INTERVALO DE CONFIANZA

I- Concepto de Intervalo de Confianza.

En el contexto de estimar un parámetro poblacional, un intervalo de confianza es un rango de valores (calculado en una muestra) en el cual se encuentra el verdadero valor del parámetro, con una probabilidad determinada.

La probabilidad de que el verdadero valor del parámetro se encuentre en el intervalo construido se denomina **nivel de confianza**, y se denota 1-α. La probabilidad de equivocarnos se llama **nivel de significancia** y se simboliza α. Generalmente se construyen intervalos con confianza 1-α=95% (o significancia α=5%). Menos frecuentes son los intervalos con α=10% o α=1%.

Para construir un intervalo de confianza, se puede comprobar que la distribución Normal Estándar cumple:

P(-1.96 < z < 1.96) = 0.95

(lo anterior se puede comprobar con una tabla de probabilidades o un programa computacional que calcule probabilidades normales).

Luego, si una variable X tiene distribución N(μ,σ), entonces el 95% de las veces se cumple:

$$-1.96 \leq \frac{\bar{X}-\mu}{\sigma}\sqrt{n} \leq 1.96$$

Despejando μ en la ecuación se tiene:

$$\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

El resultado es un intervalo que incluye al μ el 95% de las veces. Es decir, es un **intervalo de confianza al 95% para la media** μ cuando la variable X es normal y σ^2es conocido.

En forma general, Si el nivel de confianza es 1-α, y se conoce σ, el intervalo de confianza de μ es: $\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

II- Intervalo de confianza para un promedio:

Generalmente, cuando se quiere construir un intervalo de confianza para la media poblacional μ, la varianza poblacional σ^2es desconocida, por lo que el intervalo para μconstruido al final de II es muy poco práctico.

Si en el intervalo se reemplaza la desviación estándar poblacional σpor la desviación estándar muestral **s**, el intervalo de confianza toma la forma:

$$\overline{X} - 1.96 \times \frac{s}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96 \times \frac{s}{\sqrt{n}}$$

La cual es una buena aproximación para el intervalo de confianza de 95% para μcon σ^2desconocido. Esta aproximación es mejor en la medida que el tamaño muestral sea grande, es decir, *n>30*

Cuando el tamaño muestral es pequeño, (*n≤30*) el intervalo de confianza requiere utilizar la distribución t de Student (con n-1 grados de libertad, siendo n el tamaño de la muestra), en vez de la distribución normal (por ejemplo, para un intervalo de 95% de confianza, los límites del intervalo ya no serán construidos usando el valor 1,96).

En forma general si no se conoce σ se sustituye por la desviación estándar de la muestra, es decir, por s, si además el tamaño de la muestra es *n>30*. El intervalo de confianza para un nivel de confianza de 1-α, de la media poblacional es:

$$\overline{X} - Z_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\alpha/2} \frac{s}{\sqrt{n}}$$

En el caso de que *n≤30*, el intervalo será:

$$\overline{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}$$ con *n-1* grados de libertad para la distribución t

Ejemplo:
Los siguientes datos son los puntajes obtenidos para 45 personas de una escala de depresión (mayor puntaje significa mayor depresión).

2	5	6	8	8	9	9	10	11
11	11	13	13	14	14	14	14	14
14	15	15	16	16	16	16	16	16
16	16	17	17	17	18	18	18	19
19	19	19	19	19	19	19	20	20

Para construir un intervalo de confianza para el puntaje promedio poblacional, asumamos que los datos tienen distribución normal, con varianza poblacional σ^2desconocida. Como σ^2es desconocido, lo estimamos por s^2=18,7. Luego, un intervalo de confianza aproximado es:

$$\overline{X} - 1.96 \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96 \frac{S}{\sqrt{n}} \Rightarrow 14.5 - 1.96 \frac{4.3}{\sqrt{45}} \leq \mu \leq 14.5 + 1.96 \frac{4.3}{\sqrt{45}}$$

Luego, el intervalo de confianza para μ es (13,2 , 15,8). Es decir, el puntaje promedio poblacional se encuentra entre 13,2 y 15,8 con una confianza 95%.

III. Intervalo de Confianza para una Proporción.

En este caso, interesa construir un intervalo de confianza para una proporción o un porcentaje poblacional (por ejemplo, el porcentaje de personas con hipertensión, fumadoras, etc.)

Si el tamaño muestral n es grande, el Teorema Central del Límite nos asegura que:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

O bien:

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)}} \sqrt{n} \sim N(0,1)$$

Donde p es el porcentaje de personas con la característica de interés en la población (o sea, es el parámetro de interés) y $\hat{p}$ es su estimador muestral.

Luego, procediendo en forma análoga al caso de la media, podemos construir un intervalo de 95% de confianza para la proporción poblacional p.

$$\hat{p} - 1.96 \times \sqrt{\hat{p} \times (1-\hat{p})/n} \leq p \leq \hat{p} - 1.96 \times \sqrt{\hat{p} \times (1-\hat{p})/n}$$

En forma general:

$\hat{p} - Z_{\alpha/2}(\sqrt{\hat{p}(1-\hat{p})/n}) \leq p \leq \hat{p} + Z_{\alpha/2}(\sqrt{\hat{p}(1-\hat{p})/n})$ con un nivel de confianza de 1-α

Ejemplo:
En un estudio de prevalencia de factores de riesgo en una cohorte de 412 mujeres mayores de 15 años en la Región Metropolitana, se encontró que el 17.6% eran hipertensas. Un intervalo de 95% de confianza para la proporción de mujeres hipertensas en la Región Metropolitana está dado por:

$$0.176 - 1.96\sqrt{0.176(1-0.176)/412} \leq p \leq 0.176 + 1.96\sqrt{0.176(1-0.176)/412}$$

Luego, la proporción de hipertensas varía entre (0,139 , 0,212) con una confianza de 95%.

