

UNA APLICACION DE LOS DISEÑOS DE RESPUESTA ALEATORIZADA AL CONSUMO DE CANNABIS

Jesús BASULTO SANTOS

Carlos ARIAS MARTIN

1. INTRODUCCION

El objetivo del presente trabajo es, en primer lugar, describir algunos diseños de respuesta aleatorizada, en donde se han obviado todos los planteamientos matemáticos, y en segundo lugar, presentar los resultados obtenidos en la encuesta realizada a los alumnos de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Sevilla.

Con esta investigación empírica se pretendían cubrir dos objetivos fundamentales; en primer lugar, valorar la actitud de los entrevistados cuando se les realiza este tipo de encuestas, y en segundo lugar, aplicar los resultados teóricos que estábamos obteniendo.

De cualquier forma, el lector interesado en la formulación y desarrollo de los diferentes diseños, puede acudir a la tesis doctoral "Análisis Bayesiano de los Diseños de Respuesta Aleatorizada" de Arias, C. cuya lectura y defensa se realizó en la Facultad de Ciencias Económicas y Empresariales de la Universidad de Sevilla en diciembre de 1990.

2. DISEÑOS DE RESPUESTA ALEATORIZADA

Cuando se realiza una encuesta por muestreo se pueden cometer dos tipos de errores; los denominados errores debidos al muestreo y los errores ajenos al muestreo. Nuestro trabajo se va a centrar en los errores ajenos al muestreo, y más concretamente, en aquellos debidos a las propias características de los individuos que se pretenden analizar, que bien porque no estén consideradas socialmente

deseables, bien porque sean altamente personales, o bien porque sean ilegales, tienen la consideración de íntimas.

Así, cuando se pretende estimar la proporción de la población que, por ejemplo, consume cierto tipo de drogas, no es adecuado preguntar directamente a un individuo dado si posee o no esta característica íntima, puesto que tanto los porcentajes correspondientes a la no respuesta, como los debidos a contestaciones deliberadamente falsas así lo aconsejan. Para resolver estas situaciones surgen los diseños de respuesta aleatorizada, en los que a cada uno de los entrevistados elige, mediante un mecanismo de aleatorización que se le muestra, una de entre dos o más preguntas que le son presentadas, sin que el entrevistador conozca el resultado de dicha selección.

Esta técnica fue concebida por Warner (1965), y su diseño supone que la población queda dividida en dos grupos no solapados, cuya unión es el colectivo objeto de estudio, donde el primer grupo lo constituyen aquellos individuos que poseen la característica íntima, mientras que el segundo lo forman los que no la tienen. Y así, construye dos preguntas o declaraciones relacionadas, esto es;

Pertenezco al primer grupo.

Pertenezco al segundo grupo.

haciendo declarar a cada individuo seleccionado si pertenece o no al grupo que le haya correspondido después de utilizar el mecanismo aleatorio. Este diseño fue generalizado posteriormente, a partir del supuesto de que los elementos de la población, en relación con la característica íntima, se pueden dividir en k niveles exhaustivos y mutuamente excluyentes, donde al menos 1 o a lo sumo $k-1$, tienen la consideración de íntimo, por Abul-Ela et al (1967) y por Baurke y Dalenius (1973).

En la línea de proteger en mayor medida la intimidad de los entrevistados, aparece en 1967 el diseño con una pregunta no relacionada, propuesto por Horvitz et al y desarrollado notablemente por Greenberg et al. (1969). Este diseño crea una mayor incertidumbre al establecer dos preguntas o declaraciones, una íntima y otra intrascendente, donde esta última es independiente o está no relacionada con la primera, de forma que cada entrevistado debe responder si posee o no la característica que le haya correspondido al utilizar el mecanismo aleatorio. Ahora bien, el diseño es distinto según la proporción de la población que posee la característica no relacionada sea conocida o bien tenga que estimarse mediante la encuesta.

El desarrollo de este segundo supuesto ha tenido más interés, puesto que aunque la característica intrascendente es seleccionada por el analista, no debe-

mos perder de vista que el objetivo básico de los diseños de respuesta, aleatorizada es proteger la intimidad de los entrevistados, y así, debe de ser imposible verificar, para un individuo dado, si posee o no la característica no relacionada. Para la utilización de este diseño, desde el punto de vista clásico, es necesario seleccionar dos muestras aleatorias simples con reemplazamiento, independiente y no solapadas, aunque en algunos supuestos, como el de Moors (1971), en una de las muestras no se utilice mecanismo aleatorio alguno, dado que sólo se pregunta de forma directa si posee o no la característica intrascendente.

