La distribución normal como aproximación de la distribución binomial

Si se satisfacen las condiciones $np \geq 5$ y $nq \geq 5$, entonces las probabilidades de una distribución de probabilidad se pueden aproximar bastante bien utilizando una distribución normal con media $\mu = np$ y desviación estándar $\sigma = \sqrt{npq}$. Puesto que la distribución de probabilidad binomial generalmente usa solo números enteros para la variable aleatoria x, mientras que la aproximación normal es continua, debemos aplicar una "corrección por continuidad", con un número entero x representado por el intervalo de $x - 0.5$ a $x + 0.5$. Nota: en vez de utilizar una distribución normal como aproximación de una distribución de probabilidad binomial, la mayoría de las aplicaciones prácticas de la distribución binomial se pueden manejar con un programa de cómputo o calculadora.

En la unidad 4 se estableció que la distribución de probabilidad binomial tiene: 1. Un número fijos de ensayos; 2. Ensayos que son independientes; 3. Ensayos que están clasificados en dos categorías que generalmente se denominan éxito o fracaso; 4. Ensayos con la propiedad de que la probabilidad de éxito permanece constante. También se debe de recordar la siguiente notación:

n = número fijo de ensayos.

x = el número específico de éxitos en n ensayos.

p = la probabilidad de éxito en uno de n ensayos.

q = la probabilidad de fracaso en uno de n ensayos.

Requisitos

1. Se trata de una muestra aleatoria simple de tamaño n de una población donde la proporción de éxitos es p, o bien, la muestra es el resultado de realizar n ensayos independientes de un experimento binomial en el que la probabilidad de éxito es p.

2. $np \geq 5$ y $nq \geq 5$

La distribución normal como aproximación de la distribución binomial

Aproximación normal

Si se satisfacen los requisitos anteriores, entonces la distribución de probabilidad de la variable aleatoria x puede aproximarse por medio de una distribución normal con los siguientes parámetros:

$$\mu = np$$

$$\sigma = \sqrt{npq}$$

Corrección por continuidad

Cuando utilice la aproximación normal, ajuste el número entero discreto x por medio de una corrección por continuidad, de manera que x esté representada por el intervalo de $x - 0.5$ a $x + 0.5$.

Note que los requisitos incluyen la verificación de $np \geq 5$ y $nq \geq 5$. El valor mínimo de 5 es común, pero no se trata de un valor absolutamente rígido, y algunos libros de texto utilizan 10 en su lugar. Este requerimiento se incluye en el siguiente procedimiento para el uso de una distribución normal como aproximación de una distribución binomial:

Procedimiento para el uso de una distribución normal como aproximación de una distribución binomial

1. Verifique que se cumplan los dos requisitos anteriores. Si no se satisfacen ambas condiciones, entonces debe utilizar un programa de cómputo, una calculadora, una tabla o la fórmula de probabilidad binomial.

2. Obtenga los valores de los parámetros μ y σ calculando $\mu = np$ y $\sigma = \sqrt{npq}$

3. Identifique el número entero discreto x que sea relevante para el problema de probabilidad binomial. Por ejemplo, si está tratando de calcular la probabilidad de obtener al menos 1150 éxitos en 40,000 ensayos, el número entero discreto de interés es x = 1150. Primero concéntrese en el propio valor de 1150, e ignore temporalmente si busca al menos 1150 o más que 1150, menos que 1150, a lo sumo 1150 o exactamente 1150.

4. Dibuje una distribución normal centrada alrededor de μ; luego dibuje el área de una franja vertical alrededor de x. Marque el lado izquierdo de la franja con el número que es igual a $x - 0.5$, y marque el lado derecho con el número que es igual a $x + 0.5$. (Con x = 1150, por ejemplo, dibuje la franja desde 1149.5 hasta 1150.5). Considere toda el área de la franja completa para representar la probabilidad del número entero discreto x.

