

Aplicações da AAS: Vestibular Unicamp 2017

Prof. Caio Azevedo

Introdução

- Vamos considerar parte dos dados relativos ao [Vestibular Unicamp \(VU\) 2017](#).
- O VU é promovido pela Unicamp através da [Comvest](#). São duas Fases (1 e 2):
 - Fase 1: candidatos respondem a um teste de 90 itens multidisciplinares de múltipla escolha.
 - Fase 2: provas dissertativas em diferentes áreas + redação.
 - Alguns cursos demandam provas de habilidades específicas.
- Um total de 73.498 inscritos disputaram um total de 3.330 vagas, distribuídas ao longo de 83 cursos.

Introdução

- Vamos assumir que nossa população alvo é o total de inscritos, e que a população referenciada e amostrada são os que, de fato, compareceram na Fase 1 ($N = 67.179$).
- Objetivos - estimar:
 - A nota bruta (NB, total de respostas corretas) média dos alunos que compareceram à prova da Fase 1 (μ).
 - A proporção de alunos que foram aprovados para a Fase 2 (AF2) (p).
- Há diversas informações disponíveis (como o curso pretendido, variáveis sócio-econômicas, etc) que podem auxiliar no processo de inferência/planejamento amostral. Contudo, por enquanto utilizaremos AAS (com e sem reposição).

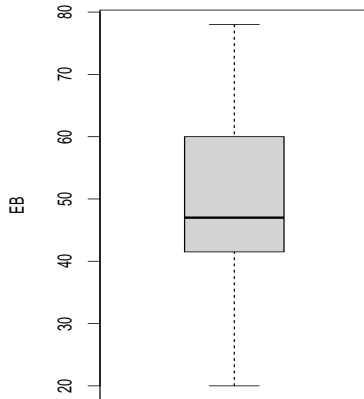
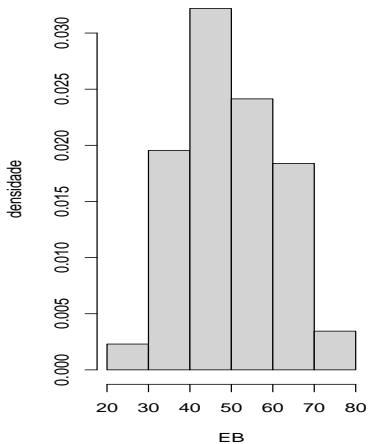
Estimação da média

- Seja $y_i, i = 1, \dots, 67.179$ a NB do i -ésimo candidato.
- À rigor, ela é uma variável aleatória mas a consideraremos como não o sendo.
- Seja $X_{ij} = 1, j = 1, 2, \dots, 90$, se o candidato i respondeu corretamente o item j e 0, caso contrário.
- Assim, $Y_i = \sum_{j=1}^{90} X_{ij}$. Embora não seja a melhor forma de abordar o problema, vamos supor que $Y_i \sim \text{binomial}(90, p)$ (para um dado i os X_{ij} não são independentes, ao longo dos j). Seria melhor considerar a dependência entre as observações ([TRI](#)) e/ou outras distribuições mais apropriadas ([link](#)).

Estimação da média

- Portanto, $\sigma^2 = \mathcal{V}(Y_i) = 90p(1 - p)$, cujo valor máximo é obtido quando $p = 1/2$. Logo $\max_p \sigma^2 = 22,5$
- Tamanho da amostra ($\gamma = 0,95 \rightarrow z_\gamma = 1,96, \delta = 1$)
 - $AAS_c : n = \frac{z_\gamma^2 \sigma^2}{\delta^2} = \frac{(1,96)^2 22,5}{1^2} \approx 87.$
 - $AAS_s : n = \frac{1}{\frac{\delta^2}{z_\gamma^2 \sigma^2} + \frac{1}{N}} = \frac{1}{\frac{1^2}{(1,96)^2 22,5} + \frac{1}{67.179}} \approx 87.$
- Usou-se a função `sample` e códigos feitos em R para, respectivamente, selecionar a amostra e gerar estimativas pontuais e intervalares.

