



Resposta aleatória: uma técnica de pesquisa para eliminar o viés de resposta evasiva Autor(es): Stanley
L. Warner Trabalho(s) revisado(is): Fonte:

Jornal do Associação Estatística Americana, Vol. 60, nº 309 (mar., 1965), pp. 63-
69

Publicado por: [Associação Estatística Americana](#) URL
estável: <http://www.jstor.org/stable/2283137> .

Acessado: 11/12/2012 12:32

Seu uso do arquivo JSTOR indica sua aceitação dos Termos e Condições de Uso, disponíveis em <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR é um serviço sem fins lucrativos que ajuda acadêmicos, pesquisadores e estudantes a descobrir, usar e desenvolver uma ampla variedade de conteúdo em um arquivo digital confiável. Usamos tecnologia da informação e ferramentas para aumentar a produtividade e facilitar novas formas de bolsa de estudos. Para obter mais informações sobre o JSTOR, entre em contato com support@jstor.org.



A American Statistical Association está colaborando com a JSTOR para digitalizar, preservar e ampliar o acesso ao *Journal of the American Statistical Association*.

RANDOMIZED RESPONSE: A SURVEY TECHNIQUE FOR ELIMINATING EVASIVE ANSWER BIAS

RESPOSTA RANDOMIZADA: UMA TÉCNICA DE PESQUISA
PARA ELIMINAR O VÍES DE RESPOSTA EVASIVA

STANLEY L. WARNER
Claremont Graduate School

For various reasons individuals in a sample survey may prefer not to confide to the interviewer the correct answers to certain questions. In such cases the individuals may elect not to reply at all or to reply with incorrect answers. The resulting evasive answer bias is ordinarily difficult to assess. In this paper it is argued that such bias is potentially removable. The individuals are given a choice of two responses to each question. One response is the correct answer and the other is a known incorrect answer. The device of randomizing this response is used to estimate the population proportion. The resulting estimator is unbiased and its mean square errors are compared with the mean square errors of conventional estimates under various assumptions about the underlying population. Estimativas imparciais de máxima verossimilhança são obtidas e seus erros quadráticos médios são comparados com os erros quadráticos médios de estimativas convencionais sob várias suposições sobre a população subjacente.

1. INTRODUCTION

FOR reasons of modesty, fear of being thought bigoted, or merely a reluctance to confide secrets to strangers, many individuals attempt to evade certain questions put to them by interviewers. In survey vernacular, these people become the "non-cooperative" group [5, pp. 235-72], either refusing to answer or giving incorrect answers. The resulting evasive answer bias is ordinarily difficult to assess. In this paper it is argued that such bias is potentially removable. The individuals are given a choice of two responses to each question. One response is the correct answer and the other is a known incorrect answer. The device of randomizing this response is used to estimate the population proportion. The resulting estimator is unbiased and its mean square errors are compared with the mean square errors of conventional estimates under various assumptions about the underlying population.

Por várias razões, os indivíduos em uma pesquisa por amostragem podem preferir não confiar ao entrevistador as respostas corretas a certas perguntas. Nessas circunstâncias, os indivíduos podem optar por não responder ou por dar respostas incorretas. O viés de resposta evasiva resultante é normalmente difícil de avaliar. Neste artigo, argumenta-se que tal viés é potencialmente removível ao permitir que os indivíduos escolham entre duas respostas para cada pergunta. Uma das respostas é a resposta correta e a outra é uma resposta incorreta conhecida. O método de randomização desta resposta é usado para estimar a proporção da população. O estimador resultante é não enviesado e seus erros quadráticos médios são comparados com os erros quadráticos médios de estimativas convencionais sob várias suposições sobre a população subjacente.

For reasons of modesty, fear of being thought bigoted, or merely a reluctance to confide secrets to strangers, many individuals attempt to evade certain questions put to them by interviewers. In survey vernacular, these people become the "non-cooperative" group [5, pp. 235-72], either refusing to answer or giving incorrect answers. The resulting evasive answer bias is ordinarily difficult to assess. In this paper it is argued that such bias is potentially removable. The individuals are given a choice of two responses to each question. One response is the correct answer and the other is a known incorrect answer. The device of randomizing this response is used to estimate the population proportion. The resulting estimator is unbiased and its mean square errors are compared with the mean square errors of conventional estimates under various assumptions about the underlying population.

