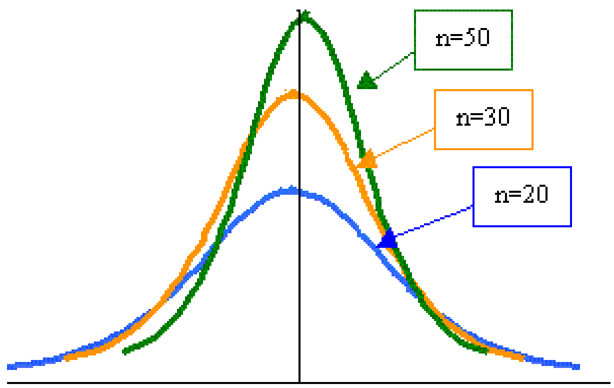
Teorema del límite central

Si se seleccionan muestras aleatorias de n observaciones de una población con media μ y desviación estándar σ, entonces, cuando n es grande, la distribución muestral de medias tendrá aproximadamente una distribución normal con una media igual a μ y una desviación estándar de $\sigma/\sqrt{n}$. La aproximación será cada vez más exacta a medida de que n sea cada vez mayor.



Ejemplo

Para la distribución muestral de medias del ejercicio pasado, encuentre:

a. El error muestral de cada media
b. La media de los errores muestrales
c. La desviación estándar de los errores muestrales.

Solución:

a. En la tabla siguiente se ven las muestras, las medias de las muestras y los errores muestrales:

Muestra	x	Error muestral, e=x- μ
(0,0)	0	0 - 3 = -3
(0,2)	1	1 - 3 = -2
(0,4)	2	2 - 3 = -1
(0,6)	3	3 – 3 = 0
(2,0)	1	1 – 3 = -2
(2,2)	2	2 – 3 = -1
(2,4)	3	3 – 3 = 0
(2,6)	4	4 – 3 = 1
(4,0)	2	2 – 3 = -1
(4,2)	3	3 – 3 = 0
(4,4)	4	4 – 3 = 1
(4,6)	5	5 – 3 = 2
(6,0)	3	3 – 3 = 0
(6,2)	4	4 – 3 = 1
(6,4)	5	5 – 3 = 2
(6,6)	6	6 – 3 = 3

b. La media de los errores muestrales es μ_e, es:

$$\mu_e = \frac{(-3)+(-2)+(-1)+0+\ldots+2+3}{16} = 0$$

c. La desviación estándar de la distribución de los errores muestrales σ_e entonces:

$$\sigma_e = \sqrt{\frac{\Sigma\left[(e-\mu_e)^2 f\right]}{N}} = \sqrt{\frac{(-3-0)^2 1+(-2-0)^2 2+(-1-0)^2 3+(0-0)^2 4+(1-0)^2 3+(2-0)^2 2+(3-0)^2 1}{16}} = 1.58$$

La desviación estándar de la distribución muestral de un estadístico se conoce como ***error estándar del estadístico****.* Para el ejercicio anterior el error estándar de la media denotado por σ_x, es 1.58. Con esto se puede demostrar que si de una población se eligen muestras de tamaño *n* ***con reemplazo***, entonces el error estándar de la media es igual a la desviación estándar de la distribución de los errores muestrales.

En general se tiene: $\sigma_x = \sigma_e$

Cuando las muestras se toman de una población pequeña y sin reemplazo, se puede usar la formula siguiente para encontrar σ_x .

$$\sigma_x = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

donde σ es la desviación estándar de la población de donde se toman las muestras, *n* es el tamaño de la muestra y *N* el de la población.

Como regla de cálculo, si el muestreo se hace sin reemplazo y el tamaño de la población es al menos 20 veces el tamaño de la muestra (N$\geq$20), entonces se puede usar la fórmula.

El factor $\sqrt{\frac{N-n}{N-1}}$ se denomina ***factor de corrección*** para una población finita.

