

PRUEBAS DE HIPÓTESIS

1 Ejemplo introductorio

Supongamos que en el mercado existe un fármaco cuyo tratamiento tiene una efectividad del 75% para combatir determinada enfermedad. El equipo de investigación del laboratorio que produce el fármaco, desarrolla un nuevo fármaco. Para probar su efectividad, suministra el mismo a 100 pacientes que padecen la enfermedad. Luego del tratamiento, 78 personas resultan recuperadas. El equipo de dirección del laboratorio debe decidir si dejar de fabricar el anterior fármaco y comenzar a fabricar el nuevo o seguir produciendo el fármaco de siempre y dejar sin efecto el nuevo.

La herramienta estadística que se utiliza para resolver este tipo de casos es lo que se llama hacer un test de hipótesis.

En este ejemplo, si tomamos un individuo que padece la enfermedad y hace el tratamiento con el nuevo fármaco y definimos

$$X = \begin{cases} 1 & \text{si la persona se recupera} \\ 0 & \text{si la persona no se recupera} \end{cases}$$

resulta que $X \sim \text{Ber}(p)$ siendo p desconocido (puesto que no conocemos la real efectividad del nuevo fármaco). En este caso contamos con 100 realizaciones independientes de X , es decir una MAS X_1, X_2, \dots, X_{100} de la cual sabemos que $X_1 + X_2 + \dots + X_{100} = 78$. Si quisiéramos estimar p , entonces calcularíamos $\hat{p} = \bar{X}_n = \frac{78}{100} = 0.78$. También podríamos hallar un intervalo de confianza para p para un determinado nivel de confianza. Pero en este caso, no queremos estimar el valor de p , lo que queremos es “tomar una decisión” entre seguir fabricando el fármaco habitual o cambiar al nuevo.

2 Definiciones

Definición 1 Dada una MAS X_1, X_2, \dots, X_n de X con distribución F_X donde alguna componente de F_X es desconocida (puede ser un parámetro o toda la distribución), realizar una prueba de hipótesis es tomar una decisión entre dos hipótesis llamadas hipótesis nula (H_0) e hipótesis alternativa (H_1). La toma de la decisión recibe el nombre de “rechazo H_0 ” o “no rechazo H_0 ”.

Observación 2 La realidad es H_0 o H_1 , pero no sabemos cuál es cierta. A partir de la información que obtenemos de la muestra se trata de tomar una decisión sobre si H_0 es cierta o H_1 es cierta, pero nos podemos equivocar en la decisión.

| decisión \ realidad | H_0 | H_1 |
|---------------------|----------|----------|
| rechazo H_0 | error | correcto |
| no rechazo H_0 | correcto | error |

En nuestro ejemplo, tenemos

H_0 : el nuevo fármaco no es mejor al actual, o sea $H_0 : p \leq 0.75$

H_1 : el nuevo fármaco es mejor al actual, o sea $H_1 : p > 0.75$.

Definición 3 *Región crítica.*

RC = zona de rechazo de H_0 , siendo RC un subconjunto de \mathbb{R}^n siendo n el tamaño de la muestra.

Observación 4 La RC la define el investigador y nos da el criterio que se toma para tomar la decisión, de modo que si $(X_1, X_2, \dots, X_n) \in RC$ la decisión es “rechazo H_0 ” y si $(X_1, X_2, \dots, X_n) \notin RC$ la decisión es “no rechazo H_0 ”.

Definición 5 Errores tipo I y II, nivel de significación de la prueba y potencia de la prueba.

Dada una prueba de hipótesis H_0 versus H_1 con región crítica RC se definen:

- Error tipo I: rechazar H_0 cuando H_0 es cierta.
- Error tipo II: no rechazar H_0 cuando H_1 es cierta.
- $\alpha = \sup_{H_0} P((X_1, X_2, \dots, X_n) \in RC) = \sup_{H_0} P(\text{rechazo } H_0 \text{ cuando } H_0 \text{ es cierto})$ se llama nivel de significación de la prueba.
- $\beta = P_{H_1}((X_1, X_2, \dots, X_n) \notin RC) = P(\text{no rechazo } H_0 \text{ cuando } H_1 \text{ es cierto})$.
- $K = 1 - \beta = P_{H_1}((X_1, X_2, \dots, X_n) \in RC) = P(\text{rechazar } H_0 \text{ cuando } H_1 \text{ es cierto})$ se llama potencia de la prueba.

Observación 6 α y β son las probabilidades de cometer errores tipo I y II respectivamente.

Comentario.

Para simplificar la comprensión en una primera lectura, es conveniente omitir lo del supremo en la definición de α . α lo pensamos como la probabilidad de tomar la decisión incorrecta en el caso de que la realidad sea H_0 y β lo pensamos como la probabilidad de tomar la decisión incorrecta en el caso en que la realidad sea H_1 .

3 Procedimiento general para realizar una prueba de hipótesis.

1. Plantear H_0 y H_1 .
2. Elegir un valor de α y construir una región crítica de nivel α es decir que cumpla $\sup_{H_0} P(RC) = \alpha$.

