UNIDAD 2

REVISIÓN DE CONCEPTOS ESTADÍSTICOS
TEMA 2: EXPERIMENTOS COMPARATIVOS SIMPLES:
MUESTREO Y DISTRUBUCIONES DE MUESTREO, ESTIMACIÓN POR INTERVALO Y PRUEBA DE HIPÓTESIS

Cuando medir cada elemento es imposible; tomar una muestra.






Juan Pablo Sucre Reyes

Cuando medir cada elemento es imposible; tomar una muestra.

Juan Pablo Sucre Reyes

Cuando medir cada elemento es imposible; tomar una muestra.

- *RECUERDE:* Elemento = entrada en la que se recolectan los datos. Población = conjunto de todos los elementos de interés. Muestra = un subconjunto de la población.
- Se selecciona una muestra para recabar datos y realizar una inferencia con el fin de responder a una pregunta de investigación acerca de una población (*estadísticos muestrales como estimaciones de parámetros poblacionales*).
- La población muestreada es aquella de la cual se extrae la muestra, y un marco es la lista de los elementos de donde se seleccionará la muestra.

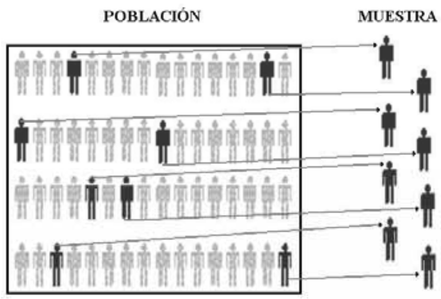
Juan Pablo Sucre Reyes

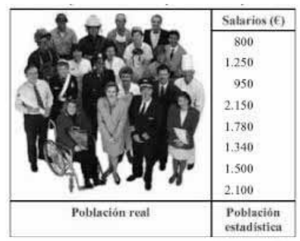
1. El problema de muestreo de Electronics Associates

•*Ejemplo: Al director de personal de Electronics Associates, Inc. (EAI) se le ha encargado elaborar un perfil de los 2500 gerentes de la empresa: sueldo medio anual y la proporción de ellos que ha completado un programa de capacitación.*

•*Solución:* Utilizando los 2 500 gerentes de la empresa se determinan los parámetros poblacionales tales como: $(\mu = \$51\,800)(\sigma = \$4\,000)\ (p = 0.60)$

• Ahora suponga que la información necesaria acerca de los 2500 gerentes no está disponible. Asuma que se empleará una muestra de ¿30 gerentes? (*menor tiempo y costo de recolectar información).*





Juan Pablo Sucre Reyes

2. Selección de una muestra

• *Muestreo de una población finita:*

MUESTREO ALEATORIO SIMPLE (POBLACION FINITA)

Una **muestra aleatoria simple** de tamaño n de una población finita de tamaño N es una muestra seleccionada de manera que cada posible muestra de tamaño n tenga la misma probabilidad de ser seleccionada.

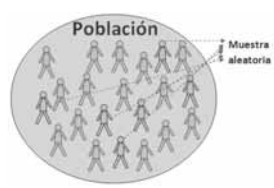
• Procedimiento: elegir los elementos para la muestra de uno en uno, así en cada paso, cada uno de los elementos que quedan tiene la misma probabilidad de ser elegido *(tabla de números aleatorios/software).*

•*Ejemplo: Listados los gerentes del 1 al 2500, se seleccionan números aleatorios de 4 dígitos hasta completar los 30 gerentes seleccionados.*

63271	59986	71744	51102	15141	80714	58683	93108	13554	79945
88547	09896	95436	79115	08303	01041	20030	63754	08459	28364
55957	57243	83865	09911	19761	66535	40102	26646	60147	15702
46276	87453	44790	67122	45573	84358	21625	16999	13385	22782

• Cualquier número aleatorio que ya ha sido usado se ignora (N° de gerente ya incluido en la muestra): *muestreo sin remplazo*.*

• Si se aceptan números aleatorios ya usados (gerentes respectivos incluidos dos o más veces): *muestreo con remplazo.*



Juan Pablo Sucre Reyes

2. Selección de una muestra

• *Muestreo de una población infinita:*

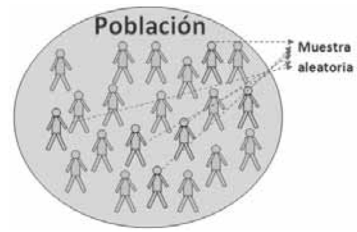
MUESTRA ALEATORIA (POBLACIÓN INFINITA)

Una **muestra aleatoria** de tamaño n de una población infinita es seleccionada de manera tal que se satisfagan las condiciones siguientes.

1. Cada elemento elegido proviene de la misma población.
2. Cada elemento es seleccionado de manera independiente.

• Población infinitamente grande o sus elementos están siendo generados por un proceso en marcha *(número infinito de elementos generados, no existe lista)*

• Poblaciones infinitas suelen asociarse con un proceso que opera continuamente a lo largo del tiempo *(partes fabricadas en una línea de producción, repetidas pruebas experimentales en un laboratorio, transacciones en un banco, llamadas que llegan a un centro de asesoría técnica y clientes que entran en una tienda).*



3. Estimación puntual

• Para estimar el valor de un parámetro poblacional se calcula la característica correspondiente de la muestra, ó *estadístico muestral.*

• Al valor numérico obtenido de $\overline{x}$ (estimación de μ), S (estimación de σ), o $\overline{p}$ (estimación de p) se le conoce como estimación puntual.

•*Ejemplo: Estimar la media poblacional μ y la desviación estándar poblacional σ de los sueldos anuales de gerentes de EAI; además de la proporción p de capacitados.*

Sueldo anual y situación respecto del programa de capacitación para una muestra

Sueldo anual ($)	Programa de capacitación	Sueldo anual ($)	Programa de capacitación
x_1 = 49094.30	Sí	x_{16} = 51766.00	Sí
x_2 = 53263.90	Sí	x_{17} = 52541.30	No
x_3 = 49643.50	Sí	x_{18} = 44980.00	Sí
x_4 = 49894.90	Sí	x_{19} = 51932.60	Sí
x_5 = 47621.60	No	x_{20} = 52973.00	Sí
x_6 = 55924.00	Sí	x_{21} = 45120.90	Sí
x_7 = 49092.30	Sí	x_{22} = 51753.00	Sí
x_8 = 51404.40	Sí	x_{23} = 54391.80	No
x_9 = 50957.70	Sí	x_{24} = 50164.20	No
x_{10} = 55109.70	Sí	x_{25} = 52973.60	No
x_{11} = 45922.60	Sí	x_{26} = 50241.30	No
x_{12} = 57268.40	No	x_{27} = 52793.90	No
x_{13} = 55688.80	Sí	x_{28} = 50979.40	Sí
x_{14} = 51564.70	No	x_{29} = 55860.90	Sí
x_{15} = 56188.20	No	x_{30} = 57309.10	No

Se calculan *estadísticos muestrales*:

$$\overline{x} = \frac{\Sigma x_i}{n} = \frac{1\,554\,420}{30} = \$51\,814$$

$$s = \sqrt{\frac{\Sigma(x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{325\,009\,260}{29}} = \$3\,348$$

Sea x = # de gerentes en n=30 que fueron capacitados, se tiene:

$$\overline{p} = \frac{x}{n} = \frac{19}{30} = 0.63$$

Resumen de las estimaciones puntuales obtenidas

Parámetro poblacional	Valor del parámetro	Estimador puntual	Estimación puntual
μ = Media poblacional de los sueldos anuales	$51 800	$\overline{x}$ = Media muestral de los sueldos anuales	$51 814
σ = Desviación estándar poblacional de los sueldos anuales	$4 000	s = Desviación estándar muestral de los sueldos anuales	$3 348
p = Proporción poblacional que ha completado el programa de capacitación	0.60	$\overline{p}$ = Proporción muestral que ha completado el programa de capacitación	0.63

3.1 Estimación puntual: consejo práctico

- La estimación puntual es una de las formas de la *inferencia estadística*, que usa un estadístico muestral para hacer una inferencia acerca de un parámetro poblacional.
- Al inferir acerca de una población basada en una muestra, se debe tener una correspondencia cerrada entre la población muestreada *(de la cual se toma realmente la muestra)* y la población objetivo *(de la cual se busca hacer inferencias)*.
- Lo deseado es que la población muestreada y la población objetivo sean idénticas. A veces, no es fácil obtener una correspondencia cerrada entre ambas poblaciones.

