

El Modelo de Respuesta Aleatorizada Mixto de Kim y Warde: Aplicación a estudiantes de la FCM – UNMSM

*José Matos*¹

*Olga Solano*²

(Recibido: 03/10/2016 - Aceptado: 07/10/2016)

Resumen: El presente trabajo de investigación trata sobre el comportamiento de los estudiantes frente a preguntas sensibles en la Facultad de Ciencias Matemáticas utilizando el Modelo de Respuesta Aleatorizada Mixto (MRAM) propuesto por Kim y Warde (2005), el cual se aplicó en una encuesta realizada en la Facultad de Ciencias Matemáticas (FCM), para investigar el comportamiento de los jóvenes estudiantes con respecto a las malas prácticas académicas, realizar copias y consumo de drogas para mejorar el rendimiento académico en un examen. La población estudiada comprende a los alumnos matriculados el Semestre 2015-I pertenecientes a las cuatro Escuelas Profesionales de la Facultad de Ciencias Matemáticas de la UNMSM. En el diseño muestral, se utilizó el muestreo aleatorio estratificado con afijación proporcional a cada estrato (Scheaffer, Mendenhall, Ott, 2007) considerando como estrato a cada una de las Escuelas Profesionales de la FCM. Para el cálculo del tamaño de muestra se consideró un límite para el error de estimación de 5% con un nivel de confianza del 95% y la información proporcionada por la Dirección Académica de la FCM, de los estudiantes matriculados el primer Semestre del año académico 2015, el tamaño de muestra fue de 300 estudiantes, repartidos proporcionalmente a las cuatro Escuelas Profesionales de la FCM.

Palabras Claves: Modelo de respuesta aleatorizada mixto, tamaño de muestra, pregunta delicada, pregunta inocua.

Mixed Randomized Response Model Kim and Warde: Application to students FCM - UNMSM

Abstract: This research deals with the behavior of students against sensitive questions at the Faculty of Mathematical Sciences using the Mixed Randomized Response Model proposed by Kim and Warde (2005), which it was applied in a survey conducted at the Faculty of Mathematical Sciences (FCM), to investigate the behavior of young students regarding poor academic practices, make copies, and drug use to improve academic performance in an examination. The study population includes students enrolled in Semester 2015-I belonging to the four professional schools of the Faculty of Mathematical Sciences of San Marcos. Stratified random sampling with proportional allocation to size of each stratum (Scheaffer, Mendenhall, Ott, 2007) was used in the sample design, considering as strata to each of the Professional Schools (EP) of the FCM. For to calculation of sample size is considered a limit to the estimation error of 5% with a confidence level of 95% and information provided by the Academic Board of the FCM, of the students enrolled the first half of the year academic 2015, the sample size was 300 students, distributed proportion to the four EP of the FCM.

Key Words: Mixed randomized response model, sample size, sensitive question, innocuous question.

¹UNMSM, Facultad de Ciencias Matemáticas, e-mail: jos_nach05@hotmail.com

²UNMSM, Facultad de Ciencias Matemáticas, e-mail: osolanod@unmsm.edu.pe

1 Introducción

Una particularidad de la realización y empleo de encuestas es que, el investigador asume como legítima y verdadera la respuesta del entrevistado en cualquiera de los temas en los que se desea preguntar, es entonces que cuando se comienzan a involucrar preguntas personales o altamente sensibles, se sospecha de la veracidad de las respuestas puesto que muchas veces las personas entrevistadas se ven inclinadas a evitar responder preguntas comprometedoras, es decir que en estos casos se presentan la no respuesta y las respuestas falsas por parte de los entrevistados. Es posible obtener las respuestas verdaderas por parte de los entrevistados utilizando otro tipo de modelos de encuestas que no sea la entrevista directa.

Existen multitud de estudios realizados para mejorar la calidad y veracidad de las respuestas obtenidas sobre temas sensibles, como el consumo de drogas, alcohol, las prácticas delictivas o fraudulentas, prácticas abortivas o hábitos sexuales o de higiene. Frecuentemente los investigadores obtienen resistencias a la participación en estos estudios u obtienen respuestas falsas de parte de los entrevistados. Warner (1965) realizó la primera propuesta para obtener respuestas válidas ante preguntas embarazosas basándose en la realización de dos preguntas mutuamente excluyentes (por ejemplo, A: “Declaré mis ingresos extraordinarios el año pasado”; B: “No declaré mis ingresos extraordinarios el año pasado”).

