INDICE

UNIDAD I

UNIDAD II

TEORIA DEL MUESTREO

Uno de los propósitos de la estadística inferencial es estimar las características poblacionales desconocidas, examinando la información obtenida de una muestra, de una población. El punto de interés es la muestra, la cual debe ser representativa de la población objeto de estudio.
Se seguirán ciertos procedimientos de selección para asegurar de que las muestras reflejen observaciones a la población de la que proceden, ya que solo se pueden hacer observaciones probabilísticas sobre una población cuando se usan muestras representativas de la misma.

Una ***población*** está formada por la totalidad de las observaciones en las cuales se tiene cierto observa.
Una ***muestra*** es un subconjunto de observaciones seleccionadas de una población.

Muestras Aleatorias

Cuando nos interesa estudiar las características de poblaciones grandes, se utilizan muestras por muchas razones; una enumeración completa de la población, llamada censo, puede ser económicamente imposible, o no se cuenta con el tiempo suficiente.

A continuación se verá algunos usos del muestreo en diversos campos:

1. *Política*. Las muestras de las opiniones de los votantes se usan para que los candidatos midan la opinión pública y el apoyo en las elecciones.
2. *Educación*. Las muestras de las calificaciones de los exámenes de estudiantes se usan para determinar la eficiencia de una técnica o programa de enseñanza.
3. *Industria*. Muestras de los productos de una línea de ensamble sirve para controlar la calidad.
4. *Medicina*. Muestras de medidas de azúcar en la sangre de pacientes diabéticos prueban la eficacia de una técnica o de un fármaco nuevo.
5. *Agricultura*. Las muestras del maíz cosechado en una parcela proyectan en la producción los efectos de un fertilizante nuevo.
6. *Gobierno*. Una muestra de opiniones de los votantes se usaría para determinar los criterios del público sobre cuestiones relacionadas con el bienestar y la seguridad nacional.

Errores en el Muestreo

Cuando se utilizan valores muestrales, o ***estadísticos*** para estimar valores poblacionales, o ***parámetros***, pueden ocurrir dos tipos generales de errores: el error muestral y el error no muestral.

El ***error muestral*** se refiere a la variación natural existente entre muestras tomadas de la misma población.

Cuando una muestra no es una copias exacta de la población; aún si se ha tenido gran cuidado para asegurar que dos muestras del mismo tamaño sean representativas de una cierta población, no esperaríamos que las dos sean idénticas en todos sus detalles. El error muestral es un concepto importante que ayudará a entender mejor la naturaleza de la estadística inferencial.

Los errores que surgen al tomar las muestras no pueden clasificarse como errores muestrales y se denominan ***errores no muestrales***.

El sesgo de las muestras es un tipo de error no muestral. El **sesgo muestral** se refiere a una tendencia sistemática inherente a un método de muestreo que da estimaciones de un parámetro que son, en promedio, menores (sesgo negativo), o mayores (sesgo positivo) que el parámetro real.

El sesgo muestral puede suprimirse, o minimizarse, usando la aleatorización.

La ***aleatorización*** se refiere a cualquier proceso de selección de una muestra de la población en el que la selección es imparcial o no está sesgada; una muestra elegida con procedimientos aleatorios se llama ***muestra aleatoria***.

Los tipos más comunes de técnicas de muestreo aleatorios son el muestreo aleatorio simple, el muestreo estratificado, el muestreo por conglomerados y el muestreo sistemático.

Si una muestra aleatoria se elige de tal forma que todos los elementos de la población tengan la misma probabilidad de ser seleccionados, la llamamos **muestra aleatoria simple**.

Ejemplo 1.1

Suponga que nos interesa elegir una muestra aleatoria de 5 estudiantes en un grupo de estadística de 20 alumnos. ${}_{20}C_5$ da el número total de formas de elegir una muestra no ordenada y este resultado es 15,504 maneras diferentes de tomar la muestra. Si listamos las 15,504 en trozos separados de papel, una tarea tremenda, luego los colocamos en un recipiente y después los revolvemos, entonces podremos tener una muestra aleatoria de 5 si seleccionamos un trozo de papel con cinco nombres. Un procedimiento más simple para elegir una muestra aleatoria sería escribir cada uno de los 20 nombres en pedazos separados de papel, colocarlos en un recipiente, revolverlos y después extraer cinco papeles al mismo tiempo.

Otro método parea obtener una muestra aleatoria de 5 estudiantes en un grupo de 20 utiliza una tabla de números aleatorios. Se puede construir la tabla usando una calculadora o una computadora. También se puede prescindir de estas y hacer la tabla escribiendo diez dígitos del 0 al 9 en tiras de papel, las colocamos en un recipiente y los revolvemos, de ahí, la primera tira seleccionada determina el primer número de la tabla, se regresa al recipiente y después de revolver otra vez se selecciona la seguida tira que determina el segundo número de la tabla; el proceso continúa hasta obtener una tabla de dígitos aleatorios con tantos números como se desee.

Hay muchas situaciones en las cuales el muestreo aleatorio simple es poco práctico, imposible o no deseado; aunque sería deseable usar muestras aleatorias simples para las encuestas nacionales de opinión sobre productos o sobre elecciones presidenciales, sería muy costoso o tardado.

El **muestreo estratificado** requiere de separar a la población según grupos que no se traslapen llamados ***estratos***, y de elegir después una muestra aleatoria simple en cada estrato. La información de las muestras aleatorias simples de cada estrato constituiría entonces una muestra global.

Ejemplo 1.2

Suponga que nos interesa obtener una muestra de las opiniones de los profesores de una gran universidad. Puede ser difícil obtener una muestra con todos los profesores, así que supongamos que elegimos una muestra aleatoria de cada colegio, o departamento académico; los estratos vendrían a ser los colegios, o departamentos académicos.

El **muestreo por conglomerados** requiere de elegir una muestra aleatoria simple de unidades heterogéneas entre sí de la población llamadas ***conglomerados.*** Cada elemento de la población pertenece exactamente a un conglomerado, y los elementos dentro de cada conglomerado son usualmente heterogéneos o disímiles.

Ejemplo 1.3

Suponga que una compañía de servicio de televisión por cable está pensando en abrir una sucursal en una ciudad grande; la compañía planea realizar un estudio para determinar el porcentaje de familias que utilizarían sus servicios, como no es práctico preguntar en cada casa, la empresa decide seleccionar una parte de la ciudad al azar, la cual forma un conglomerado.

En el muestreo por conglomerados, éstos se forman para representar, tan fielmente como sea posible, a toda la población; entonces se usa una muestra aleatoria simple de conglomerados para estudiarla. Los estudios de instituciones sociales como iglesias, hospitales, escuelas y prisiones se realizan, generalmente, con base en el muestreo por conglomerados.

El muestreo sistemático es una técnica de muestreo que requiere de una selección aleatoria inicial de observaciones seguida de otra selección de observaciones obtenida usando algún sistema o regla.

Ejemplo 1.4

Para obtener una muestra de suscriptores telefónicos en una ciudad grande, puede obtenerse primero una muestra aleatoria de los números de las páginas del directorio telefónico; al elegir el vigésimo nombre de cada página obtendríamos un muestreo sistemático, también podemos escoger un nombre de la primera página del directorio y después seleccionar cada nombre del lugar número cien a partir del ya seleccionado. Por ejemplo, podríamos seleccionar un número al azar entre los primeros 100; supongamos que el elegido es el 40, entonces seleccionamos los nombres del directorio que corresponden a los números 40, 140, 240, 340 y así sucesivamente.

Error Muestral

Cualquier medida conlleva algún error. Si se usa la media para medir, estimar, la media poblacional μ, entonces la media muestral, como medida, conlleva algún error. Por ejemplo, supongamos que se ha obtenido una muestra aleatoria de tamaño 25 de una población con media $\mu = 15$: si la media de la muestra es $x=12$, entonces a la diferencia observada $x-\mu = -3$ se le denomina el ***error muestral***. Una media muestral x puede pensarse como la suma de dos cantidades, la media poblacional μ y el error muestral; si *e* denota el error muestral, entonces:
$X = \mu + e$

Ejemplo 1.5

Se toman muestras de tamaño 2 de una población consistente en tres valores, 2, 4 y 6, para simular una población “grande” de manera que el muestreo pueda realizarse un gran número de veces, supondremos que éste se hace con reemplazo, es decir, el número elegido se reemplaza antes de seleccionar el siguiente, además, se seleccionan muestras ordenadas. En una muestra ordenada, el orden en que se seleccionan las observaciones es importante, por tanto, la muestra ordenada (2,4) es distinta de la muestra ordenada (4,2). En la muestra (4,2), se seleccionó primero 4 y después 2. La siguiente tabla contiene una lista de todas las muestras ordenadas de tamaño 2 que es posible seleccionar con reemplazo y también contiene las medioas muestrales y los correspondientes errores muestrales. La media poblacional es igual a $\mu = (2+4+6)/3 = 4$. Ver la tabla en la siguiente página.

Notese las interesantes relaciones siguientes contenidas en la tabla:

La media de la colección de medias muestrales es 4, la media de la población de la que se extraen las muestras. Si μx denota la media de todas las medias muestrales entonces tenemos:
$\mu x = (3+4+3+4+5+5+2+4+6)/9 = 4$

La suma de los errores muestrales es cero.
$e_1 + e_2 + e_3 + . . . + e_9 = (-2) + (-1) + 0 + (-1) + 0 + 1 + 0 + 1 + 2 = 0$

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

En consecuencia, si x se usa para medir, estimar, la media poblacional μ, el promedio de todos los errores muestrales es cero.

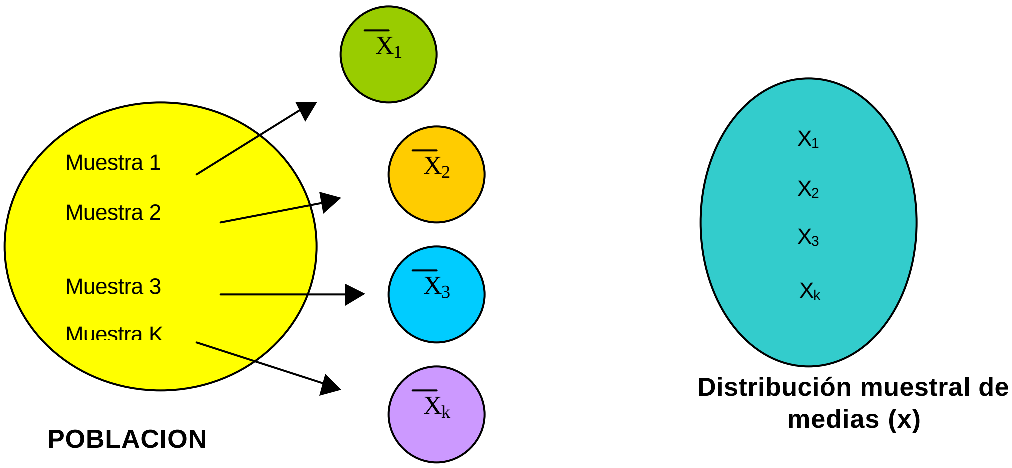
Distribuciones Muestrales

Las muestras aleatorias obtenidas de una población son, por naturaleza propia, impredecibles. No se esperaría que dos muestras aleatorias del mismo tamaño y tomadas de la misma población tenga la misma media muestral o que sean completamente parecidas; puede esperarse que cualquier estadístico, como la media muestral, calculado a partir de las medias en una muestra aleatoria, cambie su valor de una muestra a otra, por ello, se quiere estudiar la distribución de todos los valores posibles de un estadístico. Tales distribuciones serán muy importantes en el estudio de la estadística inferencial, porque las inferencias sobre las poblaciones se harán usando estadísticas muestrales. Como el análisis de las distribuciones asociadas con los estadísticos muestrales, podremos juzgar la confiabilidad de un estadístico muestral como un instrumento para hacer inferencias sobre un parámetro poblacional desconocido.

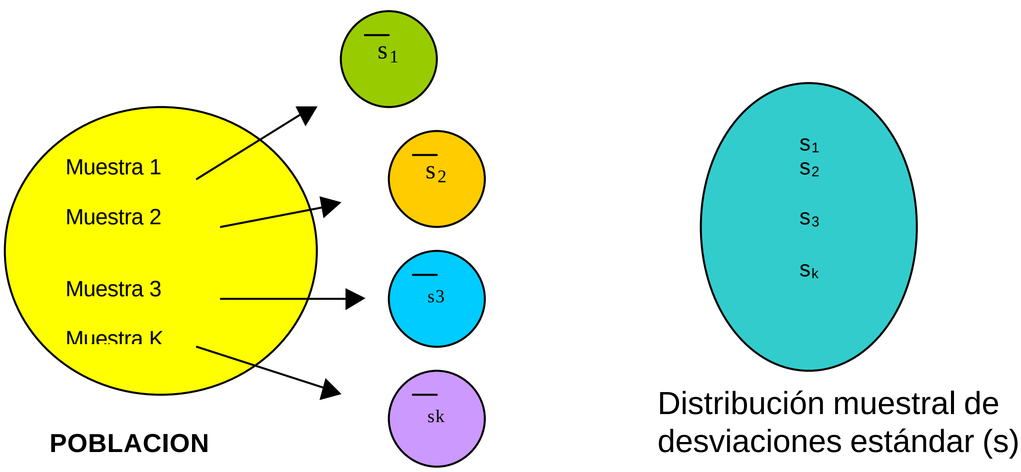
Como los valores de un estadístico, tal como x, varían de una muestra aleatoria a otra, se le puede considerar como una *variable aleatoria* con su correspondiente distribución de frecuencias.

La distribución de frecuencia de un estadístico muestral se denomina ***distribución muestral***. En general, **la distribución muestral de un estadístico** es la de todos sus valores posibles calculados a partir de muestras del mismo tamaño.

Suponga que se han seleccionado muestras aleatorias de tamaño 20 en una población grande. Se calcula la madia muestral x para cada muestra; la colección de todas estas medias muestrales recibe el nombre de **distribución muestral de medias,** lo que se puede ilustrar en la siguiente figura:



Suponga que se eligen muestras aleatorias de tamaño 20, de una población grande, y se calcula la deviación estándar de cada una. La colección de todas estas desviaciones estándar muestrales se llama **distribución muestral de la desviación estándar**, y lo podemos ver en la siguiente figura:



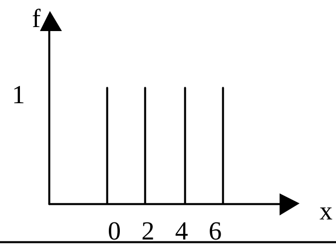
Ejemplo 1.6

Se eligen muestras ordenadas de tamaño 2, con reemplazo, de la población de valores 0, 2, 4 y 6. Encuentre:
μ, la media poblaciona.
σ, la desviación estándar poblacional.
$\mu_{x,}$ la media de la distribución muestral de medias.
σ_x, la desviación estándar de la distribución muestral de medias.
Además, grafique las frecuencias para la población y para la distribución muestral de medias.

Solución:

a) La media poblacional es:

$$m = \frac{0+2+4+6}{4} = 3$$



Gráfica de frecuencias para la población

b) La desviación estándar de la población es:

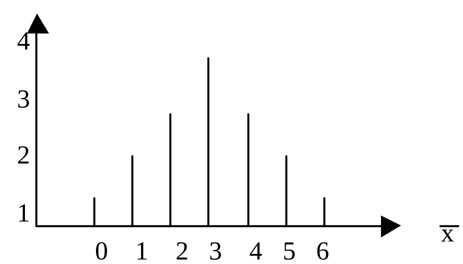
$$s = \sqrt{\frac{(0-3)^2 + (2-3)^2 + (4-3)^2 + (6-3)^2}{4}} = 2.236$$

c) A continuación se listan los elementos de la distribución muestral de la media y la correspondiente distribución de frecuencias.

Muestra	$\bar{x}$
(0,0)	0
(0,2)	1
(0,4)	2
(0,6)	3
(2,0)	1
(2,2)	2
(2,4)	3
(2,6)	4
(4,0)	2
(4,2)	3
(4,4)	4
(4,6)	5
(6,0)	3
(6,2)	4
(6,4)	5
(6,6)	6

Distribución de frecuencias de x

$\bar{x}$	f
0	1
1	2
2	3
3	4
4	3
5	2
6	1



Gráfica de frecuencias para las medias de las muestras

La media de la distribución muestral de medias es:

$$m_{\bar{x}} = \frac{\Sigma(f\bar{x})}{\Sigma f} = \frac{(0)(1)+(1)(2)+(2)(3)+(3)(4)+(4)(3)+(5)(2)+(6)(1)}{16} = \frac{48}{3} = 3$$

d) La desviación estándar de la distribución muestral de medias es:

$$s_{\bar{x}} = \sqrt{\frac{\Sigma(\bar{x}-m_{\bar{x}})^2 f}{\Sigma f}}$$

$$s_{\bar{x}} = \sqrt{\frac{(0-3)^2 1+(1-3)^2 2+(2-3)^2 3+(3-3)^2 4+(4-3)^2 3+(5-3)^2 2+(6-3)^2 1}{16}} = 1.58$$

De aquí que podamos deducir que: $s_x = \frac{s}{\sqrt{n}} = \frac{2.236}{2} = 1.58$

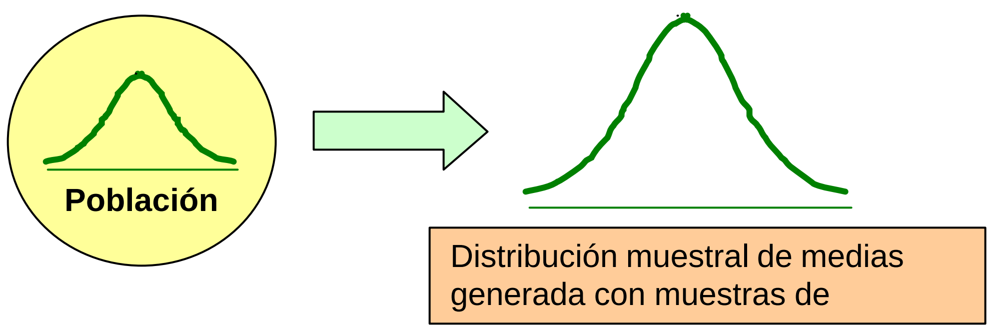
Como para cualquier variable aleatoria, la dsitribución muestral de medias tiene una media o valor esperado, una varianza y una desviación estándar, se puede demostrar que la distribución muestral de medias tiene una media igual a la media poblacional. Esto es:

$$m_{\bar{x}} = E(\bar{x}) = m = 3$$

Distribuciones muestrales

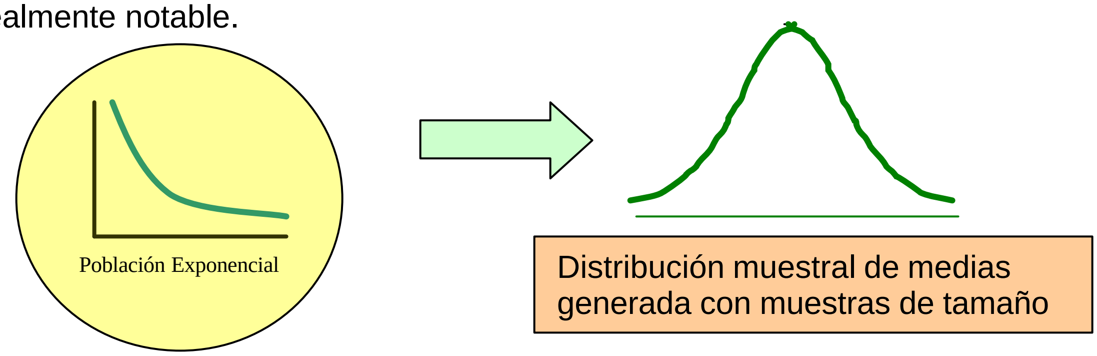
Después de haber realizado el ejercicio anterior se puede ver que una distribución muestral se genera extrayendo todas las posibles muestras del mismo tamaño de la población y calculándoles a éstas su estadístico.

Si la población de la que se extraen las muestras es normal, la distribución muestral de medias será normal sin importar el tamaño de la muestra.



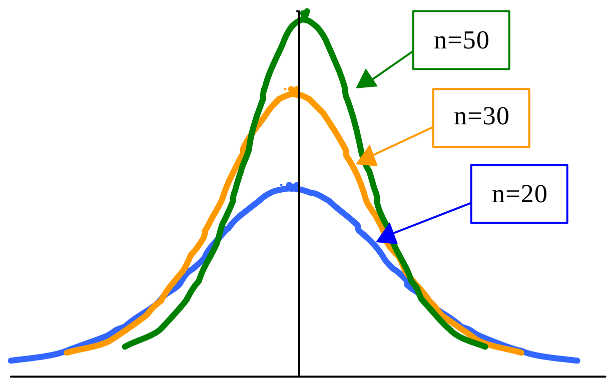
Si la población de donde se extraen las muestras no es normal, entonces el tamaño de la muestra debe ser mayor o igual a 30, para que la distribución muestral tenga una forma acampanada. Mientras mayor sea el tamaño de la muestra, más cerca estará la distribución muestral de ser normal.

Para muchos propósitos, la aproximación normal se considera buena si se cumple n=30. La forma de la disitribución muestral de medias sea aproximadamente normal, aún en casos donde la población original es bimodal, es realmente notable.



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 $s/\sqrt{n}$. La aproximación será cada vez más exacta a medida de que *n* sea cada vez mayor.



Ejemplo

Para la dsitribución 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	$\bar{x}$	Error muestral, $e=\bar{x}-\mu$
(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:

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

c) La desviación estándar de la distribución de los errores muestrales $\sigma_{e,}$ es entonces:

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

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: $s_x = s_e$

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

$$s_x = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

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 rfegla 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≥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 antiguedad en años en el trabajo de tres maestros universitarios de matemáticas:

Maestro de matemáticas	Antiguedad
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:

$$m = \frac{2+4+6}{3} = 4$$

La media de la distribución muestral es:

$$m_{\bar{x}} = \frac{5+4+3}{3} = 4$$

La desviación estándar de la población es:

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

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

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

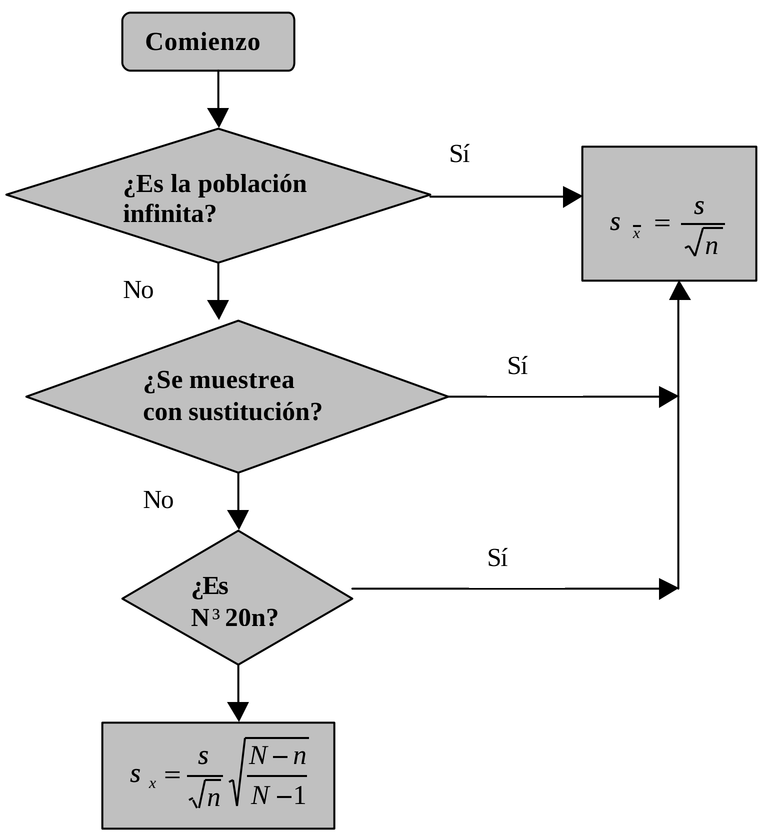
Si utilizamos la fórmula del error estándar sin el factor de correción tendriamos que:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{1.63}{\sqrt{2}} = 1.152$$

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

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

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:

$$z = \frac{x - m}{s}$$

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.

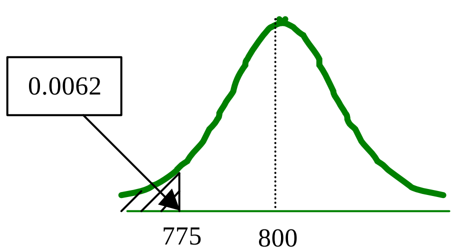
Sabemos que cuando se extraen muestras de tamaño mayor a 30 o bien de cualquier tamaño de una población normal, la distribución muestral de medias tiene un comportamiento aproximadamente normal, por lo que se puede utilizar la formula de la distribución normal con $m = m_{\bar{x}}$ y $s = s_{\bar{x}}$, entonces la fórmula para calcular la probabilidad del comportamiento del estadístico, en este caso la media de la muestra , quedaría de la siguiente manera:

$$z = \frac{\bar{x} - m}{s/\sqrt{n}}$$

y para poblaciones finitas y muestro con reemplazo:

$$z = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

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.

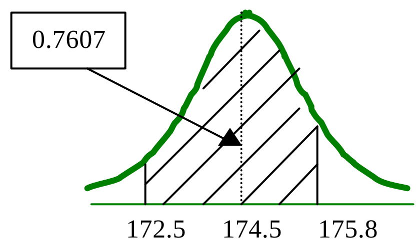
Solución:

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.

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

a)

$$z = \frac{\bar{x}-m}{\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} = \frac{172.5-174.5}{1.36} = -1.47$$



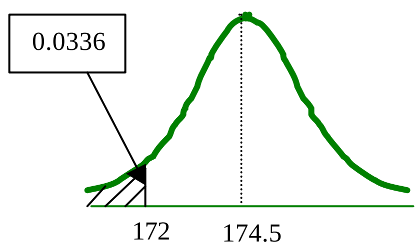
$$z = \frac{175.8-174.5}{1.36} = 0.96$$

$p(172.5 \leq \bar{x} \leq 175.8) = 0.7607$ (0.7607)(200)=152 medias muestrales

b) $z = \frac{172-174.5}{1.36} = -1.83$

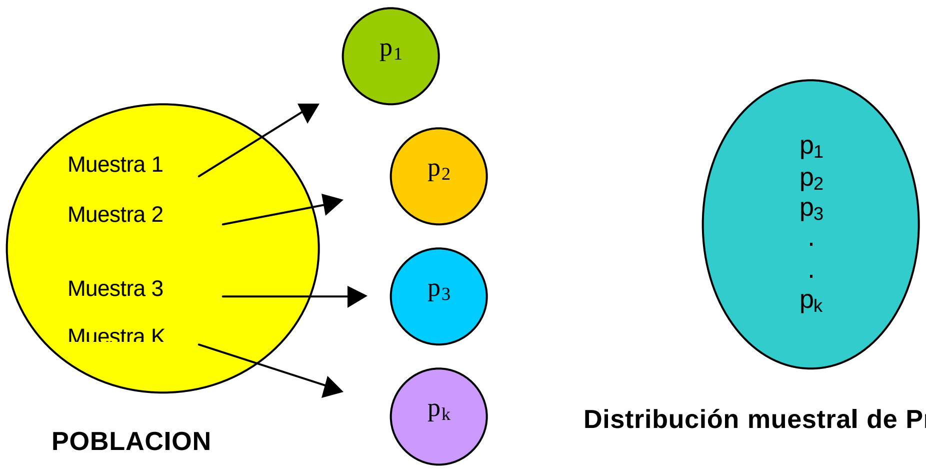
$p(\bar{x} \leq 172) = 0.00336$

(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.



