TEMA 7. Estimación por intervalos de confianza

Hasta ahora hemos visto los métodos de estimación por puntos. En estos métodos siempre se da *«un valor único»* como estimación del parámetro poblacional desconocido.

¿Cómo se supera este inconveniente?

Introduciendo un nuevo método de estimación, que llamaremos *«estimación por intervalos»*. En este método se trata de hallar dos estadísticos t_1 $(X_1, X_2,..., X_n)$ y t_2 $(X_1, X_2,..., X_n)$ que serán, por tanto, variables aleatorias como ya sabemos, y tales que para el parámetro poblacional desconocido θ, se cumpla:

$$P(t_1 \leq \theta \leq t_2) = \gamma$$

Al intervalo (t_1, t_2) se le denomina *«Intervalo de confianza»(*. (Resaltemos que los extremos t_1 y t_2 son variables aleatorias, que dependen de los valores muestrales).

7.1 Al finalizar el tema el alumno debe conocer........

- ✓ Características de los métodos de construcción de intervalos de confianza
- ✓ Intervalos de confianza en poblaciones normales.
- ✓ Intervalos de confianza en poblaciones no necesariamente normales.
- ✓ Intervalos de confianza de una proporción.
- ✓ Estimación del tamaño muestral.

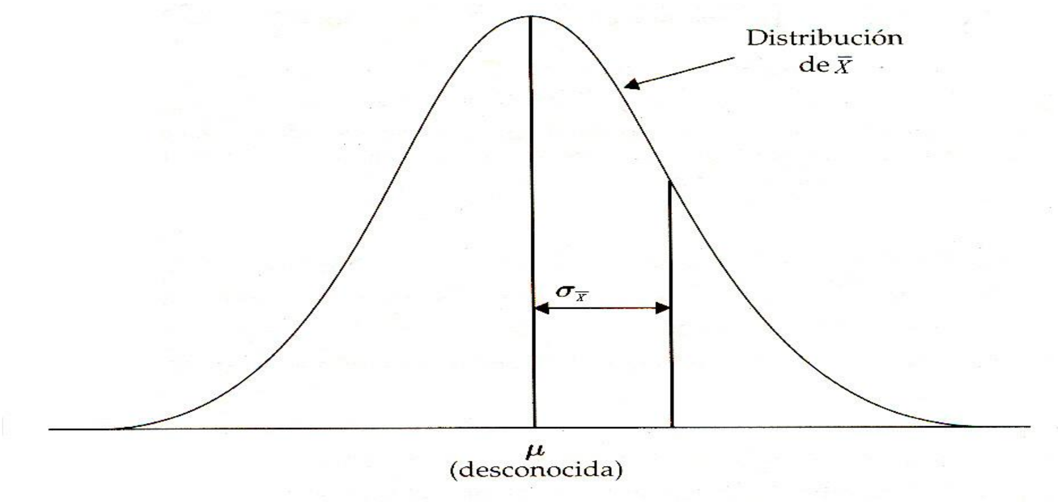
7.2 Características de los métodos de construcción de intervalos de confianza.

Supongamos que una compañía dedicada a la producción de piezas para el sector de automoción, quiere analizar la productividad de una de sus plantas industriales y para ello realiza un control de las piezas fabricadas. Mediante el control realizado sobre una muestra aleatoria de piezas, estima que el 10 % de todas las piezas son defectuosas. El gerente que se encuentra con este dato se hace la siguiente pregunta: ¿Puedo estar seguro de que el verdadero porcentaje de piezas defectuosas está entre 5 % y 15 %? Esta clase de preguntas requieren información que va más allá de la contenida en una simple estimación puntual; se trata de buscar un rango de valores entre los que posiblemente se encuentre la cantidad que se estima.

En una estimación por intervalo se define un intervalo dentro del cual puede estar el parámetro desconocido. El intervalo suele ir acompañado de una afirmación sobre el nivel de confianza que se puede asignar a su precisión, por ello se llama intervalo de confianza.

