

MODELO DE RESPOSTA ALEATORIZADA: APLICAÇÃO DO MODELO DE SIMMONS

Olga Lidia Solano DÁVILA¹
Doris Gómez TICERÁN¹
Ana María Cárdenas ROJAS¹
Rosa Ysabel Adriazola CRUZ¹
Félix Manuel Bartolo GOTARATE¹

- RESUMO: No presente trabalho apresentamos uma comparação dos resultados de uma pesquisa realizada na Universidade Nacional Maior de São Marcos (UNMSM) no Peru, para investigar a proporção de pessoas que já experimentaram pasta básica de cocaína (PBC), proporção de usuários de PBC, porcentagem de pessoas que tiveram relações sexuais com mais de duas pessoas durante toda a vida e porcentagem de pessoas que consomem álcool todo final de semana, utilizando a teoria do modelo de resposta aleatorizada, o modelo de Simmons (Greenberg et al 1969; Solano 2004) e o método convencional. Os tamanhos das amostras foram 1008 alunos para a aplicação do modelo de Simmons e 598 para o método convencional. A população em estudo compreende os alunos matriculados no semestre 2008 – II.
- PALAVRAS CHAVES: Modelo de resposta aleatorizada; mecanismo aleatório; modelo de Simmons; pergunta delicada; pergunta não relacionada.

1 Introdução

Existe uma crescente preocupação no Peru em conhecer a prevalência do consumo de drogas e do álcool especificamente. No trabalho apresentado em CEDRO (2007) fizeram uma pesquisa em nível nacional para conhecer o consumo de drogas, álcool e outras características utilizando o método convencional.

O objetivo do presente trabalho foi comparar os resultados de uma pesquisa realizada na UNMSM usando o modelo de resposta aleatorizada de Simmons (Greenberg et al 1969; Solano 2004). Apresentamos a teoria mais relevante do modelo de Simmons no contexto descrito com a finalidade de que qualquer pesquisador interessado em usar esta metodologia conheça os aspectos fundamentais da mesma.

¹ Universidad Nacional Mayor de San Marcos, Departamento de Estadística, Av. Venezuela s/n, Lima, Perú. 619-7000. E-mail: osolanod@unmsm.edu.pe / dgomezt@unmsm.edu.pe / acardenasr@unmsm.edu.pe / radriazolac@unmsm.edu.pe / fbartolog@unmsm.edu.pe

2 Modelo de resposta aleatorizada da pergunta não relacionada - modelo de Simmons

Greenberg et. al (1969) modificou o modelo original de Warner com a finalidade de que o entrevistado tenha maior disposição a colaborar, substituindo o segundo enunciado “não pertença ao grupo A” por uma pergunta referente a uma característica B, não relacionada com A. O entrevistado utiliza um mecanismo aleatório para selecionar uma das seguintes proposições que se supõe, responde corretamente:

1. Pertenço ao grupo A
2. Pertenço ao grupo B

Os parâmetros que desejamos estimar são:

π_A : Proporção dos elementos da população pertencente a A.

π_B : Proporção dos elementos da população pertencente a B .

2.1 Estimação de π_A quando π_B é conhecido

No caso quando π_B é conhecido é requerida apenas uma amostra aleatória, e as proposições são assumidas corretamente.

1. Pertenço ao grupo A, apresentada com probabilidade p
2. Pertenço ao grupo B, apresentada com probabilidade 1-p

Logo substituindo temos

$$\hat{\pi}_A = \frac{\sum_{i=1}^n x_i - (1-p)\pi_B}{p}, \quad (1)$$

onde $x_i=1$, sim o i -ésimo elemento da amostra responde sim.

$\hat{\pi}_A$ é um estimador não viciado de π_A .

A variância de $\hat{\pi}_A$ é

$$V(\hat{\pi}_A) = \frac{\mu(1-\mu)}{np^2}, \quad (2)$$

onde $\mu = p\pi_A + (1-p)\pi_B$.

