

Modelos de Respuesta Aleatoria Estratificada: Una Aplicación a Estudiantes de la Facultad de Ciencias Matemáticas de la UNMSM

*Olga Solano Dávila*¹,

*Doris Gómez Ticerán*², *Ana Cárdenas Rojas*², *Ysabel Adriazola Cruz*², *Felix Bartolo Gotarate*²,
*Blanca Martínez Portuguez*² & *Orlando Giraldo Laguna*²

Resumen: En el presente trabajo estudiamos tres Modelos de Respuestas Aleatorizada (MRA) Estratificada, utilizadas en encuestas donde se utilizan preguntas delicadas (Kim y Elam,2003). Realizamos una aplicación en el comportamiento de jóvenes de la Facultad de Ciencias Matemáticas de la UNMSM, para investigar la proporción de personas que han consumido pasta básica de cocaína (PBC) por lo menos una vez en su vida; proporción de consumidores actuales del PBC; proporción de personas que han tenido relaciones sexuales con más de dos personas durante toda su vida; la proporción de personas que consumen alcohol todos los fines de semana y la proporción de persona que han llevado o han consumido sin pagar algún producto de algún supermercado. La población en estudio comprende los alumnos matriculados en la FCM el semestre 2009-II

Palabras clave: Modelos de respuesta aleatorizada estratificada, mecanismo aleatorizado, Pregunta delicada, Pregunta no relacionada.

Stratified Randomized Response Models: An Application to Students of the Faculty of Mathematics of the UNMSM

Abstract: In the present work, we study three stratified randomized response models, used in surveys that involve sensitive question (Kim and Elam,2003). Made an application in the behavior of youths from the Faculty of Mathematical Sciences of San Marcos, to investigate the proportion of people who have consumed cocaine basic paste (CPB) at least once in life and the proportion of current users of the CPB and the proportion of people who had sex with more than two people for life and the proportion of people who consuming alcohol every weekend and the proportion of people who have been or are consumed without paying any products from a supermarket. The population in study were students matriculated in the half-yearly 2009 - II.

Key words: Stratified randomized response models, randomization device, sensitive question, unrelated question.

1. Introducción

Existen multitud de estudios realizados para mejorar la calidad y veracidad de las respuestas obtenidas a través de encuestas, sobre temas comprometidos, 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.

¹UNMSM, Departamento de Estadística, Lima- Perú, e-mail: osolanod@unmsm.edu.pe,

²UNMSM, Departamento de Estadística, Lima- Perú

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), el modelo de la pregunta no relacionada, en los estudios de Greenberg et al (1969), Moors (1971) y Lanke (1975) o el método "bogus pipeline" desarrollado por Jones y Sigall (1971), el modelo de respuesta aleatorizada para obtener datos cuantitativos presentado en Greenberg (1971), entre otros.

La técnica de respuesta aleatorizada (Chaudhuri, 1983) 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 los respondientes en ciertas conductas hacia la respuesta socialmente más deseable (Sudman y Bradburn, 1974, Shimizu y Bonham, 1971, Musch y Klauer, 2001). 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, consumo de bebidas alcohólicas (Solano, 2004), Solano etc.

El objetivo del presente trabajo es comparar el MRA estratificada de Warner con respecto a otros MRA estratificada existentes como el MRA Estratificada de la Pregunta No Relacionada y el MRA Estratificada en dos Etapas, para el tratamiento de preguntas comprometidas, con respecto al sexo, drogas, alcohol y hurto en estudiantes matriculados en la Facultad de Ciencias Matemáticas de la UNMSM.

