Amostragem por conglomerados em um único estágio (AC): parte 3 (mais aplicações)

Prof. Caio Azevedo

Exercício 7.5: Bolfarine & Bussab (2005)

- (Controle de gastos) Uma companhia que fornece carros a seus vendedores quer uma estimativa do número médio de quilômetros percorridos pelos seus carros no ano passado. A companhia tem A=12 filiais. O número de carros (B_α), a média (μ_α) e a variância (s_α^2) do número de quilômetros percorridos (em milhares), para cada filial, são dados na Tabela do slide seguinte:

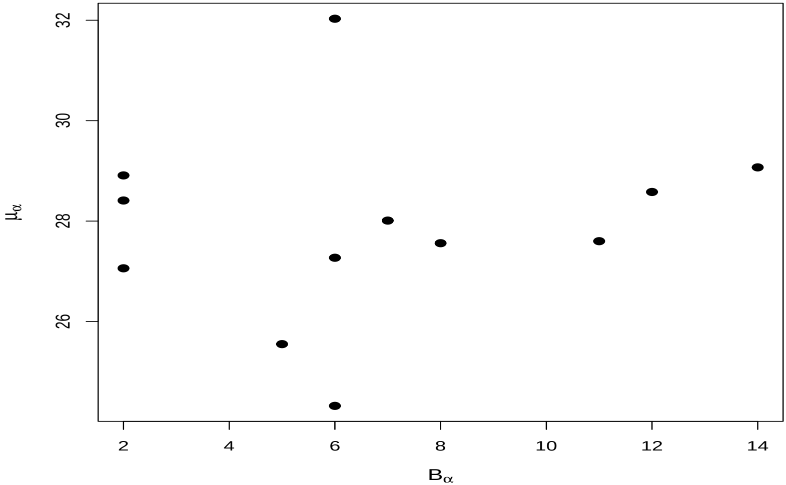
Cont. (dados)

filial	B_α	μ_α	s^2_α
1	6	24,32	5,07
2	2	27,06	5,53
3	11	27,60	6,24
4	7	28,01	6,59
5	8	27,56	6,21
6	14	29,07	6,12
7	6	32,03	5,97
8	2	28,41	6,01
9	2	28,91	5,74
10	5	25,55	6,78
11	12	28,58	5,87
12	6	27,27	5,38

Resolução

- Vamos estimar μ, via AC, considerando uma AAS_s de $a = 4$ conglomerados, comparando os respectivos resultados sob uma AAS_s com $n = 27$.
- Exercício: repetir considerando $a = 5, a = 6, a = 7$ sob AAS_c e AAS_s.
- Note que, neste caso, temos acesso as informações (de interesse) de toda a população, com exceção das observações individuais (y, que se referem aos quilômetros percorridos por cada carro).
- Faremos análises descritivas e inferenciais.

Dispersão entre B_α e μ_α (população)



Resultados populacionais

1. $Corre(B_\alpha, \mu_\alpha) = 0,136(0,313)$ (teste de nulidade, p=0,6718).
2. Modelo de regressão: $\mu_\alpha = \gamma_0 + \gamma_1 B_\alpha + E_\alpha$, $E_\alpha \overset{iid}{\sim} N(0, \sigma^2_\alpha)$ - $\widetilde{\gamma}_1 = 0,06(0,15)$ (teste de nulidade, p = 0,6718).
3. Os resultados 1) e 2), de acordo com link, sugerem que o estimador $\widehat{\mu}_{C_3}$ apresentará desempenho superior aos outros dois.

Resultados populacionais

4. Além disso, usando $s^2_{(r)} = \frac{A}{A-1}\sigma^2_{(r)}$, $r \in \{ec, ect, eq, em\}$ (veja aqui, slides 15 a 17), calculamos o CCI usando a fórmula da página 36 de link, ou seja:

	ec	ect	eq	em
ρ_{int}	0,212	0,974	0,202	0,243

5. O resultado 4) indica que, possivelmente, as inferências via AC (independentemente do estimador) serão menos acuradas do que aquela via AAS.