5. Ahora determine si el valor de x debe incluirse en la probabilidad que busca (por ejemplo, "al menos x" sí incluye a la propia x, pero "más que x" no la incluye). Después, determine si busca la probabilidad de al menos x, a lo sumo x, más que x, menos que x o exactamente x. Sombree el área a la derecha o a la izquierda de la franja, según corresponda; también sombree el interior de la franja si y solo si se incluirá a la propia x. Esta región sombreada total corresponde a la probabilidad buscada.

6. Utilice $x - 0.5$ o $x + 0.5$ en vez de x, calcule el área de la región sombreada en el paso 5 como sigue. Primero, calcule la puntuación z: $z = \frac{x-\mu}{\sigma}$ (utilizando $x - 0.5$ o $x + 0.5$ en vez de x). En segundo lugar, use esa puntuación z para encontrar el área a la izquierda del valor ajustado de x. En tercer lugar, esa área puede emplearse ahora para identificar el área sombreada correspondiente a la probabilidad deseada.

Ejemplo:

Una persona recibió por correo una encuesta de los cruceros Viking River, la cual incluía la solicitud de una dirección de correo electrónico. Suponga que el formato de encuesta se envió a 40,000 personas y que, para ese tipo de encuestas, el porcentaje de respuestas con una dirección de correo electrónico es del 3%. Si el verdadero objetivo de la encuesta era el de conseguir un banco de al menos 1150 direcciones de correo electrónico, calcule la probabilidad de obtener al menos 1150 respuestas con direcciones correctas de correo electrónico.

El problema implica una distribución binomial con un número fijo de ensayos (n = 40,000) que son independientes. Hay dos categorías para cada encuesta: se obtiene una respuesta con una dirección de correo electrónico o no se obtiene.

La distribución normal como aproximación de la distribución binomial

Se supone que la probabilidad de éxito (p =0.03) permanece constante de un ensayo a otro. Los cálculos con la fórmula de probabilidad binomial no son prácticos porque tendríamos que aplicarla 38,851 veces, una para cada valor de x desde 1150 hasta 40,000, inclusive. Las calculadoras no pueden manejar el primer cálculo para la probabilidad de exactamente 1150 éxitos (algunas calculadoras proporcionan un resultado, pero utilizan un método de aproximación en vez de cálculo exacto). La mejor estrategia consiste en utilizar el método de los seis pasos para utilizar la distribución normal para aproximar la distribución binomial.

Paso 1. Verificación requerida: aunque no sabemos cómo se seleccionaron los sujetos encuestados, procederemos suponiendo que contamos con una muestra aleatoria simple. Debemos verificar que es razonable aproximar la distribución binomial con la distribución normal, porque $np \geq 5$ y $nq \geq 5$. Con n = 40,000, p = 0.03 y q = 1-p = 0.97, verificamos las condiciones requeridas como sigue:

np = (40,000)(0.03) = 1200 (por lo tanto $np \geq 5$)

nq = (40,000)(0.97) = 38,800 (por lo tanto $nq \geq 5$)

Paso 2. Ahora procedemos a calcular los valores de los parámetros μ y σ necesarios para la distribución normal. Obtenemos lo siguiente:

$$\mu = np = (40{,}000)(0.03) = 1200$$

$$\sigma = \sqrt{npq} = \sqrt{(40{,}000)(0.03)(0.97)} = 34.117444$$

Paso 3. Buscamos la probabilidad de al menos 1150 respuestas con direcciones de correo electrónico, de manera que x = 1150 es el número entero discreto relevante para este ejemplo.

Paso 4. En la figura:

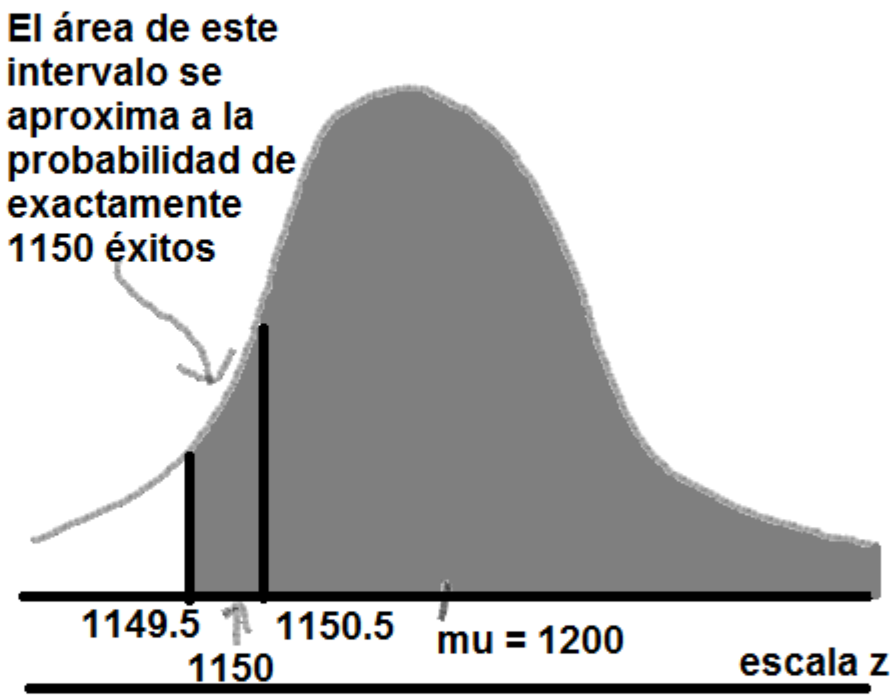

Imagen elaborada por Elsa Rivera (2015), Búsqueda de la probabilidad.

Se presenta una distribución normal con media $\mu = 1200$ y desviación estándar $\sigma = 34.117444$. Y en la figura también se observa la franja vertical desde 1149.5 hasta 1150.5.

Paso 5. Queremos encontrar la probabilidad de al menos 1150 respuestas con direcciones de correo electrónico, por lo que deseamos sombrear la franja vertical que representa 1150, así como el área ubicada a su derecha, en la figura anterior el área deseada aparece sombreada.

Paso 6. Buscamos el área ubicada a la derecha de 1149.5 en la figura, de manera que la puntuación z se obtiene utilizando los valores de μ y σ del paso 2, y el valor límite de 1149.5, como sigue:

$$z = \frac{x - \mu}{\sigma} = \frac{1149.5 - 1200}{34.117444} = -1.48$$

Al emplear la tabla de la distribución normal estándar, encontramos que z = -1.48 corresponde a un área de 0.0694, de manera que la región sombreada de la figura es 1 – 0.0694 = 0.9306.

Interpretación: existe una probabilidad de 0.9306 de obtener al menos 1150 respuestas con direcciones de correo electrónico entre un total de 40,000 encuestas enviadas. Esta probabilidad es lo suficientemente alta para concluir que es muy probable que la empresa de cruceros Viking River logre su meta de obtener al menos 1150 respuestas con direcciones de correo electrónico. Si la empresa Viking River utiliza un método de muestreo que no proporciona una muestra aleatoria simple, entonces la probabilidad resultante de 0.9306 podría ser errónea. Por ejemplo, si solo encuestaron a clientes del pasado, tienen más probabilidades de obtener una tasa de respuesta más alta, de manera que los cálculos anteriores podrían ser incorrectos. Nunca debemos olvidar la importancia de un método de muestreo adecuado.

Correcciones por continuidad

El procedimiento que implica el uso de la distribución normal como aproximación de la distribución binomial incluye una corrección por continuidad, que se define de la siguiente manera:

Definición: cuando empleamos la distribución normal (que es una distribución de probabilidad continua) como una aproximación de la distribución binomial (que es discreta), se realiza una corrección por continuidad a un número entero discreto x en la distribución binomial, representando el número entero discreto x en el intervalo de x-0.5 a x+0.5 (es decir, sumando y restando 0.5).

En el procedimiento anterior de seis pasos para utilizar una distribución normal como aproximación de una distribución binomial, los pasos 3 y 4 incorporan la corrección por continuidad.

