Índice

Tema 6

Estimación puntual y por intervalos de confianza

6.1 Introducción

Este capítulo recoge dentro de la Inferencia Estadística el problema de la estimación. Supongamos que alguna característica de los elementos de una población se puede representar mediante una variable aleatoria X de la que se conoce el tipo de distribución pero no los parámetros que la determinan. El objetivo de este tema es la obtención de un valor que pueda asignarse a cada uno de los parámetros desconocidos. Para ello se obtiene de la población la información precisa mediante una muestra aleatoria, se establece una función de los valores muestrales, estimador, y se asigna al parámetro el valor que tome esta función en una muestra concreta.

Por ejemplo, si $X \sim Be(p)$ con p desconocido ó $X \sim N(\mu, \sigma^2)$ con media y varianza desconocidos, se desea conocer de la forma más aproximada posible el valor de dichos parámetros ó de una función de ellos, es decir, se pretende dar una estimación de los parámetros. Para ello, obtenemos una muestra, considerando $X_1, X_2, ..., X_n$ m.a.s. y $x_1, x_2, ..., x_n$ los valores que se observan (realización muestral). Por tanto para cada valor que se observa, se puede considerar que hay un elemento de la m.a.s., es decir, una variable aleatoria que los determina.

La aproximación ó estimación se puede llevar a cabo de diferentes formas: desde el punto de vista de la estimación puntual, desde el punto de vista de los intervalos de confianza y mediante los contrastes de hipótesis que estudiaremos en el tema siguiente.

La estimación puntual consiste en dar un valor real a partir de un estadístico, es decir, una función de la muestra que no depende de ningún parámetro desconocido. La estimación por intervalos consiste en proporcionar todo un intervalo de valores que, con una probabilidad elevada, contiene el verdadero valor del parámetro.

6.2 Estimador

Sea X una variable aleatoria con función de distribución conocida salvo por un parámetro θ, lo que denotaremos como F_θ.

Definición 6.1 *Se define el* ***espacio paramétrico*** *como el conjunto de todos los posibles valores del parámetro desconocido θ. Se representa por Θ, $\theta \in \Theta$.*

Ejemplo 6.1 *Si $X \sim Be(p)$ entonces $\theta = p$ y $\Theta = [0,1]$.*

Definición 6.2 *Cualquier estadístico cuyos valores se utilizan para estimar un parámetro, se denomina* ***estimador*** *de θ y se representa por $\widehat{\theta}$.*

Generalmente a un estimador se le pide que tome valores dentro del espacio paramétrico.

Ejemplo 6.2 *$\bar{X} = \frac{1}{n}\sum X_i$ es un estimador de μ.*
$S_c^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$ es un estimador de σ^2.
$S^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$ es un estimador de σ^2.

6.3 Método de construcción de estimadores

6.3.1 Método de máxima verosimilitud

Este método es bastante intuitivo pues se basa en suponer que siempre ocurre lo más probable, y por ello estima el parámetro mediante aquel valor que hace más probable la muestra.

Definición 6.3 *Se define la* ***función de verosimilitud*** *de una m.a.s. de tamaño n como*

1. Caso discreto

$$L(\theta; x_1, ..., x_n) = P_\theta[X_1 = x_1, X_2 = x_2, ..., X_n = x_n] = \prod_{i=1}^{n} P_\theta[X = x_i]$$

2. Caso continuo

$$L(\theta; x_1, ..., x_n) = f_\theta(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

donde θ es un parámetro desconocido de la variable aleatoria.

Fijados $x_1, x_2, ..., x_n$, valores de la muestra conocidos, la función de verosimilitud sólo depende de θ.

