

UNIVERSIDAD NACIONAL DEL CALLAO
FACULTAD DE CIENCIAS ECONÓMICAS
INSTITUTO DE INVESTIGACIÓN INIFE

**“EL MODELO DE RESPUESTA ALEATORIZADA Y SU
EFICIENCIA EN PREGUNTAS SENSIBLES”**

RICARDO LUIS POMALAYA VERASTEGUI

(Del 01/04/2010 al 31/3/2012,

Resolución Rectoral N° 456-10-R)

2012

ÍNDICE

	Pág.
1. RESUMEN	5
2. INTRODUCCIÓN	6
3. MARCO TEÓRICO	9
3.1 Antecedentes del problema	9
3.2 Descripción de la técnica desarrollada por Stanley Warner	9
3.3 Descripción de la técnica desarrollada por Bernad G. Greenberg	11
3.3.1 Medición de variables cualitativas	12
3.3.2 Medición de variables numéricas	13
3.3.3 Opciones para p_1 y p_2	16
3.3.4 Selección de la característica numérica no sensible	17
3.3.5 Asignación para n_1 y n_2	19
3.3.6 Caso con μ_Y y σ_Y conocidas desde el inicio	20
3.4 Errores no muestrales	21
3.5 Inferencia Estadística de parámetros	23
3.6 Tablas de números aleatorios y programas informáticos	27
3.7 Estratificación de la población	29
3.7.1 Razones para el uso de Muestreo Estratificado	31
3.7.2 Afijación de la Muestra	32
3.7.3 Afijación Uniforme	33
3.7.4 Afijación Proporcional	33
3.7.5 Afijación de Mínima Varianza (o Afijación de Neyman)	34
3.7.6 Afijación Óptima	35

3.7.7	Comparación de Eficiencias según los distintos tipos de Afijación.	35
4.	MATERIALES Y MÉTODOS	37
4.1	Materiales	37
4.2	Métodos	38
4.3	Diseño de la Encuesta	39
4.3.1	Formato de la encuesta directa	40
4.3.2	Formato de encuesta con la técnica de Respuesta aleatorizada	40
4.3.3	Diseño de muestra para encuesta directa	42
4.3.4	Descripción de la técnica de encuesta directa.	44
4.3.5	Descripción de la técnica de la Respuesta Aleatorizada	44
5.	RESULTADOS	47
5.1	Estimación de la Proporción en la Encuesta Directa	47
5.2	Estimación por medio de la Técnica de Respuesta Aleatorizada	55
5.3	Comparación y Contraste de ambas Técnicas	63
5.3.1	Prueba de hipótesis para la Comparación de Proporciones en “Consumo de alimentos y bebidas en el Centro de Cómputo”	64
5.3.2	Prueba de hipótesis para la Comparación de Proporciones en “Consumo de drogas alucinógenas”	65
5.3.3	Prueba de hipótesis para la Comparación de Proporciones en “Copia en los exámenes”	67

5.3.4	Prueba de hipótesis para la Comparación de Proporciones en “Consumo de bebidas alcohólicas cerca a la UNAC”	68
5.3.5	Prueba de hipótesis para la Comparación de Proporciones en “Inasistencia a clases sin motivo justificable”	69
5.4	Respuestas Analizadas por Separado	71
5.5	Ventajas y Desventajas de cada Método de Investigación	78
6.	DISCUSIÓN	80
7.	REFERENCIALES	84
8.	APÉNDICE	86

Anexos

1. RESUMEN

Uno de los objetivos de la estadística a lo largo del tiempo, ha sido la estimación de parámetros por medio de encuestas, para conocer diversos aspectos de la población. Sin embargo, los problemas a los que se enfrentan muchos de estos estudios, es la falta de cooperación y veracidad en algunas de las respuestas, sobre todo en preguntas que se refieren a aspectos de temas sensibles o estigma social, que causan cierta resistencia por parte del entrevistado para contestarlas.

El objetivo del presente trabajo de investigación apunta a establecer los aportes de los estudios realizados para mejorar la calidad y veracidad de las respuestas obtenidas a través de encuestas, entre los cuales se encuentra el modelo de Repuesta aleatorizada, que pretende darle mayor confianza al entrevistado, para tener una mejor cooperación y un mayor grado de veracidad en sus respuestas.

En este trabajo, los datos corresponden a variables categóricas respecto a temas sensibles dirigida a la población estudiantil de la facultad de Ciencias económicas de la Universidad Nacional del Callao, semestre 2011-B y se han recopilado haciendo uso de técnicas de muestreo probabilístico.

Entre los hallazgos resultantes cabe precisar que realizada la aplicación a una muestra de estudiantes de la UNAC, comparándolo contra el método comúnmente utilizado de Encuesta directa pudo verificarse su eficacia en la reducción del error estándar de los estimadores, logrando disminuir de gran manera los problemas anteriormente mencionados.

2. INTRODUCCIÓN

La medición de parámetros por medio de encuestas, ha sido de gran utilidad a lo largo del tiempo, para conocer diversos aspectos de la población. Para que los resultados de una encuesta sean creíbles es necesario, entre otros aspectos, que las preguntas sean bien elaboradas y asumir que las respuestas sean ciertas. Para creer en los resultados de una encuesta es necesario creer también en las respuestas de las personas que han sido entrevistadas. Sin embargo, tener respuestas verídicas es difícil. Hay muchos problemas implícitos al tratar de conseguir las y de que éstas sean sinceras (Fox, 1986)

El problema que presenta la medición de parámetros por medio de encuestas es la falta de cooperación y veracidad en algunas de las respuestas, sobre todo en preguntas que se refieren a temas sensibles, que causan reparos al entrevistado para contestarlas.

Para resolver este problema, se han propuesto diversos métodos, entre los cuales se encuentra el de Respuesta Aleatorizada, que tiene como objetivo darle mayor confianza al entrevistado, para tener una mejor cooperación y un mayor grado de veracidad en sus respuestas. Este método es de gran utilidad para estudiar aspectos delicados como alcoholismo, drogadicción, actividad sexual, aborto, entre otros.

El objetivo general de esta investigación es establecer los aportes de la teoría de las técnicas de Respuesta aleatorizada para estudiar características sensibles, hacer una aplicación, así como para realizar en su caso, recomendaciones sobre la obtención de datos y calidad de la encuesta.

En el presente trabajo de investigación se realiza una aplicación de los modelos de Respuesta aleatorizada y el método convencional de entrevista directa para estudiar el comportamiento de los jóvenes de la facultad de Ciencias económicas de la Universidad Nacional del Callao, matriculados en la Universidad el semestre 2011-B, con respecto a algunos temas sensibles, a efectos de medir el grado de incidencia de los estudiantes en estos problemas, y por lo tanto la frecuencia con la que rompen ciertas normas comportamiento.

Los objetivos específicos son determinar la bondad de la técnica de Respuesta Aleatorizada. para preguntas que tiene como finalidad medir variables sensibles categóricas, diseñar una muestra probabilística representativa de la población estudiantil de la facultad de Ciencias económicas de la UNAC, y realizar una comparación de las respuestas obtenidas con preguntas directas y por el método de Respuesta Aleatorizada, ya que se cree que con el primer método, hay una tendencia a subestimarse en las respuestas de temas sensibles con respecto a las del segundo método, en donde el encuestado tiene un mayor grado de confiabilidad

Por otro lado, esta investigación es importante y se justifica por el papel esencial que tiene la estadística en el campo de la investigación, y es por ello que resulta de gran importancia contar con datos estimados con la mayor precisión posible. Para obtener dichos datos, se recurren a diversos métodos, entre los cuales se encuentran las encuestas. Sin embargo, por diversas razones, en éstas, las personas pueden elegir confiarle o no al entrevistador las respuestas correctas a ciertas preguntas.

En tales casos tienen la alternativa de no responder o responder incorrectamente a las preguntas. Esto representa un sesgo realmente difícil de evaluar. Es por ello que existen métodos que lo reducen considerablemente mediante un dispositivo para aleatorizar la respuesta del encuestado logrando mantener su privacidad. Este es el caso del método de Respuesta Aleatorizada.

Las preguntas que se realizarán en las encuestas estarán dirigidas a temas sensibles como:

- Consumo de alimentos en el Centro de cómputo de la facultad.
- Consumo de drogas.
- Fraude en los exámenes.
- Consumo de alcohol en torno al campus de la UNAC.
- Ausencia injustificada a clases.

3. MARCO TEÓRICO

Aquí se explican las bases teóricas que respaldan las investigaciones desarrolladas en la técnica de Respuesta aleatorizada, tanto para variables categóricas como para las numéricas y cuáles son los principales errores no muestrales.

3.1 Antecedentes del problema

El modelo que aplica la técnica de Respuestas Aleatorizadas, fue desarrollado en 1965 por Stanley Warner, el cual se enfocó al estudio de variables categóricas. Debido al éxito de sus investigaciones, en fechas posteriores, otros investigadores estudiaron y ampliaron su modelo, obteniendo resultados favorables para mejorar lo que ya estaba establecido. Entre las mejoras se encuentra la considerable reducción de la varianza de los estimadores.

Para cumplir los objetivos de esta investigación utilizaremos las siguientes técnicas:

1. Técnica desarrollada por Stanley Warner.
2. Técnica desarrollada por Bernad Greenberg, para datos categóricos.

3.2 Descripción de la técnica desarrollada por Stanley Warner

Una situación que probablemente nos lleve a obtener negativas para contestar o respuestas evasivas ocurre cuando una pregunta en una encuesta es sensible o sumamente personal.

Considere primero la estimación de una proporción Binomial (la proporción π_A de respuestas de personas que pertenecen a la clase A o que han cometido cierto acto).

Utilizando un ingenioso dispositivo aleatorio, Warner en 1965 mostró que es posible

estimar esta proporción sin que el encuestado revele su postura personal respecto a la pregunta. El objetivo es alentar a las personas para que den respuestas veraces conservando completamente la confidencialidad de sus respuestas.

El dispositivo aleatorio, como por ejemplo una caja con bolas blancas y negras, representa la selección de una de dos frases o preguntas, cada una con respuesta “sí” o “no”, que se le presentan al encuestado. El entrevistador no conoce la pregunta que los encuestados respondieron, pero si conoce las probabilidades relativas P y $(1-P)$ con las que se presentan las dos preguntas. El éxito del método depende totalmente en que el encuestado esté completamente convencido de que su participación no va a revelar su postura respecto a la característica sensitiva en cuestión.

En la propuesta original de Warner las dos frases son:

“Soy un miembro de la clase A.” (con probabilidad P)

“No soy miembro de la clase A.” (con probabilidad $(1-P)$)

Con una muestra aleatoria de “ n ” encuestados el entrevistador obtiene un estimado

Binomial

$\phi = \frac{m}{n}$ de la proporción ϕ de las respuestas “sí”. Si las preguntas son contestadas

verazmente, la relación entre ϕ y π_A en la población es:

$$\phi = P\pi_A + (1 - P)(1 - \pi_A) = (2P - 1)\pi_A + (1 - P)$$

Si conocemos el valor de P el estimador de la relación anterior es: $\pi_{AW} = \frac{[\phi - (1 - P)]}{(2P - 1)}$

con $P \neq \frac{1}{2}$, y donde AW significa que pertenece a la clase “A” del método de Warner.

Este es el estimador de máxima verosimilitud de π_A y por lo tanto es insesgado. Si

expresamos el complemento de ϕ de la siguiente forma: $(1 - \phi) = (2P - 1)(1 - \pi_A) + (1 - P)$

podemos obtener de manera simple:

$$V(\pi_{AW}) = \frac{\pi_A(1 - \pi_A)}{n} + \frac{P(1 - P)}{n(2P - 1)^2}$$

Este método en general es impreciso porque el entrevistador no sabe si la respuesta “sí”, significa que pertenece a la clase A o no. Sin embargo Warner demostró que con su método se obtiene un Error Cuadrático Medio (ECM) menor que el que se produce con la pregunta sensitiva directa.

3.3 Descripción de la técnica desarrollada por Bernad G. Greenberg

En ésta técnica se ampliaron y perfeccionaron las investigaciones realizadas por Warner acerca de aleatorizar las respuestas de los encuestados. Desarrollaron el método de “La segunda pregunta no relacionada”, para la medición de variables cualitativas y cuantitativas.

3.3.1 Medición de variables cualitativas.

Es un método alternativo al de Warner en el que se espera aumentar la cooperación del encuestado, dado que la segunda pregunta no es sensitiva en ningún sentido, al no estar relacionada con la primera. Un ejemplo simple sería afirmar o negar “Nací en el mes de abril”, comparada con la pregunta sensitiva “Pertenezco a la clase A”, siendo la clase “A” el grupo con la característica sensitiva que se está estudiando.

Si todos los encuestados responden la verdad, la proporción de las respuestas afirmativas se convierte en $\phi = P\pi_A + (1-P)(\pi_{NR})$, donde NR significa No Relacionada y π_{NR} es la proporción de la población encuestada que nacieron en el mes de abril. Si conocemos el

valor de π_{NR} , el valor estimado de π_{NR} es $\pi_{ANR} = \frac{\phi - (1-P)\pi_{NR}}{P}$ y la varianza es

$$V(\pi_{ANR}) = \frac{\phi(1-\phi)}{nP^2}.$$

El caso cuando la π_{NR} es conocida siempre se puede lograr. Un ejemplo simple es cuando se tiene una caja con bolas rojas, blancas y azules, con proporciones P_1, P_2, P_3 , conocidas.

Si se saca una bola roja, se contesta a la pregunta sensitiva y si se saca una bola blanca o

azul, se responde a la pregunta no sensitiva; por lo tanto $\pi_{ANR} = \frac{P_2}{P_2 + P_3}$.

$V(\pi_{ANR}) < V(\pi_{AW})$ para toda π_A y π_{NR} dado que P es mayor que $\frac{1}{3}$.

La varianza de π_{AW} es simétrica cerca de $P = \frac{1}{2}$ pero la varianza de π_{ANR} no lo es, esto es a

causa de una P pequeña, por obtener pocas respuestas de la pregunta sensible con este método.

3.3.2 Medición de variables numéricas.

En el caso de variables numéricas cuando se utilizan dos preguntas, la distribución conjunta de los resultados está comprometida a respuestas numéricas, las cuales deben de estar relacionadas, no en el tema, sino en la respuesta. Esto quiere decir que si hablamos por ejemplo de número de abortos durante la vida de una mujer, la respuesta de la pregunta no relacionada, debe arrojar un número semejante, por ejemplo el número promedio de hijos que debe tener una mujer que trabaja a tiempo completo.

Esta distribución es entonces una mezcla de dos distribuciones “puras”, la cual debe de estar separada estadísticamente para obtener estimadores de los parámetros de investigación que sean significativos. La media poblacional de las distribuciones sensibles y no sensibles se define como μ_A y μ_Y respectivamente con varianzas σ^2_A y σ^2_Y .

Si se supone que las 2 muestras son independientes de tamaño n_1 y n_2 , y que no tienen empalmes o repeticiones, definimos:

- p_i = probabilidad de ser seleccionada la pregunta sensitiva por el encuestado en la muestra i ($i = 1, 2$), $p_1 \neq p_2$,
- $1 - p_i$ = probabilidad de ser seleccionada la pregunta no sensitiva por el encuestado en la muestra i ($i = 1, 2$)
- Z_{ij} = respuesta de la j – ésima persona en la muestra i ($i = 1, 2$), $j = (1, 2, \dots, n_i)$

- $f(z)$ = función de probabilidad (o función de densidad de probabilidad) asociada con la pregunta sensitiva, con $E_f [Z] = \mu_A$,
- $g(z)$ = función de densidad de probabilidad asociada con la pregunta no sensitiva (similar a $f(z)$ en el rango de respuesta), con $E_g [Z] = \mu_Y$,
- μ_A = estimador muestral de la media de la distribución sensitiva,
- μ_Y = estimador muestral de la media de la distribución no sensitiva.

La función de probabilidad para cada elemento en una muestra es:

$$\text{Muestra 1: } \psi_1(Z_1) = p_1 f(z_1) + (1 - p_1) g(z_1) \quad (2.1a)$$

$$\text{Muestra 2: } \psi_2(Z_2) = p_2 f(z_2) + (1 - p_2) g(z_2) \quad (2.1b)$$

Entonces tenemos:

$$\mu_{Z_1} = E[Z_1] = p_1 \mu_A + (1 - p_1) \mu_Y \quad (2.2a)$$

$$\mu_{Z_2} = E[Z_2] = p_2 \mu_A + (1 - p_2) \mu_Y \quad (2.2b)$$

Al despejar μ_A y μ_Y se tiene:

$$\mu_A = \frac{(1 - p_2) \mu_{Z_1} - (1 - p_1) \mu_{Z_2}}{p_1 - p_2}$$

$$\mu_Y = \frac{p_2 \mu_{Z_1} - p_1 \mu_{Z_2}}{p_2 - p_1}$$

Si sustituimos a μ_{Z_1} y μ_{Z_2} por las medias de las respuestas de las 2 muestras, \bar{Z}_1 y

\bar{Z}_2 respectivamente, obtenemos estimadores insesgados para μ_A y μ_Y :

$$\mu_A = \frac{(1-p_2)\bar{Z}_1 - (1-p_1)\bar{Z}_2}{p_1 - p_2} \quad (2.3a)$$

$$\mu_Y = \frac{p_2\bar{Z}_1 - p_1\bar{Z}_2}{p_2 - p_1} \quad (2.3b)$$

Que tienen varianzas:

$$V(\mu_A) = \frac{1}{(p_1 - p_2)^2} \left[(1-p_2)^2 V(\bar{Z}_1) + (1-p_1)^2 V(\bar{Z}_2) \right] \quad (2.4a)$$

$$V(\mu_Y) = \frac{1}{(p_2 - p_1)^2} \left[p_2^2 V(\bar{Z}_1) + p_1^2 V(\bar{Z}_2) \right] \quad (2.4b)$$

Donde:

$$\begin{aligned} V(\bar{Z}_i) &= \frac{1}{n_i} \left[E(Z_i^2) - \{E(Z_i)\}^2 \right] = \left[\{p_i E_f(Z_i^2) + (1-p_i) E_f(Z_i^2)\} - \{p_i \mu_A + (1-p_i) \mu_Y\}^2 \right] \\ &= \frac{1}{n_i} \left[p_i \sigma_A^2 + (1-p_i) \sigma_Y^2 + p_i \mu_A^2 + (1-p_i) \mu_Y^2 - \{p_i \mu_A + (1-p_i) \mu_Y\}^2 \right] \end{aligned}$$

Por lo que:

$$V(\bar{Z}_i) = \frac{1}{n_i} \left[\sigma_Y^2 + p_i (\sigma_A^2 - \sigma_Y^2) + p_i (1-p_i) (\mu_A - \mu_Y)^2 \right] \quad i=1,2 \quad (2.5)$$

Los estimadores en la fórmula (2.3a y 2.3b) tienen la gran ventaja de ser calculados simplemente de los datos de la muestra, son insesgados y utilizan medias muestrales sin importar la naturaleza específica de $f(z)$ y $g(z)$. Las varianzas de los estimadores pueden ser estimadas convenientemente mediante la varianza de la muestra S_i^2 en (2.4a y 2.4b):

$$V(\bar{Z}_1) = \frac{S_1^2}{n_1}, \quad V(\bar{Z}_2) = \frac{S_2^2}{n_2}.$$

El diseño óptimo para una encuesta de “Respuestas Aleatorizadas” usando preguntas cuantitativas requiere de una elección apropiada para p_1 y p_2 , la selección inteligente de la pregunta no sensitiva Y , y una asignación eficiente del total de la muestra en n_1 y n_2 .

3.3.3 Opciones para p_1 y p_2

El criterio para una buena selección de las probabilidades asociadas con la distribución de bolas en el dispositivo aleatorio utilizado en la recolección de datos cuantitativos es similar a lo considerado por Greenberg en el estudio de preguntas para una respuesta dicotómica.

Si examinamos los estimadores para μ_A y μ_Y en las ecuaciones (2.3a y 2.3b) vemos que pueden obtener valores muy absurdos si los denominadores se acercan a “cero”. Una regla para evitar esta situación es escoger $p_1 + p_2 = 1$ después de haber escogido p_1 tan lejos como sea posible de 0.5 sin que se cree desconfianza en el encuestado sobre el peso del dispositivo aleatorio a favor de una pregunta en particular. La práctica indica que se

pueden obtener resultados satisfactorios con p_1 entre los valores 0.70 ó 0.80, o sus 16 complementos.

Se puede llegar a la misma conclusión para la elección de p_2 si lo que se quiere es minimizar el valor de la varianza $V(\mu_A)$ en la ecuación (2.4a y 2.4b) después de haber elegido p_1 . La expresión para la derivada de $V(\mu_A)$ con respecto a p_2 es muy parecida al caso binomial. La gráfica de la varianza dibujada contra p_2 es hiperbólica con asíntota en p_1 ; para p_1 dada, la varianza crece sin límite mientras p_2 se acerca a p_1 y decrece cuando el valor $|P_2 - P_1|$ se incrementa. Es por esto, que una vez seleccionada p_1 tan cercana a “cero” o como sea más práctico, se escoge $p_2 = 1 - p_1$. Al escoger $p_2 = 1 - p_1$, el valor de la varianza no se minimiza. En primer lugar, si $P_1 > 0.5$ el mínimo matemático ocurre cuando $p_2 = 0$, implicando que la segunda muestra nada más se va a usar para estimar μ_Y . Si dicho procedimiento no produce resultados incompatibles para estimar μ_A , es obvio que es la mejor opción.

3.3.4 Selección de la característica numérica no sensible

Una regla fundamental es que la pregunta no sensible o inocua debe de ser planteada de tal forma que la cantidad de la respuesta sea igual que la de la sensible, por ejemplo: pesos, centímetros o número de veces en que ocurre un evento.

Desde el punto de vista intuitivo se puede pensar que las funciones de densidad de las respuestas de ambas preguntas no tienen traslapes y que el cálculo de sus parámetros son

independientes. En primer lugar, si la pregunta sensitiva es contestada con números en promedio por ejemplo cerca de 50 centavos y la no sensitiva tiene respuestas con promedio

cerca de 500 pesos, se podría pensar que es ideal, porque una respuesta individual no requeriría fórmulas para estimarse. Esto es engañarse a uno mismo, ya que el encuestado no va a cooperar puesto que la respuesta puede ser clasificada automáticamente o identificada con mucha facilidad. Además, la clasificación no está hecha en base a la respuesta dada individualmente sino en grupos usando procedimientos para estimadores como en (2.3a y 2.3b). Finalmente como se ve en (2.4a y 2.4b) y (2.5), para cualquier valor $(\sigma_A, \sigma_Y, p_1, p_2, n_1, n_2)$ dados, las varianzas de los estimadores aumentan cuando $|\mu_A - \mu_Y|$ aumenta.

Una vez que se han elegido p_1 y p_2 de acuerdo al criterio anteriormente descrito, los otros parámetros manipulables son n_1, n_2, μ_Y y σ_Y^2 . No existe opción para μ_A y σ_Y^2 ya que dependen de la naturaleza de la característica sensitiva que no se conoce. Para cualquier valor de (n_1, n_2) las varianzas de los estimadores decrecen cuando decrece σ_Y^2 y $|\mu_A - \mu_Y|$.

Por eso la importancia de elegir una pregunta no sensitiva no es cuanto difiere de las respuestas de la sensitiva en significado, si no en qué tan uniforme o parecidas son las respuestas. Obviamente una opción inteligente sería elegir una pregunta no sensitiva de tal forma que μ_Y se acerque a μ_A y tenga una varianza σ_Y^2 mínima. Sin embargo si σ_Y^2 es

considerablemente menor que σ_A^2 , pueden haber pérdidas en la cooperación por parte de los encuestados. Cualquiera de las respuestas al final de la distribución de A serían

respuestas de A y esto es evidente para los encuestados más perspicaces, que posiblemente darán una respuesta evasiva o falsa en lugar de dar una respuesta verdadera que pudiera decir que pregunta están contestando. Por esta razón se recomienda que σ_Y^2 sea al menos tan grande como σ_A^2 y que todo caiga en la manipulación de $1/2 n$ y n para reducir la varianza.