Algunos ejemplos de correcciones por continuidad son:

Afirmación	Área
Al menos 8 (incluye 8 y números mayores)	A la derecha de 7.5
Más de 8 (no incluye a 8)	A la derecha de 8.5
A lo sumo 8 (incluye 8 y números menores)	A la izquierda de 8.5
Menos de 8 (no incluye a 8)	A la izquierda de 7.5
Exactamente 8	Entre 7.5 y 8.5

Ejemplo: una encuesta reciente del Pew Research Center reveló que, de 2822 adultos seleccionados al azar, 2060 (o el 73%) dijeron ser usuarios de internet. Si la proporción de los adultos que utilizan internet es en realidad de 0.75, calcule la probabilidad de que una muestra aleatoria de 2822 adultos dé por resultado exactamente 2060 usuarios de internet.

Tenemos n = 2822 sujetos entrevistados de manera independiente y x = 2060 de ellos son usuarios de internet; suponemos que la proporción de la población es p = 0.75, de lo que se deduce que q = 0.25. Utilizaremos una distribución normal para aproximar la distribución binomial.

Paso 1. Primero verificamos los requisitos. El Pew Research Center tiene la reputación de utilizar técnicas de encuestas sólidas, de manera que es razonable tratar la muestra como si fuera aleatoria simple. Ahora verificamos los requisitos $np \geq 5$ y $nq \geq 5$.

np= (2822)(0.75)=2116.5, por lo tanto $np \geq 5$

nq = (2822)(0.25) = 705.5, por lo tanto $nq \geq 5$

Paso 2. Ahora procedemos a calcular los valores de μ y σ, y obtenemos lo siguiente:

$$\mu = np = (2822)(0.75) = 2116.5$$

$$\sigma = \sqrt{npq} = \sqrt{(2822)(0.75)(0.25)} = 23.002717$$

Paso 3. Buscamos la probabilidad de exactamente 2060 usuarios de internet, de manera que el número entero discreto relevante para este ejemplo es 2060.

Paso 4. Observe la figura:

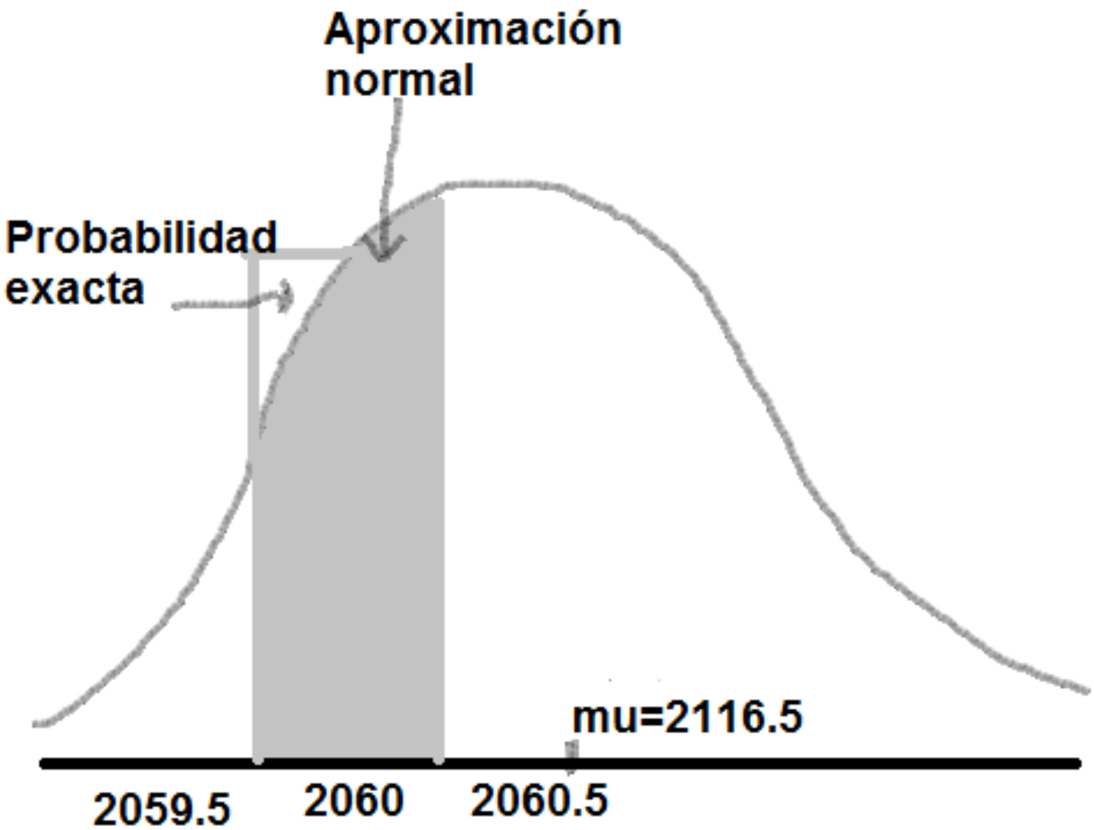

Imagen elaborada por Elsa Rivera (2015), Corrección por continuidad.

La cual presenta una distribución normal con media $\mu = 2116.5$ y desviación estándar $\sigma = 23.002717$. En la figura también se observa la franja vertical desde 2059.5 hasta 2060.5, que representa la probabilidad de exactamente 2060 usuarios de internet.

Paso 5. Como queremos encontrar la probabilidad de exactamente 2060 usuarios de internet, deseamos conocer el área sombreada en la figura anterior.

Paso 6. Para obtener la región sombreada en la figura anterior, primero calculamos el área total ubicada a la izquierda de 2060.5, y luego el área total ubicada a la izquierda de 2059.5. Después, calculamos la diferencia entre las dos áreas. Comencemos con el área total a la izquierda de 2060.5. Al utilizar la tabla de la distribución normal estándar primero debemos encontrar la puntuación z correspondiente a 2060.5. Obtenemos:

$$z = \frac{2060.5 - 2116.5}{23.002717} = -2.43$$

Utilizamos la tabla de la distribución normal para encontrar que z = -2.43 corresponde a una probabilidad de 0.0075, que es el área total a la izquierda de 2060.5. Ahora obtenemos el área a la izquierda de 2059.5, calculando primero la puntuación z correspondiente a 2059.5:

$$z = \frac{2059.5 - 2116.5}{23.002717} = -2.48$$

Utilizamos la tabla de la distribución normal estándar para encontrar que z = -2.48 corresponde a una probabilidad de 0.0066, que es el área total a la izquierda de 2059.5. El área sombreada es 0.0075-0.0066=0.0009.

Interpretación: si suponemos que el 75% de todos los adultos utilizan internet, la probabilidad de encontrar exactamente 2060 usuarios de internet entre 2822 adultos elegidos al azar es de 0.0009 (si utilizamos herramientas tecnológicas, la probabilidad es de 0.000872). Esta probabilidad nos indica que si el porcentaje de usuarios de internet en la población adulta es del 75%, entonces es sumamente improbable que obtengamos exactamente 2060 usuarios de internet al encuestar 2822 adultos, la probabilidad de encontrar cualquier número específico de usuarios de internet es muy pequeña.

Interpretación de los resultados

Cuando utilizamos una distribución normal como aproximación de la distribución binomial, nuestra meta es simplemente calcular un número de probabilidad. A menudo necesitamos hacer algún juicio con base en el valor de probabilidad. El siguiente criterio describe la aplicación de las probabilidades para distinguir resultados que pueden ocurrir fácilmente por azar de aquellos que son muy poco comunes.

La distribución normal como aproximación de la distribución binomial

Uso de las probabilidades para determinar cuando los resultados son inusuales.

Inusualmente alto: x éxitos en n ensayos es un número inusualmente alto de éxitos si P(x o más) es muy pequeña (como 0.05 o menos).

Inusualmente bajo: x éxitos en n ensayos es un número inusualmente bajo de éxitos si P(x o menos) es muy pequeña (como 0.05 o menos).

Referencia:
Triola, M., 2013, Estadística, Pearson Educación**.**