- Construcción e interpretación de los intervalos de confianza.

Podemos hacernos la siguiente pregunta; ¿Cómo podemos construir un intervalo y afirmar que confiamos al 95,5 % que ese intervalo contiene a μ, si ni siquiera sabemos cuál es al media poblacional?



Viendo el gráfico y recordando la valoración de normalidad que vimos, el 95,5 % de todas las medias muestrales se encontrarán entre $\mu \pm 2\delta$. No debemos olvidar que a partir de cualquier población podemos obtener muchas muestras diferentes de un mismo tamaño determinado, cada una de ellas con su propia media $(\overline{X}_1, \overline{X}_2, \cdots)$. Todas estas medias muestrales dan lugar a un intervalo de confianza, que podrá incluir o no a la media poblacional. Por consiguiente, si a partir de cualquier media muestral nos desplazamos dos errores típicos por encima y otros dos por debajo de esa media $(\overline{X}_1 \pm 2\delta, \overline{X}_2 \pm 2\delta, \cdots)$, podemos tener una confianza del 95,5 % de que el intervalo resultante contiene a la media poblacional desconocida.

Si tomáramos 100 muestras aleatorias de tamaño n de la misma población y calculáramos los extremos del intervalo para cada muestra, entonces esperamos que aproximadamente el 95,5% de los intervalos contendrán en su interior el verdadero valor del parámetro μ y el 4,5% restante no lo contendrán. Pero como nosotros, en la práctica, sólo tomamos una muestra aleatoria y por tanto sólo tenemos un intervalo de confianza, no conocemos si nuestro intervalo es uno del 95,5% o uno del 4,5 %, por eso hablamos de que tenemos un nivel de confianza del 95,5 %.

El valor de 95,5 % recibe el nombre de nivel de confianza, es el nivel de confianza que tenemos en que el intervalo contiene el valor del parámetro desconocido. Al valor $1-\alpha = 0{,}955$ se le llama coeficiente de confianza. El objetivo que se pretende con los intervalos de confianza es obtener un intervalo de poca amplitud y con una alta probabilidad de que el parámetro poblacional se encuentre en su interior. Así pues, elegiremos probabilidades cercanas a la unidad, que se representan por $1-\alpha$, y cuyos valores más frecuentes suelen ser: 0,90, 0,95, 0,99., que corresponden al 99%, 95% y 90%

La precisión de la estimación por intervalos vendrá caracterizada por el nivel de confianza y por la amplitud del intervalo:

- Para un coeficiente de confianza fijo, cuanto más pequeño sea el intervalo de confianza más precisa será la estimación.
- Para una misma amplitud de intervalo, cuanto mayor sea el coeficiente de confianza mayor será la precisión.

7.3 Intervalos de confianza en poblaciones normales.

- **Intervalo de confianza para la media.**

Una de las aplicaciones más corrientes de los intervalos de confianza es la de estimar la media poblacional.

- Un empresario quiere estimar el nivel medio de producción mensual, de sus industrias.

- Un gerente está interesado en conocer los ingresos anuales medios de los empleados sin cualificación y con un año de experiencia, que trabajan en empresas del mismo sector.
- Una Comunidad Autónoma quiere estimar el gasto medio mensual de las familias con más de dos hijos.

Hay un número casi infinito de situaciones que requieren una estimación de la media poblacional desconocida. Recordemos que un intervalo de confianza está comprendido entre un límite superior y un límite inferior. Si en cualquiera de los casos anteriores utilizamos como estimador la media muestral $\overline{X}$, tendremos que sumar una cantidad determinada para obtener el límite superior y restar la misma cifra para obtener el límite inferior. La pregunta que puede surgir es: ¿cuanto debemos sumar y restar?. La respuesta depende de la precisión que queramos obtener. Si deseamos construir un intervalo del 95 %. En este caso tenemos, que el coeficiente de confianza es $1-\alpha = 0{,}95$, $\alpha = 0{,}05$ y $\alpha/2 = 0{,}025$. Si utilizamos las tablas de la distribución normal y calculamos:

$$P\left(Z > z_{\alpha/2}\right) = 0{,}025$$

El valor de z que corresponde a un área de $0{,}025$ es $z_{\alpha/2} = 1{,}96$.

