Test de Hipótesis

1. **Test para la media de una población normal con varianza cononcida**

   Ejemplo: Supongamos que el consumo de nafta de los motores de cierto auto es una variable aleatoria normal con media 9 litros (cada 100 km) y desvío un litro (cada 100 kilometros). Se presenta una nueva forma de construir el motor que supuestamente disminuye el consumo de nafta del mismo. Para evaluar si esta innovación realmente mejora el rendimiento de los motores, se construyen 25 motores con el nuevo método y se los pone a funcionar en idénticas condiciones y en forma independiente. Podemos asumir que el consumo medio de nafta de un motor construido con el nuevo método es una variable aleatoria normal con esperanza desconocida y desvío igual a 1 litro (cada 100 kilometros). El consumo promedio observado por los 25 motores fue 8,9 litros cada 100 kilómetros. El promedio muestral se parece al consumo medio poblacional de nafta con la innovación, y nos dió 8,9 litros que es un poco menor al consumo sin la innovación ¿Proveen estos datos evidencia de que el consumo medio de nafta de los motores construidos con el nuevo método es menor a 9 o se debe simplemente al azar la mejoría observada?

   Para contestar la pregunta primero notemos que tenemos $X_1, X_2....X_{25}$ variables aleatorias i.i.d $X_i \sim N(\mu, 1)$ donde

   $$X_i = \text{consumo de nafta del i-ésimo motor constuido con el nuevo método, } i = 1, 2....,25.$$

   Luego, el parámetro μ representa el consumo de nafta esperado de un motor hecho con el nuevo método. Con lo cual, para contestar la pregunta, podemos pensar que nos debatimos entre dos hipótesis, una llamada hipótesis nula (H_0) o hipótesis del escéptico que establece que no hay cambios en el consumo medio de nafta debido a la nueva forma de construír el motor. La otra es llamada hipótesis alternativa (H_1) o hipótesis del investigador y establece que el consumo medio de nafta de los motores constuidos con el nuevo método es menor que 9. Esto puede escribirse así: las hipótesis a testear son

   $$H_0: \ \mu = 9 \quad \text{vs} \quad H_1: \ \mu < 9$$

   ya que nos preguntamos si tenemos evidencia para decir rechazar H_0 y creer H_1 o no tenemos evidencia suficiente.
   A las personas que desarrollaron el nuevo método para construír los motores les gustaría rechazar H_0 en favor de H_1. Notemos que hay dos tipos de errores que se pueden cometer, uno es no rechazar H_0 cuando es falsa (error tipo II) y otro rechazarla cuando es verdadera (error tipo I) . Notemos que si no rechazo H_0 y es falsa me pierdo hacer una inversión que produciría una mejora en mi producto. Sin embargo si rechazo H_0 y es verdadera voy a cambiar toda mi forma de producir cuando en realidad esto no establece ninguna mejora. Se asume que este último error (el error tipo I) es el peor y se trata de armar una regla de decisión que haga pequeña la probabilidad de cometerlo.

   Definición: Un test es una regla de decisión basada en un estadístico o función de la muestra, (en este caso de la media muestral), y en una zona de rechazo, es decir un conjunto de valores para los cuáles se rechaza la hipótesis nula H_0.

   Un test de hipótesis de nivel α es una regla de decisión entre H_0 y H_1 que verifica

$$P(\text{ rechazar } H_0|H_0 \text{ verdadero }) = \alpha$$

Típicamente α es un número chico entre 0 y 0,1.

Nota: Cuando aplicamos un test de hipótesis para decidir entre H_0 y H_1, si rechazo H_0 es en favor de H_1 (es decir, en nuestro mundo o pasa H_0 o pasa H_1. En el ejemplo de los motores, la nueva forma de construirlos no puede aumentar el consumo, o bien queda como antes o bien disminuye.)

La regla de decisión esta basada en el siguiente hecho:

Como $\overline{X}$ se parece a $\mu = E[X_1]$ si $(\overline{X} - 9)$ está LEJOS de cero entonces no creo H_0 En otras palabras,