Juan Pablo Sucre Reyes

4. Introducción a las distribuciones muestrales o de muestreo

•*Ejemplo: Suponga que el proceso de seleccionar una muestra aleatoria simple de 30 gerentes de EAI se repite una y otra vez (500 muestras), y en cada una se calculan los valores de $\overline{x}$ y de $\overline{p}$. Se tiene así:*

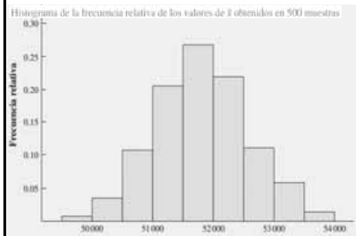
Valores de $\bar{x}$ y de $\bar{p}$ obtenidos en 500 muestras

Muestra número	Media muestral ($\bar{x}$)	Proporción muestral ($\bar{p}$)
1	51 814	0.63
2	52 670	0.70
3	51 780	0.67
4	51 588	0.53
.	.	.
.	.	.
.	.	.
500	51 752	0.50

Distribuciones de frecuencia y de frecuencia relativa de $\bar{x}$ en 500 muestras

Sueldo anual medio ($)	Frecuencia	Frecuencia relativa
49 500.00–49 999.99	2	0.004
50 000.00–50 499.99	16	0.032
50 500.00–50 999.99	52	0.104
51 000.00–51 499.99	101	0.202
51 500.00–51 999.99	133	0.266
52 000.00–52 499.99	110	0.220
52 500.00–52 999.99	54	0.108
53 000.00–53 499.99	26	0.052
53 500.00–53 999.99	6	0.012
Totals	500	1.000

Histograma de la frecuencia relativa de los valores de $\bar{x}$ obtenidos en 500 muestras



- Dado que la media muestral $\overline{x}$ es una variable aleatoria $\Rightarrow$ tiene una media o valor esperado, una desviación estándar y una distribución de probabilidad = *distribución de muestreo de x*.
- Conocer esta distribución (propiedades) permitirá hacer declaraciones de probabilidad sobre la cercanía de la media muestral $\overline{x}$ a la media poblacional μ.

•*NOTA: Idem para p. En la práctica sólo se selecciona una muestra aleatoria simple de la población*

Juan Pablo Sucre Reyes

5. Distribución de muestreo de $\overline{x}$

DISTRIBUCIÓN DE MUESTREO DE $\bar{x}$

La distribución muestral de $\bar{x}$ es la distribución de probabilidad de todos los posibles valores de la media muestral $\bar{x}$.

- Tiene un valor esperado o media, una desviación estándar y una forma característica.
- *Valor esperado de $\overline{x}$:* la media de todos los valores posibles de $\overline{x}$ (diferentes muest*ras)*

VALOR ESPERADO DE $\bar{x}$

$$E(\bar{x}) = \mu$$

donde:

$E(\bar{x})$ = valor esperado de $\bar{x}$

μ = media poblacional

- Si el valor esperado de un estimador puntual ($\overline{x}$)es igual al parámetro poblacional (μ), dicho estimador puntual es *insesgado.*
- *Desviación estándar de $\overline{x}$:* factor de corrección $\sqrt{(N-n)/(N-1)}$ *≅ 1; por lo que:*

DESVIACIÓN ESTÁNDAR DE $\bar{x}$

Población finita	Población infinita
$\sigma_{\bar{x}} = \sqrt{\dfrac{N-n}{N-1}}\left(\dfrac{\sigma}{\sqrt{n}}\right)$	$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

$\sigma_{\bar{x}}$ = desviación estándar de $\bar{x}$

σ = desviación estándar de la población

n = tamaño de la muestra

N = tamaño de la población

USAR LA EXPRESIÓN SIGUIENTE PARA CALCULAR LA DESVIACIÓN ESTÁNDAR DE $\bar{x}$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

siempre que

1. La población sea infinita; o
2. La población sea finita *y* el tamaño de la muestra sea menor o igual a 5% del tamaño de la población; es decir, $n/N \le 0.05$.

- *σ_x = error estándar de la media (desviación estándar de un estimador).*

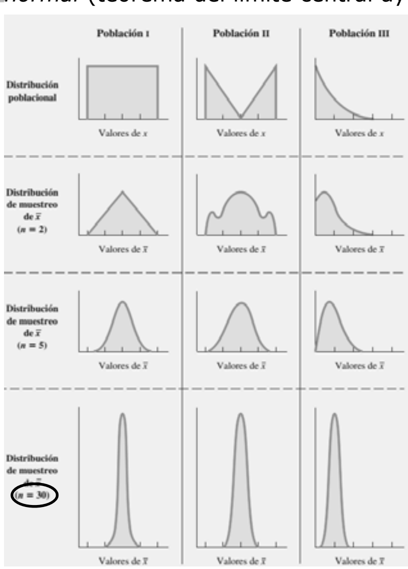
- *Ejemplo: Valor esperado y desviación estándar de $\overline{x}$ para EIA.*
- *Solución: el sueldo anual medio de gerentes de EAI es μ = 51800 ⇒ E($\overline{x}$) = 51800$.*
- *Para N=2500 gerentes (población finita) de EAI; σ=4000 $ y como n=30, se tiene n/N=30/2500= 0,012 (<5% de N), entonces:*

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{30}} = 730.3$$

Juan Pablo Sucre Reyes

5. Distribución de muestreo de $\overline{x}$

- *Forma de la distribución de muestreo de $\overline{x}$:* Se consideran dos casos: *1) La población tiene distribución normal* (cualquiera sea n); y *2) La población no tiene distribución normal* (teorema del límite central ayuda a determinar dicha forma)

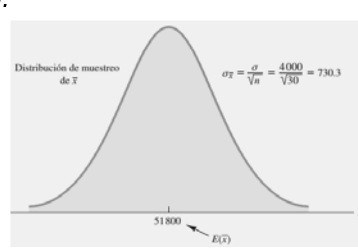


TEOREMA DEL LÍMITE CENTRAL

Cuando se seleccionan muestras aleatorias simples de tamaño n de una población, la distribución de muestreo de la media muestral $\bar{x}$ puede aproximarse mediante una *distribución normal* a medida que el tamaño de la muestra se hace grande.

- *Práctica estadística general:* la distribución de muestreo de $\overline{x}$ se puede aproximar mediante una distribución normal siempre que n ≥ 30.

- *Ejemplo: Forma de la distribución de muestreo de $\overline{x}$ para EIA.*
- *Solución: Dado el caso 2) y como n=30 gtes. ⇒ puede aproximarse mediante una distribución normal.*



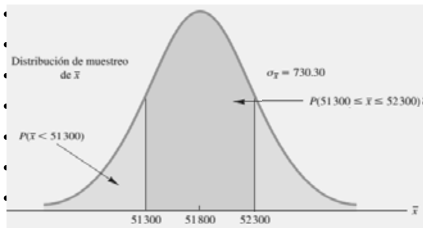
Juan Pablo Sucre Reyes

5.1 Valor práctico de la distribución de muestreo de $\bar{x}$

• Se usa para proporcionar información probabilística acerca de la diferencia entre la media muestral *x* y la media poblacional μ (*no son exactamente iguales*).

•*Ejemplo: Hallar la probabilidad de que la media muestral obtenida usando una muestra aleatoria simple de 30 gerentes de EAI se encuentre en un margen de $500 de la media poblacional.*

•*Solución:* Dado *μ = $ 51800, se desea calcular P(51300* $\leq \bar{x} \leq$ *52300)=?. Al estar distribuida normalmente, se usa* μ *y* σ *poblacionales para: E(*$\bar{x}$*)=*μ *;* σ_x*= 730,3 para estandarizar (z):*



$z = \dfrac{52300 - 51800}{730.30} = 0.68$ *y* $z = \dfrac{51300 - 51800}{730.30} = -0.68$

de donde por tabla:

P(51300 $\leq \bar{x} \leq$ *52300)= P(z* $\leq$ *0,68)-P(z* $\leq$ *-0,68) = 0,7517 - 0,2483 = 0,5034.*

Así también la probabilidad de que la diferencia entre $\bar{x}$ *y u > $500 es 1-0,5034 = 0,4966*

• *Dado que son valores cercanos, tal vez sea necesario un n mayor.*

Juan Pablo Sucre Reyes

5.2 Relación entre el tamaño de la muestra y la distribución de muestreo de $\bar{x}$

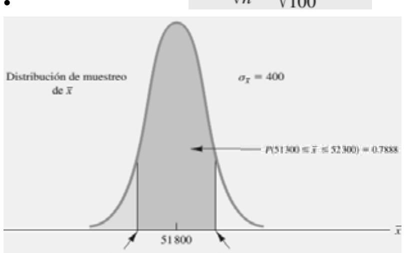
• E($\bar{x}$)= μ independientemente del tamaño de la muestra (n). Pero, en el error estándar de la media, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, siempre que n aumenta, éste disminuye.

• Una muestra de mayor tamaño proporciona mayor probabilidad de que la media muestral $\bar{x}$ esté dentro de una distancia determinada de la media poblacional μ.

