

MODELO DE RESPUESTA ALEATORIZADA DE WARNER PARA INCREMENTAR LA PROBABILIDAD DE OBTENER RESPUESTAS SINCERAS A PREGUNTAS SENSIBLES

M. Sc. Dindo Valdez Blanco

✉ dindovaldez@hotmail.com

RESUMEN

En esta investigación se estudia el modelo de respuesta aleatorizada propuesto por Warner (Warner, 1965) con el propósito de disminuir el sesgo de respuesta cuando se formulan preguntas sensibles y/o delicadas, la aplicación se realiza en la Facultad de Ciencias Puras y Naturales de la Universidad Mayor de San Andrés. Para la aplicación se considera preguntas sensibles o delicadas, como el comportamiento a realizar trampa en los exámenes y el consumo de drogas, aplicando el método de pregunta aleatorizada de Warner y el método de pregunta directa. Esencialmente el método de Warner involucra una técnica de aleatorización de tal manera que el entrevistado debe responder a las preguntas sensibles de acuerdo al resultado que arroje el método, dichos procedimientos pueden ser: juegos de monedas, maso de cartas ruletas giratorias, entre otros. La técnica de aleatorización permite calcular el estimador de la proporción de manera indirecta, dando al entrevistado el anonimato y permitiendo una respuesta sincera. Por lo tanto, se establece una relación probabilística entre una respuesta dada y la pregunta sensible. Finalmente se realiza una comparación entre el método de la pregunta directa que establece los estimadores usuales para la proporción de éxitos y el método de respuesta aleatorizada de Warner, llegando a determinar que el método de Warner es efectivo.

PALABRAS CLAVE

Pregunta sensible, modelo de respuesta aleatorizada, sesgo de respuesta, encuestas por muestreo.

ABSTRACT

In this research, the randomized response model proposed by Warner (Warner, 1965) is studied with the purpose of reducing the response bias when sensitive and / or delicate questions are asked, the application is carried out in the Faculty of Pure and Natural Sciences of the Universidad Mayor de San Andrés. For the application, sensitive or delicate questions are considered, such as behavior to cheat in exams and drug use, applying the Warner randomized question method and the direct question method. Essentially the Warner method involves a randomization technique in such a way that the interviewee must answer sensitive questions according to the result of the method, such procedures can be: coin games, deck of rotating roulette cards, among others. The randomization technique allows to calculate the estimator of the proportion indirectly, giving the interviewee anonymity and allowing a sincere answer. Therefore, a probabilistic relationship is established between a given answer and the sensitive question. Finally, a comparison is made between the direct question method that establishes the usual estimators for the proportion of successes and Warner's randomized response method, determining that Warner's method is effective.

KEYWORDS

Sensitive question, randomized response model, response bias, sample surveys.

1. INTRODUCCIÓN

En estudios sobre temáticas delicadas o muy personales, se presentan dos problemas: no dan respuesta y/o no contestan con veracidad.

Estos dos problemas generan sesgos de muestreo. Por lo cual radica la importancia de estudiar metodologías que mejoren la tasa de respuesta veraz ante este tipo de preguntas sin comprometer al entrevistado ante estas preguntas delicadas.

El objetivo principal es estudiar el modelo de respuesta aleatorizada propuesto por Warner y aplicar el mismo a los estudiantes de la Facultad de Ciencias Puras y Naturales de la Universidad Mayor de San Andrés, que se encuentran matriculados en la gestión 2019 para analizar el comportamiento de los alumnos de la facultad frente a preguntas sensibles.

2. MATERIALES Y MÉTODOS

La metodología para la presente investigación comprende de dos partes: la implementación de una encuesta y el método de Warner para la estimación de la proporción de personas con una característica sensible. La encuesta permitirá aplicar el método de entrevista directa a los estudiantes de la universidad matriculados en la gestión 2019, para comparar los resultados con el método de Warner. A partir de la encuesta se podrá analizar si el método de respuesta aleatorizada asegura realmente el anonimato de los sujetos y aumenta la probabilidad de obtener respuestas sinceras a preguntas sensibles reduciendo el error.

El Modelo de Respuesta Aleatorizada de Warner

Los individuos de una población pueden diferenciarse en cuanto si son portadores de un rasgo sensible X o no. Luego se busca la proporción π de los portadores de características sensibles en la población, donde

$$\pi = P(X = 1) \text{ y } 1 - \pi = P(X = 0),$$

también se puede describir como la probabilidad de llevar la característica sensible. A los encuestados se les presentan dos declaraciones siguiendo el siguiente esquema:

Declaración A: Soy el portador de la característica sensible X .