Ejemplo 6.3 *Sea $X \sim B(7,p)$ y X_1,X_2 m.a.s. de X. Si $x_1 = 2$, $x_2 = 3$, la función de verosimilitud de la muestra es*

$$L(p; 2, 3) = P[X_1 = 2, X_2 = 3] = P[X_1 = 2] \times P[X_2 = 3] = \binom{7}{2} p^2(1-p)^5 \times \binom{7}{3} p^3(1-p)^4$$

Definición 6.4 *Sea $L(\theta; x_1, ..., x_n)$ la función de verosimilitud para una m.a.s. $X_1, X_2, ..., X_n$. Se define el* ***estimador de máxima verosimilitud (E.M.V.)*** *de θ, $\widehat{\theta}_{MV}$, al valor de θ que maximiza $L(\theta; x_1, ..., x_n)$.*

Como interpretación podríamos decir que el estimador de máxima verosimilitud es aquel valor que hace máxima la probabilidad de que se dé la muestra obtenida.

En la práctica para determinar el estimador de máxima verosimilitud, si la función L es derivable respecto del parámetro se puede calcular la primera derivada respecto del parámetro, igualar a 0 y despejar el valor,

$$\frac{\partial}{\partial\theta} L = 0$$

comprobando posteriormente que la solución obtenida, $\widehat{\theta}$, corresponde a un máximo mediante la segunda derivada, es decir,

$$\left.\frac{\partial^2}{\partial\theta^2} L\right|_{\theta=\hat{\theta}} < 0.$$

Suele ser útil aplicar previamente logaritmo neperiano ya que al ser el logaritmo neperiano una función creciente, la función de verosimilitud alcanzará el máximo para el mismo punto que su logaritmo.

Ejemplo 6.4 *Sea $X \sim Be(p)$ y $X_1, X_2, ..., X_n$ m.a.s. con $p \in (0,1)$. Vamos a calcular el estimador de máxima verosimilitud para el parámetro p.*

$$L(\theta; x_1, ..., x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

Tomamos logaritmo neperiano de la función de verosimilitud:

$$\ln L = \sum_{i=1}^{n} x_i \ln p + (n - \sum_{i=1}^{n} x_i)\ln(1-p)$$

Calculamos la primera derivada del logaritmo de la función de verosimilitud y la igualamos a 0:

$$\frac{\partial \ln L}{\partial p} = \sum_{i=1}^{n} x_i \frac{1}{p} + (n - \sum_{i=1}^{n} x_i)\frac{1}{1-p}(-1) = 0$$

$$(1-p)\sum_{i=1}^{n} x_i - p(n - \sum_{i=1}^{n} x_i) = 0$$

$$\sum_{i=1}^{n} x_i - pn = 0$$

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}\ .$$

Calculamos la segunda derivada vemos que es negativa, por tanto es un máximo

$$\frac{\partial^2 \ln L}{\partial p^2} = -\sum_{i=1}^{n} x_i p^{-2} - (n - \sum_{i=1}^{n} x_i)(1-p)^{-2}\Bigg|_{p=\bar{x}} < 0, \qquad 0 < \bar{x} < 1.$$

Luego el resultado sería

$$\widehat{p}_{MV} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

Teorema 6.1 *Si $\widehat{\theta}$ es un E.M.V. de θ, entonces $g(\widehat{\theta})$ es un E.M.V. de $g(\theta)$.*

Ejemplo 6.5 *Si $X \sim Be(p)$ y $\widehat{p} = \bar{x}$, entonces $\widehat{q} = 1 - \widehat{p} = 1 - \bar{x}$*

6.3.2 Método de los momentos

El método de los momentos proporciona un estimador de un parámetro sustituyendo los momentos poblacionales $E[X^r] = \alpha_r$, con los momentos muestrales $m_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$ donde $r = 1, 2, ..., .$

Ejemplo 6.6 *Sea $X \sim Be(p)$ sabemos que $E[X] = p$, y que $\bar{X} = \widehat{p}$, se sustituye $E[X]$ por $\bar{X}$ y el parámetro por su estimador, y nos queda $\widehat{p} = \bar{X}$.*

Ejemplo 6.7 *Sea $X \sim Exp(\lambda)$ sabemos que $E[X] = \frac{1}{\lambda}$, luego $\bar{X} = \frac{1}{\lambda}$, y por tanto $\widehat{\lambda} = \frac{1}{\bar{X}}$.*

6.3.3 Estimadores

A continuación se muestra una tabla con los estimadores más usuales según la distribución de la variable en estudio.

