Hipótesis estadísticas

Un **test estadístico** es un procedimiento para, a partir de una muestra aleatoria y significativa, **extraer conclusiones** que permitan **aceptar o rechazar una hipótesis** previamente emitida sobre el valor de un parámetro desconocido de una población.

La hipótesis emitida se designa por $\mathbf{H_0}$ y se llama hipótesis nula.

La hipótesis contraria se designa por $\mathbf{H_1}$ y se llama hipótesis alternativa.

Contrastes de hipótesis

1. Enunciar la hipótesis nula H_0 y la alternativa H_1.

Bilateral	$H_0 = k$	$H_1 \neq k$
Unilateral	$H_0 \geq k$	$H_1 < k$
	$H_0 \leq k$	$H_1 > k$

2. A partir de un nivel de confianza $1 - \alpha$ o el de significación α. Determinar:

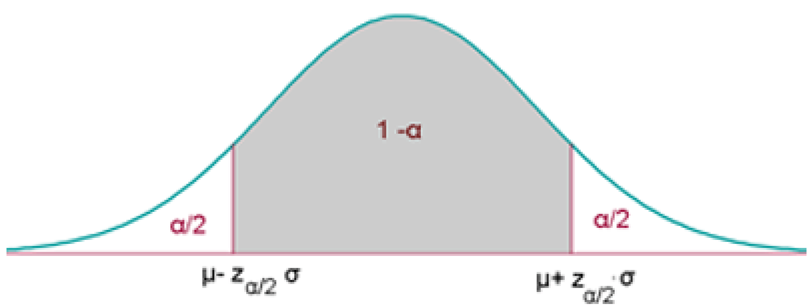
El valor $z_{\alpha/2}$ (bilaterales), o bien z_α (unilaterales)

La zona de aceptación del parámetro μ o p.

3. **Calcular: x o p', a partir de la muestra.**

4. **Si el valor del parámetro muestral está dentro de la zona de la aceptación, se acepta la hipótesis con un nivel de significación α. Si no, se rechaza**.

Contraste bilateral : Se presenta cuando la hipótesis nula es del tipo **H_0: μ = k** (o bien **H_0: p = k**) y la hipótesis alternativa, por tanto, es del tipo **H_1: μ≠ k** (o bien **H_1: p≠ k**).



El nivel de significación α se concentra en dos partes (o colas) simétricas respecto de la media. La **región de aceptación** en este caso no es más que el correspondiente intervalo de probabilidad para x o p', es decir:

$$\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

o bien:

$$\left(p - z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}},\ p + z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}\right)$$

Ejemplo Se sabe que la desviación típica de las notas de cierto examen de Matemáticas es 2,4. Para una muestra de 36 estudiantes se obtuvo una nota media de 5,6. ¿Sirven estos datos para confirmar la hipótesis de que la nota media del examen fue de 6, con un nivel de confianza del 95%?

1. Enunciamos las hipótesis nula y alternativa:

H_0 : μ = 6 La nota media no ha variado.

H_1 : μ ≠ 6 La nota media ha variado.

2. Zona de aceptación

Para **α = 0.05**, le corresponde un valor crítico: **$z_{\alpha/2}$ = 1.96**.

Determinamos el intervalo de confianza para la media:

(6 − 1,96 · 0,4 ; 6 + 1,96 · 0,4) = (5,22 ; 6,78)

3. Verificación.

Valor obtenido de la media de la muestra: **5,6** .

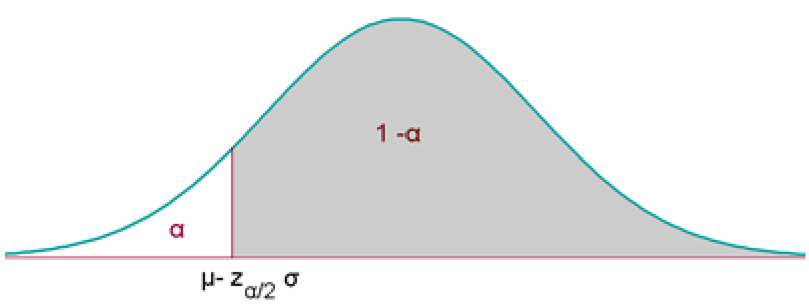
4. Decisión

Aceptamos la hipótesis nula H_0, con un nivel de significación del 5%.

