Inferencia

Estadística

(Teoría y problemas)

I. Espejo Miranda
F. Fernández Palacín
M. A. López Sánchez
M. Muñoz Márquez
A. M. Rodríguez Chía
A. Sánchez Navas
C. Valero Franco

Edita: Servicio de Publicaciones de la Universidad de Cádiz
c/ Doctor Marañón, 3. 11002 Cádiz (España)
www.uca.es/publicaciones

ISBN: 978-84-9828-131-6

Inferencia Estadística (Revisión: Marzo 2007)
I. Espejo Miranda, F. Fernández Palacín, M. A. López Sánchez, M. Muñoz Márquez, A. M. Rodríguez Chía, A. Sánchez Navas, C. Valero Franco

Capítulo 3

Estimación por intervalos de confianza

1. Introducción

En el capítulo anterior se estudiaron tanto las propiedades deseables para un estimador de un determinado parámetro poblacional, como la forma de calcularlos. Estos estimadores proporcionan, para una muestra concreta, un valor puntual que se denomina estimación puntual. Sin embargo, a pesar de la indudable utilidad de este procedimiento, en la práctica, cuando se realiza la estimación de un parámetro se necesita obtener una medida de la fiabilidad de dicha estimación, medida de la que se carece en el proceso de estimación puntual, de este modo, surge la necesidad de encontrar un método que permita calcular una región que contenga al valor del parámetro con una cierta garantía.

El capítulo se ha organizado introduciendo en primer lugar el Método del Pivote, para calcular seguidamente Intervalos de confianza de parámetros en poblaciones Normales basándose en el mismo. A continuación se han introducido otros métodos de obtención de estimadores que se han considerado interesantes, pero que dada su complejidad o su relativo uso vienen marcados con ***. En concreto, el Método asintótico basado en el Teorema central del límite va a permitir calcular intervalos en poblaciones Binomiales. El tema concluye con un epígrafe dedicado a la determinación del tamaño muestral para cumplir con el objetivo de precisión establecido.

Existen innumerables situaciones reales donde es necesario encontrar regiones en las cuales se tenga la confianza o cierto grado de seguridad de que en ellas se halle el valor de un parámetro desconocido de la población. A modo de ejemplos:

Ejemplo 3.1 *Un vendedor desea establecer la duración de la garantía de un determinado electrodoméstico, de forma que durante el período de garantía deba sustituir el menor número posible de piezas. El tiempo hasta el primer fallo, viene dado por una variable aleatoria, X, tal que, $\mathrm{E}[X] = \theta$, donde θ es un parámetro desconocido.*

Si el vendedor no quiere pagar ninguna pieza, el tiempo de garantía debería de ser nulo, pero ésto supondría una mala imagen cara al público, con los consecuentes perjuicios. Por tanto, deberá buscar una cota inferior del tiempo hasta que se produzca el primer fallo del electrodoméstico, "confiando" en que la vida media de ese electrodoméstico sea superior a esa cota inferior. Es decir, extraída una m.a.s., $\underline{X}$, de la población, para $\alpha > 0$ se busca $\underline{\theta}(\underline{X})$ tal que

$$P[\underline{\theta}(\underline{X}) \leq \theta] \quad \geq \quad 1 - \alpha.$$

Ejemplo 3.2 *Un laboratorio está interesado en estudiar la toxicidad media de un determinado producto químico, para ello quiere establecer una cota superior de dicha media y así tener cierta certeza o seguridad de que la toxicidad del producto estará por debajo de esa cota superior. Por tanto, si la toxicidad del producto viene dada por una variable aleatoria, X, tal que $\mathrm{E}[X] = \theta$ donde θ es un parámetro desconocido, se quiere obtener una cota superior del nivel de toxicidad medio "confiando" en que dicho nivel se encuentre por debajo de esa cota. Es decir, ex-*

traída una m.a.s., $\underline{X}$, de esta población, para $\alpha > 0$ se busca $\overline{\theta}(\underline{X})$ tal que

$$P[\theta \leq \overline{\theta}(\underline{X})] \quad \geq \quad 1 - \alpha.$$

Ejemplo 3.3 *Una empresa tabaquera desea estudiar el nivel medio de nicotina de sus cigarros. A la compañía le interesa que el nivel medio de nicotina se encuentre entre unos márgenes debido a que un nivel medio alto supone que el cigarro es muy perjudicial para la salud y un nivel medio bajo implica que el cigarro carece de sabor. De este modo, si el nivel de nicotina de un cigarro viene dado por una variable aleatoria, X, tal que $\mathrm{E}[X] = \theta$, donde θ es un parámetro desconocido, se desea, a partir de una m.a.s., $\underline{X}$, y para $\alpha > 0$ obtener $\underline{\theta}(\underline{X})$ y $\overline{\theta}(\underline{X})$ tal que*

$$P[\underline{\theta}(\underline{X}) \leq \theta \leq \overline{\theta}(\underline{X})] \quad \geq \quad 1 - \alpha.$$

Estos ejemplos ponen de manifiesto la necesidad que existe de construir regiones donde se tenga la "confianza" de encontrar el parámetro. Nuestro estudio se centra en el caso en que el parámetro sea unidimensional y las regiones sean intervalos, por ello, de ahora en adelante, se hablará de *intervalos de confianza*.

Dada una m.a.s. $\underline{X}$ procedente de una variable aleatoria, X, cuya distribución depende de un parámetro desconocido θ y dadas las variables aleatorias $\underline{\theta}(\underline{X})$ y $\overline{\theta}(\underline{X})$. Se define *intervalo de confianza* de nivel $1 - \alpha$ a un intervalo $[\underline{\theta}(\underline{X}), \overline{\theta}(\underline{X})]$, tal que

$$P[\underline{\theta}(\underline{X}) \leq \theta \leq \overline{\theta}(\underline{X})] \geq 1 - \alpha. \tag{3.1}$$

Nótese, que en la definición anterior se habla de nivel de confianza $1 - \alpha$, sin embargo, la probabilidad en dicha expresión es mayor o igual que $1 - \alpha$, esto se debe a que existen situaciones, como en poblaciones discretas, donde no es posible que dicha probabilidad sea exactamente $1 - \alpha$. Como se puede apreciar, los extremos del intervalo son variables aleatorias que dependen de la muestra y que toman para una realización muestral determinada, $\underline{x}$, dos valores puntuales. Así pues, el objetivo de este tema va a ser encontrar $\underline{\theta}(\underline{X})$ y $\overline{\theta}(\underline{X})$, extremos del intervalo de confianza, cumpliendo determinados criterios relativos a la calidad de dicho intervalo. Como ilustración del significado de intervalo aleatorio se tiene el siguiente ejemplo:

Ejemplo 3.4 *Sea X una variable aleatoria que sigue una distribución $U(0,\theta)$. El nivel de confianza, basado en una muestra de tamaño uno, del intervalo aleatorio $[X, 2X]$ es:*

$$P[X \leq \theta \leq 2X] = P\left[\frac{1}{2}\theta \leq X \leq \theta\right] = \tfrac{1}{2}.$$

Tras esta apreciación, hay que tener cuidado a la hora de interpretar el significado de la expresión (3.1), ya que θ es un valor desconocido pero constante, por ello, su interpretación correcta es que la probabilidad de que el intervalo aleatorio $[\underline{\theta}(\underline{X}), \overline{\theta}(\underline{X})]$ contenga el valor del parámetro, θ, es, al menos, $1-\alpha$. Por otra parte, una vez tomada una muestra, se obtiene un intervalo fijo, con lo cual no tiene sentido hablar de probabilidad, ya que el valor del parámetro pertenecerá o no a ese intervalo fijo, es decir, lo hará con probabilidad 1 ó 0. La explicación anterior justifica que tenga que hablarse en términos de "confianza" cuando se considera una muestra concreta. De esta forma, si el intervalo obtenido para una muestra concreta se ha construido con un nivel de confianza de $0'95$, se prevé que dicho intervalo contiene al valor del parámetro, ya que de cada 100 realizaciones muestrales, aproximadamente el intervalo concreto para 95 de ellas contiene dicho parámetro.

En la definición (3.1) se ha hablado de intervalo de confianza en el caso acotado, pero de igual forma, como se aprecia en los ejemplos, puede hablarse de intervalo de confianza acotado inferiormente para θ a un nivel de confianza $1-\alpha$, como el intervalo $[\underline{\theta}(\underline{X}), +\infty)$ donde $\underline{\theta}(\underline{X})$ verifica que

$$P[\underline{\theta}(\underline{X}) \leq \theta] \geq 1-\alpha,$$

y de intervalo de confianza acotado superiormente para θ a un nivel de confianza $1-\alpha$ como el intervalo $(-\infty, \overline{\theta}(\underline{X})]$, donde $\overline{\theta}(\underline{X})$ verifica que

$$P[\overline{\theta}(\underline{X}) \geq \theta] \geq 1-\alpha.$$

2. Intervalos de confianza de longitud mínima

Se puede observar que al ampliar la longitud de un intervalo de confianza aumenta también su nivel de confianza, de hecho, si se considera el intervalo $(-\infty, +\infty)$ se obtiene un intervalo a un nivel de confianza 1. Por otro lado, a un nivel de confianza prefijado, se puede comprobar que no existe un único intervalo. Por ello, se plantea el problema de elegir de entre todos los intervalos a un nivel prefijado, alguno con unas determinadas características. Desde un punto de vista práctico, el intervalo de longitud mínima es una elección interesante, ya que al conservar el nivel de confianza, éste nos da una estimación del parámetro más ajustada que el resto de intervalos del mismo nivel de confianza. Sin embargo, dicha elección presenta el problema de que este intervalo no siempre se puede calcular; en dicho caso, se recurre a una solución alternativa como puede ser la búsqueda del intervalo que tenga longitud mínima esperada, es decir, aquel que minimice la expresión

$$\mathrm{E}[\overline{\theta}(\underline{X}) - \underline{\theta}(\underline{X})].$$

Finalmente, cuando tampoco pueda resolverse este problema, el criterio más empleado consiste en el reparto equitativo del complementario del nivel de confianza entre las dos colas, es decir,

$$P[\underline{\theta}(\underline{X}) \geq \theta] = \frac{\alpha}{2}$$

$$P[\overline{\theta}(\underline{X}) \leq \theta] = \frac{\alpha}{2}\,.$$

Este criterio presenta la ventaja de que conduce a un intervalo único y que en el caso de distribución simétrica con respecto al parámetro es de longitud mínima.