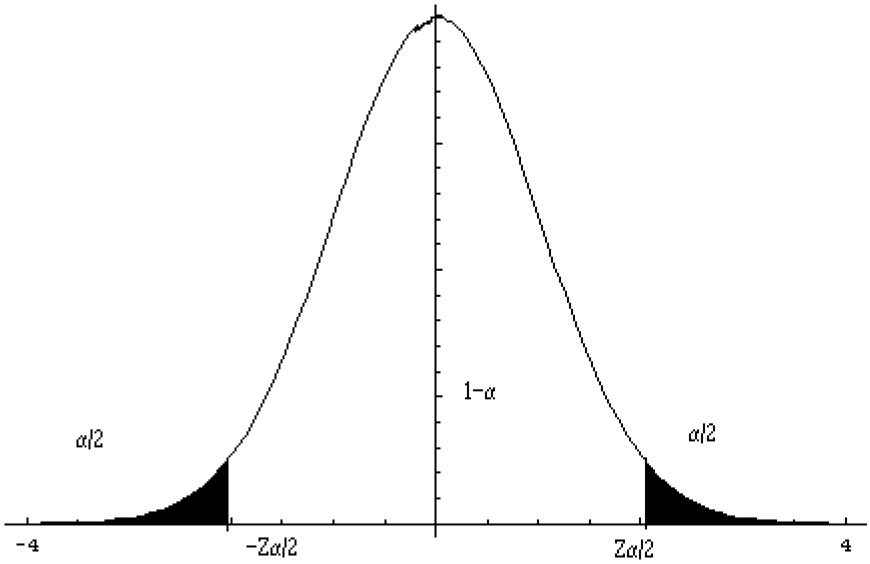
Es decir, para construir un intervalo de confianza del 95 % sólo tenemos que definir un intervalo que se extienda $1{,}96\,\delta$ por encima y por debajo de la media muestral $\overline{X}$.

Siendo δ conocida.

Supongamos que tenemos una muestra aleatoria de n observaciones procedente de una población. Si δ, la desviación típica de la población es finita conocida, y la media muestral observada es $\bar{x}$, entonces el intervalo de confianza para la media poblacional μ al nivel de confianza del $100\left(1-\alpha\right)$ viene dado por:

$$I_{\mu} = \left[\bar{x} - z_{\alpha/2}\frac{\delta}{\sqrt{n}}\,;\,\bar{x} + z_{\alpha/2}\frac{\delta}{\sqrt{n}}\right]$$

donde $z_{\alpha/2}$ es tal que:



$$P\left(Z > z_{\alpha/2}\right) = \frac{\alpha}{2}$$

y la variable aleatoria $Z \rightarrow N\,(0,1)$

Siendo δ desconocida.

El conocimiento de la desviación típica de la población puede parecer poco realista. La situación más probable es que la desviación típica de la población sea desconocida, bastará con poner en su lugar la desviación típica de la muestra.

