PROPORCIONES E INTERVALOS DE CONFIANZA

Hemos visto como comparar diferentes muestras y poblaciones a partir de sus medias. Sin embargo, es posible que no siempre tengamos la media sino una proporción. Esto es lo que vamos a abordar aquí.

PROPORCIONES

Imagine que quiere saber la proporción de estudiantes hombres que fuman. Para este efecto, toma una muestra representativa de estudiantes hombres y les pregunta si fuman. Hay dos respuestas posibles: sí o no. Podemos asignar un valor de 1 a la respuesta afirmativa y un 0 a la respuesta negativa.

La proporción de fumadores sería:

$$\text{Proporción } (\prod) = \frac{1+1+1+1\ldots0+0+0+0}{\text{Número total de la muestra}} = \frac{\text{número de fumadores}}{\text{número total de la muestra}}$$

La pregunta que nos podríamos plantear es si esto es un estimado de la proporción de fumadores en toda la población de estudiantes hombres. La respuesta sería afirmativa.

Esto es muy similar a lo postulado en el *teorema del límite central*, sólo que aquí nos estamos refiriendo a proporciones. El teorema se postularía así:

Teorema 2
Suponiendo que n es razonablemente grande, la distribución de la proporción media *p*, es aproximadamente normal cuando:

Teorema 2
(variación del Teorema del Límite Central para proporciones)

Media = $\prod$ (pi)

y

$$\text{Error estándar} = \sqrt{\left[\frac{\prod(1-\prod)}{n}\right]}$$

Notas:
1. A la proporción muestral se le denomina *p*
2. La verdadera proporción de la población que estamos intentado conocer es $\prod$ (la letra griega mayúscula 'pi',)

Si recuerda, esto es similar a la distribución de medias muestrales. Lo que implica es que, si nosotros tomamos muestras repetidas y encontramos la proporción *p* de cada muestra, entonces todas las medias de todas las *p* serán muy aproximadas a la proporción poblacional $\prod$, y la desviación estándar de la distribución de p_s será el error estándar dado en la ecuación antes mencionada.

Siempre y cuando el tamaño de la muestra sea razonablemente grande, podemos usar la proporción *p* media como una aproximación razonable de la proporción de la población en el error estándar. Veamos un ejemplo:

Ejemplo 1

Usted sabe que en promedio sólo el 90% de los pasajeros de avión que han hecho una reservación se presentan para su vuelo. Usted quiere ir a New York de último momento y se ubica en una lista de espera en el lugar 35 para un avión de 330 asientos. ¿Qué probabilidades tiene de llegar a Nueva York en este vuelo?

Primero, hay que escribir lo que sabemos. Sabemos que el tamaño de la muestra *n* es de 330 y que la proporción de la población $\prod$ es de 0.90 (90 ÷ 100). Ahora, podemos calcular la media y el error estándar de una muestra de proporciones:

Media = $\prod$ = 0.90

$$\text{Error estándar} = \sqrt{\left[\frac{\prod(1-\prod)}{n}\right]} = \sqrt{\left[\frac{0.9 \times 0.1}{330}\right]} = 0.017$$

Así que podemos decir que *p* está normalmente distribuida con una media = 0.90 y un EE = 0.017.

Ahora debemos saber que debe pasar para que usted se suba a dicho avión. Hay 330 asientos y usted está el 35º lugar en la lista de espera. La pregunta crucial es: ¿cuántos pasajeros que cuentan con una reservación se van a presentar para que usted pueda ir a Nueva York?

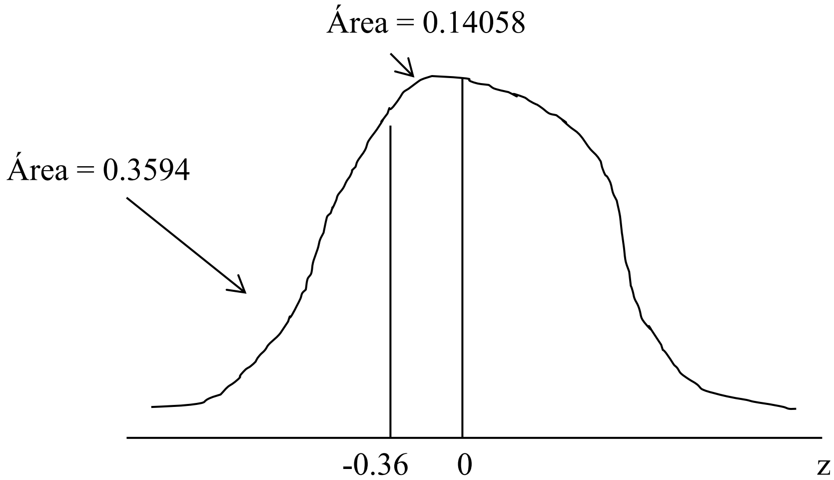
Si (330 – 35 =) 295 pasajeros o menos se presentan, usted todavía tiene oportunidad de subir a ese avión. Pero ¿cuál es la probabilidad de que tan poca gente se presente?