2.2 Estimação de π_A e π_B

Os parâmetros π_A e π_B são estimados com base em duas amostras aleatórias independentes de tamanho n_1 e n_2 , respectivamente.

Apresentam-se na Tabela 1 as proposições e probabilidades de seleção das variáveis aleatórias estudadas.

Tabela 1 - Proposições e probabilidades de seleção das variáveis aleatórias estudadas

Proposição	Probabilidade de seleção		Variável aleatória	
	amostra n_1	amostra n_2	n_1	n_2
1. Pertença ao grupo A	p_1	p_2		
2. Pertença ao grupo B	$1-p_1$	$1-p_2$	x_{1i}	x_{2j}

Os estimadores de π_A e π_B são:

$$\hat{\pi}_A = \frac{(1-p_2)\sum_{i=1}^{n_1} x_{1i} - (1-p_1)\sum_{j=1}^{n_2} x_{2j}}{p_1 - p_2} \quad \text{e} \quad \hat{\pi}_B = \frac{p_2\sum_{i=1}^{n_1} x_{1i} - p_1\sum_{j=1}^{n_2} x_{2j}}{p_2 - p_1}, \quad p_1 \neq p_2 \quad (3)$$

onde $\hat{\pi}_A$ é um estimador não viciado de π_A ,
 $\hat{\pi}_B$ é um estimador não viciado de π_B .

As variâncias de $\hat{\pi}_A$ e $\hat{\pi}_B$ são dadas por:

$$\begin{aligned} \text{Var}(\hat{\pi}_A) &= \frac{1}{(p_1 - p_2)^2} \left[\frac{(1-p_2)^2 \mu_1 (1-\mu_1)}{n_1} + \frac{(1-p_1)^2 \mu_2 (1-\mu_2)}{n_2} \right] \quad \text{e} \\ \text{Var}(\hat{\pi}_B) &= \frac{1}{(p_2 - p_1)^2} \left[\frac{p_2^2 \mu_1 (1-\mu_1)}{n_1} + \frac{p_1^2 \mu_2 (1-\mu_2)}{n_2} \right] \end{aligned} \quad (4)$$

onde $\mu_1 = p_1\pi_A + (1-p_1)\pi_B$ e $\mu_2 = p_2\pi_A + (1-p_2)\pi_B$.

2.3 Escolha de P_1 , P_2 , π_B e n

No trabalho de Greenberg, Abul-Ela, Simmons & Horvitz (1969) são apresentadas as seguintes considerações:

- 1) Na escolha de P_1 e P_2 deve-se levar em consideração a diminuição da variância e também a disposição para colaborar por parte do entrevistado. Para isto, propõem $P_1+P_2=1$ e elege-se P_1 tão longe de 0.5 quanto seja factível. Em experiências anteriores obtiveram-se resultados satisfatórios com P_1 entre 0.70 e 0.80 ou o seu complemento.
- 2) Escolha da proposição “não relacionada” π_B . Em geral recomenda-se que:
 $\pi_A < 0,5$ então escolher π_B pequeno
 $\pi_A > 0,5$ então escolher π_B grande
- 3) Determinação ótima de n_1 e n_2 no sentido de minimizar a variância de $\hat{\pi}_A$ sujeita à condição $n_1 + n_2 = n$. Esta determinação ótima é dada por

$$\frac{n_1}{n_2} = \frac{\mu_1(1 - \mu_1)(1 - p_2)^2}{\mu_2(1 - \mu_2)(1 - p_1)^2},$$

3 Aplicação do modelo de Simmons

O Modelo de Resposta Aleatorizada para a pergunta não relacionada – Modelo de Simmons – foi aplicado em uma pesquisa realizada na Universidade Nacional Maior de São Marcos no Peru, para investigar a proporção de pessoas que já experimentaram pasta básica de cocaína (PBC), proporção de usuários de PBC, porcentagem de pessoas que tiveram relações sexuais com mais de duas pessoas durante toda a vida e porcentagem de pessoas que consomem álcool todo final de semana, utilizando a teoria do modelo de resposta aleatorizada, o modelo de Simmons (Greenberg et al 1969; Solano 2004) e o método convencional. A população em estudo compreende os alunos matriculados no semestre 2008 – II.