3.3.5 Asignación para n_1 y n_2

La subdivisión óptima del total de la muestra en 2 grupos se puede basar en el principio de minimizar $V(\mu_A)$. Esto se logra tomando:

$$\frac{n_1}{n_2} = \frac{\sqrt{(1-p_2)^2 V(Z_1)}}{\sqrt{(1-p_1)^2 V(Z_2)}} = \frac{\sqrt{(1-p_2)^2 [1+p_1(\phi_1^2-1)+p_1(1-p_1)\phi_2^2]}}{\sqrt{(1-p_1)^2 [1+p_2(\phi_1^2-1)+p_2(1-p_2)\phi_2^2]}} \quad (2.6)$$

Donde $\phi_1 = \frac{\sigma_A}{\sigma_Y}$, $\phi_2 = \frac{(\mu_A - \mu_Y)}{\sigma_Y}$ y las primeras opciones de sus valores se usan para

calcular $\frac{n_1}{n_2}$, Dependemos de un punto de referencia que muchas veces nos da una

aproximación aceptable para (2.6) como:

$$\frac{n_1}{n_2} \approx \frac{p_1}{p_2} \quad \text{con} \quad p_1 + p_2 = 1 \quad (2.6)$$

Esto se debe a que cuando se eligen p_1 y p_2 que satisfagan $p_1 + p_2 = 1$ como se sugirió anteriormente, entonces podemos escribir (2.6) como:

$$\frac{n_1^2}{n_2^2} = \frac{p_1^2}{p_2^2} \left[\frac{\sigma_Y^2 + p_1 p_2 (\mu_A - \mu_Y)^2 + p_1 (\sigma_A^2 \sigma_Y^2)}{\sigma_Y^2 + p_1 p_2 (\mu_A - \mu_Y)^2 + p_2 (\sigma_A^2 \sigma_Y^2)} \right]$$

con lo cual se muestra que (2.6') es exacta cuando $\sigma_A^2 = \sigma_Y^2$, y una aproximación cercana cuando σ_A^2 y σ_Y^2 son diferentes pero razonablemente cercanas entre sí, como deberían de ser para una buena elección de la pregunta no sensitiva Y. Ya que la determinación de $\frac{n_1}{n_2}$ requiere el uso de valores intuitivos de la expresión (2.6) para los parámetros poblacionales ϕ_1 y ϕ_2 , es razonable suponer que en general (2.6') puede dar una asignación muy cercana a la óptima de (2.6).

3.3.6 Caso con μ_Y y σ_Y conocidas desde el inicio.

Como ya sabemos de las respuestas binomiales, la encuesta puede ser diseñada más eficientemente cuando el valor de μ_Y y σ_Y para una pregunta neutra se conoce de antemano. El aplicar este principio a la pregunta cuantitativa, nos dice que la segunda pregunta debe ser el número de personas que viven en una casa donde se conoce el promedio del ingreso de algún tipo de censo u otra enumeración. Cuando μ_Y y σ_Y son conocidas de antemano, no hay necesidad de 2 muestras. El estimador de μ_A y su

varianza, cuando μ_Y y σ_Y son conocidas se calculan mediante las fórmulas:

$$\mu_A \Big| \mu_Y = \frac{[\bar{Z} - (1 - \mu_Y)]}{p}$$

$$V(\mu_A \Big| \mu_Y) = \frac{V(\bar{Z})}{p^2} = \frac{V(Z)}{np^2}$$

Como en el caso binomial, hay una reducción sustancial en la varianza del parámetro μ_A cuando μ_Y y σ_Y se conocen de antemano.

3.4 Errores no muestrales

Los errores no muestrales son sesgos inevitables de las mediciones, pues no son problemas de muestreo en un sentido estricto. Los sesgos no muestrales plantean problemas profundos de la medición científica, afectan tanto el valor de la población como el valor muestral, y junto con cualquier sesgo estadístico constante, son la causa de la diferencia entre el valor de la población y el valor verdadero. Los sesgos no muestrales se dividen como se presenta en la siguiente figura:

Sesgos no muestrales
1. No observación (No cobertura, No respuesta)
2. Observación (Recolección de datos en campo, Procesamiento en oficina))

Figura 2.1 Clasificación de las causas de sesgo en una encuesta

Fuente: Leslie Kish. "Muestreo de Encuestas", Editorial Trillas, México, 1982

Entre los sesgos no muestrales podemos distinguir los sesgos de no observación de los que

se deben a la observación. La primera clase surge a partir de ocasiones en que no se

pueden obtener observaciones en algunos segmentos de la población debido a la *no cobertura* o a la *no respuesta*. En cuanto a la clasificación de no cobertura, existen dos causas:

- “No alcance”. Se presenta cuando hay unidades seleccionadas que no se pueden entrevistar con los recursos que se tienen para realizar la encuesta, por ejemplo, en encuestas de hogares puede haber secciones de difícil acceso en la división de las ciudades.
- “Fuera de casa”. Pertenece al grupo de personas que están temporalmente ausentes de su residencia habitual, en encuestas de hogares son los que temporalmente se encuentran fuera de casa. El grupo puede variar si las familias tienen hijos pequeños o algún adulto que permanezca siempre en casa y que pueda dar las respuestas.

Para la no respuesta, existe una causa principal:

- “No puede responder”. Los elementos encuestados no tienen información al respecto o no quieren proporcionarla.

El segundo tipo de sesgos de no muestreo proviene de la obtención y registro de observaciones incorrectas. Podemos distinguir dos tipos de sesgos, uno aparece en la *“Recolección de datos en el campo”*, que puede consistir en errores en entrevistas, enumeraciones, conteos o medidas, y puede darse a causa de que el mecanismo de medición, puede ser influido o impreciso o bien las preguntas son de naturaleza sensitiva y no se hacen de manera que el entrevistado se sienta libre y tranquilo de contestar la

verdad. El otro sesgo referente a la observación, es el de *“Procesamiento en la oficina”*, 22 en el cual hay errores en la codificación, la tabulación y los cálculos.

Esta investigación está enfocada a atacar principalmente la no respuesta, por lo cual se tomaron en cuenta algunas recomendaciones:

1. *El mejoramiento de los procedimientos de recolección.* Los mejoramientos indicados para la reducción de rechazos son:
 - a) Garantía del anonimato del entrevistado
 - b) Motivación para la cooperación del entrevistado
 - c) Hacer una cita previa con el entrevistado
2. *Visitas repetidas.* Ayudan a reducir los casos en los cuales el entrevistado no se encuentra.
3. *Estimación del efecto de la no respuesta.* El reporte del tamaño y las posibles causas de no respuesta es una práctica común para obtener mejores encuestas.

3.5 Inferencia Estadística de parámetros

Los métodos de Inferencia Estadística propiamente dichos son:

- a. estimación puntual
- b. estimación por intervalo, y
- c. contrastación o verificación de hipótesis

La estimación por intervalo, consiste en construir intervalos aleatorios que con una probabilidad y precisión preestablecidas incluyan a los parámetros poblacionales desconocidos, partiendo de estimadores puntuales adecuados (en general estadísticos suficientes y/o centrados) y de sus distribuciones de probabilidad.

Denotando a dicha probabilidad por $(1 - \alpha)$, la estimación por intervalo en general 23 proporcionará dos estadísticos (o valores muestrales) $\bar{\theta}_1$ y $\bar{\theta}_2$, tales que

$$\Pr(\bar{\theta}_1 < \bar{\theta}_2) = 1 - \alpha$$

esto es captan o incluyen al verdadero parámetro desconocido de la población, con una probabilidad (ex-ante o antes de obtener la muestra) o fiabilidad (ex-post o después de obtener la muestra) igual a $(1 - \alpha)$.

El significado de lo anterior es lo siguiente: antes de obtener la muestra el intervalo es aleatorio y existe una probabilidad igual a $(1 - \alpha)$ de incluir al parámetro poblacional desconocido. Seleccionada la muestra, y obtenido el intervalo, ya no hay aleatoriedad alguna sino que se tratará de dos valores concretos que incluirán o no al parámetro poblacional desconocido, pero si se repitiese dicho proceso muchas veces, aproximadamente en el $(1 - \alpha)\%$ de ellas existe la confianza o fiabilidad de que el intervalo obtenido contendrá al verdadero valor poblacional (acierto).

Así como la teoría de la estimación tiene como objetivo obtener bien un valor aproximado de cada parámetro desconocido, bien un intervalo de valores que contenga a cada uno de ellos con una determinada probabilidad, la metodología estadística de la

contrastación de hipótesis consiste en formular hipótesis o conjeturas sobre los parámetros poblacionales desconocidos, y a partir de resultados muestrales aceptarlas o rechazarlas.

Está entroncada en el Análisis de Decisiones o Teoría de la Decisión, y en general se formula en términos de una hipótesis H_0 denominada nula, pues es lo que se espera que

resultará acorde con el experimento o prueba realizada, frente a otra H denominada alternativa. Ambas pueden ser simples o compuestas según cada una de ellas abarque o contenga un único punto o no, y tras la realización del proceso de inferencia siempre

existen dos posibilidades de acierto y otras dos de equivocación, como se desprende del siguiente.

Cuadro n" 1.

Realidad \ Decisión	H_0	H_1
H_0	Acierto	Equivocación
H_1	Equivocación	Acierto

esto es, que si la realidad (o estado de la naturaleza en Teoría de la Decisión) es H_0 , y a través de una muestra se decide por H_0 hay acierto, y si se decidiera H_1 , habría equivocación o error de tipo I. Por otra parte si la realidad fuese H_1 y dada la muestra nos decidimos por H_0 , cometeremos un error o equivocación tipo II, y si nos decidiésemos por H_1 , acertaríamos.

Las dos decisiones incorrectas se denominan también errores de o del muestreo, y siempre están implícitas cuando se decide por muestreo, se hayan cometido o no errores de codificación, de clasificación, de cálculo etc. etc., durante el proceso de inferencia. Estos últimos, que pueden cometerse tanto en un análisis exhaustivo como muestral, se denominan para poder ser diferenciarlos de los anteriores, errores ajenos al muestreo.

Consecuentemente sobre las conclusiones obtenidas por muestreo, no podremos tener una seguridad total o grado de certeza del 100%, puesto que jamás conoceremos con seguridad

el estado verdadero en que se encontrará realmente parámetro poblacional. Por lo tanto lo que se debe hacer es intentar cuantificar la incertidumbre implícita en cada una de las cuatro situaciones descritas, aunque dicha cuantificación puede ser a veces muy difícil. La

metodología estadística nos asegura poder realizar dicha cuantificación en términos probabilísticos.

Denotando a las probabilidades de equivocación por α ; y β , y a las probabilidades de acierto por $(1 - \alpha)$ y $(1 - \beta)$ respectivamente, el cuadro n° 1 anterior, puede completarse de la siguiente forma (cuadros n° 2 y 3):

Cuadro N° 2

Realidad \ Decisión	H ₀	H ₁
H ₀	Acierto Probabilidades/Fiabilidades: $1 - \alpha$	Equivocación Probabilidades/riesgos: α
H ₁	Equivocación Probabilidades/riesgos: β	Acierto Probabilidades/Fiabilidades: $1 - \beta$

Cuadro N° 3

Aciertos, equivocaciones, probabilidades, riesgos y fiabilidades

Realidad \ Decisión	H ₀	H ₁
H ₀	Acierto Probabilidades = $1 - \alpha$	Error Tipo II (E ₁₁) o α Probabilidad = α
H ₁	Error Tipo II (E ₁₁) o β Probabilidad: β	Acierto Probabilidades = $1 - \beta$

A la probabilidad α de cometer un error tipo I, o bien a su máximo o cota superior si la hipótesis nula fuese compuesta, se denomina nivel de significación de la prueba.. La probabilidad β de cometer un error tipo II (o el máximo o cota superior de dichos errores

en caso de hipótesis compuestas) se conoce como riesgo de fiabilidad no garantizada, y a su complemento respecto de la unidad potencia de la prueba.

3.6 Tablas de números aleatorios y programas informáticos

Las muestras estadísticas o aleatorias son aquellas sobre las que se puede definir una distribución de probabilidad, en términos intuitivos diríamos que son las generadas por mecanismos puros de azar.

En este punto se va a dilucidar como se extraen muestras aleatorias en la realidad, y la problemática que presenta la identificación de los documentos reales correspondientes.

1. Las tablas de números o dígitos aleatorios fueron el instrumento más útil para obtener elementos al azar hasta la incorporación generalizada de las computadoras en el trabajo habitual, cosa que ha ocurrido sobre todo a partir de los años 80, y suponían una mejora sustancial respecto de cualquier mecanismo como urnas, ruletas, loterías, etc. Para realizar extracciones al azar.

Las más completas eran las de la Rand Corporation que incluían 1000000 de dígitos aleatorios.

Su construcción más o menos puede hacerse de la siguiente forma:

- a. En una urna controlada electrónicamente existen 10 bolas iguales numeradas del 0 hasta el 9.
- b. Cada vez, después de remover exactamente igual y durante el mismo tiempo la urna, se extrae electrónicamente una bola, se anota el resultado y se devuelve a la urna.

c. Los resultados se van escribiendo uno a continuación de otro en una hoja, futura tabla de números aleatorios, y cada cuatro o cinco cifras puede dejarse un blanco para que

se puedan leer más cómodamente y también para poder presentar la tabla en forma de filas y columnas.

d. Antes de dar por válida la tabla, se realizan una serie de tests estadísticos para asegurar que solo el azar ha actuado en la selección.

Hechos todos los tests oportunos las tablas pueden considerarse correctas, y los dígitos o números que allí aparecen, pueden leerse en horizontal, en vertical, el diagonal, o con cualquier regla rebuscada pero siempre manteniéndola, aunque dada la disposición de las mismas lo normal es leer en vertical a partir de cualquier fila-columna, a lo que se denomina punto de arranque. Existen reglas para inicializar el uso de la tabla, como la regla del pulgar, o bien preguntar por cuatro números a cualquier persona, los dos primeros para identificar la fila y los otros dos para la columna, etc.

Siempre que se lean cifras de una en una, cada una de ellas tendrá una probabilidad asociada de $1/10$, si se leen de dos en dos $1/100$ y así sucesivamente.

Por lo tanto la forma de seleccionar una muestra será más o menos la siguiente:

a. Seguir una determinada fila o columna de números, comenzando en cualquier lugar, considerando solo las cifras necesarias para identificar a los elementos de la población a muestrear.

b. Seleccionar los n primeros números que aparezcan entre los menores que N , y

c. Cuando se lee algún número ya seleccionado, pues las tablas son con reposición se prescinde de él y se debe buscar siguiendo el mismo procedimiento otro.

2. La mayoría de los paquetes informáticos tienen los mandatos correspondientes para 28
generar muestras aleatorias. En los programas conviene que quede identificado el punto de

arranque, mediante el mandato semilla aleatoria o equivalente, de esa forma no se necesitará guardar el listado completo de los números seleccionados por si fuera necesario. Algunos paquetes carecen de dicho mandato lo cual es un inconveniente, porque si se deseara justificar la muestra obtenida por cualquier cuestión, no habría más remedio que guardar todo el listado.

3.7 Estratificación de la población

Con la aplicación del muestreo estratificado se espera conseguir mejora importante en cuanto al tamaño muestral, sobre todo si en el proceso de estratificación se consiguen:

- estratos o subpoblaciones cuyos elementos sean lo más homogéneos entre si (esto es con la menor posible dispersión intra, o por dentro de cada estrato); y
- que entre ellos, esto es, considerando a cada uno de ellos como un bloque o subpoblación que se identifican y comparan por su promedio, lo más heterogéneos entre si, esto es con la mayor dispersión inter o, lo que es lo mismo, que sus respectivas medias difieran lo más posible, pues como dilucidaremos estas condiciones son las que optimizan el muestreo estratificado. Ahora bien siempre hay una restricción previa, la posibilidad real de estratificar en las condiciones indicadas para la optimización y el análisis del coste de la misma.

En el muestreo estratificado, una *población heterogénea* con N unidades $\{u_i\}_{i=1,2,\dots,N}$ se subdivide en L *subpoblaciones, lo más homogéneas posibles*, no solapadas denominadas

estratos $\{u_{hi}\}_{\substack{h=1,2,\dots,L \\ i=1,2,\dots,N_h}}$ de tamaños N_1, N_2, \dots, N_L . La muestra estratificada de tamaño n se obtiene seleccionando n_h elementos ($h=1,2,\dots,L$) de cada uno de los L estratos en que se subdivide la población de forma independiente. Si la muestra estratificada se obtiene seleccionando una muestra aleatoria simple en cada estrato de forma independiente, el muestreo se denomina **muestreo aleatorio estratificado**, pero en general, nada impide utilizar diferentes tipos de selección en cada estrato.

Para un estrato en particular pueden pertenecer todas sus unidades a la muestra, parte de ellas o ninguna. También puede ocurrir que para formar la muestra estratificada se obtengan elementos de todos los estratos o sólo parte de ellos. Si sabemos seguro que un determinado estrato aporta unidades para la muestra, dicho estrato se denomina **estrato correpresentado**. Por otra parte, las unidades de la población que con certeza van a pertenecer a la muestra se denominan **autorepresentadas**.

Podemos representar gráficamente la población dividida en h estratos de tamaño N_h de cada uno de los cuales seleccionamos de modo independiente n_h unidades (mediante muestreo aleatorio simple si no se especifica otra cosa) para la muestra estratificada de tamaño n .

Podemos expresar la formación de estratos en la población y la formación de la muestra estratificada de la forma siguiente:

$$\left\{ u_1 \ u_2 \ \dots \ u_N \right\} \quad \text{Se divide en } L \text{ estratos} \quad \left\{ \begin{array}{l} u_{11} \ u_{12} \ \dots \ u_{1N_1} \\ u_{21} \ u_{22} \ \dots \ u_{2N_2} \\ \dots \dots \dots \\ u_{L1} \ u_{L2} \ \dots \ u_{LN_L} \end{array} \right\} \quad \sum_{h=1}^L N_h = N$$

MUESTRA

$$\left\{ u_1 \ u_2 \ \dots \ u_n \right\} \quad \text{Se extrae en cada estrato} \quad \left\{ \begin{array}{l} u_{11} \ u_{12} \ \dots \ u_{1n_1} \\ u_{21} \ u_{22} \ \dots \ u_{2n_2} \\ \dots \dots \dots \\ u_{L1} \ u_{L2} \ \dots \ u_{Ln_L} \end{array} \right\} \quad \sum_{h=1}^L n_h = n$$

3.7.1 Razones para el uso de Muestreo Estratificado

Son diversos los motivos que aconsejan efectuar una partición de nuestra población

$\{u_i\}_{i=1,2,\dots,N}$ en L subpoblaciones, no solapadas, $\{u_{hi}\}_{\substack{h=1,2,\dots,L \\ i=1,2,\dots,N_h}}$, entre los que destacan los

siguientes:

1. El muestreo estratificado puede aportar información más precisa de algunas subpoblaciones que varían bastante en tamaño y propiedades entre sí, pero que son homogéneas dentro de sí.
2. El uso adecuado del muestreo estratificado puede generar ganancia en precisión, pues al dividir una población heterogénea en estratos homogéneos, el muestreo en estos estratos tiene poco error debido precisamente a la homogeneidad.

3. En otros casos la estratificación viene motivada por el requerimiento de estimaciones para ciertas áreas o regiones geográficas. En esta situación cada estrato será un área compacta, como por ejemplo un municipio, una provincia, una colonia de una ciudad, etc.

4. También es una razón para utilizar muestreo estratificado la existencia de una variable precisa para la estratificación cuyos valores permitan dividir convenientemente la población en estratos homogéneos. Las variables utilizadas para la estratificación deberán estar correlacionadas con las variables objeto de la investigación. Por ejemplo, si se quiere realizar estadísticas en el sector educativo puede utilizarse la variable de estratificación nivel de enseñanza, tomando como estratos los niveles de enseñanza infantil, enseñanza primaria, enseñanza secundaria obligatoria, bachillerato y enseñanza universitaria (cada estrato tiene así unas características muy peculiares que lo hacen homogéneo).

3.7.2 Afijación de la Muestra

Se llama afijación de la muestra al reparto, asignación, adjudicación, adscripción o distribución del tamaño muestral n entre los diferentes estratos. Esto es, a la determinación de los valores de n_h que verifiquen $n_1 + n_2 + \dots + n_L = n$. Pueden establecerse muchas afijaciones o maneras de repartir la muestra entre los estratos, pero las más importantes son: la afijación uniforme, la afijación proporcional, la afijación de varianza mínima y la afijación óptima.

3.7.3 Afijación Uniforme

Consiste en asignar el mismo número de unidades muestrales a cada estrato, con lo que se tomarán todos los n_h iguales a n/L , aumentando o disminuyendo este tamaño en una unidad si n no fuese múltiplo de L , esto es $n_h = \lceil n/L \rceil + 1$.

$$n_h = k \quad \forall h = 1 \dots L \Rightarrow \sum_{h=1}^L n_h = \sum_{h=1}^L k \Rightarrow n = Lk \Rightarrow f_h = \frac{n_h}{N_h} = \frac{k}{N_h}$$

Este tipo de afijación da la misma importancia a todos los estratos, en cuanto a tamaño de la muestra, con lo cual favorecerá a los estratos de menor tamaño y perjudicará a los grandes en cuanto a precisión. Sólo es conveniente en poblaciones con estratos de tamaño similar.

3.7.4 Afijación Proporcional

Consiste en asignar a cada estrato un número de unidades muestrales proporcional a su tamaño. Las n unidades de la muestra se distribuyen proporcionalmente a los tamaños de los estratos expresados en número de unidades. Tenemos:

$$n_h = N_h k \Rightarrow \underbrace{\sum_{h=1}^L n_h}_n = \sum_{h=1}^L N_h k = k \sum_{h=1}^L N_h \Rightarrow n = kN \Rightarrow k = \frac{n}{N} = f$$

$$f_h = \frac{n_h}{N_h} = \frac{N_h k}{N_h} = k = f$$

$$W_h = \frac{N_h}{N} = \frac{n_h/k}{n/k} = \frac{n_h}{n}$$

A la vista de los resultados anteriores podemos asegurar lo siguiente:

Las fracciones de muestreo en los estratos son iguales y coinciden con la fracción global de muestreo, siendo su valor la constante de proporcionalidad.

Los coeficientes de ponderación W_h se obtienen exclusivamente a partir de la muestra, pues para su cálculo sólo son necesarios valores muestrales (n_h y n).

Como $\pi_{hi} = \frac{n_h}{N_h} = k = f$ todas las unidades de la población tienen la misma probabilidad de figurar en la muestra de n unidades, es decir, estamos en el caso de *muestras autoponderadas*.

3.7.5 Afijación de Mínima Varianza (o Afijación de Neyman)

La afijación de mínima varianza, o afijación de Neyman, consiste en determinar los valores de n_h (número de unidades que se extraen del estrato h -ésimo para la muestra) de forma que para un tamaño de muestra fijo, igual a n , la varianza de los estimadores sea mínima

Donde:

$$n_h = n \cdot \frac{N_h S_h}{\sum_{h=1}^L N_h S_h}$$

Otra expresión para n_h es :

$$n_h = n \cdot \frac{N_h S_h}{\sum_{h=1}^L N_h S_h} = n \cdot \frac{\frac{N_h}{N} S_h}{\sum_{h=1}^L \frac{N_h}{N} S_h} = n \cdot \frac{W_h S_h}{\sum_{h=1}^L W_h S_h}$$

Vemos que los valores de n_h son proporcionales a los productos $N_h \cdot S_h$ y en el supuesto de que $S_h = S$, $\forall h = 1, 2, \dots, L$, esta afijación de mínima varianza coincidiría con la proporcional tal y como se ve a continuación:

$$S_h = S \Rightarrow n_h = n \cdot \frac{N_h S}{\sum_{h=1}^L N_h S} = \frac{n N_h}{N} = k N_h \quad \text{con } k = \frac{n}{N}$$

3.7.6 Afijación Óptima

La afijación óptima consiste en determinar los valores de n_h (número de unidades que se extraen del estrato h -ésimo para la muestra) de forma que para un coste fijo C la varianza de los estimadores sea mínima. El coste fijo C será la suma de los costes derivados de la selección de las unidades muestrales de los estratos, es decir, si c_h es el coste de unidad por muestreo en el estrato h , el coste total de selección de las n_h unidades muestrales en este estrato será $c_h n_h$. Sumando los costes $c_h n_h$ para los L estratos tenemos el coste total de selección de la muestra estratificada.

$$n_h = n \cdot \frac{N_h S_h / \sqrt{c_h}}{\sum_{h=1}^L N_h S_h / \sqrt{c_h}}$$

Vemos que los valores de n_h son proporcionales a los productos $N_h \cdot S_h / \sqrt{c_h}$ y en el supuesto de que $c_h = k \quad \forall h = 1, 2, \dots, L$ (coste constante en todos los estratos) la afijación óptima coincide con la de mínima varianza, y si además $S_h = S, \forall h = 1, 2, \dots, L$ la afijación óptima coincidirá con la de mínima varianza y con la proporcional.

3.7.7 Comparación de Eficiencias según los distintos tipos de Afijación.

Se realizará un estudio comparativo de la conveniencia de los distintos tipos de afijación en términos de su eficiencia medida a través el error de muestreo, o lo que es lo mismo, a través de la varianza. Por lo tanto será más eficiente aquel tipo de afijación que presente menos varianza.

El muestreo estratificado con afijación proporcional es más preciso que el muestreo aleatorio simple, produciéndose la igualdad de precisiones cuando las medias de los

estratos son todas iguales. Por lo tanto la ganancia en precisión del muestreo estratificado respecto del aleatorio simple será mayor cuanto más distintas entre sí sean las medias de los estratos, es decir, para que el muestreo estratificado sea preciso es conveniente que los estratos sean heterogéneos entre sí en media.

El muestreo estratificado con afijación de mínima varianza es más preciso que el muestreo estratificado con afijación proporcional, produciéndose la igualdad de precisiones cuando las cuasidesviaciones típicas de los estratos son todas iguales.

Por lo tanto la ganancia en precisión del muestreo estratificado con afijación de mínima varianza respecto del muestreo estratificado con afijación proporcional será mayor cuanto más distintas entre sí sean las cuasidesviaciones típicas de los estratos, es decir, para que el muestreo estratificado sea más preciso es conveniente que los estratos sean heterogéneos entre sí en desviación típica.