La proporción media *p* será de 295 ÷ 330 = 0.8939. Esta es la máxima proporción de pasajeros que usted desea que se presenten para que pueda abordar el avión. Lo que usted quiere es (a parte de subir al avión) es saber la probabilidad de que la proporción que se presente sea de 0.8939 o menos.

Ahora queremos calcular los puntajes z y el diagrama correspondiente:

$$Z = \underline{\quad p - \prod \quad} \qquad Z = \underline{0.8939 - 0.9} = -0.36$$

Error estándar 0.017



De las tablas normales, el área entre la media y z = -0.36 es de 0.14058, así que el área hacia la izquierda de Z = -0.36 será de 0.5 – 0.14058 = 0.3594.

De ese resultado usted puede concluir que tiene un 36% de probabilidades de llegar a Nueva York en ese avión.

Los Intervalos de confianza para proporciones

Cuando estimamos la media de una población a partir de una muestra, comentamos que un estimado de intervalo era preferible que un punto estimado y por lo tanto calculamos un intervalo de confianza como una medida del rango en el cual podríamos estar seguros de que se ubica la media poblacional.

De la misma forma podemos calcular intervalos de confianza para proporciones, de tal suerte que cuando estimamos una proporción de la población a partir de una muestra podemos tener una de que tan confiable es nuestro estimado.

La fórmula para un intervalo de confianza del 95% es análoga a la fórmula de una media:

95% IC = p ± (1.96 x error estándar)

Así que nuestra ecuación se verá como a continuación:

$$95\%\ \text{IC} \left\{ = p \pm \ 1.96 \sqrt{\left[\frac{p\,(1-p)}{n}\right]} \right\}$$

Al igual que con las medias, para un intervalo de confianza de 99%, hay que multiplicar el error estándar por 2.575 en lugar de 1.96. Esto hará el intervalo de confianza más amplio porque entre más amplio mayor la confianza de que la proporción de la población se ubicará dentro de dicho rango.

Ejemplo 2

Una compañía encuestadora toma una muestra al azar de 200 adultos en una ciudad y les pregunta si están de acuerdo con los planes del alcalde para construir un centro comercial en un área verde. Sólo 80 de los adultos dice que apoyan la propuesta.

a. Estime la proporción de la población adulta que apoya la propuesta y calcule un intervalo del 95% para su respuesta.
b. Calcule un intervalo de confianza de 99%. Explique porque es diferente al intervalo del 95%.
c. Un miembro del equipo del alcalde sostiene que la compañía encuestadora está equivocada y que el 60% de la población apoya la propuesta del centro comercial. ¿Considera que el/ella podría estar en lo cierto?
(OJO: ¿Cuál es la probabilidad de obtener una muestra de 200 personas con la proporción observada y que apoye la propuesta si la proporción real de la población es de 0.6?)

LA COMPROBACIÓN DE HIPÓTESIS

En ocasiones queremos saber si una muestra proviene de una población en particular o si es genuinamente diferente de dicha población.

Por ejemplo, imagine que usted trabaja para una empresa y que ésta le encargó que llevará a cabo una encuesta. Usted debe contratar los servicios de una empresa encuestadora. Una de dichas empresas que quiere conseguir el contrato sostiene que su personal tiene un coeficiente intelectual de 120 y una desviación estándar de 10. Usted decide investigar a dicha compañía y aplica un examen de inteligencia a una muestra aleatoria de 36 miembros de dicha compañía. El resultado de la muestra es de una media de CI de 114. Al parecer la muestra tiene un CI más bajo en promedio que lo que la compañía presume. Hay dos posibles razones para esto:

1. La aseveración de la compañía es cierta: su personal sí tiene un CI medio de 120. Por azar usted seleccionó una muestra con un CI más bajo.
2. La aseveración de la compañía es falsa: su personal no tiene un CI medio de 120.

La pregunta obligada es entonces, ¿cuál de las explicaciones es más probable que sea cierta? Justamente es aquí donde nos sirven las técnicas estadísticas que hacen pruebas de hipótesis.

Primero que nada, debemos escribir una hipótesis acerca de lo que esperamos si la aseveración de la compañía es verdadera. Esto es conocido como la hipótesis nula o H_0. La hipótesis nula para este ejemplo es:

H_0 = El personal de la compañía tiene un CI medio de 120

Ahora bien, en lo que respecta al resultado que usted obtuvo de la muestra, valdría la pena preguntarse si es una coincidencia que éste sea el resultado del azar. Esto es, ¿es posible que otra hipótesis alterna sea verdadera? Si esto es así, entonces pasamos al segundo paso.