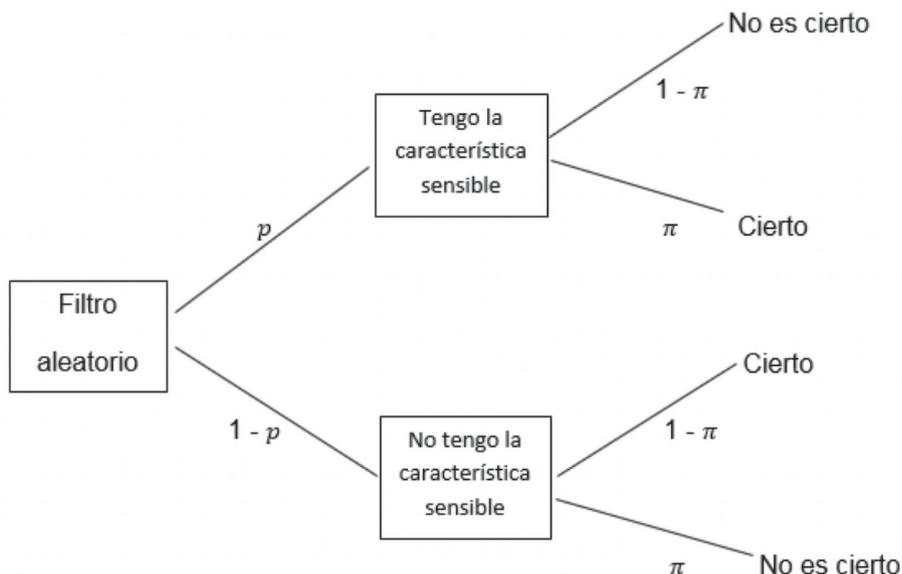
Declaración B: No soy el portador de la característica sensible X .

La selección modelo de Warner debe satisfacer dos condiciones:

1. Las probabilidades (p) de selección de las dos afirmaciones se conocen de antemano y no son iguales a 0.5.
2. El entrevistador no conoce el resultado del experimento aleatorio, solo el encuestado sabe cuál de las dos afirmaciones fue seleccionada. Luego solo indica si la declaración seleccionada se aplica a él o no. La figura 1 muestra esquemáticamente el principio de la encuesta a partir de diagrama de árbol.

Modelo de respuesta aleatorizada de Warner para incrementar la probabilidad de obtener respuestas sinceras a preguntas sensibles

Figura 1
Representación de la técnica de respuesta aleatoria según Warner (1965).



Fuente: Elaboración Propia

Cálculo de la probabilidad de tener la característica sensible λ :

$$\lambda = p\pi + (1-p)(1-\pi) ; p \neq 0.5$$

Donde

λ , probabilidad de una respuesta afirmativa

π , probabilidad de tener la característica sensible

p , probabilidad de responder la pregunta sensible

Por tanto, la estimación de la proporción de personas que tienen la característica sensible es

$$\hat{\pi} = \frac{\lambda + p - 1}{2p - 1}$$

Y su varianza es

$$V(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2}$$

Estimador de la pregunta directa

En el caso de preguntas directas, las estimaciones de la proporción de casos que responden si a la pregunta sensible son:

$$\hat{\pi} = \frac{\sum x_i}{n}$$

$$Var(\hat{\pi}) = \frac{\pi(1-\pi)}{n}$$

Con $x_i \sim \text{Bernoulli}(\pi)$.

Comparación del Modelo de respuesta aleatorizada de Warner con el modelo de pregunta directa

La equivalencia entre el modelo de Warner y el modelo de entrevista directa, se proporciona una base para comparar el modelo de Warner con el Modelo de respuesta directa.

