

RESTRINGIDA

E/CEPAL/R.254

12 de marzo de 1981

ORIGINAL: ESPAÑOL

C E P A L

Comisión Económica para América Latina

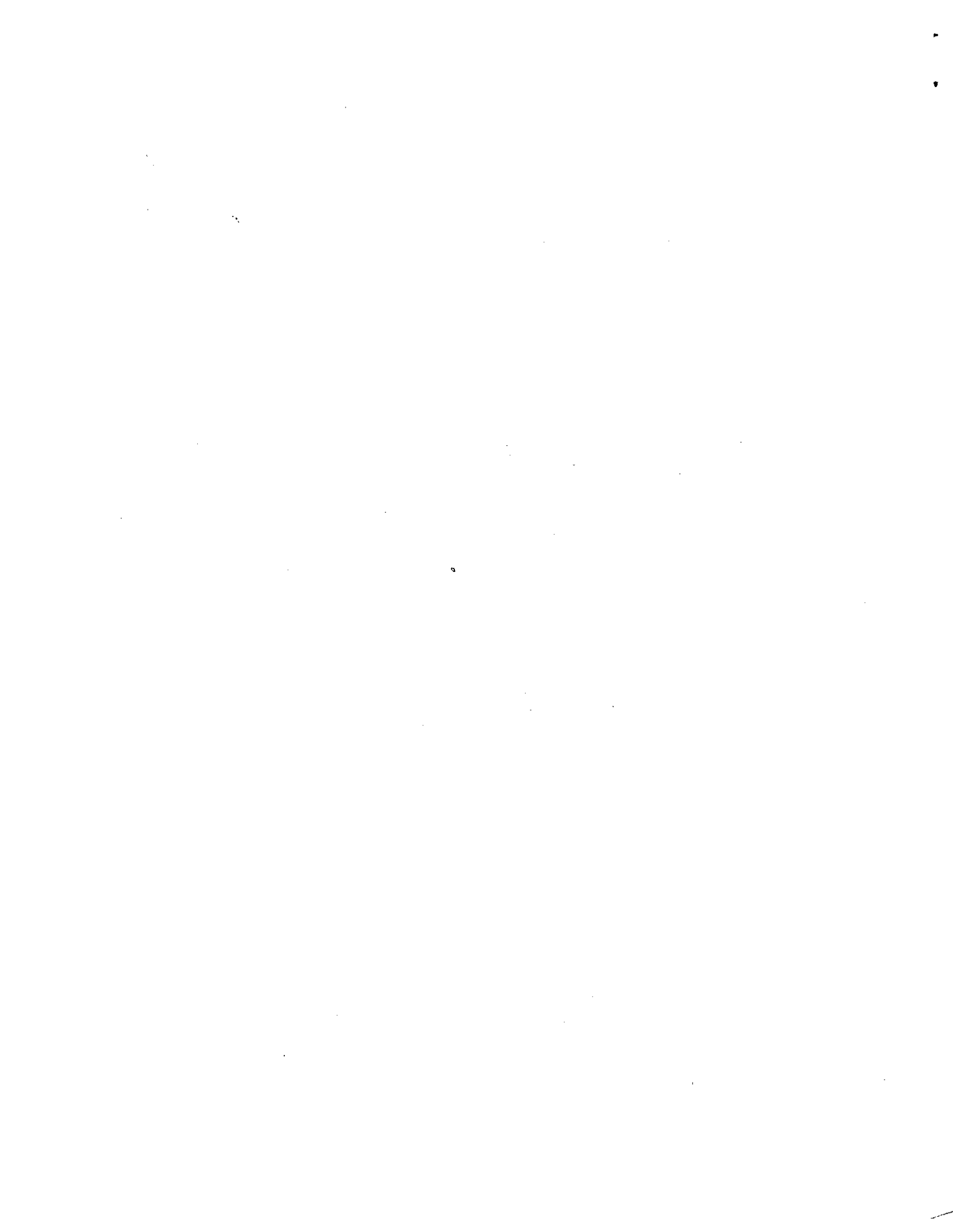
DOS PROCEDIMIENTOS DE UTILIDAD PRACTICA PARA LA EJECUCION DE
ENCUESTAS DE HOGARES

- A. Una función de fiabilidad para analizar rendimientos en ejecuciones de encuestas.
- B. Un procedimiento para generar un número aleatorio.

