Introducción a la Estadística usando Software 2025: Test de Hipótesis

Juan Piccini

LPE/IMERL

Test de Hipótesis

Juan Piccini

Indice

Generalidades

- Todos recordamos el método de demostración por absurdo o razonamiento por absurdo.
- Si queremos probar que una afirmación o una cierta hipótesis es falsa, la suponemos verdadera, y a partir de esto intentamos encontrar una contradicción (cualquiera sea).
- En estadística existe un razonamiento similar, el **razonamiento por improbable**.
- El mismo consiste en suponer que una cierta hipótesis es verdadera, y a partir de esta suposición calcular la probabilidad de observar algo tanto o más extremo que lo observado.
- Si esta probabilidad es muy baja, concluímos que la hipótesis es falsa, o al menos poco creíble.

Generalidades

- A modo de ejemplo, tenemos una moneda y suponemos que la misma es honesta (no está cargada).
- De ser verdad esto, entonces uno esperaría que en un gran número de lanzamientos la cantidad de caras y cruces sea equilibrada.
- Esto es, si llamamos p a la probabilidad de obtener cara en un lanzamiento, el suponer que la moneda es honesta se traduce en suponer $p = 1/2$.
- Por tanto si lanzamos la moneda p.ej. 100 veces y observamos 90 caras y 10 cruces, mas allá del cálculo preciso sobre la probabilidad de obtener 90 o más caras en 100 lanzamientos es claro que este resultado es poco creíble si la moneda fuese honesta.
- Volveremos pronto sobre este ejemplo.

Qué es un Test de Hipótesis?

- Un principio general en ciencia es elegir siempre la hipótesis más simple capaz de explicar la realidad observada.
- Una hipótesis se contrasta comparando sus predicciones con la realidad: si coinciden (dentro del margen de error admisible), mantendremos la hipótesis, en caso contrario la rechazaremos y buscaremos otra.
- Cuando las predicciones generadas por la hipótesis son en términos de probabilidad, la metodología empleada es la teoría de **Test de Hipótesis** o **Contraste de Hipótesis**.
- **Puede definirse como un procedimiento para juzgar si una propiedad de una población es compatible con lo observado en una muestra de dicha población.**
- Ilustraremos con un par de ejemplos.

Ejemplo1

- Supongamos que tenemos una moneda y deseamos decidir si la misma es honesta o no.
- Para ello lanzamos la moneda n veces (p. ej. 100 veces), obteniendo $M = \{x_1, x_2, \ldots, x_{100}\}$ donde las x_i son realizaciones de las V.A. $X_i \sim Ber(p)$ ($X_i = 1$ si sale cara, $X_i = 0$ sino).
- Si los lanzamientos se hacen siempre bajo las mismas condiciones, podemos asumir que M es una muestra i.id.
- Suponer que la moneda es honesta equivale a suponer $p = 1/2$. Por tanto tenemos dos escenarios: O la moneda es honesta (que es lo que deseamos comprobar), o no lo es.
- La hipótesis a comprobar suele llamarse **Hipótesis nula** o H_0, y representa la opinión que mantendremos a menos que los datos indiquen su falsedad (en el ejemplo, H_0 es que la moneda es honesta, o sea $p = 1/2$).
- La hipótesis alternativa o H_1 es el complemento de H_0 ($p \neq 1/2$).

Ejemplo1

- En este caso se trata de discernir en base a los datos si es razonable pensar que $p = 1/2$. Como la pregunta trata sobre el valor de un parámetro (el valor de p), hablaremos de **Test de Hipóteis paramétrico**.
- Si en los 100 lanzamientos se obtuviesen p.ej. 95 caras y 5 cruces, nadie dudaría en rechazar H_0, lo mismo sería para p. ej. 90/10 u 80/20.
- Y si obtenemos p.ej. 60/40, o 55/45? Rechazaríamos H_0? Es claro que hay que trazar una línea en algún lado, de modo que si rechazamos H_0 podamos hacerlo con un alto nivel de confianza en la decisión tomada.
- Notemos que estamos usando como indicador la variable aleatoria $Y = \sum_{i=1}^{100} X_i$, y si H_0 es cierta $Y \sim Bin(100, 1/2)$.

