

Aplicações da AAS: Vestibular Unicamp 2017 (Fase 2)

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 (COMVEST).

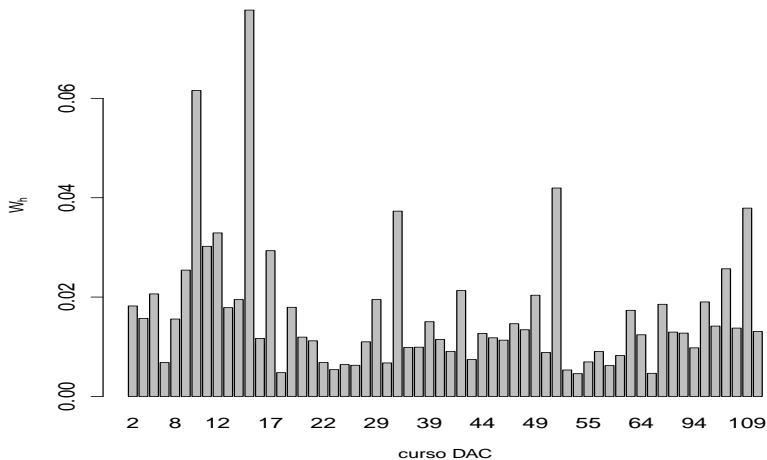
Introdução

- Em cada uma das sete provas da segunda fase, um nota no intervalo $\{0; 0,5, 1; 1,5; \dots; 4\}$ é atribuída a cada uma das seis questões. A nota final (NF) consiste na soma dessas pontuações.
- Vamos assumir que nossa população alvo é o total de alunos aprovados para a Fase 2, excluindo-se aquelas com o “curso Comvest” (veja), igual a : 37, 73, 707, 723, 726 (N=14.826). Assim, tem-se um total de 60 cursos (DAC).
- Objetivo - estimar:
 - A nota final (NF) média dos alunos acima descritos.
 - De acordo com os códigos, utilizou-se o pacote “**sampling**” para selecionar a amostra e o “**survey**” para analisar os dados.

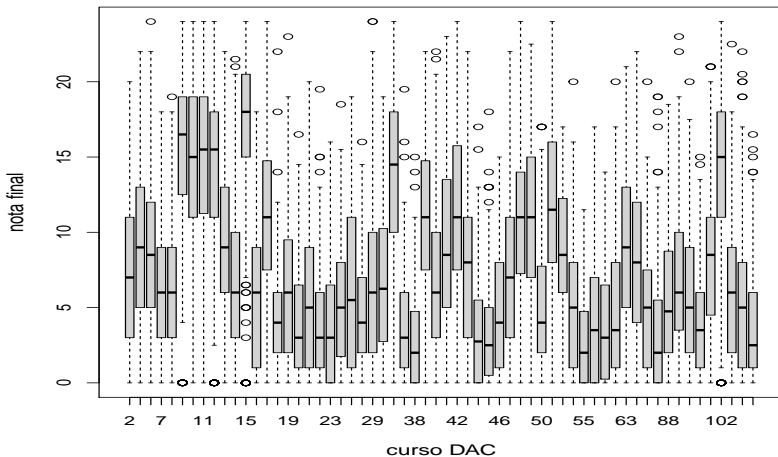
PA : planejamento amostral

- Há diversas informações disponíveis (como o curso pretendido, variáveis sócio-econômicas, etc) que podem auxiliar no processo de inferência/planejamento amostral. Vamos utilizar Amostragem Estratificada com AAS_s dentro de cada estrato, em que os estratos serão os $H = 60$ cursos.
- Tamanho da amostra: por facilidade tomaremos aproximadamente 15% da população, assim $n = 2.224$ (inicial).
- Alocação: ótima de Newman com custos iguais de seleção ao longo dos estratos (arredondando para o menor inteiro que ultrapassa o respectivo valor). Assim $n = 2.251$ (final)

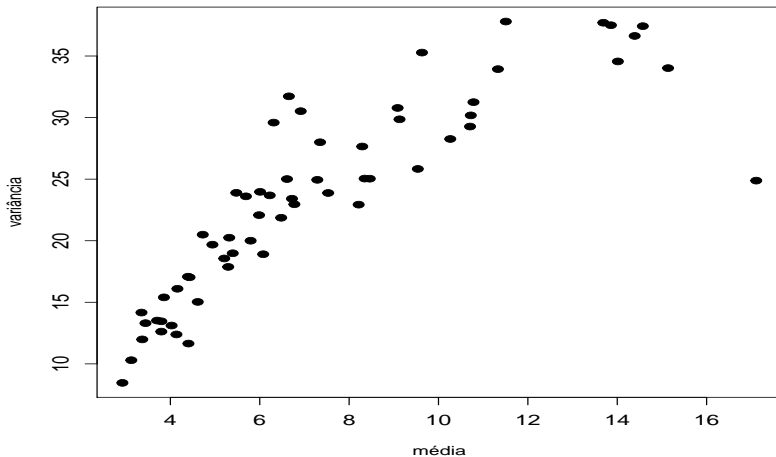
População : W_h (pesos dos estratos)