Este diseño, ha sido generalizado, con idénticos planteamientos que el de Warner, por Greenberg et al (1969), Eriksson (1973) y Bourke (1974), extendiéndose también al supuesto de dos preguntas no relacionadas en Folsom et al (1973).

Una característica fundamental de cualquier diseño de respuesta aleatorizada frente al que emplea la pregunta directa, es que su utilización conduce, bajo el supuesto de tamaños muestrales idénticos, a una disminución de la precisión del estimador, medida en términos de su desviación estándar.

Un camino para lograr incrementar la precisión del estimador sin aumentar el tamaño muestral es hacer uso de la inferencia bayesiana. Así, a partir de la información inicial sobre la proporción de individuos que poseen la característica íntima, medida en términos de una distribución de probabilidad a priori o inicial podemos ofrecer estimaciones por intervalo, más precisas que las ofrecidas por la inferencia clásica, siendo el trabajo de Winkler y Franklin (1979), el primero que aborda desde esta concepción el diseño propuesto por Warner.

Por otra parte, la función de verosimilitud resultante al aplicar estos diseños es complicada, haciendo el cálculo del estimador máximo verosímil difícil de obtener. Una forma de resolver este problema es utilizar el algoritmo EM (alternancia entre tomar esperanza y maximizar), que aunque debido a Hartley (1958), no aparece desarrollado adecuadamente hasta 1977 por Dempster et al. Así, en aquellas situaciones que producen datos no observados, se formaliza una idea intuitiva; esto es, reemplazar los valores no observados por valores estimados, estimar el parámetro, reestimar los valores no observados asumiendo que el valor estimado del parámetro es correcto, reestimar el parámetro, y así sucesivamente hasta conseguir la convergencia. La aplicación de este algoritmo a los diseños de respuesta aleatorizada, desde el punto de vista clásico, ha sido introducida por Bourke y Moran (1988).

Ahora bien, dado que para poder realizar inferencias no es suficiente aproximar la desviación estándar de la distribución a posteriori, se han utilizado diversos métodos para aproximar esta desviación estándar, que dan lugar a un conjunto de aproximaciones a la verdadera distribución final de la proporción de indivi-

duos que poseen la característica íntima, utilizando los métodos propuestos por Leonard (1975), Laird y Louis (1982) y Tierney y Kadane (1986).

3. CARACTERISTICAS DE LA ENCUESTA

El colectivo objeto de estudio estaba formado por los alumnos de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Sevilla, y el cuestionario, que se ofrece en el anexo, constaba de diez preguntas, donde dos atendían a variables de clasificación, otras dos se referían a la característica íntima que seleccionamos, mientras que con el resto de preguntas intentamos valorar la actitud de los entrevistados hacia este tipo de encuestas.

Para la elección de la característica íntima seguimos el trabajo “Los andaluces ante las drogas” (Junta de Andalucía, 1989), y así, se propuso preguntar si había consumido alguna vez, durante los últimos treinta días, Cánnabis (Porros); esto es, Haschis, Marihuana,... También, y siguiendo con el objetivo de la experimentación, a cada individuo, se le preguntaba, tanto según el diseño de Warner, como el de preguntas no relacionadas. Ahora bien, otro problema que había que resolver era la selección de la característica no relacionada. Así, se intentó aprovechar la información que suministraba el “Informe sobre actitudes, opiniones e imagen de la Universidad de Sevilla” (Diez de Castro et al, 1987), pero bien porque las variables consideradas podían tener a priori relación con el consumo de Cánnabis, bien porque podíamos conocer si el individuo poseía la característica no relacionada, tuvimos que acudir a una pregunta no considerada en el informe, y que al final fue si venía habitualmente en moto a la Facultad.

También, y siguiendo con el objetivo de la experimentación, se seleccionaron dos muestras del mismo tamaño, siendo la diferencia entre ellas la probabilidad con el que el instrumento aleatorio seleccionaba la pregunta íntima en el diseño de preguntas no relacionadas. Por lo que respecta al instrumento aleatorio, decidimos utilizar los naipes, porque pensamos que eligiendo uno que fuera conocido por los entrevistados, facilitábamos su comprensión hacia el método, lo que, en principio, nos conduciría a una mayor cooperación.

