5. ESTIMACIÓN

5.1 Planteamiento clásico. Estadístico, estimador y estimación.
5.2 Métodos de estimación Estimadores máximo-verosímiles.
5.3 Intervalos de confianza.
5.4 Principales intervalos de confianza
5.5. Aplicaciones informáticas y visualización.

5.1 Planteamiento clásico. Estadístico, estimador y estimación.

La estimación (o la estimación de parámetros) es uno de los dos problemas básicos fundamentales que se plantea la inferencia estadística (junto con el contraste de hipótesis). Básicamente consiste en aproximar, de alguna forma, el valor que toma una característica desconocida de la población a partir de la información muestral.
La resolución final del problema de forma razonablemente "precisa" se basa en criterios probabilísticos y se fundamentará, en definitiva, en el comportamiento teórico que tienen la muestras aleatorias, de acuerdo a la llamadas "distribuciones muestrales", estudiadas en el tema anterior.

El planteamiento clásico de la estimación de parámetros, en general, y de la estimación puntual, en particular, consiste en utilizar un estadístico muestral apropiado para la estimación del parámetro poblacional desconocido que se desea estimar.

Contamos, en definitiva con tres elementos:

El **parámetro** (de la población) a estimar: un valor constante y desconocido. Puede ser, la media de la población, la varianza de la población, la proporción poblacional, el coeficiente de correlación entre dos magnitudes poblacionales, etc. Lo designaremos en términos generales y mientras no nos refiramos a uno concreto por la letra griega θ (Theta)

El **estimador** del parámetro: una cierta función de la muestra genérica que se utilizará o se podrá utilizar para estimar el parámetro θ. En definitiva, un estadístico (como la media muestral, la varianza muestral etc.) y, por lo tanto, una variable aleatoria cuya distribución de probabilidad, de alguna forma, dependerá del parámetro que se desea estimar. Lo designaremos en términos generales y mientras no nos refiramos a uno concreto, por la misma letra griega ,pero con un acento circunflejo, $\hat{\theta}$, o, a veces, con otras marcas especiales.

La **estimación** del parámetro: el resultado de aplicar un determinado estimador a una muestra *observada* concreta y por lo tanto el resultado final de la estimación y, por lo tanto, un *valor*.

Para distinguir entre "estimador" (herramienta que va utilizarse) y "estimación" (resultado del uso de la herramienta) deberíamos quizás utilizar la notación:

$\hat{\theta}(X_1, X_2, ..., X_n)$ para el estimador , al tratarse de una función de la muestra genérica $(X_1, X_2, ..., X_n)$ y

Simplemente $\hat{\theta}$ para la estimación, el resultado de aplicar el estimador a una muestra concreta dónde cada X_i se ha realizado tomando un determinado valor concreto.

Sin embargo, por simplicidad habitualmente no lo haremos.

Ejemplo 0: Supongamos que queremos estimar la estatura media de los españoles y para ello vamos a seleccionar al azar y con reposición a 1000 españoles y a considerar el estadístico media muestral como estimador de la media poblacional. Una vez considerado este estimador (porque se piensa que es adecuado [esto lo veremos más adelante]) se lleva a cabo la selección y se tallan los 1000 individuos resultando que la media muestral es 173.25 cm .Pues bien, este valor sería la estimación de la media poblacional.

parámetro a estimar $\theta = \mu =$ desconocido

estimador a utilizar $\hat{\theta}(X_1, X_2, ..., X_n) = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$ variable aleatoria $\left(\text{a menudo } \bar{X} \to N(\mu, \frac{\sigma}{\sqrt{n}}) \right)$

estimación llevada a cabo al aplicar el estimador a los datos muestrales obtenidos:

$$\bar{X}_0 = \frac{\overbrace{172.45+189.35+...+178.49}^{1000}}{1000} = 173.25$$

¿Puede pensarse que el valor del parámetro desconocido será (o será probablemente) el de la estimación obtenida tras realizar la muestra? Por supuesto, que no .En problemas reales es, prácticamente, imposible acertar, sólo podemos plantearnos obtener un valor que sea lo suficientemente cercano. ¿Cuándo podremos considerar "lo suficientemente cercano"? Pues cuando la diferencia sea irrelevante.

Estrictamente (lo veremos pronto) cuando la *probabilidad de que la diferencia sea "relevante" sea "despreciable"* .O Dicho de otra de otra forma: cuando el error que vayamos a cometer sea irrelevante con una muy elevada probabilidad.

Metáfora a considerar:

Podemos pensar que el **parámetro** a estimar es una **diana**, un **estimador** es como un **rifle** o un arco y una flecha y cada posible **estimación**, un **disparo**. (La distancia y el tirador son los mismos).[1]

Explotemos el símil:

1- Aunque usemos el mismo rifle, los disparos no serán siempre iguales. Un mismo estimador aplicado a distintos datos (muestras concretas) dará distintas estimaciones.

2.- Podemos disponer de distintos rifles. Los habrá buenos y malos. Nos interesa elegir los buenos. **Habrá que estudiar qué quiere decir elegir un buen estimador**.

Aunque elijamos un buen estimador (rifle) podemos hacer un mal (disparo) pero es menos probable que si disparamos con un mal rifle.

La bondad del disparo, supuesto lo demás constante, depende de la bondad del rifle y de la suerte.

La bondad de la estimación dependerá de la bondad del estimador, y de la muestra (que es aleatoria, así que de nuevo de la suerte, pero esa suerte esta "matizada" por las distribuciones muestrales)

1.1.Propiedades (deseables) de los (buenos) estimadores

1.1.1Insesgadez y varianza pequeña

[1] La metáfora del rifle está tomada de Esteban, et. al. : "Inferencia Estadística", Ed. Garceta

Nadie querría disparar con un rifle que se desvíe hacia un lado (o hacia arriba o abajo). Un rifle que no se desvíe es un rifle que puede disparar (por culpa del viento u otras condiciones ambientales) una veces, a un lado y otras, a otro, pero ,por término, medio los disparos resultarán centrados.

Pues bien, nos interesará que los estimadores tampoco **tiendan** a desviarse. Es decir que sean **insesgados** o **centrados**.

Pero no todo está en que el rifle no se desvíe. Un rifle que tenga una gran dispersión en sus disparos, aunque por término no se desvíe, puede dar lugar, en la práctica, a errores importantes en el disparo. De la misma forma, nos interesará que los estimadores que usemos tengan la menor varianza posible.

Un estimador será insesgado o centrado si su esperanza coincide con el valor que pretende estimar:

$$\hat{\theta} \text{ es un estimador insesgado de } \theta \Leftrightarrow E(\hat{\theta})=\theta$$

Si la esperanza del estimador no coincide con el parámetro que pretende estimar diremos que es estimador es sesgado, y definiremos el sesgo (bias, en inglés) como:

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

Ejemplos:

Ejemplo 1-Sea cuál sea la población la media muestral (MAS o MI) es un estimador insesgado de la media de la población ya que $E(\bar{X}) = \mu$

Ejemplo 2- El primer dato muestral obtenido es un estimador insesgado de la media de la población ya que $E(X_1) = \mu$

(Igual ocurre con el segundo ,el tercero etc. Recordemos, en cualquier caso que la media muestral tendrá menor varianza que un solo dato muestral; será preferible, por lo tanto)

Ejemplo 3.- La proporción muestral de cierta característica es un estimador insesgado de la proporción poblacional de esa característica.

Ejemplo 4.- La varianza muestral en un MAS para cualquier población NO es un estimador insesgado de la varianza de la población. Es **sesgado** . Si bien es “asintóticamente insesgado” porque su sesgo decrece con el tamaño muestral n, y si n→∞ entonces el sesgo tiende a cero.

$$E(S^2) = \sigma^2 - \frac{\sigma^2}{n} \qquad \text{(veáse 4.3.2)}$$

$$\text{b}(S^2) = E(S^2) - \sigma^2 = \sigma^2 - \frac{\sigma^2}{n} - \sigma^2 = -\frac{\sigma^2}{n} \quad \boxed{\begin{array}{c}\text{Tiene un sesgo negativo}\\ \text{o produce infraestimación}\end{array}}$$

$$\text{pero su sesgo tiende a cero si n} \to \infty : \lim_{n\to\infty}\left(-\frac{\sigma^2}{n}\right) = 0$$

Como ilustrábamos con el ejemplo 2, más arriba, pero también, con carácter general, como argumentábamos con el símil del rifle, interesa que el estimador tenga una varianza pequeña.

Mientras que podemos llegar a aspirar que el sesgo sea cero, no es razonable aspirar a que la varianza de un estimador sea cero. Sería como un si un rifle no fallará jamás. Pero sí podemos aspirar a que la varianza de un estimador alcance el mínimo valor posible. Este valor es el valor conocido como cota de Frechet-Cramer-Rao (F-C-R), que no reproducimos aquí por simplicidad. Cuando un estimador es insesgado y además tiene varianza mínima se habla de que es un estimador **óptimo** (o eficiente en sentido absoluto).

1.1.2. Error cuadrático medio y eficiencia

Como hemos visto en el epígrafe anterior nos interesa que un estimador o un rifle nos dé estimaciones centradas, acertadas por término medio, que "caiga por dónde debe caer" pero también que "no se vaya mucho". Si conseguimos ambas cosas en términos absolutos, sesgo cero y varianza igual a la cota F-C-R , ¡enhorabuena!. Pero qué hacer si contamos con estimadores insesgados con alta varianza y estimadores con baja varianza pero sesgados ¿con cuál nos quedamos?

Parece razonable optar por aquél que nos conduzca **con escasa probabilidad** a errores de cuantía importante o dicho de otra forma que por término medio nos dé errores cuya cuantía (con independencia de su signo) sea lo más baja posible.

