

Tema 1

Introducción a la Inferencia Estadística. Estadísticos muestrales

El objetivo de este tema es introducir una serie de conceptos básicos que serán fundamentales para el desarrollo y estudio posterior, a lo largo de los siguientes temas, de la Inferencia Estadística. Previo a ello se va a hacer una pequeña introducción donde se recordarán los principales elementos asociados al estudio del Cálculo de Probabilidades, herramienta imprescindible para el estudio de la Inferencia.

1.1. Introducción

El Cálculo de Probabilidades proporciona una teoría matemática que permite analizar las propiedades de los fenómenos en los que interviene el azar. Para ello utiliza como modelo básico, común a cualquier situación aleatoria, el concepto de espacio probabilístico, (Ω, \mathcal{A}, P) , junto con el concepto de variable aleatoria definido sobre él.

1.1.1. Conceptos básicos del Cálculo de Probabilidad

- *Espacio de probabilidad o probabilístico:* (Ω, \mathcal{A}, P)
 - Ω : conjunto arbitrario.
 - \mathcal{A} : clase de subconjuntos de Ω , con estructura de σ -álgebra,
 1. $\Omega \in \mathcal{A}$,
 2. $A \in \Omega/A \in \mathcal{A} \implies \bar{A} \in \mathcal{A}$,
 3. Sean $A_n \subset \Omega/A_n \in \mathcal{A}, \forall n \implies \bigcup_{i=1}^{\infty} A_n \in \mathcal{A}$ y $\bigcap_{i=1}^{\infty} A_n \in \mathcal{A}$.
 - $P: \mathcal{A} \longrightarrow [0, 1]$ función de probabilidad,
 1. $\forall A \in \mathcal{A}, P(A) \geq 0$,

2. $P(\Omega) = 1$,

3. Sean $A_n \subset \mathcal{A}$ incompatibles dos a dos, entonces $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{i=1}^{\infty} P(A_n)$.

■ *Variable aleatoria*

$X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathbb{B}, P_X)$, función medible definida sobre un espacio de probabilidad. La función de probabilidad P_X es lo que se denomina distribución de probabilidad de X .

■ *Función de distribución de X : $F_X: \mathbb{R} \rightarrow [0, 1]$*

$$F_X(x) = P_X((-\infty, x]) = P[X^{-1}((-\infty, x])] = P[X \leq x]$$

Propiedades:

1. No decreciente
2. Continua a la derecha
3. $F_X(-\infty) = \lim_{x \rightarrow -\infty} F_X(x) = 0$ y $F_X(+\infty) = \lim_{x \rightarrow +\infty} F_X(x) = 1$

Teorema de correspondencia: Toda función $F: \mathbb{R} \rightarrow [0, 1]$ cumpliendo las tres propiedades anteriores es la función de distribución de una variable aleatoria.

■ *Clasificación de variables aleatorias*

- *Discretas:* toman valores en un conjunto finito o infinito numerable.

La función de distribución es creciente a saltos.

Función masa de probabilidad: $p_X(x) = P[X = x]$

- *Continuas:* toman valores en un conjunto infinito no numerable.

La función de distribución es creciente y continua.

Función de densidad: $f_X = F'_X(x)$

■ *Independencia de variables aleatorias*

X_1, \dots, X_n , v.a. definidas sobre (Ω, \mathcal{A}, P) , son independientes si y sólo si

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \cdot \dots \cdot F_{X_n}(x_n),$$

$$\forall x_1, \dots, x_n \in \mathbb{R}$$

El disponer de un conjunto de observaciones acerca del fenómeno considerado, en lugar del espacio probabilístico, hace abandonar el Cálculo de Probabilidades, para introducirse en el terreno de la Estadística Matemática o Inferencia Estadística, cuya finalidad es obtener información acerca de la ley de probabilidad de un fenómeno, a partir de una observación no exhaustiva del mismo.

1.1.2. Definiciones

Existen multitud de definiciones distintas para el concepto *Estadística*, como por ejemplo: “proceso de recolección de datos u observaciones, así como su tratamiento numérico; etapas que comportan considerable complicaciones técnicas”.

Otra opción es la definición que da Vic Barnett (1982): “la Estadística es la ciencia que nos indica como debe ser utilizada la información, en orden a reflejar y dar una guía de acción en situaciones prácticas que envuelven incertidumbre”.

Por último, según la RAE: “la Estadística es el estudio científico que se encarga de la recopilación, clasificación y análisis de datos numéricos sobre fenómenos específicos, buscando obtener conclusiones a partir de ellos, frecuentemente mediante el uso de la probabilidad”.

Por lo tanto, la Estadística no sólo debe encargarse de recoger la información (datos u observaciones), sino que además debe obtener conclusiones basándose en ella. A la parte de la Estadística que se encarga de esa última tarea se la denomina *Inferencia Estadística*.

Dicho de otra forma, el objetivo de la Inferencia Estadística es el análisis e interpretación de la información que aportan las observaciones, como método para obtener conclusiones sobre la ley de probabilidad asociada al fenómeno en estudio.

La Inferencia Estadística se subdivide en dos áreas de estudio denominadas Inferencia Clásica o Frecuentista e Inferencia Bayesiana. La principal diferencia conceptual entre ambas es su interpretación de lo que significa una probabilidad. Por otro lado, la Inferencia Bayesiana, en contraste a la Clásica, le da la posibilidad al investigador de incorporar la información previa que puede poseer sobre el fenómeno.

A lo largo de esta asignatura se estudiará la Inferencia Clásica con idea de que el alumno adquiera el conocimiento de los fundamentos básicos de esta materia a un nivel elemental, aunque con una perspectiva amplia.

1.2. Planteamiento de un problema de inferencia

Bajo el punto de vista clásico, un problema de Inferencia Estadística consiste en obtener conclusiones acerca del comportamiento de una o varias características de una determinada población, basándose en la observación de las mismas en un subconjunto de la población.

En todo problema de inferencia estadística existen una serie de elementos base del problema:

- La población bajo estudio (*Población*): Conjunto de elementos en el que se pretende estudiar una determinada característica. (Ω)
- La característica que se desea estudiar (*Característica*): Se suele representar por la v.a. que la cuantifica. (X)

- La muestra de que se dispone para el estudio (*Muestra*): Subconjunto de la población sobre el que se va a estudiar la característica para inferir las conclusiones sobre la misma a la población total.

1.2.1. Modelización estadística

De una manera genérica la distribución desconocida F de la variable aleatoria involucrada en un problema de Inferencia Estadística, recibe el nombre de *distribución teórica* o *distribución de la población*. Según el conocimiento que se tenga sobre la distribución teórica, se pueden planear dos situaciones:

- La forma de la función de distribución teórica es conocida, salvo el valor de uno o varios parámetros, es decir, se sabe que F pertenece a una familia de funciones de distribuciones

$$F \in \{F_\theta, \theta \in \Theta\}$$

compuesta por distribuciones de forma funcional fija y conocida, dependientes de uno o varios parámetro θ , que varían dentro de un subconjunto Θ de \mathbb{R}^k , denominado *espacio paramétrico*.

- No se conoce nada acerca de la función de distribución teórica, salvo cosas de tipo muy general, como por ejemplo que la v.a. es discreta o continua, o que existen o no momentos, pero se desconoce la forma de F .

En el primer caso, se estaría considerando un problema de *Inferencia estadística paramétrica* ya que se asume un modelo paramétrico.

En el segundo caso, se estaría considerando un modelo no paramétrico, y por tanto, sería un problema de *Inferencia estadística no paramétrica*.

Dentro del estudio de la Inferencia Clásica, se estudiará principalmente la Inferencia paramétrica, es decir, se hará inferencia sobre el/los parámetros θ , ya que es lo único que falta por conocer para determinar la distribución teórica (Temas del 1 al 8). En el tema 9 se estudiará una introducción a la Inferencia Clásica no paramétrica.

1.3. Muestra aleatoria simple

Como ya se ha comentado anteriormente el objetivo de la inferencia estadística es inferir sobre una población el estudio de una característica a partir del mismo realizado sobre una muestra o subcolección de la población. Para poder realizar correctamente dicha extensión a toda la población, la muestra considerada para el estudio debe ser aleatoria y representativa de toda la población.

El procedimiento de selección de la muestra puede conducir a diferentes tipos de muestras, pero en esta asignatura sólo se va a considerar el muestreo con reemplazamiento.

Éste tipo de muestreo consiste en seleccionar, por mecanismos aleatorios, los elementos de la población que entran a formar parte de la muestra, pero de tal forma que cuando se observa la característica que se está investigando del elemento seleccionado, éste se devuelve a la población y se selecciona el siguiente elemento de entre todos los elementos de la población. Este procedimiento permite que un elemento de la población puede ser seleccionado en más de una ocasión para formar parte de una muestra.

Para llegar al concepto de *muestra aleatoria simple* y poder dar una definición rigurosa de la misma se va a considerar una población, con función de distribución teórica F de la variable aleatoria X .

Para seleccionar una muestra aleatoria de esta población se van llevando a cabo repeticiones del experimento aleatorio que da lugar a la variable aleatoria X , siempre bajo las mismas condiciones, y anotando en cada una de ellas los valores de X . Se obtendrá así un conjunto de valores numéricos x_1, x_2, \dots, x_n que constituyen lo que se denomina una muestra de \mathcal{X} , el conjunto de posibles valores de la v.a. X . El número n de repeticiones efectuadas, y de observaciones obtenidas, se denomina *tamaño de la muestra*.

Si se consideran las n observaciones de X como (x_1, x_2, \dots, x_n) , es decir un vector de dimensión n , se pueden ver como la observación del vector aleatorio (X_1, X_2, \dots, X_n) compuesto por n v.a. independientes e idénticamente distribuidas a la v.a. X .

Al conjunto de los posibles valores del vector (x_1, x_2, \dots, x_n) se le denomina *espacio muestral* o conjunto de posibles valores de la muestra y se le denota por \mathcal{X}^n .

Una **muestra aleatoria simple**, de tamaño n , de una variable aleatoria X con distribución teórica F , es un vector (X_1, \dots, X_n) formado por n variables aleatorias independientes e idénticamente distribuidas, con distribución común F .

Al ser las v.a. X_1, \dots, X_n independientes, se tiene que la función de distribución conjunta del vector aleatorio formado por dichas variables será igual al producto de las distribuciones marginales de cada una de ellas. Además, al ser todas las variables de la muestra idénticamente distribuidas a X se tiene que la *función de distribución conjunta* de (X_1, \dots, X_n) es

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \cdot \dots \cdot F_{X_n}(x_n) = \prod_{i=1}^n F_X(x_i), \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Si X es una v.a. discreta, entonces la *función masa de probabilidad conjunta* de (X_1, \dots, X_n) es

$$P[X_1 = x_1, \dots, X_n = x_n] = \prod_{i=1}^n P[X = x_i], \quad (x_1, \dots, x_n) \in \mathcal{X}^n.$$

Si X es una v.a. continua, entonces la *función de densidad conjunta* de (X_1, \dots, X_n) es

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n f_X(x_i), \quad (x_1, \dots, x_n) \in \mathcal{X}^n.$$

Ejemplo: Obtener la función masa de probabilidad conjunta de una m.a.s. de $X \rightsquigarrow B(k_0, p)$ y la función de densidad conjunta de una m.a.s. de $X \rightsquigarrow U(a, b)$.

1.4. Función de distribución muestral

Se puede asociar a cada muestra aleatoria simple, (X_1, \dots, X_n) , de una v.a. X con función de distribución F , una *distribución muestral* F_{X_1, \dots, X_n}^* , que emule a F a partir únicamente de la información contenida en la muestra.

Puesto que $F(x) = P[X \leq x]$, para cada $x \in \mathbb{R}$, la distribución muestral se define como

$$F_{X_1, \dots, X_n}^*(x) = \frac{\text{n}^\circ \text{ de variables } X_i \leq x}{n}, \quad \forall x \in \mathbb{R}.$$

La función de distribución muestral se puede escribir también como

$$F_{X_1, \dots, X_n}^*(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i)$$

donde sumando $I_{(-\infty, x]}(X_i)$ es la siguiente v.a.

$$I_{(-\infty, x]}(X_i) = \begin{cases} 1 & \text{si } X_i \leq x \\ 0 & \text{si } X_i > x. \end{cases}$$

Propiedades:

- Para cada realización muestral, $(x_1, \dots, x_n) \in \mathcal{X}^n$, F_{x_1, \dots, x_n}^* es una función de distribución en \mathbb{R} . En particular es una función a saltos, con saltos de amplitud $1/n$ en los sucesivos valores muestrales ordenados de menor a mayor, supuestos que sean distintos, y de saltos múltiples en el caso de que varios valores muestrales coincidieran.
- $\forall x \in \mathbb{R}$, $F_{X_1, \dots, X_n}^*(x)$ es una variable aleatoria tal que $nF_{X_1, \dots, X_n}^*(x) \rightsquigarrow B(n, F(x))$:

$$E[F_{X_1, \dots, X_n}^*(x)] = F(x), \quad \text{Var}[F_{X_1, \dots, X_n}^*(x)] = \frac{F(x)(1 - F(x))}{n}.$$

- Para valores grandes de n , en virtud del Teorema Central del Límite:

$$F_{X_1, \dots, X_n}^*(x) \rightsquigarrow \mathcal{N}\left(F(x), \frac{F(x)(1 - F(x))}{n}\right)$$

Ejemplo: Dada una muestra aleatoria formada por las observaciones (3, 8, 5, 4, 5), obtener su función de distribución muestral y realizar la representación gráfica.

La distribución muestral asociada a una realización, $F_{x_1, \dots, x_n}^*(x)$ refleja exclusivamente los valores numéricos contenidos en la muestra concreta a la que está asociada y no tiene relación directa con la distribución de la población F , en el sentido de que una muestra determinada tendría la misma distribución muestral fuese cual fuese la población de origen. Pese a ello, es muy razonable esperar que la distribución muestral proporcione una imagen aproximada de la distribución de la población de la que se haya extraído la muestra. Es más, la función de distribución muestral verifica el siguiente teorema:

Teorema de Glivenko-Cantelli: Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de v.a. independientes e idénticamente distribuidas (i.i.d.) con función de distribución común F . Si F_{X_1, \dots, X_n}^* es la función muestral asociada a la m.a.s. (X_1, \dots, X_n) , se verifica que F_{X_1, \dots, X_n}^* converge casi seguramente y uniformemente a la función de distribución de X , F .

$$P\left\{\lim_{n \rightarrow +\infty} \sup_{x \in \mathbb{R}} |F_{X_1, \dots, X_n}^*(x) - F(x)| = 0\right\} = 1.$$

Con probabilidad 1, al tomar sucesivas observaciones independientes de la variable y considerar las correspondientes funciones de distribución muestrales:

$$\forall \epsilon > 0 \exists n_\epsilon / n > n_\epsilon \Rightarrow F(x) \in (F_{X_1, \dots, X_n}^*(x) - \epsilon, F_{X_1, \dots, X_n}^*(x) + \epsilon)$$

Lo cual, a efectos prácticos, implica que cuando el tamaño de la muestra crece la gráfica de la función de distribución empírica se aproxima bastante a la de la función de distribución de la población.

1.5. Estadístico muestral

Anteriormente se ha indicado que para conocer más acerca de la función de distribución de la variable X que se este estudiando, se toma una m.a.s. de la misma, (X_1, \dots, X_n) y se observan los valores que toma la muestra. Puede resultar, en ocasiones, más cómodo anotar, en vez de todo el vector de la muestra, una función de dichos resultados, como por ejemplo la suma de los mismo, $\sum_{i=1}^n X_i$. A dicha función de la m.a.s. es lo que se va a denominar *estadístico muestral* cuya ventaja es que en vez de trabajar con la muestra,

que es un vector aleatorio de dimensión n , se trabaja con una variable aleatoria o con un vector de menor dimensión. Más formalmente su definición sería:

Sea (X_1, \dots, X_n) una m.a.s. de una v.a. X . Sea $T : (\mathbb{R}^n, \mathbb{B}^n) \longrightarrow (\mathbb{R}^k, \mathbb{B}^k)$ una función medible ($T^{-1}(B) \in \mathcal{B}^n$, $\forall B \in \mathcal{B}^k$) e independiente de cualquier parámetro desconocido. A $T(X_1, \dots, X_n)$ se le denomina **estadístico muestral** asociado a la v.a. X .

Como T es una función medible, el estadístico muestral $T(X_1, \dots, X_n)$ es una v.a. de dimensión k . La función T basta con que este definida en el espacio muestral \mathcal{X}^n y no en todo \mathbb{R}^n .

Ejemplos:

1. Media muestral: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

2. Varianza muestral: $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

3. Cuasivarianza muestral: $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

4. Estadísticos ordenados:

$$X_{(1)} = \min(X_1, \dots, X_n)$$

$$X_{(n)} = \max(X_1, \dots, X_n)$$

El resto de los estadísticos ordenados, $X_{(2)}, \dots, X_{(n-1)}$ son los estadísticos muestrales que se quedan con el valor que ocupa la posición que indica su subíndice después de ordenar de menor a mayor todos los valores, es decir, por ejemplo $X_{(2)}$ dará como resultado el segundo valor más pequeño de X_1, \dots, X_n .

Ejemplo: Sea X una v.a. con distribución $B(1, p)$ con $p \in (0, 1)$. Se toma una muestra de tamaño 5, $(X_1, X_2, X_3, X_4, X_5)$, y se obtiene la siguiente observación $(0, 1, 1, 0, 0)$. Determinar el valor de los estadísticos estudiados en la observación.

1.5.1. Distribución en el muestreo de un estadístico

Ya que el estadístico muestral $T(X_1, \dots, X_n)$ es una v.a. de dimensión k , tiene una función de distribución asociada, la cual no siempre será fácil de conocer.

Se denomina *distribución en el muestreo de un estadístico* T , definido en el espacio muestral $(\mathcal{X}^n, \mathcal{B}^n)$, a la distribución de la v.a. $T(X_1, \dots, X_n)$.

Media muestral

En general, determinar la distribución del estadístico no será fácil, pero en casos en los que el estadístico sea suma de las variables, como el estadístico media muestral, puede resultar asequible. Para ello se aplica que la función generatriz de momentos, que define de forma única a una v.a., verifica la propiedad de adición, es decir, que la f.g.m. de la suma de n v.a. independientes coincide con el producto de las n f.g.m. Por lo tanto, para el estadístico media muestral se verifica:

$$M_{\bar{X}}(t) = (M_X(t/n))^n$$

teniendo en cuenta que las n v.a. de la muestra son independientes e idénticamente distribuidas a X . Identificando la distribución que tiene dicha f.g.m. se tiene la distribución del estadístico.

Ejemplo: Obtener la distribución muestral de \bar{X} para (X_1, \dots, X_n) una m.a.s. de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.

Estadísticos ordenados

En el caso de los estadísticos ordenados,

$$X_{(r)} = \begin{cases} \min(X_1, \dots, X_n) & r = 1; \\ \min(\{X_1, \dots, X_n\} - \{X_{(1)}, \dots, X_{(j-1)}\}), & r = j; \\ \max(X_1, \dots, X_n), & r = n \end{cases}$$

se tiene que la función de distribución muestral es:

$$\begin{aligned} F_{X_{(r)}}(x) &= P[X_{(r)} \leq x] = P[\text{al menos } r \text{ elementos muestrales sean } \leq x] = \\ &= \sum_{i=r}^n \binom{n}{i} (P[X \leq x])^i (P[X > x])^{n-i}, \quad x \in \mathbb{R}. \end{aligned}$$

Por lo tanto, según la v.a. X sea discreta o continua, se tiene que:

■ Discreta:

$$P[X_{(r)} = x_{(r)}] = P[X_{(r)} \leq x_{(r)}] - P[X_{(r)} < x_{(r)}] = F_{X_{(r)}}(x_{(r)}) - F_{X_{(r)}}^-(x_{(r)}).$$

■ Continua:

$$g_r(x_{(r)}) = \frac{n!}{(r-1)!(n-r)!} [F(x_{(r)})]^{r-1} [1 - F(x_{(r)})]^{n-r} f(x_{(r)}).$$

Ejemplo: Obtener las distribuciones muestrales de $X_{(1)}$ y $X_{(n)}$ para $X \rightsquigarrow U(a, b)$.

1.5.2. Estadísticos de Interés

Sea (X_1, \dots, X_n) una m.a.s. de X , v.a. con función de distribución F . Anteriormente se ha estudiado que asociada a la muestra existe una distribución denominada muestral o empírica, $F_n^*(x)$. Además de dicha distribución, sobre la muestra se pueden estudiar una serie de características:

- Momentos muestrales centrados y no centrados:

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \equiv \text{momento no centrado de orden } k \in \mathbb{N}$$

$$B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \equiv \text{momento centrado de orden } k \in \mathbb{N}$$

En particular se tiene:

$$A_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \text{ media muestral}$$

$$B_1 = 0$$

$$B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ (Varianza muestral).}$$

- Cuantiles muestrales: Para cada $p \in (0, 1)$, el cuantil de orden p , c_p , es un valor real tal que

$$F_n^*(c_p) \geq p \text{ y } F_n^*(c_p^-) \leq p$$

Se pueden expresar de la siguiente forma en función de los elementos de la muestra ordenada:

- Si $np \in \mathbb{N}$, $c_p = \frac{X_{(np)} + X_{(np+1)}}{2}$.

- En otro caso, sea $[np]$ la parte entera de np , entonces $c_p = X_{([np]+1)}$

- Función generatriz de momentos muestral:

$$M^*(t) = \frac{1}{n} \sum_{i=1}^n e^{tX_i}$$

Esta función se usa para obtener los momentos no centrados, $\left[\frac{\partial^k M^*(t)}{\partial t^k} \right]_{t=0} = A_k$.

Para los momentos no centrados se tiene

- $E[A_k] = \frac{1}{n} \sum_{i=1}^n E[X_i^k] = E[X^k]$, es decir, coincide con el momento de orden k respecto al origen de la distribución teórica. En particular,

$$E[A_1] = \mu$$

- $Var[A_k] = \frac{1}{n^2} \sum_{i=1}^n Var[X_i^k] = \frac{1}{n} Var[X^k] = \frac{1}{n} (E[X^{2k}] - E[X^k]^2)$. En particular,

$$Var[A_1] = \frac{\sigma^2}{n}.$$

Se ha denotado por $\mu = E[X]$ y por $\sigma^2 = Var[X]$.

Para los momentos centrados tiene especial interés el caso de la varianza muestral, $B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2$:

$$E[B_2] = \frac{(n-1)\sigma^2}{n}.$$

Por tanto, la esperanza de la varianza muestral no coincide con la varianza teórica. Debido a este resultado es frecuente considerar la cuasivarianza muestral, S^2 , en lugar de la varianza, ya que ella si cumple que su esperanza coincide con la varianza teórica:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} B_2 \Rightarrow$$

$$E[S^2] = E\left[\frac{n}{n-1} B_2\right] = \frac{n}{n-1} E[B_2] = \frac{n}{n-1} \frac{(n-1)\sigma^2}{n} = \sigma^2.$$

Tema 2

Distribuciones en el muestreo de poblaciones normales.

2.1. Introducción

En el tema 1 se obtuvieron algunos resultados relativos al comportamiento de algunas características muestrales como los momentos. En particular, se estudiaron:

- Resultados “exactos”, acerca de momentos de la distribución en el muestreo de algunas características muestrales.
- Resultados “de aproximación” asintóticos, sobre el comportamiento límite (para “muestras grandes”) de ciertas características.

Exceptuando los ejemplos específicos analizados, todas las consideraciones se hicieron sin presuponer ninguna distribución concreta en la población.

En este tema, se estudian algunos resultados fundamentales de la Inferencia Estadística relativos al muestreo de poblaciones normales, que se refieren a la distribución exacta (para muestras de cualquier tamaño) de estadísticos que surgen de forma natural en problemas concretos básicos de inferencia.

En este sentido, se analizan la distribución χ^2 de Pearson, t de Student y F de Snedecor, desde una doble aproximación:

1. Desde el punto de vista analítico, por particularización o construcción basada en distribuciones conocidas (normal y gamma).
2. Desde el punto de vista muestral, como distribuciones exactas relacionadas con el muestreo de poblaciones normales.

2.2. Distribuciones χ^2 de Pearson, t de Student y F de Snedecor

2.2.1. Distribución χ^2 de Pearson

Se dice que la v.a. X tiene una distribución χ^2 con n ($n \in \mathbb{N}$) grados de libertad si su función de densidad de probabilidad viene dada por

$$f(x) = \frac{1}{\Gamma(n/2) 2^{n/2}} x^{n/2-1} e^{-x/2}, \quad x > 0,$$

y se denota $X \rightsquigarrow \chi^2(n)$ ¹.

La distribución χ^2 es un caso particular de la distribución Gamma, $\Gamma(p, a)$, cuya densidad es:

$$f(x) = \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax}, \quad x > 0$$

En particular se verifica: $X \rightsquigarrow \chi^2(n) \Leftrightarrow X \rightsquigarrow \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$.

Propiedades:

- FGM: $M(t) = \frac{1}{(1-2t)^{n/2}}, \quad t < 1/2$
- Momentos: $E[X^k] = 2^k \frac{\Gamma\left(\frac{n}{2} + k\right)}{\Gamma\left(\frac{n}{2}\right)}$. En particular, se tiene: $E[X] = n$, $E[X^2] = n^2 + 2n$, $Var[X] = 2n$.
- Reproductividad: Si X_1, \dots, X_n son v.a. independientes tales que $X_i \rightsquigarrow \chi^2(k_i)$. Entonces

$$\sum_{i=1}^n X_i \rightsquigarrow \chi^2\left(\sum_{i=1}^n k_i\right).$$

- Relación con la distribución normal $\mathcal{N}(0, 1)$: Si X_1, \dots, X_n son v.a.i.i.d. con distribución común $\mathcal{N}(0, 1)$. Entonces

$$\sum_{i=1}^n X_i^2 \rightsquigarrow \chi^2(n)$$

Tablas y aproximaciones: La distribución χ^2 está tabulada para valores de n pequeños. Para n grandes, su distribución se puede aproximar por $\mathcal{N}(n, 2n)$.

Ejemplo: Calcula el valor de k o la probabilidad indicada:

¹La función Γ se define como $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$. Esta función cumple un par de propiedades: $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$; si α es un número entero, $\Gamma(\alpha) = (\alpha-1)!$.

- a) $P[\chi^2(10) \geq k] = 0.005$
- b) $P[\chi^2(45) \leq k] = 0.005$
- c) $P[\chi^2(14) \geq 21.06]$
- d) $P[\chi^2(20) \leq 12.44]$

Gráfica de la función de densidad de $\chi^2(n)$: Cumple las siguientes propiedades

- Es asimétrica a la derecha y unimodal.
- Para $n = 1$, $\lim_{x \rightarrow 0} f(x) = +\infty$ y la función de densidad es estrictamente decreciente.
- Para $n = 2$, $f(0) = 1/2$ y la función de densidad es estrictamente decreciente.
- Para $n \geq 3$, $f(0) = 0$, crece hasta la moda y luego decrece.

2.2.2. Distribución t de Student

Sean X e Y variables aleatorias independientes con distribuciones, $X \rightsquigarrow \mathcal{N}(0, 1)$ e $Y \rightsquigarrow \chi^2(n)$. Entonces, la v.a.

$$T = \frac{X}{\sqrt{Y/n}}$$

se dice que tiene una distribución t de Student con n grados de libertad. Se denota $T \rightsquigarrow t(n)$.

Propiedades:

- Función de densidad:

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad t \in \mathbb{R}$$

Para obtener esta densidad se aplica el cambio de variable indicado en la propia construcción de la distribución, $T = X/\sqrt{Y/n}$, tomando como variable auxiliar $U = Y$.

- Momentos: Sea X una v.a. con distribución $t(n)$, con $n > 1$. Entonces se tiene que existen los momentos $E[X^r]$ para $r < n$ y se verifica que,
 - Si $r < n$ e impar, se tiene que $E[X^r] = 0$.

$$\text{- Si } r < n \text{ y par, } E[X^r] = n^{r/2} \frac{\Gamma\left(\frac{r+1}{2}\right) \Gamma\left(\frac{n-r}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{n}{2}\right)}.$$

En particular, para $n > 2$, existen los momentos de primer y segundo orden,

$$E[X] = 0 \text{ y } E[X^2] = Var[x] = \frac{n}{n-2}.$$

Tablas y aproximaciones: La distribución t está tabulada para valores de n pequeños. Para n grandes, su distribución se puede aproximar por $\mathcal{N}(0, 1)$.

Ejemplo: Calcula el valor de k o la probabilidad indicada:

a) $P[t(26) \geq k] = 0.05$

b) $P[t(20) \leq k] = 0.25$

c) $P[t(26) \geq k] = 0.9$

d) $P[t(21) \geq 1.721]$

e) $P[t(11) \leq 0.697]$

f) $P[t(8) \leq -2.306]$

Gráfica de la función de densidad de $t(n)$: Cumple las siguientes propiedades

- Es similar a la de la $\mathcal{N}(0, 1)$, es decir, simétrica alrededor del cero y unimodal.
- Para $n \rightarrow +\infty$ se aproxima a la gráfica de la $\mathcal{N}(0, 1)$.
- Tiene colas mayores que las de la normal que van reduciéndose y aproximándose a las de la normal conforme n crece.
- Es más aplastada que la de la normal, es decir es platicúrtica, y su zona central va creciendo y aproximándose a la de la normal conforme n crece.

2.2.3. Distribución F de Snedecor

Sean X e Y variables aleatorias independientes con distribuciones $X \rightsquigarrow \chi^2(m)$ e $Y \rightsquigarrow \chi^2(n)$. Entonces, la v.a.

$$F = \frac{X/m}{Y/n}$$

se dice que tiene una distribución F de Snedecor con (m, n) grados de libertad. Se denota $F \rightsquigarrow F(m, n)$.

Propiedades:

- Función de densidad:

$$g(f) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} f^{m/2-1} \left(1 + \frac{m}{n}f\right)^{-\frac{m+n}{2}}, \quad f > 0.$$

Para obtener esta densidad se aplica el cambio de variable indicado en la propia construcción de la distribución, $F = \frac{X/m}{Y/n}$, tomando como variable auxiliar $U = Y$.

