Pruebas de Hipótesis

Introducción

- Muchas veces la inferencia que se debe realizar no se refiere a la estimación de un parámetro, si no que se deben tomar decisiones sobre afirmaciones hechas sobre un parámetro.

- Se debe decidir, con base en la evidencia experimental, si una afirmación (hipótesis) hecha acerca de un parámetro es falsa o verdadera.

Conceptos

- **Pruebas de hipótesis:** procedimiento basado en evidencia de la muestra y la teoría de probabilidades para determinar si la hipótesis es una afirmación razonable.

- **Hipótesis estadística:** es una afirmación o conjetura acerca de una o más poblaciones.
- **Hipótesis Nula:** es cualquier hipótesis que se desea probar. Se denota por H_0.
- **Hipótesis Alternativa:** es la hipótesis que se acepta cuando la hipótesis nula es rechazada. Se denota por H_1.

Ejemplo	Se sabe por estudios previos que los recién nacidos de cierta población tienen una talla promedio de 49,5 cm. Una enfermera estudió un grupo de 40 recién nacidos, y obtuvo una media de 53,4 cm. La enfermera desea saber si estos resultados apoyan los estudios previos. ¿Cuáles serían sus hipótesis nula y alternativa?
Solución	En esta situación la enfermera tiene un valor poblacional establecido, que es que los recién nacidos miden, en promedio, una talla de 49,5 cm. Por tanto, su hipótesis nula será: H_0: La talla media de los recién nacidos es 49,5 cm. Pero los datos recopilados sugieren que este promedio podría ser mayor que 49,5 cm, por lo que, de descartar la hipótesis nula anterior, se aceptaría la hipótesis alternativa: H_1: La talla media de los recién nacidos es mayor que 49,5 cm. Generalmente las hipótesis se expresan en términos de símbolos: H_0: $\mu = 49{,}5$ H_1: $\mu > 49{,}5$

Probabilidad

- En las pruebas de hipótesis las decisiones se toman con base en datos muestrales.

- Producto del azar, en las pruebas de hipótesis existe el riesgo de tomar una decisión equivocada, cuando sea el caso en los cuales la media muestral queda muy alejada del valor verdadero.

Tipos de errores

- **Tipo I:** es el error que se comete cuando se rechaza una hipótesis que es correcta y la probabilidad de cometer este error se denota por este error se denota por α.

- **Tipo II:** es el error que se comete cuando se acepta una hipótesis que es incorrecta y la probabilidad de cometer este error se denota β.

Ejemplo	Una empresa fabrica bombillos. Cada bombillo tiene una vida esperada de 1000 horas, pero algunos clientes se han quejado de que los bombillos se queman antes de las 1000 horas. La gerencia decide tomar una muestra y probar la hipótesis nula de que los bombillos tienen una vida media de 1000 horas, contra la hipótesis alterna de que la vida media de los bombillos es menor que dicha especificación. ¿Cómo podrían darse y qué significan los errores tipo I y tipo II en esta situación?
Solución	En esta situación los errores tipo I y tipo II podrían darse si la muestra no representa bien a la población. Esto puede darse de los modos siguientes: 1. El proceso de producción de la empresa está bien controlado, y la vida media de los bombillos es 1000 horas, pero en la muestra usada en la prueba de hipótesis se seleccionaron, por cuestión del azar, muchos bombillos con una vida media inferior a 1000 horas, por lo que se rechazó la hipótesis nula de que la vida media de los bombillos es 1000 horas, a pesar de que era verdadera. Este es el **error tipo I**. Este error llevaría a la empresa a tratar de mejorar su producción innecesariamente, lo cual le generaría costos adicionales. 2. El proceso de producción de la empresa no está bien controlado, por lo que, efectivamente, la vida media de los bombillos es inferior a 1000 horas, como lo han indicado los clientes que se han quejado, pero en la muestra, por cuestión del azar, se seleccionaron muchos bombillos con una media cercana a 1000 horas, por lo que no se rechazó la hipótesis nula, a pesar de que era falsa. Este es el **error tipo II**. Este error llevaría a la empresa a no mejorar una producción que sí requiere mejoras, por lo cual sus clientes podrían dejar de comprar sus productos.

Procedimiento para P.H. sobre la media

- **Paso I.** plantear hipótesis (nula y alternativa).
- **Paso II.** Determinar el nivel de significancia.
- **Paso III.** Identificar el estadístico de prueba.

