Capítulo 7

Inferencia estadística: estimación de parámetros.

7.1. Introducción

En este tema estudiaremos como aproximar distintos parámetros poblacionales a partir de una m.a.s. formada por observaciones independientes de una población, en los que sigue cuando digamos m.a.s. entenderemos que es una muestra aleatoria formada por observaciones independientes.

Normalmente el parámetro (por ejemplo μ, $\sigma \ldots$) tendrá distribución conocida o la aproximaremos por el T.L.C.

7.2. Estimadores

Definición 115 ***Estadístico****: Sean $X_1, \ldots, X_n$ n v.a. iid que forman una m.a.s. de una población. Un estadístico es una función de una de una muestra.*

Podemos decir que un estadístico una variable aleatoria que es función de la muestra.

Definición 116 ***Estimador puntual****: Un* estimador puntual *de un parámetro θ es un estadístico que da como resultado un único valor del que se espera que se aproxime a θ.*

Una realización del estimador $T(x_1, \ldots, x_n) = \hat{\theta}$ *en una muestra se llama* estimación puntual de parámetro.

Ejemplo 117 *Dada una m.a.s. $X_1, \ldots, X_n$ y una realización de la misma $x_1, \ldots, x_n$ los principales estimadores de los parámetros poblacionales que hemos visto son:*

Parámetro Poblacional	*Estimador(θ)*	*Estimación($\hat{\theta}$)*
μ_X	$\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$	$\overline{x} = \frac{\sum_{i=1}^n x_i}{n}$
σ_X	$\tilde{S}_X = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$	$\tilde{s}_X = \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n-1}$
p	$\hat{p}_X = \frac{\sum_1^n X_i}{n}$	$\frac{\sum_1^n x_i}{n}$

Ejemplo 118 *Consideremos una m.a.s. X_1, X_2, X_3, X_4, X_5 del lanzamiento de un dado ($n = 5$).*

Una realización de esta muestra es $x_1 = 2, x_2 = 3, x_3 = 3, x_4 = 5, x_5 = 6$.

Sabemos que, si el dado es perfecto, $\mu = 3.5$; el estadístico de esta muestra es

$$\overline{X} = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$$

y una estimación es

$$\overline{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{2 + 3 + 3 + 5 + 6}{5} = \frac{19}{5} = 3.8$$

Si queremos estimar la proporción de veces que sale 3 es $p_3 = \frac{1}{6}$
el estadístico es

$$\hat{p}_3 = \frac{\textit{frec. de 3 en la muestra}}{5}$$

y una realización será $\frac{2}{5}$.

7.2.1. Estimadores insesgados

Vamos a ver en esta sección algunas propiedades de los estimadores. La más inmediata es pedirles que a medida que se aumente el tamaño muestral se aproximen más al verdadero valor del parámetro.

Definición 119 ***Estimador insesgado*** *Sea* $\hat{\theta}$ *un estimador de un parámetro poblacional* θ. *Diremos que* $\hat{\theta}$ *es insesgado si* $E(\hat{\theta}) = \theta$.

Es este caso la estimación puntual se dice que es insesgada.

Ejemplo 120 *En el ejemplo anterior para cualquier muestra de tamaño* n, $X_1, \ldots, X_n$, *tenemos que* $E(\overline{X}) = \mu_X$ *por lo tanto* $\overline{X}$ *es un estimador insesgado de* μ_X.

Proposición 121 *Dada una m.a.s. La media, varianza y proporción muestrales son estimadores insesgados de sus correspondientes parámetros poblacionales.*

Definición 122 ***Sesgo****: Sea* $\hat{\theta}$ *un estimador puntual de un parámetro poblacional* θ, *llamaremos sesgo de* $\hat{\theta}$ *a:*

$$Sesgo(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Observación Evidentemente un estimador es insesgado si y sólo si tiene sesgo cero.

Una propiedad buena para un estimador es la carencia de sesgo. Pero podría suceder que tuviera una gran variabilidad, entonces aunque su valor central sea el verdadero valor del parámetro que se estima una realización del estadístico podría estar lejos del verdadero valor del parámetro. Parece pues interesante emplear aquellos estimadores que tengan varianza más pequeña.