- Momentos: Sea X una v.a. con distribución $F(m, n)$. Entonces se verifica que,

$$E[X^r] = \left(\frac{n}{m}\right)^r \frac{\Gamma\left(\frac{m}{2} + r\right)\Gamma\left(\frac{n}{2} - r\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}, \quad \text{para } 0 < r < \frac{n}{2}.$$

En particular, $E[X] = \frac{n}{n-2}$ si $n > 2$, $E[X^2] = \frac{n^2(m+2)}{m(n-4)(n-2)}$ si $n > 4$ y

$$Var[X] = \frac{n^2(2m+2n-4)}{m(n-2)^2(n-4)} \text{ si } n > 4.$$

- Mediante un cambio de variable se puede comprobar que:

$$- X \rightsquigarrow F(m, n) \Leftrightarrow X^{-1} \rightsquigarrow F(n, m).$$

$$- X \rightsquigarrow t(n) \Leftrightarrow X^2 \rightsquigarrow F(1, n).$$

Tablas y aproximaciones: La distribución F está tabulada para valores de m y n pequeños. Las tablas que se os han proporcionado incluyen aproximaciones para valores grandes de m y n .

Ejemplo: Calcula el valor de k o la probabilidad indicada:

- $P[F(7, 3) \leq k] = 0.95$
- $P[F(8, 4) \geq k] = 0.01$
- $P[F(2, 2) \leq 19]$
- $P[F(3, 5) \geq 12.1]$
- $P[F(60, 40) \leq k] = 0.05$

Gráfica de la función de densidad de $F(m, n)$: Cumple las siguientes propiedades

- Es asimétrica a la derecha y unimodal.
- Para $m = 1$, $\lim_{f \rightarrow 0} g(x) = +\infty$ y la función de densidad es estrictamente decreciente.
- Para $m = 2$, $g(0) = 1$ y la función de densidad es estrictamente decreciente.
- Para $m \geq 3$, $g(0) = 0$, crece hasta la moda y luego decrece.

2.3. Muestreo en una población normal unidimensional

Sea X una v.a. con distribución $\mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ y $\sigma^2 > 0$. Se considera una m.a.s. de tamaño n , X_1, \dots, X_n . Sea $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ la media muestral y $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ la cuasivarianza muestral. El objetivo de este apartado es dar las distribuciones de ambos estadísticos muestrales, para lo cual es necesario el siguiente resultado:

Teorema: Sea X_1, \dots, X_n una m.a.s. de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. Entonces, \bar{X} y $(X_1 - \bar{X}, \dots, X_n - \bar{X})$ son independientes.

Corolarios:

1. $\bar{X} \rightsquigarrow \mathcal{N}(\mu, \frac{\sigma^2}{n})$
2. (Lema de Fisher): \bar{X} y S^2 son independientes.
3. $\frac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi^2(n-1)$
4. $\frac{\bar{X} - \mu}{S/\sqrt{n}} \rightsquigarrow t(n-1)$

2.3.1. Esquemas de resultados para una muestra y uso en inferencia

$$\begin{aligned}
 \text{- Inferencia sobre } \mu : & \begin{cases} \sigma_0^2 \text{ conocida} & \frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} \rightsquigarrow \mathcal{N}(0, 1) \\ \sigma^2 \text{ desconocida} & \frac{\bar{X} - \mu}{S/\sqrt{n}} \rightsquigarrow t(n-1) \end{cases} \\
 \text{- Inferencia sobre } \sigma^2 : & \begin{cases} \mu_0 \text{ conocida} & \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2} \rightsquigarrow \chi^2(n) \\ \mu \text{ desconocida} & \frac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi^2(n-1) \end{cases}
 \end{aligned}$$

2.4. Muestreo en dos poblaciones normales unidimensionales

Sea X una v.a. con distribución $\mathcal{N}(\mu_1, \sigma_1^2)$, $\mu_1 \in \mathbb{R}$ y $\sigma_1^2 > 0$. Se considera una m.a.s. de tamaño n_1 , X_1, \dots, X_{n_1} . Sea $\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$ la media muestral y $S_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$

$\bar{X})^2$ la cuasivarianza muestral.

Sea Y otra v.a. con distribución $\mathcal{N}(\mu_2, \sigma_2^2)$, $\mu_2 \in \mathbb{R}$ y $\sigma_2^2 > 0$. Se considera una m.a.s. de tamaño n_2 , Y_1, \dots, Y_{n_2} . Sea $\bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$ la media muestral y $S_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$ la cuasivarianza muestral.

Se supone que (X_1, \dots, X_{n_1}) e (Y_1, \dots, Y_{n_2}) son independientes.

Extensión del Lema de Fisher: Los vectores (\bar{X}, \bar{Y}) y (S_1^2, S_2^2) son independientes.

Corolarios:

$$1. \frac{n_2 \sigma_2^2 \sum_{i=1}^{n_1} (X_i - \mu_1)^2}{n_1 \sigma_1^2 \sum_{i=1}^{n_2} (Y_i - \mu_2)^2} \rightsquigarrow F(n_1, n_2)$$

$$2. \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \rightsquigarrow F(n_1 - 1, n_2 - 1)$$

En particular, si $\sigma_1 = \sigma_2$, se tiene que $\frac{S_1^2}{S_2^2} \rightsquigarrow F(n_1 - 1, n_2 - 1)$

$$3. \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2}{\sigma_1^2} + \frac{(n_2-1)S_2^2}{\sigma_2^2}}} \sqrt{\frac{n_1 + n_2 - 2}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow t(n_1 + n_2 - 2)$$

En particular, si $\sigma_1 = \sigma_2$, se tiene que $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2}} \sqrt{\frac{n_1 + n_2 - 2}{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow t(n_1 + n_2 - 2)$

2.4.1. Esquemas de resultados para dos muestra y uso en inferencia

- Inferencia sobre $\mu_1 - \mu_2$ (comparación de medias):

$$\left\{ \begin{array}{ll} \sigma_1^2, \sigma_2^2 \text{ conocidas} & \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0, 1) \\ \sigma_1^2 = \sigma_2^2 = \sigma^2 \text{ desconocidas} & \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2}} \sqrt{\frac{n_1 + n_2 - 2}{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow t(n_1 + n_2 - 2) \end{array} \right.$$

- Inferencia sobre $\frac{\sigma_1^2}{\sigma_2^2}$ (comparación de varianzas):

$$\left\{ \begin{array}{ll} \mu_1, \mu_2 \text{ conocidas} & \frac{n_1 \sigma_1^2 \sum_{i=1}^{n_2} (Y_i - \mu_2)^2}{n_2 \sigma_2^2 \sum_{i=1}^{n_1} (X_i - \mu_1)^2} \rightsquigarrow F(n_2, n_1) \\ \mu_1, \mu_2 \text{ desconocidas} & \frac{\sigma_1^2 S_2^2}{\sigma_2^2 S_1^2} \rightsquigarrow F(n_2 - 1, n_1 - 1) \end{array} \right.$$

Tema 3

Suficiencia y completitud

3.1. Estadísticos suficientes

Antes de dar una definición del concepto de estadístico suficiente, se va a dar una idea intuitiva del concepto de suficiencia que facilite su comprensión.

Sea X una v.a. con distribución en una familia de distribuciones $\{F_\theta, \theta \in \Theta\}$ dependientes de un parámetro desconocido θ . El objetivo es inferir el valor de θ , para lo cual se considera una m.a.s. para, en base a la información que proporciona, estimar el valor del parámetro.

Sin embargo, en lugar de considerar la muestra, se va a trabajar con un estadístico de la misma, $T(X_1, \dots, X_n)$, que resumen la información que proporciona la muestra. Puede suceder que al resumir la información de la muestra se pierda parte relevante de la misma. Por ejemplo, si $T(X_1, \dots, X_n) = X_1$ se pierde la información que proporcionan X_2, \dots, X_n . En ocasiones dicha pérdida puede no ser relevante, según la distribución de X .

En ese sentido de no perder información relevante surge el concepto de suficiencia gracias a Fisher (1922): “Un estadístico es suficiente cuando contiene toda la información contenida en la muestra sobre el parámetro que se está considerando” (es decir, basta con usar el estadístico para inferir el valor del parámetro).

Un estadístico puede ser suficiente en una familia de distribuciones y no para otras.

Definición: Sea (X_1, \dots, X_n) una m.a.s. de $X \rightsquigarrow F \in \{F_\theta, \theta \in \Theta\}$. Un estadístico $T(X_1, \dots, X_n)$ es *suficiente* para la familia de distribuciones considerada (o suficiente para θ) si la distribución de la muestra condicionada a cualquier valor del estadístico, $T(X_1, \dots, X_n) = t$, es independiente de θ .

Notas:

- Si $T(X_1, \dots, X_n) = (X_1, \dots, X_n)$ no hay pérdida de información. Luego siempre hay un estadístico suficiente trivial, la propia muestra.

- Si se encuentra un estadístico suficiente que no sea la muestra, a partir de entonces se trabajará con él, porque es más fácil de manejar ya que resumen la información de la muestra sin perder información sobre θ .

Ejemplos:

1. Sea X_1, X_2 variables aleatorias independientes con distribución de Poisson de parámetro λ . Probar que $X_1 + X_2$ es suficiente para λ y que $X_1 + 2X_2$ no lo es.
2. En n lanzamientos de una moneda, el número de caras es suficiente para el parámetro p .

En ocasiones puede ocurrir que se disponga de una muestra y se pierda el valor de sus datos. A partir de un estadístico suficiente se puede reconstruir la muestra. Intuitivamente la idea de *reconstrucción de la muestra* es la siguiente:

Sea X una v.a. con distribución en $\{F_\theta, \theta \in \Theta\}$ y (X_1, \dots, X_n) una m.a.s. de X . Se realiza un experimento que consiste en observar la muestra y se obtienen unos valores concretos u observaciones (x_1, \dots, x_n) . Asociada a la variable X se tiene un estadístico $T(X_1, \dots, X_n)$, que al aplicarlo a los datos observados se obtiene un valor, $T(x_1, \dots, x_n) = t$.

Si se pierden los datos de la m.a.s., pero se conoce el valor del estadístico y no se puede volver a observar la variable para obtener datos equivalentes a los anteriores, para reconstruir la muestra se realiza lo siguiente: Se considera una v.a. X^* que va a tener la distribución de la v.a. $X/T = t$, o más concretamente, una m.a.s (X_1^*, \dots, X_n^*) con la misma distribución que $(X_1, \dots, X_n)/T = t$.

Si la distribución de $(X_1, \dots, X_n)/T = t$ es independiente del parámetro θ , por tanto es conocida, se pueden observar las variables (X_1^*, \dots, X_n^*) , que es una muestra equivalente a (X_1, \dots, X_n) , porque las dos muestras conducen al mismo valor de estadístico:

$$T(X_1^*, \dots, X_n^*) = T(X_1, \dots, X_n)$$

y para cada valor de t , las distribuciones son las mismas.

Ejemplo: Se supone que se ha lanzado una moneda 100 veces, y se sabe que se han obtenido 60 caras pero se han perdido los datos originales de si salió cara o cruz en cada tirada. Reconstruir la muestra.

3.1.1. Teorema de factorización de Neyman-Fisher

La definición de estadístico suficiente no es constructiva, en el sentido de que sirve para comprobar si un estadístico dado es o no suficiente, pero no indica cómo buscarlo.

El teorema de Factorización de Neyman-Fisher establece un criterio útil para la búsqueda de estadísticos suficientes, así como para probar, con mayor facilidad, si un estadístico es suficiente.

Teorema de factorización:

Sea (X_1, \dots, X_n) una m.a.s de $X \rightsquigarrow F \in \{F_\theta, \theta \in \Theta\}$. Sea f_θ la función masa de probabilidad o función de densidad de X bajo F_θ y sea f_θ^n la f.m.p. o f.d.d. de la muestra bajo F_θ .

Un estadístico $T(X_1, \dots, X_n)$ se dice que es suficiente si y sólo si, para cualquier valor de $\theta \in \Theta$,

$$f_\theta^n(x_1, \dots, x_n) = h(x_1, \dots, x_n) g_\theta(T(x_1, \dots, x_n)), \quad (x_1, \dots, x_n) \in \mathcal{X}^n$$

donde h es independiente de θ y g_θ depende de (x_1, \dots, x_n) sólo a través de $T(x_1, \dots, x_n)$.

Propiedades

1. Si $T(X_1, \dots, X_n)$ es suficiente para $\{F_\theta, \theta \in \Theta\}$, entonces $T(X_1, \dots, X_n)$ es suficiente para $\{F_\theta, \theta \in \Theta'\}$, para cualquier $\Theta' \subseteq \Theta$, es decir, si un estadístico es suficiente para una familia de distribuciones, lo es también para cualquier subfamilia suya.
2. Si $T(X_1, \dots, X_n)$ es suficiente para una familia $\{F_\theta, \theta \in \Theta\}$ y $U(X_1, \dots, X_n)$ es un estadístico tal que $T = h(U)$, entonces U es suficiente para la misma familia. (Esto no indica que una función de un estadístico suficiente sea suficiente, sino que si tengo un estadístico suficiente que es función de otro estadístico, entonces el otro estadístico también es suficiente).
3. Si $T(X_1, \dots, X_n)$ es suficiente para $\{F_\theta, \theta \in \Theta\}$, toda transformación biunívoca de T proporciona también un estadístico suficiente para la misma familia.

El teorema de factorización también se verifica si el parámetro θ es multidimensional, en cuyo caso el estadístico también es multidimensional. En particular, aunque el parámetro sea de dimensión 1, el estadístico suficiente puede tener dimensión mayor que 1. En general se verifica:

$$\text{Dimensión estadístico suficiente} \geq \text{Dimensión parámetro}$$

Ejemplo: Sea X una v.a. con distribución en $\{\mathcal{N}(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma^2 > 0\}$. Estudiar la suficiencia en los casos de :

- σ_0^2 conocida.
- μ_0 conocida.
- μ y σ^2 desconocidas.

Suficiencia minimal

Dada una familia de distribuciones $\{F_\theta, \theta \in \Theta\}$, generalmente existirán varios estadísticos suficientes para θ .

Un criterio de selección consiste en elegir estadísticos suficientes minimales (máxima reducción de los datos, sin pérdida de información sobre θ).

Va a existir siempre un estadístico suficiente que es función de todos los demás, ese estadístico es el *estadístico suficiente minimal*.

Una aplicación adecuada del teorema de factorización nos llevará a él (es único) y siempre existe en distribuciones discretas y continuas.

3.2. Familias de distribuciones completas. Estadísticos completos

El concepto de suficiencia se usa frecuentemente en conjunción con el concepto de completitud, sin embargo dicho concepto es menos intuitivo que el de suficiencia. En próximos temas se verá su utilidad cuando se estudien las propiedades de los denominados “estimadores”.

Antes de dar la definición de estadístico completo, se va a definir el concepto de familia de distribuciones completa:

Definición: Sea $\{F_\theta, \theta \in \Theta\}$ una familia de distribuciones con f.d.d. o f.m.p. $\{f_\theta(x), \theta \in \Theta\}$. Se dice que dicha familia es *completa* si para cualquier función medible unidimensional, g , tal que

$$E_\theta[g(X)] = 0, \quad \forall \theta \in \Theta$$

se tiene que

$$P_\theta[g(X) = 0] = 1, \quad \forall \theta \in \Theta$$

es decir, una familia completa es aquella que no admite funciones medibles no nulas con esperanza cero.

Definición: Un estadístico $T(X_1, \dots, X_n)$ se dice que es *completo* para la familia de distribuciones de X si para cualquier función medible unidimensional, g , se tiene:

$$E_\theta[g(T(X_1, \dots, X_n))] = 0, \quad \forall \theta \in \Theta \Rightarrow P_\theta[g(T(X_1, \dots, X_n)) = 0] = 1, \quad \forall \theta \in \Theta.$$

Como ya se ha comentado, el concepto de estadístico completo suele ir asociado al concepto de estadístico suficiente. Es más, si se pide dar un estadístico suficiente y completo, se suele tomar el estadístico suficiente que proporciona el teorema de factorización y se comprueba si dicho estadístico es completo. Además se tiene el siguiente resultado:

Ejemplos:

1. Sea (X_1, \dots, X_n) una m.a.s de la v.a. $X \rightsquigarrow \{B(1, p), p \in (0, 1)\}$. Encontrar un estadístico suficiente y completo asociado a la muestra.
2. Sea (X_1, \dots, X_n) una m.a.s de la v.a. $X \rightsquigarrow \{U(0, \theta), \theta > 0\}$. Encontrar un estadístico suficiente y completo asociado a la muestra.

3.3. Suficiencia y completitud en familias exponenciales

Sea $\{F_\theta, \theta \in \Theta\}$ una familia de distribuciones de probabilidad paramétricas con f.d.d. o f.m.p. $\{f_\theta(x), \theta \in \Theta\}$.

3.3.1. Familia exponencial uniparamétrica

Se dice que la familia de distribuciones es *exponencial uniparamétrica* si se cumple que:

1. El espacio paramétrico, Θ , es un intervalo real ($\Theta \subseteq \mathbb{R}$).
2. El conjunto de valores de la variable no depende de θ :

$$\{x, f_\theta(x) > 0\} = \chi, \quad \forall \theta \in \Theta.$$

3. Existen funciones real-valuadas $Q(\theta)$ y $D(\theta)$, definidas sobre Θ , y existen funciones medibles Borel T y S , también real valuadas, tales que

$$\forall \theta \in \Theta, \quad f_\theta(x) = \exp [Q(\theta)T(x) + D(\theta) + S(x)], \quad x \in \chi.$$

Ejemplos:

1. La $\{\mathcal{P}(\lambda), \lambda > 0\}$ es una familia exponencial uniparamétrica.
2. La $\{B(k_0, p), p \in (0, 1)\}$ es una familia exponencial uniparamétrica.

3.3.2. Familia exponencial k -paramétrica

Se dice que la familia de distribuciones es *exponencial k -paramétrica* si se cumple que:

1. El espacio paramétrico, Θ , es un intervalo de \mathbb{R}^k , ($\Theta \subseteq \mathbb{R}^k$).

3.3 Suficiencia y completitud en familias exponenciales

2. El conjunto de valores de la variable no depende de θ :

$$\{x, f_\theta(x) > 0\} = \chi, \quad \forall \theta \in \Theta.$$

3. Existen funciones real-valuadas $Q_1(\theta), \dots, Q_k(\theta)$ y $D(\theta)$, definidas sobre Θ , y existen funciones medibles Borel T_1, \dots, T_k y S , también real valuadas, tales que

$$\forall \theta \in \Theta, \quad f_\theta(x) = \exp \left[\sum_{h=1}^k Q_h(\theta) T_h(x) + D(\theta) + S(x) \right], \quad x \in \chi.$$

Ejemplo: La $\{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$ es una familia exponencial uniparamétrica en cada parámetro y bi-paramétrica en los dos parámetros.

Teorema: Si se tiene una v.a. $X \rightsquigarrow \{F_\theta, \theta \in \Theta\}$ con familia de funciones asociadas, $\{f_\theta(x), \theta \in \Theta\}$, siendo f.d.d. o f.m.p. según el caso, donde la familia de distribuciones es exponencial k -paramétrica, entonces la familia de distribuciones asociadas a (X_1, \dots, X_n) (que es una m.a.s. de X) es también exponencial k -paramétrica:

$$f_\theta^n(x_1, \dots, x_n) = \exp \left\{ \sum_{h=1}^k Q_h(\theta) \left(\sum_{i=1}^n T_h(x_i) \right) + \sum_{i=1}^n S(x_i) + nD(\theta) \right\}, \quad (x_1, \dots, x_n) \in \chi^n,$$

y se tiene:

- El estadístico $(\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i))$ es suficiente para θ .
- Si $k \leq n$ y el conjunto imagen de la función $Q(\theta) = (Q_1(\theta), \dots, Q_k(\theta))$ contiene a un abierto de \mathbb{R}^k , el estadístico $(\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i))$ es también completo.

Ejemplo: Sea X una v.a. con distribución $\mathcal{N}(\mu, \sigma^2)$ con $\mu \in \mathbb{R}$ y $\sigma^2 > 0$. Buscar un estadístico suficiente y completo basado en una muestra de tamaño n .

Tema 4

ESTIMACIÓN PUNTUAL. INSESGADEZ Y MÍNIMA VARIANZA.

4.1. Planteamiento del problema

En el tema anterior se trataron aspectos relacionados con la reducción de los datos de una muestra en términos de un estadístico. En este tema se aborda el problema de estimación del parámetro de un modelo estadístico paramétrico (o de una función de dicho parámetro) mediante un estadístico conveniente.

En general se tiene una v.a. X con distribución en una familia de distribuciones paramétricas (es decir conocidas salvo por un parámetro), $X \rightsquigarrow F \in \{F_\theta, \theta \in \Theta\}$. El objetivo es determinar la distribución de la v.a. que se estudia, que en el caso paramétrico se reduce a conocer el parámetro, es decir, inferir el verdadero valor de θ , θ_0 .

Para ello lo que se hace es coger una muestra, X_1, \dots, X_n , m.a.s. de X , y en base a la información que proporciona la muestra se aproxima el valor de θ . Por lo tanto, el problema es escoger estadísticos, $T(X_1, \dots, X_n)$, que para valores concretos de la muestra proporcionen buenas aproximaciones del parámetro θ .

Con dicho fin se van a escoger unos estadísticos particulares, que se denominan *estimadores*, de forma que cuando se sustituya la muestra aleatoria, X_1, \dots, X_n , por sus observaciones, $T(x_1, \dots, x_n)$ proporcionen una buena aproximación del parámetro desconocido.

En ocasiones en vez de estimar el parámetro θ , interesará estimar una *función paramétrica*, es decir, una transformación del parámetro, $g : \Theta \rightarrow \mathbb{R}$. En ese caso se buscará $g(\theta_0)$, el verdadero valor de $g(\theta)$, en vez de θ_0 , el verdadero valor de θ .

Ejemplo: Sea $X \rightsquigarrow B(k_0, p)$ y $p \in (0, 1)$. A partir de una m.a.s. de X indicar algún estadístico que se pueda usar para inferir p y alguna función paramétrica que pueda ser

de interés.

A la aproximación que se obtiene del parámetro a través de un estimador se la denomina estimación puntual porque lo que se obtiene al aplicarla es un valor concreto para el parámetro desconocido.

El parámetro desconocido puede ser unidimensional o multidimensional. En general se van a estudiar casos unidimensionales y, si es fácil, se generalizará al caso multidimensional.

4.1.1. Estimador

Sea X una v.a. con función de distribución en una familia de distribuciones paramétricas, $F \in \{F_\theta : \theta \in \Theta\}$ y (X_1, \dots, X_n) una m.a.s. de X .

Definición: Un estimador de θ es un estadístico, $T(X_1, \dots, X_n)$, que toma valores en Θ .

$$T : \mathcal{X}^n \rightarrow \Theta.$$

Por tanto la diferencia entre un estimador y un estadístico es que el estimador es un estadístico que en lugar de estar definido sobre \mathbb{R}^k , se le exige que tome valores en el espacio paramétrico Θ , es decir, donde toma los valores el parámetro desconocido.

Si el estimador está definido con espacio de llegada $g(\Theta)$, es decir, $T : \mathcal{X}^n \rightarrow g(\Theta)$, T es un estimador de la función paramétrica $g(\theta)$.

Para valores concretos de la muestra, x_1, \dots, x_n , $T(x_1, \dots, x_n)$ es una estimación puntual de θ o de $g(\theta)$, según el caso.

Ejemplos:

1. Sea X_1, \dots, X_n una m.a.s. de $X \rightsquigarrow \{P(\lambda), \lambda > 0\}$. \bar{X} es un estimador de λ . Es más, cualquier función medible de la muestra que sea independiente del parámetro λ y tome valores positivos es un estimador del parámetro.
2. Sea X_1, \dots, X_n una m.a.s. de $X \rightsquigarrow \{B(1, p), p \in (0, 1)\}$. \bar{X} es un estimador de p . Es más, cualquier función medible de la muestra que sea independiente del parámetro p y tome valores en el intervalo $(0, 1)$ es un estimador del parámetro.

La no unicidad del estimador de un parámetro plantea el problema de encontrar el mejor estimador. Para ello hay que establecer criterios de selección entre los estimadores para encontrar el mejor en algún sentido. Una opción es seleccionar el estimador en base a una función denominada de pérdida.

4.1.2. Función de pérdida y función de riesgo

Definición: A cualquier función $L : \Theta \times \Theta \rightarrow \mathbb{R}$, que verifique las siguientes propiedades:

- (i) $L(\theta, T) \geq 0, \forall \theta \in \Theta, T \in \Theta$.
- (ii) $L(\theta, T) = 0$, si $T = \theta$.
- (iii) $L(\theta, T) \leq L(\theta, T')$ si la distancia de T a θ es menor que la distancia de T' a θ .

se la denomina *función de pérdida*.

$L(\theta, t)$ sería la pérdida que conlleva estimar el parámetro por el valor t si su verdadero valor es θ usando la función de pérdida L .

Ejemplos:

1. $L(\theta, T) = |\theta - T|$ (error absoluto de estimación).
2. $L(\theta, T) = (\theta - T)^2$ (error cuadrático de estimación).
3. $L(\theta, T) = \left| \frac{\theta - T}{\theta} \right|$ (error relativo de estimación).

Dado un estimador $T(X_1, \dots, X_n)$ de θ , la función $L(\theta, T(X_1, \dots, X_n))$, para cada $\theta \in \Theta$, es una variable aleatoria, siempre que L sea Borel-medible.

Definición: Se define la pérdida media o *función riesgo* de un estimador como la función (del parámetro θ) que asigna a cada valor del parámetro la pérdida media asociada al estimador bajo la función de pérdida L .

$$R_T^L(\theta) = E_\theta[L(\theta, T)].$$

En particular, la función riesgo de una función paramétrica $g(\theta)$ se definiría como:

$$R_{g,T}^L(\theta) = E_\theta[L(g(\theta), T)].$$

Definición: Se dice que un estimador $T(X_1, \dots, X_n)$ es *óptimo bajo una función de pérdida* $L(\theta, T)$ si dicho estimador minimiza uniformemente la función de riesgo $R_T(\theta)$; esto es, un estimador, T , tal que, para cualquier otro, T' se tiene

$$R_T^L(\theta) \leq R_{T'}^L(\theta), \quad \forall \theta \in \Theta.$$

En general, el estimador óptimo no tiene porqué existir. Al no tener asegurada la existencia del estimador óptimo basado en la función de riesgo, el problema de estimación se puede reconsiderar mediante una de las dos siguientes vías:

1. Restringir la clase de estimadores imponiendo propiedades deseables de los mismos.
2. Introducir una nueva medida de la función de riesgo que permita ordenar totalmente la clase de todos los estimadores.

Estimación de menor error cuadrático medio (ECM)

Un criterio de comparación usual en múltiples ámbitos y aplicaciones de la estadística es el llamado criterio de menor *error cuadrático medio*. Se considera como función de pérdida:

$$L(\theta, T) = (\theta - T)^2$$

y como función de riesgo:

$$R_T(\theta) = E_\theta [(\theta - T)^2] = ECM_T(\theta).$$

Propiedades:

- El criterio del ECM tiene ventajas desde el punto de vista del manejo analítico, frente a otras funciones de riesgo.
- El ECM se interpreta como el grado de dispersión del estimador en torno al verdadero valor del parámetro, θ .
- El ECM puede descomponer en términos de la varianza y una función denominada sesgo:

$$ECM_T(\theta) = Var_\theta(T) + B_T^2(\theta)$$

donde $B_T(\theta)$ es la función denominada sesgo que se define como:

$$B_T(\theta) = E_\theta[T] - \theta.$$

- Si el estimador considerado verifica la propiedad de insesgadez, es decir, $E_\theta[T] = \theta$ o $B_T(\theta) = 0$, se verifica que el ECM coincide con la varianza del estimador.

$$ECM_T(\theta) = E_\theta [(\theta - T)^2] = Var_\theta(T).$$

4.2. Estimación insesgada de mínima varianza

Como se ha visto en la sección anterior, existe una relación sencilla e intuitiva que liga al ECM, la varianza y el sesgo de un estimador. En el problema de la búsqueda de un estimador óptimo, en algún sentido, se explota dicha relación dentro de la clase de estimadores que verifican ciertas propiedades, como la propiedad de insesgadez.

4.2.1. Estimador inssegado

Sea, como siempre a lo largo de este tema, X una v.a. con función de distribución en una familia de distribuciones paramétricas, $F \in \{F_\theta : \theta \in \Theta\}$ y (X_1, \dots, X_n) una m.a.s. de X .

Definición: Un estimador $T(X_1, \dots, X_n)$ de θ es *inssegado* o centrado en el parámetro θ si su sesgo asociado es idénticamente nulo o, equivalentemente, si:

$$E_\theta[T(X_1, \dots, X_n)] = \theta, \quad \forall \theta \in \Theta.$$

Si el estimador $T^*(X_1, \dots, X_n)$ es de una función paramétrica de θ , $g(\theta)$, se dice que es inssegado en $g(\theta)$ si:

$$E_\theta[T^*(X_1, \dots, X_n)] = g(\theta), \quad \forall \theta \in \Theta.$$

Ejemplo: Sea X_1, \dots, X_n una m.a.s. de alguna población. Probar que si existe la media de la población, $E_\theta X$, la media muestral es un estimador inssegado de la media poblacional, y si existe la varianza de la población, $Var_\theta X$, la cuasivarianza muestral es un estimador inssegado de ella.

Notas:

- Si $\theta = (\theta_1, \dots, \theta_k)$, un estimador inssegado de dicho vector de parámetros es un vector donde cada elemento del mismo es un estimador inssegado para cada parámetro θ_i componente de θ , es decir: Un estimador $T = (T_1, \dots, T_k)$ es inssegado en θ si se verifica

$$E_\theta[T_i] = \theta_i, \quad \forall i = 1, \dots, k.$$

- Para un estimador inssegado se verifica que

$$ECM_T(\theta) = Var_\theta(T(X_1, \dots, X_n)), \quad \forall \theta \in \Theta$$

- Si T es un estimador inssegado de $\theta \Rightarrow h(T)$ es estimador inssegado de $h(\theta)$, siendo h cualquier función.

Ejemplo: Sea X_1, \dots, X_n una m.a.s. de alguna población. Probar que, en general, la cuasidesviación típica no es un estimador inssegado para $\sigma_\theta = \sqrt{Var_\theta X}$.

Si h es una función lineal dicha implicación si se cumple, es decir, la inssegadez no se mantiene bajo transformaciones, en general, pero si se mantiene si la transformación es lineal.

- No tiene porque existir algún estimador inssegado de un parámetro.