Rechazo H_0 en favor de H_1 si $(\overline{X} - 9) < a$ con a un valor negativo lejano a cero, donde a es elegido para que

$$P\Big(\text{rechazar } H_0|H_0 \text{ es verdadera }\Big) = \alpha$$

Veamos como armamos un test de nivel $\alpha = 0{,}05$ en el ejemplo de los motores. Si H_0 es verdadera resulta que $E(X_i) = 9$ para todo $i = 1, 2..,25$ entonces

$$\text{Si } H_0 \text{ es verdadera, resulta que } \; Z = \frac{\overline{X} - 9}{1/5} \sim N(0,1)$$

luego tomo $z_{0,05} = -1{,}64$ y rechazo H_0 si el valor observado de Z con mi muestra, pertenece a la región

$$RR = \Big\{\frac{\overline{X} - 9}{1/5} < -1{,}64\Big\}$$

luego

$$P\Big(\text{rechazar } H_0|H_0 \text{ es verdadera}\Big) = P\Big(\frac{\overline{X} - 9}{1/5} < -1{,}64|H_0 \;\; \text{verdadera}\Big) =$$

$$= \phi(-1{,}64) = 0{,}05$$

ya que si H_0 es verdadera, entonces $Z \sim N(0,1)$. Una vez construida la región de rechazo, calculamos el valor $z_{obs} = \frac{(\overline{x}_{obs} - 9)}{1/5}$ para mi muestra puedo tomar la decisión. En el ejemplo

$$z_{obs} = \frac{\overline{x}_{obs} - 9}{1/5} = 5(8{,}9 - 9) = -0{,}5 \notin RR \text{ ya que } \;\; -0{,}5 > -1{,}64$$

luego a nivel 0,05 no hay evidencia suficiente para rechazar H_0 en favor de H_1. Es decir, los datos no muestran que el consumo esperado de nafta es menor que 9.

Si queremos que el test tenga nivel de significacion 0.1, rechazariamos H_0 si

$$\frac{\overline{X} - 9}{1/5} < -1{,}28$$

Criterio del p-valor Vamos a presentar otra regla para decidir entre

$$H_0: \ \mu = 9 \quad \text{y} \quad H_1: \ \mu < 9$$

llamada criterio del p-valor. Supongamos que no hubo mejora, es decir que vale H_0. En tal caso $\mu = 9$ y por lo tanto la variable aleatoria $Z = \frac{\left(\overline{X}-9\right)}{1/\sqrt{25}}$ debería ser normal estandar. De ser así, el valor observado de Z tendría que ser un valor típico de una normal estandar. Por las condiciones del experimento (asumimos que la innovación no pude aumentar el consumo) eso implicaría que la probabilidad de observar un valor igual a z_{obs} o menor debería ser grande si H_0 fuera verdadera. En nuestro ejemplo definimos el p-valor del test como

$$\text{p-valor} = P(Z \leq z_{obs} \ \text{ si } H_0 \text{ es verdadera})$$

El criterio del $p-$valor establece que si el $p-$valor es chico entonces rechazamos la hipótesis nula. Tomaremos como chico a un número menor a 0,1. En nuestro ejemplo $z_{obs} = \frac{\overline{X}_{obs}-9}{1/5} = 5(8{,}9-9) = -0{,}5$ y luego

$$\text{p-valor} = P(Z \leq -0{,}5 \ \text{ si } H_0 \text{ es verdadera}) = 1 - \phi(0{,}5) = 1 - 0{,}6915 = 0{,}3085 > 0{,}1$$

lo cual indica que no hay evidencia para rechazar H_0.

Como hemos visto, al seleccionar la región de rechazo controlamos la probabilidad de error tipo I, pero ¿qué ocurre con el error tipo II?.

Supongamos que en nuestro ejemplo, observamos un consumo promedio en la muestra de tamaño 25 igual a 8.9 litros cada 100 km y trabajamos con el test de nivel 0.05. En este caso, no rechazamos Ho (tampoco lo haríamos con el test de nivel 0.10) y por lo tanto, si la mejora en el motor fuese real, podríamos estar cometiendo un error de tipo II. Por ejemplo, si la modificación en el motor reduce el consumo a 8.5 litros cada 100 km, ¿cuál es la probabilidad de cometer un error tipo II?

Procedimiento general para armar un test de hipótesis de nivel α para la media de una población normal con varianza conocida y cálculo de p-valor.

Sean $X_1 ... X_n$ variables aleatorias normales independientes con $E\left(X_i\right) = \mu$ desconocida y $Var(X_i) = \sigma_0^2$ conocida (luego veremos como trabajar cuando la varianza es desconocida.)

