UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE CIENCIAS

"APLICACIONES DE PROBABILIDAD Y ESTADISTICA PARA PROBLEMAS SOCIALES"

T E S I S

QUE PARA OBTENER EL TITULO DE

A C T U A R I O

P R E S E N T A :

ARAVID TORRES GARDUÑO

DIRECTOR DE TESIS: ACT. LAURA MIRIAM QUEROL GONZALEZ

2001

VNIVERSIDAD NACIONAL
AVTONOMA DE
MEXICO

M. EN C. ELENA DE OTEYZA DE OTEYZA
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:
"Aplicaciones de Probabilidad y Estadística para Problemas Sociales"

realizado por Aravid Torres Garduño

con número de cuenta 9209848-0 , pasante de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	Act. Laura Miriam Querol González
Propietario	Mat. Javier González Rosas
Propietario	Act. María Aurora Valdez Michel
Suplente	Act. Marina Castillo Garduño
Suplente	Act. Víctor Manuel Solís Nájera

Consejo Departamental de Matemáticas

M. en C. José Antonio Flores Díaz

MATEMATICAS

INDICE

INTRODUCCIÓN

En muchas situaciones de la vida real en las que se tiene una población de estudio de interés, es necesario estudiar sólo una parte de ella, ya que por diversas causas no es posible obtener mediciones del total de los elementos de esa población. Sin embargo, en la vida real no existen criterios sólidos que permitan tomar la mejor decisión en cuanto al tamaño de esa parte de la población de estudio, por lo que es necesario identificar los elementos que están presentes en la situación real con los elementos de un modelo probabilístico.

Por ejemplo, en la vida real, generalmente se tienen personas, viviendas, hospitales, etc., a las que se les hacen diversas mediciones como la edad de la persona, el número de mujeres que habitan la vivienda o el número de niños que atiende el hospital, mientras que en el modelo probabilístico se tienen variables aleatorias y teoremas que garantizan importantes resultados siempre y cuando se cumplan condiciones muy específicas en las variables aleatorias.

Para comprender la asociación entre el mundo real y la teoría es necesario tener en cuenta por un lado que, la característica fundamental que define a una variable aleatoria, es que asume valores que no se pueden predecir y por el otro, que en una situación real no se puede predecir la edad de una persona que se va a seleccionar aleatoriamente de una población de estudio, tampoco se puede predecir su número de hijos, su nivel de escolaridad o si será hombre o mujer, etc.

En este contexto se puede identificar entonces a una variable aleatoria, con cualquier variable que se pretende medir en una persona, casa u hospital que se selecciona de manera aleatoria de una población de estudio, de tal forma que al seleccionar de manera aleatoria n elementos de esa población con el fin de medir alguna característica de interés, se tienen $X_1, X_2, \ldots, X_n$ variables aleatorias independientes. Es decir, una variable aleatoria por cada persona, casa u hospital seleccionado. De esta manera el problema de decidir cuantos elementos de una población de estudio es necesario medir en una situación real, se resuelve determinando el número n de variables aleatorias.

Este trabajo está dividido en 3 partes principales, en la primera se aborda el caso en el que la población de estudio es tan grande que puede ser considerada como infinita; En la segunda, se aborda el caso en el que la población de estudio es tan pequeña, que al efectuar el cálculo del tamaño de muestra suponiendo que la población es infinita, se obtiene un tamaño muestral mayor al de la población de estudio. En el último capítulo se presentan ejemplos de problemas reales en los que se calcula el tamaño de muestra de diferentes poblaciones de estudio. Dichas aplicaciones reflejan la utilidad de la teoría del muestreo así como el campo de trabajo del actuario en diversos estudios, como lo son: Estudios sociodemográficos, las tendencias en las votaciones para elecciones presidenciales, medición de la fecundidad, estudios sobre ingreso a la universidad, estudio de calidad de servicios de salud en aseguradoras.

El tercer capítulo tiene la importancia especial de mostrar el papel del actuario en el planteamiento lógico y solución de problemas frecuentes en la vida real.

CAPITULO I

TAMAÑO DE MUESTRA EN POBLACIONES INFINITAS

1.1 El problema de la infinitud en la práctica

En la vida real no existen poblaciones infinitas, sólo existen poblaciones tan grandes que pueden considerarse como de tamaño infinito. El ejemplo de la vida real que durante mucho tiempo se pensó que era de tamaño infinito fue el universo, sin embargo utilizando las ecuaciones de campo de Einstein y la cantidad media de materia en el universo calculada por Hubble, se calculó que el radio del universo es de 35 mil millones de años luz (Barnett, 1985), es decir, es finito. En este sentido es importante dejar claro que, las poblaciones infinitas sólo existen en los modelos teóricos que intentan explicar la vida real que nos rodea.

1.2 Ley Débil de los Grandes Números

Uno de tantos modelos teóricos basados en el concepto de infinitud es la Ley Débil de los Grandes Números. La ley afirma que si se tienen $X_1, X_2, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas con media θ finita y varianza σ^2 uniformemente acotada, entonces para toda $d > 0$ se cumple que:

$$\lim_{n \to \infty} P\left[\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \theta\right| \leq d\right] = 1$$

El resultado indica que cuando se consideran a todas las variables de la población infinita entonces la distancia entre los términos $\frac{1}{n}\sum_{k=1}^{n} X_k$ y θ es más pequeña que d y la probabilidad de que ocurra lo anterior es uno.