Ejemplos:

1. Sea X_1, \dots, X_n , con $n \geq 2$, una m.a.s. de una distribución binomial $B(1, p)$. Probar que $\frac{T^2 - T}{n^2 - n}$ con $T = \sum_{i=1}^n X_i$ es un estimador insesgado para la función paramétrica $g(p) = p^2$. Probar, además, que para $n = 1$ no existe un estimador insesgado de p^2 .
 2. Sea X una v.a. con distribución $\mathcal{P}(\lambda)$, ¿existe algún estimador insesgado para la función paramétrica $1/\lambda$ basado en una muestra de tamaño 1?
- Un estimador insesgado no tiene porque ser único.

Ejemplo: Sea X una v.a. con distribución $\mathcal{P}(\lambda)$ y X_1, \dots, X_n una m.a.s. de X . Probar que \bar{X} y S^2 son estimadores insesgados de λ .

Es más, si T_1 y T_2 son estimadores insesgados de θ , entonces $\alpha T_1 + (1 - \alpha)T_2$ con $0 \leq \alpha \leq 1$ es un estimador insesgado de θ .

Se plantea, por tanto, el problema de seleccionar un estimador óptimo dentro de la clase de estimadores insesgados. Nuestro criterio de búsqueda de los mejores estimadores será seleccionar los que tengan mínima varianza, ya que son los que están menos dispersos respecto a la media.

4.2.2. Estimador insesgado uniformemente de mínima varianza (UMVUE)

Definición: Sea X una v.a. con función de distribución en una familia de distribuciones paramétricas, $F \in \{F_\theta : \theta \in \Theta\}$ y (X_1, \dots, X_n) una m.a.s. de X . Un estimador de $g(\theta)$, $T(X_1, \dots, X_n)$, insesgado y con momento de segundo orden finito, se dice que es un *estimador insesgado uniformemente de mínima varianza* (“UMVUE”) para $g(\theta)$ si para cualquier otro estimador de $g(\theta)$, $T'(X_1, \dots, X_n)$, se tiene:

$$Var_\theta[T(X_1, \dots, X_n)] \leq Var_\theta[T'(X_1, \dots, X_n)] \quad \forall \theta \in \Theta.$$

Teorema (Unicidad del UMVUE): El UMVUE, si existe, es único, es decir, si hay dos UMVUEs son iguales con probabilidad 1.

Teorema (Linealidad del UMVUE): Si T_1 es UMVUE para una cierta función de θ , $g_1(\theta)$ y T_2 es UMVUE para otra cierta función de θ , $g_2(\theta)$ con $\theta \in \Theta$. Entonces:

1. λT_1 es UMVUE para $\lambda g_1(\theta)$ y λT_2 es UMVUE para $\lambda g_2(\theta)$,
2. $T_1 + T_2$ es UMVUE para $g_1(\theta) + g_2(\theta)$,

siendo λ cualquier valor real.

Una vez definido el UMVUE y estudiada algunas de sus propiedades, el siguiente paso es tener un método que permita su obtención de forma lo más sencilla posible. Para ello primero se estudian los Teoremas de Rao-Blackwell y Lehmann-Scheffé.

Teorema (Rao-Blackwell): Sea X una v.a. con función de distribución en una familia de distribuciones paramétricas, $F \in \{F_\theta : \theta \in \Theta\}$, y (X_1, \dots, X_n) una m.a.s. de X . Si $T(X_1, \dots, X_n)$ un estadístico suficiente para la familia, $\{F_\theta : \theta \in \Theta\}$ y $S(X_1, \dots, X_n)$ es un estimador insesgado de $g(\theta)$ con momento de segundo orden finito, entonces:

1. $E[S(X_1, \dots, X_n)/T(X_1, \dots, X_n)]$ es estimador insesgado de $g(\theta)$ y tiene momento de segundo orden finito.
2. $Var_\theta(E[S(X_1, \dots, X_n)/T(X_1, \dots, X_n)]) \leq Var_\theta(S(X_1, \dots, X_n)), \forall \theta \in \Theta$.

Teorema (Lehmann-Scheffé): Sea $T(X_1, \dots, X_n)$ es un estadístico suficiente y completo para la familia de distribuciones consideradas, $\{F_\theta : \theta \in \Theta\}$. Si $g(\theta)$ admite un estimador insesgado de segundo orden, $S(X_1, \dots, X_n)$, entonces existe el UMVUE de $g(\theta)$, que viene dado por

$$E[S(X_1, \dots, X_n)/T(X_1, \dots, X_n)].$$

Métodos para el cálculo del UMVUE: Sea $T(X_1, \dots, X_n)$ un estadístico suficiente y completo para $\{F_\theta : \theta \in \Theta\}$. El UMVUE para $g(\theta)$, si existe, se puede determinar mediante los dos siguientes procedimientos:

- Buscar cualquier estimador insesgado de $g(\theta)$ con momento de segundo orden finito, $S(X_1, \dots, X_n)$. Entonces $E[S(X_1, \dots, X_n)/T(X_1, \dots, X_n)]$ es el UMVUE.
- Buscar una función $h(T)$, siendo T un estadístico suficiente y completo, tal que $E_\theta[h(T)] = g(\theta) \forall \theta \in \Theta$, es decir, que sea insesgada en $g(\theta)$, que sea un estimador y que tenga momento de segundo orden finito. Entonces $E[h(T(X_1, \dots, X_n))/T(X_1, \dots, X_n)] = h(T(X_1, \dots, X_n))$ es el UMVUE.

Si no existiera un estadístico suficiente y completo, esto no implica que no exista UMVUE, sólo que habrá que calcularlo de otra forma.

Ejemplos

1. Sea X_1, \dots, X_n una m.a.s. con distribución $B(1, p)$, $p \in (0, 1)$. Encontrar el UMVUE para p .
2. Sea X_1, \dots, X_n una m.a.s. con distribución $\mathcal{U}(0, \theta)$, $\theta \in \mathbb{R}^+$. Encontrar el UMVUE para θ y $1/\theta$.
3. Sea X_1, \dots, X_n una m.a.s. con distribución $P(\lambda)$, $\lambda \in \mathbb{R}^+$. Encontrar, si existe, el UMVUE para la función paramétrica λ^s , $s \in \mathbb{N}$.

4.3. Estimadores eficientes

En esta sección del tema se va a estudiar el concepto de estimador eficiente con idea de, posteriormente, estudiar la relación que existe entre dichos estimadores y el UMVUE. En toda esta sección consideraremos un espacio paramétrico unidimensional para poder trabajar con el concepto de varianza en lugar del de matriz de varianzas-covarianzas.

Para dar la definición de estimador eficiente, previamente, se van a estudiar condiciones de regularidad.

4.3.1. Condiciones de regularidad de Fréchet-Crámer-Rao

Sea X una v.a. con distribución de probabilidad en la familia de distribuciones $\{F_\theta : \theta \in \Theta\}$. Sea $f_\theta(x)$ la fdp o fmp, según el caso, para cada $\theta \in \Theta \subseteq \mathbb{R}$. Se dice que esta familia de distribuciones cumple las *condiciones de regularidad* si:

- (i) Θ es un intervalo abierto de \mathbb{R} .
- (ii) El conjunto de valores de la variable, $\{x : f_\theta(x) > 0\} = \mathcal{X}$, es independiente de θ .
- (iii) $\forall x \in \mathcal{X}$, $f_\theta(x)$ es derivable respecto de θ y se verifica que

$$\int_{\mathcal{X}} \frac{\partial}{\partial \theta} f_\theta(x) dx = \frac{\partial}{\partial \theta} \int_{\mathcal{X}} f_\theta(x) dx = 0 \quad \left(\sum_{\mathcal{X}} \frac{\partial}{\partial \theta} f_\theta(x) = \frac{\partial}{\partial \theta} \sum_{\mathcal{X}} f_\theta(x) = 0 \right). \quad \forall \theta \in \Theta.$$

Esta tercera condición es equivalente a comprobar que

$$E_\theta \left[\frac{\partial \ln f_\theta(x)}{\partial \theta} \right] = 0.$$

Teorema: Si $X \rightsquigarrow F \in \{F_\theta : \theta \in \Theta\}$ cumple las condiciones de regularidad, entonces la familia de distribuciones asociada a la m.a.s. (X_1, \dots, X_n) de X también las cumple.

4.3.2. Función de información de Fisher

Definición: Sea $X \rightsquigarrow F \in \{F_\theta : \theta \in \Theta\}$, cuya familia de distribuciones es regular. Se definen las funciones

$$I_X(\theta) = E_\theta \left[\left(\frac{\partial \ln f_\theta(X)}{\partial \theta} \right)^2 \right], \quad I_{(X_1, \dots, X_n)}(\theta) = E_\theta \left[\left(\frac{\partial \ln f_\theta^n(X_1, \dots, X_n)}{\partial \theta} \right)^2 \right]$$

que se denominan *función de información de Fischer* asociada a X y a la muestra, respectivamente.

Propiedades:

- (i) $I_X \geq 0$.
- (ii) $E_\theta \left[\frac{\partial \ln f_\theta(X)}{\partial \theta} \right] = 0$ y $Var_\theta \left[\frac{\partial \ln f_\theta(X)}{\partial \theta} \right] = I_X(\theta)$, $\forall \theta \in \Theta$.
- (iii) $E_\theta \left[\frac{\partial \ln f_\theta^n(X_1, \dots, X_n)}{\partial \theta} \right] = 0$ y $Var_\theta \left[\frac{\partial \ln f_\theta^n(X_1, \dots, X_n)}{\partial \theta} \right] = I_{(X_1, \dots, X_n)}(\theta)$, $\forall \theta \in \Theta$.
- (iv) $I_{X_1, \dots, X_n}(\theta) = nI_X(\theta)$, $\forall \theta \in \Theta$. (Aditividad)

4.3.3. Desigualdad de Fréchet-Cramér-Rao

Definición: Sea (X_1, \dots, X_n) una m.a.s. de $X \rightsquigarrow F \in \{F_\theta : \theta \in \Theta\}$, cuya familia de distribuciones es regular. Un estadístico $T(X_1, \dots, X_n)$ se dice que es *regular* en el sentido de Fréchet-Cramér-Rao si verifica:

- Caso discreto

$$\frac{\partial}{\partial \theta} \sum_{(x_1, \dots, x_n) \in \mathcal{X}^n} T(x_1, \dots, x_n) f_\theta^n(x_1, \dots, x_n) = \sum_{(x_1, \dots, x_n) \in \mathcal{X}^n} T(x_1, \dots, x_n) \frac{\partial f_\theta^n(x_1, \dots, x_n)}{\partial \theta}$$

- Caso continuo

$$\frac{\partial}{\partial \theta} \int_{\mathcal{X}^n} T(x_1, \dots, x_n) f_\theta^n(x_1, \dots, x_n) dx_1 \cdots dx_n = \int_{\mathcal{X}^n} T(x_1, \dots, x_n) \frac{\partial f_\theta^n(x_1, \dots, x_n)}{\partial \theta} dx_1 \cdots dx_n$$

Es decir,

$$\frac{\partial}{\partial \theta} E_\theta [T(X_1, \dots, X_n)] = E_\theta \left[T(X_1, \dots, X_n) \frac{\partial \ln f_\theta^n(X_1, \dots, X_n)}{\partial \theta} \right]$$

Teorema (cota de Fréchet-Cramér-Rao): Sea (X_1, \dots, X_n) una m.a.s. de $X \rightsquigarrow F \in \{F_\theta : \theta \in \Theta\}$, cuya familia de distribuciones es regular con $0 < I_X(\theta) < +\infty \forall \theta \in \Theta$. Si $T(X_1, \dots, X_n)$ es un estadístico regular, de segundo orden e insesgado en una función paramétrica derivable $g(\theta)$, entonces se tiene

- (i) $Var_\theta T(X_1, \dots, X_n) \geq \frac{(g'(\theta))^2}{I_{X_1, \dots, X_n}(\theta)} \quad \forall \theta \in \Theta$.
- (ii) Para los puntos θ_0 tales que $g'(\theta_0) \neq 0$, se dará la igualdad si y sólo si $\exists a(\theta_0) \neq 0$ tal que

$$P_{\theta_0} \left[\left. \frac{\partial \ln f_\theta(X_1, \dots, X_n)}{\partial \theta} \right|_{\theta=\theta_0} = a(\theta_0) [T(X_1, \dots, X_n) - g(\theta_0)] \right] = 1$$

4.3.4. Estimador eficiente

Definición: Sea $\{F_\theta : \theta \in \Theta\}$ regular, $0 < I_X(\theta) < +\infty \forall \theta \in \Theta$ y $g(\theta)$ una función paramétrica derivable. Un estimador de $g(\theta)$, $T(X_1, \dots, X_n)$, se dice que es eficiente si es regular, insesgado y su varianza alcanza la cota de FCR para cualquier valor del parámetro, es decir,

$$\text{Var}_\theta(T(X_1, \dots, X_n)) = \frac{[g'(\theta)]^2}{I_{X_1, \dots, X_n}(\theta)}, \quad \forall \theta \in \Theta.$$

El estimador eficiente no tiene porque existir.

Lema: Sea $\{F_\theta : \theta \in \Theta\}$ regular, $0 < I_X(\theta) < +\infty \forall \theta \in \Theta$ y $g(\theta)$ una función paramétrica derivable. Entonces $g(\theta)$ admite un estimador eficiente $T(X_1, \dots, X_n)$ si:

- $g(\theta)$ es constante y en tal caso $T(X_1, \dots, X_n)$ es degenerado.
- $g(\theta)$ es estrictamente monótona: $g'(\theta) > 0 \forall \theta \in \Theta$ o $g'(\theta) < 0 \forall \theta \in \Theta$.

Teorema (caracterización de estimadores eficientes): Sea (X_1, \dots, X_n) una m.a.s. de $X \rightsquigarrow F \in \{F_\theta : \theta \in \Theta\}$ regular, con $0 < I_X(\theta) < \infty \forall \theta \in \Theta$, $g(\theta)$ una función paramétrica derivable, con $g'(\theta) \neq 0, \forall \theta \in \Theta$ y $T(X_1, \dots, X_n)$ es un estimador de $g(\theta)$. Una condición necesaria y suficiente para que T sea eficiente es:

$$\forall \theta \in \Theta, \exists a(\theta) \neq 0 \text{ tal que } \begin{cases} i) P_\theta \left[\frac{\partial \ln f_\theta^n(X_1, \dots, X_n)}{\partial \theta} = a(\theta)[T(X_1, \dots, X_n) - g(\theta)] \right] = 1 \\ ii) I_{(X_1, \dots, X_n)}(\theta) = a(\theta)g'(\theta). \end{cases}$$

Corolario 1: Si $T(X_1, \dots, X_n)$ es un estimador eficiente para $g(\theta)$, con $g'(\theta) \neq 0$, las únicas funciones paramétricas que admiten estimadores eficientes son las de la forma $ag(\theta) + b$ y los correspondientes estimadores eficientes son $aT + b$, con probabilidad 1, bajo todas las distribuciones de la familia.

Corolario 2: Si una función paramétrica admite dos estimadores eficientes, estos son iguales con probabilidad 1, bajo todas las distribuciones de la familia.

Corolario 3: Sólo existen estimadores eficientes en familias de tipo exponencial.

Corolario 4: Si $T(X_1, \dots, X_n)$ es eficiente para $g(\theta)$, entonces $T(X_1, \dots, X_n)$ es suficiente. Si además la imagen de $Q(\theta) = \int a(\theta) d\theta$ contiene a un abierto de \mathbb{R} , entonces $T(X_1, \dots, X_n)$ es completo.

Corolario 5: Si $T(X_1, \dots, X_n)$ es eficiente para $g(\theta)$, suficiente y completo, entonces $T(X_1, \dots, X_n)$ es el UMVUE de $g(\theta)$.

El recíproco no es cierto, es decir, si $T(X_1, \dots, X_n)$ es el UMVUE para $g(\theta)$ eso no implica que $T(X_1, \dots, X_n)$ sea eficiente para dicha función paramétrica.

Ejemplo: Buscar la clase de funciones paramétricas que admiten estimadores eficientes para las siguientes familias de distribuciones y calcular dichos estimadores:

1. $\{B(k_0, p), p \in (0, 1)\}$,
2. $\{\mathcal{N}(\mu, \sigma_0^2), \mu \in \mathbb{R}\}$,
3. $\{\mathcal{N}(\mu_0, \sigma^2), \sigma^2 > 0\}$.

Tema 5

ESTIMACIÓN DE MÁXIMA VEROSIMILITUD Y OTROS MÉTODOS.

5.1. Estimación de máxima verosimilitud

El método de obtención de estimadores de máxima verosimilitud es el método más usado debido, en parte, a las buenas propiedades asintóticas que tienen los estimadores obtenidos con él. No se trata de un método que proporciona un criterio de selección, sino que es un método de cálculo.

Para entender la idea principal de este método se va a estudiar, primeramente, un ejemplo:

Ejemplo: Sea X una v.a. con distribución $B(n, p)$, donde $n \in \{2, 3\}$ y $p \in \{1/2, 1/3\}$. Basándose en la observación de un valor de la variable, decidir cuál de estos valores de los parámetros corresponden a la variable bajo estudio.

Antes de indicar como se aplica el método de máxima verosimilitud, se introducen una serie de conceptos:

Definición: Sea X una v.a. con distribución en una familia paramétrica de distribuciones, $\{F_\theta, \theta \in \Theta\}$. Sea $f_\theta(x)$ la f.m.p. (caso discreto) ó la f.d.d. (caso continuo) de X . Se considera X_1, \dots, X_n una m.a.s. de X y sea $f_\theta^n(x_1, \dots, x_n)$ su f.m.p. ó f.d.d. conjunta con $\theta \in \Theta$. Para cada x_1, \dots, x_n , realización muestral, se define la *función de verosimilitud* asociada a dichos valores de la muestra como una función de θ de la siguiente forma:

$$\begin{aligned} L_{x_1, \dots, x_n} : \Theta &\rightarrow \mathbb{R}^+ \cup \{0\} \\ \theta &\rightarrow L_{x_1, \dots, x_n}(\theta) = f_\theta^n(x_1, \dots, x_n). \end{aligned}$$

Definición: Sea X una v.a. con distribución en una familia paramétrica de distribuciones, $\{F_\theta, \theta \in \Theta\}$. Sea X_1, \dots, X_n una m.a.s. de X . Un estimador $\hat{\theta}(X_1, \dots, X_n)$ de θ es *estimador de máxima verosimilitud* (EMV) de θ si:

$$\forall (x_1, \dots, x_n) \in \mathcal{X}^n, \quad L_{x_1, \dots, x_n}(\hat{\theta}(x_1, \dots, x_n)) = \max_{\theta \in \Theta} L_{x_1, \dots, x_n}(\theta).$$

Por tanto, el método de obtención de estimadores de máxima verosimilitud consiste en obtener un estimador que maximice la función de verosimilitud. Para determinarlo se deben tener en cuenta las propiedades analíticas de dicha función:

- Si la función de verosimilitud es derivable, el estimador máximo verosímil se calcula resolviendo las ecuaciones de verosimilitud, las cuales se obtienen de derivar la función de verosimilitud con respecto a cada uno de los parámetros e igualar las expresiones a cero. Lo más habitual es tomar el logaritmo neperiano de la función de verosimilitud, $\ln L_{x_1, \dots, x_n}(\theta)$, debido a que muchas distribuciones tienen exponenciales en sus expresiones, y como el logaritmo neperiano es una función creciente, no afecta al cálculo del máximo.

Por tanto si $\ln L_{x_1, \dots, x_n}(\theta)$ es derivable, se obtienen las ecuaciones de verosimilitud:

$$\frac{\partial \ln L_{x_1, \dots, x_n}(\theta)}{\partial \theta_j} = 0 \quad j = 1, \dots, k, \quad \theta = (\theta_1, \dots, \theta_k)$$

Las soluciones de estas ecuaciones son los posibles extremos de $\ln L_{x_1, \dots, x_n}(\theta)$, que pueden ser máximos o no. Si la solución es única y es un máximo y es el EMV de θ , $\hat{\theta}$. Si existen varias soluciones, se puede tomar el máximo absoluto entre ellas como $\hat{\theta}$. Una vez obtenida la solución se debe comprobar que, efectivamente, se trata de un estimador.

- Si la función de verosimilitud no es derivable, entonces hay que recurrir a otro tipo de métodos, incluso métodos numéricos, para obtener el máximo.

Ejemplos:

1. Sea X_1, \dots, X_n una m.a.s. de $X \sim \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$. Encontrar el EMV para μ y σ^2 , en el caso de un parámetro conocidos y cuando ambos parámetros son desconocidos.
2. Sea X_1, \dots, X_n una m.a.s. de X cuya f.m.p es: $P_N[X = x] = \frac{1}{N}$, $x = 1, \dots, N$ (uniforme discreta en N puntos). Calcular el EMV de N .
3. Sea X_1, \dots, X_n una m.a.s. de $X \sim \{U(\theta - 1/2, \theta + 1/2), \theta \in \mathbb{R}\}$. Calcular un EMV para θ . (El EMV no tiene que ser único).

4. Sea X una v.a. con distribución $B(1, p)$ con $p \in [1/4, 3/4]$. Calcular el EMV de p , para una muestra de tamaño 1, ver que no es insesgado y calcular su error cuadrático medio: $E[\hat{p} - p]^2$. Comprobar que el estimador $T(X) = 1/2$ es mejor que \hat{p} en el sentido del ECM. (El EMV no tiene porque ser el mejor en el sentido del menor error cuadrático medio).

5.1.1. Propiedades de los EMV

Aunque estos ejemplos muestran que los EMV no tienen por qué ser únicos, ni insesgados, ni minimizar el error cuadrático medio, sí que tienen ciertas propiedades.

Teorema 1: (Propiedades asintóticas)

Bajo condiciones bastantes generales, que incluyen las condiciones de regularidad de Fréchet-Cramér-Rao, si las ecuaciones de verosimilitud tienen solución única, $\hat{\theta}(X_1, \dots, X_n)$, esta solución satisface:

- El EMV es fuertemente consistente.

$$\hat{\theta}(X_1, \dots, X_n) \xrightarrow[n \rightarrow +\infty]{c.s.} \theta, \quad \forall \theta \in \Theta.$$

- El EMV es asintóticamente normal.

$$\hat{\theta}(X_1, \dots, X_n) \xrightarrow[n \rightarrow +\infty]{L} \mathcal{N}(\theta, 1/(nI_X(\theta))), \quad \forall \theta \in \Theta.$$

La normalidad asintótica implica que, para muestras grandes, la distribución del EMV es aproximadamente normal, de media θ , y su varianza alcanza la cota de FCR ($\text{Var}_{\theta}[\hat{\theta}(X_1, \dots, X_n)] \approx 1/(nI_X(\theta)) = 1/I_{X_1, \dots, X_n}(\theta)$).

Teorema 2: (Relación entre EMV y estadístico suficiente)

Sea X una v.a. con distribución en una familia paramétrica de distribuciones, $\{F_{\theta}, \theta \in \Theta\}$. Sea X_1, \dots, X_n una m.a.s. de X . Supongamos que la familia admite un estadístico suficiente $T(X_1, \dots, X_n)$. Entonces, si existe un EMV de θ , es una función (no constante) de $T(X_1, \dots, X_n)$.

Este resultado no implica que el EMV tenga que ser un estadístico suficiente.

Ejemplo: Sea X_1, \dots, X_n una m.a.s. de $X \sim U(\theta, \theta+1)$. Calcular el estadístico suficiente, un EMV para θ , y comprobar que no coinciden.

Teorema 3: (Relación entre EMV y estimador eficiente)

Sea X una v.a. con distribución en una familia paramétrica de distribuciones, $\{F_{\theta}, \theta \in \Theta\}$, regular con $0 < I_X(\theta) < \infty$. Sea X_1, \dots, X_n una m.a.s. de X . Si $T(X_1, \dots, X_n)$ es un estimador eficiente para θ , entonces existe un único EMV y coincide con $T(X_1, \dots, X_n)$.

5.1.2. Estimadores de máxima verosimilitud de una una función paramétrica

Sea $g : \Theta \rightarrow \Lambda$ una función paramétrica. Se puede definir el concepto de función de verosimilitud sobre Λ a partir de la función de verosimilitud definida sobre Θ de la siguiente forma:

Definición: Para cada x_1, \dots, x_n , realización muestral, se define la *función de verosimilitud* de $\lambda = g(\theta)$ asociada a dicha realización como:

$$\begin{aligned} M_{x_1, \dots, x_n} : \Lambda &\rightarrow \mathbb{R}^+ \cup \{0\} \\ \lambda &\rightarrow M_{x_1, \dots, x_n}(\lambda) = \sup_{\theta \in g^{-1}(\lambda)} L_{x_1, \dots, x_n}(\theta). \end{aligned}$$

Por otro lado, de forma análoga a como se definió el EMV de θ se puede definir el de λ .

Definición: Un estimador $\hat{\lambda}(X_1, \dots, X_n)$ de λ es *estimador de máximo verosimilitud* (EMV) de λ si:

$$\forall (x_1, \dots, x_n) \in \mathcal{X}^n, \quad M_{x_1, \dots, x_n}(\hat{\lambda}(x_1, \dots, x_n)) = \max_{\lambda \in \Lambda} M_{x_1, \dots, x_n}(\lambda).$$

Teorema de invarianza de Zehna: Sea X una v.a. con distribución en una familia paramétrica de distribuciones, $\{F_\theta, \theta \in \Theta\}$. Sea X_1, \dots, X_n una m.a.s. de X . Sea g una función medible. Si $\hat{\theta}(X_1, \dots, X_n)$ es EMV de θ , entonces $g(\hat{\theta}(X_1, \dots, X_n))$ es EMV de $g(\theta)$.

Ejemplos:

1. En el muestreo de una v.a. $X \sim \{\mathcal{P}(\lambda), \lambda > 0\}$, se obtiene que en n observaciones aparece y veces el valor 0. Obtener un EMV de λ a partir de esta información.
2. La duración de cierto tipo de lámparas es exponencial de media θ , desconocida. Después de observar el tiempo de vida de n lámparas, estimar por máxima verosimilitud la probabilidad de que la duración de una lámpara sea superior a 500 horas.

5.2. Otros métodos de estimación puntual: método de los momentos y de mínimos cuadrados

5.2.1. Método de los momentos

El método de los momentos, el cual fue introducido por K. Pearson, es el método más antiguo y sencillo para obtener estimadores de los parámetros poblacionales.

Este método consiste en estimar cualquier función medible de los momentos poblacionales por la misma función de los momentos muestrales. En particular se igualan tantos momentos muestrales como parámetros haya que estimar, a los correspondientes momentos poblacionales, que son funciones de los parámetros desconocidos, y se resuelve el sistema de ecuaciones resultante obteniéndose estimadores de los parámetros.

La idea es la siguiente: Sea X una v.a. con distribución en una familia paramétrica de distribuciones, $\{F_\theta, \theta \in \Theta \subseteq \mathbb{R}^k\}$, es decir, $\theta = (\theta_1, \dots, \theta_k)$. Sea X_1, \dots, X_n una m.a.s. de X . Sean m_1, \dots, m_k los k primeros momentos no centrados de X ,

$$m_j(\theta_1, \dots, \theta_k) = E_{(\theta_1, \dots, \theta_k)}[X^j] = \begin{cases} \sum x_i^j P_{\theta_1, \dots, \theta_k}[X = x_i] & \text{caso discreto} \\ \int x^j f_{\theta_1, \dots, \theta_k}(x) dx & \text{caso continuo} \end{cases}$$

En general m_j será una función de los k parámetros $\theta_1, \dots, \theta_k$.

Por otro lado, asociados a la muestra se pueden obtener los k primeros momentos no centrados muestrales, que son:

$$A_1 = \sum_{i=1}^n \frac{X_i}{n}, \dots, A_j = \sum_{i=1}^n \frac{X_i^j}{n}, \dots, A_k = \sum_{i=1}^n \frac{X_i^k}{n}$$

Igualando los k primeros momentos poblacionales, m_j , a los correspondientes momentos muestrales, A_j , se obtiene un sistema de k ecuaciones con k incógnitas, $\theta_1, \dots, \theta_k$,

$$\left. \begin{aligned} m_1(\theta_1, \dots, \theta_k) &= A_1 \\ &\vdots \\ m_j(\theta_1, \dots, \theta_k) &= A_j \\ &\vdots \\ m_k(\theta_1, \dots, \theta_k) &= A_k \end{aligned} \right\}$$

cuyas soluciones son los estimadores de los parámetros: $\hat{\theta}_1, \dots, \hat{\theta}_k$.

Ejemplos: Estimar mediante el método de los momentos los parámetros de las siguientes distribuciones:

1. $X \sim \mathcal{N}(\mu, \sigma^2)$.
2. $X \sim U(a, b)$.
3. $X \sim \mathcal{P}(\lambda)$.
4. $X \sim U(0, \theta)$.

La propiedad más notable de los estimadores obtenidos por el método de los momentos es la propiedad de consistencia: Si $\theta = h(m_1, \dots, m_k)$, h continua, y $A_1^{(n)}, \dots, A_k^{(n)}$ son los momentos muestrales correspondientes a una muestra de tamaño n , entonces

$$\hat{\theta}_n = h(A_1^{(n)}, \dots, A_k^{(n)}) \rightarrow h(m_1, \dots, m_k) = \theta \quad c.s. \quad (n \rightarrow \infty)$$

5.2.2. Método de mínimos cuadrados

Sea $X = \varphi(t, \theta)$ una magnitud (de interés) que depende de ciertas condiciones experimentales (t) y de ciertos parámetros desconocidos a estimar ($\theta = (\theta_1, \dots, \theta_k) \in \Theta \subseteq \mathbb{R}^k$).

En principio, si fuese posible observar X bajo distintas condiciones experimentales (t_i) se podrían procurar suficientes relaciones del tipo

$$\left. \begin{array}{l} x_1 = \varphi(t_1, \theta) \\ \vdots \\ x_n = \varphi(t_n, \theta) \end{array} \right\}$$

para determinar $\theta_1, \dots, \theta_k$ despejando en el sistema obtenido.

Sin embargo, las observaciones de X conlleva un error de medida aleatorio, ϵ , de forma que se obtendrían relaciones en términos de variables aleatorias, del tipo

$$\left. \begin{array}{l} X_1 = \varphi(t_1, \theta) + \epsilon_1 \\ \vdots \\ X_n = \varphi(t_n, \theta) + \epsilon_n \end{array} \right\}$$

Se plantea, entonces, el problema de estimar θ a partir de la muestra aleatoria X_1, \dots, X_n .

El método de mínimos cuadrados consiste en elegir el θ que minimice

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (X_i - \varphi(t_i, \theta))^2$$

Si φ es derivable respecto a θ , la solución verificará el sistema

$$\sum_{i=1}^n (X_i - \varphi(t_i, \theta)) \frac{\partial}{\partial \theta_j} \varphi(t_i, \theta) = 0, \quad j = 1, \dots, k.$$

Las propiedades de los estimadores obtenidos mediante el método de mínimos cuadrados dependen del problema particular analizado. Este método tendrá especial interés en el desarrollo del Tema 8.

