

24/4/79.

① ~~Dr. Donato~~
② ~~Dr. Boaz~~
③ DOCPAL

RESERVADO

Abril de 1979

R

INFORME DE LA MISION DE ASESORIA REALIZADA EN LA REPUBLICA
DE PANAMA DESDE EL 26 HASTA EL 31 DE MARZO DE 1979

Carlos Cavallini

Asesor Regional en Muestreo para Estadísticas
Demográficas adscrito a la CEPAL

"El autor de este informe es el único responsable de las opiniones expresadas en él. Se han presentado copias a la Oficina de Cooperación Técnica de las Naciones Unidas, la cual, a su debido tiempo, podrá comunicar al Gobierno sus propias conclusiones y recomendaciones."

REPUBLICA DE PANAMA

1. Con el objeto de continuar con el programa de investigación, relacionado con la Encuesta de Hogares de Panamá que la Dirección de Estadística y Censo - DEC - está llevando a cabo, se cumplió una misión de asesoría técnica, por un período de una semana, entre el 26 y 31 de marzo de 1978, en la Ciudad de Panamá. El lugar físico de trabajo fue en la DEC.
2. El pedido oficial para esta misión fue hecho por la DEC a CEPAL a través de la Oficina del Programa de las Naciones Unidas para el Desarrollo en Panamá - PNUD.
3. Como en la oportunidad anterior, se tuvo la suerte de contar, como contraparte nacional, con la Est. María Teresa de León, Jefa de la Sección de Muestreo, cuyos amplios conocimientos estadísticos le hacen acreedora a las eficientes técnicas de aplicación práctica que la DEC está desarrollando en el campo muestral.
4. De acuerdo con el nuevo diseño muestral elaborado en la misión anterior, ver informe de julio de 1977, se levantó la encuesta de hogares en el área urbana de la región metropolitana en marzo de 1978. Este nuevo diseño muestral contempló la aplicación del esquema de muestras replicadas, el cual permite, entre otros, i) obtener varias estimaciones para una característica dada, lo cual otorga a simple vista una apreciación de la bondad de la técnica muestral empleada; ii) poseer varias muestras representativas de la población en igual grado, las cuales pueden ser usadas en forma conjunta o en forma individual. Por ejemplo, si determinadas características, que no necesitan un tamaño muestral grande, son necesarias medirlas en algunas ocasiones, se pueden utilizar en estas ocasiones algunas de las muestras y utilizar a todas las muestras cuando además se deseen medir características que necesariamente necesitan un tamaño mayor muestral para una determinada confiabilidad; iii) fácil cálculo de las varianzas de los estimadores; iv) estimar el sesgo del entrevistador; v) hacer un análisis de la varianza de manera tal de conocer las fuentes de variación que afectan a los resultados y poder medir estas variaciones con el objeto de controlarlas.

5. En el cuadro siguiente se muestra, para 5 conceptos, los resultados obtenidos en cada una de las 10 replicaciones, el promedio por concepto de las 10 replicaciones, el coeficiente de variación de cada promedio y el promedio o tasa por concepto obtenido de la muestra sin usar replicaciones. Se acota, que la diferencia existente entre el promedio de las replicaciones y el promedio de la muestra se debió a que las replicaciones no mantuvieron el mismo tamaño muestral. En el caso en que todas las replicaciones sean de igual tamaño, estos dos promedios serán iguales.

/Concepto

Replicaciones Tasas %

Concepto											Promedio de las 10 repli- caciones %	Coefi- ciente de varia- ción %	Tasa de la mues- tra %
	1	2	3	4	5	6	7	8	9	10			
1. Activos/15 y + años	54.94	54.56	51.56	56.32	51.15	56.44	55.06	53.40	56.41	58.38	54.82	1.3090	54.69
2. Inactivos/15 y + años	45.06	45.44	48.44	43.68	48.85	43.56	44.94	46.60	43.59	41.62	45.18	1.5880	45.31
3. Desocupados/ Activos	6.50	6.80	7.14	6.33	4.69	7.68	7.14	5.42	6.42	7.83	6.60	4.6450	6.58
4. Ocupados/ Activos	93.50	93.20	92.86	93.67	95.31	92.32	92.86	94.58	93.58	92.17	93.40	.3280	93.42
5. Hombres deso- cupados/Deso- cupados	58.82	51.43	51.52	48.39	60.00	56.76	65.71	60.87	65.38	55.56	57.44	3.2240	57.14

6. De 18 variables analizadas el coeficiente de variación fluctuó entre un .1240% para la tasa de "Inactivos puros/Inactivos" y un 23.7350% para la tasa de "Imposible encontrar trabajo/Desocupados".

7. En 12 de 18 casos el coeficiente de variación fue menor al 10%.

8. Las diferencias entre las estimaciones provenientes de los promedios de las 10 replicaciones y las estimaciones provenientes de la muestra sin replicar prácticamente se pueden despreciar. La diferencia máxima absoluta se encontró para la tasa "Desocupados puros/Desocupados" que fue de .40%. (52.78% para el promedio de las 10 replicaciones y 52.38% para la estimación de la muestra).