Resultados AAS_c



Resultados AAS_c

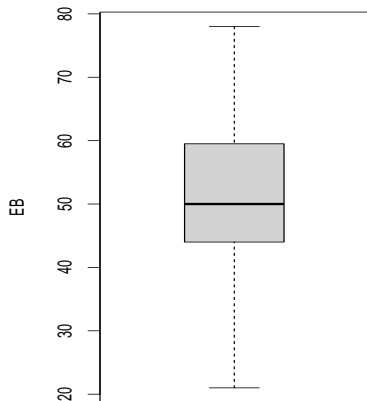
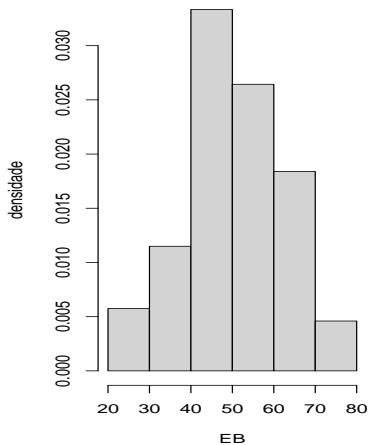
■ Medidas resumo da amostra

média	dp	var.	cv(%)	min.	med.	max.	ca	curt.
49,75	12,01	144,26	24,14	20,00	47,00	78,00	0,11	-0,74

■ Resultados inferenciais

Est.	EP	IC(95%)
49,75	1,29	[47,22;52,27]

Resultados AAS_s



Resultados AAS_5

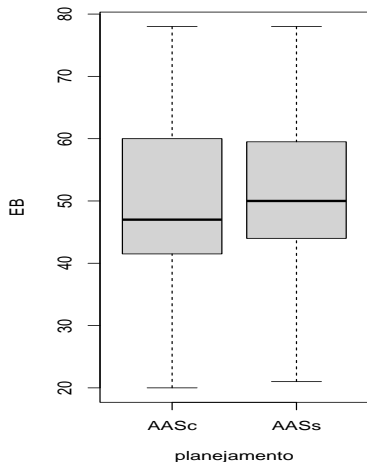
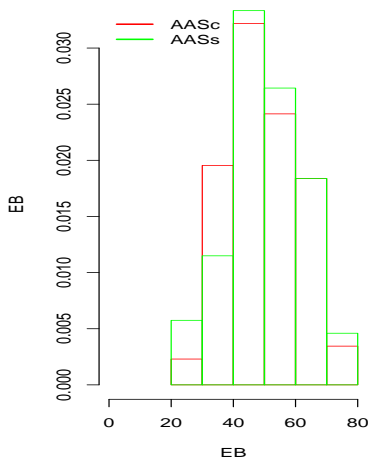
■ Medidas resumo da amostra

média	dp	var.	cv(%)	min.	med.	max.	ca	curt.
50,90	12,46	155,35	24,49	21,00	50,00	78,00	-0,07	-0,32

■ Resultados inferenciais

Est.	EP	IC(95%)
50,90	1,34	[48,28 ; 53,51]

Resultados AAS_c e AAS_s



Comparação AAS_c e AAS_s

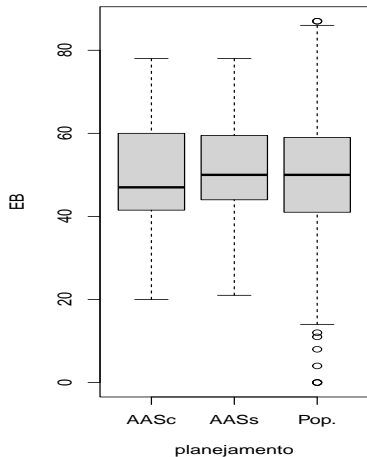
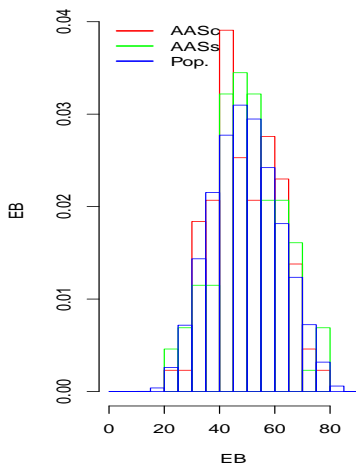
Medidas resumo da amostra

PA	média	dp	var.	cv(%)	min.	med.	max.	ca	curt.
AAS_c	49,75	12,01	144,26	24,14	20,00	47,00	78,00	0,11	-0,74
AAS_s	50,90	12,46	155,35	24,49	21,00	50,00	78,00	-0,07	-0,32

Resultados inferenciais

PA	Est.	EP	IC(95%)
AAS_c	49,75	1.29	[47,22 ; 52,27]
AAS_s	50,90	1,34	[48,28 ; 53,51]