IV. Uso de Intervalos de Confianza para verificar Hipótesis.

Los intervalos de confianza permiten verificar hipótesis planteadas respecto a parámetros poblacionales.

Por ejemplo, supongamos que se plantea la hipótesis de que el promedio de peso de nacimiento de cierta población es igual a la media nacional de 3250 gramos.

Al tomar una muestra de 30 recién nacidos de la población en estudio, se obtuvo:

$\bar{x}$=2930
s=450
n= 30

Al construir un intervalo de 95% de confianza para la media poblacional, se obtiene:

$$2930 - 1.96 \times \frac{450}{\sqrt{30}} \leq \mu \leq 2930 + 1.96 \times \frac{450}{\sqrt{30}}$$

Luego, el peso de nacimiento varía entre 2769 y 3091 gramos, con una confianza de 95%.
Como el intervalo no incluye el valor μ=3250 gramos planteado en la hipótesis, entonces esta es rechazada con confianza 95% (o un valor p menor a 0,5).

Ejercicios

1. Se sabe que el peso de los ladrillos producidos por una determinada fábrica sigue una distribución normal con una desviación típica de 0,12 kilos. En el día de hoy se extrae una muestra aleatoria de 60 ladrillos cuyo peso medio es de 4,07 kilos. Calcular un intervalo de confianza del 99% para el peso medio de los ladrillos producidos hoy.

2. Un supervisor de control de calidad en una enlatadora sabe que la cantidad exacta en cada lata varía, pues hay ciertos factores imposibles de controlar que afectan a la cantidad de llenado. El llenado medio por lata es importante. A fin de estimar la variación del llenado en la enlatadora, el supervisor escoge al azar 10 latas y pesa el contenido de cada una, obteniendo el siguiente pesaje (en onzas): 7,96 7,90 7,98 8,01 7,97 8,03 8,02 8,04 8,02. Establezca un intervalo de confianza de 90% para la verdadera media del llenado de latas en la enlatadora.

3. Es común utilizar aceros inoxidables en las plantas químicas para manejar fluidos corrosivos. Sin embargo, estos aceros tienen especial susceptibilidad al agrietamiento por corrosión causada por esfuerzos en ciertos entornos. En una muestra de 295 fallas de aleaciones de acero que ocurrieron en refinerías de petróleo y plantas petroquímicas en Japón durante los últimos 10 años, 118 se debieron a agrietamiento por corrosión causada por esfuerzos y a fatiga de corrosión. Establezca un intervalo de confianza de 95% para la verdadera proporción de fallas de aleaciones causadas por agrietamiento por corrosión debido a esfuerzos.

4. En una muestra de 40 personas ingresadas en un hospital, se encuentra que 6 de ellas padecen una determinada enfermedad. Hallar el intervalo de confianza para la proporción de dichas personas.

PRUEBA DE HIPOTESIS

A partir de esta unidad estudiaremos lo relacionado a probar diferentes tipos de hipótesis, empezando por definir que es una *hipótesis* y una *prueba de hipótesis,* enlistaremos los pasos para probar una hipótesis, y realizaremos pruebas de hipótesis relativas a la media de una población y a las medias de dos poblaciones.

¿Qué es una hipótesis?

Hipótesis es una afirmación o suposición respecto al valor de un parámetro poblacional

Son ejemplos de hipótesis, o afirmaciones hechas sobre un parámetro poblacional las siguientes:

El ingreso mensual promedio de todos los ciudadanos es $4500.00

El 20% de los delincuentes capturados son sentenciados a prisión

El 90% de las formas fiscales son llenadas correctamente

Todas estas hipótesis tienen algo en común, las poblaciones de interés son tan grandes que no es factible estudiar todos sus elementos. Como ya sabemos, una alternativa a estudiar la población entera es tomar una muestra de la población de interés. De esta manera podemos probar una afirmación para determinar si la evidencia soporta o no la afirmación.

¿Qué es una prueba de hipótesis?

Una prueba de hipótesis comienza con una afirmación o suposición acerca de un parámetro poblacional, tal como la media poblacional. Una hipótesis podría ser que la colegiatura que pagan los estudiantes universitarios de la República Mexicana es en promedio de 3000 pesos. Para comprobar esta hipótesis no podríamos contactar a todos los estudiantes universitarios de la república, el costo sería exorbitante. Para probar la validez de esta afirmación podríamos seleccionar una muestra de la población de estudiantes y basados en ciertas reglas de decisión, aceptar o rechazar la hipótesis. Si la media muestral fuera de 1000 pesos ciertamente tendríamos que rechazar la hipótesis, pero si la media muestral fuera 2990 pesos ¿podríamos asumir que la media poblacional si es de 3000 pesos?, ¿podemos atribuir al error de muestreo la diferencia de 10 pesos entre las dos medias, o es una diferencia significativa?

Prueba de hipótesis es un procedimiento basado en una evidencia muestral y la teoría de la probabilidad, usado para determinar si la hipótesis es una afirmación razonable para no ser rechazada, o es una afirmación poco razonable y ser rechazada.

Procedimiento de 4 pasos para probar una hipótesis

Hay un procedimiento de cuatro pasos que sistematizan la prueba de hipótesis. Para ilustrar el procedimiento, completemos el ejemplo anterior. Supongamos que la muestra es de 20 estudiantes y el nivel de significancia es de .05. Los cuatro pasos son los siguientes:

Paso 1. Establecer las hipótesis nula y alterna

El primer paso es establecer la hipótesis a ser probada. Esta es llamada la ***hipótesis nula***, simbolizada por H_0, el subíndice cero implica "cero diferencia". Usualmente el término "no" es encontrado en la hipótesis nula significando "no cambio". La hipótesis nula de la introducción podría ser "la colegiatura mensual promedio de los estudiantes universitarios no es diferente de 3000 pesos". Esto es lo mismo que decir "…es igual a 3000 pesos". La hipótesis nula se puede simbolizar $H_0: \mu = 3000$.