Por motivos de modestia, medo de serem considerados fanáticos ou apenas relutância em confiar segredos a estranhos, muitos indivíduos tentam fugir de certas perguntas que lhes são feitas pelos entrevistadores. No vernáculo de pesquisa, essas pessoas se tornam o grupo não cooperativo [5, pp. 235-72], recusando-se abertamente a ser pesquisado ou consentindo em ser pesquisado, mas propositalmente fornecendo respostas erradas às perguntas. Na literatura, o problema do viés de recusa [1, pp. 355-61], [2, pp. 33-6], [5, pp. 235-72], no outro caso, há o problema do viés de resposta [3, p. 69], [4, pp. 280-325].

As perguntas que as pessoas tendem a evitar são aquelas que exigem respostas muito reveladoras. Perguntas inocuas geralmente recebem boas respostas, mas perguntas que exigem afirmações pessoais ou controversas provocam resistência. Quando a resistência é encontrada, a modificação usual do método de pesquisa é simplesmente um esforço adicional por parte do entrevistador para ganhar a confiança do entrevistado. Há, entretanto, uma relutância natural do indivíduo em geral em confiar certas coisas a qualquer um - quanto mais a um estranho - e há também uma relutância natural em ter declarações confidenciais em um papel contendo seu nome e endereço. Pelo menos para algumas perguntas, provavelmente apenas ganhos limitados são possíveis ao tentar persuadir o entrevistado de que ele pode confiar ao entrevistador.

Este artigo sugere um método alternativo para aumentar a cooperação. O método é construído com base na premissa de que a cooperação deve ser naturalmente menor se as perguntas permitirem respostas que revelem menos ate para o entrevistador. Essencialmente, o método envolve o dispositivo que - para certas questões ainda não inocuas - o entrevistado responde com respostas que fornecem informações apenas com base na probabilidade. Por exemplo, uma aplicação pode envolver o entrevistado apenas fazendo uma declaração verdadeira com uma determinada probabilidade menor que 1. Nesse caso, até mesmo o entrevistado saberia apenas a probabilidade de que a resposta dada fosse verdadeira. Na medida em que esse tipo de resposta é menos revelador do que uma resposta exigida para ser verdadeira com probabilidade 1, sugere-se que esse

type of approach may encourage greater cooperation for certain survey problems. As another more detailed application of the randomized response method, the following section outlines a particular model for estimating a population proportion. Comparing the resulting estimates with those obtained by conventional methods under various assumptions about the cooperation of those interviewed. A seguir descreve um modelo específico para estimar uma proporção populacional. As estimativas resultantes são então comparadas com estimativas convencionais sob várias suposições sobre a cooperação dos entrevistados.

[illegible]

π = the true probability of A in the population,
 p = the probability that the spinner points to A , and

$$X_i = \begin{cases} 1 & \text{if the } i\text{th sample element says yes} \\ 0 & \text{if the } i\text{th sample element says no} \end{cases}$$

Then i -ésimo elemento da amostra disser sim 01 se o i -ésimo elemento da amostra disser não.

$$P(X_i = 1) = \pi p + (1 - \pi)(1 - p),$$
$$P(X_i = 0) = (1 - \pi)p + \pi(1 - p),$$
$$P(X_i = 1) = rp + (-) (1 - p)$$

and arranging the indexing of the sample so that the first n_1 report "yes" while the second $(n - n_1)$ report "no," the likelihood of the sample is

organizando a indexação da amostra de modo que o primeiro (lreporte) "sim" enquanto o segundo (n-n1) reporte "não", o probabilidade da amostra é

The log of the likelihood is

$$\log L = n_1 \log [\pi p + (1 - \pi)(1 - p)]^{-(1-p)p + 7r(l-p)} n^{-n-1} \quad (1)$$