Posteriormente, se desarrollaron otros métodos basados en Warner, como el denominado “método de alternativa forzada” de Fox y Tracy (1986), el método de la “pregunta inocua” tratado en los estudios de Campbell y Joiner (1973), o el método “bougus pipeline” desarrollado por Jones y Siggall (1971). El Modelo de Respuesta Aleatorizada (MRA) es un método especialmente diseñado para asegurar privacidad a los respondientes en el estudio de temas sensibles, delicados o embarazosos. Se intenta con ello evitar el sesgo de la respuesta de los entrevistados en ciertas conductas hacia la respuesta socialmente más deseable (Sudman y Bradburn, 1974). Se ha utilizado para analizar temas como copiar en los exámenes, insolvencia, fraudes, haber sido arrestado, conducir bajo los efectos del alcohol, tener un hijo fuera del matrimonio, aborto, etc. Es por eso que el objetivo del presente trabajo es mostrar la utilidad del Modelo de Respuesta Aleatorizada Mixto, que disminuyen la no respuesta y la respuesta falsa por parte de los entrevistados en preguntas sensibles.

2 Metodología

El modelo de respuesta aleatorizada mixto

Una muestra individual de tamaño n se selecciona por muestreo aleatorio simple con reemplazo de la población, se instruye a cada encuestado de la muestra para responder a la pregunta directa: “Soy un miembro del grupo inocuo”. Si el encuestado responde “Sí” a la pregunta directa, entonces es instruido para ir al mecanismo de aleatorización R1, que consiste en los enunciados: (i) “ Soy un miembro del grupo sensible” , y (ii) “Soy un miembro del grupo inocuo” con probabilidades preasignadas de selección P_1 y $1 - P_1$, respectivamente. Si el encuestado responde “No” a la pregunta directa, a continuación, se le indica al respondiente a utilizar el mecanismo aleatorio R2 que consiste en los enunciados: (i) “ Soy miembro del grupo sensible” , y (ii) “ Yo no soy miembro del grupo sensible” con probabilidades preasignadas P y $1 - P$, respectivamente.

Los procedimientos de la encuesta se realizan bajo el supuesto de que tanto las preguntas sensibles e inocuas no están relacionados y son independientes en el dispositivo de aleatorización R1.

Para proteger la privacidad de los encuestados, estos no deben divulgar al entrevistador la pregunta que respondieron de R1 o R2. Sea n el tamaño de la muestra frente a una pregunta directa,

n_1 y n_2 ($n_2 = n - n_1$) denotan el número de respuestas “Sí” y “No” de la muestra. Dado que todos los encuestados que utilizan el mecanismo de aleatorización R1 respondieron “Sí” a la pregunta inocua directa inicial, la proporción Y de conseguir “Sí” como respuesta de los encuestados que utilizan el mecanismo de aleatorización R1 es:

$$Y = P_1\pi_S + (1 - P_1)\pi_I = P_1\pi_S + (1 - P_1),$$

donde π_S es la proporción de respuestas “Sí” a partir del grupo sensible, π_I es la proporción de respuestas “Sí” de la pregunta inocua y P_1 es la probabilidad de pertenecer al grupo sensible en el mecanismo aleatorio R1. Un estimador insesgado de π_S , en términos de la proporción de la muestra de respuestas “Sí” de los encuestados que utilizan el mecanismo de aleatorización R1, \hat{Y} , es: $\hat{\pi}_a = \frac{\hat{Y} - (1 - P_1)}{P_1}$. Su varianza es $var(\hat{\pi}_a) = \frac{Y(1-Y)}{n_1 P_1^2} = \frac{(1 - \pi_S)[P_1\pi_S + (1 - P_1)]}{n_1 P_1}$. La proporción de respuestas “Sí” de los encuestados que utilizan el mecanismo de aleatorización R2 es de la siguiente manera: $X = P\pi_S + (1 - P)(1 - \pi_S) = (2P - 1)\pi_S + 1 - P$, donde P es la probabilidad de pertenecer al grupo sensible en el mecanismo aleatorio R2, π_S es la proporción de respuestas “Sí” a partir del grupo sensible.