Definición 123 ***Eficiencia****: Sean* $\hat{\theta}_1$ *y* $\hat{\theta}_2$ *dos estimadores de un parámetro poblacional* θ *obtenidos de la misma muestra.*

a) Diremos que $\hat{\theta}_1$ *es más eficiente que* $\hat{\theta}_2$ *si* $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$

b) Definimos la eficiencia relativa de $\hat{\theta}_2$ *respecto de* $\hat{\theta}_1$ *como*

$Eff.rel = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}$

de forma que si $Eff.rel < 1$ *entonces* $\hat{\theta}_1$ *es más eficiente que* $\hat{\theta}_2$

Ejemplo 124 *Sea* $x_1, \ldots, x_n$ *la realización ordenada de menor a mayor de una muestra de tamaño* n. *Se define la mediana muestral como*

$$Me = \begin{cases} x_{\frac{n+1}{2}} & \textit{si } n \textit{ es impar} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \textit{si } n \textit{ es par} \end{cases}$$

Como vimos en problemas la mediana es también un valor de tendencia central, pero ¿es un buen estimador de μ?

Se puede demostrar que si la población tiene distribución normal con media μ *y varianza* σ_X^2 *entonces* $E(Me) = \mu$ *y* $Var(Me) = \frac{\pi}{2}\frac{\sigma_X^2}{n} \approx \frac{1.57\sigma_X^2}{n}$

entonces $Eff.rel = \frac{Var(Me)}{Var(\overline{x})} = 1.57$

Luego si la muestra es de una población normal $\overline{X}$ *es más eficiente (un 57 % más de varianza) que la Mediana.*

Definición 125 ***Estimador más eficiente:***

Diremos que un estimador insesgado $\hat{\theta}$ *del parámetro* θ *es el estimador más eficiente si no existe ningún otro estimador insesgado que tenga menor varianza que él (también se le denomina estimador insesgado de varianza mínima).*

Ejemplo 126 [1]

- *Si la población es normal la media muestral es el estimador insesgado más eficiente de la media poblacional.*
- *Si la población es normal la varianza muestral es el estimador insesgado más eficiente de la varianza poblacional.*
- *Si la población es binomial la proporción muestral es el estimador insesgado más eficiente de la proporción poblacional.*

7.3. Métodos para calcular estimadores

Existen diversos métodos para el cálculo de estimadores:

- Método de los momentos. Momento central de orden r
  $m_r = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^r}{n}$
- El de menor error cuadrático medio
  $E((\hat{\theta}-\theta)^2)$
- Convergencia en probabilidad
  $P(|\hat{\theta}_n-\theta|<\epsilon)\to 1$
- Estimadores máximo verosímiles.

En esta sección veremos sólo este último método.

Estimadores máximo verosímiles.

Definición 127 ***Función de verosimilitud***
Sea X una v.a. tal que su distribución (densidad o función de probabilidad) depende de un parámetro desconocido λ (En el caso discreto $P_X(x;\lambda)$ y en el continuo $f_X(x;\lambda)$). Sea $X_1,\ldots X_n$ una m.a.s. de X (es decir son n v.a. iid como X) y sean $x_1,x_2,\ldots,x_n$ una realización de la muestra. Entonces la función de verosimilitud de la muestra es:

a) En el caso discreto $L(\lambda)=P_X(x_1;\lambda)\cdots P_X(x_n;\lambda)$

b) En el caso continuo $L(\lambda)=f_X(x_1;\lambda)\cdots f_X(x_n;\lambda)$

Definición 128 *Dada una función de verosimilitud $L(\lambda)$ de una muestra, sea $\hat{\lambda}=g(x_1,\ldots,x_n)$ el punto donde se alcanza en máximo de $L(\lambda)$ para la realización de la muestra $x_1,\ldots,x_n$, es decir $L(\hat{\lambda})=\max\limits_{\lambda} L(\lambda)$. Entonces definimos el estimador de máxima verosimilitud de λ como el valor:*

$$\hat{\Lambda}=g(X_1,\ldots,X_n)$$

En ocasiones es conveniente trabajar con el logaritmo de la función de verosimilitud ya que el máximo de $\log(L(\lambda))$ y $L(\lambda)$ es el mismo y suele ser más fácil de maximizar.