Ejemplo1

- Una forma de decidir es hallar (suponiendo que la moneda fuera honesta) un intervalo [a,b] que con alta probabilidad (p.ej. 90%) debería contener el valor Y. Como bajo H_0 tenemos que $Y \sim Bin(100, 1/2)$, los cálculos anteriores pueden realizarse sin dificultad, tanto si elegimos un intervalo que concentre el 90% de la probabilidad como si elegimos otro que concentre el 95% o el 99% u otro valor.
- Fijando 90%, hallamos $[a, b]$ tal que $P(Y \in [a, b]) \geq 0.90$ haciendo a=binoinv(.05,100,.5) y b=binoinv(.95,100,.5) (obtenemos el intervalo $[42, 58]$). Si observamos un número de caras fuera de dicho intervalo diremos que no parece razonable que una moneda honesta produzca semejante resultado, y podremos rechazar H_0 "con un 90% de certeza ".
- Si fijamos 95%, haciendo a=binoinv(.025,100,.5) y b=binoinv(.975,100,.5) obtenemos obtenemos a=40 y b=60.

Ejemplo 1: Observaciones

- Si bien no es imposible que en 100 lanzamientos de una moneda honesta obtengamos menos de 42 o más de 58 caras, la probabilidad es de 0.10, mientras que para menos de 40 o más de 60 caras la probabilidad es de 0.05.
- Esto significa que si repetimos N veces el experimento de lanzar 100 veces la moneda y contar la cantidad de caras, el 10% de esas N veces observaremos valores fuera del intervalo $[42, 58]$ y el 5% fuera del intervalo $[40, 60]$.
- Supongamos que en nuestros 100 lanzamientos obtuvimos 64 caras. Puede pasar que la moneda sea honesta y nuestros 100 lanzamientos sean una de esas pocas veces, o puede pasar que la moneda no sea honesta.
- Asumiendo que la naturaleza prefiere los eventos de mayor probabilidad, elegimos creer que la moneda no es honesta y rechazamos H_0 al nivel 0.95.

Ejemplo2

- Una fábrica produce un artículo cuya duración sigue una distribución normal con media de 5000 horas y desviación típica 100 horas.
- Se introducen cambios en el proceso de fabricación que pueden cambiar la media pero no la varianza.
- Para contrastar si dichos cambios han producido efectos, se toma una muestra de cuatro elementos cuyas vidas resultan ser 5010, 4750, 4826 y 4953 hs.
- Basado en estos datos, hay evidencia de un cambio en la media?
- Las dos hipótesis posibles son H_0 : no hay efectos ($\mu = 5000 hs.$) y H_1 : $\mu \neq 5000 hs.$.

Ejemplo2

- Como la suma de V.A. normales es normal, tenemos que $\overline{X}_4 \sim N(5000, \frac{\sigma^2}{4})$, por lo que, si H_0 fuese cierta, $\overline{X}_4$ debería ser como sortear al azar un valor de una distribución normal con media 5000 hs. y desvío $\sqrt{\frac{\sigma^2}{4}} = \frac{\sigma}{2} = 50$ hs.
- Recordemos que si $X \sim N(\mu, \sigma^2)$ entonces $\frac{X-\mu}{\sigma} \sim N(0,1)$.
- Como $\frac{\overline{X}-5000}{\sigma/2} = \frac{2(\overline{X}_4-5000)}{\sigma} \sim N(0,1)$ y el 95% de la masa en una normal estándar se encuentra entre -1.96 y 1.96, tenemos que $P(-1.96 \leq \frac{2(\overline{X}_4-5000)}{\sigma} \leq 1.96) = P(|\frac{2(\overline{X}_4-5000)}{\sigma}| \leq 1.96) = 0.95$
- Esto es, $P(|\overline{X}_4 - 5000| \leq 1.96 \times 50) = 0.95$. Como $1.96 \times 50 = 98$, si H_0 fuese cierta entonces en el 95% de los casos la el promedio no debería alejarse de $\mu = 5000hs.$ en más de 98 hs.

Ejemplo 2

- Esto significa que en el 95% de las veces en que hacemos el promedio de 4 artículos elegidos al azar, dicho promedio debería estar entre 4902 y 5098 hs **si H0 fuese cierta**.
- Como el promedio obtenido es 4884,75 hs. (algo muy improbable bajo H_0), tenemos dos opciones:

1. Aceptar H_0 y atribuír el resultado al azar. Notemos que esto implica creer que la baja probabilidad (5%) del resultado que obtuvimos es simple "mala suerte ", nos tocó una muestra de 4 artículos muy poco común bajo H0.
2. Rechazar H_0 y concluír que se ha producido un cambio (el nuevo proceso produce artículos cuya vida útil tiene una media menor a 5000 hs.), lo que explicaría el resultado observado.