El criterio es, entonces el de optar por el de menor **error cuadrático medio (ECM).**

Si llamamos "error del estimador" (para distinguirlo del error de estimación del que se habla en estimación por intervalo) a la diferencia entre el estimador y el parámetro, el error (del estimador) es una variable aleatoria (puesto que el estimador lo es).

$$e = e(\hat{\theta}) = \hat{\theta} - \theta$$

cuya esperanza será el sesgo : $E(e) = E(e(\hat{\theta})) = E(\hat{\theta} - \theta) = E(\hat{\theta}) - \theta = b((\hat{\theta})$

y la esperanza de su cuadrado (o error cuadrático medio) será:

$$ECM(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2 \right]$$

Obviamente es posible que un estimador que por término medio no yerre (insesgado) pueda llegar, sin embargo, a producir errores incluso muy graves que, al ser unos positivos y otros negativos, se compensen danto un promedio de cero.

En este sentido el ECM nos mide de forma precisa la "gravedad" de los errores a los que ese estimador podría conducirnos (en términos esperados o medios)

Por esta razón, y teniendo en cuenta que los errores juegan el papel de "coste" de la estimación, cuanto menor sea el error cuadrático medio de un estimador, este estimador será más **eficiente**.

$$\hat{\theta}_1 \text{ es más eficiente que } \hat{\theta}_2 \quad \Leftrightarrow \quad ECM(\hat{\theta}_1) < ECM(\hat{\theta}_2)$$

Puede obtenerse una relación que liga el sesgo y la varianza de un estimador con su ECM y por lo tanto con su eficiencia:

$$ECM(\hat{\theta}) = \left(b(\hat{\theta})\right)^2 + \text{var}(\hat{\theta})$$

Y es trivial, ver que la **eficiencia absoluta** u optimalidad coincidiría con el caso en que el sesgo fuera cero y la varianza coincidiera con la cota de Frechet-Cramer-Rao

Ejemplo 5. Supongamos que queremos estimar un parámetro θ y contamos con dos estimadores tales que el primero es insesgado y el segundo es sesgado . En concreto sabemos que sus distribuciones muestrales son,

$$estimador1 \to N(\theta,10)$$
$$estimador2 \to N((\theta+1);1)$$

y por lo tanto sus sesgos, y ECM son:

$$estimador1 \to N(\theta,10) \quad \Rightarrow b(estimador1) = 0 \quad , \quad \text{var}(estimador1) = 100$$
$$estimador2 \to N(\theta+1;1) \Rightarrow b(estimador2) = 1 \quad , \quad \text{var}(estimador1) = 1$$
$$ECM(estimador2) = 1^2 + 1 = 2 \quad < \quad ECM(estimador1) = 0^2 + 100 = 100$$

Además los errores de ambos estimadores también serán normales:

$$error\_estimador1 \to N(0,10)$$
$$error\_estimador2 \to N(1,1)$$

Ya hemos visto que el estimador es más eficiente y por lo tanto, preferible. Veamos ahora que el estimador 1 conducirá a errores “graves” con mayor probabilidad. A efectos prácticos, vamos a considerar errores graves aquellos que en valor absoluto superen las 2 unidades.
Para el primer estimador:

```
(1-(pnorm(2,0,10)-pnorm(-2,0,10)))
[1] 0.8414806
```

Para el segundo estimador:

```
(1-(pnorm(2,1,1)-pnorm(-2,1,1)))
[1] 0.1600052
```

A pesar de que el segundo estimador tiende a sobreestimar el parámetro conduce a errores graves con menor probabilidad
Considerando como graves, errores de diferentes cuantías

```
eg=c(1.5,2,2.5,3,3.5,4)
pegest1= 1-(pnorm(eg,0,10)-pnorm(-eg,0,10))
pegest2= 1-(pnorm(eg,1,1)-pnorm(-eg,1,1))
cabecera<-c("error-grave","prob.para.estim1","prob.para.estim2")
comparacion<-as.data.frame(cbind(eg,pegest1,pegest2))
colnames(comparacion)<-cabecera
comparación


    error-grave prob.para.estim1 prob.para.estim2
1         1.5        0.8807646      0.314747204
2         2.0        0.8414806      0.160005152
3         2.5        0.8025873      0.067039830
4         3.0        0.7641772      0.022781803
5         3.5        0.7263387      0.006213063
6         4.0        0.6891565      0.001350185
```

5.2 Métodos de estimación Estimadores máximo-verosímiles
En el planteamiento clásico de la estimación puntual, la elección de un estimador es un problema de decisión en el que buscamos que la elección final satisfaga de la mejor manera posible las propiedades deseables que hemos visto. Pero ¿qué métodos usar, con carácter general, para obtener buenos estimadores?. La solución general, no existe pero sí algunos métodos que sí garantizan un razonable “buen comportamiento”.

Más allá del ingenuo método de la “analogía” (estimar cualquier “cosa” poblacional, por su correspondiente “misma cosa” muestral), suelen usarse, fundamentalmente, dos: El método de los **momentos** y el método de **máxima verosimilitud**.

El método de los momentos consiste en considerar que el estimador de cada momento ordinario poblacional es el momento ordinario muestral:

$$\hat{\alpha}_r = a_r$$

Y, a partir de ahí, los distintos parámetros poblacionales, en la medida en que sean funciones de los momentos poblacionales, resultarán estimados por las correspondientes *mismas funciones* de los momentos muéstrales.

Por ejemplo:

$$\mu = \alpha_1 \Rightarrow \hat{\mu} = a_1 = \bar{X}$$
$$\sigma^2 = \alpha_2 - (\alpha_1)^2 \Rightarrow \hat{\sigma}^2 = a_2 - a_1^2 = S^2$$
$$etc.$$

El método garantiza una propiedad interesante llamada **consistencia,** que consiste (valga la redundancia) en que el estimador converge al parámetro a estimar cuando el tamaño de la muestra tiende a infinito. Sin embargo, no garantiza, en todos los casos, ni la insesgadez ni la optimalidad y tampoco “necesariamente” (como veremos), cumple con el principio de maximizar la función de probabilidad de la muestra (verosimilitud).

Además, en algunas ocasiones no da una única solución.

Por ejemplo: supongamos que queremos estimar el valor del parámetro λ de una población que sigue una distribución de Poisson. Como λ es la media de la distribución (población) pero también es la varianza, el estimador por el método de los momentos nos daría dos posibilidades : la media muestral y la varianza muestral.

5.2.Estimadores máximo-verosímilies. Método de estimación por máxima verosimilitud

El estimador máximo verosímil (EMV) de un cierto parámetro θ es aquella función de los datos muestrales (aquel estadístico) que maximice la función de verosimilitud.

Pero, ¿qué es la función de verosimilitud (Likelihood, en inglés)?

La función de verosimilitud es la función de probabilidad (de cuantía o de densidad, según el caso) de la muestra pero considerada como función del parámetro (o de los parámetros, si hay varios).

Ejemplo 6. Función de verosimilitud asociada a un muestreo aleatorio simple sobre una población normal.

Por ejemplo, si la población es Normal (N(μ,σ)) cada dato muestral obtenido al azar :

$$X_i \rightarrow N(\mu,\sigma)$$

y su función de probabilidad (densidad, en este caso) será :

$$f(X_i) = \frac{e^{-\frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$$

Si consideramos una muestra de tamaño de n con M.A.S. , cada dato muestral será una variable aleatoria estocásticamente independiente de los demás y la función de probabilidad (densidad) conjunta será el producto de las densidades (marginales):

(Aunque no hayamos estudiado con detalle las distribuciones multidimensionales de probabilidad, recordemos que la probabilidad de la intersección de sucesos **independientes** era el **producto** de sus probabilidades)

$$f(X_1, X_2,..., X_n) = \left|\text{si es M.A.S.}\right| = f(X_1)\cdot f(X_2)\cdot ...\cdot f(X_n) = \prod_{i=1}^{n} f(X_i)$$

que en nuestro caso de población normal sería:

$$f(X_1, X_2,..., X_n) = \frac{e^{-\frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}\cdot\frac{e^{-\frac{1}{2}\left(\frac{X_2-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}\cdot ...\cdot\frac{e^{-\frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} = \frac{1}{\sigma^n(2\pi)^{n/2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_n-\mu)^2}$$

Bien, vemos que esta función tiene por argumentos los valores muestrales, X_i , pero depende de los parámetros (μ,σ) . Si la consideramos, entonces como función de los parámetros estaremos hablando de la función de verosimilitud (en este caso, asociada al muestreo aleatorio simple sobre una población normal) Suele expresarse con la letra L de likelihood):

$$L(\mu,\sigma) = \frac{1}{\sigma^n(2\pi)^{n/2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_n-\mu)^2}$$

5.2.1 Sentido de la verosimilitud y de la función de verosimilitud

Para terminar de comprender el sentido que tiene considerar la función de verosimilitud y su maximización en el problema de la obtención de estimadores vamos a desarrollar un sencillo ejemplo.

Ejemplo 7. *“ejemplo de las dos monedas”*

Supongamos que tenemos dos monedas:

Una de ellas una moneda corriente con cara y cruz y que no está cargada

P(cara)= P(cruz)= 0.5

Otra, que tiene dos caras y por lo tanto la P(cara)=1 .

Alguien realiza 20 lanzamientos y nos pide que “estimemos” con qué moneda se han hecho.

Como se realizan 20 lanzamientos y se van a contar (por ejemplo) las cara que salgan, el estadístico muestral X=número de caras tendrá una distribución binomial: X→B(20,p) .