3.1 Escolha do mecanismo aleatório e da probabilidade de seleção das perguntas

O mecanismo aleatório que serviu para selecionar um de dois enunciados, cada um dos quais requer uma resposta, por parte do entrevistado, sem revelar ao entrevistador sua posição pessoal com respeito à pergunta, foi um grupo de cartões. Dez cartões da cor branca para o enunciado "pessoas que já experimentaram PBC", dez cartões da cor verde para "usuários de PBC", dez cartões da cor rosa para o enunciado "pessoas que tiveram relações sexuais com mais de duas pessoas durante toda a vida", e dez cartões da cor celeste para o enunciado de "pessoas que consomem álcool todo final de semana".

O entrevistador requer que o entrevistado escolha um cartão a partir do grupo de dez, e em seguida, sem mostrar o conteúdo, lê silenciosamente e responde em voz alta "sim" ou "não" à proposição, e logo o cartão selecionado é juntado com os outros de seu grupo.

3.2 Proporção de pessoas que já experimentaram PBC

Dez cartões da cor branca

- Oito com o enunciado: (A) “Durante minha vida tenho consumido pasta de cocaína (merla, farinha, pó), pelo menos uma vez” (Ver Figura 1)
- Dois com o enunciado: (B). “Você utiliza Messenger?” (Ver Figura 2)

Logo a probabilidade (p) de selecionar o enunciado (A) é 0,80 e a probabilidade ($1-p$) de selecionar o enunciado (B) é 0,20.

$p = 0,80$		
Durante minha vida tenho consumido pasta básica de cocaína (merla, farinha, pó), pelo menos uma vez		
Leia silenciosamente e responda em voz alta		
Sim	ou	Não

Figura 1 - Mecanismo aleatório usado na pesquisa. Modelo de cartões: Proporção de pessoas que já experimentaram PBC.

$p = 0,20$		
Você utiliza Messenger?		
Leia silenciosamente e responda em voz alta		
Sim	ou	Não

Figura 2 - Mecanismo aleatório usado na pesquisa. Modelo de cartões: Pergunta não relacionada à pergunta delicada

De forma análoga utilizou-se cartões para o enunciado de usuários de PBC, pessoas que tiveram relações sexuais com mais de duas pessoas durante toda a vida e pessoas que consomem álcool todo final de semana.

3.3 Tamanho da amostra

Para o cálculo do tamanho da amostra utilizou-se o plano de amostragem aleatória (Scheaffer; Mendenhall; Ott, 2007) com um limite para o erro de estimação de 3,2%, 95% de confiabilidade e a informação histórica proporcionada pela Direção Acadêmica da Faculdade de Ciências Matemáticas; os alunos foram selecionados aleatoriamente até completar o tamanho da amostra igual a 1008. A entrevista realizou-se de 01 a setembro a 14 de setembro do ano de 2008.

3.4 Aplicação do método convencional

A coleta da informação realizou-se no período de 15 a 29 de setembro do ano 2008, com a aplicação do questionário aos alunos matriculados no semestre 2008-II.

Para o cálculo do tamanho da amostra utilizou-se o plano de amostragem aleatória (Scheaffer; Mendenhall; Ott, 2007) com um limite para o erro de estimação de 4,2% e a informação proporcionada pela Direção Acadêmica da Faculdade de Ciências Matemáticas; os alunos foram selecionados até completar o tamanho de amostra igual a 598.