O logaritmo da probabilidade é

$$+ (n - n_1) \log [(1 - \pi)p + \pi(1 - p)], \quad (2)$$

and necessary conditions on π for a maximum are

$$\log \left[\frac{(n - n_1)(2p - 1)}{n_1(2p - 1)} \right] p + r(1 - p)],$$
 e as condições

$$\text{necessárias em } \pi \text{ para um máximo são} = \frac{\pi p + (1 - \pi)(1 - p)}{\pi p + (1 - \pi)(1 - p)}$$

or

$$\frac{(1-X)p + r(1-p) - rp + (-r)(1-p)}{\pi p + (1-\pi)(1-p)} = \frac{n_1}{n} \quad (3)$$

ou

$$7r_p + (1 - 7r)(1 - p) = \frac{n}{n} \quad (3)$$

Then, supposing $p \neq 1/2$, the maximum likelihood estimate of π is

RESPOSTA ALEATÓRIA

65

Então, supondo $p \neq 1/2$, a estimativa de máxima verossimilhança de π é

The expected value of the estimate is

O valor esperado da estimativa é

$$E\hat{\pi} = \frac{1}{2p-1} [p-1 + (1/n) \sum EX_i] \quad (5)$$

$$= \frac{1}{2p-1} [p-1 + (1/n) \sum EX_i] = \pi \quad (5)$$

and the variance of $\hat{\pi}$ is

$$\begin{aligned} \text{Var } \hat{\pi} &= \frac{n \text{Var } X_i}{(2p-1)^2 n^2} \\ &= \frac{[n \text{Var } X_i]}{(2p-1)^2 n^2} \\ &= \frac{[p(1-p)(1-p)] + [(1-p)^2 p]}{(2p-1)^2 n^2} \\ &= \frac{[p(1-p)(1-p)] + [(1-p)^2 p]}{(2p-1)^2 n^2} \\ &= \frac{[p(1-p)(1-p)] + [(1-p)^2 p]}{(2p-1)^2 n^2} \end{aligned} \quad (6)$$

Expression (5) shows $\hat{\pi}$ is an unbiased estimate of the true population proportion π . Moreover, since $\hat{\pi}$ is a maximum likelihood estimate and any useful n 's

are apt to be large, $\hat{\pi}$ may be assumed normally distributed about π with the variance indicated in expression (6). Thus, the usual confidence intervals are easily established. Expression (6) also sets out the separate dependence of the variance of $\hat{\pi}$ upon the choice of p . In fact, identifying

distribuído em torno de π com a variância indicada na expressão (6). Assim, todos os intervalos de confiança usuais são facilmente estabelecidos. A expressão (6) também estabelece a dependência separada da variância de $\hat{\pi}$ sobre a escolha de p . Na verdade, identificar

as the variance due to sampling and writing expression (6) as

$$\text{Var } \hat{\pi} = \frac{1}{n} \left[\frac{1}{16(p-1/2)^2} + \frac{1}{n} \right] \quad (7)$$

it is clear that the variance of $\hat{\pi}$ can be expressed as the sum of the variance due to sampling plus the variance due to the random device.

Two practical questions concern the estimation method implied by $\hat{\pi}$. First, how likely are people to cooperate and tell the truth when asked to respond in a randomized manner? The second question concerns the possibility of people responding outside the 0-1 range, which is remote in large samples.

The possibility of a false response outside the 0-1 range cannot be ruled out, but this possibility is remote in large samples. The probability of as people cooperating and telling the truth when solicited to respond in I? A

possibilidade de 79 obter valores fora da faixa de 0-1 não pode ser descartada, mas essa possibilidade é remota em grandes amostras.

the manner described? Second, how large a sample is required to obtain various degrees of precision by this estimate as compared to the conventional estimate?

The first question is primarily an empirical question, but the rationale for expecting better cooperation is clear. The individual being interviewed is asked various grants of precision by this estimate as compared to the conventional estimate? for less. The matter of how much less is summarized by the parameter p . Note first from expression (1) that if $p = 1/2$, the likelihood function does not even depend on π , thus, for a $p = 1/2$, the interviewee would be furnishing no information at all. Then note that if $p = 1$, the entire procedure would reduce to the conventional procedure of requiring the individual to state unreservedly whether or not he belonged to Group A. For p 's between $1/2$ and 1 (or between $1/2$ and 0), the person interviewed provides useful but not absolute information as to exactly which group he is in. In this context, the p can be thought of as describing the nature of the cooperation between the interviewer and the interviewee. As p goes from 1 to $1/2$, the burden of cooperating passes from the interviewee to the interviewer. It therefore seems reasonable to expect that for some questions at least, p 's less than 1 should induce greater cooperation on the part of the person interviewed.