La hipótesis nula es una afirmación que será aceptada si los datos de la muestra no nos proveen de evidencia convincente de que es falsa, es decir, si se acepta la hipótesis nula decimos que la evidencia no es suficiente para rechazarla pero no podemos afirmar que es verdadera.

La hipótesis alterna es la afirmación que se acepta si se rechaza la hipótesis nula. Esta hipótesis, también llamada hipótesis de investigación, se simboliza con ***Ha***. La hipótesis alterna es aceptada si la evidencia proporcionada por la muestra es suficiente para afirmar que la ***Ho*** es falsa.

En este ejemplo las hipótesis serían las siguientes:

Ho: La colegiatura promedio de los estudiantes no es diferente de 3000 pesos

Ho: $\mu = 3000$

Ha: La colegiatura promedio de los estudiantes es diferente de 3000 pesos

Ha: $\mu \neq 3000$

Paso 2. Determinar el criterio de contraste

Determinar el criterio de contraste consiste en especificar el nivel de significancia, el tipo de distribución, y los valores críticos.

Existen cuatro posibilidades al tomar una decisión respecto a una hipótesis:

	Aceptar Ho	Rechazar Ho
Ho verdadera	Decisión correcta	Error Tipo I
Ho falsa	Error Tipo II	Decisión correcta

Nivel de significancia es la probabilidad de rechazar una hipótesis nula verdadera

El **nivel de significancia** es simbolizado por **α**, y también es conocido como nivel de riesgo. Este último término es más apropiado porque es el riesgo que se toma de rechazar una hipótesis verdadera.

No hay un nivel de significancia para todos los estudios, se puede utilizar cualquier valor de probabilidad entre 0 y 1. Tradicionalmente, el nivel de .05 es aplicado a proyectos de investigación, el nivel .01 a control de calidad, y .10 a sondeos políticos. Tú como investigador debes decidir el nivel de significancia antes de colectar la muestra de datos.

El **tipo de distribución** se determinará dependiendo de la naturaleza de la hipótesis y del tamaño de la muestra. Cuando la hipótesis es relativa a medias poblacionales y las muestras son grandes ($n>30$) se utiliza la distribución normal. Cuando es relativa a la media y la muestra es chica ($n\leq30$) se utiliza la distribución t de student.

Los **valores críticos** son los valores de la variable de la distribución que limitan el área crítica, que es la parte de la curva que corresponde al nivel de significancia.

En este ejemplo el nivel de significancia es de .05, se utiliza la distribución t de student porque la muestra es pequeña, los valores críticos se encontraron de la siguiente manera

El área crítica cuando la hipótesis alterna tiene el símbolo ($\neq$) se divide en dos y se dice que el problema es de dos colas, y cada cola vale $\alpha/2$. Si la Ha tiene el signo (<) el problema es de la cola izquierda, si tiene el signo(>) es de la cola derecha, y en ambos casos la cola vale α. Este problema es de dos colas:

Paso 3. Calcular el estadístico de prueba

El estadístico de prueba es un valor obtenido de la información de la muestra para compararlo con el criterio de contraste y rechazar o aceptar la hipótesis. El estadístico de prueba cambia de acuerdo a la distribución que se utilice. En este problema el estadístico de prueba es t y se simboliza t*

Supongamos que las colegiaturas de los estudiantes universitarios entrevistados son las siguientes:

2821	3102	2398	2511	3222
2329	3109	2725	3627	2933
3822	3044	3125	2650	2741
3054	3281	2292	2952	2462

La media y la desviación estándar de la muestra son 2910 y 411.95 respectivamente, se procede enseguida a calcular el error estándar y la t*

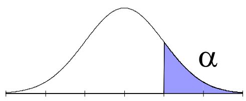
Paso 4. Tomar decisión y conclusión

Una regla de decisión es establecer las condiciones sobre las cuales la hipótesis nula es rechazada o no rechazada. Si el estadístico de prueba queda dentro de la zona crítica la hipótesis nula deberá ser rechazada. Si el estadístico de prueba queda fuera de la zona crítica la hipótesis nula no deberá ser rechazada.

En el ejemplo de las colegiaturas, como el estadístico de prueba quedó fuera de la zona crítica la hipótesis nula no puede ser rechazada. La conclusión podría ser la siguiente:

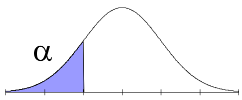
"No hay evidencia suficiente para afirmar que la colegiatura que pagan en promedio los estudiantes universitarios es diferente de 3000 pesos, en un nivel de significancia de .05"

Ejemplos Para H_1: $\mu >$ valor aceptado, la región de rechazo está dada por:



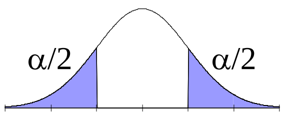
(cola derecha, z ó t)

Para $H_1 : \mu <$ valor aceptado, la región de rechazo está dada por:



(cola izquierda, z ó t)

Para $H_1 : \mu \neq$ valor aceptado, la región de rechazo es de dos colas y está dada por:



(2-colas, z ó t)

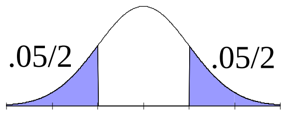
Ejemplo 1: Determine si la región de rechazo es de la cola derecha, de la cola izquierda o de dos colas.

a. $H_0 : \mu = 15$, $H_1 : \mu \neq 15$, α=.05

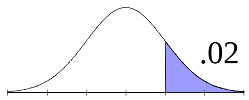
b. $H_0 : p \leq 0.7$, $H_1 : p > 0.7$, α=.02

Solución: La forma de la región de rechazo está determinada por la hipótesis alterna.

a. $H_1 : \mu \neq 15$ significa que la región está en ambas colas.



b. $H_1 : p > 7$ significa que la región está en la cola derecha.