O instrumento aplicado no método convencional foi feito da seguinte maneira:

Parte I: Aspectos gerais; características dos alunos matriculados no período de estudo.

Parte II: Aspecto específico das perguntas delicadas da pesquisa.

4 Resultados e discussão

4.1 Para obter a estimativa de π_A utilizamos o caso quando π_B é conhecido, para obter π_B

Usamos a informação histórica do censo estudantil do ano de 2007 da Universidade Nacional Maior de São Marcos.

4.2 Proporção de pessoas que já experimentaram PBC

O estimador de Simmons para a “Proporção de pessoas que já experimentaram PBC”, para o caso de $n=1008$, $\sum x_i = 232$, $\pi_B = \frac{8904}{27071} = 0,329$ (a proporção de alunos que utilizam Messenger) onde 8904 é o total de alunos que utilizam messenger e 27071 são todos os alunos no censo, substituindo na equação (1), temos que :

$$\hat{\pi}_A = \frac{\frac{\sum_{i=1}^n x_i}{n} - (1 - P)\pi_B}{P} = \frac{\frac{232}{1008} - (1 - 0,80)0,329}{0,80} = 0,21$$

A proporção de pessoas que já experimentaram PBC é 0,21. De forma análoga calculou-se a proporção de usuários de PBC, onde $\pi_B = \frac{16036}{27071} = 0,592$ (a proporção de alunos que utilizam página Web) onde 16036 é o total de alunos que utilizam página web e 27071 são todos os alunos no censo; percentagem de pessoas que tiveram relações sexuais com mais de duas pessoas durante toda a vida, onde $\pi_B = \frac{5675}{27071} = 0,21$ (a proporção de alunos que alguma vez interrompeu seus estudos universitários, onde 5675 é o total de alunos que interrompeu seus estudos universitários e 27071 são todos os alunos no censo; e percentagem de pessoas que consomem álcool todo final de semana, onde $\pi_B = \frac{8635}{27071} = 0,319$ (a proporção de alunos que trabalha atualmente) onde 8635 é o total de alunos que trabalha atualmente e 27071 são todos os alunos no censo.

Os resultados da pesquisa segundo o método utilizado encontram-se na Tabela 2.

Observa-se que utilizando o método tradicional, a proporção de usuários de PBC é 0,02, esta proporção é mais elevada quando se utiliza o método de resposta aleatorizada (0,10). De forma análoga temos que utilizando o método tradicional da modalidade de entrevista direta, 31,9% dos entrevistados tiveram relações sexuais com mais de duas pessoas durante toda a vida, esta percentagem é menor quando se utiliza o método de resposta aleatorizada (27,7%). O 25,3% dos entrevistados consomem álcool todo o final de semana, esta percentagem é menor quando utilizou-se o modelo de resposta aleatorizada (24,9%).

Tabela 2 - Resultados da pesquisa segundo o método utilizado

	MRA (Simmons)	Método convencional
Proporção de pessoas que já experimentaram PBC	0,21	0,09
Proporção de usuários de PBC	0,10	0,02
Porcentagem de pessoas que tiveram relações sexuais com mais de duas pessoas durante toda a vida	27,7%	31,9%
Porcentagem de pessoas que consomem álcool todo final de semana	24,9%	25,3%

4.3 Teste de hipóteses

Nesta seção aplicam-se testes de hipóteses para analisar os dois métodos utilizados.

Teste de hipóteses para “Proporção de pessoas que já experimentaram PBC”

Para testar que o parâmetro da proporção de pessoas que já experimentaram PBC utilizando os dois métodos são iguais, as hipóteses de interesse são

$$H_0 : \pi_s = \pi_c$$

$$H_1 : \pi_s > \pi_c$$

A estatística de prova é

$$Z_c = \frac{P_s - P_c}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} \quad e \quad P = \frac{n_s P_s + n_c P_c}{n_s + n_c}$$

Sob a hipótese nula Z_c aproxima-se de uma distribuição Normal padrão quando n aumenta.