POBLACION **Distribución muestral de Proporciones**

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 muestral de proporciones se tendría que hacer la sumatoria de la frecuencia por el valor de la proporción muestral y dividirla entre el número total de muestras. Esto es:

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

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

$$\mathbf{m_p = P}$$

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

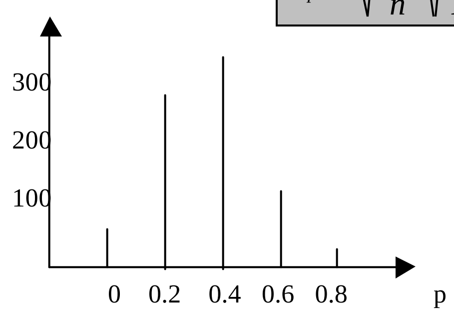
$$s_p = \sqrt{\frac{(0.8-1/3)^2*8+(0.6-1/3)^2*112+(0.4-1/3)^2*336+(0.2-1/3)^2*280+(0-1/3)^2*56}{792}} = 0.1681$$

La varianza de la distribución binomial es σ^2= npq, por lo que la varianza de la distribución muestral de proporciones es σ^2_p =(Pq)/n. Si se sustituten los valores en esta fórmula tenemos que:

$s_p = \sqrt{\frac{(1/3)(2/3)}{5}} = 0.2108$, 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:

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

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



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.

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

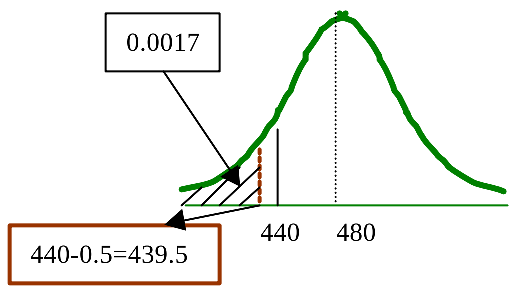
A esta fórmula se le puede agregar el factor de corrección de $\sqrt{\dfrac{N-n}{N-1}}$ si se cumple con las condiciones necesarias.

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

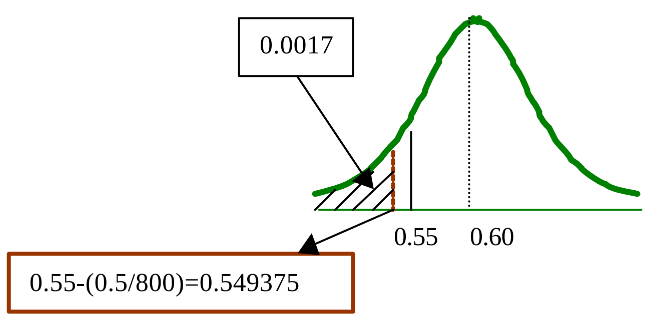


$$z=\frac{x-np}{\sqrt{npq}}=\frac{439.5-480}{\sqrt{800(0.60)(0.40)}}=-2.92$$

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) = ?



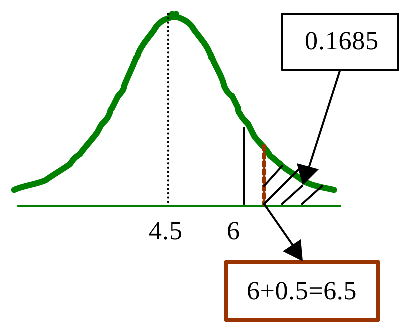
$$z=\frac{p-P}{\sqrt{\frac{Pq}{n}}}=\frac{0.549375-0.60}{\sqrt{\frac{(0.60)(0.40)}{800}}}=-2.92$$ Observe que este valor es igual al obtenido en el método de la aproximación de la distribución normal a la binomial, por lo que si lo buscamos en la tabla de “z” nos da la misma probabilidad de 0.0017.

También se debe de tomar en cuenta que el factor de corrección de 0.5 se esta 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

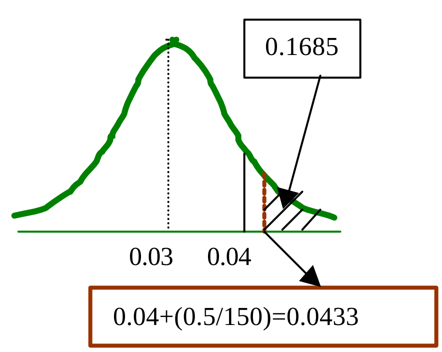


$$z = \frac{x - np}{\sqrt{npq}} = \frac{6.5 - 4.5}{\sqrt{150(0.03)(0.97)}} = 0.96$$

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) = ?



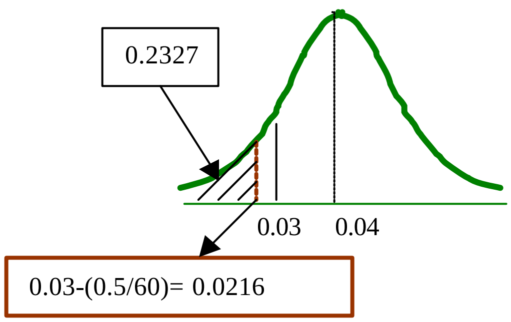
$$z=\frac{p-P}{\sqrt{\frac{Pq}{n}}}=\frac{0.0433-0.03}{\sqrt{\frac{(0.03)(0.97)}{150}}}=0.96$$ Observe que este valor es igual al obtenido y la interpretación es: existe una probabilidad del 17% de que al tomar una muestra de 150 personas se tenga una proporción mayor de 0.04 presentando una reacción adversa.

Ejemplo:
Se sabe que la verdadera proporción de los componentes defectuosos fabricadas 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.

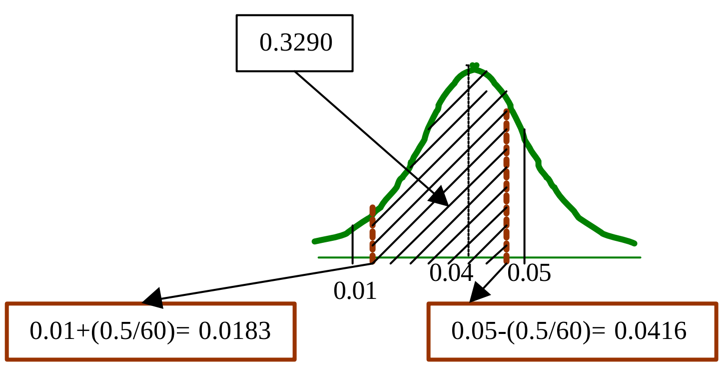
Solución:

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.

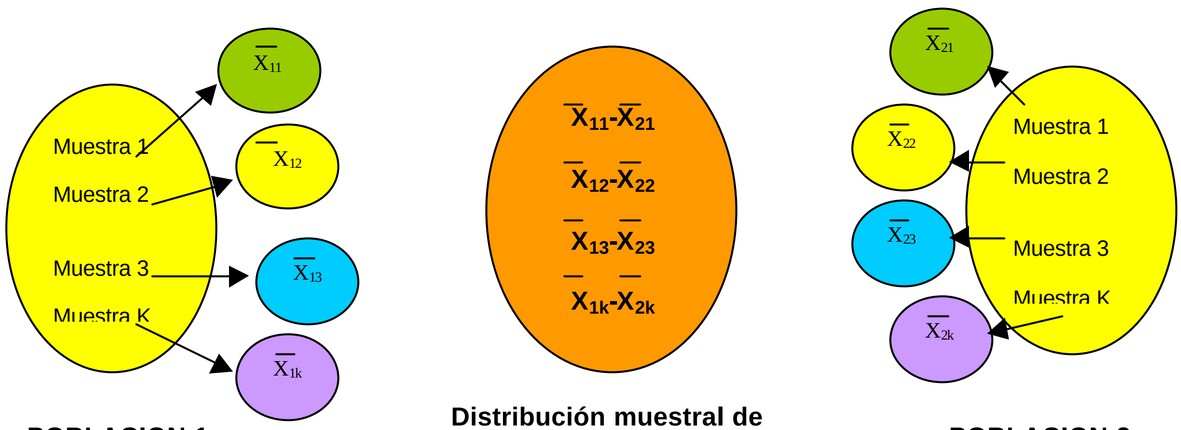
b) Datos:
n= 60 artículos
P=0.04
p= 0.01 y 0.05
p(0.01<p<0.05) = ?



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

Distribución Muestral de Diferencia de Medias

Suponga que se tienen dos poblaciones distintas, la primera con media μ_1 y desviación estándar σ_1, y la segunda con media μ_2 y desviación estándar σ_2. Más aún, se elige una muestra aleatoria de tamaño n_1 de la primera población y una muestra independiente aleatoria de tamaño n_2 de la segunda población; se calcula la media muestral para cada muestra y la diferencia entre dichas medias. La colección de todas esas diferencias se llama **distribución muestral de las diferencias entre medias** o la **distribución muestral del estadístico** $\bar{x}_1 - \bar{x}_2$.



POBLACION 1 **Distribución muestral de Diferencia de Medias** **POBLACION 2**

La distribución es aproximadamente normal para $n_1 \geq 30$ y $n_2 \geq 30$. Si las poblaciones son normales, entonces la distribución muestral de medias es normal sin importar los tamaños de las muestras.

En ejercicios anteriores se había demostrado que $m = m_{\bar{x}}$ y que $s_{\bar{x}} = s/\sqrt{n}$, por lo que no es difícil deducir que $\boxed{m_{\bar{x}1} - m_{\bar{x}2} = m_1 - m_2}$ y que $\boxed{s_{\bar{x}1-\bar{x}2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$.

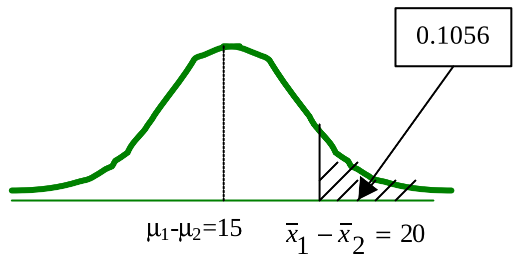
La fórmula que se utilizará para el calculo de probabilidad del estadístico de diferencia de medias es:

$$\boxed{z = \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}}$$

Ejemplo:

En un estudio para comparar los pesos promedio de niños y niñas de sexto grado en una escuela primaria se usará una muestra aleatoria de 20 niños y otra de 25 niñas. Se sabe que tanto para niños como para niñas los pesos siguen una distribución normal. El promedio de los pesos de todos los niños de sexto grado de esa escuela es de 100 libras y su desviación estándar es de 14.142, mientras que el promedio de los pesos de todas las niñas del sexto grado de esa escuela es de 85 libras y su desviación estándar es de 12.247 libras. Si $\bar{x}_1$ representa el promedio de los pesos de 20 niños y $\bar{x}_2$ es el promedio de los pesos de una muestra de 25 niñas, encuentre la probabilidad de que el promedio de los pesos de los 20 niños sea al menos 20 libras más grande que el de las 25 niñas.

Solución:
Datos:
μ_1 = 100 libras
μ_2 = 85 libras
σ_1 = 14.142 libras
σ_2 = 12.247 libras
n_1 = 20 niños
n_2 = 25 niñas



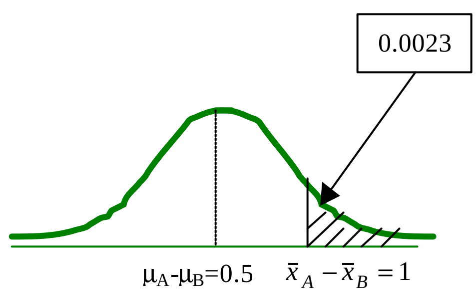
$p(\bar{x}_1 - \bar{x}_2 \rangle 20)$ = ?

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}}} = \frac{20 - (100 - 85)}{\sqrt{\frac{(14.142)^2}{20} + \frac{(12.247)^2}{25}}} = 1.25$$

Por lo tanto, la probabilidad de que el promedio de los pesos de la muestra de niños sea al menos 20 libras más grande que el de la muestra de las niñas es 0.1056.

Ejemplo:
Uno de los principales fabricantes de televisores compra los tubos de rayos catódicos a dos compañías. Los tubos de la compañía A tienen una vida media de 7.2 años con una desviación estándar de 0.8 años, mientras que los de la B tienen una vida media de 6.7 años con una desviación estándar de 0.7. Determine la probabilidad de que una muestra aleatoria de 34 tubos de la compañía A tenga una vida promedio de al menos un año más que la de una muestra aleatoria de 40 tubos de la compañía B.

Solución:
Datos:
μ_A = 7.2 años
μ_B = 6.7 años
σ_A = 0.8 años
σ_B = 0.7 años
n_A = 34 tubos
n_B = 40 tubos



$p(\bar{x}_A - \bar{x}_B \rangle 1)$ = ?

$$z = \frac{(\bar{x}_A - \bar{x}_B) - (m_A - m_B)}{\sqrt{\frac{s_A^{\,2}}{n_A} + \frac{s_B^{\,2}}{n_B}}} = \frac{1 - (7.2 - 6.7)}{\sqrt{\frac{(0.8)^2}{34} + \frac{(0.7)^2}{40}}} = 2.84$$

Ejemplo:

Se prueba el rendimiento en km/L de 2 tipos de gasolina, encontrándose una desviación estándar de 1.23km/L para la primera gasolina y una desviación estándar de 1.37km/L para la segunda gasolina; se prueba la primera gasolina en 35 autos y la segunda en 42 autos.

a) ¿Cuál es la probabilidad de que la primera gasolina de un rendimiento promedio mayor de 0.45km/L que la segunda gasolina?

b) ¿Cuál es la probabilidad de que la diferencia en rendimientos promedio se encuentre entre 0.65 y 0.83km/L a favor de la gasolina 1?.

Solución:

En este ejercicio no se cuenta con los parámetros de las medias en ninguna de las dos poblaciones, por lo que se supondrán que son iguales.

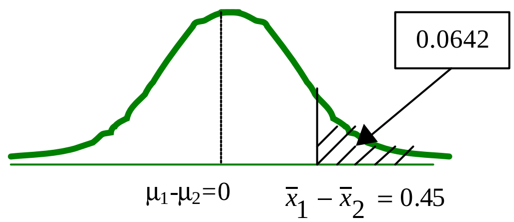
Datos:

$\sigma_1 = 1.23$ Km/Lto

$\sigma_2 = 1.37$ Km/Lto

$n_1 = 35$ autos

$n_2 = 42$ autos



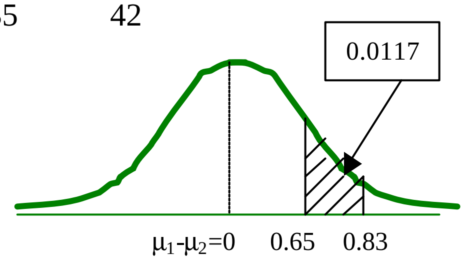
a) $p(\bar{x}_1 - \bar{x}_2 \rangle 0.45)$ = ?

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{0.45 - 0}{\sqrt{\frac{(1.23)^2}{35} + \frac{(1.37)^2}{42}}} = 1.52$$

b) $p(0.65 \leq \bar{x}_1 - \bar{x}_2 \leq 0.83) = ?$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{0.65 - 0}{\sqrt{\frac{(1.23)^2}{35} + \frac{(1.37)^2}{42}}} = 2.19$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{0.83 - 0}{\sqrt{\frac{(1.23)^2}{35} + \frac{(1.37)^2}{42}}} = 2.80$$



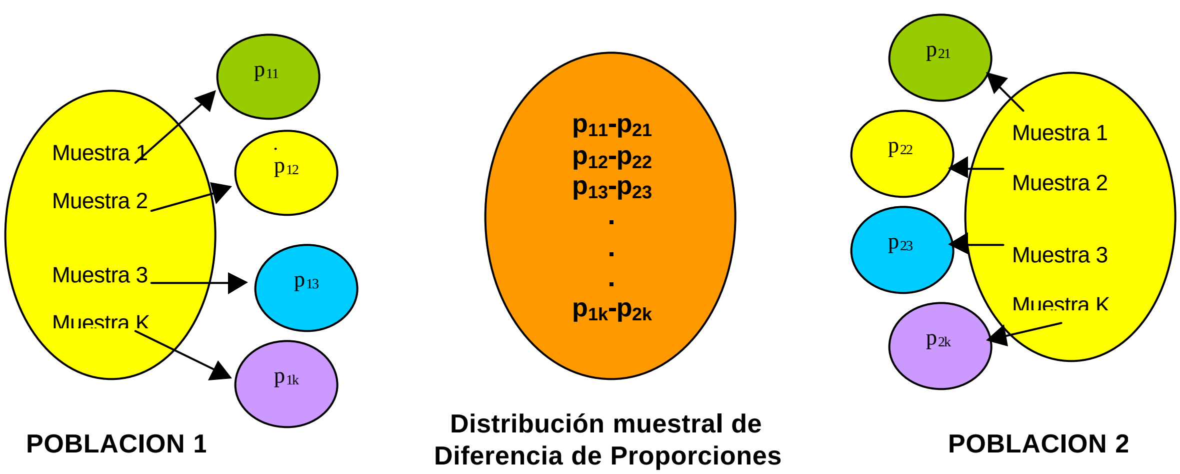
La probabilidad de que la diferencia en rendimientos promedio en las muestras se encuentre entre 0.65 y 0.83 Km/Lto a favor de la gasolina 1 es de 0.0117.

Distribución Muestral de Diferencia de Proporciones

Muchas aplicaciones involucran poblaciones de datos cualitativos que deben compararse utilizando proporciones o porcentajes. A continuación se citan algunos ejemplos:

- Educación.- ¿Es mayor la proporción de los estudiantes que aprueban matemáticas que las de los que aprueban inglés?
- Medicina.- ¿Es menor el porcentaje de los usuarios del medicamento A que presentan una reacción adversa que el de los usuarios del fármaco B que también presentan una reacción de ese tipo?
- Administración.- ¿Hay diferencia entre los porcentajes de hombres y mujeres en posiciones gerenciales.
- Ingeniería.- ¿Existe diferencia entre la proporción de artículos defectuosos que genera la máquina A a los que genera la máquina B?

Cuando el muestreo procede de dos poblaciones binomiales y se trabaja con dos proporciones muestrales, la distribución muestral de diferencia de proporciones es aproximadamente normal para tamaños de muestra grande ($n_1p_1 \geq 5$, $n_1q_1 \geq 5$, $n_2p_2 \geq 5$ y $n_2q_2 \geq 5$). Entonces p_1 y p_2 tienen distribuciones muestrales aproximadamente normales, así que su diferencia p_1-p_2 también tiene una distribución muestral aproximadamente normal.



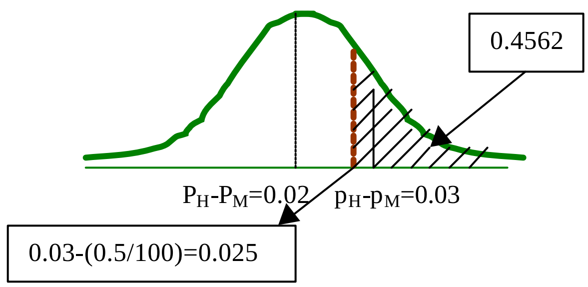
Cuando se estudió a la distribución muestral de proporciones se comprobó que $P = m_p$ y que $s_p = \sqrt{\frac{Pq}{n}}$, por lo que no es difícil deducir que $\boxed{m_{p1} - m_{p2} = P_1 - P_2}$ y que $\boxed{s_{p1-p2} = \sqrt{\frac{P_1q_1}{n_1} + \frac{P_2q_2}{n_2}}}$.

La fórmula que se utilizará para el calculo de probabilidad del estadístico de diferencia de proporciones es:

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

Ejemplo:
Los hombres y mujeres adultos radicados en una ciudad grande del norte difieren en sus opiniones sobre la promulgación de la pena de muerte para personas culpables de asesinato. Se cree que el 12% de los hombres adultos están a favor de la pena de muerte, mientras que sólo 10% de las mujeres adultas lo están. Si se pregunta a dos muestras aleatorias de 100 hombres y 100 mujeres su opinión sobre la promulgación de la pena de muerte, determine la probabilidad de que el porcentaje de hombres a favor sea al menos 3% mayor que el de las mujeres.

Solución:
Datos:
$P_H = 0.12$
$P_M = 0.10$
$n_H = 100$
$n_M = 100$
$p(p_H - p_M \geq 0.03) = ?$



Se recuerda que se está incluyendo el factor de corrección de 0.5 por ser una distribución binomial y se está utilizando la distribución normal.

$$z = \frac{(p_H - p_M) - (P_H - P_M)}{\sqrt{\frac{P_H q_H}{n_H} + \frac{P_M q_M}{n_M}}} = \frac{0.025 - (0.12 - 0.10)}{\sqrt{\frac{(0.12)(0.88)}{100} + \frac{(0.10)(0.90)}{100}}} = 0.11$$

Se concluye que la probabilidad de que el porcentaje de hombres a favor de la pena de muerte, al menos 3% mayor que el de mujeres es de 0.4562.

Ejemplo:
Una encuesta del Boston College constó de 320 trabajadores de Michigan que fueron despedidos entre 1979 y 1984, encontró que 20% habían estado sin trabajo durante por lo menos dos años. Supóngase que tuviera que seleccionar otra muestra aleatoria de 320 trabajadores de entre todos los empleados despedidos entre 1979 y 1984. ¿Cuál sería la probabilidad de que su porcentaje muestral de trabajadores sin empleo durante por lo menos dos años, difiera del porcentaje obtenido en la encuesta de Boston College, en 5% o más?

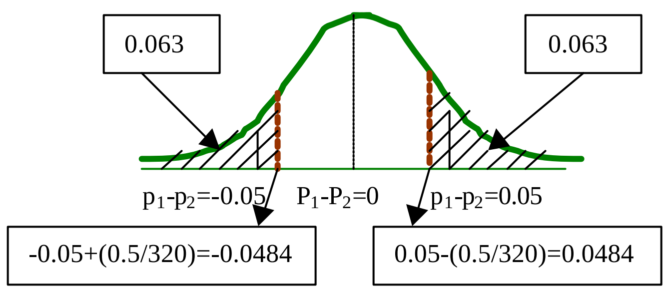
Solución:
En este ejercicio se cuenta únicamente con una población, de la cual se están extrayendo dos muestras y se quiere saber la probabilidad de la diferencia de los

porcentajes en esas dos muestras, por lo que se debe de utilizar la distribución muestral de proporciones con $P_1 = P_2$, ya que es una misma población.
Otra de las situaciones con la cual nos topamos es que desconocemos la proporción de trabajadores despedidos entre 1979 y 1984 que estuvieron desempleados por un período de por lo menos dos años, sólo se conoce la $p_1 = 0.20$ ya que al tomar una muestra de 320 trabajadores se observó esa proporción.

En la fórmula de la distribución muestral de proporciones para el cálculo de probabilidad se necesita saber las proporciones de las poblaciones, las cuales en este ejercicio las desconocemos, por lo que se utilizará el valor de 0.20 como una estimación puntual de P. En el siguiente tema se abordará el tema de estimación estadística y se comprenderá el porque estamos utilizando de esa manera el dato.

También debe de comprenderse la pregunta que nos hace este problema, ¿cuál sería la probabilidad de que su porcentaje muestral de trabajadores sin empleo durante por lo menos dos años, ***difiera*** del porcentaje obtenido en la encuesta de Boston College, en 5% o más?, la palabra ***difiera*** quiere decir que puede existir una diferencia a favor de la muestra uno, o a favor de la muestra dos, por lo que se tendrán que calcular dos áreas en la distribución y al final sumarlas.

Datos:
$p_1 = 0.20$
$n_1 = 320$ trabajadores
$n_2 = 320$ trabajadores
$P_1 = P_2$



$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} = \frac{0.0484 - 0}{\sqrt{\frac{(0.20)(0.80)}{320} + \frac{(0.20)(0.80)}{320}}} = 1.53$$

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} = \frac{-0.0484 - 0}{\sqrt{\frac{(0.20)(0.80)}{320} + \frac{(0.20)(0.80)}{320}}} = -1.53$$

$$p(-0.05 \geq p_1 - p_2 \geq 0.05) = 0.063 + 0.063 = 0.1260$$

La probabilidad de que su proporcion muestral de trabajadores sin empleo durante por lo menos dos años, ***difiera*** del porcentaje obtenido en la encuesta de Boston College, en 0.05 o más es de 0.1260.

Ejemplo:

Se sabe que 3 de cada 6 productos fabricados por la máquina 1 son defectuosos y que 2 de cada 5 objetos fabricados por la máquina 2 son defectuosos; se toman muestras de 120 objetos de cada máquina:

a) ¿cuál es la probabilidad de que la proporción de artículos defectuosos de la máquina 2 rebase a la máquina 1 en por lo menos 0.10?

b) ¿cuál es la probabilidad de que la proporción de artículos defectuosos de la máquina 1 rebase a la máquina 2 en por lo menos 0.15?

Solución:

Datos:

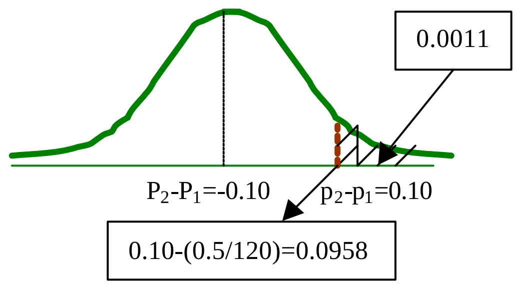
$P_1 = 3/6 = 0.5$

$P_2 = 2/5 = 0.4$

$n_1 = 120$ objetos

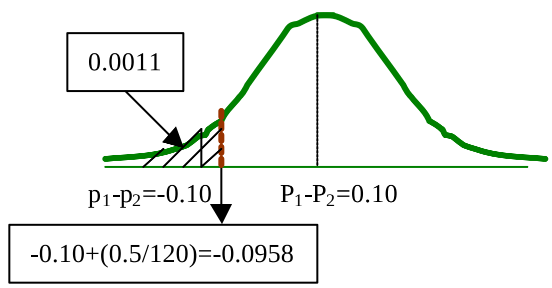
$n_2 = 120$ objetos

a) $p(p_2 - p_1 \geq 0.10) = ?$



$$z = \frac{(p_2 - p_1) - (P_2 - P_1)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} = \frac{0.0958 - (-0.10)}{\sqrt{\frac{(0.50)(0.50)}{120} + \frac{(0.40)(0.60)}{120}}} = 3.06$$

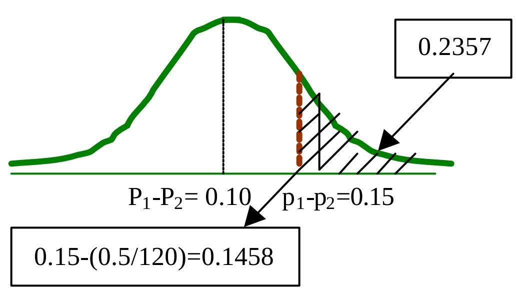
Otra manera de hacer este ejercicio es poner $P_1 - P_2$:



$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} = \frac{-0.0958 - 0.10}{\sqrt{\frac{(0.50)(0.50)}{120} + \frac{(0.40)(0.60)}{120}}} = -3.06$$

La probabilidad de que exista una diferencia de proporciones de artículos defectuosos de por lo menos 10% a favor de la máquina 2 es de 0.0011.

b) $p(p_1 - p_2 \geq 0.15) = ?$



$$z=\frac{(p_1-p_2)-(P_1-P_2)}{\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}}}=\frac{0.1458-0.10}{\sqrt{\frac{(0.50)(0.50)}{120}+\frac{(0.40)(0.60)}{120}}}=0.72$$

La probabilidad de que exista una diferencia de proporciones de artículos defectuosos de por lo menos 15% a favor de la máquina 1 es de 0.2357.