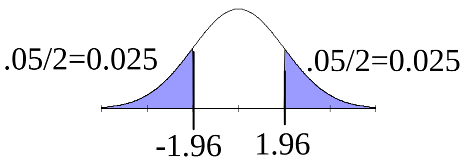


Ejemplo 2: En el Ejemplo 1a, presumamos que la región de rechazo es parte de la curva normal estándar. Complete el dibujo de la región crítica para los valores α siguientes:

a. $\alpha = .05$

Solución:

a. Del ejemplo 1(*a*), tenemos:



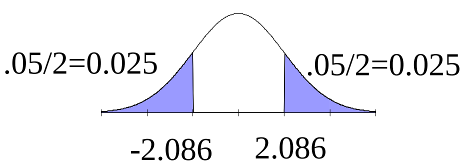
De la tabla de la distribución normal, la P(Z<z) =.025 corresponde a un valor Z= -1.96. Por simetría la P(Z>z)=.025 corresponde a Z= 1.96.

Ejemplo 3: En el ejemplo 1a, presumamos que la región de rechazo es parte de la curva *t*. Complete el dibujo de la región de rechazo para:

a. $\alpha = .05$ y $\upsilon = 14$

Solución:

a. Del ejemplo 1(a), $\alpha = .05$, y $\upsilon = 14$, tenemos:



De la tabla de la distribución t, la P(T<t) =.025 corresponde a un valor t= -2.086. Por simetría la P(T>t)=.025 corresponde a t= 2.086.

Ejemplo 4: Establezca las hipótesis nula y alterna.

a. Las millas por galón (mpg) promedio de un nuevo modelo de automóvil es 32.

b. Más del 65% de los empleados de un colegio aportan a Fondos Unidos.

c. En promedio, los empleados de cierta compañía viven a no más de 15 millas de la misma.

d. Al menos un 60% de la población adulta de una comunidad votará en las próximas elecciones Presidenciales.

e. El peso promedio de un pollo para asar es de al menos cuatro libras.

Solución:

a. $H_0 : \mu = 32$, $H_1 : \mu \neq 32$ **b.** $H_0 : p \geq .65$, $H_1 : p < .65$ **c.** $H_0 : \mu \leq 15$, $H_1 : \mu > 15$

d. $H_0 : p \geq .6$, $H_1 : p < .6$ **e.** $H_0 : \mu \geq 4$, $H_1 : \mu < 4$

EJERCICIOS

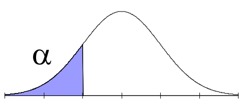
En los ejercicios (1-6) determine si la región de rechazo para la hipótesis nula está en la cola izquierda, en la cola derecha, o ambas colas. Para el nivel de significancia α dibuje la región de rechazo.

1. $H_0 : \mu \leq 11;\ H_1 : \mu > 11$ **2.** $H_0 : \mu \geq 5.8;\ H_1 : \mu < 5.8$

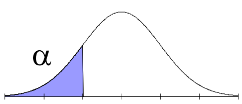
3. $H_0 : p = 0.4;\ H_1 : p \neq 0.4$ **4.** $H_0 : \mu = 110;\ H_1 : \mu \neq 110$

5. $H_0 : p \geq 0.3;\ H_1 : p < 0.3$ **6.** $H_0 : p \geq 0.8;\ H_1 : p < 0.8$

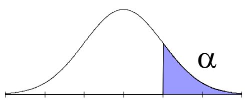
En los ejercicios (7 - 18) complete la región de rechazo (encuentre el valor de z y t).

7. 

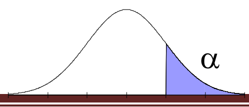
a) z, si $\alpha = .05$ b) *t*, si $\alpha = .025$ y $\upsilon = 9$

8. 

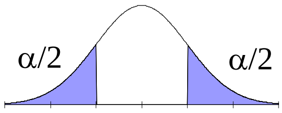
a) z, si $\alpha = .01$ b) *t*, si $\alpha = .05$ y $\upsilon = 13$

9. 

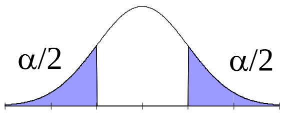
a) *z*, si $\alpha = .02$ b) *t*, si $\alpha = .01$ y $\upsilon = 5$

10. 

a) *z*, si $\alpha = .025$ b) *t*, si $\alpha = .01$ y $\upsilon = 9$

11. 

a) z, si $\alpha = .05$ b) t, si $\alpha=.05$ y $\upsilon = 10$

12. 

a) z, si $\alpha = .01$ b) t, si $\alpha=0.1$ y $\upsilon = 7$

En los ejercicios (13 - 18) establezca las hipótesis nula y alterna.

13. Los automóviles estacionados en el estacionamiento de periodo prolongado del aeropuerto internacional de Baltimore permanecen un promedio de 2.5 días.

14**.** Una nueva marca de llantas radiales dura en promedio más de 48,000 millas.

15. El balance promedio de una cuenta de cheques en el First State Bank es de al menos $150.

16. Se reclama que al menos el 60% de las compras realizadas en cierta tienda por departamentos son artículos de especiales.

17**.** Se reclama que el 20% de los graduados de cierto colegio privado solicitan admisión a escuelas de medicina.

18**.** Un dentista reclama que el 5% de sus pacientes sufren enfermedades en las encías.