Resultados inferenciais (amostra)

- $Corre(b_{\alpha}, \widetilde{\mu}_{\alpha}) = -0,375(0,6243)$ (teste de nulidade, p=0,6718).
- Modelo de regressão: $\widetilde{\mu}_{\alpha} = \gamma_0 + \gamma_1 b_{\alpha} + E_{\alpha}$, $E_{\alpha} \overset{iid}{\sim} N(0, \sigma^2_{\alpha})$ - $\widetilde{\gamma}_1 = -0,31(0,55)$ (teste de nulidade, p = 0,6243).
- Estimativas pontuais e intervalares:

Estatística	$\widehat{\mu}_{C_1}$	$\widehat{\mu}_{C_2}$	$\widehat{\mu}_{C_3}$
Est.	20,83	26,78	27,00
EP	14,09	0,79	0,76
Vício (esperado)	0,00	1,27	-1,00
Vício (observado)	-7,17	-1,22	-1,00
EQM	198,57	2,23	1,57

Continuação

- O vício (observado) só pode ser calculado pois temos acesso ao parâmetro de interesse.
- Escolhemos o estimador $\widehat{\mu}_{C_3}$, assim:

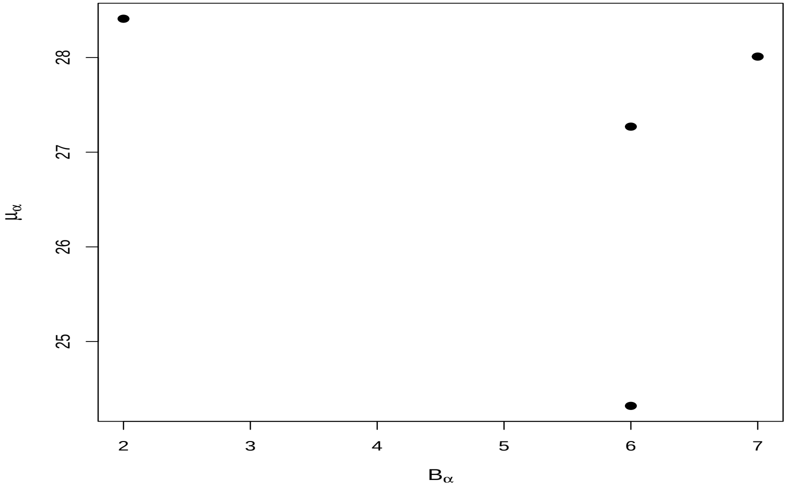
Estimativa	EP	IC(95%)
27,00	0,79	[25,52 ; 28,48]

- Temos também que $\widetilde{\rho}_{C_3} = 0,188$.

Continuação

- Além disso, $EPA = 1 + \widetilde{\rho}_{C_3}\left(\overline{B} - 1\right) = 2,08$ (em relação ao PA AAS_s). Também, $EPA = \dfrac{V_{AC_2}\left(\widehat{\mu}_{C_3}\right)}{V_{A_2}\left(\widehat{\mu}\right)} = 2,08$.
- Contudo, neste caso, não temos como utilizar AAS_s haja vista que não temos informações em relação aos carros de cada filial (unidade elementar).

