

Aplicação AAS_C e AAS_S : Exercício 2.2 - Bolfarine & Bussab (2005)

Prof. Caio Azevedo

Introdução ao problema

- Considere a Tabela 2.8 do livro “Elementos de amostragem” (relativa ao Exercício 2.2), como população alvo, referenciada e amostrada, e a variável (y) que representa o número de apartamentos alugados por condomínio, interpretando os espaços em branco como zeros.
- A população acima é composta por $N=180$ condomínios, dos quais também temos informações sobre a variável (x), que representa o número de apartamentos por condomínio.

Introdução ao problema

- O objetivo consiste em estimar a média (μ_y) e total (τ_y) populacionais dessa variável, bem como a proporção (p) de condomínios com mais de 20 apartamentos alugados.
- Novamente, compararemos os resultados sob os planos AAS_c e AAS_s (veja no site do curso os slides correspondentes à esses planos).
- Utilizaremos os pacotes `sampling` e `survey` do R ([link](#)).

- Inicialmente, para a obtenção do tamanho amostral é útil termos informação sobre a variância da população.
 - Uma maneira de proceder nesse caso seria obter uma amostra piloto. Consideraremos o interesse em estimar a média, inicialmente.
 - Uma outra forma seria levantar informações junto à imobiliárias, corretores, síndicos, administradoras de condomínio etc.
- Considerando $n = 10$ ($\approx 5,56$ % do tamanho da população), coletou-se uma amostra piloto usando a função `srswr` do pacote `sampling` do R, obtendo-se $\mathbf{y} = (28, 10, 8, 0, 19, 48, 13, 69, 6, 2)'$, em que $\bar{y} = 20,30$ e $\tilde{s}_y^2 = 495,79$.

AAS_S

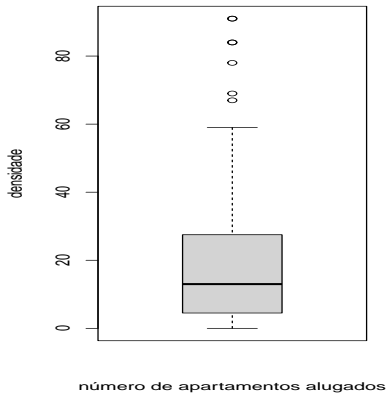
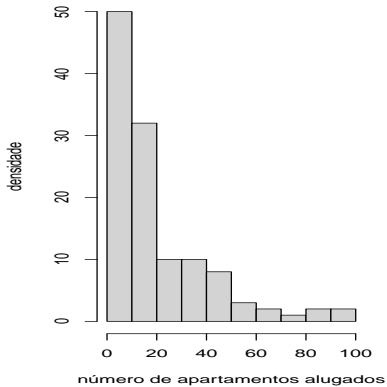
- De posse dessas estimativas, considerando um nível de confiança de 95% e precisão $\delta = 4$ (diferença de 4 apartamentos alugados, entre a estimativa e o valor verdadeiro), obtemos $n = \frac{1,96^2 495,79}{4^2} \approx 120$.
- Note que escolheu-se um valor de δ não muito baixo e, mesmo assim, obtivemos um tamanho amostral grande em comparação ao tamanho populacional.
- Isso se deveu à alta variabilidade observada na amostra piloto e pelo fato de estarmos usando AAS_C. Para valores de δ menores do que esse poderíamos ter um tamanho amostral maior do que o tamanho populacional (o que pode ocorrer nesse tipo de planejamento).

- Utilizando novamente a função `srswr` coletou-se uma amostra de tamanho 110 da população restante ($180-10=170$), a qual fora unida a amostra piloto, obtendo-se $n = 120$.
- Medidas resumo da amostra

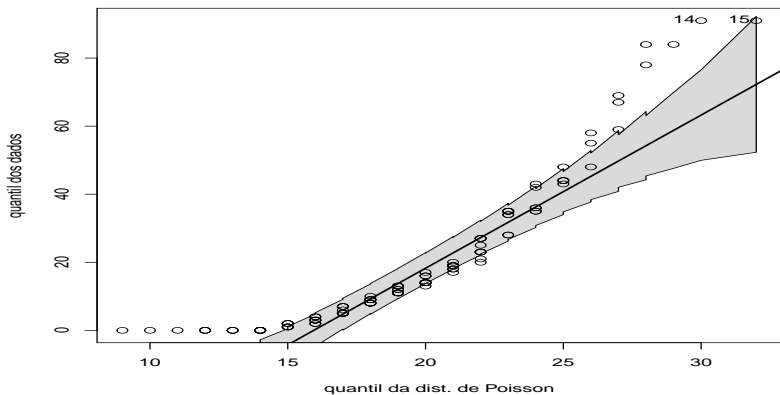
média	dp	var.	cv(%)	min.	med.	max.	ca	curt.
19,58	21,11	445,54	107,78	0,00	13,00	91,00	1,58	2,17

Vemos indícios de superdispersão, o que pode indicar que o modelo de Poisson (abordagem clássica) pode não ser apropriado.