Así pues, cada individuo selecciona, en primer lugar, mediante el instrumento de aleatorización una de estas dos preguntas:

- ¿Pertenece al grupo de los consumidores?
- ¿Pertenece al grupo de los no consumidores?

respondiendo simplemente SI o NO a la pregunta que le haya correspondido mientras que al aplicar el diseño de preguntas no relacionada, selecciona nuevamente mediante el mecanismo de aleatorización una de estas dos preguntas:

¿Pertenece al grupo de los consumidores?

¿Viene en moto habitualmente a la Facultad?

respondiendo nuevamente SI o NO a la pregunta seleccionada, donde en cualquier caso, se considera consumidor todo aquél que haya consumido alguna vez durante los últimos treinta días, Cánabis.

4. CARACTERISTICAS DE LA MUESTRA

Dado que el número de alumnos matriculados en la Facultad, en el curso académico 1989-90 ascendía a 4.290, y puesto que lo que se pretende estimar es una proporción, aplicamos la fórmula habitual para determinar el tamaño muestra, ascendiendo éste a 366. En función de la información disponible, estratificamos la población por el curso superior en que se encontraba matriculado, realizando la afijación de forma proporcional al tamaño de cada estrato. Las encuestas se realizaron desde el día 26 de marzo al 4 de abril de 1990.

El número de encuestas realizadas ascendió a 275 (75,13%), siendo la distribución por muestra:

Muestra 1: 137 (74,86%)

Muestra 2: 138 (75,41%)

Las características de las muestras, según las variables de clasificación, son las que aparecen en las tablas I y II:

TABLA I

Edad (años)	Total		Muestra 1		Muestra 2	
	Nº	%	Nº	%	Nº	%
17-18	50	18,19	21	15,33	29	21,01
19	55	20,00	30	21,90	25	18,13
20	53	19,27	27	19,71	26	18,84
21-22	64	23,27	35	25,55	29	21,01
>22	53	19,27	24	17,51	29	21,01
Total	275	100,00	137	100,00	138	100,00

TABLA II

Sexo	Total		Muestra 1		Muestra 2	
	Nº	%	Nº	%	Nº	%
Mujeres	104	37,82	53	38,69	51	36,96
Hombres	171	62,18	84	61,31	87	63,04
Total	275	100,00	137	100,00	138	100,00

De las dos tablas anteriores podemos concluir que no se aprecian diferencias notables entre las dos muestras, en lo que se refiere a estas dos variables de clasificación.

5. VALORACION DE LOS ENCUESTADOS SOBRE LOS DISEÑOS PROPUESTOS

Después de que el alumno hubiera respondido SI o NO a la pregunta seleccionada según el diseño de Warner, se le preguntaba si creía que su intimidad quedaba adecuadamente protegida al utilizar la baraja en lugar de preguntarle directamente. A continuación, y después de cambiarle la baraja, debía responder nuevamente SI o NO a la pregunta impresa en la carta seleccionada por él, según el diseño de preguntas no relacionadas, y nuevamente se le hacía la misma pregunta sobre la protección de su intimidad. Los resultados obtenidos son los que se muestran en las tablas III y IV.

Así, podemos observar que el 87% de los encuestados considera que su intimidad queda adecuadamente protegida, sin diferencias notables según el diseño aplicado, aunque este porcentaje es ligeramente superior, para ambos diseños, en la muestra 1.

TABLA III

Warner	Total		Muestra 1		Muestra 2	
	Nº	%	Nº	%	Nº	%
SI	240	87,27	123	89,78	117	84,78
NO	35	12,73	14	10,22	21	15,22
Total	275	100,00	137	100,00	138	100,00

TABLA IV

No relacionadas	Total		Muestra 1		Muestra 2	
	Nº	%	Nº	%	Nº	%
SI	238	86,54	124	90,51	114	82,61
NO	37	13,46	13	9,49	24	17,39
Total	275	100,00	137	100,00	138	100,00

También, se pidió que valorasen cual de los dos diseños propuestos protegía más su intimidad. De esta forma, y teniendo en cuenta sus respuestas sobre cada uno de ellos por separado, podemos ofrecer la tabla V, donde se observa que para el 58% de los entrevistados no existe diferencia entre ambos diseños, considerando que ambos protegen de igual forma su intimidad, mientras que existe un porcentaje algo superior de alumnos que piensan que es mejor el diseño de preguntas no relacionadas que el de Warner; que alrededor del 9% considera que ninguno de los dos diseños es suficiente para que su intimidad quede protegida, y que únicamente hay tres individuos que a pesar de haber manifestado que ninguno de los diseños propuestos protege adecuadamente su intimidad, se inclinan a valorar uno superior a otro.