Distribución Muestral de Número de Defectos

En el control de calidad y específicamente en los gráficos de control "c" se aplica esta distribución, la cual consiste en que al extraer un artículo contabilicemos el número de defectos que tiene ese artículo.

Esta distribución muestral proviene de la distribución de Poisson, en la cual le media es λ y que en este caso es el número promedio de defectos por unidad. Como ya es conocido la varianza de la distribución de Poisson es igual a λ por lo que se puede deducir la formula de la siguiente manera:

$$z=\frac{x-l}{\sqrt{l}}$$

Para la distribución muestral de número de defectos la nomenclatura utilizada es:

c = número defectos por unidad de inspección

C = número de defectos promedio por unidad de inspección

Se debe de recordar que la distribución de Poisson es una distribución discreta, y se esta utilizando la aproximación de la normal a la Poisson, debiendo aplicar el factor de corrección de ± 0.5 según sea el caso. La formula para la dsitribución muestral de número de defectos quedaría de la siguiente manera:

$$z=\frac{c-C}{\sqrt{C}}$$

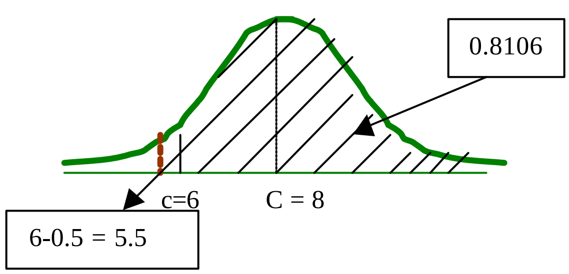
Ejemplo:

En cierta empresa se fabrican productos con un promedio de 8 defectos por unidad. Determine la probabilidad de que el próximo producto inspeccionado tenga un número de defectos:

a) Mayor o igual a 6
b) Exactamente 7
c) Como máximo 9

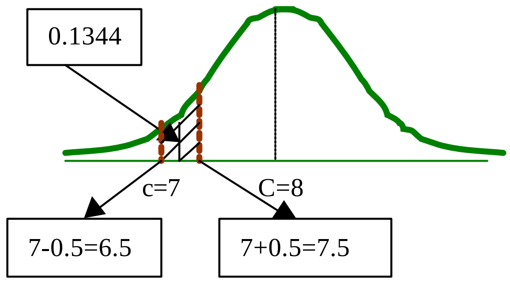
a)

$$z=\frac{c-C}{\sqrt{C}}=\frac{5.5-8}{\sqrt{8}}=0.88$$



La probabilidad de que el siguiente producto inspeccionado tenga por lo menos 6 defectos es de 0.8106.

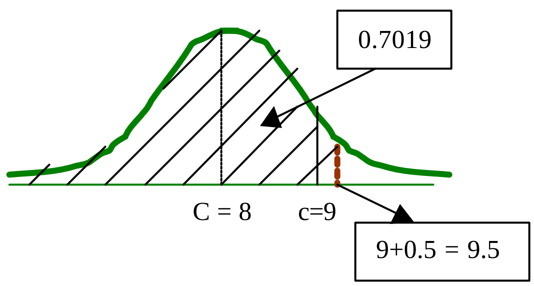
b) $z=\dfrac{c-C}{\sqrt{C}}=\dfrac{6.5-8}{\sqrt{8}}=0.53$ $\qquad z=\dfrac{7.5-8}{\sqrt{8}}0.17$



La probabilidad de que el siguiente producto inspeccionado tenga exactamente 7 defectos es de 0.1344.

c)

$z=\dfrac{c-C}{\sqrt{C}}=\dfrac{9.5-8}{\sqrt{8}}=0.53$



La probabilidad de que el siguiente producto inspeccionado tenga a lo más 9 defectos es de 0.7019.

Problemas propuestos

1. Se sabe que la resistencia a la ruptura de cierto tipo de cuerda se distribuye normalmente con media de 2000 libras y una varianza de 25,000 lbs^2. Si se selecciona una muestra aleatoria de 100 cuerdas; determine la probabilidad de que en esa muestra:
   a) La resistencia media encontrada sea de por lo menos 1958 libras.
   b) La resistencia media se mayor de 2080 libras.

2. Como parte de un proyecto general de mejoramiento de la calidad, un fabricante textil decide controlar el número de imperfecciones encontradas en cada pieza de tela. Se estima que el número promedio de imperfecciones por cada pieza de tela es de 12, determine la probabilidad de que en la próxima pieza de tela fabricada se encuentren:
   a) Entre 10 y 12 imperfecciones.
   b) Menos de 9 y más de 15 imperfecciones.

3. En una prueba de aptitud la puntuación media de los estudiantes es de 72 puntos y la desviación estándar es de 8 puntos. ¿Cuál es la probabilidad de que dos grupos de estudiantes, formados de 28 y 36 estudiantes, respectivamente, difieran en su puntuación media en:
   a) 3 ó más puntos.

b) 6 o más puntos.
c) Entre 2 y 5 puntos.

4. Un especialista en genética ha detectado que el 26% de los hombres y el 24% de las mujeres de cierta región del país tiene un leve desorden sanguíneo; si se toman muestras de 150 hombres y 150 mujeres, determine la probabilidad de que la diferencia muestral de proporciones que tienen ese leve desorden sanguíneo sea de:
   a) Menos de 0.035 a favor de los hombres.
   b) Entre 0.01 y 0.04 a favor de los hombres.

5. Una urna contiene 80 bolas de las que 60% son rojas y 40% blancas. De un total de 50 muestras de 20 bolas cada una, sacadas de la urna con remplazamiento, ¿en cuántas cabe esperar
   a) Igual número de bolas rojas y blancas?
   b) 12 bolas rojas y 8 blancas?
   c) 8 bolas rojas y 12 blancas?
   d) 10 ó mas bolas blancas?

6. Los pesos de 1500 cojinetes de bolas se distribuyen normalmente con media de 2.40 onzas y desviación estándar de 0.048 onzas. Si se extraen 300 muestras de tamaño 36 de esta población, determinar la media esperada y la desviación estándar de la distribución muestral de medias si el muestreo se hace:
   a) Con remplazamiento
   b) Sin remplazamiento

7. La vida media de una máquina para hacer pasta es de siete años, con una desviación estándar de un año. Suponga que las vidas de estas máquinas siguen aproximadamente una distribución normal, encuentre:
   a) La probabilidad de que la vida media de una muestra aleatoria de 9 de estas máquinas caiga entre 6.4 y 7.2 años.
   b) El valor de la $\bar{x}$ a la derecha del cual caería el 15% de las medias calculadas de muestras aleatorias de tamaño nueve.

8. Se llevan a cabo dos experimentos independientes en lo que se comparan dos tipos diferentes de pintura. Se pintan 18 especímenes con el tipo A y en cada uno se registra el tiempo de secado en horas. Lo mismo se hace con el tipo B. Se sabe que las desviaciones estándar de la población son ambas 1.0. Suponga que el tiempo medio de secado es igual para los dos tipo de pintura. Encuentre la probabilidad de que la diferencia de medias en el tiempo de secado sea mayor a uno a favor de la pintura A.

Respuestas a los problemas propuestos:

1. a) 0.9960	b) 0		
2. a) 0.3221	b) 0.3122		
3. a) 0.2150	b) 0.0064	c) 0.4504	
4. a) 0.2227	b) 0.2848		
5. a) 6	b) 9	c) 2	d) 12

6. a) $m_{\bar{x}} = 22.4, s_{\bar{x}} = 0.008$ b) $m_{\bar{x}} = 22.4, s_{\bar{x}} =$ ligeramente menor que 0.008

7. a) 0.6898 b) 7.35

8. 0.0013

ESTIMACION

El objetivo principal de la estadística inferencial es la **estimación**, esto es que mediante el estudio de una muestra de una población se quiere generalizar las conclusiones al total de la misma. Como vimos en la sección anterior, los estadísticos varían mucho dentro de sus distribuciones muestrales, y mientras menor sea el error estándar de un estadístico, más cercanos serán unos de otros sus valores.

Existen dos tipos de estimaciones para parámetros; puntuales y por intervalo. Una **estimación puntual** es un único valor estadístico y se usa para estimar un parámetro. El estadístico usado se denomina **estimador**.
Una **estimación por intervalo** es un rango, generalmente de ancho finito, que se espera que contenga el parámetro.

Estimación Puntual

La inferencia estadística está casi siempre concentrada en obtener algún tipo de conclusión acerca de uno o más parámetros (características poblacionales). Para hacerlo, se requiere que un investigador obtenga datos muestrales de cada una de las poblaciones en estudio. Entonces, las conclusiones pueden estar basadas en los valores calculados de varias cantidades muestrales . Po ejemplo, representamos con μ (parámetro) el verdadero promedio de resistencia a la ruptura de conexiones de alambres utilizados para unir obleas de semiconductores. Podría tomarse una muestra aleatoria de 10 conexiones para determinar la resistencia a la ruptura de cada una, y la media muestral de la resistencia a la ruptura $\bar{x}$ se podía emplear para sacar una conclusión acerca del valor de μ. De forma similar, si σ^2 es la varianza de la distribución de resistencia a la ruptura, el valor de la varianza muestral s^2 se podría utilizar pra inferir algo acerca de σ^2.

Cuando se analizan conceptos generales y métodos de inferencia es conveniente tener un símbolo genérico para el parámetro de interés. Se utilizará la letra griega θ para este propósito. *El objetivo de la estimación puntual es seleccionar sólo un número, basados en datos de la muestra, que represente el valor más razonable de* q.
Una muestra aleatoria de 3 baterías para calculadora podría presentar duraciones observadas en horas de x_1=5.0, x_2=6.4 y x_3=5.9. El valor calculado de la duración media muestral es $\bar{x}$ = 5.77, y es razonable considerar 5.77 como el valor más adecuado de μ.

Una **estimación puntual** de un parámetro θ es un sólo número que se puede considerar como el valor más razonable de θ. La estimación puntual se obtiene al seleccionar una estadística apropiada y calcular su valor a partir de datos de la muestra dada. La estadística seleccionada se llama **estimador puntual** de θ.

El símbolo $\hat{\theta}$ (theta sombrero) suele utilizarse para representar el estimador de θ y la estimación puntual resultante de una muestra dada. Entonces $\hat{\mu} = \bar{x}$ se lee como "el estimador puntual de μ es la media muestral $\bar{x}$". El enunciado "la estimación puntual de μ es 5.77" se puede escribir en forma abreviada $\hat{\mu} = 5.77$.

Ejemplo:

En el futuro habrá cada vez más interés en desarrollar aleaciones de Mg de bajo costo, para varios procesos de fundición. En consecuencia, es importante contar con métodos prácticos para determinar varias propiedades mecánicas de esas aleaciones. Examine la siguiente muestra de mediciones del módulo de elasticidad obtenidos de un proceso de fundición a presión:

44.2 43.9 44.7 44.2 44.0 43.8 44.6 43.1

Suponga que esas observaciones son el resultado de una muestra aleatoria. Se desea estimar la varianza poblacional σ^2. Un estimador natural es la varianza muestral:

$$\hat{\sigma}^2 = s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{(44.2 - 44.0625)^2 + (43.9 - 44.0625)^2 + \ldots + (43.1 - 44.0625)^2}{8-1} = 0.251$$

En el mejor de los casos, se encontrará un estimador $\hat{\theta}$ para el cual $\hat{\theta} = \theta$ siempre. Sin embargo, $\hat{\theta}$ es una función de las X_i muestrales, por lo que en sí misma una variable aleatoria.

$$\hat{\theta} = \theta + \text{error de estimación}$$

entonces el estimador preciso sería uno que produzca sólo pequeñas diferencias de estimación, de modo que los valores estimados se acerquen al valor verdadero.

Propiedades de un Buen Estimador

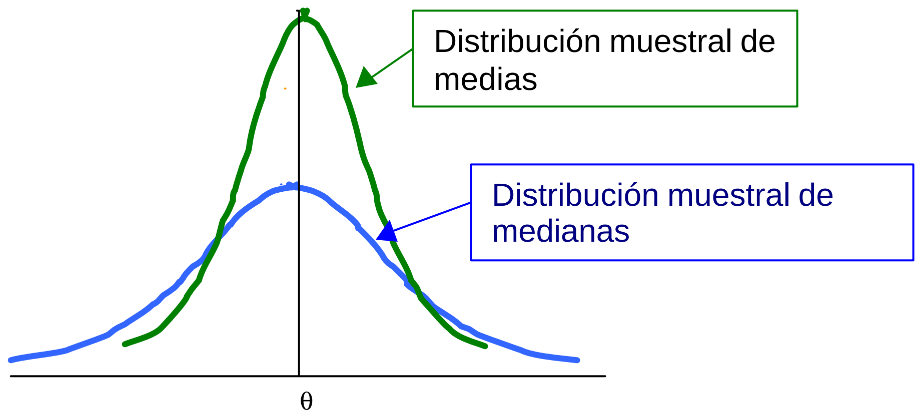
Insesgado.- Se dice que un estimador puntual $\hat{\theta}$ es un estimador insesgado de θ si $E(\hat{\theta}) = \theta$, para todo valor posible de θ. En otras palabras, un estimador insesgado es aquel para el cual la media de la distribución muestral es el parámetro estimado. Si se usa la media muestral $\bar{x}$ para estimar la media poblacional μ, se sabe que la $\mu_{\bar{x}} = \mu$, por lo tanto la media es un estimador insesgado.

Eficiente o con varianza mínima.- Suponga que $\hat{\theta}_1$ y $\hat{\theta}_2$ son dos estimadores insesgados de θ. Entonces, aun cuando la distribución de cada estimador esté centrada en el valor verdadero de θ, las dispersiones de las distribuciones alrededor del valor verdadero pueden ser diferentes.

Entre todos los estimadores de θ que son insesgados, seleccione al que tenga varianza mínima. El $\hat{\theta}$ resultante recibe el nombre de **estimador insesgado con varianza mínima** (MVUE, minimum variance unbiased estimator) de θ.

En otras palabras, la eficiencia se refiere al tamaño de error estándar de la estadística. Si comparamos dos estaíisticas de una muestra del mismo tamaño y tratamos de decidir cual de ellas es un estimador mas eficiente, escogeríamos la

estadística que tuviera el menor error estándar, o la menor desviación estándar de la distribución de muestreo.
Tiene sentido pensar que un estimador con un error estándar menor tendrá una mayor oportunidad de producir una estimación mas cercana al parámetro de población que se esta considerando.



Como se puede observar las dos distribuciones tienen un mismo valor en el parámetro sólo que la distribución muestral de medias tiene una menor varianza, por lo que la media se convierte en un estimador eficiente e insesgado.

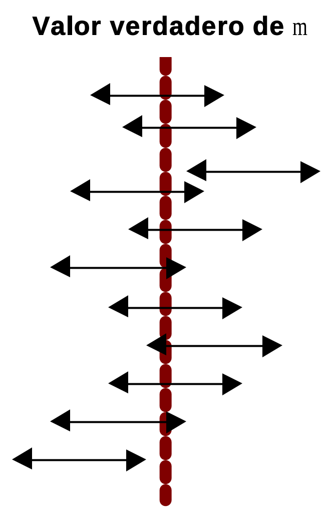
Coherencia.- Una estadística es un estimador coherente de un parámetro de población, si al aumentar el tamaño de la muestra se tiene casi la certeza de que el valor de la estadística se aproxima bastante al valor del parámetro de la población. Si un estimador es coherente se vuelve mas confiable si tenemos tamaños de muestras mas grandes.

Suficiencia.- Un estimador es suficiente si utiliza una cantidad de la información contenida de la muestra que ningún otro estimador podría extraer información adicional de la muestra sobre el parámetro de la población que se esta estimando.
Es decir se pretende que al extraer la muestra el estadístico calculado contenga toda la información de esa muestra. Por ejemplo, cuando se calcula la media de la muestra, se necesitan todos los datos. Cuando se calcula la mediana de una muestra sólo se utiliza a un dato o a dos. Esto es solo el dato o los datos del centro son los que van a representar la muestra. Con esto se deduce que si utilizamos a todos los datos de la muestra como es en el caso de la media, la varianza, desviación estándar, etc; se tendrá un estimador suficiente.

Estimación por Intervalos

Un estimado puntual, por ser un sólo número, no proporciona por sí mismo información alguna sobre la precisión y confiabilidad de la estimación. Por ejemplo, imagine que se usa el estadístico $\bar{x}$ para calcular un estimado puntual de la resistencia real a la ruptura de toallas de papel de cierta marca, y suponga que $\bar{x}$ = 9322.7. Debido a la variabilidad de la muestra, nunca se tendrá el caso de que $\bar{x}$ =μ. El estimado puntual nada dice sobre lo cercano que esta de μ. Una alternativa para reportar un solo valor del parámetro que se esté estimando es calcular e informar todo un intervalo de valores factibles, un *estimado de intervalo o intervalo de confianza (IC)*. Un ***intervalo de confianza*** se calcula siempre seleccionando primero un ***nivel de confianza***, que es una medida de el grado de fiabilidad en el intervalo. Un intervalo de confianza con un nivel de confianza de 95% de la resistencia real promedio a la ruptura podría tener un límite inferior de 9162.5 y uno superior de 9482.9. Entonces, en un nivel de confianza de 95%, es posible tener cualquier valor de μ entre 9162.5 y 9482.9. Un nivel de confianza de 95% implica que 95% de todas las muestras daría lugar a un intervalo que incluye μ o cualquier otro parámetro que se esté estimando, y sólo 5% de las muestras producirá un intervalo erróneo. Cuanto mayor sea el nivel de confianza podremos creer que el valor del parámetro que se estima está dentro del intervalo.

Una interpretación correcta de la "confianza de 95%" radica en la interpretación frecuente de probabilidad a largo plazo: decir que un evento A tiene una probabilidad de 0.95, es decir que si el experimento donde A está definido re realiza una y otra vez, a largo plazo A ocurrirá 95% de las veces. Para este caso el 95% de los intervalos de confianza calculados contendrán a μ.



Esta es una construcción repetida de intervalos de confianza de 95% y se puede observar que de los 11 intervalos calculados sólo el tercero y el último no contienen el valor de μ.

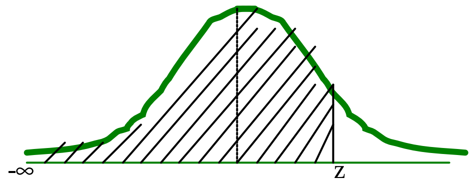
De acuerdo con esta interpretación, el nivel de confianza de 95% no es tanto un enunciado sobre cualquier intervalo en particular, más bien se refiere a lo que sucedería si se tuvieran que construir un gran número de intervalos semejantes.

Encontrar z a partir de un nivel de confianza

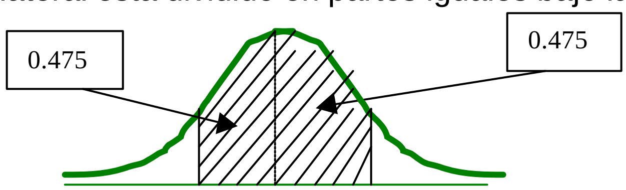
Existen varias tablas en las cuales podemos encontrar el valor de z, según sea el área proporcionada por la misma. En esta sección se realizará un ejemplo para encontrar el valor de z utilizando tres tablas diferentes.

Ejemplo:
Encuentre el valor de z para un nivel de confianza del 95%.

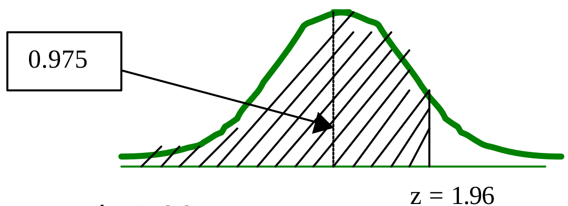
Solución 1:
Se utilizará la tabla que tiene el área bajo la curva de -∞ hasta z. Si lo vemos gráficamente sería:



El nivel de confianza bilateral está dividido en partes iguales bajo la curva:

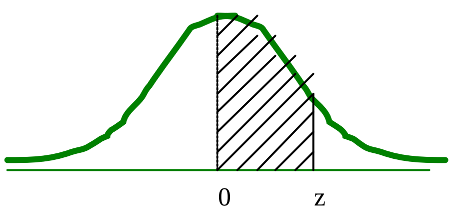


En base a la tabla que se esta utilizando, se tendrá que buscar el área de 0.975, ya que cada extremo o cola de la curva tiene un valor de 0.025.



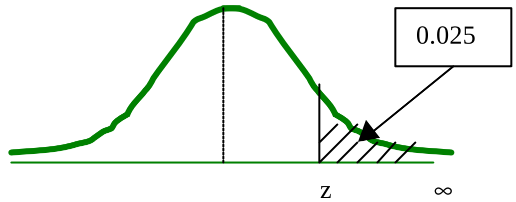
Por lo que el valor de z es de 1.96.

Solución 2:
Si se utiliza una tabla en donde el área bajo la curva es de 0 a z:



En este caso sólo se tendrá que buscar adentro de la tabla el área de 0.475 y el resultado del valor de z será el mismo, para este ejemplo 1.96.

Solución 3:
Para la tabla en donde el área bajo la curva va desde z hasta ∞:



Se busca el valor de 0.025 para encontrar z de 1.96.

Independientemente del valor del Nivel de Confianza este será el procedimiento a seguir para localizar a z. En el caso de que no se encuentre el valor exacto se tendrá que interpolar.

Estimación para la Media

Es conocido de nosotros durante este curso, que en base a la distribución muestral de medias que se generó en el tema anterior, la formula para el calculo de probabilidad es la siguiente: $z = \frac{\bar{x} - m}{s / \sqrt{n}}$. Como en este caso no conocemos el parámetro y lo queremos estimar por medio de la media de la muestra, sólo se despejará μ de la formula anterior, quedando lo siguiente:

$$m = \bar{x} \pm \frac{zs}{\sqrt{n}}$$

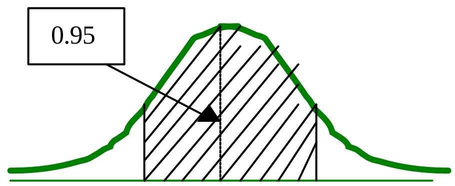
De esta formula se puede observar que tanto el tamaño de la muestra como el valor de z se conocerán. Z se puede obtener de la tabla de la distribución normal a partir del nivel de confianza establecido. Pero en ocasiones se desconoce σ por lo que en esos casos lo correcto es utilizar otra distribución llamada "t" de student si la población de donde provienen los datos es normal.
Para el caso de tamaños de muestra grande se puede utilizar una estimación puntual de la desviación estándar, es decir igualar la desviación estándar de la muestra a la de la población (s=σ).

Ejemplos:

1. Se encuentra que la concentración promedio de zinc que se saca del agua a partir de una muestra de mediciones de zinc en 36 sitios diferentes es de 2.6 gramos por mililitro. Encuentre los intervalos de confianza de 95% y 99% para la concentración media de zinc en el río. Suponga que la desviación estándar de la población es 0.3.

Solución:
La estimación puntual de μ es $\bar{x}$ = 2.6. El valor de z para un nivel de confianza del 95% es 1.96, por lo tanto:

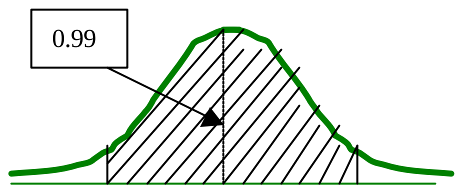
$$m = \bar{x} \pm \frac{zs}{\sqrt{n}} = 2.6 \pm \frac{(1.96)(0.3)}{\sqrt{36}} = 2.50\,y\,2.70$$



$$m_{\min} = 2.50 \qquad m_{\max} = 2.70$$

Para un nivel de confianza de 99% el valor de z es de 2.575 por lo que el intervalo será más amplio:

$$m = \bar{x} \pm \frac{zs}{\sqrt{n}} = 2.6 \pm \frac{(2.575)(0.3)}{\sqrt{36}} = 2.47\,y\,2.73$$



$$m_{\min} = 2.47 \qquad m_{\max} = 2.73$$

El intervalo de confianza proporciona una estimación de la presición de nuestra estimación puntual. Si μ es realmente el valor central de intervalo, entonces $\bar{x}$ estima μ sin error. La mayor parte de las veces, sin embargo, $\bar{x}$ no será exactamente igual a μ y la estimación puntual es errónea. La magnitud de este error será el valor absoluto de la diferencia entre μ y $\bar{x}$, y podemos tener el nivel de confianza de que esta diferencia no excederá $\frac{zs}{\sqrt{n}}$.

Como se puede observar en los resultados del ejercicio se tiene un error de estimación mayor cuando el nivel de confianza es del 99% y más pequeño cuando se reduce a un nivel de confianza del 95%.

2. Una empresa eléctrica fabrica focos que tienen una duración aproximadamente distribuida de forma normal con una desviación estándar de 40 horas. Si una muestra de 30 focos tiene una duración promedio de 780 horas, encuentre un intervalos de confianza de 96% para la media de la población de todos los focos que produce esta empresa.

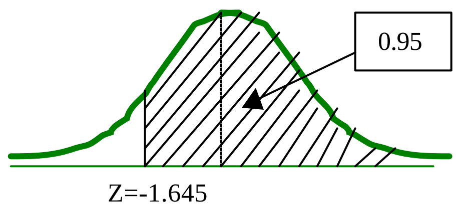
Solución:

$$\bar{x} - \frac{zs}{\sqrt{n}} < m < \bar{x} + \frac{zs}{\sqrt{n}} = 780 - \frac{(2.054)(40)}{\sqrt{30}} < m < 780 + \frac{(2.054)(40)}{\sqrt{30}} = 765 < m < 795$$

Con un nivel de confianza del 96% se sabe que la duración media de los focos que produce la empresa está entre 765 y 765 horas.