En realidad podemos ver que:

$$V_{MAS}(\bar{x}) \geq V_{MEP}(\bar{x}) \geq V_{MEMV}(\bar{x})$$

Lo que permite asegurar que en general el muestreo estratificado con afijación de mínima varianza es más preciso que el muestreo estratificado con afijación proporcional y que el aleatorio simple, siendo además el estratificado con afijación proporcional más preciso que el aleatorio simple.

4.1 Materiales

En este capítulo se presenta los materiales utilizados en el presente trabajo y las características técnicas de la metodología. El objetivo es presentar de manera ordenada y detallada, tanto las estrategias metodológicas desarrolladas durante la realización de esta investigación, como la descripción del proceso de encuesta directa y de respuesta aleatorizada para hacer las cuantificaciones requeridas en el marco de un proceso inferencial que incluye a un conjunto de preguntas sensibles.

En esta investigación, lo fundamental es analizar los aportes de la estadística y las diversas técnicas de muestreo probabilístico para la realización de encuestas confiables, analizando la fiabilidad de los estimadores, así como para realizar en su caso recomendaciones sobre eficiencia, y mejoras posibles a introducir en la recolección de datos.

Universo. El Universo o población aquí está constituido por los 1550 alumnos matriculados en el Semestre 2011-B, en la Universidad Nacional del Callao.

Unidad de muestreo: Cada alumno matriculado en el Semestre 2011-B, en la Universidad Nacional del Callao.

Tamaño de la muestra. Para determinar el tamaño de la muestra, se utiliza la técnica del muestreo aleatorio simple, para la estimación de proporciones con un error no mayor de 0.05 y con una probabilidad de 95%.

Los datos corresponden a observaciones de las siguientes variables categóricas nominales:

1. Consumo de alimentos en el Centro de cómputo de la facultad.
2. Consumo de drogas.
3. Fraude en los exámenes.
4. Consumo de alcohol en torno al campus de la UNAC.
5. Ausencia injustificada a clases.

Para el procesamiento de los datos se ha utilizado una computadora compatible pentium V 1300 mhz, donde se usaron el procesador de texto Word, impresora Epson 5600, y los datos fueron procesados con el programa estadístico SPSS versión 18, mientras que los diferentes tamaños de muestra fueron calculados con el software Excel.

4.2 Métodos

Aquí se presenta la metodología que se siguió para aplicar el cuestionario de preguntas a raves de la encuesta directa. De igual manera, se presenta una breve descripción de proceso diseñado para la aplicación de las preguntas del método de “Respuestas Aleatorizadas”.

- a) Los datos se han recopilado mediante una encuesta directa y a través de la técnica de respuesta aleatorizada.
- b) Para el tratamiento de los datos correspondientes a las variables categóricas, se ha utilizado las técnicas estadísticas descriptivas, como son tablas de frecuencias relativas, porcentajes y medidas de resumen.

c) Para la estimación de los tamaños de muestra, de los intervalos de confianza y demás inferencia estadística, se han utilizado técnicas estadísticas de muestreo basados tanto en la metodología estadística clásica, así como en las técnicas de respuesta aleatorizada, tal como se indicó en el proyecto de investigación.

4.3 Diseño de la Encuesta.

En este capítulo discutimos los pasos que seguimos para el diseño de las encuestas, tanto de la encuesta directa como para la Respuesta Aleatorizada. De igual forma se explica brevemente la técnica para aleatorizar las preguntas en el cuestionario.

Con la finalidad de comprobar la eficacia de método de Respuesta Aleatorizada este trabajo de investigación incluye el diseño de dos tipos de encuesta. Las preguntas en ambas encuestas, serán las mismas; sin embargo, para el método de Respuestas Aleatorizadas, será necesario tener una serie de preguntas no sensibles, con probabilidades conocidas. Las dos encuestas se aplicarán de manera distinta, de acuerdo a las necesidades de cada una, como lo veremos a continuación.

4.3.1 Formato de la encuesta directa.

El cuestionario que se aplicará de manera directa, tiene como finalidad la medición de 5 variables categóricas. Las 5 preguntas miden variables cualitativas y tienen como única opción de respuesta SI o NO.

El cuestionario con preguntas directas es el siguiente:

Encuesta anónima			
Marca con una X tu respuesta			
SEXO	<input type="checkbox"/> Masculino	<input type="checkbox"/> Femenino	
1. ¿Sueles consumir alimentos o bebidas en la sala de computo?	<input type="checkbox"/> SI	<input type="checkbox"/> NO	
2. ¿Alguna vez has consumido drogas alucinógenas?	<input type="checkbox"/> SI	<input type="checkbox"/> NO	
3. ¿Algunas veces has copiado en los exámenes?	<input type="checkbox"/> Si	<input type="checkbox"/> NO	
4. ¿Has consumido bebidas alcohólicas, cerca a la Universidad?	<input type="checkbox"/> SI	<input type="checkbox"/> NO	
5. ¿En este ciclo, has faltado a clases sin motivo justificable?	<input type="checkbox"/> SI	<input type="checkbox"/> NO	

Figura 3.1 Preguntas de la Encuesta Directa
Fuente: Elaboración propia

Este cuestionario será aplicado a la muestra de estudiantes hombres y mujeres, de la forma convencional o directa.

4.3.2 Formato de encuesta con la técnica de Respuesta Aleatorizada.

Para la técnica de “Respuesta Aleatorizada”, se aplica el método de “Segunda pregunta no relacionada”, para lo cual es necesario un cuestionario adicional con las siguientes características:

- Las preguntas no deben de ser sensibles.
- No deben estar relacionadas con el tema de la pregunta “sensible”.
- Para cada una de las preguntas que miden variables cualitativas, las respuestas deben tener una probabilidad conocida.

El cuestionario con las características anteriores, se colocará a la par del cuestionario de preguntas “sensibles”, quedando de la siguiente forma:

Sexo	<input type="text" value="Femenino"/>	Encuesta anónima		
	<input type="text" value="Masculino"/>	Marcar con un X su respuesta		
Juego 1	NEGRO	¿Sueles consumir alimentos o bebidas en la sala de cómputo?	<input type="checkbox"/>	<input type="checkbox"/>
	ROJO	¿Naciste en el mes de julio?	<input type="checkbox"/>	<input type="checkbox"/>
Juego 2	NEGRO	¿Alguna vez has consumido drogas alucinógenas (marihuana, PBC, etc)?	<input type="checkbox"/>	<input type="checkbox"/>
	ROJO	¿Tu DNI termina en número 2?	<input type="checkbox"/>	<input type="checkbox"/>
Juego 3	NEGRO	¿Algunas veces has copiado en los exámenes?	<input type="checkbox"/>	<input type="checkbox"/>
	ROJO	¿Naciste en el mes de diciembre?	<input type="checkbox"/>	<input type="checkbox"/>
Juego 4	NEGRO	¿Has consumido bebidas alcohólicas, cerca a la Universidad?	<input type="checkbox"/>	<input type="checkbox"/>
	ROJO	¿Naciste en el mes de enero?	<input type="checkbox"/>	<input type="checkbox"/>
Juego 5	NEGRO	¿En este ciclo, has faltado a clases sin motivo justificable?	<input type="checkbox"/>	<input type="checkbox"/>
	ROJO	¿Tu DNI termina en número 5?	<input type="checkbox"/>	<input type="checkbox"/>

Figura 3.3 Preguntas de la encuesta para Respuestas aleatorizadas.
Fuente: Elaboración propia.

Para la estimación de resultados, será necesario tomar en cuenta las probabilidades del cuestionario de preguntas no sensibles. Por tal motivo, presentamos la tabla:

Tabla 3.1 Probabilidades de preguntas no sensibles.

#	Pregunta	Probabilidad de Respuesta
1	¿Naciste en el mes de julio?	1/12
2	¿Tu DNI termina en número 2?	1/10
3	¿Naciste en el mes de diciembre?	1/12
4	¿Naciste en el mes de enero?	1/12
5	¿Tu DNI termina en número 5?	1/10

Fuente: Elaboración Propia

4.3.3 Diseño de muestra para encuesta directa.

Para la encuesta directa, se diseñó una muestra aleatoria con un error máximo de 4.7% y una confiabilidad de 95%

. La fórmula que nos permitirá determinar el tamaño muestral es la siguiente:

$$n = \frac{N z_{\alpha/2}^2 P(1-P)}{(N-1)e^2 + z_{\alpha/2}^2 P(1-P)}$$

donde

$z_{\alpha/2}$: z correspondiente al nivel de confianza elegido

P: proporción de una categoría de la variable. Si no se sabe nada acerca de P, usar 0.5.

e: error máximo

N: tamaño de la población

N	1550
Z	1.96
Z ²	3.8416
p	0.5
1-p	0.5
N-1	1549
E	0.047
E ²	0.002209
Z ² p (1-p)	0.9604
n	340

Por lo tanto, el cuestionario de preguntas se aplicara a una muestra de 304 estudiantes en el proceso de recoger la información mediante la encuesta directa.

Para determinar el tamaño de la muestra para la aplicación de la técnica de “Respuesta Aleatorizada”, se usó el método de muestreo aleatorio simple para la estimación de proporciones con un error no mayor de 4.2% a efectos de tener un tamaño de muestra mayor que en el caso de la encuesta directa y con una confiabilidad de 95%, resultando en un tamaño de 404 estudiantes.

4.3.4 Descripción de la técnica de encuesta directa.

La encuesta directa, como su nombre lo dice, consta de preguntas directas hechas por un entrevistador. Para este trabajo de investigación, la forma de aplicación es de manera escrita.

La encuesta directa se aplicó a una muestra representativa de la población de estudiantes de la facultad de Ciencias económicas de la Universidad Nacional del callao. Los cuestionarios fueron contestados por escrito marcando la respuesta de las preguntas categóricas con un aspa como se observa en la Figura 4.1 por el propio entrevistado y posteriormente depositados en un sobre para darles mayor confidencialidad.

4.3.5 Descripción de la técnica de la “Respuesta Aleatorizada”

Para la aplicación de la encuesta por medio de la técnica de “Respuesta Aleatorizada”, y con la finalidad de darle más confianza al encuestado, se les entregó una cartilla de instrucciones, con el siguiente contenido:

INSTRUCCIONES

PASO 1	Baraja las cartas que te van a entregar
PASO 2	Selecciona una carta al azar y no se la muestres a NADIE

PASO 3	<p>La pregunta que contestarás en cada juego, dependerá del COLOR de la carta que obtengas:</p> <p>Si es color negro, contesta a la pregunta NEGRO.</p> <p>Si es color rojo, contesta a la pregunta ROJO.</p> <p>Este procedimiento lo repetirás para cada juego.</p>
--------	---

IMPORTANTE: Tus respuestas son totalmente confidenciales.

Figura 4.2 Hoja de instrucciones para Captura de Datos

Fuente: Elaboración Propia

El entrevistado únicamente tuvo que poner en la hoja de preguntas un aspa en la respuesta que elija y después de haber terminado de llenar los espacios con las repuestas, se depositaba la ficha de encuesta en una bolsa. El desarrollo de la técnica de “Respuesta Aleatorizada” es necesario utilizar un proceso aleatorio que nos ayude a aleatorizar las respuestas.

El procedimiento aleatorio que se eligió es un grupo de 6 cartas de una baraja inglesa, con 4 cartas negras y 2 rojas y que va a dar lugar a un juego fácil de ejecutar.

Este juego de cartas está dividido de la siguiente manera:

Color	Numero de cartas
Negro	4
Rojo	2
TOTAL	6

Fuente: Elaboración Propia

Tabla 3.2 División de cartas usadas.

La anterior forma de distribución nos permite conocer fácilmente la probabilidad de que cualquier carta sea escogida al azar.

Como podemos apreciar en las Figuras 3.3 y 3.4, las preguntas sensibles están precedidas de la palabra Negro, y las no sensibles de la palabra Rojo. Cuando se aplica la encuesta a cada estudiante de la muestra, se le entrega el grupo de cartas, para que lo revuelva como se hace usualmente. Después se le pide que elija una carta al azar. Si la carta que elija es de color negro, las preguntas que tendrá que contestar serán las sensibles. Por otra parte si la carta que elija es de color rojo, las preguntas que tendrá que contestar serán las no sensibles.

De esta forma conocemos la probabilidad de que nos conteste a las preguntas sensibles, que es de $4/6$, y como consecuencia, la probabilidad de que nos conteste las preguntas no sensibles es de $2/6$. Con esto garantizamos que la probabilidad de que el cuestionario que del cual nos interesa conocer las respuestas, tiene más probabilidad de ser contestado.

5. RESULTADOS

Los resultados que se obtuvieron después de la aplicación de las encuestas son sumamente importantes, ya que a partir de éstos podemos hacer inferencias, comparaciones y contrastes. Es por ello que este capítulo presenta en forma detallada los resultados obtenidos tanto de la Encuesta Directa como de la encuesta de “Respuesta Aleatorizada” con lo que se cumple el objetivo de establecer los aportes de la teoría de las técnicas de “Respuesta Aleatorizada” en el estudio de características sensibles, y la de demostrar mediante la inferencia estadística, la hipótesis que afirma que cuando se utiliza el método de “Encuesta Directa” para medir variables sensibles, hay una tendencia a subestimarse los parámetros poblacionales en comparación al obtenido cuando se usa el método de Respuesta Aleatorizada, así como para realizar en su caso, recomendaciones sobre la obtención de datos y calidad de la encuesta.

5.1 Estimación de la Proporción en la Encuesta Directa.

Para la estimación de la proporción de estudiantes que dijeron SI abiertamente a las preguntas que miden las características sensibles categóricas, se realizaron los siguientes pasos:

1. Se contabilizaron el número de respuestas afirmativas de cada pregunta
2. Esa cantidad fue dividida entre el tamaño de la muestra aleatoria.

Los resultados obtenidos se muestran en la Tabla 5.1, en donde podemos observar que las proporciones de algunas preguntas son relativamente bajas, lo que nos dice que esas preguntas son consideradas por los encuestados mucho más sensibles que las demás, como por ejemplo las preguntas 1 y 2. Pero también es importante mencionar que en las preguntas 3, 4, y 5 se tienen proporciones altas, lo cual significa que esas preguntas que se consideraban sensibles, en realidad para los estudiantes no lo son tanto.

Tabla 5.1 Estimación de proporciones en la encuesta directa: Variables Cualitativas

Estadísticos

		¿Sueles consumir alimentos o bebidas en la sala de cómputo?	¿Alguna vez has consumido drogas alucinógenas?	¿Algunas veces has copiado en los exámenes?	¿Has consumido bebidas alcohólicas, cerca a la Universidad?	¿En este ciclo, has faltado a clases sin motivo justificable?
N	Válidos	340	340	340	340	340
	Perdidos	0	0	0	0	0

sexo

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	Masculino	156	45.9	45.9
	Femenino	184	54.1	100.0
	Total	340	100.0	

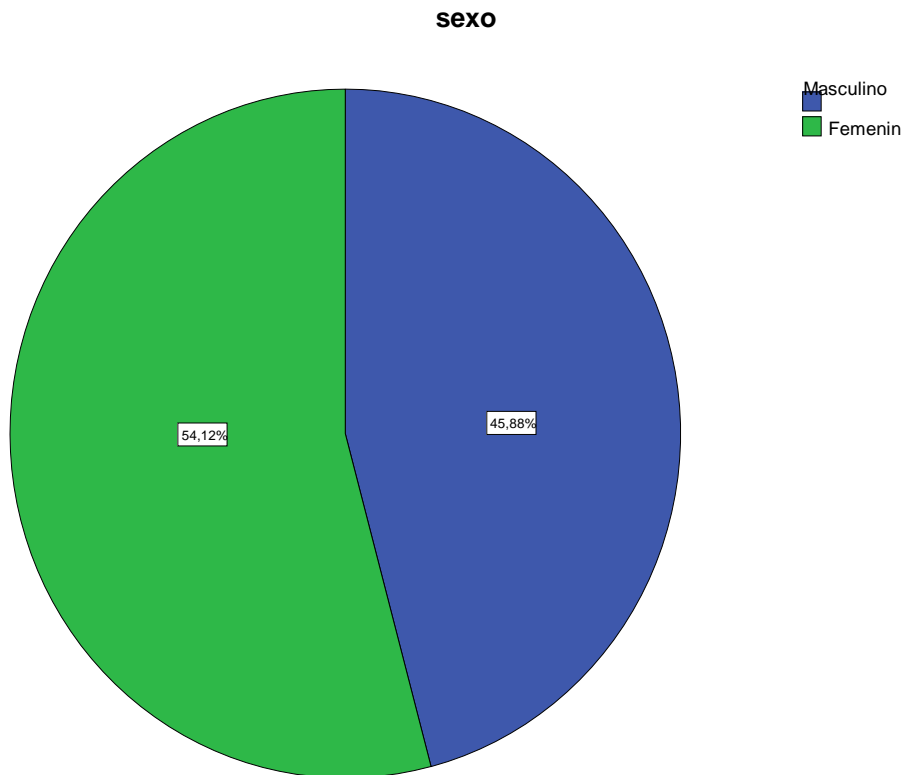


Gráfico 5.1

El 45.9% de la muestra está compuesta por varones, mientras que el restante 54.1% son mujeres.

¿Sueles consumir alimentos o bebidas en la sala de cómputo?

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	SI	109	32.1	32.1
	NO	231	67.9	100.0
	Total	340	100.0	

Tabla 5.2

Se observa tanto en la tabla como en el gráfico 5.2 que el 32.1% de los encuestados suelen consumir alimentos o bebidas en el centro de cómputo.

¿Sueles consumir alimentos o bebidas en la sala de computo?

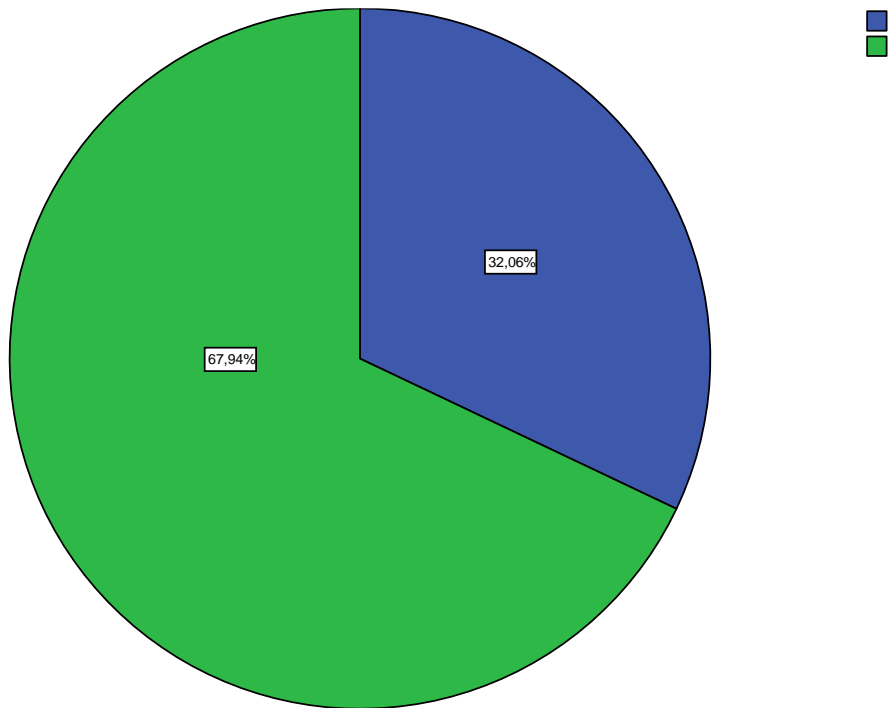


Grafico 5.2

¿Alguna vez has consumido drogas alucinógenas?

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	SI	28	8.2	8.2
	NO	312	91.8	100.0
	Total	340	100.0	

Tabla 5.3

Se observa tanto en la tabla como en el grafico 5.3 que el 8.2% de los encuestados suelen consumir drogas alucinógenas.

¿Alguna vez has consumido drogas alucinógenas?

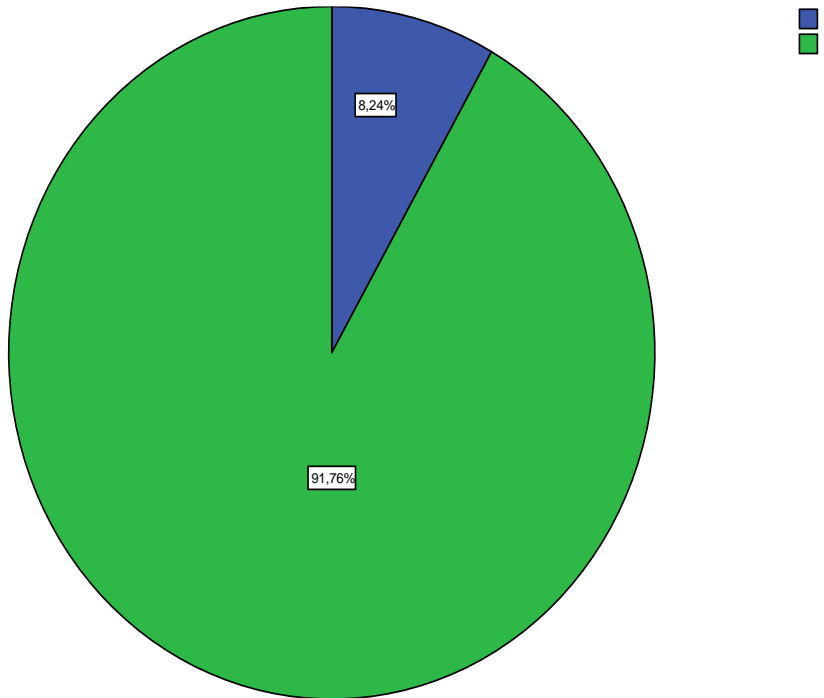


Gráfico 5.3

¿Algunas veces has copiado en los exámenes?

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	SI	266	78.2	78.2
	NO	74	21.8	100.0
	Total	340	100.0	

Tabla 5.4

Se observa tanto en la tabla como en el gráfico 5.4 que el 78.2% de los encuestados han copiado alguna vez en los exámenes.

¿Algunas veces has copiado en los exámenes?

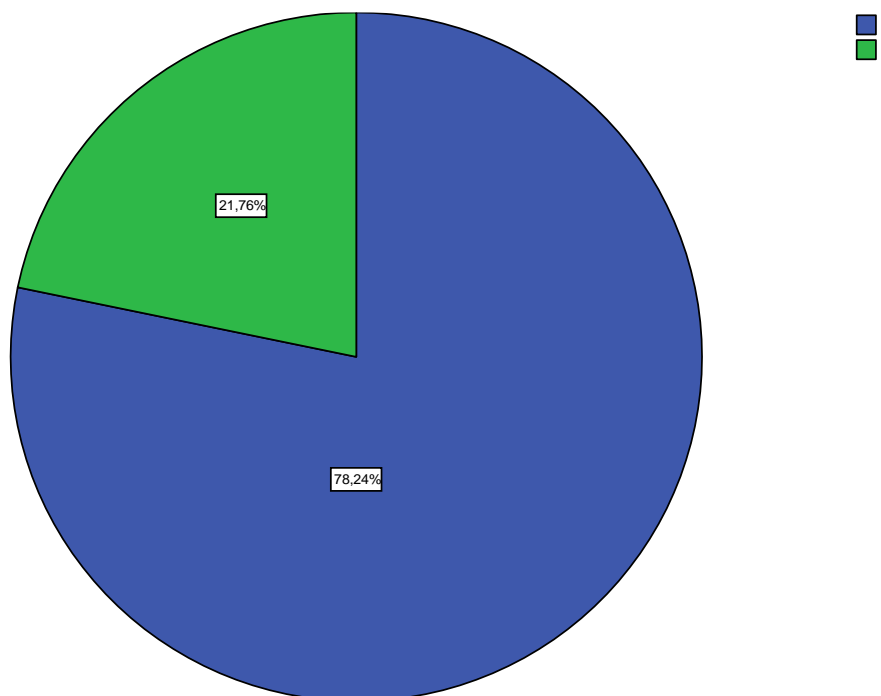


Gráfico 5.4

¿Has consumido bebidas alcohólicas, cerca a la Universidad?

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	SI	182	53.5	53.5
	NO	158	46.5	100.0
	Total	340	100.0	

Tabla 5.5

Se observa tanto en la tabla como en el gráfico 5.5 que el 53.5 % de los encuestados han consumido bebidas alcohólicas cerca a la universidad.

¿Has consumido bebidas alcohólicas, cerca a la Universidad?