Este resultado es la base para calcular el tamaño de muestra en una población infinita porque si se considera una $d > 0$ y un número de variables aleatorias menor a la población infinita entonces se tiene que

$$P\left[\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \theta\right| \leq d\right] = 1 - \alpha$$

La expresión $\frac{1}{n}\sum_{k=1}^{n} X_k$ se conoce como un estimador de θ y se denota como $\hat{\theta}$.

1.3 Variables aleatorias independientes

En la práctica, es común, que al considerar el valor de una característica de interés dentro la población de estudio, el valor obtenido en algún elemento de dicha población, no tiene ninguna relación con el valor obtenido en ningún otro elemento de la misma población. De modo que cuando se asocian variables aleatorias a los elementos de la población se puede considerar que dichas variables aleatorias son independientes entre sí

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas con media θ y varianza σ^2 finitas.

Es claro que $\hat{\theta}$ es una variable aleatoria que tiene una media y una varianza, las cuales se calculan de la siguiente manera:

$$
\begin{aligned}
E(\hat{\theta}) &= E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n} E\left(\sum_{k=1}^{n} X_k\right) \\
&= \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \frac{1}{n} n\theta = \theta
\end{aligned}
$$

$$
VAR(\hat{\theta}) = VAR\left(\frac{\sum_{k=1}^{n} X_k}{n}\right) = \frac{1}{n^2} VAR\left(\sum_{i=k}^{n} X_k\right)
$$

Como las variables aleatorias son independientes entonces

$$
\begin{aligned}
VAR(\hat{\theta}) &= \frac{\sum_{k=1}^{n} VAR(X_k)}{n^2} = n\frac{\sigma^2}{n^2} \\
&= \frac{\sigma^2}{n} \qquad (1.1)
\end{aligned}
$$

De esta manera al establecer que la diferencia entre el valor estimado del parámetro y su valor real sea menor que una cantidad d y además al fijar una probabilidad $1-\alpha$, entonces como se había mencionado la ley Débil de los Grandes Números asegura la existencia de una n que satisface lo siguiente:

$$P\left(\left|\hat{\theta}-\theta\right| \leq d\right)=1-\alpha \qquad (1.2)$$

Lo anterior quiere decir que cuando se calcula $\hat{\theta}$ con esa n, la desviación del estimador $\hat{\theta}$ con respecto al valor real θ es menor que d y la probabilidad de que esto suceda es $1-\alpha$. La expresión (1.2) es equivalente a:

$$P\left(-d \leq (\hat{\theta}-\theta) \leq d\right)=1-\alpha \qquad (1.3)$$

También es importante observar por otro lado que, la expresión $\hat{\theta}-\theta$ es también una variable aleatoria que se puede dividir entre el número $\frac{\sigma}{\sqrt{n}}$ que es la desviación estándar del estimador $\hat{\theta}$

A partir de este hecho y de la ecuación (1.3) se tiene que:

$$P\left(\frac{-d}{\sigma/\sqrt{n}} \leq \frac{(\hat{\theta}-\theta)}{\sigma/\sqrt{n}} \leq \frac{d}{\sigma/\sqrt{n}}\right)=1-\alpha \qquad (1.4)$$

donde:

n es el número de personas que se seleccionarán de la población de estudio, y

σ representa la desviación estándar de las variables aleatorias X_i

La ventaja de la expresión (1.4) es que por el Teorema Central del Límite[1], cuando n es muy grande, la variable aleatoria $\frac{(\hat{\theta}-\theta)}{\sigma/\sqrt{n}}$ se distribuye de acuerdo a una ley de distribución Normal estándar con media cero y varianza uno, entonces para $1-\alpha$ fija, se pueden encontrar dos números reales denotados como $-Z_{\alpha/2}$ y $Z_{\alpha/2}$ tales que:

$$P\left(-Z_{\alpha/2} \leq \frac{(\hat{\theta}-\theta)}{\sigma/\sqrt{n}} \leq Z_{\alpha/2}\right) = 1-\alpha \qquad (1.5)$$

donde:

$-Z_{\alpha/2}$ representa un número real tal que, la probabilidad de que los valores de la variable aleatoria normal estándar sean menores que este valor es $\alpha/2$, y

$Z_{\alpha/2}$ representa un número real tal que, la probabilidad de que los valores de la variable aleatoria normal estándar sean valores mayores que este número es de $\alpha/2$

---

[1] Sean $X_1, X_2, \ldots, X_n$, n variables aleatorias independientes e idénticamente distribuidas con una distribución de probabilidad no especificada y que tiene una media μ y una varianza σ^2 finita.

El promedio muestral $\bar{X} = (X_1 + X_2 + \cdots X_n)/n$ tiene una distribución con media μ y varianza σ^2/n que tiende hacia una distribución normal conforme n tiende a infinito.

En otras palabras, la variable aleatoria $(\bar{X}-\mu)/(\sigma/\sqrt{n})$ tiene como límite una distribución normal estándar. (Canavos, 1991)

Ahora bien como las expresiones (1.4) y (1.5) se refieren a la misma variable aleatoria y a la misma probabilidad, entonces los extremos de los intervalos de las probabilidades (1.4) y (1.5) deben de ser iguales respectivamente, es decir:

$$\frac{-d}{\sigma/\sqrt{n}} = -Z_{\alpha/2} \tag{1.6}$$

$$\frac{d}{\sigma/\sqrt{n}} = Z_{\alpha/2} \tag{1.7}$$

Entonces despejando a n de la expresión (1.6) o (1.7) se obtiene el valor de n y queda de la siguiente forma:

$$\boxed{n = \sigma^2\left(Z_{\alpha/2}/d\right)^2} \tag{1.8}$$

donde:

σ^2 es la varianza de cualquiera de las variables aleatorias $X_1, X_2, \cdots, X_n$.

d es la desviación máxima del estimador $\hat{\theta}$ y el valor real θ.