3. La prueba de corte sesgado es el procedimiento más aceptado para evaluar la calidad de una unión entre un material de reparación y su sustrato de concreto. El artículo “Testing the Bond Between Repair Materials and Concrete Substrate” informa que, en cierta investigación, se obtuvo una resistencia promedio muestral de 17.17 N/mm^2, con una muestra de 48 observaciones de resistencia al corte, y la desviación estándar muestral fue 3.28 N/mm^2. Utilice un nivel de confianza <u>inferior</u> del 95% para estimar la media real de la resistencia al corte.

Solución:
En este ejercicio se nos presentan dos situaciones diferentes a los ejercicios anteriores. La primera que desconoce la desviación estándar de la población y la segunda que nos piden un intervalo de confianza unilateral.
El primer caso ya se había comentado y se solucionará utilizando la desviación estándar de la muestra como estimación puntual de sigma.
Para el intervalo de confianza unilateral, se cargará el área bajo la curva hacia un solo lado como sigue:



$$\bar{x} - \frac{zs}{\sqrt{n}} = 17.17 - \frac{(1.654)(3.38)}{\sqrt{48}} = 16.39$$

Esto quiere decir que con un nivel de confianza de 95%, el valor de la media está en el intervalo (16.39, ∞).

Estimación de una Proporción

Un estimador puntual de la proporción P en un experimento binomial está dado por la estadística P=X/N, donde x representa el número de éxitos en *n* pruebas. Por tanto, la proporción de la muestra p =x/n se utiuñlizará como estimador puntual del parámetro P.

Si no se espera que la proporción P desconocida esté demasiado cerca de 0 ó de 1, se puede establecer un intervalo de confianza para P al considerar la distribución muestral de proporciones.

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

Al despejar P de esta ecuación nos queda:

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

En este despeje podemos observar que se necesita el valor del parámetro P y es precisamente lo que queremos estimar, por lo que lo sustituiremos por la proporción de la muestra *p* siempre y cuando el tamaño de muestra no sea pequeño.

$$P = p \pm z\sqrt{\frac{pq}{n}}$$

Cuando *n* es pequeña y la proporción desconocida P se considera cercana a 0 ó a 1, el procedimiento del intervalo de confianza que se establece aquí no es confiable, por tanto, no se debe utilizar. Para estar seguro, se debe requerir que np ó nq sea mayor o igual a 5.

El error de estimación será la diferencia absoluta entre p y P, y podemos tener el nivel de confianza de que esta diferencia no excederá $z\sqrt{\frac{pq}{n}}$.

Ejemplos:

1. Un fabricante de reproductores de discos compactos utiliza un conjunto de pruebas amplias para evaluar la función eléctrica de su producto. Todos los reproductores de discos compactos deben pasar todas las pruebas antes de venderse. Una muestra aleatoria de 500 reproductores tiene como resultado 15 que fallan en una o más pruebas. Encuentre un intervalo de confianza de 90% para la proporción de los reproductores de discos compactos de la población que no pasan todas las pruebas.

Solución:
n=500
p = 15/500 = 0.03
z(0.90) = 1.645

$$P = p \pm z\sqrt{\frac{pq}{n}} = 0.03 \pm (1.645)\sqrt{\frac{(0.03)(0.97)}{500}}$$

$$0.0237 < P < 0.0376$$

Se sabe con un nivel de confianza del 90% que la proporción de discos defectuosos que no pasan la prueba en esa población esta entre 0.0237 y 0.0376.

2. En una muestra de 400 pilas tipo B fabricadas por la Everlast Company, se encontraron 20 defectuosas. Si la proporción *p* de pilas defectuosas en esa muestra se usa para estimar *P*, que vendrá a ser la proporción verdadera de todas las pilas defectuosas tipo B fabricadas por la Everlast Company, encuentre el máximo error de estimación ε tal que se pueda tener un 95% de confianza en que *P* dista menos de ε de *p.*

Solución:
p=x/n = 20/400=0.05
z(0.95)=1.96

$$e = z\sqrt{\frac{pq}{n}} = 1.96\sqrt{\frac{(0.05)(0.95)}{400}} = 0.021$$

Si p=0.05 se usa para estimar P, podemos tener un 95% de confianza en que P dista menos de 0.021 de *p.* En otras palabras, si p=0.05 se usa para erstimar P, el error máximo de estimación será aproximadamente 0.021 con un nivel de confianza del 95%.

Para calcular el intervalo de confianza se tendría:

$$p \pm e = 0.05 \pm 0.021$$

Esto da por resultado dos valores, (0.029, 0.071). Con un nivel de confianza del 95% se sabe que la proporción de pulas defectuosas de esta compañía está entre 0.029 y 0.071.

Si se requiere un menor error con un mismo nivel de confianza sólo se necesita aumentar el tamaño de la muestra.

3. En un estudio de 300 accidentes de automóvil en una ciudad específica, 60 tuvieron consecuencias fatales. Con base en esta muestra, construya un intervalo del 90% de confianza para aproximar la proporción de todos los accidentes automovilísticos que en esa ciudad tienen consecuencias fatales.

Solución:
P= 60/300 = 0.20
Z(0.90) = 1.645

$$P = p \pm z\sqrt{\frac{pq}{n}} = 0.20 \pm (1.645)\sqrt{\frac{(0.20)(0.80)}{300}} = 0.20 \pm 0.038$$

$$0.162<P<0.238$$

Estimación de la Diferencia entre dos Medias

Si se tienen dos poblaciones con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2, respectivamente, un estimador puntual de la diferencia entre μ_1 y μ_2 está dado por la estadística $\bar{x}_1 - \bar{x}_2$. Por tanto. Para obtener una estimación puntual de μ_1- μ_2, se seleccionan dos muestras aleatorias independientes, una de cada población, de tamaño n_1 y n_2, se calcula la diferencia $\bar{x}_1 - \bar{x}_2$, de las medias muestrales.
Recordando a la distribución muestral de diferencia de medias:

$$z=\frac{(\bar{x}_1-\bar{x}_2)-(m_1-m_2)}{\sqrt{\frac{s_1^{\ 2}}{n_1}+\frac{s_2^{\ 2}}{n_2}}}$$

Al despejar de esta ecuación μ_1- μ_2 se tiene:

$$m_1-m_2=(\bar{x}_1-\bar{x}_2)\pm z\sqrt{\frac{s_1^{\ 2}}{n_1}+\frac{s_2^{\ 2}}{n_2}}$$

En el caso en que se desconozcan las varianzas de la población y los tamaños de muestra sean mayores a 30 se podrá utilizar la varianza de la muestra como una estimación puntual.

Ejemplos:

1. Se lleva a cabo un experimento en que se comparan dos tipos de motores, A y B. Se mide el rendimiento en millas por galón de gasolina. Se realizan 50 experimentos con el motor tipo A y 75 con el motor tipo B. La gasolina que se utiliza y las demás condiciones se mantienen constantes. El rendimiento promedio de gasolina para el motor A es de 36 millas por galón y el promedio para el motor B es 24 millas por galón. Encuentre un intervalo de confianza de 96% sobre la diferencia promedio real para los motores A y B. Suponga que las desviaciones estándar poblacionales son 6 y 8 para los motores A y B respectivamente.

Solución:

Es deseable que la diferencia de medias sea positiva por lo que se recomienda restar la media mayor menos la media menor. En este caso será la media del motor B menos la media del motor A.

El valor de z para un nivel de confianza del 96% es de 2.05.

$$m_B-m_A=(\bar{x}_B-\bar{x}_A)\pm z\sqrt{\frac{s_A^{\ 2}}{n_A}+\frac{s_B^{\ 2}}{n_B}}=(42-36)\pm 2.05\sqrt{\frac{36}{50}+\frac{64}{75}}$$

$$3.43<\mu_B-\mu_A<8.57$$

La interpretación de este ejemplo sería que con un nivel de confianza del 96% la diferencia del rendimiento promedio esta entre 3.43 y 8.57 millas por galón a favor del motor B. Esto quiere decir que el motor B da mas rendimiento promedio que el motor A, ya que los dos valores del intervalo son positivos.

2. Una compañía de taxis trata de decidir si comprar neumáticos de la marca A o de la B para su flotilla de taxis. Para estimar la diferencia de las dos marcas, se lleva a cabo un experimento utilizando 12 de cada marca. Los neumáticos se utilizan hasta que se desgastan, dando como resultado promedio para la marca A 36,300 kilómetros y para la marca B 38,100 kilómetros. Calcule un intervalo de confianza de 95% para la diferencia promedio de las dos marcas, si se sabe que las poblaciones se distribuyen

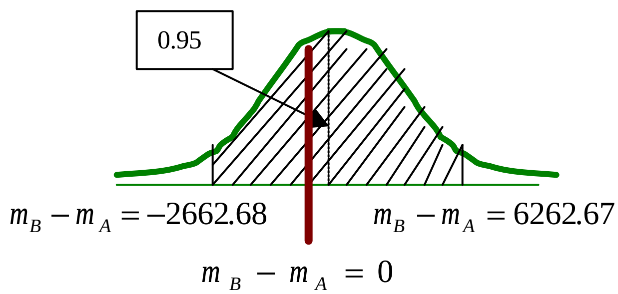
de forma aproximadamente normal con desviación estándar de 5000 kilómetros para la marca A y 6100 kilómetros para la marca B.

Solución:

$$m_B - m_A = (\bar{x}_B - \bar{x}_A) \pm z\sqrt{\frac{s_A^{\;2}}{n_A} + \frac{s_B^{\;2}}{n_B}} = (38100 - 36300) \pm 1.96\sqrt{\frac{5000^2}{12} + \frac{6100^2}{12}}$$

$-2662.68 < \mu_B - \mu_A < 6262.67$

Gráficamente:



Como el intervalo contiene el valor “cero”, no hay razón para creer que el promedio de duración del neumático de la marca B es mayor al de la marca A, pues el cero nos está indicando que pueden tener la misma duración promedio.

Estimación de la Diferencia de dos Proporciones

En la sección anterior se vio el tema de la generación de las distribuciones muestrales, en donde se tenía el valor de los parámetros, se seleccionaban dos muestras y podíamos calcular la probabilidad del comportamiento de los estadísticos. Para este caso en particular se utilizará la distribución muestral de diferencia de proporciones para la estimación de las misma. Recordando la formula:

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

Despejando P_1-P_2 de esta ecuación:

$$P_1 - P_2 = (p_1 - p_2) \pm z\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$$

Aquí se tiene el mismo caso que en la estimación de una proporción, ya que al hacer el despeje nos queda las dos proporciones poblacionales y es precisamente lo que queremos estimar, por lo que se utilizarán las proporciones de la muestra como estimadores puntuales:

$$P_1 - P_2 = (p_1 - p_2) \pm z\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

Ejemplos:

1. Se considera cierto cambio en un proceso de fabricación de partes componentes. Se toman muestras del procedimiento existente y del nuevo

para determinar si éste tiene como resultado una mejoría. Si se encuentra que 75 de 1500 artículos del procedimiento actual son defectuosos y 80 de 2000 artículos del procedimiento nuevo también lo son, encuentre un intervalo de confianza de 90% para la diferencia real en la fracción de defectuosos entre el proceso actual y el nuevo.

Solución:
Sean P_1 y P_2 las proporciones reales de defectuosos para los procesos actual y nuevo, respectivamente. De aquí, p_1=75/1500 = 0.05 y p_2 = 80/2000 = 0.04. con el uso de la tabla encontramos que z para un nivel de confianza del 90% es de 1.645.

$$P_1 - P_2 = (p_1 - p_2) \pm z\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} = (0.05 - 0.04) \pm 1.645\sqrt{\frac{(0.05)(0.95)}{1500} + \frac{(0.04)(0.96)}{2000}}$$

-0.0017$<P_1-P_2<0.0217$

Como el intervalo contiene el valor de cero, no hay razón para creer que el nuevo procedimiento producirá una disminución significativa en la proporción de artículos defectuosos comparado con el método existente.

2. Un artículo relacionado con la salud, reporta los siguientes datos sobre la incidencia de disfunciones importantes entre recién nacidos con madres fumadoras de marihuana y de madres que no la fumaban:

	Usuaria	**No Usuaria**
Tamaño Muestral	1246	11178
Número de disfunciones	42	294
Proporción muestral	0.0337	0.0263

Encuentre el intervalo de confianza del 99% para la diferencia de proporciones.

Solución:
Representemos P_1 la proporción de nacimientos donde aparecen disfunciones entre todas las madres que fuman marihuana y definamos P_2, de manera similar, para las no fumadoras. El valor de z para un 99% de confianza es de 2.58.

$$P_1 - P_2 = (p_1 - p_2) \pm z\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} = (0.0337 - 0.0263) \pm 2.58\sqrt{\frac{(0.0337)(0.9663)}{1246} + \frac{(0.0263)(0.9737)}{11178}}$$

-0.0064$<P_1-P_2<0.0212$

Este intervalo es bastante angosto, lo cual sugiere que P_1-P_2 ha sido estimado de manera precisa.

Determinación de Tamaños de Muestra para Estimaciones

Al iniciar cualquier investigación, la primer pregunta que surge es: ¿de qué tamaño debe ser la o las muestras?. La respuesta a esta pregunta la veremos en esta sección, con conceptos que ya se han visto a través de este material.

Cálculo del Tamaño de la Muestra para Estimar una Media

¿Qué tan grande debe ser una muestra si la media muestral se va a usar para estimar la media poblacional?. La respuesta depende del error estándar de la media, si este fuera cero, entonces se necesitaría una sola media que será igual necesariamente a la media poblacional desconocida μ, porque $\sigma = 0$. Este caso extremo no se encuentra en la práctica, pero refuerza el hecho de que mientras menor sea el error estándar de la media, menor es el tamaño de muestra necesario para lograr un cierto grado de precisión.

Se estableció antes que una forma de disminuir el error de estimación es aumentar el tamaño de la muestra, si éste incluye el total de la población, entonces $|\bar{x} - m|$ sería igual a cero. Con esto en mente, parece razonable que para un nivel de confianza fijo, sea posible determinar un tamaño de la muestra tal que el error de estimación sea tan pequeño como queramos, para ser mas preciso, dado un nivel de confianza y un error fijo de estimación ε, se puede escoger un tamaño de muestra *n* tal que $P(|\bar{x} - m| < \varepsilon)$ = Nivel de confianza. Con el propósito de determinar *n.* El error máximo de estimación esta dado por:

$$\varepsilon = \frac{zs}{\sqrt{n}}$$

Si se eleva al cuadrado ambos lados de esta ecuación y se despeja *n* de la ecuación resultante, obtenemos:

$$n = \left(\frac{zs}{e}\right)^2$$

Como *n* debe de ser un número entero, redondeamos hacia arriba todos los resultados fraccionarios.

En el caso de que se tenga una población finita y un muestreo sin reemplazo, el error de estimación se convierte en:

$$\varepsilon = \frac{zs}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

De nuevo se eleva al cuadrado ambos lados y se despeja la *n*, obteniendo:

$$n = \frac{z^2 s^2 N}{e^2 (N-1) + z^2 s^2}$$

Ejemplos:

1. Un biólogo quiere estimar el peso promedio de los ciervos cazados en el estado de Maryland. Un estudio anterior de diez ciervos cazados mostró que la desviación estándar de sus pesos es de 12.2 libras. ¿Qué tan grande debe ser una muestra para que el biólogo tenga el 95% de confianza de que el error de estimación es a lo más de 4 libras?

Solución:

$$n=\left(\frac{zs}{e}\right)^2=\left(\frac{(1.96)(12.2)}{4}\right)^2=35.736$$

En consecuencia, si el tamaño de la muestra es 36, se puede tener un 95% de confianza en que μ difiere en menos de 4 libras de $\bar{x}$.

2. Una empresa eléctrica fabrica focos que tienen una duración aproximadamente normal con una desviación estándar de 40 horas. ¿De qué tamaño se necesita una muestra si se desea tener 96% de confianza que la media real esté dentro de 10 horas de la media real?

$$n=\left(\frac{zs}{e}\right)^2=\left(\frac{(2.053)(40)}{10}\right)^2=67.43$$

Se necesita una muestra de 68 focos para estimar la media de la población y tener un error máximo de 10 horas.

¿Qué pasaría si en lugar de tener un error de estimación de 10 horas sólo se requiere un error de 5 horas?

$$n=\left(\frac{zs}{e}\right)^2=\left(\frac{(2.053)(40)}{5}\right)^2=269.74$$

Se puede observar como el tamaño de la muestra aumenta, pero esto tiene como beneficio una estimación más exacta.

3. Suponga que en el ejercicio anterior se tiene una población de 300 focos, y se desea saber de que tamaño debe de ser la muestra. El muestreo se realizará sin reemplazo.

Solución:

Como se tiene una población finita y un muestreo sin reemplazo es necesario utilizar la formula con el factor de corrección.

$$n=\frac{z^2 s^2 N}{e^2(N-1)+z^2 s^2}=\frac{(2.053)^2(40)^2(300)}{(10)^2(300-1)+(2.053)^2(40)^2}=55.21$$

Si se tiene una población finita de 300 focos sólo se tiene que extraer de la población una muestra sin reemplazo de 56 focos para poder estimar la duración media de los focos restantes con un error máximo de 10 horas.

Cálculo del Tamaño de la Muestra para Estimar una Proporción

Se desea saber que tan grande se requiere que sea una muestra para asegurar que el error al estimar P sea menor que una cantidad específica ε.

$$e=z\sqrt{\frac{pq}{n}}$$

Elevando al cuadrado la ecuación anterior se despeja n y nos queda:

$$n=\frac{z^2 pq}{e^2}$$

Esta fórmula está algo engañosa, pues debemos utilizar *p* para determinar el tamaño de la muestra, pero *p* se calcula a partir de la muestra. Existen ocasiones en las cuales se tiene una idea del comportamiento de la proporción de la población y ese valor se puede sustituir en la fórmula, pero si no se sabe nada referente a esa proporción entonces se tienen dos opciones:

- Tomar una muestra preliminar mayor o igual a 30 para proporcionar una estimación de P. Después con el uso de la fórmula se podría determinar de forma aproximada cuántas observaciones se necesitan para proporcionar el grado de precisión que se desea.
- Tomar el valor de p como 0.5 ya que sustituyendo este en la fórmula se obtiene el tamaño de muestra mayor posible. Observe el siguiente ejemplo:

Se desconoce el valor de P, por lo que se utilizarán diferentes valores y se sustituirán en la formula para observar los diferentes tamaños de muestras. El nivel de confianza que se utilizará es del 95% con un error de estimación de 0.30.

p	$\frac{z^2pq}{e^2}$	**n**
0.10	$\frac{(1.96)^2(0.10)(0.90)}{(0.30)^2}$	3.84
0.20	$\frac{(1.96)^2(0.20)(0.80)}{(0.30)^2}$	6.82
0.30	$\frac{(1.96)^2(0.30)(0.70)}{(0.30)^2}$	8.96
0.40	$\frac{(1.96)^2(0.40)(0.60)}{(0.30)^2}$	10.24
0.50	$\frac{(1.96)^2(0.50)(0.50)}{(0.30)^2}$	10.67
0.60	$\frac{(1.96)^2(0.60)(0.40)}{(0.30)^2}$	10.24
0.70	$\frac{(1.96)^2(0.70)(0.30)}{(0.30)^2}$	8.96
0.80	$\frac{(1.96)^2(0.80)(0.20)}{(0.30)^2}$	6.82
0.90	$\frac{(1.96)^2(0.90)(0.10)}{(0.30)^2}$	3.84

Como se puede observar en la tabla anterior cuando P vale 0.5 el tamaño de la muestra alcanza su máximo valor.

En el caso de que se tenga una población finita y un muestreo sin reemplazo, el error de estimación se convierte en:

$$e = z\sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}$$

De nuevo se eleva al cuadrado ambos lados y se despeja la *n*, obteniendo:

$$n=\frac{z^2pqN}{e^2(N-1)+z^2pq}$$

Ejemplos:

1. En una muestra aleatoria de 500 familias que tienen televisores en la ciudad de Hamilton, Canadá, se encuentra que 340 están suscritas a HBO. ¿Qué tan grande se requiere que sea una muestra si se quiere tener 95% de confianza de que la estimación de P esté dentro de 0.02?

Solución:
Se tratarán a las 500 familias como una muestra preliminar que proporciona una estimación de p=340/500=0.68.

$$n=\frac{z^2pq}{e^2}=\frac{(1.96)^2(0.68)(0.32)}{(0.02)^2}=2090$$

Por lo tanto si basamos nuestra estimación de P sobre una muestra aleatoria de tamaño 2090, se puede tener una confianza de 95% de que nuestra proporción muestral no diferirá de la proporción real por más de 0.02.

2. Una legisladora estatal desea encuestar a los residentes de su distrito para conocer qué proporción del electorado conoce la opinión de ella, respecto al uso de fondos estatales para pagar abortos. ¿Qué tamaño de muestra se necesita si se requiere un confianza del 95% y un error máximo de estimación de 0.10?

Solución:
En este problema, se desconoce totalmente la proporción de residentes que conoce la opinión de la legisladora, por lo que se utilizará un valor de 0.5 para *p.*

$$n=\frac{z^2pq}{e^2}=\frac{(1.96)^2(0.50)(0.50)}{(0.10)^2}=96.04$$

Se requiere un tamaño de muestra de 97 residentes para que con una confianza del 95% la estimación tenga un error máximo de 0.10.

Cálculo del Tamaño de la Muestra para Estimar la Diferencia de Medias

Si se recuerda a la distribución muestral de diferencia de medias se tiene que error esta dado por:

$$e=z\sqrt{\frac{s^2{}_1}{n_1}+\frac{s^2{}_2}{n_2}}$$

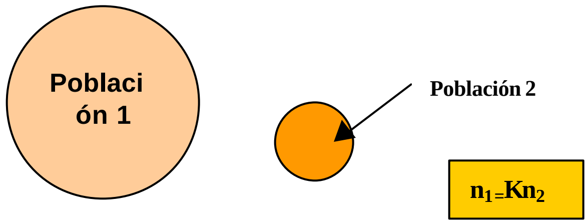
En esta ecuación se nos pueden presentar dos casos:

- Los tamaños de muestra son iguales.
- Los tamaño de muestra son diferentes .

Para el primer caso no se tiene ningún problema, se eleva al cuadrado la ecuación y se despeja n ya que n_1 es igual a n_2.

$$n = \frac{z^2\left(s^2{}_1 + s^2{}_2\right)}{e^2}$$

Para el segundo caso se pondrá una n en función de la otra. Este caso se utiliza cuando las poblaciones son de diferente tamaño y se sabe que una es K veces mayor que la otra.



$$n_2 = \frac{z^2\left(s^2{}_1 + ks^2{}_2\right)}{ke^2}$$

Ejemplo:

Un director de personal quiere comparar la efectividad de dos métodos de entrenamiento para trabajadores industriales a fin de efectuar cierta operación de montaje. Se divide un número de operarios en dos grupos iguales: el primero recibe el método de entrenamiento 1, y el segundo, el método 2. Cada uno realizará la operación de montaje y se registrará el tiempo de trabajo. Se espera que las mediciones para ambos grupos tengan una desviación estándar aproximadamente de 2 minutos. Si se desea que la estimación de la diferencia en tiempo medio de montaje sea correcta hasta por un minuto, con una probabilidad igual a 0.95, ¿cuántos trabajadores se tienen que incluir en cada grupo de entrenamiento?

$$n = \frac{z^2\left(s^2{}_1 + s^2{}_2\right)}{e^2} = \frac{\left(1.96^2\right)\left(2^2 + 2^2\right)}{1^2} = 31$$

Cada grupo debe contener aproximadamente 31 empleados.

Cálculo del Tamaño de la Muestra para Estimar la Diferencia de Proporciones

Si se recuerda a la distribución muestral de diferencia de medias se tiene que error esta dado por:

$$e = z\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

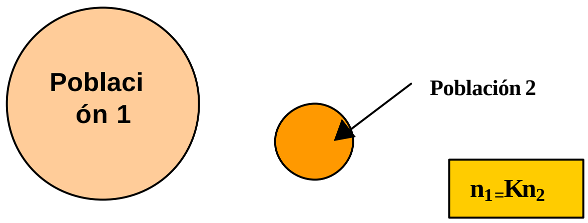
En esta ecuación se nos pueden presentar dos casos:

- Los tamaños de muestra son iguales.
- Los tamaño de muestra son diferentes .

Para el primer caso no se tiene ningún problema, se eleva al cuadrado la ecuación y se despeja n ya que n_1 es igual a n_2.

$$n = \frac{z^2 \left(p_1 q_1 + p_2 q_2 \right)}{e^2}$$

Para el segundo caso se pondrá una *n* en función de la otra. Este caso se utiliza cuando las poblaciones son de diferente tamaño y se sabe que una es *K* veces mayor que la otra.



$$n_2 = \frac{z^2 \left(p_1 q_1 + k p_2 q_2 \right)}{k e^2}$$

Ejemplo:

Una compañía de productos alimenticios contrató a una empresa de investigación de mercadotecnia , para muestrear dos mercados, I y II, a fin de comparar las proporciones de consumidores que prefieren la comida congelada de la compañía con los productos de sus competidores. No hay información previa acerca de la magnitud de las proporciones P_1 y P_2. Si la empresa de productos alimenticios quiere estimar la diferencia dentro de 0.04, con una probabilidad de 0.95, ¿ cuántos consumidores habrá que muestrear en cada mercado?

$$n = \frac{z^2 \left(p_1 q_1 + p_2 q_2 \right)}{e^2} = \frac{\left(1.96^2\right)\left[(0.5)(0.5) + (0.5)(0.5)\right]}{0.04^2} = 1200.5$$

Se tendrá que realizar encuestas a 1201 consumidores de cada mercado para tener una estimación con una confianza del 95% y un error máximo de 0.04.

Problemas propuestos

1. Se probó una muestra aleatoria de 400 cinescopios de televisor y se encontraron 40 defectuosos. Estime el intervalo que contiene, con un coeficiente de confianza de 0.90, a la verdadera fracción de elementos defectuosos.

2. Se planea realizar un estudio de tiempos para estimar el tiempo medio de un trabajo, exacto dentro de 4 segundos y con una probabilidad de 0.90, para terminar un trabajo de montaje. Si la experiencia previa sugiere que $\sigma = 16$ seg. mide la variación en el tiempo de montaje entre un trabajador y otro al realizar una sola operación de montaje, ¿cuántos operarios habrá que incluir en la muestra?

3. El decano registró debidamente el porcentaje de calificaciones D y F otorgadas a los estudiantes por dos profesores universitarios de matemáticas. El profesor I alcanzó un 32%, contra un 21% para el profesor II, con 200 y 180 estudiantes, respectivamente. Estime la diferencia entre los

porcentajes de calificaciones D y F otorgadas por los dos profesores. Utilice un nivel de confianza del 95% e interprete los resultados.

4. Suponga que se quiere estimar la producción media por hora, en un proceso que produce antibiótico. Se observa el proceso durante 100 períodos de una hora, seleccionados al azar y se obtiene una media de 34 onzas por hora con una desviación estándar de 3 onzas por hora. Estime la producción media por hora para el proceso, utilizando un nivel de confianza del 95%.