TABLA V

Diseño	Warner				Total	
	SI		NO			
	SI	NO	SI	NO	Nº	%
Warner	33	7	0	2	42	15,27
No relac.	38	0	9	1	48	17,45
Ninguno	0	0	0	25	25	9,10
Por igual	160	0	0	0	160	58,18
Total	231	7	9	28	275	100,00

Además, hemos pretendido que los entrevistados indicaran cuál de los diseños propuestos les parecía más claro. En este sentido ofrecemos la tabla VI.

Si bien los diseños con preguntas no relacionadas surgieron, fundamentalmente, para obtener una actitud más favorable por parte de los entrevistados, puesto que crea una incertidumbre sobre si se está respondiendo a la pregunta íntima o a la no relacionada, podemos comprobar que este hecho no queda debidamente

TABLA VI

Diseño Claridad	Total		Muestra 1		Muestra 2	
	Nº	%	Nº	%	Nº	%
Warner	125	45,45	57	41,61	68	49,28
No relac.	19	6,91	14	10,22	5	3,62
Ninguno	60	21,82	31	22,63	29	21,01
Por igual	71	25,82	35	25,54	36	26,09
Total	275	100,00	137	100,00	138	100,00

reflejado en la tabla V, puesto que únicamente el 17% de los entrevistados se manifiestan en este sentido. Ahora bien, nosotros pensábamos que el diseño de Warner podía crear cierta confusión en la pregunta complementaria; o sea, en si pertenece al grupo de los no consumidores. Sin embargo, el 45% de los alumnos eligen como diseño más claro el de Warner, frente al 7% que se inclinan por el de preguntas no relacionadas. Creemos que estas situaciones se deben al hecho de los entrevistados no comprenden que tiene que ver el ser consumidor o no, con venir habitualmente en moto a la facultad o no. Y precisamente esta falta de claridad en la comprensión del diseño, que no de las preguntas, es lo que ha motivado los resultados expuestos en la tabla V y sobre todo en la VI.

También, debemos resaltar que el 22% de los entrevistados no considera que ninguno de los diseños sea claro, y ello a pesar de que a cada entrevistado se le iba mostrando la composición de las barajas, indicándole que fuera el mismo el que barajara los naipes, que extrajera uno, que sólo debía ser visto por él, y que después de contestar SI o NO, volviera a introducir el naipe en la baraja, moviendo de nuevo las cartas para que nadie pudiera saber la pregunta que le había correspondido. Nuevamente nos inclinamos a pensar que este resultado se debe, en gran medida, en la falta de comprensión del diseño, que no de las preguntas ni de la partición que se realiza en cada caso.

Por último, en las tablas VII y VIII, mostramos los resultados obtenidos sobre dos hechos que consideramos esenciales; por una parte si piensan que su respuesta se vería afectada si su nombre y su teléfono tuviera que aparecer en la encuesta (recordemos que la encuesta era anónima y que únicamente se expuso una lista de los alumnos aleatoriamente seleccionados), y por otra si estarían dispuestos a contestar, de forma veraz, a la pregunta de si es consumidor o no, sin instrumento aleatorio alguno.

De las tablas VII y VIII, podemos resaltar que el 86% de los alumnos consideran que su respuesta no se vería afectada por el hecho de que su nombre y su teléfono tuviera que aparecer en la encuesta, siendo este porcentaje ligeramente

TABLA VII

Nombre Teléfono	Total		Muestra 1		Muestra 2	
	Nº	%	Nº	%	Nº	%
SI	38	13,82	15	10,95	23	16,67
NO	237	86,18	122	89,05	115	83,33
Total	275	100,00	137	100,00	138	100,00

TABLA VIII

Pregunta directa	Total		Muestra 1		Muestra 2	
	Nº	%	Nº	%	Nº	%
SI	254	92,36	127	92,70	127	92,03
NO	21	7,64	10	7,30	11	7,97
Total	275	100,00	137	100,00	138	100,00

superior en la muestra 1, mientras que el 92% estarían dispuestos a manifestarnos directamente si son consumidores o no, siendo el comportamiento prácticamente idéntico en ambas muestras.

6. ESTUDIO DE LA PROPORCION DE CONSUMIDORES

En primer lugar se va a considerar las respuestas obtenidas al aplicar el diseño de Warner, que se recogen en la tabla IX.