Supongamos que tenemos una muestra aleatoria de n observaciones procedentes de una población. Si δ es desconocida, y la media y la desviación típica muestral observadas son $\bar{x}$ y s respectivamente, entonces el intervalo de confianza para la media poblacional μ al nivel de confianza del $100\,(1-\alpha)$ viene dado por:

$$I_{\mu} = \left[\bar{x} - z_{\alpha/2}\,\frac{s}{\sqrt{n}}\,;\,\bar{x} + z_{\alpha/2}\,\frac{s}{\sqrt{n}}\right]$$

donde $z_{\alpha/2}$ es tal que:

$$P\left(Z > z_{\alpha/2}\right) = \frac{\alpha}{2}$$

y la variable aleatoria $Z \rightarrow N\,(0,1)$

Estas técnicas de estimación para la media poblacional se pueden utilizar:

- Cuando la población sigue una distribución normal o bien cuando se supone que la distribución de la población es desconocida.
- Cuando se conoce δ o bien cuando se desconoce.

Pero en todos los casos hay una circunstancia común: siempre se debe elegir una muestra aleatoria simple suficientemente grande procedente de la población $(n \geq 30)$. En muchas ocasiones encontraremos que obtener una muestra grande es poco práctico e incluso imposible. Un ejemplo puede ser el de una compañía farmacéutica, que para probar un nuevo medicamento no puede encontrar a más de 30 personas que estén dispuestas a probarlo.
Cuando hay que tomar una muestra pequeña $(n < 30)$, la distribución normal no siempre es la adecuada, en este caso las distribuciones que se seguirán dependerá de si δ es conocida o desconocida.

Supongamos que tenemos una muestra aleatoria de n observaciones de una distribución $n(\mu,\delta)$. Si δ es desconocida, y la media y la desviación típica muestral observadas son $\bar{x}$ y s respectivamente, entonces el intervalo de confianza para la media poblacional μ al nivel de confianza del $100\,(1-\alpha)$ viene dado por:

$$I_{\mu} = \left[\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}}\, ;\, \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right]$$

donde $t_{\alpha/2}$ es tal que:

$$P\left(t_{n-1} > t_{\alpha/2} \right) = \frac{\alpha}{2}$$

y la variable aleatoria t_{n-1} sigue una distribución t-Student con $(n-1)$ grados de libertad.

- **Intervalo de confianza para la varianza.**

En ocasiones se requieren también estimaciones, mediante intervalos de confianaza, para la varianza de la población. Veamos como ejemplos:

- El director comercial de una industria conservera, está interesado en que el peso neto del producto incluido en el interior de la lata tenga poca variabilidad, para que no existan diferencias entre el peso real y el peso anunciado en la etiqueta.
- Un gerente está interesado en que exista poca variabilidad en los ingresos anuales de los empleados con un año de experiencia, que trabajan en su empresa.

Siendo la µ desconocida.

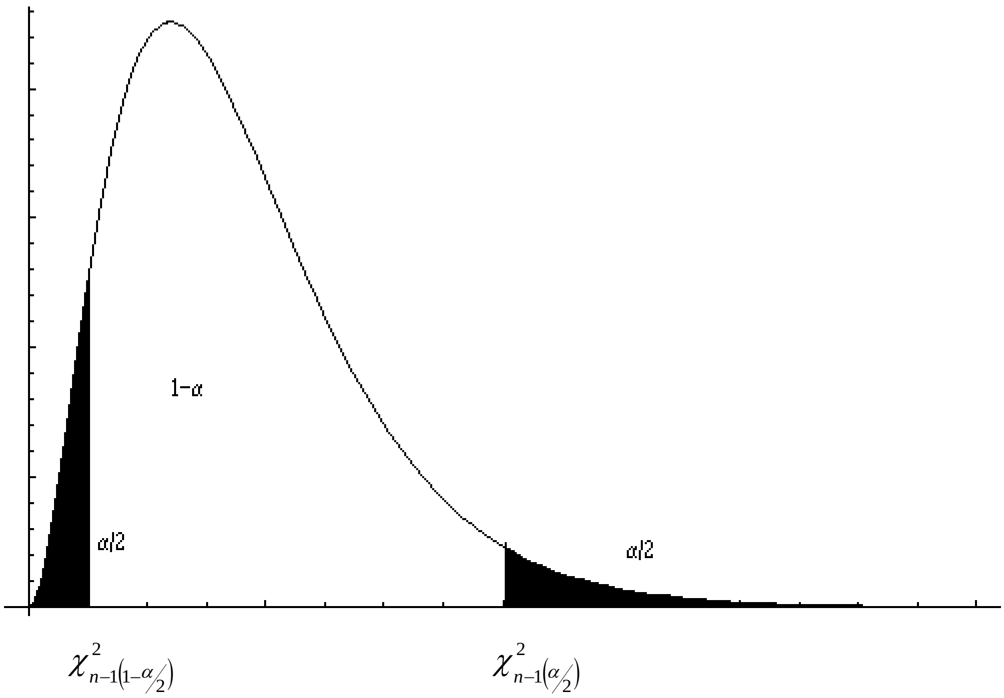
Supongamos que tenemos una muestra aleatoria de n observaciones de una población $N(\mu, \delta)$. Si δ es desconocida, y la varianza muestral observada es s^2.Usaremos un estadístico que dependa de σ^2, siendo su distribución no dependiente de la varianza.