A continuación se dan algunos procedimientos para obtener intervalos de confianza en las situaciones que usualmente se presentan.

3. Método del pivote

Se considera una m.a.s. $\underline{X}$ procedente de una población definida por una variable aleatoria X, cuya distribución dependa de un parámetro desconocido θ. El objetivo de esta sección va a ser desarrollar un método para calcular intervalos de confianza a partir de una función de la muestra que contenga al parámetro y cuya distribución no dependa de él. A continuación, se muestra un ejemplo que ilustra este procedimiento.

Ejemplo 3.5 *Sea $\underline{X}$ una m.a.s. extraída de una $N(\mu, 2)$, se busca un intervalo de confianza para μ a un nivel de confianza $1-\alpha$. Para ello, se sabe que $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$, por tanto*

$$\frac{\overline{X}-\mu}{2}\sqrt{n} \quad \sim \quad N(0,1),$$

con lo cual se pueden tomar las constantes $k_1(\alpha)$ y $k_2(\alpha)$ verificando

$$P\left[k_1(\alpha) \leq \frac{\overline{X}-\mu}{2}\sqrt{n} \leq k_2(\alpha)\right] \quad = \quad 1-\alpha,$$

de donde se obtiene que

$$P\left[\overline{X}-k_2(\alpha)\frac{2}{\sqrt{n}} \leq \mu \leq \overline{X}-k_1(\alpha)\frac{2}{\sqrt{n}}\right] \quad = \quad 1-\alpha$$

y por tanto un intervalo de confianza a nivel $1-\alpha$ para μ es

$$\left[\overline{X} - k_2(\alpha)\frac{2}{\sqrt{n}}\ ,\ \overline{X} - k_1(\alpha)\frac{2}{\sqrt{n}}\right].$$

Obsérvese que $k_1(\alpha)$ y $k_2(\alpha)$ son constantes cuyo valor depende del valor escogido α.

Se dice que $T(\underline{x};\theta)$ es un *pivote* o *cantidad pivotal* si $T(\underline{x};\theta)$ es una función monótona en θ para todo valor muestral $\underline{x}$, la ecuación $\lambda = T(\underline{x};\theta)$ tiene solución para todo λ y la distribución de $T(\underline{X};\theta)$ es independiente de θ.

Si existe $T(\underline{x};\theta)$ pivote se puede construir un intervalo de confianza para θ a cualquier nivel.

Ejemplo 3.6 *Sea $\underline{X}$ una m.a.s. extraída de una población con distribución $U(0,\theta)$. Se quiere encontrar un intervalo de confianza para θ a un nivel de significación $1-\alpha$. Para ello se considera como estimador de θ*

a $\hat{\theta}(\underline{X}) = \max\{X_1, \ldots, X_n\}$ *que se sabe tiene una función de distribución*

$$F_{\hat{\theta}(\underline{X})}(t) \quad = \quad \left(\frac{t}{\theta}\right)^n$$

que al ser función de distribución de una variable aleatoria continua verifica que

$$F_{\hat{\theta}(\underline{X})}(\hat{\theta}(\underline{X})) \quad \sim \quad U(0,1),$$

con lo cual se pueden encontrar $k_1(\alpha)$ *y* $k_2(\alpha)$ *tales que*

$$P[k_1(\alpha) \leq F_{\hat{\theta}}(\hat{\theta}(\underline{X})) \leq k_2(\alpha)] \quad = \quad 1-\alpha.$$

Por simplicidad se toma $k_1(\alpha) = \frac{\alpha}{2}$ *y* $k_2(\alpha) = 1 - \frac{\alpha}{2}$. *Resolviendo las siguientes ecuaciones*

$$\begin{aligned} \left(\frac{\hat{\theta}(\underline{X})}{\theta}\right)^n &= \frac{\alpha}{2} \\ \left(\frac{\hat{\theta}(\underline{X})}{\theta}\right)^n &= 1 - \frac{\alpha}{2}, \end{aligned}$$

se obtiene que;

$$\begin{aligned} \theta &= \frac{\hat{\theta}(\underline{X})}{(\frac{\alpha}{2})^{\frac{1}{n}}} \\ \theta &= \frac{\hat{\theta}(\underline{X})}{(1-\frac{\alpha}{2})^{\frac{1}{n}}}, \end{aligned}$$

con lo cual un intervalo de confianza a un nivel $1-\alpha$ *para* θ *es,*

$$I_{1-\alpha}(\theta) \quad = \quad \left[\frac{\hat{\theta}(\underline{X})}{(1-\frac{\alpha}{2})^{\frac{1}{n}}}, \frac{\hat{\theta}(\underline{X})}{(\frac{\alpha}{2})^{\frac{1}{n}}}\right].$$

Hay que hacer notar que este procedimiento no conduce a un único intervalo de confianza, ya que $k_1(\alpha)$ y $k_2(\alpha)$ se pueden escoger de formas diferentes para que cumplan

$$P[k_1(\alpha) \leq T(\underline{X};\theta) \leq k_2(\alpha)] = 1-\alpha$$

de lo cual puede deducirse que existen diferentes $\underline{\theta}(\underline{X})$ y $\overline{\theta}(\underline{X})$ tal que

$$P[\underline{\theta}(\underline{X}) \leq \theta \leq \overline{\theta}(\underline{X})] = 1 - \alpha.$$

Como ya se comentó en la sección anterior, $k_1(\alpha)$ y $k_2(\alpha)$ se eligen de manera que $\underline{\theta}(\underline{X}) - \overline{\theta}(\underline{X})$ sea mínima, con lo cual, se habrá obtenido un intervalo de confianza a un nivel $1-\alpha$ de longitud mínima construido a partir de $T(\underline{X};\theta)$. Sin embargo, no podrá decirse que es un intervalo de longitud mínima de entre todos los intervalos de confianza a nivel $1-\alpha$, ya que podría existir otro pivote T^* del cual se obtuviera un intervalo más pequeño.

4. Intervalos de confianza en poblaciones Normales

Debido a la importancia que tienen las poblaciones Normales, se ha dedicado este apartado al estudio de los intervalos de confianza para sus parámetros. Por otro lado, la facilidad del cálculo de cantidades pivotales que presentan estas poblaciones hacen recomendable la obtención de estos intervalos de confianza a través del método pivotal.

En esta sección se tratan tanto los intervalos de confianza en una población como en dos poblaciones Normales. En ambos casos, dependiendo del parámetro para el cual se busca un intervalo de confianza y del conocimiento o no de los otros parámetros, se presentan diferentes situaciones que a continuación van a ser estudiadas. En primer lugar se analizan las distintas situaciones para el caso de una población X que sigue una Normal de media μ y varianza σ^2 y de la cual se extrae una m.a.s., $\underline{X}$, de tamaño n. Posteriormente se estudia el caso de dos poblaciones Normales de medias μ_1 y μ_2, varianzas σ_1^2 y σ_2^2 y de las cuales se extraen dos m.a.s., $\underline{X}$ e $\underline{Y}$, de tamaños n_1 y n_2, respectivamente.

Las tablas 3.1 y 3.2 resumen los resultados que se van a obtener en lo que sigue.

4.1. Intervalo de confianza para la media, conocida la varianza

Debido a que se quiere encontrar un intervalo de confianza para la media y se sabe que $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$, puede elegirse como pivote la tipificación de dicha variable aleatoria, es decir,

$$T(\underline{X};\theta) = \frac{\overline{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1),$$

con lo cual, dado un nivel de confianza $1-\alpha$, para una variable aleatoria $Z \sim N(0,1)$ se pretende encontrar $k_1(\alpha)$ y $k_2(\alpha)$ (que para un mejor entendimiento serán denotados por $k_1 = k_1(\alpha)$ y $k_2 = k_2(\alpha)$), tales que

$$P[k_1 \leq Z \leq k_2] = 1-\alpha.$$

Dados α_1, $\alpha_2 \geq 0$ tales que $\alpha_1+\alpha_2 = \alpha$ (α_1 y α_2 representan el reparto de la probabilidad α entre las dos colas), k_1 y k_2 se obtendrán a partir de las igualdades

$$\begin{aligned} P[Z \leq k_1] &= \alpha_1 \\ P[Z \geq k_2] &= \alpha_2. \end{aligned}$$

Una vez calculados k_1 y k_2 se obtiene que

$$\begin{aligned} 1-\alpha &= P\left[k_1 \leq \frac{\overline{X}-\mu}{\sigma}\sqrt{n} \leq k_2\right] \\ &= P\left[k_1\frac{\sigma}{\sqrt{n}} \leq \overline{X}-\mu \leq k_2\frac{\sigma}{\sqrt{n}}\right] \\ &= P\left[\overline{X}-k_2\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}-k_1\frac{\sigma}{\sqrt{n}}\right]. \end{aligned}$$

Se observa que para cada elección de α_1 y α_2 tales que $\alpha_1+\alpha_2 = \alpha$, se obtiene un intervalo diferente; con lo cual, como se decía anteriormente, se escogerá de entre todos ellos el de longitud mínima para este pivote, siempre que ello sea posible. Es decir,

$$\begin{array}{ll} \text{Min} & \overline{X}-k_1\frac{\sigma}{\sqrt{n}} - (\overline{X}-k_2\frac{\sigma}{\sqrt{n}}) = \frac{\sigma}{\sqrt{n}}(k_2-k_1) \\ \text{Sujeto a} & F_Z(k_2)-F_Z(k_1) = 1-\alpha \end{array}$$

donde F_Z es la función de distribución de una $N(0,1)$.