Ejemplo: Para estimar la aceleración de la gravedad en una ciudad, θ , se deja caer un objeto durante tiempos t_1, \dots, t_n y se mide el espacio recorrido en cada tiempo. Si X_i representa la medida correspondiente al espacio recorrido en el tiempo t_i , con error de medida ϵ_i , estimar θ por el método de mínimos cuadrados. ($e = \frac{1}{2} at^2$)

Tema 6

ESTIMACIÓN POR INTERVALOS DE CONFIANZA

6.1. Planteamiento del problema y conceptos básicos

En primer lugar se recuerda que el problema que se esta intentando resolver es la inferencia del parámetro, ó parámetros, desconocido de una familia de funciones paramétricas, a la cual pertenece la distribución de la variable aleatoria bajo estudio.

Hasta ahora se ha afrontado dicha inferencia mediante una estimación puntual. Sin embargo, existe otra forma de resolver dicho problema, mediante la inferencia por intervalos de confianza, es decir, en lugar de dar un valor para el parámetro, se da un intervalo, que contiene al parámetro con una cierta probabilidad.

Definición: Sea X una variable aleatoria con distribución en la familia $\{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^k\}$ y sea X_1, \dots, X_n una muestra aleatoria de X . Sea \mathcal{X}^n es espacio muestral de los posibles valores de la muestra.

Sea S una función definida:

$$S : \mathcal{X}^n \longrightarrow \mathcal{P}(\Theta) \quad \text{independiente de } \theta.$$

A $S(X_1, \dots, X_n)$ se le denomina *conjunto aleatorio*.

Existe cierta similitud entre la definición de estimador y la de conjunto aleatorio, pero la principal diferencia es que fijada una muestra, x_1, \dots, x_n , $T(x_1, \dots, x_n)$ es un valor concreto mientras que $S(x_1, \dots, x_n)$ es un conjunto que es un subespacio del espacio paramétrico.

En el caso particular en que el conjunto aleatorio sea un intervalo, $S(X_1, \dots, X_n) = (I_1(X_1, \dots, X_n), I_2(X_1, \dots, X_n))$, se le denomina *intervalo aleatorio*.

Definición: Un intervalo aleatorio $(I_1(X_1, \dots, X_n), I_2(X_1, \dots, X_n))$, se dice que es un

6.1 Planteamiento del problema

intervalo de confianza para el parámetro θ al nivel de confianza $1 - \alpha$ si

$$\forall \theta \in \Theta \quad P_\theta[I_1(X_1, \dots, X_n) \leq \theta \leq I_2(X_1, \dots, X_n)] \geq 1 - \alpha$$

es decir, sea cual sea el valor del parámetro la probabilidad de que el intervalo aleatorio contenga al parámetro es mayor o igual que $1 - \alpha$.

Según la definición frecuentista de probabilidad, esto puede interpretarse como que el $(1 - \alpha)100\%$ de las veces que se observe una muestra de tamaño n , el intervalo concreto obtenido, $(I_1(x_1, \dots, x_n), I_2(x_1, \dots, x_n))$, contendrá al verdadero valor del parámetro; por tanto, cada vez que tomamos una muestra, tenemos una confianza del $(1 - \alpha)100\%$ de que así ocurre. En general conviene tomar un valor de α pequeño, para que la probabilidad de que el intervalo contenga al verdadero valor del parámetro sea casi 1.

Los intervalos de confianza se pueden dividir en dos grupos:

- Intervalos de confianza bilaterales: Son aquellos donde los dos extremos del intervalo toman valores finitos,

$$S(X_1, \dots, X_n) = (I_1(X_1, \dots, X_n), I_2(X_1, \dots, X_n)).$$

- Intervalos de confianza unilaterales: Son aquellos donde uno de los extremos del intervalo no toma un valor finito,

$$S(X_1, \dots, X_n) = (-\infty, I_2(X_1, \dots, X_n)) \quad \text{ó} \quad S(X_1, \dots, X_n) = (I_1(X_1, \dots, X_n), +\infty).$$

A partir de los intervalos de confianza unilaterales se pueden obtener cotas de confianza para un nivel de confianza fijado:

- Cota superior de confianza: A partir del intervalo unilateral

$$\forall \theta \in \Theta \quad P_\theta[-\infty < \theta \leq I_2(X_1, \dots, X_n)] \geq 1 - \alpha$$

se obtiene $P_\theta[I_2(X_1, \dots, X_n) \geq \theta] \geq 1 - \alpha$, por lo tanto $I_2(X_1, \dots, X_n)$ es una cota de confianza superior al nivel de confianza $1 - \alpha$.

- Cota inferior de confianza: A partir del intervalo unilateral

$$\forall \theta \in \Theta \quad P_\theta[I_1(X_1, \dots, X_n) \leq \theta < +\infty] \geq 1 - \alpha$$

se obtiene $P_\theta[I_1(X_1, \dots, X_n) \leq \theta] \geq 1 - \alpha$, por lo tanto $I_1(X_1, \dots, X_n)$ es una cota de confianza inferior al nivel de confianza $1 - \alpha$.

Ejemplo: Sea X_1, \dots, X_n una m.a.s. de $X \rightsquigarrow \{\mathcal{N}(\mu, \sigma_0^2), \mu \in \mathbb{R}\}$. Construir un intervalo de confianza para μ al nivel de confianza $1 - \alpha$ de la forma $(\bar{X} - C_1, \bar{X} + C_2)$. Repetir la construcción del intervalo bajo la hipótesis de que σ^2 es desconocida.

En el ejemplo anterior se ha obtenido un intervalo de confianza para μ al nivel de confianza $1 - \alpha$, imponiendo cierta estructura en el mismo. Pero no es la única opción de intervalo de confianza en dicho caso. Por ejemplo, si se considera $1 - \alpha = 0.95$, $(\bar{X} \mp 2\sigma_0/\sqrt{n})$ es otro intervalo de confianza al mismo nivel. Sin embargo, el obtenido en el ejemplo, $(\bar{X} \mp 1.96\sigma_0/\sqrt{n})$ sería un intervalo más informativo debido a que su amplitud, ó longitud, es menor.

Por tanto sería conveniente establecer criterios para comparar entre distintos intervalos de confianza, al mismo nivel de confianza, y poder determinar que intervalo es el óptimo.

6.1.1. Intervalos de menor longitud esperada

Como se ha visto anteriormente, una forma de comparar intervalos de confianza puede ser mediante la longitud de los mismos, entendiendo por longitud la amplitud del intervalo:

$$L = I_2(X_1, \dots, X_n) - I_1(X_1, \dots, X_n).$$

Por su definición, la longitud es una variable aleatoria, por lo tanto, para ser más exactos se utilizará la longitud esperada, que proporciona un valor numérico con el que poder comparar los intervalos.

Definición: Se dice que un intervalo de confianza, $(I_1(x_1, \dots, x_n), I_2(x_1, \dots, x_n))$, al nivel de confianza $1 - \alpha$, tiene *menor longitud esperada uniforme* si para cualquier otro intervalo de confianza, $(I'_1(x_1, \dots, x_n), I'_2(x_1, \dots, x_n))$, al mismo nivel de confianza, se tiene que

$$\forall \theta \in \Theta \quad E_\theta[I_2(X_1, \dots, X_n) - I_1(X_1, \dots, X_n)] \leq E_\theta[I'_2(X_1, \dots, X_n) - I'_1(X_1, \dots, X_n)].$$

En general no tiene porque existir un intervalo de confianza con mínima longitud esperada uniformemente para todos los valores del parámetro. Además, de existir, puede variar al cambiar el parámetro.

6.2. Métodos de construcción

Una vez dada la definición de intervalo de confianza y estudiado algún criterio que permita compara entre dos intervalos con el mismo nivel de confianza, es necesario el estudio de métodos de construcción de intervalos de confianza. En este tema se van a estudiar dos métodos distintos entre los que existen. El primero de ellos va a poder usarse bajo condiciones muy generales, aunque tendrá proporcionará intervalos de longitud muy amplia. El segundo es más restrictivo en sus hipótesis pero puede proporcionar intervalos con mínima longitud esperada uniforme y es el que hoy en día está implementado en los paquetes estadísticos para determinar los intervalos de confianza bajo normalidad.

6.2.1. Intervalos de confianza obtenidos mediante la desigualdad de Chebychev

La desigualdad de Chebychev dice que para cualquier v.a. X de segundo orden se verifica:

$$P[|X - E[X]| < k] \geq 1 - \frac{Var(X)}{k^2}, \quad \forall k > 0$$

A partir de esta desigualdad se puede obtener un I.C. para cualquier parámetro de la siguiente forma. Sea $T(X_1, \dots, X_n)$ un estimador insesgado del parámetro θ con varianza uniformemente acotada, $(E_\theta[T(X_1, \dots, X_n)] = \theta$ y $Var_\theta[T(X_1, \dots, X_n)] \leq c, \forall \theta \in \Theta$). Si le aplicamos la desigualdad de Chebychev se tiene:

$$P[|T(X_1, \dots, X_n) - E_\theta[T(X_1, \dots, X_n)]| < k] \geq 1 - \frac{Var_\theta(T(X_1, \dots, X_n))}{k^2}, \quad \forall k > 0$$

Como el estimador es insesgado y tiene varianza uniformemente acotada operando adecuadamente se tiene,

$$P[T(X_1, \dots, X_n) - k < \theta < T(X_1, \dots, X_n) + k] \geq 1 - \frac{c}{k^2}, \quad \forall k > 0$$

Por lo tanto, fijado un nivel de confianza $1 - \alpha$, se obtiene el intervalo de confianza para θ ,

$$\left(T(X_1, \dots, X_n) - \sqrt{\frac{c}{\alpha}}, T(X_1, \dots, X_n) + \sqrt{\frac{c}{\alpha}} \right).$$

En general la ventaja que tiene este método es que se puede aplicar a cualquier variable, ya que no se le impone nada a ella, pero tiene el inconveniente de que suele proporcionar intervalos muy grandes.

Ejemplo: Obtener, utilizando la desigualdad de Chebychev, un intervalo de confianza para la media poblacional, μ , basado en una muestra de tamaño n de una variable aleatoria cuya varianza, σ^2 , es conocida.

6.2.2. Método de la cantidad pivotal

Sea X_1, \dots, X_n una m.a.s. de $X \rightsquigarrow F \in \{F_\theta, \theta \in \Theta \subset \mathbb{R}\}$. Sea \mathcal{X}^n el espacio muestral.

Definición: Una función de la muestra y del parámetro, $T : \mathcal{X}^n \times \Theta \rightarrow \mathbb{R}$, que sea una variable aleatoria cuya distribución es independiente del parámetro, θ , se denomina *función pivote*.

Teorema: Sea $T(X_1, \dots, X_n; \theta)$ una función pivote. Si se cumplen las siguientes condiciones:

1. $T(X_1, \dots, X_n; \theta)$ es estrictamente monótona en θ .
2. Sea Λ la imagen de T . $\forall \lambda \in \Lambda$, $T(X_1, \dots, X_n; \theta) = \lambda$ tiene solución en θ .

Entonces, se puede construir un intervalo de confianza para θ a cualquier nivel de confianza.

Descripción del método: Fijado un nivel de confianza $1 - \alpha$, con $0 < \alpha < 1$, se quiere construir un intervalo $(I_1(X_1, \dots, X_n), I_2(X_1, \dots, X_n))$ tal que:

$$P_\theta[I_1(X_1, \dots, X_n) < \theta < I_2(X_1, \dots, X_n)] \geq 1 - \alpha$$

es decir, un intervalo de confianza al nivel establecido. Sea $T(X_1, \dots, X_n; \theta)$ una función pivote en las condiciones del teorema anterior. Entonces se buscan dos valores $\lambda_1(\alpha)$ y $\lambda_2(\alpha)$ tales que:

$$P_\theta[\lambda_1(\alpha) < T(X_1, \dots, X_n; \theta) < \lambda_2(\alpha)] \geq 1 - \alpha.$$

Por cierta hipótesis realizada, se tiene garantizada la existencia de solución, en θ , para el sistema:

$$\left. \begin{array}{l} T(X_1, \dots, X_n; \theta) = \lambda_1(\alpha) \\ T(X_1, \dots, X_n; \theta) = \lambda_2(\alpha) \end{array} \right\} \Rightarrow \text{Solución: } \left\{ \begin{array}{l} \hat{\theta}_1(X_1, \dots, X_n) \\ \hat{\theta}_2(X_1, \dots, X_n) \end{array} \right.$$

Por tanto, y teniendo en cuenta que la función pivote T , fijada la muestra, es estrictamente monótona, se tiene:

- Si T es creciente:

$$\begin{aligned} P_\theta[\lambda_1(\alpha) < T(X_1, \dots, X_n, \theta) < \lambda_2(\alpha)] &= \\ P_\theta[\hat{\theta}_1(X_1, \dots, X_n) < \theta < \hat{\theta}_2(X_1, \dots, X_n)] &\geq 1 - \alpha \quad \forall \theta \in \Theta \end{aligned}$$

es decir, el intervalo de confianza, al nivel de confianza $1 - \alpha$, viene determinado por $(\hat{\theta}_1(X_1, \dots, X_n), \hat{\theta}_2(X_1, \dots, X_n))$.

- Si T es decreciente:

$$\begin{aligned} P_\theta[\lambda_1(\alpha) < T(X_1, \dots, X_n, \theta) < \lambda_2(\alpha)] &= \\ P_\theta[\hat{\theta}_2(X_1, \dots, X_n) < \theta < \hat{\theta}_1(X_1, \dots, X_n)] &\geq 1 - \alpha \quad \forall \theta \in \Theta \end{aligned}$$

es decir, el intervalo de confianza, al nivel de confianza $1 - \alpha$, viene determinado por $(\hat{\theta}_2(X_1, \dots, X_n), \hat{\theta}_1(X_1, \dots, X_n))$.

Determinación de pivotes en distribuciones continuas

- Si X es de tipo continuo con función de distribución F_θ , la siguiente función es un pivote para θ :

$$T(X_1, \dots, X_n; \theta) = -2 \sum_{i=1}^n \ln F_\theta(X_i) \rightsquigarrow \chi^2(2n).$$

- Si $S(X_1, \dots, X_n)$ es un estadístico con distribución de tipo continuo y F_θ^S es su función de distribución, la siguiente función constituye un pivote para θ :

$$T(X_1, \dots, X_n; \theta) = F_\theta^S(S(X_1, \dots, X_n)) \rightsquigarrow U(0, 1).$$

Ejemplo: Sea X una variable con función de densidad $f_\theta(x) = \theta x^{\theta-1}$ $0 < x < 1$. Determinar un intervalo de confianza para θ utilizando el método de la cantidad pivotal.

6.3. Intervalos de confianza para los parámetros de una población normal

En esta sección vamos a determinar los intervalos de confianza para los parámetros de una población normal usando el método de construcción el del pivote y como criterio de selección el de menor longitud esperada uniforme.

6.3.1. Intervalos de confianza para la media, μ , de una normal con varianza conocida, σ_0^2 .

Sea (X_1, \dots, X_n) una muestra aleatoria simple de la una variable $X \rightsquigarrow \mathcal{N}(\mu, \sigma_0^2)$. Consideramos como función pivote

$$T(X_1, \dots, X_n; \mu) = \frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} \rightsquigarrow \mathcal{N}(0, 1)$$

Efectivamente es una función de la muestra y el parámetro con distribución independiente de μ . Además dicha función cumple todas las condiciones del teorema estudiado en el método del pivote o de la cantidad pivotal:

1. $\frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}}$ es estrictamente decreciente en μ .
2. $\frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} = \lambda \Rightarrow \mu = \bar{X} - \lambda \frac{\sigma_0}{\sqrt{n}}.$

Por lo tanto, aplicando el método obtenemos un intervalo de confianza para μ al nivel de confianza $1 - \alpha$:

$$\left(\bar{X} - \lambda_2 \frac{\sigma_0}{\sqrt{n}}, \bar{X} - \lambda_1 \frac{\sigma_0}{\sqrt{n}} \right)$$

donde λ_1 y λ_2 verifican

$$P_\mu \left(\lambda_1 < \frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} < \lambda_2 \right) = 1 - \alpha,$$

o equivalentemente

$$P(\lambda_1 < Z < \lambda_2) = 1 - \alpha \Leftrightarrow F_Z(\lambda_2) - F_Z(\lambda_1) = 1 - \alpha \quad (F_Z(z) = P[Z \leq z]).$$

Ahora debemos buscar entre todos estos intervalos el que tenga menor longitud esperada uniforme, es decir que minimize $(\forall \mu \in \mathbb{R})$ la longitud media:

$$E_\mu \left[\left(\bar{X} - \lambda_1 \frac{\sigma_0}{\sqrt{n}} \right) - \left(\bar{X} - \lambda_2 \frac{\sigma_0}{\sqrt{n}} \right) \right] = (\lambda_2 - \lambda_1) \frac{\sigma_0}{\sqrt{n}}$$

bajo la restricción $F_Z(\lambda_2) - F_Z(\lambda_1) = 1 - \alpha$. Una forma de resolver este problema es aplicar el método de los multiplicadores de Lagrange y, por lo tanto, minimizar la siguiente función:

$$F(\lambda_1, \lambda_2) = \lambda_2 - \lambda_1 - \lambda [F_Z(\lambda_2) - F_Z(\lambda_1) - (1 - \alpha)]$$

donde no se ha considerado la parte constante de la longitud media por no afectar al procedimiento de minimización.

Derivando y resolviendo el sistema de ecuaciones normales al que se llega se obtiene como solución $f_Z(\lambda_1) = f_Z(\lambda_2)$, donde se ha denotado por f_Z a la función de densidad de la normal estándar. Por lo tanto, λ_1 y λ_2 deben ser dos valores de la $\mathcal{N}(0, 1)$ que tengan el mismo valor en su función de densidad. Teniendo en cuenta la simetría de la densidad normal eso sólo puede ocurrir si $\lambda_1 = \lambda_2$ ó si $\lambda_1 = -\lambda_2$. La primera de dichas opciones se descarta por no verificar la restricción $(F_Z(\lambda_1) - F_Z(\lambda_1) = 0 \neq 1 - \alpha)$. Así que la única posible solución al problema de minimización planteado es $\lambda_1 = -\lambda_2$. Si ahora tenemos en cuenta la restricción $F_Z(\lambda_2) - F_Z(\lambda_1) = 1 - \alpha$ se tiene que

$$\lambda_1 = -z_{\alpha/2} \quad \text{y} \quad \lambda_2 = z_{\alpha/2}$$

siendo $P[Z > z_{\alpha/2}] = \alpha/2$.

De esta forma se ha obtenido el intervalo de confianza para μ de menor longitud media uniformemente, al nivel de confianza $1 - \alpha$:

$$\left(\bar{X} - z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}} \right)$$

Ejemplo: Comparar este intervalo con el que se obtuvo como ejemplo utilizando la desigualdad de Chebychev.

Intervalos de confianza unilaterales

Para obtener intervalos de confianza unilaterales, como ya se ha visto anteriormente, uno de los extremos se considera que no toma un valor finito.

- Si $\lambda_1 = -\infty$, entonces λ_2 debe verificar

$$P(-\infty < Z < \lambda_2) = 1 - \alpha \Rightarrow P(Z < \lambda_2) = 1 - \alpha \Rightarrow P(Z > \lambda_2) = \alpha \Rightarrow \lambda_2 = z_\alpha$$

y el intervalo quedaría

$$\left(\bar{X} - z_\alpha \frac{\sigma_0}{\sqrt{n}}, +\infty \right).$$

- Si $\lambda_2 = +\infty$, entonces λ_1 debe verificar

$$P(\lambda_1 < Z < +\infty) = 1 - \alpha \Rightarrow P(Z > \lambda_1) = 1 - \alpha \Rightarrow P(Z > -\lambda_1) = \alpha \Rightarrow \lambda_1 = -z_\alpha$$

y el intervalo quedaría

$$\left(-\infty, \bar{X} + z_\alpha \frac{\sigma_0}{\sqrt{n}} \right).$$

Como es evidente estos intervalos cumplen tener nivel de confianza $1 - \alpha$, pero su longitud no es finita.

6.3.2. Intervalos de confianza para la media, μ , de una normal con varianza desconocida, σ^2 .

Sea (X_1, \dots, X_n) una muestra aleatoria simple de la una variable $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. Consideramos como función pivote

$$T(X_1, \dots, X_n; \mu) = \frac{\bar{X} - \mu}{S/\sqrt{n}} \rightsquigarrow t(n-1)$$

Dicha función, al igual que en el caso analizado anteriormente, cumple todas las condiciones para poder aplicar el método de la cantidad pivotal. Así que, siguiendo un procedimiento análogo y teniendo en cuenta que la distribución t de Student tiene las mismas propiedades de simetría con respecto al origen que la $\mathcal{N}(0, 1)$, se llega a que el intervalo

de confianza para μ de menor longitud media uniformemente, al nivel de confianza $1 - \alpha$ es:

$$\left(\bar{X} - t_{n-1;\alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1;\alpha/2} \frac{S}{\sqrt{n}} \right)$$

Intervalos de confianza unilaterales

Al igual que antes, los intervalos unilaterales son:

- Para $\lambda_1 = -\infty$

$$\left(\bar{X} - t_{n-1;\alpha} \frac{S}{\sqrt{n}}, +\infty \right).$$

- Para $\lambda_2 = +\infty$

$$\left(-\infty, \bar{X} + t_{n-1;\alpha} \frac{S}{\sqrt{n}} \right).$$

6.3.3. Intervalos de confianza para la varianza, σ^2 , de una normal con media conocida, μ_0 .

Sea (X_1, \dots, X_n) una muestra aleatoria simple de la una variable $X \rightsquigarrow \mathcal{N}(\mu_0, \sigma^2)$. Consideramos como función pivote

$$T(X_1, \dots, X_n; \sigma^2) = \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2} \rightsquigarrow \chi^2(n)$$

Dicha función tiene distribución independiente de σ^2 y cumple todas las condiciones del teorema estudiado en el método del pivote o de la cantidad pivotal:

1. $\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2}$ es estrictamente decreciente en σ^2 .
2. $\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2} = \lambda \Rightarrow \sigma^2 = \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\lambda}$.

Por lo tanto, aplicando el método obtenemos un intervalo de confianza para σ^2 al nivel de confianza $1 - \alpha$:

$$\left(\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\lambda_2}, \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\lambda_1} \right)$$

donde λ_1 y λ_2 verifican

$$P_{\sigma^2} \left(\lambda_1 < \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2} < \lambda_2 \right) = 1 - \alpha,$$

o equivalentemente

$$P(\lambda_1 < T < \lambda_2) = 1 - \alpha \Leftrightarrow F_T(\lambda_2) - F_T(\lambda_1) = 1 - \alpha \quad (T \rightsquigarrow \chi^2(n), F_T(t) = P[T \leq t]).$$

Ahora debemos buscar entre todos estos intervalos el que tenga menor longitud esperada uniforme, es decir que minimize $(\forall \sigma^2 \in \mathbb{R}^+)$ la longitud media:

$$E_{\sigma^2} \left[\left(\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\lambda_1} \right) - \left(\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\lambda_2} \right) \right] = \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) E_{\sigma^2} \left[\sum_{i=1}^n (X_i - \mu_0)^2 \right]$$

bajo la restricción $F_T(\lambda_2) - F_T(\lambda_1) = 1 - \alpha$. Aplicando el método de los multiplicadores de Lagrange se debe minimizar la siguiente función:

$$F(\lambda_1, \lambda_2) = \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) - \lambda [F_T(\lambda_2) - F_T(\lambda_1) - (1 - \alpha)]$$

donde no se ha considerado la parte constante de la longitud media por no afectar al procedimiento de minimización.

Derivando y resolviendo el sistema de ecuaciones normales al que se llega se obtiene como solución

$$\frac{f_T(\lambda_1)}{f_T(\lambda_2)} = \frac{\lambda_2^2}{\lambda_1^2},$$

donde se ha denotado por f_T a la función de densidad de la distribución $\chi^2(n)$. En este caso, la asimetría de la distribución hace que los valores λ_1 y λ_2 no sean los de colas iguales. Sin embargo, la diferencia no es suficientemente importante (sobre todo para grandes muestras), y en la práctica se usa el intervalo de colas iguales:

$$\lambda_1 = \chi_{n;1-\alpha/2}^2 \quad \text{y} \quad \lambda_2 = \chi_{n;\alpha/2}^2$$

siendo $P[\chi_n^2 > \chi_{n;\alpha/2}^2] = \alpha/2$.

De esta forma se ha obtenido un intervalo de confianza para σ^2 , al nivel de confianza $1 - \alpha$:

$$\left(\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n;\alpha/2}^2}, \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n;1-\alpha/2}^2} \right)$$

Intervalos de confianza unilaterales

Para obtener intervalos de confianza unilaterales, como ya se ha visto anteriormente, uno de los extremos se considera que no toma un valor finito. En este caso hay que tener en cuenta la distribución χ^2 sólo está definida en \mathbb{R}^+ , con lo que el mínimo valor que se puede considerar para el extremo inferior del intervalo es el 0.

- Si $\lambda_1 = 0$, entonces λ_2 debe verificar

$$P(0 < T < \lambda_2) = 1 - \alpha \Rightarrow P(T < \lambda_2) = 1 - \alpha \Rightarrow P(T > \lambda_2) = \alpha \Rightarrow \lambda_2 = \chi_{n;\alpha}^2$$

y el intervalo quedaría

$$\left(\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n;\alpha}^2}, +\infty \right).$$

- Si $\lambda_2 = +\infty$, entonces λ_1 debe verificar

$$P(\lambda_1 < T < +\infty) = 1 - \alpha \Rightarrow P(T > \lambda_1) = 1 - \alpha \Rightarrow \lambda_1 = \chi_{n;1-\alpha}^2$$

y el intervalo quedaría

$$\left(0, \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n;1-\alpha}^2} \right).$$

6.3.4. Intervalos de confianza para la varianza, σ^2 , de una normal con media desconocida, μ .

Sea (X_1, \dots, X_n) una muestra aleatoria simple de la una variable $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. Consideramos como función pivote

$$T(X_1, \dots, X_n; \sigma^2) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi^2(n-1)$$

Dicha función, al igual que en el caso analizado anteriormente, cumple todas las condiciones para poder aplicar el método de la cantidad pivotal. Así que, siguiendo un procedimiento análogo y teniendo en cuenta que la distribución que se tiene es la misma salvo por los grados de libertad ($\chi^2(n-1)$), se llega a que un intervalo de confianza para σ^2 , al nivel de confianza $1 - \alpha$ es:

$$\left(\frac{(n-1)S^2}{\chi_{n-1;\alpha/2}^2}, \frac{(n-1)S^2}{\chi_{n-1;1-\alpha/2}^2} \right)$$

Intervalos de confianza unilaterales

Al igual que antes, los intervalos unilaterales son:

- Para $\lambda_1 = 0$

$$\left(\frac{(n-1)S^2}{\chi_{n-1;\alpha}^2}, +\infty \right).$$

- Para $\lambda_2 = +\infty$

$$\left(0, \frac{(n-1)S^2}{\chi_{n-1;1-\alpha}^2} \right).$$

6.4. Intervalos de confianza para los parámetros de dos poblaciones normales

Ahora vamos a determinar los intervalos de confianza para los parámetros de dos poblaciones normales, basados en muestras aleatorias simples de las variables bajo estudio, usando el mismo método y criterio que en la sección anterior.

6.4.1. Intervalos de confianza para la diferencia de medias, $\mu_1 - \mu_2$ de dos normales con varianzas conocidas, σ_1^2, σ_2^2 .

Sea (X_1, \dots, X_{n_1}) una muestra aleatoria simple de la una variable $X \rightsquigarrow \mathcal{N}(\mu_1, \sigma_1^2)$ y sea (Y_1, \dots, Y_{n_2}) una muestra aleatoria simple de la una variable $Y \rightsquigarrow \mathcal{N}(\mu_2, \sigma_2^2)$. Asumimos σ_1^2 y σ_2^2 conocidas. En este caso consideramos como función pivote

$$T(X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}; \mu_1 - \mu_2) = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0, 1).$$

Esta función cumple, además de tener distribución independiente de $\mu_1 - \mu_2$:

1. $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ es estrictamente decreciente en $\mu_1 - \mu_2$.
2. $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \lambda \Rightarrow \mu_1 - \mu_2 = \bar{X} - \bar{Y} - \lambda \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$

Por lo tanto, aplicando el método obtenemos un intervalo de confianza para $\mu_1 - \mu_2$ al nivel de confianza $1 - \alpha$:

$$\left(\bar{X} - \bar{Y} - \lambda_2 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{X} - \bar{Y} - \lambda_1 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

donde λ_1 y λ_2 verifican

$$P_{\mu_1, \mu_2} \left(\lambda_1 < \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < \lambda_2 \right) = 1 - \alpha,$$

o equivalentemente

$$P(\lambda_1 < Z < \lambda_2) = 1 - \alpha \Leftrightarrow F_Z(\lambda_2) - F_Z(\lambda_1) = 1 - \alpha \quad (F_Z(z) = P[Z \leq z]).$$

El problema de buscar, entre estos intervalos, el que minimice uniformemente la longitud media es exactamente equivalente al caso del intervalo de confianza para μ con σ^2 conocida estudiado en la sección 6.3.1. y la solución es por tanto:

$$\lambda_1 = -z_{\alpha/2} \quad \text{y} \quad \lambda_2 = z_{\alpha/2}$$

siendo $P[Z > z_{\alpha/2}] = \alpha/2$.

De esta forma se obtiene que el intervalo de confianza para $\mu_1 - \mu_2$ de menor longitud media uniformemente, al nivel de confianza $1 - \alpha$, es:

$$\left(\bar{X} - \bar{Y} - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{X} - \bar{Y} + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

Intervalos de confianza unilaterales

Análogamente, los intervalos de confianza unilaterales son.:

- Si $\lambda_1 = -\infty$, entonces el intervalo quedaría

$$\left(\bar{X} - \bar{Y} - z_{\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, +\infty \right).$$

- Si $\lambda_2 = +\infty$, entonces el intervalo quedaría

$$\left(-\infty, \bar{X} - \bar{Y} + z_{\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right).$$

6.4.2. Intervalos de confianza para la diferencia de medias, $\mu_1 - \mu_2$ de dos normales con varianzas desconocidas e iguales, σ^2 .