3. Tomar la decisión (si la muestra cumple la RC , rechazo H_0 y si la muestra no cumple la RC no rechazo H_0).

4 Notas y comentarios.

1. Si la RC está bien diseñada, bajo ciertas hipótesis se puede probar que $\alpha \rightarrow 0$ y $\beta \rightarrow 0$ cuando $n \rightarrow +\infty$ (notar que tanto α y β dependen de n). ¿Cómo se interpreta?
2. En muchas aplicaciones se cuenta con la información de determinada muestra, por lo que n será fijo, entonces se suele elegir el valor de α , se obtiene la RC de nivel α , se toma la decisión y luego el valor de β se puede calcular.
3. El valor de α (nivel de significación de la prueba) juega un papel similar al del nivel de confianza en un intervalo de confianza $(1 - \alpha)$. Se suele elegir un valor pequeño que nos asegure que la decisión luego de tomada haya sido tomada mediante un procedimiento que tenga una probabilidad de error tipo I pequeña, pero si α es demasiado pequeño aumenta drásticamente el valor de β y como podemos equivocarnos al tomar la decisión no es conveniente.
4. Siguiendo al comentario anterior, el costo que tiene elegir un valor de α muy pequeño en el intervalo de confianza es en la pérdida de precisión en el intervalo (aumenta la longitud del intervalo). Si elegimos $\alpha = 0$, entonces para obtener un intervalo de confianza al nivel 100%, el intervalo tendría longitud infinita. El costo en una prueba de hipótesis de tomar un α demasiado pequeño, radica en el aumento del β . En particular si $\alpha = 0$, la única forma de no equivocarnos nunca si H_0 es cierto, es tomando la estrategia de no rechazar nunca H_0 , pero si esa es nuestra estrategia, entonces si H_1 es cierto me equivocaré siempre por lo que tendremos $\beta = 1$.
5. ¿Cómo se construye la región crítica? Depende mucho de cada prueba de hipótesis, si bien existe algún teorema en general como el de Neyman Pearson, la región crítica se construye generalmente a través de métodos intuitivos, en algún sentido similar a la construcción de los intervalos de confianza.
6. Más adelante dejamos algunas regiones críticas “tabuladas” para muchos casos especiales que ocurren en la práctica con frecuencia.

5 Resolución del ejemplo introductorio.

Vamos a resolver el problema planteado en la introducción al nivel del 5%. Encontraremos una región crítica de modo “intuitivo” (que la aproximaremos

mediante el empleo del TCL, la ajustaremos para que tenga nivel $\alpha = 0.05$, y luego tomaremos la decisión.

En nuestro caso tenemos $H_0 : p \leq 0.75$ versus $H_1 : p > 0.75$. Vamos a resolver primero un problema simplificado más sencillo y luego veremos que la solución de este problema más sencillo tiene la misma solución que nuestro problema original.

Problema simplificado.

Con los mismos datos, supongamos que queremos testear $H_0 : p = 0.75$ versus $H_1 : p > 0.75$ (H_0 y H_1 no tienen por qué ser complementarios, sí deben ser incompatibles entre sí porque expresan o bien una realidad o bien otra).

El paso 1 del procedimiento ya está (planteamiento de las hipótesis). Para construir la región crítica que nos indica el paso 2 podemos razonar como sigue.

Por la ley de los grandes números, sabemos que $\bar{X}_n \xrightarrow{c.s.} p$ (porque $\bar{X}_n \xrightarrow{c.s.} \mu$ y en este caso $\mu = E(X) = p$ por ser la esperanza de una Bernoulli). Entonces \bar{X}_n tomará valores cercanos al verdadero valor de p que desconocemos. Entonces si \bar{X}_n es mucho más grande que 0.75, es razonable rechazar H_0 , si \bar{X}_n mucho menor que 0.75 es razonable no rechazar H_0 . Y si \bar{X}_n toma valores cercanos a 0.75 es razonable no rechazar H_0 . Entonces parece natural definir una región crítica de la forma

$$RC = \{(X_1, X_2, \dots, X_n) \in \mathbb{R}^n : \bar{X}_n \geq 0.75 + k\}$$

donde el valor de k es el umbral más allá del cual consideramos que \bar{X}_n es mucho más grande que 0.75. Por lo tanto tenemos la región crítica a menos del valor de k . Para hallar el valor de k planteamos la ecuación (una ecuación con una incógnita)

$$P_{H_0}(\bar{X}_n \geq 0.75 + k) = \alpha, \text{ o sea } P_{H_0}(\bar{X}_n \geq 0.75 + k) = 0.05.$$

Para poder resolver la ecuación necesitamos la distribución de \bar{X}_n que si n es grande se aproxima por una normal por el TCL. Más explícitamente, en este caso

$$\bar{X}_n \text{ es aproximadamente } N(\mu, \sigma^2/n) = N(0.75; 0.001875)$$

en donde en la última igualdad se utilizó que $\mu = E(X) = p = 0.75$ (porque se calcula suponiendo H_0 cierto), $\sigma^2 = V(X) = p(1-p) = 0.75 \times 0.25 = 0.1875$ y $n = 100$.

Entonces $P_{H_0}(\bar{X}_n \geq 0.75 + k) \stackrel{\text{aprox}}{=} 1 - \phi((0.75 + k - 0.75)/\sqrt{0.001875}) = 1 - \phi(k/\sqrt{0.001875}) = 0.05$ de donde se deduce que $\phi(k/\sqrt{0.001875}) = 0.95$, entonces $k/\sqrt{0.001875} = 1.645$ por lo que $k = 1.645 \times \sqrt{0.001875} = 0.07123$ de modo que $0.75 + k = 0.82123$.