Estimar que moneda se ha usado es equivalente a estimar p. Con la particularidad que el parámetro sólo puede tomar dos valores o bien p=0.5 o bien p=1.

Supongamos que se llevan a cabo los lanzamientos y salen 20 caras X=20

¿Cuál sería la probabilidad (función de cuantía en este caso) de este resultado muestral?

$$P(X=20)=\binom{20}{20}p^{20}q^{0}=p^{20}$$

Que obviamente depende de p (el parámetro)

Si p= 0.5 la probabilidad de obtener el resultado (que se ha obtenido) será 0.5^{20} es prácticamente despreciable

Si p=1 la probabilidad de obtener el resulta será 1^{20}= 1

Obviamente la opción p= 1 es , teniendo en cuenta el resultado obtenido, mucho “más verosímil” que la opción p=0.5

(Obsérvese que si el número de caras hubiera sido otro (15, por ejemplo) la opción más verosímil habría sido la de p=0.5 ya que la verosimilitud de p=1 habría sido cero)

Este es el sentido de que la función de probabilidad de la muestra cuando se considera como dependiente de los parámetros se llame verosimilitud (apariencia de verdad, según el DRAE).

(Nota: Si recordamos el teorema de Bayes llamábamos verosimilitudes a las probabilidades del suceso B (suceso-compatible-con-las-opciones que había ocurrido, de hecho) condicionada a las distintas opciones (opciones, cuyas probabilidades a posteriori de la experiencia hecha queríamos conocer) . Y, efectivamente si consideramos las probabilidades condicionadas como relativas a cada opción nos hablan de su verosimilitud de forma igual que en nuestro ejemplo)

Bueno, en este ejemplo, hemos considerado sólo dos opciones y, además una de ellas extrema (p=1), y, por si fuera poco , el resultado “muestral” también era extremo.

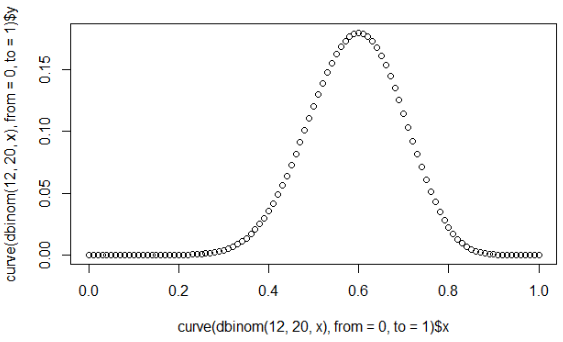
¿Qué pasaría si tuviéramos que estimar la probabilidad de cara de la moneda (cualquier valor entre 0 y 1 podría ser posible) y aún no supiéramos qué ocurre en la muestra? Pues esa es, precisamente, la situación en la que nos encontramos en un problema de la obtención de un estimador.

p toma un valor desconocido que queremos estimar (mediante un estimador)

Hacemos una experimento (MAS) de n lanzamientos Si nos salen X caras, ¿qué función de X elegiríamos como estimador de p? La opción de máxima verosimilitud (EMV) nos propone aquella función de los datos muestrales que maximice la verosimilitud (Que haga más probable lo que “efectivamente” ha pasado)

Con un pequeño script como el de abajo podemos representar la función de verosimilitud en [0,1], fijando n (tamaño muestral) y las caras (valor muestral de éxitos). (n=20 y caras=6, en este caso, pero puede modificarse a voluntad)

En este caso, el valor máximo se alcanza, en p= 0.6 que es la proporción muestral caras $\hat{p} = \frac{X}{n}$. (Precisamente, la proporción muestral será el EMV de la proporción poblacional supuesto un M.A.S.)

n=20 caras= 12 plot(curve(dbinom(caras,20,x),from=0, to=1))	

5.2.2.Algunos estimadores maximo-verosímiles

Para cada parámetro o parámetros a estimar y para cada distribución de la población la obtención del estimador máximo-verosímil será, obviamente, diferente. Vamos a ver aquí algunos estimadores EMV sin ánimo de ser exhaustivos ni tampoco de perdernos en las cuestiones técnicas de cálculo.

Antes de ver algunos ejemplos hagamos una consideración “casi” general a la hora de maximizar la función de verosimilitud.

Suele ser preferible maximizar el logaritmo de la función de verosimilitud, más que la propia función de verosimilitud. Recordemos que se trata de una función positiva (la verosimilitud lo es; es una función de densidad o cuantía, al fin y al cabo) y el logaritmo de una función positiva tiene sus máximos y mínimos en los mismos puntos. Por otra parte, ocurre que casi todas las funciones de densidad o de cuantía de los modelos más importantes incorporan expresiones exponenciales que al tomar logaritmos se simplifican. Y, finalmente, la función de probabilidad de la muestra es el producto de las funciones de probabilidad de cada dato muestral; por lo que, si tomamos logaritmos, (el logaritmo del producto es la suma de logaritmos) y la expresión también se simplificará.

Ejemplo 8 . *EMV del parámetro λ de una población de Poisson. A obtener tras un M.A.S. de tamaño n.*

La función de cuantía de la distribución de Poisson es :

$P(X) = \frac{e^{\lambda}\lambda^{X}}{X!}$ Por tanto para cada datos muestral : $P(X_i) = \frac{e^{\lambda}\lambda^{X_i}}{X_i!}$ y para toda la muestra (MAS), al ser los datos muestrales independientes:

$$P(X_1, X_2, \ldots X_n) = \frac{e^{\lambda}\lambda^{X_1}}{X_1!} \cdot \frac{e^{\lambda}\lambda^{X_2}}{X_2!} \cdot \ldots \cdot \frac{e^{\lambda}\lambda^{X_n}}{X_n!} = \frac{e^{n\lambda}\lambda^{\sum_{i=1}^{n} X_i}}{\prod_{i=1}^{n} X_i!} = \frac{e^{n\lambda}\lambda^{n\bar{X}}}{\prod_{i=1}^{n} X_i!}$$

Si la consideramos como función de λ ésta será la función de verosimilitud.

$L(\lambda)=\dfrac{e^{n\lambda}\lambda^{n\bar{X}}}{\prod_{i=1}^{n} X_i\,!}$ Obtener el EMV de λ es obtener el que maximiza esa función.

Podemos ver que maximizar $L(\lambda)$ equivale a maximizar su numerador ya que el denominador no depende de λ .Y ,siguiendo la estrategia general, maximizar este numerador es equivalente a maximizar su logaritmo:

$$\max\left(L(\lambda)=\frac{e^{-n\lambda}\lambda^{n\bar{X}}}{\prod_{i=1}^{n} X_i\,!}\right)\Rightarrow \max\left(e^{-n\lambda}\lambda^{n\bar{X}}\right)\Rightarrow \max\left(\ln(e^{-n\lambda}\lambda^{n\bar{X}})\right)\Rightarrow \max\left(n\lambda+n\bar{X}\ln(\lambda)\right)$$

igualamos a cero la primera derivada respecto a λ y obtenemos el valor que maximiza la expresión y , por lo tanto , la verosimilitud :

$$0=\frac{\partial\left(-n\lambda+n\bar{X}\ln(\lambda)\right)}{\partial\lambda}=-n+\frac{n\bar{X}}{\lambda}=0\Rightarrow\hat{\lambda}=\frac{n\bar{X}}{n}=\bar{X}$$

De modo que el estimador máximo-verosímil de λ es la media muestral.

Ejemplo 9. *EMV de la media y la varianza de una población normal (MAS de tamaño n)*

Retomamos (del ejemplo 6) la función de densidad conjunta de la muestra equivalente a la verosimilitud de μ y de σ :

$L(\mu,\sigma)=\dfrac{1}{\sigma^n(2\pi)^{n/2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(X_n-\mu\right)^2}$ Nos interesará considerar su logaritmo:

$$\ln\left(L(\mu,\sigma)\right)=-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(X_n-\mu\right)^2-n\ln(\sigma)-\frac{n}{2}\ln(2\pi)$$

Como el último término no depende de los parámetros maximizar ln(L(μ,σ)) equivale a maximizar : $-\dfrac{1}{2\sigma^2}\sum_{i=1}^{n}\left(X_n-\mu\right)^2-n\ln(\sigma)$ Para ello igualaremos las dos derivadas parciales a cero y despejaremos:

$$\left.\begin{array}{l}\dfrac{\partial\left(-\dfrac{1}{2\sigma^2}\displaystyle\sum_{i=1}^{n}\left(X_i-\mu\right)^2-n\ln(\sigma)\right)}{\partial\mu}=0\\ \dfrac{\partial\left(-\dfrac{1}{2\sigma^2}\displaystyle\sum_{i=1}^{n}\left(X_i-\mu\right)^2-n\ln(\sigma)\right)}{\partial\sigma}=0\end{array}\right\}\quad\left.\begin{array}{l}-\dfrac{1}{2\sigma^2}\cdot 2\cdot\displaystyle\sum_{i=1}^{n}\left(X_i-\mu\right)(-1)=0\\ -\dfrac{1}{2}\displaystyle\sum_{i=1}^{n}\left(X_i-\mu\right)^2\cdot(-2)\cdot\dfrac{1}{\sigma^3}-\dfrac{n}{\sigma}=0\end{array}\right\}$$

$$\sum_{i=1}^{n}\left(X_i-\mu\right)=0\Rightarrow\sum_{i=1}^{n}X_i-n\mu=0\Rightarrow\boxed{\hat{\mu}=\bar{X}}$$

$$\sum_{i=1}^{n}\left(X_i-\mu\right)^2-n\sigma^2=0\Rightarrow\left|\text{tras sustituir de la ec. de arriba}\right|\boxed{S^2=\hat{\sigma}^2}$$

Es decir que los EMV de la media y la varianza de una población normal son la media y la varianza muestrales.