Segundo paso: escribimos una hipótesis alterna, conocida como H_A o H_1. En el ejemplo en cuestión estamos pensando que la afirmación de la compañía es falsa, así que escribimos:

H_1 = El personal de la compañía no tiene un CI medio de 120.

Lo que puede suceder es que encontremos un rango de valores muestrales que no nos permitan rechazar la hipótesis nula, pero si el valor de la población se ubica fuera de dicho rango entonces decidiremos no aceptar la hipótesis nula. Lo que nos interesa, en suma, es rechazar la hipótesis nula.

LAS PRUEBAS DE SIGNIFICACIÓN
(1) EL MÉTODO DEL INTERVALO DE CONFIANZA

El proceso de examinar la hipótesis nula a la luz de evidencia muestral se le conoce como prueba de significación. Esto lo podemos hacer con los intervalos de confianza (aunque no son los más usados, nos permiten entender lo que está pasando).

Pasos para seguir:
Paso 1. Escribir las hipótesis: hipótesis nula e hipótesis alterna. Aquí se tiene que escribir en términos de la media poblacional.

Paso 2. Encontrar el estadístico para la muestra. Calcular la media para la muestra.

Paso 3. Encontrar un intervalo de confianza. El 95% para la media muestral.

Paso 4. Sacar una conclusión. ¿Cae el valor de la población dentro de este rango? Si la respuesta es afirmativa entonces no podemos rechazar la hipótesis nula; si la respuesta es negativa entonces podemos rechazarla.

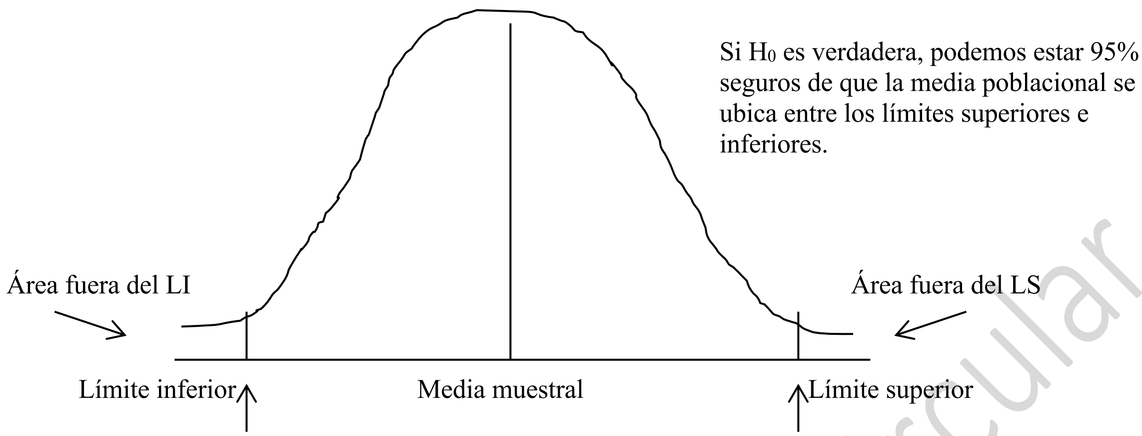
Ejemplo 1

1. $H_0 = \mu = 120$ (el personal de la compañía tiene un CI medio de 120).
   $H_1 = \mu \neq 120$ (el personal de la compañía no tiene un CI medio de 120)

2. El CI medio en nuestra muestra es de 114.

$$\bar{\bar{x}} = 114$$

3. Supongamos que estamos listos para estar equivocados 1 vez en 20 (equivocados 5% de las veces y acertados en un 95% de las veces). Entonces para una distribución normal, queremos encontrar los valores en los que en 95% la media poblacional se ubica aquí si la hipótesis nula es cierta. Esto se puede mostrar así:



Ahora bien, si estamos calculando un intervalo de confianza del 95%, entonces usamos el valor de 1.96. También tenemos que la desviación estándar, de acuerdo con la compañía, fue de 10. Por lo tanto, el intervalo de confianza será:

$$95\%\ \text{IC} = \text{media muestral} \pm \left| \frac{1.96 \ \text{x} \ \text{desviación estándar}}{\sqrt{n}} \right|$$

$$95\%\ \text{IC} = 114 \pm \left| \frac{1.96 \ \text{x} \ 10}{\sqrt{36}} \right| = 114 \pm 3.27 = (110.73,\ 117.27)$$

4. Conclusión. La media poblacional, 120, no se ubica dentro del rango (110.73, 117.27). Es muy poco probable que la muestra haya venido de una población con un CI de 120. Por lo tanto, podemos rechazar la H_0. Concluimos que el personal en la compañía no tiene un CI medio de 120. La muestra es significativamente diferente a lo que aseveraba la compañía, lo suficiente para que creamos que no nos estaban diciendo la verdad.