Carlos Cavallini */

Asesor Regional en Muestreo para
Estadísticas Demográficas
adscrito a la CEPAL

*/ Las opiniones expresadas en este documento son de exclusiva responsabilidad del autor y pueden no coincidir con las de la Organización.



A. Una función de fiabilidad para analizar rendimientos en ejecución de encuestas

1. En las distintas etapas que conforman una encuesta de hogares, pueden aparecer fallas que perturben la eficiencia de la investigación. Como ejemplo de estas fallas podemos citar: El rechazo, de un hogar seleccionado, a ser entrevistado; la pérdida de un formulario con información; la tergiversación de un dato en la entrada al proceso por computadora; etc.

Bajo el supuesto de que la probabilidad de la presentación de la falla sea pequeña, menor o igual a .1, para un número grande de casos, la ley de distribución de estas fallas vendrá dada por la fórmula asintótica de Poisson,

$$P(v, k) = \frac{(hv)^k e^{-hv}}{k!} \quad 0 \leq k \leq \infty \quad (1)$$

que indica la probabilidad de que en v casos se obtengan k fallas, siendo h una constante de la intensidad de la falla. Se llama intensidad de la falla el promedio de fallas estimadas por caso.

Se observa que $\sum_{k=0}^{\infty} P(v, k) = 1$ por ser $\sum_{k=0}^{\infty} \frac{(hv)^k}{k!} = e^{hv}$

Función de fiabilidad

2. Definiremos a la función de fiabilidad como a una función que nos permita conocer la probabilidad de que un determinado trabajo se cumpla sin fallas.

Si en la fórmula de Poisson, en (1), hacemos $k = 0$, obtenemos

$$P_v = e^{-hv} \quad 0 \leq v \leq \infty \quad (2)$$

que es una función exponencial de fiabilidad que permite hallar la probabilidad de hacer un trabajo sin fallas en un intervalo que comprenda una duración de v casos. Los supuestos de esta función son los mismos de la fórmula de Poisson.

3. La probabilidad de tener por lo menos una falla en el intervalo $(0 ; v)$, Q_v , es, por tanto

$$Q_v = 1 - e^{-hv} \quad (3)$$

En el campo continuo Q_v es una función integral que determina la probabilidad de fallo, es decir

$$Q_v = \int_0^v f(v) dv \quad (4)$$

de donde

$$\begin{aligned} f(v) &= h e^{-hv} \\ &= h P_v \end{aligned} \quad (5)$$

es la función exponencial de la distribución de probabilidad de la variable v , que cumple la condición

$$\int_0^{\infty} f(v) dv = 1 \quad (6)$$

4. Asimismo

$$\begin{aligned} \int_{v_0}^{v_1} f(v) dv &= Q_{v_1} - Q_{v_0} \\ &= P_{v_0} - P_{v_1} \end{aligned} \quad (7)$$