2. Metodología

2.1. El Modelo de Respuesta Aleatorizada

En investigaciones realizadas en el pasado, en las cuales se involucran preguntas altamente delicadas o personales, se detecta muchas veces que las personas no están inclinadas a responder con honestidad o simplemente señalan una respuesta negativa. Esto es, la no respuesta, respuestas evasivas y respuestas falsas son tan comunes en la práctica que difícilmente pueden ser controladas. Para solucionar este problema es posible utilizar procedimientos alternativos, si nuestra intención es obtener datos confiables en esta clase de investigación, en lugar de realizar pesquisas por la modalidad de entrevista directa. La primera técnica de Respuesta Aleatorizada fue introducida por Warner (1965), con el objetivo de obtener datos confiables; para esto se diseñan hábilmente las preguntas de tal manera que no se revela la identidad del entrevistado durante la entrevista. El entrevistador debe utilizar un mecanismo aleatorio (urna con bolas, flecha giratoria, una moneda, etc.) para seleccionar una de dos preguntas (pertenezco al grupo A o no pertenezco al grupo A), cada una de las cuales requiere de una respuesta "Sí" o "No" por parte del entrevistado sin revelar al entrevistador su posición personal con respecto a la pregunta.

El modelo estima la proporción de elementos de la población que pertenece al grupo A a partir de una muestra aleatoria con reposición.

Con el propósito de que el entrevistado tenga una mayor disposición a colaborar, Greenberg, Abul-Ela, Simmons y Horvitz (1969) modificaron el modelo original de Warner, sustituyendo la segunda proposición ("No pertenezco al grupo A"), por otra pregunta referente a una característica B, no relacionada con A.

En consecuencia el entrevistado debe seleccionar a través de un procedimiento aleatorio (flecha giratoria, una moneda, urna con bolas, etc.) una de las dos siguientes proposiciones a las que se supone responde correctamente.

1. Pertenezco al grupo A.

2. Pertenezco al grupo B .

Los parámetros que se desea estimar son: π_A y π_B (Proporción de elementos de la población que pertenecen al grupo A y B , respectivamente) en base a dos muestras aleatorias independientes de tamaño n_1 y n_2 respectivamente, esto se realiza cuando π_B es desconocido.

En el caso que π_B fuese conocido, solo se requiere de una muestra aleatoria, y las proposiciones a las que se supone correctamente son:

1. Pertenezco al grupo A , presentada con probabilidad P .
2. Pertenezco al grupo B , presentada con probabilidad $1 - P$.

Donde

$$x_i = \begin{cases} 1, & \text{si } i\text{-ésimo elemento de la muestra responde "Sí"} \\ 0, & \text{si } i\text{-ésimo elemento de la muestra responde "No"} \end{cases}$$

$$P(x_i = 1) = P\pi_A + (1 - P)\pi_B$$

Y utilizando el método de momentos se tiene

$$E(x_i = 1) = P(x_i = 1) = \frac{\sum_{i=1}^n x_i}{n}$$

Luego reemplazando tenemos

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

Que es el estimador de la proporción de elementos de la población que pertenecen al grupo A (π_A).

Estimador insesgado de π_A

$$E(\hat{\pi}) = \frac{E(x_i) - (1 - P)\pi_B}{P}$$

$$E(\hat{\pi}_A) = \frac{P\pi_A + (1 - P)\pi_B - (1 - P)\pi_B}{P} = \pi_A$$

$\hat{\pi}_A$, es un estimador de π_A .

Varianza de π_A

$$Var(\hat{\pi}_A) = \frac{\mu(1 - \mu)}{nP^2}, \text{ donde } \mu = P\pi_A + (1 - P)\pi_B$$

2.2. El Modelo de Respuesta Aleatorizada Estratificada de Warner

En el Muestreo Aleatorizado estratificado la población a investigar se divide en estratos. Se selecciona una muestra aleatoria simple con reemplazamiento de cada estrato para conseguir todo el beneficio de estratificación, se asume que el número de unidades en cada estrato es conocido.

En el Modelo de Warner estratificada y el Modelo de Respuesta Aleatorizada de la pregunta no relacionada, un respondiente individual en la muestra del estrato i es instruido para usar el mecanismo aleatorizado R_i el cual consiste de una tarjeta con pregunta sensible (S) con probabilidad P_i y una tarjeta con pregunta negativa (\bar{S}) con probabilidad $1 - P_i$. El respondiente

o entrevistado responde la pregunta con un "sí" o "no" sin reportar que tarjeta ha elegido y que pregunta está respondiendo. Un respondiente de diferente estrato llevará a cabo la asignación al azar de diferentes mecanismos aleatorizados, el cual tiene diferentes probabilidades preasignadas. Bajo la suposición que las respuestas reportadas "sí" o "no" son verdaderas y P_i son fijadas por el investigador, la probabilidad de responder "sí" en el estrato " i " por el modelo de Respuesta Aleatorizada de Warner estratificado es:

$$Z_i = p_i \pi_{s_i} + (1 - p_i)(1 - \pi_{s_i}), \text{ para } i = 1, 2, \dots, k,$$

donde

Z_i : es la proporsición de respuestas "sí" es el estrato i

P_i : Probabilidad de seleccionar la pregunta sensible en el estrato i

π_{s_i} : es la proporción de respondientes en la población con la pregunta sensible en el estrato i

n_i : denota el número de unidades en la muestra de estratos i , n : número total de unidades en las muestras de todos los estratos

Estimador de π_{s_i}

El estimador de π_{s_i} es expresado de la siguiente manera:

$$\hat{\pi}_{s_i} = \frac{Z_i}{2(P_i - 1)} + \frac{(P_i - 1)}{(2P_i - 1)},$$

donde $Z_i = \frac{\sum_{i=1}^n X_i}{n_i}$.

Sea n_i que denota el número de unidades en las muestras de cada estrato entonces $n = \sum_{i=1}^k n_i$, Kim y Warde (2003) muestra la afijación óptima de n para n_1, n_2, \dots, n_{k-1} y n_k para derivar la mínima varianza de π_{s_i} , el cual es el siguiente:

$$\frac{n_i}{n} = \frac{w_i \left[\pi_{s_i}(1 - \pi_{s_i}) + \frac{P_i(1 - P_i)}{(2P_i - 1)^2} \right]^{1/2}}{\sum_{i=1}^k w_i \left[\pi_{s_i}(1 + \pi_{s_i}) + \frac{P_i(1 - p_i)}{(2P_i - 1)^2} \right]^{1/2}}$$

Donde $w_i = \frac{N_i}{N}$ para $i = 1, 2, \dots, k$ (N es el tamaño de la población y N_i es el total de elementos o unidades en el estrato i). Bajo la suposición de que $n_i = n \left(\frac{N_i}{N} \right)$ y $P_i = P$ para todo i , Kim y Warde (2003) muestra la mínima varianza de $\hat{\pi}_s$ como sigue:

$$Var(\hat{\pi}_s) = \frac{1}{n} \left[\sum_{i=1}^k w_i \left\{ \pi_{s_i}(1 - \pi_{s_i}) + \frac{2(1 - p)}{(2p - 1)^2} \right\}^{1/2} \right]^2$$

2.3. El Modelo de Respuesta Aleatorizada Estratificado de la Pregunta no Relacionada

Para el Modelo de Respuesta Aleatorizada Estratificado de la pregunta no relacionada, la probabilidad de obtener una respuesta "sí" en el estrato " i " es:

$$Z_i = P_i \pi_{s_i} + (1 - p_i) \pi_N, \text{ para } i = 1, 2, \dots, k,$$

donde π_N : se asume conocido y es la proporción de respondientes con la pregunta no sensible en el estrato i .

El Estimador de π_{s_i}

El estimador de π_{s_i} en el Modelo de la Pregunta No relacionada es expresado de la siguiente manera:

$$\hat{\pi}_{s_i} = \frac{Z_i - (1 - P_i)\pi_N}{P_i},$$

donde $Z_i = \frac{\sum_{i=1}^n X_i}{n_i}$.

Kim y Elam (2003a) muestra la afijación óptima de n para n_1, n_2, \dots, n_{k-1} y n_k , para derivar la mínima varianza de π_s , el cual es el siguiente:

$$\frac{n_i}{n} = \frac{\frac{w_i}{p_i} \sqrt{z_i(1 - z_i)}}{\sum_{i=1}^k \frac{w_i}{P_i} \sqrt{Z_i(1 - Z_i)}}.$$

Kim y ELam(2003a) también muestra la varianza mínima de $\hat{\pi}_s$ como sigue:

$$var(\hat{\pi}_s) = \frac{1}{n} \left[\sum_{i=1}^k \frac{w_i}{P_i} \sqrt{z_i(1 - z_i)} \right]^2$$