Dispersão entre B_α e μ_α (amostra)

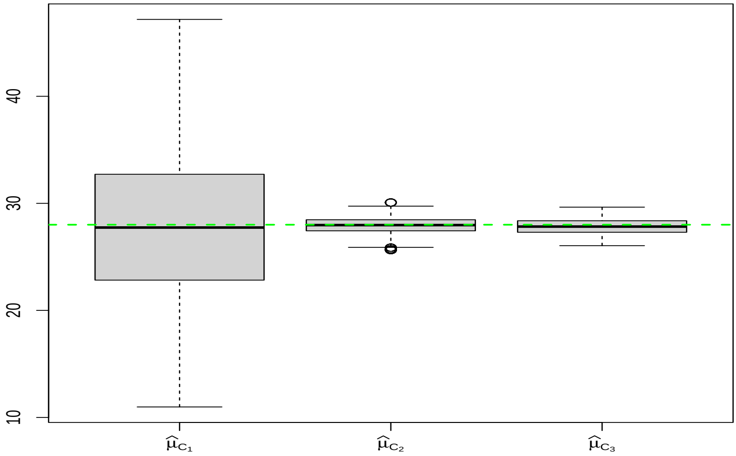


Pequeno estudo de simulação

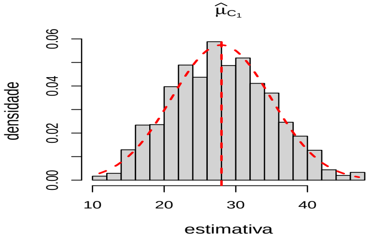
- Foram sorteadas R = 5.000 amostras de tamanho a = 4 (AC_2) e, para cada amostra, calculamos a estimativa para cada estimador.
- Construímos box-plots, histogramas e medidas resumo. :

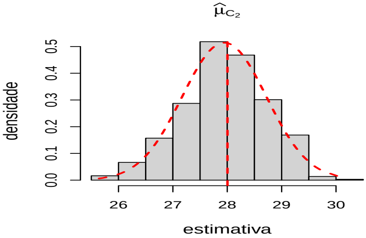
Estatística	$\widehat{\mu}_{C_1}$	$\widehat{\mu}_{C_2}$	$\widehat{\mu}_{C_3}$
média	27,97	27,95	27,86
variância	48,33	0,60	0,58
vício	-0,03	-0,06	-0,15
eqm	48,33	0,61	0,61
rqeqm	6,95	0,78	0,78

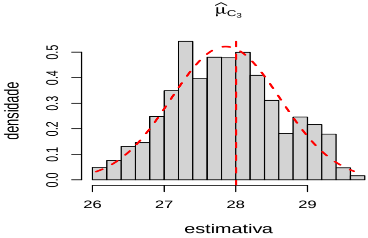
Boxplots



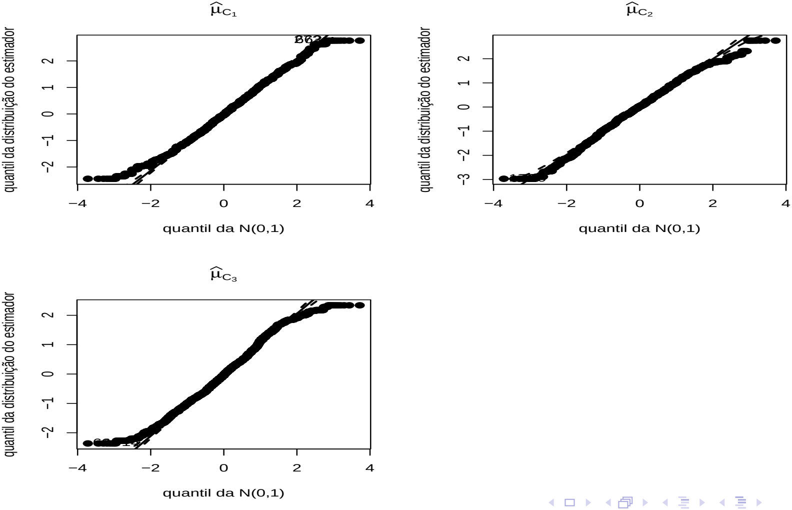
Histogramas







QQ plots



$\widehat{\mu}_{C_1}$
$\widehat{\mu}_{C_2}$
$\widehat{\mu}_{C_3}$
quantil da distribuição do estimador
quantil da N(0,1)
−4 −2 0 2 4

