

T.4 ESTIMACIÓN POR INTERVALO

1. Introducción

2. Construcción de intervalos

2.1. Intervalo de confianza para la media de una población cualquiera, conocida la varianza.

ejemplo 1

2.2. Intervalo de confianza para la media de una población normal con varianza conocida.

ejemplo 2

2.3. Intervalo de Confianza para la Media de una Población Normal de Varianza desconocida (muestras pequeñas) . *ejemplo 3*

2.4. Intervalo de confianza para la diferencia de medias con poblaciones normales y varianzas conocidas

2.5. Intervalo de Confianza para la proporción de una característica. *ejemplo 4*

2.6. Intervalo de confianza para el cociente de varianzas en poblaciones normales.

2.7. Otros intervalos de interés.

2.8. Esquema de actuación para los intervalos para la media

3. Cálculo del Tamaño muestral para obtener un error de estimación prefijado. *ejemplo 5*

1. Introducción

La "estimación por intervalo" consiste en determinar un par de valores a y b , tales que constituidos en intervalo $[a, b]$; y para una probabilidad $1-\alpha$ prefijada (nivel de confianza) se verifique en relación al parámetro θ a estimar se cumpla:

$$P\{\theta \in [a, b]\} = 1 - \alpha \quad \text{ó en otros términos} \quad P(a \leq \theta \leq b) = 1 - \alpha$$

Podemos considerar el **nivel de confianza** ($1-\alpha$) que hemos prefijado para la expresión anterior como la probabilidad que existe (antes de tomar la muestra) de que el intervalo a construir a partir de la muestra incluya el verdadero valor del parámetro a estimar. Refleja la "confianza" en la "construcción" del intervalo y de que éste tras concretar la muestra contendrá el valor a estimar. De ahí que en términos numéricos dicho nivel o probabilidad haya de tomar un valor alto (0.9,0.95,0.99).

Evidentemente el complementario al nivel de confianza; es decir α , nivel de **significación** supondrá las probabilidades de cometer el error de no dar por incluido el

verdadero valor del parámetro a estimar en un intervalo en el que realmente si está. De ahí y dado que se trata de un error posible a cometer, su cuantificación en términos de probabilidad sea muy pequeña (0.1,0.05,0.005,..).

En relación a lo anterior. Obviamente, cuanto mayor sea el nivel de confianza prefijado la amplitud del intervalo de estimación será también mayor y por tanto la estimación será menos *precisa*.

Existen para cualquier distribución una infinidad de intervalos a los cuales les corresponde la misma probabilidad y por tanto habrá una infinidad de intervalos, IN, que verifiquen que

$P(\theta \in IN) = 1 - \alpha$ lógicamente nosotros buscamos una estimación lo más precisa posible; es decir ,de todos los intervalos que verifican la anterior expresión $P(\theta \in IN) = 1 - \alpha$ el de menor amplitud .En este sentido, es sencillo ver que si la distribución es simétrica y unimodal ,de todos los intervalos isoprobables, el de menor amplitud (que coincidirá con el de mayor densidad media de probabilidad) es el intervalo centrado en la media .De acuerdo con esto ,si la distribución que consideramos es simétrica la determinación del intervalo de estimación es relativamente sencilla.

La construcción de intervalos específicos depende de las características de la población (normal o no, etc.), de los parámetros o combinaciones de parámetros a los que se les construye (media, varianza, proporción, coeficiente de correlación, diferencias de medias,...), tamaño muestral y parámetros poblacionales conocidos. De ello se deduce que según dichas circunstancias la construcción de intervalos variará, si bien es cierto que el patrón de trabajo para su construcción permanece invariable.

2.- Construcción de intervalos

2.1. Intervalo de confianza para la media de una población cualquiera, conocida la varianza.

Las circunstancias específicas para la construcción de este intervalo son las siguientes:

Intervalo para μ

Conocida σ (o la varianza)

Distribución poblacional desconocida.

