Semana 3

Intervalos de confianza y valores p *Valentín Vergara Hidd*

Como vimos en el documento anterior, parte de la prueba de hipótesis implica tomar una decisión: aceptamos H_0 o no, tratando de evitar cometer un error de tipo I, o lo que es equivalente, minimizando α.

Existen dos formas de tomar esta decisión, ambas a partir de lo que se obtiene en los tres pasos previos. Los intervalos de confianza nos dan un rango en que el parámetro que se está estimando se encuentra en la población; en tanto que el valor p es interpretado como la probabilidad de aceptar H_0.

---

1. Revisitando la distribución normal

Pensemos por un momento que observamos una variable aleatoria X, y a partir de ella, obtenemos N muestras de tamaño n. Del documento anterior, sabemos que mientras mayor sea N, la distribución de $\overline{x}$ se va a acercar a una distribución normal con media μ y desviación estándar $\frac{\sigma}{\sqrt{n}}$. En este caso, 68 % de las observaciones de dicha variable van a estar en el intervalo:

$$\left[\mu - \frac{\sigma}{\sqrt{n}}, \mu + \frac{\sigma}{\sqrt{n}}\right].$$

Incluso, podemos extender el argumento e indicar que el 95 % de los casos (aproximadamente) se encuentra en el intervalo

$$\left[\mu - \frac{2\sigma}{\sqrt{n}}, \mu + \frac{2\sigma}{\sqrt{n}}\right].$$

En ambos casos, definimos un intervalo donde es bastante probable que cualquier muestra tomada al azar de la variable X, arroje un valor de $\overline{x}$ que se encuentre dentro de él. Algo similar es lo que tenemos que hacer para crear intervalos de confianza y usarlos en nuestras estimaciones (en nuestra *generación* de modelos).

2. Intervalos de confianza

Al utilizar los pasos de la prueba de hipótesis (en el documento anterior), estamos usando una herramienta que nos permite decidir si un *valor propuesto* por el investigador es razonable para extrapolar a toda la población. En otras palabras, si un modelo refleja correctamente o no la realidad. Sin embargo, es muy difícil que a partir de una muestra, por muy grande que esta sea, podamos estimar *exactamente* el valor del parámetro poblacional de interés.

Ejemplo 1: Encuestas de opinión

En más de alguna oportunidad hemos visto en algún periódico o revista algo así como: *el 27 % de las personas apoya la opción A por sobre la opción B, con un margen de error del 3 %*.
Esto significa que existe una proporción *real* de personas, p, que prefieren la opción A a la opción B. El dato del margen de error nos indica que esta proporción no se conoce exactamente, pero un buen estimador es situarla en el intervalo [24 %, 30 %]

2.1. Construcción de intervalos de confianza

Para ordenar un poco el procedimiento de la construcción de intervalos de confianza, Lynch (2013) propone los siguientes pasos, coonsiderando el problema del intervalo de confianza de una media muestral $\overline{x}$:

1. Decidir un valor de α[1]. Usualmente aceptamos $\alpha = 0,05$
2. Encontrar el valor de la variable aleatoria Z para el que se deje un área de $\frac{\alpha}{2}$ en cada *cola* de la distribución. Llamaremos a este valor $Z_{\alpha/2}$
3. Calcular el error estándar de la media. Esto se hace a través de la expresión

$$\text{sem}_{\text{muestra}} = \frac{\sigma}{\sqrt{n}} \tag{1}$$

4. Construir el estimador del intervalo de confianza a partir de:

$$\overline{x} \pm \left(Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

El paso 2 requiere una explicación más detallada, para lo que debemos nuevamente revisar la distribución normal.

2.2. ¿Cómo obtenemos $Z_{\alpha/2}$?