Sea (X_1, \dots, X_{n_1}) una muestra aleatoria simple de la una variable $X \rightsquigarrow \mathcal{N}(\mu_1, \sigma^2)$ y sea (Y_1, \dots, Y_{n_2}) una muestra aleatoria simple de la una variable $Y \rightsquigarrow \mathcal{N}(\mu_2, \sigma^2)$. En este caso se ha asumido que las varianzas de ambas distribuciones son desconocidas pero iguales. La función pivote a considerar en este caso es

$$T(X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}; \mu_1 - \mu_2) = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow t(n_1 + n_2 - 2)$$

donde $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$.

Esta función, al igual que en el caso analizado anteriormente, cumple todas las condiciones para poder aplicar el método de la cantidad pivotal. Así que, siguiendo un procedimiento análogo y, al igual que en el caso analizado en la sección 6.3.2., teniendo en cuenta que la distribución t de Student tiene las mismas propiedades de simetría con respecto al origen que la $\mathcal{N}(0, 1)$, se llega a que el intervalo de confianza para $\mu_1 - \mu_2$ de menor longitud media uniformemente, al nivel de confianza $1 - \alpha$ es:

$$\left(\bar{X} - \bar{Y} - t_{n_1+n_2-2; \alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{X} - \bar{Y} + t_{n_1+n_2-2; \alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

Intervalos de confianza unilaterales

Los intervalos de confianza unilaterales son:

- Si $\lambda_1 = -\infty$, entonces el intervalo quedaría

$$\left(\bar{X} - \bar{Y} - t_{n_1+n_2-2; \alpha} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, +\infty \right).$$

- Si $\lambda_2 = +\infty$, entonces el intervalo quedaría

$$\left(-\infty, \bar{X} - \bar{Y} + t_{n_1+n_2-2; \alpha} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right).$$

6.4.3. Intervalos de confianza para el cociente de varianzas, σ_1^2/σ_2^2 de dos normales con medias conocidas, μ_1, μ_2 .

Sea (X_1, \dots, X_{n_1}) una muestra aleatoria simple de la una variable $X \rightsquigarrow \mathcal{N}(\mu_1, \sigma_1^2)$ y sea (Y_1, \dots, Y_{n_2}) una muestra aleatoria simple de la una variable $Y \rightsquigarrow \mathcal{N}(\mu_2, \sigma_2^2)$. Asumimos μ_1 y μ_2 conocidas y consideramos como función pivote:

$$T(X_1, \dots, X_{n_1}; \sigma_1^2/\sigma_2^2) = \frac{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2/n_2 \sigma_2^2}{\sum_{i=1}^{n_1} (X_i - \mu_1)^2/n_1 \sigma_1^2} \rightsquigarrow \mathcal{F}(n_2, n_1).$$

Dicha función tiene distribución independiente de σ_1^2/σ_2^2 y cumple todas las condiciones del teorema estudiado en el método del pivote o de la cantidad pivotal:

1. $\frac{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2/n_2 \sigma_2^2}{\sum_{i=1}^{n_1} (X_i - \mu_1)^2/n_1 \sigma_1^2} = \frac{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2/n_2}{\sum_{i=1}^{n_1} (X_i - \mu_1)^2/n_1} \frac{\sigma_1^2}{\sigma_2^2}$ es estrictamente creciente en σ_1^2/σ_2^2 .

$$2. \frac{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2 / n_2}{\sum_{i=1}^{n_1} (X_i - \mu_1)^2 / n_1} \frac{\sigma_1^2}{\sigma_2^2} = \lambda \Rightarrow \frac{\sigma_1^2}{\sigma_2^2} = \lambda \frac{\sum_{i=1}^{n_1} (X_i - \mu_1)^2 / n_1}{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2 / n_2}.$$

Por lo tanto, aplicando el método obtenemos un intervalo de confianza para $\frac{\sigma_1^2}{\sigma_2^2}$ al nivel de confianza $1 - \alpha$:

$$\left(\lambda_1 \frac{\sum_{i=1}^{n_1} (X_i - \mu_1)^2 / n_1}{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2 / n_2}, \lambda_2 \frac{\sum_{i=1}^{n_1} (X_i - \mu_1)^2 / n_1}{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2 / n_2} \right)$$

donde λ_1 y λ_2 verifican

$$P_{\sigma_1^2, \sigma_2^2} \left(\lambda_1 < \frac{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2 / n_2 \sigma_2^2}{\sum_{i=1}^{n_1} (X_i - \mu_1)^2 / n_1 \sigma_1^2} < \lambda_2 \right) = 1 - \alpha,$$

o equivalentemente

$$P(\lambda_1 < T < \lambda_2) = 1 - \alpha \Leftrightarrow F_T(\lambda_2) - F_T(\lambda_1) = 1 - \alpha \quad (T \rightsquigarrow F(n_2, n_1), F_T(t) = P[T \leq t]).$$

El problema de buscar, entre estos intervalos, el que minimice uniformemente la longitud media es exactamente equivalente al caso del intervalo de confianza para μ con σ^2 conocida estudiado en la sección 6.3.1. cuya función a minimizar era $\lambda_2 - \lambda_1$ bajo la restricción $F_T(\lambda_2) - F_T(\lambda_1) = 1 - \alpha$, cuya solución vimos que era $f_T(\lambda_2) = f_T(\lambda_1)$ sólo que siendo ahora f_T la función de densidad de la distribución F de Snedecor que tiene las mismas propiedades de asimetría que la χ^2 de Pearson. Por ello, al igual que en los casos estudiado en los intervalos para varianzas de una población normal, el intervalo de mínima longitud esperada uniformemente no es de colas iguales pero en la práctica suele utilizarse el que sí las tiene, es decir:

$$\lambda_1 = F_{n_2, n_1; 1-\alpha/2} \quad \text{y} \quad \lambda_2 = F_{n_2, n_1; \alpha/2}$$

siendo $P[F_{n_2, n_1} > F_{n_2, n_1; \alpha/2}] = \alpha/2$.

De esta forma se ha obtenido un intervalo de confianza para σ_1^2/σ_2^2 , al nivel de confianza $1 - \alpha$:

$$\left(F_{n_2, n_1; 1-\alpha/2} \frac{\sum_{i=1}^{n_1} (X_i - \mu_1)^2 / n_1}{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2 / n_2}, F_{n_2, n_1; \alpha/2} \frac{\sum_{i=1}^{n_1} (X_i - \mu_1)^2 / n_1}{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2 / n_2} \right)$$

Intervalos de confianza unilaterales

Para obtener intervalos de confianza unilaterales, como ya se ha visto anteriormente, uno de los extremos se considera que no toma un valor finito. En este caso, de nuevo, hay que tener en cuenta la distribución F sólo está definida en \mathbb{R}^+ , con lo que el mínimo valor que se puede considerar para el extremo inferior del intervalo es el 0.

- Si $\lambda_1 = 0$, entonces λ_2 debe verificar

$$P(0 < T < \lambda_2) = 1 - \alpha \Rightarrow P(T < \lambda_2) = 1 - \alpha \Rightarrow P(T > \lambda_2) = \alpha \Rightarrow \lambda_2 = F_{n_2, n_1; \alpha}$$

y el intervalo quedaría

$$\left(0, F_{n_2, n_1; \alpha} \frac{\sum_{i=1}^{n_1} (X_i - \mu_1)^2 / n_1}{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2 / n_2} \right).$$

- Si $\lambda_2 = +\infty$, entonces λ_1 debe verificar

$$P(\lambda_1 < T < +\infty) = 1 - \alpha \Rightarrow P(T > \lambda_1) = 1 - \alpha \Rightarrow \lambda_1 = F_{n_2, n_1; 1-\alpha}$$

y el intervalo quedaría

$$\left(F_{n_2, n_1; 1-\alpha} \frac{\sum_{i=1}^{n_1} (X_i - \mu_1)^2 / n_1}{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2 / n_2}, +\infty \right).$$

6.4.4. Intervalos de confianza para el cociente de varianzas, σ_1^2 / σ_2^2 de dos normales con medias desconocidas, μ_1, μ_2 .

Sea (X_1, \dots, X_{n_1}) una muestra aleatoria simple de la una variable $X \rightsquigarrow \mathcal{N}(\mu_1, \sigma^2)$ y sea (Y_1, \dots, Y_{n_2}) una muestra aleatoria simple de la una variable $Y \rightsquigarrow \mathcal{N}(\mu_2, \sigma^2)$. La función pivote a considerar en este caso es

$$T(X_1, \dots, X_n; \sigma_1^2 / \sigma_2^2) = \frac{S_2^2 / \sigma_2^2}{S_1^2 / \sigma_1^2} \rightsquigarrow F(n_2 - 1, n_1 - 1).$$

Esta función, al igual que en el caso analizado anteriormente, cumple todas las condiciones para poder aplicar el método de la cantidad pivotal. Así que, siguiendo un procedimiento análogo al caso analizado en la sección anterior se llega a que un intervalo de confianza para σ_1^2 / σ_2^2 , al nivel de confianza $1 - \alpha$ es:

$$\left(F_{n_2-1, n_1-1; 1-\alpha/2} \frac{S_1^2}{S_2^2}, F_{n_2-1, n_1-1; \alpha/2} \frac{S_1^2}{S_2^2} \right)$$

Intervalos de confianza unilaterales

Los intervalos de confianza unilaterales, como en el caso anterior, son:.

- Si $\lambda_1 = 0$, entonces el intervalo quedaría

$$\left(0, F_{n_2-1, n_1-1; \alpha} \frac{S_1^2}{S_2^2} \right).$$

- Si $\lambda_2 = +\infty$, entonces el intervalo quedaría

$$\left(F_{n_2-1, n_1-1; 1-\alpha} \frac{S_1^2}{S_2^2}, +\infty \right).$$

Tema 7

CONTRASTE DE HIPÓTESIS

7.1. Planteamiento del problema y conceptos básicos

Hasta ahora se ha estudiado, para un problema de inferencia paramétrica, la estimación puntual y la estimación por intervalos. Una tercera opción, que se puede plantear para este tipo de problemas, es el contraste de hipótesis.

Esta opción consiste en plantear una hipótesis sobre el parámetro y decidir según la información de la muestra si dicha hipótesis es rechazada por la muestra o no. Debe tenerse en cuenta que la distribución de la variable, de la que se tiene la muestra, depende de si la hipótesis planteada es cierta o no.

Ejemplo: Se tiene una moneda con probabilidad p de que salga cara y se desea saber si dicha probabilidad es $1/3$. Plantear el problema de contraste asociado.

Definición: Sea $X \rightsquigarrow F \in \{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^k\}$. Una *hipótesis* (paramétrica) es una declaración (afirmación o conjetura) acerca del parámetro desconocido θ . Por ejemplo, $\theta \in \Theta_0$ con $\Theta_0 \subset \Theta$.

Cuando se formula una hipótesis sobre el parámetro, se establece una partición sobre el espacio paramétrico, entre los valores de la hipótesis que se formula y el resto de los valores: $\Theta = \Theta_0 \cup \Theta_1$, $\Theta_0 \cap \Theta_1 = \emptyset$.

Ejemplo: $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$ Si la hipótesis que se desea plantear es $\mu = 3$, entonces separo $\Theta = \{3\} \cup \mathbb{R} - \{3\}$. Otra opción sería plantear la hipótesis $\mu > 0$, entonces separo $\Theta = \mathbb{R}^+ \cup \mathbb{R}^-$.

La hipótesis no sólo establece una partición sobre el espacio paramétrico, sino también sobre la familia de distribuciones: $\{F_\theta; \theta \in \Theta\} = \{F_\theta; \theta \in \Theta_0\} \cup \{F_\theta; \theta \in \Theta_1\}$.

Esto indica que toda hipótesis que se hace sobre el parámetro lleva implícita otra hipótesis que es la negación de la primera.

Usualmente, a la primera hipótesis que se hace se le llama *hipótesis nula*, y se denota

por

$$H_0 : \theta \in \Theta_0 \subset \Theta$$

A la negación de dicha hipótesis, se le llama *hipótesis alternativa*, y se denota por

$$H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$$

Un problema de contraste de hipótesis se leería: Se contrasta la hipótesis nula $H_0 : \theta \in \Theta_0$, frente a la hipótesis alternativa $H_1 : \theta \in \Theta_1$. Bajo el punto de vista clásico, la H_0 se formula como aquella hipótesis en la que se tiene mayor confianza y debe tenerse en cuenta que H_0 y H_1 no son intercambiables.

Si el subconjunto que define la hipótesis, Θ_0 (Θ_1), contiene un solo punto, entonces esa hipótesis, H_0 (H_1), se denomina *hipótesis simple*. En caso contrario, cuando conste de más de un punto, se denomina *hipótesis compuesta*.

Ejemplos:

1. Sea $X \rightsquigarrow B(1, p)$. Se pueden plantear, por ejemplo, los siguientes contrastes de hipótesis:

$$\begin{aligned} & - \begin{cases} H_0 : p = 1/2 \\ H_1 : p \neq 1/2 \end{cases} \\ & - \begin{cases} H_0 : p \leq 1/2 \\ H_1 : p > 1/2 \end{cases} \\ & - \begin{cases} H_0 : p = 1/2 \\ H_1 : p = 1/3 \end{cases} \end{aligned}$$

¿Qué tipo de contraste son?. (El último contraste implica que $X \rightsquigarrow \{B(1, p); p \in \{1/2, 1/3\}\}$, ya que si el espacio paramétrico no se especifica de partida es la unión de las dos hipótesis.)

2. Si se tiene $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. ¿Qué tipo de contraste es el siguiente?

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

Una vez estudiado como se plantea un problema de contraste de hipótesis, el siguiente paso es buscar un procedimiento que permita llegar a una solución del problema. Dicho procedimiento debe estar basado en las observaciones de la variable aleatoria cuya distribución depende de que la hipótesis formulada sea cierta o no.

Definición: Un *test de hipótesis* es un estadístico, $\varphi(X_1, \dots, X_n)$, con valores en $[0, 1]$, que especifica la probabilidad de rechazar H_0 para cada realización muestral.

Los test de hipótesis se pueden clasificar en dos tipos:

- Test No Aleatorizados: Un test no aleatorizado va a ser un procedimiento, mediante el cual, una vez observada la variable aleatoria, una o varias veces, según el tamaño de la muestra, se decide rechazar la hipótesis nula con probabilidad 1 ó se decide rechazarla con probabilidad 0.

Matemáticamente hablando, se considera \mathcal{X}^n , el espacio muestral asociado a la muestra de tamaño n , (X_1, \dots, X_n) , y el test consiste en elegir una *región crítica o de rechazo* $\mathcal{C} \subset \mathcal{X}^n$ tal que si:

$(X_1, \dots, X_n) \in \mathcal{C}$, entonces se rechaza la hipótesis nula H_0 con probabilidad 1.

$(X_1, \dots, X_n) \notin \mathcal{C}$, entonces se rechaza H_0 con probabilidad 0 o equivalentemente no se rechaza H_0 con probabilidad 1.

Esto se puede escribir, de la siguiente forma, mediante una función $\varphi : \mathcal{X}^n \rightarrow \{0, 1\}$, donde $\varphi(X_1, \dots, X_n) = \begin{cases} 1 & \text{si } (X_1, \dots, X_n) \in \mathcal{C} \\ 0 & \text{si } (X_1, \dots, X_n) \notin \mathcal{C} \end{cases}$

Entonces, la región crítica $\mathcal{C} = \varphi^{-1}(1) = \{(x_1, \dots, x_n) \in \mathcal{X}^n / \text{se rechaza } H_0\}$. Al complementario de \mathcal{C} se le llama región de aceptación del test, $\bar{\mathcal{C}}$.

Este tipo de test sólo rechaza o no rechaza radicalmente ya que φ sólo puede tomar los valores 0 y 1.

- Test Aleatorizados: Un test aleatorizado va a ser un procedimiento, mediante el cual, una vez observada la variable aleatoria, una o varias veces, según el tamaño de la muestra, se decide rechazar la hipótesis nula con una cierta probabilidad que dependerá del valor observado.

La diferencia con el test no aleatorizado es que φ no va a estar definida en $\{0, 1\}$, sino $\varphi : \mathcal{X}^n \rightarrow [0, 1]$, donde, como ya se ha indicado, $\varphi(X_1, \dots, X_n)$ expresa la probabilidad de rechazar la hipótesis nula H_0 para cada realización muestral. Además, $\forall \theta \in \Theta$, $E_\theta \varphi(X_1, \dots, X_n)$ va a representar la probabilidad media o esperanza de rechazar H_0 cuando θ es el verdadero valor del parámetro.

La aplicación práctica de este tipo de test requiere de procedimientos auxiliares para tomar la decisión final. Sin embargo, son de gran interés teórico ya que permiten dar la solución óptima, (siempre desde el punto de vista clásico), a muchos problemas de contraste.

En relación con los problemas de diseño y elección de posibles test, se van a definir los siguientes conceptos: tipos de errores, nivel de significación, tamaño de un test y función de potencia.

7.1 Planteamiento del problema

- Tipos de errores: Los test llevan asociado dos tipos de errores:
 - Error de tipo I: Se rechaza la hipótesis nula cuando es cierta.
 - Error de tipo II: No se rechaza la hipótesis nula cuando es falsa.

	H_0 verdadera	H_0 falsa
No rechazar H_0	Correcto	Error Tipo II
Rechazar H_0	Error Tipo I	Correcto

En test no aleatorizados, la probabilidad de cometer un error de tipo I sería $P_\theta[(X_1, \dots, X_n) \in \mathcal{C}] = P_\theta(\mathcal{C})$ con $\theta \in \Theta_0$ y la probabilidad de cometer un error de tipo II sería $P_\theta[(X_1, \dots, X_n) \in \bar{\mathcal{C}}] = P_\theta(\bar{\mathcal{C}})$ con $\theta \in \Theta_1$.

Idealmente interesaría encontrar una región crítica, \mathcal{C} , que minimice los errores Tipo I y Tipo II. En la práctica, esto no será posible (salvo en situaciones muy específicas), por lo que el enfoque que se hace es el siguiente:

1. Se acota o limita la probabilidad de error de tipo I a un nivel, α , prefijado. Evidentemente se van a considerar valores de α pequeños como 0.1, 0.05 ó 0.01.
 2. Se trata de minimizar entonces la probabilidad de error de tipo II, entre los test que son admisibles.
- Función de potencia: Para un test φ , se define la *función de potencia* como la función que asocia a cada parámetro, θ , la probabilidad media de rechazar H_0 cuando el (verdadero) valor del parámetro es θ .

$$\begin{aligned}\beta_\varphi : \Theta &\rightarrow [0, 1] \\ \theta &\rightarrow \beta_\varphi(\theta) = E_\theta[\varphi(X_1, \dots, X_n)]\end{aligned}$$

En el caso de un test φ no aleatorizado, esto se puede expresar,

$$\begin{aligned}\beta_\varphi : \Theta &\rightarrow [0, 1] \\ \theta &\rightarrow \beta_\varphi(\theta) = P_\theta[(X_1, \dots, X_n) \in \mathcal{C}]\end{aligned}$$

- Tamaño de un test: Para un test φ , se define el *tamaño del test* como la máxima probabilidad media de cometer un error de tipo I con dicho test, es decir

$$\sup_{\theta \in \Theta_0} E_\theta[\varphi(X_1, \dots, X_n)] = \sup_{\theta \in \Theta_0} \beta_\varphi(\theta)$$

En el caso de un test φ no aleatorizado, esto se puede expresar,

$$\sup_{\theta \in \Theta_0} P_\theta[(X_1, \dots, X_n) \in \mathcal{C}]$$

- Nivel de significación: Se dice que φ es un test de hipótesis con *nivel de significación* α ($\in [0, 1]$) si su tamaño es menor o igual que α (cota superior de las probabilidades medias de cometer error de tipo I), es decir:

$$\forall \theta \in \Theta_0, \beta_{\varphi}(\theta) = E_{\theta}[\varphi(X_1, \dots, X_n)] \leq \alpha.$$

En el caso de un test φ no aleatorizado, esto se puede expresar,

$$\forall \theta \in \Theta_0 P_{\theta}[(X_1, \dots, X_n) \in \mathcal{C}] \leq \alpha.$$

Planteado un problema de contraste de hipótesis, el objetivo será, dado un nivel de significación α prefijado, encontrar un test φ con dicho nivel de significación, es decir con error de tipo I menor o igual a α , que maximice la función potencia en todos los valores de la hipótesis alternativa, es decir, que minimice el error de tipo II.

En general puede ocurrir que existan más de un test que verifique dichas condiciones. En dicho caso se debe seleccionar el más potente, entendiendo que dados dos test φ y φ^* , ambos con nivel de significación α , se dice que φ es *más potente* que φ^* , para un valor $\theta \in \Theta_1$, si $\beta_{\varphi}(\theta) \geq \beta_{\varphi^*}(\theta)$.

Definición: Dado un nivel de significación α fijo, se dice que un test φ es *uniformemente más potente* (UMP), a ese nivel, si se cumplen dos cosas:

1. φ tiene nivel de significación α , es decir

$$E_{\theta}\varphi(X_1, \dots, X_n) \leq \alpha \quad \forall \theta \in \Theta_0.$$

2. Para cualquier otro test φ^* , con nivel de significación α , se cumple que la función potencia en φ es mayor que en φ^* , $\forall \theta \in \Theta_1$,

$$\beta_{\varphi}(\theta) \geq \beta_{\varphi^*}(\theta), \forall \theta \in \Theta_1.$$

En general no tiene porque existir un test UMP para un problema dado (esta relación establece un orden parcial).

Ejemplos:

1. Sea $X \rightsquigarrow \mathcal{N}(\mu, 1)$. Formular el problema de contrastar $\mu = 15$ frente a $\mu = 18$ y resolverlo de forma intuitiva usando un test no aleatorizado basado en \bar{X} .
2. Sea (X_1, \dots, X_5) una m.a.s. de $X \rightsquigarrow \{B(1, p); p \in [0, 1]\}$. Construir un test para el contraste $H_0 : p = 1/2$ frente a $H_1 : p \neq 1/2$ con nivel de significación 0.1.

7.2. Test de Neyman-Pearson

Lema de Neyman-Pearson:

Sea $X \rightsquigarrow F \in \{F_\theta : \theta \in \Theta = \{\theta_0, \theta_1\}\}$, y sea f_θ la fmp o la fdp de X , según X sea discreta o continua. Por simplicidad, se denota $f_0 \equiv f_{\theta_0}$ y $f_1 \equiv f_{\theta_1}$. Sea X_1, \dots, X_n , una m.a.s. de X y se plantea el siguiente problema de contraste de hipótesis:

$$\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta = \theta_1. \end{cases}$$

a) Sea test φ un test de la forma

$$\varphi(X_1, \dots, X_n) = \begin{cases} 1 & \text{si } f_1^n(X_1, \dots, X_n) > k f_0^n(X_1, \dots, X_n), \\ \gamma(X_1, \dots, X_n) & \text{si } f_1^n(X_1, \dots, X_n) = k f_0^n(X_1, \dots, X_n), \\ 0 & \text{si } f_1^n(X_1, \dots, X_n) < k f_0^n(X_1, \dots, X_n), \end{cases}$$

para alguna constante $k \in \mathbb{R}^+ \cup \{0\}$, y $\gamma(x_1, \dots, x_n) \in [0, 1]$. Si $\varphi(X_1, \dots, X_n)$ tiene tamaño α , es de máxima potencia a nivel de significación α . Un test de esta forma se denomina test de Neyman-Pearson.

- b) Para todo $\alpha \in (0, 1]$ existe un test de Neyman-Pearson de tamaño α , con $\gamma(X_1, \dots, X_n) = \gamma$ constante.
- c) Si $\varphi'(X_1, \dots, X_n)$ es un test de tamaño $\alpha \in (0, 1]$ y es de máxima potencia a nivel de significación α , $\varphi'(X_1, \dots, X_n)$ es un test de Neyman-Pearson.
- d) El test de Neyman-Pearson de tamaño $\alpha \in (0, 1]$ con $\gamma(X_1, \dots, X_n) = \gamma$ constante es único con probabilidad uno bajo P_{θ_0} y bajo P_{θ_1} .
- e) El test de máxima potencia entre todos los de nivel de significación 0 (tamaño 0) es

$$\varphi_0(X_1, \dots, X_n) = \begin{cases} 1 & \text{si } f_0(X_1, \dots, X_n) = 0, \\ 0 & \text{si } f_0(X_1, \dots, X_n) > 0. \end{cases}$$

Ejemplos:

1. Sea (X_1, \dots, X_5) una m.a.s. de $X \rightsquigarrow B(1, p)$. Encontrar el test más potente de tamaño $\alpha = 0.05$ para resolver el problema de contraste:

$$\begin{cases} H_0 : p = 1/2 \\ H_1 : p = 3/4 \end{cases}$$

2. Sea (X_1, \dots, X_n) una m.a.s. de X cuya función de densidad viene dada por $f_\theta(x) = e^{\theta-x}$ $x \geq \theta$. Encontrar el test más potente de tamaño α para resolver el problema de contraste:

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

7.3. Test de la razón de verosimilitudes

Los test de la razón de verosimilitud están basados en el Principio de Máxima Verosimilitud. Sea $X \rightsquigarrow \{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^k\}$ y se considera el siguiente contraste: $\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}$ con $\Theta = \Theta_0 \cup \Theta_1$. Sea X_1, \dots, X_n una m.a.s. de X . Se puede construir la siguiente función:

$$\lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta_0} L_{x_1, \dots, x_n}(\theta)}{\sup_{\theta \in \Theta} L_{x_1, \dots, x_n}(\theta)} \quad \forall (x_1, \dots, x_n) \in \chi^n$$

donde $L_{x_1, \dots, x_n}(\theta) = f_\theta(x_1, \dots, x_n)$ es la función de verosimilitud asociada a la muestra, (x_1, \dots, x_n) . Es evidente que $0 \leq \lambda(x_1, \dots, x_n) \leq 1$ puesto que el denominador es el supremo en todo el espacio paramétrico, luego será mayor o igual que el numerador.

El *test de la razón de verosimilitud* (TRV) será:

$$\varphi(X_1, \dots, X_n) = \begin{cases} 1 & \lambda(X_1, \dots, X_n) < c \\ 0 & \lambda(X_1, \dots, X_n) \geq c \end{cases}$$

siendo $c \in (0, 1]$ una contante que se determina imponiendo el tamaño o el nivel de significación.

El TRV es un test no aleatorizado, por tanto habrá veces que se pueda calcular el tamaño y otras en que se tomará el más próximo. Si se opta por aleatorizar el test, ya no será el TRV, sino la aleatorización de él.

Propiedades del TRV

Si $T(X_1, \dots, X_n)$ es un estadístico suficiente para la familia $\{F_\theta : \theta \in \Theta\}$, entonces el TRV es función de $T(X_1, \dots, X_n)$.

Ejemplo: Sea X la variable aleatoria que mide el número de accidentes por semana, $X \rightsquigarrow P(\theta)$, $\theta \leq 2.5$. Después de modificaciones en la vía se desea comprobar si la media de accidentes sigue sin superar 2.5 para lo cual se realiza un estudio del número de accidentes en 4 semanas obteniendo los valores 4, 3, 2 y 3. Resolver el problema de contraste planteado al nivel de significación de 0.05.