•*Ejemplo: Hallar la probabilidad de que la media muestral obtenida usando una muestra aleatoria simple de 100 gerentes de EAI se encuentre en un margen de $500 de la media poblacional.*

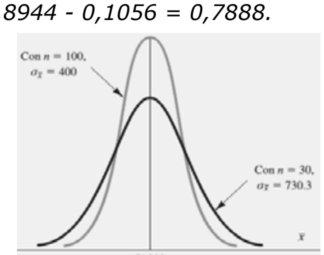
•*Solución:* Dado *μ = $ 51800, se desea calcular P(51300* $\leq \bar{x} \leq$ *52300)=?. Al estar distribuida normalmente, se usa* μ *y* σ *poblacionales para: E(x)=*μ *;* σ_x*= 400 para estandarizar (z):* $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{4000}{\sqrt{100}} = 400$

$z = \dfrac{52300 - 51800}{400} = 1.25$ *y* $z = \dfrac{51300 - 51800}{400} = -1.25$



de donde por tabla:

P(51300 $\leq x \leq$ *52300)= P(z* $\leq$ *1,25)-P(z* $\leq$ *-1,25) = 0,8944 - 0,1056 = 0,7888.*



Juan Pablo Sucre Reyes

6. Distribución de muestreo de $\overline{p}$

DISTRIBUCIÓN DE MUESTREO DE $\overline{p}$

La distribución de muestreo de $\overline{p}$ es la distribución de probabilidad de todos los posibles valores de la proporción muestral $\overline{p}$.

$$\overline{p} = \frac{x}{n}$$

- Tiene un valor esperado o media, una desviación estándar y una forma característica.
- *Valor esperado de $\overline{p}$:* la media de todos los valores posibles de $\overline{x}$ (diferentes mues*tras)*

VALOR ESPERADO DE $\overline{p}$

$$E(\overline{p}) = p$$

donde

$E(\overline{p})$ = valor esperado de $\overline{p}$

p = proporción poblacional

- Como *$E(\overline{p}) = p$*, $\overline{p}$ es un *estimador insesgado* de *p*.
- *Desviación estándar de $\overline{p}$:* factor de corrección $\sqrt{(N-n)/(N-1)} \cong$ *1; por lo que:*

DESVIACIÓN ESTÁNDAR DE $\overline{p}$

Población finita

$$\sigma_{\overline{p}} = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{p(1-p)}{n}}$$

Población infinita

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}}$$

USAR LA EXPRESIÓN SIGUIENTE PARA CALCULAR LA DESVIACIÓN ESTÁNDAR DE $\overline{p}$

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}}$$

siempre que

1. La población sea infinita; o
2. La población sea finita y el tamaño de la muestra sea menor o igual a 5% del tamaño de la población; es decir, $n/N \leq 0.05$.

- *σ_p = error estándar de la proporción (desviación estándar de un estimador).*

- *Ejemplo: Valor esperado y desviación estándar de $\overline{p}$ para EIA.*
- *Solución: la proporción de gerentes capacitados de EAI es p = 0,60 $\Rightarrow$ $E(\overline{p})$ = 0,60.*
- *Para N=2500 gerentes (población finita) de EAI; como n=30, se tiene n/N=30/2500= 0,012 (<5% de N), entonces:*

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.60(1-0.60)}{30}} = 0.0894$$

Juan Pablo Sucre Reyes

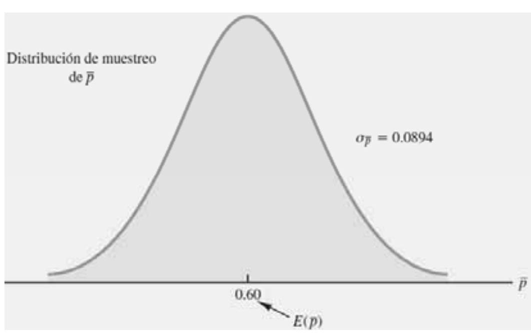
6. Distribución de muestreo de $\overline{p}$

- *Forma de la distribución de muestreo de $\overline{p}$:* la distribución de muestreo de $\overline{p}$ es una distribución de probabilidad discreta binomial: *probabilidad de cada x/n = la de x*.

La distribución de muestreo de $\overline{p}$ se aproxima mediante una distribución normal, siempre que $np \geq 5$ y $n(1-p) \geq 5$.

- *Práctica estadística general:* al estimar p, n es suficientemente grande para permitir usar la aproximación normal para la distribución de muestreo de *$\overline{p}$.*

- *Ejemplo: Forma de la distribución de muestreo de p para EIA.*
- *Solución: Dado n=30 gtes. y p = 0,60 $\Rightarrow$ np =30(0,60)= 18 y n(1-p)=30(0,40)= 12 $\Rightarrow$ puede aproximarse mediante una distribución normal.*



Juan Pablo Sucre Reyes

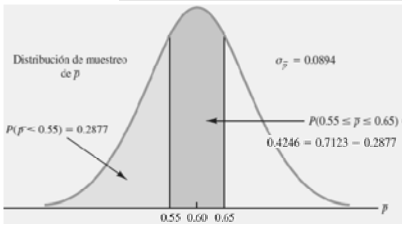
6.1 Valor práctico de la distribución de muestreo de $\bar{p}$

• Se usa para proporcionar información probabilística acerca de la diferencia entre la proporción muestral $\bar{p}$ y la proporción poblacional p (*no son exactamente iguales*).

•*Ejemplo: Hallar la probabilidad de que la proporción muestral obtenida usando una muestra aleatoria simple de 30 gerentes de EAI se encuentre en un margen de 0,05 de proporción poblacional de los gerentes capacitados.*

•*Solución:* Dado *p = 0,60, se desea calcular* $P(0{,}55 \leq \bar{p} \leq 0{,}65)$*=?. Al estar distribuida normalmente, se usa p poblacional para:* $E(\bar{p})=p$; $\sigma_p = 0{,}0894$ *para estandarizar (z):*

• $z = (0.65 - 0.60)/0.0894 = 0.56.$ *y* $z = (0.55 - 0.60)/0.0894 = -0.56.$



de donde por tabla:
$P(0{,}55 \leq \bar{p} \leq 0{,}65) = P(z \leq 0{,}56) - P(z \leq -0{,}56)$
$= 0{,}7123 - 0{,}2877 = 0{,}4246.$
Así también la probabilidad de que la diferencia entre $\bar{x}$ *y u > \$500 es 1-0,5034 = 0,4966*

•*Ejemplo variación n: Si se aumenta el tamaño n=100, el error estándar de la proporción* $\bar{p}$ *cambia. Así:*

$$\sigma_{\bar{p}} = \sqrt{\frac{0.60(1-0.60)}{100}} = 0.049$$

•*Siendo los valores z:* $z = (0.65 - 0.60)/0.049 = 1.02.$ *y* $z = (0.55 - 0.60)/0.049 = -1.02.$

•*de donde por tabla:* $P(0{,}55 \leq p \leq 0{,}65) = P(z \leq 1{,}02) - P(z \leq -1{,}02) = 0{,}8461 - 0{,}1539 = 0{,}6922.$ *(a mayor n mayor probabilidad)*

Juan Pablo Sucre Reyes

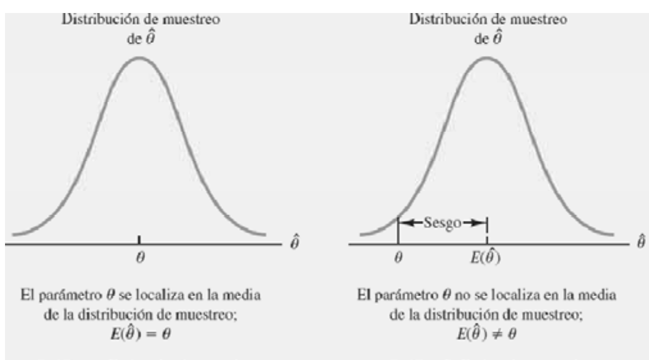
7. Propiedades de los estimadores puntuales

• Antes de usar un estadístico muestral como estimador puntual, se verifica si éste tiene ciertas propiedades (3) que corresponden a un buen estimador puntual. Sea:

θ = parámetro poblacional de interés
$\hat{\theta}$ = estadístico muestral o estimador puntual de θ

•*1. Insesgadez:* el valor esperado del estadístico muestral = parámetro poblacional.

El estadístico muestral $\hat{\theta}$ es un estimador insesgado del parámetro poblacional θ si
$$E(\hat{\theta}) = \theta$$

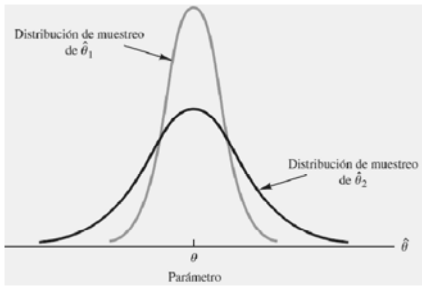


Gráfica *A*. Estimador insesgado **Gráfica *B*. Estimador sesgado**

Juan Pablo Sucre Reyes

7. Propiedades de los estimadores puntuales

•*2. Eficiencia:* al usar una muestra aleatoria simple de n elementos para obtener dos estimadores puntuales insesgados de un mismo parámetro poblacional; se preferirá aquel con el menor error estándar (estimaciones más cercanas al parámetro poblacional) = mayor eficiencia relativa.



•*3. Consistencia:* un estimador puntual es consistente si su valor tiende a estar más cerca del parámetro poblacional a medida que n aumenta. Un n grande tiende a proporcionar mejor estimación puntual que un n pequeño.

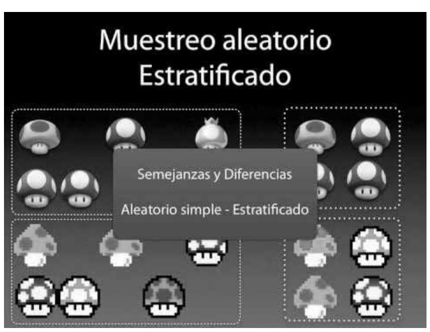
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

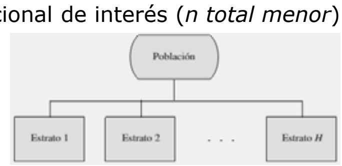
8. Otros procedimientos de muestreo

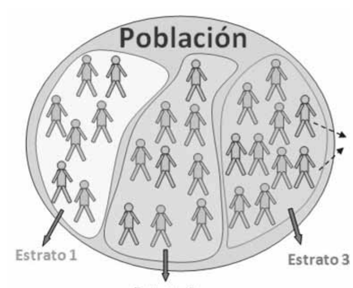
• Al estudiar las distribuciones de muestreo de $\bar{x}$ y de $\bar{p}$ se ha usado el muestreo aleatorio simple. Sin embargo, no es el único método *(probabilístico)* que existe.