Nivel de confianza dado $1-\alpha$

Tamaño muestral desconocido luego nos colocamos en el peor de los casos, es decir pequeño. Partiendo del conocido teorema de Markov:

$$P(g(x) \leq c) \geq 1 - \frac{E[g(x)]}{c}$$

donde $g(x)$ es una función cualquiera de la variable aleatoria x , y dicha función g está definida NO negativa, siendo c una constante cualquiera. Así:

$$\text{definiendo } g(x) = \frac{(\bar{x} - \mu)^2}{\frac{\sigma^2}{n}}$$

$g(x)$ es, evidentemente, no negativa y tomando $c = H^2$

tendremos en aplicación de Markov :

$$P\left(\frac{(\bar{x} - \mu)^2}{\frac{\sigma^2}{n}} \leq H^2\right) \geq 1 - \frac{E\left[\frac{(\bar{x} - \mu)^2}{\frac{\sigma^2}{n}}\right]}{H^2}$$

dado que :

$$E[\bar{x}] = \mu$$

$$D^2[\bar{x}] = \frac{\sigma^2}{n}$$

tendremos que:

$$E\left[\frac{(\bar{x} - \mu)^2}{\frac{\sigma^2}{n}}\right] = \frac{E[(\bar{x} - \mu)^2]}{\frac{\sigma^2}{n}} = \frac{D^2[\bar{x}]}{\frac{\sigma^2}{n}} = \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n}} = 1$$

transponiendo este resultado al enunciado general :

$$P\left(\frac{(\bar{x} - \mu)^2}{\frac{\sigma^2}{n}} \leq H^2\right) \geq 1 - \frac{1}{H^2}$$

y tomando la raíz cuadrada tendremos

$$P\left(-H \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq H\right) \geq 1 - \frac{1}{H^2}$$

Y despejando para centrar el parámetro a estimar μ :

$$P\left(\bar{x} - \frac{\sigma}{\sqrt{n}} H \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} H\right) \geq 1 - \frac{1}{H^2}$$

si queremos establecer un nivel de confianza $1-\alpha$ igualaremos éste a

$$1 - \frac{1}{H^2} \quad \text{de manera que}$$

$$H = \frac{1}{\sqrt{\alpha}}$$

por lo que en función del nivel de confianza el intervalo quedaría :

$$P\left(\bar{x} - \frac{\sigma}{\sqrt{n \alpha}} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n \alpha}}\right) \geq 1 - \alpha \quad \text{o bien :}$$

$$\mu \in \left[\bar{x} - \frac{\sigma}{\sqrt{n \alpha}}; \bar{x} + \frac{\sigma}{\sqrt{n \alpha}} \right] \quad \text{con más de } 1-\alpha \text{ de confianza}$$

ejemplo 1

En población cuya distribución se desconoce se obtiene una muestra (m.a.s.) de 2000 valores de la que resulta una media de 225 y una desviación típica de 10. Suponiendo que la varianza muestral coincide con la poblacional, estimar un intervalo para la media de la población con un nivel de confianza del 95%.

Tendríamos $1-\alpha=0.95$ luego $\alpha=0.05$; $S=10=\sigma$; $n=2000$;
con $\bar{x} = 225$

aplicando

$$\mu \in \left[\bar{x} - \frac{\sigma}{\sqrt{n \alpha}}; \bar{x} + \frac{\sigma}{\sqrt{n \alpha}} \right] \Rightarrow \mu \in [224; 226]$$

con más de 0.95 de confianza.

2.2-Intervalo de confianza para la media de una población normal con varianza conocida.

Las circunstancias específicas para la construcción de este intervalo son las siguientes:

Intervalo para μ
 Conocida σ (o la varianza)
 Distribución poblacional **normal**.
 Nivel de confianza dado $1-\alpha$

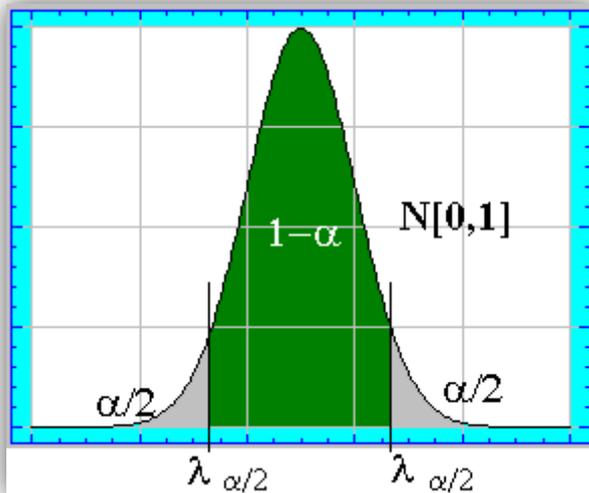
Tamaño muestral desconocido luego nos colocamos en el peor de los casos, es decir pequeño.