Ejemplo 129 *Sea $X_1,\ldots X_n$ una muestra con observaciones independientes, de una población Bernouilli, por ejemplo se pregunta a 100 personas si votarán al partido P.X. en las próximas elecciones y se anota un 1 si lo votan y cero en cualquier otro caso. Sea $\hat{\theta}=T(X_1,\ldots,X_n)$ un estimador cualquiera. Sea p la proporción poblacional de personas que votarán a P.X. Entonces*

$$P(X_i=1)=p \text{ y } P(X_i=0)=1-p=q,$$

o lo que es lo mismo

[1] Más concretamente estos estimadores son del tipo UMVUE del acrónimo inglés "Uniformly Minimum Variance Unbiased Estimator": Estimadores insesgados uniformemente de mínima varianza".

$$P(X = x_i) = p^{x_i} q^{1-x_i} \text{ si } x_i = 0, 1$$

Como las observaciones son independientes. la función de verosimilitud es:

$$L(p) = P_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X_1 = x_1,\ldots,X_n = x_n) = P(X_1 = x_1)\cdots P(X_n = x_n) =$$
$$p^{x_1} q^{1-x_1} \cdots p^{x_n} q^{1-x_n} = p^{\sum_{i=1}^n x_i} q^{\sum_{i=1}^n (1-x_i)} = p^{\sum_{i=1}^n x_i} q^{n-\sum_{i=1}^n x_i}$$

entonces el valor de p que hace máxima esta probabilidad es el más verosímil o el de máxima verosimilitud de esta muestra.

El problema se reduce a estudiar qué valor de p maximiza

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

Tomando logaritmos

$$\log\left(p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}\right) = \sum_{i=1}^n x_i \log p + (n - \sum_{i=1}^n x_i)\log(1-p)$$

Derivando respecto de p

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

Despejando

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

por lo tanto

$$p = \frac{\sum_{i=1}^n x_i}{n}$$

luego el estimador máximo verosímil de p es la proporción muestral, que es el que maximiza la función de verosimilitud $L(p)$.

De modo similar se puede definir los estimadores máximo verosímiles cuando el número de parámetros no conocidos de la distribución son más de uno.

7.4. Estimación por intervalos

Una estimación por intervalos de un parámetro poblacional es una regla para determinar un rango o un intervalo donde, con cierta probabilidad, se encuentre el verdadero valor del parámetro. La estimación correspondiente se llama estimación por intervalo. Más formalmente:

Definición 130 *Sea θ un parámetro, el intervalo (A, B) es un intervalo de confianza del $(1-\alpha)\dot{1}00\,\%$ para el parámetro θ si*

$$P(A < \theta < B) = 1 - \alpha.$$

El valor $1-\alpha$ recibe el nombre de nivel de confianza, α es la "cola"de probabilidad sobrante que normalmente se reparte por igual $(\alpha/2)$ a cada lado del intervalo. Es muy frecuente que el nivel de confianza se dé en tanto por ciento.

En lo que sigue daremos distintas maneras de calcular o aproximar intervalos de confianza para distintos parámetros.

7.4.1. Intervalo de confianza para la media de una población normal: varianza poblacional conocida

Sea $X_1, \ldots, X_n$ una m.a.s. de una v.a. X con distribución normal y $Var(X) = \sigma^2$ conocida.
Encontremos un intervalo de confianza al *nivel de confianza* del 90 % para la media poblacional μ.

Sabemos por el tema anterior que bajo estas condiciones la variable $Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ sigue una distribución normal estándar pues es una trasformación lineal de una combinación lineal de variables normales e independientes..