Para encontrar dicho intervalo se recurre al método de los multiplicadores de Lagrange. A partir de la función

$$\psi(k_1,k_2,\lambda)=\frac{\sigma}{\sqrt{n}}(k_2-k_1)+\lambda\left(F_Z(k_2)-F_Z(k_1)-(1-\alpha)\right),$$

se obtiene el siguiente sistema de ecuaciones:

$$\begin{aligned}
\frac{\partial\psi(k_1,k_2,\lambda)}{\partial k_2} &= \frac{\sigma}{\sqrt{n}}+\lambda f_Z(k_2)=0\\
\frac{\partial\psi(k_1,k_2,\lambda)}{\partial k_1} &= -\frac{\sigma}{\sqrt{n}}-\lambda f_Z(k_1)=0\\
\frac{\partial\psi(k_1,k_2,\lambda)}{\partial \lambda} &= F_Z(k_2)-F_z(k_1)-(1-\alpha)=0,
\end{aligned}$$

siendo f_Z la función de densidad de una $N(0,1)$. Sumando las dos primeras ecuaciones y operando se obtiene que

$$e^{-\frac{1}{2}k_2^2}=e^{-\frac{1}{2}k_1^2},$$

de donde se deduce que

$$k_1^2=k_2^2.$$

Por tanto, las soluciones son:

1. $k_1=k_2$, esta solución no es válida pues se tendría un intervalo de longitud nula.
2. $k_1=-k_2$, con lo cual $k_2=Z_{1-\frac{\alpha}{2}}$, donde $Z_{1-\frac{\alpha}{2}}$, verifica que $F_Z(Z_{1-\frac{\alpha}{2}})=1-\frac{\alpha}{2}$.

Es decir, el intervalo de confianza más pequeño coincide con el obtenido por el reparto equitativo de α entre ambas colas; lo cual era esperable ya que la distribución Normal es simétrica respecto a su media. El intervalo en forma explícita viene dado por la expresión

$$I_{1-\alpha}(\mu)=\left[\overline{X}-Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\ ,\ \overline{X}+Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right].$$

Ejemplo 3.7 *Con el fin de estudiar el número medio de flexiones continuadas que pueden realizar sus alumnos, un profesor de educación física somete a 80 de ellos, elegidos aleatoriamente, a una prueba. Los resultados fueron los siguientes:*

Flexiones	35	41	46	48	50	52	53	54	56	60
Alumnos	5	6	2	10	15	6	11	10	5	5

Se sabe que el número de flexiones se distribuye según una Normal de varianza poblacional $7'5$.

Para construir un intervalo de confianza al 95 % para la media del número de flexiones, se tiene que la media muestral es $\overline{x} = 49'78$ *y que* $Z_{1-\frac{\alpha}{2}} = 1'96$. *Por tanto, el intervalo obtenido para esta muestra concreta, viene dado por*

$$\begin{aligned} I_{0'95}(\mu) &= \left[49'78 \pm 1'96\sqrt{\tfrac{7'5}{80}}\right] \\ &= [49'18, 50'38]. \end{aligned}$$

4.2. Intervalo de confianza para la media, desconocida la varianza

Denotando por S_c^2 a la cuasivarianza muestral y usando que $\overline{X}$ y S_c^2 son independientes, se tiene que

$$\left.\begin{array}{l} \dfrac{\overline{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1) \\ (n-1)\dfrac{S_c^2}{\sigma^2} \sim \chi^2_{n-1} \end{array}\right\} \Longrightarrow \frac{\overline{X}-\mu}{S_c}\sqrt{n} \sim t_{n-1},$$

donde t_{n-1} representa la distribución t-student con $n-1$ grados de libertad.

Se puede observar que $T(\underline{X}, \mu) = \frac{\overline{X}-\mu}{S_c}\sqrt{n}$ es un pivote, con lo cual, puede usarse para obtener un intervalo de confianza para la media de una población Normal cuando la varianza es desconocida. Operando igual que en el caso anterior se tendría

$$I_{1-\alpha}(\mu) = \left[\overline{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}} \; , \; \overline{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}}\right],$$

que expresado en términos de la varianza, S^2,

$$I_{1-\alpha}(\mu) = \left[\overline{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n-1}} \; , \; \overline{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n-1}}\right].$$

Ejemplo 3.8 *A partir de una muestra de 20 linternas cuyos periodos de duración (en horas) han sido*

503	480	345	427	386	432	429	378	440	434
429	436	451	466	394	422	412	507	433	480

se quiere obtener un intervalo de confianza al 95 % para la vida media de una población de linternas que se distribuye normalmente.

Teniendo en cuenta que $\overline{x} = 434'2$, $S_c = 40'63$ *y que para* $\alpha = 0'05$ *y* $n = 20$ *es* $t_{n-1,1-\frac{\alpha}{2}} = 2'093$, *se tiene que un intervalo de confianza al 95 % para la vida media de las linternas es*

$$\begin{aligned} I_{0'95}(\mu) &= \left[434'2 \pm 2'093 \tfrac{40'63}{\sqrt{20}}\right] \\ &= [415'18, 453'21]. \end{aligned}$$

4.3. Intervalo de confianza para la varianza, conocida la media

En este caso, puesto que $\frac{X_i-\mu}{\sigma} \sim N(0,1)$ y $\frac{X_i-\mu}{\sigma}$, para $i = 1, \ldots, n$, son independientes dos a dos, se tiene que

$$T(\underline{X};\theta) = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n,$$

donde χ^2_n representa la distribución Chi–cuadrado con n grados de libertad. Utilizando $T(\underline{X};\theta)$ como pivote y definiendo $S^2_\mu = \dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$, el intervalo de confianza a un nivel $1-\alpha$ viene dado por

$$I_{1-\alpha}(\sigma^2) = \left[\frac{nS_\mu^2}{\chi^2_{n,1-\frac{\alpha}{2}}} \; , \; \frac{nS_\mu^2}{\chi^2_{n,\frac{\alpha}{2}}} \right].$$

4.4. Intervalo de confianza para la varianza, desconocida la media

Por el Teorema de Fisher se tiene que

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Razonando de igual forma que en el apartado anterior se obtiene que el intervalo de confianza a un nivel $1-\alpha$ para σ^2 es

$$I_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)S_c^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \; , \; \frac{(n-1)S_c^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \right],$$

que expresado en términos de la varianza, S^2, queda

$$I_{1-\alpha}(\sigma^2) = \left[\frac{nS^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \; , \; \frac{nS^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \right].$$

Ejemplo 3.9 *Se sabe que el peso por comprimido de un cierto preparado farmacéutico se distribuye según una Normal. Con el objeto de estudiar la varianza de la distribución, se extrae una m.a.s. de 6 artículos. Sabiendo que la varianza muestral es igual a 40, se pretende estimar la varianza poblacional mediante un intervalo de confianza al 90 %.*

Puesto que μ es desconocida, un intervalo de confianza para σ^2 viene dado por

$$I_{1-\alpha}(\sigma^2) \quad = \quad \left[\frac{nS^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \; , \; \frac{nS^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \right],$$

donde $\alpha = 0'1$, $n = 6$, *y* $S^2 = 40$. *Así,*

$$\chi^2_{5,0'95} = 11'07 \quad y \quad \chi^2_{5,0'05} = 1'145;$$

con lo cual,

$$\begin{aligned} I_{0'90}(\sigma^2) &= \left[\frac{6 \cdot 40}{11'07}, \frac{6 \cdot 40}{1'145}\right] \\ &= [21'68, 209'61]. \end{aligned}$$

4.5. Intervalo de confianza para la diferencia de medias de muestras apareadas

Sean $\underline{X}$ e $\underline{Y}$ dos m.a.s. de tamaño n y apareadas, de tal forma que la primera procede de una población $N(\mu_1, \sigma_1)$ y la segunda de una población $N(\mu_2, \sigma_2)$.

Antes de proporcionar el intervalo para la diferencia de medias de estas dos poblaciones, se hace necesario indicar qué se entiende por muestras apareadas.

Se dice que dos muestras $\underline{X}$ e $\underline{Y}$ están *apareadas* cuando los datos de las muestras vienen por parejas, uno de cada una de ellas, de manera que cada individuo proporciona dos observaciones.

Ejemplo 3.10 *Para estudiar los efectos de un determinado fármaco para adelgazar, se selecciona aleatoriamente 6 personas y se toma nota de sus pesos antes y después de administrarles el medicamento.*

Antes	72'0	73'5	70'0	71'5	76'0	80'5
Después	73'0	74'5	74'0	74'5	75'0	82'0

Como puede observarse, los datos vienen por parejas: peso antes y después, dos datos por individuo. Parece lógico que los datos se encuentren relacionados entre sí.

En los casos de muestras apareadas, el modo de proceder para obtener un intervalo de confianza para la diferencia de medias es considerar una única muestra formada por la diferencia de los pares de valores, $\underline{D} = \underline{X} - \underline{Y}$, reduciendo así el problema a encontrar un intervalo de confianza para la media de una población.