A partir de sus varianzas:

$$Var(\hat{\pi}_w) = Var(\hat{\pi}_D) + \frac{p(1-p)}{n(2p-1)^2}$$

Donde $Var(\hat{\pi}_w)$ es la varianza del Modelo de Warner. Utilizando el criterio de varianzas se obtiene lo siguiente:

$$Var(\hat{\pi}_w) - Var(\hat{\pi}_D) = \frac{p}{n(1-p)(2p-1)^2} h_{WT}(p|\pi)$$

Con:

$$h_{WT}(p|\pi) = (4\pi - 3)p^2 + (2 - 4\pi)p + \pi$$

Para encontrar los valores de p para diferentes valores de π primero se resuelve la ecuación de segundo grado de la función $h_{WT}(p|\pi)$ en términos de p :

$$p = \frac{2\pi - 1 \pm \sqrt{1 - \pi}}{4\pi - 3} \quad ; \quad \pi \neq \frac{3}{4}$$

Eficiencia relativa del Modelo de Warner y el Modelo Directo

En este sentido se considera que la eficiencia relativa del modelo de Warner ($p \neq 0.5$) para el

modelo directo; es decir,

$$ER_{W \rightarrow D}(\pi, p) = \frac{Var(\hat{\pi}_W)}{Var(\hat{\pi}_D)} = 1 + \frac{p(1-p)}{\pi(1-\pi)(2p-1)^2}$$

La eficiencia relativa es el cociente de varianzas y es independiente del tamaño de la muestra n , y solo depende de los parámetros π y p , para el modelo de Warner como para el modelo Directo como se muestra a continuación en la tabla 1, se observa que a medida que la probabilidad de responder a la pregunta sensible se aproxima a 0.5, la eficiencia relativa del método directo aumenta exponencialmente, la eficiencia del modelo directo es mayor que la del modelo de Warner, en particular cuando ($0.48 \leq p < 0.50$) que es el rango óptimo para que la privacidad de los encuestados este protegido, la eficiencia del modelo directo es de aproximadamente 2500 hasta 6942 veces mayor que del modelo de Warner.

Tabla 1

Eficiencia Relativa del método de Ward en relación al método directo para varias combinaciones de π y p

π	p									
	0,10	0,20	0,30	0,40	0,49	0,51	0,60	0,70	0,80	0,90
0,10	2,56	5,94	15,58	67,67	6942,67	6942,67	67,67	15,58	5,94	2,56
0,20	1,88	3,78	9,20	38,50	3905,69	3905,69	38,50	9,20	3,78	1,88
0,30	1,67	3,12	7,25	29,57	2976,00	2976,00	29,57	7,25	3,12	1,67
0,40	1,59	2,85	6,47	26,00	2604,13	2604,13	26,00	6,47	2,85	1,59
0,49	1,56	2,78	6,25	25,01	2501,00	2501,00	25,01	6,25	2,78	1,56
0,51	1,56	2,78	6,25	25,01	2501,00	2501,00	25,01	6,25	2,78	1,56
0,60	1,59	2,85	6,47	26,00	2604,13	2604,13	26,00	6,47	2,85	1,59
0,70	1,67	3,12	7,25	29,57	2976,00	2976,00	29,57	7,25	3,12	1,67
0,80	1,88	3,78	9,20	38,50	3905,69	3905,69	38,50	9,20	3,78	1,88
0,90	2,56	5,94	15,58	67,67	6942,67	6942,67	67,67	15,58	5,94	2,56

Fuente: Elaboración Propia

Modelo de respuesta aleatorizada de Warner para incrementar la probabilidad de obtener respuestas sinceras a preguntas sensibles

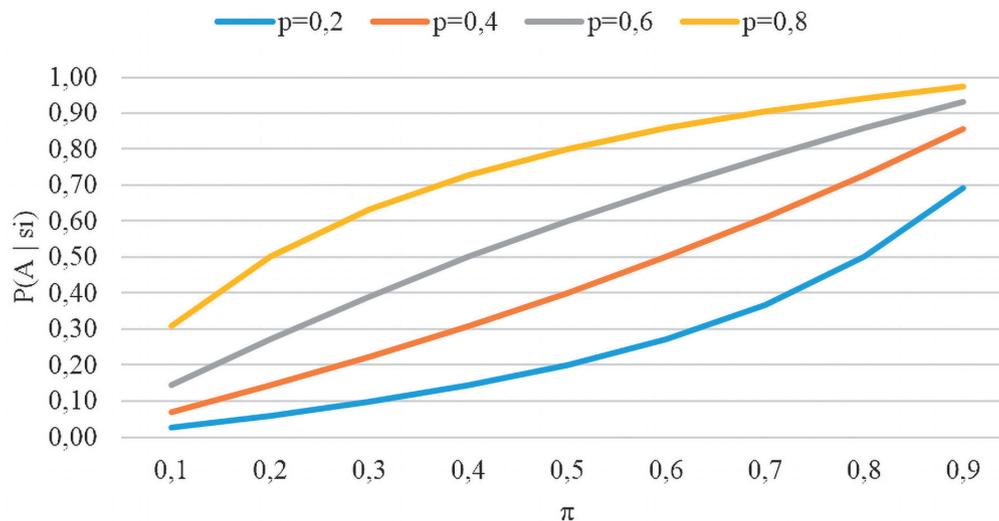
Grado de protección de la privacidad

$$GPP = P(A|si) = \frac{\pi p}{\pi p + (1 - \pi)(1 - p)}$$

Se define la probabilidad condicional, que el encuestado realmente posea la característica sensible dado que responde afirmativamente.