$Z_{\alpha/2}$ es un número real determinado con base en la probabilidad $1-\alpha$ y,

n representa el tamaño de la muestra o bien el número de individuos que es necesario seleccionar de la población de estudio.

La expresión (1.8) representa entonces la solución al problema de calcular el tamaño de una parte de una población infinita de interés, e indica que para calcularla es necesario conocer la varianza σ^2 y establecer una desviación y una probabilidad. La desviación mide la cercanía del estimador $\hat{\theta}$ con el valor real θ y esta se define como la precisión del estimador. Por otra parte la probabilidad asociada al número de veces que el estimador se desviará en una cantidad menor a la precisión, se conoce como la confiabilidad del estimador.

Sin embargo, la solución tiene un grave problema, pues como se dijo el tamaño de muestra depende de la confiabilidad, de la precisión y de la varianza σ^2 de las variables aleatorias X_i. El problema de la confiabilidad y de la precisión no es tan difícil de resolver, ya que dichos valores el investigador los establece, pero el problema de σ^2 no es tan sencillo de resolver, ya que en la práctica generalmente θ y σ^2 son desconocidos.

El problema tiene solución para algunas variables aleatorias en las que se puede demostrar que la varianza de la expresión (1.8) se puede acotar con un valor conocido.

1.3.1 El caso con ley de distribución Bernoulli

En muchos casos el objetivo se centra en calcular el porcentaje de personas que tienen alguna característica de interés en una población de estudio. En este caso, con cada una de las personas de la población se define una variable que vale uno si tiene la característica de interés y cero si no la tiene.

En la teoría probabilística una variable con estas características se conoce como variable aleatoria con ley de distribución Bernoulli[2], cuya media es la probabilidad de que la variable aleatoria asuma el valor de uno.

De esta manera, al sustituir en (1.8) el valor de σ^2 para el caso en donde cada una de las X_i son independientes entre sí y tiene una distribución de tipo Bernoulli se obtiene que:

$$\boxed{n = PQ\left(\frac{Z_{\alpha/2}}{d}\right)^2} \qquad (1.9)$$

donde:

PQ es la varianza de cualquiera de las variables aleatorias $X_1, X_2, \ldots, X_n$

P representa la proporción o porcentaje de interés

Q es el complemento de P es decir $Q = (1-P)$

d es la precisión deseada

$Z_{\alpha/2}$ es el valor de una distribución normal estándar asociado con la confiabilidad $1-\alpha$ deseada, y

n representa el tamaño de la muestra o el número de elementos que es necesario seleccionar de la población de estudio.

---

[2] La distribución Bernoulli esta dada por:

$$\begin{aligned} p(x) &= p \text{ si } x = 1 \\ &= 1 - p \text{ si } x = 0 \\ &= \text{en otro caso (Leithold, 1982)} \end{aligned}$$

Como se observa en la expresión (1.9) el tamaño de muestra n depende de manera directamente proporcional al valor P. Lo cual indica que entre más grande sea el valor de P mayor será el tamaño de muestra n y el problema se centra en determinar si existe un valor máximo para una confiabilidad y precisión fijas.

Como el tamaño de muestra en (1.9) es una función de P al dejar la confiabilidad y la precisión constante, se tiene que:

$$n = f(P) = PQ\left(Z_{\alpha/2}/d\right)^2$$

si $K = \left(Z_{\alpha/2}/d\right)^2$ entonces

$$n = P(1-P)K \qquad , K > 0$$

$$= KP - KP^2$$

Si ahora se deriva $f(P)$ con respecto a P y se encuentra el valor de la derivada en donde ésta se hace cero, entonces se obtiene un valor crítico[3] el cual puede ser un mínimo o un máximo. Es decir:

$$f'(P) = K - 2KP$$

[3] Si c es un número en el dominio de la función f y si $f'(c) = 0$ ó $f'(c)$ no existe, entonces c se llama un número crítico de f

Igualando la derivada a cero se tiene que:

$$K - 2KP = 0$$

$$K(1 - 2P) = 0$$

como $K > 0$ entonces

$$(1 - 2P) = 0$$

$$P = 1/2$$

Para verificar que se trata de un máximo[4] se obtiene la segunda derivada de la función $f'(P) = K - 2KP$:

$$f''(P) = -2K < 0$$

La cual es menor que cero para toda P, en particular para $P = \frac{1}{2}$ lo que implica que en dicho valor la función tamaño de muestra alcanza el máximo[5], de donde se concluye que si el cálculo de n se hace sustituyendo el valor de $P = \frac{1}{2}$ en la ecuación (1.9), se obtiene que n es igual a:

[4] Sea c un número crítico de una función f en la cuál $f'(c) = 0$, y f existe para todos los valores de x en algún intervalo abierto que contenga a c. Sí $f''(c) = 0$ existe y
(i) si $f''(c) < 0$, entonces f tiene un valor máximo relativo en c
(ii) si $f''(c) > 0$, entonces f tiene un valor mínimo relativo en c

[5] La gráfica de $y = f(x)$ es cóncava hacia abajo en todo el intervalo en que $y'' < 0$ (Thomas/Finney, 1987)

$$\boxed{n = \frac{1}{4}\left(\frac{Z_{\alpha/2}}{d}\right)^2} \qquad (1.10)$$

Que tiene dos características muy importantes, en primer lugar que es mayor o igual a la n que se obtiene con (1.9) y en segundo lugar, que sólo depende de d y $Z_{\alpha/2}$ que son valores que el investigador establece.