Ejemplo 3.11 *Si se quisiera construir un intervalo de confianza para la diferencia de medias de los datos del ejemplo anterior, suponiendo que ambas son m.a.s. procedentes de poblaciones Normales, bastaría considerar una nueva muestra:*

$$\underline{D} = \underline{X} - \underline{Y},$$

siendo $\underline{X}$ los pesos antes del tratamiento y $\underline{Y}$ los pesos después del mismo. Así, los valores de la nueva muestra D de tamaño $n = 6$ son

$$-1 \quad -1 \quad -4 \quad -3 \quad 1 \quad -1'5,$$

cuya media muestral es $\overline{x} = -1'58$ y su cuasivarianza $S_c^2 = 3'04$. El intervalo de confianza para la diferencia de medias viene dado por

$$I_{1-\alpha}(\mu_D) = \left[\overline{D} - t_{n-1,1-\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}},\, \overline{D} + t_{n-1,1-\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}}\right].$$

Para $\alpha = 0'05$, se tiene $t_{n-1,1-\frac{\alpha}{2}} = t_{5,0'975} = 2'57$ y el intervalo queda

$$I_{0'95}(\mu_D) = [-3'41,\, 0'25].$$

4.6. Intervalo de confianza para la diferencia de medias de muestras independientes

Sean ahora dos m.a.s. $\underline{X}$ e $\underline{Y}$ de tamaños n_1 y n_2, respectivamente, independientes entre sí, de tal forma que la primera procede de una población $N(\mu_1, \sigma_1)$ y la segunda de una población $N(\mu_2, \sigma_2)$. Usando que $\overline{X}$ y $\overline{Y}$ son independientes, se sabe que

$$\left.\begin{array}{l} \overline{X} \sim N\left(\mu_1, \dfrac{\sigma_1}{\sqrt{n_1}}\right) \\ \overline{Y} \sim N\left(\mu_2, \dfrac{\sigma_2}{\sqrt{n_2}}\right) \end{array}\right\} \Longrightarrow \overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right),$$

y usando que $S^2_{c_1}$ y $S^2_{c_2}$ son independientes, se tiene que

$$\left.\begin{array}{l}(n_1-1)\dfrac{S^2_{c_1}}{\sigma_1^2}\sim\chi^2_{n_1-1}\\(n_2-1)\dfrac{S^2_{c_2}}{\sigma_2^2}\sim\chi^2_{n_2-1}\end{array}\right\}\Longrightarrow(n_1-1)\frac{S^2_{c_1}}{\sigma_1^2}+(n_2-1)\frac{S^2_{c_2}}{\sigma_2^2}\sim\chi^2_{n_1+n_2-2},$$

donde $S^2_{c_1}$ y $S^2_{c_2}$ son las cuasivarianzas muestrales para las muestras $\underline{X}$ e $\underline{Y}$ respectivamente.

El pivote que permite construir el intervalo de confianza para la diferencia de medias de ambas poblaciones, se construye basándose en los resultados anteriores, y depende en gran medida del conocimiento o no de las varianzas poblacionales.

4.6.1. Intervalo de confianza cuando las varianzas son conocidas

Se sabe que

$$\overline{X}-\overline{Y}\sim N\left(\mu_1-\mu_2,\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right),$$

con lo cual se puede tomar como pivote

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1),$$

de donde siguiendo la metodología anterior se obtiene que

$$P\left[\overline{X}-\overline{Y}-Z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\sigma_1^2}{n_1}+\tfrac{\sigma_2^2}{n_2}}\leq\mu_1-\mu_2\leq\overline{X}-\overline{Y}+Z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\sigma_1^2}{n_1}+\tfrac{\sigma_2^2}{n_2}}\right]=$$
$$=1-\alpha$$

por tanto el intervalo buscado es

$$I_{1-\alpha}(\mu_1-\mu_2)=$$
$$=\left[\overline{X}-\overline{Y}-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ \overline{X}-\overline{Y}+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right].$$

Ejemplo 3.12 *Se quiere estudiar la diferencia de las vidas medias de dos tipos de lámparas. Para ello, se toma una muestra de 150 lámparas de tipo H y otra, independiente de la anterior, de 200 lámparas de tipo N, obteniéndose que las de tipo H tienen una vida media de 1400 horas y una desviación típica de 120, y que las de tipo N tienen una vida media de 1200 horas y desviación típica 80.*

Para estimar la diferencia de medias se construye un intervalo de confianza al 95 %, que viene dado por

$$\left[\overline{X}_H - \overline{Y}_N \pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_H^2}{n_H} + \frac{\sigma_N^2}{n_N}}\right].$$

Sustituyendo en la expresión anterior, se obtiene

$$\left[1400 - 1200 \pm 1'96\sqrt{\frac{120^2}{150} + \frac{80^2}{200}}\right],$$

y por tanto,

$$I_{0'95}(\mu_H - \mu_N) \quad = \quad [177'8, 222'2].$$

4.6.2. Intervalo de confianza cuando las varianzas son desconocidas e iguales

Como $S_{c_1}^2$ y $S_{c_2}^2$ son independientes, se sabe que

$$\left.\begin{array}{l} (n_1-1)\dfrac{S_{c_1}^2}{\sigma^2} \sim \chi_{n_1-1}^2 \\ (n_2-1)\dfrac{S_{c_2}^2}{\sigma^2} \sim \chi_{n_2-1}^2 \end{array}\right\} \Longrightarrow \frac{(n_1-1)S_{c_1}^2 + (n_2-1)S_{c_2}^2}{\sigma^2} \sim \chi_{n_1+n_2-2}^2$$

y puesto que $\overline{X}$ y $S_{c_1}^2$, así como, $\overline{Y}$ y $S_{c_2}^2$ son independientes, se tiene el pivote

$$T(\underline{X}, \underline{Y}, \mu_1 - \mu_2) =$$

$$= \frac{\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sigma}\left(\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)^{-1}}{\sqrt{\frac{(n_1-1)S_{c_1}^2+(n_2-1)S_{c_2}^2}{\sigma^2(n_1+n_2-2)}}}$$

$$= \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_{c_1}^2 + (n_2-1)S_{c_2}^2}{n_1+n_2-2}}} \left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)^{-1} \sim t_{n_1+n_2-2},$$

obteniéndose como intervalo de confianza

$$I_{1-\alpha}(\mu_1 - \mu_2) =$$
$$= \left[\overline{X} - \overline{Y} \pm t_{n_1+n_2-2,1-\frac{\alpha}{2}} \sqrt{\frac{(n_1-1)S_{c_1}^2 + (n_2-1)S_{c_2}^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right]$$

que expresado en función de la varianza muestral queda

$$I_{1-\alpha}(\mu_1 - \mu_2) =$$
$$= \left[\overline{X} - \overline{Y} \pm t_{n_1+n_2-2,1-\frac{\alpha}{2}} \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right].$$

Ejemplo 3.13 *De una población $N(\mu_1, \sigma^2)$, se extrae una m.a.s. de tamaño 10, tal que la media muestral es 4'1 y la varianza muestral es 6'09. De otra población $N(\mu_2, \sigma^2)$ se toma otra m.a.s. de tamaño 16 e independiente de la anterior, cuya media y varianza muestrales son 3'875 y 3'609, respectivamente. Se quiere obtener un intervalo de confianza del 95 % para la diferencia de medias poblacionales.*

Puesto que la varianzas poblacionales son desconocidas pero iguales, el intervalo de confianza para la diferencia de medias viene dado por

$$\left[\overline{X} - \overline{Y} \pm t_{n_1+n_2-2,1-\frac{\alpha}{2}} \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right].$$

Para $\alpha = 0'05$ se tiene que $t_{24,0'975} = 2'0639$. Así,

$$\left[4'1 - 3'875 \pm 2'0639 \sqrt{\frac{10 \cdot 6'09 + 16 \cdot 3'609}{24} \left(\frac{1}{10} + \frac{1}{16} \right)} \right].$$

Por tanto,

$$I_{0'95}(\mu_1 - \mu_2) \quad = \quad [-1'6248, 2'0748].$$

4.6.3. Intervalo de confianza cuando las varianzas son desconocidas y distintas

En este caso, puesto que $\overline{X}$ y $S_{c_1}^2$, así como, $\overline{Y}$ y $S_{c_2}^2$ son independientes, se tiene que

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1),$$

así como

$$(n_1-1)\frac{S_{c_1}^2}{\sigma_1^2}+(n_2-1)\frac{S_{c_2}^2}{\sigma_2} \sim \chi^2_{n_1+n_2-2}.$$

Por tanto

$$\frac{\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}}{\sqrt{\frac{(n_1-1)\frac{S_{c_1}^2}{\sigma_1^2}+(n_2-1)\frac{S_{c_2}^2}{\sigma_2^2}}{n_1+n_2-2}}} \sim t_{n_1+n_2-2},$$

pero como se ve el estadístico depende de σ_1 y σ_2 por lo que se recurre a la aproximación de Welch, en función de la cual el estadístico

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_{c_1}^2}{n_1}+\frac{S_{c_2}^2}{n_2}}},$$

tiene una distribución aproximada $t_{a,1-\frac{\alpha}{2}}$ siendo a un factor corrector que se calcula tomando el entero más próximo a

$$a=\frac{\left(\frac{S_{c_1}^2}{n_1}+\frac{S_{c_2}^2}{n_2}\right)^2}{\frac{1}{n_1+1}\left(\frac{S_{c_1}^2}{n_1}\right)^2+\frac{1}{n_2+1}\left(\frac{S_{c_2}^2}{n_2}\right)^2}-2$$

y donde el intervalo de confianza para $\mu_1-\mu_2$ viene dado por

$$I_{1-\alpha}(\mu_1-\mu_2)=\left[\overline{X}-\overline{Y}\pm t_{a,1-\frac{\alpha}{2}}\sqrt{\frac{S_{c_1}^2}{n_1}+\frac{S_{c_2}^2}{n_2}}\right].$$

Ejemplo 3.14 *Para realizar un estudio sobre la hipertensión y sus consecuencias, se toman dos muestras de 13 y 16 pacientes de ciudades distintas. Los datos muestrales obtenidos fueron los siguientes:*

$$\begin{aligned} \overline{x_1} &= 166\,mm. \quad S_{c_1} = 28\,mm. \\ \overline{x_2} &= 164'7\,mm. \quad S_{c_2} = 7\,mm. \end{aligned}$$

Supuesto que ambas poblaciones son Normales y que sus varianzas son desconocidas y distintas, se quiere determinar un intervalo de confianza al 95 % *para la diferencia de medias.*

Lo primero es calcular el valor de a.