AAS_S



AAS_S



■ Estimativas

- Média: $\bar{y} = 19,58$, $\tilde{s}^2 = 445,54$, $\widehat{EP}(\bar{Y}) = \sqrt{445,54/120} = 1,93$,
 $IC(95\%) = [15,91; 23,36]$.
- Total: $\tilde{\tau} = 180\bar{y} = 3525,00$, $\widehat{EP}(\bar{Y}) = 180\sqrt{445,54/120} = 346,83$,
 $IC(95\%) = [2845,21; 4204,79]$.

AAS_S

- Considere agora o problema de estimar a proporção (p) de condomínios com mais de 20 apartamentos alugados e a variável $w_i(W_i)$ (população (amostra)) que assume o valor 1 se o condomínio i possui mais que 20 apartamentos alugados e 0 se não.
- Podemos proceder como anteriormente para a média obtendo, inicialmente, informações de interesse, a partir de uma amostra piloto.
- Novamente, considerando $n = 10$, coletou-se uma amostra piloto usando a função `srswr` do pacote `sampling` do R e obteve-se $\mathbf{w}_p = (1, 0, 0, 0, 0, 1, 0, 1, 0, 0)'$, em que $\tilde{p} = 0,30$ e $\tilde{s}^2 = \tilde{p}(1 - \tilde{p}) = 0,21$.

- Com base nessas informações, considerando um nível de confiança de 95% e precisão $\delta = 0,1$, temos que $n = \frac{1,96^2 0,21}{0,10^2} \approx 81$.
- Utilizando novamente a função srswr coletou-se uma amostra de tamanho 71, da população restante ($180-10=170$), e juntou-se com a amostra piloto obtendo-se: $\tilde{p} = 34,56\%$,
 $\widetilde{EP}(\hat{p}) = 0,2261/80 = 0,002826$, $IC(95\%) = [24,14\%; 44,99\%]$.

- Adotando, agora, um plano do tipo AAS_s, iremos seguir os mesmos passos feitos sob AAS_c e, depois, comparar os resultados.
- Considerando $n = 10$, coletou-se uma amostra piloto usando a função `srswor` do pacote `sampling` do R e obteve-se $\mathbf{y} = (19, 48, 48, 36, 3, 12, 25, 19, 43, 0)'$, em que $\bar{y} = 25,30$ e $\tilde{s}^2 = 316,90$

- Com base nessas informações, um nível de confiança de 95% e precisão $\delta = 4$ (diferença de 4 apartamentos alugados, entre a estimativa e o valor verdadeiro), assim como na AAS_C, tem-se que

$$n = \frac{1}{\frac{316,904^2}{1,96^2} + \frac{1}{180}} \approx 54.$$

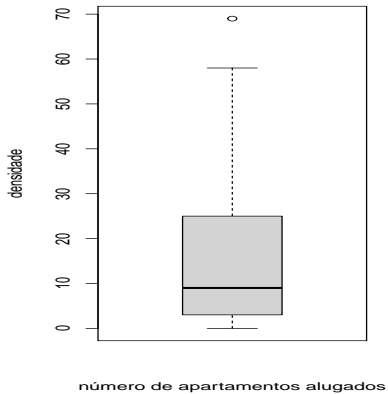
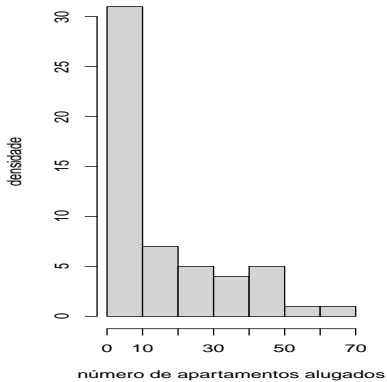
- Note que, para uma mesma precisão $\delta = 4$ obteve-se um tamanho amostral muito menor do que aquele obtido no plano AAS_C.

- Utilizando novamente a função `srswor` coletou-se uma amostra de tamanho 110 da população restante ($180-10=170$), a qual fora unida a amostra piloto, obtendo-se $n = 54$.

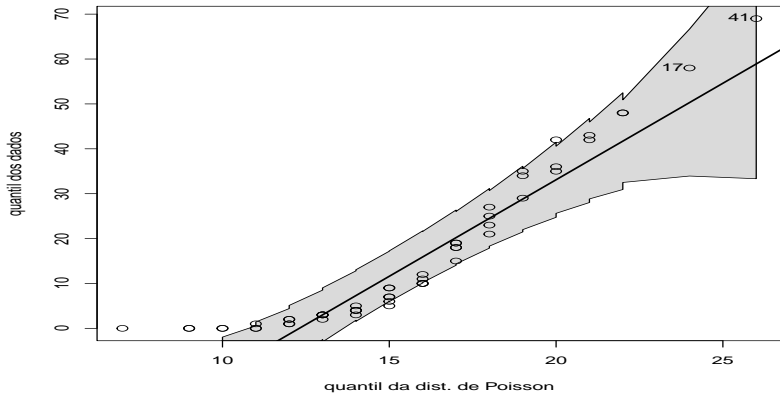
média	dp	var.	cv(%)	min.	med.	max.	ca	curt.
15,57	17,21	296,10	110,49	0,00	9,00	69,00	1,18	0,52

Vemos indícios de superdispersão, o que pode indicar que o modelo de Poisson (abordagem clássica) pode não ser apropriado.