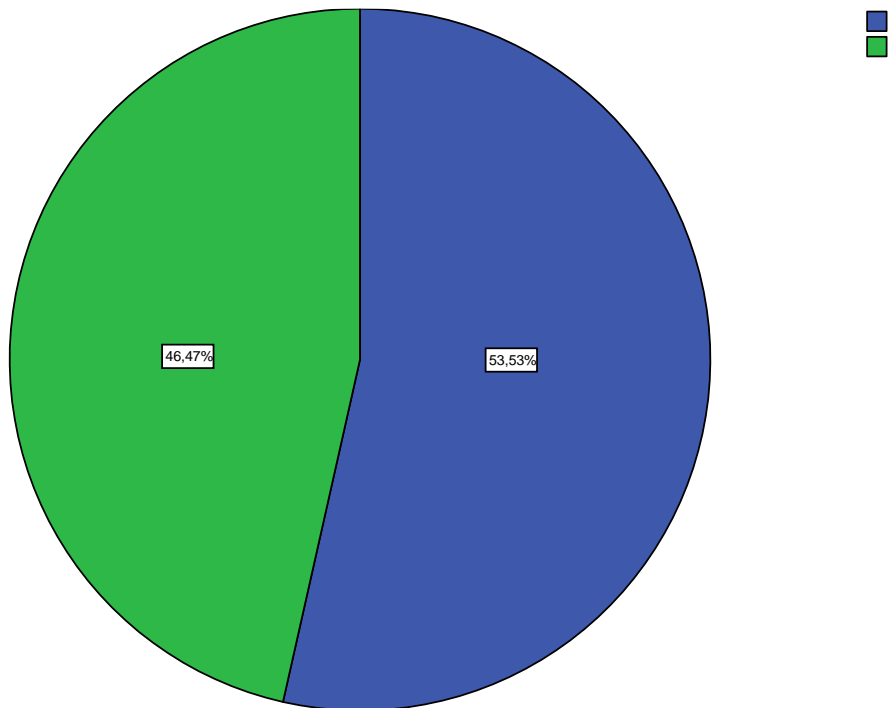


Gráfico 5.5

¿En este ciclo, has faltado a clases sin motivo justificable?

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	SI	212	62.4	62.4
	NO	128	37.6	100.0
	Total	340	100.0	

Tabla 5.6

Se observa tanto en la tabla como en el gráfico 5.6 que el 62.4 % de los encuestados han faltado a clases en este ciclo, sin motivo justificable.

¿En este ciclo, has faltado a clases sin motivo justificable?

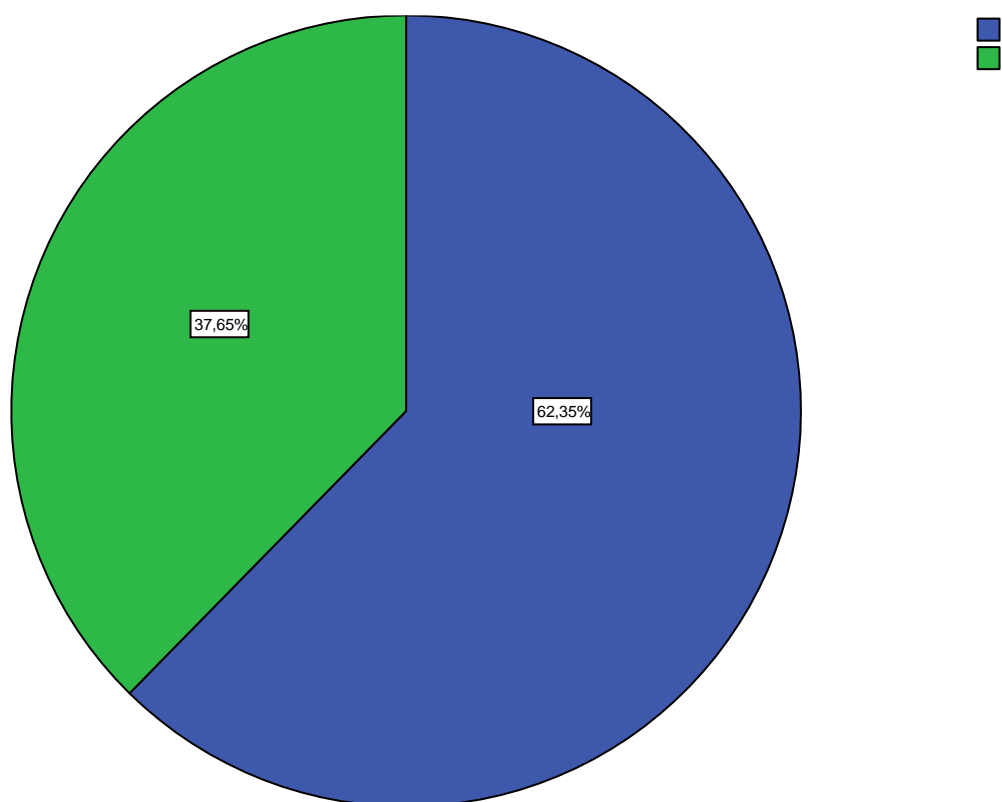


Tabla 5.7 Resumen de Proporciones estimadas

Número de pregunta	Número de respuestas Si	proporción
1	109	0.320588235
2	28	0.082352941
3	266	0.782352941
4	182	0.535294118
5	212	0.623529412

Fuente: Elaboración propia

n_{estrato2}

n_{estrato4}

5.2 Estimación por medio de la Técnica de Respuesta Aleatorizada

Para la estimación de las proporciones para variables cualitativas en la Técnica de “Respuesta Aleatorizada”, se realizaron los siguientes pasos:

1. Por medio del programa estadístico SPSS, podemos saber fácil y rápidamente el número de estudiantes de la muestra que respondieron afirmativamente a cada una de la preguntas.
2. Se asignaron los datos conocidos de las probabilidades de las preguntas no sensibles, así como los datos obtenidos del programa para cada una de las preguntas sensibles como se presentan en la Tabla 5.15.

Tabla 5.8 Resumen de casos válidos

	SEXO	¿Sueles consumir alimentos o bebidas en la sala de cómputo? / ¿Naciste en el mes de julio?	¿Alguna vez has consumido drogas alucinógenas? / ¿Tu DNI termina en número 2?	¿Algunas veces has copiado en los exámenes? / ¿Naciste en el mes de diciembre?	¿Has consumido bebidas alcohólicas, cerca a la Universidad? / ¿Naciste en el mes de enero?	¿En este ciclo, has faltado a clases sin motivo justificable? / ¿Tu DNI termina en número 5?
N	Válidos	404	404	404	404	404
	Perdidos	0	0	0	0	0

Tabla 5.9 Género de los encuestados

	Frecuencia	Porcentaje	Porcentaje acumulado
Válidos Masculino	197	48.8	48.8
Femenino	207	51.2	100.0
Total	404	100.0	

La muestra consta de 404 alumnos. Varones son 197 y 207 mujeres.

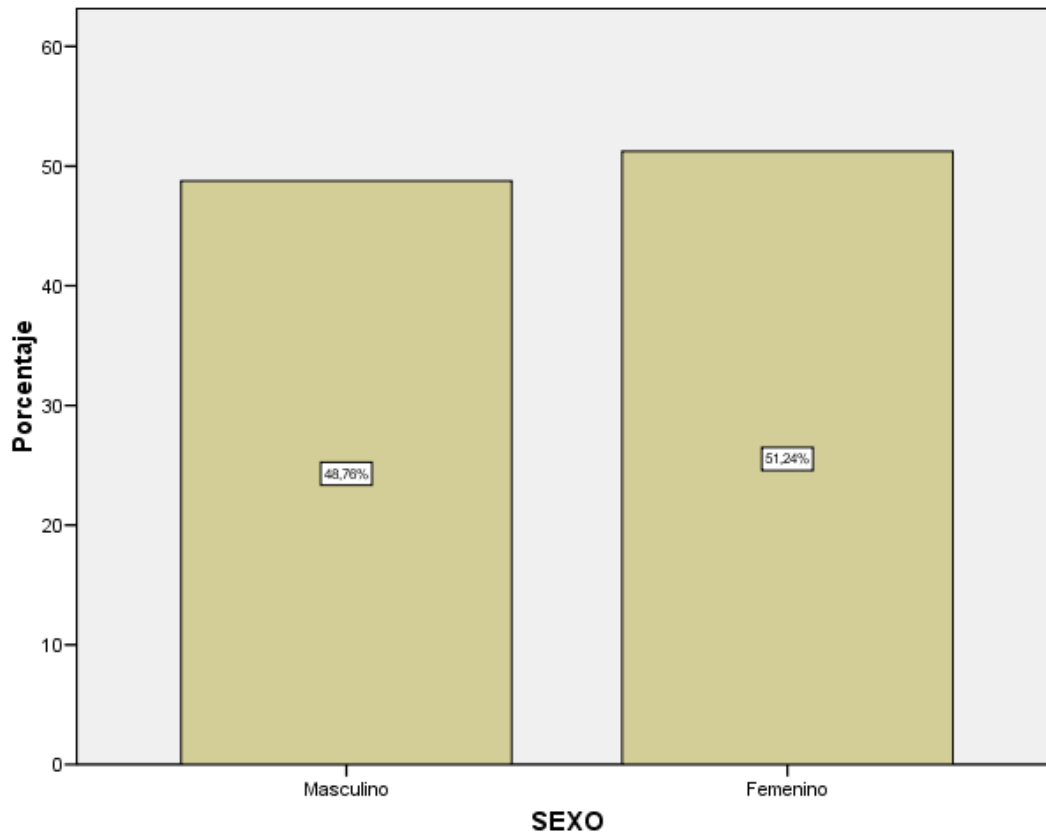


Grafico 5.9

¿Sueles consumir alimentos o bebidas en la sala de cómputo? / ¿Naciste en el mes de julio?

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	SI	152	37.6	37.6
	NO	252	62.4	100.0
	Total	404	100.0	

Tabla 5.10

Con el método de respuesta aleatorizada, se observa que el 37.6% de los estudiantes contestaron afirmativamente a la pregunta que se refiere al consumo de alimentos o bebidas en la sala de cómputo o el haber nacido en el mes de julio.

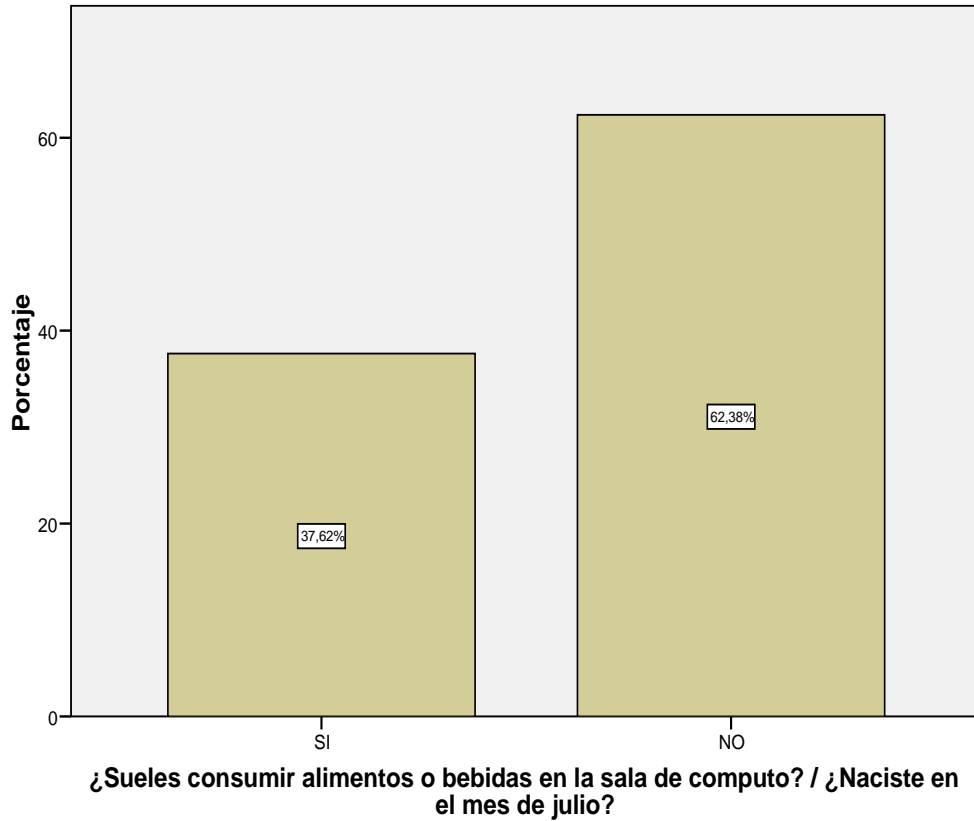


Grafico 5.10

¿Alguna vez has consumido drogas alucinógenas? / ¿Tu DNI termina en número 2?

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	SI	75	18.6	18.6
	NO	329	81.4	100.0
	Total	404	100.0	

Tabla 5.11

Con el método de respuesta aleatorizada, se observa que el 18.6% de los estudiantes contestaron afirmativamente a la pregunta que se refiere al consumo de drogas alucinógenas o si su DNI termina en número 2.

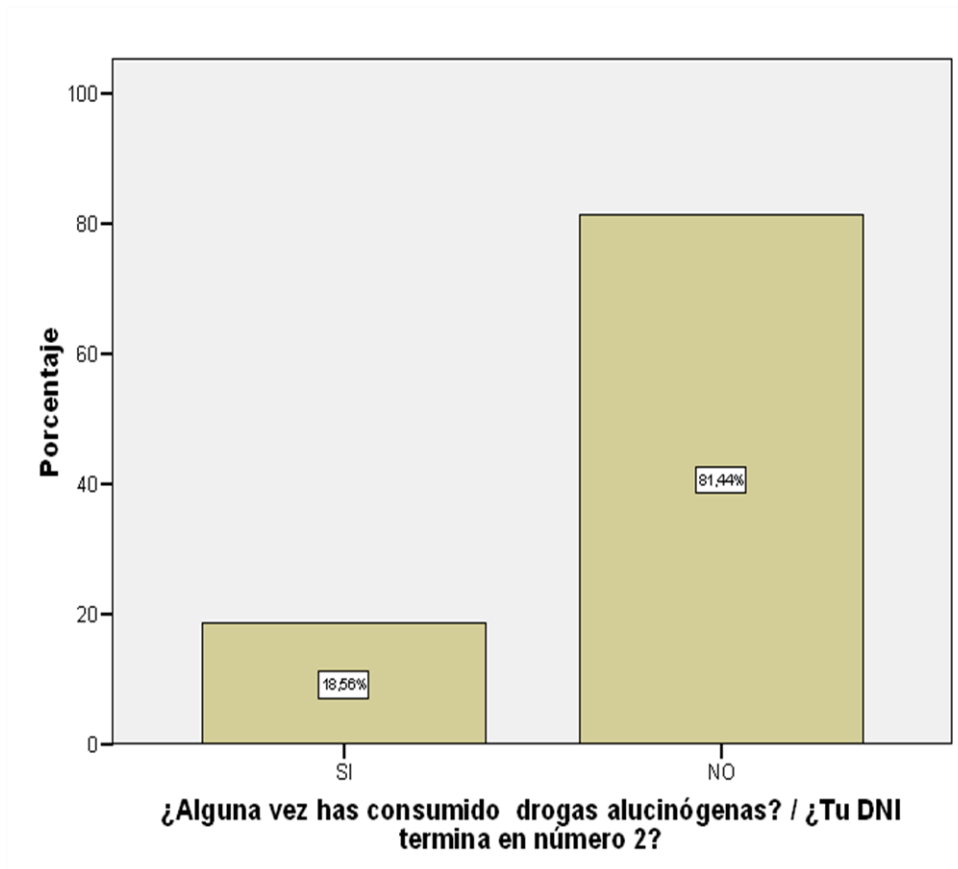


Grafico 5.11

¿Algunas veces has copiado en los exámenes? / ¿Naciste en el mes de diciembre?

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	SI	233	57.7	57.7
	NO	171	42.3	100.0
	Total	404	100.0	

Tabla 5.12

Con el método de respuesta aleatorizada, se observa que el 57.7% de los estudiantes contestaron afirmativamente a la pregunta que se refiere al haber copiado en los exámenes o el haber nacido en el mes de diciembre.

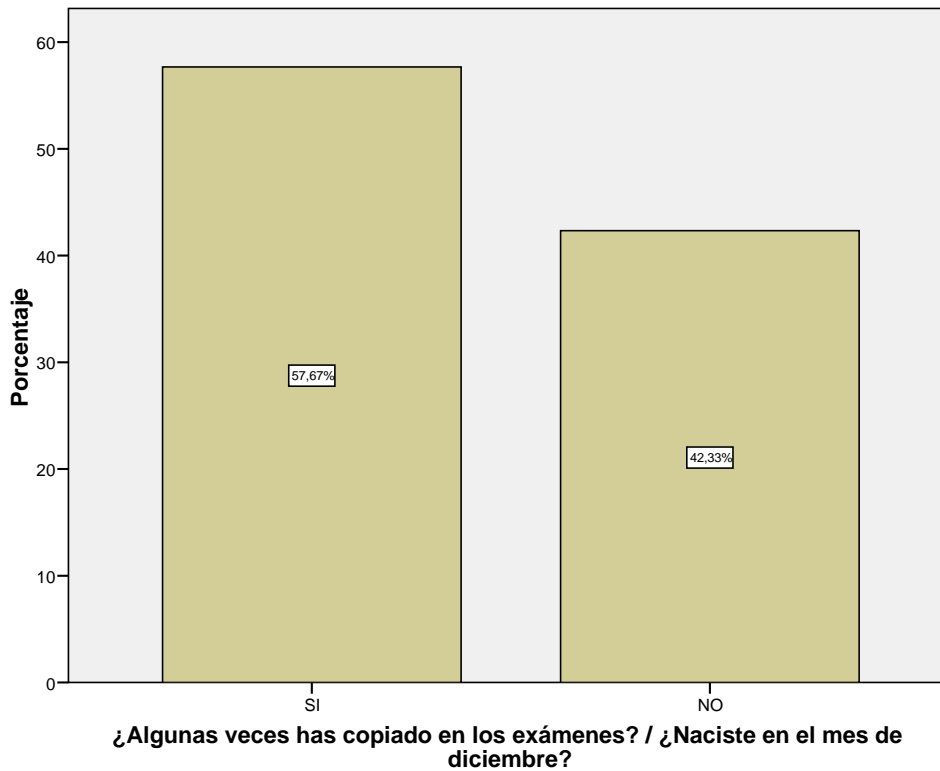


Grafico 5.12

¿Has consumido bebidas alcohólicas, cerca a la Universidad? / ¿Naciste en el mes de enero?

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	SI	178	44.1	44.1
	NO	226	55.9	100.0
Total		404	100.0	

Tabla 5.13

Con el método de respuesta aleatorizada, se observa que el 44.1% de los estudiantes contestaron afirmativamente a la pregunta que se refiere al consumo de bebidas alcohólicas, cerca a la universidad o el haber nacido en el mes de enero.

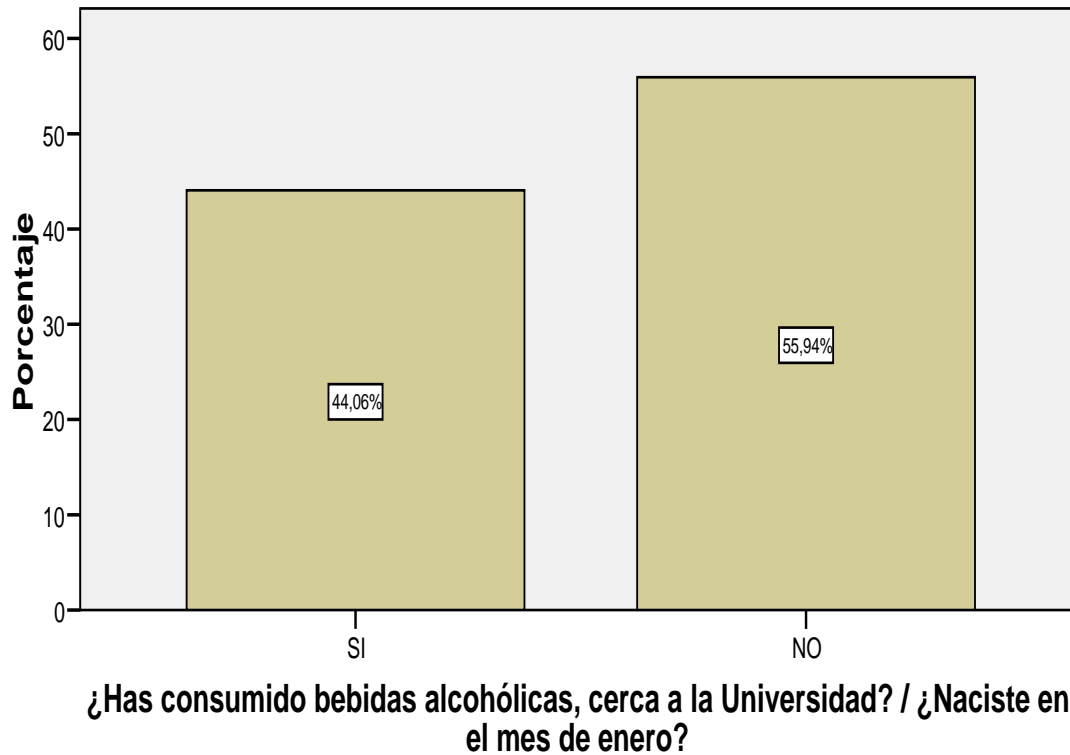


Gráfico 5.13

¿En este ciclo, has faltado a clases sin motivo justificable? / ¿Tu DNI termina en número 5?

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	SI	199	49.3	49.3
	NO	205	50.7	100.0
	Total	404	100.0	

Tabla 5.14

Con el método de respuesta aleatorizada, se observa que el 49.3% de los estudiantes contestaron afirmativamente a la pregunta que se refiere al haber faltado a clases sin motivo justificable o si su DNI termina en número 5.

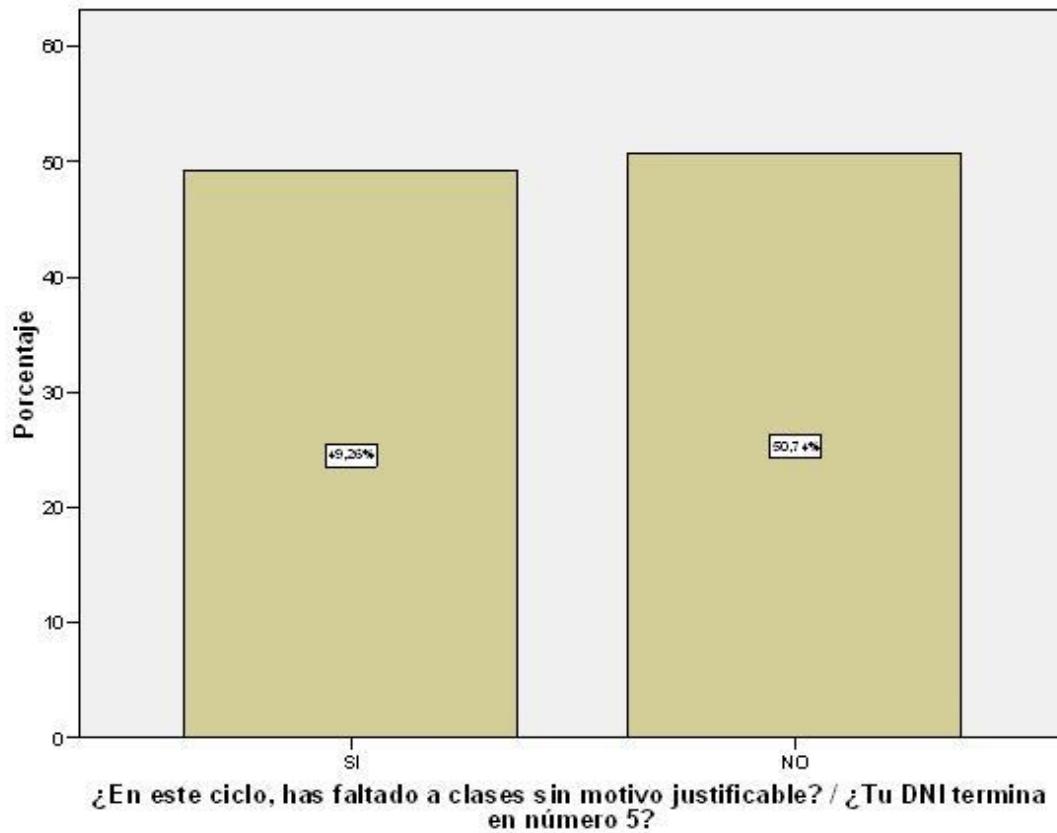


Grafico 5.14

Tabla 5.15 Datos obtenidos de la encuesta y probabilidades conocidas.

Tamaño de la muestra = 404 estudiantes				
número de pregunta	Respuestas si en la muestra	p	1-p	Probabilidad de la característica no sensitiva
1	152	0.67	0.33	0.0833
2	75	0.67	0.33	0.1
3	233	0.67	0.33	0.0833
4	178	0.67	0.33	0.0833
5	199	0.67	0.33	0.1

Fuente: Elaboración propia

3. El cálculo de la proporción se presenta en la Tabla 5.16, para la muestra total utilizando la fórmula que aparece arriba de la columna, las cuales ya fueron explicadas en el capítulo 2

Tabla 5.16 Estimación de las proporciones

Número de Pregunta	Proporción	Estimado de la característica sensible
1	0.3762376	0.52052033
2	0.1856436	0.22782622
3	0.5767327	0.81976668
4	0.4405941	0.61657472
5	0.4925743	0.68593173

Fuente: Elaboración propia

4. Al tener las proporciones estimadas a partir de la muestra, se procede a calcular las varianzas de los estimadores de las preguntas sensibles, las cuales se presentan en la Tabla 5.17.

Tabla 5.17 Varianzas definitivas.

