

**MINISTERIO DE HACIENDA  
INSTITUTO NACIONAL DE ESTADISTICA  
SUBDIRECCION DE ESTADISTICAS SOCIALES  
DIVISION ENCUESTA INTEGRADA DE HOGARES (EIH)**

**SEMINARIO  
RESULTADOS DE EMPLEO  
SERIE 1989 - 1995**

**METODOLOGIA DE AJUSTE DE LOS  
FACTORES DE PONDERACION**

**RAFAEL TRIGUEROS MEJIAS  
CONSULTOR EXTERNO DE ESTADISTICA**

**LA PAZ, 12 DE JULIO DE 1996**

# ENCUESTA INTEGRADA DE HOGARES

## METODOLOGIA

### AJUSTE DE LOS FACTORES DE PONDERACION

#### INTRODUCCION

El Instituto Nacional de Estadística ha realizado Encuestas de Hogares en forma anual hasta la fecha a partir de 1989, con el propósito de obtener principalmente información relativa a empleo, las cuales estaban dirigidas a la población residentes en las ciudades capitales del país y en la ciudad de El Alto.

Durante el transcurso de la ejecución de las encuestas de hogares (1989-1995) se utilizaron marcos muestrales distintos y diferentes períodos de referencia para obtener dicha información.

Lo indicado en el párrafo anterior dio origen a algunas variaciones en cuanto a cobertura en cada una de las ciudades de estudio y, por tal razón se consideró necesario realizar pruebas estadísticas para evaluar la validez de cada estimación antes de iniciar la construcción de la serie.

La metodología planteada fue diseñada, asesorada y supervisada por el Consultor Externo en Estadística, Dr. Rafael Trigueros, el cual recomendó utilizar la estructura paramétrica del Censo del 92 por sexo y por tramos decenales de edad, para cada ciudad con la finalidad de conocer si las estimaciones muestrales estadísticamente se pueden considerar como provenientes de la misma población o no.

#### LA PRUEBA KOLMOGOROV-SMIRNOV PARA LAS MUESTRAS EXPANDIDAS

La prueba de Kolmogorov-Smirnov puede rechazar como cierta el que una muestra ha sido extraída de una población con una distribución estructural dada.

Esta prueba estudia los puntos de coincidencia entre dos distribuciones acumulativas muestrales. Si una muestra ha sido seleccionada de un

Marco Muestral que represente a una población, es de esperar que las estimaciones resultantes de la muestra y la población representada, tenga la misma estructura. Un alejamiento excesivo de las distribuciones acumulativas sugiere que corresponden a poblaciones diferentes. Una desviación bastante grande en las distribuciones acumulativas, es razón de rechazar, que estas distribuciones muestrales previenen de una misma población ( $H_0$ , hipótesis nula).

## EL METODO

Al aplicar la prueba de Kolmogorov-Smirnov, se construye la distribución de frecuencias acumulativas de ambas distribuciones, usando los mismos intervalos para las dos distribuciones, se resta para cada intervalo cada frecuencia acumulada de la otra. La prueba examina la mas grande de estas desviaciones (diferencias en valor absoluto) observadas.

Sea;

$F_{n_i}(X_{ij})$ : La frecuencia acumulativa que se ha observado en el intervalo  $j$  de la distribución  $i$ ,

$$F_{n_i}(X_{ij}) = \frac{N_{ij}}{n_i}$$

Donde;

$i$ : Distribución muestral  $i$ ;  $i = 1, 2$

$j$ : Intervalo  $j$  en la Distribución  $i$ ;  $j = 1, 2, \dots, k$

$N_{ij}$ : Es la frecuencia acumulada de  $X_{ij}$ , en la distribución  $i$  del intervalo  $j$ .

$n_i$ : Es el total de observaciones, en la distribución  $i$ .

La prueba examina la diferencia:

$$D = \max |F_{n_1}(X_{1j}) - F_{n_2}(X_{2j})|$$

Para una prueba de dos colas encontramos el valor máximo absoluto de  $D$ , es decir que la desviación máxima es independiente de su dirección. Donde la hipótesis  $H_1$  (hipótesis alterna), indica que las dos distribuciones proceden de diferentes poblaciones.

## APLICACION

### PUNTO 1

Comparar las estructuras de edades de la población, por sexo con los resultados obtenidos por ciudad en las diferentes rondas, mediante la Prueba no paramétrica de Kolmogorov-Smirnov, esta prueba permite comparar dos poblaciones, la reflejada por la muestra de la EIH con el Censo Nacional de Población y Vivienda de 1992.