Ejemplo:

Suponga que la tabla siguiente muestra la antigüedad en años en el trabajo de tres maestros universitarios de matemáticas:

Maestro de matemáticas	**Antigüedad**
A	6
B	4
C	2

Suponga además que se seleccionan muestras aleatorias de tamaño 2 sin reemplazo. Calcule la antigüedad media para cada muestra, la media de la distribución muestral y el error estándar, o la desviación estándar de la distribución muestral.

Solución:

Se pueden tener ${}_3C_2 = 3$ muestras posibles. La tabla lista todas las muestras posibles de tamaño 2, con sus respectivas medias muestrales.

Muestras	**Antigüedad**	**Media Muestral**
A,B	(6,4)	5
A,C	(6,2)	4
B,C	(4,2)	3

La media poblacional es: $\mu = \dfrac{2+4+6}{3} = 4$

La media de la distribución muestral es: $\mu_{\bar{x}} = \dfrac{5+4+3}{3} = 4$

La desviación estándar de la población es:

$$\sigma = \sqrt{\frac{(6-4)^2 + (4-4)^2 + (2-4)^2}{3}} = 1.63$$

El error estándar o la desviación estándar de la distribución muestral es:

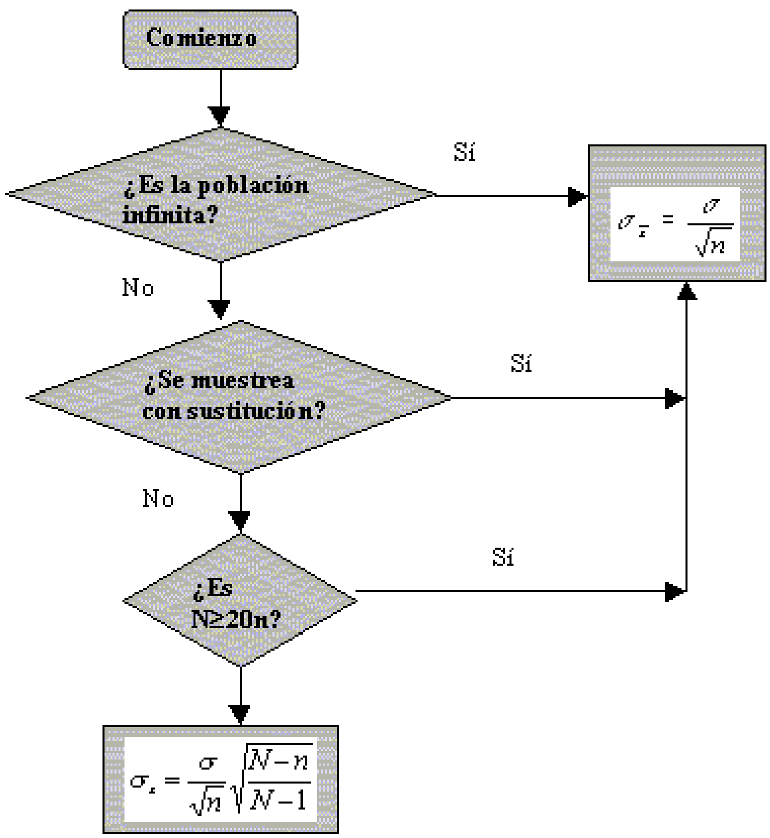
$$\sigma_{\bar{x}} = \sqrt{\frac{(5-4)^2 + (4-4)^2 + (3-4)^2}{3}} = 0.816$$

Si utilizamos la fórmula del error estándar sin el factor de corrección tendríamos que: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.63}{\sqrt{2}} = 1.152$

Por lo que observamos que este valor no es el verdadero. Agregando el factor de corrección obtendremos el valor correcto:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{1.63}{\sqrt{2}}\sqrt{\frac{3-2}{3-1}} = 0.816$$

El diagrama de flujo resume las decisiones que deben tomarse cuando se calcula el valor del error estándar:



Distribución Muestral de Medias

Si recordamos a la distribución normal, esta es una distribución continua, en forma de campana en donde la media, la mediana y la moda tienen un mismo valor y es simétrica.