2.4. Modelo de Respuesta Aleatorizada Estratificada en dos Etapas

En el Modelo de Respuesta Aleatorizada estratificada en dos etapas, en la primera etapa de la encuesta el entrevistador requiere de un respondiente en la muestra del estrato i para usar el mecanismo aleatorizado R_{1i} el cual consiste de una tarjeta con pregunta sensible (S) con probabilidad M_i y otra tarjeta con el enunciado "Ir al mecanismo aleatorizado R_{2i} en la segunda etapa" con probabilidad $1 - M_i$. Los respondientes en la segunda etapa del estrato " i " son instruidos para usar el mecanismo aleatorizado R_{2i} , el cual consiste de una pregunta sensible (S) con probabilidad P_i y la pregunta negativa (\bar{S}) con probabilidad $1 - P_i$. Los respondientes responden la pregunta con un "sí" o "no" sin reportar que pregunta ha respondido para proteger la privacidad de respondiente bajo la suposición que M_i y P_i son fijados por el investigador, la probabilidad de una respuesta "sí" en el estrato i es:

$$Y_i = M_i\pi_{s_i} + (1 - M_i)[P_i\pi_{s_i} + (1 - P_i)(1 - \pi_{s_i})] \quad \text{para } i = 1, 2, \dots, k,$$

donde

Y_i : proporción de respuesta "sí" en el estrato i

M_i : Probabilidad de seleccionar la pregunta sensible en el estrato i , en la primera etapa

π_{s_i} : es la proporción de respondientes en la población con la pregunta sensible en el estrato i

P_i : Probabilidad de seleccionar la pregunta delicada en el estrato i , en la segunda etapa.

El Estimador de π_{s_i}

El estimador de π_{s_i} en el Modelo de Respuesta Aleatorizada en dos etapas es expresado de la siguiente manera:

$$\hat{\pi}_{s_i} = \frac{Y_i - (1 - P_i)^2}{(4P_i - 2P_i^2 - 1)},$$

Kim y Elam (2003b) muestra la afijación óptima de n para n_1, n_2, \dots, n_{k-1} y n_k para derivar la varianza mínima de π_s como sigue:

$$\frac{n_i}{n} = \frac{w_i \left\{ \pi_{s_i}(1 - \pi_{s_i}) + \frac{(1 - M_i)(1 - P_i)[1 - (1 - M_i)(1 - P_i)]}{[2P_i - 1 + 2M_i(1 - P_i)]^2} \right\}^{1/2}}{\sum_{i=1}^k \left\{ w_i \pi_{s_i}(1 - \pi_{s_i}) + \frac{(1 - M_i)(1 - P_i)[1 - (1 - M_i)(1 - P_i)]}{[2P_i - 1 + 2M_i(1 - P_i)]^2} \right\}^{1/2}}$$

Kim y Elam (2003b) también muestran la varianza mínima del estimador π_s como sigue

$$VAR(\hat{\pi}_s) = \frac{1}{n} \left[\sum_{i=1}^k w_i \left\{ \pi_{s_i}(1 - \pi_{s_i}) + \frac{(1 - M_i)(1 - P_i)[1 - (1 - M_i)(1 - P_i)]}{[2P_i - 1 + 2M_i(1 - P_i)]^2} \right\}^{1/2} \right]^2$$

3. Aplicación

El estudio fue realizado en la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos. De acuerdo a Registros Académicos de la Dirección Académica los Alumnos Matriculados el Semestre 2009 - II en la Facultad de Ciencias Matemáticas de la UNMSM, está dividida en cuatro Escuelas académico Profesionales, de acuerdo a esta información se decidió considerar a cada Escuela Académico Profesional como un estrato, en total tenemos cuatro estratos o Escuelas.

El esquema de muestreo que se utilizó fue el Muestreo Aleatorio Estratificado, con afijación proporcional de acuerdo a la cantidad de alumnos en cada uno de los estratos donde cada estrato es una Escuela Académico Profesional de la Facultad de Ciencias Matemáticas de la UNMSM. Se utilizó el Muestreo Sistemático Estratificado (Cochran, 1977; Des, 1980), con un nivel de confianza del 95 % y un margen de error del 4,5 %, el tamaño de muestra fue de 301 alumnos matriculados el Semestre 2009-II. El Modelo de Respuesta Aleatorizada requiere un tamaño de muestra más grande que el método convencional. La distribución de la muestra por Escuela Académico Profesional se presenta en el Cuadro 1.