AAS_S



AAS_S



■ Estimativas

- Média: $\bar{y} = 15,57$, $\widetilde{EP}(\bar{Y}) = \sqrt{(1 - 54/180)296,10/54} = 1,96$,
 $IC(95\%) = [11,73; 19,41]$.
- Total: $\tilde{\tau} = 180\bar{y} = 2803,33$,
 $\widetilde{EP}(\hat{\tau}) = \sqrt{(180^2)(1 - 54/180)296,10/54} = 352,64$,
 $IC(95\%) = [2112,15; 3494,51]$.

- Considere agora o problema de estimar a proporção (p) de condomínios com mais de 20 apartamentos alugados e a variável $w_i(W_i)$ (população (amostra)) que assume o valor 1 se o condomínio i possui mais que 20 apartamentos alugados e 0 se não. Aqui, coletamos também inicialmente informações de uma amostra piloto.
- Novamente, considerando $n = 10$, coletou-se uma amostra piloto usando a função `srswor` do pacote `sampling` do R, obtendo-se $\mathbf{w} = (0, 1, 1, 1, 0, 0, 1, 0, 1, 0)'$, em que $\tilde{p} = 0,50$ e $\tilde{s}^2 = \tilde{p}(1 - \tilde{p}) = 0,25$.

- Com base nessas informações, considerando um nível de confiança de 95% e precisão $\delta = 0,1$, temos que $n = \frac{1}{\frac{0,1^2}{1,96^2 \cdot 0,25} + \frac{1}{180}} \approx 63$.
- Utilizando novamente a função `srswor` coletou-se uma amostra de tamanho 53 da população restante ($180-10=170$), unindo-a à amostra piloto, obtendo-se: $\tilde{p} = 38,10\%$, $\widetilde{EP}(\hat{p}) = 0,0497$, $IC(95\%) = [28,35\%; 47,84\%]$.

Comparação dos resultados

- Agora com os resultados obtidos pelos dois planos amostrais vamos compará-los com aqueles obtidos a partir da população.
- Para a média temos que $\mu_y = 18,69$, e se pode notar, através da Tabela 1, que ambos os intervalos contém o parâmetro populacional.

Tabela: IC(95%) para a média obtidos de dois planos amostrais

PA	IC(95%)	CI	EPA*	n
AAS _c	[15,91;23,36]	7,55	1,00	120
AAS _s	[11,73;19,41]	7,67	0,70	54

* supondo a mesma amostra

Comparação dos resultados

- Embora o IC obtido pelo AAS_c tenha um comprimento menor do que aquele obtido pelo AAS_s foi necessário um tamanho amostral maior para a mesma precisão, o que na prática poderia ser mais custoso. Também, o EPA foi menor do que 1. Assim, novamente, temos evidências empíricas com respeito à superioridade do PA AAS_s em relação ao AAS_c .

Comparação dos resultados

- Para o total populacional temos que $\tau_y = 3363$, nota-se através da Tabela 2 que ambos os intervalos contém o parâmetro populacional.
- Aqui os comentários feitos para a média, também se aplicam.

Tabela: IC(95%) para o total obtido de dois planos amostrais

PA	IC(95%)	CI	EPA*	n
AAS _c	[2845,21 ;4204,78]	1359,57	1,00	120
AAS _s	[2112,15;3494,51]	1382,36	0,70	54

* supondo a mesma amostra

Comparação dos resultados

- Para a proporção de condomínios com mais de 20 apartamentos alugados temos que $p_y = 32,22\%$, nota-se através da Tabela 3 que ambos os intervalos contém o parâmetro populacional.

Tabela: IC(95%) para proporção de condomínios com mais de 20 apartamentos alugados obtido de dois planos amostrais

PA	IC(95%)	CI	EPA*	n
AAS _c	[28,40%; 47,79%]	20,84%	1,00	81
AAS _s	[23,00%; 43,66%]	19,50	0,65	63

* supondo a mesma amostra

Comparação dos resultados

- Aqui o IC obtido pelo AAS_c tem um comprimento maior do que aquele obtido pelo AAS_s , mas os dois não diferem muito, diferentemente do que fora observado para a média. Também, o tamanho amostral sob AAS_s for bem menor, bem como o EPA indica uma enorme vantagem deste plano, em relação ao AAS_c .