5. Un ingeniero de control de calidad quiere estimar la fracción de elementos defectuosos en un gran lote de lámparas. Por la experiencia, cree que la fracción real de defectuosos tendría que andar alrededor de 0.2. ¿Qué tan grande tendría que seleccionar la muestra si se quiere estimar la fracción real, exacta dentro de 0.01, utilizando un nivel de confianza fe 95%?

6. Se seleccionaron dos muestras de 400 tubos electrónicos, de cada una de dos líneas de producción, A y B. De la línea A se obtuvieron 40 tubos defectuosos y de la B 80. Estime la diferencia real en las fracciones de defectuosos para las dos líneas, con un coeficiente de confianza de 0.90 e interprete los resultados.

7. Se tienen que seleccionar muestras aleatorias independientes de $n_1=n_2=n$ observaciones de cada una de dos poblaciones binomiales, 1 y 2. Si se desea estimar la diferencia entre los dos parámetros binomiales, exacta dentro de 0.05, con una probabilidad de 0.98. ¿qué tan grande tendría que ser n?. No se tiene información anterior acerca de los valores P_1 y P_2, pero se quiere estar seguro de tener un número adecuado de observaciones en la muestra.

8. Se llevan a cabo pruebas de resistencia a la tensión sobre dos diferentes clases de largueros de aluminio utilizados en la fabricación de alas de aeroplanos comerciales. De la experiencia pasada con el proceso de fabricación se supone que las desviaciones estándar de las resistencias a la tensión son conocidas. La desviación estándar del larguero 1 es de 1.0 Kg/mm^2 y la del larguero 2 es de 1.5 Kg/mm^2. Se sabe que el comportamiento de las resistencias a la tensión de las dos clases de largueros son aproximadamente normal. Se toma una muestra de 10 largueros del tipo 1 obteniéndose una media de 87.6 Kg/mm^2, y otra de tamaño 12 para el larguero 2 obteniéndose una media de 74.5 Kg/mm^2. Estime un intervalo de confianza del 90% para la diferencia en la resistencia a la tensión promedio.

9. Se quiere estudiar la tasa de combustión de dos propelentes sólidos utilizados en los sistemas de escape de emergencia de aeroplanos. Se sabe que la tasa de combustión de los dos propelentes tiene aproximadamente la

misma desviación estándar; esto es $\sigma_1=\sigma_2$ = 3 cm/s. ¿Qué tamaño de muestra debe utilizarse en cada población si se desea que el error en la estimación de la diferencia entre las medias de las tasas de combustión sea menor que 4 cm/s con una confianza del 99%?.

Respuesta a los Problemas propuestos

1. $0.07532 \leq P \leq 0.1246$
2. n= 44
3. $0.0222 \leq P_1 - P_2 \leq 0.1978$
4. $33.412 \leq \mu \leq 34.588$
5. n= 6147
6. $0.059 \leq P_B - P_A \leq 0.141$
7. n= 1086
8. $12.22 \leq \mu_1 - \mu_2 \leq 13.98$
9. n= 8

U N I D A D I I

PRUEBA DE HIPOTESIS

Las secciones anteriores han mostrado cómo puede estimarse un parámetro a partir de los datos contenidos en una muestra. Puede encontrarse ya sea un sólo número (estimador puntual) o un intervalo de valores posibles (intervalo de confianza). Sin embargo, muchos problemas de ingeniería, ciencia, y administración, requieren que se tome una decisión entre aceptar o rechazar una proposición sobre algún parámetro. Esta proposición recibe el nombre de **hipótesis.** Este es uno de los aspectos más útiles de la inferencia estadística, puesto que muchos tipos de problemas de toma de decisiones, pruebas o experimentos en el mundo de la ingeniería, pueden formularse como problemas de prueba de hipótesis.

Una **hipótesis estadística** es una proposición o supuesto sobre los parámetros de una o más poblaciones.

Suponga que se tiene interés en la rapidez de combustión de un agente propulsor sólido utilizado en los sistemas de salida de emergencia para la tripulación de aeronaves. El interés se centra sobre la rapidez de combustión promedio. De manera específica, el interés recae en decir si la rapidez de combustión promedio es o no 50 cm/s. Esto puede expresarse de manera formal como

$$H_o;\ \mu = 50 \text{ cm/s}$$
$$H_1;\ \mu \neq 50 \text{ cm/s}$$

La proposición H_o; μ = 50 cm/s, se conoce como **hipótesis nula**, mientras que la proposición H_1; $\mu \neq$ 50 cm/s, recibe el nombre de **hipótesis alternativa.** Puesto que la hipótesis alternativa especifica valores de μ que pueden ser mayores o menores que 50 cm/s, también se conoce como **hipótesis alternativa bilateral.** En algunas situaciones, lo que se desea es formular una **hipótesis alternativa unilateral**, como en

$$H_o;\ \mu = 50 \text{ cm/s} \qquad\qquad H_o;\ \mu = 50 \text{ cm/s}$$
$$\text{ó}$$
$$H_1;\ \mu < 50 \text{ cm/s} \qquad\qquad H_1;\ \mu > 50 \text{ cm/s}$$

Es importante recordar que las hipótesis siempre son proposiciones sobre la población o distribución bajo estudio, no proposiciones sobre la muestra. Por lo general, el valor del parámetro de la población especificado en la hipótesis nula se determina en una de tres maneras diferentes:

1. Puede ser resultado de la experiencia pasada o del conocimiento del proceso, entonces el objetivo de la prueba de hipótesis usualmente es determinar si ha cambiado el valor del parámetro.

2. Puede obtenerse a partir de alguna teoría o modelo que se relaciona con el proceso bajo estudio. En este caso, el objetivo de la prueba de hipótesis es verificar la teoría o modelo.

3. Cuando el valor del parámetro proviene de consideraciones externas, tales como las especificaciones de diseño o ingeniería, o de obligaciones contractuales. En esta situación, el objetivo usual de la prueba de hipótesis es probar el cumplimiento de las especificaciones.

Un procedimiento que conduce a una decisión sobre una hipótesis en particular recibe el nombre de **prueba de hipótesis**. Los procedimientos de prueba de hipótesis dependen del empleo de la información contenida en la muestra aleatoria de la población de interés. Si esta información es consistente con la hipótesis, se concluye que ésta es verdadera; sin embargo si esta información es inconsistente con la hipótesis, se concluye que esta es falsa. Debe hacerse hincapié en que la verdad o falsedad de una hipótesis en particular nunca puede conocerse con certidumbre, a menos que pueda examinarse a toda la población. Usualmente esto es imposible en muchas situaciones prácticas. Por tanto, es necesario desarrollar un procedimiento de prueba de hipótesis teniendo en cuenta la probabilidad de llegar a una conclusión equivocada.

La **hipótesis nula,** representada por H_o, es la afirmación sobre una o más características de poblaciones que al inicio se supone cierta (es decir, la "creencia a priori").

La **hipótesis alternativa,** representada por H_1, es la afirmación contradictoria a H_o, y ésta es la hipótesis del investigador.

La hipótesis nula se rechaza en favor de la hipótesis alternativa, sólo si la evidencia muestral sugiere que H_o es falsa. Si la muestra no contradice decididamente a H_o, se continúa creyendo en la validez de la hipótesis nula. Entonces, las dos conclusiones posibles de un análisis por prueba de hipótesis son ***rechazar H_o o no rechazar H_o.***

Prueba de una Hipótesis Estadística

Para ilustrar los conceptos generales, considere el problema de la rapidez de combustión del agente propulsor presentado con anterioridad. La hipótesis nula es que la rapidez promedio de combustión es 50 cm/s, mientras que la hipótesis alternativa es que ésta no es igual a 50 cm/s. Esto es, se desea probar:

$$H_o;\ \mu = 50 \text{ cm/s}$$
$$H_1;\ \mu \neq 50 \text{ cm/s}$$

Supóngase que se realiza una prueba sobre una muestra de 10 especímenes, y que se observa cual es la rapidez de combustión promedio muestral. La media muestral es un estimador de la media verdadera de la población. Un valor de la media muestral $\bar{x}$ que este próximo al valor hipotético μ = 50 cm/s es una evidencia de que el verdadero valor de la media μ es realmente 50 cm/s; esto es, tal evidencia apoya la hipótesis nula H_o. Por otra parte, una media muestral muy diferente de 50 cm/s constituye una evidencia que apoya la hipótesis alternativa H_1. Por tanto, en este caso, la media muestral es el estadístico de prueba.

La media muestral puede tomar muchos valores diferentes. Supóngase que si $48.5 \leq \bar{x} \leq 51.5$, entonces no se rechaza la hipótesis nula H_o; μ = 50 cm/s, y que si $\bar{x}$ <48.5 ó $\bar{x}$ >51.5, entonces se acepta la hipótesis alternativa H_1; $\mu \neq$ 50 cm/s. Los valores de $\bar{x}$ que son menores que 48.5 o mayores que 51.5 constituyen la ***región crítica*** de la prueba, mientras que todos los valores que están en el intervalo $48.5 \leq \bar{x} \leq 51.5$ forman la ***región de aceptación***. Las fronteras entre las regiones crítica y de aceptación reciben el nombre de ***valores críticos.*** La costumbre es establecer conclusiones con respecto a la hipótesis nula H_o. Por tanto, se rechaza H_o en favor de H_1 si el estadístico de prueba cae en la región crítica, de lo contrario, no se rechaza H_o.

Este procedimiento de decisión puede conducir a una de dos conclusiones erróneas. Por ejemplo, es posible que el valor verdadero de la rapidez promedio de combustión del agente propulsor sea igual a 50 cm/s. Sin embargo, para todos los especímenes bajo prueba, bien puede observarse un valor del estadístico de prueba $\bar{x}$ que cae en la región crítica. En este caso, la hipótesis nula H_o será rechazada en favor de la alternativa H_1 cuando, de hecho, H_o en realidad es verdadera. Este tipo de conclusión equivocada se conoce como ***error tipo I.***

El **error tipo I** se define como el rechazo de la hipótesis nula H_o cuando ésta es verdadera. También es conocido como α **ó nivel de significancia.**

Si tuviéramos un nivel de confianza del 95% entonces el nivel de significancia sería del 5%. Análogamente si se tiene un nivel de confianza del 90% entonces el nivel de significancia sería del 10%.

Ahora supóngase que la verdadera rapidez promedio de combustión es diferente de 50 cm/s, aunque la media muestral $\bar{x}$ caiga dentro de la región de aceptación. En este caso se acepta H_o cuando ésta es falsa. Este tipo de conclusión recibe el nombre de ***error tipo II.***

El **error tipo II ó error** β se define como la aceptación de la hipótesis nula cuando ésta es falsa.

Por tanto, al probar cualquier hipótesis estadística, existen cuatro situaciones diferentes que determinan si la decisión final es correcta o errónea.

Decisión	H_o es verdadera	H_o es falsa
Aceptar H_o	No hay error	Error tipo II ó β
Rechazar H_o	Error tipo I ó α	No hay error

1. Los errores tipo I y tipo II están relacionados. Una disminución en la probabilidad de uno por lo general tiene como resultado un aumento en la probabilidad del otro.
2. El tamaño de la región crítica, y por tanto la probabilidad de cometer un error tipo I, siempre se puede reducir al ajustar el o los valores críticos.
3. Un aumento en el tamaño muestral n reducirá α y β de forma simultánea.
4. Si la hipótesis nula es falsa, β es un máximo cuando el valor real del parámetro se aproxima al hipotético. Entre más grande sea la distancia entre el valor real y el valor hipotético, será menor β.

PASOS PARA ESTABLECER UN ENSAYO DE HIPOTESIS INDEPENDIENTEMENTE DE LA DISTRIBUCION QUE SE ESTE TRATANDO

1. Interpretar correctamente hacia que distribución muestral se ajustan los datos del enunciado.

2. Interpretar correctamente los datos del enunciado diferenciando los parámetros de los estadísticos. Así mismo se debe determinar en este punto información implícita como el tipo de muestreo y si la población es finita o infinita.

3. Establecer simultáneamente el ensayo de hipótesis y el planteamiento gráfico del problema. El ensayo de hipótesis está en función de ***parámetros*** ya que se quiere evaluar el universo de donde proviene la muestra. En este punto se determina el tipo de ensayo (unilateral o bilateral).

4. Establecer la regla de decisión. Esta se puede establecer en función del valor crítico, el cual se obtiene dependiendo del valor de α (Error tipo I o nivel de significancia) o en función del estadístico límite de la distribución muestral. Cada una de las hipótesis deberá ser argumentada correctamente para tomar la decisión, la cual estará en función de la hipótesis nula o H_o.

5. Calcular el estadístico real, y situarlo para tomar la decisión.

6. Justificar la toma de decisión y concluir.

Tipos de Ensayo

Se pueden presentar tres tipos de ensayo de hipótesis que son:

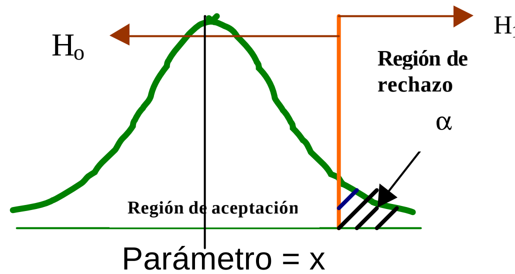
- Unilateral Derecho
- Unilateral Izquierdo
- Bilateral

Dependiendo de la evaluación que se quiera hacer se seleccionará el tipo de ensayo.

- Unilateral Derecho. El investigador desea comprobar la hipótesis de un aumento en el parámetro, en este caso el nivel de significancia se carga todo hacia el lado derecho, para definir las regiones de aceptación y de rechazo.

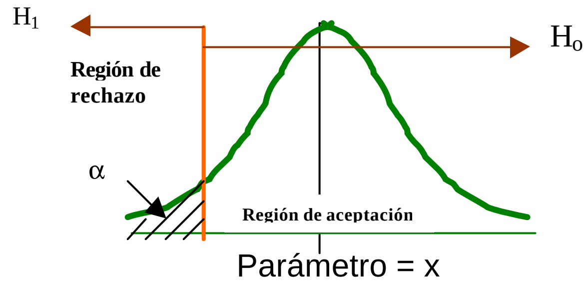
  Ensayo de hipótesis:

  H_o; Parámetro $\leq$ x
  H_1; Parámetro $>$ x



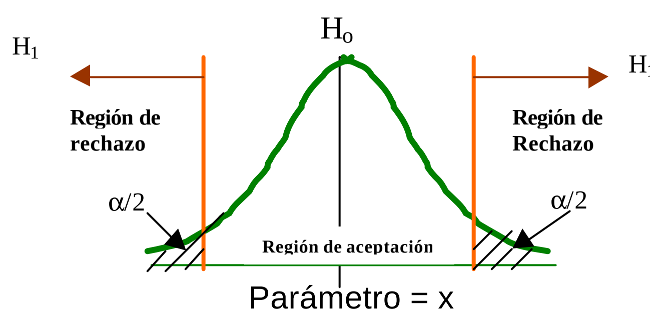
- Unilateral Izquierdo: El investigador desea comprobar la hipótesis de una disminución en el parámetro, en este caso el nivel de significancia se carga todo hacia el lado izquierdo, para definir las regiones de aceptación y de rechazo.
  Ensayo de hipótesis:

  H_o; Parámetro $\geq$ x
  H_1; Parámetro $<$ x



- Bilateral: El investigador desea comprobar la hipótesis de un cambio en el parámetro. El nivel de significancia se divide en dos y existen dos regiones de rechazo.
  Ensayo de hipótesis:

  H_o; Parámetro = x
  H_1; Parámetro $\neq$ x



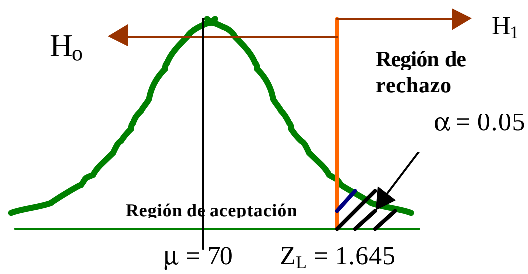
Para realizar los ejemplos y ejercicios de ensayo de hipótesis se recomienda seguir los pasos mencionados anteriormente. Los ejemplos siguientes se solucionarán por los pasos recomendados, teniéndose una variedad de problemas en donde se incluirán a todas las distribuciones muestrales que se han visto hasta aquí.

Ejemplos:

1. Una muestra aleatoria de 100 muertes registradas en Estados Unidos el año pasado muestra una vida promedio de 71.8 años. Suponga una desviación estándar poblacional de 8.9 años, ¿esto parece indicar que la vida media hoy en día es mayor que 70 años? Utilice un nivel de significancia de 0.05.

Solución:

1. Se trata de una distribución muestral de medias con desviación estándar conocida.
2. Datos:
   μ=70 años
   σ = 8.9 años
   $\bar{x}$ = 71.8 años
   n = 100
   α = 0.05
3. Ensayo de hipótesis
   H_o; μ = 70 años.
   H_1; μ > 70 años.



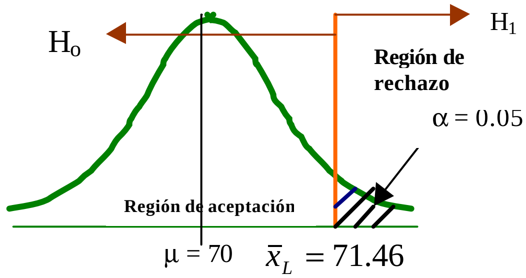
4. Regla de decisión:
   Si $z_R \leq$ 1.645 no se rechaza H_o.
   Si $z_R >$ 1.645 se rechaza H_o.
5. Cálculos:

$$Z_R = \frac{\bar{x}_R - m}{s/\sqrt{n}} = \frac{71.8 - 70}{8.9/\sqrt{100}} = 2.02$$

6. Justificación y decisión.
   Como 2.02 >1.645 se rechaza H_o y se concluye con un nivel de significancia del 0.05 que la vida media hoy en día es mayor que 70 años.

Existe otra manera de resolver este ejercicio, tomando la decisión en base al estadístico real, en este caso la media de la muestra. De la formula de la distribución muestral de medias se despeja la media de la muestra:

$$Z_L = \frac{\bar{x}_L - m}{s/\sqrt{n}} \qquad \bar{x}_L = m + \frac{Z_l s}{\sqrt{n}} = 70 + \frac{(1.645)(8.9)}{\sqrt{100}} = 71.46$$



Regla de decisión:
Si $\bar{x}_R \leq 71.46$ No se rechaza H_o
Si $\bar{x}_R > 71.46$ Se rechaza H_o

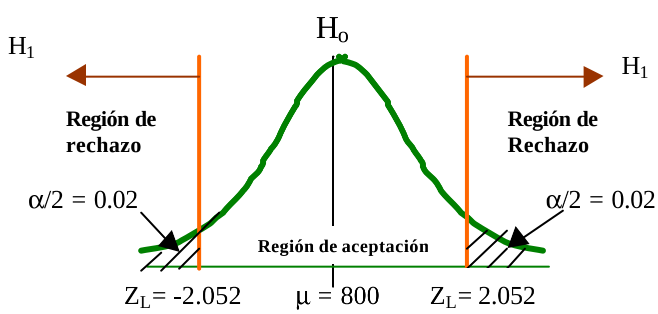
Como la media de la muestral es de 71.8 años y es mayor al valor de la media muestral límite de 71.46 por lo tanto se rechaza H_o y se llega a la misma conclusión.

2. Una empresa eléctrica fabrica focos que tienen una duración que se distribuye de forma aproximadamente normal con una media de 800 horas y una desviación estándar de 40 horas. Si una muestra aleatoria de 30 focos tiene una duración promedio de 788 horas, ¿muestran los datos suficiente evidencia para decir que la duración media ha cambiado? Utilice un nivel de significancia del 0.04.

Solución:
1. Se trata de una distribución muestral de medias con desviación estándar conocida.

2. Datos:
   μ=800 horas
   σ = 40 horas
   $\bar{x}$ = 788 horas
   n = 30
   α = 0.04

3. Ensayo de hipótesis
   H_o; μ = 800 horas
   H_1; $\mu \neq$ 800 horas



4. Regla de Decisión:
   Si $-2.052 \leq Z_R \leq 2.052$ No se rechaza H_o
   Si $Z_R < -2.052$ ó si $Z_R > 2.052$ Se rechaza H_o

5. Cálculos:

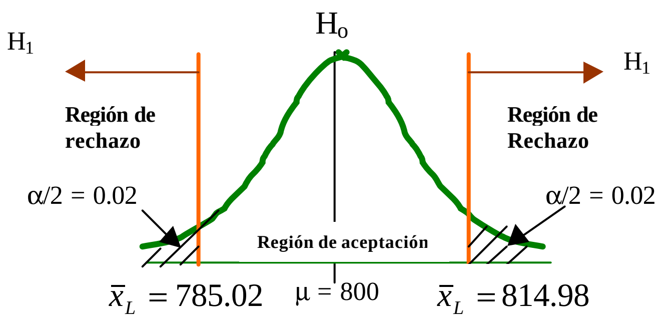
$$Z_R = \frac{\bar{x}_R - m}{s/\sqrt{n}} = \frac{788 - 800}{40/\sqrt{30}} = -1.643$$

6. Justificación y decisión:

Como $-2.052 \leq$ **-1.643** ≤ 2.052 por lo tanto, no se rechaza H_o y se concluye con un nivel de significancia del 0.04 que la duración media de los focos no ha cambiado.

Solución por el otro método:

$$\bar{x}_L = m \pm \frac{Z_l s}{\sqrt{n}} = 800 \pm \frac{(2.052)(40)}{\sqrt{30}} = 785.02 \text{ y } 814.98$$



Regla de decisión:
Si $785.02 \leq \bar{x}_R \leq 814.98$ No se rechaza H_o
Si $\bar{x}_R < 785.02$ ó $\bar{x}_R > 814.98$ se rechaza H_o

Como la $\bar{x}_R$ = 788 horas, entonces no se rechaza H_o y se concluye que la duración media de los focos no ha cambiado.

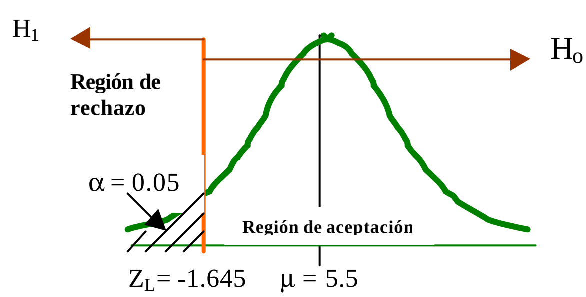
3. Una muestra aleatoria de 64 bolsas de palomitas de maíz pesan, en pomedio 5.23 onzas con una desviación estándar de 0.24 onzas. Pruebe la hipótesis de que $\mu = 5.5$ onzas contra al hipótesis alternativa, $\mu < 5.5$ onzas en el nivel de significamcia de 0.05.

Solución:

1. Se trata de una distribución muestral de medias con desviación estándar desconocida, pero como el tamaño de muestra es mayor a 30 se puede tomar la desviación muestral como un estimador puntual para la poblacional.

2. Datos:
   μ= 5.5 onzas
   s= 0.24 onzas
   $\bar{x}$ = 5.23 onzas
   n = 64
   α = 0.05

3. Ensayo de hipótesis
   H_o; μ = 5.5 onzas
   H_1; μ < 5.5 onzas



4. Regla de decisión:
   Si $Z_R \geq -1.645$ No se rechaza H_o
   Si $Z_R < -1.645$ Se rechaza H_o

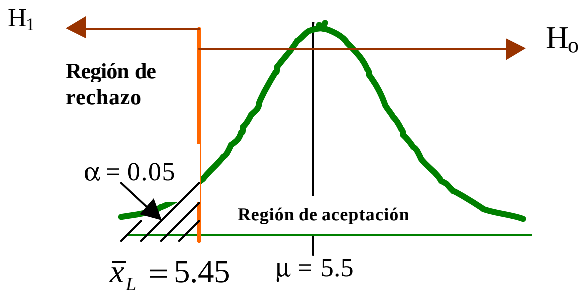
5. Cálculos:

$$Z_R = \frac{\bar{x}_R - m}{s/\sqrt{n}} = \frac{5.23 - 5.5}{0.24/\sqrt{64}} = -9$$

6. Justificación y decisión:
   Como $-9 < -1.645$ por lo tanto se rechaza H_o y se concluye con un nivel de significancia del 0.05 que las bolsas de palomitas pesan en promedio menos de 5.5 onzas.

Solución por el otro método:

$$\bar{x}_L = m - \frac{Z_l\, s}{\sqrt{n}} = 5.5 - \frac{(1.645)(0.24)}{\sqrt{64}} = 5.45$$



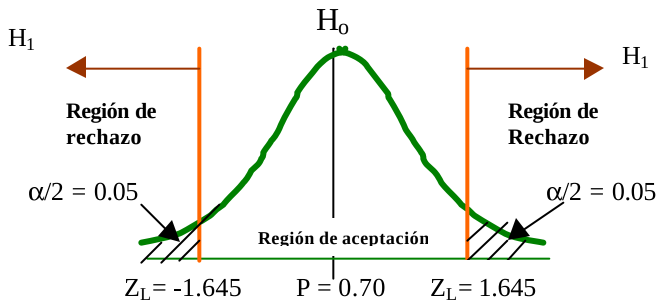
Regla de decisión:
Si $\bar{x}_R \geq 5.45$ No se Rechaza H_o
Si $\bar{x}_R < 5.45$ Se rechaza H_o

Como la $\bar{x}_R = 5.23$ y este valor es menor que 5.45 pot lo tanto se rechaza H_o.

4. Un constructor afirma que se instalan bombas de calor en 70% de todas las casas que se construyen hoy en día en la ciudad de Richmond. ¿Estaría de acuerdo con esta afirmación si una investigación de casas nuevas en esta ciudad muestra que 8 de 15 tienen instaladas bombas de calor? Utilice un nivel de significancia de 0.10.

Solución:
1. Se trata de una distribución muestral de proporciones.
2. Datos:
   P= 0.70
   p = 8/15 = 0.5333
   n = 15
   $\alpha = 0.10$

3. Ensayo de hipótesis
   H_o; $P = 0.70$
   H_1; $P \neq 0.70$



4. Regla de Decisión:
   Si $-1.645 \leq Z_R \leq 1.645$ No se rechaza H_o
   Si $Z_R < -1.645$ ó si $Z_R > 1.645$ Se rechaza H_o

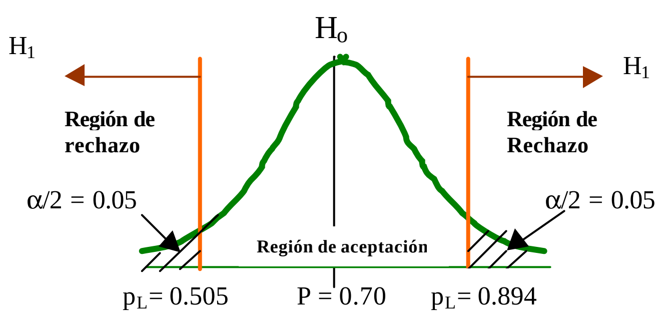
5. Cálculos:

$$Z_R = \frac{p - P}{\sqrt{\frac{Pq}{n}}} = \frac{0.533 - 0.70}{\sqrt{\frac{(0.70)(0.30)}{15}}} = -1.41$$

6. Justificación y decisión:
   Como $-1.645 \leq$ **-1.41** ≤ 1.645 No se rechaza H_o y se concluye con un nivel de significancia de 0.10 que la afirmación del constructor es cierta.