Ejercicio 1 .

Finalmente dejamos como ejercicio la obtención del EMV de la proporción de una característica, que como ya se "intuyó" será la proporción muestral.

5.3. Intervalos de Confianza. Estimación por intervalo(s de confianza).

En la estimación por intervalos de confianza el objetivo es obtener un intervalo, de la menor amplitud posible (lo más preciso, posible) que, "a priori", con una probabilidad alta contenga al verdadero (y desconocido) valor del parámetro.

Antes de introducir el planteamiento general del problema, veamos, lo que queremos decir, con un ejemplo.

Ejemplo 10.

Supongamos que queremos estimar el gasto medio anual de las familias españolas. Consideramos también, (con bastante razonabilidad) que podemos suponer que la variable se distribuye con una distribución normal. Y, por último, con el único fin de simplificar la situación, vamos a suponer (esto ya no es muy razonable) que conocemos la varianza de esta variable en la población y que es de 100 €^{2}

Para estimar el gasto medio, realizaremos un M.A.S. de 100 familias. Una estimación (puntual) del gasto medio, sería considerar el valor que tome un "buen **estimador**" de la media de una población. El EMV de la media poblacional es la media muestral, además sabemos que ese estimador es insesgado y sabemos que se distribuye según una normal.

Llevamos a cabo, en la práctica, la muestra, y resulta que la media muestral es de 630€. Esta es una **estimación** puntual. Será una, quizás, buena aproximación, pero, casi con total seguridad, el valor buscado será otro.

Sin embargo si afirmáramos que el gasto medio población está entre 620 y 640 €, muy probablemente acertaríamos. En esta idea se va a basar la estimación por intervalo.

La elección del intervalo se hará de forma que el método de construcción del mismo dé ciertas garantías de éxito. A esa garantía se la llama nivel de confianza y se basará en la probabilidad asociada a un intervalo de la distribución (muestral) del estadístico utilizado.

Veámoslo: Supongamos que queremos hacer la estimación por intervalo con una confianza del 95 %.

Sabemos que la distribución del estadístico es :

$$\bar{X} \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}}) \Rightarrow \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0,1)$$

Nos planteamos, ahora, construir un intervalo en el que "caiga" el valor tipificado del estadístico con una probabilidad del 95%

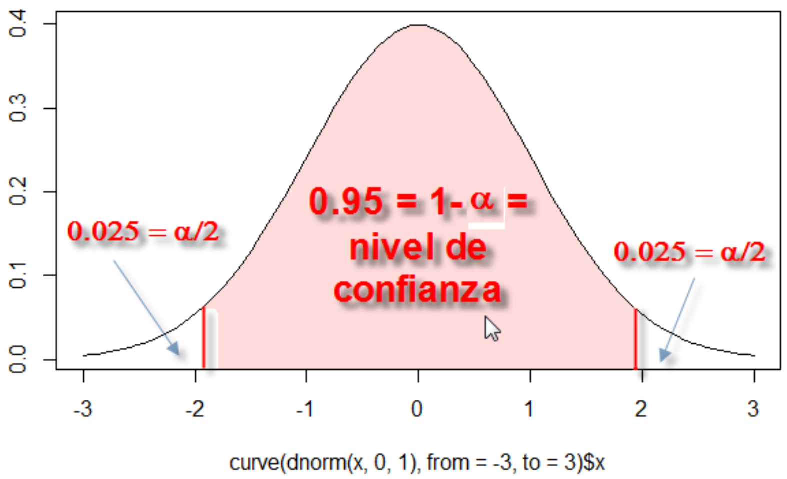
Si llamamos $z_{\alpha/2}$ al valor que, en una normal tipificada deja una cola a la derecha de $\alpha/2$, esto es, que cumple que $P(Z > z_{\alpha/2})= \alpha/2$

```
z.alfa.medios <- qnorm(α/2,0,1,lower.tail=FALSE)
```

El intervalo: $[-z_{\alpha/2}, z_{\alpha/2}]$ cumplirá que $P(-z_{\alpha/2} < Z < z_{\alpha/2})=1 - \alpha$

Para $1 - \alpha= 0.95$ tenemos que $z_{\alpha/2}=z_{0.025}= 1.96$ obtenido de la tabla, Caest o en R según:

```
qnorm(0.025,0,1,lower.tail=FALSE)
```



En nuestro caso quedará como:

$$P(-1.96 < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < 1.96) = 0.95 \Rightarrow$$

$$P(-1.96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < 1.96 \cdot \frac{\sigma}{\sqrt{n}}) = 0.95 \Rightarrow P(\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}) = 0.95$$

Como n=10 y σ^2=100, sustituyendo los valores "realmente obtenidos en la muestra" (la media muestral obtenida era 630€), nos llevaría a un intervalo: [628.04, 630.96] en el que estimaríamos que *está incluido el auténtico valor del gasto medio* de las familias españolas **con una confianza del 95%.**

Habitualmente el resultado se expresaría como **630€ con un error de ±1.96€ y una confianza del 95 %**

Estimación que para una confianza bastante alta, ha resultado ser bastante precisa (poco margen de error , poca amplitud del intervalo de estimación, poco **error de estimación**)

5.3.1.Planteamiento general de la construcción de intervalos de confianza

La obtención de un intervalo de confianza para la estimación de un parámetro arranca con la fijación del nivel de confianza con que se quiere trabajar. A este nivel de confianza se le designa por la expresión << $1-\alpha$ >> (por razones de congruencia terminológica con el contraste de hipótesis [tema 6]). Y en la fase inicial de proceso coincidirá con la probabilidad con la que trabajaremos.

1. El primer paso es la consideración de un cierto estadístico ," T ", cuya distribución sea conocida y se relacione con el parámetro a estimar , " θ " .
2. Después obtendremos un intervalo [A , B] que contenga al estadístico T con la probabilidad pedida.
   En realidad el intervalo [A,B] es ,más bien,
   $[A(\theta,(1-\alpha),(X_1,X_2,\dots,X_n)\ ),\ B((\theta,(1-\alpha),(X_1,X_2,\dots,X_n))\ ]$ ya que dependerá del parámetro, de la muestra, de la confianza.
   Una vez obtenido el intervalo tal que $P(A < T < B) = 1-\alpha$
3. Despejamos para obtener a partir de la expresión probabilística anterior un intervalo para el parámetro θ.
   $P(A^* < \theta < B^*) = 1-\alpha$ donde A* y B* serán los nuevos extremos del intervalo al despejar θ ; que ,ahora dependerán del estadístico T, de la confianza y de la muestra . Será, más bien : $[A^*(T,(1-\alpha),(X_1,X_2,\dots,X_n)\ ),\ B^*((T,(1-\alpha),(X_1,X_2,\dots,X_n))\ ]$.
   Antes de realizar la muestra concreta, el intervalo (sus extremos) dependen de los valores muestrales (del estadístico T y quizá de alguna de otro característica muestral). El intervalo es, por lo tanto, aleatorio. Y *tiene sentido* hablar de probabilidad.
4. Finalmente REALIZAMOS la muestra y sustituimos todos los valores numéricos obtenidos en la expresión del intervalo de probabilidad anterior. Entonces el intervalo ***ha dejado de ser aleatorio***, es un intervalo concreto, ya no podemos hablar de probabilidad. No podemos decir que el intervalo (ya numérico) obtenido incluirá al parámetro con una "probabilidad" de … Simplemente el intervalo, o acertará, incluyendo al parámetro, o fallará, no incluyéndolo. Pero tenemos una "importante" **confianza** en que lo incluirá porque el desarrollo teórico se ha hecho a niveles de probabilidad altos. ¿Cuánta confianza depositamos en acertar con el intervalo obtenido? Una confianza igual a la probabilidad con la hemos trabajado.

En el ejemplo 10, anterior la confianza fijada era del 95 % (α=0.025)
1. El estadístico era la media muestral cuya distribución se conoce (Normal) y se relaciona con μ.
2. El intervalo [A,B] será :

$$P(\mu - 1.96\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$, o bien ,

$$P(-1.96 < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96) = 0.95$$, si consideramos la versión tipificada de $\bar{X}$

3.El intervalo [A*,B*] lo obtenemos despejando μ

$$P(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$ .Observamos que el intervalo es aleatorio puesto que depende de la media muestral que es una v. aleatoria. Por lo tanto tiene sentido hablar de que la probabilidad es del 95% (La probabilidad de que el intervalo, que es aleatorio, incluya a μ , que es desconocido pero constante)

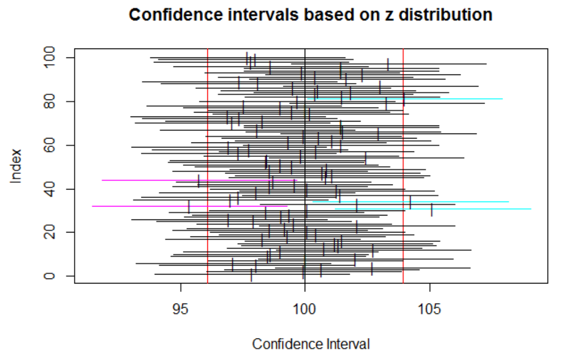
4. Al obtener la muestra, estimamos que μ pertenecerá al intervalo (numérico) obtenido (al sustituir) con una confianza (ya no hay nada aleatorio) del 95%:

$\mu \in$[628.04, 630.96] con el 95 % de confianza.

El paquete de R TeachingDemos a través de su función ci.examp( )
Permite hace repeticiones de Intervalos de confianza para visualizar que es el intervalo el que es aleatorio y ,una vez fijado, habremos acertado o no y confiamos en haberlo hecho en la medida en que la probabilidad, antes de llevarlo a cabo, era alta.