Observaciones

A partir de estos dos ejemplos podemos identificar puntos en común:

- A partir de los datos construímos Y (**el estadístico del test**), que es la V.A. mediante la cual decidiremos.
- Bajo H_0 el estadístico Y tiene una cierta distribución ($Bin(100, 1/2)$ en el ejemplo 1, $N(5000, 50)$ en el ejemplo 2) con la cual podemos luego calcular probabilidades.
- Dado un nivel α (p.ej. 90% en el ejemplo 1, 95% en el ejemplo 2) se fija una **Región crítica** R_α o región de rechazo (caer fuera de determinado intervalo, esto es, que el valor de Y caiga en $[42, 58]^c$ en el ejemplo 1, que caiga en $[4902, 5098]^c$ en el ejemplo 2).
- Esta región crítica cumple $P(Y \in R_\alpha) \leq \alpha$.

Observaciones

- Si el valor que toma Y con los datos concretos (que llamaremos $\hat{Y}$) cae en la zona crítica, rechazamos H_0, caso contrario no rechazamos H_0.
- Hemos fijado de antemano el valor α, el grado de evidencia necesario para rechazar H_0.
- Cuanto más creamos que H_0 es cierta, más evidencia en contra hará falta para rechazarla.
- El valor $\hat{Y}$ que toma Y depende del juego de datos. Con otra muestra podríamos perfectamente obtener otro valor $\hat{Y}$ que nos llevara a tomar otra decisión.
- A continuación algunas definiciones.

Hipótesis estadística

- Una **Hipótesis estadística** es una suposición que determina total o parcialmente la distribución de probabilidad de una o varias V.A.
- Dichas hipótesis pueden clasificarse en
  1. **paramétricas:** La suposición se refiere al valor o valores de uno parámetro o parámetros.
  2. **no paramétricas:** La suposición se refiere a la forma de la distribución de la V.A.
- El segundo tipo se vincula con la estadística descriptiva ya vista. P.ej. es razonable suponer que la distribución de una cierta variable es normal? Si tenemos dos conjuntos de datos, es razonable suponer que provienen de la misma distribución?
- Una **hipótesis simple** es $\theta = \theta_0$, mientras que una **hipótesis compuesta** es p.ej. $\theta > \theta_0$ o $\theta \in [a, b]$

La hipótesis nula

- Llamamos **Hipótesis nula** (H_0) a la hipótesis que se contrasta. Es la hipótesis que mantendremos a menos que los datos indiquen su falsedad, por lo que debe entenderse como "neutra ".
- H_0 nunca se considera probada, aunque puede ser rechazada por los datos.

  P. ej. dadas dos poblaciones A y B, la hipótesis A=B puede ser rechazada encontrando elementos de A y B distintos, pero no puede ser demostrada excepto estudiando todos los elementos de ambas poblaciones, lo que puede ser imposible.
- Análogamente, la hipótesis de que dos poblaciones tienen la misma media puede rechazarse cuando ambas medias difieran mucho, analizando muestras suficientemente grandes de ambas poblaciones, pero no puede ser demostrada mediante muestreo (puede suceder que las medias difieran en un valor pequeño, imperceptible en el muestreo).

La hipótesis nula

- H_0 se elige habitualmente según el **principio de simplicidad científica (navaja de Occam)**, que puede resumirse como que solamente debemos abandonar un modelo simple a favor de otro más complejo cuando la evidencia a favor de este último sea fuerte.

- Por tanto en los T.H. paramétricos, H_0 suele ser que el parámetro es igual a un valor concreto que se toma como referencia.

- Cuando comparamos poblaciones, H_0 es que las poblaciones son iguales.

- Cuando investigamos la forma de de la distribución, H_0 suele ser que los datos siguen una distribución conocida (Normal, Poisson, Exponencial, etc.)

