Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação Científica
Departamento de Estatística - ME430

Trabalho de ME430
Questão 2 da Lista IV

Grupo
Victor 206493, Jordão 170844, Nicole 204186, Leticia 201357
Prof. Dr. Caio Azevedo

Campinas
2018

1 Introdução

Este trabalho consiste na aplicação de técnicas aprendidas na disciplina ME430 - Técnicas de Amostragem - em um conjunto de dados da COMVEST. Esse conjunto de dados possui informações de 73498 candidatos ao vestibular 2017 e seus desempenhos nele. O objetivo é estimar i) a média da pontuação total de cada cadiadato, ii) a proporção de candidatos que cursaram todo o ensino médio em escola pública e iii) o total de quartos nas casas de todos os candidatos. As subseções *Média*, *Proporção* e *Total* nas seções *Análise Descritiva* e *Análise Inferencial* se referem a, respectivamente, i), ii) e iii). As amostragens realizadas foram artificiais, visto que há acesso às informações de todos os elementos da população (candidatos ao vestibular).

2 Análise Descritiva

2.1 Média

Inicialmente, para estimar μ, a pontuação média de todos os candidatos na 1ª fase, foi coletada uma amostra piloto de tamanho 100 a fim de estimar a variância $\sigma_\mu^2 = \frac{1}{N}\sum_{i=1}^{N} y_i$ da pontuação y_i dos candidatos, onde N é o tamanho da população (73498), e simular estratificações nessa amostra, buscando determinar o plano amostral mais adequado.

A amostra piloto resultou em uma estimativa de $\hat{s_\mu^2} = 252,1086$ para $s_\mu^2 = \frac{n\sigma^2}{n-1}$. De posse dessa informação, para realizar uma amostragem aleatória simples sem reposição (AASs), com um erro de estimativa $\delta = 5$, é preciso um tamanho de amostra n de pelo menos $\left(\frac{\delta^2}{\hat{s^2} z_{0.95}} + \frac{1}{N}\right)^{-1}$ para garantir que $P(|\mu - \hat{\mu}| \geq \delta) \geq 0.95$ [1], onde $z_{0.95}$ é o 95-quantil da normal padrão. Portanto, para os dados coletados, são necessárias pelo menos 137 unidades amostrais.

Por outro lado, uma amostragem estratificada (AE) com alocação proporcional (AP) e H estratos, descritos a seguir, requer $\frac{z_{0.95}^2}{\delta^2}\sum_{i=1}^{H}\frac{N_h}{N}\hat{\sigma^2}$ (OU SERIA S^2 AQUI?) para garantir que $P(|\mu - \hat{\mu}| \geq \delta) \geq 0.95$ [1], onde $\delta = 5$, N_h é o tamanho populacional do h-ésimo estrato e $\hat{\sigma_h^2}$ é a estimativa da variância da pontuação do h-ésimo estrato. Para os dados observados, isso significa que são necessárias XX unidades amostrais, número menor que no caso AASs.

É possível tornar a amostragem ainda mais robusta, utilizando a alocação ótima de Neyman (AON), que minimiza a variância da estimativa qunado o custo de amostragem é homogêneo entre os estratos. Usando AON, temos que $n_h = n\frac{N_h\sigma_h}{\sum_{i=1}^{H} N_h\sigma_h}$. Como σ_h não são valores conhecidos, foi usado $\hat{\sigma_h^2}$.

2.2 Proporção

2.3 Total

3 Análise Inferencial

3.1 Média

3.2 Proporção

3.3 Total

4 Conlcusões

Referências

[1] Heleno Bolfarine. *Elementos de amostragem.* Blucher, 2005.