The question of the sample size required for a given level of precision also depends on the parameter p . If a p close to 1 (or close to 0) is adequate to insure cooperation, then a smaller sample size is required than if a p close to $1/2$ is required to insure cooperation. Values of p close to $1/2$ convey less information from each interview, thus they also imply either a larger variance of the estimate or a larger sample size. Substituting values of p in expression (6) sets out the precise relation. As an example, supposing a $\pi = .5$ and a p halfway between

$1/2$ and 1 , the person interviewed provides useful but not absolute information as to exactly which group he is in. In this context, the p can be thought of as describing the nature of the cooperation between the interviewer and the interviewee. As p goes from 1 to $1/2$, the burden of cooperating passes from the interviewee to the interviewer. It therefore seems reasonable to expect that for some questions at least, p 's less than 1 should induce greater cooperation on the part of the person interviewed.

The more pertinent comparisons are between the randomized estimates and regular estimates under the assumption that the regular estimates are handicapped by less than 100 per cent truthfulness. Suppose that in a regular survey all consent to be surveyed, but members of Group A tell the truth only with probability $7r$, and members of Group B tell the truth only with probability $7b$. Then, if $Y_i = 1$ or 0 according as the i th member of the sample reports he is or is not in Group A, the conventional estimate of the true population proportion π is

conforme o i -ésimo membro da amostra relatar que ele está ou não no Grupo A, a estimativa convencional da verdadeira proporção populacional π é

$$\hat{\pi} = \frac{\sum Y_i}{n} \quad (8)$$

The expected value, response bias [3, p. 89], and variance of this regular estimate are given by

O valor esperado, viés de resposta [3, p. 89], e a variância desta estimativa regular são dadas por

$$EV = 7rTa + [(1 - 7r)(1 - Tb)], \quad (9)$$

$$\text{Bias } \hat{\pi} = E(\hat{\pi} - \pi)$$

RESPOSTA ALEATÓRIA

$$= \pi[T_a + T_b - 2] + [1 - T_b], \text{ and}$$
$$\text{Viés } \hat{\pi} = E(\hat{\pi} - \pi)$$
$$\text{Var } \hat{\pi} = \frac{[\pi T_a + (1 - \pi)(1 - T_b)][1 - \pi T_a - (1 - \pi)(1 - T_b)]}{n[T_a + T_b - 2] + [1 - T_b]}. \quad (14)$$

67

Tables 1 and 2 then compare the mean square errors (the variance plus the square of the bias) of the randomized and regular methods of estimation under the assumption that the interviewed individuals tell the truth in the randomized method but only tell the truth in the non-random method with probabilities given by T_a and T_b . The left hand two columns of each table indicate various values of T_a and T_b and the right hand two columns of each table indicate various values of π .

TABLE 1. COMPARISON OF RANDOMIZED AND REGULAR ESTIMATES FOR TRUE PROBABILITY OF $A = .6$ AND $n = 1000$

TABELA 1. COMPARAÇÃO DE ESTIMATIVAS RANDOMIZADAS E REGULARES PARA VERDADEIRA PROBABILIDADE DE $A = .6$ AND $n = 1000$			Mean Square Error Randomized			
Regular Estimates			Mean Square Error Regular			
Probability of Truth T_a Estimativas Regulares		Bias	Erro Quadrado Médio Randomizado			
			$p = .6$	$p = .7$	$p = .8$	$p = .9$
Probabilidade da verdade para T_a	Probabilidade da verdade para T_b	Bias				
.90	1.00	-.03	5.45	1.36	.60	.33
.70	1.00	-.06	1.62	.40	.18	.10
.95	1.00	-.18	.19	.05	.02	.01
.50	1.00	-.30	5.45	1.36	.60	.33
.80	1.00	-.06	.07	.02	.01	.00
1.00	.95	.02	9.82	2.44	1.08	.60
1.00	.90	.04	3.41	.95	.37	.21
1.00	.80	.12	.43	.11	.05	.03
1.00	.70	.20	9.82	2.44	1.08	.60
1.00	.50	.01	18.25	4.51	2.00	1.11
1.00	.95	-.01	9.70	2.41	1.06	.59
1.00	.90	-.02	1.62	.40	.18	.10
1.00	.80	-.01	18.25	4.51	2.00	1.11
.90	.90	-.02	9.70	2.41	1.06	.59
.70	.70	-.06	1.62	.40	.18	.10
.50	.50	-.10	.61	.15	.07	.04