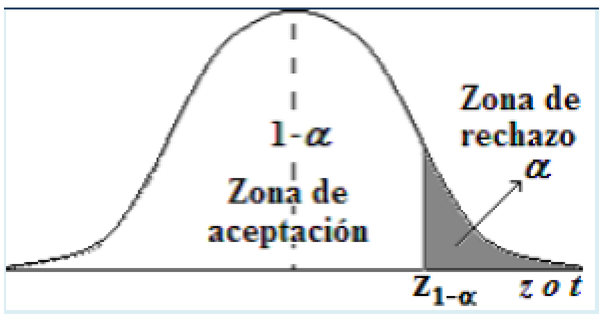
$$z_c = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}} \quad \text{con } n \geq 30 \text{ con } \sigma \text{ conocida o con } n < 30 \text{ y } \sigma \text{ conocida}$$

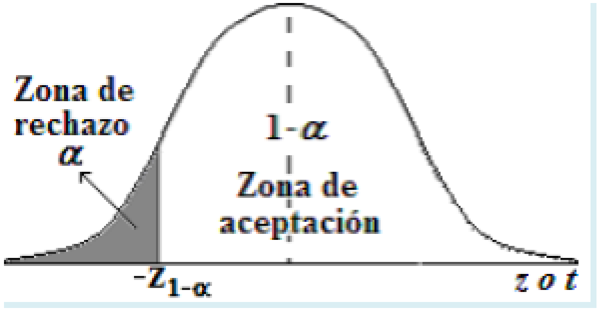
$$z_c = \frac{\overline{x} - \mu}{s / \sqrt{n}} \quad \text{con } n \geq 30 \text{ con } \sigma \text{ desconocida}$$

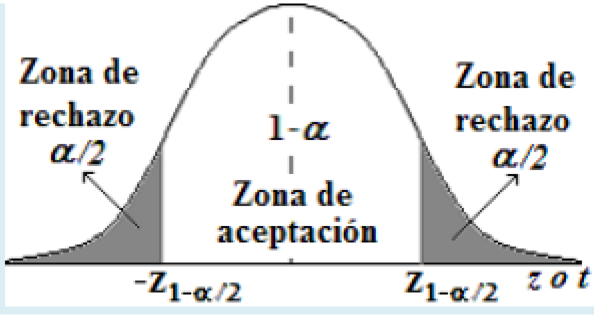
$$t_c = \frac{\overline{x} - \mu}{s / \sqrt{n}} \quad \text{con } n < 30 \text{ y } \sigma \text{ desconocida}$$

Procedimiento para P.H. sobre la media

- **Paso IV.** Se formula una regla para tomar la decisión.







- **Paso V.** Se toma una muestra y se llega a una decisión: se acepta o se rechaza la hipótesis nula.

P.H. para la proporción poblacional

- Se debe comprobar, con base en los resultados obtenidos en una muestra, si el valor verdadero de una proporción es igual a una constante determinada o si las proporciones de dos o más poblaciones son iguales.

- Estadístico de prueba para la proporción poblacional:

x = número de eventos observados
n = tamaño de la muestra
P = proporción poblacional
Q = complemento de P es decir, $Q = 1 - P$

$$z = \frac{x - nP}{\sqrt{nPQ}}$$

P.H. para la diferencia entre medias

- Procedimiento para probar si la diferencia observada entre dos medias muestrales es estadísticamente significativa.

- Ejemplo: probar que la productividad en dos plantas ubicados en distintos países es diferente, o en una tiende hacer mayor que en el otro. Es importante determinar si las diferencias son significativas, o si pueden ser atribuidas al azar.

Ejemplo

Una empresa posee operaciones en dos países distintos y en cada país posee una planta de producción. En uno de los países se tienen 2000 empleados y en el otro 3000. En ambas plantas se realizan los mismos procesos, pero se ha observado que, según algunos datos muestrales, la productividad de los operarios tiende a ser mayor en uno de los países que en el otro.

Los datos recopilados se muestran en la tabla (la media y la desviación estándar se expresan en número de unidades producidas correctamente por hora):

Planta de producción	Tamaño de muestra n	Media $\overline{x}$	Desviación estándar σ
En el país 1	40	22	3,1
En el país 2	50	31	4,2

Determine, a un nivel de significación del 5%, si se presenta diferencia entre los dos promedios.

Solución

Se tiene que $n_1 = 40$, $n_2 = 50$, $\bar{x}_1 = 22$, $\bar{x}_2 = 31$, $\sigma_1 = 3{,}1$ y $\sigma_2 = 4{,}2$.

El ejercicio busca determinar si existe diferencia, por lo que se tendrá que probar si $\delta = 0$.

Además, se indica que $\alpha = 0{,}05$.

Entonces, se plantea la hipótesis nula como:

$$H_0: \mu_1 - \mu_2 = 0$$

Como $\bar{x}_1 = 22 < \bar{x}_2 = 31$, entonces se formula la hipótesis alternativa como:

$$H_1: \mu_1 - \mu_2 < 0$$

Se usa Z porque aunque las desviaciones estándar poblacionales son desconocidas, se tienen muestras grandes ($n \geq 30$):

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{(22 - 31) - 0}{\sqrt{\dfrac{3{,}1^2}{40} + \dfrac{4{,}2^2}{50}}} = -11{,}69$$

De la tabla normal con un nivel de significación del 5% se obtiene $Z_\alpha = 1{,}645$. El valor de Z calculado con la fórmula es menor que el Z_α por lo tanto cae en la zona de rechazo de la hipótesis nula. Se concluye que se rechaza H_o con $\alpha = 0{,}05$. Se puede decir que existe evidencia suficiente para creer que la productividad en el primer país es menor que en el segundo.