1.3.2 El caso con ley de distribución Binomial

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas con ley de distribución Binomial[6], con media θ y varianza σ^2 y que se obtienen al sumar m variables aleatorias de Bernoulli, es decir:

$$X_i = Y_1 + Y_2 + \cdots + Y_m$$

donde:

Y_i tiene una distribución Bernoulli con media P $(1 \le i \le m)$, y

X_i tiene una distribución Binomial, con media mP y varianza mPQ

[6] Sea X una variable aleatoria que representa el número de éxitos en n ensayos y p la probabilidad de éxito con cualquiera de éstos. Entonces X tiene una distribución Binomial con función de probabilidad

$$p(x; n, p) = \begin{cases} \dfrac{n!}{(n-x)!\,x!} p^x (1-p)^{n-x} & x = 0,1,2,\ldots,n, \\ 0 \quad en \quad otro \quad caso & 0 \le p \le 1, \quad para \quad n \quad entero \end{cases}$$ (Canavos, 1991).

Por lo que, si en la expresión (1.8) se sustituye el valor de σ^2 para el caso en donde cada una de las X_i tiene una distribución de tipo Binomial se obtiene:

$$\boxed{n = mPQ\left(\frac{Z_{\alpha/2}}{d}\right)^2} \qquad (1.11)$$

donde:

mPQ es la varianza de cualquiera de las variables aleatorias

$X_1, X_2, \ldots, X_n$

m representa el número variables aleatorias de Bernoulli

P representa la media de las variables de tipo Bernoulli

Q es el complemento de P es decir $Q = (1 - P)$

d es la precisión deseada

$Z_{\alpha/2}$ es el valor de una distribución normal estándar asociada con la confiabilidad $1-\alpha$ deseada, y

n representa el tamaño de la muestra o el número de elementos que es necesario seleccionar de la población de estudio.

Como se observa en la expresión (1.11) el tamaño de muestra n depende de manera directamente proporcional al valor P ya que todo lo demás es constante. Es decir:

$$n = f(P) = mPQ\left(\frac{Z_{\alpha/2}}{d}\right)^2$$

si $K = m\left(\frac{Z_{\alpha/2}}{d}\right)^2$ entonces

$$n = P(1-P)K, \quad K > 0$$

Que como se demostró en la sección anterior alcanza el máximo en $P = 1/2$ y al sustituirse en la expresión (1.11) se tiene que:

$$\boxed{n = \frac{1}{4}m\left(\frac{Z_{\alpha/2}}{d}\right)^2}$$

que son valores establecidos por el investigador.

1.4 Variables aleatorias no independientes

Es importante aclarar que hasta aquí, el cálculo del tamaño de muestra está resuelto solamente cuando las variables aleatorias son independientes.

El supuesto de independencia ha sido crucial en la solución del problema del tamaño de muestra, en primer lugar porque son válidos el Teorema Central del Límite y la Ley Débil de los Grandes Números y en segundo lugar porque el desarrollo de la expresión de la varianza de θ es muy sencillo. Sin embargo en algunas situaciones las variables aleatorias no son independientes y entonces es necesario encontrar la solución del cálculo del tamaño de muestra en estos casos.

1.4.1 Varianza del estimador de la media θ

Supóngase que $X_1, X_2, \ldots, X_n$ son variables aleatorias idénticamente distribuidas con media θ y varianza σ^2 pero no independientes entre sí. Entonces:

$$VAR(\theta) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left[\sum_{i=1}^{n}\sigma^2 + 2\sum_{i<j}\sum Cov(X_i, X_j)\right] \tag{1.18}$$

Por definición de covarianza se tiene que:

$$COV(X_i, X_j) = E[(X_i - \mu)(X_j - \mu)]$$

$$= E(X_i X_j - \mu X_i - \mu X_j + \mu^2)$$

$$= E(X_i X_j) - \mu E(X_i) - \mu E(X_j) + E(\mu^2)$$

$$= E(X_i X_j) - \mu^2 - \mu^2 + \mu^2$$

$$= E(X_i X_j) - \mu^2 \tag{1.19}$$

Ahora bien, como para la unidad i-ésima la probabilidad de selección es de $1/N$ y para la unidad j-ésima la probabilidad de selección (después de la unidad i) es de $1/(N-1)$ por tratarse de un muestreo aleatorio simple sin reemplazo, se tiene al desarrollar la expresión (1.19) en términos de sumatorias:

$$COV(X_i,X_j)=E(X_iX_j)-\mu^2=\sum_{i\neq j}^{N}X_iX_j\left[\frac{1}{N(N-1)}\right]-\frac{1}{N^2}\left(\sum_{i=1}^{N}X_i\right)^2$$

$$=\frac{1}{N}\left[\sum_{i\neq j}^{N}\frac{X_iX_j}{N-1}-\frac{1}{N}\left(\sum_{i=1}^{N}X_i\right)^2\right] \quad (1.20)$$