9. De acuerdo con la distribución t para 9 grados de libertad se obtuvieron los siguientes valores de confiabilidad para la característica desocupados/activos,

t	Nivel de rechazo	Límites $\hat{z} = \hat{p} \pm t \hat{\sigma}(\hat{p})$	
		mínimo	máximo
3.250	.01	.0559	.0761
4.781	.001	.0512	.0808

donde $\hat{p} = .066$ y $\hat{\sigma}(\hat{p}) = .0031$

por ser $\hat{cv}(\hat{p}) = \frac{\hat{\sigma}(\hat{p})}{\hat{p}} = \frac{.0031}{.066} = .04645$

de acuerdo con los resultados de la investigación. Se observa que al nivel de rechazo del 1% las replicaciones 5, 6, 8 y 10 caen fuera de los límites mientras que al nivel de rechazo del 1°/100 sólo la replicación 5 cae fuera de los límites de confiabilidad.

10. En función de estos resultados y de los coeficientes de variación obtenidos, se propuso reducir el tamaño muestral, dado que el esquema aplicado de replicar la investigación resultó altamente sensitivo.

/11. Siendo

11. Siendo $\hat{p} = .066$ y $\hat{\sigma}(\hat{p}) = .0031$, el desvío absoluto d , aceptado con una confiabilidad del 95%, es

$$d = t \hat{\sigma}(\hat{p}) = 2 (.0031) = .0062$$

y el desvío relativo e es

$$e = \frac{d}{\hat{p}} = \frac{.0062}{.066} = .09$$

menor al 10%, lo cual es significativamente pequeño para este tipo de investigación.

12. Por tanto, se decidió reducir en un cuarto el tamaño muestral. Es decir, reducir el tamaño de 3 200 viviendas a 2 400. En este caso es

$$d^2 = t^2 \frac{p \cdot q}{n} = 4 \cdot \frac{(.066) (.934)}{2400} = .0001$$

o sea,

$$d = .01$$

lo cual es realista. Por otro lado se espera trabajar con un coeficiente de variación estimado en alrededor de un 7%, por ser

$$\hat{cv}(\hat{p}) = \sqrt{\frac{p}{np}} = \sqrt{\frac{.934}{2400 (.066)}} \approx 7\%$$

Sugerencias

13. Dado los resultados obtenidos en el área urbana de la región metropolitana, los cuales han permitido adecuar el tamaño muestral de viviendas a niveles más óptimos sin desmerecer la confiabilidad esperada de la investigación, lo cual implica un ahorro efectivo de costo, se sugiere,

i) mantener el mismo esquema del muestreo replicado con las modificaciones propuestas; en la próxima encuesta a llevarse a cabo en el mes de agosto de 1979 en el área urbana de la región metropolitana;

ii) en base a los resultados que se obtengan en agosto de 1979 presentar un informe con las conclusiones y sugerencias que el esquema muestral replicado modificado haya merecido,

/iii) extender

iii) extender el esquema muestral replicado, a partir de agosto de 1979, al resto del área urbana y al área rural,

iv) estudiar la posibilidad de publicar, posiblemente en forma conjunta con la División de Estadística y Análisis Cuantitativo de la CEPAL, un documento sobre el método aplicado, el cual podría ser de amplio interés para las Oficinas Nacionales de Estadística del área latinoamericana.

Análisis de la Varianza

14. Existen situaciones en las cuales un resultado o una observación está sujeto a numerosas fuentes de variación. La técnica estadística del análisis de la varianza permite separar y estimar estas fuentes de variación.

15. De acuerdo con el esquema operativo adoptado en esta investigación cada valor observado es la suma de cinco partes: i) una media parámetro; ii) un desvío del efecto de la replicación; iii) un desvío del efecto del entrevistador; iv) un desvío del efecto de la unidad primaria de selección UP y v) un desvío del efecto de un elemento aleatorio que se distribuye normalmente con media cero y desvío estándar σ .

16. El modelo lineal aditivo estadístico planteado es el siguiente,

$$y_{ijlh} = u + t_j + b_{ij} + d_{ijl} + e_{ijlh}$$

donde

$$e_{ijlh} \sim N(0, \sigma^2)$$

con

$$h = \overline{1;v}$$

vivienda

$$l = \overline{1;w}$$

unidad primaria - UP -

$$j = \overline{1;k}$$

replicación

$$i = \overline{1;m}$$

entrevistador

/siendo

siendo

y_{ijlh}

valor observado en la $ijlh$ - vivienda

u

media parámetro

t_j

efecto entre replicaciones

b_{ij}

efecto entre entrevistadores dentro de la replicación

d_{ijl}

efecto entre UP dentro de la celda entrevistador - replicación

e_{ijlh}

efecto aleatorio debido fundamentalmente a las características peculiares de la $ijlh$ - vivienda

17. Los componentes de la varianza los obtenemos elaborando la siguiente identidad algebraica,

$$\begin{aligned}
 & \sum_i^m \sum_j^k \sum_l^w \sum_h^v (y_{ijlh} - \bar{y})^2 \\
 &= \sum_i^m \sum_j^k \sum_l^w \sum_h^v (y_{ijlh} - \bar{y}_{ijl})^2 \\
 &+ v \sum_i^m \sum_j^k \sum_l^w (\bar{y}_{ijl} - \bar{y}_{ij})^2 \\
 &+ v w \sum_i^m \sum_j^k (\bar{y}_{ij} - \bar{y}_j)^2
 \end{aligned}$$

$$+ v w m \sum_j^k (\bar{y}_j - \bar{y})^2$$

+ 0 (los dobles productos se anulan por haber en cada sumando sumas de desvíos iguales a cero).