•*1. Muestreo aleatorio estratificado:* los elementos de la población primero se dividen en grupos ó *estratos* (base elegida a discreción) con elementos lo más homogéneos. Luego se toma una muestra aleatoria simple de cada estrato.

• Existen fórmulas para combinar los resultados de las muestras de varios estratos individuales en una estimación del parámetro poblacional de interés (*n total menor*).





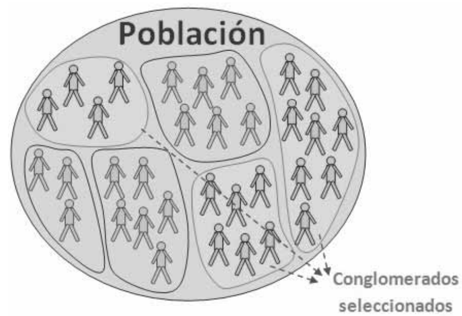


8. Otros procedimientos de muestreo

•*2. Muestreo por conglomerados:* los elementos de la población primero se dividen en grupos separados: conglomerados o clusters. Se toma una muestra aleatoria simple de los conglomerados (*elementos en cada conglomerado muestreado forman la muestra*).

• Mejores resultados si los elementos dentro los conglomerados no son semejantes (conglomerado = *representación a pequeña escala, de* la *N* completa).

• Ideal para muestreo de áreas (n total mayor, pero a un menor costo).

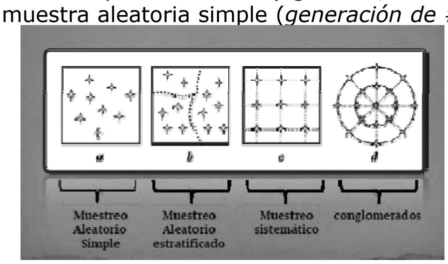


Juan Pablo Sucre Reyes

8. Otros procedimientos de muestreo

•*3. Muestreo sistemático:* el primer elemento que se selecciona es elegido al azar (*muestra aleatoria simple*) y luego se muestrea uno de cada N/n elementos de la población *(lista de los elementos de N en orden aleatorio).*

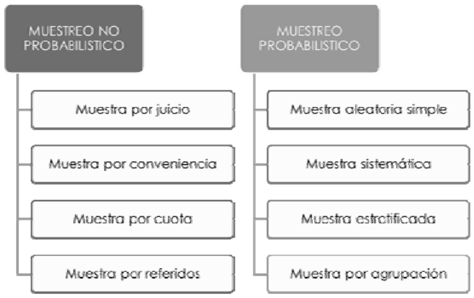
• Útil con poblaciones muy grandes, donde se necesita mucho tiempo para tomar una muestra aleatoria simple (*generación de #s aleatorios, selección de la lista*).

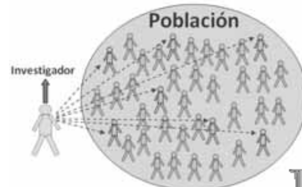


Juan Pablo Sucre Reyes

8.1 Otros procedimientos de muestreo no probabilístico

- No permite plantear afirmaciones probabilísticas acerca del error asociado con el uso de los resultados muestrales al hacer inferencias de la población.
- *4. Muestreo por conveniencia:* Los elementos se incluyen sin tener una probabilidad previamente especificada o conocida de que sean incorporados en n. Es más fácil seleccionarla y recabar datos; pero imposible evaluar su representatividad (N).
- *5. Muestreo subjetivo:* un experto en un asunto selecciona elementos de la población a los que considera los más representativos. Manera fácil de seleccionar una muestra, pero la calidad de los resultados muestrales depende del juicio del experto.



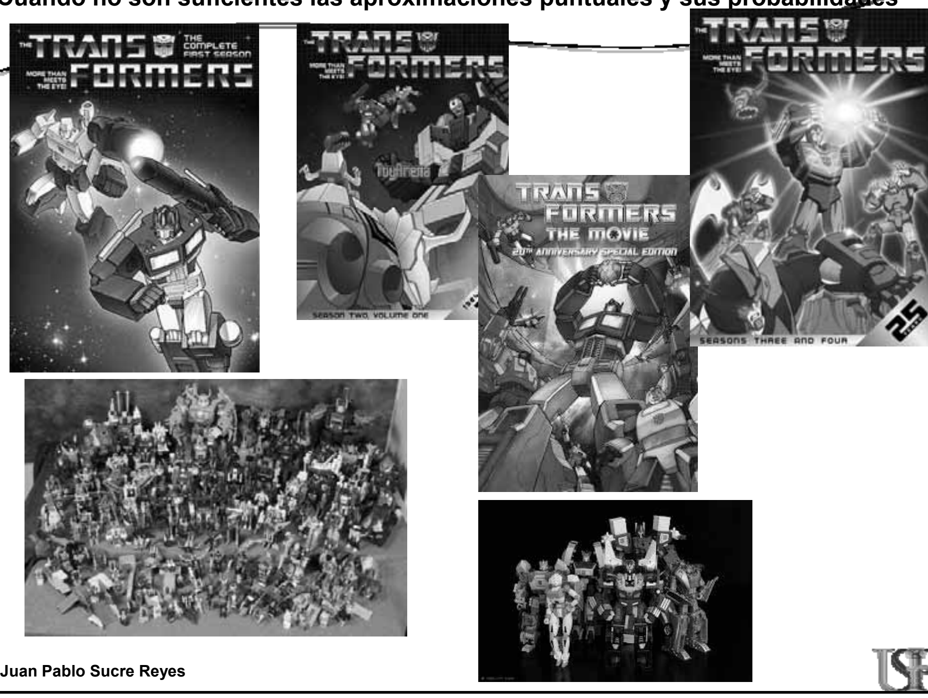


Juan Pablo Sucre Reyes

Cuando no son suficientes las aproximaciones puntuales y sus probabilidades

Juan Pablo Sucre Reyes

Cuando no son suficientes las aproximaciones puntuales y sus probabilidades



Juan Pablo Sucre Reyes

Cuando no son suficientes las aproximaciones puntuales y sus probabilidades

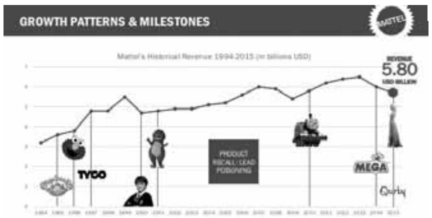
MATTEL PLAY GROUND PRODUCTIONS

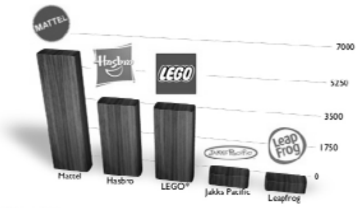
Juan Pablo Sucre Reyes

Cuando no son suficientes las aproximaciones puntuales y sus probabilidades

- *RECUERDE:* un estimador puntual es un estadístico muestral usado para estimar un parámetro poblacional: $\bar{x}$ (estimación de μ), S (estimación de σ), o $\bar{p}$ (estimación de p)
- Al no poder obtener el valor exacto del parámetro poblacional, se suele calcular una *estimación por intervalo* cuya formar general es: Estimación puntual $\pm$ margen de error
- Así entonces se tiene para μ y p :

$\bar{x} \pm$ margen de error $\bar{p} \pm$ margen de error





Juan Pablo Sucre Reyes

1. Media poblacional: σ conocida

- Se cuenta con una gran cantidad de datos históricos para calcular σ antes de tomar la muestra y calcular S (aplicaciones control de calidad).

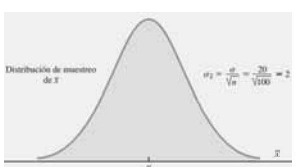
ESTIMACIÓN POR INTERVALO DE LA MEDIA POBLACIONAL: σ CONOCIDA

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

donde $(1-\alpha)$ es el coeficiente de confianza y $z_{\alpha/2}$ es el valor de z que proporciona un área $\alpha/2$ en la cola superior de la distribución de probabilidad normal estándar.

- *Ejemplo: Lloyd's Department Store selecciona una muestra aleatoria simple de 100 clientes para estimar la cantidad que gastan en cada visita a la tienda. Históricamente σ =$20 y además se tiene una distribución normal. Ésta semana, la $\bar{x}$ = $82. Se pide construir un intervalo de confianza de 95%.*
- *Solución:* Dado el coeficiente de confianza $(1-\alpha)=0{,}95$ y, por tanto, α = 0,05. En la tabla z $z_{0.025} = 1.96$. Entonces se tiene así: y el intervalo de confianza de 95% va de 82+3,92 = 78,08 $ a 82-3,92 = 85,92 $.

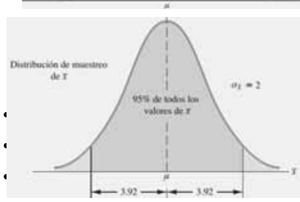
$$82 \pm 1.96\frac{20}{\sqrt{100}}$$
$$82 \pm 3.92$$



Se pueden considerar otros niveles de confianza y por ende otros intervalos de confianza:

Nivel de confianza	α	$\alpha/2$	$z_{\alpha/2}$
90%	0.10	0.05	1.645
95%	0.05	0.025	1.960
99%	0.01	0.005	2.576

$$82 \pm 1.645\frac{20}{\sqrt{100}} \quad 82 \pm 3.29$$
$$82 \pm 2.576\frac{20}{\sqrt{100}} \quad 82 \pm 5.15$$



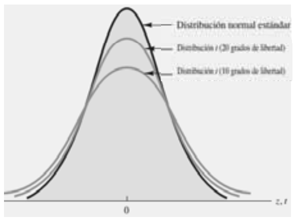
A mayor confianza $\Rightarrow$ mayor margen de error (amplitud).