La prueba consiste en determinar si los resultados de ambas poblaciones (muestras y Censo) en cuanto a la distribución de estructuras de edades, provienen de una población que tiene la misma distribución, realizando un contraste de las hipótesis siguiente (como se indico anteriormente):

$H_0$ : Las distribuciones de las muestras provienen de una misma población.

$H_1$ : Las distribuciones de las muestras no provienen de una misma población

Aplicando el estadístico que plantean Kolmogorov-Smirnov,

$$KS = \frac{k_{\alpha}}{\sqrt{n}}$$

Donde:

- $\alpha$  : Es el nivel de significancia
- $k$  : Es el valor de la tabla para esta prueba correspondiente a  $n$  grados de libertad, (Siendo este valor  $k = 1.63$  con un nivel de confianza del 95%)
- $n$  : Tamaño de la muestra (nivel persona)

Este estadístico es comparable con la diferencia absoluta de cada uno de los valores de la distribución relativa acumulada, en la población muestral expandida, respecto de la distribución relativa acumulada de la población censal, rechazando la hipótesis nula cuando alguno de esos valores sea mayor o igual  $KS$ .

Las pruebas de hipótesis se las realizaron separadamente por ciudad y por ronda, para poder observar si cada ciudad estudiada en el tiempo, tiene la misma distribución, a la distribución que presenta los resultados del Censo.

En todos los casos estudiados la  $H_0$  no pudo ser rechazada, por lo cual se ajustaran los resultados muestrales. En este punto se considero oportuno ajustar las estructuras muestrales a las censales, con el objeto de obtener una serie de datos libres de los efectos de las diferencias estructurales provenientes del muestreo.

### AJUSTE DE LOS FACTORES DE PONDERACION

Con las frecuencias relativas de la distribución del censo, y con las frecuencias absolutas de la distribución expandida a partir de la muestra, usando los mismos intervalos para ambas distribuciones, se multiplica para cada intervalo la frecuencia relativa de la distribución del censo con la frecuencia absoluta de la distribución expandida, obteniendo de esta manera la distribución con base en la estructura de la distribución del Censo.

#### DISTRIBUCION CENSAL

Intervalos de clase	Frecuencia Absoluta	Frecuencia Relativa
$L_{i-1} - L_i$	$n_i$	$f_i$
$L_1 - L_2$	$n_1$	$f_1$
$L_2 - L_3$	$n_2$	$f_2$
$L_3 - L_4$	$n_3$	$f_3$
.	.	.
.	.	.
$L_{i-1} - L_i$	$n_i$	$f_i$
.	.	.
.	.	.
$L_{k-1} - L_k$	$n_k$	$f_k$
Total	N	1.00

Entonces con la relación,

$$n_i' = f_i \times n_i$$

se construye una nueva distribución, la cual se denominará "distribución de la población expandida de acuerdo a la estructura del Censo".

Intervalos de clase	Frecuencia	Frecuencia
	Absoluta Expandida	Absoluta Expan. Censo
$L_{i-1} - L_i$	$n'_i$	$n''_i = n'_i * f_i$
$L_1 - L_2$	$n'_1$	$n''_1$
$L_2 - L_3$	$n'_2$	$n''_2$
$L_3 - L_4$	$n'_3$	$n''_3$
·	·	·
·	·	·
$L_{i-1} - L_i$	$n'_i$	$n''_i$
·	·	·
·	·	·
$L_{k-1} - L_k$	$n'_k$	$n''_k$
Total	$N'$	$N''$

Donde;

$f(X_i)$ : La frecuencia relativa del intervalo de clase  $i$  en la distribución base de la estructura (censo),

$$f(X_i) = \frac{n_i}{N}$$

$N$ : Es el total de la población en la distribución del censo

$n_i$ : La frecuencia absoluta en el intervalo  $i$ , en la distribución del censo

$n''_i$ : La frecuencia absoluta en el intervalo  $i$ , en la distribución en base a la estructura (censo)

$$N'' = \sum n''_i$$

$n'_i$ : La frecuencia absoluta en el intervalo  $i$ , en la distribución expandida a partir de la muestra

$$N' = \sum n'_i$$

$N'$ : Es la estimación del total de la población (muestra expandida), con  $N' = N''$

Una vez obtenida la distribución de la población expandida de acuerdo a la estructura del Censo, se procede a construir la distribución estimada a partir de la proyección, esto es;