Entonces la región crítica que tiene nivel 5% es

$$RC = \{(X_1, X_2, \dots, X_n) \in \mathbb{R}^n : \bar{X}_n \geq 0.82123\}.$$

En el paso 3 tomamos la decisión: con los datos de nuestra muestra tenemos $\bar{X}_n = 0.78 \not\geq 0.82123$, entonces no se cumple la RC por lo tanto la decisión es “no rechazar H_0 ”.

Resolución del caso general.

En nuestro caso real, tenemos $H_0 : p \leq 0.75$ versus $H_1 : p > 0.75$. Cuando vamos a hallar el valor de k , se calcula suponiendo H_0 cierto, pero a diferencia del caso simplificado ahora cuando H_0 es cierto, no sabemos el valor de p , aquí es donde entra en juego el supremo bajo H_0 cierto. Ahora nuestra ecuación con una incógnita pasa a ser $\sup_{p \leq 0.75} P_{H_0}(\bar{X}_n \geq 0.75 + k) = \alpha$. La manera de obtener este supremo es haciendo el mismo cálculo que en el caso simplificado, pero en función de p , es decir que calcularíamos

$$P_{H_0}(\bar{X}_n \geq 0.75 + k) \stackrel{\text{aprox}}{=} 1 - \phi\left(\frac{(0.75 + k - p)/\sqrt{p(1-p)/100}}{1}\right)$$

en función de p y calcularíamos el máximo de dicha función para los valores de $p \leq 0.75$.

Se puede probar que dicha función es creciente por lo que el supremo se obtiene para $p = 0.75$ quedando exactamente el mismo cálculo que en el caso simplificado.

Conclusión. La región crítica de nivel 5% quedó

$$\{(X_1, X_2, \dots, X_n) \in \mathbb{R}^n : \bar{X}_n \geq 0.82123\}$$

y como los datos de nuestra muestra no la verifican, no rechazamos H_0 .

Observación 7 En nuestro caso, $H_0 : p \leq 0.75$ versus $H_1 : p > 0.75$, de alguna manera el caso “difícil” a distinguir entre H_0 y H_1 es cuando $p = 0.75$, es cuando la hipótesis nula más se parece a la alternativa.

Cálculo de la probabilidad de error del tipo II.

Una vez tenemos diseñada la RC de modo que tenga el nivel de significación deseado (α), es posible calcular la probabilidad de error tipo II (β).

En el ejemplo anterior, es decir para la prueba $H_0 : p \leq 0.75$ versus $H_1 : p > 0.75$ de nivel $\alpha = 0.05$, si el verdadero porcentaje de éxito del nuevo fármaco es del 84%, ¿cuál es la probabilidad de error tipo II? ¿Y la potencia? Hallar la función de potencia para la prueba.

Como la región crítica de nivel $\alpha = 0.05$ nos quedó

$$\{(X_1, X_2, \dots, X_n) \in \mathbb{R}^n : \bar{X}_n \geq 0.82123\},$$

entonces ahora queremos calcular

$$\beta = P_{H_1}(\bar{X}_n < 0.82123) \text{ donde la alternativa es } p = 0.84.$$

Aplicamos nuevamente el TCL y nos queda que \bar{X}_n es aproximadamente $N(\mu, \sigma^2/n) = N(0.84; 0.001344)$ ¿por qué? chequearlo.

Entonces

$$\beta = P_{H_1}(\bar{X}_n < 0.82123) \stackrel{\text{aprox}}{=} \phi\left(\frac{(0.82123 - 0.84)/\sqrt{0.001344}}{1}\right) =$$

$$\phi(-0.512) = 1 - \phi(0.512) = 0.3043.$$

Si el verdadero p es $p = 0.84$, la potencia es

$$K = P_{H_1}(RC) = P_{H_1}(\bar{X}_n \geq 0.82123) = 1 - \beta = 1 - 0.3043 = 0.6957.$$

Función de potencia.

La potencia es por definición $1 - \beta$, es decir la probabilidad de “no equivocarnos en la decisión si es que H_1 es cierto”. Recién calculamos β en el supuesto de que la alternativa es $p = 0.84$, pero si lo dejamos en función de p , queda β en función de p y por lo tanto la potencia en función de p . Como $H_1 : p > 0.75$, entonces no sabemos el valor de p por lo que $\beta(p) = P_{H_1}(\bar{X}_n < 0.82123)$ queda en función de p . Se calcula como cuando hallamos el k para ajustar al nivel de significación. Aproximamos \bar{X}_n es aproximadamente $N(\mu, \sigma^2/n) = N(p, p(1-p)/100)$ por lo que $K(p) = 1 - \beta(p) =$

$$P_{H_1}(\bar{X}_n > 0.82123) = 1 - \phi\left(10 \times \frac{0.82123 - p}{\sqrt{p(1-p)}}\right) = 1 - \phi\left(\frac{8.2123 - 10p}{\sqrt{p(1-p)}}\right)$$

definido para $p > 0.75$.

6 Notas y comentarios.

1. En la práctica, como ya se dijo, para una muestra dada el valor de n está fijo, por lo que puede elegirse el valor de α tan pequeño como se quiera, pero β puede llegar a ser grande como en el ejemplo anterior, situación que no es deseable. En algunos casos, es posible definir de antemano los valores de α y β y obtener el n adecuado (suficientemente grande) de modo de que ambas probabilidades de error sean arbitrariamente pequeñas. Más adelante habrá un ejercicio al respecto.
2. ¿Por qué no se dice “acepto H_0 ” que es más fácil y directo que “no rechazo H_0 ”?