$$\begin{aligned} a &= \frac{\left(\frac{S_{c_1}^2}{n_1} + \frac{S_{c_2}^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{S_{c_1}^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{S_{c_2}^2}{n_2}\right)^2} - 2 \\ &= 13'4. \end{aligned}$$

Luego se toma $a = 13$. Por otra parte, puesto que $t_{13,0'975} = 2'16$, el intervalo buscado es

$$\begin{aligned} I_{0'95}(\mu_1 - \mu_2) &= \left[166 - 164'7 \pm 2'16\sqrt{\tfrac{784}{13} + \tfrac{49}{16}}\right] \\ &= [-15'89, 18'49]. \end{aligned}$$

4.7. Intervalo de confianza para el cociente de varianzas

Al ser S_{c_1} y S_{c_2} independientes, se tiene que

$$\left.\begin{array}{l} (n_1-1)\dfrac{S_{c_1}^2}{\sigma_1^2} \sim \chi^2_{n_1-1} \\ (n_2-1)\dfrac{S_{c_2}^2}{\sigma_2^2} \sim \chi^2_{n_2-1} \end{array}\right\} \Longrightarrow \frac{\frac{n_1-1}{n_1-1}\frac{S_{c_1}^2}{\sigma_1^2}}{\frac{n_2-1}{n_2-1}\frac{S_{c_2}^2}{\sigma_2^2}} = \frac{\frac{S_{c_1}^2}{\sigma_1^2}}{\frac{S_{c_2}^2}{\sigma_2^2}} \sim F_{n_1-1,n_2-1},$$

con lo cual, hay que determinar k_1 y k_2 que verifiquen la igualdad

$$P[k_1 \leq F_{n_1-1,n_2-1} \leq k_2] = 1 - \alpha.$$

Usando el método del reparto equitativo del nivel de significación, se obtiene que el intervalo de confianza buscado es

$$I_{1-\alpha}\left(\frac{\sigma_1^2}{\sigma_2^2}\right) = \left[\frac{1}{F_{n_1-1,n_2-1,1-\frac{\alpha}{2}}}\frac{S_{c_1}^2}{S_{c_2}^2}, \frac{1}{F_{n_1-1,n_2-1,\frac{\alpha}{2}}}\frac{S_{c_1}^2}{S_{c_2}^2}\right],$$

y puesto que $S_c^2 = \frac{n}{n-1}S^2$, se tiene el intervalo

$$I_{1-\alpha}\left(\frac{\sigma_1^2}{\sigma_2^2}\right) = \left[\frac{1}{F_{n_1-1,n_2-1,1-\frac{\alpha}{2}}}\frac{\frac{n_1}{n_1-1}S_1^2}{\frac{n_2}{n_2-1}S_2^2}, \frac{1}{F_{n_1-1,n_2-1,\frac{\alpha}{2}}}\frac{\frac{n_1}{n_1-1}S_1^2}{\frac{n_2}{n_2-1}S_2^2}\right].$$

Ejemplo 3.15 *Con el fin de estudiar el gasto de combustible de dos motos procedentes de dos compañías diferentes, C1 y C2, se seleccionan al azar 9 motos de la compañía C1 y 12 de la C2. Las de la compañía C1 proporcionan una media de* $18\,km$ *recorridos por cada litro de combustible, con una cuasivarianza de* $1'1\,km^2/l^2$ *y las de la compañía C2, una media de* $15\,km/l$ *y una cuasivarianza de* $2'9\,km^2/l^2$.

Sabiendo que la distancia recorrida por cada litro de combustible se distribuye normalmente en las dos compañías, se pretende obtener un intervalo de confianza al 90 % para el cociente de varianzas. Llamando $S_{c_1}^2$ *y* $S_{c_2}^2$ *a las cuasivarianzas muestrales de las motos de las compañías 1 y 2 y teniendo en cuenta que* $\alpha = 0'1$, *se tiene que para* $n_1 = 9$ *y* $n_2 = 12$ *es*

$$F_{n_1-1,n_2-1,\frac{\alpha}{2}} = 0'34 \quad y \quad F_{n_1-1,n_2-1,1-\frac{\alpha}{2}} = 2'95.$$

Así pues, un intervalo de confianza al 90 % para el cociente de varianzas viene dado por

$$\begin{aligned} I_{0'9}\left(\tfrac{\sigma_1^2}{\sigma_2^2}\right) &= \left[\frac{1}{2'95}\frac{1'1}{2'9}, \frac{1}{0'34}\frac{1'1}{2'9}\right] \\ &= [0'13, 1'12]. \end{aligned}$$

5.*** Método basado en la desigualdad de Tchebychev

En el método del pivote se parte del conocimiento, salvo parámetros, de la distribución de la variable aleatoria que define la población. Sin embargo, en esta sección se aborda una metodología que permite obtener un intervalo de confianza para un parámetro de la población conociendo únicamente la media y varianza del estimador de dicho parámetro. Para ello, se usará la desigualdad de Tchebychev, la cual dice que dada una variable aleatoria, X, tal que $\mathrm{E}[X] = \mu$ y $\mathrm{V}[X] = \sigma^2$, se verifica que

$$P[|X - \mu| \leq k\sigma] > 1 - \frac{1}{k^2}.$$

Sea $\hat{\theta}(\underline{X})$ un estimador del parámetro que se quiere estudiar, usando la desigualdad anterior se puede encontrar un intervalo de confianza con una cota inferior del nivel de confianza prefijado. Así pues, se verifica que

$$P\left[|\hat{\theta} - \mathrm{E}[\hat{\theta}]| \leq k\sqrt{\mathrm{V}[\hat{\theta}]}\,\right] > 1 - \frac{1}{k^2}.$$

En el caso de que el estimador sea insesgado se podrá obtener un intervalo con nivel de confianza de al menos $1 - \alpha$ que viene dado a partir de las expresiones

$$1 - \frac{1}{k^2} = 1 - \alpha \Rightarrow k = \frac{1}{\sqrt{\alpha}}$$

$$P\left[\hat{\theta} - \frac{1}{\sqrt{\alpha}}\sqrt{\mathrm{V}[\hat{\theta}]} \leq \theta \leq \hat{\theta} + \frac{1}{\sqrt{\alpha}}\sqrt{\mathrm{V}[\hat{\theta}]}\,\right] \geq 1 - \alpha.$$

De lo cual se deduce que un intervalo de confianza es

$$I_{1-\alpha}(\theta) = \left[\hat{\theta} - \sqrt{\frac{\mathrm{V}[\hat{\theta}]}{\alpha}}\ ,\ \hat{\theta} + \sqrt{\frac{\mathrm{V}[\hat{\theta}]}{\alpha}}\,\right].$$

En el caso particular de que se quiera encontrar un intervalo de confianza para la media de una población de la cual se conoce la varianza, puede tomarse como estimador $\overline{X}$. De esta forma se obtiene el intervalo

$$I_{1-\alpha}(\mu) = \left[\overline{X} - \frac{\sigma}{\sqrt{n\alpha}}, \overline{X} + \frac{\sigma}{\sqrt{n\alpha}}\right].$$

Ejemplo 3.16 *Se quiere comparar el intervalo obtenido en este método con el obtenido usando el método del pivote bajo la hipótesis de Normalidad. En este caso, el intervalo venía dado por*

$$\left[\overline{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\ ,\ \overline{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right],$$

con lo cual la amplitud de este intervalo es

$$2 \cdot Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\,,$$

mientras que para el método basado en la desigualdad de Tchebychev, la amplitud del intervalo es

$$2 \cdot \frac{1}{\sqrt{\alpha}} \frac{\sigma}{\sqrt{n}}\ .$$

Para el caso de un intervalo de confianza a un nivel de $0'95$ *y puesto que* $Z_{1-\frac{\alpha}{2}} = 1'96$ *y* $\frac{1}{\sqrt{\alpha}} = 4'47$*, se puede deducir que aunque el método aproximado es aplicable en situaciones muy generales, presenta la desventaja de no proporcionar un intervalo con buenas propiedades.*

6.*** Método asintótico basado en el Teorema Central del Límite

Esta sección está dedicada a la búsqueda de intervalos de confianza para la media de una población de la cual se posee una muestra de gran tamaño, es decir, se va a construir un intervalo de confianza asintótico para la media.

Cuando se busca el intervalo de confianza para la media de una población, el estimador natural es la media muestral $\overline{X}$. Sin embargo, puede suceder que se desconozca su distribución y consecuentemente no se pueda calcular dicho intervalo. Para superar esta dificultad, se utiliza el Teorema Central del Límite.

Antes de continuar es necesario introducir el concepto de convergencia en distribución o en ley.

Dada $(X_n)_{n\in\mathcal{N}}$ una sucesión de variables aleatorias con función de distribución F_n. Se dice que X_n *converge en ley* o *en distribución* a una variable aleatoria X con función de distribución F, si $F_n(x) \to F(x)$ en todo punto de continuidad de F, este tipo de convergencia se denota por $X_n \xrightarrow{d} X$ o $X_n \xrightarrow{l} X$.

Así mismo, será de gran utilidad el siguiente teorema conocido como *Teorema de Linderberg–Lévy.*

Dadas $X_1, \ldots, X_n$, variables aleatorias independientes e idénticamente distribuidas, tales que, $\mathrm{E}[X] = \mu$ y $\mathrm{V}[X] = \sigma^2 < \infty$. Entonces:

$$\frac{\overline{X} - \mu}{\sigma}\sqrt{n} \xrightarrow{d} N(0, 1).$$

Por tanto, siempre que el objetivo sea obtener un intervalo de confianza para la media, se puede aplicar este teorema y usar las mismas técnicas empleadas para obtener el intervalo de confianza para la media de una población Normal, con la salvedad de que en este caso el nivel de confianza de este intervalo no será exacto, sino aproximado.