Comparação entre os planos e os dados reais (DR)



Comparação entre os planos e os dados reais (DR)

Medidas resumo

PA	média	dp	var.	cv(%)	min.	med.	max.	ca	curt.
AAS _c	49,75	12,01	144,26	24,14	20,00	47,00	78,00	0,11	-0,74
AAS _s	50,90	12,46	155,35	24,49	21,00	50,00	78,00	-0,07	-0,32
DR	50,08	12,25	150,16	24,47	0,00	50,00	87,00	0,10	-0,44

Resultados inferenciais

PA	Est.	EP	IC(95%)	vício	var.	eqm
AAS _c	49,75	1,29	[47,22 ; 52,27]	-0,34	1,66	1,77
AAS _s	50,90	1,34	[48,28 ; 53,51]	0,81	1,78	2,44

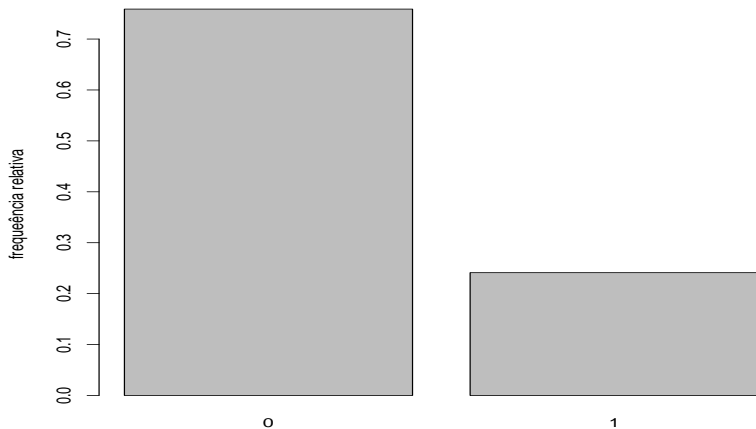
Estimação da proporção

- Seja $y_i, i = 1, \dots, 67.179$ igual ao 1 se o i -ésimo candidato fora aprovado para a Fase 2 e 0 caso contrário.
- À rigor, ela é uma variável aleatória mas a consideraremos como não o sendo.
- Uma alternativa para se determinar o tamanho da amostra seria tentar identificar (por exemplo via vestibulares anteriores) o(s) valor(es) mais prováveis de p .
- Contudo, neste caso, consideraremos o pior cenário em termos de variabilidade $p = 1/2$.

Estimação da proporção

- Tamanho da amostra ($\gamma = 0,95 \rightarrow z_\gamma = 1,96, \delta = 0,02$)
 - $AAS_c : n = \frac{z_\gamma^2 \sigma^2}{\delta^2} = \frac{(1,96)^2 0,25}{0,02^2} \approx 2.401.$
 - $AAS_s : n = \frac{1}{\frac{\delta^2}{z_\gamma^2 \sigma^2} + \frac{1}{N}} = \frac{1}{\frac{0,02^2}{(1,96)^2 0,25} + \frac{1}{67.179}} \approx 2.319.$
- Usou-se a função `sample` e códigos feitos em R para, respectivamente, selecionar a amostra e gerar estimativas pontuais e intervalares.

Resultados AAS_c



EB



Resultados AAS_c

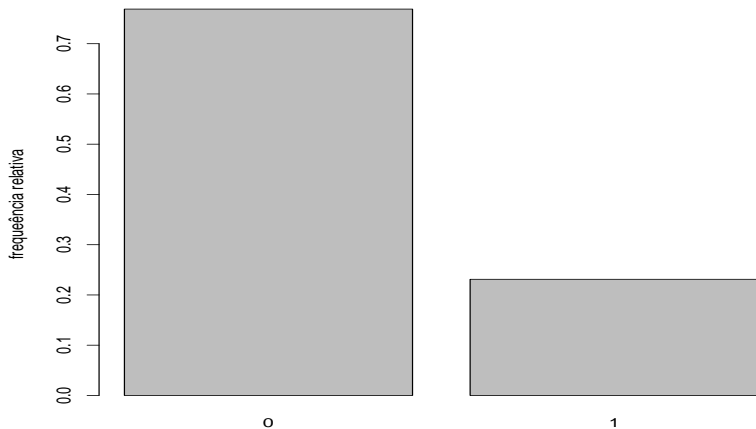
■ Medidas resumo da amostra

média	dp	var.	cv(%)	min.	med.	max.	ca	curt.
0,24	0,43	0,18	177,43	0,00	0,00	1,00	1,21	-0,54