Con esta distribución podíamos calcular la probabilidad de algún evento relacionado con la variable aleatoria, mediante la siguiente fórmula:

$$z = \frac{x - \mu}{\sigma}$$

En donde z es una variable estandarizada con media igual a cero y varianza igual a uno. Con esta fórmula se pueden a hacer los cálculos de probabilidad para cualquier ejercicio, utilizando la tabla de la distribución z.

Sabemos que cuando se extraen muestras de tamaño mayor a 30 o bien de cualquier tamaño de una población normal, la distribución muestral de medias tiene un comportamiento aproximadamente normal, por lo que se puede utilizar la fórmula de la distribución normal con $\mu = \mu_x$ y $\sigma = \sigma_x$, entonces la fórmula para calcular la probabilidad del comportamiento del estadístico, en este caso la media de la muestra, quedaría de la siguiente manera:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

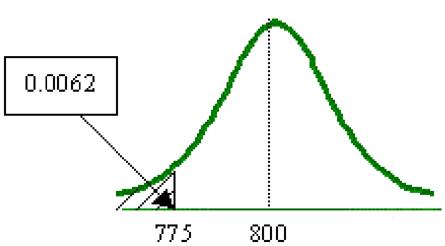
y para poblaciones finitas y muestro con reemplazo:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n} \sqrt{\frac{N - n}{N - 1}}}$$

Ejemplo:

Una empresa eléctrica fabrica focos que tienen una duración que se distribuye aproximadamente en forma normal, con media de 800 horas y desviación estándar de 40 horas. Encuentre la probabilidad de que una muestra aleatoria de 16 focos tenga una vida promedio de menos de 775 horas.

Solución:



$$z = \frac{775 - 800}{40 / \sqrt{16}} = -2.5$$

Este valor se busca en la tabla de z

$$P(\bar{x} \leq 775) = P(z \leq -2.5) = 0.0062$$

La interpretación sería que la probabilidad de que la media de la muestra de 16 focos sea menor a 775 horas es de 0.0062.

Ejemplo:

Las estaturas de 1000 estudiantes están distribuidas aproximadamente en forma normal con una media de 174.5 centímetros y una desviación estándar de 6.9 centímetros. Si se extraen 200 muestras aleatorias de tamaño 25 sin reemplazo de esta población, determine:

a. El número de las medias muestrales que caen entre 172.5 y 175.8 centímetros.
b. El número de medias muestrales que caen por debajo de 172 centímetros.

Solución:

Como se puede observar en este ejercicio se cuenta con una población finita y un muestreo sin reemplazo, por lo que se tendrá que agregar el factor de corrección. Se procederá a calcular el denominador de Z para sólo sustituirlo en cada inciso.

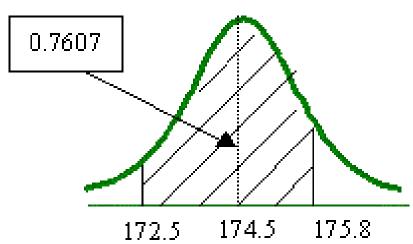
$$\sigma/\sqrt{n}\sqrt{\frac{N-n}{N-1}} = 6.9/\sqrt{25}\sqrt{\frac{1000-25}{1000-1}} = 1.36$$

a.