Ahora considerando la expresión

$$\sum_{i\neq j}^{N}X_iX_j \quad (1.21)$$

la cual se puede escribir como sigue

$$\begin{aligned}\sum_{i\neq j}^{N}X_iX_j = {} & X_2X_1+X_3X_1+\ldots+X_NX_1+\\ & X_1X_2+X_3X_2+\ldots+X_NX_2+\\ & X_1X_3+X_2X_3+\ldots+X_NX_3+\\ & \vdots\\ & X_1X_N+X_2X_N+\ldots+X_{N-1}X_N\end{aligned}$$

$$=\sum_{i\neq 1}^{N}X_iX_1+\sum_{i\neq 2}^{N}X_iX_2+\sum_{i\neq 3}^{N}X_iX_3+\cdots+\sum_{i\neq N}^{N}X_iX_N$$

$$=\sum_{i\neq j}^{N}\sum_{j=1}^{N}X_iX_j \quad (1.22)$$

Sustituyendo este resultado en (1.20) se llega a que

$$\frac{1}{N}\left[\sum_{i\neq j}^{N}\frac{X_iX_j}{N-1}-\frac{1}{N}\left(\sum_{i=1}^{N}X_i\right)^2\right]=\frac{1}{N}\left[\sum_{i\neq j}^{N}\sum_{j=1}^{N}\frac{X_iX_j}{N-1}-\frac{1}{N}\left(\sum_{i=1}^{N}X_i\right)^2\right] \quad (1.23)$$

Sea $T_1 = X_1X_2 + X_1X_3 + \cdots + X_1X_N$

$T_2 = X_2X_1 + X_2X_3 + \cdots + X_2X_N$

$\vdots$

$T_N = X_1X_n + \cdots + X_NX_{N-1}$

Entonces la expresión (1.23) se escribe como

$$= \frac{1}{N}\left[\frac{1}{N-1}(T_1 + T_2 + \cdots + T_N) - \frac{1}{N}\left(\sum_{i=1}^{N} X_i\right)^2\right]$$

Donde al sumar y restar $X_1^2 + \cdots X_N^2$ se tiene

$$= \frac{1}{N}\left[\frac{1}{N-1}\left(X_1^2 + T_1 + X_2^2 + T_2 + \cdots + X_n^2 + T_N - \{X_1^2 + \cdots + X_N^2\}\right) - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}\right]$$

Al hacer $U = X_1 + X_2 + \cdots X_N$ la expresión anterior queda

$$= \frac{1}{N}\left[\frac{1}{N-1}\left[X_1U + X_2U + \cdots + X_NU - \left(X_1^2 + \cdots + X_N^2\right)\right] - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}\right]$$

Al factorizar a U se tiene que

$$\frac{1}{N}\left[\sum_{i\neq j}^{N}\sum_{j=1}^{N}\frac{X_iX_j}{N-1}-\frac{1}{N}\left(\sum_{i=1}^{N}X_i\right)^2\right]=\frac{1}{N}\left[\frac{1}{N-1}\left[U^2-\left(X_1^2+\cdots+X_N^2\right)\right]-\frac{\left(\sum_{i=1}^{N}X_i\right)^2}{N}\right]$$

$$=\frac{1}{N}\left[\frac{1}{N-1}\left[(X_1+\cdots+X_N)(X_1+\cdots+X_N)-\left(X_1^2+\cdots+X_N^2\right)\right]-\frac{\left(\sum_{i=1}^{N}X_i\right)^2}{N}\right]$$

Por lo que se obtiene que

$$COV\left(X_i,X_j\right)=\frac{1}{N}\left[\frac{\left(\sum_{i=1}^{N}X_i\right)^2-\sum_{i=1}^{N}X_i^2}{N-1}-\frac{1}{N}\left(\sum_{i=1}^{N}X_i\right)^2\right]$$

$$=\frac{1}{N}\left[\frac{\left(\sum_{i=1}^{N}X_i\right)^2}{N-1}-\frac{\sum_{i=1}^{N}X_i^2}{N-1}-\frac{\left(\sum_{i=1}^{N}X_i\right)^2}{N}\right]$$

$$=\frac{1}{N}\left[-\frac{\sum_{i=1}^{N}X_i^2}{N-1}+\frac{\left(\sum_{i=1}^{N}X_i\right)^2}{N-1}-\frac{\left(\sum_{i=1}^{N}X_i\right)^2}{N}\right]$$

$$= -\frac{1}{N}\left[\frac{\sum_{i=1}^{N} X_i^2}{N-1} - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N-1} + \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}\right]$$

factorizando $\left(\sum_{i=1}^{N} X_i\right)^2$ en la expresión anterior se llega a

$$= -\frac{1}{N}\left[\frac{\sum_{i=1}^{N} X_i^2}{N-1} + \left(\frac{-1}{N-1} + \frac{1}{N}\right)\left(\sum_{i=1}^{N} X_i\right)^2\right]$$

$$= -\frac{1}{N}\left[\frac{\sum_{i=1}^{N} X_i^2}{N-1} + \left(\frac{-N+N-1}{N(N-1)}\right)\left(\sum_{i=1}^{N} X_i\right)^2\right]$$

$$= -\frac{1}{N}\left[\frac{1}{N-1}\sum_{i=1}^{N} X_i^2 - \frac{1}{N(N-1)}\left(\sum_{i=1}^{N} X_i\right)^2\right]$$

$$= -\left[\frac{1}{N(N-1)}\sum_{i=1}^{N} X_i^2 - \frac{1}{N^2(N-1)}\left(\sum_{i=1}^{N} X_i\right)^2\right]$$

$$= -\frac{1}{N-1}\left[\left(\frac{\sum_{i=1}^{N} X_i^2}{N}\right) - \left(\frac{\sum_{i=1}^{N} X_i}{N}\right)^2\right] \qquad (1.25)$$