PRUEBA DE UNA MEDIA

Hay dos métodos. Uno usa la región de rechazo y el otro usa los valores *P*.

a. Método de la Región de Rechazo (Método 1)

Dejemos que μ_o sea el valor reclamado (aceptado) de la media.

***Paso* 1** Establezca la hipótesis: $H_o: \quad \mu = \mu_o$

$H_1: \quad \mu > \mu_o$ ó

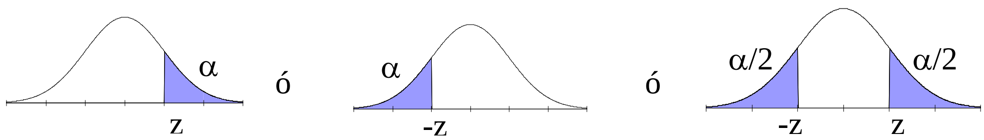
$\mu < \mu_o$ ó

$\mu \neq \mu_o$

***Paso* 2** Lea el enunciado y determine si se conoce o no el valor de σ y si el tamaño de la muestra es grande ($n \geq 30$) o pequeño ($n < 30$).

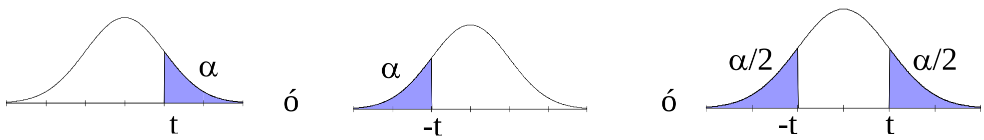
***Paso* 3** Use el nivel de significancia (α) y dibuje la región de rechazo en:

a. la curva normal estándar si conoce σ o si n es grande. La región de rechazo es igual a la hipótesis alterna.



($H_1: \mu > \mu_o$) ($H_1: \mu < \mu_o$) ($H_1: \mu \neq \mu_o$)

b. la curva t para σ desconocida y n pequeña, así la región de rechazo puede ser:



($H_1: \mu > \mu_o$) ($H_1: \mu < \mu_o$) ($H_1: \mu \neq \mu_o$)

***Paso* 4** Calcule el valor z ó t de la media muestral z ó $t = \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}}$,

Donde $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$.

***Paso* 5** Dibuje el valor de z ó t obtenido en el Paso 4 en el dibujo de la región de rechazo.

***Paso* 6** Si el valor de z o el de t quedan dentro de la región de rechazo (cola o colas), entonces rechace H_0. Si el valor de z o el de t caen fuera de la región de rechazo, entonces no rechace H_0.

***Paso* 7** Escriba la conclusión de la prueba.

Ejemplo 1: Prueba de Hipótesis $H_0: \quad \mu = 18$

$H_1: \quad \mu > 18$

Presuma que $\sigma = 5$, $\overline{X} = 19$, $n = 100$ y $\alpha = .05$

Solución:

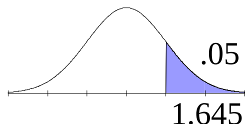
Paso 1 H_0 : $\mu = 18$

H_1 : $\mu > 18$

Paso 2 Dado que conocemos σ, así que usamos la curva normal estándar.

Paso 3 Con $\alpha = .05$, la región de rechazo es:

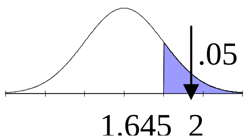
De la tabla $z = 1.645$.



Paso 4 $$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{100}} = 0.5$$

$$z = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{19 - 18}{0.5} = 2$$

Paso 5 Dibujando $z = 2$ en el dibujo hecho en el Paso 3 tenemos:



Paso 6 Dado que $z = 2$ cae dentro de la región de rechazo, de esta forma rechazamos H_o.

Paso 7 Hay evidencia para creer que $\mu > 18$.

b. Método del valor P (Método 2)

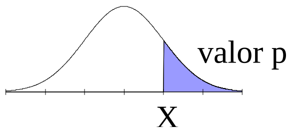
Este método es una pequeña variación del método anterior. Es más útil cuando se usa una computadora. Este método usa el **valor *P*** para probar la hipótesis nula.

Valor *P*

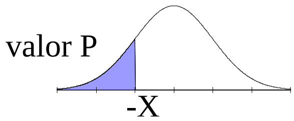
El valor *P* es la probabilidad de que la media (o cualquier otra estadística de prueba) pueda asumir un valor extremo o más extremo que la media muestral (o la estadística de prueba correspondiente) si la hipótesis nula es cierta.

Comentario: *Los valores de la media que son extremos o más extremos que la media muestral se indican por la hipótesis alterna. Dado que la probabilidad esta dada por el área bajo la curva de densidad, así el valor P también esta dado por el área.*

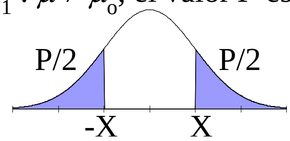
Ejemplos: 1. Para $H_1 : \mu > \mu_o$, el valor *P* esta dado por:



2. Para $H_1 : \mu < \mu_o$, el valor *P* esta dado por:



3. Para $H_1 : \mu \neq \mu_o$, el valor *P* esta dado por:



***Paso* 1** Establezca las hipótesis : $H_o : \quad \mu = \mu_o$

$H_1 : \quad \mu > \mu_o$ ó

$\mu < \mu_o$ ó

$\mu \neq \mu_o$

***Paso* 2** Lee el enunciado y determine si σ es conocida o desconocida y si el tamaño de muestra es grande o pequeño. Por consiguiente seleccione *z* ó *t*.

***Paso* 3** Calcule $z = \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}}$ si σ esta dada o *n* es grande; y

$t = \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}}$ si σ es desconocida y *n* es pequeña.