Número de pregunta	Varianza
1	0.000129404
2	0.000083361
3	0.000134604
4	0.000135904
5	0.000137820

5.3 Comparación y Contraste de ambas Técnicas.

Uno de los objetivos de este trabajo de investigación es comprobar la eficacia de la Técnica de “Respuesta Aleatorizada”, por medio de la comparación y el contraste de las estimaciones obtenidas con el método de Encuesta Directa contra los de la “Aleatorizada”. Es por ello que presentamos a continuación una serie de tablas que nos muestran los resultados de ambas técnicas, tanto proporciones como varianzas estimadas de cada variable cualitativa.

Tabla 5. 18 Comparación de Técnicas. (Preguntas Cualitativas).

Preguntas sensibles	Proporción Encuesta directa	Proporción Respuesta aleatorizada	Varianza Encuesta Directa	Varianza Respuesta aleatorizada
1. ¿Sueles consumir alimentos o bebidas en la sala de cómputo?	0.32058824	0.52052	0.0006406	0.000129404
2. ¿Alguna vez has consumido drogas alucinógenas?	0.08235294	0.227826	0.0002222	0.000083361
3. ¿Algunas veces has copiado en los exámenes?	0.78235294	0.819767	0.0005008	0.000134604
4. ¿Has consumido bebidas alcohólicas, cerca a la Universidad?	0.53529412	0.616575	0.0007316	0.000135904
5. ¿En este ciclo, has faltado a clases sin motivo justificable?	0.62352941	0.685932	0.0006904	0.000137820

Fuente: Elaboración propia

podemos notar que no son muy grandes las diferencias en cuanto a los valores que tienen las proporciones de las preguntas 3 y 5, que son las referentes al copiar en los exámenes y faltar a clases injustificadamente.

también se aprecia fácilmente la diferencia entre las proporciones resultantes de ambas técnicas, ya que las obtenidas por medio de Encuesta Directa, son más pequeñas que las obtenidas por medio de método de “Respuestas Aleatorizadas”. Esto se traduce en que con la primera técnica, la población encuestada no acepta su participación en ciertas acciones etiquetadas y rechazadas por la sociedad en general.

A continuación se realiza cinco pruebas de hipótesis estadística en los cuales la hipótesis nula es que no existe diferencia significativa en las proporciones poblacionales de ambos métodos de encuesta, hipótesis que es rechazada en cada caso, concluyéndose que el parámetro del modelo de respuesta aleatorizada es significativamente mayor al nivel de significación de 0.05.

5.3.1 Prueba de hipótesis para la Comparación de Proporciones en “Consumo de alimentos y bebidas en el Centro de Cómputo”

1. Hipótesis

$H_0: \pi_R = \pi_D$: Los parámetros de “Consumo de alimentos y bebida en el Centro de computo” en ambos métodos son iguales

$H_1: \pi_R > \pi_D$ El parámetro del Modelo de Respuesta Aleatorizada es mayor que el Método directo.

2. Estadística de Prueba

$$P = \frac{n P_R + n P_D}{n_R + n_D}$$

$$Z_c = \frac{P_R - P_D}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}}$$

Bajo la hipótesis nula, Z_c tiene una distribución aproximadamente $N(0,1)$ a medida que n va aumentando.

3. Cálculo de la Estadística de Prueba

$$P_R = 0,52 \quad P_D = 0,32, \quad n_R = 404 \text{ y } n_D = 340$$

$$P = \frac{n_s P_R + n_c P_D}{n_s + n_c} = \frac{404(0,52) + 340(0,32)}{404 + 340} = 0,42$$

$$Z_c = \frac{P_R - P_D}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} = \frac{0,52 - 0,32}{\sqrt{0,42(1-0,42)(1/404 + 1/340)}} = 5,49$$

4. Regla de decisión

Para un nivel de significancia del 5%, $Z_{1-\alpha} = Z_{0,95} = 1,645$, luego como $Z_c = 5,49 > 1,645$, rechazamos H_0 , y concluimos que el parámetro de “Consumo de alimentos y bebidas en el Centro de computo” considerando el Modelo de Respuesta Aleatorizada es significativamente mayor que el Método Convencional o Directo, esto quiere decir que el Modelo de Respuesta Aleatorizada es más eficiente que el Método directo.

5.3.2 Prueba de hipótesis para la Comparación de Proporciones en “Consumo de drogas alucinógenas”

1. Hipótesis

$H_0 : \pi_R = \pi_D$: Los parámetros de “Consumo de drogas alucinógenas en ambos métodos son iguales

$H_1 : \pi_R > \pi_D$ El parámetro del Modelo de Respuesta Aleatorizada es mayor que el Método directo.

2. Estadística de Prueba

$$P = \frac{n P_R + n P_D}{n_R + n_D} \quad Z_c = \frac{P_R - P_D}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}}$$

Bajo la hipótesis nula, Z_c tiene una distribución aproximadamente $N(0,1)$ a medida que n va aumentando.

3. Cálculo de la Estadística de Prueba

$$P_R = 0,28 \quad P_D = 0,08, \quad n_R = 404 \quad \text{y} \quad n_D = 340$$

$$P = \frac{n_s P_R + n_c P_D}{n_s + n_c} = \frac{404(0,28) + 340(0,08)}{404 + 340} = 0,16$$

$$Z_c = \frac{P_R - P_D}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} = \frac{0,28 - 0,08}{\sqrt{0,16(1-0,16)(1/404 + 1/340)}} = 5,36$$

4. Regla de decisión

Para un nivel de significancia del 5%, $Z_{1-\alpha} = Z_{0,95} = 1,645$, luego como

$Z_c = 5,36 > 1,645$, rechazamos H_0 , y concluimos que el parámetro de “Consumo de drogas alucinógenas” considerando el Modelo de Respuesta Aleatorizada es significativamente mayor que el Método Convencional o Directo, esto quiere decir que el Modelo de Respuesta Aleatorizada es más eficiente que el Método directo.

5.3.3 Prueba de hipótesis para la Comparación de Proporciones en “Copia en los exámenes”

1. Hipótesis

$H_0 : \pi_R = \pi_D$: Los parámetros de “Copia en los exámenes” en ambos métodos son iguales

$H_1 : \pi_R > \pi_D$ El parámetro del Modelo de Respuesta Aleatorizada es mayor que el Método directo.

2. Estadística de Prueba

$$Z_c = \frac{P_R - P_D}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}}$$

$$P = \frac{n P_R + n P_D}{n_R + n_D}$$

Bajo la hipótesis nula, Z_c tiene una distribución aproximadamente $N(0,1)$ a medida que n va aumentando.

3. Cálculo de la Estadística de Prueba

$$P_R = 0,82 \quad P_D = 0,78, \quad n_R = 404 \quad \text{y} \quad n_D = 340$$

$$P = \frac{n_s P_R + n_c P_D}{n_s + n_c} = \frac{404(0,82) + 340(0,78)}{404 + 340} = 0,80$$

$$Z_c = \frac{P_R - P_D}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} = \frac{0,82 - 0,78}{\sqrt{0,80(1-0,80)(1/404 + 1/340)}} = 1,26$$

4. Regla de decisión

Para un nivel de significancia del 5%, $Z_{1-\alpha} = Z_{0,95} = 1,645$, luego como

$Z_c = 1,26 < 1,645$, no se rechaza H_0 , y concluimos que el parámetro de “Copia en los exámenes” considerando el Modelo de Respuesta Aleatorizada no es significativamente mayor que el Método Convencional o Directo. Esta pregunta no ha sido considerada como muy sensible por los estudiantes.

5.3.4 Prueba de hipótesis para la Comparación de Proporciones en “Consumo de bebidas alcohólicas cerca a la UNAC”

1. Hipótesis

$H_0 : \pi_R = \pi_D$: Los parámetros de “Consumo de bebidas alcohólicas cerca a la UNAC” en ambos métodos son iguales

$H_1 : \pi_R > \pi_D$ El parámetro del Modelo de Respuesta Aleatorizada es mayor que el Método directo.

2. Estadística de Prueba

$$Z_c = \frac{P_R - P_D}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}}$$

$$P = \frac{n_R P_R + n_D P_D}{n_R + n_D}$$

Bajo la hipótesis nula, Z_c tiene una distribución aproximadamente $N(0,1)$ a medida que n va aumentando.

3. Cálculo de la Estadística de Prueba

$$P_R = 0,62 \quad P_D = 0,53, \quad n_R = 404 \quad \text{y} \quad n_D = 340$$

$$P = \frac{n_s P_R + n_c P_D}{n_s + n_c} = \frac{404(0,62) + 340(0,53)}{404 + 340} = 0,58$$

$$Z_c = \frac{P_R - P_D}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} = \frac{0,52 - 0,32}{\sqrt{0,58(1-0,58)(1/404 + 1/340)}} = 2,23$$

4. Regla de decisión

Para un nivel de significancia del 5%, $Z_{1-\alpha} = Z_{0,95} = 1,645$, luego como

$Z_c = 2,23 > 1,645$, rechazamos H_0 , y concluimos que el parámetro de “Consumo de bebidas alcohólicas cerca a la UNAC” considerando el Modelo de Respuesta Aleatorizada es

significativamente mayor que el Método Convencional o Directo, esto quiere decir que el Modelo de Respuesta Aleatorizada es más eficiente que el Método directo.

5.3.5 Prueba de hipótesis para la Comparación de Proporciones en “Inasistencia a clases sin motivo justificable”

1. Hipótesis

$H_0 : \pi_R = \pi_D$: Los parámetros de “Inasistencia a clases sin motivo justificable” en ambos métodos son iguales

$H_1 : \pi_R > \pi_D$ El parámetro del Modelo de Respuesta Aleatorizada es mayor que el Método directo.

2. Estadística de Prueba

$$Z_c = \frac{P_R - P_D}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}}$$

$$P = \frac{n P_R + n P_D}{n_R + n_D}$$

Bajo la hipótesis nula, Z_c tiene una distribución aproximadamente $N(0,1)$ a medida que n va aumentando.

3. Cálculo de la Estadística de Prueba

$$P_R = 0,68 \quad P_D = 0,62, \quad n_R = 404 \quad \text{y} \quad n_D = 340$$

$$P = \frac{n_s P_R + n_c P_D}{n_s + n_c} = \frac{404(0,68) + 340(0,62)}{404 + 340} = 0,66$$

$$Z_c = \frac{P_R - P_D}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} = \frac{0,52 - 0,32}{\sqrt{0,66(1 - 0,66)(1/404 + 1/340)}} = 1,77$$

4. Regla de decisión

Para un nivel de significancia del 5%, $Z_{1-\alpha} = Z_{0,95} = 1,645$, luego como

$Z_c = 1,77 > 1,645$, rechazamos H_0 , y concluimos que el parámetro de “Inasistencia a clases sin motivo justificable” considerando el Modelo de Respuesta Aleatorizada es significativamente mayor que el Método Convencional o Directo, esto quiere decir que el Modelo de Respuesta Aleatorizada es más eficiente que el Método directo.

Por lo tanto, como puede observarse hasta este punto, se ha demostrado la hipótesis planteada que afirma que cuando se utiliza el método de Encuesta directa para medir variables sensibles, hay una tendencia a subestimarse el resultado y tener menor eficiencia en comparación al que se obtiene cuando se usa un método de Respuesta aleatorizada.

5.4 Respuestas Analizadas por Separado.

En la sección anterior se mencionaron de manera muy breve los resultados obtenidos en las encuestas, pero es necesario hacer una comparación más detallada de cada una de las preguntas.

El consumo de alimentos y bebidas en la sala de cómputo es el tema que se discute en la primera pregunta.

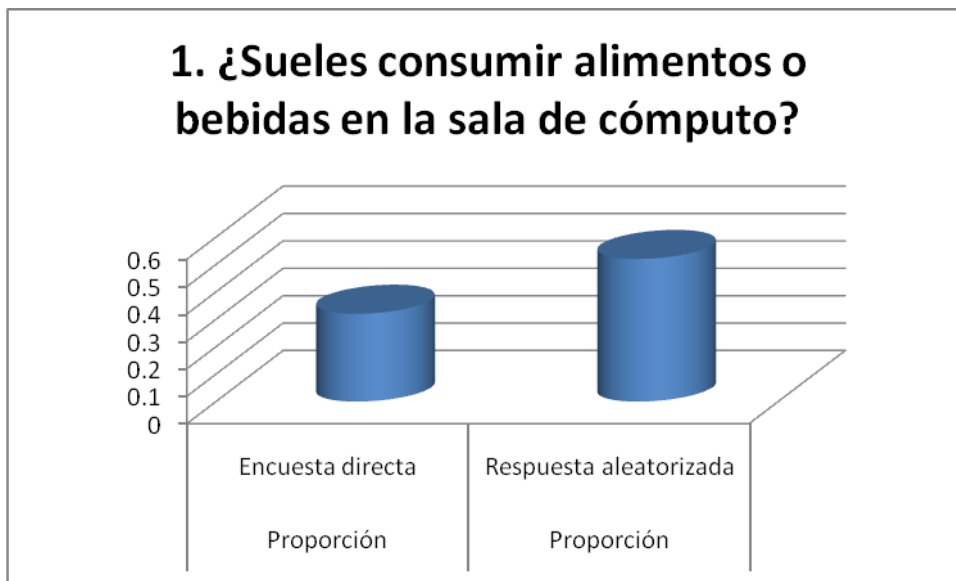


Gráfico 5.15. Elaboración propia.

Se observa que el 32% de los encuestados afirman que suelen consumir alimentos y bebidas en la sala de cómputo, a través de la encuesta directa, mientras que con la Respuesta aleatorizada este porcentaje sube significativamente a 52%. La varianza de la proporción es 0.0006406 con la encuesta directa, mientras que con la respuesta aleatorizada es 0.0001294.

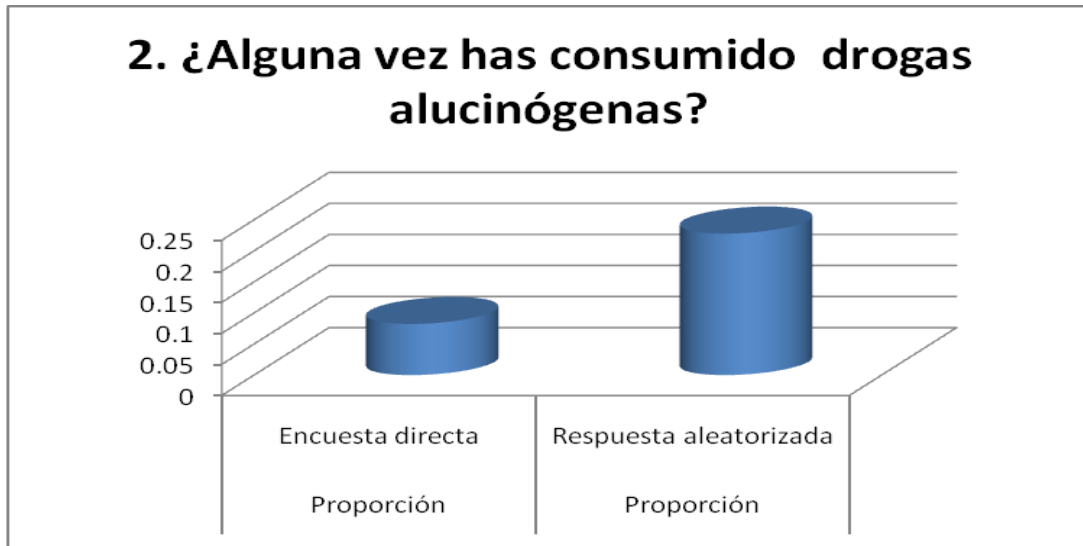


Gráfico 5.16. Elaboración propia.

Se observa que el 8.2% de los encuestados afirman que alguna vez han consumido drogas alucinógenas, a través de la encuesta directa, mientras que con la Respuesta aleatorizada este porcentaje sube significativamente a 22.7%. La varianza de la proporción es 0.000227 con la encuesta directa, mientras que con la respuesta aleatorizada es 0.0000833.

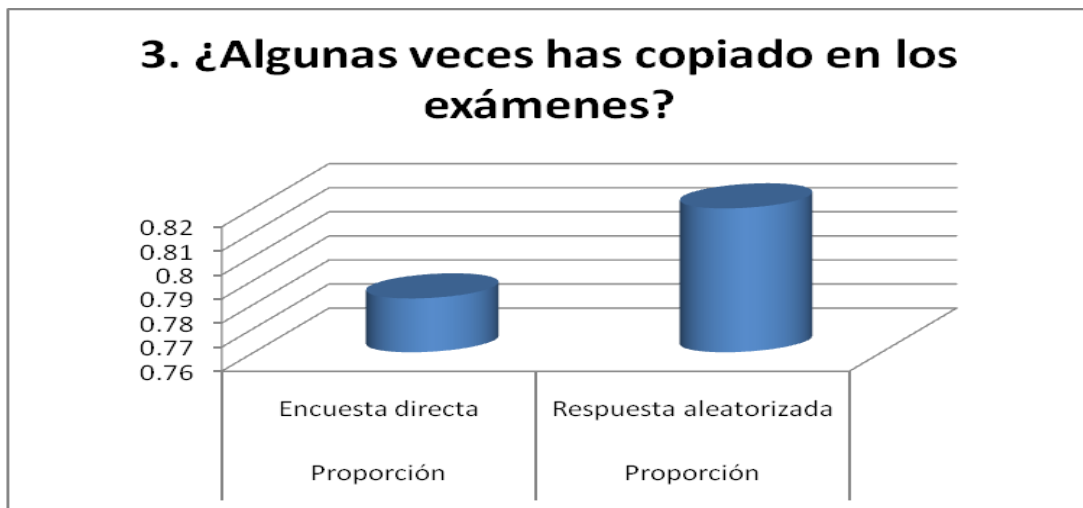


Gráfico 5.17. Elaboración propia.

Se observa que el 78.2% de los encuestados afirman que alguna vez han copiado en los exámenes, a través de la encuesta directa, mientras que con la Respuesta aleatorizada este porcentaje sube significativamente a 81.9%. La varianza de la proporción es 0.00073 con la encuesta directa, mientras que con la respuesta aleatorizada es 0.0001346.

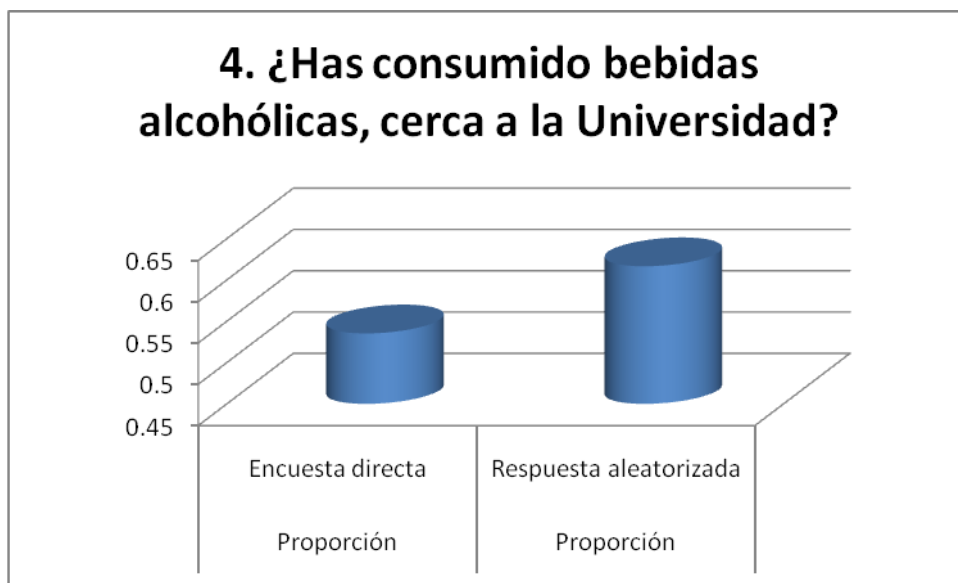


Gráfico 5.18. Elaboración propia.

Se observa que el 53.5% de los encuestados afirman que alguna vez han consumido bebidas alcohólicas, cerca a la universidad, a través de la encuesta directa, mientras que con la Respuesta aleatorizada este porcentaje sube significativamente a 61.6%. La varianza de la proporción es 0.00073 con la encuesta directa, mientras que con la respuesta aleatorizada es 0.000125.

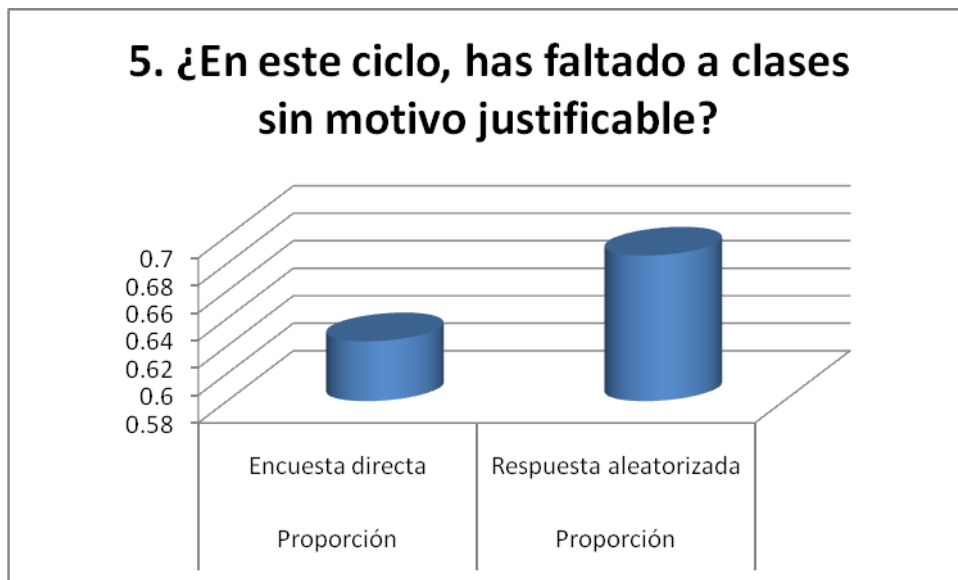


Gráfico 5.19. Elaboración propia.

Se observa que el 62.3% de los encuestados afirman que alguna vez han faltado a clases sin motivo justificable en el presente ciclo académico, a través de la encuesta directa, mientras que con la Respuesta aleatorizada este porcentaje sube significativamente a 68.5%. La varianza de la proporción es 0.0006904 con la encuesta directa, mientras que con la respuesta aleatorizada es 0.0001378.

Tabla 5.19 ¿Sueles consumir alimentos o bebidas en la sala de computo? * sexo

			sexo		Total
			Masculino	Femenino	
¿Sueles consumir alimentos o bebidas en la sala de computo?	SI	Recuento	61	48	109
		% de ¿Sueles consumir alimentos o bebidas en la sala de computo?	56.0%	44.0%	100.0%
	NO	Recuento	95	136	231
		% de ¿Sueles consumir alimentos o bebidas en la sala de computo?	41.1%	58.9%	100.0%
Total		Recuento	156	184	340
		% de ¿Sueles consumir alimentos o bebidas en la sala de computo?	45.9%	54.1%	100.0%

Como podemos observar en las proporciones de la encuesta directa, el 32.1% de los estudiantes suelen consumir alimentos y bebidas en la sala de computo. El 56.0% de los que suelen consumir alimentos y bebidas en la sala de computo son varones, mientras que el restante 44.0% son mujeres.

Tabla 5.20 ¿Alguna vez has consumido drogas alucinógenas? * sexo

			sexo		Total
			Masculino	Femenino	
¿Alguna vez has consumido drogas alucinógenas?	SI	Recuento	23	5	28
		% de ¿Alguna vez has consumido drogas alucinógenas?	82.1%	17.9%	100.0%
	NO	Recuento	133	179	312
		% de ¿Alguna vez has consumido drogas alucinógenas?	42.6%	57.4%	100.0%
Total		Recuento	156	184	340
		% de ¿Alguna vez has consumido drogas alucinógenas?	45.9%	54.1%	100.0%

Como podemos observar en las proporciones de la encuesta directa, el 8.2% de los estudiantes alguna vez han consumido drogas alucinógenas. El 82.1 % de los que alguna vez han consumido drogas alucinógenas son varones, mientras que el restante 17.9% son mujeres.

Tabla 5.21 ¿Algunas veces has copiado en los exámenes? * sexo

			sexo		Total
			Masculino	Femenino	
¿Algunas veces has copiado en los exámenes?	SI	Recuento	132	134	266
		% de ¿Algunas veces has copiado en los exámenes?	49.6%	50.4%	100.0%
	NO	Recuento	24	50	74
		% de ¿Algunas veces has copiado en los exámenes?	32.4%	67.6%	100.0%
Total		Recuento	156	184	340
		% de ¿Algunas veces has copiado en los exámenes?	45.9%	54.1%	100.0%

Como podemos observar en las proporciones de la encuesta directa, el 78.2% de los estudiantes algunas veces han copiado en los exámenes. El 49.6% de los que algunas veces han copiado en los exámenes son varones, mientras que el restante 50.4% son mujeres.