Un estimador insesgado de π_S , en términos de la proporción, de la muestra, de respuestas “Sí” de los encuestados que utilizan el mecanismo de aleatorización R2, \hat{X} , es: $\hat{\pi}_b = \frac{\hat{X} - (1 - P)}{2P - 1}$. La varianza de $\hat{\pi}_b$ es: $var(\hat{\pi}_b) = \frac{X(1-X)}{n_2 P^2} = \frac{\pi_S(1 - \pi_S)}{n - n_1} + \frac{P(1-P)}{(n - n_1)(2P - 1)^2}$.

El estimador de π_S , en términos de las proporciones de la muestra de respuestas “Sí”, \hat{Y} y \hat{X} , es $\hat{\pi}_m = \frac{n_1}{n}\hat{\pi}_a + \frac{n_2}{n}\hat{\pi}_b = \frac{n_1}{n}\hat{\pi}_a + \frac{(n - n_1)}{n}\hat{\pi}_b$ con $0 < \frac{n_1}{n} < 1$. Como ambos $\hat{\pi}_a$ y $\hat{\pi}_b$ son estimadores insesgados, el valor esperado de $\hat{\pi}_m$ es $E(\hat{\pi}_m) = E\left(\frac{n_1}{n}\hat{\pi}_a + \frac{(n - n_1)}{n}\hat{\pi}_b\right) = \frac{n_1}{n}\pi_S + \frac{n - n_1}{n}\pi_S = \pi_S$.

Puesto que los dos dispositivos de aleatorización son independientes, obtenemos la varianza de $\hat{\pi}_m$ es $var(\hat{\pi}_m) = \frac{n_1}{n^2} \left[\frac{(1 - \pi_S)[P_1\pi_S + (1 - P_1)]}{P_1} \right] + \frac{n - n_1}{n^2} \left[\pi_S(1 - \pi_S) + \frac{P(1 - P)}{(2P - 1)^2} \right]$.

Para que la varianza del estimador en el MRAM sea más pequeña, tenemos que asignar más encuestados al mecanismo de aleatorización R1 que al mecanismo de aleatorización R2. Ya que n_1 y $n_2 = n - n_1$ en el modelo propuesto son aleatorios (es decir, no fijos) las variables, como se ha señalado por un árbitro, a veces es difícil de conseguir una mejora en la eficiencia de este modelo en comparación con el modelo de Moors (1971). Sin embargo, si un investigador elige cuidadosamente una pregunta directa inocua en el modelo, entonces el problema puede ser rectificado.

3 Resultados y Discusión

El presente estudio se realizó en la Facultad de Ciencias Matemáticas de la UNMSM. Se aplicó el modelo de respuesta aleatorizado mixto que fue propuesto por Kim & Warde (2005) en una encuesta sobre preguntas sensibles con temas como prácticas académicas indebidas, copiar, suplantación de información y uso de drogas para mejorar el desempeño académico. Para el cálculo del tamaño de muestra se utilizó el esquema de muestreo estratificado (Scheaffer, Mendenhall, Ott, 2007) con afijación proporcional al tamaño de cada estrato, donde cada estrato es una Escuela Profesional, con un límite para el error de estimación de 5% y la información proporcionada por la Dirección Académica. Los alumnos fueron seleccionados aleatoriamente del listado de alumnos matriculados en el Semestre 2015-I, hasta completar el tamaño de muestra igual a 300. De acuerdo a Registros Académicos de la Dirección Académica los Alumnos Matriculados el Semestre 2015 - I en la Facultad de Ciencias Matemáticas de la UNMSM, está dividida en cuatro Escuelas Profesionales (ver Tabla 1).

Tabla 1: Distribución de los estudiantes de Pre-Grado según Escuela Académico profesional
Semestre 2015-I - FCM-UNMSM

Escuela Acémico Profesional	Frecuencia	Porcentaje
Matemática	458	30,11 %
Estadística	291	19,13 %
Investigación Operativa	471	30,97 %
Computación Científica	301	19,79 %
Total	1521	100,00 %

Fuente: Dirección Académica - FCM - UNMSM

La entrevista se realizó del 08 de junio al 13 de julio del año 2015. La distribución de la muestra por Escuela Profesional se muestra en la Tabla 2.