Sea  $N_{\text{proy}}$ , la proyección de la población al momento de la realización de la encuesta,

$$n_{i \text{ proy}} = n'_i \times C_R$$

con;

$$C_R = \frac{N_{\text{proy}}}{N'}$$

Siendo,  $C_R$  el cambio relativo existente entre el total de la población proyectada y el total de la población expandida, y,

$$\sum n_{i \text{ proy}} = N_{\text{proy}}$$

Con esta distribución se procede a determinar el cambio en el factor de expansión que será en un primer ajuste para cada grupo (individuos) que pertenece al intervalo de clase, siendo el cambio del factor de expansión el siguiente:

$$C_{\text{EXP}} = \frac{n_{i \text{ proy}}}{n'_i}$$

Dada la distribución de frecuencias estimada a partir de la proyección:

Intervalos de clase $L_{i-1} - L_i$	Frecuencia Abs. por Proyección $n_{i_{proy}}$	Cambio en el Factor de Expansión $C_{i_{FEXP}} = n_{i_{proy}}/n_i$
$L_1 - L_2$	$n_{1_{proy}}$	$C_{1_{FEXP}}$
$L_2 - L_3$	$n_{2_{proy}}$	$C_{2_{FEXP}}$
$L_3 - L_4$	$n_{3_{proy}}$	$C_{3_{FEXP}}$
.	.	.
.	.	.
$L_{i-1} - L_i$	$n_{i_{proy}}$	$C_{i_{FEXP}}$
.	.	.
.	.	.
$L_{k-1} - L_k$	$n_{k_{proy}}$	$C_{k_{FEXP}}$
Total	$N_{proy}$	

Para un segundo ajuste, y con los tamaños de muestra que se tienen dentro de cada intervalo de clase en la distribución muestral, se obtiene el número de individuos, donde su factor de expansión ajustado se verá afectado en más o menos en una unidad, es decir;

Sea  $n_{i_{mus}}$ , la frecuencia absoluta en el intervalo  $i$ , de la distribución muestral,

$$(C_{i_{FEXP}} - 1) \times n_{i_{mus}}$$

La distribución de estos cambios se realizó en forma sistemática, con arranque aleatorio en un intervalo de selección.



## APLICACION

### PUNTO 2.

En caso de rechazarse la hipótesis nula, es decir de encontrarse diferencias estadísticas significativas en las distribuciones a través del tiempo, lo lógico hubiese sido eliminar de la serie las encuestas defectuosas.

Para uniformizar la serie en si, se optó por corregir todas las estructuras de las ciudades en estudio en las diferentes rondas utilizando las censales. Además de este ajuste, se procedió a usar como "Techo" de cada estimación la proyección de la estimación de población para cada fecha.

Para lograr nueva distribución con la proyección, se siguieron los siguientes pasos:

- a. Se obtiene una nueva distribución tomando como base la estructura del Censo. Multiplicando el valor correspondiente a la frecuencia absoluta de cada grupo de edad de población muestral, expandida con la frecuencia relativa de la población censal, a la cual denominaremos "distribución de la población expandida de acuerdo a la estructura del Censo".
- b. Con el cociente entre la proyección y el total de la población muestral expandida, obtenemos la variación de cambio, multiplicándose éste a cada grupo de edad de la distribución obtenida en el paso anterior. Esto para la construcción de la "distribución estimada con la proyección", la cual nos servirá para obtener los cambios en el factor de expansión original.
- c. Los cambios relativos en el factor de expansión para un primer ajuste es el cociente entre cada uno de los valores de la distribución muestral expandida y la nueva distribución estimada con la proyección, los cuales se multiplicaran a cada factor de expansión original.

Como un segundo ajuste y de acuerdo a los tamaños de muestra a nivel persona, obtenemos el número de personas donde se afectara

el Factor de Expansión Ajustado, incrementándose o decrementándose según sea el caso en una unidad. La distribución de estos cambios se los realiza en forma sistemática, con arranque aleatorio dentro de un intervalo de selección.

### EJEMPLO DE APLICACION

Para este ejemplo, tomamos la distribución de la población por sexo y por tramos de edad de la Primera Ronda y del Censo, para la ciudad de Sucre.

#### BOLIVIA POBLACION POR EDAD Y SEXO

PRIMERA RONDA 1989  
CIUDAD: SUCRE

PROYECCION A 1989 = 115068  
MUESTRA 1989 = 1482

GRUPOS EDAD Y SEXO	POBLACION MUESTRAL EXPANDIDA	POBLACION EXPANDIDA ACUMULADA	RELATIVA ACUMULADA (E)
0- 9 MASC	14,349	14,349	0.136
10-19	11,565	25,914	0.245
20-29	8,138	34,052	0.322
30-39	7,496	41,548	0.393
40-49	3,498	45,046	0.426
50-59	2,142	47,188	0.446
60-69	2,213	49,401	0.467
70 Y +	1,857	51,258	0.484
0- 9 FEM	14,135	65,393	0.618
10-19	11,779	77,172	0.729
20-29	10,352	87,524	0.827
30-39	7,496	95,020	0.898
40-49	4,069	99,089	0.937
50-59	2,784	101,873	0.963
60-69	2,713	104,586	0.989
70 Y +	1,214	105,800	1.000