Contraste unilateral: Caso 1 La **hipótesis nula** es del tipo **H_0: μ ≥ k** (o bien **H_0: p ≥ k**). La **hipótesis alternativa**, por tanto, es del tipo **H_1: μ < k** (o bien **H_1: p < k**).

Valores críticos

1 − α	α	z_α
0.90	0.10	1.28
0.95	0.05	1.645
0.99	0.01	2.33



El nivel de significación **α** se concentra en una parte o cola.

La región de aceptación en este caso será:

$$\left(\mu - z_{\alpha} \bullet \frac{\sigma}{\sqrt{n}},\ \infty\right)$$

o bien:

$$\left(p - z_{\alpha} \bullet \sqrt{\frac{pq}{n}},\ \infty\right)$$

Ejemplo : Un sociólogo ha pronosticado, que en una determinada ciudad, el nivel de abstención en las próximas elecciones será del 40% como mínimo. Se elige al azar una muestra aleatoria de 200 individuos, con derecho a voto, 75 de los cuales estarían dispuestos a votar. Determinar con un nivel de significación del 1%, si se puede admitir el pronóstico.

1. Enunciamos las hipótesis nula y alternativa:

H_0 : μ ≥ 0.40 La abstención será como mínimo del 40%.

H_1 : μ < 0.40 La abstención será como máximo del 40%;

2. Zona de aceptación

Para **α = 0.01**, le corresponde un valor crítico: **z_α = 2.33**.

Determinamos el intervalo de confianza para la media:

$$\left(0.4-2.33\cdot\sqrt{\frac{0.4\cdot0.6}{200}},\ \infty\right)=(0.3192;\ \infty)$$

3. Verificación.

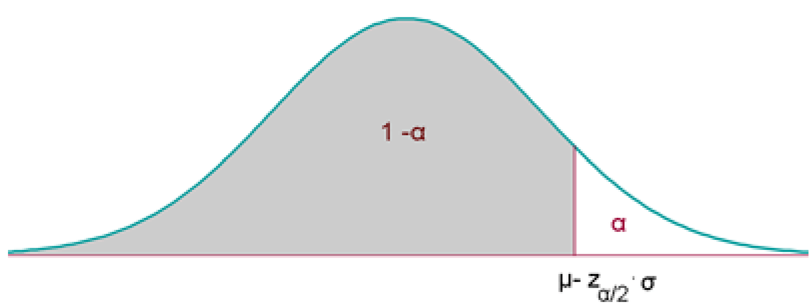
$$p'=\frac{125}{200}=0.625$$

4. Decisión

Aceptamos la hipótesis nula H_0. Podemos afirmar, con un nivel de significación del 1%, que la La abstención será como mínimo del 40%.

Caso 2La hipótesis nula es del tipo **H_0: $\mu \leq k$** (o bien **H_0: $p \leq k$**).

La hipótesis alternativa, por tanto, es del tipo **H_1: $\mu > k$** (o bien **H_1: $p > k$**).



El nivel de significación α se concentra en la otra parte o cola.

La región de aceptación en este caso será:

$$\left(-\infty,\ \mu + z_{\alpha} \bullet \frac{\sigma}{\sqrt{n}}\right)$$

o bien:

$$\left(-\infty\ ,\ p + z_{\alpha} \bullet \sqrt{\frac{pq}{n}}\right)$$

Ejemplo: Un informe indica que el precio medio del pasaje de avión entre Ushuaia y Córdoba es, como máximo, de 120 dólares con una desviación típica de 40 dólares. Se toma una muestra de 100 viajeros y se obtiene que la media de los precios de sus pasajes es de 128 dólares. ¿Se puede aceptar, con un nivel de significación igual a 0,1, la afirmación de partida?

1. Enunciamos las hipótesis nula y alternativa:

H_0 : μ ≤ 120

H_1 : μ > 120

2. Zona de aceptación

Para **α = 0.1**, le corresponde un valor crítico: **z_α = 1.28** .