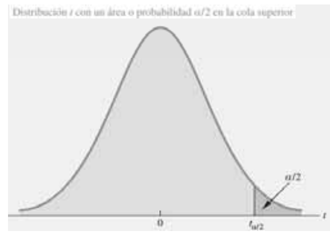
NOTA: la calidad de la aproximación depende de la distribución de N *(normal=exacto)* y de n *($\geq$30).*

Juan Pablo Sucre Reyes

2. Media poblacional: σ desconocida

- No se cuenta con una buena estimación de σ. Al usar S para estimar σ, el margen de error y la estimación por intervalo de μ se basan en la *distribución de probabilidad t*.
- Ésta es una familia de distribuciones de probabilidad, y c/u depende de sus *grados de libertad*; mientras más grande más se aproxima a la distribución normal estándar.
- La notación $t_{\alpha/2}$ representa el valor de t que deja un área de $\alpha/2$ en la cola superior de la distribución t (*TABLA t).*





- Así, se tiene:

ESTIMACIÓN POR INTERVALO DE LA MEDIA POBLACIONAL: σ DESCONOCIDA

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

donde s es la desviación estándar muestral, $(1-\alpha)$ es el coeficiente de confianza y $t_{\alpha/2}$ es el valor de t que proporciona un área $\alpha/2$ en la cola superior de la distribución t con $n-1$ grados de libertad.

- Recordando que:

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$$

Juan Pablo Sucre Reyes

2. Media poblacional: σ desconocida

•*Ejemplo: Construir un intervalo de confianza de 95% para la media del adeudo en las tarjetas de crédito en la población de familias de Estados Unidos a partir de una muestra de n=70 familias. No se cuenta con una estimación previa de σ.*

•*Solución:* Se usan los datos muestrales para estimar tanto μ ($\bar{x}$= *\$9312*) como *σ (S= \$4007).* Con $(1-\alpha)=0{,}95$ y, por tanto, $t_{0,025}$ con n-1=69 grados de libertad se tiene:

$$t_{0.025} = 1.995.$$

9430	14661	7159	9071	9691	11032
7535	12195	8137	3603	11448	6525
4078	10544	9467	16804	8279	5239
5604	13659	12595	13479	5649	6195
5179	7061	7917	14044	11298	12584
4416	6245	11346	6817	4353	15415
10676	13021	12806	6845	3467	15917
1627	9719	4972	10493	6191	12591
10112	2200	11356	615	12851	9743
6567	10746	7117	13627	5337	10324
13627	12744	9465	12557	8372	
18719	5742	19263	6232	7445	

y el intervalo de confianza de 95% es:

$$9312 \pm 1.995\frac{4007}{\sqrt{70}} \qquad 9312 \pm 955$$

o está entre *\$8357 y \$10267*.

Grados de libertad	Área en la cola superior 0.20	0.10	0.05	0.025	0.01	0.005
67	0.847	1.294	1.668	1.996	2.383	2.651
68	0.847	1.294	1.668	1.995	2.382	2.650
69	0.847	1.294	1.667	1.995	2.382	2.649

•*NOTA:* la calidad de la aproximación depende de la distribución de N: a) Normal = *exacto)* y se puede usar con cualquier tamaño de muestra b) No normal = sólo aproximado y depende del tamaño de n *(≥ 30 suficiente; aunque poblaciones muy sesgadas ≥ 50, ó si N es más o menos simétrica ≥ 15).*

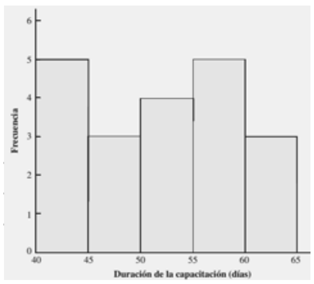
Juan Pablo Sucre Reyes

2.1 Media poblacional: σ desconocida con muestra pequeña

•*Ejemplo: Estimación de la media poblacional del tiempo requerido para que los empleados de mantenimiento de Scheer Industries completen la capacitación asistida por computadora. Considere una n=20 individuos que siguen el programa con los datos del tiempo (días) necesitados para completarlo y su respectivo histograma.*

•*Solución:* Se usan los datos muestrales para estimar tanto μ ($\bar{x}$= *51,5 días*) como *σ (S= 6,84 días).* Con (1-α)=0,95 y, por tanto, $t_{0,025}$ con n-1=19 grados de libertad se tiene $t_{0.025} = 2.093$ y el intervalo de confianza de 95% es: $51.5 \pm 2.093\left(\frac{6.84}{\sqrt{20}}\right)$ 51.5 ± 3.2 o está entre *48,3 y 54,7 días*.

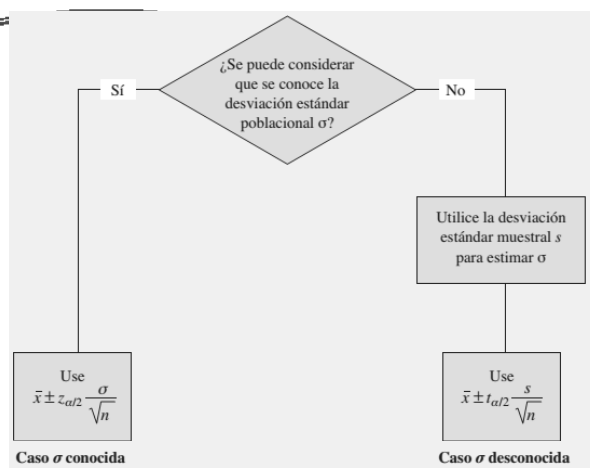
52	59	54	42
44	50	42	48
55	54	60	55
44	62	62	57
45	46	43	56



NOTA: El histograma, junto con la opinión del analista, es la última opción para decidir si es adecuado usar $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$ *para obtener una estimación por intervalo.*

Juan Pablo Sucre Reyes

2.2 Resumen de los procedimientos de estimación por intervalo



NOTA: En la mayoría de las aplicaciones, n $\geq$ 30 es adecuado. Sin embargo, si la población tiene distribución normal o aprox. normal, n puede ser menor.
Con σ desconocida y si la distribución de la población es muy sesgada o existen observaciones atípicas, se recomienda que n $\geq$ 50.

Juan Pablo Sucre Reyes

3. Determinación del tamaño de la muestra

• Cómo elegir un n suficientemente grande para obtener un margen de error deseado *(E)* al nivel de confianza elegido (1-α; más común 95% $\Rightarrow z_{0,025}$ =1,96).

• Es necesario contar con el valor de σ. Si no, se puede usar un valor preliminar o un valor planeado de σ *(estimación estudios anteriores, S de estudio piloto, Rango/4)*

TAMAÑO DE LA MUESTRA PARA UNA ESTIMACIÓN POR INTERVALO DE LA MEDIA POBLACIONAL

$$n = \frac{(z_{\alpha/2})^2\sigma^2}{E^2}$$

• *Ejemplo: En un estudio previo para investigar el costo de la renta de automóviles en Estados Unidos se encontró que el costo medio de rentar un vehículo mediano era aproximadamente 55 \$/día con un S=9,65 \$/día. Al diseñar un nuevo estudio, se especificó que la media poblacional de las rentas por día debía estimarse con un margen de error de \$2 con un nivel de 95% de confianza.*

•*Solúción:* Al utilizar \$9.65 como valor planeado de σ, tenemos:

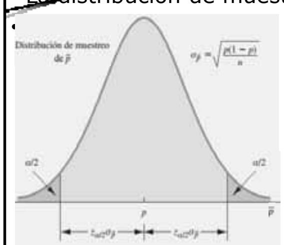
$$n = \frac{(z_{\alpha/2})^2\sigma^2}{E^2} = \frac{(1.96)^2(9.65)^2}{2^2} = 89.43$$

• el tamaño de n aconsejable es 90 rentas de automóviles medianos.

4. Proporción poblacional *p*

• La estimación por intervalo de la proporción poblacional *p*, es: $\bar{p} \pm$ margen de error

• La distribución de muestreo de *p* se aproxima a una normal si np≥5 y n(1-p)≥5 y así:



Pero al no conocer *p,* se usa $\bar{p}$*,* de modo que:

ESTIMACIÓN POR INTERVALO DE UNA PROPORCIÓN POBLACIONAL

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

donde $1 - \alpha$ es el coeficiente de confianza y $z_{\alpha/2}$ es el valor de z que deja un área $\alpha/2$ en la cola superior de la distribución normal estándar.

•*Ejemplo: Un estudio encuestó a 900 mujeres golfistas para conocer su opinión acerca de cómo se les trataba en los cursos de golf. En el estudio se encontró que 396 estaban satisfechas con la disponibilidad de horarios de salida. Estime un intervalo de confianza del 95% para la proporción poblacional de golfistas satisfechas.*

•*Solución: la proporción de golfistas satisfechas es* $\bar{p}$ *= 396/900=0,44. Con* (1-α)=0,95 y, por tanto, α = 0,05. En la tabla z $z_{0.025} = 1.96.$, y entonces el intervalo de confianza de 95% de p es: $0.44 \pm 1.96\sqrt{\frac{0.44(1-0.44)}{900}}$ 0.44 ± 0.0324 ó de 0,4076 a 0,4724.