Tabla 5.22 ¿Has consumido bebidas alcohólicas, cerca a la Universidad? * sexo

			sexo		Total
			Masculino	Femenino	
¿Has consumido bebidas alcohólicas, cerca a la Universidad?	SI	Recuento	119	63	182
		% de ¿Has consumido bebidas alcohólicas, cerca a la Universidad?	65.4%	34.6%	100.0%
	NO	Recuento	37	121	158
		% de ¿Has consumido bebidas alcohólicas, cerca a la Universidad?	23.4%	76.6%	100.0%
Total		Recuento	156	184	340
		% de ¿Has consumido bebidas alcohólicas, cerca a la Universidad?	45.9%	54.1%	100.0%

Como podemos observar en las proporciones de la encuesta directa, el 53.5% de los estudiantes han consumido bebidas alcohólicas cerca a la universidad. El 65.4% de los que han consumido bebidas alcohólicas cerca a la universidad son varones, mientras que el restante 34.6% son mujeres.

Tabla 5.23 ¿En este ciclo, has faltado a clases sin motivo justificable? * sexo

			sexo		Total
			Masculino	Femenino	
¿En este ciclo, has faltado a clases sin motivo justificable?	SI	Recuento	112	100	212
		% de ¿En este ciclo, has faltado a clases sin motivo justificable?	52.8%	47.2%	100.0%
	NO	Recuento	44	84	128
		% de ¿En este ciclo, has faltado a clases sin motivo justificable?	34.4%	65.6%	100.0%
Total		Recuento	156	184	340
		% de ¿En este ciclo, has faltado a clases sin motivo justificable?	45.9%	54.1%	100.0%

Como podemos observar en las proporciones de la encuesta directa, el 62.3% de los estudiantes han faltado en este ciclo sin motivo justificable. El 52.8 % de los que han faltado en este ciclo sin motivo justificable son varones, mientras que el restante 47.2 % son mujeres.

5.5 Ventajas y Desventajas de cada Método de Investigación.

La principal ventaja que tiene la encuesta directa sobre la de Respuesta Aleatorizada es el tiempo que se invierte, especialmente para el encuestado, ya que cuando se levantan las encuestas, las personas no tardan más de 2 minutos en promedio en marcar sus respuestas, y depositar su encuesta en una sobre, lo que la hace muy práctica y rápida. Mientras que para contestar la encuesta “Aleatorizada” tuvieron que dedicarle por lo menos 1 minuto a leer y seguir las instrucciones, para después recién llenar la hoja de respuestas.

La desventaja en el método de Respuesta aleatorizada se presenta para los investigadores, porque hay que invertirle un tiempo mayor en el trabajo de campo, toda vez que requiere precisar al encuestado los pasos a seguir a efectos de la forma en que se va a registrar las respuesta.

Otra desventaja que tiene la encuesta Directa y que consideramos la más importante es que las personas se sienten invadidas en su privacidad por lo fuerte del contexto de las preguntas y en muchos casos las llenaron de puras respuestas negativas, confirmando de cierta forma que las preguntas al ser de estigma social no se pueden preguntar abiertamente.

La desventaja anterior de la encuesta Directa, representa la principal ventaja del método de Respuesta Aleatorizada, ya que los encuestados al no sentirse invadidos en su privacidad y tener un poco más de confianza en que no conocemos sus respuestas, no hay cuestionarios de puros ceros , claro que hay que recordar que son estimaciones, las cuales tienen cierto margen de error, pero que nos dan una idea más amplia de lo que sucede en la totalidad de la facultad de Ciencias económicas.

6. DISCUSIÓN

En esta última sección de la investigación se discuten las conclusiones a las que llegamos después de haber realizado los cálculos y así poder dar recomendaciones para las personas que más adelante quieran continuar con aplicaciones del método de Respuesta aleatorizada.

A tenor de los resultados obtenidos del análisis de las inferencias estadísticas, la primera de las notas que se puede apuntar es la observación que las proporciones estimadas de las preguntas sensibles de la encuesta directa, presenta valores siempre menores a los observados en la encuesta mediante la respuesta aleatorizada, de manera que en base en estos resultados puntuales, podría asumirse que existe una subestimación de los parámetros poblacionales.

Las encuestas mediante la Respuesta aleatorizada tienen más o menos el mismo comportamiento que las directas, pero podemos observar un incremento considerable de las proporciones llegando a la conclusión, a través de la inferencia estadística con un nivel de significación de cinco por ciento, que cuando se trata de temas sensibles, las proporciones de respuestas afirmativas son significativamente subestimadas

Una de las observaciones importantes en este trabajo es que el 68.56% de los alumnos manifiestan que han faltado a clases sin motivo justificado y esto se debe seguramente a que no hay un control efectivo de las inasistencias a clase por parte del docente u autoridad

correspondiente, lo cual incide también en el rendimiento académico del alumno. Aquí se

sugiere que la asistencia a clase debe formar parte de la evaluación de las asignaturas que se imparten en esta facultad.

Los resultados hallados en este trabajo son compatibles a los postulados por los autores revisados en el marco teórico, como Cochran (1997), quienes sintetizan de manera esquemática la importancia de la magnitud del error estándar de los estimadores de las proporciones poblacionales, en todo caso resultaron mucho menores con el método de respuesta aleatorizada, lo cual incide en intervalos de confianza más precisos.

Los resultados arrojados por la investigación, son de mucha importancia para analizar el cumplimiento de cada uno de los objetivos de la presente investigación. En primer lugar se observa que con el primer objetivo relacionado con los aportes de la teoría de las técnicas de Respuesta Aleatorizada para estudiar características sensibles, se han presentado las principales características y bases teóricas de ambas técnicas, y los posibles errores no muestrales a los que podíamos enfrentarnos en el levantamiento de encuestas. Principalmente se dieron a conocer los antecedentes y la evolución del método de Respuesta Aleatorizada, de manera que se ha cumplido con el primer objetivo específico.

En el siguiente objetivo se tenía como finalidad diseñar una muestra probabilística representativa de la población estudiantil de la facultad de ciencias económicas por medio del muestreo aleatorio. Este punto fue parte importante de la investigación, ya que se obtuvo un 95% de confianza de tener el número necesario de estudiantes dentro de la muestra a efectos de lograr inferencias confiables. Por consiguiente se considera cumplido el segundo

objetivo.

Con respecto a la estimación de parámetros poblacionales a partir de encuestas sobre temas sensibles o de estigma social, se planteó la hipótesis que cuando se utiliza el método de Encuesta directa para medir variables sensibles, hay una tendencia a subestimarse el resultado en comparación al obtenido cuando se usa el método de Respuesta aleatorizada, afirmación que ha sido comprobada a partir de los resultados hallados. Se pudo observar que hubo un mayor grado de respuesta por parte de los estudiantes cuando les fue aplicado el método aleatorizado, ya que tenían mayor confianza en la confidencialidad de sus respuestas.

Es importante mencionar que el método de Respuesta Aleatorizada proporciona estimaciones con errores típicos pequeños, por lo que si bien no puede afirmarse que los resultados obtenidos resulten 100% verdaderos, son muy cercanos a la realidad.

Es importante mencionar que para el levantamiento de encuestas por el método de “Respuesta Aleatorizada”, es necesario contar con un número grande de personas que se involucren en el trabajo de campo, ya que en cada encuesta se invierten de 10 a 12 minutos. Si consideramos un número grande de muestra, la inversión y total de tiempo es muy elevada.

De igual manera es importante capacitar a cada encuestador, y la inversión monetaria

puede incrementarse de manera considerable; sin embargo consideremos que al capturar los datos en una computadora la depuración de resultados es prácticamente innecesaria, además de que el análisis se facilita de manera considerable, reduciendo al máximo el tiempo de entrega de resultados.

Para terminar, un hecho palpable es que en la actualidad, existen algunos aspectos que están favoreciendo el uso y difusión de los métodos de muestreo aplicado a la investigación en las Ciencias sociales en general., tales como la generalización del uso de los medios informáticos, STATA, SAS, SPSS y otros que han facilitado el uso de técnicas y métodos de análisis sofisticados. También mencionar que la difusión de nuevas tecnologías, aunado a la disponibilidad de información estadística, han motivado que los investigadores basen sus decisiones en estudios cuantitativos bien fundamentados, lo que ha provocado un incremento en el uso del muestreo aplicado.

7. REFERENCIALES

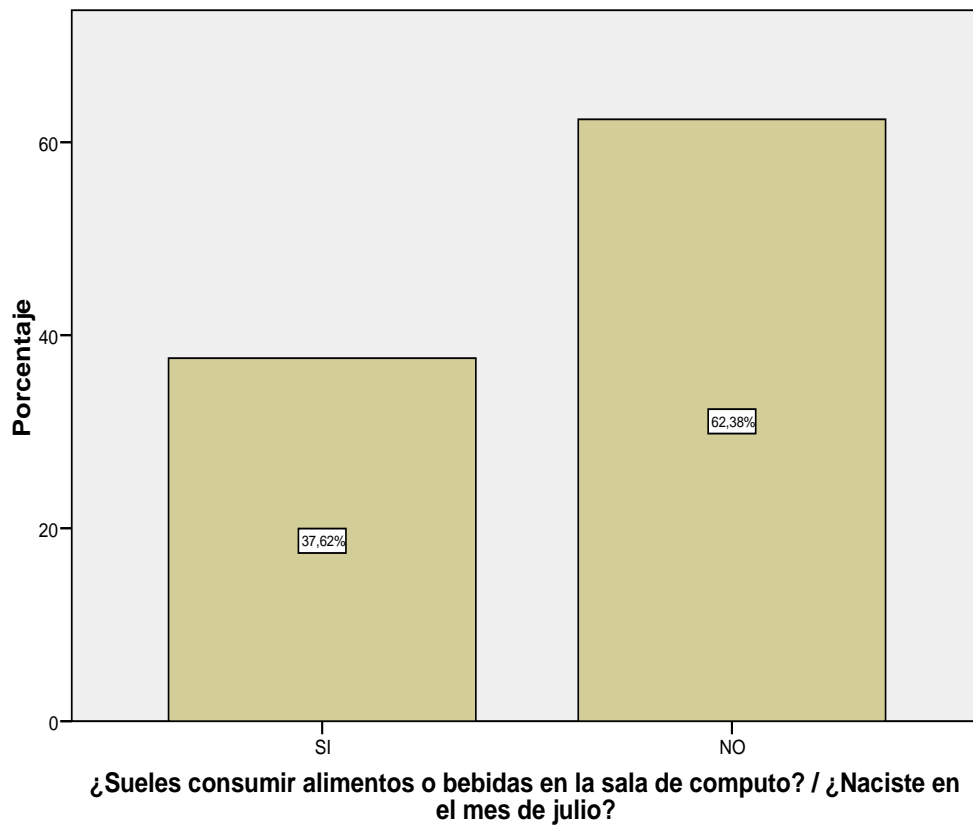
1. CAMBELL, C. and JOINER, B. how to get the answer without being sure you've asked the question. *The American Statistician*, diciembre, 1973, 229-231
2. CHAUDHIRI, A; MUKERJEE, R. *Randomized response: theory and techniques*. New York: Marcel, 1983.161p.
3. COCHRAN, G.W. *Técnicas de muestreo*: México: Continental, 1997. 513p.
4. DES, R. *Teoría del muestreo*. México: Fondo de Cultura Económica, 1980. 305p.
5. GREENBERG, B. G. et al. The unrelated question randomized response model: theoretical frame work. *J. Am. Stat. Assoc.*, Alexandria, v.64, p. 520 -539, 1969.
6. GREENBERG, B. G. et al. Aplication on the randomized response technique in obtaining quantitative data. *J. Am. Stat. Assoc.*, Alexandria, v. 71, n.353-, p72-73, p.980-83,1975
7. FOX, J.A: ; Tracy, P.E. *Randomized Response: A Method for sensitive Surveys*. Sage University Paper on Quantitative Applications in the Social Sciencies, 58. Beverly Hills: Sage Publications, 1986.
8. KISH, Leslie. *Muestreo de encuestas*, 3ª reimp, México, Trillas, 1982, 736 p.
9. LANKE, S. On the choice of the unrelated question in Simmons version of randomized response model. *J. Stat. Assoc.*, New York, v.70, p.80-83,1975.
10. MANN, C.R. Tell me your secret - random response methodology. *Advertising y marketing research professionals*, jun. 1999.
11. MENDEZ, I; Eslava, G. y P. Romero (2004)/ “Conceptos Básicos de Muestreo”; *Monografías; Vol 12; No.27*. IIMAS, UNAM

12. MOORS, J.J. A Optimization of the unrelated question randomized response model.
J. Stat. Assoc., Mew York, v. 66, n. 361, p- 627-629, 1971.
13. MUSCH, J.BRODER, A. and KLAUER, K.C. Improving Survey Research on the
World-Wide Web using the Randomized Response Technique Dimensions of Internet
Science, 2001.
- 14 .SHIMIZU, I.M.; BONHAM, G. S. Randomized response technique in a national
survey. J. Am. Stat. Assoc. , New York, v. 66, n. 361, p. 627-629, 1971.
15. WARNER, S.L. Randomized response: A survey technique for elimination evasive
answer bias. J. Stat. Assoc., New York, v. 60, p.63-69, 1965.

8. APÉNDICE

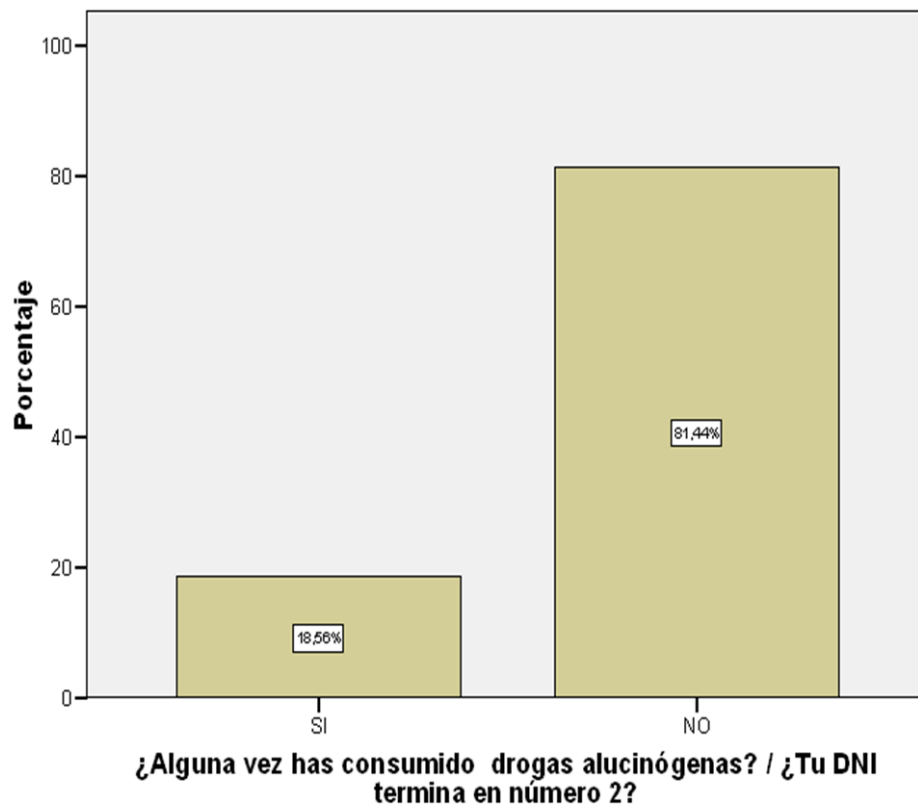
Contenido: Elaboración propia del autor

8.1 Apéndice 1



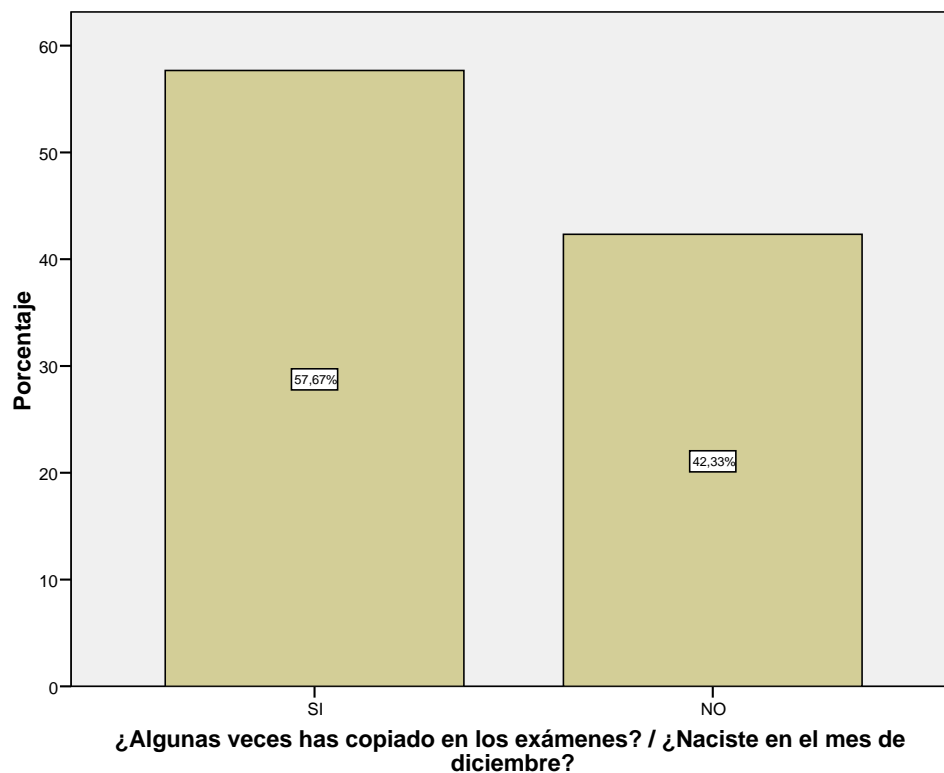
Fuente: Elaboración propia

8.2 Apéndice 2



Fuente: Elaboración propia

8.3 Apéndice 3



Fuente: Elaboración propia

8.4 Apéndice 4

Distribuciones discretas útiles en el muestreo

. Distribución Binaria o de Bernoulli

Se denomina así a la distribución de probabilidad cuya variable aleatoria sólo toma dos valores: el 0 y el 1 con probabilidades q y p , donde evidentemente $p+q = 1$.

Esta distribución modeliza adecuadamente los experimentos con dos resultados presencia o no de un atributo, éxito o fracaso, pieza correcta o incorrecta, procedimiento administrativo-contable adecuado o inadecuado, control que se cumple o se incumple etc.

Su función de cuantía $P(x)$ y de distribución respectivamente son:

$$P(x) = \begin{cases} q & x=0 \\ p & x=1 \\ 0 & \text{demás casos} \end{cases}$$

. Distribución Binomial

Se denomina distribución binomial de parámetros n y p y se denota por $B(n,p)$ a la distribución de probabilidad de la v.a. X con función de cuantía.

$$P(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & x=0, 1, 2, \dots, n \\ 0 & \text{demás casos} \end{cases}$$

Sus parámetros más importantes, media y varianza tienen por expresión:

$$\begin{aligned} \mu &= E(X) = n \cdot p \\ V(X) &= \sigma^2 = n \cdot p \cdot q \end{aligned}$$

Empíricamente hablando la distribución binomial se genera obtener una muestra aleatoria con reposición de una distribución binaria, corresponda a una población empírica tanto finita como infinita.

Distribución Hipergeométrica

Se define por distribución hipergeométrica de parámetros (N,n,p) y se denota $H(N,n,p)$ la distribución de la v.a. X , cuya función de cuantía es:

$$P(x) = \begin{cases} \frac{\binom{Np}{x} \binom{N(1-p)}{n-x}}{\binom{N}{n}} & \text{para } x = 0, 1, 2, \dots, n \\ 0 & \text{demás casos} \end{cases}$$

Se demuestra que:

$$\mu = np, y$$
$$V(X) = \frac{N-n}{N-1} \cdot npq$$

Esta distribución, análogamente a la distribución binomial, se genera a partir de una población binaria cuando el muestreo es sin reposición, por lo tanto ya no se tratará de pruebas idénticas e independientes. Cuando N toma un valor muy elevado, en general mayor a 10.000, se demuestra que las probabilidades de la distribución hipergeométrica pueden aproximarse por las de la Binomial.

Distribución de Poisson

Es la distribución de v.a. X, con función de cuantía:

$$P(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & x = 0, 1, 2, \dots \\ 0 & \text{demás casos} \end{cases}$$

siendo el parámetro λ . la media como la varianza de la misma.

La generación de esta distribución, por el correspondiente proceso de Poisson, es bastante más complejo que un proceso de pruebas repetidas con reposición (binomial) o sin reposición (hipergeométrica) y hay que proceder mediante ecuaciones diferenciales y/o en diferencias.

Esta distribución también se utiliza en Auditoria, ya que se parte del principio de que el número de errores debe ser muy pequeño con respecto al total de documentos, y por lo tanto la probabilidad de encontrar algún error en una muestra tiende a ser muy baja.

8.5 Apéndice 5

Algunos teoremas de interés

. Teorema de adición de distribuciones normales. Un teorema muy importante que se verifica cuando las distribuciones son normales es que la suma o convolución de un número finito n de distribuciones normales $X_i = N(\mu_i, \sigma_i^2)$ para $i = 1, 2, 3, \dots, n$ independientes o deducidas de la misma distribución multinormal, es otra distribución Normal X de media suma de las medias y varianza suma de las varianzas, esto es

$$X = \sum_{i=1}^n X_i \rightarrow N\left(\sum_{i=1}^n \mu_i; \sum_{i=1}^n \sigma_i^2\right)$$

Este teorema es importantísimo, pero requiere que las distribuciones convolucionadas sean normales. Además, como la combinación lineal de normales también es normal y la media muestral, cuando el muestreo es con reposición es una combinación lineal de n distribuciones normales, permite fácilmente obtener, bajo dichos supuestos, la distribución en el muestreo de la media muestral.

Cuando la población no es Normal, y eso es el caso más usual en Auditoría de Estados Financieros, no puede asegurarse lo anterior y por ello hay que recurrir a otros teoremas y requisitos como son el Teorema Central de límite y el Teorema de Tchebychev.

El Teorema Central de Límite (TCL) indica las condiciones necesarias y suficientes para la convergencia de una sucesión de variables aleatorias, no necesariamente normales, a la distribución Normal. Este teorema ha sido formulado bajo diferentes requerimientos. Uno

de ellos implica que la sucesión de variables aleatorias que se convolucionan son todas iguales (y por lo tanto con la misma media y varianza) e independientes, cuyo caso pionero es cuando todas ellas son binomiales y fue demostrado por De Moivre.

Puede demostrarse la convergencia de una suma o convolución de n variables a la distribución Normal si se verifican las condiciones aquí indicadas y el número de sumandos es relativamente grande, en general mayor o igual que 30.

8.6 Apéndice 6

Deducción de las expresiones de la media y la varianza de la media muestral de una muestra de tamaño n obtenida de una población finita de tamaño N

1. Si de una población finita U sobre la que se ha definido una variable X, se obtiene una muestra aleatoria de tamaño n, con o sin reposición, sobre dicha muestra se pueden calcular diferentes estadísticos muestrales como la media muestral, varianza muestral, desviación típica muestral, cuasidesviación típica muestral, etc. En este epígrafe vamos a deducir las expresiones matemáticas de la media y varianza del estadístico media muestral:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

y teniendo en cuenta las propiedades de la esperanza matemática aludidas en A.1.2, fácilmente se deduce que:

$$E(\bar{x}) = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{n\mu}{n} = \mu$$

coincide con la propia media poblacional, independientemente de que el muestreo sea con o sin reposición. En consecuencia el estadístico media muestral \bar{x} es un estimador insesgado de la media poblacional μ .

2. Así como la deducción de $E(\bar{x})$ ha sido casi inmediata, la de $V(\bar{x})$ es mas compleja. Recordando las propiedades de la varianza de una suma y del producto de una constante por una variable, se sigue que:

$$V(\bar{x}) = V\left[\frac{1}{n} \sum x_i\right] = \frac{1}{n^2} V\left(\sum x_i\right) = \frac{1}{n^2} \left[\sum_{i=1}^n V(x_i) + \sum_{i \neq j}^n Cov(x_i, x_j) \right]$$

luego para el cálculo de se precisa el previo de las $Cov(x_i, x_j)$, o covarianza de cualquier par de distribuciones marginales bidimensionales (X_i, X_j) asociadas al muestreo.

En el caso de muestreo con reposición, las distribuciones marginales son independientes y por lo tanto las covarianzas son nulas, y consecuentemente la correspondiente varianza de la media muestral será:

$$V(\bar{x}) = \frac{1}{n^2} \left[\sum_{i=1}^n V(x_i) + \sum_{i \neq j}^n Cov(x_i, x_j) \right] = \frac{1}{n^2} V\left(\sum x_i\right) = \frac{\sigma^2}{n}$$

luego la varianza de la media muestra] de una muestra con reposición de tamaño n, obtenida de una población finita, coincide con la correspondiente a una muestra aleatoria simple obtenida de una población infinita o distribución de probabilidad, como puede verse en cualquier curso de Estadística General (véase por ejemplo Escuder y Murgui (1995)).