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum\left(x_i - \bar{x}\right)}{\sigma^2} \rightarrow \chi^2_{n-1},$$

Entonces el intervalo de confianza para la varianza poblacional δ^2 al nivel de confianza del $100(1-\alpha)$ viene dado por:

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

donde $\chi^2_{n-1,1-\alpha/2}$ es tal que:



$$P\left(\chi^2_{n-1} \geq \chi^2_{n-1,\,1-\alpha/2}\right) = 1 - \frac{\alpha}{2}$$

y $\chi^2_{n-1,\,\alpha/2}$ es tal que:

$$P\left(\chi^2_{n-1} \geq \chi^2_{n-1,\,\alpha/2}\right) = \frac{\alpha}{2}$$

y la variable aleatoria χ^2_{n-1} sigue una distribución χ^2 de Pearson con n-1 grados de libertad.

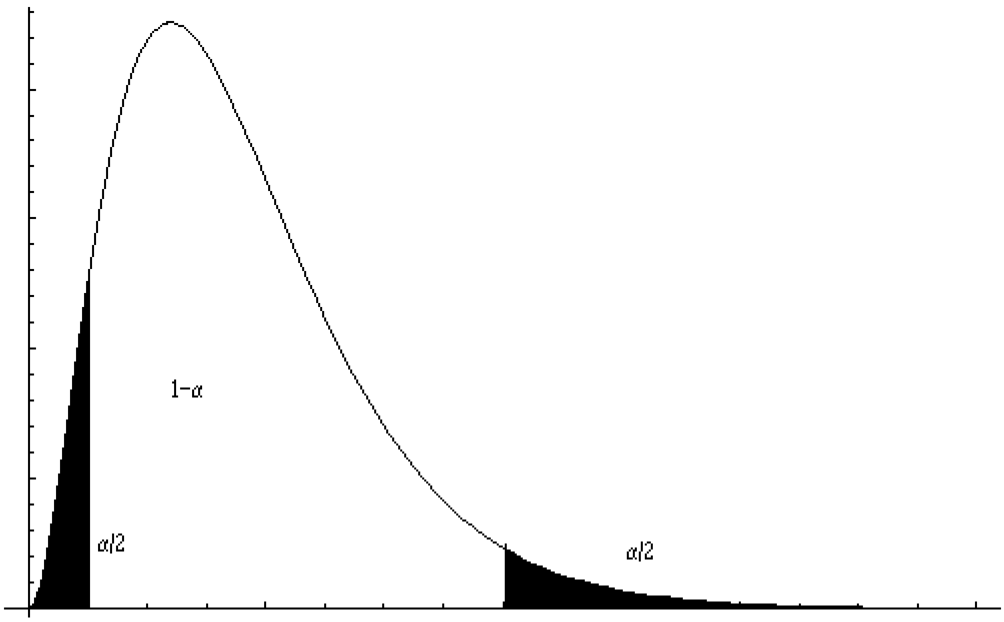
Siendo µ conocida.

