ISSN: 2317-0840

Uma medida empírica para reduzir o vício no planejamento de amostragem aleatória simples e estratificada causado pela ausência de resposta

Mariano M. Espinosa[1†], Arthur C. Rezende[2], Lucas M. Castelo[3], Marcus V. D. Moura[4]

[1]*Professor Doutor do Departamento de Estatística, Instituto de Ciências Exatas e da Terra, Universidade Federal de Mato Grosso – UFMT.*
[2]*Graduando em Estatística, Universidade Federal de Mato Grosso - UFMT. E-mail: kia.rezende@hotmail.com.*
[3]*Graduando em Estatística, Universidade Federal de Mato Grosso - UFMT. E-mail: lucasmatoscastelo@hotmail.com.*
[4]*Graduando em Estatística, Universidade Federal de Mato Grosso - UFMT. E-mail: marcusduarte.98@hotmail.com.*

Resumo: *O objetivo do presente trabalho é apresentar um procedimento empírico para reduzir o vício no planejamento de amostragem aleatória simples e estratificada causado pela ausência de resposta. Trata-se de um estudo metodológico sobre o planejamento amostral e determinação do tamanho da amostra em diversas áreas do saber. Uma vez que nas pesquisas científicas geralmente são utilizadas amostras para inferir sobre uma determinada população. No entanto, a ausência de resposta pode introduzir vício no processo de amostragem e alterar as inferências obtidas por meio das informações amostrais. Assim, para evitar o problema da ausência de resposta causado por ausência no momento da entrevista e recusas. O acréscimo do tamanho da amostra durante o planejamento amostral é um procedimento geralmente adotado, para reduzir o vício ocasionado pela ausência de resposta. Para isso no presente estudo o tamanho de amostra é divido por 0,85, assim se devem realizar retornos ao domicilio até alcançar uma taxa mínima de resposta estimada igual a 85%.*

Palavras chave: População; Amostra; Amostragem; Ausência de resposta.

Abstract: *The objective of the present work is to present an empirical procedure to reduce the addiction in the planning of simple and stratified random sampling caused by the absence of response. This is a methodological study on sample planning and sample size determination in several areas of knowledge. Since in scientific research samples are usually used to infer about a given population. However, the absence of response may introduce an error in the sampling process and change the inferences obtained through the sample information. Thus, to avoid the problem of the absence of response caused by absence at the time of the interview and refusals. the addition of the sample size during sample planning is a generally adopted procedure to reduce the lack of response. For this in the present study the sample size is divided by 0.85, so returns should be made to the home until an estimated minimum response rate of 85% is reached.*

Keywords: Population; Sample; Sampling; Absence of response.

Introdução

A obtenção de uma amostra adequada é essencial para poder generalizar as conclusões obtidas, além de garantir uma boa representatividade da população de uma determinada pesquisa. Assim sendo, uma das etapas importantes do planejamento da pesquisa é a coleta dos dados e dentro desta etapa o planejamento amostral e determinação do tamanho da amostra são fundamentais para fornecer resultados estatisticamente confiáveis. Nas pesquisas, em geral, utiliza-se a amostragem para inferir sobre os parâmetros populacionais, devido, sobretudo, ao tempo e ao custo.

No entanto, a ausência de resposta pode introduzir vício no processo de amostragem e alterar as inferências obtidas por meio das informações amostrais (DA SILVA, 2001). Portanto, o objetivo do presente trabalho é apresentar um procedimento empírico para reduzir o vício no procedimento de amostragem causado pela ausência de resposta.

[†]Autor correspondente: marianom@cdp.ufmt.br.

Estimação de parâmetro por intervalo

Particularmente na determinação do tamanho de amostra n, o grau de confiança é de fundamental importância. Quando um parâmetro ou um estimador é determinado, deve-se avaliar a magnitude do erro de estimação, já que o valor do estimador por ponto de um parâmetro varia de uma amostra para outra. Para esta avaliação, em geral são utilizados intervalos de confiança para um parâmetro θ com coeficiente de confiança aproximadamente igual a *1-*α.