• En %; entre 40,76% y 47,24% de las golfistas están satisfechas con los horarios.

4.1 Determinación del tamaño de la muestra

• Útil para obtener una estimación de p con una precisión determinada (E).

• Sea E (margen de error deseado): $E = z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ entonces se tiene: $n = \frac{(z_{\alpha/2})^2\bar{p}(1-\bar{p})}{E^2}$

• Como no se conocerá $\bar{p}$ hasta que se tome la muestra, no es posible usar esta fórmula. Se necesita así, un valor planeado de $\bar{p}$ útil (p*) para hacer este cálculo:

TAMAÑO DE LA MUESTRA PARA UNA ESTIMACIÓN POR INTERVALO DE LA PROPORCIÓN POBLACIONAL

$$n = \frac{(z_{\alpha/2})^2 p^*(1-p^*)}{E^2}$$

1. Se utiliza la proporción poblacional de una muestra previa de las mismas unidades o de unidades similares.
2. Se toma un estudio piloto y se elige una muestra preliminar. La proporción muestral de esta muestra se usa como valor planeado, p^*.
3. Se utiliza el criterio o una "mejor aproximación" para el valor de p^*.
4. Si no es aplicable ninguna de las alternativas anteriores, se emplea como valor planeado $p^* = 0.50$.

•*Ejemplo 1: Se llevar a cabo otro estudio para determinar la proporción actual en la población de golfistas que está satisfecha con la disponibilidad de horarios de salida. Halle el tamaño n si se desea un margen de error de 0,025 a 95% de confianza.*

•*Solución: Se usa como valor planeado p* el resultado del estudio anterior, $\bar{p}$=0,44.*

• *Así:* $n = \frac{(z_{\alpha/2})^2 p^*(1-p^*)}{E^2} = \frac{(1.96)^2(0.44)(1-0.44)}{(0.025)^2} = 1514.5$ *o 1515 mujeres golfistas.*

•*NOTA: Cuando no se cuenta con ninguna otra información, se usa p*=0,50; que da el mayor tamaño de muestra recomendable (suficiente para obtener el E deseado).*

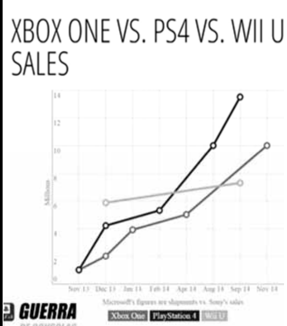
p^*	$p^*(1-p^*)$
0.10	(0.10)(0.90) = 0.09
0.30	(0.30)(0.70) = 0.21
0.40	(0.40)(0.60) = 0.24
0.50	(0.50)(0.50) = 0.25
0.60	(0.60)(0.40) = 0.24
0.70	(0.70)(0.30) = 0.21
0.90	(0.90)(0.10) = 0.09

Ejemplo 2: Repetir el ejemplo 1, para p=0,50:*

$n = \frac{(z_{\alpha/2})^2 p^*(1-p^*)}{E^2} = \frac{(1.96)^2(0.50)(1-0.50)}{(0.025)^2} = 1536.6$ *o 1537 mujeres golfistas*

Juan Pablo Sucre Reyes

Probar una afirmación acerca del valor de un parámetro poblacional



Juan Pablo Sucre Reyes

Probar una afirmación acerca del valor de un parámetro poblacional

- En las pruebas de hipótesis se hace un supuesto tentativo acerca del parámetro poblacional = *hipótesis nula* (H_0). Luego se define la *hipótesis alternativa* (H_a), que contradice lo que establece la H_0.
- En este procedimiento se usan datos de una muestra para probar dos afirmaciones contrarias indicadas por H_0 y H_a.

1. Formulación de las hipótesis nula y alternativa

- Si bien H_0 es un supuesto tentativo acerca de un parámetro poblacional (μ ó p) y H_a contradice a H_0; a veces es más fácil identificar la H_a primero y luego desarrollar la H_0.

- *a) La H_a como hipótesis de investigación:* intento de obtener evidencia en apoyo de una hipótesis de investigación: se comienza con H_a (*conclusión a sustentar*).
- *Ejemplo: Una automotríz desarrolló un nuevo sistema de inyección de combustible para dar un mejor rendimiento en millas por galón de gasolina (actualmente 24). Se busca un sustento estadístico para concluir que éste rinde más que el sistema actual.*
- *Solución:* $H_0: \mu \leq 24$ *Si los resultados muestrales llevan a rechazar H_0 se puede inferir*
- $H_a: \mu > 24$ *que H_a es verdadera.*

1. Formulación de las hipótesis nula y alternativa

•*b) La H_0 como un supuesto para ser rebatido:* inicia con el supuesto de que el valor de un parámetro poblacional es verdadero: se comienza con H_0 (*conclusión a rebatir*).

•*Ejemplo 1: Un fabricante de bebidas refrescantes afirma en su etiqueta que contienen 67,6 onzas de líquido. Se considerará correcta la leyenda si la μ (peso de llenado) es por lo menos de 67,6 onzas de líquido. Un ente estatal afirma que esto no es así.*

•*Solución:* $H_0\colon \mu \geq 67.6$ *Si los resultados muestrales llevan a rechazar H_0 se puede inferir*

• $H_a\colon \mu < 67.6$ *que H_a es verdadera.*

•*Ejemplo 2: Si a la empresa le interesa reajustar la operación de llenado a las 67,6 onzas de líquido programadas; se debe saber si el llenado es insuficiente o en exceso. Así:* $H_0\colon \mu = 67.6$ *Si los resultados muestrales llevan a rechazar H_0 se puede inferir*

• $H_a\colon \mu \neq 67.6$ *que H_a es verdadera.*

•*Resumen de las formas para H_0 y H_a* (para μ y p)

$H_0\colon \mu \geq \mu_0$	$H_0\colon \mu \leq \mu_0$	$H_0\colon \mu = \mu_0$
$H_a\colon \mu < \mu_0$	$H_a\colon \mu > \mu_0$	$H_a\colon \mu \neq \mu_0$
pruebas de una cola		*prueba de dos colas.*

Juan Pablo Sucre Reyes

2. Errores tipo Iy tipo II

• H_0 y H_a son afirmaciones opuestas acerca de la población y sólo una es verdadera.

• *Ideal:* aceptar H_0 cuando sea verdadera y rechazar H_0 cuando H_a sea verdadera.

• Desafortunadamente, las conclusiones correctas no siempre son posibles. Como la prueba de hipótesis se basa en una información muestral, puede existir error:

Errores y conclusiones correctas en las pruebas de hipótesis

		Condición poblacional	
		H_0 verdadera	**H_a verdadera**
Conclusión	**H_0 es aceptada**	Conclusión correcta	Error tipo II
	H_0 es rechazada	Error tipo I	Conclusión correcta

• *Ejemplo: Una automotríz desarrolló un nuevo sistema de inyección de combustible …*

$H_0\colon \mu \leq 24$ *Error tipo I: Afirmar que el nuevo sistema rinde más cuando no lo hace.*

$H_a\colon \mu > 24$ *Error tipo II: Afirmar que el nuevo sistema no es mejor que el actual cuando*

• *si rinde más.*

NIVEL DE SIGNIFICANCIA
Consiste en la probabilidad de cometer un error tipo I cuando la hipótesis nula es verdadera como igualdad.

• *Se denota como α (valores típicos 0,05 y 0,01); y debe especificarse en la prueba de hipótesis (controla la probabilidad de cometer error tipo I: pruebas de significancia)*

• *Si el costo de cometer error tipo I es alto, α debe ser pequeño y viceversa.*

•*NOTA: Aunque es poco común controlar el error tipo II, es posible.*

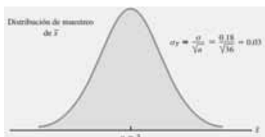
Juan Pablo Sucre Reyes

3. Media poblacional: σ conocida

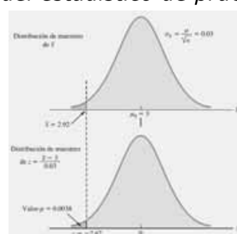
- a) Prueba de una cola:

	Prueba de cola inferior	Prueba de cola superior
Hipótesis	$H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_a: \mu > \mu_0$
Estadístico de prueba	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
Regla de rechazo: método del valor-*p*	Rechazar H_0 si el valor-$p \leq \alpha$	Rechazar H_0 si el valor-$p \leq \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $z \leq -z_\alpha$	Rechazar H_0 si $z \geq z_\alpha$

- *Ejemplo: En la etiqueta de una lata grande de Hilltop Coffee se dice que contiene 3 lb. La FTC considera que si la μ del peso de llenado sea por lo menos de 3 libras por lata, el consumidor está protegido. Se selecciona una n= 36 latas y se calcula la media muestral $\bar{x}$=2,92 lb. La población tiene una distrib. normal y su σ = 0,18. La FTC asume un riesgo de 1% de cometer un error tipo I.*
- *Solución: Paso 1: Establecer H_0 y H_a :* $H_0: \mu \geq 3$, $H_a: \mu < 3$ donde μ_0 =3
- *Paso 2: Especificar el nivel de significancia: siendo α = 0,01.*
- *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba z:*



$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.92 - 3}{0.18/\sqrt{36}} = -2.67$$



- *Ahora se debe evaluar z para decidir rechazar H_0*
- *Existen 2 métodos: valor-p y valor crítico.*

Juan Pablo Sucre Reyes

3.1 Prueba de una cola

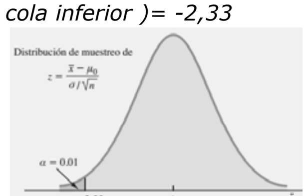
- *Método del valor-p*

VALOR-*p*
Es una probabilidad que aporta una medida de la evidencia suministrada por la muestra contra la hipótesis nula. Valores-*p* pequeños indican una evidencia mayor contra H_0.