Recuerden que en un documento previo definimos Z como una variable con distribución normal y parámetros $\mu = 0$ y $\sigma = 1$. Si la graficamos, se debería ver más o menos como curva negra en la Figura 1. Puesto que esta curva se obtiene de la función de densidad de la distribución normal[2], que para simplificar llamaremos genéricamente como $f(x)$; podemos concluir algunas cosas.

Lo primero es que el área total bajo la curva es 1, por lo que al ser una curva simétrica, implica que el área entre $x = 0$ y $x \to \infty^+$ es 0,5. Luego, la suma de ambos segmentos coloreados en rojo en la Figura 1 equivale a α; por lo que el área entre $x = Z_{\alpha/2}$ y $x \to \infty^+$ equivale a 0.025.

¿Para qué sirve toda la información del párrafo anterior? A partir de ella, deducimos que el área entre $x = 0$ y $x = Z_{\alpha/2}$ se obtiene mediante la expresión:

$$\int_0^{Z_{\alpha/2}} f(x)dx = 0,5 - 0,025 = 0,475 \tag{2}$$

[1] Esto es, la probabilidad máxima de cometer un error tipo I (ver documento anterior)
[2] Nuevamente, para refrescar la memoria, sugiero revisar los documentos anteriores

Un poco de álgebra más tarde, concluímos que $Z_{\alpha/2} = 1,96$. Afortunadamente, después de un par de siglos de uso de estas cosas, existen recursos que facilitan el trabajo. Probablemente cuando ustedes aprendieron esto, al igual que yo, se acostumbraron a revisar una tabla con las áreas bajo la curva de la distribución normal, como por ejemplo esta. En el video que acompaña este documento, les voy a recordar como se lee esta tabla, pero también les mostraré como se calcula directamente en SPSS.

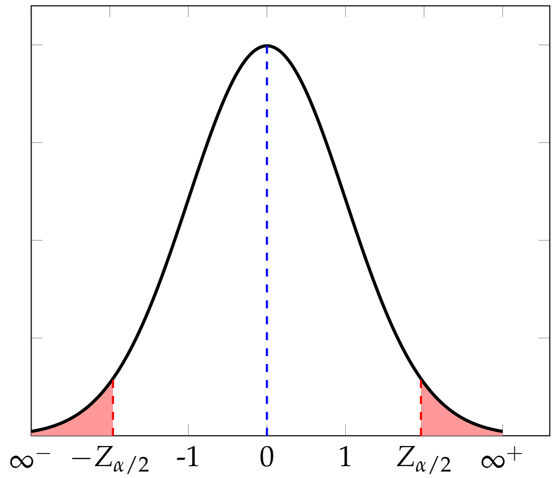

Figura 1: Variable aleatoria Z, con $\alpha = 0,05$

Un ejercicio interesante es retomar un ejemplo presentado en un documento anterior, para obtener un intervalo de confianza.

2.3. Retomando el ejemplo de los puntajes PSU

Ejemplo 2: Puntajes PSU

En un video anterior, les comenté que la PSU está diseñada para ser una prueba cuya distribución de puntajes tiene $\mu = 550$ y $\sigma = 110$.
Sólo sabiendo eso, supongamos que contamos con los puntajes de toda una generación que egresa de un colegio, con 63 estudiantes. Para esta generación, si X es la variable definida como el puntaje PSU, medimos $\overline{x} = 588,7$ y $s_x = 83,9$.

La idea es calcular el intervalo de confianza que contiene los puntajes de la población a partir de la muestra descrita en el ejemplo.

El primer paso es definir $\alpha = 0,05$. Luego, el segundo paso implica algo que ya calculamos, $Z_{\alpha/2} = 1,96$. Para el tercer paso, utilizamos:

$$\text{sem}_{\text{muestra}} = \frac{110}{\sqrt{63}} \approx 13,8586$$

Con este resultado, podemos calcular el intervalo de confianza, paso 4.

$$\overline{x} \pm \left(Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 588,7 \pm 13,8586 \implies [574,8414; 602,5586]$$

Obviamente, no es un buen estimador, porque incluso utilizando el intervalo de confianza, encontramos que este no incluye el valor de μ. Esto muestra que si partimos con un valor sesgado, probablemente lleguemos a un intervalo de confianza que no sea muy informativo. Para remediar esto, siempre podemos complementar los resultados de un intervalo de confianza con un valor p.