Se sabe que $Var(Y)=E(Y^2)-E^2(Y)$ por lo que al expresarlo en sumatorias se tiene que

$$\frac{\sum_{i=1}^{N}(X_i-\theta)^2}{N}=\left(\frac{\sum_{i=1}^{N}X_i^2}{N}\right)-\left(\frac{\sum_{i=1}^{N}X_i}{N}\right)^2 \tag{1.26}$$

por lo que de (1.25) y (1.26) se concluye que:

$$-\frac{1}{N-1}\left[\left(\frac{\sum_{i=1}^{N}X_i^2}{N}\right)-\left(\frac{\sum_{i=1}^{N}X_i}{N}\right)^2\right]=-\left(\frac{\sum_{i=1}^{N}(X_i-\theta)^2}{N}\right)\left(\frac{1}{N-1}\right) \tag{1.27}$$

Y como $\frac{1}{N}\sum_{i=1}^{N}(X_i-\theta)^2=\sigma^2$ entonces (1.27) queda como

$$-\frac{1}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}(X_i-\theta)^2\right]=-\frac{1}{N-1}\sigma^2 \tag{1.28}$$

que es finalmente la covarianza entre X_1 y X_2, por lo que al sustituir en (1.18) se llega a que

$$VAR(\theta)=\frac{1}{n^2}\left[\sum_{i=1}^{n}\sigma^2+2\sum_{i<j}\sum\frac{-\sigma^2}{N-1}\right] \tag{1.30}$$

en donde los valores σ^2 y $-\sigma^2/_{N-1}$ no dependen de los índices de las sumatorias. En el primer caso es claro que se suma n veces σ^2, sin embargo en el segundo es necesario considerar lo siguiente:

Se tiene que $\sum_{i<j}\sum \frac{-\sigma^2}{N-1} = \sum_{i=1}^{n}\sum_{j=i+1}^{n}\frac{-\sigma^2}{N-1}$ y el problema consiste en determinar cuantas veces se suma el factor $\frac{-\sigma^2}{N-1}$.

La siguiente matriz muestra el número total de sumandos si se tratara de la sumatoria $\sum_{i=1}^{n}\sum_{j=1}^{n}$

$$\begin{matrix} (1,1) & (1,2) & (1,3) & & (1,n) \\ (2,1) & (2,2) & (2,3) & & (2,n) \\ & & \vdots & \cdots & \\ (n-1,1) & (n-1,2) & (n-1,3) & & (n-1,n) \\ (n,1) & (n,2) & (n,3) & & (n,n) \end{matrix}$$

Para el caso de este estudio la sumatoria es $\sum_{i=1}^{n}\sum_{j=i+1}^{n}$ donde la condición $j = i+1$ sólo se cumple para los elementos que se encuentran por arriba de la diagonal de la matriz. Entonces el número de elementos que hay por encima de la diagonal de la matriz es el número que se quiere calcular

El número de sumandos de $\sum_{i=1}^{n}\sum_{j=i+1}^{n}$ es igual al número de parejas (i,j) tales que $i<j$ lo cual es equivalente a obtener el total de subconjuntos de tamaño 2 tomados de n elementos

$\binom{n}{r}=\frac{n!}{(n-r)r!}$ indica el número total de formas de agrupar n elementos en subgrupos de r elementos, es decir, el total de formas posibles de seleccionar n elementos de r en r de forma que para este caso la expresión usada es

$$\binom{n}{2}=\frac{n!}{(n-2)!2!}=\frac{n!}{(n-2)!2}=\frac{n(n-1)(n-2)!}{(n-2)!2}=\frac{n(n-1)}{2} \tag{1.31}$$

por lo que (1.29) queda de la siguiente manera

$$VAR(\theta)=\frac{1}{n^2}\left\{n\sigma^2-\frac{2\sigma^2}{N-1}\left[\frac{n(n-1)}{2}\right]\right\} \tag{1.32}$$

$$=\frac{\sigma^2}{n}-\frac{\sigma^2(n-1)}{n(N-1)}$$

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

$$n = mPQ\left(\frac{k}{d}\right)^2$$

$$= \sigma^2 \left(\frac{N-n}{n(N-1)} \right) \tag{1.33}$$

Por lo tanto:

$$V(\hat{\theta}) = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right) \tag{1.35}$$

donde:

σ^2 es la varianza de cualquiera de las variables aleatorias
X_i,
N es el tamaño de la población de estudio, y
n es el tamaño de la muestra

Ahora es necesario analizar lo que sucede con la expresión (1.35) en el caso particular en que $N \to \infty$ por ser este uno de los supuestos en el desarrollo de las expresiones de este capítulo. Al calcular el límite se tiene lo siguiente:

$$\underset{N\to\infty}{Lim}\, \sigma_{\theta} = \underset{N\to\infty}{Lim} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$= \underset{N\to\infty}{Lim} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N\left(1-\frac{n}{N}\right)}{N\left(1-\frac{1}{N}\right)}}$$

$$= \frac{\sigma}{\sqrt{n}}$$

Que es la desviación estándar obtenida para el caso donde se supone independencia entre cada una de las variables aleatorias.