El intervalo de confianza para la varianza de una población normal, suponiendo que μ es conocida puede obtenerse razonando análogamente al caso anterior. Al ser la media μ conocida no hay que estimarla pues:

$$\frac{\sum (x_i - \mu)^2}{\sigma^2} \rightarrow \chi^2_n$$

El intervalo de confianza para la varianza poblacional σ^2 al nivel de confianza del $100\,(1-\alpha)$ viene dado por:

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



$\chi^2_{n\left(1-\alpha/2\right)}$ $\chi^2_{n\left(\alpha/2\right)}$

7.4 Intervalos de confianza en poblaciones no necesariamente normales.

- **Utilizando la desigualdad de Chebychev.**

Para obtener un intervalo para la μ con varianza poblacional conocida**.**

$$E(\bar{X}) = \mu \qquad V(\bar{X}) = \frac{\sigma^2}{n}$$

$$P\left(|\bar{x}-\mu| \le k\right) \ge 1 - \frac{V(\bar{X})}{k^2} = 1 - \frac{\sigma^2}{nk^2}$$

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

Por lo tanto tenemos que:

$$P\left(|\bar{X}-\mu| \le \frac{\sigma}{\sqrt{n\alpha}}\right) \ge 1-\alpha$$

y el intervalo será el siguiente:

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n\alpha}}; \bar{X} + \frac{\sigma}{\sqrt{n\alpha}}\right]$$

7.5 Intervalos de confianza de una proporción.

Otra aplicación de los intervalos de confianza es la de estimar la proporción poblacional.

- Un empresario puede estar interesado en conocer el porcentaje de sus productos que son defectuosos, frente al porcentaje de productos no defectuosos.
- Un gerente puede estar interesado en conocer la proporción de clientes que pagan a crédito frente a quienes pagan al contado.

En cada uno de estos resultados solo hay dos situaciones posibles. Por lo tanto, lo que nos interesa es la proporción de respuestas que se clasifica en uno de estos dos resultados.

Situándonos en el marco de la distribución binomial, $\hat{p}_x$ la proporción de éxitos de una muestra aleatoria de n observaciones, si n es grande, entonces la variable aleatoria:

$$Z = \frac{\hat{p}_x - p}{\sqrt{\frac{p\,q}{n}}} \quad \rightarrow \quad N(0,1)$$

Este resultado no nos permite por sí solo el cálculo de intervalos de confianza para la proporción poblacional, ya que el denominador depende del parámetro p desconocido. Sin embargo, si el tamaño de la muestra es grande, podemos conseguir una buena aproximación sustituyendo p por su estimador puntual $\hat{p}_x$, es decir:

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

Sea una población $B(1,p)$, sea $\hat{p}_x$ la proporción de éxitos de una muestra aleatoria de n observaciones. Entonces, si n es grande, el intervalo de confianza para la proporción poblacional al nivel de confianza del $100\,(1-\alpha)$ viene dado por:

$$I_p = \left[\hat{p}_x - z_{\alpha/2} \sqrt{\frac{\hat{p}_x\,(1-\hat{p}_x)}{n}} \;;\qquad \hat{p}_x + z_{\alpha/2} \sqrt{\frac{\hat{p}_x\,(1-\hat{p}_x)}{n}} \right]$$

donde $z_{\alpha/2}$ es tal que:

$$P\left(Z > z_{\alpha/2} \right) = \frac{\alpha}{2}$$

y la variable aleatoria $Z \rightarrow N\,(0,1)$

Si tenemos una muestra pequeña, lo expresado hasta ahora posee una mayor dificultad. En este caso se utiliza un método gráfico, en él para cada muestra y en consecuencia para cada valor de $\hat{p}_x = \frac{x}{n}$ se tienen dos soluciones, que en el plano cartesiano vendrán expresadas por dos puntos uno inferior y otro superior constituyendo éstos los extremos del intervalo.