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}\sqrt{\frac{N-n}{N-1}}} = \frac{172.5-174.5}{1.36} = -1.47$$

$$z = \frac{175.8-174.5}{1.36} = 0.96$$

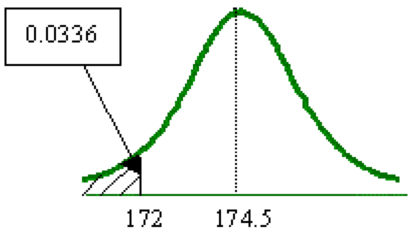
$$p(172.5 \leq \bar{x} \leq 175.8) = 0.7607$$



(0.7607)(200)=152 medias muestrales

b. $z = \dfrac{172 - 174.5}{1.36} = -1.83$

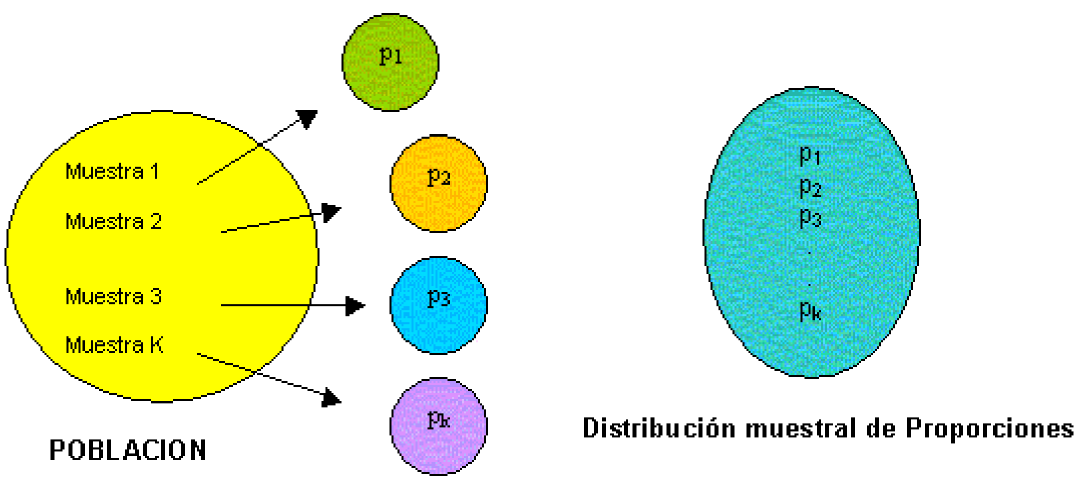
$p(\bar{x} \leq 172) = 0.00336$



(0.0336)(200)= 7 medias muestrales

Distribución muestral de Proporciones

Existen ocasiones en las cuales no estamos interesados en la media de la muestra, sino que queremos investigar la proporción de artículos defectuosos o la proporción de alumnos reprobados en la muestra. La distribución muestral de proporciones es la adecuada para dar respuesta a estas situaciones. Esta distribución se genera de igual manera que la distribución muestral de medias, a excepción de que al extraer las muestras de la población se calcula el estadístico proporción (p=x/n en donde "x" es el número de éxitos u observaciones de interés y "n" el tamaño de la muestra) en lugar del estadísitico media.



Una población binomial está estrechamente relacionada con la distribución muestral de proporciones; una población binomial es una colección de éxitos y fracasos, mientras que una distribución muestral de proporciones contiene las posibilidades o proporciones de todos los números posibles de éxitos en un experimento binomial, y como consecuencia de esta relación, las afirmaciones probabilísticas referentes a la proporción muestral pueden evaluarse usando la aproximación normal a la binomial, siempre que $np \geq 5$ y $n(1-p) \geq 5$. Cualquier evento se puede convertir en una proporción si se divide el número obtenido entre el número de intentos.

Generación de la Distribución Muestral de Proporciones

Suponga que se cuenta con un lote de 12 piezas, el cual tiene 4 artículos defectuosos. Se van a seleccionar 5 artículos al azar de ese lote sin reemplazo. Genere la distribución muestral de proporciones para el número de piezas defectuosas.

Como se puede observar en este ejercicio la Proporción de artículos defectuosos de esta población es 4/12=1/3. Por lo que podemos decir que el 33% de las piezas de este lote están defectuosas.