$N(\mu, \sigma^2)$	$\widehat{\mu} = \bar{X}$ $\hat{\sigma}^2 = S_c^2$
$N(\mu, \sigma^2)$, μ conocida	$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$
$Be(p)$	$\hat{p} = \bar{X}$
$B(N, p)$, N conocido	$\hat{p} = \bar{X}/N$

$\mathcal{P}(\lambda)$	$\hat{\lambda} = \bar{X}$
$Ge(p)$	$\hat{p} = 1/(\bar{X}+1)$
$E(\lambda)$	$\hat{\lambda} = 1/\bar{X}$
$U(0, \theta)$,	$\hat{\theta} = 2\bar{X}$

6.4 Propiedades de los estimadores

Cuando se llevan a cabo estimaciones se pretende que el resultado se halle lo más cerca posible del valor desconocido del parámetro. El objetivo de las propiedades que vamos a estudiar en este apartado es obtener un"buen" estimador que aproxime o estime al parámetro θ.

Definición 6.5 *Sea $X_1, X_2, ..., X_n$ una m.a.s. de una población descrita por una variable aleatoria X con F.d.D. F_θ. Sea $T = T(X_1, X_2, ..., X_n)$ un estimador de θ. T es un **estimador insesgado** de θ si $E[T(X_1, X_2, ..., X_n)] = \theta$. Si $E[T(X_1, X_2, ..., X_n)] = \theta + b(\theta)$, se dice que T es un **estimador sesgado** de θ con sesgo $b(\theta)$.*

Ejemplo 6.8 *Sea $X_1, X_2, ..., X_n$ una m.a.s. de una población descrita por una variable aleatoria X con media μ y varianza σ^2 finitas, y $T(X_1, X_2, ..., X_n) = \bar{X}$. Sabemos que en estas condiciones*

$$E[\bar{X}] = \mu,$$

de donde $\bar{X}$ es un estimador insesgado de μ.

Ejemplo 6.9 *Sea $X_1, X_2, ..., X_n$ una m.a.s. de una población descrita por una variable aleatoria X con media μ y varianza σ^2 finitas, y $T(X_1, X_2, ..., X_n) = S_c^2$. Sabemos que en estas condiciones*

$$E[S_c^2] = \sigma^2,$$

de donde S_c^2 es un estimador insesgado de σ^2.

Ejemplo 6.10 *Sea $X_1, X_2, ..., X_n$ una m.a.s. de una población descrita por una variable aleatoria X con media μ y varianza σ^2 finitas, y $T(X_1, X_2, ..., X_n) = S^2$. Sabemos que en estas condiciones*

$$E[S^2] = \sigma^2 \frac{n-1}{n} = \sigma^2 - \frac{1}{n}\sigma^2,$$

de donde S^2 es un estimador sesgado de σ^2 con sesgo $b(\sigma) = -\sigma^2/n$.

Definición 6.6 *Sea $X_1, X_2, ..., X_n$ una m.a.s. de una población descrita por una variable aleatoria X con F.d.D. F_θ, para algún $\theta \in \Theta$. Sean $T = T(X_1, X_2, ..., X_n)$ y $S = S(X_1, X_2, ..., X_n)$ dos estimadores insesgados de θ. Se dice que T es más* ***eficiente*** *que S para estimar θ si*

$$Var[T] \leq Var[S] \; .$$

Ejemplo 6.11 *Sea X_1, X_2, X_3 m.a.s. de X tal que $E[X] = \mu$ y $Var[X] = \sigma^2$. Sean $T_1 = \frac{X_1+X_2+X_3}{3}$ y $T_2 = \frac{X_1+2X_2+X_3}{4}$ dos estadísticos. Vamos a estudiar cuál es más eficiente.*