Por supuesto que es válido decir “acepto H_0 ”, sin embargo es más correcto decir “no rechazo H_0 ”, porque en las pruebas de hipótesis no hay una simetría en cuanto a las hipótesis nula y alternativa. Esto puede deducirse por varios lados. Si miramos el comentario anterior, vimos que podemos elegir α arbitrariamente pequeño, pero β luego se calcula, es decir que podemos controlar tanto como queramos el nivel de error tipo I, pero no el error tipo II. Si pensamos en el ejemplo que hemos resuelto, el porcentaje de eficacia del nuevo fármaco en la muestra fue superior al 75%, sin embargo no rechazamos H_0 . Esto ocurre porque para rechazar H_0 necesitamos un porcentaje de eficacia en la muestra claramente mayor (en nuestro caso necesitaríamos más del 82%), de lo contrario “no tenemos suficiente evidencia” para descartar que $p = 0.75$, pues el 78% muestral

está cerca del 75%. En síntesis, y siguiendo con nuestro ejemplo, si la muestra hubiera arrojado 83 pacientes recuperados, tendríamos suficiente evidencia de que H_0 es falso. Pero con 78 pacientes recuperados “no tenemos suficiente evidencia de que H_0 sea cierto, sino que no tenemos la suficiente evidencia como para rechazarlo”.

3. De acuerdo al comentario realizado en el punto anterior y a grandes rasgos, podemos plantear la siguiente interpretación de la decisión en una prueba de hipótesis.

| Decisión | Interpretación |
|------------------|--|
| rechazo H_0 | Tenemos suficiente evidencia de que H_0 no es cierto |
| no rechazo H_0 | No tenemos suficiente evidencia de que H_0 sea falso |

4. ¿Cómo se determina a quién llamarle H_0 y a quién H_1 ?

Sabemos que tomemos la decisión que tomemos, podemos equivocarnos. Sabemos que podemos controlar el error tipo I y no el error tipo II, eso hace que en la práctica tomemos en cuenta cuál es el error más grave de los que podemos cometer y a ese error le llamamos error tipo I, y de acuerdo a la definición de error tipo I, deducimos a quién nos conviene llamarle H_0 y a quién H_1 . Qué significa el error más grave, es subjetivo depende de cada caso. Volvamos al ejemplo para entender mejor lo anteriormente dicho. Queremos saber si el nuevo fármaco es mejor o no que el otro, entonces las hipótesis serán $p \leq 0.75$ versus $p > 0.75$ pero no sabemos aún a quién llamarle H_0 y a quién H_1 . Los posibles errores que podemos cometer son

A: decidir que el nuevo fármaco es mejor cuando en realidad no era mejor

B: decidir que el nuevo fármaco no es mejor cuando en realidad es mejor.

¿Cuál es el peor error? Supongamos que integramos el equipo gerencial del laboratorio. Decidir que el nuevo fármaco es mejor, implicaría cambiar el fármaco actual por el nuevo. Eso podría implicar muchos cambios en contratos, podríamos prescindir de los empleados especializados en la fabricación del fármaco actual y contratar otros especializados en el nuevo fármaco, se podría necesitar cambiar determinados aparatos útiles para el fármaco actual y cambiarlos por otros que sean útiles para fabricar el nuevo fármaco. Además habría que hacer una campaña publicitaria para avisar a la población que ahora se tiene un nuevo fármaco que es mejor que el anterior. En contraposición, decidir que el nuevo fármaco no es mejor que el anterior, queda todo como está y no se corren riesgos de entrar en costos innecesarios. Si tomamos este criterio donde se toman en cuenta los costos para el laboratorio, diríamos que el “peor error” es el A. Entonces, podemos concluir que el error tipo I sería el A. Entonces

A: decidir que el nuevo fármaco es mejor cuando en realidad no era mejor.

Error tipo I: rechazar H_0 cuando H_0 es cierto.

Entonces la prueba será $H_0 : p \leq 0.75$ versus $H_1 : p > 0.75$.

mejor malo conocido que bueno por conocer

5. Siguiendo un poco con el comentario anterior, en muchas ocasiones H_0 es la hipótesis más “conservadora” en el sentido que la decisión de no rechazar H_0 implica no hacer grandes cambios, mientras que para hacer mayores cambios o movimientos debemos tener suficiente evidencia empírica. Históricamente se llama a H_0 la hipótesis nula, justamente porque su no rechazo implica “efectos nulos”.

Ejercicio.

Si X_1, X_2, \dots, X_n es una MAS de $X \sim N(\mu, \sigma^2 = 9)$, se considera la prueba $H_0 : \mu = 8$ versus $H_1 : \mu = 10$. Mediante una región crítica de la forma $\{\bar{X}_n \geq k\}$.

1. Pensar el por qué es razonable plantear una región crítica como la mencionada.
2. ¿De qué tamaño debería ser la muestra para que $\alpha = 0.05$ y $\beta = 0.08$?
3. Si una muestra del tamaño hallado en la pregunta anterior arrojó un promedio muestral de 8.9 ¿Cuál es la decisión?