***Paso* 4** Use el valor *z* ó *t* anterior para indicar la región de valores extremos de la media. La región dependerá de la hipótesis alterna (ver ejemplos anteriores).

***Paso* 5** El valor *P* está dado por el área de la cola sombreada o la suma de las áreas de las colas.

***Paso* 6** Si el valor *P* es < α, entonces rechazo $H_{o.}$

Si el valor *P* es ≥ α, entonces no rechazo H_o.

***Paso* 7** Escriba la conclusión de la prueba.

Ejemplo 2: Pruebe la hipótesis $H_0 : \quad \mu = 18$

$H_1 : \quad \mu > 18$

Presuma que $\sigma = 5$, $\overline{X} = 19$, $n = 100$, y $\alpha = .05$

Solución:

Paso 1 H_0 : $\mu = 18$

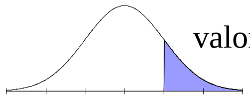
H_1 : $\mu > 18$

Paso 2 Como σ es conocida, usamos la curva *z*.

Paso 3 $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{5}{\sqrt{100}} = 0.5$

$z = \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \dfrac{19 - 18}{0.5} = 2$

Paso 4 La región que representa los valores extremos es:



Paso 5 Valor *P* = el área de la cola sombreada

= 0.5 – .4772

= .0228

Paso 6 Dado que el valor $P < .05$, entonces rechazamos la hipótesis nula.

Paso 7 Hay evidencia para creer que $\mu > 18$.

EJERCICIOS

En los ejercicios (1-5) use el método de la región de rechazo para probar las hipótesis:

1. H_0 : $\mu = 400$

H_1 : $\mu > 400$, $\sigma = 32$, $\overline{X} = 413$, $n = 36$, y $\alpha = 0.05$

2. H_0 : $\mu = 4.8$

H_1 : $\mu \neq 4.8$, $\sigma = 0.8$, $\overline{X} = 4.7$, $n = 81$, y $\alpha = 0.05$

3. H_0 : $\mu = 7.4$

H_1 : $\mu < 7.4$, $\sigma = 1.1$, $\overline{X} = 7.25$, $n = 400$, y $\alpha = 0.01$

4. H_0 : $\mu = 0.9$

H_1 : $\mu < 0.9$, $\sigma = 0.2$, $\overline{X} = 0.65$, $n = 36$, y $\alpha = 0.05$

5. H_0 : $\mu = 5.5$

H_1 : $\mu \neq 5.5$, $s = 0.8$, $\overline{X} = 5.2$, $n = 36$, y $\alpha = 0.01$

En los ejercicios (6-10) use el método del valo-P para probar las hipótesis:

6. $H_0: \ \mu = 325$

$H_1: \ \mu < 325, \ s = 3, \quad \overline{X} = 323, \quad n = 9, \ y \ \alpha = 0.1$

7. $H_0: \ \mu = 50$

$H_1: \ \mu \neq 50, \ s = 10, \quad \overline{X} = 46, \quad n = 25, \ y \ \alpha = 0.01$

8. $H_0: \ \mu = 75$

$H_1: \ \mu > 75, \ s = 4, \quad \overline{X} = 72.5, \quad n = 9, \ y \ \alpha = 0.1$

9. $H_0: \ \mu = 850$

$H_1: \ \mu \neq 850, \ s = 1.4, \quad \overline{X} = 849, \quad n = 4, \ y \ \alpha = 0.05$

10. $H_0: \ \mu = 15$

$H_1: \ \mu < 15, \ s = 0.5, \quad \overline{X} = 14.7, \quad n = 50, \ y \ \alpha = 0.01$

En los ejercicios (11-15) use el método de la región de rechazo para probar las hipótesis.

1**1.** La estatura media de la población de varones de 15 años de cierto estado es de 67.4 pulgadas con una DE de 1.3 pulgadas. Una muestra aleatoria de 400 varones de 15 años de edad de una de las ciudades de éste estado dio una estatura media de 68 pulgadas. ¿Es esto una indicación de que los varones de 15 años de ésta ciudad son más altos que la estatura promedio del estado? Pruebe la hipótesis con un nivel de significancia $\alpha = .05$.

1**2.** Un manufacturero afirma que la fortaleza promedio requerida para romper un resorte es de 575 libras con una DE de 10 libras. Un distribuidor probó una muestra de 36 resortes escogidos al azar y encontró una fuerza promedio de 571 libras. ¿Debe el distribuidor considerar la posibilidad de que surjan quejas de resortes débiles? Use $\alpha = .05$.

13. Una máquina de ejercicios de un gimnasio esta diseñada para resistir hasta 65 libras. La máquina es usada por niños en las edades de 8 hasta 10 años. Para probar la seguridad de la máquina se midieron los pesos de 36 niños en una muestra aleatoria. Los datos de la muestra revelaron que el peso promedio de los niños (de 8 a 10 años) fue de 67 libras con una DE de 8. ¿Se debe considerar la máquina como segura usando un nivel de significancia de 0.01?

14. Los resultados en las pruebas del SAT de los estudiantes de cierto pueblo están normalmente distribuidas con un promedio de 625 y una DE de 100. Se tomó una muestra aleatoria de 64 estudiantes, quienes tomaron la prueba del SAT después de haber tomado unas tutorías. Los resultados produjeron un puntaje promedio de 632 en la sección de matemáticas. Pruebe a un $\alpha = 0.05$ si el resultado promedio en matemáticas del SAT es significativamente mayor de 625 cuando se ofrecen las tutorías.

15. El balance promedio de las cuentas de ahorros durante 1995 en el banco First State fue de $1300. Una muestra aleatoria de 45 cuentas de ahorros promediaron $1,350 con una DE de $80 durante 1996. Usando un nivel de significancia $\alpha = 0.1$, ¿podemos concluir que el balance promedio de las cuentas de ahorros durante 1996 difiere del balance de las cuantas de ahorros durante 1995?