18. En nuestro caso la simbología empleada corresponde a los siguientes conceptos,

y_{ijlh}

proporción de desocupados en la
ijlh - vivienda

$$\bar{y}_{ijl} = \frac{1}{v} \sum_h^v y_{ijlh}$$

media de la proporción de desocupados
en la ijl - UP

$$\bar{y}_{ij} = \frac{1}{v w} \sum_l^w \sum_h^v y_{ijlh}$$

$$= \frac{1}{w} \sum_l^w \bar{y}_{ijl}$$

media de la proporción de desocupados
en la ij - celda

$$\bar{y}_j = \frac{1}{v w m} \sum_i^m \sum_l^w \sum_h^v y_{ijlh}$$

$$= \frac{1}{m} \sum_i^m \bar{y}_{ij}$$

media de la proporción de desocupados
en la j - replicación

$$\bar{y} = \frac{1}{v w m k} \sum_j^k \sum_i^m \sum_l^w \sum_h^v y_{ijlh}$$

$$= \frac{1}{k} \sum_j^k \bar{y}_j$$

media total de la proporción de
desocupados en el esquema

19. Las fuentes de variación, la suma de los cuadrados, los grados de libertad y los cuadrados medios pueden presentarse en la siguiente forma de tabulado,

Fuentes de variación	Suma de los cuadrados	Grados de libertad	Cuadrados medios
Entre replicaciones	$A = v w m \sum_j^k (\bar{y}_j - \bar{y})^2$	$k - 1$	$\frac{A}{k - 1}$
Entre entrevistadores dentro de las replicaciones	$B = v w \sum_i^m \sum_j^k (\bar{y}_{ij} - \bar{y}_j)^2$	$(m-1) k$	$\frac{B}{(m-1) k}$
Entre UP dentro de las celdas	$C = v \sum_i^m \sum_j^k \sum_l^w (\bar{y}_{ijl} - \bar{y}_{ij})^2$	$(w-1) mk$	$\frac{C}{(w-1) mk}$
Dentro de la UP	$D = \sum_i^m \sum_j^k \sum_l^w \sum_h^v (y_{ijlh} - \bar{y}_{ijl})^2$	$(v-1) wmk$	$\frac{D}{(v-1) wmk}$
Total	$\sum_i^m \sum_j^k \sum_l^w \sum_h^v (y_{ijlh} - \bar{y})^2$	$mkwv - 1$	

20. La variación total es la suma de las variaciones de las 4 fuentes de variación. Es decir, la variación total ha sido descompuesta en componentes aditivos de variación. Cada uno de estos componentes responde a un efecto o fuente de variación. El análisis de la varianza permite estimar la contribución que hace a la varianza total cada uno de estos componentes de variación. Luego de estimarlos se puede aplicar una docimasia o prueba de hipótesis para conocer si esta contribución es significativa. Conociendo cuáles son las fuentes de variación que influyen significativamente en la variación total de un resultado se está en condiciones de poder, por ejemplo, decrecer el tamaño de las unidades muestrales que tienen un pequeño componente de variación y aumentar el tamaño de las unidades muestrales que posean un componente de variación grande. De esta manera se logrará mayor eficiencia de información por unidad de costo.

21. En la presente investigación se desea conocer, entre otros, en qué grado la variación total del estimador se debe,

i) a la variación entre replicaciones. Si esta variación es pequeña significativamente, se estará en condiciones de estudiar la reducción del tamaño muestral por replicación. En este caso, reducir el número de replicaciones tal vez no resulte práctico, dado que los grados de libertad son pocos, $k = 10$;

ii) a la variación entre los entrevistadores. Si este componente es significativamente grande, habrá que estudiar las causas de ello, dado que el mismo podría deberse al sesgo introducido por uno o varios entrevistadores;

iii) a la variación entre UP. En este caso se estará en condiciones de adecuar el tamaño muestral de UP;

iv) a una variación aleatoria entre las viviendas dentro de las UP. Esta componente servirá para adecuar el tamaño muestral de viviendas dentro de las UP.

22. Con la información levantada en la encuesta del área metropolitana, en marzo de 1978, la Sección de Muestreo elaborará un análisis de varianza, el cual permitirá analizar las variaciones que influyen en la variación total de los resultados y tomar decisiones con respecto al diseño muestral empleado.