Conocemos que la media muestral se distribuye

$$\bar{x} \Rightarrow N\left[\mu, \frac{\sigma}{\sqrt{n}}\right]$$

luego tipificando

$$L\left(\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}\right) \Rightarrow N[0,1]$$



Como se ha comentado, al ser la normal reducida una distribución simétrica y unimodal, el intervalo de menor amplitud y de probabilidad

$1 - \alpha$ será el intervalo centrado en la media, es decir:

el intervalo $\left[-\lambda_{\alpha/2}, \lambda_{\alpha/2}\right]$.

Donde $\lambda_{\alpha/2}$ es el valor de la tabla de la $N[0;1]$ que haga que

$$P(\lambda_{\alpha/2}) = 1 - \frac{\alpha}{2}$$

Es decir el valor de la normal reducida que deje a su derecha una cola de probabilidad de $\alpha/2$. Así el valor $\lambda_{\alpha/2}$ será el valor simétrico de $-\lambda_{\alpha/2}$ (con signo negativo) y dejará a su izquierda una cola de $\alpha/2$. De esta forma entre, $-\lambda_{\alpha/2}$ y $\lambda_{\alpha/2}$ queda encerrada una probabilidad de $1 - \alpha$: sería así el intervalo de menor amplitud:

$$P\left(-\lambda_{\alpha/2} \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \lambda_{\alpha/2}\right) = 1 - \alpha$$

despejando la media poblacional tendríamos:

$$P\left(\bar{x} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

nota: si el muestreo fuera sin reposición, por lo tanto no m.a.s. y si irrestricto; el intervalo se vería afectado por el factor corrector de poblaciones finitas. Resultando, y sirva este ejemplo para comprobar lo que ocurriría en otros intervalos de otras características, de la siguiente forma :

$$P\left(\bar{x} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{x} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}\right) = 1 - \alpha$$

siendo N el tamaño de la población.

nota: Aunque no se conozca σ (desviación poblacional) si la muestra es bastante grande, $n > 30$ es habitual considerar la desviación típica muestral, S, como si fuera la poblacional y aplicar el intervalo de estimación obtenido arriba.

ejemplo 2

Realizar la estimación de μ del ejemplo 1 considerando ahora que la población es normal

Tendríamos $1-\alpha = 0.95$ luego $\alpha = 0.05$; $S=10=\sigma$ (muestra grande $n>30$); $n=2000$;
 $\bar{x} = 225$; población normal.

Aplicando el intervalo anterior:

$$P\left(\bar{x} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

el resultado sería : $\mu \in [224'56 , 225'44]$ con el 95 % de confianza.

2.3.-Intervalo de Confianza para la Media de una Población Normal de Varianza desconocida (muestras pequeñas)

Las circunstancias específicas para la construcción de este intervalo son las siguientes:

Intervalo para μ

desconocida σ (o la varianza) dado que n es pequeña no podemos tomar S como σ

Distribución poblacional **normal**.

Nivel de confianza dado $1-\alpha$

Tamaño muestral desconocido luego nos colocamos en el peor de los casos, es decir pequeño.

Del estudio de las distribuciones muestrales conocemos que:

$$L\left(\frac{\bar{x} - \mu}{S} \sqrt{n-1}\right) \Rightarrow t_{n-1}$$

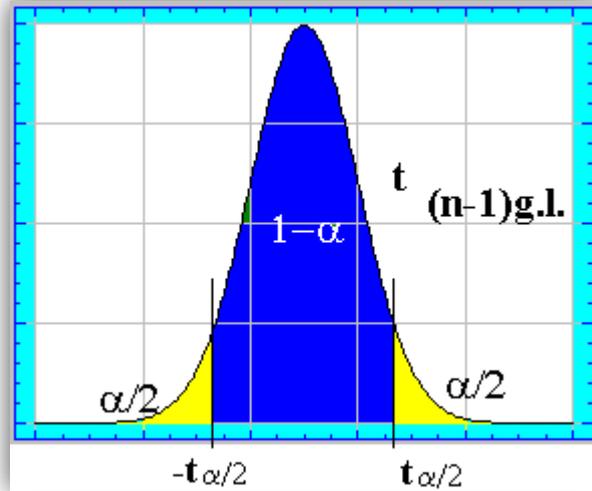
Como la distribución de t de student es una distribución simétrica, unimodal y centrada en 0, de todos los intervalos que verifiquen que:

$P(t \in IN) = 1 - \alpha$ el de menor amplitud será:

$$IN \in \left[-t_{\alpha/2}; t_{\alpha/2} \right]$$

Donde $t_{\alpha/2}$ es el correspondiente valor de la variable t para n - 1 grados de libertad y nivel de significación α . Dicho intervalo quedaría:

$$P \left(-t_{\alpha/2} \leq \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n-1}}} \leq t_{\alpha/2} \right) = 1 - \alpha$$



para un nivel de confianza prefijado y despejando tendríamos como intervalo

$$P \left(\bar{x} - t_{\alpha/2} \frac{S}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{\alpha/2} \frac{S}{\sqrt{n-1}} \right) = 1 - \alpha$$

recuérdese que en el caso de muestreo irrestricto sería de aplicación el antes nombrado factor corrector de poblaciones finitas.

ejemplo 3

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)

conocemos que según la información que poseemos, estamos ante:

normal; n=10 (muestra pequeña); S=4(poblacional desconocida); media muestral=100 ;

$1-\alpha = 0.95$ luego $\alpha = 0.05$ con lo que

$$t_{\alpha/2(9 \text{ g.l.})} = 2,26 \quad \text{(ir a tabla de la t)}$$

y dado que el intervalo a utilizar (nada dicen de irrestricto ; luego m.a.s) será:

$$P\left(\bar{x} - t_{\alpha/2} \frac{S}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{\alpha/2} \frac{S}{\sqrt{n-1}}\right) = 1 - \alpha$$

resultando : $\mu \in [96'99;103'01]$ con el 95 % de confianza

2.4-Intervalo de confianza para la diferencia de medias con poblaciones normales y varianzas conocidas.

Las circunstancias específicas para la construcción de este intervalo son las siguientes:

Intervalo para $\mu_x - \mu_y$

Conocidas σ_x, σ_y ó desconocidas con tamaños muestrales grandes

Distribuciones poblacionales **normales**.

Nivel de confianza dado $1-\alpha$

Tamaños muestrales n_x, n_y .

Conocemos que

$$L(\bar{x} - \bar{y}) \rightarrow N\left[\mu_x - \mu_y; \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right]$$

luego la tipificando tendremos :

$$\frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \rightarrow N[0,1]$$

para un nivel de confianza dado $1-\alpha$ que genera los valores extremos $\pm \lambda_{\alpha/2}$ de un intervalo centrado en dicha $N[0 ; 1]$; como ya vimos tendríamos :

$$P \left(-\lambda_{\alpha/2} \leq \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \leq \lambda_{\alpha/2} \right) = 1 - \alpha$$

despejando:

$$P \left((\bar{x} - \bar{y}) - \lambda_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \leq (\mu_x - \mu_y) \leq (\bar{x} - \bar{y}) + \lambda_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right) = 1 - \alpha$$

Evidentemente y como en los casos anteriores se ha supuesto m.a.s., en caso de irrestricto se aplicaría el f.c.p.f..

Si la varianza no fuera conocida pero la muestra fuera grande (superior a treinta, cada una) tomaríamos como varianza poblacional su homónima muestral

2.5.-Intervalo de Confianza para la proporción de una característica.

Deseamos estimar la proporción con la que se da una característica en una determinada población, esta característica es dicotómica por lo que o bien se posee o bien no . El intervalo se plantea, como todos con un nivel de confianza $1-\alpha$ prefijado. Realizando, claro está, un muestreo de tamaño n , que en principio consideramos aleatorio simple.

Del estudio de las distribuciones conocemos el comportamiento de la proporción con la que una característica \hat{p} se da en la muestra:

$$L(\hat{p}) \Rightarrow N \left[p, \sqrt{\frac{pq}{n}} \right]$$

siendo p la proporción con la que la característica aparece en la población ;
"q" lógicamente su complementario y n el tamaño muestral .