7 Regiones críticas conocidas

En esta sección se incluyen regiones críticas conocidas para algunas pruebas de hipótesis paramétricas. En las siguientes pruebas de hipótesis α representa el nivel de significación de la prueba.

Se recuerda que la notación $z_p = \phi^{-1}(1 - p)$ representa aquel valor tal que el área a la derecha de z_p es igual a p , y análogamente con $t_{p,k}$ y $\chi_{p,k}^2$. Se recuerda también que $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

7.1 Población normal

En esta subsección se tiene X_1, X_2, \dots, X_n MAS de $X \sim N(\mu, \sigma^2)$.

7.1.1 Pruebas sobre la media

1. Caso σ^2 conocido.

•

$$H_0 : \mu \leq \mu_0$$

$$H_1 : \mu > \mu_0$$

$$RC = \left\{ \bar{X}_n \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha \right\}$$

•

$$\bar{H}_0 : \mu \geq \mu_0$$

$$H_1 : \mu < \mu_0$$

$$RC = \left\{ \bar{X}_n \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_\alpha \right\}$$

•

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu \neq \mu_0 \end{aligned} \quad RC = \left\{ |\bar{X}_n - \mu_0| \geq \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right\}$$

2. Caso σ^2 desconocido.

•

$$\begin{aligned} H_0 &: \mu \leq \mu_0 \\ H_1 &: \mu > \mu_0 \end{aligned} \quad RC = \left\{ \bar{X}_n \geq \mu_0 + \frac{S_n}{\sqrt{n}} t_{\alpha, n-1} \right\}$$

•

$$\begin{aligned} H_0 &: \mu \geq \mu_0 \\ H_1 &: \mu < \mu_0 \end{aligned} \quad RC = \left\{ \bar{X}_n \leq \mu_0 - \frac{S_n}{\sqrt{n}} t_{\alpha, n-1} \right\}$$

•

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu \neq \mu_0 \end{aligned} \quad RC = \left\{ |\bar{X}_n - \mu_0| \geq \frac{S_n}{\sqrt{n}} t_{\alpha/2, n-1} \right\}$$

7.1.2 Pruebas sobre la varianza

X_1, X_2, \dots, X_n MAS de $X \sim N(\mu, \sigma^2)$.

•

$$\begin{aligned} H_0 &: \sigma^2 \leq \sigma_0^2 \\ H_1 &: \sigma^2 > \sigma_0^2 \end{aligned} \quad RC = \left\{ S_n^2 \geq \frac{\sigma_0^2}{n} \chi_{\alpha, n-1}^2 \right\}$$

•

$$\begin{aligned} H_0 &: \sigma^2 \geq \sigma_0^2 \\ H_1 &: \sigma^2 < \sigma_0^2 \end{aligned} \quad RC = \left\{ S_n^2 \leq \frac{\sigma_0^2}{n} \chi_{1-\alpha, n-1}^2 \right\}$$

•

$$\begin{aligned} H_0 &: \sigma^2 = \sigma_0^2 \\ H_1 &: \sigma^2 \neq \sigma_0^2 \end{aligned} \quad RC = \left\{ S_n^2 \leq \frac{\sigma_0^2}{n} \chi_{1-\alpha/2, n-1}^2 \right\} \cup \left\{ S_n^2 \geq \frac{\sigma_0^2}{n} \chi_{\alpha/2, n-1}^2 \right\}$$

7.2 Pruebas sobre la media, población cualquiera

X_1, X_2, \dots, X_n MAS de X con distribución cualquiera (tal que exista su varianza) siendo n grande. Las regiones críticas son de nivel aproximado α (por aplicación del TCL).

•

$$\begin{aligned} H_0 &: \mu \leq \mu_0 \\ H_1 &: \mu > \mu_0 \end{aligned} \quad RC = \left\{ \bar{X}_n \geq \mu_0 + \frac{S_n}{\sqrt{n}} z_\alpha \right\}$$

•

$$\begin{aligned} H_0 &: \mu \geq \mu_0 \\ H_1 &: \mu < \mu_0 \end{aligned} \quad RC = \left\{ \bar{X}_n \leq \mu_0 - \frac{S_n}{\sqrt{n}} z_\alpha \right\}$$

•

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu \neq \mu_0 \end{aligned} \quad RC = \left\{ |\bar{X}_n - \mu_0| \geq \frac{S_n}{\sqrt{n}} z_{\alpha/2} \right\}.$$

7.3 Población Bernoulli

X_1, X_2, \dots, X_n MAS de $X \sim \text{Ber}(p)$ con n grande, las regiones críticas son aproximadas por aplicación del TCL.

•

$$\begin{aligned} H_0 &: p \leq p_0 \\ H_1 &: p > p_0 \end{aligned} \quad RC = \left\{ \bar{X}_n \geq p_0 + \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}} z_\alpha \right\}$$

•

$$\begin{aligned} H_0 &: p \geq p_0 \\ H_1 &: p < p_0 \end{aligned} \quad RC = \left\{ \bar{X}_n \leq p_0 - \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}} z_\alpha \right\}$$

•

$$\begin{aligned} H_0 &: p = p_0 \\ H_1 &: p \neq p_0 \end{aligned} \quad RC = \left\{ |\bar{X}_n - p_0| \geq \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}} z_{\alpha/2} \right\}$$

Ejercicio.

Resolver el ejemplo introductorio con ayuda de las regiones críticas conocidas descritas anteriormente.