Tabla 2: Distribución de la muestra según Escuela Académico Profesional
Semestre 2015-I - FCM-UNMSM

Escuela Académico Profesional	Femenino	Masculino	Total
Matemática	17	73	90
Estadística	26	32	58
Investigación Operativa	41	52	93
Computación Científica	15	44	59
Total	99	201	300

Instrumento de recolección de la información

Elección del procedimiento aleatorizado

El procedimiento aleatorio que se utiliza para seleccionar uno de dos enunciados, cada uno de los cuales requiere de una respuesta “SI” o “No” por parte del entrevistado sin revelar al entrevistador su posición personal respecto a la pregunta, es un grupo de tarjetas. Para el mecanismo aleatorio 1 y 2, se utilizó las siguientes tarjetas:

- Diez tarjetas de color verde para “copia en exámenes”.
- Diez tarjetas de color amarillo para “uso de notas en un examen”
- Diez tarjetas de color plomo para “uso de drogas”
- Diez tarjetas de color púrpura para “usurpación de información”

Procedimiento

El entrevistado es requerido por el entrevistador a elegir la tarjeta de la pregunta inocua, que contiene el siguiente enunciado : “¿Su D.N.I. termina en número impar (1,3,5,7 ó 9)?, si su respuesta es afirmativa va al mecanismo aleatorio 1, elige una tarjeta de un grupo de diez tarjetas, y luego, sin mostrarle el contenido al entrevistado, lee en silencio y responde en voz alta “ SI” o “ No” , procediendo inmediatamente a juntar la tarjeta seleccionada con las demás de su grupo. Si su respuesta es negativa a la pregunta inocua, va al mecanismo aleatorio 2, y elige una tarjeta de un grupo de diez tarjetas, y luego, sin mostrarle el contenido, lee en silencio y responde en voz

alta “ SI” o “ No” , procediendo inmediatamente a juntar la tarjeta seleccionada con las demás de su grupo.

Para aplicar el Modelo de Respuesta Aleatorizada Mixto se utilizó las siguientes tarjetas con los diferentes enunciados, para el mecanismo aleatorio 1 y 2:

Copia en los exámenes

Diez tarjetas de color verde

- Ocho con el enunciado: A)“ ¿Ud. ha copiado de otros estudiantes durante un examen (ó práctica) por lo menos una vez?”
- Dos con la pregunta “ Inocua” B) ¿Su D.N.I. termina en número impar (1,3,5,7 ó 9)?

Uso de notas en un examen

Diez tarjetas de color amarillo

1. Ocho con el enunciado: A)“ ¿Ud. ha utilizado notas en un examen (incluyendo notas en los teléfonos móviles, calculadoras u otros similares) por lo menos una vez?
2. Dos con la pregunta “ Inocua” B) ¿Su D.N.I. termina en número impar (1,3,5,7 ó 9)?

Uso de drogas

Diez tarjetas de color plomo

- Ocho con el enunciado: A) ¿Ud. Ha utilizado drogas para mejorar su rendimiento académico en una práctica o un examen por lo menos una vez?
- Dos con la pregunta “ Inocua” B)¿Su D.N.I. termina en número impar (1,3,5,7 ó 9)?

Usurpación de información

Diez tarjetas de color púrpura

- Ocho con el enunciado: A) ¿Ud. ha presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona por lo menos una vez?
- Dos con la pregunta “ Inocua” B) ¿Su D.N.I. termina en número impar (1,3,5,7 ó 9)?

En cada caso la probabilidad (P) de seleccionar el enunciado A) es 0,80 y la probabilidad (1-P) de seleccionar el enunciado B) es 0,20.

