RESTRINGIDA

E/CEPAL/R.246
2 de marzo de 1981

ORIGINAL: ESPAÑOL

C E P A L

Comisión Económica para América Latina

EL EFECTO DEL ENTREVISTADOR EN LAS ENCUESTAS DE HOGARES

Carlos Cavallini */
Asesor Regional en Muestreo para
Estadísticas Demográficas

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

81-1-0092-170

EL EFECTO DEL ENTREVISTADOR EN LAS ENCUESTAS DE HOGARES

1. En las encuestas donde se usan entrevistadores para recoger la información, se hace necesario conocer el grado de estandardización que los mismos tienen. Si el trabajo de los entrevistadores no es estándar, es decir, si existen diferencias entre ellos, con respecto a la forma de llevar a cabo el trabajo o diferencias en la interpretación de los conceptos que se quieren medir o diferencias en el tratamiento de la unidad a medir, los resultados que se obtengan estarán influídos por los criterios propios del entrevistador.

El entrevistador puede así introducir, en la información, un factor de error, que se conoce como el efecto del entrevistador. Mahalanobis considera que estas modificaciones introducidas, se deben a la personalidad, al entrenamiento y a la educación del entrevistador, por lo cual le llama la ecuación personal del entrevistador. El efecto del entrevistador puede, por tanto, ser muy distinto de un entrevistador a otro.

2. Para estudiar el efecto del entrevistador supongamos,

i) tener una super-población compuesta de M unidades muestrales. Esta super-población ha sido dividida, aleatoriamente, en K poblaciones de N unidades cada una, siendo $M = NK$, donde K es la población de entrevistadores,

ii) de las K poblaciones se selecciona una muestra de k poblaciones y dentro de cada población de N unidades se selecciona una muestra de n unidades,

iii) se tiene así constituído un diseño muestral completamente aleatorizado de k muestras de n unidades cada una, estando cada muestra afectada por un entrevistador,

iv) este diseño es equivalente a seleccionar una muestra de m unidades de la super-población de M unidades y luego dividir aleatoriamente la muestra en k submuestras interpenetrantes, cada una del mismo número de unidades $n = \frac{m}{k}$. Luego, seleccionar una muestra de k entrevistadores de una población de K entrevistadores y asignar aleatoriamente un entrevistador a cada submuestra,

v) dentro de la i - población, $Y_{ij} \sim N(u_i; \sigma_o^2)$

donde

Y_{ij} valor real de ij - unidad con $i = \overline{1;K}$ y $j = \overline{1;N}$

$u_i = \frac{1}{N} \sum_j^N Y_{ij}$ media parámetro real de la i-población

siendo la varianza, σ_o^2, igual en las K poblaciones, supuesto de homosedasticidad.

El símbolo $\sim$ significa "se distribuye como",

vi) las unidades muestrales seleccionadas son medidas una sola vez por el entrevistador correspondiente. Para el valor observado en la ij - unidad se establece el siguiente modelo

$$y_{ij} = Y_{ij} + g_i \qquad (1)$$

donde

y_{ij} valor obtenido por el i - entrevistador en la j - unidad, con $i = \overline{1;k}$ y $j = \overline{1;n}$

Y_{ij} valor real de la ij - unidad

g_i efecto debido al i - entrevistador

vii) el efecto del i - entrevistador, g_i, es constante en cada unidad medida por el i - entrevistador, con $\underset{j}{E}\, g_i = g_i$,

viii) luego

$$\bar{y}_i = \bar{Y}_i + g_i \qquad (2)$$

donde

$\bar{y}_i = \frac{1}{n} \sum_j^n y_{ij}$ media muestral obtenida por el i - entrevistador

$\bar{Y}_i = \frac{1}{n} \sum_{j}^{n} Y_{ij}$ media real muestral en la i - submuestra,

ix) el efecto del entrevistador es variable entre los entrevistadores, siendo,

$$\bar{y} = \bar{Y} + \bar{g} \qquad (3)$$

donde

$\bar{y} = \frac{1}{k} \sum_{i}^{k} \bar{y}_i$ media muestral obtenida en el diseño muestral