7.4. Contrastes sobre los parámetros de una población normal

Sea $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, y X_1, \dots, X_n una m.a.s. Se pueden plantear los siguientes test según sean conocidos o no los parámetros de la distribución:

7.4 Contrastes sobre los parámetros de una población normal

- Test sobre la media de una distribución normal con varianza conocida:

Fijado un nivel de significación, α , se calcula el valor del estadístico

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0, 1) \Rightarrow z_{exp} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Contraste	Región Crítica del TRV de tamaño α Se rechaza H_0 a nivel α si
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$z_{exp} < -z_{\alpha/2}$ o $z_{exp} > z_{\alpha/2}$ $\alpha \in [0, 1]$ equivalentemente $ z_{exp} > z_{\alpha/2}$
$H_0 : \mu \leq \mu_0$ $H_1 : \mu > \mu_0$	$z_{exp} > z_{\alpha}$ $\alpha \leq 1/2$
$H_0 : \mu \geq \mu_0$ $H_1 : \mu < \mu_0$	$z_{exp} < -z_{\alpha}(z_{1-\alpha})$ $\alpha \leq 1/2$

(Ver páginas 5 y 6 del resumen)

- Test sobre la media de una distribución normal con varianza desconocida:

Fijado un nivel de significación, α , se calcula el valor del estadístico

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \rightsquigarrow t_{n-1} \Rightarrow t_{exp} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Contraste	Región Crítica del TRV de tamaño α Se rechaza H_0 a nivel α si
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$t_{exp} \leq -t_{n-1;\alpha/2}$ o $t_{exp} \geq t_{n-1;\alpha/2}$ $\alpha \in [0, 1]$ equivalentemente $ t_{exp} > t_{n-1;\alpha/2}$
$H_0 : \mu \leq \mu_0$ $H_1 : \mu > \mu_0$	$t_{exp} > t_{n-1;\alpha}$ $\alpha \leq 1/2$
$H_0 : \mu \geq \mu_0$ $H_1 : \mu < \mu_0$	$t_{exp} < -t_{n-1;\alpha}(t_{n-1;1-\alpha})$ $\alpha \leq 1/2$

(Ver páginas 7 y 8 del resumen)

- Test sobre la varianza de una distribución normal con media conocida:

Fijado un nivel de significación, α , se calcula el valor del estadístico

$$\chi^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \rightsquigarrow \chi_n^2 \Rightarrow \chi_{exp}^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2}$$

Contraste	Región Crítica del TRV de tamaño $\approx \alpha$ Se rechaza H_0 a nivel α si
$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	$\chi_{exp}^2 < \chi_{n;1-\alpha/2}^2$ o $\chi_{exp}^2 > \chi_{n;\alpha/2}^2$ $\alpha \in [0, 1]$
$H_0 : \sigma^2 \leq \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$	$\chi_{exp}^2 > \chi_{n;\alpha}^2$ $\alpha \leq P(Y > n) (Y \rightsquigarrow \chi^2(n))$
$H_0 : \sigma^2 \geq \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$	$\chi_{exp}^2 < \chi_{n;1-\alpha}^2$ $\alpha \leq P(Y \leq n) (Y \rightsquigarrow \chi^2(n))$

(Ver páginas 9 y 10 del resumen)

- Test sobre la varianza de una distribución normal con media desconocida:

Fijado un nivel de significación, α , se calcula el valor del estadístico

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2 \Rightarrow \chi_{exp}^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Contraste	Región Crítica del TRV de tamaño $\approx \alpha$ Se rechaza H_0 a nivel α si
$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	$\chi_{exp}^2 < \chi_{n-1;1-\alpha/2}^2$ o $\chi_{exp}^2 > \chi_{n-1;\alpha/2}^2$ $\alpha \in [0, 1]$
$H_0 : \sigma^2 \leq \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$	$\chi_{exp}^2 > \chi_{n-1;\alpha}^2$ $\alpha \leq P(Y > n) (Y \rightsquigarrow \chi^2(n-1))$
$H_0 : \sigma^2 \geq \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$	$\chi_{exp}^2 < \chi_{n-1;1-\alpha}^2$ $\alpha \leq P(Y \leq n) (Y \rightsquigarrow \chi^2(n-1))$

(Ver páginas 11 y 12 del resumen)

7.5. Contrastes sobre los parámetros de dos poblaciones normales

Sean $X \rightsquigarrow \mathcal{N}(\mu_1, \sigma_1^2)$ e $Y \rightsquigarrow \mathcal{N}(\mu_2, \sigma_2^2)$, y sus muestras aleatorias X_1, \dots, X_{n_1} e Y_1, \dots, Y_{n_2} . Se pueden plantear los siguientes test según sean conocidos o no los parámetros de las distribuciones:

- Test para comparar las medias de dos distribuciones normales con varianzas conocidas: Fijado un nivel de significación, α , se calcula el valor del estadístico

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0, 1) \Rightarrow z_{exp} = \frac{\bar{x} - \bar{y} - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

7.5 Contrastes sobre los parámetros de dos poblaciones normales

Contraste	Región Crítica Se rechaza H_0 a nivel α si
$H_0 : \mu_1 - \mu_2 = \mu_0$ $H_1 : \mu_1 - \mu_2 \neq \mu_0$	$z_{exp} < -z_{\alpha/2}$ o $z_{exp} > z_{\alpha/2}$ equivalentemente $ z_{exp} > z_{\alpha/2}$
$H_0 : \mu_1 - \mu_2 \leq \mu_0$ $H_1 : \mu_1 - \mu_2 > \mu_0$	$z_{exp} > z_{\alpha}$
$H_0 : \mu_1 - \mu_2 \geq \mu_0$ $H_1 : \mu_1 - \mu_2 < \mu_0$	$z_{exp} < -z_{\alpha}(z_{1-\alpha})$

- Test para comparar las medias de dos distribuciones normales con varianzas desconocidas pero iguales: Fijado un nivel de significación, α , se calcula el valor del estadístico

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow t_{n_1+n_2-2} \Rightarrow t_{exp} = \frac{\bar{x} - \bar{y} - \mu_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\text{con } S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}.$$

Contraste	Región Crítica Se rechaza H_0 a nivel α si
$H_0 : \mu_1 - \mu_2 = \mu_0$ $H_1 : \mu_x - \mu_y \neq \mu_0$	$t_{exp} < -t_{n_1+n_2-2;\alpha/2}$ o $t_{exp} > t_{n_1+n_2-2;\alpha/2}$ equivalentemente $ t_{exp} > t_{n_1+n_2-2;\alpha/2}$
$H_0 : \mu_1 - \mu_2 \leq \mu_0$ $H_1 : \mu_1 - \mu_2 > \mu_0$	$t_{exp} > t_{n_1+n_2-2;\alpha}$
$H_0 : \mu_1 - \mu_2 \geq \mu_0$ $H_1 : \mu_1 - \mu_2 < \mu_0$	$t_{exp} < -t_{n_1+n_2-2;\alpha}(t_{n_1+n_2-2;1-\alpha})$

- Test para comparar las varianzas de dos distribuciones normales con medias conocidas: Fijado un nivel de significación, α , se calcula el valor del estadístico

$$F = \frac{\sigma_1^2 \sum_{i=1}^{n_2} (Y_i - \mu_2)^2 / n_2}{\sigma_2^2 \sum_{i=1}^{n_1} (X_i - \mu_1)^2 / n_1} \rightsquigarrow F_{n_2, n_1} \Rightarrow F_{exp} = \frac{\sum_{i=1}^{n_2} (y_i - \mu_2)^2 / n_2}{\sum_{i=1}^{n_1} (x_i - \mu_1)^2 / n_1}$$

Contraste	Región Crítica Se rechaza H_0 a nivel α si
$H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 \neq \sigma_2^2$	$F_{exp} < F_{n_2, n_1; 1-\alpha/2}$ o $F_{exp} > F_{n_2, n_1; \alpha/2}$
$H_0 : \sigma_1^2 \leq \sigma_2^2$ $H_1 : \sigma_1^2 > \sigma_2^2$	$F_{exp} > F_{n_2, n_1; \alpha}$
$H_0 : \sigma_1^2 \geq \sigma_2^2$ $H_1 : \sigma_1^2 < \sigma_2^2$	$F_{exp} < F_{n_2, n_1; 1-\alpha}$

- Test para comparar las varianzas de dos distribuciones normales con medias desconocidas: Fijado un nivel de significación, α , se calcula el valor del estadístico

$$F = \frac{\sigma_1^2}{\sigma_2^2} \cdot \frac{S_2^2}{S_1^2} \rightsquigarrow F_{n_2-1, n_1-1} \Rightarrow F_{exp} = \frac{s_2^2}{s_1^2}$$

Contraste	Región Crítica Se rechaza H_0 a nivel α si
$H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 \neq \sigma_2^2$	$F_{exp} < F_{n_2-1, n_1-1; 1-\alpha/2}$ o $F_{exp} > F_{n_2-1, n_1-1; \alpha/2}$
$H_0 : \sigma_1^2 \leq \sigma_2^2$ $H_1 : \sigma_1^2 > \sigma_2^2$	$F_{exp} > F_{n_2-1, n_1-1; \alpha}$
$H_0 : \sigma_1^2 \geq \sigma_2^2$ $H_1 : \sigma_1^2 < \sigma_2^2$	$F_{exp} < F_{n_2-1, n_1-1; 1-\alpha}$

7.6. Dualidad entre estimación por intervalos y contraste de hipótesis

Se puede establecer una relación entre los intervalos de confianza para un problema y la región de aceptación si se resuelve el problema mediante un test.

Sea $X \rightsquigarrow \{F_\theta; \theta \in \Theta\}$ y X_1, \dots, X_n una m.a.s. de X . Según se ha visto en el tema anterior un problema de estimación puede resolverse mediante, un intervalo de confianza, ó de forma más general mediante una región de confianza, $(S(x_1, \dots, x_n))$, al nivel de confianza $1 - \alpha$, que es un subconjunto del espacio paramétrico tal que verifica:

$$P_\theta[S(X_1, \dots, X_n) \ni \theta] \geq 1 - \alpha \quad \forall \theta \in \Theta$$

Por otro lado, se ha visto en este tema que si se plantea el problema de estimación como un contraste de hipótesis del tipo $H_0 : \theta = \theta_0$ (frente a $H_1 : \theta \neq \theta_0$), al nivel de significación α , el problema se resuelve dando un test ó, equivalentemente, dando una región de rechazo, $\mathcal{C}(\theta_0)$. Dicha región es un subconjunto del espacio de posibles valores de la muestra \mathcal{X}^n , de forma que:

$$P_{\theta_0}[(X_1, \dots, X_n) \in \mathcal{C}(\theta_0)] = (\text{Error de tipo I}) \leq \alpha$$

El test también puede expresarse en función de la región complementaria a la de rechazo, es decir la región de aceptación que podemos denotar $A(\theta_0)$:

$$\varphi_{\theta_0}(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } (x_1, \dots, x_n) \notin \mathcal{A}(\theta_0) \\ 0 & \text{si } (x_1, \dots, x_n) \in \mathcal{A}(\theta_0) \end{cases}$$

$$P_{\theta_0}[(X_1, \dots, X_n) \in A(\theta_0)] \geq 1 - \alpha.$$

Teorema: Cada uno de los test $\varphi_{\theta_0}(X_1, \dots, X_n)$ aplicado al problema de contrastar $H_0 : \theta = \theta_0$ (frente a $H_1 : \theta \neq \theta_0$), tiene nivel de significación α si y sólo si $S(X_1, \dots, X_n)$ es una región de confianza para θ al nivel de confianza $1 - \alpha$.

Es decir, existe una relación entre ambas formas de resolver el problema de estimación.

■ Contraste de Hipótesis \rightarrow Intervalo de Confianza

Si se sabe resolver el problema de contraste de hipótesis, del tipo planteado anteriormente, para cada valor $\theta \in \Theta$, entonces se conocen conjuntos del tipo $A(\theta)$ verificado $P_{\theta_0}[(X_1, \dots, X_n) \in A(\theta_0)] \geq 1 - \alpha$. Se puede construir un intervalo de confianza, al nivel de confianza $1 - \alpha$, como sigue:

$$S(x_1, \dots, x_n) = \{\theta \in \Theta : (x_1, \dots, x_n) \in A(\theta)\}.$$

■ Intervalo de Confianza \rightarrow Contraste de Hipótesis

Si se sabe resolver el problema mediante un intervalo de confianza, $S(x_1, \dots, x_n)$, al nivel de confianza $1 - \alpha$ y se desea resolver el problema de contraste, del tipo planteado anteriormente, se construye la región de aceptación, al nivel de significación α , de la siguiente forma:

$$A(\theta_0) = \{(x_1, \dots, x_n) \in \mathcal{X}^n : S(x_1, \dots, x_n) \ni \theta_0\}$$

Ejemplo: Sea $X \rightsquigarrow \mathcal{N}(\mu, \sigma_0^2)$. Obtener el intervalo de confianza para μ al nivel de confianza $1 - \alpha$ a partir del test de razón de verosimilitud para el contraste $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$.

Siguiendo un razonamiento análogo al del ejemplo se puede estudiar la dualidad entre los intervalos de confianza para poblaciones normales, estudiados en el tema anterior, y los contrastes de hipótesis para esas mismas poblaciones, estudiados en este tema (ver pág. 13 del resumen).

Para los contrastes de igualdad de medias o de varianzas estudiados, resueltos por el test de razón de verosimilitud, se pueden encontrar también sus duales (ver pág. 14 del resumen). Por ejemplo, si se contrasta $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$, este contraste es equivalente al contraste $\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$, o de forma más general $\begin{cases} H_0 : \mu_1 - \mu_2 = \delta \\ H_1 : \mu_1 - \mu_2 \neq \delta \end{cases}$, con $\delta \in \mathbb{R}$, el cual se pueden resolver rechazando la hipótesis nula si δ no pertenece al intervalo de confianza de menor longitud esperada para la diferencia de medias. De forma equivalente, si se contrasta $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$, este contraste es equivalente a contrastar

$\left\{ \begin{array}{l} H_0 : \sigma_1^2/\sigma_2^2 = 1 \\ H_1 : \sigma_1^2/\sigma_2^2 \neq 1 \end{array} \right\}$ ó $\left\{ \begin{array}{l} H_0 : \sigma_2^2/\sigma_1^2 = 1 \\ H_1 : \sigma_2^2/\sigma_1^2 \neq 1 \end{array} \right\}$, los cuales pueden resolverse rechazando la hipótesis nula si 1 no pertenece al intervalo de confianza de menor longitud esperada para el cociente de varianzas adecuado.

Para los contrastes de hipótesis con alternativas unilaterales, es decir, para los contrastes del tipo $\left\{ \begin{array}{l} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{array} \right\}$ ó $\left\{ \begin{array}{l} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{array} \right\}$ también se pueden obtener los intervalos de confianza duales al problema de resolver el contraste de hipótesis planteado, sólo que en este caso los intervalos serán de tipo unilateral, $(a, +\infty)$ ó $(-\infty, b)$ (ver pág. 13 y 14 del resumen).

Ejemplo: Sea $X \rightsquigarrow \mathcal{N}(\mu, \sigma_0^2)$. Obtener el intervalo de confianza para μ al nivel de confianza $1 - \alpha$ a partir del test de razón de verosimilitud para el contraste $\left\{ \begin{array}{l} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{array} \right\}$.

7.7. Resolución de contrastes de hipótesis mediante el p – valor

Otra opción para resolver un problema de contraste hipótesis mediante un test no aleatorizado es usar el denominado p – nivel o p – valor asociado a los datos. Este método no depende del nivel de significación para el cual se desea resolver el contraste lo que lo diferencia de usar la región de rechazo ya que para determinar esta hace falta dicho valor.

Definición: El p – valor es el mínimo nivel de significación para el cual el valor del estadístico en la muestra cae en la región de rechazo siendo la hipótesis nula verdadera.

Si $T(X_1, \dots, X_n)$ es el estadístico en función del cual está definido el test que resuelve el contraste y $T_{exp} = T(x_1, \dots, x_n)$ es el valor del estadístico en la muestra observada, el p – valor se determina mediante el siguiente cálculo:

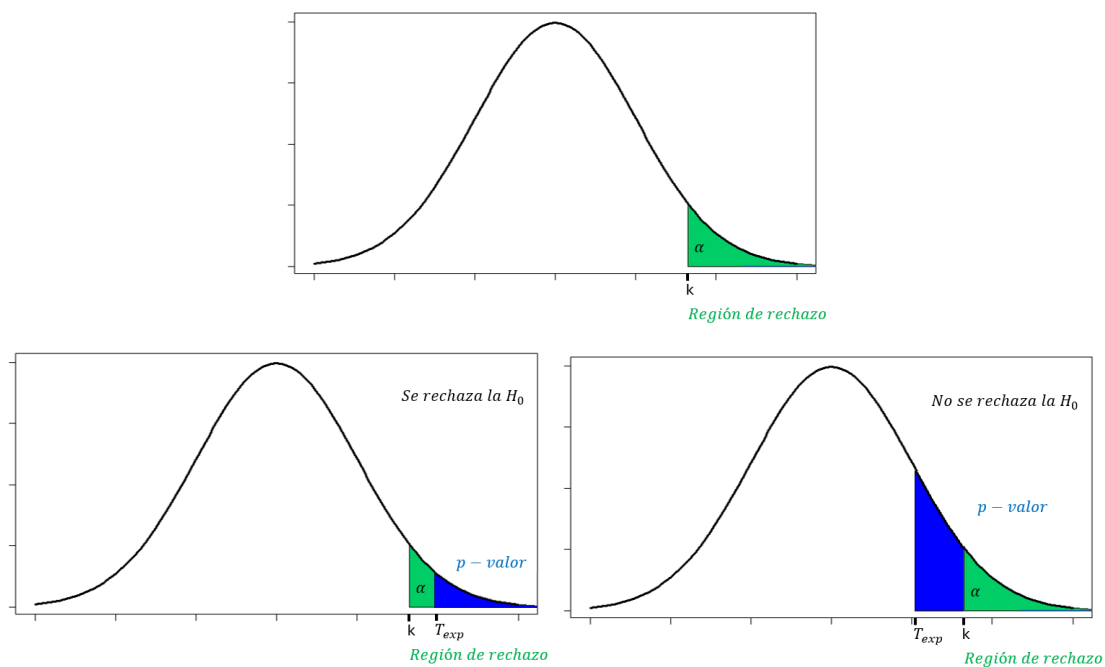
- Si la región de rechazo es de cola a la derecha,

$$T_{exp} \geq k \text{ con } P[T(X_1, \dots, X_n) \geq k] \leq \alpha,$$

se verifica

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

7.7 Resolución de contrastes de hipótesis mediante el p - valor

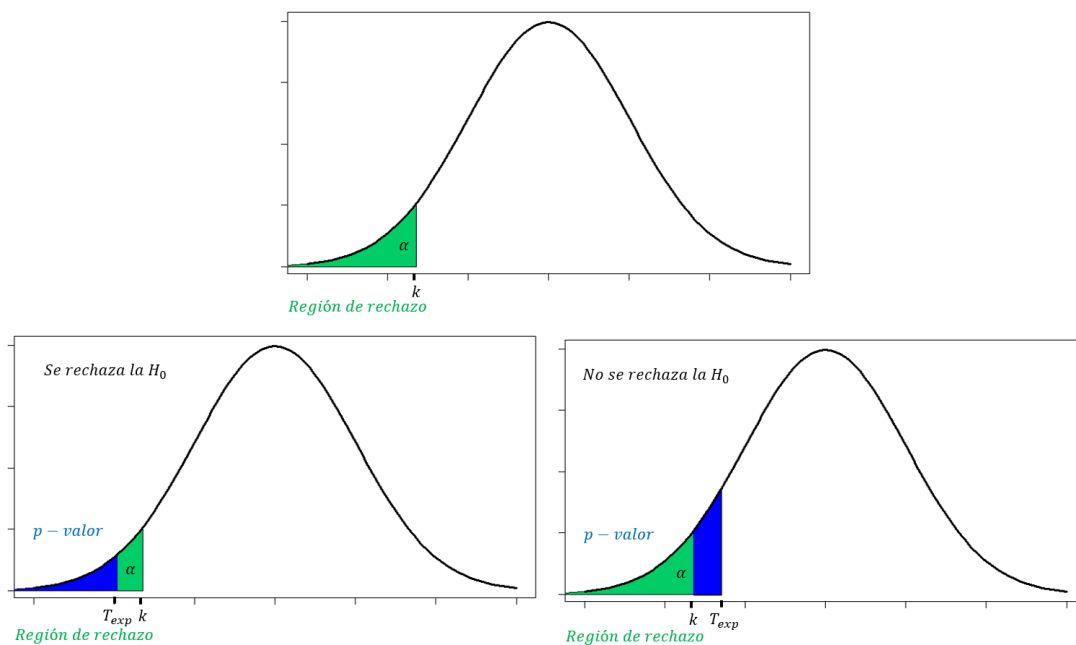


- Si la región de rechazo es de cola a la izquierda,

$$T_{exp} \leq k \text{ con } P[T(X_1, \dots, X_n) \leq k] \leq \alpha,$$

se verifica

$$p - valor = P_{H_0}(T \leq T_{exp}).$$



- Si la región de rechazo es de dos colas,

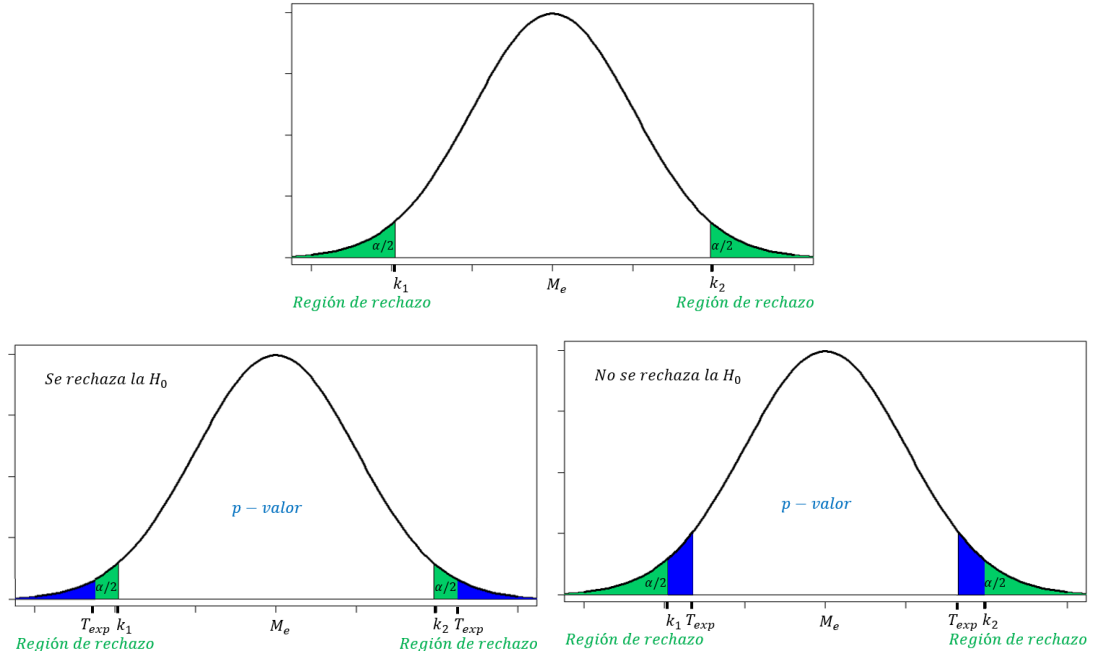
$$T_{exp} \leq k_1 \text{ o } T_{exp} \geq k_2, \text{ con } P[T(X_1, \dots, X_n) \leq k_1] + P[T(X_1, \dots, X_n) \geq k_2] \leq \alpha$$

se verifica

$$p - \text{valor} = \begin{cases} 2P_{H_0}[T(X_1, \dots, X_n) \geq T_{exp}] & \text{si } T_{exp} \geq Me \\ 2P_{H_0}[T(X_1, \dots, X_n) \leq T_{exp}] & \text{si } T_{exp} \leq Me \end{cases}$$

siendo Me la mediana de la distribución del estadístico bajo hipótesis nula

$$P_{H_0}[T(X_1, \dots, X_n) \leq Me] = \frac{1}{2}.$$



En general la forma de interpretar el $p - \text{valor}$ es que para niveles de significación menores que el p-valor, el test no rechaza H_0 mientras que para niveles de significación mayores o iguales, el test rechaza H_0 .

En la práctica, la conclusión del contraste en base al $p - \text{valor}$ sería la siguiente:

- Si el $p - \text{valor}$ es grande (mayor a 0.1), no se rechaza H_0 , ya que para rechazar en base al él se deben tomar niveles de significación muy altos, es decir, asumir un error de tipo I grande. En ese caso se suele indicar que el contraste no es significativo.

- Si el $p - valor$ es pequeño (menor a 0.01), se rechaza H_0 , ya que para los niveles de significación más estándar (0.01, 0.05 y 0.1) el $p - valor$ quedaría por debajo. En ese caso se suele indicar que el contraste es significativo para niveles de significación mayores o iguales a 0.01.
- Para valores intermedios, hay que tener en cuenta el nivel de significación con que se quiere trabajar.

Si se toma un nivel de significación de 0.05, para rechazar H_0 se debe obtener un $p - valor$ inferior a dicho nivel, es decir si el $p - valor < 0.05$ el test rechazaría H_0 . Si el $p - valor$ está entre 0.05 y 0.1, no se puede rechazar la H_0 pero como el $p - valor$ está cerca del valor de comparación, se suele indicar que el contraste está próximo a la significación. Si el $p - valor > 0.1$ el test no rechazaría H_0 .

Hay que tratar cada situación en particular aunque, normalmente, lo aconsejable en los casos donde se obtenga un $p - valor$ próximo a la significación por encima, sería poder tomar más datos y rehacer los cálculos para que el test pueda ser concluyente.

Tema 8

INTRODUCCIÓN A TEORÍA GENERAL DE MODELOS LINEALES: REGRESIÓN Y ANÁLISIS DE LA VARIANZA

8.1. Descripción del modelo lineal general. Modelo de Gauss-Markov.

La teoría de los modelos lineales proporciona las bases para tratar una gran cantidad de problemas reales en los que se pretende estudiar el comportamiento de una variable aleatoria en términos de otras variables (aleatorias o no).

Estadísticamente, se trata de hacer inferencia sobre los parámetros que definen las relaciones entre las variables con el fin de precisar dichas relaciones.

Este planteamiento se resuelve adecuadamente en el contexto del llamado Modelo Lineal General (GLM), bajo especificaciones y criterios de estimación apropiados.

Definición: Un *modelo lineal* es una expresión de la forma $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, donde $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$ es un vector (columna) aleatorio n -dimensional, observable, \mathbf{X} es una matriz de orden $n \times k$, con $k < n$, a que se denota por matriz de diseño, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^T$ es un vector desconocido, no observable, al que se conoce como vector de efectos y $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$ es un vector aleatorio, no observable, que se denomina vector de errores:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1k} \\ \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Si se denota por x_j a la columna j -ésima de la matriz \mathbf{X} , se puede escribir que $\mathbf{Y} = \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$, donde se puede comprobar que cada β_j pondera el efecto (influencia) de la columna x_j en \mathbf{Y} , de ahí que se denomine a $\boldsymbol{\beta}$ vector de efectos.

Los modelos lineales se pueden clasificar, atendiendo a que el vector $\boldsymbol{\beta}$ sea no aleatorio o aleatorio, en modelos de efectos fijos y modelos de efectos aleatorios. En el caso de los modelos fijos, los cuales centrarán nuestro estudio, al ser \mathbf{X} conocida y $\boldsymbol{\beta}$ no aleatorio, ε es una perturbación aleatoria en la que se engloban todos los factores aleatorios que intervienen en el comportamiento de \mathbf{Y} .

La teoría de modelos se puede generalizar al caso en que \mathbf{X} sea aleatoria, y atendiendo al rango de \mathbf{X} , los modelos lineales se pueden clasificar en modelos de rango máximo o completo, si $rg(\mathbf{X}) = k$, y modelos lineales de rango no máximo, si $rg(\mathbf{X}) < k$.

Dentro de los modelos lineales, este tema se va a centrar en el estudio del modelo de Gauss-Markov, es decir, un modelo lineal de efectos fijos y tal que las componentes del vector $\boldsymbol{\varepsilon}$ sean variables aleatorias de segundo orden, centradas, incorreladas y homocedásticas, es decir, $\forall i = 1, \dots, n$ se verifica: $E[\varepsilon_i] = 0$, $E[\varepsilon_i^2] = \sigma^2$ y $E[\varepsilon_i \varepsilon_j] = 0$ $i \neq j$.

Las tres condiciones impuestas al vector $\boldsymbol{\varepsilon}$ se pueden resumir en las dos siguientes: $E[\boldsymbol{\varepsilon}] = 0$ y $E[\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T] = \sigma^2 I_n$.

El objetivo del estudio de estos modelos será hacer inferencia sobre las componentes indeterminadas del modelo: el vector de efectos $\boldsymbol{\beta}$ y σ^2 a partir de una observación del vector \mathbf{Y} .

8.2. Estimación de un modelo de Gauss-Markov

Sea $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ un modelo lineal de Gauss-Markov, es decir verificando $E[\boldsymbol{\varepsilon}] = 0$ y $E[\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T] = \sigma^2 I_n$. Cada componente del vector \mathbf{Y} verifica

$$Y_i = \sum_{j=1}^k x_{ij} \beta_j + \varepsilon_i, \quad i = 1, \dots, n \text{ con } E[\varepsilon_i] = 0, \quad Var[\varepsilon_i] = \sigma^2 \text{ y } Cov[\varepsilon_i, \varepsilon_j] = 0, \quad i \neq j.$$

Para estimar el modelo hay que estimar el vector de efectos del mismo y la varianza.

8.2.1. Estimación mínimo cuadrática del vector $\boldsymbol{\beta}$

El método de mínimos cuadrados consiste en minimizar la suma de los cuadrados de los errores cometidos al aproximar \mathbf{Y} por $\mathbf{X}\boldsymbol{\beta}$, esto es, se trata de minimizar la función

$$S^2(\boldsymbol{\beta}) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k x_{ij} \beta_j \right)^2 = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

Para ello se deriva la función $S^2(\boldsymbol{\beta})$ con respecto a cada una de las componentes del vector $\boldsymbol{\beta}$

$$\frac{\partial \sum_{i=1}^n \varepsilon_i^2}{\partial \beta_h} = -2 \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k x_{ij} \beta_j \right) x_{ih}, \quad h = 1, \dots, k.$$

Igualando a cero estas derivadas se obtienen las denominadas ecuaciones normales:

$$\sum_{i=1}^n Y_i x_{ih} = \sum_{i=1}^n \sum_{j=1}^k x_{ij} x_{ih} \beta_j, \quad h = 1, \dots, k,$$

que matricialmente se puede expresar como $\mathbf{X}^T \mathbf{Y} = (\mathbf{X}^T \mathbf{X}) \boldsymbol{\beta}$.

Cualquier solución de las ecuaciones normales que de un mínimo absoluto de $S^2(\boldsymbol{\beta})$ se denomina un estimador de mínimos cuadrados de $\boldsymbol{\beta}$ y se denota $\hat{\boldsymbol{\beta}}(\mathbf{Y}) = (\hat{\beta}_1(\mathbf{Y}), \dots, \hat{\beta}_k(\mathbf{Y}))$ o para simplificar $\hat{\boldsymbol{\beta}}$.

Propiedades:

- Existencia: existe, al menos un estimador de mínimos cuadrados de $\boldsymbol{\beta}$.
- Unicidad: no está garantizada, salvo si el modelo es de rango máximo en cuyo caso el único estimador mínimo cuadrático viene dado por:

$$\hat{\boldsymbol{\beta}} := \hat{\boldsymbol{\beta}}(\mathbf{Y}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

el cual es una función lineal de \mathbf{Y} con

$$E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta} \quad \text{Cov}[\hat{\boldsymbol{\beta}}] = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}.$$

8.2.2. Estimación de funciones lineales de $\boldsymbol{\beta}$. Teorema de Gauss-Markov

Definición: Una función escalar de las componentes de $\boldsymbol{\beta}$, $\psi(\boldsymbol{\beta})$, se dice que es *estimable* si admite un estimador insesgado función lineal de las componentes de \mathbf{Y} :

$$\psi(\boldsymbol{\beta}) \text{ estimable} \Leftrightarrow \exists \mathbf{c} \in \mathbb{R}^n / E[\mathbf{c}^T \mathbf{Y}] = \psi(\boldsymbol{\beta}), \quad \forall \boldsymbol{\beta}.$$

Caracterización de funciones estimables:

- $\psi(\boldsymbol{\beta})$ es estimable $\Leftrightarrow \psi(\boldsymbol{\beta}) = \mathbf{c}^T \mathbf{X} \boldsymbol{\beta}$, $\mathbf{c} \in \mathbb{R}^n$.
- Si el modelo es de rango máximo, $\psi(\boldsymbol{\beta})$ es estimable $\Leftrightarrow \psi(\boldsymbol{\beta}) = \mathbf{a}^T \boldsymbol{\beta}$, $\mathbf{a} \in \mathbb{R}^k$.

En particular, cualquier componente del vector $\mathbf{Y} = \mathbf{X} \boldsymbol{\beta}$ es estimable ya que

$$Y_i = (\mathbf{X} \boldsymbol{\beta})_i = (0, \dots, 1, \dots, 0) \mathbf{X} \boldsymbol{\beta}.$$

Teorema de Gauss-Markov: Toda función estimable, $\mathbf{a}^T \boldsymbol{\beta}$, admite un único estimador insesgado uniformemente de mínima varianza en la clase de estimadores lineales insesgados. Dicho estimador es $\mathbf{a}^T \hat{\boldsymbol{\beta}}(\mathbf{Y})$, y se denomina estimador de mínimos cuadrados de $\mathbf{a}^T \boldsymbol{\beta}$.