Ejemplo 131 *Comencemos calculando un intervalo centrado en* 0 *para esta* Z *que tenga probabilidad* 0.975.

$$0.975 = P(-\delta < Z < \delta) = F_Z(\delta) - F_Z(-\delta) = 2F_Z(\delta) - 1$$

Entonces

$$F_Z(\delta) = \frac{1.975}{2} = 0.9875$$

mirando en las tablas de la distribución normal estándar, entonces $F_Z(2.24) = 0.9875$ *y por lo tanto* $\delta = 2.24$
Luego $P(-2.24 < Z < 2.24) = 0.975$
En resumen, hemos obtenido lo siguiente

$$0.975 = P(-2.24 < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < 2.24) =$$

$$P(\overline{X} - 2.24\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 2.24\frac{\sigma}{\sqrt{n}})$$

Hemos encontrado un intervalo de confianza para μ*, y además la probabilidad de que* μ *se encuentre en el intervalo*
$\left(\overline{X} - 2.24\frac{\sigma}{\sqrt{n}}, \overline{X} + 2.24\frac{\sigma}{\sqrt{n}}\right)$
es 0.975; luego es un intervalo de confianza con nivel de confianza 97.5 %

Ejemplo 132 *Supongamos que tenemos una muestra con* $n = 16$ *de una v.a. normal de forma que* $\overline{x} = 20$*, y la desviación típica poblacional es conocida* $\sigma = 4$*. Entonces un intervalo de confianza al 97.5 % para* μ *será:*

$$\left(20 - \frac{(2.24)4}{\sqrt{16}}, 20 + \frac{(2.24)4}{\sqrt{16}}\right)$$

La probabilidad con que el verdadero valor del parámetro μ *se encuentra en el intervalo* $(17.76, 22.24)$ *es* 0.975*, o lo que es lo mismo:*
$P(17.76 < \mu < 22.24) = 0.975$

Interpretación: *En el 97.5 % de la muestras de tamaño 16 el verdadero valor del parámetro* μ *se encontrará dentro del intervalo correspondiente.*

En general si tenemos una m.a.s. $X_1, \ldots, X_n$ de una población normal (representado por la v.a. X) con distribución normal de media μ y varianza conocida σ^2 el intervalo de confianza para μ al nivel de confianza $(1-\alpha)\cdot 100\,\%$ será:

$$1 - \alpha = P(z_{\alpha/2} < Z < z_{1-\alpha/2}) = P(z_{\alpha/2} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{1-\alpha/2}) =$$

$$P(z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}) = P(\overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$$

Resumen: Intervalo de confianza para μ: σ^2 conocida.

Condiciones:

a) Población Normal con media μ y varianza σ^2 conocida

b) Muestra aleatoria de tamaño n

Entonces el intervalo de confianza del $100(1-\alpha)\,\%$ para μ es:

$$\left(\overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

donde $z_{\frac{\alpha}{2}}$ es el cuantil $\frac{\alpha}{2}$, es decir $P(Z \leq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$, cuando Z tiene distribución normal estándar, mientras que $z_{1-\frac{\alpha}{2}}$ es el cuantil $1-\frac{\alpha}{2}$, es decir $P(Z \leq z_{1-\frac{\alpha}{2}}) = 1-\frac{\alpha}{2}$, cuando Z tiene distribución normal estándar. Notemos que $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$

Ejemplo 133 *Para discutir la conveniencia de aumentar sus instalaciones una empresa desea estimar la demanda que espera recibir. Para ello, selecciona al azar a diez de sus clientes, observando el número de unidades demandadas en el último año por éstos se distribuye de la forma siguiente:*

Núm. Unidades	*Núm. Clientes*	*Unidades× Clientes*
1.000	*1*	*1.000*
1.002	*2*	*2.004*
1.004	*1*	*1.004*
1.006	*2*	*2.012*
1.008	*1*	*1.008*
1.010	*2*	*2.020*
1.012	*1*	*1.012*
Total	*10*	*10.06*

Supongamos que la demanda sigue una distribución normal con varianza poblacional conocida $\sigma^2 = 16$ y que se espera que en el futuro siga comportándose como en el periodo anterior, calcular un intervalo de confianza al 90 % para la media de la demanda futura.