Tipificando obtendríamos:

$$\frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \Rightarrow N[0,1]$$

Como en los casos anteriores y dado que utilizamos la $N[0,1]$; para un nivel de confianza dado $1-\alpha$, los valores de dicha normal que generan un intervalo centrado corresponderían a $\pm \lambda_{\alpha/2}$ por lo que quedaría el siguiente intervalo:

$$P\left(-\lambda_{\frac{1}{2}} \leq \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \leq \lambda_{\frac{1}{2}}\right) = 1 - \alpha$$

despejando la proporción poblacional :

$$P\left(\hat{p} - \lambda_{\frac{1}{2}} \frac{\sqrt{pq}}{\sqrt{n}} \leq p \leq \hat{p} + \lambda_{\frac{1}{2}} \frac{\sqrt{pq}}{\sqrt{n}}\right) = 1 - \alpha$$

que es intervalo que estábamos buscando.

En el caso de que el muestreo fuera irrestricto su expresión sería, tras aplicar el f.c.p.f., la siguiente:

$$P\left(\hat{p} - \lambda_{\frac{1}{2}} \frac{\sqrt{pq}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \leq p \leq \hat{p} + \lambda_{\frac{1}{2}} \frac{\sqrt{pq}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}\right) = 1 - \alpha$$

Como se ha podido comprobar en ambas expresiones de intervalo está contenida y por tanto es necesaria para su construcción la proporción poblacional "p". Parece ilógico que si queremos estimar dicha proporción la conozcamos y por tanto la podamos utilizar. El problema puede resolverse de varias formas según el caso:

- Si el tamaño muestral es muy grande podríamos utilizar como proporciones poblacionales las obtenidas por la muestra; es decir tomar $\hat{p}\hat{q}$ en lugar de pq . Evidentemente si el intervalo se pre-construye antes de que se realice la muestra (por ejemplo para calcular el tamaño muestral necesario para un determinado error, como veremos después) éste no será el método aconsejable.
- En el caso de poseer información "fiable" de la proporción de la población, bien por estimaciones anteriores recientes, bien por pre-muestreos (piloto) , podríamos utilizar dichas estimaciones como los valores reales de p y por tanto de $q=1-p$.
- Lo más habitual, para resolver este problema, consiste en "colocarse" en la situación más desfavorable posible en cuanto a los valores de dicha proporción poblacional. Dicha situación no es otra que valorar $p=q=0.5$, con lo que se consigue establecer como cierta la máxima varianza en la población (es fácil demostrar que dadas las características de p y q el máximo de $p \cdot q$ se obtiene precisamente con $p=q=0.5$). Adoptando esta medida, en la construcción del intervalo, se consigue que la amplitud de éste sea mayor, menos precisa, y por tanto para una misma amplitud necesitará un mayor tamaño de la muestra que el realizado con otra p , pero resuelve el problema del desconocimiento de p y augura, al darse mayor amplitud, más "fiabilidad". (obsérvese que en casi-todas las encuestas de opinión la ficha técnica incluye el mencionado $p=q=0.5$).

ejemplo 4

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 %.

conocemos:

$$1-\alpha = 0.95; n=100 \text{ (grande) proporción muestral} = \hat{p} = 0.25$$

el intervalo sería :

$$P\left(\hat{p} - \lambda_{\alpha/2} \frac{\sqrt{pq}}{\sqrt{n}} \leq p \leq \hat{p} + \lambda_{\alpha/2} \frac{\sqrt{pq}}{\sqrt{n}}\right) = 1 - \alpha$$

donde el valor de $\pm\lambda_{\alpha/2} = 1.96$ según tabla $N[0,1]$ (ir a tabla de la normal) y 0.95 de confianza.

Desconocemos la proporción poblacional p ; dos opciones

- $p = \hat{p} = 0.25$ dado que la muestra es grande ; que aplicada en el intervalo daría que : la proporción de penetración en el mercado está entre el 16'51 % y el 33'48 % con una confianza del 95 % (ir a script de realización)
- $p=q=0.5$ poniéndonos en el caso de varianza máxima; en el caso por tanto más desfavorable. En este caso la proporción de penetración en el mercado estaría entre el 15.2% y el 34.8 con una confianza del 95%; como se puede apreciar el intervalo tiene más holgura que el realizado por el método anterior. (ir a script de realización)

2.6-Intervalo de confianza para el cociente de varianzas en poblaciones normales.