Debido a que en la mayoría de las situaciones reales que se presentan la varianza poblacional es desconocida, en este método asintótico la varianza poblacional, se aproxima por la muestral, obteniéndose que

$$\frac{\overline{X} - \mu}{S}\sqrt{n} \xrightarrow{d} N(0, 1).$$

6.1. Intervalos de confianza para la proporción

Considérese una m.a.s. $\underline{X}$ extraída de una población definida por una variable aleatoria X, distribuida según una Bernouilli de parámetro p. Si la variable aleatoria toma valores 0 y 1, el estimador de máxima verosimilitud del parámetro p es

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}.$$

Puesto que

$$\begin{aligned} \mathrm{E}[\hat{p}] &= p \\ \mathrm{V}[\hat{p}] &= \frac{p(1-p)}{n}, \end{aligned}$$

puede deducirse, aplicando una variante del Teorema de Linderberg–Levy para este tipo de distribuciones, que

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} N(0,1).$$

A partir de este resultado se puede construir un intervalo de confianza, bien directamente, o a través de la doble aproximación, donde $p(1-p)$ es sustituido por su estimador $\hat{p}(1-\hat{p})$.

Véase qué intervalos se obtienen por ambos métodos:

1. Sin utilizar la doble aproximación y haciendo $k = Z_{1-\frac{\alpha}{2}}$.

$$\begin{aligned} 1-\alpha &= P[-k \leq N(0,1) \leq k] \\ &= P\left[-k \leq \frac{\hat{p}-p}{\sqrt{p(1-p)}}\sqrt{n} \leq k\right] \\ &= P\left[(\hat{p}-p)^2 n \leq k^2 p(1-p)\right] \\ &= P\left[(\hat{p}^2+p^2-2p\hat{p})n \leq \frac{k^2 p(1-p)}{n}\right] \\ &= P\left[\left(1+\frac{k^2}{n}\right)p^2 - \left(2\hat{p}+\frac{k^2}{n}\right)p + \hat{p}^2 \leq 0\right] \end{aligned}$$

$$p = \frac{(2\hat{p}+\frac{k^2}{n}) \pm \sqrt{(2\hat{p}+\frac{k^2}{n})^2 - 4(1+\frac{k^2}{n})\hat{p}^2}}{2(1+\frac{k^2}{n})}.$$

Por tanto el intervalo es

$$I_{1-\alpha}(p) = \left[\frac{(2\hat{p}+\frac{k^2}{n}) \pm \sqrt{(2\hat{p}+\frac{k^2}{n})^2 - 4(1+\frac{k^2}{n})\hat{p}^2}}{2(1+\frac{k^2}{n})}\right],$$

o equivalentemente

$$I_{1-\alpha}(p) = \left[\frac{\hat{p}}{1+\frac{k^2}{n}} + \frac{1}{2(1+\frac{n}{k^2})} \pm \frac{\frac{k}{\sqrt{n}}\sqrt{4\hat{p}(1-\hat{p})+\frac{k}{n^2}}}{2(1+\frac{k^2}{n})} \right].$$

2. Utilizando la doble aproximación y el mismo valor para $k = Z_{1-\frac{\alpha}{2}}$.

$$\begin{aligned} 1-\alpha &= P[-k \leq N(0,1) \leq k] = \\ &= P\left[-k \leq \frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})}}\sqrt{n} \leq k\right] = \\ &= P\left[\hat{p} - \frac{k}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})} \leq p \leq \hat{p} + \frac{k}{n}\sqrt{\hat{p}(1-\hat{p})}\right], \end{aligned}$$

de donde se deduce que el intervalo es

$$I_{1-\alpha}(p) = \left[\hat{p} \pm \frac{k}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})}\right].$$

Como puede observarse el primer intervalo obtenido converge al segundo.

Ejemplo 3.17 *En unas elecciones, el candidato A desea estimar, al 95 % de confianza, la proporción de votantes que están a su favor. Con este fin, toma una muestra aleatoria de 100 votantes, observando que el 55 % son partidarios suyos, obteniendo un intervalo de confianza de sus probabilidades de triunfo igual a*

$$\begin{aligned} I_{0'95}(p) &= \left[\hat{p} \pm Z_{0'975}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \\ &= \left[0'55 \pm 1'96\sqrt{\frac{0'55 \cdot 0'45}{100}}\right] \\ &= [0'55 \pm 0'1] \\ &= [0'45, 0'65]. \end{aligned}$$

6.2. Intervalo de confianza para la diferencia de proporciones

Se consideran dos muestras aleatorias simples $\underline{X}$ e $\underline{Y}$, de tamaños n_1 y n_2 e independientes entre sí, extraídas de poblaciones con distribuciones Bernouilli de parámetros p_1 y p_2 respectivamente. El objetivo consiste en encontrar un intervalo de confianza para la diferencia de sus proporciones. Como en el caso anterior, se obtiene que los estimadores máximo verosímiles para ambas muestras son:

$$\hat{p}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i \quad \hat{p}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$$

verificándose que

$$\mathrm{E}[\hat{p}_1] = p_1 \quad \mathrm{V}[\hat{p}_1] = \frac{p_1(1-p_1)}{n_1} \quad \hat{p}_1 \xrightarrow{d} N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}}\right)$$

y

$$\mathrm{E}[\hat{p}_2] = p_2 \quad \mathrm{V}[\hat{p}_2] = \frac{p_2(1-p_2)}{n_2} \quad \hat{p}_2 \xrightarrow{d} N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{n_2}}\right).$$

Puesto que ambas muestras son independientes se tiene que

$$\hat{p}_1 - \hat{p}_2 \xrightarrow{d} N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

de lo cual puede deducirse usando la doble aproximación la expresión

$$P\left[-Z_{1-\frac{\alpha}{2}} \leq \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \leq Z_{1-\frac{\alpha}{2}}\right] = 1 - \alpha,$$

siendo el intervalo de confianza buscado igual a

$$I_{1-\alpha}(p_1 - p_2) = \left[\hat{p}_1 - \hat{p}_2 \pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right].$$

Ejemplo 3.18 *Una determinada empresa quiere saber si su nuevo producto tendrá más aceptación en la población adulta o entre los jóvenes. Para ello, considera una m.a.s. de 400 adultos y 600 jóvenes, observando que sólo a 100 adultos y 300 jóvenes les había gustado su innovador producto. Para comparar las proporciones de adultos y jóvenes a los que les gusta el producto, a un nivel de confianza del 99 %, se considera el intervalo de confianza*

$$\left[\hat{p}_1 - \hat{p}_2 \pm Z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \tfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right].$$

Si se considera

$$\begin{aligned} p_1 &= \text{proporción de jóvenes a los que gusta} \\ p_2 &= \text{proporción de adultos a los que gusta} \end{aligned}$$

entonces

$$\hat{p}_1 = \tfrac{300}{600} = 0'5 \quad y \quad \hat{p}_2 = \tfrac{100}{400} = 0'25,$$

con lo que el intervalo queda

$$\left[0'5 - 0'25 \pm 2'58\sqrt{\tfrac{0'5 \cdot 0'5}{600} + \tfrac{0'25 \cdot 0'75}{400}}\right],$$

es decir,

$$I_{0'99}(p_1 - p_2) = [0'19, 0'31].$$

7.*** Intervalo asintótico para cualquier parámetro

En el apartado anterior, se estudió la construcción de un intervalo de confianza para la media de una población a través de métodos asintóticos. En esta sección, se extiende el método anterior a cualquier parámetro del cual se disponga un estimador máximo verosímil. Para ello, se considera el siguiente resultado.

Teorema 3.1 *Si se verifican las condiciones de Fisher–Wolfowitz el es-*

timador máximo–verosímil de θ, $\hat{\theta}_{MV}$, es asintóticamente Normal:

$$\frac{\hat{\theta}_{MV}(\underline{X}) - \theta}{\sqrt{(I(\theta))^{-\frac{1}{2}}}} \xrightarrow{d} N(0,1),$$

donde $I(\theta) = -\,\mathrm{E}\left[\frac{\partial^2 \log f(x,\theta)}{\partial \theta^2}\right]$.

Así, dada una m.a.s. procedente de una población cuya distribución depende de un parámetro θ desconocido y suponiendo conocido su estimador de máxima verosimilitud, el intervalo de confianza para dicho parámetro viene dado por la expresión

$$\begin{aligned} 1-\alpha &= P\left[-Z_{1-\frac{\alpha}{2}} \leq \frac{\hat{\theta}_{MV}(\underline{X}) - \theta}{\sqrt{I(\theta)}} \leq Z_{1-\frac{\alpha}{2}}\right] \\ &= P\left[\hat{\theta}_{MV}(\underline{X}) - Z_{1-\frac{\alpha}{2}}\sqrt{I(\theta)} \leq \theta \leq \hat{\theta}_{MV}(\underline{X}) + Z_{1-\frac{\alpha}{2}}\sqrt{I(\theta)}\right], \end{aligned}$$

obteniéndose el intervalo

$$I_{1-\alpha}(\theta) = \left[\hat{\theta}_{MV}(\underline{X}) - Z_{1-\frac{\alpha}{2}}\sqrt{I(\theta)}\ ,\ \hat{\theta}_{MV} + Z_{1-\frac{\alpha}{2}}\sqrt{I(\theta)}\right].$$

Ejemplo 3.19 *Sea $\underline{X}$ una m.a.s. extraída de una población de Poisson de parámetro desconocido λ. Se sabe que el estimador de máxima verosimilitud para λ es*

$$\hat{\lambda}_{MV}(\underline{X}) = \overline{X}$$

y la cantidad de información de Fisher para esta distribución es

$$I(\lambda) = \frac{n}{\lambda},$$

de lo cual se deduce que

$$\frac{\overline{X}-\lambda}{\sqrt{\lambda}}\sqrt{n} \xrightarrow{d} N(0,1),$$

y donde, o bien, puede sustituirse λ, como se vio anteriormente, en el denominador por un estimador

consistente como es $\overline{X}$, o resolverse directamente.
Si se sustituye, se obtiene que:

$$\frac{\overline{X}-\lambda}{\sqrt{\overline{X}}}\sqrt{n} \quad \xrightarrow{d} \quad N(0,1)$$

y de aquí, el intervalo de confianza para λ es

$$I_{1-\alpha}(\lambda) \quad = \quad \left[\overline{X} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}}{n}}\ ,\ \overline{X} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}}{n}}\right].$$

8. Determinación del tamaño muestral

Una vez estudiados diferentes métodos de construcción de un intervalo de confianza, se pone de manifiesto la importancia del tamaño muestral en los procesos inferenciales y más concretamente, en la construcción de intervalos de confianza para la media de una población. En esta sección, el objetivo va a consistir en fijar el tamaño muestral para que el error cometido en el proceso de estimación de dichos intervalos sea menor que una cantidad prefijada.

