Tres Mitos en Investigación de Mercados

Mito 2: Diferencias entre Números (Medias o Porcentajes)

Marcelo De Fuentes G. Merc

En este segundo de tres artículos abordaremos uno de los "mitos" con los que me enfrento más frecuentemente al momento de presentar los estudios de mercado: diferencias estadísticas. Los comentarios a este respecto son más o menos en el siguiente tenor:

- "Como la formulación **A** obtuvo un 53% de preferencias y la **B** tan sólo un 48%, entonces debemos utilizar la primera porque es mejor que la segunda".
- "Nuestro *top of mind* en Guadalajara es del 35% y en Monterrey del 39%, por lo tanto estamos mejor en conocimiento en Monterrey que en Guadalajara".
- "El porcentaje de la población entre 18 y 24 años que consume mi producto es del 55% y entre 25 a 35 años es del 50%, lo cual indica que elaboramos un producto para jóvenes".

Considero que ante estas afirmaciones que día con día escuchamos en nuestro honroso medio de investigación de mercados debemos hacer referencia a dos grandes hombres, muy distintos entre sí, pero con una claridad de mente envidiable: San Agustín y Thomas Carlyle (uno de los grandes pensadores sajones) que decían:

"Errar es humano, perseverar en el error es diabólico" ***San Agustín***

"La mayor equivocación es imaginar que nunca nos equivocamos"

Thomas Carlyle

Para adentrarnos al tema de por qué los tres puntos (y muchos más que puedan ustedes pensar que van bajo la misma tónica) son un error que nos puede orillar a tomar decisiones absolutamente equivocadas, debemos recordar que al establecer una muestra y no censar a la población objetivo, estamos asumiendo un margen de error directamente atribuible al tamaño de la misma, error estadístico que puede ser fácilmente computable (ver artículo previo sobre el tamaño de la muestra en boletín 23).

Sabemos que además de esta "desviación" de los resultados por diseño muestral existen los que se conoce como errores no muestrales, mismos que son mucho más difíciles de calcular estadísticamente y a los cuales no haremos alusión en este documento.

El margen de error tiene efecto directo sobre los resultados que obtengamos a una pregunta específica de un cuestionario. Para poner esto de forma más clara permítanme utilizar un ejemplo: pensemos que hacemos un estudio de refrescos en la ciudad de México con 156 entrevistas, las cuales tienen un margen de error del +/- 8% (ver artículo sobre tamaño de muestra) y obtenemos los siguientes resultados en la preferencia:

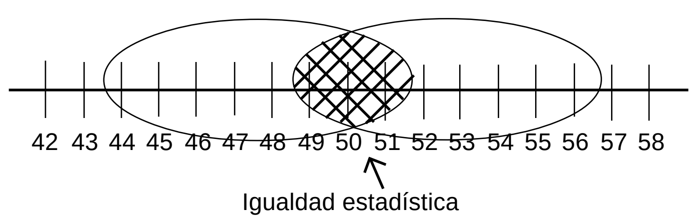
Refresco **A** = 48%

Refresco **B** = 53%

De manera muy sencilla en un esquema de muestreo simple y aplicando el margen de error de forma directa a estos resultados tendremos que la variabilidad de estas cifras, por error muestral, es la siguiente:

	Límite inferior	RESULTADO	Límite superior
Refresco **A**	43.24%	48.0%	50.76%
Refresco **B**	48.76%	53.0%	57.24%

Si en una recta numérica graficamos lo anterior tendremos algo como esto:



Aunque en la tabla era claro que los resultados de ambos refrescos se traslapaban por efecto del margen de error, la gráfica nos deja ver que existe un área que conocemos como de “igualdad estadística”; es decir, si bien es cierto que la aritmética básica nos indica que el 53% es mayor que el 48%, la estadística pone en un plano muy distinto los resultados y nos lleva a la interpretación de que ambos son iguales o que no existe diferencia, estadísticamente hablando. Como conclusión podemos decir que el refresco A tiene la misma preferencia que el B y no existe un ganador claro a este respecto.

En este sencillo ejemplo tan sólo hemos hablado de un dato, pero en la vida real los estudios están atiborrados de preguntas, segmentos, cruces, etcétera, que hacen materialmente imposible seguir este proceso “manual” para saber si dos porcentajes tienen una zona de igualdad estadística o no. Más aún, al intentar obtener o aplicar este mismo proceso pero en el caso de medias aritméticas las cosas se ponen más complicadas porque intervienen indicadores tales como la desviación estándar o varianza (no confundir entre ellos porque son conceptos distintos) que definitivamente dificultan el proceso de “picar piedra”.