Las circunstancias específicas para la construcción de este intervalo son las siguientes:

Intervalo para el cociente de dos varianzas poblacionales $\frac{\sigma_1^2}{\sigma_2^2}$

Realizamos dos muestras aleatorias simples de tamaños n y m

De dichas muestras se extraen las varianzas muestrales S_1^2, S_2^2

Conocemos que las poblaciones 1 y 2 para cuya razón de varianzas queremos construir el intervalo son Normales.

Lógicamente prefijamos un nivel de confianza de $1-\alpha$.

Conocemos según vimos cuando estudiamos las distribuciones muestrales de normales que:

por el lema de Fisher - Cochran :

$$\frac{n S^2}{\sigma^2} \Rightarrow \chi_{n-1}^2$$

para las dos poblaciones que tenemos se dará .

$$\frac{n S_1^2}{\sigma_1^2} \Rightarrow \chi_{n-1}^2 \quad \frac{m S_2^2}{\sigma_2^2} \Rightarrow \chi_{m-1}^2$$

nótese que el tamaño muestral de la muestra de 1ª es n y la de 2ª es m

Dado que nos interesa la razón de varianzas y en aplicación de la expresión de la F de Snedecor:

$$\frac{\frac{n S_1^2}{\sigma_1^2}}{\frac{m S_2^2}{\sigma_2^2}} \Rightarrow F(n-1)(m-1)$$

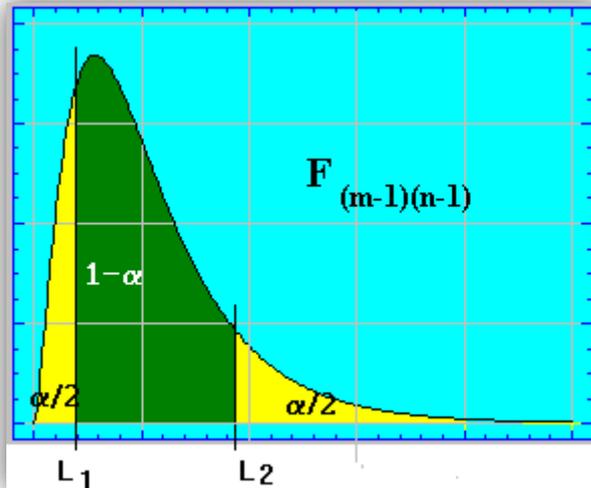
arreglando el cociente

$$\frac{n S_1^2 \sigma_2^2 (m-1)}{m S_2^2 \sigma_1^2 (n-1)}$$

dado que nos interesa el cociente inverso $\frac{\sigma_1^2}{\sigma_2^2}$ tendríamos que:

$$\frac{m S_2^2 \sigma_1^2 (n-1)}{n S_1^2 \sigma_2^2 (m-1)} \Rightarrow F(m-1)(n-1) gl$$

donde esta incluida la razón de varianzas para la cual queremos crear un intervalo .



Si hemos establecido un nivel de confianza de $1-\alpha$ el intervalo para una $F(m-1)(n-1)$ vendría dado por las constantes L_1 y L_2 , adoptándose el criterio simplificador, próximo al de mayor longitud, de considerar:

$$P(L_1 \leq F(m-1)(n-1) \leq L_2) = 1 - \alpha$$

de manera que

$$P(F(m-1)(n-1) \geq L_1) = 1 - \alpha + \frac{\alpha}{2} = 1 - \frac{\alpha}{2}$$

y por otra parte:

$$P(F(m-1)(n-1) \geq L_2) = \frac{\alpha}{2}$$

una vez determinados los valores de L_1 y L_2 tendríamos el intervalo

$$P\left(L_1 \leq \frac{m S_2^2 \sigma_1^2 (n-1)}{n S_1^2 \sigma_2^2 (m-1)} \leq L_2\right) = 1 - \alpha$$

despejando la razón que nos interesa tendríamos :

$$P\left(L_1 \frac{n(m-1) S_1^2}{m(n-1) S_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq L_2 \frac{n(m-1) S_1^2}{m(n-1) S_2^2}\right) = 1 - \alpha$$

Evidentemente si planteamos medir la diferencia entre las varianzas, cuanto más próximo se la razón a la unidad menor diferencia habrá entre las varianzas y lógicamente cuando la razón entre estas difiera mucho de 1, la diferencia entre varianzas será mas ostensible.