Cuadro 1: Método de Respuesta Aleatorizada Estratificada: Distribución de la muestra según Escuela Académico Profesional, Semestre 2009-II, FCM-UNMSM

Escuela Académica Profesional	Frecuencia
Matemática	88
Estadística	57
Investigación Operativa	95
Computación Científica	61
Total	301

Instrumentos 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 "Sí" o "No" por parte del entrevistado sin revelar al entrevistador su posición personal respecto a la pregunta, es un GRUPO DE TARJETAS.

- Diez tarjetas de color blanco para el enunciado de "Prevalencia de Vida de Drogas", y
- Diez tarjetas de color palo rosa para el enunciado de "Uso actual de Drogas".

- Diez tarjetas de color rosado para la pregunta de “Prevalencia de sexo”,
- Diez tarjetas de color celeste para “Uso actual de alcohol”, y
- Diez tarjetas de color naranja para “Prevalencia de hurto”.

El entrevistado es requerido por el entrevistador a elegir una tarjeta del grupo de diez, y luego, sin mostrarle el contenido, leerla en silencio y responder en voz alta “sí” o “no”, procediendo inmediatamente a juntar la tarjeta seleccionada con las demás de su grupo.

Para aplicar el **Modelo de Respuesta Aleatorizada Estratificada de Warner** se utilizó las siguientes tarjetas con los diferentes enunciados:

Prevalencia de vida de drogas

Diez tarjetas de color Blanco

- Ocho con el enunciado:

A) “Durante mi vida he consumido, pasta básica de cocaína (pasta, pay o pastel) por lo menos una vez”

- Dos con el enunciado:

B) “Durante mi vida jamás he consumido pasta básica de cocaína (pasta, pay o pastel) por lo menos una vez”

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

Uso actual de drogas

Diez tarjetas de color Palo Rosa

- Ocho con el enunciado:

A) “Actualmente soy consumidor de pasta básica de cocaína (pasta, pay o pastel)”

- Dos con el enunciado;

B) “Actualmente no soy consumidor de pasta básica de cocaína (pasta, pay o pastel)”

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

Prevalencia de sexo

Diez tarjetas de color Rosado

- Ocho con el enunciado:

A) Durante mi vida he mantenido relaciones sexuales con más de dos personas (simultáneamente o no)

- Dos con el enunciado:

B) Durante mi vida jamás he mantenido relaciones sexuales con más de dos personas.

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

Consumo actual de alcohol

Diez tarjetas de color Celeste

- Ocho con el enunciado:

A) ¿Consume alcohol (cerveza, vino, sangría, etc) con frecuencia? (todos los fines de semana)

- Dos con el enunciado;

B) No consumo alcohol (cerveza, vino, sangría, etc.) con frecuencia (todos los fines de semana)

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

Prevalencia de hurto

Diez tarjetas de color Naranja

- Ocho con el enunciado:

A) He llevado o he consumido sin pagar algún producto de algún supermercado

- Dos con el enunciado;

B) Jamás he llevado o he consumido sin pagar algún producto de algún supermercado

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

Para aplicar el **Modelo de Respuesta Aleatorizada Estratificada de la Pregunta No Relacionada** se utilizó las siguientes tarjetas con los diferentes enunciados:

Prevalencia de vida de drogas

Diez tarjetas de color Blanco

- Ocho con el enunciado:

A) "Durante mi vida he consumido, pasta básica de cocaína (pasta, pay o pastel) por lo menos una vez"

- Dos con el enunciado:

B) ¿Utiliza Messenger?

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

Uso actual de drogas

Diez tarjetas de color Palo Rosa

- Ocho con el enunciado:

A) "Durante mi vida he consumido, pasta básica de cocaína (pasta, pay o pastel) por lo menos una vez"

- Dos con el enunciado;

B) ¿Utiliza ud. Página Web?