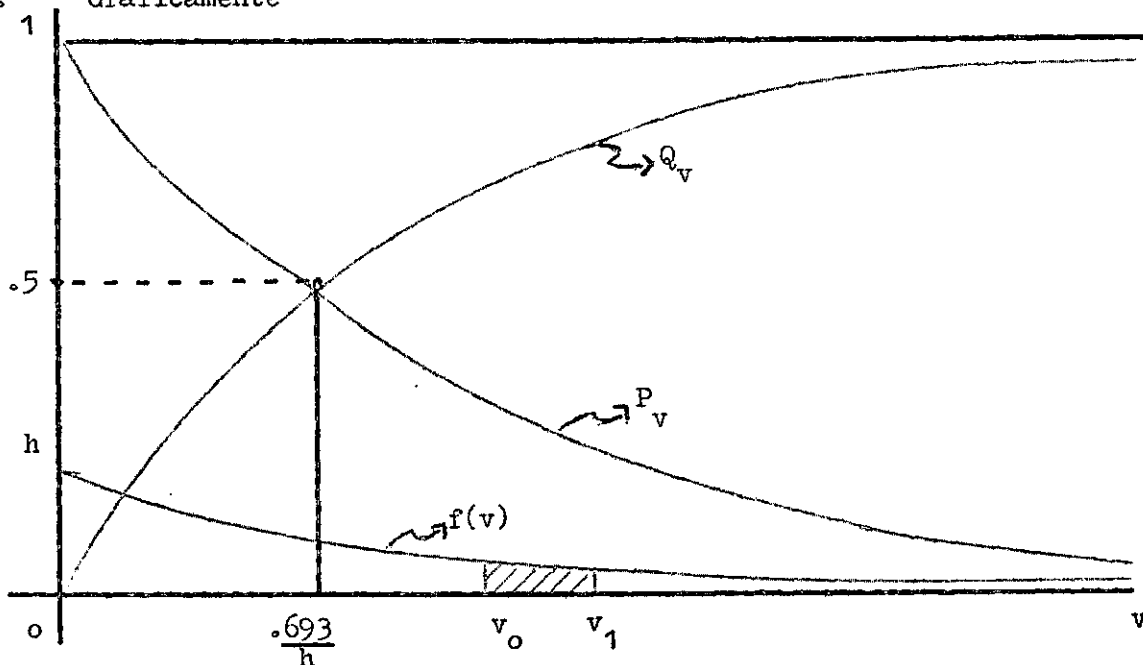
y designando con t al intervalo $v_1 - v_0$ es $v_1 = t + v_0$ y siendo por (18)

$P_{v_1} = P_t P_{v_0}$ la (7) queda $P_{v_0} (1 - P_t)$ o sea

$$\int_{v_0}^{v_1} f(v) dv = P_{v_0} Q_t \quad (8)$$

probabilidad que se trabaje sin fallo un número v_0 de casos y que se produzca fallo en el intervalo t .

5. Gráficamente



La probabilidad de tener por lo menos una falla en el intervalo $(0 ; v_0)$ es

$$Q_{v_0} = \int_0^{v_0} f(v) dv = \text{Prob}(v \leq v_0) \quad (9)$$

que es la probabilidad de fallo.

La probabilidad de no tener fallas en el intervalo $(0 ; v_0)$ es

$$P_{v_0} = 1 - \int_0^{v_0} f(v) dv \quad (10)$$

o también

$$P_{v_0} = \int_{v_0}^{\infty} f(v) dv = \text{Prob}(v > v_0) \quad (11)$$

que es la probabilidad de no tener fallas o función de fiabilidad.

En el continuo la $\text{Prob}(v = v_0) = 0$.

6. La esperanza matemática de v se define

$$E(v) = \int_0^{\infty} v f(v) dv \quad (12)$$

Integrando por partes, obtenemos

$$E(v) = \frac{1}{h} \quad (13)$$

Es decir, la esperanza matemática de la distribución exponencial es la inversa de h . Si se desconoce el parámetro h , el mismo se puede estimar con

$$h = \frac{1}{\bar{v}} \quad (14)$$

donde \bar{v} es el promedio de casos por unidad de falla y h es la intensidad de la falla por caso.

7. La varianza de v es

$$V(v) = \int_0^{\infty} v^2 f(v) dv - E^2(v) \quad (15)$$

Integrando por partes

$$V(v) = \frac{1}{h^2} \quad (16)$$

Por tanto, el desvío medio cuadrático, $\sigma(v)$, es

$$\sigma(v) = \frac{1}{h} \quad (17)$$

La esperanza matemática y el desvío medio cuadrático de la distribución exponencial son iguales entre sí.

8. La ley exponencial de la fiabilidad, e^{-hv} , es simple y la probabilidad de trabajar sin fallas en un intervalo depende solamente del número de casos v .

Por ser, de (2),

$$P_v = (P_1)^v \quad (18)$$

la probabilidad condicional de que el trabajo no tenga fallas en v_1 casos, a condición de que no haya tenido fallas en los v_0 casos precedentes, $v_1 > v_0$, es

$$\text{Prob}(v_1 / v_0) = \frac{\text{Prob}(v_1, v_0)}{\text{Prob}(v_0)} = \text{Prob}(v_1) \quad (19)$$

Es decir, el trabajo sin falla en el pasado es independiente del trabajo sin falla en el futuro.