Ya que cada componente de $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$ es estimable, por el Teorema de Gauss-Markov, existe su estimador de mínimos cuadrados y es la correspondiente componente del vector $\mathbf{X}\hat{\boldsymbol{\beta}}$.

$$Y_i = (\mathbf{X}\boldsymbol{\beta})_i = (0, \dots, 1, \dots, 0)\mathbf{X}\boldsymbol{\beta} \Rightarrow \hat{Y}_i = \widehat{(\mathbf{X}\boldsymbol{\beta})}_i = (0, \dots, 1, \dots, 0)\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{X}\hat{\boldsymbol{\beta}})_i.$$

Puesto que estos estimadores son insesgados se tiene $E[\hat{\mathbf{Y}}] = E[\mathbf{X}\hat{\boldsymbol{\beta}}] = \mathbf{X}\boldsymbol{\beta}$.

8.2.3. Modelo estimado, residuos y estimación de la varianza

Según acabamos de ver, si denotamos por $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_k)^T$ a $\hat{\boldsymbol{\beta}}(\mathbf{Y})$, el estimador de mínimos cuadrados de $\boldsymbol{\beta}$, la estimación del modelo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$ viene dada por

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \Rightarrow \hat{Y}_i = \sum_{j=1}^k x_{ij}\hat{\beta}_j, \quad i = 1, \dots, n.$$

Además, \hat{Y}_i , $i = 1, \dots, n$, es el estimador lineal insesgado de mínima varianza de $E[Y_i] = \sum_{j=1}^k x_{ij}\beta_j$ al ser el estimador de mínimos cuadrados y ser insesgado en dicha función.

Definición: Se define el *vector de residuos mínimo cuadráticos* como el que resulta de hacer la diferencia entre el verdadero valor del modelo y el aproximado según el modelo estimado:

$$\mathbf{R} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

cuyas componentes son $\mathbf{R}_i = \mathbf{Y}_i - \hat{\mathbf{Y}}_i$, $i = 1, \dots, n$.

Propiedades:

- Los residuos son variables aleatorias con media nula, $E[R_i] = 0$, $\forall i = 1, \dots, n$, ya que

$$E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta} \text{ y } E[\hat{\mathbf{Y}}] = \mathbf{X}\boldsymbol{\beta}.$$

- El vector de residuos es ortogonal a los vectores columna de \mathbf{X} , $\mathbf{X}^T \mathbf{R} = 0$, ya que al ser $\hat{\boldsymbol{\beta}}$ un estimador de mínimos cuadrados, verifica el sistema de ecuaciones normales $\mathbf{X}^T \mathbf{Y} = (\mathbf{X}^T \mathbf{X})\boldsymbol{\beta}$.
- El vector de residuos es ortogonal al vector estimado, $\hat{\mathbf{Y}}^T \mathbf{R} = 0$. Esto se deduce de la propiedad anterior y de que $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

Definición: Se define la *varianza residual* como el cociente entre la suma de cuadrados de los residuos y el número de residuos linealmente independientes.

$$S_R^2 = \frac{\sum_{i=1}^n R_i^2}{n-r} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-r} = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n-r} = \frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{n-r} = \frac{\|\mathbf{R}\|^2}{n-r}$$

El número de residuos linealmente independientes viene determinado por el rango de la matriz \mathbf{X} , $r = \text{Rango}(\mathbf{X})$, ya que al verificarse que $\mathbf{X}^T \mathbf{R} = 0$, las k columnas de \mathbf{X} determinan k relaciones lineales entre los residuos igualadas a cero. Por lo tanto, el rango de \mathbf{X} indica cuantas de dichas relaciones no son proporcionales y, por tanto, cuantas son linealmente dependientes. De ahí que el número de residuos linealmente independientes es $n - r$.

Además, puede probarse que la varianza residual, S_r^2 , es un estimador insesgado de σ^2 .

8.3. Inferencia bajo hipótesis de normalidad

8.3.1. Distribución normal N-dimensional

Sean

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{pmatrix}$$

Se dice que el vector aleatorio $X = (X_1, \dots, X_n)^T$ tiene una distribución normal n -dimensional de media μ y matriz de varianzas-covarianzas Σ , definida positiva, y se notará $\mathcal{N}_n(\mu, \Sigma)$, si y sólo si su función de densidad es

$$f(x) = \frac{1}{(2\pi)^{n/2} [\det(\Sigma)]^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}, \quad x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$$

Propiedades

Si $X \rightsquigarrow \mathcal{N}_n(\mu, \Sigma)$, $X = (X_1, \dots, X_n)^T$

1. $E[X] = \mu$ y $Cov[X] = \Sigma$.
2. Su función generatriz de momentos será: $M(t) = e^{t^T \mu + \frac{t^T \Sigma t}{2}}$.
3. Las distribuciones marginales de cualquier dimensión son normales y, en particular, $X_i \rightsquigarrow \mathcal{N}(\mu_i, \sigma_{ii})$.

4. Si $\gamma_{n \times 1}$ es un vector constante $\Rightarrow X + \gamma \sim N_n(\mu + \gamma, \Sigma)$.
5. X_1, \dots, X_n son independientes $\Leftrightarrow \Sigma$ es diagonal, es decir, $\sigma_{ij} = 0, \forall i \neq j$,

8.3.2. Inferencia

Si en un modelo de Gauss-Markov añadimos la condición de que los residuos sean normales, además de centrados, independientes y homocedásticos, es decir si consideramos el modelo:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ con } \boldsymbol{\varepsilon} \rightsquigarrow \mathcal{N}_n(0, \sigma^2 I_{n \times n})$$

estamos considerando de forma equivalente que cada componente del vector \mathbf{Y} sigue una distribución normal unidimensional

$$Y_i = \sum_{j=1}^k x_{ij}\beta_j + \varepsilon_i \rightsquigarrow \mathcal{N}\left(\sum_{j=1}^k x_{ij}\beta_j, \sigma^2\right)$$

y que las componentes del vector $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ son independientes. Por lo tanto

$$\mathbf{Y} \rightsquigarrow \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I_{n \times n})$$

Estimadores máximo verosímiles

Para determinar los estimadores máximo verosímiles del modelo, es decir del vector de efectos, $\boldsymbol{\beta}$, y del parámetro σ^2 , hay que determinar primero la función de verosimilitud:

$$\mathbf{y} \in \mathbb{R}^n \rightarrow L_{\mathbf{y}}(\boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2}{2\sigma^2} \right\}.$$

- Puesto que maximizar $L_{\mathbf{y}}(\boldsymbol{\beta}, \sigma^2)$ en $\boldsymbol{\beta}$ es equivalente a mín $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$ en ese mismo parámetro, el estimador máximo verosímil de $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, será el estimador mínimo cuadrático.
- Para obtener el estimador máximo verosímil de σ^2 hay que, tras tomar logaritmo de la verosimilitud, resolver:

$$\frac{\partial \ln L_{\mathbf{y}}(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = 0,$$

de donde se deduce que:

$$\hat{\sigma}^2(\mathbf{y}) = \frac{\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{n} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} = \frac{\sum_{i=1}^n R_i^2}{n} = \frac{(n-r)S_R^2}{n}.$$

Test de razón de verosimilitud para la hipótesis lineal general

Definición: Se denomina *hipótesis lineal general* a la hipótesis formulada por $H_0 : \mathbf{C}\boldsymbol{\beta} = 0$ siendo $\mathbf{C}_{q \times k}$ una matriz conocida de rango $q (\leq k)$.

$$\begin{cases} c_{11}\beta_1 + \cdots + c_{1k}\beta_k = 0 \\ \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ c_{q1}\beta_1 + \cdots + c_{qk}\beta_k = 0 \end{cases}$$

Si todas las componentes del vector $\mathbf{C}\boldsymbol{\beta}$ son estimables, entonces el test de razón de verosimilitud, de tamaño α , que resuelve el contraste

$$\begin{cases} H_0 : \mathbf{C}\boldsymbol{\beta} = 0 \\ H_1 : \mathbf{C}\boldsymbol{\beta} \neq 0 \end{cases}$$

viene dado por

$$\varphi(\mathbf{Y}) = \begin{cases} 1 & F(\mathbf{Y}) > F_{q,n-r;\alpha} \\ 0 & F(\mathbf{Y}) \leq F_{q,n-r;\alpha} \end{cases}$$

donde

$$F(\mathbf{Y}) = \frac{n-r}{q} \left(\frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}^0\|^2 - \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2} \right)$$

siendo $\hat{\boldsymbol{\beta}}^0$ el estimador máximo verosímil de $\boldsymbol{\beta}$ bajo H_0 .

Si se desea resolver el contraste sin establecer previamente el nivel de significación con el que se desea trabajar, se debe calcular:

$$p - \text{valor} = P[F_{q,n-r} > F_{exp}] \text{ siendo } F_{exp} = \frac{n-r}{q} \left(\frac{\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^0\|^2 - \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2} \right).$$

8.4. Modelo de regresión lineal simple

El problema de regresión consiste en determinar una función que aplicada a una serie de variables aleatorias X_1, \dots, X_k permita predecir, con la mayor precisión posible, los valores de otra variable Y según los valores de X_1, \dots, X_k : $Y = \varphi(X_1, \dots, X_k) + E$ (E debe ser una variable aleatoria que expresa el error cometido al predecir Y por $\varphi(X_1, \dots, X_k)$).

Las variables X_1, \dots, X_k son variables explicativas, independientes o regresoras que fijamos al principio. La variable Y es la variable explicada, dependiente o de respuesta. Además, como se trata de un modelo univariante, se considera que la variable Y es unidimensional. Finalmente, al considerarse un modelo de regresión lineal, se asume que φ es

lineal, es decir, X_1, \dots, X_k influyen en la respuesta de forma lineal. El resto de los factores que pueden influir en Y se engloban en E .

Al tratarse de un modelo simple, el problema se centra en el caso $k = 1$, es decir hay una única variable explicativa X y una variable explicada Y , ambas unidimensionales, y se asume el modelo de regresión lineal simple:

$$Y = \beta_0 + \beta_1 X + E$$

y se resolverá el problema de regresión desde el punto de vista empírico, basándose en observaciones de la variable, (x_1, \dots, x_n) . El objetivo va a ser estimar el modelo a partir de las observaciones de la variable Y correspondiente a cada una de las x_i .

8.4.1. Planteamiento del modelo

Sean X e Y dos variables aleatorias tales que $E[Y^2] < +\infty$ y, fijado un valor arbitrario, $X = x$ se cumple:

$$E[Y/X = x] = \beta_0 + \beta_1 x, \quad \text{Var}[Y/X = x] = \sigma^2.$$

Sean x_1, \dots, x_n un conjunto de valores de X (se exigen al menos dos valores distintos). Supongamos que para cada $i = 1, \dots, n$, Y_i es la observación aleatoria de la variable Y supuesto que $X = x_i$. Entonces $\forall i = 1, \dots, n$

$$E[Y_i] = E[Y/X = x_i] = \beta_0 + \beta_1 x_i, \quad \text{Var}[Y_i] = \text{Var}[Y/X = x_i] = \sigma^2$$

Por tanto, $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ con $E[\varepsilon_i] = 0$ y $\text{Var}[\varepsilon_i] = \sigma^2$, $i = 1, \dots, n$. Es decir, se tiene

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

que es un modelo lineal de rango máximo, al exigir que $\exists x_i \neq x_j$. Si las observaciones se toman de forma independiente, entonces las variables $\varepsilon_1, \dots, \varepsilon_n$ son independientes, y centradas, por lo que el modelo es de Gauss-Markov de rango máximo.

8.4.2. Estimación de los parámetros del modelo

Como el modelo es de Gauss-Markov de rango máximo, sabemos que es estimador mínimo cuadrático de β es $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$, de donde se puede ver que la estimación de los parámetros del modelo, β_0 y β_1 , es:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} \quad \hat{\beta}_1 = \frac{\sigma_{xY}}{\sigma_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Propiedades:

1. Los estimadores mínimos cuadrados son único (por ser el modelo de rango máximo) y son funciones lineales de las componentes de \mathbf{Y} .
2. Los estimadores mínimo cuadrados son insesgados: $E[\hat{\beta}_0] = \beta_0$ y $E[\hat{\beta}_1] = \beta_1$.
3. $Var[\hat{\beta}_0] = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{n\sigma_x^2} \right)$ y $Var[\hat{\beta}_1] = \sigma^2 \frac{1}{n\sigma_x^2}$.
4. $Cov[\hat{\beta}_0, \hat{\beta}_1] = -\sigma^2 \frac{\bar{x}}{n\sigma_x^2}$.
5. La estimación del modelo es $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, por lo tanto, la estimación de cada componente $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ es

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \bar{Y} + \frac{\sigma_{xY}}{\sigma_x^2} (x_i - \bar{x}), \quad i = 1, \dots, n.$$

6. El vector de residuos, $\mathbf{R} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, en este caso está compuesto por

$$R_i = Y_i - \hat{Y}_i = Y_i - \bar{Y} - \frac{\sigma_{xY}}{\sigma_x^2} (x_i - \bar{x}), \quad i = 1, \dots, n.$$

Además, sabemos $\mathbf{X}^T \mathbf{R} = 0$ y $\hat{\mathbf{Y}}^T \mathbf{R} = 0$, por lo tanto:

- $\mathbf{X}^T \mathbf{R} = 0 \Rightarrow \sum_{i=1}^n R_i = 0 \Rightarrow \sum_{i=1}^n R_i/n = 0 \Rightarrow \sum_{i=1}^n \hat{Y}_i/n = \bar{Y}$
- $\hat{\mathbf{Y}}^T \mathbf{R} = \sum_{i=1}^n \hat{Y}_i R_i = 0$

7. El estimador de σ^2 es la varianza residual:

$$\hat{\sigma}^2 = S_R^2 = \frac{\sum_{i=1}^n R_i^2}{n-2} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2} = \frac{\sum_{i=1}^n (Y_i - \bar{Y} - \frac{\sigma_{xY}}{\sigma_x^2} (x_i - \bar{x}))^2}{n-2}.$$

8.4.3. Análisis de la bondad

La bondad del modelo ajustado depende en magnitud de la componente residual. Ésta es una variable aleatoria, por tanto, una forma de medirla es a través de la varianza residual S_R^2 , que no debe olvidarse que es la estimación de σ^2 .

El inconveniente que presenta esta medida de la bondad es que tiene dimensión y, por tanto, no es válida para hacer comparaciones en general. Esto motiva la búsqueda de otra medida de la bondad que sea adimensional. Para ello se estudia la descomposición de la variabilidad.

Descomposición de la variabilidad

La variabilidad total (VT) de las observaciones aleatorias Y_1, \dots, Y_n viene determinada por $\sum_{i=1}^n (Y_i - \bar{Y})^2 = n\sigma_Y^2$. Esta VT se puede descomponer en dos términos:

- La variabilidad explicada (VE), que es la medida de la variabilidad de las estimaciones $\hat{Y}_1, \dots, \hat{Y}_n$: $\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = \frac{n\sigma_{xY}^2}{\sigma_x^2}$.
- La variabilidad no explicada (VNE), que es la variabilidad de la parte residual: $\sum_{i=1}^n R_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = (n-2)S_R^2$.

La descomposición de la variabilidad, por tanto, viene dada por: VT=VE+VNE.

Coefficiente de determinación

El coeficiente de determinación se define como la proporción de la variabilidad total explicada por el modelo de regresión:

$$R^2 = \frac{\text{VE}}{\text{VT}} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{n\sigma_{xY}^2}{\sigma_x^2 n\sigma_Y^2} = \frac{\sigma_{xY}^2}{\sigma_x^2 \sigma_Y^2}$$

Propiedades:

1. R^2 es adimensional, por lo tanto es una medida de la bondad adecuada para hacer comparaciones.
2. $0 \leq R^2 \leq 1$.
3. Cuanto mayor sea R^2 , menores serán los residuos y, por tanto, mejor será el modelo.

8.4.4. Predicción con el modelo estimado

Utilizando el modelo estimado puede obtenerse una predicción para Y basada en el valor que toma X . Para ello, teniendo en cuenta que si $X = x_p$ entonces

$$Y_p = \beta_0 + \beta_1 x_p + \varepsilon_p; \quad E[\varepsilon_p] = 0, \quad \text{Var}[\varepsilon_p] = \sigma^2$$

se tiene que la predicción de Y para $X = x_p$ según el modelo estimado viene dada por:

$$\hat{Y}_p = \hat{\beta}_0 + \hat{\beta}_1 x_p = \bar{Y} + \frac{\sigma_{xY}}{\sigma_x^2} (x_p - \bar{x}).$$

Propiedades:

- $E[\hat{Y}_p] = E[Y_p]$.
- $Var[\hat{Y}_p] = \sigma^2 \left[\frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\sigma_x^2} \right]$

Para medir la bondad de las predicciones se va a utilizar el error cuadrático medio.

$$ECM = E[\hat{Y}_p - Y_p]^2 = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\sigma_x^2} \right]$$

A saber, el ECM será menor cuanto más se aproxime x_p a \bar{x} .

8.4.5. Aplicación práctica: recta de regresión estimada

La recta de regresión estimada a partir de $(x_1, y_1), \dots, (x_n, y_n)$, observaciones independientes del vector aleatorio (X, Y) , se obtiene a partir del modelo de regresión teórico, obtenido anteriormente, con los valores x_1, \dots, x_n :

$$\hat{Y}_i = \bar{Y} + \frac{\sigma_{xy}}{\sigma_x^2}(x_i - \bar{x})$$

La estimación concreta del modelo, para $Y_1 = y_1, \dots, Y_n = y_n$, viene dada por

$$\hat{y}_i = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x_i - \bar{x}), \quad i = 1 \dots, n.$$

Por tanto, la recta de regresión estimada a partir de las observaciones (x_i, y_i) será

$$y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x}).$$

La recta de regresión teórica sería

$$y = E[Y] + \frac{Cov[X, Y]}{Var[X]}(x - E[X]).$$

Coefficiente de determinación

La estimación del coeficiente de determinación lineal R^2 es

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2}$$

y su correspondiente poblacional viene dado por

$$\rho_{XY}^2 = \frac{Cov^2[X, Y]}{Var[X]Var[Y]}.$$

Predicción

La recta de regresión estimada permite predecir valores de la variable Y a partir de la variable X con valores que no han intervenido en la predicción. Para ello basta con considerar un valor arbitrario x_p de X , y la recta de regresión estimada a partir de las observaciones $(x_1, y_1), \dots, (x_n, y_n)$ proporcionará una predicción de Y :

$$\hat{y}_p = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2} (x_p - \bar{x}).$$

Puesto que para medir la bondad de las predicciones se usa el ECM que depende de σ^2 , si lo desconocemos se puede estimar el ECM estimando σ^2 por la varianza residual s_R^2 :

$$\widehat{ECM}(\hat{Y}_p) = s_R^2 \left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\sigma_x^2} \right].$$

8.4.6. Contraste de regresión bajo hipótesis de normalidad

El contraste de regresión trata de resolver el problema de ver si la variable X ejerce o no influencia lineal sobre la variable Y , es decir, se quiere contrastar que la variable X no tiene inferencia lineal sobre la variable Y . Para ello se plantea el contraste de hipótesis nula

$$H_0 : \beta_1 = 0$$

en el modelo $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1, \dots, n$, modelo lineal con hipótesis de normalidad ($\varepsilon_i \rightsquigarrow \mathcal{N}(0, \sigma^2)$) de rango máximo.

Para resolver este contraste se aplica el test de la razón de verosimilitud (TRV) para la hipótesis lineal general, ya que el modelo de regresión lineal es un MLG de rango máximo.

En este caso se tiene $H_0 : \mathbf{C}\boldsymbol{\beta} = 0$ con $\mathbf{C} = (0, 1)$ y $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$. Luego $q = \text{rang}(\mathbf{C}) = 1$, $r = \text{rang}(\mathbf{X}) = 2$ y, según se estudió anteriormente, el TRV sería:

$$\varphi(\mathbf{Y}) = \begin{cases} 1 & F(\mathbf{Y}) > F_{q, n-r; \alpha} = F_{1, n-2; \alpha} \\ 0 & F(\mathbf{Y}) \leq F_{q, n-r; \alpha} = F_{1, n-2; \alpha} \end{cases}$$

con

$$F(\mathbf{Y}) = \frac{n-r}{q} \left(\frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}^0\|^2 - \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2} \right) = (n-2) \frac{VT - VNE}{VNE} = \frac{VE}{VNE/(n-2)} = \frac{VE}{S_R^2}.$$

Puesto que $R^2 = VE/VT$, $F(\mathbf{Y}) = (n-2) \frac{R^2}{1-R^2}$, creciente en R^2 . Por lo tanto, grandes valores de R^2 conducen al rechazo de H_0 , lo que concuerda con la definición de R^2 como medida de la bondad del modelo estimado.

Si se desea resolver el contraste sin establecer previamente el nivel de significación con el que se desea trabajar, se debe calcular:

$$p - \text{valor} = P[F_{1,n-2} > F_{exp}] \text{ siendo } F_{exp} = \frac{ve}{s_R^2} = \frac{n\sigma_{xy}^2/\sigma_x^2}{s_R^2} = (n-2) \frac{r^2}{1-r^2}.$$

8.5. Análisis de la varianza de una vía

Si se desea medir el efecto de un supuesto factor de variación sobre el comportamiento de una variable aleatoria se consideran muestras de la variable aleatoria bajo distintos niveles del factor de variación y se descompone la variabilidad total de los datos en dos sumandos: uno que exprese la variabilidad entre las muestras y otro que exprese la variabilidad de los datos por ser observaciones de una variable aleatoria.

Ejemplo: Si se desea estudiar como afecta el tipo de alimentación en el peso de ciertos individuos, el factor de variabilidad sería el tipo de alimentación y la variable aleatoria el peso. Para dicho estudio se mediría el peso de los individuos, tras dividirlos en k grupos, y dar a cada grupo un tipo de alimentación diferente. La variabilidad que se apreciaría en los datos obtenidos puede venir dada por ser observaciones de una variable aleatoria o por ser individuos de distintos grupos.

El análisis de la varianza de una vía estudia esta variabilidad. Para ello se asumen dos supuestos, dados por Fisher:

- i) La variable de interés no está afectada por factores distintos del que es objeto de estudio.
- ii) En cada uno de los k niveles del factor de variación, la variable aleatoria se distribuye normalmente y con varianza común (homocedasticidad).

El problema de analizar la variabilidad se puede estudiar mediante el contraste del efecto de variación en los k grupos lo que se puede reducir a comparar las medias de k poblaciones normales con varianza común. Para ello se considera $(Y_{i1}, \dots, Y_{in_i})$ m.a.s. de $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, $i = 1, \dots, k$, donde cada Y_i es la variable aleatoria que esta influida por el i -ésimo factor de variación. Las muestras de cada variable Y_i deben ser independientes entre si.

Cada una de las variables de la muestra se puede descomponer en dos sumandos $Y_{ij} = \mu_i + \varepsilon_{ij}$, siendo ε_{ij} una variable aleatoria con distribución $\mathcal{N}(0, \sigma^2)$, $i = 1, \dots, k$; $j = 1, \dots, n_i$; $n = \sum_{i=1}^k n_i$. De esta forma, considerando el vector \mathbf{Y} , la matriz \mathbf{X} de rango k , el vector $\boldsymbol{\mu}$ y el vector $\boldsymbol{\varepsilon}$ siguientes, el problema de análisis de la varianza de una vía se puede ver como un modelo lineal general. Es más, el modelo $\mathbf{Y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$ es un modelo de Gauss-Markov de rango máximo, k .

$$\mathbf{Y} = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ Y_{k1} \\ \vdots \\ Y_{kn_k} \end{pmatrix}_{n \times 1} \quad \mathbf{X} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}_{n \times k} \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}_{k \times 1} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \\ \vdots \\ \varepsilon_{k1} \\ \vdots \\ \varepsilon_{kn_k} \end{pmatrix}_{n \times 1}$$

8.5.1. Problema de contraste

Como se ha indicado, el problema de análisis de una vía se puede reducir a un contraste de igualdad de medias del tipo: $H_0 : \mu_1 = \dots = \mu_k$, el cual equivale a contrastar $H_0 : \mu_1 - \mu_2 = 0, \dots, \mu_1 - \mu_k = 0$ (donde se establecen $k - 1$ relaciones lineales entre las componentes de $\boldsymbol{\mu}$). Es decir, el contraste que se quiere estudiar es una hipótesis lineal general del tipo:

$$H_0 : \mathbf{C}\boldsymbol{\mu} = 0, \quad \mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ \vdots & 0 & -1 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}_{(k-1) \times k} \quad q = \text{rang}(\mathbf{C}) = k - 1.$$

El test de razón de verosimilitud que resuelve este contraste es, de nuevo, el TRV para el MLG, es decir:

$$\varphi(\mathbf{Y}) = \begin{cases} 1 & F(\mathbf{Y}) > F_{q, n-r; \alpha} = F_{k-1, n-k; \alpha} \\ 0 & F(\mathbf{Y}) \leq F_{q, n-r; \alpha} = F_{k-1, n-k; \alpha} \end{cases}$$

con

$$F(\mathbf{Y}) = \frac{n - k}{k - 1} \left(\frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\mu}}^0\|^2 - \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\mu}}\|^2}{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\mu}}\|^2} \right)$$

donde $\hat{\boldsymbol{\mu}}$ es el EMV de $\boldsymbol{\mu}$ y $\hat{\boldsymbol{\mu}}^0$ es el EMV de $\boldsymbol{\mu}$ bajo H_0 .

Dichos estimadores son fácilmente deducibles teniendo en cuenta que se está trabajando con distribuciones normales.

Estimador máximo verosímil (mínimos cuadrados) de $\boldsymbol{\mu}$

El parámetro $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^T$ es la media de la distribución del vector $\mathbf{Y} \rightsquigarrow \mathcal{N}_k(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_k)$, es decir, es el parámetro media de una normal multivariante. Pero, además, \mathbf{Y} es un vector formado por k variables aleatorias independientes con distribuciones $\mathcal{N}(\mu_i, \sigma^2)$, donde cada componente del vector $\boldsymbol{\mu}$ es la media de una variable aleatoria normal unidimensional.

Según se ha estudiado anteriormente, el estimador máximo verosímil (EMV) del parámetro media de una distribución normal es la media muestral y, por tanto, como en este caso se tiene Y_{i1}, \dots, Y_{in_i} m.a.s. de Y_i :

$$\hat{\mu}_i = \bar{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}.$$

Luego el EMV de $\boldsymbol{\mu}$ es: $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \dots, \hat{\mu}_k)^T = (\bar{Y}_1, \dots, \bar{Y}_k)^T$.

Estimador máximo verosímil (mínimos cuadrados) de $\boldsymbol{\mu}$ bajo H_0

Bajo $H_0 : \mu_1 = \dots = \mu_k$, el parámetro $\boldsymbol{\mu}$ se puede ver como un vector formado por k componentes iguales, a las que también se las puede llamar μ , es decir, $\boldsymbol{\mu} = (\mu, \dots, \mu)^T$. Por lo tanto, se tiene que μ es un parámetro unidimensional que determina la media de la distribución de $Y \rightsquigarrow \mathcal{N}_k(\mu \mathbf{1}_k, \sigma^2 \mathbf{I}_k)$. Pero, en este caso, Y es un vector formado por k variables aleatorias independientes e idénticamente distribuidas $\mathcal{N}(\mu, \sigma^2)$, por lo tanto μ es la media de una variable aleatoria normal unidimensional.

Al igual que antes, se sabe que el EMV del parámetro media de una distribución normal es la media muestral y, por tanto, como en este caso se tiene $Y_{11}, \dots, Y_{1n_1}, \dots, Y_{k1}, \dots, Y_{kn_k}$ m.a.s. de $\mathcal{N}(\mu, \sigma^2)$:

$$\hat{\mu} = \bar{Y} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}}{n} \quad \text{donde } n = \sum_{i=1}^k n_i$$

Luego el EMV de $\boldsymbol{\mu}$ bajo H_0 es: $\hat{\boldsymbol{\mu}}^0 = (\hat{\mu}, \dots, \hat{\mu})^T = (\bar{Y}, \dots, \bar{Y})^T$.

El estadístico del contraste se puede reescribir, sustituyendo los EMV obtenidos y quedaría:

$$F(\mathbf{Y}) = \frac{n-k}{k-1} \left(\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 - \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2} \right) (\rightsquigarrow F(k-1, n-k) \text{ bajo } H_0).$$

8.5.2. Descomposición de la variabilidad de las observaciones

Para estudiar la descomposición de la variabilidad, primero se debe recordar que estamos considerando el modelo lineal $\mathbf{Y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$, cuyo modelo estimado es $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\mu}}$, es

decir $\hat{Y}_{ij} = \hat{\mu}_i = \bar{Y}_i$. Luego, el error que se comete al estimar Y_{ij} por \hat{Y}_{ij} viene dado por $R_{ij} = Y_{ij} - \bar{Y}_i$.

Por otro lado, anteriormente se ha estudiado que la variabilidad total de una serie de observaciones aleatorias, $Y_{11}, \dots, Y_{1n_1}, \dots, Y_{k1}, \dots, Y_{kn_k}$, viene determinada por:

$$VT = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2$$

Dichas observaciones aleatorias son las m.a.s., independientes, de las variables $Y_i \rightsquigarrow \mathcal{N}(\mu_i, \sigma^2)$.

Al igual que anteriormente, la VT se puede descomponer como suma de dos términos,

$$VT = VE + VNE$$

- La variabilidad explicada (VE) por el modelo, que mide las variabilidades entre los grupos (también se llama variabilidad entre grupos):

$$VE = \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2.$$

- La variabilidad no explicada (VNE) por el modelo, que es la suma de las variabilidades dentro de cada grupo:

$$VNE = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} R_{ij}^2.$$

Teniendo en cuenta estos nuevos términos, el estadístico del contraste se puede reescribir como:

$$F(\mathbf{Y}) = \frac{n-k}{k-1} \left(\frac{VT - VNE}{VNE} \right) = \frac{n-k}{k-1} \frac{VE}{VNE} = \frac{VE/(k-1)}{VNE/(n-k)} = \frac{S_E^2}{S_R^2}$$

donde

$$S_E^2 = \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2 / (k-1)$$

denota la varianza entre los grupos y

$$S_R^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 / (n-k)$$

la varianza residual.