El número posible de muestras de tamaño 5 a extraer de una población de 12 elementos es ${}_{12}C_5$=792, las cuales se pueden desglosar de la siguiente manera:

Artículos Buenos	Artículos Malos	Proporción de artículos defectuoso	Número de maneras en las que se puede obtener la muestra
1	4	4/5=0.8	${}_8C_1*{}_4C_4=8$
2	3	3/5=0.6	${}_8C_2*{}_4C_3=112$
3	2	2/5=0.4	${}_8C_3*{}_4C_2=336$
4	1	1/5=0.2	${}_8C_4*{}_4C_1=280$
5	0	0/5=0	8C5*4C0=56
Total	**792**		

Para calcular la media de la distribución muestra de proporciones se tendría que hacer la sumatoria de la frecuencia por el valor de la proporción muestra y dividirla entre el número total de muestras. Esto es:

$$\mu_p = \frac{(0.8*8)+(0.6*112)+(0.4*336)+(0.2*280)+(0*56)}{792} = \frac{1}{3} = 0.3333$$

Como podemos observar la media de la distribución muestra de proporciones es igual a la Proporción de la población.

$$\mu_p = P$$

También se puede calcular la desviación estándar de la distribución muestra de proporciones:

$$\sigma_p = \sqrt{\frac{(0.8-1/3)^2*8+(0.6-1/3)^2*112+(0.4-1/3)^2*336+(0.2-1/3)^2*280+(0-1/3)^2*56}{792}} = 0.1681$$

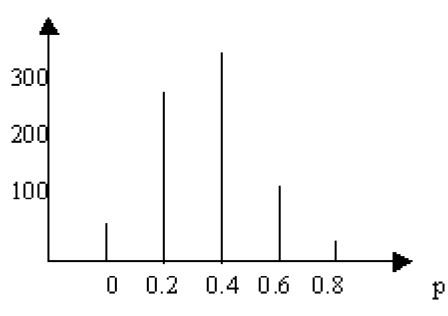
La varianza de la distribución binomial es $\sigma^2 = npq$ por lo que la varianza de la distribución muestral de proporciones es $\sigma_P^2 = \frac{Pq}{n}$. Si se sustituyen los valores en esta fórmula tenemos que:

$$\sigma_p = \sqrt{\frac{(1/3)(2/3)}{5}} = 0.2108$$

, este valor no coincide con el de 0.1681, ya que nos falta agregar el factor de corrección para una población finita y un muestreo sin reemplazo:

$$\sigma_p = \sqrt{\frac{(1/3)(2/3)}{5}}\sqrt{\frac{12-5}{12-1}} = 0.1681$$

$$\sigma_p = \sqrt{\frac{Pq}{n}}\sqrt{\frac{N-n}{N-1}}$$



Gráfica de frecuencias para las proporciones de las muestras

La fórmula que se utilizará para el cálculo de probabilidad en una distribución muestral de proporciones está basada en la aproximación de la distribución normal a la binomial. Esta fórmula nos servirá para calcular la probabilidad del comportamiento de la proporción en la muestra.

$$z = \frac{p - P}{\sqrt{\frac{Pq}{n}}}$$

A esta fórmula se le puede agregar el factor de corrección de $\sqrt{\frac{N-n}{N-1}}$ si se cumple con las condiciones necesarias.

Ejemplo:

Se ha determinado que 60% de los estudiantes de una universidad grande fuman cigarrillos. Se toma una muestra aleatoria de 800 estudiantes. Calcule la probabilidad de que la proporción de la muestra de la gente que fuma cigarrillos sea menor que 0.55.

Solución:

Este ejercicio se puede solucionar por dos métodos. El primero puede ser con la aproximación de la distribución normal a la binomial y el segundo utilizando la fórmula de la distribución muestral de proporciones.

Aproximación de la distribución normal a la binomial:

Datos:

n=800 estudiantes

p=0.60

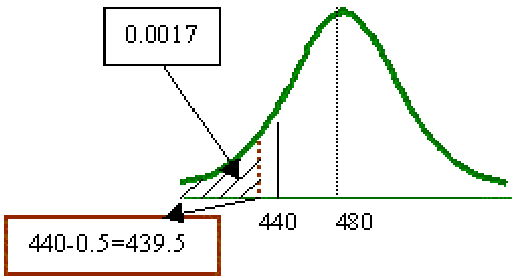
x= (.55)(800) = 440 estudiantes

p(x< 440) = ?

