

Censo Nacional de Población, Hogares y Viviendas 2010

Censo del Bicentenario

Base de datos REDATAM

Errores muestrales en el procesamiento del cuestionario censal ampliado

Serie Base de datos Censo 2010
Noviembre 2015

Errores muestrales en el procesamiento del cuestionario censal ampliado

El cuestionario ampliado del Censo Nacional de Población, Hogares y Viviendas 2010 (Censo 2010) se aplicó a una muestra probabilística de segmentos censales. Como en los censos anteriores de Argentina donde se utilizó un cuestionario básico y uno ampliado, en el caso del Censo 2010 el formulario ampliado se aplicó a una selección de viviendas en las localidades de más de 50.000 habitantes. En el resto de las localidades todas las viviendas fueron censadas con el cuestionario ampliado. Por lo tanto, las estimaciones que surgen del cuestionario ampliado para dominios donde se aplicó muestra están afectadas por el llamado *error muestral*. Cuanto menor sea el dominio al cual hacen referencia las estimaciones, mayor será el *error relativo*. Como en toda encuesta por muestreo, este error afecta no solo las estimaciones de ciertas variables de interés sino la estimación de *los totales de población, varones, mujeres, etcétera*. Estos totales de referencia deben ser tomados del procesamiento del cuestionario básico. Si el usuario está interesado en cuantificar o estimar *el total* de un conjunto específico de personas u hogares con una determinada característica medida en el cuestionario ampliado, el procedimiento recomendable es estimar las proporciones con el cuestionario ampliado para un determinado interés y luego aplicar este porcentaje o proporción a la población de referencia que surge del cuestionario básico.

En el caso de las estimaciones que refieren a grandes dominios, como el total nacional o provincial, el usuario puede estimar el error muestral o el Coeficiente de Variación (CV) de las estimaciones mediante las tablas de errores de muestreo que se presentan en el Anexo Metodológico “*Precisión de las estimaciones del cuestionario ampliado*”.

Ejemplos prácticos para el cálculo de errores muestrales

Ejemplo 1

En la provincia de Salta, el total de población de 60 años y más que posee cobertura de salud (ver pregunta 13 del cuestionario ampliado), según el cuadro P12-P. Provincia de Salta. Población en viviendas particulares por tipo de cobertura de salud, según sexo y grupo de edad. Año 2010 es de 101.466 personas. Si buscamos en la tabla de *errores muestrales* de la provincia de Salta, el CV para la estimación más cercana es de 1,2%. El total de población de 60 años y más en la provincia de Salta, según el cuadro P8-P. Provincia de Salta. Población de 3 años y más en viviendas particulares por utilización de computadora, según sexo y edad. Año 2010 es de 120.066 personas con un CV asociado de 1,2%.

$\hat{Y}$ = Total de población estimado de 60 años o más que tiene cobertura de salud = 101.466

$\hat{X}$ = Total de población estimado de 60 años o más = 120.066

X = Total de población (según cuestionario básico) de 60 años o más = 120.304

$$\hat{R} = \frac{\hat{Y}}{\hat{X}} = \frac{\text{Total de población estimado de 60 años o más que tiene cobertura de salud}}{\text{Total de población estimado de 60 años o más}}$$

$$\hat{R} = \frac{101.466}{120.066} = 0,84$$

CV($\hat{Y}$) = 1,2% (correspondiente a la estimación de 101.466)

CV($\hat{X}$) = 1,2% (correspondiente a la estimación de 120.066)

Establecemos la siguiente estimación del total de población de 60 años y más con cobertura de salud a partir del *estimador de razón*:

$$\hat{Y}_R = \frac{\hat{Y}}{\hat{X}} \cdot X$$

$$\hat{Y}_R = \frac{101.466}{120.066} \cdot 120.304$$

$$\hat{Y}_R = 101.667$$

Con esto podemos calcular el CV aplicando la fórmula anteriormente mencionada:

$$CV(\hat{Y}_R) = \sqrt{CV^2(\hat{Y}) + CV^2(\hat{X})}$$

$$CV(\hat{Y}_R) = \sqrt{1,2^2 + 1,2^2}$$

$$CV(\hat{Y}_R) = 1,7\%$$

El *error muestral* lo podemos despejar de la siguiente fórmula:

$$CV(\hat{Y}_R) = 100 \cdot \frac{\text{Error Muestral}}{\text{Estimación}}$$

$$\text{Error Muestral} = \frac{CV(\hat{Y}_R) \cdot \text{Estimación}}{100}$$

Por lo tanto

$$\text{Error Muestral} = \frac{1,7\% \cdot 101.667}{100} = 1.725$$

La fórmula para el cálculo del *intervalo de confianza* al 95% es:

$$IC(\hat{Y}_R) = [(\hat{Y}_R - 1,96 \cdot \text{Error Muestral}); (\hat{Y}_R + 1,96 \cdot \text{Error Muestral})]$$

Por lo tanto

$$IC(\hat{Y}_R) = [(101.667 - 1,96 \cdot 1.725); (101.667 + 1,96 \cdot 1.725)]$$

$$IC(\hat{Y}_R) = [98.285; 105.049]$$

Ejemplo 2

En la provincia de Salta, el total de población de 10 a 14 años que posee alguna *dificultad permanente* (ver pregunta 14 del cuestionario ampliado) según el cuadro *P9-P. Provincia de Salta. Población en viviendas particulares, población con dificultad o limitación permanente y prevalencia de la dificultad o limitación permanente, según sexo y grupo de edad. Año 2010* es de 11.731 personas. Si buscamos en la tabla *errores muestrales de Salta*, el CV para la estimación más cercana es de 3%. El total de población de 10 a 14 años en la provincia de Salta, según el cuadro *P8-P. Provincia de Salta. Población de 3 años y más en viviendas particulares por utilización de computadora, según sexo y edad. Año 2010* es de 131.732 personas con un CV asociado de 1,2%.

$\hat{Y}$ = Total de población estimado de 10 a 14 años que tiene alguna discapacidad permanente = 11.731

$\hat{X}$ = Total de población estimado de 10 a 14 años = 131.732

X = Total de población (según cuestionario básico) de 10 a 14 años = 132.153

$$\hat{R} = \frac{\hat{Y}}{\hat{X}} = \frac{\text{Total de población estimado de 10 a 14 años que tiene alguna discapacidad permanente}}{\text{Total de población estimado de 10 a 14 años}}$$

$$\hat{R} = \frac{11.731}{131.732} = 0,09$$

$CV(\hat{Y}) = 3\%$ (correspondiente a la estimación de 11.731)

$CV(\hat{X}) = 1,2\%$ (correspondiente a la estimación de 131.732)

Establecemos la siguiente estimación del Total de población de 60 años y más con cobertura de salud a partir del *estimador de razón*:

$$\hat{Y}_R = \frac{\hat{Y}}{\hat{X}} \cdot X$$

$$\hat{Y}_R = \frac{11.731}{131.732} \cdot 132.153$$

$$\hat{Y}_R = 11.768$$

Con esto podemos calcular el CV aplicando la fórmula anteriormente mencionada:

$$CV(\hat{Y}_R) = \sqrt{CV^2(\hat{Y}) + CV^2(\hat{X})}$$

$$CV(\hat{Y}_R) = \sqrt{3^2 + 1,2^2}$$

$$CV(\hat{Y}_R) = 3,2\%$$

El *error muestral* lo podemos despejar de la siguiente fórmula:

$$CV(\hat{Y}_R) = 100 \cdot \frac{\text{Error Muestral}}{\text{Estimación}}$$

$$\text{Error Muestral} = \frac{CV(\hat{Y}_R) \cdot \text{Estimación}}{100}$$

Por lo tanto

$$\text{Error Muestral} = \frac{3,2\% \cdot 11.768}{100} = 380$$

La fórmula para el cálculo del *intervalo de confianza* al 95% es:

$$IC(\hat{Y}_R) = [(\hat{Y}_R - 1,96 \cdot \text{Error Muestral}); (\hat{Y}_R + 1,96 \cdot \text{Error Muestral})]$$

Por lo tanto

$$IC(\hat{Y}_R) = [(11.768 - 1,96 \cdot 380); (11.768 + 1,96 \cdot 380)]$$

$$IC(\hat{Y}_R) = [(11.023; 12.513)]$$