Así, dependiendo del conocimiento o no de la varianza se pueden distinguir los siguientes casos:

8.1. Determinación del tamaño muestral para estimar la media, conocida la varianza

Tanto en el caso que la población sea una población $N(\mu,\sigma)$, como en el caso de que el tamaño muestral sea suficientemente grande se ha visto que

$$P\left[\overline{X} - \frac{Z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + \frac{Z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right] = 1-\alpha,$$

o equivalentemente

$$P\left[|\overline{X}-\mu| \leq \frac{Z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right] = 1-\alpha,$$

con lo cual, una medida del error cometido en la estimación de la media viene dada por $|\overline{X}-\mu|$, siendo $\dfrac{Z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}$ una cota de dicho error.

Por tanto, se puede calcular el tamaño muestral para que el error absoluto cometido en la estimación sea a lo sumo una cantidad prefijada ε, de la forma

$$n = \frac{Z^2_{1-\frac{\alpha}{2}}\sigma^2}{\varepsilon^2}.$$

Ejemplo 3.20 *Se pretende estimar la media μ de una población Normal de varianza $17'64$ y se quiere tener una confianza del 95 % de que el error absoluto de estimación sea menor de $0'05$. Determínese el tamaño de la muestra.*

Para $\alpha = 0'05$ es $Z_{1-\frac{\alpha}{2}} = 1'96$. Por tanto, para $\varepsilon = 0'05$ se tiene que

$$n = \frac{(1'96)^2 \cdot 17'64}{(0'05)^2} = 27106.$$

8.2. Determinación del tamaño muestral para estimar la media, desconocida la varianza

De igual forma que en el caso anterior, se obtiene que un intervalo de confianza a nivel $1-\alpha$ para la media de una población es

$$P\left[\overline{X} - t_{n-1,1-\frac{\alpha}{2}}\sqrt{\frac{S_c^2}{n}} \leq \mu \leq \overline{X} + t_{n-1,1-\frac{\alpha}{2}}\sqrt{\frac{S_c^2}{n}}\right] = 1-\alpha,$$

o equivalentemente

$$P\left[|\overline{X} - \mu| \leq t_{n-1,1-\frac{\alpha}{2}}\sqrt{\frac{S_c^2}{n}}\right] = 1-\alpha.$$

Razonando de igual forma que en el apartado anterior, el tamaño muestral necesario para obtener un error de estimación menor que una cantidad prefijada, ε, debe ser

$$\frac{t^2_{n-1,1-\frac{\alpha}{2}}S_c^2}{\varepsilon^2}.$$

8.3. Determinación del tamaño muestral para la proporción

En las secciones anteriores se estudió que el intervalo de confianza para la proporción de una población viene dada por

$$P\left[\hat{p} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = 1-\alpha,$$

o equivalentemente

$$P\left[|\hat{p} - p| \leq Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = 1-\alpha.$$

Siguiendo el razonamiento de los apartados anteriores se obtiene que el tamaño necesario para cometer un error menor que una cantidad prefijada, ε, debe ser mayor que

$$\frac{Z^2_{1-\frac{\alpha}{2}}\hat{p}(1-\hat{p})}{\varepsilon^2}.$$

Ejemplo 3.21 *Con el fin de organizar su producción, una determinada fábrica emprende una investigación para conocer la proporción de consumidores que adquieren su producto. Se quiere que el error de estimación máximo sea del 3 % con una confianza del 95 %, así que se trata de averiguar cuál debe ser el tamaño de la muestra para que se cumplan estos objetivos.*

Se puede observar que nada se sabe acerca de $\hat{p}$*. En esta ocasión y puesto que* $\max \hat{p}(1-\hat{p}) = \frac{1}{4}$*, se tendrá que*

$$\frac{\frac{1}{4}Z^2_{1-\frac{\alpha}{2}}}{\varepsilon^2} \leq n.$$

Como $\varepsilon = 0'03$ *y* $\alpha = 0'05$*, se tiene que* $Z_{\frac{\alpha}{2}} = 1'96$*. Así,*

$$0'25\frac{(1'96)^2}{(0'03)^2} = 1067'11 \leq n.$$

Por tanto, el tamaño de la muestra ha de ser como mínimo de 1068 consumidores.

9. Tablas de Intervalos de Confianza

Distribución	Parámetro	Casos	Intervalo
Normal $\mathcal{N}(\mu, \sigma^2)$	μ	σ conocida	$\bar{x} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
		σ desconocida	$n < 30$: $\bar{x} \pm t_{n-1,1-\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}}$ $n \geq 30$: $\bar{x} \pm Z_{1-\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}}$
Desconocida $n \geq 30$	μ	σ conocida	$\bar{x} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
		σ desconocida	$\bar{x} \pm Z_{1-\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}}$
Bernoulli $\mathcal{B}(p)$	p	Muestras grandes	$\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$ $\hat{p} = \bar{x}$ $\hat{q} = 1 - \bar{x}$
Normal $\mathcal{N}(\mu, \sigma^2)$	σ^2	μ conocida	$\left(\frac{nS_\mu^2}{\chi^2{}_{n,1-\frac{\alpha}{2}}}, \frac{nS_\mu^2}{\chi^2{}_{n,\frac{\alpha}{2}}} \right)$ $S_\mu^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$
		μ desconocida	$\left(\frac{(n-1)S_c^2}{\chi^2{}_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)S_c^2}{\chi^2{}_{n-1,\frac{\alpha}{2}}} \right)$
Poisson $\mathcal{P}(\lambda)$	λ	Muestras grandes	$\bar{x} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{x}}{n}}$

Tabla 3.1: Intervalos de confianza para una población

Distribución	Parámetro	Casos	Intervalo
Normales Indep. $\mathcal{N}(\mu_1, \sigma_1^2)$ $\mathcal{N}(\mu_2, \sigma_2^2)$	$\mu_1 - \mu_2$	σ_1, σ_2 conocidas	$(\bar{x}_1 - \bar{x}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
		$\sigma_1 = \sigma_2$ desconocidas	$\bar{x}_1 - \bar{x}_2 \pm t_{m,1-\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ con $m = n_1 + n_2 - 2$ $S_p^2 = \frac{(n_1 - 1)S_{c_1}^2 + (n_2 - 1)S_{c_2}^2}{n_1 + n_2 - 2}$
		$\sigma_1 \neq \sigma_2$ desconocidas	$\bar{x}_1 - \bar{x}_2 \pm t_{a,1-\frac{\alpha}{2}} \sqrt{\frac{S_{c_1}^2}{n_1} + \frac{S_{c_2}^2}{n_2}}$ $a = \frac{\left(S_{c_1}^2/n_1 + S_{c_2}^2/n_2\right)^2}{\frac{\left(S_{c_1}^2/n_1\right)^2}{n_1+1} + \frac{\left(S_{c_2}^2/n_2\right)^2}{n_2+1}} - 2$
Normales Depen. $\mathcal{N}(\mu_1, \sigma_1^2)$ $\mathcal{N}(\mu_2, \sigma_2^2)$	$\mu_1 - \mu_2$	σ_d conocida	$\bar{d} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma_d}{\sqrt{n}} \quad \bar{d} = \bar{x}_1 - \bar{x}_2$
		σ_d desconocida	$n < 30$: $\bar{d} \pm t_{n-1,1-\frac{\alpha}{2}} \frac{S_{c_d}}{\sqrt{n}}$ $n \geq 30$: $\bar{d} \pm Z_{1-\frac{\alpha}{2}} \frac{S_{c_d}}{\sqrt{n}}$
Normales $\mathcal{N}(\mu_1, \sigma_1^2)$ $\mathcal{N}(\mu_2, \sigma_2^2)$	$\frac{\sigma_1^2}{\sigma_2^2}$	μ_1, μ_2 desconocidas	$\left(\frac{S_{c_1}^2/S_{c_2}^2}{F_{n_1-1,n_2-1,1-\frac{\alpha}{2}}}, \frac{S_{c_1}^2/S_{c_2}^2}{F_{n_1-1,n_2-1,\frac{\alpha}{2}}} \right)$
Bernoulli $\mathcal{B}(p_1)$ $\mathcal{B}(p_2)$	$p_1 - p_2$	$n_1\bar{x}_1, n_2\bar{x}_2$ $n_1(1-\bar{x}_1)$ $n_2(1-\bar{x}_2)$ n_1p_1, n_2p_2 $n_1q_1, n_2q_2 \geq 5$	$(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$ $\hat{p}_1 = \bar{x_1} \qquad \hat{p}_2 = \bar{x_2}$ $\hat{q}_1 = 1 - \bar{x_1} \qquad \hat{q}_2 = 1 - \bar{x_2}$

Tabla 3.2: Intervalos de confianza para dos poblaciones

10. Ejercicios

10.1. Ejercicios resueltos

3.1 Se han generado aleatoriamente 20 datos extraídos de una población $N(0,4)$, obteniéndose que $\overline{X} = -0'052783$ y $S_c^2 = 3'17325$. Obtener un intervalo de confianza a un nivel de confianza $1-\alpha$ para el caso en que $\sigma = 1$ y en el caso en que no se conozca su valor. De igual forma, calcúlense los intervalos de confianza para σ en el caso de que $\mu = 0$ y en el caso de que se desconozca su valor.