Solución por el otro método:

$$p_L = P \pm z_L \sqrt{\frac{Pq}{n}} = 0.70 \pm 1.645\sqrt{\frac{(0.70)(0.30)}{15}}$$ = 0.505 y 0.894



Regla de decisión:
Si $0.505 \leq p_R \leq 0.894$ No se rechaza H_o
Si $p_R < 0.505$ ó si $Z_R > 0.894$ Se rechaza H_o

Como el valor del estadístico real es de 0.533 por lo tanto no se rechaza H_o y se llega a la misma conclusión.

5. Un fabricante de semiconductores produce controladores que se emplean en aplicaciones de motores automovilísticos. El cliente requiere que la fracción de controladores defectuosos en uno de los pasos de manufactura críticos no sea mayor que 0.05, y que el fabricante demuestre esta característica del proceso de fabricación con este nivel de calidad, utilizando $\alpha = 0.05$. El

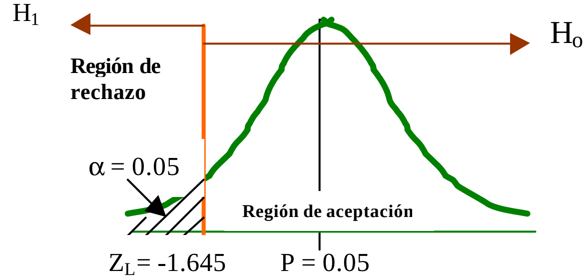
fabricante de semiconductores toma una muestra aleatoria de 200 dispositivos y encuentra que cuatro de ellos son defectuosos. ¿El fabricante puede demostrar al cliente la calidad del proceso?

Solución:

1. Se trata de una distribución muestral de proporciones.

2. Datos:
   P= 0.05
   p = 4/200 = 0.02
   n = 200
   α = 0.05

3. Ensayo de hipótesis
   H_o; P = 0.05
   H_1; P < 0.05



4. Regla de decisión:
   Si $Z_R \geq -1.645$ No se rechaza H_o
   Si $Z_R < -1.645$ Se rechaza H_o

5. Cálculos:

$$Z_R = \frac{p-P}{\sqrt{\frac{Pq}{n}}} = \frac{0.02-0.05}{\sqrt{\frac{(0.05)(0.95)}{200}}} = -1.946$$

6. Justificación y decisión:
   Puesto que –1.946<-1.645, se rechaza H_o y se concluye con un nivel de significancia del 0.05 que la fracción de artículos defectuosos es menor que 0.05.

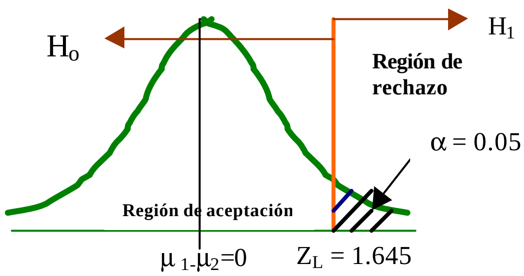
6. Un diseñador de productos está interesado en reducir el tiempo de secado de una pintura tapaporos. Se prueban dos fórmulas de pintura; la fórmula 1 tiene el contenido químico estándar, y la fórmula 2 tiene un nuevo ingrediente secante que debe reducir el tiempo de secado. De la experiencia se sabe que la desviación estándar del tiempo de secado es ocho minutos, y esta variabilidad inherente no debe verse afectada por la adición del nuevo ingrediente. Se pintan diez especímenes con la fórmula 1, y otros diez con la fórmula 2. Los dos tiempos promedio de secado muestrales son 121 min y 112 min respectivamente. ¿A qué conclusiones puede llegar el diseñador del producto sobre la eficacia del nuevo ingrediente, utilizando α = 0.05?

Solución:

1. Se trata de una distribución muestral de diferencia de medias con desviación estándar conocida.

2. Datos:
   $\sigma_1=\sigma_2= 8$
   $\bar{x}_1 = 121 \min$
   $\bar{x}_2 = 112 \min$
   $n_1=n_2= 10$
   $\alpha = 0.05$

3. Ensayo de hipótesis
   H_o; $\mu_1-\mu_2 = 0$
   H_1; $\mu_1-\mu_2 > 0$ Se desea rechazar H_o si el nuevo ingrediente disminuye el tiempo promedio de secado, por eso se pone la diferencia mayor a cero o sea positiva para poder probar que μ_2 es menor que μ_1.



4. Regla de decisión:
   Si $z_R \leq 1.645$ no se rechaza H_o.
   Si $z_R > 1.645$ se rechaza H_o.

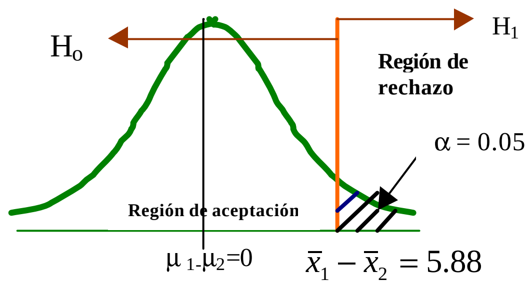
5. Cálculos:

$$Z_R = \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(121 - 112) - 0}{\sqrt{\frac{8^2}{10} + \frac{8^2}{10}}} = 2.52$$

6. Justificación y decisión:
   Puesto que 2.52>1.645, se rechaza H_o, y se concluye con un nivel de significancia de 0.05 que la adición del nuevo ingrediente a la pintura si disminuye de manera significativa el tiempo promedio de secado.

Solución por el otro método:

$$(\bar{x}_1 - \bar{x}_2)_L = (m_1 - m_2) + z\sqrt{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}} = 0 + 1.645\sqrt{\frac{8^2}{10} + \frac{8^2}{10}} = 5.88$$



Regla de decisión:
Si $(\bar{x}_1 - \bar{x}_2)_R \leq 5.88$ No se rechaza H_o
Si $(\bar{x}_1 - \bar{x}_2)_R > 5.88$ Se rechaza H_o

Puesto que $(\bar{x}_1 - \bar{x}_2)_R$ = 121-112 = 9 y este número es mayor a 5.88 por lo tanto se rechaza H_o.

7. Se utilizan dos máquinas para llenar botellas de plástico con un volumen neto de 16.0 onzas. Las distribuciones de los volúmenes de llenado pueden suponerse normales, con desviaciones estándar σ_1= 0.020 y σ_2 = 0.025 onzas. Un miembro del grupo de ingeniería de calidad sospecha que el volumen neto de llenado de ambas máquinas es el mismo, sin importar si éste es o no de 16 onzas. De cada máquina se toma una muestra aleatoria de 10 botellas. ¿Se encuentra el ingeniero en lo correcto? Utilice α = 0.05

MAQUINA 1		MAQUINA 2	
16.03	16.01	16.02	16.03
16.04	15.96	15.97	16.04
16.05	15.98	15.96	16.02
16.05	16.02	16.01	16.01
16.02	15.99	15.99	16.00

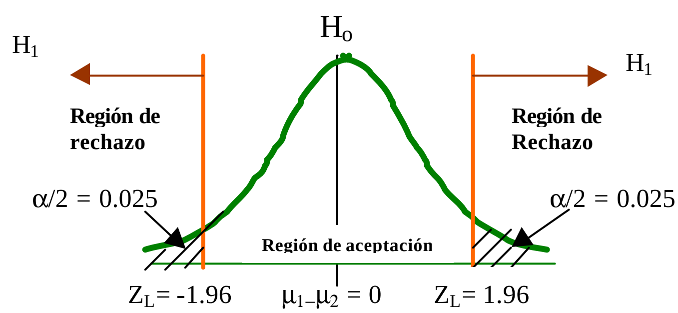
Solución:

1. Se trata de una distribución muestral de diferencia de medias con desviación estándar conocida.

2. Datos:
   σ_1= 0.020
   σ_2= 0.025
   $\bar{x}_1 = 16.015$ Este dato se obtuvo calculando la media de los datos en la máquina 1.

$\bar{x}_2 = 16.005$ Este dato se obtuvo calculando la media de los datos en la máquina 2.
$n_1 = n_2 = 10$
$\alpha = 0.05$

3. Ensayo de hipótesis
   H_o; $\mu_1 - \mu_2 = 0$
   H_1; $\mu_1 - \mu_2 \neq 0$ Si se cae en H_o se podrá probar que el volumen de llenado es el mismo en las dos máquinas.



4. Regla de Decisión:
   Si $-1.96 \leq Z_R \leq 1.96$ No se rechaza H_o
   Si $Z_R < -1.96$ ó si $Z_R > 1.96$ Se rechaza H_o

5. Cálculos:

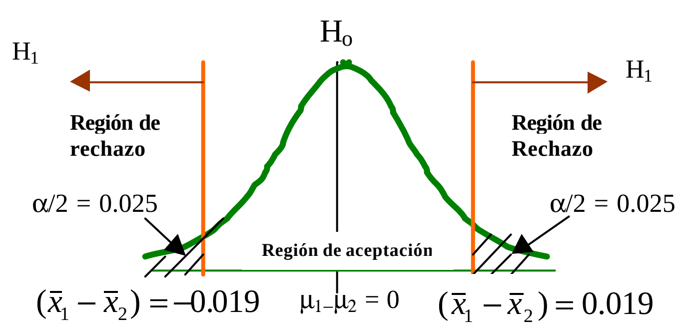
$$Z_R = \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(16.015 - 16.005) - 0}{\sqrt{\frac{0.020^2}{10} + \frac{0.025^2}{10}}} = 0.987$$

6. Justificación y decisión:

Como $-1.96 \leq$ **0.987** ≤ 1.96 entonces no se rechaza H_o y se concluye con un nivel de significancia de 0.05 que las dos máquinas tienen en promedio la misma cantidad de llenado.

Solución por el otro método:

$$(\bar{x}_1 - \bar{x}_2)_L = (m_1 - m_2) \pm z\sqrt{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}} = 0 \pm 1.96\sqrt{\frac{0.020^2}{10} + \frac{0.025^2}{10}} = \text{-0.019 y 0.019}$$



Regla de decisión:
Si –0-019$\leq(\bar{x}_1-\bar{x}_2)_R\leq$ 0.019 No se rechaza H_o
Si $(\bar{x}_1-\bar{x}_2)_R<$ -0.019 ó $(\bar{x}_1-\bar{x}_2)_R>$ 0.019 Se rechaza H_o

Como $(\bar{x}_1-\bar{x}_2)_R=$ 16.015 – 16.005 = 0.01, entonces cae en la región de aceptación y no se rechaza H_o.

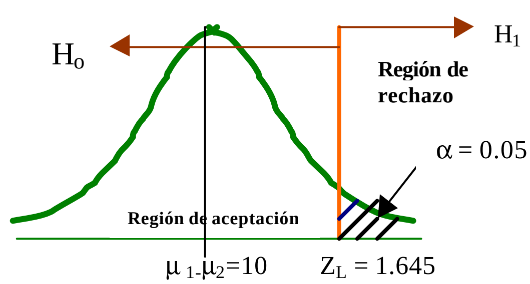
8. Existen dos tipos de plástico apropiados para su uso por un fabricante de componentes electrónicos. La tensión de ruptura de ese plástico es un parámetro importante . Se sabe que $\sigma_1=\sigma_2=$ 1.0 psi. De una muestra aleatoria de tamaño 10 y 12 para cada plástico respectivamente, se tiene una media de 162.5 para el plástico 1 y de 155 para el plástico 2. La compañía no adoptará el plástico 1 a menos que la tensión de ruptura de éste exceda a la del plástico 2 al menos por 10 psi. Con base a la información contenida en la muestra, ¿la compañía deberá utilizar el plástico 1? Utilice α = 0.05 para llegar a una decisión.

Solución:
1. Se trata de una distribución muestral de diferencia de medias con desviación estándar conocida.

2. Datos:
   $\sigma_1=\sigma_2=$ 1.0 psi
   $\bar{x}_1=162.5\,psi$
   $\bar{x}_2=155\,psi$
   $n_1=$ 10
   $n_2=$ 12
   α = 0.05

3. Ensayo de hipótesis

   H_o; $\mu_1-\mu_2$ = 10
   H_1; $\mu_1-\mu_2$ > 10 Se desea rechazar H_o si la media del plástico 1 supera a la media del plástico 2 en por lo menos 10 psi.



4. Regla de decisión:
   Si $z_R \leq 1.645$ no se rechaza H_o.
   Si $z_R > 1.645$ se rechaza H_o.

5. Cálculos:

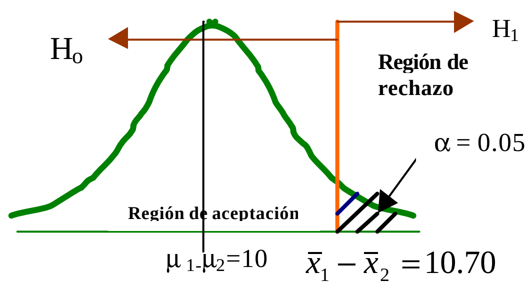
$$Z_R = \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{(162.5 - 155) - 10}{\sqrt{\dfrac{1^2}{10} + \dfrac{1^2}{12}}} = -5.83$$

6. Justificación y decisión:

No existe evidencia suficiente para apoyar el uso del plástico 1 ya que $-5.83 \leq 1.645$, por lo tanto no se rechaza H_o.

Solución por el otro método:

$$(\bar{x}_1 - \bar{x}_2)_L = (m_1 - m_2) + z\sqrt{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}} = 10 + 1.645\sqrt{\frac{1^2}{10} + \frac{1^2}{12}} = 10.70$$



Regla de decisión:

Si $(\bar{x}_1 - \bar{x}_2)_R \leq 10.70$ No se rechaza H_o

Si $(\bar{x}_1 - \bar{x}_2)_R > 10.70$ Se rechaza H_o

Puesto que $(\bar{x}_1 - \bar{x}_2)_R$ = 162.5-155 = 7.5 y este número es no es mayor a 10.7 por lo tanto no se rechaza H_o.

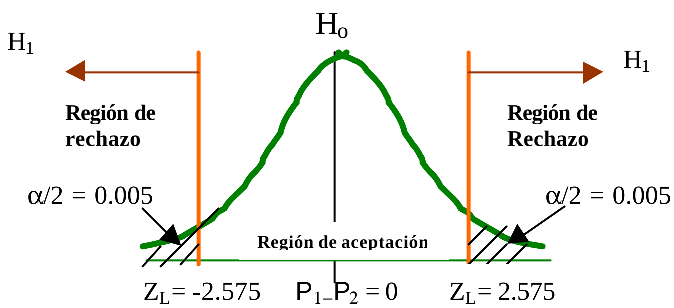
9. Se evalúan dos tipos diferentes de soluciones para pulir, para su posible uso en una operación de pulido en la fabricación de lentes intraoculares utilizados en el ojo humano después de una cirugía de cataratas. Se pulen 300 lentes con la primera solución y, de éstos, 253 no presentaron defectos inducidos por el pulido. Después se pulen otros 300 lentes con la segunda solución, de los cuales 196 resultan satisfactorios. ¿Existe alguna razón para creer que las dos soluciones para pulir son diferentes? Utilice $\alpha = 0.01$

Solución:

1. Se trata de una distribución muestral de diferencia de proporciones.

2. Datos:
   p_1= 253/300= 0.8433
   p_2 = 196/300= 0.6533
   n_1=n_2 = 300

3. Ensayo de hipótesis:
   H_o; P_1-P_2 = 0
   H_1; P_1-$P_2 \neq 0$



4. Regla de Decisión:
   Si –2.575≤ Z_R≤ 2.575 No se rechaza H_o
   Si Z_R < -2.575 ó si Z_R > 2.575 Se rechaza H_o

5. Cálculos:

$$Z_R = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

En esta fórmula se puede observar que en el denominador se tienen a las proporciones poblacionales o sea los parámetros, los cuales no se conocen, por lo que en el ensayo de hipótesis la fórmula para poder calcular la Z_R cambia, estimando a el parámetro común P de la siguiente forma:

$$P = \frac{x_1 + x_2}{n_1 + n_2} \text{ ó bien } P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

Entonces la fórmula de Z_R quedaría de la siguiente manera:

$$Z_R = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Se calculará el valor de P:

$$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{253 + 196}{300 + 300} = 0.7483$$

$$Z_R = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.8433 - 0.6533) - 0}{\sqrt{(0.7483)(0.2517)\left(\frac{1}{300} + \frac{1}{300}\right)}} = 5.36$$

6. Justificación y decisión:
   Puesto que 5.36>2.575, se rechaza la hipótesis nula y se concluye con un nivel de significancia de 0.01 que los dos fluidos para pulir son diferentes.

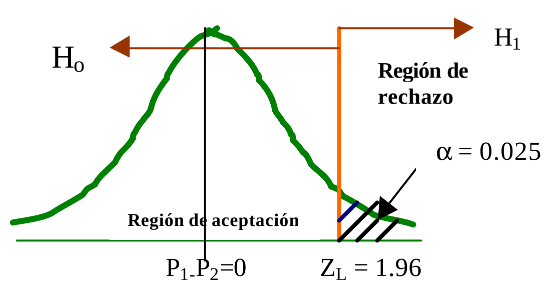
10. Se tomará el voto entre los residentes de una ciudad y el condado circundante para determinar si se debe construir una planta química propuesta. El lugar de construcción está dentro de los límites de la ciudad y por esta razón muchos votantes del condado consideran que la propuesta pasará debido a la gran proporción de votantes que favorecen la construcción. Para determinar si hay una diferencia significativa en la proporción de votantes de la ciudad y votantes del condado que favorecen la propuesta, se realiza una encuesta. Si 120 de 200 votantes de la ciudad favorecen la propuesta y 240 de 500 residentes del condado también lo hacen, ¿estaría de acuerdo en que la proporción de votantes de la ciudad que favorecen la propuesta es más alto que la proporción de votantes del condado? Utilice un nivel de significancia de 0.025.

Solución:

1. Se trata de una distribución muestral de diferencia de proporciones.

2. Datos:
   p_1= 120/200= 0.60
   p_2 = 240/500= 0.48
   n_1 = 200
   n_2 = 500

3. Ensayo de hipótesis:
   H_o; P_1-P_2 = 0
   H_1; P_1-P_2 > 0



4. Regla de decisión:
   Si $z_R \leq 1.96$ no se rechaza H_o.
   Si $z_R > 1.96$ se rechaza H_o.

5. Cálculos:
   Se calculará el valor de P:

$$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{120 + 240}{200 + 500} = 0.51 \; P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{120 + 240}{200 + 500} = 0.51$$

$$Z_R = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.60 - 0.48) - 0}{\sqrt{(0.51)(0.49)\left(\frac{1}{200} + \frac{1}{500}\right)}} = 2.9$$

6. Justificación y decisión:
   Puesto que 2.9>1.96, se rechaza la hipótesis nula y se concluye con un nivel de significancia de 0.025 que la proporción de votantes de la ciudad a favor de la propuesta es más alta que la proporción de votantes del condado.

Uso de valores P para la toma de decisiones

Al probar hipótesis en las que la estadística de prueba es discreta, la región crítica se puede elegir de forma arbitraria y determinar su tamaño. Si α es demasiado grande, se puede reducir al hacer un ajuste en el valor crítico. Puede ser necesario aumentar el tamaño de la muestra para compensar la disminución que ocurre de manera automática en la potencia de la prueba (probabilidad de rechazar H_o dado que una alternativa específica es verdadera).

Por generaciones enteras de análisis estadístico, se ha hecho costumbre elegir un nivel de significancia de 0.05 ó 0.01 y seleccionar la región crítica en consecuencia. Entonces, por supuesto, el rechazo o no rechazo estricto de H_o dependerá de esa región crítica. En la estadística aplicada los usuarios han adoptado de forma extensa la aproximación del ***valor P.*** La aproximación se diseña para dar al usuario una alternativa a la simple conclusión de “rechazo” o “no rechazo”.

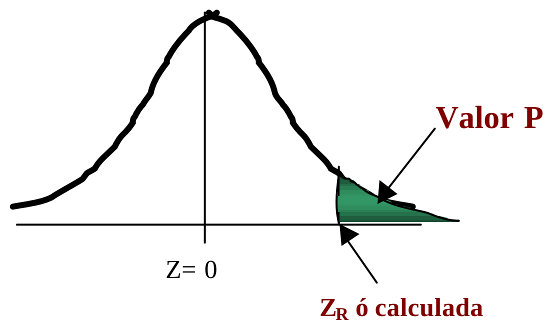
La aproximación del ***valor P*** como ayuda en la toma de decisiones es bastante natural pues casi todos los paquetes de computadora que proporcionan el cálculo de prueba de hipótesis entregan valores de P junto con valores de la estadística de la prueba apropiada.

- Un ***valor P*** es el nivel (de significancia) más bajo en el que el valor observado de la estadística de prueba es significativo.

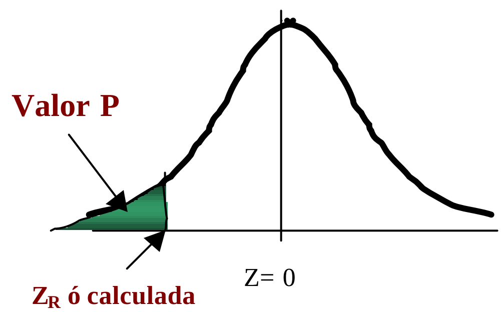
- El ***valor P*** es el nivel de significancia más pequeño que conduce al rechazo de la hipótesis nula H_o.
- El ***valor P*** es el mínimo nivel de significancia en el cual H_o sería rechazada cuando se utiliza un procedimiento de prueba especificado con un conjunto dado de información. Una vez que el valor de P se haya determinado, la conclusión en cualquier nivel α particular resulta de comparar el valor P con α:

1. Valor P $\leq \alpha \Rightarrow$ rechazar H_o al nivel α.
2. Valor P $> \alpha \Rightarrow$ No rechazar H_o al nivel α.

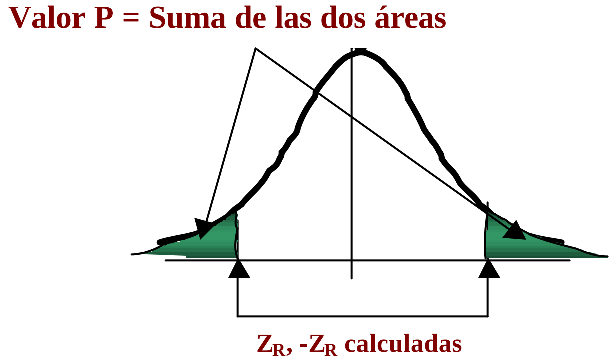
Ensayo Unilateral Derecho:



Ensayo Unilateral Izquierdo:



Ensayo Bilateral:

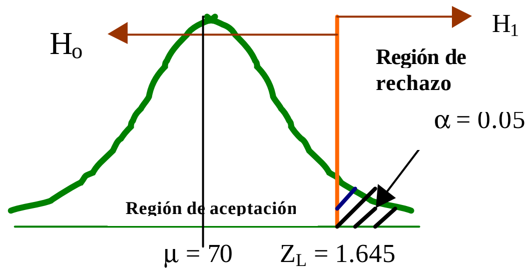


Ejemplos:

1. Calcular el valor de P para el primer ejemplo de ensayo de hipótesis en donde se quería probar que la edad media de los habitantes de Estados Unidos es superior a 70 años.

Solución:

1. Ensayo de hipótesis
   H_o; μ = 70 años.
   H_1; μ > 70 años.

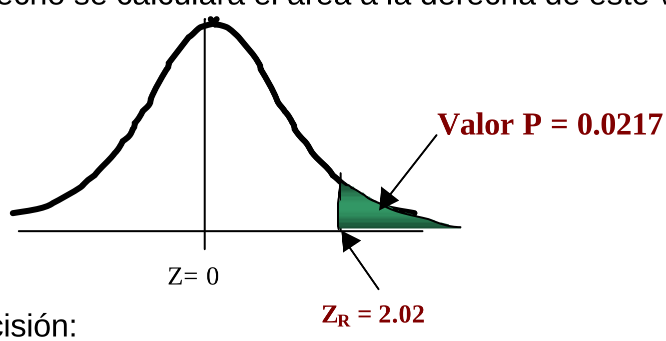


2. Regla de decisión:
   Si P≤ 0.05 se rechaza H_o.
   Si P > 0.05 No se rechaza H_o.

3. Cálculos:

$$Z_R = \frac{\bar{x}_R - m}{s/\sqrt{n}} = \frac{71.8 - 70}{8.9/\sqrt{100}} = 2.02$$

Esta es el valor de Z que se utilizará para calcular el valor de P, como es un ensayo unilateral derecho se calculará el área a la derecha de este valor.



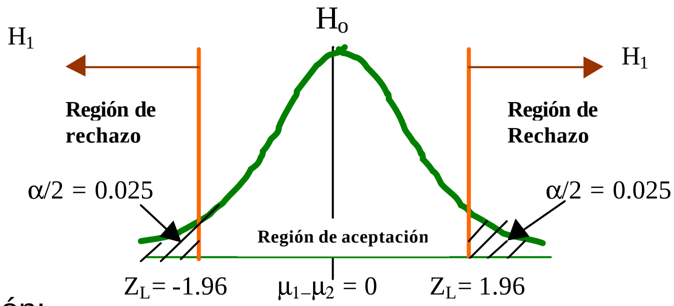
4. Justificación y decisión:
   Como el valor de P es 0.217 y es menor al valor del nivel de significancia de 0.05 por lo tanto se rechaza H_0, y se concluye que la edad media de los habitantes es mayor a 70 años.

2. Calcular el valor de P para el ejemplo 7 de esta sección en donde se tiene dos máquinas y se quiere ver si tienen la misma cantidad promedio de llenado en las botellas de plástico.

Solución:

1. Ensayo de hipótesis
   H_o; μ_1-μ_2 = 0

H_1; μ_1-$\mu_2 \neq 0$ Si se cae en H_o se podrá probar que el volumen de llenado es el mismo en las dos máquinas.



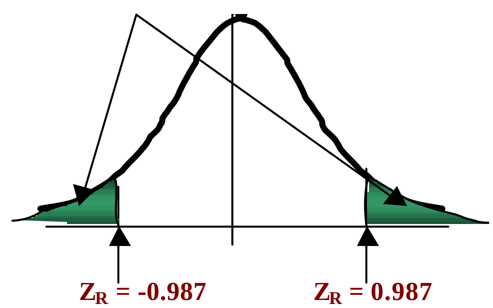
2. Regla de Decisión:
   Si P≤ 0.05 Se rechaza H_o
   Si P > 0.05 No se rechaza H_o

3. Cálculos:

$$Z_R = \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(16.015 - 16.005) - 0}{\sqrt{\frac{0.020^2}{10} + \frac{0.025^2}{10}}} = 0.987$$

Como este es un ensayo bilateral se procederá a calcular el valor de P mediante el valor de la Z_R, positiva y negativa y luego se sumarán las áreas.