2.7-Otros intervalos de interés.

Evidentemente es posible la realización o construcción de otros intervalos que pueden revestir interés. Dado que algunos son consecuencias de los ya estudiados o bien, otros,

en los que su utilidad es muy ocasional, no los analizamos en profundizar. No obstante cabe destacar:

- -Intervalo para la diferencia de proporciones. Su formulación es idéntica a la del de diferencia de medias teniendo la precaución de considerar como media muestral a la proporción muestral y como desviación al producto de p por q .
- -Intervalo para la diferencia de medias con misma varianza poblacional. Evidentemente si bien tiene una formulación propia, su resolución solo requiere de la utilización del antes mencionado de diferencia de medias con la precaución de que ambas desviaciones que se utilizan son lógicamente iguales.
- -Intervalo para diferencia de medias con varianzas desconocidas y muestras pequeñas. En este caso su construcción acarrearía la utilización de la t de student cómo en el caso del de la media con varianza desconocida. Su utilidad es mínima pues pensemos que entre ambos valores muestrales han de sumar menos de, aproximadamente, 30 lo que parece que sea un valor fácilmente superable para cualquier muestreo serio. Si la muestra o muestras fueran grandes (más de 30; lo habitual) el intervalo pasa a ser el de varianza conocida que es, evidentemente, el más utilizado y por ello lo hemos desarrollado.

2.8. Esquema de actuación para los intervalos para la media

A modo de resumen se plantean los diversos intervalos para la media de la población según las características que se den para su construcción: así

tipo de población	¿ varianza conocida ?	tamaño muestral	intervalo con <i>m.a.s.</i>
normal	sí	cualquiera	$P\left(\bar{x} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$
		n > 30	$P\left(\bar{x} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$ tomar como σ ; S
	no	n < 30	$P\left(\bar{x} - t_{\alpha/2} \frac{S}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{\alpha/2} \frac{S}{\sqrt{n-1}}\right) = 1 - \alpha$
		n > 20	$P\left(\bar{x} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$
no normal	sí	n < 20	$P\left(\bar{x} - \frac{\sigma}{\sqrt{n\alpha}} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n\alpha}}\right) \geq 1 - \alpha$
		n > 50	$P\left(\bar{x} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$ tomar como σ ; S
	no	n < 50	sin solución

3. Cálculo del Tamaño muestral para obtener un error de estimación prefijado

Dado un nivel de confianza $1 - \alpha$, el intervalo de confianza para la media de una distribución normal con varianza conocida es:

$$P\left(\bar{x} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Podemos establecer el error "e" como la amplitud del intervalo es decir:

$$e = 2\left(\left|\bar{x} - \mu\right|\right) = 2\left(\lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

determinar, por tanto, el tamaño muestral consistirá en:
una vez establecido el error que se quiere cometer, despejar "n" de la expresión que anteriormente hemos establecido :

$$\sqrt{n} = \frac{2 \lambda_{\alpha/2} \sigma}{e} \rightarrow n = \frac{4 \lambda_{\alpha/2}^2 \sigma^2}{e^2}$$

Hemos indicado que el error e, es la amplitud del intervalo; en muchos manuales se especifica como error= $\pm e$ el semi-intervalo ó la semi-holgura. Que no es más que lo mismo, pues dos veces la semi-holgura es, evidentemente, la holgura total, por tanto su amplitud. En otros se trata al error e como simplemente la semi-holgura; la semi-amplitud. Lógicamente todos los planteamientos son correctos; solo hemos de poner cuidado a la hora de concluir la expresión correcta del tamaño muestral.

Así: un error de $e = \pm 4$ supone un error=amplitud del intervalo=8 y el error medido como semi-amplitud sería de 4.