En los ejercicios (16 - 20) use el método del valor-*P* para probar la hipótesis.

16. La pizzería Mama Jane afirma que el tiempo promedio de entregas es de 20 minutos. Cierto competidor quiere debatir tal afirmación demostrando que en una muestra aleatoria de 16 entregas, el tiempo promedio de entrega fue de 31 minutos con una DE de 10 minutos. Use un $\alpha = 0.01$.

17. El centro de cómputos del área académica de cierta universidad esta solicitando más apoyo financiero de parte de la administración. La solicitud se basa en el reclamo de que al presente el tiempo promedio que un estudiante espera para usar una computadora es mayor de 20 minutos. El Vicepresidente de Finanzas tomó una muestra aleatoria de 16 estudiantes quienes esperaron en el centro para usar una computadora. El calculó el tiempo promedio de espera en 16 minutos con una DE de 7 minutos. Usando un nivel de significancia de 10%, ¿debe el Vicepresidente recomendar más dinero para el centro?

18. La cadena de restaurantes Big Burger reclama que los empleados de más antigüedad trabajan un promedio de 16 horas por semana. Una muestra de 10 empleados de mayor antigüedad trabajando en uno de su restaurantes produjo un tiempo promedio de trabajo por semana de 21 horas con una DE de 5 hrs. Use un nivel de significancia de 1%.

19. Un manufacturero de automóviles anuncia que su último modelo de automóvil compacto promedia más de 40 mpg en la carretera. Para probar esta afirmación, una agencia independiente de protección de los consumidores probó 25 de éstos automóviles. El rendimiento promedio fue 41.3 mpg con una DE de 1.4 mpg en la carretera. Con un nivel de significancia de 5%, pruebe la promoción del manufacturero.

20. Cierto manufacturero de llantas introdujo una nueva llanta cuya vida promedio es mayor de 60,000 millas. Un laboratorio de prueba independiente probó 5 de éstas llantas y los resultados del largo de vida de cada una son los siguientes (en miles de millas):

62, 63, 59, 56, 61

Use un nivel de significancia de 1% para retar el reclamo del manufacturero.

PRUEBA DE UNA PROPORCIÓN

Un político puede estar interesado en conocer si ha habido un aumento en la proporción (porcentaje) de votantes que lo favorecen en las próximas elecciones; una productor de cereales puede querer conocer si ha ocurrido o no una baja en la proporción de clientes que prefieren su marca de cereal; un hospital desea confirmar el reclamo de un manufacturero de medicamentos quien afirma éste cura al 80% de los usuarios. Estos ejemplos son algunas de las situaciones donde nos interesa probar alguna afirmación referente a una proporción. El procedimiento para probar una proporción en una población normal es casi igual al usado para las medias.

¿Cómo probar una proporción?

Podemos usar cualquiera de los siguientes:

1. Método de la región de rechazo (Método 1) ó
2. Método del valor P (Método 2)

A. MÉTODO DE LA REGIÓN DE RECHAZO (MÉTODO 1)

Digamos que p_o es la proporción aceptada o reclamada.

***Paso* 1** Establezca las hipótesis.

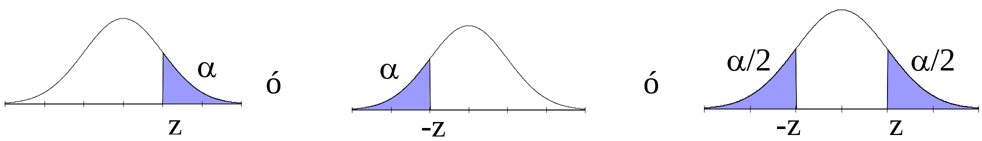
$H_o: \quad p = p_o$

$H_1: \quad p > p_o$ ó

$\quad\quad p < p_o$ ó

$\quad\quad p \neq p_o$

***Paso* 2** Use el nivel de significancia (α) y dibuje la región de rechazo en la curva normal estándar (curva z).



$(H_1 : \mu > \mu_o)$ $(H_1 : \mu < \mu_o)$ $(H_1 : \mu \neq \mu_o)$

***Paso* 3** Calcule el valor z para la proporción muestral $\left(\bar{p} = \dfrac{x}{n}\right)$ usando la fórmula

$$Z = \frac{\bar{p} - p_0}{\sigma_p}, \quad \sigma_p = \sqrt{\frac{p_0(1-p_0)}{n}}$$

***Paso* 4** Dibuje este valor de *z* en el diagrama de la región de rechazo (Paso 2).

***Paso* 5** Si el valor *z* cae dentro de la región de rechazo (sombreada), entonces rechace H_o. Si cae fuera de la región sombreada, entonces no rechace H_o.

***Paso* 6** Escriba la conclusión de la prueba.

Ejemplo 1: Prueba la hipótesis $H_0 : p = 0.4$

$H_1 : p \neq 0.4$

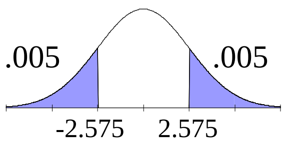
Presuma que = 0.45, n = 200, y α = .01.

Solución:

Paso 1 $H_0 : p = 0.4$

$H_1 : p \neq 0.4$

Paso 2 Usando α = .01, el diagrama de la región de rechazo es:

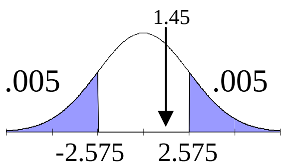


Paso 3 Calculando el valor *z* para la proporción muestral $\bar{p}$ = 0.45), obtenemos:

$$\sigma_p = \sqrt{\frac{0.4(1-0.4)}{200}} = 0.0346$$

$$Z = \frac{0.45 - 0.4}{0.0346} = 1.45$$

Paso 4 Dibujando z = 1.45 en el diagrama de la región de rechazo (Paso 2) obtenemos:



Paso 5 Como el valor *z* está fuera de la región de rechazo (sombreada), por lo tanto no rechazamos H_o.