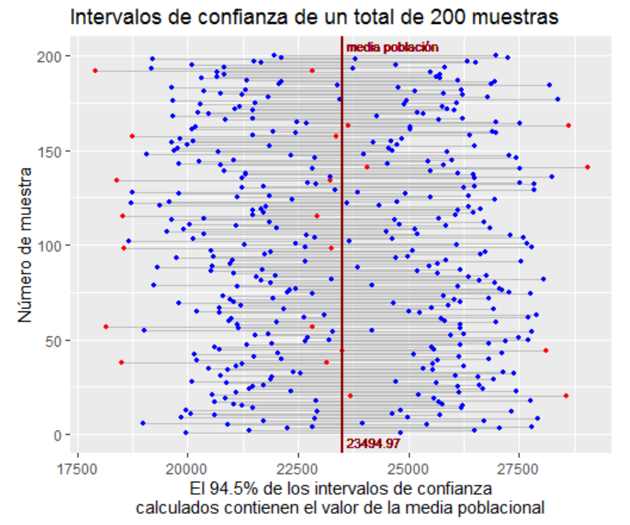
```
ci.examp(mean.sim = 100, sd = 10, n = 25, reps = 50, conf.level = 0.95,
         method = "z", lower.conf = (1 - conf.level)/2,
                   upper.conf = 1 - (1 - conf.level)/2 )

# mean.sim      Media de la población
# sd            Desviación típica de la población
# n             Tamaño de la muestra
# reps          Numero de repeticiones o muestra distintas a tomar
# conf.level    Nivel de confianza.
# method        'z', 't', or 'both', si los intervalos se basan en la normal, la t, or ambas
# lower.conf    Quantile para el límite inferior del intervalo
# upper.conf    Quantile para el límite superior del intervalo (  ídem)
# seed          semilla aleatoria
```

```
library(TeachingDemos)
ci.examp(100,10,25,100,0.95,method="z")

# los intervalos negros incluyen el parámetro
# los intervalos fucsia no lo incluyen (por debajo)
# los intervalos  cian no  lo incluyen ( por arriba)

# si no fijamos la semilla aleatoria y lo
# repetimos saldrán otros resultados distintos
```



Similarmente en el paquete {estadística} encontramos la función: nivel.confianza {estadistica}   Esta función simula una población de tamaño 100,000 de la que se extraen diversas muestras y construye los correspondientes intervalos de confianza. El objetivo es transmitir el concepto de nivel de confianza y su uso es:

```
nivel.confianza(min.pob = 2000,  max.pob =
45000,  muestras = 200,  n = 100, confianza =
0.95, grafico = TRUE, exportar = FALSE,
eplicar = FALSE)
nivel.confianza()
```



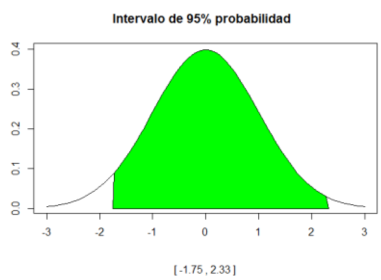
A este planteamiento general de 4 puntos debemos hacerle alguna consideración adicional más:

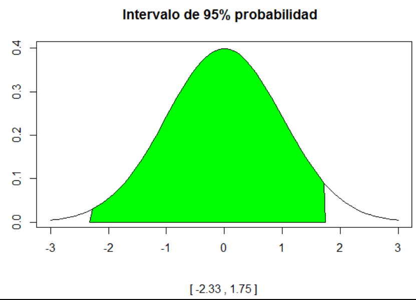
1) El estadístico (o variable aleatoria derivada de algún estadístico) que utilicemos debe ser adecuado para nuestro objetivo. Esto va a requerir, en cada caso, ver si con la información disponible, o en vías de estar disponible, nuestra elección es capaz de dar cuenta de toda la información que la muestra ofrezca sobre el parámetro.

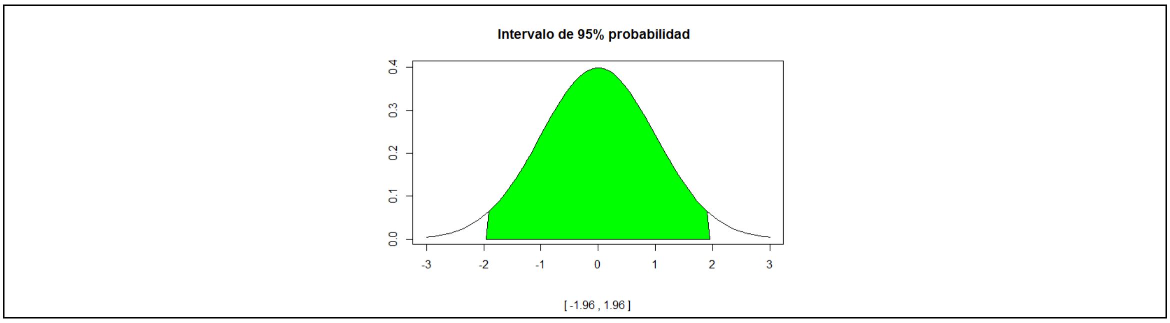
2) Una vez encontrado el estadístico hay que considerar que obtener un intervalo de probabilidad $1 - \alpha$ para el mismo no tiene una solución única (de hecho, suele haber infinitas)

Por ejemplo en una N(0,1) los intervalos :
$[-z_{0,01},z_{0.04}]$, $[-z_{0,02},z_{0.03}]$, $[-z_{0,025},z_{0.025}]$, $[-z_{0,03},z_{0.02}]$, $[-z_{0,04},z_{0.01}]$ y otros muchos acotan todos ellos en su interior una probabilidad de 0.95
Ya que todos ellos dejan dos colas que suman una probabilidad de 0.05 , y por tanto en el centro queda una probabilidad de 0.95

```
colaizquierda=0.01; coladerecha=0.04     # pueden cambiarse por las otras combinaciones
li=qnorm(colaizquierda,0,1)      # límite inferior del intervalo
ls=qnorm(colaizquierda,0,1,lower.tail=FALSE)  # límite superior del intervalo
x <- seq(-3,3,length=100)
hx <- dnorm(x,0,1)      #densidad de la normal entre -3 y 3

plot(x, hx, type="l", xlab =" ", ylab=" ", main="Intervalo de 95% probabilidad”,
sub=paste("[",round(li,2),",",round(ls,2),"]"))

i <- x >= li & x <= ls
lines(x, hx)
polygon(c(li,x[i],ls), c(0,hx[i],0), col="green")  # colorea el intervalo
```



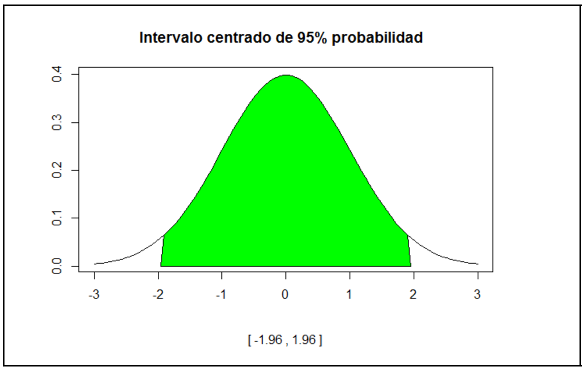


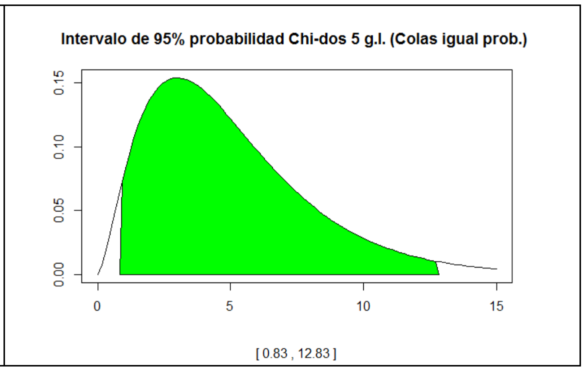


De los distintos criterios posibles para decantarse por una u otra solución los más importantes son:

1) Optar por un intervalo de estimación con mínima amplitud (lo que supone una estimación más precisa).En el caso de distribuciones simétricas y unimodales (Normal, t de Student, por ejemplo) supone optar por un intervalo centrado en la media). En el caso de distribuciones de una sola cola, χ^2 o F la solución depende de los grados de libertad de modo que no resulta operativa.

2) Optar por un intervalo que deja igual probabilidad a la cola de la derecha que a la de la izquierda. En el caso de una distribución unimodal y simétrica esta opción equivale a la primera. En el caso de distribuciones de una sola cola, χ^2 o F, esta opción aún no equivaliendo a la primera da un resultado de parecida precisión y resulta preferible por su generalidad.





Por último, antes de obtener los intervalos de confianza, más habituales, hagamos un comentario general sobre los factores que afectan a la amplitud de un intervalo de confianza y, por ello, a su **precisión.** (Cuanto más amplio resulte el IC menos precisa será la estimación).
Los dos factores que, con carácter general, influyen en la amplitud de los IC, son el **nivel de confianza** y el **tamaño de la muestra**. Ceteris paribus, a mayor nivel de confianza, mayor amplitud (y por tanto menor precisión). Y, de nuevo, ceteris paribus, a mayor tamaño de la muestra menor amplitud (mayor precisión).
Obviamente, el ideal de estimación sería la mayor confianza posible y la menor amplitud (mayor precisión o menor error) posible. Sin embargo van en sentido contario. Alcanzar un compromiso entre alta confianza y alta precisión requiere trabajar con un **adecuado tamaño muestral**. En definitiva, el tamaño de la muestra suele ir de la mano con el coste asociado del estudio y nos definirá los límites (de confianza y precisión) de las alternativas posibles.