8 p-valor para una prueba de hipótesis

De acuerdo a lo previamente visto, la decisión en toda prueba de hipótesis depende del valor de α elegido, ya que el umbral a superar depende del valor de α (recordar que se halla el valor de c tal que $P_{H_0}(T \geq c) = \alpha$ y luego con el c hallado se toma la decisión. Por lo tanto, si tomamos la decisión al 5% y queremos ahora tomar la decisión al 1% o 10% deberíamos recalculamos el c para luego tomar nuevamente la decisión. El p-valor nos servirá para tomar la decisión para todo valor de α . Dicho de otra forma, si conocemos el p-valor, conoceremos la decisión para todo valor de α .

Definición 8 Dada una prueba de hipótesis cualquiera H_0 versus H_1 cuya región crítica es de la forma

$$\{(X_1, X_2, \dots, X_n) \in \mathbb{R}^n : T(X_1, X_2, \dots, X_n) \geq c\}$$

siendo T un estadístico y c una constante. Si tenemos la muestra realizada (x_1, x_2, \dots, x_n) y le llamamos $t_0 = T(x_1, x_2, \dots, x_n)$ entonces

$$p - \text{valor} = P_{H_0}(T \geq t_0).$$

Proposición 9

Dada una prueba de hipótesis cualquiera H_0 versus H_1 cuya región crítica es de la forma

$$\{(X_1, X_2, \dots, X_n) \in \mathbb{R}^n : T(X_1, X_2, \dots, X_n) \geq c\}$$

siendo T un estadístico y c una constante. Entonces se cumple que

$$\text{rechazo } H_0 \text{ si y sólo si } \alpha \geq p - \text{valor}.$$

Demostración.

Dado α le llamo c al valor tal que $P_{H_0}(T \geq c) = \alpha$. Entonces $\alpha \geq p - \text{valor}$ si y sólo si $P_{H_0}(T \geq c) \geq P_{H_0}(T \geq t_0) = p - \text{valor}$ lo cual ocurre si y sólo si $c \leq t_0$ lo cual ocurre si y sólo si rechazo H_0 . ■

Ejemplo.

Si X_1, X_2, \dots, X_{25} es una MAS de $X \sim N(\mu, \sigma^2 = 4)$ de la cual se obtuvo $\bar{x} = 1.3$. Queremos testear $H_0 : \mu = 1$, $H_1 : \mu > 1$. Hallar el p-valor y tomar la decisión al 5%.

La región crítica es de la forma $\{\bar{X}_n \geq \mu_0 + \frac{\sigma}{\sqrt{n}}z_\alpha\} = \{\bar{X}_n \geq cte\}$ entonces $T = \bar{X}_n$, $t_0 = 1.3$.

Entonces $P_{H_0}(\bar{X}_n \geq 1.3) = 1 - \phi\left(\frac{1.3-1}{2/5}\right) = 1 - \phi(0.75) = 1 - \phi(0.773) = 0.227$. Tener en cuenta que para el cálculo anterior se usó que \bar{X}_n distribuye $N(\mu, \sigma^2/n)$ y como el cálculo se realiza bajo H_0 cierto, se tiene que \bar{X}_n distribuye $N(1, 4/25)$.

Como $\alpha = 0.05 < 0.227$ resulta que no rechazamos la hipótesis nula al 5%.

Observación 10 *En general se trabaja con valores de α pequeños, comunemente $\alpha \leq 0.1$ por lo que cuando un p-valor toma un valor tan grande como en el ejemplo, no rechazaremos H_0 cualquiera sea ese valor pequeño considerado de α .*

Ejercicio.

Con los mismos datos del ejemplo anterior calcular el p-valor y tomar la decisión al 5% para las hipótesis $H_0 : \mu = 1.5$, $H_1 : \mu < 1.5$ y luego para $H_0 : \mu = 1$, $H_1 : \mu \neq 1$.

9 Pruebas de bondad de ajuste

Cuando X_1, X_2, \dots, X_n es una MAS de X con distribución F_X , en general no conocemos la distribución de los datos. Ya vimos como aplicación de la ley de los grandes números que podemos estimar a la función F_X mediante la distribución empírica definida como

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \geq x\}} = \frac{\text{cantidad de observaciones } \leq x}{n}.$$

Por suerte existen técnicas estadísticas para realizar una prueba de hipótesis en donde en H_0 se plantea la distribución verdadera de los datos. A este tipo de pruebas se le llaman pruebas de bondad de ajuste. Por ejemplo $H_0 : X$ distribuye $N(\mu, \sigma^2)$ versus $H_1 : \text{no } H_0$ es una prueba de hipótesis así como $H_0 : X$ distribuye $\text{Exp}(\lambda = 2)$ versus $H_1 : \text{no } H_0$. O sea que en H_0 se especifica la distribución de la variable que observamos, donde pueden haber parámetros en la misma o no como en los ejemplos mencionados recién. Veremos dos test de bondad de ajuste: el llamado de Kolmogorov–Smirnov y el de Lilliefors de normalidad.

9.1 Prueba de Kolmogorov–Smirnov

Si X_1, X_2, \dots, X_n MAS de $X \sim F_X$ donde F_X es desconocida y dada una función de distribución F_0 continua y completamente conocida (es decir sin parámetros desconocidos). Se plantea

$H_0 : F_X = F_0$ (es decir $X \sim F_0$) versus

$H_1 : \text{no } H_0$.