Los intervalos de confianza para la proporción poblacional están centrados en la proporción muestral. Se puede comprobar también que, siempre que las demás condiciones permanezcan intactas, cuanto mayor sea el tamaño muestral n, menor será la longitud del intervalo de confianza. Este hecho refleja el incremento de la precisión de la información sobre la proporción poblacional obtenida a medida que aumenta el tamaño muestral.

7.6 Estimación del tamaño muestral.

Hasta ahora hemos desarrollado métodos para construir intervalos de confianza para un parámetro poblacional, basándonos en la información contenida en una muestra determinada. Siguiendo este proceso, un investigador puede creer que el intervalo de confianza resultante es demasiado amplio, enfrentándose así a un grado de incertidumbre poco deseable. Normalmente la única forma de reducir esta incertidumbre consiste en tomar una muestra con un tamaño mayor.

En algunas situaciones el investigador puede ser capaz de fijar por adelantado la amplitud del intervalo de confianza, eligiendo un tamaño muestral lo suficientemente grande como para garantizar dicha amplitud.

- **Tamaño de la muestra para estimar la media de una población normal y varianza poblacional conocida.**

En este caso sabíamos que el intervalo de confianza era:

$$I_{\mu}=\left[\bar{x}-z_{\alpha/2}\frac{\delta}{\sqrt{n}}\,;\,\bar{x}+z_{\alpha/2}\frac{\delta}{\sqrt{n}}\right]$$

Siendo la amplitud del intervalo:

$$L=\left(\bar{x}+z_{\alpha/2}\frac{\delta}{\sqrt{n}}\right)-\left(\bar{x}-z_{\alpha/2}\frac{\delta}{\sqrt{n}}\right)=2\,z_{\alpha/2}\frac{\delta}{\sqrt{n}}$$

Si previamente, se fija la longitud del intervalo L y deseamos conocer el tamaño de la muestra para obtener ese intervalo al nivel de confianza del $100(1-\alpha)\%$, bastará despejar n de la expresión.

Supongamos que disponemos de una muestra aleatoria procedente de una población normal con varianza conocida δ^2. Entonces, un intervalo de confianza del $100\,(1-\alpha)\,\%$ para la media poblacional tiene una amplitud L, si el número de observaciones es:

$$n = 4\ \frac{z^2_{\alpha/2}\ \delta^2}{L^2}$$

donde $z_{\alpha/2}$ es tal que:

$$P\left(Z > z_{\alpha/2}\right) = \frac{\alpha}{2}$$

y la variable aleatoria $Z \to N\,(0{,}1)$

Naturalmente, el número de observaciones de la muestra tiene que ser entero. Si el número n resultante no es entero, lo redondearemos por exceso para garantizar que nuestro intervalo de confianza no exceda la amplitud requerida.

También podemos hacer el siguiente razonamiento, si la media μ fuera el valor central del intervalo que hemos construido, entonces $\bar{x}$ estimaría puntualmente a μ sin error alguno. Pero generalmente $\bar{x}$ no será exactamente igual a μ y entonces se comete un error, $e = \left|\bar{x} - \mu\right|$ y como máximo será:

$$e = z_{\alpha/2}\ \frac{\delta}{\sqrt{n}}$$

Entonces si queremos determinar el tamaño de la muestra necesario para obtener un intervalo de confianza para la media poblacional μ, admitiendo un error e, despejando de la expresión anterior tenemos:

$$n = \frac{z^2_{\alpha/2}\ \delta^2}{e^2}$$

- **Tamaño de la muestra para estimar la media de una población normal y varianza poblacional desconocida**.

