

Marco Metodológico de Muestreo para Reportes Storecheck

Existen dos tipos básicos de reporte que se generará utilizando la información recabada por Storecheck semanalmente.

- 1) Reportes indicando medidas a nivel tienda. Estos reportes incluyen la Rotación y las Distribuciones Horizontal y Vertical por tienda.
- 2) Reportes indicando medidas a nivel global, tales como la participación de mercado de las distintas categorías, marcas, SKU's, etc.

Para cada uno de estos tipos de reporte se cuenta con un esquema de muestreo polietápico en donde se identifican estratos de interés y para cada uno de estos estratos se seleccionan aleatoriamente conglomerados de unidades muestrales. Para cada conglomerado se muestrea el total de unidades muestrales.

Los estratos de interés que demuestran un claro incremento de la variabilidad entre los mismos se presentan a continuación. Para los mismos estratos, la variabilidad al interior de los mismos se reduce significativamente. En la siguiente tabla se presentan las desviaciones estándar en los diferentes estratos y a nivel global para: unidades, precios e importes.

	Desviación Estándar		
	unidades	precios	importes
HABITACIONAL POPULAR	2.2	13.0	15.9
HABITACIONAL RESIDENCIAL	1.0	16.6	15.0
MIXTO INDUSTRIAL	0.6	16.4	7.8
MIXTO OFICINAS	1.2	16.8	19.7
OFICINAS	2.5	7.6	26.2
GLOBAL	5.4	19.1	34.3

Tal como puede observarse, la variabilidad crece significativamente cuando la información se considera a nivel global. Esto indica que los estratos están seleccionados eficientemente.

Al interior de cada estrato se seleccionan conglomerados de manera aleatoria para asegurar una cobertura geográfica uniforme. Los conglomerados son tiendas de abarrotes, mientras que las unidades muestrales en cada conglomerado están representadas por las compras realizadas por cada cliente de la tienda, es decir, cada evento de compra.

Tamaño de la muestra

La muestra actualmente consta de 92 tiendas en el Área Metropolitana de la Ciudad de México.

Considerando que se tratara de un muestreo aleatorio simple de 92 tiendas en el AMCM, el error en las estimaciones para el primer tipo de reporte se presenta a continuación.

De acuerdo al teorema central del límite, la distribución de la rotación (ventas promedio) sería la siguiente:

Supongamos que R representa la rotación de interés en un periodo particular y que se cuenta con una muestra de 92 tiendas para estimarla. Las unidades muestrales, es decir, cada evento de compra, deben ser agregadas para contar con una medida de la venta de cierto producto a nivel "tienda". Entonces:

$$r = \frac{\sum_{i=1}^{92} u_i}{92}$$

donde u_i representa a las unidades vendidas en cada tienda y r es el promedio muestral que estima la rotación R .

Entonces, R , la rotación en la población se distribuye como una normal con los siguientes parámetros:

$$N\left(r, \sigma/\sqrt{92}\right)$$

Donde σ representa la desviación estándar de la rotación en la población. Este valor se puede estimar como:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{92} (u_i - r)^2}{92}}$$

El intervalo de confianza al 95% para \bar{x} estará dado por:

$$\left(r - 1.96 \frac{\sigma}{\sqrt{92}}, r + 1.96 \frac{\sigma}{\sqrt{92}} \right)$$

La teoría anterior es válida cuando se trata de muestreo aleatorio simple. Sin embargo, como se mencionó en la sección anterior, el muestreo de las 92 tiendas es estratificado, lo que necesariamente tendrá que reducir la varianza. En este sentido, los intervalos de confianza proporcionados por la aproximación normal tendrán una reducción significativa en tamaño.

En el caso de estimación de proporciones (ie proporción de tiendas que venden la marca "x"), el error máximo se calcula de la siguiente manera:

$$e = \sqrt{\frac{p(1-p)}{92}} = \sqrt{\frac{0.25}{92}} = 0.102$$

donde p representa a la proporción de interés. La varianza máxima para las proporciones está dada cuando $p=0.5$, por lo que el error máximo para la muestra de 92 está dado por la expresión anterior.

Esto significa que los intervalos de confianza al 95% para la estimación, \hat{p} , de la proporción están dados por:

$$\left(\hat{p} - 0.102, \hat{p} + 0.102 \right)$$

En el caso de los reportes a nivel global, el tamaño de la muestra cambia significativamente ya que las unidades muestrales no necesitan ser agregadas para contar con una medida a nivel tienda. Se requiere únicamente una medida a nivel general.

Las fórmulas arriba mencionadas cambiarán significativamente en términos del error ya que el tamaño de la muestra se incrementa. Si se venden m unidades de un SKU en particular para un periodo dado en las 92 tiendas, las fórmulas presentadas anteriormente cambiarán de la siguiente forma.

Por el teorema central del límite, la distribución de x , el parámetro de interés estará aproximada por:

$$N\left(\bar{x}, \sigma/\sqrt{m}\right)$$

donde \bar{x} representa la estimación en la muestra de la cantidad de interés, σ es la varianza muestral y m el total de eventos de compra de del SKU en cuestión. A medida que crece m , la desviación estándar de la distribución tiende a volverse cada vez más pequeña.

El intervalo de confianza al 95% para \bar{x} estará dado por:

$$\left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{m}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{m}} \right)$$

Para la estimación de proporciones, el error máximo se calcula de la siguiente manera:

$$e = \sqrt{\frac{p(1-p)}{m}} = \sqrt{\frac{0.25}{m}}$$

De la misma manera, a medida que crece m , el error anterior tenderá a minimizarse.

El intervalo de confianza para \hat{p} será:

$$\left(\hat{p} - \sqrt{\frac{0.25}{m}}, \hat{p} + \sqrt{\frac{0.25}{m}} \right)$$