Si denominamos π_A a la proporción de consumidores, tenemos que aplicar las técnicas clásicas, la estimación máximo verosímil e insesgada de esta proporción

TABLA IX

Warner	Total		Muestra 1		Muestra 2	
	Nº	%	Nº	%	Nº	%
SI	89	32,36	45	32,85	44	31,88
NO	186	67,64	92	67,15	94	68,12
Total	275	100,00	137	100,00	138	100,00

$\hat{\pi}_A$, así como la de su desviación estándar, $\hat{\sigma}_A$, es, dado que la probabilidad con que el instrumento aleatorio selecciona la pregunta íntima es de 0,7 y teniendo en cuenta tanto cada una de las muestras como el total.

$$\begin{array}{ll} \text{Para la muestra 1: } \hat{\pi}_A = 0,071 & \hat{\sigma}_A = 0,1005 \\ \text{Para la muestra 2: } \hat{\pi}_A = 0,047 & \hat{\sigma}_A = 0,0995 \\ \text{Para el total: } \hat{\pi}_A = 0,059 & \hat{\sigma}_A = 0,0707 \end{array}$$

Estos resultados ponen de manifiesto la sensibilidad de este diseño frente a pequeñas variaciones en el número de respuestas afirmativas, así como la gran variabilidad de las estimaciones.

TABLA X

No relacionadas	Muestra 1		Muestra 2	
	Nº	%	Nº	%
SI	18	13,14	17	12,32
NO	119	86,86	121	87,68
Total	137	100,00	138	100,00

Considerando a continuación los resultados obtenidos al aplicar el diseño de preguntas no relacionadas, que aparecen en la tabla X, y teniendo en cuenta que la probabilidad con que el mecanismo de aleatorización selecciona la pregunta íntima es 0,7 en la muestra 1 y 0,6 en la muestra 2, y suponiendo que la proporción de alumnos que acuden habitualmente en moto a la facultad, π_B , es desconocida, podemos obtener las siguientes estimaciones máximo verosímiles e insesgadas, según las técnicas clásicas:

$$\hat{\pi}_{A,B} = 0,156 \qquad \hat{\sigma}_{A,B} = 0,1431$$

$$\hat{\pi}_{B,A} = 0,074 \qquad \hat{\sigma}_{B,A} = 0,2623$$

donde $\hat{\pi}_{A,B}$ y $\hat{\pi}_{B,A}$ son las estimaciones de π_A y π_B , respectivamente, bajo el supuesto de que se desconocen estas proporciones poblacionales, y donde $\hat{\sigma}_{A,B}$ y $\hat{\sigma}_{B,A}$ son las correspondientes estimaciones de σ_A y σ_B , respectivamente, bajo el mismo supuesto.

La proporción de consumidores supera en mucho a la obtenida mediante el diseño de Warner, aunque el valor tan elevado de la desviación estándar puede

poner en duda la validez de esta estimación. Ahora bien, es evidente que se puede obtener información sobre la proporción de alumnos que acuden habitualmente en moto al Centro. Así, se contaron durante los ocho días en que se realizaron las encuestas el número de motos que se encontraban en el aparcamiento de la Facultad, tanto en el turno de la mañana como en el de la tarde, y suponiendo que cada moto transporta a un alumno y medio, y a partir del número medio de estos vehículos se ha estimado esta proporción, que ha alcanzado un valor de 0,077. De esta forma, y aplicando nuevamente las técnicas clásicas, podemos obtener las siguientes estimaciones máximo verosímiles e insesgadas para cada una de las muestras:

$$\begin{array}{ll} \text{Para la muestra 1: } \hat{\pi}_{A/B} = 0,155 & \hat{\sigma}_{A/B} = 0,0412 \\ \text{Para la muestra 2: } \hat{\pi}_{A/B} = 0,154 & \hat{\sigma}_{A/B} = 0,0469 \end{array}$$

donde $\hat{\pi}_{A/B}$ y $\hat{\sigma}_{A/B}$ son las correspondientes estimaciones de π_A y σ_A , bajo el supuesto de que π_B es conocida.

Podemos observar que la proporción de consumidores sigue situándose en torno al 15%, mientras que la desviación estándar ha disminuído notablemente.

Ahora bien, afirmar que la proporción de alumnos que acuden en moto a la Facultad es 0,077 parece, en principio, muy arriesgado. También debemos tener en cuenta que el trabajo “Los andaluces ante las drogas” (Junta de Andalucía, 1989), estima en un 12,9% la proporción de individuos que, con estudios medios-superiores, han consumido alguna vez durante los últimos seis meses cannabis.