En primer lugar comprobamos si son insesgados

$$E[T_1] = E\left[\frac{X_1 + X_2 + X_3}{3}\right] = \frac{1}{3}\sum_{i=1}^{n} E[X_i] = \frac{1}{3}3\mu = \mu$$

$$E[T_2] = E\left[\frac{X_1 + 2X_2 + X_3}{4}\right] = \frac{E[X_1] + 2E[X_2] + E[X_3]}{4} = \frac{4}{4}\mu = \mu$$

Por lo tanto ambos estadísticos son insesgados de μ. ¿Cuál de los dos es preferible como estimador de μ? Para contestar vamos a calcular la eficiencia.

$$Var(T_1) = \frac{1}{9}\sum_{i=1}^{n} Var(X_i) = \frac{3\sigma^2}{9} = \frac{\sigma^2}{3}$$

$$Var(T_2) = \frac{1}{16}[Var(X_1) + 4Var(X_2) + Var(X_3)] = \frac{6\sigma^2}{16} = \frac{3\sigma^2}{8}$$

Hemos comprobado que la $Var(T_1) \leq Var(T_2)$, luego T_1 es más eficiente que T_2.

Nota 6.1 *Si los estimadores no son insesgados, la comparación se hace mediante el error cuadrático medio (ECM) y será preferible aquel estimador en el que el error cuadrático medio sea menor.*

$$ECM(T) = E[(T - \theta)^2]$$

6.5 Estimación por Intervalos de Confianza

Ya vimos que la Inferencia Estadística podía ser abordada mediante varias estrategias (estimación puntual, estimación por intervalos, contrastes de hipótesis). En este apartado estudiaremos el problema de la estimación por intervalos de confianza.

Supongamos que queremos estudiar una población descrita por una v.a. X, con F.d.D. F y sea θ un parámetro desconocido de la población. Sea $X_1, X_2, ..., X_n$ una m.a.s. de la población. El problema de la estimación por intervalos consiste en la búsqueda de un intervalo que sea función de la muestra y que contenga al verdadero y desconocido valor del parámetro con una probabilidad especificada de antemano.

Definición 6.7 *Sea $X_1, X_2, ..., X_n$ una m.a.s. de una población descrita por una v.a. X con F.d.D. F_θ, para algún $\theta \in \Theta$. Sean $T_1 = T_1(X_1, X_2, ..., X_n)$ y $T_2 = T_2(X_1, X_2, ..., X_n)$ dos estadísticos tales que $T_1 \leq T_2$ y*

$$P[T_1 < \theta < T_2] = 1 - \alpha, \tag{6.1}$$

*donde $\alpha \in (0, 1)$ es una cantidad fija que no depende de θ. Entonces se dice que (T_1, T_2) es un **intervalo aleatorio de confianza** al nivel $1-\alpha$ para θ. A $1-\alpha$ se le denomina **nivel de confianza**.*

En la práctica se suele tomar $1 - \alpha = 0.90$, 0.95 , 0.99

Para cada realización de la muestra, esto es, para cada valor observado de la muestra, $(x_1, x_2, ..., x_n)$, obtendremos un intervalo, $(T_1(x_1, x_2, ..., x_n), T_2(x_1, x_2, ..., x_n))$, denominado *intervalo de confianza.* Nótese que cada intervalo de confianza resultante de una experimentación, puede o no contener al parámetro θ. La condición (6.1) dice que un $100(1 - \alpha)\%$ de los intervalos que se calculen contendrán al verdadero valor del parámetro.

6.5.1 Intervalos de confianza en una población normal

En este apartado la información muestral siempre procederá de poblaciones con distribución normal.

- **Intervalo de confianza para μ con σ^2 conocida**

De una población desconocemos la media μ y deseamos estimarla a partir de la media muestral $\bar{x}$ obtenida en una muestra de tamaño n.

La distribución muestral de $\bar{X}$ está centrada en μ y en la mayoría de las aplicaciones la varianza es más pequeña que la de otros estimadores cualesquiera de μ. La media muestral $\bar{x}$ se utilizará como estimación puntual para la media de la población μ.