Ejemplo 10 (Continuación). Como en nuestro ejemplo inicial el intervalo de confianza era para un nivel de 1–α el intervalo era:

$$\mu \in \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ con } 1-\alpha \text{ de confianza}$$

Que para el 95 % nos daba $z_{0.025}$= 1.96 y con n= 100 nos llevaba a un intervalo de estimación de [628.04, 630.96] con un error de estimación de ±1.96
(recordemos que n= 100 y σ=10)
Si aumentamos la confianza al 99 % tendremos que $z_{0.005}$= 2.576, que nos lleva a una estimación de [627.42, 632.58] con un error de estimación de ± 2.576
Si queremos trabajar con el 99 % de confianza pero no cometer un error mayor que el teníamos antes (±1.96) tendríamos que aumentar el tamaño de la muestra:

$$1.96 = 2.567 \frac{10}{\sqrt{n}} \Rightarrow n = \left(\frac{2.567 \cdot 10}{1.96} \right)^2 = 171.53 \text{ es decir 172 familias a muestrear (72 más)}$$

5.3.2.Obtención de intervalos de confianza
5.3.2.1. I.C. para la media una población normal (varianza poblacional conocida)
Es el caso del ejemplo de referencia anterior. Las circunstancias específicas para la construcción de este intervalo son las siguientes: Intervalo para μ, conocida σ, distribución poblacional normal, nivel de confianza dado 1–α.

Conocemos que la media muestral se distribuye según
$\bar{X} \to N(\mu, \frac{\sigma}{\sqrt{n}}) \Rightarrow \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,1)$ a partir de ahí obtener el intervalo es sencillo y aplicando el esquema general de actuación:

Despejando del intervalo de probabilidad:

$P(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 0.95$ y de aquí el intervalo de confianza será:

$$\boxed{\mu \in \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ con } 1-\alpha \text{ de confianza}}$$

Si no se tratara de una MAS sino de MI (tamaño de la población, N, pequeño) la distribución de la media muestral se vería afecta en su desviación típica por la raíz cuadrada del Factor Corrector para Poblaciones Finitas por lo que al incorporarlo al intervalo, despejar y obtener el I.C. nos resultaría :

$$\mu \in \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \text{ con } 1-\alpha \text{ de confianza}$$

Como esta consideración para el caso de poblaciones finitas **siempre** va a ser así, en los próximos casos la omitiremos, dándola por supuesta.

Nos queda hacer una pequeña discusión sobre el hecho de que conozcamos σ :
Estando en un contexto en el que vamos a estimar μ y, por lo tanto, no lo conocemos, parece absurdo “conocer σ” ; y lo es. Sin embargo, como veremos, el intervalo tiene su

uso en aquellos casos en que la muestra sea lo suficientemente grande (n > 30) sin más que considerar en la expresión que la varianza de la población coincide con la de la muestra.

5.3.2.2. I.C. para la media una población normal con varianza poblacional desconocida o, a efectos prácticos, n pequeño (n< 30)

Las circunstancias específicas para la construcción de este intervalo son las siguientes: Intervalo para μ , σ desconocida, Distribución poblacional **normal**, Nivel de confianza dado, $1-\alpha$.

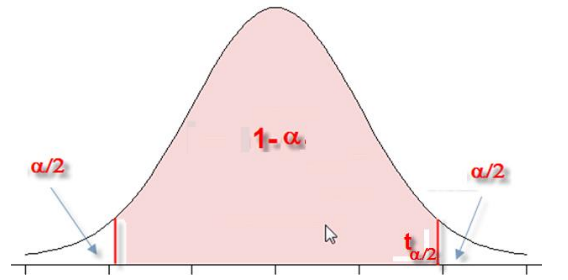
Conocemos que $\frac{\bar{X}-\mu}{S}\sqrt{n-1} \to t_{n-1}$. A partir de aquí obtenemos un intervalo de probabilidad $1-\alpha$ para el "estadístico t":

$$P(-t_{\alpha/2} < \frac{\bar{X}-\mu}{S}\sqrt{n-1} < t_{\alpha/2}) = 1-\alpha$$ y despejando μ:

$$P(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n-1}} < \mu < \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n-1}}) = 1-\alpha$$ como intervalo de probabilidad y

finalmente : $$\boxed{\mu \in \left[\bar{X} \pm t_{\alpha/2}\frac{S}{\sqrt{n-1}}\right] \text{ con } 1-\alpha \text{ de confianza}}$$

donde el valor crítico $t_{\alpha/2}$ es el que deja una cola de probabilidad de $\alpha/2$ en una t de Student con n-1 grados de libertad.



Fijémonos en que si $n \to \infty$ entonces $t_{\alpha/2} \to z_{\alpha/2}$ por la convergencia de la t a la normal.

De igual forma se $n \to \infty$ la raíz de n-1 y la raíz de n son casi iguales y la expresión de I.C. tiende a la del I.C. con varianza conocida, como anticipábamos en el punto anterior, sin más que considerar que la desviación típica muestral como σ .

Ejemplo 11: Las ventas diarias de cierta oficina comercial se supone que siguen una distribución normal. Para estimar el volumen medio de ventas por día se realiza una muestra de 10 días escogidos al azar ,resultando que la media de las ventas de esos 10 días es 100 u.m. con una desviación típica de 4 u.m. Dar un intervalo de estimación para el volumen medio de ventas por día con una confianza del 95 % . (ir a script de realización en CEACES o CAEST)

Estamos ante:

Población normal; n=10 (muestra pequeña);S=4(poblacional desconocida); media muestral=100 ; $1-\alpha = 0.95$, luego $\alpha = 0.05$, con lo que $t_{\alpha/2}$ (9 g.l.) = 2.262 (según tabla) (ir a tabla de la t) (En R : qt(0.975,9) o qt(0.025,9,lower.tail=FALSE)) y dado que el intervalo a utilizar (nada dicen de irrestricto; luego m.a.s) será

$$\mu \in \left[\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n-1}} \right] \text{ con } 1-\alpha \text{ de confianza}$$ resultando :

μ ∈ [96'99;103'01] con el 95 % de confianza

En el paquete {estadística} la función ic.media permite obtener intervalos de confianza para la media de una población normal o no y con varianza conocida o no. Con la que podemos replicar los ejemplos 10 y 11 anteriores

```
ic.media(x,  variable = NULL, introducir = FALSE,  poblacion = c("normal","desconocida"),  var_pob = c("conocida","desconocida"), confianza = 0.95, grafico = FALSE)
ic.media(introducir = TRUE,poblacion  ="normal",var_pob = "conocida",confianza=0.95)
ic.media(introducir = TRUE,poblacion  ="normal",var_pob = "conocida",confianza=0.99)
ic.media(introducir = TRUE,poblacion  ="normal",var_pob = "desconocida",confianza=0.95)
```

```
> ic.media(introducir = TRUE,poblacion  ="normal",var_pob = "conocida",confianza=0.95)
Introducir el tamaño de la muestra: 100
Introducir el valor de la media muestral: 630
Introducir el valor de la varianza poblacional: 100
  limite_inferior limite_superior
1          628.04          631.96
> ic.media(introducir = TRUE,poblacion  ="normal",var_pob = "conocida",confianza=0.99)
Introducir el tamaño de la muestra: 100
Introducir el valor de la media muestral: 630
Introducir el valor de la varianza poblacional: 100
  limite_inferior limite_superior
1        627.4242        632.5758
> ic.media(introducir = TRUE,poblacion  ="normal",var_pob = "desconocida",confianza=0.95)
Introducir el tamaño de la muestra: 10
Introducir el valor de la media muestral: 100
Selecciona el valor que quieres utilizar?
 1. "Varianza muestral"
 2. "Cuasivarianza muestral"
1
Introducir el valor de la varianza muestral: 16
  limite_inferior limite_superior
1        96.98379        103.0162
```

Nótese que como pide introducir la varianza ésta es 16 (cuadrado de la desv.típica)

5.3.2.3. I.C. para la diferencia de dos medias de dos poblaciones normales (varianzas poblacionales conocidas) a partir de dos MAS independientes

Partiendo de la distribución del estadístico diferencia de medias:

$$\bar{X} - \bar{Y} \rightarrow N\left(\mu_x - \mu_y ; \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right)$$ Podemos obtener el intercalo de probabilidad 1–α:

$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} < z_{\alpha/2} \right) = 1-\alpha$$ y despejando $\mu_x - \mu_y$:

$$P\left((\bar{X} - \bar{Y}) - z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} < \mu_x - \mu_y < (\bar{X} - \bar{Y}) + z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right) = 1-\alpha$$

que finalmente una vez consideradas las dos muestras concretas nos llevará aun intervalo de confianza :

$$\boxed{\left(\mu_x-\mu_y\right)\in\left[(\bar{X}-\bar{Y})\pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}\right] \text{ con } 1-\alpha \text{ de confianza}}$$

En la práctica, el hecho de que las varianza sean conocidas vendrá a "traducirse" como tamaños muestrales grandes y consideraremos las varianzas muestrales como si fueran las poblacionales.

Ejemplo 12:
Queremos conocer la diferencia entre las ventas medias diarias de dos de nuestros supermercados ubicados en ciudades distintas. Para ello obtenemos información aleatoria de 300 días de nuestro supermercado de Avila, resultando: ventas medias diarias 20 u.m. desviación típica 5 .u.m. La información resultante de 250 días aleatorios en nuestro supermercado de Badajoz fue: media de ventas 15 u.m. desviación 8 u.m. . Si para conocer de diferencia entre las ventas medias construimos un intervalo de confianza con nivel de confianza del 90%. Estimar el error que podemos cometer al intentar conocer dicha diferencia.