Se plantea la región crítica de la forma

$$\left\{ (X_1, X_2, \dots, X_n) : \sup_{x \in \mathbb{R}} |F_n^*(x) - F_0(x)| \geq c \right\}.$$

La idea en la cual está basada la región crítica radica en que sabemos que $F_n^*(x)$ converge a la verdadera $F_X(x)$ para todo x (que es desconocida). Entonces si la verdadera distribución fuera F_0 la distribución empírica debería converger en todo punto a $F_0(x)$. Por lo tanto si n es grande, la distribución empírica

debería estar cerca de F_0 en todo punto, por lo que si existiera algún punto en el cual la diferencia entre la empírica y F_0 fuera grande, deberíamos rechazar la hipótesis nula. Eso lo podemos medir mediante el supremo de la diferencia que es el estadístico de la prueba. Para poder hallar el valor de la constante c para que la región considerada tenga nivel α , deberíamos resolver la ecuación

$$P_{H_0} \left(\sup_{x \in \mathbb{R}} |F_n^*(x) - F_0(x)| \geq c \right) = \alpha$$

para lo cual es necesario saber la distribución del estadístico $T = \sup_{x \in \mathbb{R}} |F_n^*(x) - F_0(x)|$. Esa distribución para n fijo no es conocida, pero sí hay un resultado asintótico del estilo del TCL que nos da la distribución límite. El resultado siguiente sería una especie de TCL para el supremo de la diferencia (en lugar de para un promedio muestral). El teorema siguiente se deduce a partir de un resultado mucho más potente y general que es conocido con el nombre de principio de invarianza de Donsker.

Teorema 11

Si X_1, X_2, \dots, X_n es una MAS de X con distribución F_0 , entonces

$$\lim_{n \rightarrow +\infty} P \left(\sqrt{n} \sup_{x \in \mathbb{R}} |F_n^*(x) - F_0(x)| \leq x \right) = 1 - 2 \sum_{k=1}^{+\infty} (-1)^{k-1} e^{-2k^2 x^2} \text{ para todo } x.$$

A partir de la función que aparece a la derecha de la igualdad, se pueden obtener de manera aproximada tanto los valores de c para que la región tenga nivel α , así como el p-valor de la prueba. Para valores de n pequeños, los valores críticos están tabulados en la tabla llamada de Kolmogorov–Smirnov que será utilizada para tomar la decisión cuando implementemos esta prueba de hipótesis.

9.1.1 Cálculo del estadístico de prueba

Sabemos que F_0 es una función no decreciente, y F_n^* también es no decreciente, pero constante a trozos. Recordar que si ordenamos la muestra de menor a mayor y les llamamos a dichos valores $X_1^* < X_2^*, \dots < X_n^*$ (al ser la distribución continua hay probabilidad cero de observar igualdades entre las observaciones) tenemos que $F_n^*(x) = i/n$ para todo $x \in [X_i^*, X_{i+1}^*)$. Si nos restringimos al intervalo $x \in [X_i^*, X_{i+1}^*)$, tenemos que calcular

$$\sup_{x \in [X_i^*, X_{i+1}^*)} |i/n - F_0(x)|$$

pero observando que F_0 es una función creciente dicho extremo se puede dar en alguno de los extremos del intervalo, por lo que bastará con evaluar en ambos extremos y tomar el mayor valor. Concluimos entonces que

$$\sup_{x \in [X_i^*, X_{i+1}^*)} |i/n - F_0(x)| = \max\{|i/n - F_0(X_i^*)|, |i/n - F_0(X_{i+1}^*)|\}.$$

Esta última igualdad es válida para los valores de i tales que $1 \leq i \leq n-1$. Por otro lado, observamos que en los intervalos extremos $(-\infty, X_1^*)$ y $[X_n^*, +\infty)$ tenemos que

$$\sup_{x \in (-\infty, X_1^*)} |F_n^*(x) - F_0(x)| = F_0(X_1^*)$$

y

$$\sup_{x \in [X_n^*, +\infty)} |F_n^*(x) - F_0(x)| = 1 - F_0(X_n^*).$$

Finalmente, teniendo en cuenta que para buscar el supremo sobre todos los $x \in \mathbb{R}$ basta considerar una partición en intervalos de los reales, obtener el supremo en cada intervalo y luego tomar el máximo de ellos, obtenemos que

$$\begin{aligned} & \sup_{x \in \mathbb{R}} |F_n^*(x) - F_0(x)| = \\ & \max\left\{ \sup_{x \in (-\infty, X_1^*)} |F_n^*(x) - F_0(x)|, \dots, \sup_{x \in [X_i^*, X_{i+1}^*)} |i/n - F_0(x)|, \dots, \sup_{x \in [X_n^*, +\infty)} |F_n^*(x) - F_0(x)| \right\} = \\ & \max\{F_0(X_1^*), \dots, \max\{|i/n - F_0(X_i^*)|, |i/n - F_0(X_{i+1}^*)|\}, \dots, 1 - F_0(X_n^*)\}. \end{aligned}$$

En definitiva, la fórmula anterior la podemos compactificar en la siguiente:

$$\sup_{x \in \mathbb{R}} |F_n^*(x) - F_0(x)| = \max \left\{ \max_{1 \leq i \leq n} |i/n - F_0(X_i^*)|, \max_{1 \leq i \leq n} |(i-1)/n - F_0(X_i^*)| \right\}$$

9.1.2 Procedimiento de cálculo del estadístico

De acuerdo a lo observado en la subsección anterior, tenemos el siguiente procedimiento para el cálculo del estadístico T .