Consideramos como hipótesis nula a la hipótesisque involucra que no hay cambios debido al nuevo método. Es decir $H_0 : \mu = \mu_0$. La hipótesis alternativa establece que hubo cambios, o es la hipótesis que se quiere ver si se tiene evidencia para probarla. La hipótesis alternativa varía según el problema. Hay tres escenarios excluyentes posibles. Si denotamos con H_1 a la hipótesis alternativa tenemos

$$H_1 : \mu < \mu_0, \qquad H_1 : \mu > \mu_0, \qquad H_1 : \mu \neq \mu_0.$$

Queremos una regla de decisión entre H_0 y H_1 que verifique que la probabilidad de rechazar H_0 cuando sea verdadera sea un valor prefijado α. Esta regla de decisión esta basada en

Como $\overline{X}$ se parece a μ, si $Z = \frac{\left(\overline{X}-\mu_0\right)}{\sigma_0/\sqrt{n}}$ está LEJOS de cero entonces no creo H_0

$$Z = \frac{\left(\overline{X} - \mu_0\right)}{\sigma_0/\sqrt{n}} \sim N(0,1) \text{ bajo } H_0$$

A la variable aletoria $Z = \frac{\left(\overline{X}-\mu_0\right)}{\sigma_0/\sqrt{n}}$ la llamamos estadístico del test.

Dado α un número entre 0 y 0,1, el criterio para decidir entre H_0 y H_1 basados en un test de nivel α es el siguiente:

a) Si testeamos $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$ tomamos $Z = \frac{\left(\overline{X}-\mu_0\right)}{\sigma_0/\sqrt{n}}$ y consideramos la siguiente región llamada región de rechazo

$$RR = \{Z : Z \geq z_\alpha\}$$

donde z_α verifica $\phi(z_\alpha) = 1 - \alpha$. Luego calculamos z_{obs} con la muestra (es el valor del Z para los datos observados). Si z_{obs} pertenece a RR rechazamos H_0, caso contrario no hay evidencia para rechazar.
Definimos el p-valor como:

$$\text{p-valor} = P(Z \geq z_{obs} | H_0 \text{ es verdadera}) = 1 - \phi(z_{obs}),$$

si este valor es chico (menor a 0,1) rechazaremos la hipótesis nula, de lo contrario no hay evidencia para rechazar H_0.

b) Si testeamos $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$ tomamos $Z = \frac{\left(\overline{X}-\mu_0\right)}{\sigma_0/\sqrt{n}}$ y consideramos la siguiente región llamada región de rechazo

$$RR = \{Z : Z \leq -z_\alpha\}$$

donde z_α verifica $\phi(z_\alpha) = 1 - \alpha$. Luego calculamos z_{obs} con la muestra. Si z_{obs} pertenece a RR rechazamos H_0, caso contrario no hay evidencia para rechazar.
Definimos el p-valor como:

$$\text{p-valor} = P(Z \leq z_{obs} \text{ si } H_0 \text{ es verdadera}) = \phi(z_{obs}),$$

si este valor es chico (menor a 0,1) rechazaremos la hipótesis nula, de lo contrario no hay evidencia para rechazar H_0.

c) Si testeamos $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ tomamos $Z = \frac{\left(\overline{X}-\mu_0\right)}{\sigma_0/\sqrt{n}}$ y consideramos la siguiente región llamada región de rechazo

$$RR = \{Z : |Z| \geq z_{\frac{\alpha}{2}}\}$$

donde $z_{\frac{\alpha}{2}}$ verifica $\phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$. Luego calculamos z_{obs} con la muestra. Si z_{obs} pertenece a RR rechazamos H_0, caso contrario no hay evidencia para rechazar.

Definimos el p-valor como:

$$\text{p-valor} = 2P(Z \geq |z_{obs}| \text{ si } H_0 \text{ es verdadera}) = 2(1 - \phi(|z_{obs}|)),$$

si este valor es chico (menor a 0,1) rechazaremos la hipótesis nula, de lo contrario no hay evidencia para rechazar H_0.

Notar que si la hipótesis nula es cierta

$$Z = \frac{\left(\overline{X} - \mu_0\right)}{\sigma_0/\sqrt{n}} \approxeq_{\text{BAJO } H_0} N(0,1)$$

En los tres casos se verifica: $P(\text{rechazar } H_0 \text{ cuando es verdadera}) = \alpha$. Recordemos que cuando rechazamos H_0 y es verdadera cometemos un error que se llama error tipo I. A la probabilidad de error tipo I la llamamos nivel del test. Todas estas reglas tienen nivel α. El otro error que podemos cometer es no rechazar H_0 cuando es falsa (error tipo II) y típicamente lo calculamos en un valor específico de μ perteneciente a la hipótesis alternatíva.

Observación Se puede ver que si uno quisiera testear $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$ o $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$ vale la misma regla de decisión presentada para testear $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$ y $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$ respectivamente.