■ Resultados inferenciais

Est.	EP	IC(95%)
24,11	0,87	[22,40 ; 25,83]

Resultados AAS_5



EB

Resultados AAS_5

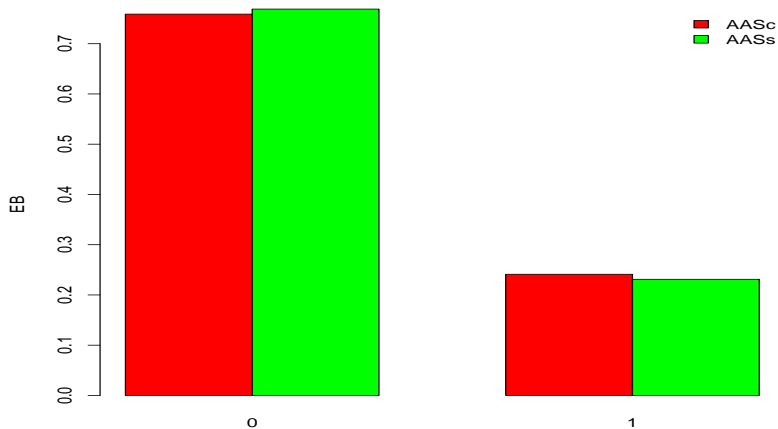
■ Medidas resumo da amostra

média	dp	var.	cv(%)	min.	med.	max.	ca	curt.
0,23	0,42	0,18	182,43	0,00	0,00	1,00	1,27	-0,38

■ Resultados inferenciais

Est.	EP	IC(95%)
23,11	0,86	[21,43 ; 24,80]

Resultados AAS_c e AAS_s



EB

Comparação AAS_c e AAS_s

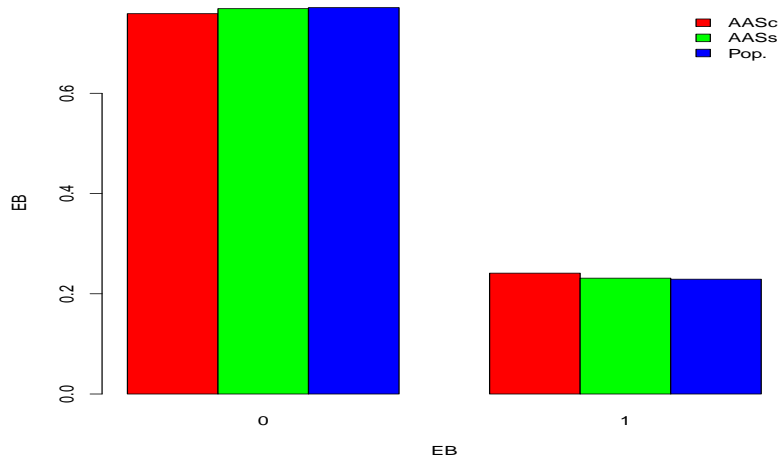
Medidas resumo da amostra

PA	média	dp	var.	cv(%)	min.	med.	max.	ca	curt.
AAS_c	0,24	0,43	0,18	177,43	0,00	0,00	1,00	1,21	-0,54
AAS_s	0,23	0,42	0,18	182,43	0,00	0,00	1,00	1,27	-0,38

Resultados inferenciais

PA	Est.	EP	IC(95%)
AAS_c	24,11	0,87	[22,40 ; 25,83]
AAS_s	23,11	0,86	[21,43 ; 24,80]

Comparação entre os planos e os dados reais (DR)



Comparação entre os planos e os dados reais (DR)

Medidas resumo

PA	média	dp	var.	cv(%)	min.	med.	max.	ca	curt.
AAS _c	0,24	0,43	0,18	177,43	0,00	0,00	1,00	1,21	-0,54
AAS _s	0,23	0,42	0,18	182,43	0,00	0,00	1,00	1,27	-0,38
DR	0,23	0,42	0,18	183,44	0,00	0,00	1,00	1,29	-0,34

Resultados inferenciais

PA	Est.	EP	IC(95%)	vício	var.	eqm
AAS _c	24,11	0,87	[22,40 ; 25,83]	1,21	0,01	0,02
AAS _s	23,11	0,86	[21,43 ; 24,80]	0,20	0,01	0,01