La estimación de la diferencia la obtenderemos de aplicar el estimador " diferencia de medias muestrales" a los datos obtenidos : 20-15 = 5 u.m (a favor del supermercado de Ávila). Para ver el error de la estimación trabajando con un nivel de confianza del 90% construimos el I.C:

$$\left(\mu_x-\mu_y\right)\in\left[(\bar{X}-\bar{Y})\pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}\right] \text{ con } 1-\alpha \text{ de confianza } :$$

Donde :
$\bar{X}-\bar{Y}$ = 5 (ya lo hemos dicho) ;
$z_{\alpha/2}$= 1.645 (De las tablas, la Caest o en R qnorm(0.95,0,1)
n_x=300, n_y=250 y asumiremos las varianzas poblaciones como iguales a las muestrales : σ_x^2 = 25 ; σ_y^2 = 64

$$\left(\mu_x-\mu_y\right)\in\left[5\pm 1.645\sqrt{\frac{25}{300}+\frac{64}{250}}\right] \text{ con 90 de confianza}$$

Esto es $\left(\mu_x-\mu_y\right)\in\left[5\pm 0.95825\right]$ u.m. con el 90% de confianza de forma que el error será de $\pm$0.96 u.m con una confianza del 90%

5.3.2.4. I.C. para la diferencia de dos medias de dos poblaciones normales (varianza poblacional común pero desconocida) a partir de dos MAS independientes.

En esta situación (en la práctica cuando $n_x+n_y < 32$) tendremos que usar el estadístico t para la diferencia de medias:

$$\frac{\bar{X}-\bar{Y}-(\mu_x-\mu_y)}{\sqrt{n_x+n_y}\cdot\sqrt{n_xS_x^2+n_yS_y^2}}\sqrt{n_x\cdot n_y}\cdot\sqrt{n_x+n_y-2}\rightarrow t_{n_x+n_y-2}$$

Que nos llevará a un I.C :

$$\mu_x - \mu_y \in \left[\left(\bar{X} - \bar{Y} \right) \pm t_{\alpha/2} \frac{\sqrt{n_x + n_y} \cdot \sqrt{n_x S_x^2 + n_y S_y^2}}{\sqrt{n_x \cdot n_y} \cdot \sqrt{n_x + n_y - 2}} \right] \text{ con una confianza de } 1-\alpha$$

En el paquete {estadística} la función ic.diferencia.medias permite obtener intervalos de confianza para la diferencia de dos media de poblaciones normales o no y con varianzas conocidas o no y en este segundo caso consideradas iguales o no. (por defecto : normales, conocida y distintas)

```
ic.diferencia.medias(x, variable = NULL,introducir = FALSE, poblacion = c("normal","desconocida"),
var_pob = c("conocida","desconocida"), iguales = FALSE, confianza = 0.95)
```

```
# replicamos el ejemplo 12
ic.diferencia.medias(introducir = TRUE,confianza=0.9)
> ic.diferencia.medias(introducir = TRUE,confianza=0.9)
Introducir el tamaño de la muestra 1: 300
Introducir el valor de la media muestral 1: 250
Introducir el tamaño de la muestra 2:
> ic.diferencia.medias(introducir = TRUE,confianza=0.9)
Introducir el tamaño de la muestra 1: 300
Introducir el valor de la media muestral 1: 20
Introducir el tamaño de la muestra 2: 250
Introducir el valor de la media muestral 2: 15
Introducir el valor de la varianza poblacional 1: 25
Introducir el valor de la varianza poblacional 2: 64
  limite_inferior limite_superior
1        4.041835        5.958165
#y si el tamaño de las muestras fuera pequeño, nx=15 y ny=10 (intervalo t) suponiendo  la misma varianza poblacional
ic.diferencia.medias(introducir= TRUE, var_pob = "desconocida",iguales= TRUE,confiaza=0.95)
> ic.diferencia.medias(introducir= TRUE, var_pob = "desconocida",iguales= TRUE,confianza=0.95)
Introducir el tamaño de la muestra 1: 15
Introducir el valor de la media muestral 1: 20
Introducir el tamaño de la muestra 2: 10
Introducir el valor de la media muestral 2: 15
Selecciona el valor que quieres utilizar:
 1. "Varianza muestral"
 2. "Cuasivarianza muestral"
1
Introducir el valor de la varianza muestral 1: 25
Introducir el valor de la varianza muestral 2: 64
  limite_inferior limite_superior
1      -0.6102484        10.61025

#y considerando diferente varianza poblacional
ic.diferencia.medias(introducir= TRUE, var_pob = "desconocida",confiaza=0.95)
> ic.diferencia.medias(introducir= TRUE, var_pob = "desconocida",confianza=0.95)
Introducir el tamaño de la muestra 1: 15
Introducir el valor de la media muestral 1: 20
Introducir el tamaño de la muestra 2: 10
Introducir el valor de la media muestral 2: 15
Selecciona el valor que quieres utilizar:
 1. "Varianza muestral"
 2. "Cuasivarianza muestral"
1
Introducir el valor de la varianza muestral 1: 25
Introducir el valor de la varianza muestral 2: 64
  limite_inferior limite_superior
1       -1.357591        11.35759
```

5.3.2.5.I.C. Para la proporción, p , de una característica

A partir de la distribución de la proporción muestral

$\hat{p} = \frac{X}{n} \to N(p, \sqrt{\frac{pq}{n}})$ Podemos obtener un intervalo de probabilidad 1– α:

$P(-z_{\alpha/2} < \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} < z_{\alpha/2}) = 1-\alpha$ y desde aquí despejar la proporción poblacional:

$P(\hat{p} - z_{\alpha/2}\sqrt{\frac{pq}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{pq}{n}}) = 1-\alpha$ para obtener después el IC.

Al sustituir los valores muestrales para obtener el intervalo de confianza, nos encontraremos, sin embargo, con un problema : para poder estimar p necesitamos conocer p , ya que nos aparece en las cantidades que se suman/restan a la proporción muestral para obtener los extremos del intervalo.

Hay dos estrategias posibles (en realidad tres) :
1.- Hacer alguna consideración adicional sobre p. Quizás sepamos aproximadamente su valor. O conozcamos una cota superior del mismo.
2.- Tomar el valor muestral.
En realidad la opción que suele utilizarse es un caso "especial" de 1 (de ahí lo tres estrategias):
Ponernos en "el peor caso posible" entendiendo por tal, áquel que nos conduciría al intervalo de estimación con mayor amplitud (máximo error de estimación) .Proponer una solución así nos permite hacer una estimación que se cumplirá con una confianza igual, o **mayor incluso**, a la que estamos utilizando.
¿Cuál es este peor caso posible?: p=q =0.5 . Máxima varianza en la binomial, situación que maximizaría el factor que aparece multiplicando en la **horquilla** de la estimación.
En consecuencia:
Según la opción 2 (la menos habitual y menos recomendable) el IC quedaría como:

$$p \in \left[\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \right] \text{con una confianza de } 1-\alpha$$

Según la opción 1 considerando un valor "supuesto para p" de p*

$$p \in \left[\hat{p} \pm z_{\alpha/2}\sqrt{\frac{p^*q^*}{n}} \right] \text{con una confianza de } 1-\alpha$$

Según la opción habitualmente preferida :

$$\boxed{p \in \left[\hat{p} \pm z_{\alpha/2}\sqrt{\frac{0.5\cdot 0.5}{n}} \right] \text{con una confianza de } 1-\alpha}$$

Ejemplo 13 :
En una investigación comercial se muestrea a 100 individuos resultando que 25 de ellos han comprado nuestro producto .Dar un intervalo para la proporción de penetración en el mercado con una probabilidad (nivel de confianza) del 95 %.

Poniéndonos en el caso más desfavorable (p=q=0.5)

$$p \in \left[\hat{p} \pm z_{\alpha/2}\sqrt{\frac{0.5\cdot 0.5}{n}} \right] \text{con una confianza de } 1-\alpha$$

Como ya sabemos que, para $1-\alpha=0.95$, $z_{\alpha/2}$=1.96 , n= 100 y $\hat{p}$ =0.25 nos llevará a una estimación por IC. :

$$p \in \left[0.25 \pm 1.96\sqrt{\frac{0.5 \cdot 0.5}{100}} \right]$$ con una confianza del 95% ,esto es: $p \in [0.25 \pm 0.098]$ con el 95 de confianza (es un error de 9.8 puntos porcentuales que es mucho pero hay que pensar que la muestra es pequeña 100 encuestados)

Nota: es muy habitual que en muchas encuestas para estimar proporciones se trabaje con 95.5 % de confianza (y p=q=0.5) la razón es que en tal caso el valor crítico $z_{\alpha/2}$ es aproximadamente 2 que se simplifica con la raíz de 0.5×0.5 y el error acaba siendo el recíproco de la raíz de n