Donde A indica que el entrevistado posee la característica sensible.

Figura 2
Grado de protección de la privacidad para los que responden Si



Fuente: Elaboración Propia

Figura 3

Modelo de preguntas con respuesta aleatorizada de Warner

A continuación, contesta la opción **A** si tu cédula de identidad termina en 1 o 2, en caso contrario contesta la opción **B**.

opción A	opción B
Nunca hice trampa en un examen de la universidad	Alguna vez hice trampa en un examen de la universidad
SI <input type="radio"/>	NO <input type="radio"/>

A continuación, contesta la opción **A** si tu cédula de identidad termina en 8 o 9, en caso contrario contesta la opción **B**.

opción A	opción B
Nunca he consumido drogas	He consumido drogas en alguna ocasión
SI <input type="radio"/>	NO <input type="radio"/>

Fuente: Elaboración Propia

3. RESULTADOS

Tabla 2

Estimaciones de las preguntas sensibles por el método de Warner y el método Directo

Pregunta sensible	Método de Warner		Método Directo	
	π	IC	π	IC
¿Alguna vez hiciste trampa en un examen de la universidad?	0,5017	0,4290 - 0,5740	0,4160	0,3605 - 0,4715
¿Has consumido drogas en alguna ocasión?	0,1667	0,1002 - 0,2332	0,0594	0,0327 - 0,0860

Fuente: Elaboración Propia

Tabla 3

Estimaciones de las preguntas sensibles por el método de Warner por sexo

Pregunta sensible	Masculino		Femenino	
	π	IC	π	IC
¿Alguna vez hiciste trampa en un examen de la universidad?	0,4817	0,3808 - 0,5825	0,5233	0,4190 - 0,6276
¿Has consumido drogas en alguna ocasión?	0,2333	0,1377 - 0,3289	0,0950	0,0038 - 0,1861

Fuente: Elaboración Propia

4. CONCLUSIONES

1. La aplicación del modelo de respuesta aleatorizada de Warner ha demostrado que es una técnica que permite obtener mejores resultados en encuestas con preguntas sensibles.
2. La estimación de la proporción de estudiantes que hacen “trampa en los exámenes” utilizando el estimador de Warner es 50,17% mientras que en el diseño de entrevista directa es de 41,6%.
3. La estimación de “Consumo de drogas en alguna ocasión” utilizando el estimador de Warner es de 16,67%, mientras que con el diseño de entrevista directa es de 5,94%.
4. La desventaja principal del método de Warner radica en lo que respecta a la capacitación de los encuestadores y el tiempo que requiere explicar al entrevistado el cuestionario.

A la fecha existen otros métodos de estimación frente a preguntas sensibles, algunos de estos modelos son: los modelos de respuesta no aleatorizada y el modelo triangular de respuesta no aleatorizada. Se recomienda estudiar dichos métodos para futuros trabajos de investigación.

BIBLIOGRAFÍA

1. Basulto, J. (1982). El diseño de respuesta aleatorizada de Warner. Un modelo de superpoblación., (96), 51 a 62.
2. Guo-Liang Tian, J.-W. Y. (2007). A new non-randomized model for analysis sensitive questions with binary outcomes. *Statistics in medicine*, 26(23), 4238-52. <http://doi.org/10.1002/sim.2863>.
3. Liang, T. G., & Lai, T. M. (2008). Two new models for survey sampling with sensitive characteristic: desing and analysis. *Metrica*, 251-263.
4. Nayak, T. K. (1994). On randomized response surveys for estimating a proportion. *Communications in Statistics - Theory and Methods*, 23(11), 3303-3321. <http://doi.org/10.1080/03610929408831448>.
5. Warner, S. L. (1965). Randomized response: a survey tecnique for eliminating evasive answer bias. *Journal of Applied Psychology* (Vol. 60).