Solución: Tenemos las siguientes condiciones:

- Población de demandas normal varianza $\sigma^2 = 16$ conocida
- Muestra aleatoria de tamaño $n = 10$

Podemos entonces aplicar la formula anterior para $1-\alpha = 0.9$, de donde $\alpha = 0.1$, entonces $\frac{\alpha}{2} = 0.05$ y $1-\frac{\alpha}{2} = 0.95$

Calculemos la media aritmética de las observaciones

$$\overline{x} = \frac{10.06}{10} = 1.006,$$

entonces el intervalo es

$$\left(1.006 + z_{0.05}\frac{4}{10}, 1.006 + z_{1-0.05}\frac{4}{10}\right).$$

Mirando en las tablas de la normal $P(Z \leq 1.65) = 0.9505 \approx 0.95$ entonces $z_{0.95} = 1.65$, y $z_{0.05} = -1.65$ sustituyendo tenemos que

$z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 1.65\frac{4}{10} = 0.66$ $z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = -1.65\frac{4}{10} = 0.66$

por lo que el intervalo de confianza del 90 % para la media de la demanda es :

$$(1.006 - 0.66, 1.006 + 0.66) = (0.346, 1.666)$$

Lo que quiere decir que en el 90 % de la ocasiones en que tomemos una muestra de tamaño 10 la demanda media está comprendida entre 0.346 y 1.666

Amplitud del intervalo de confianza

Como de todos es conocido la amplitud (longitud) de un intervalo es la diferencia entre sus extremos superior e inferior. En el caso anterior la amplitud A será

$A = \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} - \left(\overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

El *error* máximo, al nivel $(1-\alpha)$, que cometemos al estimar μ por $\overline{X}$ será la mitad de la amplitud del intervalo de confianza $z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

Si queremos calcular el tamaño n de la muestra para asegurarnos que el intervalo de confianza para μ al nivel $(1-\alpha)$ tiene amplitud prefijada A (o un error $\frac{A}{2}$) se puede despejar así:

$n = \left(2z_{1-\frac{\alpha}{2}} \frac{\sigma}{A}\right)^2$

Observaciones:

- El intervalo está centrado en $\overline{X}$.
- Para n y $1-\alpha$ fijos si la varianza poblacional aumenta entonces A aumenta.
- Para una varianza poblacional conocida y $1-\alpha$ fijos si n aumenta entonces A disminuye.
- Para una varianza poblacional conocida y n fijos si $1-\alpha$ aumenta entonces A aumenta.

7.4.2. Intervalo de confianza para la media poblacional: tamaños muestrales grandes

Condiciones:

- Población con media μ y varianza σ^2 conocida o si no se estima por S^2
- Muestra aleatoria de tamaño n grande (criterio $n \geq 30$)

Entonces el intervalo de confianza del $100(1-\alpha)\,\%$ para μ es:

$$\left(\overline{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right)$$

En caso de que σ sea conocida pondremos σ en lugar de S

Ejemplo 134 *Se tomó una muestra de 147 expertos en investigación de mercados y se les pidió que calificasen en una escala de 1 (totalmente en desacuerdo) a 10 (totalmente de acuerdo) la siguiente afirmación: "A veces utilizo técnicas de investigación que garantizan la obtención de los resultados que mi cliente o jefe desea". La calificación media de la muestra fue* 6.06 *y la desviación típica muestral fue 1.43. Se pide calcular un intervalo de confianza al 90 % para la media de las puntuaciones.*

Solución: El enunciado no nos asegura que la población sea normal pero como el tamaño de la población es grande podemos aplicar el resultado anterior.

Tenemos $n = 147$, $S = 1.43$, $1-\alpha = 0.1$ entonces $\frac{\alpha}{2} = 0.05$ y por lo tanto $z_{1-0.05} \approx 1.65$

El intervalo para la media poblacional de las puntuaciones al nivel de confianza del $90\,\%$ es

$\left(6.06 - 1.65\frac{1.43}{\sqrt{147}}, 6.06 + 1.65\frac{1.43}{\sqrt{147}}\right) =$

$(5.8654, 6.2546)$

Distribución t de Student

Si queremos calcular un intervalo de confianza para μ en una población normal con varianza poblacional desconocida necesitamos una nueva distribución: la t de Student.

Dada una muestra de n observaciones con media muestral $\overline{X}$ y desviación típica muestral $\tilde{S}_X$ procedente de una población normal con media μ la variable aleatoria:

$$t = \frac{\overline{X} - \mu}{\frac{\tilde{S}_X}{\sqrt{n}}}$$

sigue una distribución t de Student con $n-1$ grados de libertad.