Al utilizar las fórmulas de los estimadores de la proporción de respondientes en la población con la pregunta sensible del Modelo de Respuesta Aleatorizada Mixta presentado anteriormente, se obtuvieron los siguientes resultados:

Tabla 3: Resultados del Modelo de Respuesta Aleatorizada Mixto según estimación del porcentaje de las “ preguntas sensibles” y error de estimación

Porcentaje de estudiantes que han	Estimación	Error estándar
copiado durante un examen por lo menos una vez	62,4	0,0418
utilizado notas en un examen por lo menos una vez	52,8	0,0427
utilizado drogas para mejorar su rendimiento académico	3,6	0,0356
presentado copia adaptada del trabajo de otra persona	46,6	0,0429

En la Tabla 3 se observa que utilizando MRAM los alumnos que han copiado por lo menos una vez en un examen o práctica calificada es de 62,4 %; los estudiantes que utilizan notas en un examen por lo menos una vez es 52,8 %; al respecto de los estudiantes que han consumido drogas para mejorar el rendimiento académico por lo menos una vez es de 3,6 % y observamos también que los estudiantes que han presentado un documento con el contenido de otro trabajo por lo menos una vez es de 46,6 %.

4 Conclusiones

De los resultados se concluye que el modelo de respuesta aleatorizada, fue útil y provechosa en una investigación de la realidad de nuestra sociedad identificar el comportamiento en los estudiantes de la FCM frente a preguntas sensibles.

Una desventaja de aplicar el MRAM es el tiempo en que se demora al explicar el procedimiento al realizar la entrevista, y también el posible tiempo y costo de capacitación para que los entrevistadores lo hagan de la mejor manera y se obtenga la información correctamente. Se recomienda seguir experimentando estos modelos, en muestras más grandes y en temas en donde la pregunta sea efectivamente altamente sensible, para obtener la respuesta verdadera de los entrevistados frente a preguntas sensibles en diversos temas sociales.

5 Agradecimiento

Los autores expresamos nuestro agradecimiento al Consejo Superior de Investigaciones de la Universidad Nacional Mayor de San Marcos por el apoyo financiero para la ejecución del estudio motivo de la presente publicación.

Referencias Bibliográficas

- [1] Fox, J.A., Tracy, P.E.(1986). *Randomized response: A method for sensitive surveys*. Sage University Paper on Quantitative Applications in the Social Sciences, 58, Beverly Hills: Sage Publications.
- [2] Ayesha, N.(2010). *A New Mixed Randomized Response Model*. International Journal of Business and Social Science. Vol 1, N^o 1.
- [3] Basulto, J. (1982). *El diseño de respuesta aleatorizada de Warner: Un modelo de superpoblación*. Estadística Española N^o 96, 51- 62.
- [4] Bouza, C.N., Herrera, C.,Pasha,M.G. (2010). *A review of randomized responses procedures: the qualitative variable case*. Revista Investigación Operacional. Vol 31, N^o 3: 240-247.
- [5] Campbell, C., Joiner, B.L.(1973). *How to get answers without being sure your're asked the question?*The American Statistician. N^o 27, 229-232.
- [6] Cobo, R.B. (2013). *Respuesta Aleatoria y Técnicas de preguntas indirectas*. Departamento de Estadística e Investigación Operativa. Universidad de Granada.
- [7] Jones, E.E., Sigall, H.(1971). *The bogus pipeline: A new paradigm for measuring affectand attitude*.Psychological Bulletin, Vol 76, 349-364.
- [8] Kim, J.M., Warde, W. D. (2005) *A mixed randomized response model* . Journal of Statistical Planning and Inference. Vol 133, N^o 1, 211-221.
- [9] Saha, A.(2005). *Kim and Warde's mixed randomized response technique for complex surveys*.Journal of Modern Applied Statistical Methods. Vol 4, N^o 2, 538-544.
- [10] Scheaffer, R., Mendenhall, W. y Ott, L. (2007). *Elementos de muestreo*. Grupo Edit. International Thomson Paraninfo S.A., Madrid, España.
- [11] Solano, O., Cárdenas, A., Adriazola, Y., Bartolo, F., Martinez, B., Giraldo, O. (2010). *Modelos de Respuesta aleatoria estratificada: Una aplicación a estudiantes de la Facultad de Ciencias Matematicas de la UNMSM*. Pesquimat, Revista de la F.C.M. de la UNMSM. Vol xiii, N^o1, 71-85.
- [12] Usman, A., Oshungade, I.O.(2012). *A Mixed-Stratified Randomized Response Model for HIV Seroprevalence Surveys*. Research Journal of Mathematics and Statistics Vol 4, N^o3, 70-75.
- [13] Warner, S.L. (1965). *Randomized response: a survey technique for eliminating evasive answer bias*. Journal of the American Statistical Association. Vol 60, 63-69.