Para todos los intervalos que hemos visto es posible el cálculo del tamaño muestral si prefija el error admisible; no vamos a realizar el desarrollo de todos. Nos centramos por su importancia en el tamaño muestral necesario para la construcción de un intervalo de confianza para la proporción de una característica con un determinado error prefijado, así

Si establecemos construir un intervalo de confianza para la p de una característica con un determinado nivel de confianza $1-\alpha$; y un determinado error "e" de un tanto por uno; y con muestreo aleatorio simple. Tendremos que:

dado que el intervalo es

$$P\left(\hat{p} - \lambda_{\alpha/2} \frac{\sqrt{pq}}{\sqrt{n}} \leq p \leq \hat{p} + \lambda_{\alpha/2} \frac{\sqrt{pq}}{\sqrt{n}}\right) = 1 - \alpha$$

y dado que el error es e =amplitud del intervalo = $2\lambda_{\alpha/2} \sqrt{\frac{pq}{n}}$ despejando n

obtendríamos

$$\sqrt{n} = \frac{2 \lambda_{\alpha/2} \sqrt{pq}}{e} \rightarrow n = \frac{4 \lambda_{\alpha/2}^2 \cdot p \cdot q}{e^2}$$

Como es habitual el cálculo de n se realiza previamente al muestreo y precisamente se realiza el muestreo para conseguir información sobre la proporción p con la que se da una característica en una población , por lo que parece difícil que se conozca dicha p

necesaria para conseguir el tamaño muestral para estimar precisamente p , de ahí que poniéndonos en el caso más desfavorable (varianza poblacional máxima) tomemos $p=q=0.5$; habitual precisión que se hace en las fichas técnicas de las encuestas.

Recurrentemente y por facilidad operativa suele plantearse, para este tipo de cálculos de tamaños muestrales, el nivel de confianza en el 95,5 %, la razón es que los valores de

$$\lambda_{\frac{1}{2}} = \lambda_{0.0225} \cong 2 ;$$

(ir a tabla de la normal) por lo que para una semi-holgura determinada de error la expresión sería

$$n = \frac{4 \cdot 0.5 \cdot 0.5}{e_1^2} = \frac{1}{e_1^2}$$

siendo, recordemos, e_1 la semi-amplitud en tanto por uno.

ejemplo 5

Calcular el tamaño muestral necesario para llevar a cabo un m.a.s. para determinar un intervalo para la proporción de personas vegetarianas si el error que estamos dispuestos a admitir es del más menos 3%. A) nivel de confianza 90%. B) nivel de confianza 95.5%.

A) $e = \pm 3\% = \pm 0.03 = 0.06$ amplitud del intervalo

(ir a script de realización)

nivel de confianza $1-\alpha = 0.9$ por lo que (ir a tabla de la normal) $\lambda_{\alpha/2} = \lambda_{0.05} = 1.645$
1.645

luego tomando $p=q=0.5$

$$n = \frac{4 \lambda_{\frac{\alpha}{2}}^2 \cdot p \cdot q}{e^2} = \frac{4 \cdot 2.706 \cdot 0.5 \cdot 0.5}{0.0036} \cong 752$$

B) la semi-holgura sería 0.03 semi-amplitud
(ir a script de realización)

nivel de confianza $1-\alpha = 0.995$ por lo que

$\lambda_{\frac{\alpha}{2}} \cong 2$ (ir a tabla de la normal) y, finalmente:

$$n = \frac{4 \cdot 0.5 \cdot 0.5}{e_1^2} = \frac{1}{e_1^2} = \frac{1}{0.0009} \cong 1111$$

Si el intervalo a construir parte de un muestreo irrestricto (sin reemplazamiento) el cálculo del tamaño muestral partiría de la expresión:

$$P\left(\hat{p} - \lambda_{\alpha/2} \frac{\sqrt{pq}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \leq p \leq \hat{p} + \lambda_{\alpha/2} \frac{\sqrt{pq}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}\right) = 1 - \alpha$$

conocido el error "e" que estamos dispuestos a conocer y conocida también la población (su tamaño), N, así como el nivel de confianza . Tendríamos que

$$e = 2 \left(\lambda_{\alpha/2} \frac{\sqrt{pq}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right)$$

de donde despejaríamos n para conseguir el tamaño muestral. Es evidente que es mucho más cómodo despejar n para cada caso específico que plantear una expresión general, por lo que recomendamos que así se haga. Las afirmaciones que se vertieron sobre "p" y su valor en el caso de m.a.s. son de aplicación, también, en este caso.