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

Prevalencia de sexo

Diez tarjetas de color Rosado

- Ocho con el enunciado:

A) ¿Ha mantenido relaciones sexuales con más de dos personas a lo largo de su vida?

- Dos con el enunciado;

B) ¿Alguna vez interrumpió sus estudios universitarios?

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

Consumo actual de alcohol

Diez tarjetas de color Celeste

- Ocho con el enunciado:

A) ¿Consume alcohol (cerveza, vino, sangría, etc) con frecuencia? (todos los fines de semana)

- Dos con el enunciado;

B) ¿Trabaja ud. actualmente?

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

Prevalencia de hurto

Diez tarjetas de color Naranja

- Ocho con el enunciado:

A) He llevado o he consumido sin pagar algún producto de algún supermercado

- Dos con el enunciado;

B) ¿Ud. eligió su carrera profesional por vocación?

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

Para aplicar el **Modelo de Respuesta Aleatorizada Estratificada en dos Etapas** se utilizó las siguientes tarjetas con los diferentes enunciados:

Prevalencia de vida de drogas

En la **Primera Etapa** se utilizó:

Diez tarjetas de color Blanco

- Ocho con el enunciado:

A) "Durante mi vida he consumido, pasta básica de cocaína (pasta, pay o pastel) por lo menos una vez"

- Dos con el enunciado:

B) "Ir al mecanismo aleatorizado en la Etapa 2"

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

En la **Segunda Etapa** se utilizó:

Diez tarjetas de color Blanco

- Ocho con el enunciado:

A) “Durante mi vida he consumido, pasta básica de cocaína (pasta, pay o pastel) por lo menos una vez”

- Dos con el enunciado:

B) “Durante mi vida jamás he consumido pasta básica de cocaína (pasta, pay o pastel)”

Uso actual de drogas

En la **Primera Etapa** se utilizó:

Diez tarjetas de color Palo Rosa

- Ocho con el enunciado:

C) “Durante mi vida he consumido, pasta básica de cocaína (pasta, pay o pastel) por lo menos una vez”

- Dos con el enunciado:

B) Ir al mecanismo aleatorizado en la Etapa 2

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

En la **Segunda Etapa** se utilizó:

Diez tarjetas de color Palo Rosa

- Ocho con el enunciado:

A) “Durante mi vida he consumido, pasta básica de cocaína (pasta, pay o pastel) por lo menos una vez”

- Dos con el enunciado:

B) “Durante mi vida **jamás** he mantenido relaciones sexuales con más de dos personas”

Prevalencia de sexo

En la **Primera Etapa** se utilizó:

Diez tarjetas de color Rosado

- Ocho con el enunciado:

¿Ha mantenido relaciones sexuales con más de dos personas a lo largo de su vida?

- Dos con el enunciado :

B) Ir al mecanismo aleatorizado en la Etapa 2

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

En la **Segunda Etapa** se utilizó:

Diez tarjetas de color Rosado

- Ocho con el enunciado:

¿Ha mantenido relaciones sexuales con más de dos personas a lo largo de su vida?

- Dos con el enunciado:

C) Durante mi vida **jamás** he mantenido relaciones sexuales con más de dos personas

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

Consumo actual alcohol

En la **Primera Etapa** se utilizó:

Diez tarjetas de color Celeste

- Ocho con el enunciado:

A) ¿Consume alcohol (cerveza, vino, sangría, etc) con frecuencia? (todos los fines de semana)

- Dos con el enunciado;

B) Ir al mecanismo aleatorizado en la Etapa 2

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

En la **Segunda Etapa** se utilizó:

Diez tarjetas de color Celeste

- Ocho con el enunciado:

A) ¿Consume alcohol (cerveza, vino, sangría, etc) con frecuencia? (todos los fines de semana)

- Dos con el enunciado;

B) Ir al mecanismo aleatorizado en la Etapa 2

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

Prevalencia de hurto

En la **Primera Etapa** se utilizó:

Diez tarjetas de color Naranja

- Ocho con el enunciado:

A) He llevado o he consumido sin pagar algún producto de algún supermercado

- Dos con el enunciado;

B) Ir al mecanismo aleatorizado en la Etapa 2

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

En la **Segunda Etapa** se utilizó:

Diez tarjetas de color Naranja

- Ocho con el enunciado:

A) He llevado o he consumido sin pagar algún producto de algún supermercado

- Dos con el enunciado;

B) Jamás he llevado o he consumido sin pagar algún producto de algún supermercado

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

4. Resultados

Utilizando las formulas de los estimadores de la proporción de respondientes en la población con la pregunta sensible del Modelo de Respuesta Aleatorizada Estratificada de Warner, Pregunta No relacionada y el Modelo de Respuesta Aleatorizada Estratificada en dos Etapas y las varianzas de los estimadores, presentadas anteriormente, se obtuvieron los siguientes resultados:

Cuadro 3: Modelos de Respuesta Aleatorizada Estratificada según la estimación de Prevalencia de Drogas y Varianza del Estimador

MRA	Prevalencia de Drogas	Varianza del Estimador
Warner	0,22 %	0,0005
Pregunta no relacionada	19,20 %	0,0007
Dos etapas	1,26 %	0,00003

Cuadro 4: Modelos de Respuesta Aleatorizada Estratificada según la estimación de Uso Actual de Drogas y Varianza del Estimador

MRA	Uso Actual de Drogas	Varianza del Estimador
Warner	0,0 %	-
Pregunta no relacionada	9,81 %	0,0006
Dos etapas	0,12 %	0,00002

Cuadro 5: Modelos de Respuesta Aleatorizada Estratificada según la estimación de Prevalencia de Sexo y Varianza del Estimador

MRA	Prevalencia de Sexo	Varianza del Estimador
Warner	0,22 %	0,00037
Pregunta no relacionada	18,9 %	0,0006
Dos etapas	11,56 %	0,00012

Cuadro 6: Modelos de Respuesta Aleatorizada Estratificada según la estimación del Porcentaje de personas que consumen alcohol y Varianza del Estimador

MRA	Consumo de Alcohol	Varianza del Estimador
Warner	0 %	-
Pregunta no relacionada	15,27 %	0,00063
Dos etapas	10,09 %	0,00010

Cuadro 7: Modelos de Respuesta Aleatorizada Estratificada según la estimación de Prevalencia de hurto y Varianza del Estimador

MRA	Prevalencia de Hurto	Varianza del Estimador
Warner	8,79 %	0,00053
Pregunta no relacionada	15,32 %	0,00072
Dos etapas	10,44 %	0,00010

5. Discusión

La Técnica de Respuesta Aleatorizada Estratificada nos permitió poner en práctica una serie de técnicas estadísticas de muestreo, que resultaron eficientes en la selección y aplicación de la encuesta.

Utilizando el Modelo de Respuesta Aleatorizada Estratificada de la pregunta no relacionada, el 19,20 % de los entrevistados han consumido en su vida pasta básica de cocaína por lo menos una vez un su vida, este porcentaje es menor cuando se utiliza el Modelo de Respuesta Aleatorizada Estratificada de las dos etapas (1,26 %), con el Modelo de Warner solo el 0,22 % de las personas entrevistadas han consumido en su vida pasta básica de cocaína por los menos una vez en su vida.

Con el Modelo de Respuesta Aleatorizada Estratificada de la pregunta no relacionada, el 9,81 % de los entrevistados son consumidores actuales de la pasta básica de cocaína, este porcentaje es menor cuando se utiliza el Modelo de Respuesta Aleatorizada Estratificada de las dos etapas (0,12 %), con el Modelo de Warner solo el 0,12 % de las personas entrevistadas consumen actualmente pasta básica de cocaína.

Utilizando el Modelo de Respuesta Aleatorizada Estratificada de la pregunta no relacionada, el 18,9 % de los entrevistados ha mantenido relaciones sexuales con más de dos personas a lo largo de su vida, este porcentaje es menor cuando se utiliza el Modelo de Respuesta Aleatorizada Estratificada de las dos etapas (11,56 %) , con el Modelo de Warner solo el 0,22 % de las personas entrevistadas ha mantenido relaciones sexuales con más de dos personas a lo largo de su vida.