1.4.2 El caso con ley de distribución arbitraria

Dado que las variables $X_1, X_2, \ldots, X_n$ no son independientes entre sí, entonces no es posible utilizar el Teorema Central del Límite como en el caso donde las variables son independientes, pero por el teorema de Tchebysheff se sabe que para la variable aleatoria $\hat{\theta}$ tal que $E(\hat{\theta}) = \theta$ y $Var(\hat{\theta}) = \sigma^2$ son finitos. Se cumple que:

$$P\left(\left|\hat{\theta} - \theta\right| \leq k\sigma/\sqrt{n}\right) \geq 1 - \frac{1}{k^2} \tag{1.36}$$

ya que la media de $\hat{\theta}$ es θ y su desviación estándar como se demostró en la sección anterior es $\sigma/\sqrt{n}$. La expresión (1.36) quiere decir que la probabilidad de que una variable aleatoria se aleje no más de k desviaciones estándar de θ es mayor o igual a $1 - (1/k^2)$ para algún valor de $k \geq 1$.

Además si $\alpha = 1/k^2$ entonces $k = 1/\sqrt{\alpha}$ y al sustituir el valor de k en el lado derecho de (1.36) se tiene que:

$$P\left(\left|\hat{\theta} - \theta\right| \leq k\sigma/\sqrt{n}\right) \geq 1 - \alpha \tag{1.37}$$

Por lo tanto, se puede y se desea encontrar una $d > 0$ tal que:

$$P\left(\left|\hat{\theta}-\theta\right|\leq d\right)\geq 1-\alpha \qquad (1.38)$$

por consecuencia de (1.37) y (1.38) se tiene que:

$$d = k\,\sigma/\sqrt{n} \qquad (1.39)$$

De manera que al despejar a n queda como

$$\boxed{n=\sigma^2\left(\frac{k}{d}\right)^2} \qquad (1.40)$$

donde:

σ^2 es la varianza de cualquiera de las variables aleatorias $X_1, X_2, \cdots, X_n$

d es la desviación máxima del estimador $\hat{\theta}$ y el valor real θ.

k es un número real que determina que la confiabilidad del estimador sea al menos $1-\alpha$ y,

n representa el tamaño de la muestra o bien el número de individuos que es necesario seleccionar de la población de estudio.

1.4.3 El caso con ley de distribución Bernoulli y Binomial

En el caso en que las X_i son no independientes entre sí y tienen una distribución Bernoulli, al sustituir en (1.40) el valor de σ^2 se tiene que:

$$\boxed{n=PQ\left(\frac{k}{d}\right)^2} \qquad (1.41)$$

Como la expresión (1.41) depende sólo de P al sustituir Q por $(1-P)$ y dejar constante k y d, entonces

$$n = f(P) = \left(P - P^2\right)R, \quad R = \left(\frac{k}{d}\right)^2$$

que como se demostró en la sección 1.3.1 la expresión anterior alcanza su máximo en $P = \frac{1}{2}$, de tal manera que al sustituir en (1.41) se puede calcular el tamaño de muestra.

El otro caso es cuando $X_1, X_2, \ldots, X_n$ son variables aleatorias no independientes e idénticamente distribuidas con ley de distribución Binomial, con media θ y varianza σ^2.

En este caso, si en la expresión (1.40) se sustituye el valor de σ^2 se obtiene:

$$\boxed{n = mPQ\left(\frac{k}{d}\right)^2} \tag{1.42}$$

Expresando la ecuación (1.42) como una función de P y dejando constante tanto la precisión como la confiabilidad se tiene

$$n = f(P) = \left(P - P^2\right)T, \quad T = \left(\frac{k}{d}\right)^2$$

que de acuerdo al resultado de la sección 1.3.2 alcanza el máximo también en $P = \frac{1}{2}$.

CAPITULO 2

TAMAÑO DE MUESTRA EN POBLACIONES FINITAS

2.1 El problema en poblaciones finitas pequeñas

El mundo que nos rodea se puede clasificar en grande y pequeño, ambos desde hace muchos años han sido motivo de estudio. Numerosos científicos han dedicado parte de su vida al estudio de lo más pequeño, el átomo y sus componentes y también al estudio de lo más grande, el universo.

En términos de poblaciones de estudio de la vida real, la barrera entre lo grande y lo pequeño es endeble, porque no existe algún criterio único para determinar si 570 y 596 son grandes o pequeños, o si uno es pequeño y el otro es grande.

Por otro lado en muchas situaciones al calcular el tamaño de muestra con las fórmulas desarrolladas en el capítulo anterior, resulta que el tamaño de muestra es más grande que el número de elementos de la población de estudio. Esto se debe a que las poblaciones además de ser finitas, son pequeñas.

El inconveniente para el desarrollo del cálculo del tamaño de muestra es que no son válidos ni el Teorema Central del Límite, ni la Ley Débil de los Grandes Números que son las bases para calcular el tamaño de muestra en poblaciones consideradas como de tamaño infinito, por lo que es necesario desarrollar la teoría para resolver el cálculo del tamaño de muestra en este tipo de casos y además garantizar que el tamaño de la muestra que se obtenga sea menor o igual al tamaño de la población.

Precisamente el criterio utilizado en el presente trabajo para definir cuando una población es finita pequeña, tiene su base en el cálculo que arroja la fórmula para el caso de poblaciones consideradas como infinitas. Si el tamaño de muestra arrojado por esta fórmula es mayor que la población de estudio, entonces se puede considerar que la población es pequeña y utilizar la teoría desarrollada en este capítulo.