REGLA PARA EL RECHAZO USANDO EL VALOR-*p*
Rechazar H_0 si el valor-$p \leq \alpha$

- *Paso 4. Emplear el valor del estadístico de prueba para calcular el valor-p:*
- *El valor-p (probabilidad de que z ≤ - 2,67) = 0,0038 (Tablas z)*
- *Paso 5. Regla de decisión:*
- *Como 0,0038 < α =0, 01 ⇒ H_0 es rechazada (el consumidor de café es engañado)*

- *Método del valor crítico* — Rechazar H_0 si $z \leq -z_\alpha$
- *Paso 4. Utilizar α para determinar el valor crítico ($-z_\alpha$) y la regla de rechazo:*
- *$-z_\alpha$(valor de z que corresponde a un área de α=0,01 en la cola inferior)= -2,33*
- *Paso 5. Emplear el valor del estadístico de prueba z y*
- *la regla de rechazo para determinar si H_0 es rechazada.*
- *Como z=-2,67 < -2,33, H_0 es rechazada (Hilltop Coffee*
- *está llenando las latas de manera deficiente).*



- *NOTA: En una prueba de cola superior: para el valor-p determinar el área bajo la curva a la derecha del estadístico de prueba; para valor crítico* H_0 es rechazada si $z \geq z_\alpha$

Juan Pablo Sucre Reyes

3.2 Prueba de dos colas

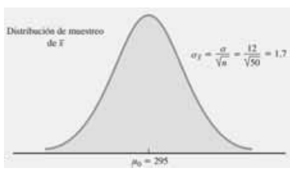
	Prueba de dos colas
Hipótesis	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$
Estadístico de prueba	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
Regla de rechazo: método del valor-*p*	Rechazar H_0 si el valor-$p \leq \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $z \leq -z_{\alpha/2}$ o si $z \geq z_{\alpha/2}$

•*Ejemplo: La USGA, establece reglas para los fabricantes de equipos de golf a ser aceptados en sus eventos. MaxFlight fabrica pelotas de golf con una distancia media de recorrido de 295 yardas. A veces el proceso se desajusta y se fabrican pelotas con una distancia media de recorrido diferente (</>). Se toman muestras periódicas de 50 pelotas para monitorear la manufactura y en una de ellas $\bar{x}$ = 297,6,. Se elige α=0,05 como nivel de significancia y se sabe que σ=12.*

•*Solución: Paso 1: Establecer H_0 y H_a :* $H_0: \mu = 295$, $H_a: \mu \neq 295$ donde μ_0 =295.

• *Paso 2: Especificar el nivel de significancia: siendo α = 0,05.*

• *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba z:*



$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{297.6 - 295}{12/\sqrt{50}} = 1.53$$

Juan Pablo Sucre Reyes

3.2 Prueba de dos colas

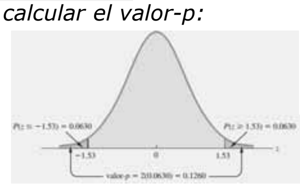
•*Método del valor-p*

CÁLCULO DEL VALOR-*p* EN UNA PRUEBA DE DOS COLAS

1. Determine el valor del estadístico de prueba z.
2. Si el valor del estadístico de prueba está en la cola superior ($z > 0$), encuentre el área bajo la curva normal estándar a la derecha de z; si está en la cola inferior ($z < 0$), localice el área bajo la curva normal estándar a la izquierda de z.
3. Duplique el área, o probabilidad, en la cola, obtenida en el paso 2 y determine el valor-*p*.

•*Paso 4. Emplear el valor del estadístico de prueba para calcular el valor-p:*

• *El valor-p = 2(0,0630) = 0,1260. (Tablas z)*



•*Paso 5. Regla de decisión:*

• *Como 0,1260>α =0,05 $\Rightarrow$ H_0 no es rechazada (el proceso de manufactura está bien)*

Juan Pablo Sucre Reyes

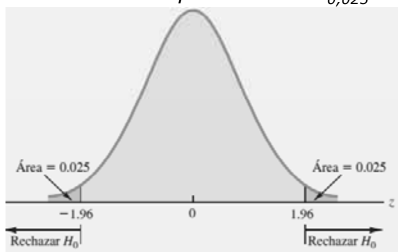
3.2 Prueba de dos colas

•*Método del valor crítico*

•*Paso 4. Utilizar α para determinar los valores críticos y la regla de rechazo:*

Rechazar H_0 si
$z \leq -z_{\alpha/2}$
o si $z \geq z_{\alpha/2}$

•*Si α =0,05; en cada cola, el área más allá del valor crítico es $\alpha/2$=0,025. En tabla z los valores críticos para el estadístico de prueba son -$z_{0,025}$=-1,96 y $z_{0,025}$=1,96.*



•*Paso 5. Emplear el valor del estadístico de prueba z y la regla de rechazo para determinar si H_0 es rechazada.*

• *Como z=1,53 no es < -1,96, ni tampoco es > 1,96 $\Rightarrow$ H_0 no es rechazada (el proceso de manufactura está bien).*

•*NOTA: Generalmente n$\geq$30 es adecuado. Si N tiene distribución normal la prueba es exacta y n puede ser cualquiera. Si N no es normal pero es simétrica, n debe ser mínimamente 15.*

3.3 Relación entre estimación por intervalo y prueba de hipótesis

MÉTODO DEL INTERVALO DE CONFIANZA PARA PROBAR UNA HIPÓTESIS DE LA FORMA

$$H_0: \mu = \mu_0$$
$$H_a: \mu \neq \mu_0$$

1. Seleccionar de la población una muestra aleatoria simple y emplear el valor de la media muestral $\bar{x}$ para obtener un intervalo de confianza de la media poblacional μ.

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

2. Si el intervalo de confianza contiene el valor hipotético μ_0, H_0 no es rechazada. En caso contrario, H_0 es rechazada.[2]

Ejemplo: Prueba de hipótesis para MaxFlight y sus pelotas de golf.
Solución: Se tiene el intervalo de 95% de confianza:

$H_0: \mu = 295$
$H_a: \mu \neq 295$

$$297.6 \pm 1.96\frac{12}{\sqrt{50}} \qquad 297.6 \pm 3.3 \qquad 294.3 \text{ a } 300.9$$

Como el μ_0 =295 está en dicho intervalo, H_0 *no es rechazada (el proceso de manufactura está bien).*

4. Media poblacional: σ desconocida

- Para esta prueba, la $\bar{x}$ se usa como estimación de μ y la S, como estimación de σ.
- Además, el estadístico de prueba tiene distribución t *(n-1 g. de libertad).* $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

•*a) Prueba de una cola:*

	Prueba de cola inferior	Prueba de cola superior
Hipótesis	$H_0: \mu \ge \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu \le \mu_0$ $H_a: \mu > \mu_0$
Estadístico de prueba	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
Regla de rechazo: método del valor-*p*	Rechazar H_0 si el valor-$p \le \alpha$	Rechazar H_0 si el valor-$p \le \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $t \le -t_\alpha$	Rechazar H_0 si $t \ge t_\alpha$

•*Ejemplo: Una revista (viajes de negocios) clasifica aeropuertos internacionales con base en la evaluación de la población de viajeros de negocios (0-10).Aquellos con una media>7 son considerados de servicio superior. Se toma una muestra de 60 viajeros de cada terminal aeroportuaria. En la tomada en Heathrow (Londres) la $\bar{x}$=7,25 y la S=1,052. Con una significancia de 5% diga si el servicio superior lo tiene Heathrow.*

•*Solución: Paso 1: Establecer H_0 y H_a :* $H_0: \mu \le 7$ $H_a: \mu > 7$ donde μ_0 =7.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{7.25 - 7}{1.052/\sqrt{50}} = 1.84$$

- *Paso 2: Especificar el nivel de significancia: siendo α = 0,05.*
- *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba t:*

•*Método del valor-p* Minitab 15

•*Paso 4. Emplear el valor del estadístico de prueba para calcular el valor-p:*

•*n-1= 59 gl; el valor-p (área a la derecha de t=1,84) es 0,025<v-p<0,05 (Tabla t)*

•*Paso 5. Regla de decisión:*

- *Como valor-p < α =0,05 ⇒ H_0 es rechazada (Heathrow tiene servicio superior)*

Área en la cola superior	0.20	0.10	0.05	0.025	0.01	0.005
Valor *t* (59 *gl*)	0.848	1.296	1.671	2.001	2.391	2.662

4.1 Prueba de una cola

•*Ejemplo: Una revista (viajes de negocios) clasifica aeropuertos internacionales con base en la evaluación de la población de viajeros de negocios (0-10).Aquellos con una media>7 son considerados de servicio superior. Se toma una muestra de 60 viajeros de cada terminal aeroportuaria. En la tomada en Heathrow (Londres) la x=7,25 y la S=1,052. Con una significancia de 5% diga si el servicio superior lo tiene Heathrow.*

•*Solución: Paso 1: Establecer H_0 y H_a :* $H_0: \mu \le 7$ $H_a: \mu > 7$ donde μ_0 =7.