8.7 Apéndice 7

Deducción de la expresión del tamaño muestral para la estimación de la media y del total de una población finita

Se obtuvo que para cualquier distribución normal, se verifica:

$$P\left(X - z_{\frac{\alpha}{2}}\sigma_x, X + z_{\frac{\alpha}{2}}\sigma_x\right) = 1 - \alpha$$

Por lo tanto, dado que en Auditoria las poblaciones en general no pueden Considerarse normales, admitiendo que el estadístico media muestral \bar{X} , correspondiente a una muestra aleatoria sin reposición de una población finita de tamaño X , puede modelizarse aproximadamente por una distribución normal por una aplicación "débil" del TCL (véase A.1.4, ya que dicho teorema requiere independencia), teniendo en cuenta que, como acabamos de demostrar (A.2.1 y A.2.2) su media y varianza son respectivamente:

$$E(\bar{X}) = \mu$$

$$V(\bar{X}) = \sigma_{\frac{2}{X}}^2 = \frac{N-n}{N-1} \frac{\sigma^2}{N} = (1-F) \frac{\sigma^2}{n}$$

Puede asumirse que la expresión:

$$\begin{aligned} & \Pr\left(\bar{X} - z_{\frac{\alpha}{2}}\sigma_x < \mu < \bar{X} + z_{\frac{\alpha}{2}}\sigma_x\right) = \\ & = \Pr\left(N\bar{X} - Nz_{\frac{\alpha}{2}}\sqrt{\frac{N-n}{N-1}}\frac{\sigma^2}{n} < N\mu < N\bar{X} + Nz_{\frac{\alpha}{2}}\sqrt{\frac{N-n}{N-1}}\frac{\sigma^2}{n}\right) \end{aligned}$$

tendrá una probabilidad aproximada de $1 - \alpha$. Llamando error o precisión total PT de la estimación a la semiamplitud de dicho intervalo, esto es:

$$PT = N z_{\frac{\alpha}{2}} \sqrt{\frac{N - n}{N - 1} \frac{\sigma^2}{n}}$$

y denotando por E a la precisión unitaria o cociente entre PT y N, esto es $E = PT/N$, elevando al cuadrado tendremos que:

$$E^2 = z_{\frac{\alpha}{2}}^2 \left(\frac{N - n}{N - 1} \right) \frac{\sigma^2}{n}$$

y haciendo operaciones se sigue que:

$$\begin{aligned} E^2(N - 1)n &= z_{\alpha/2}^2(N - n)\sigma^2 = z_{\alpha/2}^2 N \sigma^2 - z_{\alpha/2}^2 n \sigma^2 \\ n[E^2(N - 1) + z_{\alpha/2}^2 \sigma^2] &= z_{\alpha/2}^2 N \sigma^2 \end{aligned}$$

de donde despejando n se obtiene la fórmula para el cálculo del tamaño muestral que es:

$$n = \frac{1}{\frac{1}{z_{\alpha/2}^2} \left(\frac{E}{\sigma} \right)^2 + \frac{1}{N}}$$

fórmula que en general se toma para el cálculo del tamaño muestral de una prueba sustantiva.

8.8 Apéndice 8

Distribuciones empíricas o de frecuencias

1. Sea una población U de N elementos o documentos cualesquiera: u_1, u_2, \dots, u_N ; sobre los que hemos definido una transformación en R o variable X que representa a los diferentes importantes monetarios x_i $i=1,2,\dots,N$; de cada documento u_i , de la que se extrae una muestra aleatoria de tamaño n.

Sus medidas de posición y dispersión o parámetros fundamentales son: la media, la varianza y desviación típica, que denotaremos por μ, σ^2 y σ respectivamente, y se definen mediante:

$$\mu = \sum_{i=1}^N x_i \frac{1}{N}, \quad \sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}, \quad \text{y} \quad \sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

En el caso de un población finita, además de la varianza y desviación típica, tiene sentido definir como medidas de dispersión las llamadas cuasivarianza y cuasidesviación típica, que son:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N-1} = \frac{N}{N-1} \sigma^2, \quad \text{y} \quad \sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N-1}} = \sqrt{\frac{N}{N-1}} \sigma$$

En la varianza se verifica la siguiente descomposición:

$$\sigma^2 = \sum_{i=1}^N (x_i^2 - 2\mu x_i + \mu^2) \frac{1}{N} = \sum_{i=1}^N x_i^2 \frac{1}{N} - 2\mu^2 + \mu^2 = \sum_{i=1}^N x_i^2 \frac{1}{N} - \mu^2$$

2. Si sobre la población finita se consideraran a la vez varias variables tendríamos una distribución multidimensional. De las distribuciones multidimensionales se pueden deducir distribuciones marginales y Condicionadas, a las que no vamos a referirnos en este apéndice, aunque Si introduciremos el concepto de covarianza (véase A. 1.2), pues dicho concepto se necesita para desarrollos que demostraremos en el apéndice A.2.

3. Al obtener muestras de tamaño n , las medidas anteriores se denominan estadísticos muestrales, y se definen y denotan mediante:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}; \quad s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}; \quad s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{n}{n-1} s^2; \quad y \quad s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{n}{n-1}} s$$

8.8 Apéndice 8

Distribuciones de probabilidad. Características generales

Dos son las funciones matemáticas más importantes que pueden definirse sobre una distribución discreta. Una es $P(x)$ o probabilidad de que la variable aleatoria X tome exactamente el valor x , a la que se denomina función de cuantía, que sería el correlato formal de la función de frecuencias relativas de una distribución empírica; y la otra, denotada por $F(x)$, que expresa la probabilidad total acumulada hasta el punto x (incluido), a la que se denomina función de distribución y es el correlato de la función de frecuencias acumuladas. En definitiva, si simbolizamos probabilidad por Pr , las dos funciones anteriores son respectivamente:

$$P(x) = \Pr(X = x) \quad F(x) = \Pr(X \leq x)$$

Los parámetros más importantes de toda distribución de probabilidad son la media o esperanza matemática $E(X)$, que se define como suma, o integral en el caso continuo, de todos los posibles valores de la variable por sus probabilidades respectivas; y la varianza $V(X)$ como suma (o integral) de las diferencias entre cada valor de la variable y su media elevadas al cuadrado y multiplicadas por sus respectivas probabilidades, esto es

$$E(X) = \mu = \sum x \Pr(x) \quad \text{y} \quad V(X) = \sigma^2 = \sum (x - \mu)^2 \Pr(x)$$

En el caso de variable continua, el sumatorio se transforma en una integral y la función de cuantía en la función de densidad, esto es

$$E(X) = \mu = \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx \quad \text{y} \quad V(X) = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x) \cdot dx$$

ANEXOS

ANEXO 1

Base de datos. Encuesta Directa(a)

	ficha	sexo	¿Sueles consumir alimentos o bebidas en la sala de cómputo?	¿Alguna vez has consumido drogas alucinógenas?	¿Algunas veces has copiado en los exámenes?	¿Has consumido bebidas alcohólicas, cerca a la Universidad?	¿En este ciclo, has faltado a clases sin motivo justificable?
1	1	Femenino	SI	NO	SI	NO	SI
2	2	Femenino	NO	NO	SI	NO	SI
3	3	Masculino	SI	NO	SI	SI	SI
4	4	Masculino	SI	NO	SI	SI	NO
5	5	Masculino	NO	NO	SI	SI	NO
6	6	Femenino	SI	NO	SI	SI	SI
7	7	Masculino	NO	NO	SI	NO	NO
8	8	Femenino	SI	NO	SI	NO	NO
9	9	Masculino	SI	NO	SI	SI	NO
10	10	Masculino	SI	SI	SI	NO	NO
11	11	Masculino	NO	NO	SI	SI	NO
12	12	Femenino	SI	NO	NO	NO	NO
13	13	Femenino	SI	NO	SI	NO	NO
14	14	Masculino	NO	NO	SI	NO	NO
15	15	Masculino	NO	NO	SI	SI	NO
16	16	Masculino	SI	NO	NO	SI	SI
17	17	Masculino	NO	NO	SI	SI	NO
18	18	Femenino	SI	NO	SI	SI	NO
19	19	Masculino	NO	NO	SI	SI	SI
20	20	Masculino	NO	SI	SI	NO	NO
21	21	Masculino	NO	NO	SI	NO	NO
22	22	Masculino	NO	NO	SI	SI	SI
23	23	Femenino	NO	NO	SI	NO	SI
24	24	Masculino	NO	NO	NO	SI	SI
25	25	Femenino	NO	NO	NO	SI	NO
26	26	Femenino	NO	NO	SI	NO	SI
27	27	Femenino	NO	NO	NO	NO	SI
28	28	Femenino	NO	NO	SI	NO	NO
29	29	Femenino	NO	NO	SI	NO	SI
30	30	Femenino	NO	NO	SI	SI	NO

31	31	Femenino	SI	NO	SI	NO	SI
32	32	Masculino	SI	NO	SI	SI	SI
33	33	Masculino	SI	NO	SI	SI	SI
34	34	Femenino	NO	NO	SI	NO	NO
35	35	Femenino	NO	NO	SI	SI	SI
36	36	Femenino	NO	NO	SI	NO	SI
37	37	Masculino	NO	NO	NO	SI	NO
38	38	Femenino	NO	NO	SI	NO	NO
39	39	Masculino	SI	NO	NO	NO	NO
40	40	Masculino	SI	NO	SI	SI	SI
41	41	Masculino	NO	NO	SI	SI	SI
42	42	Masculino	NO	NO	SI	SI	SI
43	43	Masculino	SI	NO	SI	NO	NO
44	44	Masculino	NO	NO	SI	SI	SI
45	45	Masculino	NO	SI	SI	SI	SI
46	46	Masculino	NO	NO	SI	SI	SI
47	47	Masculino	NO	NO	SI	NO	SI
48	48	Femenino	NO	NO	SI	NO	NO
49	49	Femenino	NO	NO	NO	SI	NO
50	50	Masculino	SI	NO	SI	SI	NO
51	51	Masculino	NO	NO	SI	NO	SI
52	52	Femenino	SI	NO	SI	NO	SI
53	53	Femenino	NO	NO	NO	NO	NO
54	54	Femenino	NO	NO	NO	NO	SI
55	55	Femenino	NO	NO	NO	NO	NO
56	56	Femenino	NO	NO	NO	NO	SI
57	57	Masculino	SI	NO	NO	NO	SI
58	58	Femenino	NO	NO	NO	NO	NO
59	59	Masculino	SI	NO	NO	NO	SI
60	60	Femenino	NO	NO	NO	NO	SI
61	61	Masculino	NO	NO	SI	SI	SI
62	62	Femenino	NO	NO	NO	NO	NO
63	63	Femenino	SI	NO	SI	NO	NO
64	64	Femenino	SI	NO	SI	SI	NO
65	65	Femenino	SI	NO	SI	NO	SI
66	66	Femenino	NO	NO	SI	NO	NO
67	67	Femenino	SI	NO	SI	NO	SI
68	68	Femenino	SI	NO	SI	SI	NO
69	69	Femenino	NO	NO	SI	NO	SI
70	70	Femenino	NO	NO	SI	SI	SI
71	71	Femenino	SI	NO	SI	NO	NO
72	72	Masculino	SI	NO	SI	SI	SI
73	73	Masculino	SI	NO	SI	SI	SI
74	74	Masculino	NO	NO	SI	SI	NO
75	75	Femenino	NO	NO	NO	NO	NO
76	76	Femenino	SI	NO	SI	NO	SI
77	77	Femenino	NO	NO	SI	NO	SI
78	78	Femenino	NO	NO	SI	NO	SI

79	79	Femenino	SI	NO	SI	NO	NO
80	80	Femenino	NO	NO	SI	NO	SI
81	81	Masculino	NO	NO	SI	SI	NO
82	82	Masculino	NO	NO	SI	SI	SI
83	83	Masculino	SI	NO	SI	SI	SI
84	84	Femenino	NO	SI	SI	SI	NO
85	85	Masculino	SI	SI	SI	SI	SI
86	86	Masculino	SI	NO	SI	SI	SI
87	87	Masculino	NO	NO	SI	SI	SI
88	88	Masculino	NO	NO	SI	SI	SI
89	89	Femenino	NO	NO	SI	NO	NO
90	90	Masculino	NO	NO	SI	SI	SI
91	91	Masculino	NO	NO	SI	SI	SI
92	92	Femenino	SI	NO	SI	SI	SI
93	93	Masculino	SI	NO	SI	SI	SI
94	94	Femenino	NO	NO	SI	NO	SI
95	95	Masculino	SI	SI	SI	SI	SI
96	96	Femenino	SI	NO	SI	SI	SI
97	97	Femenino	NO	NO	NO	NO	NO
98	98	Masculino	NO	SI	SI	SI	SI
99	99	Femenino	NO	NO	SI	SI	SI
100	100	Masculino	NO	NO	SI	SI	SI
101	101	Femenino	NO	NO	NO	SI	NO
102	102	Femenino	NO	NO	SI	NO	NO
103	103	Femenino	NO	NO	NO	NO	NO
104	104	Masculino	SI	NO	SI	NO	SI
105	105	Femenino	NO	NO	SI	NO	NO
106	106	Femenino	NO	NO	NO	NO	SI
107	107	Femenino	NO	NO	SI	NO	NO
108	108	Masculino	SI	NO	NO	SI	NO
109	109	Femenino	SI	NO	SI	NO	NO
110	110	Masculino	NO	NO	NO	NO	SI
111	111	Masculino	NO	NO	SI	SI	NO
112	112	Masculino	NO	NO	SI	SI	SI
113	113	Masculino	NO	NO	SI	SI	NO
114	114	Masculino	NO	NO	SI	SI	SI
115	115	Masculino	NO	SI	SI	SI	SI
116	116	Masculino	NO	NO	SI	NO	SI
117	117	Femenino	NO	NO	SI	SI	SI
118	118	Masculino	NO	NO	SI	SI	SI
119	119	Masculino	NO	NO	SI	SI	SI
120	120	Masculino	SI	NO	SI	SI	SI
121	121	Masculino	NO	NO	SI	SI	SI
122	122	Masculino	SI	NO	SI	SI	SI
123	123	Masculino	SI	NO	SI	SI	SI
124	124	Femenino	NO	NO	SI	SI	SI
125	125	Masculino	SI	NO	NO	SI	SI
126	126	Femenino	NO	NO	NO	NO	NO
127	127	Masculino	SI	NO	SI	SI	SI

128	128	Femenino	NO	NO	NO	NO	SI
129	129	Femenino	SI	NO	NO	SI	SI
130	130	Masculino	NO	NO	SI	SI	SI
131	131	Masculino	NO	NO	NO	SI	NO
132	132	Femenino	NO	NO	NO	NO	NO
133	133	Masculino	NO	NO	NO	SI	NO
134	134	Femenino	NO	NO	SI	SI	SI
135	135	Masculino	NO	SI	SI	SI	SI
136	136	Masculino	NO	NO	SI	NO	SI
137	137	Femenino	NO	NO	NO	NO	NO
138	138	Femenino	NO	NO	SI	NO	NO
139	139	Masculino	NO	NO	SI	SI	NO
140	140	Masculino	NO	NO	SI	SI	SI
141	141	Femenino	NO	NO	SI	SI	NO
142	142	Femenino	NO	NO	SI	NO	NO
143	143	Femenino	SI	NO	SI	SI	SI
144	144	Femenino	NO	NO	SI	SI	SI
145	145	Masculino	NO	NO	SI	SI	SI
146	146	Masculino	SI	SI	SI	SI	SI
147	147	Femenino	NO	NO	SI	SI	SI
148	148	Masculino	NO	NO	SI	SI	SI
149	149	Femenino	NO	NO	NO	NO	SI
150	150	Masculino	NO	NO	SI	NO	SI
151	151	Masculino	SI	NO	SI	SI	SI
152	152	Masculino	SI	NO	SI	SI	SI
153	153	Femenino	NO	NO	SI	NO	SI
154	154	Femenino	NO	NO	SI	NO	NO
155	155	Masculino	SI	NO	NO	NO	NO
156	156	Masculino	NO	NO	SI	NO	SI
157	157	Femenino	NO	NO	SI	NO	SI
158	158	Femenino	NO	NO	NO	NO	SI
159	159	Femenino	NO	NO	SI	NO	NO
160	160	Masculino	SI	NO	SI	SI	SI
161	161	Femenino	SI	NO	SI	NO	SI
162	162	Femenino	NO	NO	SI	SI	SI
163	163	Masculino	SI	NO	SI	SI	SI
164	164	Femenino	SI	NO	SI	NO	NO
165	165	Femenino	NO	NO	SI	NO	NO
166	166	Femenino	SI	NO	SI	NO	SI
167	167	Femenino	NO	NO	SI	NO	SI
168	168	Femenino	NO	NO	SI	NO	SI
169	169	Femenino	NO	NO	SI	NO	SI
170	170	Masculino	NO	NO	SI	SI	SI
171	171	Femenino	NO	NO	SI	SI	NO
172	172	Femenino	NO	NO	NO	NO	SI
173	173	Masculino	SI	NO	SI	NO	SI
174	174	Masculino	NO	SI	SI	SI	SI
175	175	Masculino	SI	NO	SI	SI	SI
176	176	Masculino	SI	NO	SI	SI	SI

177	177	Femenino	SI	NO	SI	NO	SI
178	178	Masculino	SI	NO	SI	SI	SI
179	179	Masculino	NO	NO	SI	SI	SI
180	180	Masculino	NO	SI	SI	SI	SI
181	181	Masculino	NO	NO	SI	SI	SI
182	182	Femenino	NO	NO	SI	NO	NO
183	183	Femenino	NO	NO	SI	SI	SI
184	184	Femenino	SI	NO	SI	SI	NO
185	185	Femenino	NO	NO	NO	NO	SI
186	186	Femenino	NO	NO	NO	SI	SI
187	187	Masculino	NO	NO	SI	SI	SI
188	188	Masculino	NO	NO	SI	SI	SI
189	189	Masculino	SI	NO	SI	SI	SI
190	190	Masculino	NO	NO	SI	SI	SI
191	191	Masculino	NO	NO	SI	NO	SI
192	192	Femenino	NO	NO	SI	NO	NO
193	193	Femenino	NO	NO	SI	SI	SI
194	194	Masculino	NO	NO	SI	SI	SI
195	195	Masculino	NO	NO	SI	NO	SI
196	196	Femenino	NO	NO	SI	NO	SI
197	197	Masculino	NO	NO	SI	SI	SI
198	198	Femenino	SI	NO	SI	NO	NO
199	199	Femenino	NO	NO	SI	NO	NO
200	200	Masculino	NO	NO	SI	SI	SI
201	201	Masculino	NO	NO	SI	SI	SI
202	202	Femenino	NO	NO	SI	NO	NO
203	203	Femenino	NO	NO	SI	SI	SI
204	204	Masculino	SI	NO	SI	SI	SI
205	205	Femenino	NO	NO	SI	SI	NO
206	206	Femenino	NO	NO	NO	SI	NO
207	207	Masculino	NO	NO	SI	SI	SI
208	208	Masculino	NO	SI	NO	SI	SI
209	209	Femenino	NO	NO	SI	SI	NO
210	210	Femenino	NO	NO	SI	NO	NO
211	211	Masculino	SI	NO	SI	SI	SI
212	212	Masculino	NO	NO	NO	NO	SI
213	213	Masculino	NO	NO	SI	NO	NO
214	214	Femenino	NO	NO	NO	NO	NO
215	215	Femenino	SI	NO	NO	NO	SI
216	216	Femenino	NO	NO	NO	NO	SI
217	217	Femenino	NO	NO	NO	NO	SI
218	218	Femenino	NO	NO	NO	NO	NO
219	219	Masculino	NO	NO	SI	SI	SI
220	220	Masculino	NO	NO	SI	NO	NO
221	221	Femenino	NO	NO	SI	NO	NO
222	222	Femenino	NO	NO	SI	NO	NO
223	223	Femenino	SI	NO	SI	SI	NO
224	224	Femenino	SI	NO	NO	SI	SI
225	225	Femenino	NO	NO	NO	NO	NO

226	226	Femenino	NO	NO	NO	NO	NO
227	227	Masculino	SI	NO	NO	NO	NO
228	228	Masculino	SI	NO	SI	SI	SI
229	229	Masculino	NO	NO	SI	SI	SI
230	230	Femenino	NO	NO	SI	SI	SI
231	231	Masculino	NO	NO	NO	SI	SI
232	232	Masculino	NO	NO	SI	SI	NO
233	233	Masculino	NO	NO	NO	SI	NO
234	234	Masculino	NO	NO	SI	NO	NO
235	235	Masculino	NO	SI	SI	SI	NO
236	236	Femenino	SI	NO	SI	SI	SI
237	237	Femenino	SI	NO	SI	SI	SI
238	238	Femenino	NO	NO	SI	SI	SI
239	239	Femenino	NO	NO	SI	SI	SI
240	240	Femenino	SI	NO	SI	SI	SI
241	241	Masculino	NO	SI	SI	SI	SI
242	242	Femenino	NO	NO	SI	NO	NO
243	243	Femenino	NO	NO	NO	NO	NO
244	244	Masculino	SI	NO	SI	SI	SI
245	245	Masculino	SI	NO	SI	NO	SI
246	246	Femenino	SI	NO	SI	NO	SI
247	247	Femenino	SI	NO	SI	NO	NO
248	248	Femenino	NO	NO	SI	NO	SI
249	249	Femenino	NO	NO	SI	NO	NO
250	250	Femenino	NO	NO	SI	NO	NO
251	251	Femenino	NO	NO	NO	NO	SI
252	252	Femenino	NO	NO	SI	SI	SI
253	253	Femenino	SI	NO	SI	NO	SI
254	254	Femenino	SI	NO	SI	SI	SI
255	255	Masculino	NO	NO	SI	NO	SI
256	256	Femenino	NO	NO	SI	NO	SI
257	257	Femenino	NO	NO	SI	NO	NO
258	258	Masculino	NO	NO	SI	SI	SI
259	259	Femenino	NO	NO	NO	NO	SI
260	260	Femenino	NO	NO	SI	NO	SI
261	261	Femenino	NO	NO	NO	NO	NO
262	262	Femenino	SI	SI	SI	SI	SI
263	263	Masculino	SI	SI	SI	SI	SI
264	264	Femenino	NO	SI	SI	NO	SI
265	265	Femenino	SI	SI	SI	SI	SI
266	266	Femenino	SI	NO	SI	SI	SI
267	267	Masculino	SI	NO	NO	NO	NO
268	268	Masculino	SI	SI	SI	SI	SI
269	269	Masculino	NO	NO	NO	NO	NO
270	270	Masculino	SI	NO	NO	SI	SI
271	271	Femenino	NO	NO	NO	SI	SI
272	272	Masculino	NO	NO	SI	NO	NO
273	273	Masculino	SI	NO	SI	SI	SI
274	274	Femenino	NO	NO	SI	NO	SI

275	275	Masculino	SI	NO	NO	NO	NO
276	276	Masculino	SI	SI	SI	SI	SI
277	277	Masculino	NO	NO	SI	SI	SI
278	278	Femenino	NO	NO	NO	NO	SI
279	279	Masculino	SI	NO	SI	NO	NO
280	280	Masculino	SI	NO	SI	SI	SI
281	281	Masculino	SI	SI	SI	SI	SI
282	282	Masculino	NO	NO	SI	SI	NO
283	283	Femenino	NO	NO	SI	SI	SI
284	284	Masculino	SI	NO	NO	SI	NO
285	285	Femenino	NO	NO	SI	SI	NO
286	286	Femenino	NO	NO	SI	SI	NO
287	287	Masculino	SI	SI	SI	SI	SI
288	288	Masculino	SI	SI	SI	SI	SI
289	289	Masculino	SI	SI	SI	SI	SI
290	290	Femenino	NO	NO	SI	SI	SI
291	291	Femenino	NO	NO	SI	SI	NO
292	292	Femenino	NO	NO	NO	SI	NO
293	293	Femenino	SI	NO	SI	NO	SI
294	294	Masculino	NO	NO	SI	NO	SI
295	295	Femenino	NO	NO	SI	NO	NO
296	296	Femenino	SI	NO	SI	SI	NO
297	297	Masculino	NO	NO	SI	NO	SI
298	298	Femenino	NO	NO	SI	NO	NO
299	299	Femenino	NO	NO	SI	NO	SI
300	300	Femenino	NO	NO	SI	NO	SI
301	301	Masculino	NO	NO	NO	NO	NO
302	302	Masculino	NO	NO	SI	SI	SI
303	303	Masculino	NO	NO	SI	SI	NO
304	304	Femenino	SI	NO	SI	NO	SI
305	305	Femenino	NO	NO	SI	SI	SI
306	306	Masculino	NO	SI	NO	SI	SI
307	307	Femenino	SI	NO	SI	NO	SI
308	308	Femenino	NO	NO	SI	SI	SI
309	309	Masculino	NO	NO	SI	SI	NO
310	310	Femenino	NO	NO	SI	NO	SI
311	311	Femenino	NO	NO	NO	NO	SI
312	312	Masculino	SI	NO	SI	SI	SI
313	313	Masculino	NO	SI	SI	SI	SI
314	314	Femenino	NO	NO	SI	NO	NO
315	315	Femenino	NO	NO	NO	SI	NO
316	316	Femenino	NO	NO	SI	NO	NO
317	317	Masculino	NO	NO	SI	SI	SI
318	318	Femenino	NO	NO	SI	NO	NO
319	319	Masculino	SI	NO	SI	SI	NO
320	320	Femenino	NO	NO	NO	NO	NO
321	321	Femenino	SI	NO	SI	NO	NO
322	322	Femenino	NO	NO	NO	NO	NO
323	323	Masculino	NO	NO	SI	SI	NO

324		324	Femenino	NO	NO	SI	NO	SI
325		325	Femenino	NO	NO	SI	NO	NO
326		326	Femenino	NO	NO	SI	SI	SI
327		327	Femenino	SI	NO	SI	NO	NO
328		328	Masculino	NO	NO	SI	NO	NO
329		329	Femenino	SI	NO	SI	NO	SI
330		330	Masculino	SI	NO	SI	SI	SI
331		331	Femenino	NO	NO	SI	NO	SI
332		332	Femenino	NO	SI	NO	NO	SI
333		333	Femenino	NO	NO	SI	SI	SI
334		334	Femenino	SI	NO	NO	SI	SI
335		335	Femenino	NO	NO	NO	SI	NO
336		336	Femenino	NO	NO	NO	SI	NO
337		337	Masculino	NO	NO	SI	SI	SI
338		338	Femenino	NO	NO	SI	NO	SI
339		339	Femenino	NO	NO	SI	SI	SI
340		340	Femenino	NO	NO	SI	SI	SI
Total	N	340	340	340	340	340	340	340

a Limitado a los primeros 400 casos.