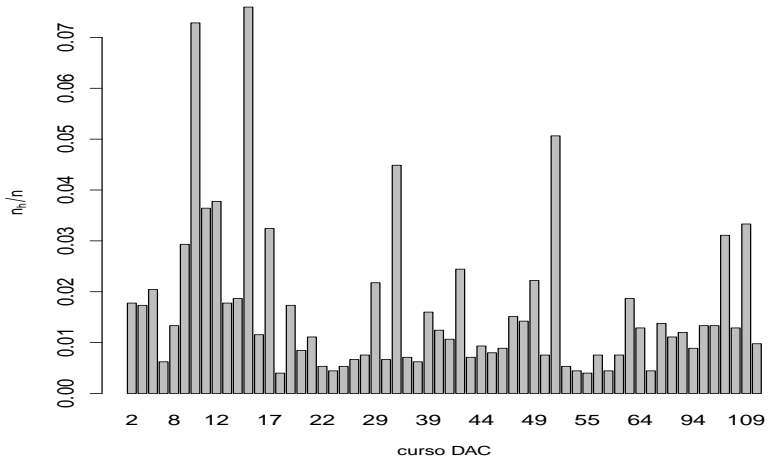
População : Box-plot da NF por estrato



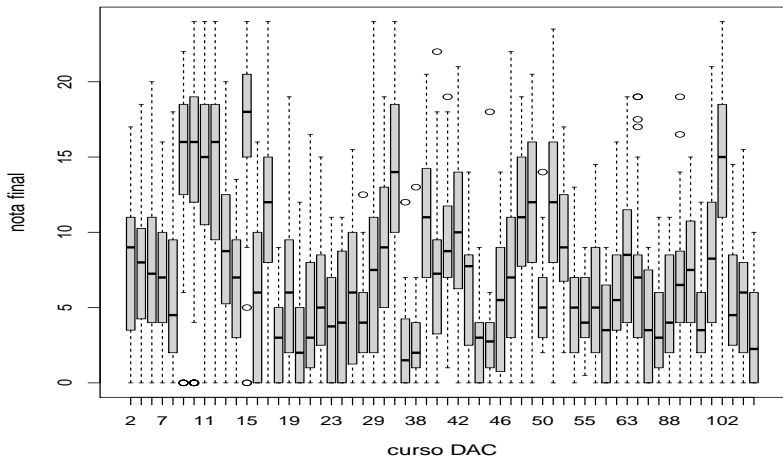
População : médias \times variâncias da NF por estrato



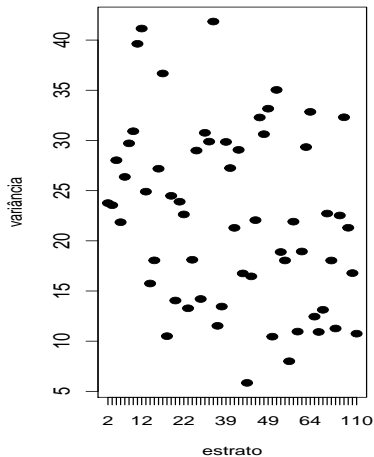
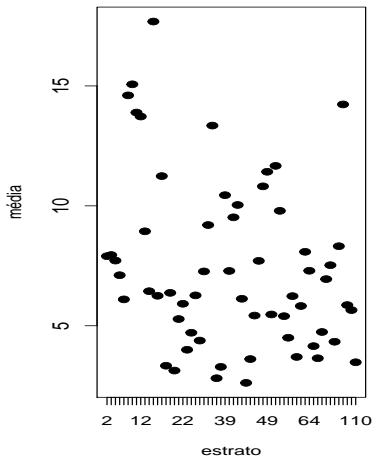
Amostra : n_h/n (“pesos” dos estratos na amostra)