Consideramos una muestra que se selecciona a partir de una distribución normal, a falta de ésta, si n es suficientemente grande, podemos establecer un intervalo de confianza para μ al considerar la distribución muestral de $\bar{X}$. Como vimos en el tema anterior, la distribución de $\bar{X}$ es una normal del media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$. De acuerdo con esto, sabemos que

$$P\left[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right] = 1 - \alpha,$$

donde

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

Por ello

$$P\left[-z_{1-\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha.$$

Al multiplicar cada término en la desigualdad por $\frac{\sigma}{\sqrt{n}}$, y después restar $\bar{X}$ de cada término y multiplicar por -1, obtenemos

$$P\left[\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha.$$

Si $\bar{X}$ es la media de una muestra aleatoria de tamaño n de una población con varianza σ^2 conocida, un intervalo de confianza de $(1-\alpha)\%$ para μ viene dado por

$$\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}.$$

- **Intervalo de confianza para μ con σ^2 desconocida**

De una población deseamos estimar la media cuando la varianza se desconoce. Por el tema anterior sabemos, si tenemos una muestra aleatoria a partir de una distribución normal, entonces la variable aleatoria

$$T = \frac{\bar{X}-\mu}{\frac{S_c}{\sqrt{n}}}$$

tiene una distribución t de Student con n-1 grados de libertad. El procedimiento es el mismo que cuando se conoce σ^2, excepto que σ^2 al ser desconocido se sustituye por la cuasivarianza muestral S_c^2 y la distribución normal se reemplaza por la distribución t de Student. En este caso podemos decir que

$$P\left[-t_{n-1,1-\frac{\alpha}{2}} < T < t_{n-1,1-\frac{\alpha}{2}}\right] = 1-\alpha,$$

de donde obtenemos que

$$P\left[\bar{X} - t_{n-1,1-\frac{\alpha}{2}}\frac{S_c}{\sqrt{n}} < \mu < \bar{X} + t_{n-1,1-\frac{\alpha}{2}}\frac{S_c}{\sqrt{n}}\right] = 1-\alpha.$$

Si $\bar{X}$ y S_c^2 son la media y la cuasivarianza de una muestra aleatoria de una población normal con varianza σ^2 desconocida, un intervalo de confianza de $(1-\alpha)\%$ para μ viene dado por

$$\bar{X} - t_{n-1,1-\frac{\alpha}{2}}\frac{S_c}{\sqrt{n}} < \mu < \bar{X} + t_{n-1,1-\frac{\alpha}{2}}\frac{S_c}{\sqrt{n}}$$

Resumiendo:

- Si σ^2 es conocida $\quad I.C(\mu; 1-\alpha) = \left(\bar{X} \mp z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$
- Si σ^2 es desconocida $\quad I.C(\mu; 1-\alpha) = \left(\bar{X} \mp t_{n-1,1-\frac{\alpha}{2}}\frac{S_c}{\sqrt{n}}\right)$

- Generalmente, cuando la distribución no es normal con σ^2 desconocida y tamaño de muestra grande ($n > 30$), se puede utilizar el siguiente intervalo de confianza para estimar la media μ, (si la varianza σ^2 es desconocida)

$$I.C(\mu \; ; \; 1-\alpha) = \left(\bar{X} \mp z_{1-\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}} \right)$$

- **Intervalo de confianza para σ^2, (si μ es desconocida)**

Se obtiene por un procedimiento similar al utilizado en los casos anteriores

$$I.C\left(\sigma^2 \; ; \; 1-\alpha\right) = \left(\frac{(n-1)S_c^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \quad ; \quad \frac{(n-1)S_c^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \right), \quad \text{con} \quad S_c^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \; .$$

6.5.2 Intervalos de confianza en dos poblaciones normales

Sean $X_1, \ldots, X_n$ m.a. de $N(\mu_1; \sigma_1^2)$; $Y_1, \ldots, Y_m$ m.a. de $N(\mu_2; \sigma_2^2)$ independientes de tamaño n y m respectivamente.