La solución es bastante sencilla y hasta cierto punto primaria: utilizar pruebas de hipótesis, ya sea a través del estadístico *Z* o *t* (tradicionalmente la decisión de usar uno u otro dependía del tamaño de la muestra; para efectos prácticos el común denominador en estudios de mercado es *Z)*. Recordemos que la manera de plantear esto es a través de una hipótesis nula (generalmente denominada como H_0) y una alternativa (H_1) en donde el objetivo es aceptar la H_0 (y por lo tanto rechazar la H_1) o rechazar la H_0 (y por lo tanto aceptar la H_1). Sintácticamente se escribe de la siguiente manera:

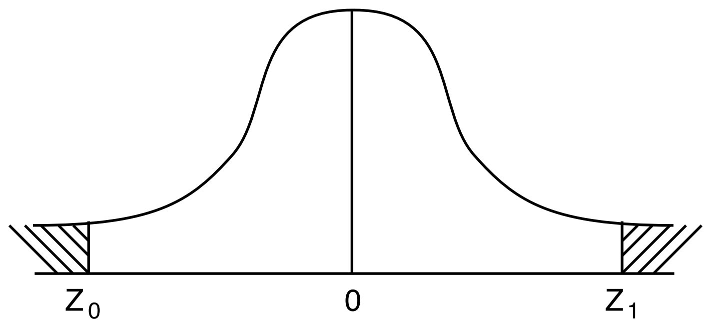
H_0: El porcentaje / media A es = al porcentaje / media B

H_1: El porcentaje / media B es ≠ al porcentaje / media B

A nivel semántico decimos que si aceptamos la hipótesis nula (H_0) es que el porcentaje/media A es **estadísticamente** igual al porcentaje/media B; es decir, entre ambos no existe una diferencia estadística. En este sentido, de manera automática al aceptar H_0 estamos rechazando H_1. Exactamente la interpretación contraria deberíamos dar si rechazamos H_0.

La aplicación de este proceso es bastante sencilla y se resume en los siguientes pasos:

1. Estandarizar la curva de distribución de probabilidad; esto es, tener resultados con media 0 y desviación estándar 1 cuya expresión es: N ~ (0, 1).



2. Determinar el valor *Z* de tablas (como ya lo mencionamos, en algunos casos se utiliza *t).*

3. Calcular los límites inferior y superior que establecerán las zonas de aceptación y rechazo de la hipótesis nula o H_0. Para ello se utiliza la fórmula que a continuación desarrollamos:

$$\pm Z \sqrt{\frac{pq}{n_1} + \frac{pq}{n_2}}$$

$$p = \frac{P_1 n_1 + P_2 n_2}{n_1 + n_2}$$

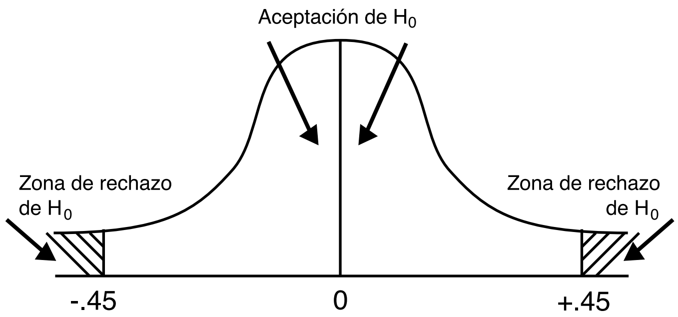
$$q = (1-p)$$

P_1 = porcentaje 1
P_2 = porcentaje 2
n_1 = muestra 1
n_2 = muestra 2

4. Hacer la diferencia de los porcentajes que estamos probando (en nuestro ejemplo sería 0.53 – 0.48 = 0.05).

5. Ya que hemos calculado los límites y conocemos la diferencia entre los porcentajes, tenemos que identificar si el número resultante de la diferencia (0.05) se encuentra en la zona de aceptación o de rechazo de la hipótesis nula.

Si en el ejemplo de refrescos del cual estamos hablando encontramos que los límites superior e inferior se encuentran en +/- 0.45 y la diferencia entre los porcentajes es 0.05, entonces la gráfica se vería de la siguiente forma:



Vean que el valor de la diferencia entre los porcentajes se encuentra en la zona de aceptación de H_0, por lo que podemos decir que estadísticamente el porcentaje **A** es igual al porcentaje **B**.

A todo lo escrito en este muy breve y somero artículo podemos hacer una serie de importantes acotaciones como:

- Las diferencias entre medias se obtienen bajo una fórmula distinta que incluye las desviaciones estándar.
- Mucho se ha escrito y discutido sobre las diferencias entre grupos que provienen de una misma muestra; por ejemplo cuando queremos comparar un estudio que fue realizado entre México, Guadalajara, Monterrey, Puebla y Querétaro, en donde cada ciudad no debe ser tomada como evento independiente.
- Existen otros métodos para obtener diferencias entre grupos como es el ANOVA (Analisys of Variance).

Pero la intención del documento no es otra que motivar al lector a no leer o interpretar de manera directa resultados de un estudio de mercado, sino a dar un pequeño paso más en el manejo estadístico de los datos. De ninguna manera pretendo aquí hacer una cobertura extensa (y mucho menos detallada) de las diferentes herramientas existentes para el manejo de diferencias estadísticas, sino simplemente llamar la atención, con el método más simple y rudimentario, de aquellos que interpretan, como: proveedores, clientes o académicos; investigaciones derivadas de esquemas muestrales.

En el entendido que todo lo aquí expuesto no es nada nuevo bajo el sol, y mucho menos algún tipo de descubrimiento de estadística atómico-nuclear-cibernética-cuántica (¿??????), los espero en el próximo DDT en donde encontrarán el tercer y último mito, cerrando con una frase del maravilloso escultor francés que le da expresión y sentimiento humano a sus obras, Auguste Rodin:

“Yo no inventé nada: yo redescubro”
Auguste Rodin