Proposición 135 *La distribución t de Student es similar a la normal si el número de grados de libertad es grande. Su función de densidad es simétrica respecto al origen como la de la normal estándar. Es decir si t_ν es una v.a. que sigue la distribución t de Student con ν g.l. entonces:*

$$P(t_\nu \leq -t) = 1 - P(t_\nu \leq t)$$

Notación: Sea t_ν una v.a. que sigue una distribución t de Student con ν g.l. Denotaremos por $t_{\nu,\alpha}$ al valor para el que se verifica que:

$$P(t_\nu \leq t_{\nu,\alpha}) = \alpha.$$

Luego $t_{\nu,\alpha}$ es el α cuantil de una t de Student con ν g.l. y $t_{\nu,\alpha} = -t_{\nu,1-\alpha}$.

7.4.3. Intervalo de confianza para la media de una población normal: varianza poblacional desconocida

Condiciones:

- Muestra aleatoria de n observaciones independientes
- Población normal varianza desconocida

Entonces si $\overline{X}$ y $\tilde{S}_X$ son respectivamente la media y la desviación típica muestrales un intervalo de confianza al nivel $(1-\alpha)100\,\%$ para la media de la población μ es:

$$\left(\overline{X} + t_{n-1,\frac{\alpha}{2}} \frac{\tilde{S}_X}{\sqrt{n}}, \overline{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{\tilde{S}_X}{\sqrt{n}}\right)$$

Siendo $t_{n-1,\frac{\alpha}{2}}$ y $t_{n-1,\frac{\alpha}{2}}$ los cuantiles de una v.a. t_{n-1} con distribución t de Student con n-1 g.l., respectivamente.

Ejercicio Demostrar que la probabilidad con que μ se encuentra en el intervalo anterior es $1-\alpha$

Ejemplo 136 *Un fabricante de cartuchos de tinta para impresoras afirma en su publicidad que sus cartuchos imprimirán un promedio de 500 páginas*; donde el asterisco remite a una nota a pie de página donde afirma que: “* `Datos técnicos: Muestra mensual de tamaño` $n = 25$ `población supuesta normal nivel de confianza del 90%`*”.*

Una organización de consumidores desea comprobar estas afirmaciones y toma también una muestra al azar de tamaño $n = 25$ obteniendo como media $\overline{x} = 518$ páginas y una desviación estándar $\tilde{S}_X = 40$. Comprobar que con esta muestra la media poblacional que afirma el fabricante cae dentro del intervalo de confianza del 90 %

Solución: El problema se reduce a calcular, bajo las condiciones que afirma el fabricante el intervalo de confianza para μ con $\alpha = 0.1$.

Mirando en las tablas de la t de Student para $n-1 = 24$ g.l. tenemos que $t_{n-1}, 1-\frac{\alpha}{2} = t_{24,1-0.05} = 1.711$
El intervalo para la media al 90 % es

$$\left(518 - 1.711 \frac{40}{\sqrt{25}}, 518 + 1.711 \frac{40}{\sqrt{25}}\right) = (504.312, 531.688)\,.$$

Es este caso la afirmación del fabricante queda contradicha por la muestra pues 500 cae fuera del intervalo. En cualquier caso se equivoca a favor del consumidor.

7.4.4. Intervalos de confianza para una proporción

El procedimiento es similar al caso de las medias. Comencemos con un ejemplo.

Ejemplo 137 *En una muestra aleatoria de 500 familias que poseen televisores en una ciudad se encontró que 340 se habían suscrito al canal TEVE. Encontrar un intervalo de confianza del 95 % para la proporción actual de familias de esta ciudad que están suscritas a TEVE.*

Tenemos una población binomial donde los éxitos son las familias que tienen contrato con TEVE. Sea X el número de familias contratadas con TEVE entre una muestra aleatoria de tamaño n. Entonces X sigue una distribución binomial con n repeticiones y probabilidad de éxito p (proporción poblacional de familias contratadas a TEVE). Si llamamos $\hat{p}_X = \frac{X}{n}$ a la proporción muestral, sabemos que $Z = \frac{\hat{p}_X - p}{\sqrt{\frac{p(1-p)}{n}}}$ sigue aproximadamente una distribución normal estándar.