POBLACION CENSAL

GRUPOS EDAD Y SEXO	POBLACION CENSAL	POBLACION ACUMULADA	RELATIVA ACUMULADA CENSO (C)	DIFERENCIAS ABSOLUTAS E-C	LIMITE SUPERIOR DIFERENCIA
0- 9 MASC	16,713	16,713	0.127	0.008	0.042
10-19	15,665	32,378	0.246	(0.002)	0.042
20-29	11,863	44,241	0.337	(0.015)	0.042
30-39	7,033	51,274	0.390	0.002	0.042
40-49	4,619	55,893	0.425	0.000	0.042
50-59	2,529	58,422	0.445	0.001	0.042
60-69	1,870	60,292	0.459	0.008	0.042
70 Y +	1,398	61,690	0.470	0.015	0.042
0- 9 FEM	16,804	78,494	0.597	0.021	0.042
10-19	17,113	95,607	0.728	0.002	0.042
20-29	13,764	109,371	0.832	(0.005)	0.042
30-39	8,235	117,606	0.895	0.003	0.042
40-49	5,514	123,120	0.937	(0.001)	0.042
50-59	3,459	126,579	0.963	(0.001)	0.042
60-69	2,609	129,188	0.983	0.005	0.042
70 Y +	2,196	131,384	1.000	0.000	0.042

Aplicando las Hipótesis planteadas anteriormente y realizando los cálculos correspondientes, se tiene:

Tamaño de la muestra (nivel persona)

$$n = 1482$$

$$\sqrt{n} = 38.50$$

El valor K con un nivel de confianza del 95%,  $k = 1.63$ , esto es;

$$KS = \frac{1.63}{38.50}$$

$$KS = 0.042$$

Puesto que KS es mayor a cada valor de la columna E-C, rechazamos la hipótesis alterna  $H_1$ , y concluimos que la distribución muestral expandida proviene de la Población Censal.

Posteriormente para uniformizar la serie se procedió a la construcción de las distribuciones, siendo en primera instancia una distribución de acuerdo a la estructura del censo y a la variación de cambio de la proyección respecto de la población muestral expandida, obtenemos la distribución nueva con proyección, la cual nos permitirá obtener el cambio relativo en el factor de expansión original en un primer ajuste. Donde el cambio relativo en el factor de expansión es igual al cociente entre las frecuencias absolutas de la distribución nueva con proyección y la distribución expandida original de cada grupo de edad.

GRUPOS EDAD Y SEXO	DISTRIB INICIAL	DISTRIBUCION NUEVA DE ACUERDO A LA ESTRUCTURA CENSAL		DISTRIBUCION NUEVA CON PROYECCION			
	POBLACION MUESTRA EXPANDIDA	POBLACION EXPANDIDA ESTRU CENSO	POBLACION ACUMULADA ESTRU CENSO	ESTIMACION NUEVA CON PROYEC	TAMAÑO MUESTRA PERSONAS	CAMBIO REL EN EL FAC- TOR DE EXP	FAC +1-1
0- 9 M	14,349	13,459	13,459	14,637	201	1.020	4
10-19	11,565	12,615	26,073	13,720	162	1.186	30
20-29	8,138	9,553	35,626	10,390	114	1.277	32
30-39	7,496	5,663	41,290	6,160	105	0.822	-19
40-49	3,498	3,720	45,009	4,045	49	1.156	8
50-59	2,142	2,037	47,046	2,215	30	1.034	1
60-69	2,213	1,506	48,552	1,638	31	0.740	-8
70 Y +	1,857	1,126	49,677	1,224	26	0.659	-9
0- 9 F	14,135	13,532	63,209	14,717	198	1.041	8
10-19	11,779	13,781	76,990	14,988	165	1.272	45
20-29	10,352	11,084	88,074	12,055	145	1.164	24
30-39	7,496	6,631	94,705	7,212	105	0.962	-4
40-49	4,069	4,440	99,145	4,829	57	1.187	11
50-59	2,784	2,785	101,931	3,029	39	1.088	3
60-69	2,713	2,101	104,032	2,285	38	0.842	-6
70 Y +	1,214	1,768	105,800	1,923	17	1.584	10
TOTAL	105,800	105,800		115,068	1,482		

Por último de acuerdo a la columna FAC + 1-1 obtenemos el número de personas donde se afectara el factor para el segundo ajuste, denominándose a los nuevos factores, Factores de Expansión Ajustados.