Ejemplo 1

La presión sanguínea es la tensión ejercida por la sangre circulante sobre las paredes de los vasos sanguíneos. Dicha tensión varía durante el ciclo cardíaco. Al valor máximo de presión obtenido en un ciclo cardíaco se lo llama presión arterial sistólica. Se sabe que la presión sistólica de la población general de mujeres adultas en descanso es una variable aletoria con esperanza 100 mmHg y desvío estandar 15 mmHg (unidades medidas en milímetros del mercurio). Un grupo de nutricionistas sospecha que el sobrepeso aumenta la esperanza de la presión sistólica. Con el fin de validar su supuesto, se mide la presión a 64 mujeres adultas con sobrepeso obteniendose una media muestral de 104.89 mmHg.

a) Basados en un test de nivel 0.01 ¿Hay evidencia de que la esperanza de la presión sistólica de las mujeres con sobrepeso es mayor que la de la población en general? Explicitar las variables y parámetros de interés, las hipótesis del test, el estadístico del test y su distribución bajo la hipótesis nula.

b) Calcular el p-valor del test

c) Si la esperanza de la presión sistólica de las mujeres con sobrepeso fuera de 102 mmHg. ¿Cuál es la probabilidad de tomar una decisión equivocada? (Error de tipo II)

2. **Test para la media de una población normal con varianza descononcida**

En el caso en que la varianza de las X_i no sea conocida hacemos exactamente lo mismo que acabamos de ver, pero usando como estadístico del test a $Z = \frac{\left(\overline{X}-\mu_0\right)}{S/\sqrt{n}}$ donde S es el desvío muestral y en cada una de las RR (regiones de rechazo), en lugar de usar z_α, $-z_\alpha$ y $z_{\alpha/2}$ usaremos $t_{n-1,\alpha}$, $-t_{n-1,\alpha}$ y $t_{n-1,\alpha/2}$ ya que ahora, si H_0, es verdadera

$$\frac{\left(\overline{X}-\mu_0\right)}{S/\sqrt{n}} \sim T_{n-1}.$$

Ejemplo 2 En la construcción de un edificio debe usarse un concreto que tenga una tensión media mayor a 300 psi. Para verificar si el concreto preparado a partir del cemento "Loma Blanca" cumple con este requerimiento, se realizan 15 mediciones en forma independiente de la tensión de este concreto. Se observa una media muestral de 304 psi y un desvío estándar muestral de 10 psi. El constructor está dispuesto a correr un riesgo del 5 % de comprar cemento "Loma Blanca" cuando éste produce un concreto que no cumple con las especificaciones. Suponiendo que los datos están normalmente distribuidos: Plantear el test correspondiente. ¿Qué decisión se toma?

3. **Test para la varianza de una población normal**

4. **Test para proporciones**

Sean $X_1, \cdots, X_n$ son i.i.d, $X_i \sim Ber(p)$ donde $p =$ probabilidad de éxito . Resulta que $E\Big(X_i\Big) = p$ y $V\Big(X_i\Big) = p(1-p)$

Aplicando TCL obtenemos

$$Z = \sqrt{n}\frac{\Big(\overline{X} - p\Big)}{\sqrt{p(1-p)}} \sim_{aprox} N(0,1)$$

Un test de nivel aproximado α es testear la hipótesis nula $H_0 : p = p_0$. versus H_1 la hipótesis alternativa. Para hipótesis alternativas tenemos

$$H_1 : p < p_0, \qquad H_1 : p > p_0, \qquad H_1 : p \neq p_0.$$

Esta regla de decisión esta basada en el estadístico

$$Z = \sqrt{n}\frac{\Big(\overline{X} - p_0\Big)}{\sqrt{p_0(1-p_0)}}$$

El cual si H_0 es cierto tiene distribución aproximadamente $N(0,1)$.

Ejercicio 3 Se supone que 1 de cada 10 fumadores prefiere la marca A. Después de una campaña publicitaria en cierta región de ventas, se entrevistó a 200 fumadores para determinar la efectividad de la campaña. El resultado de esta encuesta mostró que 26 personas preferían la marca A.

a) Basado en un test de nivel aproximado 0,05 ¿Indican estos datos un aumento en la preferencia por la marca A?

b) Calcular el valor p.

c) ¿Cuál es la probabilidad aproximada de decidir que la campaña publicitaria no fue efectiva, cuando en realidad la proporción de preferencia por la marca A después de la campaña es 0.15?