Media= np= (800)(0.60)= 480

$$z = \frac{x - np}{\sqrt{npq}} = \frac{439.5 - 480}{\sqrt{800(0.60)(0.40)}} = -2.92$$

p(x< 440) = 0.0017. Este valor significa que existe una probabilidad del 0.17% de que al extraer una muestra de 800 estudiantes, menos de 440 fuman cigarrillos.



Distribución Muestral de Proporciones

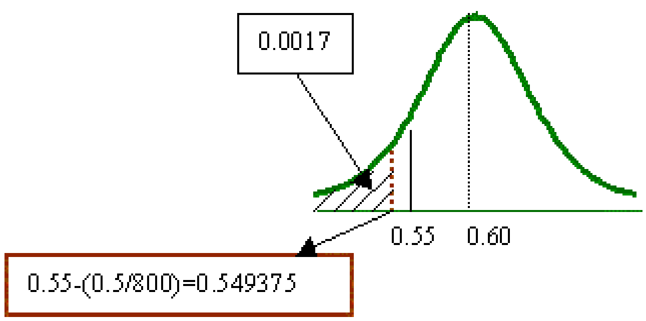
Datos:

n=800 estudiantes

P=0.60

p= 0.55

p(p< 0.55) = ?



$$z = \frac{p - P}{\sqrt{\frac{Pq}{n}}} = \frac{0.549375 - 0.60}{\sqrt{\frac{(0.60)(0.40)}{800}}} = -2.92$$

Observe que este valor es igual al obtenido en el método de la aproximación de la distribución normal a la binomial, por lo que si lo buscamos en la tabla de "z" nos da la misma probabilidad de 0.0017. También se debe de tomar en cuenta que el factor de corrección de 0.5 se está dividiendo entre el tamaño de la muestra, ya que estamos hablando de una proporción.

La interpretación en esta solución, estaría enfocada a la proporción de la muestra, por lo que diríamos que ***la probabilidad de que al extraer una muestra de 800 estudiantes de esa universidad, la proporción de estudiantes que fuman cigarrillos sea menor al 55% es del 0.17%.***

Ejemplo:

Un medicamento para malestar estomacal tiene la advertencia de que algunos usuarios pueden presentar una reacción adversa a él, más aún, se piensa que alrededor del 3% de los usuarios tienen tal reacción. Si una muestra aleatoria de 150 personas con malestar estomacal usa el medicamento, encuentre la probabilidad de que la proporción de la muestra de los usuarios que realmente presentan una reacción adversa, exceda el 4%.

a. Resolverlo mediante la aproximación de la normal a la binomial
b. Resolverlo con la distribución muestral de proporciones

a. Aproximación de la distribución normal a la binomial:

   Datos:

   n=150 personas

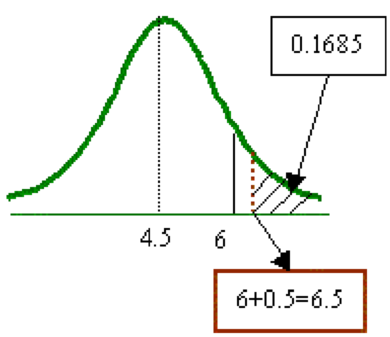
   p=0.03

   x= (0.04)(150) = 6 personas

   p(x>6) = ?