Pero como es evidente no conocemos p así que no tenemos más remedio que aproximar el denominador

$$\sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}$$

Si la muestra es grande $Z = \frac{\hat{p}_X - p}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}}$ seguirá siendo aproximadamente normal estándar.

Intervalos de confianza para la proporción poblacional:(muestras grandes)

Condiciones:

- Una muestra aleatoria de tamaño n grande.
- Población Bernouilli con proporción de éxitos p (desconocida)

Bajo estas condiciones y si $\hat{p}_X$ es la proporción de éxitos en la muestra, un intervalo de confianza al nivel $(1-\alpha)100\,\%$ es

$$\left(\hat{p}_X - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}, \hat{p}_X + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}\right)$$

Criterio: los intervalos de confianza anteriores son fiables si $n \geq 40$.
Observaciones

- El intervalo de confianza anterior está centrado en la proporción muestral.
- Cuando n crece se reduce la amplitud del intervalo de confianza.
- La amplitud del intervalo de confianza es $A = z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}$
- De la fórmula anterior no podemos determinar el tamaño de la muestral sin conocer $\hat{p}_X$ así que nos podremos en el caso peor:

  El máximo de
  $\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}$
  se alcanza en $\hat{p}_X = 0.5$ y en este caso
  $\sqrt{\frac{0.5(1-0.5)}{n}}$ por lo tanto en el peor de los casos[2]
  $n = \frac{0.25 z^2_{1-\frac{\alpha}{2}}}{A^2}$.

7.4.5. Intervalo de confianza para la varianza de una población normal

Recordemos que si tenemos una población normal con varianza σ^2 y una muestra aleatoria de tamaño n de esta población con varianza muestral $\tilde{S}_X^2$ entonces el estadístico

$$\chi^2_{n-1} = \frac{(n-1)S_X^2}{\sigma^2}$$

sigue una distribución χ^2 con $n-1$ g.l.

Notación Si χ^2_ν es una v.a. que tiene distribución χ^2 con ν g.l. denotaremos por $\chi^2_{\nu,\alpha}$ al valor que verifica:

[2] Por esto en las especificaciones o detalles técnicos de las encuestas se suele leer, por ejemplo: “Universo población Balear mayor de 18 años. Encuesta telefónica, selección aleatoria, de tamaño mil, error en las proporciones $\pm 3\,\%$ con una confianza del $95\,\%$ supuesto que $p = q = \frac{1}{2}$”

$$P(\chi^2_\nu \leq \chi^2_{\nu,\alpha}) = \alpha$$

es decir el cuantil $\frac{\alpha}{2}$ de una v.a. con distribución χ^2_ν. Estos valores están tabulados para distintos g.l. en la tabla de la distribución χ^2.

Ejemplo 138 *Sea* χ^2_{10} *una v.a. que tiene distribución* χ^2 *con* 10 *g.l. Entonces* $\chi^2_{10,0.995} = 25.19$ *y* $\chi^2_{10,0.005} = 2.16$*, es decir*

$$P(\chi^2_{10} \leq 25.19) = 0.995 \text{ y } P(\chi^2_{10} \leq 2.16) = 0.005$$

Además tendremos que

$$P(2.16 \leq \chi^2_{10} \leq 25.19) = P(\chi^2_{10} \leq 25.19) - P(\chi^2_{10} \leq 2.16) = (1 - 0.005) - (1 - 0.995) = 0.995 - 0.005 = 0.99$$

En general dado α entre 0 y 1 tendremos que

$$1 - \alpha = P(\chi^2_{\nu,\frac{\alpha}{2}} \leq \chi^2_\nu \leq \chi^2_{\nu,1-\frac{\alpha}{2}})$$