Paso 6 La proporción en la población es 0.4.

B. MÉTODO DEL VALOR P (MÉTODO 2)

Dejemos que p_0 sea la proporción aceptada o reclamada.

***Paso* 1** Establezca las hipótesis: $H_0: \quad p = p_0$

$H_1: \quad p > p_0$ ó

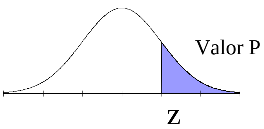
$p < p_0$ ó

$p \neq p_0$

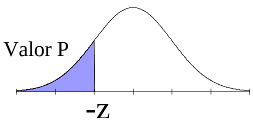
***Paso* 2** Calcule el valor z para la proporción muestral $\left(\overline{p} = \dfrac{x}{n}\right)$ usando la fórmula:

$$Z = \frac{\overline{p} - p_0}{\sigma_p}, \text{ donde } \sigma_p = \sqrt{\frac{p_0(1-p_0)}{n}} .$$

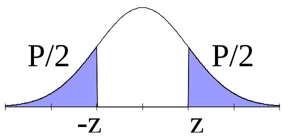
***Paso* 3** Usando la hipótesis alterna dibuja la región bajo la curva z que representa los valores extremos.



ó



ó



$(H_1 : \mu > \mu_o)$ $\qquad$ $(H_1 : \mu < \mu_o)$ $\qquad$ $(H_1 : \mu \neq \mu_o)$

***Paso* 4** El valor P = al área de la cola sombreada (s) en el Paso 3.

***Paso* 5** Si el valor $P < \alpha$, entonces rechaza H_0

Si el valor $P \geq \alpha$, entonces no rechaces H_0.

***Paso* 6** Escribe la conclusión de la prueba.

Ejemplo 1: Pruebe la hipótesis $H_0: p = 0.4$

$H_1: p \neq 0.4$

Presuma que = 0.45, $n = 200$, y $\alpha = 0.01$.

Solución:

Paso 1 $H_0: p = 0.4$

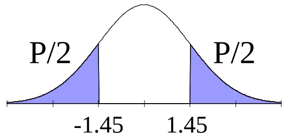
$H_1: p \neq 0.4$

Paso 2 Calculando el valor z de $\overline{p}$, obtenemos

$$\sigma_p = \sqrt{\frac{0.4(1-0.4)}{200}} = 0.0346$$

$$Z = \frac{0.45 - 0.4}{0.0346} = 1.45$$

Paso 3 La región bajo la curva z que contiene los valores extremos de es



Paso 4 El valor P = suma de las áreas de las regiones sombreadas en el Paso 3.

= 2(el área a la derecha de 1.45)
= 2(0.5 – .4265)
= 0.147

Paso 5 Como el valor P es mayor que α, entonces no podemos rechazar H_0.

Paso 6 La proporción en la población es 0.4.

EJERCICIOS

En los ejercicios (1-5) use el método de la región de rechazo para probar la hipótesis.

1. H_0 : 0.6

$H_1 : p \neq 0.6$, = 0.65, $n = 100$, y $\alpha = 0.01$

2. $H_0 : p = 0.29$

$H_1 : p \neq 0.29$, = 0.26, $n = 90$, y $\alpha = 0.01$

3. $H_0 : p = 0.36$

$H_1 : p < 0.36$, = 0.34, $n = 630$, y $\alpha = 0.05$

4. Un manufacturero de juguetes Tailandés reclama que solo un 10% de los osos de juguete hechos para hablar están defectuosos. Cuatrocientos de éstos juguetes se sometieron a prueba de forma aleatoria y se encontró que 50 estaban defectuosos. Pruebe el reclamo del manufacturero con un nivel de significancia de 5%.

5. Una agencia de empleos afirma que el 80% de todas las solicitudes hechas por mujeres con hijos prefieren trabajos a tiempo parcial. En una muestra aleatoria de 200 solicitantes mujeres con niños, se encontró que 110 prefirieron trabajos a tiempo parcial. Pruebe la hipótesis de la agencia con un nivel de significancia de 5%.

En los ejercicios (6 - 10) use el método del valor-*p* para pruebas de hipótesis.

6. $H_0 : p = 0.2$

$H_1 : p > 0.2$, = 0.245, $n = 400$, y $\alpha = 0.01$

7. $H_0 : p = 0.55$

$H_1 : p < 0.55,\ x = 175,\ n = 300,$ y $\alpha = 0.05$

8. $H_0 : p = 0.2$

$H_1 : p \neq 0.2,\ x = 235,\ n = 1000,$ y $\alpha = 0.02$

9. Nacionalmente, un 16 % de los hogares tiene una computadora personal. En una muestra aleatoria de 80 hogares en Baltimore, solo 13 poseían una computadora personal. Con un nivel de significancia de 5%, pruebe si el porciento de hogares en Baltimore que tienen computadoras personales es menor que el porcentaje nacional.

10. El registrador de cierta universidad ha dicho que esta dispuesto a permitir una sección del curso ESTADISTICA una vez a la semana si más del 65% de los estudiantes matriculados en el curso expresan que prefieren el curso una vez a la semana, en vez de dos veces a la semana. En una muestra aleatoria de 40 estudiantes, 26 indicaron su preferencia de una vez a la semana. Usando un nivel de significancia de 0.01, debe el registrador autorizar el ofrecimiento del curso ESTADISTICA una vez a la semana?