Valor P = 0.1618 + 0.1618 = 0.3236



Como el valor de P es mayor al de α, se no se rechaza H_0, y se concluye que las maquinas tienen el mismo llenado promedio.

3. Se afirma que un automóvil se maneja en promedio más de 20,000 kilómetros por año. Para probar esta afirmación, se pide a una muestra de 100 propietarios de automóviles que lleven un registro de los kilómetros que viajen. ¿Está de acuerdo con esta afirmación si la muestra aleatoria tiene un promedio de 23,500 kilómetros y una desviación estándar de 3900 kilómetros? Utilice un valor P para su conclusión.

Solución:
En este ejercicio no nos manejan ningún valor de α, por lo que se procederá a plantear el ensayo y luego calcular z para poder conocer el valor de P y llegar a una conclusión.

1. Ensayo de hipótesis
   H_o; μ = 20,000 kilómetros.
   H_1; μ > 20,000 kilómetros.

2. Cálculos:

$$Z_R = \frac{\overline{x}_R - m}{s/\sqrt{n}} = \frac{23500 - 20000}{3900/\sqrt{100}} = 8.97$$

3. Decisión.
   Se observa que este valor de Z es muy grande, ni siquiera se encuentra en la tabla, entonces quiere decir que el área a la derecha de ese valor es **cero** y este sería el valor de P, por lo que no apoya a la hipótesis nula y se concluye que los automóviles se manejan en promedio más de 20,000 kilómetros por año.

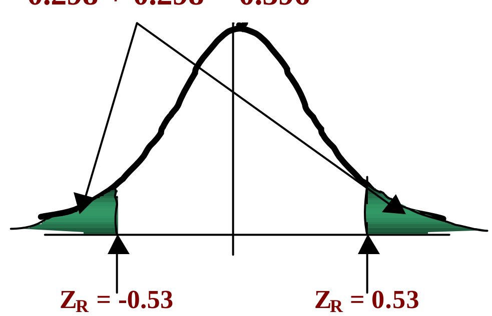
4. Se estudia la fracción de circuitos integrados defectuosos producidos en un proceso de fotolitografía. Para ello se somete a prueba una muestra de 300 circuitos, en la que 13 son defectuosos. Utilice los datos para probar H_o: P=0.05 contra H_1: P≠ 0.05. Utilice un valor de P para su conclusión.

Solución:

1. Ensayo de hipótesis
   H_o; P = 0.05
   H_1; P ≠ 0.05
2. Cálculos:

$$Z_R = \frac{p - P}{\sqrt{\frac{Pq}{n}}} = \frac{0.043 - 0.05}{\sqrt{\frac{(0.05)(0.95)}{300}}} = -0.53$$

Valor P = 0.298 + 0.298 = 0.596



3. Decisión:
   Este valor de P de 0.596 es muy grande por lo que se concluye que la fracción defectuosa de circuitos integrados es de 0.05, o sea no se rechaza H_o.

ERROR TIPO II ó b

Al evaluar un procedimiento de prueba de hipótesis, también es importante examinar la probabilidad del error tipo II, el cual se denota por β. Esto es,

$$\beta = \text{P(error tipo II)} = \text{P(aceptar } H_o / \ H_o \text{ es falsa)}$$

Para calcular β se debe tener una hipótesis alternativa específica; esto es, debe tenerse un valor particular del parámetro. Por ejemplo, supóngase que es importante rechazar la hipótesis nula H_o: $\mu = 50$ cada vez que la rapidez promedio de combustión μ es mayor que 52 cm/s o menor que 48 cm/s. Para ello, puede calcularse la probabilidad β de un error tipo II para los valores $\mu = 52$ y $\mu = 48$, y utilizar este resultado para averiguar algo con respecto a la forma en que se desempeñará la prueba. De manera específica, ¿cómo trabajará el procedimiento de prueba si se desea detectar, esto es, rechazar H_o, para un valor medio de $\mu = 52$ ó $\mu = 48$? Dada la simetría, sólo es necesario evaluar uno de los dos casos, esto es, encontrar la probabilidad de aceptar la hipótesis nula H_o: $\mu = 50$ cuando el valor verdadero es $\mu = 52$.

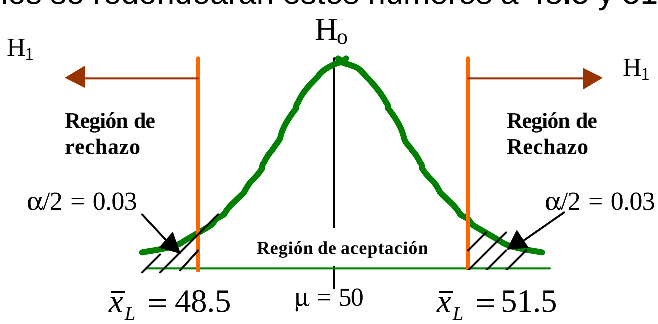
Para hacer este cálculo se tendrá un tamaño de muestra de 10 y una desviación estándar de la población de 2.5 cm/s. Además se evaluará el error tipo II con un nivel de significancia de 0.06.

H_o: $\mu = 50$
H_1: $\mu \neq 50$

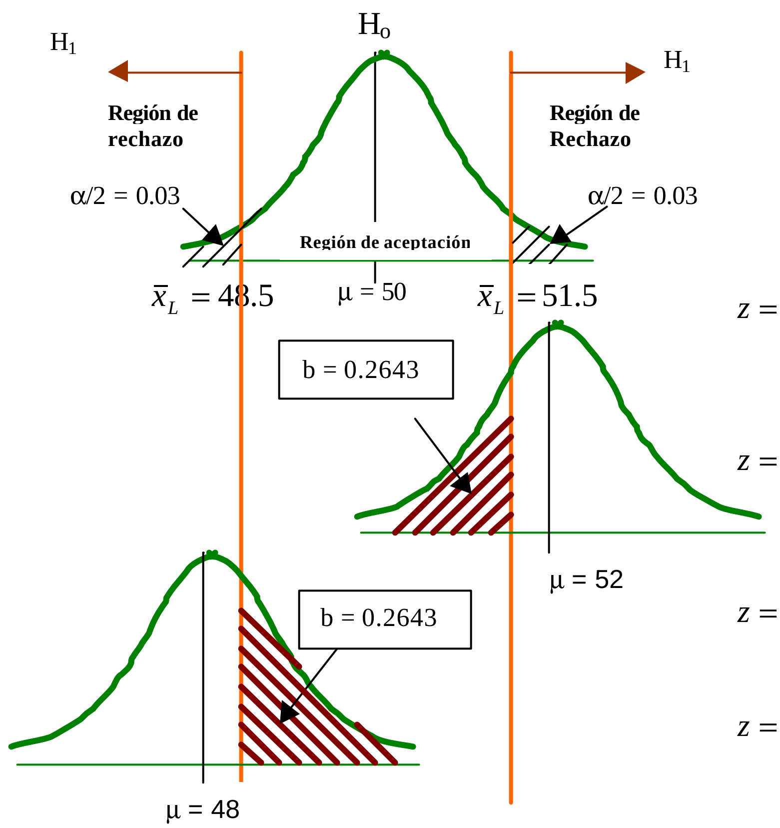
Como ya sabemos se trata de un ensayo bilateral por lo que se tendrá que calcular el valor del estadístico $\bar{x}_L$ de la siguiente manera:

$$\bar{x}_L = m \pm \frac{Z_l s}{\sqrt{n}} = 50 \pm \frac{(1.88)(2.5)}{\sqrt{10}} = 48.51 \text{ y } 51.48$$

Para facilitar los cálculos se redondearán estos números a 48.5 y 51.5



Para poder comprender mejor el cálculo del error tipo II se delimitará el área de la región de aceptación con dos líneas ya que es bilateral y se evaluará la probabilidad de caer en esa área cuando la media tiene un valor de 52 y de 48.



$$z = \frac{48.5 - 52}{2.5/\sqrt{10}} = -4.43$$

$$z = \frac{51.5 - 52}{2.5/\sqrt{10}} = -0.63$$

$$z = \frac{48.5 - 48}{2.5/\sqrt{10}} = 0.63$$

$$z = \frac{51.5 - 48}{2.5/\sqrt{10}} = 4.43$$

Como se puede observar en cada calculo del valor β se tuvieron que evaluar los dos valores de z. En el primer calculo de β se tiene un valor de z=-4.43, esto quiere decir que no existe área del lado izquierdo del 48.5, por lo que β sólo será el área que corresponda a la z=-0.63. Lo mismo pasa con el segundo cálculo de β. Como las medias de 52 y 48 son equidistantes del 50 por este motivo los valores del error tipo II son los mismos.

En caso que no estén equidistantes se tienen que calcular por separado y calcular los valores correspondientes de z porque en ocasiones se tiene un área que no está dentro de la región de aceptación, la cual no se tiene que tomar en cuenta para evaluar al error tipo II.

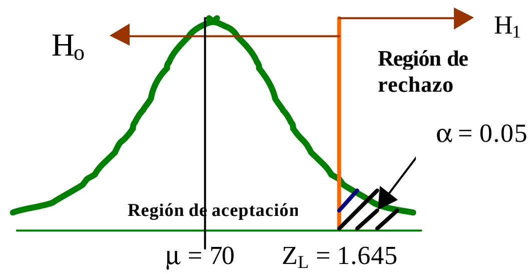
A continuación se procederá a generar algunas curvas características de operación para evaluar al error tipo II, entre más se aleja el valor verdadero de la media de la media de la hipótesis nula, menor es la probabilidad del error tipo II para un tamaño de muestra y nivel de significancia dadas. A medida que el tamaño de la muestra aumenta la probabilidad de cometer el error tipo II disminuye. Esto se observará en los ejercicios siguientes.

Ejemplos:

1. Generar una curva característica de operación para el ejercicio número 1 de la sección de ensayo de hipótesis con las siguientes medias supuestas:
   μ = 70.5, 71, 71.5, 72, 72.5, 73, 73.5, y 74.

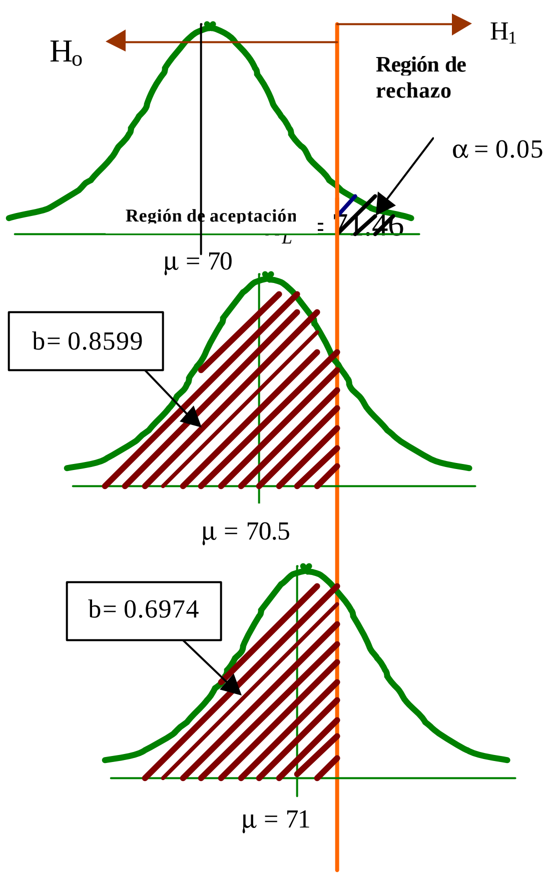
2. Datos:
   μ=70 años
   σ = 8.9 años
   $\bar{x}$ = 71.8 años
   n = 100
   α = 0.05

3. Ensayo de hipótesis
   H_o; μ = 70 años.
   H_1; μ > 70 años.



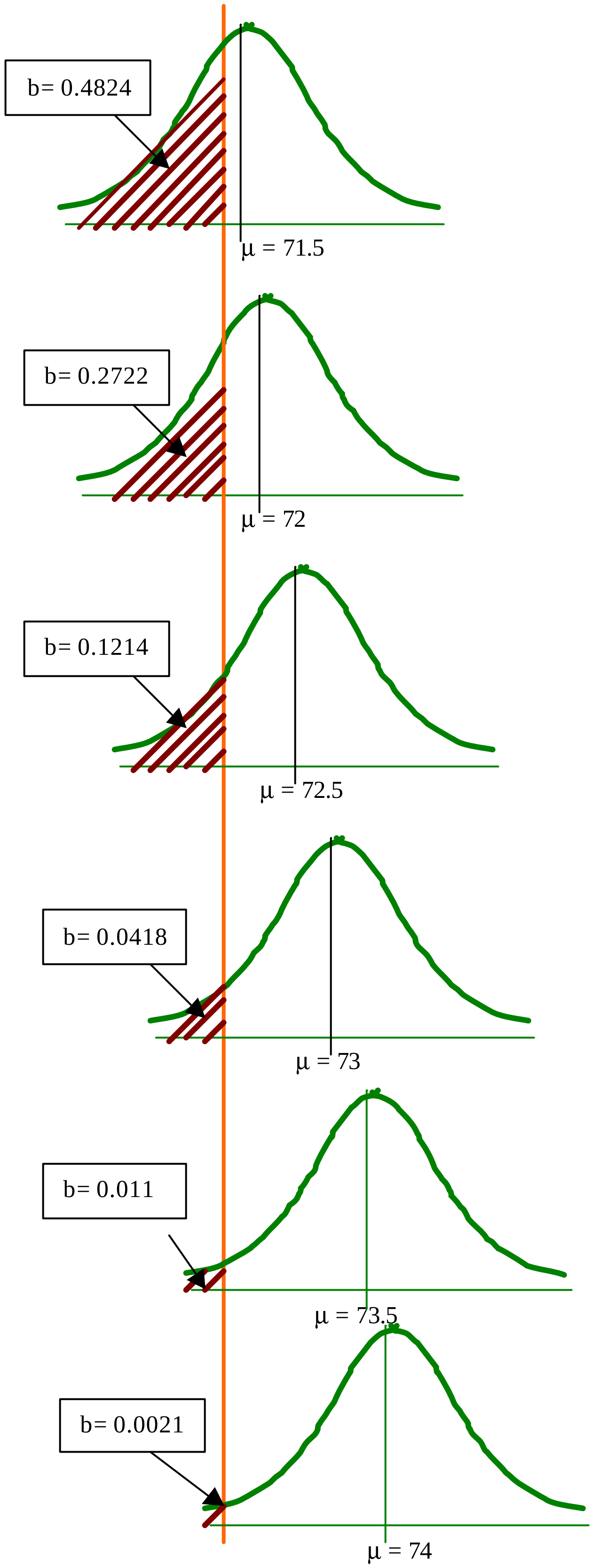
Se calculará el estadístico límite:

$$\bar{x}_L = m + \frac{Z_l s}{\sqrt{n}} = 70 + \frac{(1.645)(8.9)}{\sqrt{100}} = 71.46$$



$$z = \frac{71.46 - 70.5}{8.9/\sqrt{100}} = 1.08$$

$$z = \frac{71.46 - 71}{8.9/\sqrt{100}} = 0.517$$



$$z = \frac{71.46 - 71.5}{8.9 / \sqrt{100}} = -0.044$$

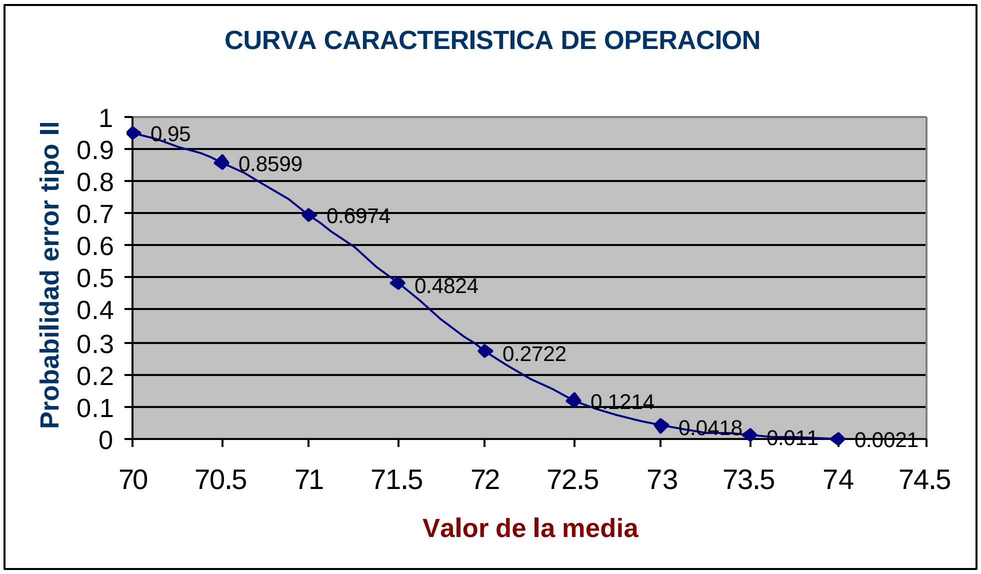
$$z = \frac{71.46 - 72}{8.9 / \sqrt{100}} = -0.606$$

$$z = \frac{71.46 - 72.5}{8.9 / \sqrt{100}} = -1.168$$

$$z = \frac{71.46 - 73}{8.9 / \sqrt{100}} = -1.73$$

$$z = \frac{71.46 - 73.5}{8.9 / \sqrt{100}} = -2.29$$

$$z = \frac{71.46 - 74}{8.9 / \sqrt{100}} = -2.85$$



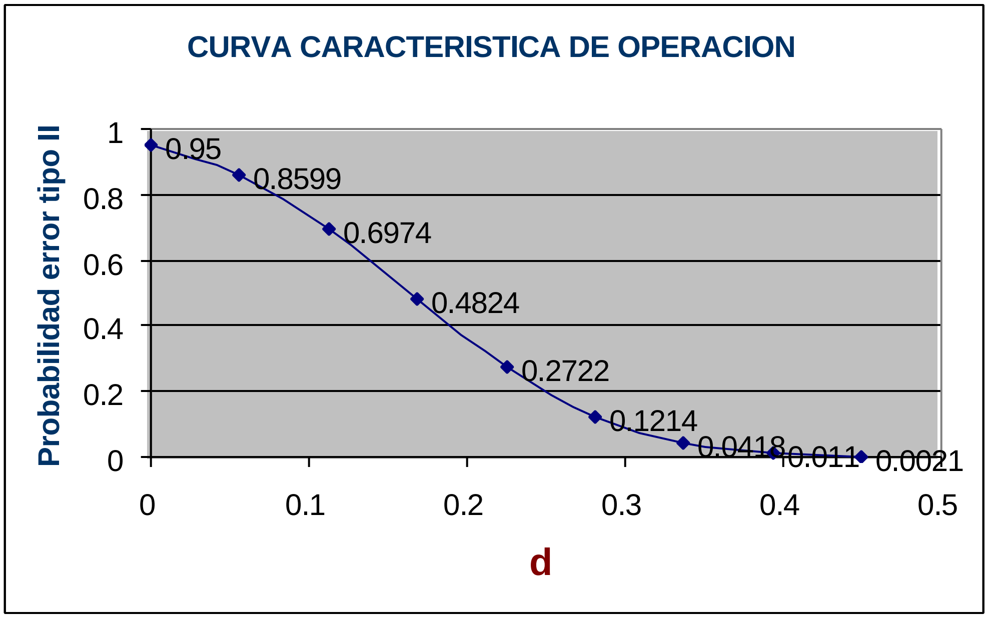
En la mayoría de los libros de estadística existen las curvas características de operación para diferentes tamaños de muestra y éstas se proporcionan tanto para α = 0.05 como para α = 0.01 (son las más comunes). Para poder utilizar las curvas se define un parámetro llamado *d*, que estandariza para cualquier valor de μ y σ:

$$d = \frac{|m - m_o|}{s} = \frac{|d|}{s}$$

Si se quisiera consultar en un libro, ¿cuál es la probabilidad de cometer el error tipo II ó β cuando la media verdadera es de 72?; se tendría que calcular el valor de *d* y buscar en las curvas la que pertenezca a un tamaño de muestra de 100 con un α = 0.05.

$$d = \frac{|72 - 70|}{8.9} = \frac{|2|}{8.9} = 0.2247$$

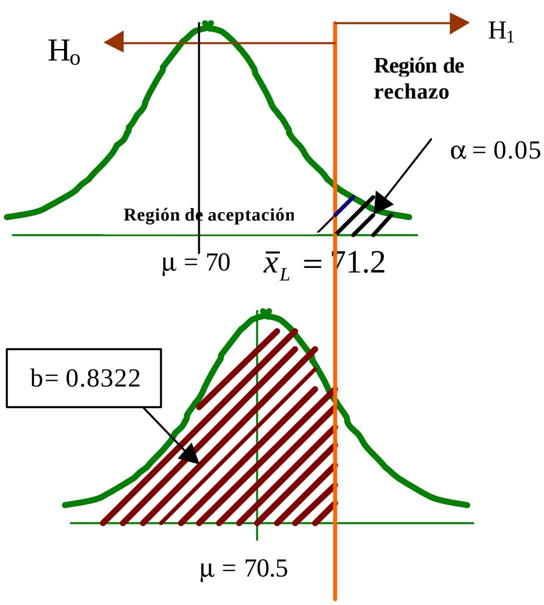
Este valor se encuentra en el eje de las x. Si se transforma la curva característica de operación con el valor de *d* quedaría de la siguiente manera:



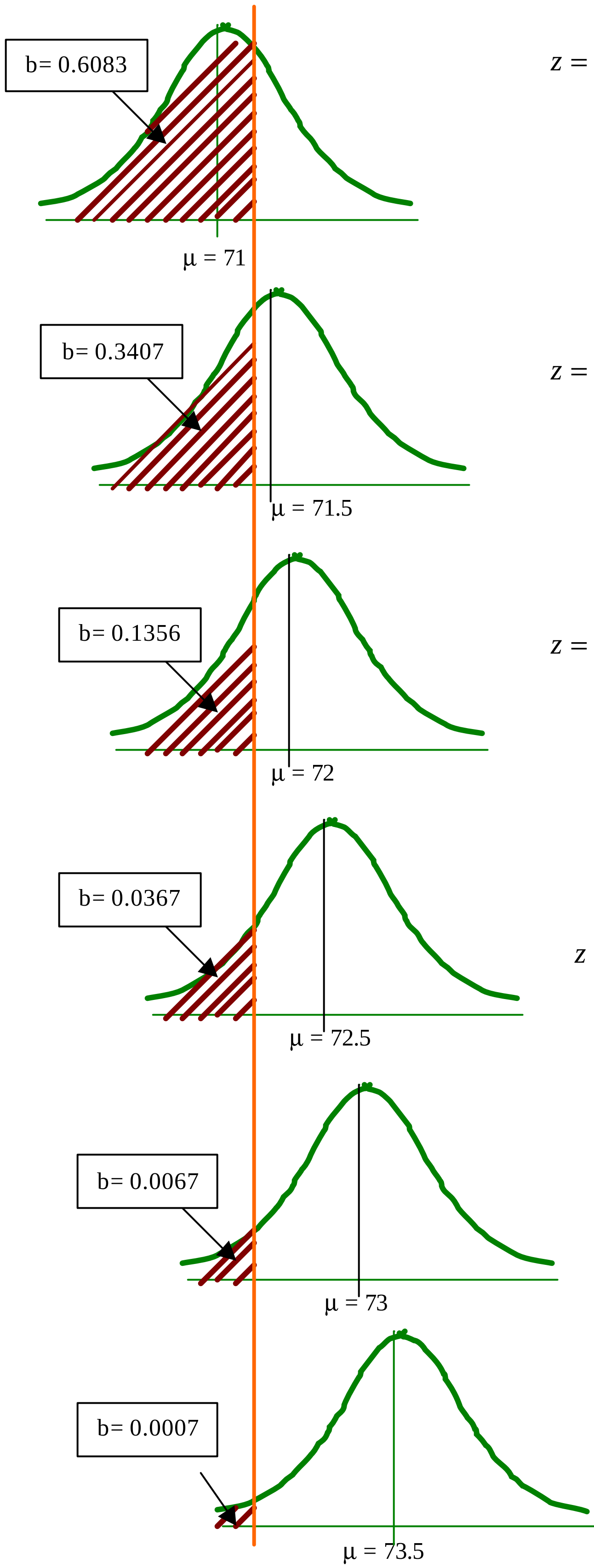
Se comentó anteriormente que si el tamaño de la muestra aumenta los dos tipos de errores α y β disminuyen. Para probar esto y específicamente en lo que se refiere al error tipo II se realizará el ejercicio anterior suponiendo que en lugar de tener 100 personas, el tamaño de la muestra aumenta a 150 personas.