El intervalo para la media de una población normal con δ desconocida, viene dado por:

$$I_{\mu} = \left[\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} \; ; \; \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right]$$

Siendo la amplitud del intervalo:

$$L = \left(\bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right) - \left(\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} \right) = 2\, t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Si previamente, se fija la longitud del intervalo L y deseamos conocer el tamaño de la muestra para obtener ese intervalo al nivel de confianza del $100\,(1-\alpha)\,\%$, bastará despejar n de la expresión.

Supongamos que disponemos de una muestra aleatoria procedente de una población normal con varianza desconocida δ^2 . Entonces, un intervalo de confianza del $100\,(1-\alpha)\,\%$ para la media poblacional tiene una amplitud, si el número de observaciones es:

$$n = 4 \, \frac{t_{\alpha/2}^{2} \, s^{2}}{L^{2}}$$

donde $t_{\alpha/2}$ es tal que:

$$P\left(t_{n-1} > z_{\alpha/2} \right) = \frac{\alpha}{2}$$

y la variable aleatoria t_{n-1} sigue una distribución t-Student con $(n-1)$ grados de libertad.

Si en lugar de utilizar la amplitud L del intervalo utilizamos el error, entonces el tamaño de la muestra será:

$$n = \frac{t_{\alpha/2}^{2} \, s^{2}}{e^{2}}$$

- **Tamaño de la muestra para estimar la proporción de una población.**

Sabemos que el intervalo al nivel de confianza del $100\,(1-\alpha)\,\%$ para la proporción poblacional es:

$$I_p = \left[\hat{p}_x - z_{\alpha/2} \sqrt{\frac{\hat{p}_x (1-\hat{p}_x)}{n}} \;;\quad \hat{p}_x + z_{\alpha/2} \sqrt{\frac{\hat{p}_x (1-\hat{p}_x)}{n}} \right]$$

La longitud del intervalo es:

$$L = 2\, z_{\alpha/2} \sqrt{\frac{\hat{p}_x (1-\hat{p}_x)}{n}}$$

y despejando de esta expresión el valor de n tendremos:

$$n = 4\, \frac{z_{\alpha/2}^2\, \hat{p}_x (1-\hat{p}_x)}{L^2}$$

De forma análoga a los dos casos anteriores si utilizamos el error $e = \left| \hat{p}_x - p \right|$, el cual como máximo será:

$$n = \frac{z_{\alpha/2}^2\, \hat{p}_x (1-\hat{p}_x)}{e^2}$$

7.7 Resumen y preguntas frecuentes:

- Objetivo de la estimación por intervalos de confianza.
- ¿ De qué depende la precisión en la estimación por intervalos de confianza?.
- ¿Qué métodos se pueden utilizar para la estimación por intervalos de confianza?¿Cuál de ellos es preferible utilizar?.Razone la respuesta.
- Supongamos una muestra aleatoria simple $(X_1, \ldots, X_{20})$, de una distribución $N(\mu,\ 100)$, la media muestral observada es $\overline{X} = 1000$.¿Cuál será el intervalo de confianza para la media poblacional, con un nivel de confianza del 90%?.
- ¿Qué efecto tiene en un intervalo de confianza una disminución en el tamaño de la muestra?¿ y una disminución en el nivel de confianza?.
- Supongamos una muestra aleatoria simple $(X_1, \ldots, X_{20})$, de una distribución $N(\mu,\ \sigma)$, la media muestral observada es $\overline{X} = 1000$ y la correspondiente varianza muestral es $s = 200$.¿Cuál será el intervalo de confianza para la media poblacional, con un nivel de confianza del 90%?.¿Cuál será el intervalo de confianza para la varianza poblacional, con un nivel de confianza del 90%?.
- ¿Cuándo podemos decir que se comete un error en la estimación? Razone la

respuesta.

- ¿Cuál será el tamaño de la muestra para estimar la media μ de una población normal, con desviación típica σ desconocida, si previamente fijamos la longitud del intervalo de confianza en un valor L dado?.