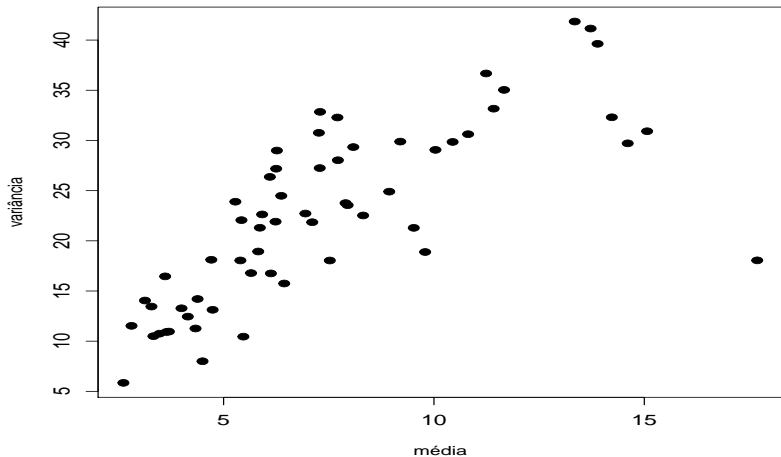
Amostra : Box-plot da NF por estrato



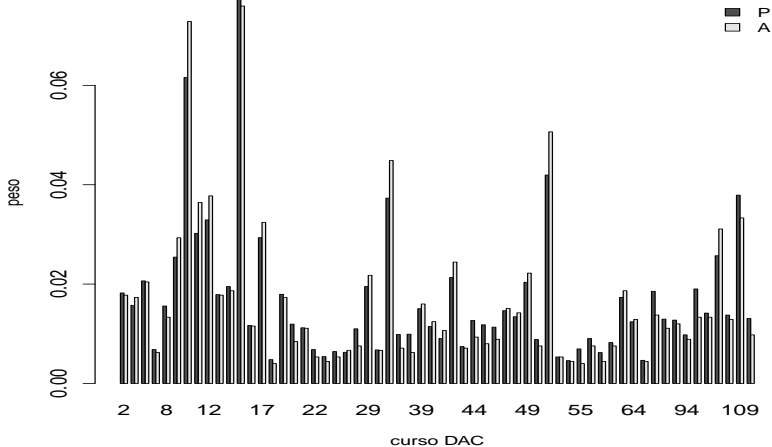
Amostra : médias e variâncias da NF por estrato



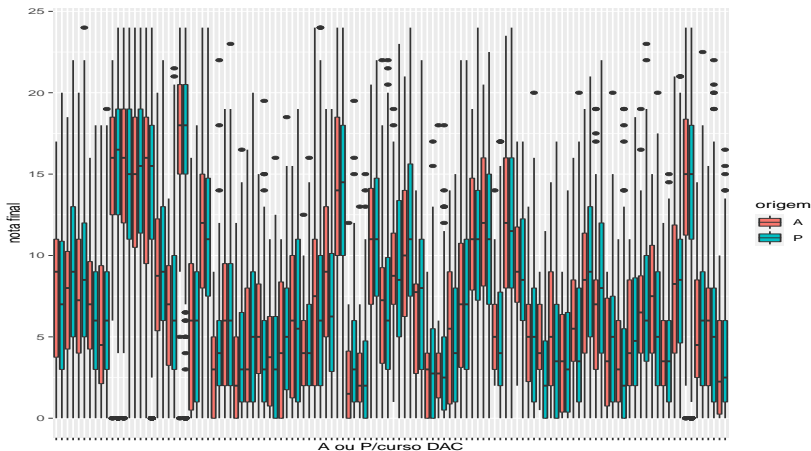
Amostra : médias \times variâncias da NF por estrato



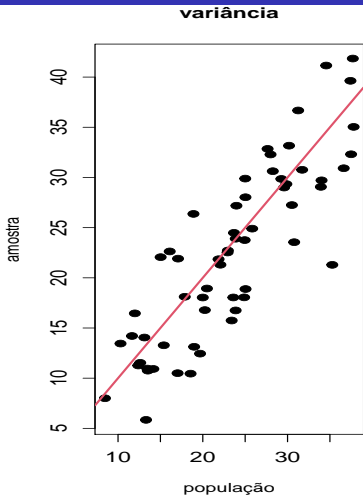
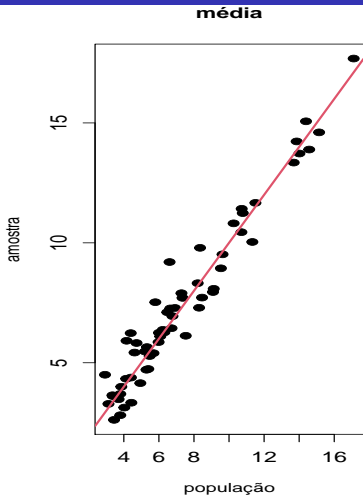
População \times Amostra : (pesos dos estratos)



População \times Amostra : Box-plot da NF por estrato



Popul. \times Amost. : médias e variâncias da NF por estrato



Características dos estratos (população e amostra)

- Temos que:

- População: $s_{d'}^2 = 26,29$, $s_{e'}^2 = 18,59$ e $\frac{s_{d'}^2}{s_{d'}^2 + s_{e'}^2} = 0,59$.

- Amostra: $\tilde{s}_{d'}^2 = 25,07$, $\tilde{s}_{e'}^2 = 19,09$ e $\frac{\tilde{s}_{d'}^2}{\tilde{s}_{d'}^2 + \tilde{s}_{e'}^2} = 0,57$.

- Apesar de, tanto em relação à população quanto em relação à amostra, a variabilidade dentro dos estratos (d') ser maior do que a entre estratos (e'), aparentemente, por conta das diferenças vistas anteriormente, entre os estratos, o processo de estratificação pode nos conduzir a resultados mais acurados do que aqueles eventualmente obtidos via AAS.

Resultados inferenciais

- Estimativas para a média (valor verdadeiro, $\mu = 9,33$):

Estimativa	EP	IC(95%)
9,41	0,10	[9,22;9,60]

- Temos os resultados para o EPA ($\frac{AE}{AAS}$):
 - Com reposição: 0,47.
 - Sem reposição: 0,56.
- Portanto, vemos que a AE levou à estimativas muito mais acuradas do que aquelas que seriam obtidas via AAS (com/sem reposição).