Usualmente para construir intervalos de confiança são utilizadas quantidades pivotais. Uma quantidade pivotal $Q(\tilde{x};\theta)$ para um parâmetro θ é uma variável aleatória cuja distribuição não depende de θ (BOLFARINE E SANDOVAL, 2001). Assim, para cada probabilidade desejada $1-\alpha$, onde $0<\alpha<1$, podemos encontrar os limites a e b na distribuição de $Q(\tilde{x};\theta)$ tais que:

$$P[a \leq Q(\tilde{x};\theta) \leq b] = 1-\alpha . \tag{1.1}$$

Observar que, uma vez que $Q(\tilde{x};\theta)$ é independente do parâmetro θ, os limites a e b também não dependem de θ. Então, se para cada $\tilde{x}$ observado existirem os valores $q_1(\tilde{x})$ e $q_2(\tilde{x})$, de modo que $a \leq Q(\tilde{x};\theta) \leq b$ se e somente se $q_1(\tilde{x}) \leq \theta \leq q_2(\tilde{x})$ e, considerando a expressão 1.1,

$$P[q_1(\tilde{x}) \leq \theta \leq q_2(\tilde{x})] = 1-\alpha . \tag{1.2}$$

Assim, o conjunto $[q_1(\tilde{x}); q_2(\tilde{x})]$ é um intervalo de $100(1-\alpha)\%$ de confiança para o parâmetro θ, onde os valores, $q_1(\tilde{x})$ e $q_2(\tilde{x})$ são chamados limite inferior e superior de confiança, respectivamente, e $1-\alpha$ é chamado de coeficiente de confiança. Assim, a expressão 1.2 fornece uma estimação por intervalo dos valores possíveis para o parâmetro θ de uma população.

Normalidade assintótica e intervalos de confiança para a média

Seja $x_1, x_2, \cdots, x_n$ uma amostra aleatória de uma variável aleatória X com distribuição normal com média μ e variância σ^2. Assumindo σ^2 conhecido, então uma quantidade pivotal para a média μ com n grande, é dado por:

$$Q(\tilde{x};\mu) = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}, \tag{1.3}$$

a qual utilizando o Teorema Central do Limite tem distribuição normal padrão.

Assim, por 1.1 e considerando 1.3, temos um intervalo dado por:

$$P\left[a \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq b\right] = 1-\alpha . \tag{1.4}$$

Portanto, existem infinitos valores de a e b que satisfazem a expressão 1.4. Além disso, pela simetria da distribuição normal padrão, o intervalo com menor comprimento que satisfaz tal expressão é aquele dado por a e b, cuja área à direita de b é igual a área à esquerda de a, que é igual a $\alpha/2$. Assim, $a=-z_{\alpha/2}$ e $b=z_{\alpha/2}$, onde $P[Z \leq z_{\alpha/2}] = 1-\alpha/2$. Assim,

$$P\left[-z_{\alpha/2} \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right] = 1-\alpha . \tag{1.5}$$

Considerando que em geral σ^2 é desconhecido, ele é substituído por seu estimador não viciado s^2, que para n grande é bem próximo de σ^2.

Determinação do tamanho da amostra para estimar a média da população

Para determinar o tamanho da amostra necessário para estimar a média da população μ, com um mínimo de erro possível, d, suponha que a média amostral $\bar{x}$ é considerada como um estimador pontual da média da população e considere um intervalo de $100(1-\alpha)\%$ de confiança para μ com n grande, dado por:

$$n = \left(\frac{Z_{\alpha/2}\sigma}{d}\right)^2, \tag{1.6}$$

No caso de amostragem aleatória simples sem reposição a variância pode ser reescrita da seguinte maneira:

$$Var(\bar{x}) = (1-f)\frac{s^2}{n} = \frac{s^2}{n/(1-f)} = \frac{s^2}{n'}, \tag{1.7}$$

com $f = n/N$ e $n' = n/(1-f)$, assim obtém-se uma expressão análoga à da amostragem aleatória simples com reposição, isto é,

$$n' = \frac{Z_{\alpha/2}^2 s^2}{d^2}. \tag{1.8}$$

Observar que, a Expressão 1.8 pode ser utilizada para populações muito grandes ou infinitas, na prática, para *N* desconhecido.