La hipótesis alternativa

- Cuando rechazamos H_0 implícitamente estamos aceptando una hipótesis alternativa o H_1.
- Suponiendo H_0 del tipo $\theta = \theta_0$, los casos más importantes de H_1 son:
  1. Desconocemos en qué dirección puede ser falsa H_0 y por tanto tendremos $H_1 : \theta \neq \theta_0$ (Test de Hipótesis **bilateral**).
  2. Sabemos que si $\theta \neq \theta_0$ forzosamente $\theta > \theta_0$ (o bien $\theta < \theta_0$). Llamamos a esto un Test de Hipótesis **unilateral**.

Metodología

- Establecida por Fisher, Neyman y Pearson entre 1920 y 1933, su lógica es la de un juicio penal, donde $H_0 =$ "inocente "y solamente se rechazará si la evidencia es contundente (o sea que los datos observados son muy improbables si fuese cierta H_0).
- Esto presupone:
  1. Definir H_0 y H_1
  2. Definir un estadístico Y para el test (es una función de los datos, una cuenta que hacemos con los datos, una variable aleatoria) cuyo valor servirá como medida para evaluar si lo observado se comporta como debería si H_0 fuese cierta.
  3. Decidir cuando una discrepancia se considera inadmisible, esto es, cuando la discrepancia es demasiado grande para ser creíble como producto del azar (esto es, fijar $\alpha \in (0, 1)$).

Metodología

4 Trazada la línea a partir de la cual una discrepancia es demasiado grande (algo que bajo H_0 sea de probabilidad muy baja), hallamos la región crítica del test, o sea el conjunto de valores del estadístico Y que tienen muy baja probabilidad de ser observados bajo H_0.

5 Tomar los datos, calcular el valor $\hat{Y}$ del estadístico Y del test y ver si la discrepancia es aceptable o no. Si lo es, no rechazamos H_0, caso contrario la rechazamos y aceptamos H_1

- Si el valor observado $\hat{Y}$ del estadístico cae en dicha región crítica, rechazaremos H_0.

Errores tipo I y tipo II

- Existen dos tipos de errores que podemos cometer:
- Error tipo I, que consiste en rechazar H_0 cuando la misma es verdadera (condenar a un inocente).
- Error tipo II, que consiste en no rechazar H_0 cuando la misma es falsa (absolver a un culpable).
- La probabilidad de cometer un error tipo I es precisamente α, el nivel de significación del test, mientras que la probabilidad de cometer un error tipo II se suele llamar β.
- Se llama **potencia del test** a $1 - \beta$ (la probabilidad de haber rechazado bien H_0).
- Si llamamos $P(A|B)$ a la probabilidad de decidirnos por A cuando de hecho B es la opción correcta, tenemos que $\alpha = P(H_1|H_0)$, $1 - \alpha = P(H_0|H_0)$, $\beta = P(H_0|H_1)$ y $1 - \beta = P(H_1|H_1)$.

Errores tipo I y tipo II

Decisión ↓Realidad→		
	H_0	H_1
H_0	$1-\alpha$	β
H_1	α	$1-\beta$

- Los valores α y β se comportan en forma opuesta: cuando uno decrece el otro crece. No se puede lograr que ambos sean simultáneamente pequeños.
- Esto obliga a elegir cual de los dos mantendremos controlados. Se elige controlar α (que sea pequeño), o sea que cuando rechacemos H_0 lo haremos con la "conciencia tranquila ".
- Esto nos lleva al concepto de p-valor.

El p-valor

- **El p-valor** puede interpretarse como el nivel de credibilidad que tiene H_0.
- Dicho en otras maneras:
- Es la probabilidad bajo H_0 de obtener un valor de Y tanto o más extremo al obtenido a partir de los datos, $\hat{Y}$.
- Es la probabilidad de obtener una discrepancia tanto o más extrema que la observada en la muestra cuando H_0 es cierta.
- Es la probabilidad bajo H_0 de observar algo tanto o más extremo que lo observado.

El p-valor

- Si $\hat{Y}$ es el valor observado del estadístico, $p = P(Y \geq \hat{Y}|H_0)$, por tanto p se determina a partir de la muestra, no se fija a priori.
- Cuanto menor sea p, menor es la probabilidad de aparición de una discrepancia como la observada, menor es la credibilidad de H_0.
- Fijado un nivel de significación α (habitualmente $\alpha = 0.10$, aunque también se usa $\alpha = 0.05$ o $\alpha = 0.01$), si el p-valor es inferior a α se rechaza H_0.
- La mayor parte de los tests implementados en paquetes de software proporcionan el p-valor correspondiente a los datos con los que se performa el test, correspondiendo al usuario decidir si se rechaza o no H_0. Asimismo utilizan regiones críticas que para un α dado tienen el menor β posible (máxima potencia).