- *Paso 2: Especificar el nivel de significancia: siendo α = 0,05.*
- *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba t:*

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{7.25 - 7}{1.052/\sqrt{50}} = 1.84$$

•*Método del valor crítico*

•*Paso 4. Utilizar α para determinar el valor crítico (t_α) y la regla de rechazo:* Rechazar H_0 si $t \ge t_\alpha$

- *t_α (área de α = 0,05 en la cola superior)= 1,671*

Área en la cola superior	0.20	0.10	0.05	0.025	0.01	0.005
Valor *t* (59 *gl*)	0.848	1.296	1.671	2.001	2.391	2.662

•*Paso 5. Emplear el valor del estadístico de prueba t y la regla de rechazo para determinar si H_0 es rechazada.*

- *Como 1,84 > 1,671, H_0 es rechazada (Heathrow tiene un servicio superior).*

4.2 Prueba de dos colas

•

	Prueba de dos colas
Hipótesis	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$
Estadístico de prueba	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
Regla de rechazo: método del valor-*p*	Rechazar H_0 si el valor-$p \leq \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $t \leq -t_{\alpha/2}$ o si $t \geq t_{\alpha/2}$

•*Ejemplo: Holiday Toys distribuye a más de 1000 puntos de venta. Al planear su producción, su gerente de Marketing espera que su nuevo juguete tenga una demanda de 40 unidades en promedio por punto de venta. Por ello se toma una muestra de 25 puntos de venta y se les pide anticipar la cantidad que solicitarán, de donde x=37,4 y S=11,79 unidades (histograma simétrico). Al 5% de significancia pruebe lo anterior.*

•*Solución: Paso 1: Establecer H_0 y H_a :* $H_0: \mu = 40$ $H_a: \mu \neq 40$ donde μ_0 =40.

• *Paso 2: Especificar el nivel de significancia: siendo α = 0,05.*

• *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba t:*

•*Método del valor-p* $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{37.4 - 40}{11.79/\sqrt{25}} = -1.10$

•*Paso 4. Emplear el valor del estadístico de prueba para calcular el valor-p:*

• *A n-1= 24 gl; el valor-p (2 x área a la derecha de t=1,10; por simetría con -1,10) es 0,20<valor-p<0,40 (Tabla t)*

Área en la cola superior	0.20	0.10	0.05	0.025	0.01	0.005
Valor *t* (24 *gl*)	0.857	1.318	1.711	2.064	2.492	2.797

•*Paso 5. Regla de decisión:*

• *Como valor-p > α =0,05 $\Rightarrow H_0$ no es rechazada (Holiday Toys debe seguir su plan de producción).*

Juan Pablo Sucre Reyes

4.2 Prueba de dos colas

•*Ejemplo: Holiday Toys distribuye a más de 1000 puntos de venta. Al planear su producción, su gerente de Marketing espera que su nuevo juguete tenga una demanda de 40 unidades en promedio por punto de venta. Por ello se toma una muestra de 25 puntos de venta y se les pide anticipar la cantidad que solicitarán, de donde x=37,4 y S=11,79 unidades (histograma simétrico). Al 5% de significancia pruebe lo anterior.*

•*Solución: Paso 1: Establecer H_0 y H_a :* $H_0: \mu = 40$ $H_a: \mu \neq 40$ donde μ_0 =40.

• *Paso 2: Especificar el nivel de significancia: siendo α = 0,05.*

• *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba t:*

•*Método del valor crítico* $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{37.4 - 40}{11.79/\sqrt{25}} = -1.10$

•*Paso 4. Utilizar α para determinar los valores críticos y la regla de rechazo:*

•*Si α =0,05; en cada cola, el área es α/2=0,025. En tabla t los valores críticos son: $t_{0,025}$=2,064 y -$t_{0,025}$=-2,064.*

Área en la cola superior	0.20	0.10	0.05	0.025	0.01	0.005
Valor *t* (24 *gl*)	0.857	1.318	1.711	2.064	2.492	2.797

Rechazar H_0 si $t \leq -t_{\alpha/2}$ o si $t \geq t_{\alpha/2}$

•*Paso 5. Emplear el valor del estadístico d prueba z y la regla de rechazo para determinar si H_0 es rechazada.*

• *Como t=-1,10 no es < -2,064, ni tampoco es > 2,064 $\Rightarrow H_0$ no es rechazada (Holiday Toys debe seguir su plan de producción).*

•*NOTA: Generalmente n$\geq$30 es adecuado. Si N tiene distribución aprox. normal n puede ser < 15. Si N es sesgado o tiene observaciones atípicas, n debe ser $\geq$ 50.*

Juan Pablo Sucre Reyes

5. Proporción poblacional

• p_0 es el valor hipotético para p. Las 3 formas de una prueba de hipótesis para p son:

	Prueba de cola inferior	Prueba de cola superior	Prueba de dos colas
Hipótesis	$H_0: p \geq p_0$ $H_a: p < p_0$	$H_0: p \leq p_0$ $H_a: p > p_0$	$H_0: p = p_0$ $H_a: p \neq p_0$
Estadístico de prueba	$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Regla de rechazo: método del valor-p	Rechazar H_0 si el valor-$p \leq \alpha$	Rechazar H_0 si el valor-$p \leq \alpha$	Rechazar H_0 si el valor-$p \leq \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $z \leq -z_\alpha$	Rechazar H_0 si $z \geq z_\alpha$	Rechazar H_0 si $z \leq -z_{\alpha/2}$ o si $z \geq z_{\alpha/2}$

•Ejemplo: En años anteriores, 20% de los golfistas del campo Pine Creek eran mujeres. Para aumentar esta proporción, se realizó una promoción especial para atraerlas. Un mes después, el gerente pidió un estudio estadístico con 5% de error para determinar si la proporción había aumentado. De una muestra aleatoria de 400, 100 son mujeres.

•Solución: Paso 1: Establecer H_0 y H_a : $H_0: p \leq 0.20$, $H_a: p > 0.20$ donde p_0 =0,20. $\bar{p} = \frac{100}{400} = 0.25$

• Paso 2: Especificar el nivel de significancia: siendo α = 0,05.

• Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba z:

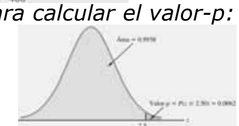
$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.25 - 0.20}{\sqrt{\frac{0.20(1-0.20)}{400}}} = \frac{0.05}{0.02} = 2.50$$

•Método del valor-p

•Paso 4. Emplear el valor del estadístico de prueba para calcular el valor-p:

• valor-p (área a la derecha de z=2,50 ó z$\geq$ 2,50)

•es 1–0,9938=0,0062. (Tabla z)



•Paso 5. Regla de decisión:

• Como valor-p 0,0062 < α =0,05 $\Rightarrow$ H_0 es rechazada (la proporción si aumentó)

Juan Pablo Sucre Reyes

5. Proporción poblacional

•Ejemplo: En años anteriores, 20% de los golfistas del campo Pine Creek eran mujeres. Para aumentar esta proporción, se realizó una promoción especial para atraerlas. Un mes después, el gerente pidió un estudio estadístico con 5% de error para determinar si la proporción había aumentado. De una muestra aleatoria de 400, 100 son mujeres.

•Solución: Paso 1: Establecer H_0 y H_a : $H_0: p \leq 0.20$, $H_a: p > 0.20$ donde p_0 =0,20. $\bar{p} = \frac{100}{400} = 0.25$

• Paso 2: Especificar el nivel de significancia: siendo α = 0,05.

• Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba z:

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.25 - 0.20}{\sqrt{\frac{0.20(1-0.20)}{400}}} = \frac{0.05}{0.02} = 2.50$$

•Método del valor crítico

•Paso 4. Utilizar α para determinar el valor crítico (z_α) y la regla de rechazo: Rechazar H_0 si $z \geq z_\alpha$

• z_α (área de α = 0,05 en la cola superior)= 1,645 (Tabla z)

•Paso 5. Emplear el valor del estadístico de prueba z y la regla de rechazo para determinar si H_0 es rechazada.

• Como 2,50 > 1,645, H_0 es rechazada (la proporción de jugadoras si aumentó).

•NOTA: Se supone que np $\geq$ 5 y n(1-p) $\geq$ 5, con lo cual se puede usar una distribución normal como aproximación a la distribución de muestreo de $\bar{p}$.

Juan Pablo Sucre Reyes

6. Prueba de hipótesis y toma de decisiones

- Las anteriores pruebas de hipótesis son *pruebas de significancia* (controlan la probabilidad de cometer error tipo I (α), pero no tipo II. Por tanto, se recomienda decir *"no rechazar H_0"* más que "aceptar H_0" (*riesgo de error tipo II :* aceptar lo falso).
- *"No rechazar H_0"* describe una evidencia estadística no concluyente que posterga una decisión o una acción hasta mayor investigación y pruebas.
- Si se debe actuar, se recomienda controlar además la probabilidad de cometer un error tipo II *(otras pruebas y ajustes de n)* y así se puede concluir con *"aceptar H_0" o "rechazar H_0".*

Juan Pablo Sucre Reyes

GRACIAS POR SU ATENCIÓN…..

Juan Pablo Sucre Reyes