Portanto, para obter o tamanho efetivo da amostra, utilizar a seguinte expressão:

$$n = \frac{n'}{(1+n'/N)}. \tag{1.9}$$

Assim, substituindo 1.8 em 1.9 tem-se:

$$n = \frac{N(s^2)}{N(d^2/Z_{\alpha/2}^2)+s^2}, \tag{1.10}$$

na qual $d^* = d^2/Z_{\alpha/2}^2$, σ^2 é a variância da população, d é o limite para o erro de estimação, estipulado pelo pesquisador e $Z_{\alpha/2}$ é obtido da tabela da distribuição normal padronizada. Observar que na prática em geral o valor de σ^2 é estimado por s^2. Assim, a expressão 1.10 é utilizada para determinar o tamanho da amostra de populações finitas de tamanho *N*, isto é, para *N* conhecido.

A expressão 1.11 é usada para determinar o tamanho dos estratos para a amostra previamente calculada. Onde W_k é a fração populacional no estrato *k* ($W_k = N_k/N$).

$$n_k = n \times \frac{N_k}{N} = nW_k, \tag{1.11}$$

No entanto, é necessário considerar a possibilidade de perdas durante a coleta dos dados, geralmente ocasionadas pela não cobertura e pela ausência de resposta. Segundo Kish (1985) a ausência de resposta refere-se à falha em obter observações sobre alguns elementos selecionados e designados para a amostra; pode ser devido a recusas, ausência no momento da entrevista, questionários não devolvidos ou perdidos, etc. Diferente da não resposta, a não cobertura denota falha em incluir algumas unidades, ou seções inteiras, da população no planejamento da pesquisa ou no planejamento amostragem. Assim, em pesquisas bem planejadas e delineamentos amostrais bem elaborados os erros de coberturas em geral são mínimos. Por tal motivo a finalidade deste trabalho será a ausência de resposta.

Ausência de resposta

A ausência de resposta pode introduzir vicio no processo de amostragem e alterar as inferências obtidas por meio das informações amostrais (DA SILVA, 2001).

Para um melhor entendimento sobre o efeito da ausência de resposta em uma população de tamanho N, suponha que a mesma seja dividida em duas partes na qual N_1 seja o número de respostas obtidas e N_2 o número de respostas não obtidas. Então,

$$N = N_1 + N_2. \tag{1.12}$$

Uma vez que, não existe informação na amostra para a segunda parte da população (N_2), deve-se minimizar a proporção ou número de não respondentes, para reduzir a presença do vício, de tal maneira que o mesmo seja desprezível (KISH, 1965).

Assim, para evitar o problema da ausência de resposta causado por ausência no momento da entrevista e recusas, existem diversos trabalhos (KISH, 1965; COCHRAN, 1977). Os quais enfatizam a importância de um bom planejamento da pesquisa, para reduzir a proporção de perdas e desta forma manter o tamanho da amostra determinado segundo um nível de precisão especificado. Tendo planejado adequadamente, o acréscimo do tamanho da amostra durante o planejamento amostral é um procedimento geralmente adotado, o qual pelo menos, não mudará a precisão dos níveis especificados (KISH, 1965).

O retorno do entrevistador é outro procedimento bastante utilizado para minimizar a proporção dos não respondentes. O número de retornos varia segundo a natureza das variáveis estudadas. No entanto, usualmente após o terceiro retorno, as estimativas para as variáveis demográficas tornam-se estáveis e o vício pode ser considerado desprezível (SILVA, 2001). Segundo Kish (1965) se devem realizar retornos ao domicilio até alcançar uma taxa mínima de resposta estimada igual a 85%. Assim a expressão 1.10 pode ser escrita da seguinte maneira:

$$n^* = \left(\frac{N(s^2)}{N(d^2 / Z_{\alpha/2}^2) + s^2} \right) \Big/ 0{,}85 \tag{1.13}$$

O tamanho de amostra necessário para estimar uma proporção p, também pode ser determinado ao substituir σ ou s nas Expressões 1.6, 1.10 e 1.13 pela quantidade $\sqrt{p(1-p)}$. Neste caso, a expressão 1.13 pode ser expressa da seguinte maneira:

$$n^* = \left(\frac{N(p(1-p))}{N(d^2 / Z_{\alpha/2}^2) + p(1-p)} \right) \Big/ 0{,}85. \tag{1.14}$$

Resultado e discussão

Para mostrar a aplicação deste estudo foi considerada a população constituída pelos moradores dos municípios de Alta Floresta, Diamantino, Sinop e Sorriso da área urbana, no Estado de Mato Grosso, pertencentes à área de influência da BR 163. Estes municípios foram selecionados por possuírem escritórios regionais de saúde, exceto Sorriso que foi escolhido por proximidade geográfica do município de Sinop, além de questões políticas e de apoio logístico.