9. Por otro lado, dado P_v se puede obtener v de la fórmula

$$v = \frac{\ln P_v}{-h} \quad (20)$$

10. Ejemplo 1: Por experiencias pasadas, se ha estimado que de cada 100 visitas a los hogares se obtiene un promedio de 4 rechazos. Es decir, que el promedio de visitas por rechazo es, aproximadamente, $\bar{v} = 25$. Por tanto, la intensidad del rechazo por visita se estima en $h = .04$. La función exponencial de probabilidad es

$$f(v) = .04 e^{-.04v} \quad v \geq 0 \quad (21)$$

Se quiere conocer la probabilidad de que un entrevistador realice
i) 10 visitas sin tener rechazos; ii) 50 visitas sin tener rechazos.
Utilizando la fórmula (2)

$$P_{10} = e^{-.4} = .67032$$

$$P_{50} = e^{-2} = .13534$$

La probabilidad de que un entrevistador no tenga rechazos al hacer 10 visitas es .67, y de que no tenga rechazos al hacer 50 visitas es .14.

Se observa que $(P_{10})^5 = P_{50}$.

11. Analicemos, ahora, el caso de que exista un efecto entre la intensidad del fallo h y el número de casos v , es decir

$$h = f(v) \quad (22)$$

Esto puede suceder, por ejemplo, con los entrevistadores en las encuestas de hogares, que a medida que ellos adquieren más experiencia en la forma de llevar a cabo las entrevistas, el h puede disminuir.

Supongamos que h adopte la siguiente función exponencial

$$h = a v^{-b} \quad (23)$$

luego, la nueva función exponencial de fiabilidad, P_v^0 , vendrá dada por

$$\begin{aligned} P_v^0 &= e^{-a v^{-b} v} \\ &= e^{-a v^{1-b}} \end{aligned} \quad (24)$$

o sea

$$\ln P_v^0 = -a v^{1-b} \quad (25)$$

12. Ejemplo 2: En una encuesta se ha estimado que en las 10 primeras visitas se espera una intensidad de rechazo $h = .10$ pero que al llegar a las 100 visitas se espera que $h = .02$. Se desea saber cual es la probabilidad de no tener rechazos en v visitas.

De (23) se obtiene

$$\left. \begin{aligned} .10 &= a 10^{-b} \\ .02 &= a 100^{-b} \end{aligned} \right\} \quad (26)$$

resolviendo

$$\begin{aligned} a &= .50 \\ b &= .70 \end{aligned} \quad (27)$$

por tanto

$$\ln P_v^0 = -.50 v^{.30} \quad (28)$$

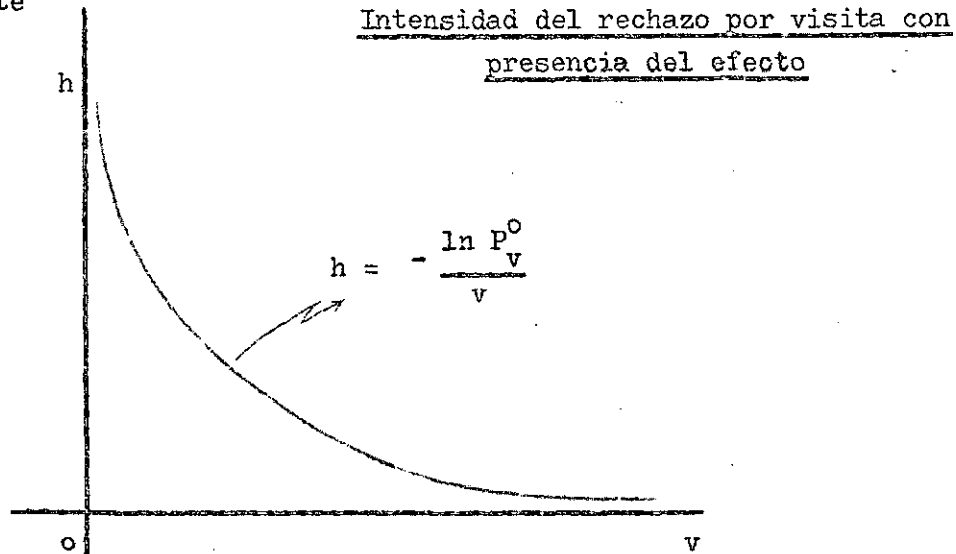
Para distintos valores de v las funciones de fiabilidad P_v^0 y P_v son