3. Valores p

Esta debe ser la medida más popular –y controversial– que vemos al simplemente leer cualquier artículo científico. Generalmente en la sección de resultados se presenta, directa o indirectamente, una lista de coeficientes con sus respectivos valores p. Coloquialmente, estamos acostumbrados a pensar en los valores p como la *probabilidad de aceptar* H_0. Esto no es completamente equivocado, sólo es impreciso.

Veamos de manera más formal, qué vamos a entender por un valor p. La idea parte desde el capítulo anterior. Cuando nos encontramos con un resultado proveniente de una muestra, ¿qué tan probable es que ese resultado esté en la cola de la distribución de las muestras de una población?

Definición 1: Valores p

El **valor** p de una observación x de la variable aleatoria X con una distribución de probabilidad conocida, equivale a la **suma de probabilidades** para valores igual o más extremos que x.
Precisando, *valores más extremos* se refiere a valores $x' \leq x$ cuando x está a la izquierda de la distribución; y a $x' \geq x$ cuando está a la derecha.

Siguiendo el ejemplo 2, analicemos si el promedio de la muestra $\overline{x} = 588,7$ es un valor que *probablemente* podamos asumir que es el promedio de toda la población. Implícito en la afirmación anterior está la hipótesis nula y alternativas

$$\begin{aligned} H_0: &\quad \overline{x} = \mu \\ H_1: &\quad \overline{x} \neq \mu \end{aligned}$$

Previamente hemos establecido que todas las medias que provienen de una muestra tienen una distribución normal centrada en μ y con desviación estándar $\frac{\sigma}{\sqrt{n}}$, donde n es el tamaño de las muestras. Si $f(x)$ es la función de probabilidad de la distribución normal[3], lo que buscamos es:

$$p = \Pr(x \geq 588,7|\mu = 550, \sigma = 110) = \int_{588,7}^{\infty^+} f(x|\mu = 550, \sigma = 110)dx \tag{3}$$

Como resulta dificil evaluar la ecuación 3; lo que podemos hacer es evaluar la siguiente expresión:

$$p = 0,5 - \int_{0}^{588,7} f(x|\mu = 550, \sigma = 110)dx = 0,362487 \tag{4}$$

[3]Una vez más, no estamos escribiendo la función para no complicar la expresión. Si se quiere conocer la función de probabilidad de una distribución normal, basta con revisar los documentos anteriores.

Aproximemos el valor obtenido en 4 a 0,36; o a 36 %. Esto significa, que la probabilidad de cometer un error tipo I al afirmar que $\overline{x} = 588,7 = \mu$ es de 36 %. Es una probabilidad muy alta como para aceptar el resultado como válido, considerando que establecimos como 5 % lo máximo permitido.

4. Finalizando

Como he mostrado en este documento, es útil contar con herramientas para apoyar el proceso de toma de decisiones, que a su vez es parte de la prueba de hipótesis. Dos de estas herramientas son los intervalos de confianza y los valores p. Ambos tienen ventajas y desventajas, así como amplios debates respecto a su pertinencia o no dentro de distintos diseños de investigación. Mi impresión al respecto es que son herramientas poderosísimas por sí solas, pero en combinación nos entregan resultados extremadamente confiables. Como veremos más adelante, es precisamente esto lo que hace cualquier software estadístico: nos entrega los valores para construir un intervalo de confianza, al mismo tiempo que un valor p para saber si es pertinente usar crear dicho intervalo.

Referencias

Lynch, Scott. 2013. *Using statistics in social research : a concise approach*. New York: Springer.