostra



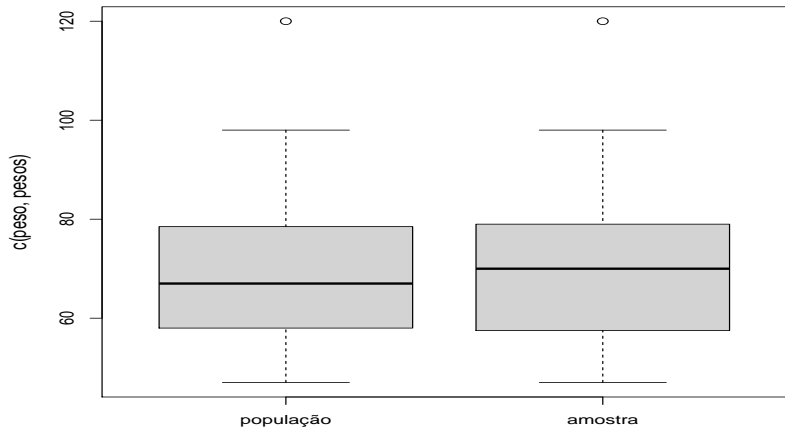
Valor verdadeiro, estimativa pontual e intervalar



Estimação da média do peso

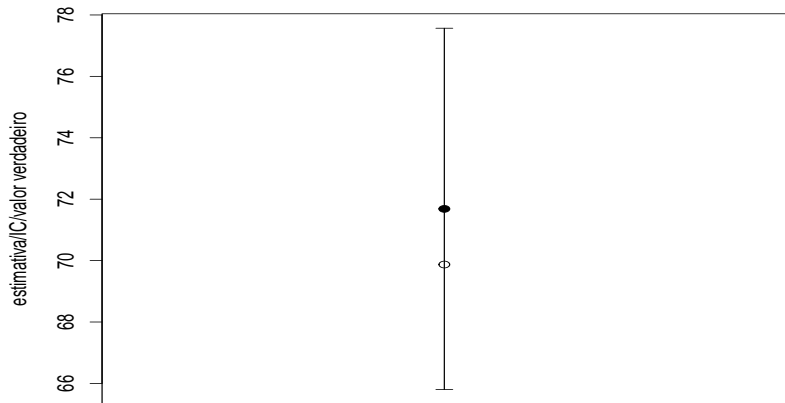
- $\delta = 5,00(\text{kg})$ (precisão), $\tilde{s}^2 = 230,38\text{kg}$ (populacional), $\gamma = 0,95$ (nível de confiança), $N = 39$.
- $n = \frac{1}{\frac{\delta^2}{s^2 z_\gamma^2} + \frac{1}{N}} = 19$.
- Resultados: $\tilde{\mu} = 73,32\text{kg}$; $IC(95\%) = [67,95\text{kg}; 78,68\text{kg}]$

Box-plot dos pesos: população \times amostra



`as.factor(c(rep(1, N), rep(2, n)))`

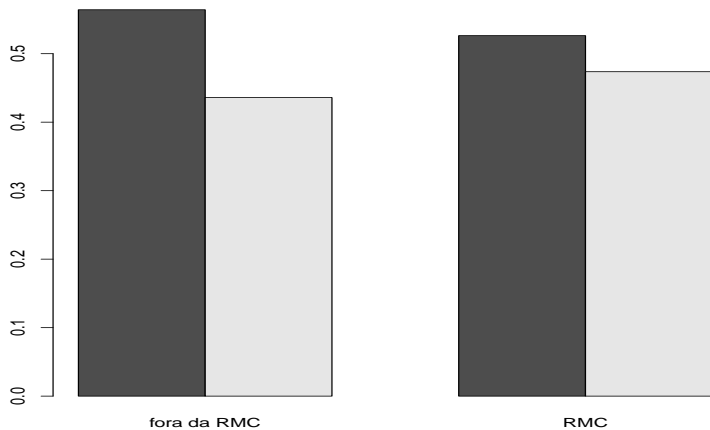
Valor verdadeiro, estimativa pontual e intervalar



Estimação da proporção de oriundos da RMC

- $\delta = 0,05$ (precisão), $\gamma = 0,95$ (nível de confiança), $N = 39$.
- Se usássemos $n = \frac{1}{\frac{4\delta^2}{z_\gamma^2} + \frac{1}{N}} = 35$, ou seja, quase toda a população.
- Dessa forma, vamos considerar $n = 19$ (tamanho determinado para a estimação da média do peso) e verificaremos qual será a precisão (erro da estimativa).
- $\delta = z_\gamma \sqrt{(1-f) \frac{pq}{n}} = 1,96 \sqrt{(1 - \frac{19}{39}) \frac{pq}{19}}$
- Resultados: $\tilde{p} = 0,32$, $IC(95\%) = [0,17; 0,46]$.
- Assim, $\delta = 0,15$.

Box-plot dos pesos: população \times amostra



Valor verdadeiro, estimativa pontual e intervalar