$\bar{Y} = \frac{1}{k} \sum_{i}^{k} Y_i$ media muestral real del diseño muestral

$\bar{g} = \frac{1}{k} \sum_{i}^{k} g_i$ efecto promedio de los entrevistadores del diseño muestral

x) el efecto del entrevistador tiene un carácter aditivo en el valor observado y_{ij},

xi) entre las poblaciones $g_i \sim N\ (G\ ;\sigma_1^2)$ donde,

$$\underset{i}{E}\ g_i = G \quad y \quad \sigma_1^2 = \frac{1}{K} \sum_{i}^{K} (g_i - G)^2$$

3. Restando (3) de (2) se puede escribir

$$\bar{y}_i - \bar{y} = (\bar{Y}_i - \bar{Y}) + (g_i - \bar{g}) \qquad (4)$$

o sea, que el efecto total observado en la i - submuestra ha sido descompuesto en un efecto propio de los valores reales, no explicable, más un efecto debido al i - entrevistador. Lo que se desea conocer, en rigor, es si los entrevistadores trabajan con efectos propios que son, entre los entrevistadores, significativamente diferentes. Es decir, si se está trabajando con varas de medición distintas. Si así fuera, significaría que la información recogida está afectada por una variación

debida al efecto del entrevistador. En este caso habría que recurrir a un control exhaustivo de los entrevistadores para detectar a aquellos que no realizan una labor normal y que necesitan ser re-entrenados. El efecto del entrevistador es distinto al sesgo que pueda tener el estimador. 1/ El efecto del entrevistador tiene que ver con la existencia de una variación significativa entre las medias observadas por los distintos entrevistadores. De allí que al efecto del entrevistador también se le llama la variabilidad del entrevistador. Para estimar la significación de este efecto, se plantea la hipótesis de nulidad, de que los entrevistadores no afectan distintamente a la información que recogen. Es decir, se plantea la hipótesis

$$Ho : g_1 = g_2 = \dots = g_i = \dots = g_K$$

o lo mismo

$$Ho : \sigma_1^2 = 0$$

Hallando los cuadrados, sumando sobre i y dividiendo por (k - 1), la expresión (4) puede ser escrita como

$$\text{var}(\bar{y}_i) = \text{var}(\bar{Y}_i) + \text{var}(g_i) + 2\,\text{cov}(\bar{Y}_i ; g_i) \quad (5)$$

donde "var" significa "varianza estimada de" y "cov" significa "covarianza estimada de". Dado el supuesto que el efecto g_i es constante en el i - entrevistador, cualquiera sea el valor de Y_{ij}, la correlación entre g_i e Y_{ij} es nula. Por tanto, suponemos independencia entre $\bar{Y}_i$ y g_i y expresar que cov $(\bar{Y}_i ; g_i) = 0$, con lo cual la (5) queda

$$\text{var}(\bar{y}_i) = \text{var}(\bar{Y}_i) + \text{var}(g_i) \quad (6)$$

Es decir, hemos descompuesto la varianza estimada de las medias observadas en 2 componentes aditivos, que son, la varianza estimada de las medias reales y la varianza estimada del efecto del entrevistador.

1/ Zarcovich, S.S., Quality of Statistical Data, FAO, Rome 1966

4. En este punto conviene observar que ambas varianzas estimadas, var $(\bar{y}_i)$ y var $(\bar{Y}_i)$, serán iguales, sólo si el efecto del entrevistador no existe, es decir si var $(g_i) = 0$. Ahora bien, pareciera que ambas varianzas estimadas de medias se pudieran escribir como la varianza estimada de la correspondiente variable dividida por el número de observaciones. En este caso la varianza de la variable y_{ij}, o sea de $Y_{ij} + g_i$, es la misma varianza que de la variable Y_{ij}, dado que g_i es constante para los distintos j de un i. Pero esto sólo es válido cuando las variaciones de la variable son debidas a causas aleatorias, no explicables por el modelo estadístico establecido. Dicho de otra forma, var ($\bar{x}$) es igual a $\frac{\text{var}(x_j)}{n}$, siendo $\bar{x} = \frac{1}{n} \sum_j^n x_j$, siempre que la varianza de las $\bar{x}$'s se deba a las variaciones propias del muestreo y de que las variaciones de la x_j se deban a las variaciones propias de la población. Pero si las $\bar{x}$'s se ven perturbadas por un factor ajeno a la población, que puede explicarse, la igualdad no es válida.