Solución: Los resultados obtenidos vienen reflejados en la tabla 3.3.

Intervalos ($\alpha = 0'05$)	$\overline{X} = -0'052783$ $S_c^2 = 3'17325$	Amplitud del intervalo
$\sigma = 1$	$I_{1-\alpha}(\mu) = [-0'437, 0'331]$	$\Delta = 0'768$
σ desconocida	$I_{1-\alpha}(\mu) = [-0'407, 0'317]$	$\Delta = 0'724$
$\mu = 0$	$I_{1-\alpha}(\sigma^2) = [2'427, 4'236]$	$\Delta = 1'809$
μ desconocida	$I_{1-\alpha}(\sigma^2) = [2'449, 4'275]$	$\Delta = 1'826$

Tabla 3.3: Resultados: Distribución $N(\mu,\sigma)$

Hay que señalar que el intervalo encontrado cuando σ es desconocida es más pequeño que el encontrado cuando es conocida. Esto se debe a que cuando σ es desconocida se toma la cuasivarianza (o varianza), que en este caso vale $S_c^2 = 3'17325$, que es más pequeño que el valor de $\sigma^2 = 4$.

3.2 Encuéntrense intervalos de confianza para la Exponencial de parámetro λ por los métodos del pivote, la desigualdad de Tchebychev y los métodos asintóticos I y II. Compárense los resultados obtenidos para un nivel de confianza del 95 % cuando

a) la muestra es de tamaño 100 y la media muestral $\overline{X} = 0'560001$

b) la muestra es de tamaño 10000 y la media muestral $\overline{X} = 0'502409$.

Solución: Los resultados obtenidos vienen reflejados en la tabla 3.4, donde la amplitud del intervalo se denota por Δ.

Método	Intervalo de confianza	$n = 100$ $\overline{X} = 0'56000$	$n = 10000$ $\overline{X} = 0'50241$
Pivotal	$\left[\frac{\sum_{i=1}^{n} X_i}{g_{1,n,1-\frac{\alpha}{2}}}, \frac{\sum_{i=1}^{n} X_i}{g_{1,n,\frac{\alpha}{2}}} \right]$	$[0'463, 0'69]$ $\Delta = 0'227$	$[0'473, 0'554]$ $\Delta = 0'081$
Desigualdad de Tchebychev	$\left[\frac{\overline{X}}{1+\frac{1}{\sqrt{\alpha n}}}, \frac{\overline{X}}{1-\frac{1}{\sqrt{\alpha n}}} \right]$	$[0'387, 1'013]$ $\Delta = 0'626$	$[0'481, 0'526]$ $\Delta = 0'045$
I. Asintótico I (T.C.L.)	$\left[\frac{\overline{X}}{1+\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}}, \frac{\overline{X}}{1-\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}} \right]$	$[0'468, 0'697]$ $\Delta = 0'229$	$[0'493, 0'512]$ $\Delta = 0'019$
I. Asintótico II (M.V.)	$\left[\overline{X}\left(1-\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right), \overline{X}\left(1-\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) \right]$	$[0'45, 0'67]$ $\Delta = 0'22$	$[0'493, 0'512]$ $\Delta = 0'019$

Tabla 3.4: Resultados: Distribución Exp(λ) ($\alpha = 0'05$)

10.2. Ejercicios propuestos

3.1. Un fabricante diseña un experimento para estimar la tensión de ruptura media de una fibra. Para ello, observa las tensiones de ruptura, en libras, de 16 hilos de dicha fibra seleccionados aleatoriamente. Las tensiones son $20'8$, $20'6$, $21'0$, $20'9$, $19'9$, $20'2$, $19'8$, $19'6$, $20'9$, $21'1$, $20'4$, $20'6$, $19'7$, $19'6$, $20'3$, $20'7$.

Si la tensión de ruptura se distribuye según una Normal de desviación típica $\sigma = 0'45$ libras, constrúyase un intervalo al 98 % para el valor real de la tensión de ruptura promedio de la fibra.

3.2. El ayuntamiento de una determinada ciudad está interesado en estimar la cantidad promedio de dinero que gastan los turistas durante su estancia en la ciudad. Una encuesta llevada a cabo entre una muestra aleatoria de turistas obtuvo los siguientes datos expresados en euros: 150, 175, 163, 148, 142, 189, 135, 174, 168, 152, 158, 184, 134, 146, 155, 163. Suponiendo que la cantidad gastada al día es una variable aleatoria Normal, obténganse los intervalos de confianza para el promedio de dinero que gastan los turistas al día, estimados al 90, 95 y 98 %.

3.3. A partir de una muestra de 26 embotelladoras de agua, se observa que el número medio de botellas llenas es de $71'2$ por minuto y que su varianza es de $13'4$. Suponiendo Normalidad, calcule un intervalo de confianza del 95 % para el número medio de botellas rellenas.

3.4. Se está realizando un estudio para determinar el grado de precisión de las medidas efectuadas por un aparato. Para ello, se realizan 10 medidas, observándose que presentan una desviación típica de $0'23$ unidades. Suponiendo Normalidad, obténgase un intervalo de confianza al 99 % para la desviación típica de las medidas llevadas a cabo por el aparato.

3.5. Dos universidades siguen métodos distintos a la hora de matricular a sus alumnos. Para comparar el tiempo que los alumnos tardan en completar los trámites de matrícula se seleccionó al azar una muestra de 100 alumnos de cada universidad, obteniéndose los siguientes resultados, expresados en minutos,

$$\overline{x}_1 = 50'2; \quad \overline{x}_2 = 52'9; \quad S_1 = 4'8; \quad S_2 = 5'4.$$

Supuesto que ambas muestras son independientes y procedentes de poblaciones Normales, obténganse los intervalos al 90, 95 y 99 % para la diferencia de las medias del tiempo de matrícula.

3.6. Un agricultor siembra dos tipos de tomates híbridos en cinco parcelas diferentes. Las producciones, en quintales métricos por hectáreas son las siguientes:

	1	2	3	4	5
Híbrido I	90	85	95	76	80
Híbrido II	84	87	90	92	90

Si se supone que las poblaciones son Normales:

a) Construya un intervalo de confianza del 90 % para la diferencia entre las producciones medias.

b) Construya un intervalo de confianza del 90 % para el cociente de las varianzas.

3.7. Para estudiar la diferencia de estaturas medias, medidas en centímetros, de estudiantes varones en las facultades de ciencias de Cádiz y Málaga, se toma una muestra aleatoria de 15 estudiantes en cada facultad, obteniéndose:

Cádiz	182	170	175	167	171	174	181	169
	174	174	170	176	168	178	180	
Málaga	181	173	177	170	170	175	169	169
	171	173	177	182	179	165	174	

Obtenga el intervalo de confianza al 99 % para la diferencia de estaturas medias entre ambos colectivos de estudiantes. Se supone que las estaturas siguen una distribución Normal y que las varianzas poblacionales son iguales.

3.8. Se está realizando un estudio sobre la evolución del nivel de colesterol de las personas, para lo cual se seleccionan 10 individuos al azar y se les somete a una nueva dieta alimenticia durante seis meses, tras la cual se les volvió a medir el nivel de colesterol en mg/dl. Suponiendo Normalidad, obtenga un intervalo de confianza al 90 % para la diferencia de medias.

Antes	200	156	178	241	240	256	245	220	235	200
Después	190	145	160	240	240	255	230	200	210	195

3.9. Una fábrica produce barras de hierro cuya longitud sigue una distribución Normal. A partir de la muestra

$$100'9, 101'2, 100'2, 100'4, 99'8, 100'1, 101'5, 100'4, 101'7, 99'5.$$

a) Encuentre un intervalo de confianza para la longitud media.

b) Tras revisar la maquinaria, se obtuvo una nueva muestra:

$$99'7, 100'7, 97'8, 98'8, 101'4, 100'3, 98'7, 101'1, 99'4, 99'5.$$

Estudie si se produjo algún cambio en la longitud media de la barras.

3.10. Partiendo de una m.a.s. de tamaño n, construya un intervalo de confianza utilizando la desigualdad de Tchebychev con un nivel $1-\alpha$ para el parámetro θ de las siguientes ditribuciones:
a) $B(\theta)$.
b) $U(0,\theta)$.
c) $N(0,\theta)$.

3.11. En un comercio se recibe un lote de 200 artículos de los cuales 8 están defectuosos. Obténganse intervalos de confianza al 90, 95 y 99 % para la proporción de artículos defectuosos.

3.12. En una población de 10000 niños se desea hacer una campaña de vacunación. Se quiere saber cuántas vacunas deben preverse, con un 95 % de confianza, si de una m.a.s. de 90 encuestados 30 estaban vacunados.

3.13. A partir de una muestra de tamaño 100, cuya media fue $0'37$, obtenga un intervalo de confianza del $92'5$ % para el parámetro de una distribución $B(1,p)$.

3.14. A partir de una muestra de 150 enfermos escogidos entre los admitidos en un hospital durante un periodo de tres años, se observó que 129 tenían algún tipo de seguro hospitalario. En un segundo hospital, se tomó otra muestra de 160 individuos, extraída de forma similar, de los cuales 144 tenían algún tipo de seguro. Encuentre los intervalos al 90, 95 y 99 % de confianza para la diferencia de proporciones.

3.15. Con el propósito de estudiar la cantidad de nicotina de una determinada marca de cigarrillos se toma una muestra de 100 de ellos, encontrándose una media de 26 mg. Se sabe que la cantidad de nicotina se distribuye normalmente, y que su desviación típica es de 8 mg.

a) Obtenga un intervalo de confianza para el contenido medio en nicotina al 99 %.

b) Estudie cuál debe ser el tamaño de la muestra para que la amplitud del intervalo disminuya en 2 mg.

3.16. Determine el tamaño muestral necesario para estimar la media de una población Normal con varianza igual a 12 y un 90 % de confianza, de manera que el error en la estimación no sea mayor de $0'01$.