A seguir são apresentadas as populações e domicílios de cada município (Tabela 1 e 2).

Tabela 1 População total por município para a área urbana de quatro municípios do Estado de Mato Grosso.

Município	Alta Floresta	Diamantino	Sinop	Sorriso	Total
Total	40091	15895	93696	58363	208045

Fonte: IBGE - Censos Demográficos e Contagem Populacional; para os anos intercensitários, estimativas preliminares dos totais populacionais, estratificadas por idade e sexo pelo MS/SE/Datasus, 2010.

Utilizando-se a Expressão 1.13 e as informações da Tabela 1 foi determinado o tamanho da amostra na população estudada, considerando o nível de confiança de 95%, isto é, $z_{\alpha/2}$ = 1,96, uma proporção de 0,5% (p=0,5), um erro de estimação de 2,5% (d=0,025) e uma taxa minima de resposta estimada igual a 85%. Assim o tamanho de amostra da população considerada corrigida e com acréscimo foi de aproximadamente de 1795 individuos. Também foi determinado o tamanho da mostra por municipio (n_i) proporcional ao tamanhoda população estes valores são apresentados na Tabela 2.

Tabela 2 Tamanho da população e amostra por município para a área urbana dos quatro municípios do Estado de Mato Grosso considerados, Cuiabá MT, 2019.

Município	Alta Floresta	Diamantino	Sinop	Sorriso	Total
N_i	40091	15895	93696	58363	208045
n_i	294	117	687	428	1525
n_i^*	346	137	808	504	1795
Taxa mínima de resposta estimada.	0,85	0,85	0,85	0,85	0,85

Fonte: IBGE - Censos Demográficos e Contagem Populacional; para os anos intercensitários, estimativas preliminares dos totais populacionais, estratificadas por idade e sexo pelo MS/SE/Datasus, 2010.

Observar que a proporção aqui foi a mesma por estratos, no entanto se existe a proporção ou prevalência por estrato e o custo esta informação também pode ser utilizada no planejamento amostral.

Conclusão

No presente estudo conclui-se que em levantamentos amostrais de estudos que pretendam utilizar inferências estatísticas, a amostragem do tipo probabilística deverá ser utilizada e quando a população considerada não seja homogênea, recomenda-se a utilização do método de amostragem aleatória estratificada.

Além disso, para reduzir o vício no planejamento amostral dos métodos de amostragem aleatória simples e estratificada causado pela ausência de resposta no momento da entrevista e recusas, um acréscimo no tamanho da amostra durante o planejamento amostral pode ser considerado, por exemplo, um acréscimo obtido ao dividir o tamanho da amostra por 0,85. Assim, foi concluído que a utilização destes métodos conjuntamente possibilitou determinar o tamanho preciso da amostra, em cada estrato considerado, assim como reduzir a variância dos estimadores a serem determinados.

Referências bibliográficas

BOLFARINE, H.; BUSSAB, W. O. Elementos de amostragem. Editora Edgar Blücher. São Paulo, 2005, 274p

COCHRAN, William Gemmell. Sampling techniques. 3rd ed. New York: John Wiley & Sons, 1977.

IBGE. Censos Demográficos e Contagem Populacional; para os anos intercensitários, estimas preliminares dos totais populacionais, estratificado por idade e sexo pelo MS/SE/DATASUS. Mato Grosso, 2005

KISH, L. Survey sampling. John Wiley & Sons, inc., New York, 1965, 664p.

SILVA, N.N. Amostragem probabilística: um curso introdutório. São Paulo: EDUSP; 2001, 120p.