5. En nuestro caso $Y_{ij} \sim N(u_i ; \sigma_o^2)$ e $y_{ij} \sim N(u_i + g_i ; \sigma_o^2)$, pero si bien ambas variables se distribuyen con una igual σ_o^2, sólo var $(\bar{Y}_i) = \frac{\sigma_o^2}{n}$. Por tanto, la (6) puede ser escrita

$$\text{var}(\bar{y}) = \frac{s_o^2}{n} + \text{var}(g_i) \qquad (7)$$

donde $s_o^2 = \frac{1}{k(n-1)} \sum_i^k \sum_j^n (Y_{ij} - \bar{Y}_i)^2$ es la varianza estimada de σ_o^2.

Como no se conocen los desvíos $(Y_{ij} - \bar{Y}_i)$ se puede calcular s_o^2 utilizando los desvíos observados $(y_{ij} - \bar{y}_i)$, pues ambos son iguales por ser g_i constante para el i - entrevistador.

6. Dado que s_o^2 es un estimador de σ_o^2 y var (g_i) es un estimador de σ_1^2 podemos usar el cociente de n var $(\bar{y}_i)$ sobre s_o^2, que se distribuye como una F con (k-1) y k (n-1) grados de libertad, para estimar $\frac{\sigma_o^2 + n\sigma_1^2}{\sigma_o^2}$.

Si la hipótesis de nulidad H_o: $\sigma_1^2 = 0$ es cierta, luego F estima a 1. Si la hipótesis de nulidad es falsa, luego F estima a una cantidad mayor que 1.

Comparando el valor de F obtenido con los correspondientes valores de F_o tabulados, para un determinado nivel de significación, se puede conocer la probabilidad de tener un F mayor o igual a F_o, y adoptar un criterio para rechazar o no rechazar la H_o.

7. Lo expuesto se puede resumir en el siguiente cuadro del análisis de las variaciones.

Fuente de variación	Grados de libertad	Suma de cuadrados	Suma de cuadrados medios	Parámetros estimados por la suma de cuadrados medios
Entre entrevistadores	k - 1	$B = n \sum_i^k (\bar{y}_i - \bar{y})^2$	$\frac{B}{k-1}$	$\sigma_o^2 + n\sigma_1^2$
Dentro de la submuestra	k(n-1)	$A = \sum_i^k \sum_j^n (y_{ij} - y_i)^2$	$\frac{A}{k(n-1)}$	σ_o^2
Total	nk - 1	A + B	-	-

8. El análisis de la varianza aquí presentado se utiliza para establecer si el efecto de los entrevistadores, que tiene distintos niveles, ejerce una influencia considerable en los resultados que se estudian. La idea fundamental de este análisis es la comparación de la "variación del efecto", ocasionada por la acción de los entrevistadores y la "variación residual" debida a causas aleatorias no explicables. Si la diferencia entre estas dos variaciones es significativa ello indicará que los resultados estarían afectados por la influencia de los entrevistadores. En este caso se hará necesario aclarar cual de los niveles ejerce una acción máxima.

9. Bibliografía

- Dunn, O.J., Clark, V.A. Applied Statistics: Analysis of Variance and Regression, John Wiley and Sons, New York, 1974.
- Davies, O.L., Statistical Methods in Research and Production, Oliver and Boyd, London, 1958.
- Konijn, H.S., Statistical Theory of Sample Survey Design and Analysis, North-Holland Publishing Company, Amsterdam, 1973.