Se calculará el estadístico límite:

$$\bar{x}_L = m + \frac{Z_l s}{\sqrt{n}} = 70 + \frac{(1.645)(8.9)}{\sqrt{150}} = 71.2$$



$$z = \frac{71.2 - 70.5}{8.9/\sqrt{150}} = 0.963$$



$$z = \frac{71.2 - 71}{8.9 / \sqrt{150}} = 0.275$$

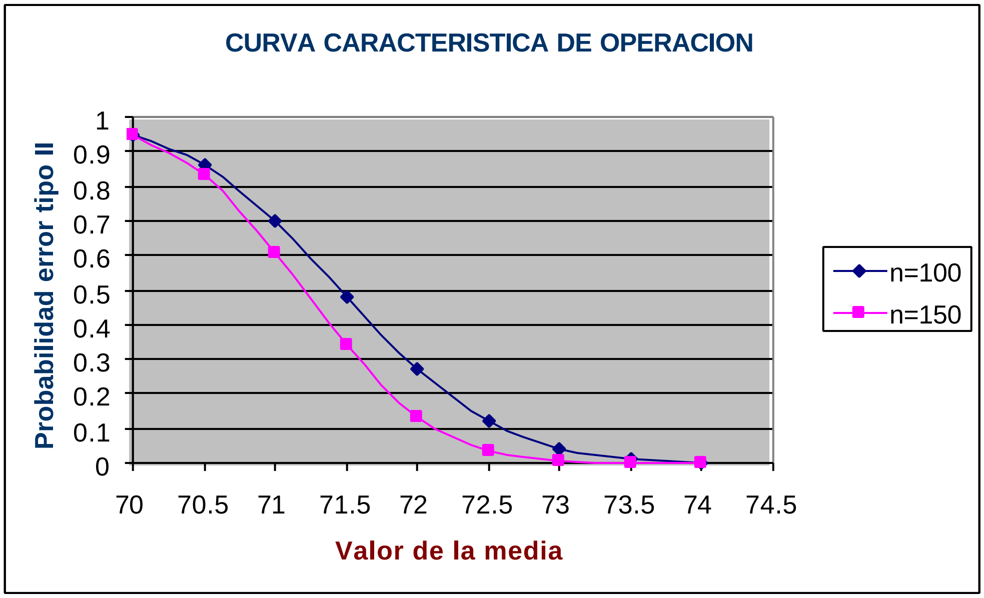
$$z = \frac{71.2 - 71.5}{8.9 / \sqrt{150}} = -0.412$$

$$z = \frac{71.2 - 72}{8.9 / \sqrt{150}} = -1.10$$

$$z = \frac{71.2 - 72.5}{8.9 / \sqrt{150}} = -1.79$$

$$z = \frac{71.2 - 73}{8.9 / \sqrt{150}} = -2.47$$

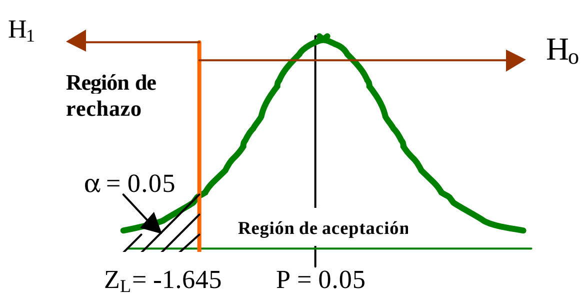
$$z = \frac{71.2 - 73.5}{8.9 / \sqrt{150}} = -3.16$$



3. Generar una curva característica de operación (CCO) para el ejercicio 5 de ensayo de hipótesis. Suponer los siguientes valores de P; 0.04, 0.03, 0.025, 0.02 y 0.01. Enseguida se proporciona la información necesaria para realizar la CCO:

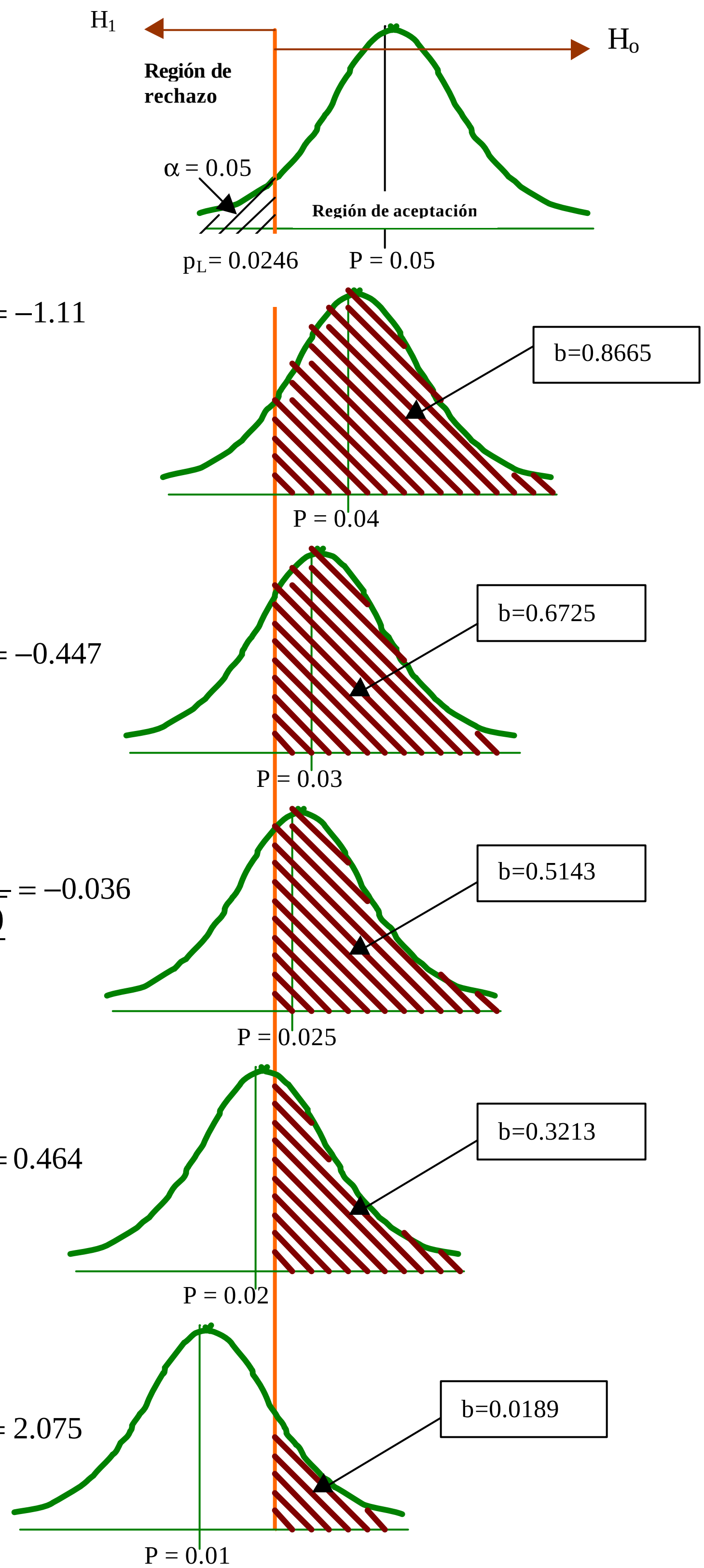
   Datos:
   P= 0.05
   p = 4/200 = 0.02
   n = 200
   α = 0.05

   Ensayo de hipótesis
   H_o; P = 0.05
   H_1; P < 0.05



Solución:
Se procederá a calcular el estadístico límite p_L:

$$p_L = P - z\sqrt{\frac{Pq}{n}} = 0.05 - 1.645\sqrt{\frac{(0.05)(0.95)}{200}} = 0.0246$$



$$z=\frac{0.0246-0.04}{\sqrt{\frac{(0.04)(0.96)}{200}}}=-1.11$$

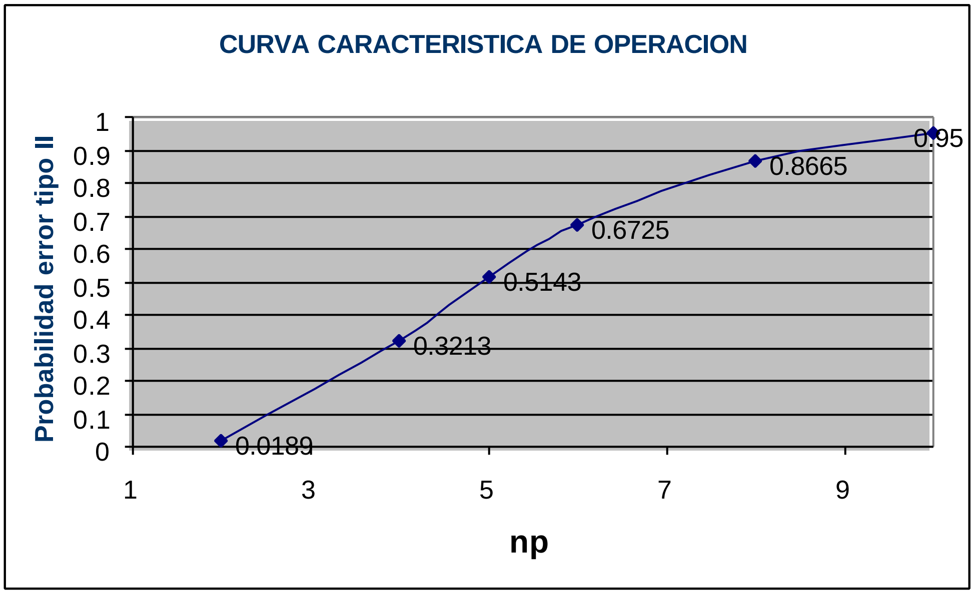
$$z=\frac{0.0246-0.03}{\sqrt{\frac{(0.03)(0.97)}{200}}}=-0.447$$

$$z=\frac{0.0246-0.025}{\sqrt{\frac{(0.025)(0.975)}{200}}}=-0.036$$

$$z=\frac{0.0246-0.02}{\sqrt{\frac{(0.02)(0.98)}{200}}}=0.464$$

$$z=\frac{0.0246-0.01}{\sqrt{\frac{(0.01)(0.99)}{200}}}=2.075$$

En una distribución muestral de proporciones, para graficar la CCO, se necesita calcular el valor de *np*, que es el que irá en el eje de las x para estandarizar la curva.



4. Genere un CCO para el ejercicio número 6 de la sección anterior. Suponga las siguientes diferencias de medias: μ_1-μ_2 =2, 4, 6, 7, 9, 12 y 14.

   Datos:

   $\sigma_1=\sigma_2= 8$

   $\bar{x}_1 = 121 \text{ min}$

   $\bar{x}_2 = 112 \text{ min}$

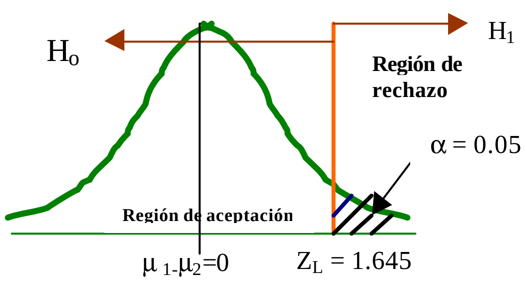
   $n_1=n_2= 10$

   $\alpha = 0.05$

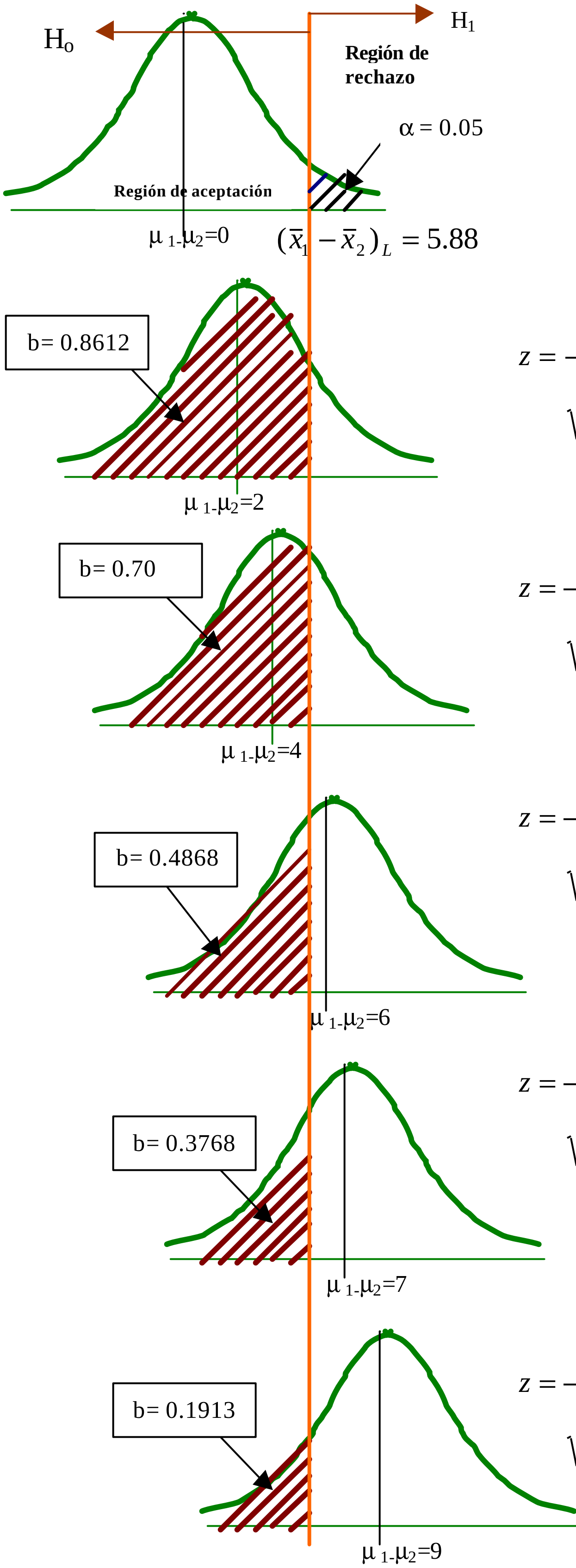
   Ensayo de hipótesis

   H_o; $\mu_1-\mu_2 = 0$

   H_1; $\mu_1-\mu_2 > 0$



$$(\bar{x}_1 - \bar{x}_2)_L = (m_1 - m_2) + z\sqrt{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}} = 0 + 1.645\sqrt{\frac{8^2}{10} + \frac{8^2}{10}} = 5.88$$



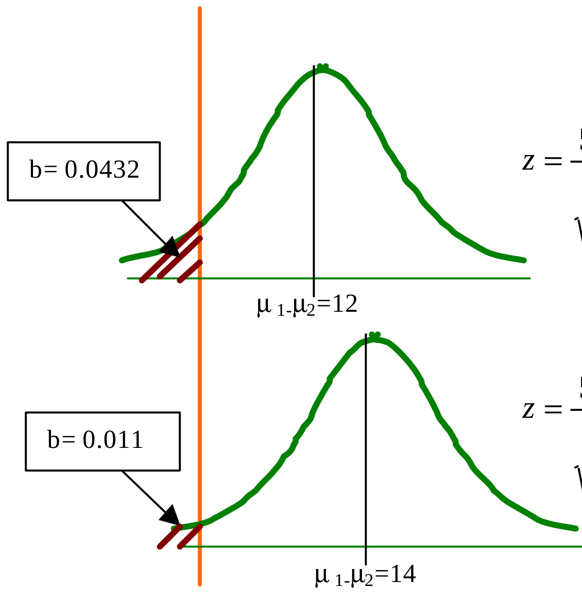
$$z = \frac{5.88 - 2}{\sqrt{\frac{8^2}{10} + \frac{8^2}{10}}} = 1.086$$

$$z = \frac{5.88 - 4}{\sqrt{\frac{8^2}{10} + \frac{8^2}{10}}} = 0.526$$

$$z = \frac{5.88 - 6}{\sqrt{\frac{8^2}{10} + \frac{8^2}{10}}} = -0.033$$

$$z = \frac{5.88 - 7}{\sqrt{\frac{8^2}{10} + \frac{8^2}{10}}} = -0.313$$

$$z = \frac{5.88 - 9}{\sqrt{\frac{8^2}{10} + \frac{8^2}{10}}} = -0.873$$

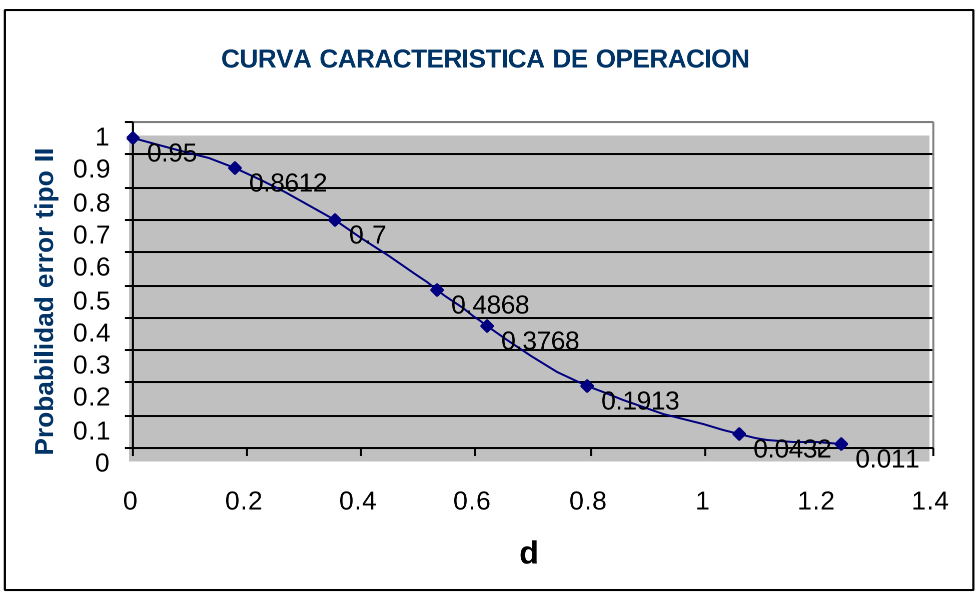


$$z=\frac{5.88-12}{\sqrt{\frac{8^2}{10}+\frac{8^2}{10}}}=-1.714$$

$$z=\frac{5.88-14}{\sqrt{\frac{8^2}{10}+\frac{8^2}{10}}}=-2.274$$

Para graficar la curva se utilizará el valor de *d*, el cual para una distribución muestral de diferencia de medias tiene la siguiente fórmula:

$$d=\frac{|m_1-m_2|}{\sqrt{s_1^2+s_2^2}}=\frac{|d|}{\sqrt{s_1^2+s_2^2}}$$



En los libros de estadística lo que se acostumbra en algunos de los ejercicios es preguntar sólo un punto de la CCO, por lo que a continuación se resolverán dos problemas tipo.

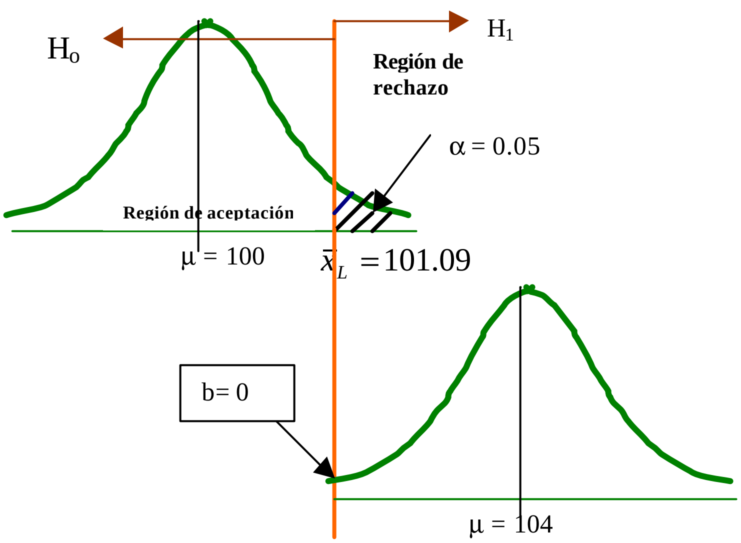
5. Se require que la tensión de ruptura de un hilo utilizado en la fabricación de material de tapicería se al menos de 100 psi. La experiencia ha indicado que la desviación estándar de la tensión de ruptura es de 2 psi. Se prueba una

muestra aleatoria de nueve especímenes, y la tensión de ruptura promedio observada en ella es de 98 psi. ¿Cual es la probabilidad de aceptar la hipótesis nula con un $\alpha = 0.05$ si la tensión promedio de ruptura verdadera de la fibra es 104 psi?

Solución:
Ensayo de hipótesis:
H_o; $\mu = 100$
H_1; $\mu > 100$

Se calcula el estadístico límite:

$$\bar{x}_L = m + \frac{Z_l s}{\sqrt{n}} = 100 + \frac{(1.645)(2)}{\sqrt{9}} = 101.09$$



$$z = \frac{104 - 100}{2/\sqrt{10}} = 6.32$$

6. Del ejercicio número 7 de la sección anterior encontrar el error tipo II ó β suponiendo que la diferencia verdadera entre las medias de las máquinas es fe 0.03

Datos:
$\sigma_1 = 0.020$
$\sigma_2 = 0.025$
$\bar{x}_1 = 16.015$
$\bar{x}_2 = 16.005$
$n_1 = n_2 = 10$
$\alpha = 0.05$

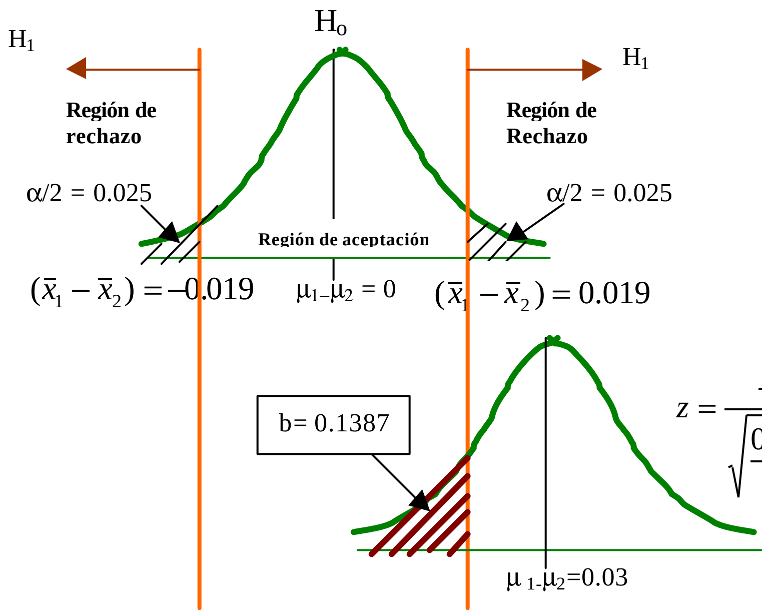
Solución:

Ensayo de hipótesis

H_o; μ_1-μ_2 = 0

H_1; μ_1-$\mu_2 \neq$ 0

$$(\bar{x}_1-\bar{x}_2)_L = (m_1-m_2) \pm z\sqrt{\frac{s_1^{\,2}}{n_1}+\frac{s_1^{\,2}}{n_2}} = 0 \pm 1.96\sqrt{\frac{0.020^2}{10}+\frac{0.025^2}{10}} = \text{-0.019 y 0.019}$$



$$z = \frac{-0.019-0.03}{\sqrt{\frac{0.020^2}{10}+\frac{0.025^2}{10}}} = -4.83$$

$$z = \frac{0.019-0.03}{\sqrt{\frac{0.020^2}{10}+\frac{0.025^2}{10}}} = -1.086$$

Por ser bilateral se calcularon dos valores de z, y como se puede observar del lado izquierdo de –0.019 ya no se encuentra área, por lo que el error tipo II sólo será el área a la izquierda del valor de la diferencia del estadístico límite 0.019.

Problemas propuestos

1. En un estudio para estimar la proporción de residentes de cierta ciudad y sus suburbios que están a favor de la construcción de una planta de energía nuclear, se encuentra que 63 de 100 residentes urbanos están a favor de la construcción mientras que sólo 59 de 125 residentes suburbanos la favorecen. ¿Hay una diferencia significativa entre la proporción de residentes urbanos y suburbanos que favorecen la construcción de la planta nuclear? Use un valor de P para su conclusión.

2. Una compañía petrolera afirma que un quinto de las casas en cierta ciudad se calientan con petróleo. ¿Tenemos razón en dudar de esta afirmación si, en una muestra aleatoria de 1000 casas en esta ciudad, se encuentra que 136 se calientan con petróleo? Utilice un nivel de significancia de 0.01.

3. Se sabe que la duración, en horas, de un foco de 75 watts tiene una distribución aproximadamente normal, con una desviación estándar de 25 horas. Se toma una muestra aleatoria de 20 focos, la cual resulta tener una duración promedio de 1014 horas.
   a) ¿Existe evidencia que apoye la afirmación de que la duración promedio del foco es mayor que 1000 horas? Utilice un $\alpha = 0.05$.
   b) ¿Cual es el valor P para la prueba?
   c) ¿Cuál es el valor de β para la prueba del inciso a) si la verdadera duración promedio del foco es de 1050 horas?

4. Se estudia la tasa de combustión de dos propelentes sólidos utilizados en los sistemas de escape de emergencia de aeroplanos. Se sabe que la tasa de combustión de los dos propelentes tiene aproximadamente la misma desviación estándar de 3 cm/s. Se prueban dos muestras aleatorias de 20 especímenes cada una, obteniéndose medias de 18 y 24 cm/s respectivamente.
   a) Pruebe la hipótesis de que los dos combustibles sólidos tienen la misma rapidez promedio de combustión. Utilice un $\alpha = 0.05$.
   b) ¿Cuál es el valor de P de la prueba?
   c) ¿Cuál es el valor de β para la prueba del inciso a) si la verdadera diferencia en la rapidez promedio de combustión es 2.5 cm/s?

5. Un artículo publicado en *Fortune* afirma que casi la mitad de todos los ingenieros continúan sus estudios académicos después de obtener la licenciatura. Un artículo publicado en *Engineering Horizons* indica que 117 de 484 recién graduados planean continuar sus estudios.
   a) ¿Los datos publicados en *Engineering Horizons* son consistentes con los publicados en *Fortune*?
   b) Encuentre el valor de P de la prueba.

6. En un invierno con epidemia de gripe, una compañía farmacéutica bien conocida estudió 2000 bebes para determinar si la nueva medicina de la compañía era efectiva después de dos días. Entre 120 bebes que tenían gripe y se les administró la medicina, 29 se curaron dentro de dos días. Entre 280 bebés que tenían gripe pero que no recibieron la medicina, 56 se curaron dentro de dos días. ¿Hay alguna indicación significativa que apoye la afirmación de la compañía de la efectividad de la medicina? Calcule el valor P.

7. Se lanza 20 veces una moneda, con un resultado de cinco caras. ¿Esta es suficiente evidencia para rechazar la hipótesis de que la moneda esta balanceada a favor de la alternativa de que las caras ocurren menos de 50% de las veces.? Realice la prueba con un nivel de significancia de 0.03 y cite un valor P.

8. Se supone que los neumáticos para automóvil de cierto tipo recién comprados deben llenarse a una presión de 30 lb/pulg^2. Se representa con μ el verdadero promedio de presión. Encuentre el valor P asociado con cada valor del estadístico z dado para probar H_o; $\mu = 30$ contra H_1; $\mu \neq 30$.
   a) 2.10 b) –1.75 c) –0.55 d) 1.41 e) –5.3

9. Se realizó un experimento para comparar la resistencia a la fractura del acero con níquel maragizado, con el acero de pureza comercial del mismo tipo. Para 32 especímenes, la resistencia promedio muestral fue de 65.6 para el acero de alta pureza, mientras que se obtuvo una media muestral de 59.8 en 38 especímenes del acero comercial. Debido que el acero de alta pureza es más costoso, su uso para cierta aplicación puede justificarse sólo si su resistencia a la fractura excede la del acero de pureza comercial en más de 5. Suponga que ambas distribuciones de resistencias son normales.
   a) Si se supone que σ_1 = 1.2 y σ_2 = 1.1, pruebe las hipótesis pertinentes usando α = 0.001.
   b) Calcule β para la prueba del inciso anterior cuando $\mu_1 - \mu_2 = 6$.

10. Se cree que la portada y la naturaleza de la primera pregunta de encuestas por correo influyen en la tasa de respuesta. Un artículo probó esta teoría al experimentar con diferentes diseños de portadas. Una portada sencilla, y la otra utilizó la figura de un paracaidista. Los investigadores especularon que la tasa de devolución sería menor para la portada sencilla.

Portada	**Número de envíos**	**Número de devoluciones**
Sencilla	207	104
Paracaidista	213	109

   ¿Esta información apoya la hipótesis de los investigadores? Haga la prueba con un nivel de significancia de 0.10, calculando primero un valor P.

Respuesta a los Problemas propuestos

1. z= 2.40; sí, P=0.01
2. P<0.0001; concluir que menos de 1/5 de las casas se calientan con petróleo.
3. a) z = 2.50; se rechaza H_o b) P = 0.0062 c) 0
4. a) Se Rechaza H_o, z= -6.32 b) 0 c) 0.248
5. a) Se rechaza H_o, z= -11.36 b) valor P = 0
6. No se rechaza H_o, z= 0.93, valor de P = 0.1762
7. Rechazar H_o. Valor P = 0.0207
8. a) 0.0358 b) 0.0802 c) 0.5824 d) 0.1586 e) 0
9. a) z=2.89, no se debe usar el acero de alta pureza o se no se rechaza H_o.
   b) 0.2981
10. Valor P = 0.4247, no se rechaza H_o.