- **Intervalo de Confianza para la diferencia de medias, (si $\sigma_1^2 = \sigma_2^2$ desconocidas):**

$$I.C(\mu_1 - \mu_2; 1-\alpha) = \left((\bar{X} - \bar{Y}) \mp t_{n+m-2,1-\frac{\alpha}{2}} \sqrt{\frac{(n-1)S_{c_1}^2 + (m-1)S_{c_2}^2}{n+m-2} \left(\frac{1}{n} + \frac{1}{m} \right)} \right)$$

- **Intervalo de Confianza para la diferencia de medias, (si $\sigma_1^2 \neq \sigma_2^2$ desconocidas):**

$$I.C(\mu_1 - \mu_2 \; ; 1-\alpha) = \left((\bar{X} - \bar{Y}) \mp t_{\nu,1-\frac{\alpha}{2}} \sqrt{\frac{S_{c_1}^2}{n} + \frac{S_{c_2}^2}{m}} \right), \quad \text{con} \quad \nu = \frac{\left(\frac{S_{c_1}^2}{n} + \frac{S_{c_2}^2}{m} \right)^2}{\left(\frac{S_{c_1}^2}{n} \right)^2 \frac{1}{n-1} + \left(\frac{S_{c_2}^2}{m} \right)^2 \frac{1}{m-1}} - 2 \; .$$

- **Intervalo de Confianza para el cociente de varianzas, (si μ_1, μ_2 desconocidas):**

$$I.C\left(\frac{\sigma_1^2}{\sigma_2^2} \; ; \; 1-\alpha \right) = \left(\frac{S_{c_1}^2}{S_{c_2}^2} \frac{1}{F_{n-1,m-1,1-\frac{\alpha}{2}}} \quad ; \quad \frac{S_{c_1}^2}{S_{c_2}^2} \frac{1}{F_{n-1,m-1,\frac{\alpha}{2}}} \right)$$

6.5.3 Intervalos de confianza para muestras relacionadas

A diferencia del caso de dos poblaciones independientes, para cada individuo de la muestra se observan las dos variables y tendremos situaciones del tipo antes-después, izquierda-derecha, etc. Por ejemplo, medimos la variable tiempo de ejecución de dos algoritmos sobre un mismo conjunto de datos. Bajo esta condición, vamos a calcular la diferencia entre las dos variables y trabajamos con la variable diferencia distribuida normalmente. Para la variable diferencia calculamos la media, $\bar{d}$, y la cuasivarianza S_{cd}^2. De forma similar a los apartados anteriores, se deduce que el intervalo de confianza para la diferencia de medias, μ_D, al $(1-\alpha)\%$ es

$$\bar{d} - t_{n-1,1-\frac{\alpha}{2}} \frac{S_{cd}}{\sqrt{n}} < \mu_D < \bar{d} + t_{n-1,1-\frac{\alpha}{2}} \frac{S_{cd}}{\sqrt{n}}$$

6.5.4 Intervalos de confianza para proporciones

En este apartado queremos calcular un intervalo de confianza para estimar la proporción binomial p de éxitos; por ejemplo dentro del control de calidad nos puede interesar la proporción de artículos defectuosos producidos en una línea de producción.

Por el teorema central del límite, para n suficientemente grande, $\widehat{p}$ (proporción de éxitos de la muestra) está distribuida normalmente con media

$$\mu_{\widehat{p}} = p,$$

y varianza

$$\sigma_{\widehat{p}}^2 = \frac{pq}{n}.$$

Sabemos que

$$P\left[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha,$$

donde

$$Z = \frac{\widehat{p} - p}{\sqrt{\frac{pq}{n}}}$$

y al sustituir Z , tenemos

$$P\left[\widehat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{pq}}{n}} < p < \widehat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{pq}}{n}}\right] = 1-\alpha.$$