Por ello, y dado que las distribuciones beta parecen ser razonables aproximaciones para una extensa variedad de tipos de información a priori sobre la proporción de la población que posee cierta característica, proponemos como distribución inicial del parámetro π_A , la distribución beta $B(a_1, b_1)$, y como distribución inicial del parámetro π_B , la distribución beta $B(a_2, b_2)$. Así, si suponemos que la moda de π_A es 0,12 y su desviación estándar la ciframos en 0,04, podemos determinar, aproximadamente, los valores de a_1 y b_1 , resolviendo el sistema:

$$\begin{aligned} 0,12 &= (a_1 - 1) / (a_1 + b_1 - 2) \\ 1/0,04^2 &= - \left[\frac{\partial^2}{\partial \pi_A^2} \ln f(\pi_A / a_1, b_1) \right] \Big|_{\pi_A = 0,12} \end{aligned}$$

siendo entonces $a_1=8,92$ y $b_1=59,08$. Análogamente, considerando que la moda de π_B es 0,077 y su desviación estándar 0,05; tenemos que $a_2=3,19$ y $b_2=27,24$.

De esta forma, la moda de la distribución final de la proporción de individuos que tienen la consideración de consumidores, π_A , utilizando el algoritmo EM, la desviación estándar de esta distribución final, σ_A , que se calcula a partir de las aproximaciones de los dos primeros momentos de la distribución final (Tierney y Kdane, 1986), así como la mejor aproximación a la distribución final, son:

Diseño de Warner:

Muestra 1:	$\hat{\pi}_A = 0,114$	$\hat{\sigma}_A = 0,0359B (9,99; 71,08)$
Muestra 2:	$\hat{\pi}_A = 0,111$	$\hat{\sigma}_A = 0,0354B (9,89; 72,46)$
Total:	$\hat{\pi}_A = 0,106$	$\hat{\sigma}_A = 0,0323B (11,15; 86,08)$

Diseño de preguntas no relacionadas:

Muestra 1:	$\hat{\pi}_A = 0,136$	$\hat{\sigma}_A = 0,0337N (0,136; 0,033)$
Muestra 2:	$\hat{\pi}_A = 0,132$	$\hat{\sigma}_A = 0,0361N (0,132; 0,035)$

donde $N(\mu; \sigma)$ representa una distribución normal de media μ y desviación estándar σ .

Podemos observar que al utilizar las técnicas bayesianas, bajo el supuesto que las distribuciones iniciales de los parámetros son correctas, las estimaciones de las desviaciones estándares obtenidas alcanzan unos valores que podemos considerar adecuados, aunque la diferencia de las modas estimadas según el diseño empleado se sigue poniendo de manifiesto, aunque para el de Warner, estas diferencias se han recortado notablemente. Así pues, y puesto que mediante el diseño de preguntas no relacionadas se ha estimado que la proporción de alumnos que acuden habitualmente en moto a la facultad, utilizando las técnicas clásicas, prácticamente es igual a nuestra estimación basada en el número de motos aparcadas, nos inclinamos a pensar que este diseño ofrece mejores resultados.

También, y puesto que la información utilizada para estimar la distribución inicial de la proporción de consumidores, se obtuvo en el año 1987, y además no se ajusta exactamente a nuestra encuesta, puesto que nuestros alumnos todavía están estudiando y que la consideración de consumidor está referida a los últimos treinta días y no a los seis meses anteriores, preferimos considerar como distribución inicial de este parámetro una beta $B(1;1)$, aunque se incremente la estimación de la desviación estándar. De esta forma, bajo este último supuesto tenemos:

Muestra 1:	$\hat{\pi}_A = 0,155$	$\hat{\sigma}_A = 0,0442N (0,155; 0,045)$
Muestra 2:	$\hat{\pi}_A = 0,154$	$\hat{\sigma}_A = 0,0539N (0,154; 0,055)$

7. CONCLUSIONES DE LA ENCUESTA

La experimentación que hemos llevado a cabo en la Facultad de Ciencia Económicas y Empresariales de la Universidad de Sevilla ha perseguido fundamentalmente valorar la actitud de los entrevistados cuando se les realiza un encuesta que incorpora un instrumento aleatorio, y además aplicar los desarrollos teóricos que hasta la fecha hemos desarrollado en lo concerniente a la aplicación del algoritmo EM y las aproximaciones a la distribución final cuando se utilizan técnicas bayesianas.