Observación

- Si rechazamos H_0 cuando de hecho no debíamos hacerlo (cometemos un error tipo I), esto tendrá consecuencias que dependen del problema y de H_0.
- Supongamos un control de calidad de un determinado artículo (p.ej. licuadoras). Fijamos $\alpha = 0.10$. Si $H_0 =$ "el lote es bueno ", entonces con probabilidad $= 0.10$ estaremos rechazando un lote bueno (perdemos de enviarlo al mercado).
- Si $H_0 =$ "el lote es malo ", con probabilidad 0.10 estaremos enviando al mercado un lote malo. En este caso habrá que sopesar qué importa más, si lo que nos perdemos de vender o el buen nombre de la marca.
- Si hablamos de medicamentos, entonces las consecuencias son bien distintas: en un caso nos perdemos de vender un lote bueno, en el otro arriesgamos la salud de la gente al distribuír un lote malo. En este último caso es inadmisible usar $\alpha = 0.10$, usaremos $\alpha = 0.01$ o incluso menor.

Poblaciones normales: Test sobre la media

- Supongamos una muestra i.id $M = \{x_1, \ldots, x_n\}$ **con distribución normal** $N(\mu, \sigma^2)$ de parámetro μ desconocido.
- Deseamos contrastar $\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$
- Tenemos dos casos:
  a) Varianza σ^2 conocida
  b) Varianza σ^2 desconocida

En a) tenemos que bajo H_0 el promedio $\overline{X}_n$ es una V.A. normal con media μ desconocida y varianza $\sigma^2_{\overline{X}_n} = \frac{\sigma^2}{n}$ conocida, de donde el estadístico $Y = \frac{\overline{X}_n - \mu_0}{\sqrt{\sigma^2_{\overline{X}_n}}} = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma}$ tendrá distribución normal estándar.

- Por tanto, fijado $\alpha \in (0,1)$, la región crítica (o de rechazo) al nivel α será el complemento del intervalo $R_\alpha = (-z_{\alpha/2}, z_{\alpha/2})^c$ tal que $P(Y \in R_\alpha = \alpha)$, o sea $P(Y \notin R_\alpha = 1 - \alpha)$

Poblaciones normales: Test sobre la media

- Esto es, no rechazaremos H_0 si

$$-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \overline{X}_n - \mu_0 \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, si|\overline{X}_n - \mu_0| \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad (1)$$

En b) tenemos que bajo H_0 el promedio $\overline{X}_n$ es una V.A. normal con media μ y varianza σ^2 desconocidas, por lo que el estadístico $Y = \frac{\overline{X}_n - \mu_0}{\sqrt{s_n^2}} = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{s_n}$ tendrá distribución t de Student con n-1 grados de libertad ($s_n = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{X}_n)^2}{n-1}}$).

La región de aceptación será

$$|\overline{X}_n - \mu_0| \leq t_{\alpha/2}\frac{s_n}{\sqrt{n}} \quad (2)$$

donde $t_{\alpha/2}$ tal que $P(d \in (-t_{\alpha/2}, t_{\alpha/2}) = 1 - \alpha)$.

Ejemplos: a)Varianza conocida

- Generamos 16 datos $\sim N(5,1)$: x=normrnd(5,1,1,16), obtenemos

5.670	5.432	6.435	3.922	4.459	5.351	5.428	4.401
4.968	5.046	6.992	5.066	2.916	6.017	5.435	3.891

- La media es M=mean(x)=5.089, hacemos el test $\begin{cases} H_0: & \mu = M \\ H_1: & \mu \neq M \end{cases}$ usando la función z-test de octave, z-test(x,M,1), obtenemos un p-valor =1 (no rechazamos H_0). Con M=4, z-test(x,4,1) nos da pval = 1.308e-05 (la muestra no viene de una normal con M=4).
- Si mantenemos M=4 y hacemos test unilateral $\begin{cases} H_0: & \mu = 4 \\ H_1: & \mu < 4 \end{cases}$ obtenemos: pval = 0.999 (entre $\mu = 4$ y $\mu < 4$ se queda con la primera).
- Si hacemos $\begin{cases} H_0: & \mu = 4 \\ H_1: & \mu > 4 \end{cases}$ obtenemos :pval= 6.541e-06 (prefiere H_1)