O cálculo da estatística é $Z_c=6,19$. Para um nível de significância de 5%, $Z_{1-\alpha} = Z_{0,95} = 1,645$, logo como $Z_c = 6,19 > 1,645$, rejeita-se a hipótese nula, isto é, que o parâmetro “Proporção de pessoas que já experimentaram PBC” do Modelo de Resposta Aleatorizada é maior que o Método Convencional.

De forma análoga os resultados para o teste de hipóteses para a proporção de usuários de PBC é $Z_c = 6,27$ maior a 1,645, rejeita-se a hipótese nula, isto é, que o parâmetro “Proporção de usuários de PBC” do Modelo de Resposta Aleatorizada é maior que o Método Tradicional. Similarmente os resultados para o teste de hipóteses para proporção de pessoas que tiveram relações sexuais com mais de duas pessoas durante toda a vida é $Z_c=-1,79$ é menor que 1,645 e não rejeita-se a hipótese nula, isto é, que o parâmetro de “Proporção de pessoas que tiveram relações sexuais com mais de duas pessoas durante toda a vida”, estimado pelo Modelo de Resposta Aleatorizada é igual ao do método tradicional pela modalidade de entrevista direta.

Os resultados de teste de hipóteses para a proporção de pessoas que consomem álcool em todo final de semana foi $Z_c=-0,18$ menor que 1,645, não rejeita-se a hipótese nula, isto é, o parâmetro de “Proporção de pessoas que consomem álcool todo final de semana” estimadas pelo Modelo de Resposta Aleatorizada é igual ao do método convencional.

Conclusões

A técnica de resposta aleatorizada permitiu implementar uma série de técnicas de amostragem estatística, que mostrou-se eficiente na seleção e execução da pesquisa.

A estimativa da “Proporção de pessoas que já experimentaram PBC” com o Modelo de Resposta Aleatorizada foi de 0,21, a estimativa de “Proporção de usuários de PBC” foi de 0,10; a estimativa de “Porcentagem de pessoas que tiveram relações sexuais com mais de duas pessoas durante toda a vida” foi de 27,7% e a estimativa de “Porcentagem de pessoas que consomem álcool todo final de semana” foi de 24,9%.

Da aplicação dos testes de hipóteses, tanto para “Proporção de pessoas que já experimentaram PBC” e “Proporção de usuários de PBC”, pode-se concluir que as estimativas da população de “Proporção de pessoas que já experimentaram PBC” e “Proporção de usuários de PBC” pela técnica do MRA de Simmons são maiores que o método tradicional de entrevista direta, para um nível de significância de 5%.

Da aplicação dos testes de hipóteses, tanto para a “Proporção de pessoas que tiveram relações sexuais com mais de duas pessoas durante toda a vida” e “Proporção de pessoas que consomem álcool todo final de semana”, pode-se concluir que não existem evidências suficientes para sugerir que as estimativas de “Proporção de pessoas que tiveram relações sexuais com mais de duas pessoas durante toda a vida” e “Proporção de pessoas que consomem álcool todo final de semana” sejam maiores usando o método de MRA de Simmons que o método tradicional da entrevista direta, a um nível de significância de 5%.

Embora a técnica de resposta aleatorizada de Simmons não apresente taxas mais elevadas de resposta no caso de “Porcentagem de pessoas que tiveram relações sexuais com mais de duas pessoas durante toda a vida” - já que pelo método de entrevista direta a estimativa foi ligeiramente superior -, recomenda-se continuar experimentando estes modelos, e especialmente o modelo de Simmons em amostras maiores e em casos quando a questão é realmente muito sensível ou muito comprometedor. A técnica de resposta aleatorizada é mais eficiente quando utilizada em pesquisas onde o problema é altamente sensível e para o qual se requer um tamanho de amostra maior que o método tradicional.