Si tenemos una muestra de tamaño n de una población normal con desviación típica muestral $\tilde{S}^2_X$, dado un nivel de confianza $1 - \alpha$ tendremos que $\chi^2_{n-1} = \frac{(n-1)\tilde{S}^2_X}{\sigma^2}$ y entonces:

$$1 - \alpha = P(\chi^2_{n-1,\frac{\alpha}{2}} \leq \chi^2_{n-1} \leq \chi^2_{n-1,1-\frac{\alpha}{2}}) =$$

$$P(\chi^2_{n-1,\frac{\alpha}{2}} \leq \frac{(n-1)S^2_X}{\sigma^2} \leq \chi^2_{n-1,1-\frac{\alpha}{2}}) = P(\frac{(n-1)\tilde{S}^2_X}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)\tilde{S}^2_X}{\chi^2_{n-1,\frac{\alpha}{2}}})$$

Luego, bajo estas condiciones, un intervalo de confianza para la varianza poblacional del $(1 - \alpha)100\,\%$ es

$$\left(\frac{(n-1)\tilde{S}^2_X}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)\tilde{S}^2_X}{\chi^2_{n-1,\frac{\alpha}{2}}}\right).$$

Intervalo de confianza para la varianza de una población normal

Condiciones

- Población normal
- Muestra aleatoria de tamaño n con varianza muestral S^2_X

Entonces un intervalo de confianza del $(1 - \alpha)100\,\%$ es

$$\left(\frac{(n-1)\tilde{S}^2_X}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)\tilde{S}^2_X}{\chi^2_{n-1,\frac{\alpha}{2}}}\right)$$

Donde $\chi^2_{n-1,\frac{\alpha}{2}}$ es el valor que verifica

$$P(\chi^2_{n-1} < \chi^2_{n-1,\frac{\alpha}{2}}) = \frac{\alpha}{2}$$

y

$$\chi^2_{n-1,1-\frac{\alpha}{2}}$$

es el valor tal que

$$P(\chi^2_{n-1} \leq \chi^2_{n-1,1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$

donde χ^2_{n-1} es una v.a. que sigue una distribución χ^2 con $n - 1$ g.l.
Observación El intervalo de confianza para σ^2 no está centrado en $\tilde{S}^2_X$.

Ejemplo 139 *Una cadena de hoteles tiene una* Línea 900 *para recibir reservas telefónicas. Un índice de la calidad del servicio es el tiempo de espera, el tiempo que transcurre desde que el teléfono suena por primera vez hasta que el operador responde. El estándar de la cadena es que el tiempo promedio de espera no debe ser superior a 30 segundos además se supone que la distribución del tiempo de espera será aproximadamente normal. La cadena tiene inspectores que visitan los distintos hoteles y verifican todos los aspectos del servicio. Estas personas realizan cada semana 30 llamadas para hacer reservas y anotan, entre otros indicadores el tiempo de espera en cad una de ellas. En una semana los tiempos de espera en segundos son:*

12, 13, 13, 14, 14, 14, 15, 15, 16, 17, 17, 18, 18, 19, 19, 25, 25, 26, 27, 30, 33, 34, 35, 40, 40, 51, 51, 58, 59, 83

Calcular un intervalo de confianza para la varianza y la desviación poblacionales al nivel $95\,\%$.

Solución: *Sea* X *el tiempo de espera. Haciendo los cálculos tenemos que (redondeando al segundo decimal):* $\overline{X} = 28.37$ *y* $\tilde{s}_X = 17.37$

Como $1-\alpha = 0.95$ *tenemos que* $\frac{\alpha}{2} = 0.025$*, entonces mirando en las tablas de la* χ^2 *(y redondeando también al segundo decimal)*

$$\chi^2_{n-1,\frac{\alpha}{2}} = \chi^2_{29,0.975} = 45.72 \text{ y } \chi_{n-1}, 1-\frac{\alpha}{2}^2 = \chi^2_{29,0.025} = 16.05.$$

Por lo tanto un intervalo de confianza del 95 % para σ^2 *es*

$$\left(\frac{(30-1)(17.37)^2}{45.72}, \frac{(30-1)(17.37)^2}{16.05}\right) = (191.38, 545.16)$$

Es decir $P(191.28 \leq \sigma^2 \leq 545.16) = 0.95$ *y operando tenemos que*

$$P(\sqrt{191.28} \leq \sigma \leq \sqrt{545.16}) = 0.95,$$

luego un intervalo de confianza del 95 % para σ *es*

$$(13.83, 23.35)\,.$$