Así pues, debemos mencionar en primer lugar que nuestros alumnos consideran que su intimidad queda adecuadamente protegida al utilizar alguno de los dos diseños propuestos, sin que existan diferencias notables entre ellos. Ahora bien este colectivo alcanza un cierto nivel cultural, y eso en principio, tal y como afirman algunos autores, favorece notablemente este aspecto. Sin embargo, creemos que la utilización de un instrumento aleatorio tan fácilmente comprensible como las cartas, supondría, sin importar sobre manera el nivel cultural del entrevistado, su comprensión hacia el hecho de que es imposible que el entrevistado conozca la pregunta que le ha correspondido. También resaltar que los alumnos han manifestado que consideran más claro el diseño de Warner que el de preguntas no relacionada, aunque tal y como hemos mencionado creemos que esta situación se debe a que no han comprendido este último diseño, puesto que no encontraban relación alguna entre ambas características, aunque, evidentemente la hipótesis de independencia entre ellas es esencial para el desarrollo del diseño. Por último, mencionar que, en principio, los entrevistados no se muestran contrarios a que su nombre y teléfono aparezca en la encuesta, y que parecen dispuestos a contestarnos directamente si son consumidores, en los términos en los que se ha definido.

En segundo lugar, y a la vista de los resultados obtenidos en la estimación tanto de la proporción de consumidores como de su desviación estándar, la incorporación de las técnicas bayesianas, cuando la información inicial sobre uno, o ambos, de los parámetros es adecuada, reduce notablemente la sensibilidad que ha presentado el diseño de Warner hacia el número de respuestas afirmativas, y logra, en ambos diseños, una mayor eficiencia, medida en términos de la característica no relacionada, o sea, la proporción de alumnos que acuden habitualmente en moto a la Facultad, ha sido fundamental para considerar más adecuados los resultados obtenidos por el diseño de preguntas no relacionadas que por el de Warner. Por último debemos resaltar que al emplear las técnicas bayesianas en el diseño de preguntas no relacionadas, se reducen notablemente los costes de análisis, ya que únicamente hace falta una muestra, situación que, cuando se

utilizan las técnicas clásicas, sólo se puede presentar si se conoce exactamente la proporción de individuos que poseen la característica no relacionada, y que no podemos considerar como general para la mayoría de los estudios.

8. CONSIDERACIONES FINALES

La sociedad actual presenta tres grandes problemas, como son el terrorismo, el paro y la droga. Precisamente sobre el consumo de drogas debemos realizar una breve reflexión, puesto que si no disponemos de datos sobre el consumo, no se podrá realizar una adecuada prevención, ya que la información es el punto de partida imprescindible para su planeamiento.

En esta línea, es el propio Comisionado para la droga de la Consejería de Salud y Servicios Sociales de la Junta de Andalucía quien afirma que estos datos no se toman, a veces, con la suficiente fiabilidad (Los andaluces ante las drogas, 1989). Y es precisamente en este punto donde las técnicas estadísticas deben ser aplicadas para lograr una mayor fiabilidad en la información recogida mediante una muestra. Entonces, aunque el diseño de la misma sea correcto, no se suelen considerar aquellos errores que son ajenas al muestreo, esto es, las respuestas deliberadamente falsas y la propia falta de respuesta.

De esta forma, al considerar aquellas drogas que son ilegales, o que están consideradas socialmente indeseables, se debe facilitar al entrevistado de un mecanismo que proteja su intimidad, con lo que se estará contribuyendo a facilitar su cooperación, logrando así reducir dichos errores.

Ahora bien, debemos resaltar que los diseños de respuesta aleatorizada incrementan los costes de una encuesta, tanto por la vía de aumentar el tamaño muestral necesario para conseguir la misma eficiencia en los estimadores, como por la de incrementar el tiempo para que el entrevistador realice la encuesta, puesto que es imprescindible que al entrevistado se le presente el mecanismo de aleatorización y que compruebe que su intimidad queda adecuadamente protegida.

Por otro lado, también se tienen que considerar las limitaciones que presentan los diseños de respuesta aleatorizada, pues para estimar los valores de una variable policotómica o de una variable continua, la aplicación de éstos es, normalmente, compleja.

A pesar de todo ello, creemos que la utilización de estos diseños conduciría a obtener resultados más fiables y precisos que los conseguidos mediante la aplicación de las técnicas clásicas, siempre y cuando, se pretenda obtener información sobre aquellas drogas que bien estén mal consideradas socialmente o bien sean ilegales.