Determinamos el intervalo de confianza:

$$\left(-\infty; 120 + 1.28\frac{40}{\sqrt{100}}\right) = (-\infty; 125.12)$$

3. Verificación.

Valor obtenido de la media de la muestra: **128 dólares** .

4. Decisión

No aceptamos la hipótesis nula H_0. Con un nivel de

Errores de tipo I y tipo II

Error de tipo I. Se comete cuando la **hipótesis nula es verdadera** y, como consecuencia del contraste, **se rechaza**.

Error de tipo II. Se comete cuando la **hipótesis nula es falsa** y, como consecuencia del contraste **se acepta**.

H_0	Verdadera	Falsa
Aceptar	**Decisión correcta** **Probabilidad = $1 - \alpha$**	**Decisión incorrecta:** **ERROR DE TIPO II**

Rechazar	**ERROR DE TIPO I** **Probabilidad = α**	**Decisión correcta**

La **probabilidad** de cometer **Error de tipo I** es el **nivel de significación α**.

La probabilidad de cometer **Error de tipo II** depende del verdadero valor del parámetro. Se hace **tanto menor cuanto mayor sea n**.

1.-Un criador de pollos sabe por experiencia que el peso de los pollos de cinco meses es 4,35 libras. Los pesos siguen una distribución normal. Para tratar de aumentar el peso de dichas aves se le agrega un aditivo al alimento. En una muestra de pollos de cinco meses se obtuvieron los siguientes pesos (en libras).

2.-Una empresa que se dedica a hacer en cuestas se queja de que un agente realiza en promedio 53 encuestas por semana. Se ha introducido una forma más moderna de realizar las encuetas y la empresa quiere evaluar su efectividad. Los números de encuestas realizadas en una semana por una muestra aleatoria de agentes son:53 57 55 50 58 54 60 52 59 62 60 60 51 59 56 . En el nivel de significancia 0,05, puede concluirse que la cantidad media de entrevistas realizadas por los agentes es superior a 53 por semana? Evalúe el valor p.

Soluciones

1

n = 10

u = 4,35

Xmed = 43,68/10 = 4,368

Peso librasX	X-Xmed	(X-Xmed)^2	X^2
4,41	0,042	0,001764	19,4481
4,37	0,002	4E-06	19,0969
4,33	-0,038	0,001444	18,7489
4,35	-0,018	0,000324	18,9225
4,3	-0,068	0,004624	18,49
4,39	0,022	0,000484	19,2721
4,36	-0,008	6,4E-05	19,0096
4,38	0,012	0,000144	19,1844
4,4	0,032	0,001024	19,36
4,39	0,022	0,000484	19,2721
43,68		0,01036	190,8046

X med = 4,368

$$s = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}} = \sqrt{\frac{0.01036}{10-1}}$$

$$s = \sqrt{0.00115} = 0.0339$$

Planteamiento de hipótesis

$$H_o : u \leq 4.35$$
$$H_1 : u > 4.35$$

a) Prueba de una cola

b) nivel de significancia 0,01

c) Estadístico de prueba

$$t = \frac{\bar{X}-u}{s/\sqrt{n}} = \frac{4.368-4.35}{0.0339/\sqrt{10}} = \frac{0.018}{0.01072} = 1.68$$

Area = 0,4535

d) Plantear la regla de desición:

alfa = 0,01 y gl = n - 1 =10 -1 = 9

Si t > 2,821 Se rechaza Ho y se acepta H1

Tomar la decisión:

Como t (1,68) < 2,821se acepta la hipótesis nula y se rechaza H1 y se concluye que el aditivo no aumenta el peso medio de los pollos en un 4,35.