<u>v</u>	<u>P_v^0</u>	<u>h</u>	<u>P_v</u>	
			<u>h = .02</u>	<u>h = .10</u>
0	1	-	1	1
1	.61	.50	.98	.90
10	.37	.10	.82	.37
100	.14	.02	.14	$5(10)^{-5}$
200	.09	.01	.02	$2(10)^{-9}$
300	.06	.009	.002	$9(10)^{-14}$
1 000	.02	.004	$2(10)^{-9}$	$4(10)^{-44}$

13. Es decir, cuando existe un efecto anterior que se va proyectando en el trabajo futuro, la función de fiabilidad P_v^0 es más realista. En el Ejemplo 2, que supone una disminución de la intensidad de los rechazos a través del aumento del número de visitas, se observa, que bajo este efecto, la probabilidad de que un entrevistador realice 1 000 visitas sin rechazos

es igual a .02, mientras que si no se toma en cuenta este efecto, la probabilidad de hacer 1 000 visitas sin rechazos es prácticamente imposible. Asimismo, se nota en P_v^0 , la disminución de la intensidad del rechazo por visita, a medida que aumenta el número de visitas, mientras que en P_v el h es constante en cada columna. El h se obtiene de la (20).

Gráficamente



14. Ejemplo 3: En una encuesta se han dado las siguientes reglas,
- i) si no se encuentra al informante en 2 visitas, se sustituye al hogar,
 - ii) si un informante rechaza 3 veces, se sustituye al hogar,
 - iii) si no se encuentra a la vivienda en la primera visita, se la sustituye,
- y las siguientes definiciones,

n	tamaño muestral de hogares
A_1	probabilidad de no encontrar al informante en la primera visita
A_2	probabilidad de no encontrar al informante en la segunda visita
B_1	probabilidad de obtener un primer rechazo por parte del informante
B_2	probabilidad de obtener un segundo rechazo por parte del informante
B_3	probabilidad de obtener un tercer rechazo por parte del informante
C	probabilidad de detectar un error en la información recogida que origine una nueva visita al hogar

$h = A_1 + A_1 A_2 + B_1 + B_1 B_2 + B_1 B_2 B_3 + C$ intensidad de la falla por visita al hogar

D probabilidad de no encontrar la vivienda del hogar seleccionado

$m = n(A_1 A_2 + B_1 B_2 B_3 + D)$ hogares sustituidos.

luego, el número de visitas, v , que se estima que deberá realizarse para obtener la información de n hogares, es

$$v = (n + m) (1 + h) \quad (29)$$

Se desea estimar, por razones de presupuesto, la cantidad de visitas que deberán llevarse a cabo para obtener la información real sobre 10 000 hogares.

Se conoce que

$$A_1 = .05, A_2 = .001, B_1 = .02, B_2 = .25, B_3 = .50$$

$$C = .003, \text{ luego } h = .08055$$

Además, $D = .01$, y siendo $n = 10\ 000$ es $m \doteq 126$

Por tanto

$$v = 10\ 126 (1.08055) \doteq 10\ 942$$

es decir, se estima que deberán realizarse unas 10 942 visitas para obtener la información real de 10 000 hogares.

En este ejemplo se ha considerado, que de todos los hogares sustituidos se ha conseguido la información. Caso contrario, habría que contemplar a $k = n (A_1 A_2 + B_1 B_2 B_3 + D)^2$, es decir los hogares que se espera sustituir de m . En este caso la (29) queda

$$v = (n + m + k) (1 + h) \quad (30)$$

B. Un Procedimiento Práctico para Generar un Número Aleatorio

1. En los trabajos operativos, muchas veces se hace necesario, en el terreno, generar números aleatorios. Esto sucede, habitualmente, en las áreas rurales, cuando se actualizan las unidades seleccionadas de primera etapa, o áreas geográficas, y luego se deben seleccionar, dentro de dichas áreas seleccionadas, las unidades o viviendas que han sido enumeradas en la actualización. En estos casos, el entrevistador debe a veces permanecer en el terreno hasta completar el trabajo de enumeración, selección y entrevista de viviendas. Por tanto, se ve enfrentado, una vez hecho el trabajo de enumeración, a tener que seleccionar un número dado de viviendas. Para ello, generalmente, y en función del número de viviendas a seleccionar, divide a las viviendas enumeradas en intervalos de selección. Luego debe generar, dentro de un intervalo, un arranque aleatorio, el cual le permitirá, en forma sistemática, seleccionar el número establecido de viviendas.