Resumiendo, un intervalo de confianza para el parámetro p de una binomial al $(1-\alpha)\%$

$$I.C(p\ ;\ 1-\alpha) = \left(\hat{p} \mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{pq}}{n}}\right)$$

6.5.5 Intervalos de confianza para la diferencia de proporciones

Sean $X_1, \ldots, X_n$ m.a. de $Be(p_1)$; $Y_1, \ldots, Y_m$ m.a. de $Be(p_2)$ independientes, con n y m respectivamente. Se desea estimar la diferencia entre los dos parámetros de las binomiales p_1 y p_2. Sabemos que, si n y m son elevados, $\widehat{p}_1$ y $\widehat{p}_2$ están distribuidos cada uno de forma aproximadamente normal, con medias p_1 y p_2 y varianzas $\frac{p_1 q_1}{n}$ y $\frac{p_2 q_2}{m}$, respectivamente. Al seleccionar muestras independientes de las dos poblaciones, las variables $\widehat{p}_1$ y $\widehat{p}_2$ serán independientes, y $\widehat{p}_1 - \widehat{p}_2$ está distribuida de forma aproximadamente normal con media

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2,$$

y varianza

$$\sigma^2_{\hat{p}_1 - \hat{p}_2} = \frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}.$$

Igual que en casos anteriores, tenemos que,

$$P[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}] = 1 - \alpha,$$

donde

$$Z = \frac{(\widehat{p}_1 - \widehat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}}}$$

sustituyendo Z, obtenemos un intervalo de confianza para la diferencia de proporciones al $(1-\alpha)\%$

$$I.C(p_1 - p_2; 1 - \alpha) = \left((\hat{p}_1 - \hat{p}_2) \mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n} + \frac{\hat{p}_2 \hat{q}_2}{m}} \right)$$

6.6 Aplicación

Consideremos la siguiente aplicación, en ella se tenían las medidas realizadas por Albert Michelson en 1879 sobre la velocidad de la luz en el aire. Se consideró como población X : velocidad de la luz en el aire, $X \sim N(\mu, \sigma^2)$.

- Estimemos los parámetros por intervalos de confianza.

El número de observaciones es $n = 100$. A partir de los datos teníamos: $\bar{x}_n = 852.40$; $s_c^2 = 6242.667$; $s_c = \sqrt{6242.667} = 79.01$.

- Intervalo de Confianza para μ a nivel de confianza de 95%:

$$I.C(\mu; 1 - \alpha) = \left(\bar{x} \mp t_{n-1, 1-\frac{\alpha}{2}} \frac{s_c}{\sqrt{n}} \right) = \left(852.40 - 1.962339 \frac{79.01}{\sqrt{100}} \ , \ 852.40 + 1.962339 \frac{79.01}{\sqrt{100}} \right)$$
$$= (852.40 - 15.504 \ , \ 852.40 + 15.504) = (836.896 \ , \ 867.904) \ .$$

Se ha aproximado $t_{n-1,0.975} = t_{99,0.975} \approx t_{100,0.975} = 1.962339$.

- Intervalo de Confianza para σ^2 a nivel de confianza de 95%:

$$I.C\left(\sigma^2 \ ; \ 1 - \alpha\right) = \left(\frac{(n-1)s_c^2}{\chi^2_{n-1, \, 1-\frac{\alpha}{2}}} \quad ; \quad \frac{(n-1)s_c^2}{\chi^2_{n-1, \frac{\alpha}{2}}} \right)$$

$$= \left(\frac{99 \times 6242.667}{129.561197} \ , \ \frac{99 \times 6242.667}{74.221927} \right) = (4770.13 \ , \ 8326.70) \ .$$

Se han aproximado

$\chi^2_{n-1,0.025} = \chi^2_{99,0.025} \approx \chi^2_{100,0.025} = 74.221927$

$\chi^2_{n-1,0.975} = \chi^2_{99,0.975} \approx \chi^2_{100,0.975} = 129.561197 \quad .$