En el paquete {estadística} la función ic.proporcion permite obtener intervalos de confianza para una proporción ,permitiendo la consideración de que el muestreo es irrestricto (por defecto, no lo es) pero **sólo permite** la solución del intervalo aproximando p por el valor muestral ; es decir, **la estrategia 2** de las consideradas arriba.

```
ic.proporcion(x,variable = NULL,introducir = FALSE, irrestricto = FALSE, confianza = 0.95,  grafico = FALSE)
```

```
#el ejemplo anterior pero con la estrategia quedaría
ic.proporcion(introducir= T,grafico=T)
> ic.proporcion(introducir= T,grafico=T)
[1] "Intervalo de confianza de una proporción. El tamaño de la muestra es grande."
Introducir el tamaño de la muestra: 100
Introducir el valor de la proporción muestral: 0.25
¿Quieres aproximar el valor de p por la proporción muestral?
 1. "Sí"
 2. "No"
1
[1] "Como n es suficientemente grande, se aproxima el valor de p poblacional por su estimación puntual (p muestral)"
[[1]]
  limite_inferior limite_superior
1       0.1651311       0.3348689
```


```
# para aplicar la estrategia 1 (p=q=0.5) podemos construir una función de usuario , ad-hoc
iconf.proporcion=function(n,p,confianza=0.95){
    errorest=qnorm((1-confianza)/2,lower.tail=FALSE)*0.5/sqrt(n)
    sol = c(p-errorest,p+errorest)
    print("el intervalo de confianza para la proporción es ")
    print (paste("[",sol[1],",",sol[2],"]"))
    print("")
    print("habiéndose considerado p=q=0.5")
    print("")
   print(paste("el error de estimación es de +/- ",errorest))
  }
iconf.proporcion(100,0.25)
> iconf.proporcion(100,0.25)
[1] "el intervalo de confianza para la proporción es "
[1] "[ 0.152001800772997 , 0.347998199227003 ]"
[1] ""
[1] "habiéndose considerado p=q=0.5"
[1] ""
[1] "el error de estimación es de +/-  0.0979981992270027"
```

5.3.2.6.I.C. Para la diferencia de dos proporciones

Basándonos en la distribución de la diferencia de dos proporciones muestrales

$$\hat{p}_x - \hat{p}_y \rightarrow N(p_x - p_y, \sqrt{\frac{p_x q_x}{n_x} + \frac{p_x q_x}{n_x}})$$ podemos obtener el correpondiente intervalo de probabilidad $1-\alpha$:

$$P\left(\left(\hat{p}_x-\hat{p}_y\right)-z_{\alpha/2}\sqrt{\frac{p_xq_x}{n_x}+\frac{p_yq_y}{n_y}}<p_x-p_y<\left(\hat{p}_x-\hat{p}_y\right)+z_{\alpha/2}\sqrt{\frac{p_xq_x}{n_x}+\frac{p_yq_y}{n_y}}\right)=1-\alpha$$

Una vez obtenidos los datos y haciendo la misma consideración que en el caso de una sola proporción respecto a p=q=0.5 nos quedaría el IC:

$$\boxed{p_x-p_y\in\left[\left(\hat{p}_x-\hat{p}_y\right)\pm z_{\alpha/2}\sqrt{\frac{0.5\cdot 0.5}{n_x}+\frac{0.5\cdot 0.5}{n_y}}\right]\text{ con una confianza de }1-\alpha}$$

Al igual que antes la función ic.diferencia.proporciones del paqueta {estadística} nos dará el intervalo de estimación considerando los valores de las p muestrales (estrategia 2) si queremos usar la estrategia 1 (px=py=qx=qy=0.5) tendremos que usar una función ad-hoc de usuario.

Ejemplo: para estimar la diferencia entre la proporción de votantes satisfechos con la acción del gobierno antes y después de la última decisión de política económica tomada por éste se han hecho dos encuestas que arrojan la siguiente información. Llevar a cabo la estimación con un 99% de confianza

Encuesta	Proporción muestral de satisfechos	Tamaño de la muestra
Antes de la acción	0,3	400
Después de la acción	0,28	324

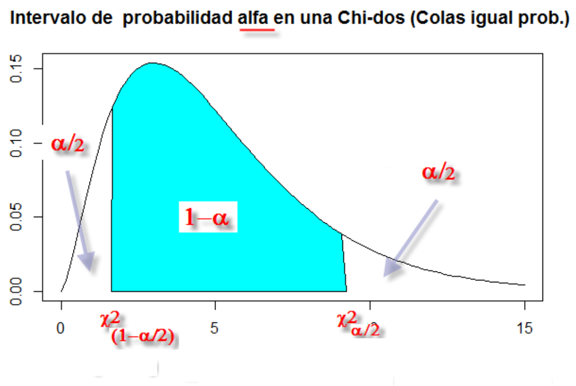
```
#siguiendo la estrategia de considerar las p muestrales // paquete {estadística}
ic.diferencia.proporciones(introducir=T,confianza=0.99)
> ic.diferencia.proporciones(introducir=T,confianza=0.99)
[1] "Primero vas a introducir los datos de la muestra 1 y a continuacipn introducirás los datos de la muestra 2"
[1] "Si los datos provienen de encuestas realizadas antes y después de una determinada acción, introduce primero los da
tos de la encuesta realizada después de dicha acción"
Introducir el tamaño de la muestra 1: 400
Introducir el valor de la proporción muestral 1: 0.3
Introducir el tamaño de la muestra 2: 324
Introducir el valor de la proporción muestral 2: 0.28
  limite_inferior limite_superior
1     -0.06724506       0.1072451
#siguiendo la estrategia de p1=p2=q1=q2=0.5 //función ad-hoc de usuario
iconf.diferencia.proporciones=function(n1,p1,n2,p2,confianza=0.95){
   alfamedios=(1-confianza)/2
   errorest=qnorm(alfamedios,lower.tail =FALSE )*sqrt((0.25/n1)+(0.25/n2))
   sol=c(p1-p2-errorest,p1-p2+errorest)
   print (paste("[",sol[1],",",sol[2],"]"))
   print("")
   print("habiéndose considerado p=q=0.5")
   print("")
  print(paste("el error de estimación es de +/- ",errorest))
 }
iconf.diferencia.proporciones(400,0.3,324,0.28,confianza=0.99)
> iconf.diferencia.proporciones(400,0.3,324,0.28,confianza=0.99)
[1] "[ -0.0762617751649517 , 0.116261775164952 ]"
[1] ""
[1] "habiéndose considerado p=q=0.5"
[1] ""
[1] "el error de estimación es de +/-  0.0962617751649517"
```

5.3.2.7. IC para la varianza de una población normal

Partiendo de la distribución muestral de la suma de cuadrados estandarizada:

$\frac{nS^2}{\sigma^2} \to \chi^2_{n-1}$ y de lo comentado para la construcción de I.C. en el caso de que la distribución fuera de una sola cola tendremos que empezar por considerar el intervalo para esta variable como:

$$P(\chi^2_{(1-\alpha/2)} < \frac{nS^2}{\sigma^2} < \chi^2_{\alpha/2}) = 1-\alpha$$



Y despejando:

$$P(\frac{1}{\chi^2_{\alpha/2}} < \frac{\sigma^2}{nS^2} < \frac{1}{\chi^2_{(1-\alpha/2)}}) = 1-\alpha \Rightarrow P(\frac{nS^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{nS^2}{\chi^2_{(1-\alpha/2)}}) = 1-\alpha$$

Y resultando, tras tomar la muestra un IC de :

$$\boxed{\sigma^2 \in \left[\frac{nS^2}{\chi^2_{\alpha/2}}, \frac{nS^2}{\chi^2_{(1-\alpha/2)}}\right] \text{ con confianza } 1-\alpha}$$

La función ic.varianza del paquete {estadística} nos permite automatizar este intervalo. Por ejemplo para obtener un IC. del 99% para la varianza de una población normal a partir de una muestra de tamaño 50 en la que la varianza muestral ha resultado ser 25 tendríamos :

```
ic.varianza(introducir=T,confianza=0.99)
> ic.varianza(introducir=T,confianza=0.99)
Introducir el tamaño de la muestra: 50
Selecciona el valor que quieres utilizar:
 1. "Varianza muestral"
 2. "Cuasivarianza muestral"
1
Introduce el valor de la varianza muestral: 25
[1] "Intervalo de confianza para la varianza poblacional, supuesta desconocida la media poblacional."
  limite_inferior limite_superior
1        15.97838        45.87266
```

5.3.3 Determinación del tamaño muestral

Suele ser habitual que, antes de obtener la muestra y realizar el estudio que se tenga previsto, nos planteemos determinar el tamaño muestral que deberá tener la muestra para que, trabajando con un determinado nivel de confianza, hagamos unas estimaciones (por intervalo) que tengan una determinada amplitud en su intervalo. Esta amplitud no es otra cosa que el **error de la estimación** habitualmente representada por la horquilla ± E.

En todos los intervalos esta horquilla depende (además de la confianza) de algunos factores relacionados con la variabilidad (muestral, poblacional, supuesta o máxima) y del tamaño de la muestra.

Conocidos todos los demás factores influyentes en el error de la estimación y fijado el nivel de confianza nos bastará con despejar.

La función *muestra* del paquete {estadística} permite automatizar la solución del problema

Ejemplo 14.
En un sondeo electoral se quiere estimar los porcentajes de votos de los partidos con una confianza del 95.5 % y con error de ± 1% Supuesto un MAS ¿a cuántos electores habrá que sondear?
El intervalo es :

$$\boxed{p \in \left[\hat{p} \pm z_{\alpha/2}\sqrt{\frac{0.5 \cdot 0.5}{n}} \right] \text{con una confianza de } 1-\alpha}$$

de forma que si el error es de ± 1% = ± 0.01 en tanto por 1 , el valor crítico $z_{\alpha/2}$ para el 95.5 % de confianza será : $z_{0.0225} = 2.0046$ (aunque a menudo se aproxima por 2) De forma que :

$$\pm 0.01 = \pm z_{\alpha/2}\sqrt{\frac{0.5 \cdot 0.5}{n}} = \pm 2 \cdot \frac{0.5}{\sqrt{n}} \Rightarrow n = \left(\frac{1}{0.01} \right)^2 = 10000$$