Ejemplos: b)Varianza desconocida

- Para los mismos datos, hacemos $\begin{cases} H_0: & \mu = M \\ H_1: & \mu \neq M \end{cases}$ mediante el t-test de octave: t-test(x,4) , obteniendo: pval=6.397e-04 (rechaza H_0).
- Si hacemos $\begin{cases} H_0: & \mu = 4 \\ H_1: & \mu < 4 \end{cases}$ obtenemos: pval: 0.99968 (entre $\mu = 4$ y $\mu < 4$ se queda con la primera).
- Si hacemos $\begin{cases} H_0: & \mu = 4 \\ H_1: & \mu > 4 \end{cases}$ obtenemos: pval 3.1986e-04 (rechaza enfáticamente $\mu = 4$ frente a $\mu > 4$)

Test sobre la varianza

- En este caso tenemos $\begin{cases} H_0: & \sigma^2 = \sigma_0^2 \\ H_1: & \sigma^2 \neq \sigma_0^2 \end{cases}$

Bajo H_0, el estadístico $Y = \frac{(n-1)s^2}{\sigma_0^2}$ tiene distribución χ^2_{n-1} (Ji-cuadrado o Chi-cuadrado con n-1 grados de libertad).

- Como esta distribución no es simétrica, dado $\alpha \in (0,1)$ hallamos los valores $\chi^2_{1-\alpha/2}$ y $\chi^2_{\alpha/2}$ tal que $P(Y \in (\chi^2_{1-\alpha/2}, \chi^2_{\alpha/2}) = 1-\alpha$.

- Para $H_1: \sigma^2 > \sigma_0^2$ la región de rechazo es $\hat{Y} \geq \chi^2_{\alpha/2}$, mientras que para $H_1: \sigma^2 < \sigma_0^2$ es $\hat{Y} \leq \chi^2_{1-\alpha/2}$

Ejemplo

- Se espera que la resistencia (Kg/cm2) de cierto material se distribuya normalmente con una media de 220 y un desvío típico de 7.75. Se toma una muestra de 9 elementos, obteniendo $M = \{203, 229, 215, 220, 223, 233, 208, 228, 209\}$ y s=10.524.
- El contraste para la varianza es $\begin{cases} H_0: & \sigma^2 = 7.75^2 \\ H_1: & \sigma^2 > 7.75^2 \end{cases}$
- Para $H_1 : \sigma^2 > \sigma_0^2$ la región de rechazo es $Y = \frac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{\alpha/2}$, donde la distribución es χ^2_8 con 8 grados de libertad.
- El valor del estadístico es $\hat{Y} = 14.75$, y el p-valor es $P(\chi^2_8 \geq 14.75)$, que calculamos haciendo p=chi2cdf(14.75,8) (o sea $P(\chi^2_8 \leq 14.75)$). Obtenemos p=0.93560, por lo que el p-valor es 1-0.93560=0.06443.
- Si tomamos $\alpha = 0.05$ o 0.01 no rechazaremos H_0, pero si elegimos $\alpha = 0.10$ tendríamos que rechazar H_0.

Test de Hipótesis mediante simulaciones

- La gran facilidad de cálculo disponible hoy día nos permite una forma adicional de realizar tests de hipótesis.
- Siguiendo el concepto de p-valor, supongamos una muestra de la que conocemos la distribución y H_0 expresa alguna propiedad de la población de la que proviene dicha muestra.
- Podemos simular entonces muchas muestras "sintéticas "con la misma distribución de nuestros datos y para cada una de estas muestras simuladas calcular la propiedad en cuestión.
- Luego vemos como se distribuye dicha propiedad en esa población simulada y vemos que tan probable es observar un valor tanto o más extremo que el obtenido de nuestros datos.
- Si dicha probabilidad es muy baja, rechazaremos H_0.