TABLE 2. COMPARISON OF RANDOMIZED AND REGULAR ESTIMATES FOR TRUE PROBABILITY OF $A = .5$ AND $n = 1000$

TABELA 2. COMPARAÇÃO DE ESTIMATIVAS RANDOMIZADAS E REGULARES PARA VERDADEIRA PROBABILIDADE DE $A = .5$ AND $n = 1000$			Mean Square Error Randomized			
Regular Estimates			Mean Square Error Regular			
Probability of Truth T_a Estimativas Regulares		Bias	Erro Quadrado Médio Randomizado			
			$p = .6$	$p = .7$	$p = .8$	$p = .9$
Probabilidade da verdade para T_a	Probabilidade da verdade para T_b	Bias				
.90	1.00	-.03	7.15	1.79	.79	.45
.70	1.00	-.05	2.28	.57	.25	.14
.95	1.00	-.15	.28	.07	.03	.02
.50	1.00	-.25	7.15	1.79	.79	.45
.80	1.00	-.05	.10	.03	.01	.01
.95	.95	.00	25.00	6.25	2.78	1.56
.90	.90	.00	25.00	6.25	2.78	1.56
.70	.90	.00	25.00	6.25	2.78	1.56
.50	.90	.00	25.00	6.25	2.78	1.56
.90	.50	.00	25.00	6.25	2.78	1.56
.70	.70	.00	25.00	6.25	2.78	1.56
.50	.50	.00	25.00	6.25	2.78	1.56

paired values for T_a and T_b . The third column shows the bias of the non-random method, and the remaining columns exhibit the ratios of the mean square errors of the unpaired estimates to the mean square errors of the regular estimates for various values of p . Tables 1 and 2 are respectively appropriate for the cases where the true probability of A is 0.6 and 0.5. The sample size is set at 1000 in each case.

3. CONCLUSIONS

[illegible][illegible]

A estimativa de máxima verossimilhança para o último esquema tem a mesma forma e a mesma variância da estimativa da Secção 2. Há

2 Como antes, ap de i não fornece nenhuma informação, ap de 1 fornece informações completas e outros p's fornecem informações dependendo de quão longe eles estão de J. É uma característica da natureza dicotômica da população dizer a verdade .2 do tempo é equivalente a dizer a verdade 0.8 do tempo.

thus a question as to which of these or other equivalent randomized models is to be preferred from the standpoint of increasing cooperation.

Finally, it is a question to be noted that it is easy to extend the randomized response techniques to estimate distributions of points other than that appropriate to a simple dichotomous variable. As one example, the technique could be applied to estimate the normal distribution of a variable through the previous derivation of an estimator of the normal sample variance. This is completely a technical problem of application. In this case one can interview a right-hand sample of classes to make five separate randomized responses concerning whether or not he was in each of the five separate classes. Just as with the proportion problem, it is clear that other randomized response methods may be imagined for this more general estimation problem. And just as with the proportion problem, the question of which specific technique will prove superior is a matter for empirical investigation.

4. ACKNOWLEDGMENTS

Assim como no problema da proporção, a questão de qual técnica específica se mostrará superior é uma questão para investigação empírica.

4. ACKNOWLEDGMENTS

4. AGRADECIMENTOS

REFERENCES

Estou em dívida com o árbitro por sugestões úteis.

- [1] Cochran, W. G., *Sampling Techniques, Second Edition*. New York: John Wiley and Sons, Inc., 1963.

REFERÊNCIAS

- [2] Deming, W. E., *Some Theory of Sampling*. New York: John Wiley and Sons, Inc., 1950.
 [3] Cochran, WG, *Sampling Techniques*, Segunda edição. Nova York: John Wiley and Sons, Inc., 1963.
 [12] Deming, WE, *Alguns Teoriada Amostragem*. Nova York: John Wiley and Sons, Inc., 1950.
 [31] Hansen, MH, Hurwitz, WN e Madow, WG, *Sample Survey Methods and Theory*, Volume I. Nova York: John Wiley and Sons, Inc., 1953.
 [51] Stephan, FF, e McCarthy, PJ, *Sampling Opinions*. New York: John Wiley and Sons, Inc., 1963.
 [52] Hansen, MH, Hurwitz, WN e Madow, WG, *Metodos de pesquisa de amostra e Teoria*, Volume II. Nova York: John Wiley and Sons, Inc., 1953.
 [5] Stephan, FF, e McCarthy, PJ, *Sampling Opinions*. Nova York: John Wiley e Filhos, Inc., 1963.