ANEXO 2:

Base de datos - Respuesta aleatorizada(a)

	Ficha	SEXO	¿Sueles consumir alimentos o bebidas en la sala de cómputo? / ¿Naciste en el mes de julio?	¿Alguna vez has consumido drogas alucinógenas? / ¿Tu DNI termina en número 2?	¿Algunas veces has copiado en los exámenes? / ¿Naciste en el mes de diciembre?	¿Has consumido bebidas alcohólicas, cerca a la Universidad? / ¿Naciste en el mes de enero?	¿En este ciclo, has faltado a clases sin motivo justificable? / ¿Tu DNI termina en número 5?
1	1	Masculino	NO	NO	SI	SI	NO
2	2	Masculino	NO	NO	SI	NO	NO
3	3	Masculino	SI	NO	NO	NO	NO
4	4	Masculino	NO	NO	SI	SI	NO
5	5	Masculino	NO	SI	NO	NO	SI
6	6	Femenino	SI	NO	SI	SI	SI
7	7	Femenino	NO	NO	NO	NO	NO
8	8	Femenino	NO	NO	SI	NO	SI
9	9	Masculino	NO	NO	SI	SI	SI
10	10	Femenino	SI	NO	NO	SI	NO
11	11	Masculino	SI	NO	SI	SI	NO
12	12	Masculino	NO	NO	SI	SI	NO
13	13	Masculino	NO	SI	SI	SI	NO
14	14	Femenino	NO	NO	SI	NO	NO
15	15	Femenino	NO	NO	SI	NO	NO
16	16	Femenino	SI	NO	SI	NO	SI
17	17	Femenino	NO	NO	SI	NO	NO
18	18	Femenino	NO	NO	SI	SI	SI
19	19	Femenino	SI	NO	SI	NO	NO
20	20	Masculino	NO	SI	SI	NO	NO
21	21	Masculino	NO	NO	SI	SI	SI
22	22	Masculino	NO	SI	SI	NO	NO
23	23	Masculino	NO	NO	SI	SI	NO
24	24	Masculino	NO	NO	NO	NO	SI
25	25	Masculino	NO	NO	SI	NO	SI

26	26	Masculino	NO	NO	SI	SI	SI
27	27	Masculino	SI	NO	SI	SI	SI
28	28	Femenino	NO	NO	SI	NO	NO
29	29	Masculino	NO	NO	SI	SI	NO
30	30	Femenino	SI	NO	SI	SI	NO
31	31	Masculino	NO	NO	SI	SI	SI
32	32	Masculino	NO	NO	NO	SI	NO
33	33	Femenino	SI	NO	NO	NO	SI
34	34	Femenino	SI	SI	NO	SI	SI
35	35	Femenino	NO	SI	SI	NO	SI
36	36	Femenino	SI	SI	SI	NO	NO
37	37	Masculino	NO	NO	SI	SI	NO
38	38	Femenino	NO	NO	SI	NO	NO
39	39	Femenino	NO	NO	SI	NO	NO
40	40	Femenino	NO	SI	NO	NO	NO
41	41	Masculino	NO	NO	NO	SI	NO
42	42	Femenino	NO	NO	SI	NO	NO
43	43	Masculino	NO	NO	NO	NO	NO
44	44	Femenino	NO	NO	SI	NO	NO
45	45	Masculino	NO	NO	SI	NO	SI
46	46	Femenino	NO	NO	NO	NO	NO
47	47	Masculino	SI	NO	SI	NO	SI
48	48	Masculino	SI	NO	SI	SI	SI
49	49	Femenino	NO	NO	NO	NO	SI
50	50	Masculino	SI	NO	NO	NO	NO
51	51	Masculino	NO	NO	SI	SI	SI
52	52	Masculino	NO	NO	SI	NO	NO
53	53	Masculino	NO	NO	SI	SI	NO
54	54	Femenino	NO	NO	NO	NO	NO
55	55	Femenino	NO	SI	SI	NO	NO
56	56	Masculino	SI	NO	SI	SI	SI
57	57	Masculino	NO	NO	SI	NO	SI
58	58	Masculino	NO	NO	SI	SI	NO
59	59	Masculino	NO	NO	NO	SI	SI
60	60	Masculino	NO	SI	SI	NO	SI
61	61	Masculino	SI	NO	NO	SI	NO
62	62	Femenino	SI	NO	NO	SI	SI
63	63	Masculino	NO	NO	SI	SI	NO
64	64	Femenino	NO	SI	NO	NO	SI
65	65	Femenino	NO	NO	SI	NO	NO
66	66	Masculino	SI	SI	NO	NO	SI
67	67	Masculino	SI	NO	NO	NO	SI
68	68	Femenino	SI	SI	NO	NO	NO
69	69	Femenino	SI	NO	SI	NO	SI
70	70	Femenino	NO	NO	NO	NO	NO
71	71	Femenino	NO	NO	SI	NO	SI
72	72	Masculino	NO	NO	NO	NO	SI
73	73	Masculino	SI	NO	SI	NO	NO
74	74	Masculino	NO	NO	SI	SI	NO

75	75	Femenino	NO	NO	SI	SI	SI
76	76	Femenino	NO	NO	NO	NO	NO
77	77	Masculino	NO	NO	SI	NO	NO
78	78	Masculino	NO	NO	NO	NO	SI
79	79	Masculino	NO	NO	SI	SI	NO
80	80	Femenino	SI	NO	SI	SI	SI
81	81	Femenino	SI	NO	NO	NO	SI
82	82	Femenino	NO	NO	SI	NO	NO
83	83	Masculino	NO	NO	NO	SI	SI
84	84	Femenino	NO	NO	SI	NO	NO
85	85	Masculino	SI	NO	NO	NO	NO
86	86	Masculino	NO	SI	SI	SI	SI
87	87	Masculino	NO	SI	NO	SI	SI
88	88	Femenino	NO	NO	NO	NO	SI
89	89	Femenino	NO	NO	SI	NO	SI
90	90	Masculino	SI	NO	SI	SI	SI
91	91	Masculino	SI	NO	NO	NO	SI
92	92	Femenino	NO	NO	SI	NO	NO
93	93	Femenino	NO	NO	SI	SI	SI
94	94	Femenino	NO	NO	SI	NO	NO
95	95	Femenino	SI	NO	NO	NO	SI
96	96	Femenino	NO	NO	NO	NO	NO
97	97	Femenino	SI	NO	SI	NO	SI
98	98	Femenino	NO	NO	SI	NO	NO
99	99	Femenino	SI	NO	NO	SI	NO
100	100	Masculino	SI	SI	NO	SI	SI
101	101	Femenino	SI	NO	NO	NO	SI
102	102	Masculino	NO	NO	NO	SI	SI
103	103	Femenino	NO	NO	SI	NO	NO
104	104	Masculino	NO	NO	SI	SI	SI
105	105	Femenino	SI	NO	NO	NO	SI
106	106	Femenino	NO	NO	SI	NO	SI
107	107	Femenino	NO	NO	NO	NO	NO
108	108	Femenino	SI	SI	NO	NO	SI
109	109	Femenino	NO	NO	SI	NO	NO
110	110	Femenino	SI	NO	SI	NO	SI
111	111	Masculino	SI	NO	NO	SI	SI
112	112	Masculino	NO	NO	SI	NO	NO
113	113	Femenino	SI	NO	NO	NO	NO
114	114	Femenino	SI	SI	NO	SI	SI
115	115	Masculino	SI	SI	SI	NO	NO
116	116	Femenino	NO	NO	NO	SI	SI
117	117	Masculino	SI	SI	NO	NO	NO
118	118	Femenino	SI	SI	SI	NO	SI
119	119	Femenino	NO	NO	NO	SI	SI
120	120	Masculino	NO	NO	SI	SI	SI
121	121	Masculino	NO	NO	SI	NO	NO
122	122	Masculino	SI	NO	SI	NO	SI
123	123	Femenino	NO	SI	SI	NO	NO

124	124	Masculino	SI	SI	SI	NO	SI
125	125	Femenino	SI	NO	NO	SI	SI
126	126	Masculino	SI	NO	NO	SI	SI
127	127	Masculino	SI	NO	SI	NO	SI
128	128	Masculino	SI	NO	SI	NO	NO
129	129	Femenino	SI	NO	NO	SI	NO
130	130	Masculino	SI	NO	NO	NO	NO
131	131	Femenino	NO	NO	NO	NO	SI
132	132	Femenino	SI	NO	NO	SI	NO
133	133	Femenino	NO	NO	SI	SI	NO
134	134	Femenino	NO	NO	SI	NO	SI
135	135	Femenino	NO	NO	NO	SI	SI
136	136	Femenino	NO	NO	NO	SI	NO
137	137	Femenino	NO	NO	NO	NO	SI
138	138	Femenino	NO	NO	NO	NO	SI
139	139	Femenino	SI	NO	NO	NO	SI
140	140	Femenino	NO	NO	NO	NO	SI
141	141	Femenino	NO	SI	NO	SI	NO
142	142	Masculino	NO	NO	SI	SI	SI
143	143	Masculino	NO	NO	SI	NO	NO
144	144	Masculino	NO	NO	NO	NO	SI
145	145	Masculino	NO	SI	NO	NO	NO
146	146	Femenino	SI	NO	NO	NO	SI
147	147	Masculino	NO	NO	NO	NO	NO
148	148	Masculino	NO	NO	SI	SI	NO
149	149	Masculino	NO	NO	SI	NO	NO
150	150	Masculino	NO	NO	SI	SI	SI
151	151	Masculino	NO	NO	NO	SI	SI
152	152	Femenino	NO	NO	SI	NO	SI
153	153	Masculino	NO	NO	SI	SI	NO
154	154	Femenino	SI	NO	NO	NO	NO
155	155	Femenino	NO	NO	SI	NO	NO
156	156	Femenino	NO	NO	NO	NO	SI
157	157	Masculino	SI	NO	NO	NO	SI
158	158	Femenino	SI	NO	NO	NO	SI
159	159	Femenino	NO	NO	NO	SI	SI
160	160	Femenino	SI	NO	NO	NO	SI
161	161	Masculino	NO	SI	NO	SI	NO
162	162	Femenino	NO	NO	NO	SI	SI
163	163	Masculino	NO	NO	NO	NO	SI
164	164	Masculino	SI	NO	NO	SI	NO
165	165	Masculino	NO	NO	SI	NO	SI
166	166	Femenino	NO	SI	NO	SI	NO
167	167	Masculino	NO	SI	SI	NO	NO
168	168	Femenino	SI	NO	SI	NO	NO
169	169	Masculino	NO	NO	SI	NO	NO
170	170	Masculino	NO	NO	NO	SI	NO
171	171	Masculino	NO	SI	SI	NO	SI
172	172	Masculino	NO	NO	SI	NO	NO

173	173	Femenino	SI	NO	SI	NO	NO
174	174	Masculino	NO	SI	NO	SI	SI
175	175	Femenino	NO	NO	SI	SI	SI
176	176	Masculino	NO	NO	NO	NO	SI
177	177	Masculino	NO	NO	SI	NO	NO
178	178	Masculino	NO	NO	SI	NO	SI
179	179	Masculino	NO	NO	SI	SI	NO
180	180	Femenino	SI	NO	SI	SI	NO
181	181	Femenino	NO	SI	SI	SI	NO
182	182	Femenino	NO	NO	SI	SI	SI
183	183	Femenino	NO	NO	NO	SI	SI
184	184	Masculino	SI	NO	SI	SI	SI
185	185	Masculino	NO	SI	SI	NO	SI
186	186	Masculino	NO	NO	SI	SI	SI
187	187	Femenino	SI	NO	SI	SI	SI
188	188	Masculino	SI	SI	NO	NO	NO
189	189	Femenino	SI	NO	SI	SI	SI
190	190	Masculino	NO	NO	SI	SI	SI
191	191	Masculino	SI	NO	SI	NO	NO
192	192	Masculino	NO	NO	NO	NO	NO
193	193	Femenino	SI	NO	SI	NO	SI
194	194	Femenino	NO	NO	SI	NO	SI
195	195	Femenino	SI	NO	SI	SI	SI
196	196	Femenino	SI	NO	SI	SI	SI
197	197	Femenino	NO	NO	SI	NO	NO
198	198	Femenino	SI	NO	SI	SI	SI
199	199	Masculino	SI	NO	SI	SI	SI
200	200	Masculino	NO	NO	SI	SI	SI
201	201	Masculino	SI	SI	SI	NO	SI
202	202	Masculino	NO	NO	NO	NO	SI
203	203	Femenino	NO	NO	NO	SI	SI
204	204	Femenino	SI	NO	SI	SI	SI
205	205	Masculino	NO	NO	SI	SI	NO
206	206	Femenino	NO	NO	SI	NO	NO
207	207	Femenino	NO	NO	NO	NO	NO
208	208	Femenino	SI	NO	SI	NO	NO
209	209	Masculino	SI	NO	SI	SI	NO
210	210	Femenino	SI	NO	SI	NO	SI
211	211	Masculino	SI	NO	NO	SI	NO
212	212	Femenino	NO	NO	SI	NO	NO
213	213	Femenino	SI	NO	NO	NO	SI
214	214	Femenino	NO	NO	SI	NO	SI
215	215	Masculino	SI	NO	SI	SI	NO
216	216	Femenino	SI	SI	SI	NO	SI
217	217	Femenino	NO	NO	SI	NO	NO
218	218	Masculino	NO	SI	NO	SI	SI
219	219	Masculino	NO	NO	SI	NO	NO
220	220	Femenino	SI	NO	SI	NO	SI
221	221	Femenino	SI	NO	NO	SI	NO

222	222	Masculino	NO	NO	NO	SI	SI
223	223	Masculino	NO	NO	SI	SI	SI
224	224	Masculino	NO	NO	NO	SI	NO
225	225	Masculino	NO	SI	SI	NO	SI
226	226	Femenino	SI	NO	NO	NO	SI
227	227	Femenino	SI	NO	SI	SI	SI
228	228	Femenino	SI	SI	SI	NO	SI
229	229	Masculino	NO	SI	SI	NO	NO
230	230	Femenino	NO	SI	NO	SI	SI
231	231	Masculino	SI	NO	NO	SI	NO
232	232	Masculino	SI	SI	NO	NO	NO
233	233	Femenino	NO	NO	SI	SI	NO
234	234	Femenino	NO	NO	SI	SI	NO
235	235	Masculino	SI	NO	SI	NO	SI
236	236	Femenino	NO	NO	SI	SI	NO
237	237	Femenino	SI	NO	SI	SI	NO
238	238	Masculino	NO	NO	NO	SI	NO
239	239	Femenino	NO	NO	SI	NO	SI
240	240	Masculino	SI	SI	NO	NO	NO
241	241	Femenino	NO	SI	SI	NO	NO
242	242	Femenino	NO	NO	SI	NO	SI
243	243	Masculino	NO	NO	NO	SI	SI
244	244	Femenino	NO	NO	SI	SI	SI
245	245	Masculino	NO	NO	SI	NO	SI
246	246	Femenino	NO	NO	SI	NO	NO
247	247	Masculino	NO	NO	NO	NO	SI
248	248	Femenino	SI	NO	NO	NO	SI
249	249	Femenino	SI	NO	SI	NO	NO
250	250	Masculino	NO	NO	NO	NO	NO
251	251	Femenino	NO	NO	NO	NO	SI
252	252	Masculino	SI	SI	NO	SI	NO
253	253	Masculino	NO	NO	NO	SI	SI
254	254	Femenino	SI	NO	SI	SI	NO
255	255	Femenino	NO	NO	SI	SI	NO
256	256	Masculino	SI	NO	SI	SI	NO
257	257	Masculino	NO	NO	SI	NO	NO
258	258	Masculino	SI	SI	SI	NO	NO
259	259	Masculino	SI	NO	SI	NO	SI
260	260	Masculino	SI	NO	SI	SI	NO
261	261	Masculino	SI	NO	NO	NO	SI
262	262	Femenino	SI	SI	SI	NO	SI
263	263	Femenino	SI	NO	SI	NO	SI
264	264	Femenino	SI	NO	NO	SI	NO
265	265	Masculino	NO	SI	NO	NO	NO
266	266	Femenino	NO	SI	SI	SI	NO
267	267	Masculino	NO	NO	SI	SI	NO
268	268	Femenino	NO	SI	NO	SI	SI
269	269	Femenino	SI	SI	NO	SI	NO
270	270	Masculino	SI	SI	NO	SI	SI

271	271	Masculino	SI	SI	NO	SI	NO
272	272	Masculino	NO	SI	NO	SI	NO
273	273	Femenino	NO	NO	NO	SI	SI
274	274	Femenino	SI	SI	NO	SI	SI
275	275	Femenino	SI	SI	NO	SI	NO
276	276	Masculino	NO	NO	SI	SI	NO
277	277	Femenino	SI	NO	SI	SI	NO
278	278	Femenino	NO	NO	NO	NO	NO
279	279	Femenino	SI	NO	NO	NO	SI
280	280	Masculino	NO	NO	SI	NO	NO
281	281	Masculino	SI	NO	NO	SI	NO
282	282	Masculino	NO	NO	SI	NO	NO
283	283	Masculino	NO	NO	NO	SI	SI
284	284	Femenino	NO	NO	SI	NO	SI
285	285	Masculino	NO	NO	NO	NO	NO
286	286	Femenino	SI	NO	NO	SI	SI
287	287	Masculino	NO	NO	NO	SI	NO
288	288	Masculino	NO	NO	SI	SI	SI
289	289	Masculino	NO	NO	SI	SI	SI
290	290	Masculino	NO	NO	SI	NO	NO
291	291	Masculino	NO	NO	SI	NO	SI
292	292	Masculino	SI	NO	NO	SI	SI
293	293	Masculino	NO	SI	NO	NO	NO
294	294	Femenino	NO	NO	NO	NO	SI
295	295	Femenino	NO	NO	SI	NO	NO
296	296	Femenino	NO	NO	NO	SI	NO
297	297	Femenino	NO	NO	SI	NO	SI
298	298	Femenino	NO	NO	SI	NO	SI
299	299	Femenino	NO	NO	NO	NO	NO
300	300	Femenino	NO	NO	NO	NO	NO
301	301	Masculino	SI	NO	SI	SI	NO
302	302	Masculino	NO	NO	SI	SI	NO
303	303	Femenino	SI	NO	SI	NO	SI
304	304	Femenino	NO	SI	NO	NO	SI
305	305	Femenino	NO	NO	NO	NO	SI
306	306	Femenino	SI	NO	NO	NO	NO
307	307	Femenino	NO	NO	NO	NO	NO
308	308	Femenino	NO	NO	SI	NO	NO
309	309	Masculino	SI	NO	SI	NO	NO
310	310	Femenino	SI	NO	SI	NO	NO
311	311	Masculino	SI	NO	NO	SI	NO
312	312	Masculino	NO	NO	NO	NO	NO
313	313	Masculino	SI	NO	SI	NO	NO
314	314	Femenino	SI	SI	SI	SI	SI
315	315	Masculino	SI	NO	SI	NO	SI
316	316	Femenino	NO	NO	SI	SI	NO
317	317	Masculino	NO	NO	NO	SI	NO
318	318	Femenino	SI	NO	SI	NO	NO
319	319	Masculino	NO	NO	SI	NO	NO

320	320	Masculino	NO	NO	SI	SI	SI
321	321	Femenino	SI	NO	SI	NO	SI
322	322	Masculino	NO	SI	SI	NO	SI
323	323	Femenino	NO	NO	NO	NO	NO
324	324	Femenino	SI	SI	NO	NO	NO
325	325	Femenino	SI	SI	SI	NO	SI
326	326	Masculino	NO	NO	NO	NO	NO
327	327	Masculino	SI	SI	NO	NO	NO
328	328	Masculino	NO	NO	NO	SI	NO
329	329	Masculino	NO	NO	SI	SI	SI
330	330	Masculino	NO	NO	NO	NO	NO
331	331	Femenino	SI	NO	NO	NO	SI
332	332	Femenino	SI	SI	SI	SI	SI
333	333	Femenino	NO	NO	SI	NO	NO
334	334	Masculino	NO	NO	SI	NO	NO
335	335	Masculino	SI	NO	NO	SI	SI
336	336	Masculino	SI	NO	SI	SI	SI
337	337	Femenino	NO	NO	NO	SI	SI
338	338	Femenino	NO	SI	SI	NO	NO
339	339	Masculino	NO	SI	NO	SI	SI
340	340	Femenino	SI	NO	SI	SI	NO
341	341	Masculino	NO	NO	NO	SI	NO
342	342	Femenino	NO	NO	SI	SI	SI
343	343	Masculino	NO	SI	NO	NO	NO
344	344	Femenino	SI	NO	NO	SI	SI
345	345	Masculino	NO	NO	SI	SI	SI
346	346	Femenino	SI	NO	SI	SI	NO
347	347	Masculino	SI	SI	SI	NO	SI
348	348	Femenino	NO	NO	NO	SI	NO
349	349	Masculino	SI	NO	NO	SI	NO
350	350	Femenino	NO	NO	SI	SI	SI
351	351	Masculino	NO	NO	SI	SI	NO
352	352	Masculino	NO	SI	NO	NO	SI
353	353	Masculino	SI	SI	SI	SI	SI
354	354	Masculino	SI	SI	NO	NO	NO
355	355	Femenino	NO	NO	NO	SI	SI
356	356	Femenino	NO	NO	SI	NO	NO
357	357	Masculino	SI	NO	NO	NO	NO
358	358	Femenino	NO	NO	NO	SI	NO
359	359	Masculino	NO	NO	SI	NO	NO
360	360	Femenino	NO	NO	SI	NO	NO
361	361	Femenino	SI	NO	SI	NO	NO
362	362	Masculino	SI	NO	SI	SI	SI
363	363	Masculino	SI	NO	NO	NO	SI
364	364	Femenino	SI	NO	NO	NO	SI
365	365	Masculino	NO	NO	NO	SI	SI
366	366	Femenino	NO	NO	NO	SI	SI
367	367	Masculino	SI	NO	NO	NO	NO
368	368	Femenino	SI	NO	NO	NO	NO

369	369	Femenino	NO	NO	SI	NO	SI
370	370	Femenino	NO	NO	SI	SI	SI
371	371	Femenino	NO	NO	SI	NO	NO
372	372	Femenino	NO	NO	SI	NO	NO
373	373	Femenino	NO	NO	SI	SI	SI
374	374	Femenino	NO	NO	SI	SI	SI
375	375	Femenino	NO	NO	SI	NO	SI
376	376	Femenino	NO	NO	NO	NO	SI
377	377	Masculino	NO	NO	SI	NO	NO
378	378	Masculino	SI	NO	SI	SI	NO
379	379	Masculino	NO	NO	SI	SI	SI
380	380	Masculino	NO	SI	SI	SI	NO
381	381	Masculino	SI	SI	SI	NO	SI
382	382	Masculino	SI	NO	SI	SI	SI
383	383	Femenino	SI	SI	NO	NO	SI
384	384	Masculino	SI	SI	NO	NO	SI
385	385	Femenino	NO	NO	SI	SI	NO
386	386	Masculino	SI	NO	NO	SI	NO
387	387	Femenino	NO	NO	SI	NO	NO
388	388	Masculino	NO	NO	SI	SI	NO
389	389	Masculino	NO	NO	SI	NO	NO
390	390	Femenino	SI	NO	SI	NO	SI
391	391	Masculino	NO	SI	SI	SI	SI
392	392	Femenino	NO	NO	NO	NO	NO
393	393	Masculino	NO	NO	SI	SI	SI
394	394	Masculino	NO	NO	SI	SI	NO
395	395	Masculino	NO	NO	SI	NO	NO
396	396	Femenino	NO	NO	SI	SI	SI
397	397	Femenino	NO	NO	NO	SI	SI
398	398	Femenino	NO	NO	NO	NO	SI
399	399	Femenino	NO	NO	SI	SI	SI
400	400	Femenino	NO	NO	NO	NO	NO
401	401	Femenino	SI	NO	NO	NO	SI
402	402	Femenino	NO	NO	NO	NO	NO
403	403	Femenino	SI	NO	SI	SI	NO
404	404	Femenino	NO	NO	SI	NO	NO
Total	N	404	404	404	404	404	404

a Limitado a los primeros 500 casos.