Ejemplo

Se tiene los siguientes datos sobre el tiempo entre fallas (en horas) de un cierto sistema:

1	43.21	24.62	152.31	65.27	51.13
2	100.74	198.66	3.45	69.41	263.85
3	158.99	45.31	56.63	250.37	231.83
4	379.75	148.93	2.85	97.61	173.03
5	37.06	118.45	50.00	53.26	102.71

Table: Tiempo entre fallas

Se presume que la distribución de los datos es $\mathcal{E}(\lambda)$ y se desea contrastar la hipótesis $\begin{cases} H_0: & \lambda = 100 \\ H_1: & \lambda > 100 \end{cases}$

Ejemplo

- Para ello estimamos λ a partir de la muestra, obteniendo $\hat{\lambda} = 115.18$
- Generamos entonces muchas muestras de 25 datos provenientes de una distribución exponencial de parámetro $\lambda = 100$ y para cada una de esas muestras estimamos λ
- Hacemos luego el histograma de dichas estimaciones y vemos donde cae el valor $\hat{\lambda} = 115.18$ obtenido de nuestros 25 datos (opcional).
- Calculamos la probabilidad de obtener un valor tanto o más alejado de $\lambda = 100$ que el obtenido de los datos.

Ejemplo

1. El promedio de nuestros 25 datos lo calculamos mediante: mean(datos), obteniendo:ans = 115.18.
2. Luego generamos 1000 muestras de 25 elementos cada una provenientes de una distribución exponencial de parámetro $\lambda = 100$ mediante:
   muestra=exprnd(100,1000,25); (el ; del final es para que no nos muestre dicha matriz).
3. Ahora calculamos la media para cada fila mediante:
   medias=mean(muestra,2);
4. Esto nos da las medias de cada una de las 1000 filas, es una matriz con 1000 filas y una columna, lo que podemos ver si hacemos: size(medias), nos dirá que el tamaño de medias es 1000 x 1.
5. Buscamos cuantas de esas 1000 medias superan el valor 115.18 de nuestros datos, mediante :
   size(find(medias >=115.18)), obtenemos 221 x 1, esto nos dice que en 221 de las 1000 muestras simuladas la media fue mayor o igual a

Ejemplo

- Como 221/1000=0.221, nuestra muestra no es un fenómeno raro bajo H_0.
- Esto es, hay una probabilidad del 22.1% de que una distribución exponencial de parámetro $\lambda = 100$ produzca una muestra de 25 datos con un valor $\hat{\lambda}$ tanto o más alejado del valor $\lambda = 100$ (que propone H0) que la muestra que observamos.
- Como dicha probabilidad es mayor a 0.10 o 0.05 o 0.01 (los valores habituales de rechazo), no rechazamos H_0.
- Generamos un histograma de 30 subintervalos de las 1000 medias mediante: hist(medias,30), luego fijamos dicha figura mediante: hold on;
- Agregamos una línea vertical roja en el valor 115.18 mediante: line([115.18,115.18],ylim,'LineWidth',2,'Color','r')
- El histograma se muestra en la figura (1).

Ejemplo

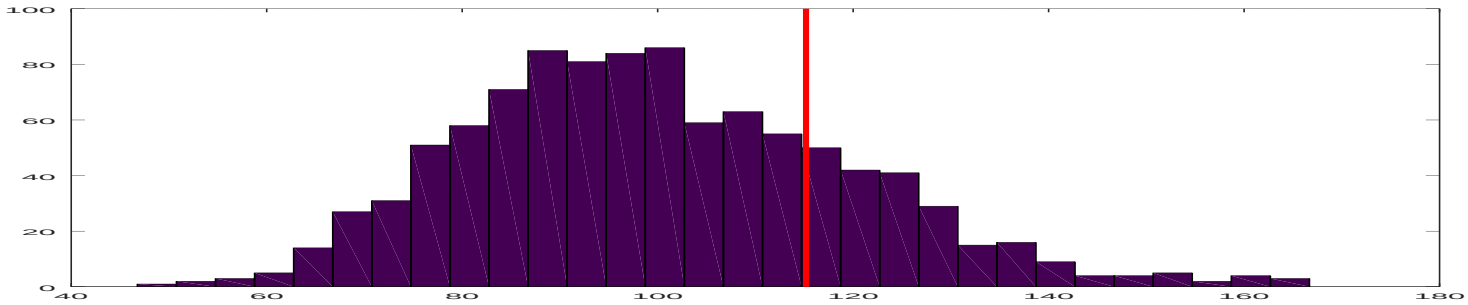

Figure: 1

- Puede verse que la proporción de muestras cuyos promedios fueron iguales o mayores al observado en los datos es apreciable (22.1%).
- Esto nos dice que es perfectamente razonable suponer que nuestros 25 datos provienen de la distribución que H_0 propone.