Usando el Modelo de Respuesta Aleatorizada Estratificada de la pregunta no relacionada, el 15,27 % de los entrevistados consume actualmente alcohol todos los fines de semana, este porcentaje es menor cuando se utiliza el Modelo de Respuesta Aleatorizada Estratificada de las dos etapas (10,09 %), con el Modelo de Warner ninguna persona entrevistada consume actualmente alcohol todos los fines de semana.

Utilizando el Modelo de Respuesta Aleatorizada Estratificada de la pregunta no relacionada, el 15,32 % de los entrevistados ha llevado o ha consumido sin pagar algún producto de algún Supermercado, este porcentaje es menor cuando se utiliza el Modelo de Respuesta Aleatorizada Estratificada de las dos etapas (10,44 %), con el Modelo de Warner el 8,79 % de las personas entrevistadas ha llevado o ha consumido sin pagar algún producto de algún Supermercado

De los resultados se concluye que el Modelo de Respuesta Aleatorizada de la Pregunta No relacionada fue la mejor ya que reportó tasas mayores de respuestas en todos los casos considerados, es por eso que se recomienda seguir experimentando estos modelos, y sobre todo el Modelo de la Pregunta No Relacionada Estratificada en muestras más grandes y en temas en donde la cuestión sea efectivamente altamente sensible o muy comprometedor.

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] Cambell, C. and Joiner, B. (1973). How to get the answer without being sure you've asked the question. *The American Statistician*, diciembre : 229-231.
- [2] Chaudhuri, A; Mukerjee, R. (1983) *Randomized response: theory and techniques*. Marcel, New York.
- [3] Cochran, G.W. (1977). *Técnicas de muestreo*. Continental, México.
- [4] Des, R. (1980). *Teoría del muestreo*. Fondo de Cultura Económica, México.
- [5] Fox, J.A; Tracy, P.E.(1986). *Randomized Response: A Method for sensitive Surveys*. Sage University Paper on Quantitative Applications in the Social Sciencies, 58. Beverly Hills: Sage Publications.
- [6] Greenberg, B. G. et al. (1969). The unrelated question randomized response model: theoretical frame work. *J. Am. Stat. Assoc.*, Alexandria Vol. 64, 520-539.
- [7] Greenberg, B. G. et al. (1971). Application on the randomized response technique in obtaining quantitative data. *J. Am. Stat. Assoc.*, Alexandria, Vol. 71, N° 353: 243-250.
- [8] Kim, J.M. and Elam, M.E. (2003). Comparison and analysis of stratified randomized response models. *Joint Statistical Meetings. Section on Surveys Research Methods. J. Am. Stat. Assoc.* 2158-2161.
- [9] Kim J.M. and Elam, M.E. (2003a) Estratified unrelated question randomized response model. *Kournal of Statistical Planning and Inference*, in review.
- [10] Kim, J.M. and Elam, M.E.(2003b) A two-stage stratified Warner's randomized response model using optimal allocation. *Metrika*, in review.
- [11] KIM, J.M. and Elam, M.E.(2003). A stratified Warner's randomized responde model. *Journal of Statistical Planning and Inference*, in press.
- [12] Lanke, S. (1975). On the choice of the unrelated question in Simmons version of randomized response model. *J. Stat. Assoc.* Vol.70, N°1: 80-83.
- [13] Mann, C.R. (1999). Tell me your secret - random response methodology. *Advertising y marketing research professionals*, jun.
- [14] Moors, J.J. A. (1971). Optimization of the unrelated question randomized response model. *J. Stat. Assoc.* Vol. 66, N° 361 : 627-629.
- [15] Musch, J.; Broder, A. and Klauer, K.C. (2001). Improving survey research on the world-wide web using the Randomized Response technique dimensions of internet science.

- [16] Shimizu, I.M.; Bonham, G. S. (1971). Randomized response technique in a national survey. J. Am. Stat. Assoc. Vol. 66, N° 361 : 627-629.
- [17] Solano, O. (2004). Modelos de respuesta aleatorizada para variáveis quantitativas: Modelo de Greenberg. Vol. 22, N°3: 47-56.
- [18] Warner, S.L. (1965). Randomized response: A survey technique for elimination evasive answer bias. J. Stat. Assoc. Vol. 60 : 63-69.