9. BIBLIOGRAFIA

Antes de ofrecer las referencias bibliográficas, queremos **dejar** constancia de nuestro agradecimiento a Susana León Sánchez y a Vicenta María Márquez de la Plata y Cuevas, por su inestimable colaboración en el diseño, tabulación y análisis de la encuesta.

- ABUL-ELA, A. A., GREENBERG, B. G. and HORVITZ, D. G. (1967). A multiproportion randomized response model. *J. Amer. Statist. Assoc.* 62, 990-1008.
- ARIAS, C. (1990). Análisis bayesiano de los diseños de respuesta aleatorizada. *Tesis Doctoral*. Universidad de Sevilla. Pendiente de publicación.
- BOURKE, P.D. (1974) Multi-proportions randomized response using the unrelated question technique Report 74, Errors in Surveys Research Project. *Inst. of Statist., University of Stockholm*.
- BOURKE, P.D. (1974) Multi-proportions randomized response using the unrelated question technique Report 68, Errors in Surveys Research Project. *Inst. of Statist., University of Stockholm*.
- BOURKE, P.D. and MORAN M.A. (1988). Estimating proportions from randomized response data using the EM algorithm. *J. Amer. Statist. Assoc.* 83, 964-968.
- DEMPSTER, A. P. LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. B* 39, 1-38.
- DIEZ DE CASTRO, E. MARTIN, E., SERRANO F. y BASULTO, J. (1988). *Informe sobre actitudes opiniones e imagen de la Universidad de Sevilla*. Universidad de Sevilla.
- ERIKSSON, S.A. (1973). A new model for randomized response. *Internat. Statist. Rev.*, 41, 101-113.
- FOLSOM, R. E., GREENBERG, B. G., HORVITZ, D. G. and ABERNATHY, J.R. (1973). The two alternate questions randomized response model for human surveys. *J. Amer. Statist. Assoc.*, 68, 525-530.
- GREENBERG, B. G., ABDEL-LATIF, A., ABUL-ELA, A. A., SIMMONS, W. R. and HORVITZ, D.G (1969). The unrelated question randomized response model: theoretical framework. *J. Amer. Statist. Assoc.* 64, 520-539.
- HARTLEY, H. O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics*, 14, 174-194.
- HORVITZ, D. G. SHAH, B. V. and SIMMONS, W. R. (1967). The unrelated question randomized response model. *Proc. ASA. Soc. Statist. Sec.*, 65-72.

10. ANEXO

Questionario nº 1

Fecha: ___ / ___ / ___

1. Edad:

2. Sexo: 1 Hombre 2 Mujer

3. Supuesto que tiene la consideración de CONSUMIDOR todo aquel que haya consumido alguna vez, durante los últimos 30 días, Cannabis, esto es. Haschis, Marihuana, Porros, póngase donde el entrevistador no pueda verle, baraje las cartas y extraiga una al azar, que sólo debe ser vista por usted. Conteste SI o NO a la declaración que aparece impresa en la carta extraída.

RESPUESTA:

Introduzca la carta en la baraja, y barájeelas para que nadie sepa la carta que usted ha extraído.

4. ¿Cree usted que su intimidad queda adecuadamente protegida al utilizar la baraja en lugar de preguntarle directamente? 1 SI 2 NO

5. El entrevistador va a cambiarle la baraja. Bajo el supuesto referido en la cuestión cuarta, conteste SI o NO a la declaración que aparece impresa en la carta extraída.

RESPUESTA:

Introduzca la carta en la baraja, y barájeelas para que nadie sepa la carta que usted ha extraído.

6. ¿Cree usted que su intimidad queda adecuadamente protegida al utilizar la baraja en lugar de preguntarle directamente?

1 SI 2 NO

7. ¿Cuál de los dos métodos propuestos cree que es más claro?

1 El primero 2 El segundo 3 Ninguno 4 Los dos por igual

8. ¿Cuál de los dos métodos propuestos cree que protege más su intimidad?

1 El primero 2 El segundo 3 Ninguno 4 Los dos por igual

9. Si su nombre y su teléfono tuviera que aparecer en el cuestionario, ¿Cree que se vería afectada la respuesta?

1 SI 2 NO

10. Si se le preguntara si usted es consumidor o no lo es, ¿Estaría dispuesto a contestar, de forma veraz, a la pregunta?

1 SI 2 NO