   Media = np= (150)(0.03)= 4.5

$$z = \frac{x - np}{\sqrt{npq}} = \frac{6.5 - 4.5}{\sqrt{150(0.03)(0.97)}} = 0.96$$



   p(x>6) = 0.1685. Este valor significa que existe una probabilidad del 17% de que al extraer una muestra de 150 personas, mas de 6 presentarán una reacción adversa.

b. Distribución Muestral de Proporciones

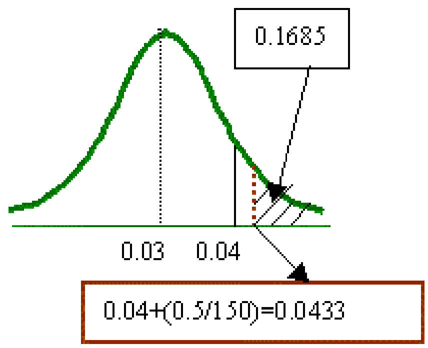
Datos:

n=150 personas

P=0.03

p= 0.04

p (p>0.04) =?



$$z = \frac{p - P}{\sqrt{\frac{Pq}{n}}} = \frac{0.0433 - 0.03}{\sqrt{\frac{(0.03)(0.97)}{150}}} = 0.96$$

Observe que este valor es igual al obtenido y la interpretación es: existe una probabilidad del 17% de que al tomar una muestra de 150 personas se tenga una proporción mayor de 0.04 presentando una reacción adversa.

Ejemplo:

Se sabe que la verdadera proporción de los componentes defectuosos fabricados por una firma es de 4%, y encuentre la probabilidad de que una muestra aleatoria de tamaño 60 tenga:

a. Menos del 3% de los componentes defectuosos.
b. Más del 1% pero menos del 5% de partes defectuosas.

Solución:

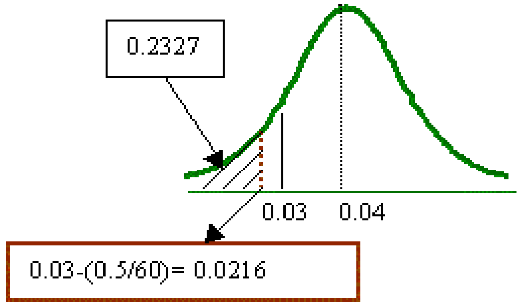
a. Datos:

   n= 60 artículos

   P=0.04

   p= 0.03

p(p<0.03) = ?



$$z = \frac{p - P}{\sqrt{\frac{Pq}{n}}} = \frac{0.0216 - 0.04}{\sqrt{\frac{(0.04)(0.96)}{60}}} = -0.73$$

La probabilidad de que en una muestra de 60 artículos exista una proporción menor de 0.03 artículos defectuosos es de 0.2327.

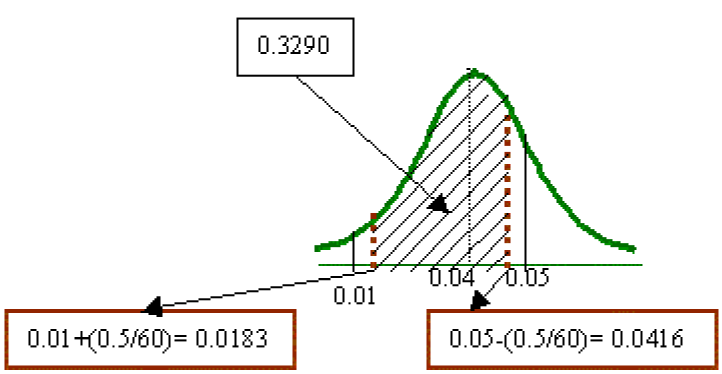
b. Datos:

n= 60 artículos

P=0.04

p= 0.01 y 0.05

p(0.01<p<0.05) = ?



$$z = \frac{p - P}{\sqrt{\frac{Pq}{n}}} = \frac{0.0183 - 0.04}{\sqrt{\frac{(0.04)(0.96)}{60}}} = -0.86 \quad z = \frac{p - P}{\sqrt{\frac{Pq}{n}}} = \frac{0.0416 - 0.04}{\sqrt{\frac{(0.04)(0.96)}{60}}} = 0.06$$