Valor p = 1,68 es 0,4535

p = 0,50 - 0,4535 = 0,046

2.-

u = 53
n = 15
Xmed = 56,4

$$s=\sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}}=\sqrt{\frac{1956}{15-1}}=\sqrt{13971}=3.73$$

Planteamiento de hipótesis

$$H_o : u \le 53$$
$$H_1 : u > 53$$

a) Prueba de una cola
b) nivel de significancia 0,05
c) Estadístico de prueba

$$t=\frac{\bar{X}-u}{s/\sqrt{n}}=\frac{56.4-53}{3.73/\sqrt{15}}=\frac{3.4}{0.963}=3.53$$

# encuestas	X-Xmed	(X-Xmed)^2	X^2
53	-3,4	11,56	2809
57	0,6	0,36	3249
50	-6,4	40,96	2500
55	-1,4	1,96	3025
58	1,6	2,56	3364
54	-2,4	5,76	2916
60	3,6	12,96	3600
52	-4,4	19,36	2704
59	2,6	6,76	3481
62	5,6	31,36	3844
60	3,6	12,96	3600
60	3,6	12,96	3600
51	-5,4	29,16	2601
59	2,6	6,76	3481
56	-0,4	0,16	3136
846		195,6	47910

Xmed = 846/15 = 56,4
Area = 0,4989

d) Plantear la regla de desición:
alfa = 0,05 y gl = n - 1 =15 -1 = 14

Si t < 1,761 Se rechaza Ho y se acepta H1

Tomar la decisión:

Como t (3,53) > 1,761 se rechaza la hipótesis nula y se acepta H1 y se concluye que la cantidad media de entrevistas realizadas por los agentes es mayor a 53 por semana.

Valor p = 1,761 es 0,4989
p = 0,50 - 0,4989 = 0,0011

3.- Lisa Mendez es directora de presupuesto en la empresa New Process Company, desea comparar los gastos diarios de transporte del equipo de ventas y del personal de cobranza. Recopiló la siguiente información muestral (importe en dólares).

Ventas ($)	131	135	146	165	136	142	
Cobranza ($)	130	102	129	143	149	120	139

Al nivel de significancia de 0,10, puede concluirse que los gastos medios diarios del equipo de ventas son mayores? cuál es el valor p?

Solución

ventas	cobranzas	
131	130	
135	102	
146	129	
165	143	
136	149	
142	120	
	139	
142,5	130,285714	promedio
12,2433656	15,7872764	desviación st
n = 6	n = 7	
alfa = 0,10		

Prueba de hipótesis

$$H_o : u1 \leq u2$$
$$H_1 : u1 > u2$$

a) Es esta una prueba de una o de dos colas

Esta es una prueba a una sola cola

b) Establezca la regla de desición

Si Z > que el valor crítico, se rechaza la hipótesis nula y se acepta H1

c) Calcule el valor del estadístico de prueba

Datos:

n1 = 6 n2 = 7

Prom 1 = 142,5 Prom 2 = 130,3

s1 = 12,2 s2 = 15,8

Alfa = 0.10

Grados de libertad = 6+7 - 2 =11 =1,363 de la tabla con 0,10

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s_p{}^2 = \frac{(6-1)148.84 + (7-1)249.64}{6+7-2} = \frac{744,2 + 1497.84}{11} = 203.82$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p{}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$t = \frac{142,5 - 130,3}{\sqrt{203.82\left(\frac{1}{6} + \frac{1}{7}\right)}} = \frac{12,2}{7,819} = 1,56$$

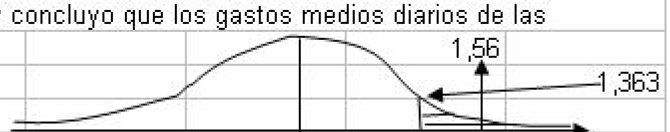
d) Cuál es la desición respecto a la hipótesis nula

t calculado 1,56 es mayor que 1,363 rechazo la hipóteis nula

Por lo tanto rechazo la hipótesis nula y concluyo que los gastos medios diarios de las ventas realizadas son mayores.

e) Cual es su valor p

El valor se encuentra entre 0,10 y 0,5



4.-De una población se toma una muestra de 40 observaciones. La media muestral es de 102 y la desviación estándar 5. De otra población se toma una muestra de 50 observaciones. La media mustral es ahora 99 y la desviación estándar es 6. Realice la siguiente prueba de hipótesis usando como nivel de significancia 0,04.