2. Cuando no se cuenta con una tabla de números al azar, un procedimiento práctico para generar un número aleatorio, acotado dentro de determinados límites, puede ser el siguiente, utilizando, por ejemplo, una moneda o similar:

i) Simbolizando con A al número aleatorio a generar, éste debe cumplir con la condición que $1 \leq A \leq N$ siendo 1 y N los límites inferior y superior, respectivamente, de selección.

ii) Dado que la condición del evento "arrojar una moneda" puede solamente producir el suceso binomial "cara o cruz", la cantidad de grupos distintos que puede producir el suceso, arrojando la moneda n veces, es 2^n .

iii) Cada grupo identificará, así, a un número aleatorio en forma biunívoca.

iv) Simbolizando con E_n al evento, definimos

$$E_n = \left. \begin{array}{l} 1 \text{ si sale cara} \\ \\ 0 \text{ si sale cruz} \end{array} \right\} \quad (1)$$

v) El número aleatorio A responderá a la ecuación

$$A = 1 + \sum_{n=1}^n 2^{n-1} E_n \quad n \geq 1 \quad (2)$$

Para $E_n = 1$ es $A = 2^n$, valor máximo de A.

vi) La determinación de n, número de veces que debe efectuarse el evento E_n , se obtiene de la siguiente condición,

$$2^{n-1} \leq N - 1 < 2^n \quad (3)$$

Prácticamente, es el n por defecto, del intervalo que comprende a $N - 1$ el desarrollo de la serie 2^{n-1} .

vii) En este caso, si el número aleatorio generado, A, resulta mayor a N, el experimento se debe repetir.

4. Para no tener que, ocasionalmente, repetir la experiencia, es posible hallar el número aleatorio B, el cual responde a la siguiente expresión,

$$B = \frac{(N-1)(A-1)}{2^n - 1} + 1 \quad (4)$$

que siempre cumple la condición $1 \leq B \leq N$

5. Ejemplo: Generar un número aleatorio entre 1 y 174.

$$N = 174$$

$$N - 1 = 173$$

$$n = 8, \text{ que cumple con la condición (3) dado}$$

que $2^7 = 128$ y $2^8 = 256$. Supongamos que arrojando la moneda 8 veces se obtengan los sucesos: CXXXCXCC (donde C = cara y X = cruz). Luego,

$$A = 1 + 2^0 + 2^4 + 2^6 + 2^7 = 210$$

y

$$B = \frac{173(209)}{255} + 1 = 142.79$$

o sea

$$B = 143$$

6. Si los sucesos hubiesen sido, por ejemplo, CCCCCCCC, luego

$$\begin{aligned} A &= 1 + 2^0 + 2^1 + 2^2 + 2^3 + 2^4 + 2^5 + 2^6 + 2^7 \\ &= 256 \end{aligned}$$

y

$$\begin{aligned} B &= \frac{173 (255)}{255} + 1 = \\ &= 174 \end{aligned}$$

o sea, el límite superior del intervalo de selección.

7. Si los sucesos hubiesen sido, por ejemplo XXXXXXXX, luego,

$$A = 1$$

y

$$\begin{aligned} B &= \frac{173 (0)}{255} + 1 \\ &= 1 \end{aligned}$$

o sea, el límite inferior del intervalo de selección.

8. Debido, en algunos casos, al redondeo de B, algunos números del intervalo de selección, pueden tener más de una posibilidad de ser seleccionados, pero para los fines prácticos que se persiguen este hecho se puede despreciar. Si se quiere obviar esta circunstancia, habrá que utilizar como número de selección a A, el cual otorga la misma posibilidad a todos los números del intervalo de selección.