Por tanto, el test de la razón de verosimilitud quedaría:

$$\varphi(\mathbf{Y}) = \begin{cases} 1 & \text{si } \frac{S_E^2}{S_R^2} > F_{k-1, n-k; \alpha} \\ 0 & \text{si } \frac{S_E^2}{S_R^2} \leq F_{k-1, n-k; \alpha} \end{cases}$$

Si se desea resolver el contraste sin establecer previamente el nivel de significación con el que se desea trabajar, se debe calcular:

$$p\text{-valor} = P[F_{k-1, n-k} > F_{exp}] \text{ siendo } F_{exp} = \frac{s_E^2}{s_R^2} = \frac{ve/(k-1)}{vne/(n-k)} = \frac{\sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2 / (k-1)}{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 / (n-k)}.$$

Tema 9

INTRODUCCIÓN A LA INFERENCIA NO PARAMÉTRICA

Hasta ahora se suponía conocida la distribución de la variable bajo estudio salvo algún parámetro y, por tanto, se aplicaban conceptos de inferencia paramétrica. No obstante, en la práctica, no se suele conocer la forma funcional de la distribución y, en dicho caso, no se puede hacer inferencia paramétrica.

Por ello es necesario poder obtener conclusiones sobre la distribución de la variable aleatoria a partir de las observaciones sin conocer la forma funcional, lo cual es uno de los problemas que se estudian dentro de la inferencia no paramétrica. Aunque no se conozca la forma funcional si se suele tener información de tipo general sobre la variable: si es discreta o continua, simetría, curtosis, etc.

Aunque existen procesos de estimación no paramétricos, en este tema se van a estudiar solamente cuestiones de contrastes de hipótesis.

9.1. Problemas de bondad de ajuste

El problema de bondad de ajuste trata de decidir, en base a la información que proporciona una m.a.s. de una variable aleatoria, si se puede admitir que la distribución de la variable es una concreta (ejemplo: $\mathcal{N}(0, 1)$, $\exp(3)$) o bien si pertenece aun cierto tipo de distribuciones (ejemplo: normal, exponencial). Es decir, es un problema de bondad de ajuste de los datos observados a una distribución especificada.

En particular se va a contrastar si una muestra proviene de una población con una función de distribución específica, F_0 , frente a que dicha función de distribución sea diferente ($F(x) \neq F_0(x)$ para algún $x \in \mathbb{R}$).

Las dos soluciones más frecuentes para resolver este contraste son:

- Test χ^2 de Pearson (1900). Es el primero que surge históricamente y se puede aplicar a variables de tipo discreto, continuo y cualitativo. Este test lo único que tiene en cuenta es una clasificación de las observaciones muestrales en distintas categorías.
- Test de Kolmogorov-Smirnov. Este test se basa en el teorema de Glivenko-Cantelli.

9.1.1. Test χ^2 de Pearson

Sea (X_1, \dots, X_n) una m.a.s. de una variable aleatoria X que se distribuye según una función de distribución F que es completamente desconocida.

A) Hipótesis nula simple: $\begin{cases} H_0 : F = F_0 \\ H_1 : F \neq F_0 \end{cases}$

Para resolver este problema se parte el recorrido de la función teórica correspondiente a F_0 en k subconjuntos A_1, \dots, A_k de probabilidad no nula y se consideran las siguientes probabilidades $p_i^0 = P_{F_0}[X \in A_i] > 0$, $i = 1, \dots, k$. Sea N_i el número de observaciones muestrales en cada A_i , $i = 1, \dots, k$. Entonces el estadístico para el contraste es

$$\chi^2(N_1, \dots, N_k) = \sum_{i=1}^k \frac{(N_i - np_i^0)^2}{np_i^0} \left(= -n + \sum_{i=1}^k \frac{N_i^2}{np_i^0} \right)$$

donde np_i^0 es el número de observaciones muestrales que cabría esperar en A_i si H_0 es cierta.

Este estadístico es una medida de la discrepancia entre las observaciones reales en cada clase y el número que debería de haber si H_0 fuera cierta.

Pearson demostró que bajo H_0 este estadístico tiene distribución asintótica,

$$\chi^2(X_1, \dots, X_n) \rightsquigarrow_{n \rightarrow \infty} \chi^2(k-1).$$

El **Test asintótico para $H_0 : F = F_0$** , de tamaño α , es:

$$\varphi(X_1, \dots, X_n) = \begin{cases} 1 & \chi^2(N_1, \dots, N_k) \geq \chi_{k-1;\alpha}^2 \\ 0 & \chi^2(N_1, \dots, N_k) < \chi_{k-1;\alpha}^2 \end{cases}$$

con

$$p - \text{valor} = P_{H_0}[\chi^2(N_1, \dots, N_k) \geq \chi_{exp}^2] \approx_{n \rightarrow \infty} P[\chi^2(k-1) \geq \chi_{exp}^2]$$

siendo χ_{exp}^2 el valor de estadístico en la muestra observada. Entonces, se rechaza H_0 si el $p - \text{valor}$ es menor o igual que α .

Notas:

- Este test, como ya se ha especificado, es un test asintótico, por lo tanto, habrá que especificar cómo tiene que ser n para poder usar el test. Usualmente se considera como restricción para la aplicación del test que $np_i^0 \geq 5, i = 1, \dots, k$.
- Si se plantean como hipótesis nula $H_0 : F = F_0$ ó $H'_0 : F = F'_0$, y tanto F_0 como F'_0 asignan la misma probabilidad a todos los $A_i, i = 1, \dots, k$, este test no distingue entre F_0 y F'_0 . Para solucionar este problema se considerarán al menos 5 clases. En tal caso $\exists i : p_i^0 \leq 1/5 \Rightarrow n \geq 25$. (Para hacer particiones “buenas” se pueden usar los percentiles).
- Este test es aplicable a cualquier tipo de variable cuyos valores puedan clasificarse en un número finito de categorías pero es más apropiado para variables cualitativas ya que ellas son propiamente categóricas.

B) Hipótesis nula compuesta:

En muchos casos, la hipótesis nula H_0 no especifica una única distribución F_0 , sino una familia de distribuciones posibles (p.e., una normal con parámetros desconocidos, etc), dependientes de uno o varios parámetros. En dicho caso no se puede aplicar directamente el test χ^2 . Será necesario tener una estimación previa de los parámetros. Por tanto:

- Primero se estiman los parámetros de la familia especificada en H_0 , usualmente por máxima verosimilitud.
- Después se aplica el test con los parámetros ya estimados.

La distribución del estadístico del contraste depende de cómo se hayan obtenido las estimaciones. En particular, si los parámetros se estiman partiendo de observaciones independientes de las que se van a usar para el problema de contraste, se usa χ^2 igual que antes. Sin embargo, si se usan para el contraste los mismos datos que para la estimación, los \hat{p}_i dependen de las observaciones, y la distribución del estadístico bajo H_0 varía:

$$\hat{\chi}^2(N_1, \dots, N_k) = \sum_{i=1}^k \frac{(N_i - n\hat{p}_i)^2}{n\hat{p}_i} \longrightarrow \chi^2(k - q - 1)$$

donde q es el número de parámetros estimados.

Este test tiene una serie de inconvenientes:

- Es un test asintótico y, por tanto, aproximado.

- No trata los datos individualmente, sino por categorías. Por tanto, no usa toda la información contenida en la muestra. Por ello no es un buen test para variables aleatorias continuas.

El siguiente test es un test exacto, no asintótico, válido para variables aleatorias continuas y trata todos los datos de forma individual.

Ejemplo: Se recoge una muestra aleatoria simple de 30 tornillos producidos por cierta máquina y se mide su longitud, obteniéndose:

10.39	10.66	10.12	10.32	10.25	10.52	10.83	10.72	10.28	10.35
10.46	10.54	10.23	10.18	10.62	10.49	10.61	10.64	10.29	10.78
10.81	10.34	10.75	10.41	10.53	10.31	10.47	10.43	10.57	10.74

Contrastar si estos datos avalan que la distribución de la longitud de los tornillos es normal.

9.1.2. Test de Kolmogorov-Smirnov

El test de Kolmogorov-Smirnov se basa en el teorema de Glivenko-Cantelli que, como ya se estudió, proporciona la convergencia casi segura uniformemente de la función de distribución muestral o empírica (F_{X_1, \dots, X_n}^*) a la función de distribución de la variable aleatoria (F) .

Para resolver el problema planteado se considera (X_1, \dots, X_n) una m.a.s. de una variable aleatoria X continua que se distribuye según una función de distribución F que es completamente desconocida. El contraste a resolver es

$$\begin{cases} H_0 : F = F_0 \\ H_1 : F \neq F_0 \end{cases}$$

Para resolver este problema se usa el estadístico de Kolmogorov-Smirnov

$$D(X_1, \dots, X_n) = \sup_{x \in \mathbb{R}} |F_{X_1, \dots, X_n}^*(x) - F_0(x)|$$

el cual proporciona una medida de la discrepancia entre F_{X_1, \dots, X_n}^* y F_0 . Por tanto, teniendo en cuenta que la distribución muestral converge uniformemente a la distribución teórica, se rechazará la hipótesis nula, si el valor de $D(X_1, \dots, X_n)$ es grande. Es decir, el test de Kolmogorov-Smirnov sería:

$$\varphi(X_1, \dots, X_n) = \begin{cases} 1 & D(X_1, \dots, X_n) \geq d_\alpha \\ 0 & D(X_1, \dots, X_n) < d_\alpha \end{cases}$$

donde d_α verifica:

$$P_{H_0} (D(X_1, \dots, X_n) \geq d_\alpha) = \alpha$$

y

$$p - \text{valor} = P_{H_0}[D(X_1, \dots, X_n) \geq D_{exp}]$$

siendo D_{exp} el valor de estadístico en la muestra observada.

Teorema: Si F_0 es continua:

- (a) La distribución de $D(X_1, \dots, X_n)$ es independientes de F_0 .
- (b) $D(X_1, \dots, X_n) \rightsquigarrow_{H_0} Z$ de Kolmogorov.
- (c) Si las n observaciones son distintas, entonces

$$D_{exp} = \max\left\{\max_{x_i}[F_{X_1, \dots, X_n}^*(x_i) - F_0(x_i)], \max_{x_i}[F_0(x_i) - F_{X_1, \dots, X_n}^*(x_i)]\right\}$$

Notas:

1. Existe otra expresión de $D(X_1, \dots, X_n)$ para observaciones iguales, aunque esto es poco probable por ser la distribución continua. Si ocurriera, por redondeos, se eliminan del estudio los elementos iguales para asegurarse que las observaciones sean distintas.
2. Si la distribución con la que se quiere comparar no está totalmente determinada, al igual que en el test χ^2 , se pueden estimar los parámetros de la distribución, lo cual varía la distribución de $D(X_1, \dots, X_n)$. Hay modificaciones del test de Kolmogorov-Smirnov, como el test de Lilliefors, para estos casos. Sin embargo, otra opción aceptable es usar el test de la χ^2 para dichos casos.

Ejemplo: Se supone que el tiempo de reacción a un determinado compuesto se distribuye según una $\mathcal{N}(10.5; 0.15^2)$. Contrastar si los siguientes datos, obtenidos en un muestreo aleatorio simple de 10 individuos a los que se ha administrado el compuesto, proporcionan evidencia para rechazar esta hipótesis:

10.39 10.66 10.12 10.32 10.25 10.52 10.83 10.72 10.28 10.35.

9.2. Problema de localización

Se van a usar tests de localización para resolver problemas de contrastes de hipótesis relativos a medidas de posición (mediana o cuantiles en general). En concreto, los posibles contrastes sobre la mediana son:

$$\left\{ \begin{array}{l} H_0 : M_X = m \\ H_1 : M_X \neq m \end{array} \right\} \quad \left\{ \begin{array}{l} H_0 : M_X = m \\ H_1 : M_X > m \end{array} \right\} \quad \left\{ \begin{array}{l} H_0 : M_X = m \\ H_1 : M_X < m \end{array} \right\}$$

y se van a utilizar dos tests para resolverlos:

- Test de los signos de Fisher: se aplica sólo a variables aleatorias de tipo continuo y se generaliza fácilmente a contraste de hipótesis sobre cuantiles.
- Test de rangos signados de Wilcoxon: usa más información muestral que el de Fisher pero exige, además de continuidad, simetría en la distribución.

9.2.1. Test de los signos de Fisher

Sea X_1, \dots, X_n m.a.s. de $X \sim F$ continua (desconocida).

Idea intuitiva: Es de esperar que, si H_0 es cierta, aproximadamente la mitad de los valores muestrales queda por encima de m y la otra mitad por debajo (concuerda con la idea de convergencia de cuantiles muestrales a cuantiles poblacionales), también con el hecho de que $F(M_X) = 1/2$.

Se puede definir el estadístico del test de los signos:

$$\begin{aligned} T(X_1, \dots, X_n) &= \text{número de observaciones muestrales mayores que } m \\ &= \text{n}^\circ \text{ de signos positivos en } (X_i - m) \rightsquigarrow_{H_0} B(n, 1/2) \end{aligned}$$

Las regiones crítica del test, según el contraste plateado son:

- Para $H_1 : M_X > m$, se rechaza H_0 para

$$T_{exp}(= T(x_1, \dots, x_n)) \geq k : P_{H_0}[T(X_1, \dots, X_n) \geq k] \leq \alpha.$$

Otra opción es calcular el p -valor $= P_{H_0}[T(X_1, \dots, X_n) \geq T_{exp}]$.

- Para $H_1 : M_X < m$, se rechaza H_0 para

$$T_{exp} \leq k : P_{H_0}[T(X_1, \dots, X_n) \leq k] \leq \alpha.$$

Otra opción es calcular el p -valor $= P_{H_0}[T(X_1, \dots, X_n) \leq T_{exp}]$.

- Para $H_1 : M_X \neq m$, se rechaza H_0 para

$$T_{exp} \leq k \text{ ó } T_{exp} \geq n - k : P_{H_0}[T(X_1, \dots, X_n) \leq k] \leq \alpha/2.$$

Otra opción es calcular el

$$p\text{-valor} = \begin{cases} 2P_{H_0}[T(X_1, \dots, X_n) \leq T_{exp}] & \text{si } T_{exp} \leq n/2 \\ 2P_{H_0}[T(X_1, \dots, X_n) \geq T_{exp}] & \text{si } T_{exp} \geq n/2 \end{cases}$$

Notas:

1. Este test se puede aleatorizar.
2. Si algún valor de la muestra coincide con m , ($X_i - m = 0$), dicho dato se elimina y se reajusta el tamaño de la muestra, n .
3. Para n grande ($n \geq 20$) se puede emplear la aproximación normal a la distribución binomial para determinar los puntos críticos (k), pero sería un test asintótico:

$$B(n, p) \approx N(np, npq) \Rightarrow \frac{2T(X_1, \dots, X_n) - n}{\sqrt{n}} \sim N(0, 1)$$

4. El test se puede generaliza a tests sobre cuantiles de cualquier orden.

Ejemplo: Una empresa que tradicionalmente comenzaba su actividad diaria a las 9 h. ha cambiado su horario para abrir a las 8 h. y se pregunta si ello ha afectado significativamente al retraso de sus empleados. Es aceptable pensar que la forma de la distribución de los retrasos no ha variado con el cambio de horario, pero se teme que se haya desplazado hacia la derecha, lo cual supondrá un incremento del tiempo perdido. Se sabe, además, que la mediana de los retrasos de los empleados era inicialmente de 5 minutos. Con el cambio de horario se selecciona a 12 empleados y se observa, en determinados días, los siguientes retrasos (en minutos):

2.5, 1.2, 7, 1.8, 8.3, 6.8, 5.2, 3.4, 4.7, 6.2, 9.1, 5.2

A partir de estos datos, contrastar la hipótesis de que la distribución de los retrasos no ha variado con el cambio de horario.

9.2.2. Test de los rangos signados de Wilcoxon

Este test sólo se puede aplicar en el caso en que se conoce que la distribución, además de continua, es simétrica. Wilcoxon propuso un test para contrastar $H_0 : M_X = m$, que además de tener en cuenta la diferencia, tiene en cuenta la magnitud de la misma, por lo que es un test muy potente. El problema de este test es que necesita que la distribución sea simétrica y que los datos sean exactos.

Sea X_1, \dots, X_n una m.a.s. de X , v.a. con distribución continua y simétrica (alrededor de la mediana) y sea $D_i = X_i - m$, $i = 1, \dots, n$. Si algún D_i es 0, se elimina ese dato y se reajusta el número de datos n .

El método propuesto por Wilcoxon consiste en ordenar de forma creciente los valores absolutos de estas diferencias ($|D_i|$) y anotar el rango o lugar que ocupan ($r(|D_i|)$), de ahí

el nombre del test. Si hubiera empates, es decir si hubiera datos repetidos, se le asigna a cada uno el promedio de los rangos.

Basándose en esta idea, se define el estadístico del test de Wilcoxon como:

$$T^+(X_1, \dots, X_n) = \text{suma de los rangos de los } D_i \text{ positivos} = \sum_{i=1}^n r(|D_i|)I\{D_i > 0\}$$

La distribución de $T^+(X_1, \dots, X_n)$ bajo H_0 es simétrica en torno a la media $\frac{n(n+1)}{4}$ y viene dada por:

$$P[T^+(X_1, \dots, X_n) = t] = P\left[T^+(X_1, \dots, X_n) = \frac{n(n+1)}{2} - t\right]$$

Si $n \leq 15$, la distribución bajo H_0 está tabulada para ambas colas. Si $n > 15$, se puede aproximar asintóticamente la distribución de $T^+(X_1, \dots, X_n)$, bajo H_0 , por

$$T^+(X_1, \dots, X_n) \approx \mathcal{N}\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right).$$

Las regiones críticas del test según el contraste planteado son:

- Si $H_1 : M_X > m$ es cierta, cabe esperar que haya más diferencias mayores que 0 que menores que 0 y que la magnitud de las que son mayores que 0 sea mayor que la magnitud de las que son menores que 0. Por tanto se rechaza H_0 si $T^+(X_1, \dots, X_n)$ es grande: Para un nivel de significación α se rechaza H_0 si

$$T_{exp}^+ (= T^+(x_1, \dots, x_n)) \geq k$$

con $P[T^+(X_1, \dots, X_n) \geq k] \leq \alpha$.

En general se calcula el p -valor $= P_{H_0}[T^+(X_1, \dots, X_n) \geq T_{exp}^+]$.

- Si $H_1 : M_X < m$ es cierta, cabe esperar lo contrario que antes. Por tanto se rechaza H_0 si $T^+(X_1, \dots, X_n)$ es pequeño: Para un nivel de significación α se rechaza H_0 si

$$T_{exp}^+ \leq k$$

con $P[T^+(X_1, \dots, X_n) \leq k] \leq \alpha$.

En general se calcula el p -valor $= P_{H_0}[T^+(X_1, \dots, X_n) \leq T_{exp}^+]$.

- Si $H_1 : M_X \neq m$ es cierta, se rechaza H_0 para valores pequeños o grandes de $T^+(X_1, \dots, X_n)$: Para un nivel de significación α se rechaza H_0 si

$$T_{exp}^+ \leq k \text{ o } T_{exp}^+ \geq \frac{n(n+1)}{2} - k$$

con $P[T^+(X_1, \dots, X_n) \leq k] \leq \alpha/2$.

En general se calcula el p – *valor*:

$$p - \text{valor} = \begin{cases} 2P_{H_0}[T^+(X_1, \dots, X_n) \geq T_{exp}^+] & \text{si } T_{exp}^+ \geq \frac{n(n+1)}{4} \\ 2P_{H_0}[T^+(X_1, \dots, X_n) \leq T_{exp}^+] & \text{si } T_{exp}^+ \leq \frac{n(n+1)}{4} \end{cases}$$

Ejemplo: A partir de los datos del ejemplo anterior, y suponiendo que la distribución de los retrasos es simétrica, contrastar la hipótesis de que ésta no varía con el cambio de horario.

Nota: Los tests de localización se usan también para contrastar la hipótesis de homogeneidad de las distribuciones correspondientes a dos muestras apareadas o relacionadas cuando se tiene constancia de que las distribuciones tienen la misma forma funcional pero una está desplazada respecto de la otra. En dicho caso se toma la variable diferencia y $H_0 : M_{X-Y} = 0$. Para aplicar cada test deberá comprobarse primeramente que se está bajo las condiciones necesarias.

9.3. Problema de independencia: test χ^2

El problema de independencia relativo a dos muestras trata, como su nombre indica, de ver si dos variables, referidas a una misma población son independientes o no. Sean X e Y dos características poblacionales distintas. Se va a contrastar:

$$\begin{cases} H_0 : X \text{ e } Y \text{ son independientes} \\ H_1 : X \text{ e } Y \text{ no son independientes} \end{cases}$$

Para resolver este contraste se va a utilizar el test χ^2 de independencia.

Test χ^2 de independencia

Sean X e Y dos variables cualitativas, teniendo X las categorías A_1, \dots, A_m , e Y las B_1, \dots, B_k . Para resolver el problema de contraste $H_0 : X$ e Y son independientes, se toma una m.a.s. de individuos y se clasifican los individuos según las categorías de X e Y . Sea N_{ij} el número de individuos de la muestra que presentan las categorías A_i y B_j , $\forall i = 1, \dots, m, \forall j = 1, \dots, k$.

Con dichos datos muestrales se construye la tabla de contingencia muestral (reparto de la muestra por categorías bidimensionales, totales, marginales y total global).

9.3 Problema de independencia: χ^2

$X \backslash Y$	B_1	B_2	\dots	B_k	Totales
A_1	N_{11}	N_{12}	\dots	N_{1k}	$N_{1\cdot}$
A_2	N_{21}	N_{22}	\dots	N_{2k}	$N_{2\cdot}$
\vdots	\vdots	\vdots	\ddots	\vdots	
A_m	N_{m1}	N_{m2}	\dots	N_{mk}	$N_{m\cdot}$
Totales	$N_{\cdot 1}$	$N_{\cdot 2}$	\dots	$N_{\cdot k}$	n

donde $N_{i\cdot} = \sum_{j=1}^k N_{ij}$ y $N_{\cdot j} = \sum_{i=1}^m N_{ij}$.

Sean $P_{ij} = P[X \in A_i, Y \in B_j]$, $P_{i\cdot} = P[X \in A_i]$ y $P_{\cdot j} = P[Y \in B_j]$, $\forall i = 1, \dots, m$, $\forall j = 1, \dots, k$. El estadístico del contraste es:

$$\sum_{i=1}^m \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

siendo O_{ij} la frecuencia observada y E_{ij} la frecuencia esperada.

Una forma de plantear la hipótesis nula, basada en las frecuencias, es $H_0 : P_{ij} = P_{i\cdot}P_{\cdot j}$, ya que la independencia de dos variables se caracteriza porque su frecuencia conjunta sea igual al producto de sus frecuencias marginales. Por lo tanto, bajo H_0 la frecuencia esperada $E_{ij} = nP_{ij}$ quedaría $E_{ij} = nP_{i\cdot}P_{\cdot j}$.

Por otro lado, ya que las frecuencias teóricas P_{ij} , $P_{i\cdot}$ y $P_{\cdot j}$, en general, no son conocidas, se trabajará con sus estimaciones máximo verosímiles: $\hat{P}_{ij} = N_{ij}/n$, $\hat{P}_{i\cdot} = N_{i\cdot}/n$ y $\hat{P}_{\cdot j} = N_{\cdot j}/n$. Finalmente, la frecuencia observada $O_{ij} = N_{ij}$. Teniendo en cuenta todo esto se llega a que el estadístico del contraste es:

$$\chi^2(N_{ij}) = \sum_{i=1}^m \sum_{j=1}^k \frac{(N_{ij} - \frac{N_{i\cdot}N_{\cdot j}}{n})^2}{\frac{N_{i\cdot}N_{\cdot j}}{n}}$$

que bajo H_0 tiene la distribución asintótica $\chi^2_{(m-1)(k-1)}$.

Al nivel de significación α se rechaza H_0 si

$$\chi_{exp}^2 \geq \chi_{(m-1)(k-1); \alpha}^2,$$

siendo $\chi_{exp}^2 = \chi^2(x_1, \dots, x_n)$ y $P_{H_0}[\chi^2(N_{ij}) \geq \chi_{(m-1)(k-1); \alpha}^2] = \alpha$.

El p-valor es $P_{H_0}[\chi^2(N_{ij}) \geq \chi_{exp}^2]$.

Notas

- Los requisitos mínimos para poder usar la distribución asintótica indicada son:
 - Las frecuencias esperadas deben ser mayores o iguales que 2.

$$\frac{n_{i\cdot}n_{\cdot j}}{n} \geq 2$$

- Hay que asegurarse que al menos el 80 % de las frecuencias esperadas sea mayores o iguales a 5.

Si no es así, se debe aumentar el tamaño de la muestra.

- Para aplicar el test de χ^2 de independencia a dos v.a. cualesquiera, no cualitativas, se agrupan los valores de cada variable en un número finito de categorías que respeten las condiciones necesarias para poder aplicar el test.

Ejemplo: Para estudiar si el grupo sanguíneo de los individuos tiene relación con la predisposición a la diabetes, se han seleccionado al azar 400 sujetos a los que se ha determinado el grupo sanguíneo y el nivel de glucosa en sangre en idénticas condiciones experimentales. Clasificando la segunda medida en tres niveles, los resultados han sido:

Grupo \ Nivel	Bajo	Medio	Alto
O	137	86	35
A	42	23	11
B	19	17	7
AB	14	7	2

Contrastar, al nivel de significación 0.05, si ambas variables son independientes.

9.4. Problema de homogeneidad: test χ^2

El problema de homogeneidad consiste en estudiar si una serie de poblaciones se comportan de la misma forma frente a una determinada característica. Para ello se toman m.a.s. de cada población, se mide la característica de interés en ellas y se trata de contrastar si todas las muestras proceden de variables con la misma distribución teórica. Se va a contrastar:

$$\begin{cases} H_0 : F_1 = \dots = F_m \\ H_1 : \text{Alguna distribución es distinta} \end{cases}$$

Para resolver dicho contraste se va a utilizar el test χ^2 de homogeneidad.

Test χ^2 de homogeneidad

Este test se debe aplicar, en un principio, a características de tipo cualitativo. Se van a suponer m poblaciones, m muestras aleatorias simples, de tamaños n_1, \dots, n_m y que, en todos los casos, las variables pueden tomar valores en k categorías A_1, \dots, A_k . Sean N_{ij} el número de observaciones de la muestra i -ésima que presenta la modalidad A_j , $i = 1, \dots, m$, $j = 1, \dots, k$, $N_{.j} = \sum_{i=1}^m N_{ij}$, $n_i = \sum_{j=1}^k N_{ij}$ y $n = \sum_{i=1}^m n_i$.

Con dichos datos muestrales se construye la tabla de contingencia muestral

9.4 Problema de homogeneidad: χ^2

Muestras \ Categorías	A_1	A_2	\dots	A_k	
1	N_{11}	N_{12}	\dots	N_{1k}	n_1
2	N_{21}	N_{22}	\dots	N_{2k}	n_2
\vdots	\vdots	\vdots	\ddots	\vdots	
m	N_{m1}	N_{m2}	\dots	N_{mk}	n_m
	$N_{.1}$	$N_{.2}$	\dots	$N_{.k}$	n

Si se denota por P_{ij} a la probabilidad de que un individuo de la muestra i -ésima presente la modalidad A_j , la hipótesis nula del contraste se puede escribir como: $H_0 : P_{1j} = P_{2j} = \dots = P_{mj} (= P_{.j})$, $j = 1, \dots, k$. Desde el punto de vista paramétrico, este contraste se puede ver como contrastar la igualdad de los parámetros p de m multinomiales de dimensión $k - 1$, que se puede resolver mediante el test de la razón de verosimilitud. Como se está estudiando el caso no paramétrico, se va a resolver el contraste con el test χ^2 , pero ambos test son asintóticamente equivalentes.

Al igual que antes, el estadístico del contraste es:

$$\sum_{i=1}^m \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

siendo O_{ij} la frecuencia observada y E_{ij} la frecuencia esperada.

En este caso, bajo H_0 la frecuencia esperada es $E_{ij} = n_i P_{.j}$, pero como no son conocidas, se trabajará con sus estimaciones máximo verosímiles: $\hat{P}_{.j} = N_{.j}/n$. Finalmente, al igual que antes, la frecuencia observada es $O_{ij} = N_{ij}$, con lo que el estadístico del contraste queda:

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(N_{ij} - \frac{n_i N_{.j}}{n})^2}{\frac{n_i N_{.j}}{n}}$$

que, bajo H_0 tiene la distribución asintótica $\chi^2_{(m-1)(k-1)}$.

Al nivel de significación α se rechaza H_0 si

$$\chi_{exp}^2 \geq \chi_{(m-1)(k-1); \alpha}^2,$$

siendo $\chi_{exp}^2 = \chi^2(x_1, \dots, x_n)$ y $P_{H_0}[\chi^2(N_{ij}) \geq \chi_{(m-1)(k-1); \alpha}^2] = \alpha$.

El p-valor es $P_{H_0}[\chi^2(N_{ij}) \geq \chi_{exp}^2]$.

Notas:

- Los requisitos mínimos para poder usar la distribución asintótica indicada son:
 - Los tamaños muestrales en cada población deben ser como mínimo de 20.

$$n_i \geq 20, \forall i = 1, \dots, m.$$

- Las frecuencias esperadas deben ser mayores o iguales que 2.

$$\frac{n_i n_{.j}}{n} \geq 2$$

- Hay que asegurarse que no más del 20% de las frecuencias esperadas sea menores a 5.

Si no es así, se debe aumentar el tamaño de las muestras.

- A pesar de que existe una gran analogía con el test de independencia, el problema que resuelve este otro test es totalmente distinto.
- Para aplicar el test χ^2 de homogeneidad a m variables aleatorias cualesquiera, es decir, si se tienen X_1, \dots, X_m variables aleatorias cualesquiera y se desea contrastar $H_0 : F_1 = \dots = F_m$, se particiona el rango de valores comunes a todas las variables en k subconjuntos o modalidades (A_j) de probabilidad no nula bajo todas las distribuciones. Se considera P_{ij} = probabilidad de que la variable $X_i \in A_j$, $\forall i = 1, \dots, m$, $\forall j = 1, \dots, k$ y se toma una m.a.s. de cada variable. Como las muestras son independientes se puede aplicar el test χ^2 de homogeneidad a ellas para resolver el contraste planteado.
- Para variables no cualitativas hay tests mucho mejores, ya que el test χ^2 no utiliza los datos, sino la pertenencia a algunos intervalos.

Ejemplo: Contrastar, a partir de los resultados de la siguiente tabla, si los distintos grupos sanguíneos se presentan con la misma frecuencia en tres grupos étnicos diferentes:

Raza \ Grupo	O	A	B	AB
1	32	11	7	2
2	47	13	17	9
3	23	7	9	6