Diferencia entre medias (muest. pequeñas)

- Cuando n_1 o n_2, o ambas, son menores de 30 y se desconocen las varianzas poblacionales, se usa el estadístico *t*, siempre que se pueda suponer razonablemente que las poblaciones son normales y que $\sigma_1 = \sigma_2 = \sigma$. El estadístico de prueba corresponde a:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- Los grados de libertad son $gl = n_1 + n_2 - 2$. se emplea la tabla *t* para obtener los valores t tabulares.

Ejemplo

Una empresa desea capacitar en gestión de proyectos a todos sus profesionales. Una muestra de 15 empleados realizó la capacitación y efectuaron la prueba para obtener la certificación en dicho tema. De los 15 colaboradores que hicieron, 9 la realizaron en modalidad presencial (asistiendo a clases) y 6 la efectuaron en línea (a través de internet). La tabla muestra los resultados obtenidos en la prueba final de cada curso.

Presencial	79	88	54	81	73	56	79	64	58
En línea	70	80	72	52	70	61			

El departamento de recursos humanos desea saber si una modalidad de estudio es más efectiva que la otra. Utilice un nivel de significación del 5%.

Solución

Se tienen los datos:

Modalidad presencial: $n_1 = 9$, $\bar{x}_1 = 70{,}2$, $s_1 = 12{,}5$
Modalidad en línea: $n_2 = 6$, $\bar{x}_2 = 67{,}50$, $s_2 = 9{,}71$

Además, $\alpha = 0{,}05$.

Como no se especifica el valor de la diferencia, puede suponerse que va a ser cero, por lo que $\delta = 0$. Además como $\bar{x}_1 > \bar{x}_2$, se plantean las hipótesis como:

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 - \mu_2 > 0$$

También, podrían plantearse las hipótesis como:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 > \mu_2$$

Debido a que las desviaciones estándar poblacionales son desconocidas y se tienen muestras pequeñas ($n < 30$) se usa t. Para esto se supone que las poblaciones son normales y que $\sigma_1 = \sigma_2$. Se calcula:

$$\sigma^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(9-1)12{,}5^2 + (6-1)9{,}71^2}{9+6-2} = 132{,}42$$

Luego se calcula t:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{(70{,}2 - 67{,}5) - 0}{\sqrt{\frac{132{,}42}{9} + \frac{132{,}42}{6}}} = 0{,}45$$

Aplicando la distribución t:

$$gl = n_1 + n_2 - 2 = 9 + 6 - 2 = 13$$

Entonces de la tabla con $\alpha = 0{,}05$, se obtiene $t_\alpha = 1{,}771$.

El valor de t calculado con la fórmula es menor que el t_α, por lo tanto, cae en la zona de aceptación de la hipótesis nula. Se acepta H_o con $\alpha = 0{,}05$. Se puede decir que la diferencia entre los promedios de ambas modalidades de estudio no es estadísticamente significativa.

Pruebas para la diferencia de 2 proporciones

Ejemplo	Un investigador en el área de tecnología quiere determinar si hay diferencias en el uso de las redes sociales en internet entre hombres y mujeres. Para este fin toma una muestra de 40 hombres y 50 mujeres, y obtuvo que de ellos 28 hombres empleaban a diario al menos una de estas redes y 25 mujeres también usaban a diario al menos una de las redes. Con base en esos datos y a un nivel de significancia de 5%, ¿puede concluirse que existe diferencia significativa entre hombres y mujeres e n cuanto a su frecuencia de uso de las redes sociales en internet?
Solución	Se cuenta con la siguiente información: Hombres: $p_1 = 28/40 = 0{,}70$ Mujeres: $p_2 = 25/50 = 0{,}50$ Se plantean las hipótesis: H_o: $p_1 = p_2$ H_1: $p_1 > p_2$ Primero se calculan p y q:

$$p = \frac{p_1 + p_2}{n_1 + n_2} = \frac{28 + 25}{40 + 50} = 0{,}59$$

$$q = 1 - 0{,}5889 = 0{,}41$$

Luego se calcula z:

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0{,}70 - 0{,}50}{\sqrt{0{,}59 \cdot 0{,}41\left(\frac{1}{40} + \frac{1}{50}\right)}} = 1{,}92$$

De la tabla se obtiene $Z_{\alpha/2} = 1{,}96$. Como puede verse en el gráfico, el valor de Z_c cae en la zona de aceptación de H_0, por lo tanto se acepta H_0 con $\alpha = 0{,}05$. La diferencia no es estadísticamente significativa. La proporción de hombres que emplea a diario las redes sociales en internet no es significativamente diferente de la proporción de mujeres que realizan esta actividad.

Bibliografía

- Rodríguez Franco, Jesús, Pierdant Rodríguez, Alberto Isaac. Estadística para administración. (Primera Edición). México: Grupo Editorial Patria. (2014).
- Leandro Oviedo Gabriel. Estadística y Probabilidad con aplicaciones. (Primera Edición). Costa Rica: Publitex Grupo Editorial S.A. (2014).