Comentários

- Como esperado, o melhor estimador foi $\widehat{\mu}_{C_3}$.
- Empiricamente:
  - Todos os estimadores apresentaram vício próximo de zero.
  - Os estimadores $\widehat{\mu}_{C_2}$ e $\widehat{\mu}_{C_3}$ apresentaram variâncias semelhantes e bem menores do que aquelas relativas ao estimador $\widehat{\mu}_{C_1}$.
  - A suposição de normalidade não parece ser razoável para a distribuição de nenhum estimador.
- Para a amostra selecionada, o estimador $\widehat{\mu}_{C_1}$ apresentou a pior performance.
- Exercício: Utilizar os pacotes "sampling" e "survey" para resolver o exercício.

Exercício 7.7: Bolfarine & Bussab (2005)

Considere uma população $\mathcal{U}$ com $N = 12$ elementos divididos em $A = 3$ conglomerados. Os valores de $y_{\alpha i}$ (resposta) correspondentes aos 3 conglomerados são:

α	$y_{\alpha i}$	B_α	μ_α	σ^2_α
1	0,1	2	0,5	0,25
2	1,2,2,3	4	2,0	0,50
3	3, 3, 4, 4, 5, 5 6	6	4,0	2/3

Resultados populacionais

- $\rho_{int} = 0,997$.
- $Corre(B_{\alpha}, \mu_{\alpha}) = 0,572$.
- Os resultados indicam:
  - Uma possível perda, em termos de precisão, das inferências oriundas da *AC* em relação àquelas oriundas da *AAS*.
  - Uma provável superioridade do estimador $\widehat{\mu}_{C_2}$ em relação aos outros dois $(\widehat{\mu}_{C_1}, \widehat{\mu}_{C_3})$

Resultados populacionais/seleção da amostra

- Utilizamos os pacotes “sampling” (para selecionar a amostra) e “survey” (para analisar os dados via PA AC_2).
- Para o estimador $\widehat{\mu}_{C_2}$ utilizamos a função “svymean” nos dados originais, enquanto que para os estimadores $\widehat{\mu}_{C_1}$ e $\widehat{\mu}_{C_3}$ utilizamos, respectivamente, a função “svymean” para os totais e as médias dos conglomerados (em uma base com informações somente relativas aos conglomerados).

Resultados amostrais

- Estimativas pontuais e intervalares:

Estimador	Est.	EP	IC(95%)
$\widehat{\mu}_{C_1}$	4,00	1,15	[6,95 ; 25,05]
$\widehat{\mu}_{C_2}$	3,20	0,55	[2,11 ; 4,29]
$\widehat{\mu}_{C_3}$	3,00	0,58	[1,87 ; 4,13]

- CCI:

$\widehat{\rho}_{C1}$	$\widehat{\rho}_{C2}$	$\widehat{\rho}_{C3}$
0,790	0,403	0, 430

- EPA:

$EPA(\widehat{\mu}_{C1})$	$EPA(\widehat{\mu}_{C2})$	$EPA(\widehat{\mu}_{C3})$
4,16	2,61	2,72

Comentários

- O melhor estimador foi o $\widehat{\mu}_{C_2}$ $(EP(\widehat{\mu}_{C_2}) = 2,61; B(\widehat{\mu}_{C_2}) = -0,08)$.
- Novamente, a divisão em conglomerados, em questão, leva à uma inferência menos precisa do que aquela obtenível com uma amosta de igual tamanho, sob AAS.
- Nesse caso, possivelmente, os resultados assintóticos vistos anteriormente não são válidos $(a = 2; A = 3)$.
- Os vieses foram, respectivamente (observado, estimado, observado): $\widehat{B}(\widehat{\mu}_{C_1}) = 1,2550; \widehat{B}(\widehat{\mu}_{C_2}) = -0,08, \widehat{B}(\widehat{\mu}_{C_3}) = 0,450$.