Ho: u1 = u2

Ho: u1 ≠ u2

a) Es esta una prueba de una o de dos colas?

Esta es una prueba de hipótesis de dos colas

b) Establezca la regla de decisión

Si Z > que le valor crítico, se rechaza la hipótesis nula y se acepta la hipótesis alternativa

c) Calcule el valor del estadístico de prueba

Si Z > que el valor crítico, se rechaza la hipótesis nula y se acepta H1

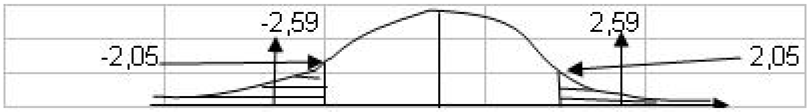
Datos:	
n1 = 40	n2 = 50
Prom 1 = 102	Prom 2 = 99
s1 = 5	s2 = 6

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$z = \frac{102-99}{\sqrt{\frac{(5)^2}{40} + \frac{(6)^2}{50}}} = \frac{3}{\sqrt{0,625+0,72}} = \frac{3}{\sqrt{1,345}} = 2,59$$

Como su valor calculado Z (2,59) > 2,05; se rechaza la hipótesis nula y se acepta la hipótesis alternativa

Si Z tabulada es 0,5 - 0,02 = 0,48 este valor en la tabla es 2,05



Cuál es el valor p?

Z = 2,59 Area 0,4952

0,5 - 0,4952 = 0,0048 * 2 = 0,0096

Prueba de Hipótesis para Medias con Distribución T de Student

La Prueba de Hipótesis para medias usando Distribución t de Student se usa cuando se cumplen las siguientes dos condiciones:
Es posible calcular las media y la desviación estándar a partir de la muestra. El tamaño de la muestra es menor a 30.

El procedimiento obedece a los 5 pasos esenciales:

Paso 1:
Plantear Hipótesis Nula (Ho) e Hipótesis Alternativa (Hi).
La Hipótesis alternativa plantea matemáticamente lo que queremos demostrar.
La Hipótesis nula plantea exactamente lo contrario.
Paso 2:
Determinar Nivel de Significancia. (Rango de aceptación de hipótesis alternativa)

α

Casi siempre lo proporciona el problema, y normalmente se considera:
- 0.05 para proyectos de investigación
- 0.01 para aseguramiento de calidad
-0.10 para encuestas de mercadotecnia y políticas.

Paso 3:
Evidencia de la Muestra. Se calcula la media y la desviación estándar a partir de la muestra.
Paso 4:
Se aplica la Distribución t de Student:

$$t^* = \frac{\bar{X} - \mu}{\frac{S_x}{\sqrt{n}}}$$

$$Grados\ de\ Libertad = \ df = n - 1$$

Sabiendo que:

$$\bar{X} = Media$$

$$\mu = Valor\ a\ \ analizar$$

$$S_x = Desviación\ Estándar$$

$$\bar{X} = Media$$

$$n = Tamaño\ de\ muestra$$

Paso 5:
En base a la evidencia disponible se buscan las regiones de aceptación o rechazo.

Conclusión:
Se acepta o rechaza la Hipótesis Nula.
Se concluye de acuerdo ala información de la Hipótesis Alternativa.
<u>Ejemplo:</u>

Se aplica una prueba de autoestima a 25 personas quienes obtienen una calificación promedio de 62.1 con una desviación estándar de 5.83 Se sabe que el valor correcto de la prueba debe ser mayor a 60. ¿Existe suficiente evidencia para comprobar que no hay problemas de autoestima en el grupo seleccionado?

Considera un nivel de significancia de 0.05

Paso 1:

Hipótesis Alternativa (Hi) :
Lo que se quiere comprobar

El grupo no tiene problemas de autoestima.

Valor de autoestima mayor a 60.

Valor de Autoestima > 60

Hipótesis Nula (Ho):
Lo contrario a la Hipótesis Alternativa

El grupo tiene problemas de autoestima.

Valor de autoestima < = 60.