Agradecimentos

Os autores agradecem ao Conselho Superior de Investigações da Universidade Nacional Mayor de São Marcos (Peru) pelo suporte financeiro.

DÁVILA, O. L.; GÓMEZ, D. A.; CÁRDENAS, A. M.; ADRIAZOLA, R. Y.; BARTOLO, F. Randomized response models: Simmons Model Application. *Rev. Bras. Biom.*, São Paulo, v.28, n.4, p.43-51, 2010.

- *ABSTRACT: In the present work we present a comparison of the results in a survey conducted in the University National Mayor de San Marcos in Peruvian to investigate the proportion of people who have tried cocaine basic paste, proportion of users of PBC, percentage of people who have had sex with more than two people over a lifetime, percentage of people who drink alcohol every weekend, using theory the randomized response model (RRM). The RRM, o Simmons model (Greenberg et al 1969; Solano 2004) and conventional method. The sample size was 1008 students to apply or Simmons model and 598 for the*

conventional method. The population in study were students matriculated in the half-yearly 2008 - II.

- **KEYWORDS:** Randomized response model; random mechanism; Simmons model; sensitive question; unrelated question

Referências

CHAUDHRI, A; MUKERJEE, R. *Randomized response: theory and techniques*. New York: Marcel Dekker, 1983. 161p.

CEDRO. *Epidemiologia de drogas en la población urbana peruana*. Encuesta de hogares. Lima-Perú, 2007.134p. (Monografía de Investigación, 15).

COCHRAN, G. W. *Técnicas de muestreo*. México: Compañía Editorial Continental, 2000. 513p.

DES, R. *Teoría del muestreo*. México: Fondo de Cultura Económica, 1980. 305p.

GREENBERG, B. G.; ABUL-ELA, A. A.; SIMMONS W.R.; HORVITZ, D. G. The unrelated question randomized response model: theoretical frame work, *J. Am. Stat. Assoc.*, New York, v.64, p.520-539,1969.

GREENBERG, B. G.; KUEBLER, R. R.; ABERNATHY J.R.;HORVITZ, D.G. Aplicación de la randomized response technique in obtaining quantitative data, *J. Am. Stat. Assoc.*, New York, v.66, n.334, p.243-250,1971.

LANKE, S. On the choice of the unrelated question in Simmons version of randomized response model. *J. Am. Stat. Assoc.*, New York, v.70, p.80-83, 1975.

LIU, P.T.; CHOW, L.P. A new discrete quantitative randomized response model. *J. Am. Stat. Assoc.*, New York, v.71, n.353, p.72-73,1976.

MAHMOOD, M.; SINGH, S.; HORN, S. On the confidentiality guaranteed under randomized response sampling: A comparison with several new techniques. *Biom. J.*, v.40, Berlin, n.2, p.237-242, 1998.

MOORS, J. J. A Optimization of the unrelated question randomized response model. *J. Am. Stat. Assoc.*, *J. Am. Stat. Assoc.*, New York, v.66, n.361, p.627-629,1971.

SCHEAFFER, R.; MENDENHALL, W. Y; OTT, L. *Elementos de muestreo*. México: Iberoamérica, 2007. 321p.

SHIMIZU, I.M.; BONHAM, G.S. Randomized response technique in a national survey. *J. Am. Stat. Assoc.*, *J. Am. Stat. Assoc.*, New York, v.73, n.361, p.35-39,1978.

SOLANO, O.L. Modelos de resposta aleatorizada para variáveis quantitativas: Modelo de Greenberg. *Rev. Mat. Estat.*, São Paulo, v.22, n.3, p.47-56, 2004.

WARNER, S.L. Randomized response: A survey technique for elimination evasive answer bias. *J. Am. Stat. Assoc.*, New York, v.60, p.63-69, 1965.

Recebido em 30.03.2010.

Aprovado após revisão 3.11.2010.