Paso 1. Ordenamos los datos de menor a mayor ($X_1^* < X_2^*, \dots < X_n^*$).

Paso 2. Calculamos todos los valores de la forma $|i/n - F_0(X_i^*)|$ con i variando entre 1 y n .

Paso 3. Calculamos todos los valores de la forma $|(i-1)/n - F_0(X_i^*)|$ con i variando entre 1 y n .

Paso 4. T es el valor máximo entre todos los puntos hallados en los pasos 2 y 3.

Dichos pasos pueden ser llevados a cabo mediante una tabla como sigue.

| i | X_i^* | $ i/n - F_0(X_i^*) $ | $ (i-1)/n - F_0(X_i^*) $ |
|----------|----------|----------------------|--------------------------|
| 1 | | | |
| 2 | | | |
| \vdots | \vdots | \vdots | \vdots |
| n | | | |

En definitiva tendríamos que el valor de nuestro estadístico T es el mayor valor entre el máximo de la columna 3 y el máximo de la columna 4.

Ejemplo.

Supongamos que tenemos los siguientes datos 0.8, 1, 0.6, 0.5, 0.7.

Queremos saber si es razonable suponer que los datos siguen una distribución exponencial con $\lambda = 1$. O sea que realizaremos el test

$H_0 : X \sim \text{Exp}(\lambda = 1)$ versus

$H_1 : \text{no } H_0$.

Trabajaremos al nivel 5%.

En este caso tenemos que $F_0(x) = 1 - e^{-x}$ para $x \geq 0$, $n = 5$ (tamaño de la muestra)

| i | X_i^* | $ i/5 - (1 - e^{-X_i^*}) $ | $ (i-1)/5 - (1 - e^{-X_i^*}) $ |
|-----|---------|----------------------------|--------------------------------|
| 1 | 0.5 | 0.1935 | 0.3935 |
| 2 | 0.6 | 0.0512 | 0.2512 |
| 3 | 0.7 | 0.0966 | 0.1034 |
| 4 | 0.8 | 0.2493 | 0.0493 |
| 5 | 1 | 0.3679 | 0.1679 |

el máximo de la primer columna es 0.3679, el máximo de la segunda columna

es 0.3935. Entonces $T = \max\{0.3679, 0.3935\} = 0.3935$.

Ahora vamos a la tabla para el test de Kolmogorov–Smirnov. Nos restringimos a la línea número 5 de la tabla, dado que $n = 5$. Como el nivel de significación de nuestra prueba es $\alpha = 0.05$, la tabla nos dice que el valor crítico es 0.565.

Eso significa que

$$P_{H_0} \left(\sup_{x \in \mathbb{R}} |F_n^*(x) - F_0(x)| \geq 0.565 \right) = 0.05.$$

Como en nuestro caso obtuvimos $T = \sup_{x \in \mathbb{R}} |F_n^*(x) - F_0(x)| = 0.3935$, resulta que $0.3935 \not\geq 0.565$, o sea que no se cumple la región crítica, por lo que la decisión es no rechazar H_0 , o sea que no tenemos suficiente evidencia como para rechazar la hipótesis de que los datos provengan de una variable con distribución $\text{Exp}(\lambda = 1)$. ¿Qué se puede decir del p-valor de esta prueba?

9.2 Test de normalidad de Lilliefors

El test de Lilliefors de normalidad, es un test donde en H_0 se plantea la hipótesis de que los datos tienen distribución normal. Es decir que dada X_1, X_2, \dots, X_n MAS de X con distribución desconocida, planteamos

$H_0 : X \sim N(\mu, \sigma^2)$ para determinados μ y σ^2 versus

$H_1 : \text{no } H_0$.

Este test fue propuesto por Hubert Lilliefors en un artículo publicado en 1967 bajo el nombre “On the Komogorov–Smirnov Test for Normality with Mean and Variance Unknown” en JASA (Journal of the American Statstical Association).

El test en su implementación es similar al de Kolmogorov–Smirnov. La región crítica es

$$\left\{ (X_1, X_2, \dots, X_n) : \sup_{x \in \mathbb{R}} \left| F_n^*(x) - \phi\left(\frac{x - \bar{X}_n}{s_n}\right) \right| \geq c \right\}$$

siendo $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Notar que a este estimador de la varianza lo estamos indicando con la letra “s” minúscula (a diferencia de la “S” mayúscula utilizada para las fórmulas de intervalos de confianza y pruebas de hipótesis paramétricas. Observamos que para el cálculo del estadístico en esta prueba se procede de forma idéntica al de Kolmogorov–Smirnov, cambiando $F_0(x)$ por $\phi\left(\frac{x - \bar{X}_n}{s_n}\right)$. Una vez realizado el cálculo del estadístico T , se toma la decisión con ayuda de la tabla llamada de Lilliefors.

Ejercicio.

Con los datos del ejemplo anterior, realizar la prueba de Lilliefors al 5% para ver si los datos 0.8, 1, 0.6, 0.5, 0.7 pueden ser considerados con distribución normal. ¿Cómo se interpreta este resultado junto con el no rechazo de que provienen de una distribución exponencial?