Paso 2:

Determinar nivel de significancia:

$\alpha = 0.05$

Paso 3:

Evidencia Muestral

$\bar{X} = 62.1 \quad y \quad S_x = 5.83$

Paso 4:
Aplicando la Distribución de Probabilidad

Calculando t*:

$$t^* = \frac{\bar{X} - \mu}{\frac{S_x}{\sqrt{n}}}$$

$$t^* = \frac{62.1 - 60}{\frac{5.83}{\sqrt{25}}} = \frac{2.1}{1.166} = 1.8$$

$$df = n - 1 = 25 - 1 = 24$$

Paso 5: Toma de Decisión.

Se buscan las regiones de aceptación o rechazo.

Buscando en la Tabla t:

Nivel de significancia 0.05

Grados de Libertad: 25 -1 = 24

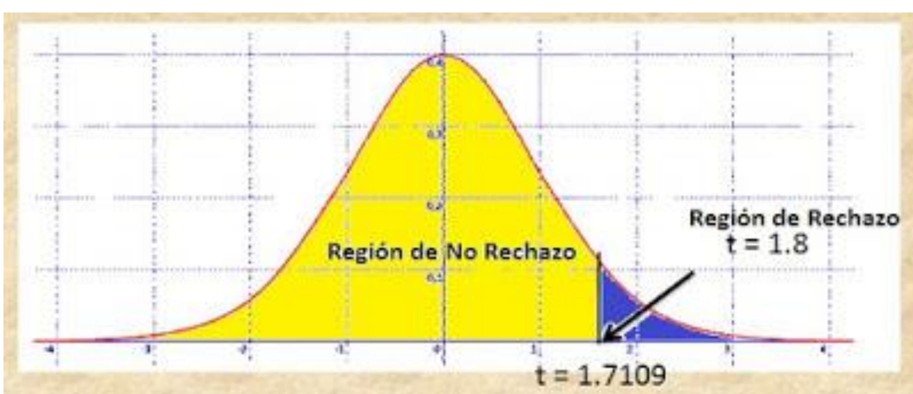
Tabla de valores criticos de t para una cola

Grados de Libertad	0.25	0.10	0.05	0.025	0
1	1.0000	3.0777	[redacted]	12.7062	31.
2	0.8165	1.8856		4.3027	6.9
3	0.7649	1.6377		3.1824	4.5
4	0.7407	1.5332		2.7764	3.7
5	0.7267	1.4759		2.5706	3.3
6	0.7176	1.4398		2.4469	3.1
7	0.7111	1.4149		2.3646	2.9
8	0.7064	1.3968		2.3060	2.8
9	0.7027	1.3830		2.2622	2.8
10	0.6998	1.3722		2.2281	2.7
11	0.6974	1.3634		2.2010	2.7
12	0.6955	1.3562		2.1788	2.6
13	0.6938	1.3502		2.1604	2.6
14	0.6924	1.3450		2.1448	2.6
15	0.6912	1.3406		2.1315	2.6
16	0.6901	1.3368		2.1199	2.5
17	0.6892	1.3334		2.1098	2.5
18	0.6884	1.3304		2.1009	2.5
19	0.6876	1.3277		2.0930	2.5
20	0.6870	1.3253		2.0860	2.5
21	0.6864	1.3232		2.0796	2.5
22	0.6858	1.3212		2.0739	2.5
23	0.6853	1.3195		2.0687	2.4
24	[redacted]		1.7109	2.0639	2.4
25	0.6844	1.3163	1.7081	2.0595	2.4

Valor crítico de t = 1.7109

[(Para ver la tabla de valores críticos de t)]()

Colocando en la campana el valor crítico de t y el valor de t calculado en el paso 4 (t=1.8)



Se trata de un problema de una cola, por lo que únicamente tenemos una región de rechazo.

La región de rechazo se ubica de acuerdo al signo de la Hipótesis Alternativa. (>)

El valor de t calculado se encuentra en la Región de Rechazo.

CONCLUSIÓN

SE RECHAZA HIPÓTESIS NULA

SE ACEPTA HIPÓTESIS ALTERNATIVA

De acuerdo a la muestra, existe suficiente evidencia para demostrar que el grupo no tiene problemas de autoestima con un nivel de significancia de 0.05.