2.2 Caso con ley de distribución arbitraria

Supóngase que se tienen $X_1, X_2, \ldots, X_n$ variables aleatorias idénticamente distribuidas con media θ y varianza σ^2.

Para poder calcular el tamaño de muestra en poblaciones pequeñas y garantizar además que este sea menor que la población de estudio es necesario utilizar una expresión que involucre a la n muestral y a la N poblacional.

La expresión (1.35) se desarrolló para el caso de variables aleatorias no independientes[1], sin embargo el mismo resultado se puede obtener

[1] Ya que no se supone que $COV(X_i, X_j) = 0$

sin utilizar este supuesto. Es claro que en este caso, no es necesario tender la N a infinito, por lo que la desviación estándar de $\hat{\theta}$ es

$$\sigma_{\theta} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \tag{2.35}$$

donde:

σ_{θ} es la desviación estándar de cualquiera de las variables aleatorias X_i,

N es el tamaño de la población de estudio, y

n es el tamaño de la muestra

Por el teorema de Tchebysheff se tiene que:

$$P\left(\left|\hat{\theta}-\theta\right| \leq \frac{k\sigma}{\sqrt{n}}\left(\frac{N-n}{N-1}\right)^{1/2}\right) \geq 1-\frac{1}{k^2} \tag{2.36}$$

ya que la media de $\hat{\theta}$ es θ y su desviación estándar esta dada por la expresión (2.35)

De donde si $\alpha = 1/k^2$ entonces $k = 1/\sqrt{\alpha}$ por lo que al sustituir del lado derecho queda

$$P\left(\left|\hat{\theta}-\theta\right| \leq k\frac{\sigma}{\sqrt{n}}\left(\frac{N-n}{N-1}\right)^{1/2}\right) \geq 1-\alpha \tag{2.37}$$

Por otro lado se sabe que existe $d > 0$ tal que

$$P\left(\left|\hat{\theta} - \theta\right| \leq d\right) \geq 1 - \alpha \tag{2.38}$$

como en (2.37) y (2.38) es la misma variable aleatoria y la misma probabilidad entonces

$$d = k\frac{\sigma}{\sqrt{n}}\left(\frac{N-n}{N-1}\right)^{1/2} \tag{2.39}$$

Al elevar al cuadrado ambos lados de la igualdad se tiene que

$$d^2 = k^2\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

$$d^2 = \frac{\sigma^2 k^2 N - \sigma^2 k^2 n}{nN - n}$$

Multiplicando por $\frac{1/n}{1/n}$ en ambos lados de la igualdad queda

$$d^2 = \frac{\frac{\sigma^2 k^2 N}{n} - \sigma^2 k^2}{N-1}$$

$$\frac{d^2}{k^2}(N-1) + \sigma^2 = \frac{\sigma^2 N}{n}$$

$$n=\frac{\sigma^2 N}{\frac{d^2}{k^2}(N-1)+\sigma^2}=\frac{\sigma^2}{\frac{d^2}{k^2}-\frac{1}{N}\frac{d^2}{k^2}+\frac{\sigma^2}{N}}$$

$$=\frac{\sigma^2}{\frac{d^2}{k^2}-\frac{1}{N}\frac{d^2}{k^2}+\frac{\sigma^2}{N}}=\frac{\sigma^2\left(\frac{k}{d}\right)^2}{1-\frac{1}{N}+\frac{\sigma^2}{N}\left(\frac{k}{d}\right)^2}$$

De donde finalmente al factorizar $\frac{1}{N}$ el tamaño de muestra n queda como

$$\boxed{n=\frac{\sigma^2\left(\frac{k}{d}\right)^2}{1+\frac{1}{N}\left[\sigma^2\left(\frac{k}{d}\right)^2-1\right]}} \qquad (2.40)$$

donde:

σ^2 es la varianza de cualquiera de las variables X_i

d es la desviación máxima del estimador y el valor real θ.

k es un número real determinado con base en la probabilidad $1-\alpha$ N representa el número de elementos en la población de estudio, y

n representa el tamaño de la muestra o bien el número de individuos que es necesario seleccionar de la población de estudio.

2.3 Demostración de que la n muestral es menor o igual que la N poblacional

Para facilitar la notación supóngase que $A=\sigma^2\left(\frac{k}{d}\right)^2$ por lo que $A>0$, ya que $n>0$, $N>0, \sigma^2>0$ y $k>0$

Al sustituir A en (2.40) se tiene que

$$n=\frac{A}{1+\frac{1}{N}(A-1)}$$

$$n\left[1+\frac{1}{N}(A-1)\right]=A$$

$$n+\frac{n}{N}(A-1)=A$$

$$n+\frac{An}{N}-\frac{n}{N}=A$$

$$n+\frac{An}{N}=A+\frac{n}{N}$$

Multiplicando por $\frac{1}{n}$ en los 2 lados de la igualdad

$$1+\frac{A}{N}=\frac{A}{n}+\frac{1}{N}$$

$$1+\frac{A}{N}-\frac{A}{n}=\frac{1}{N}$$

$$1+A\left(\frac{1}{N}-\frac{1}{n}\right)=\frac{1}{N}$$

$$A\left(\frac{1}{N}-\frac{1}{n}\right)=\frac{1}{N}-1$$

$$\frac{1}{N}-\frac{1}{n}=\frac{1}{NA}-\frac{1}{A} \qquad (2.41)$$

Es claro que $NA \geq A$ de tal manera que al considerar inversos

$$\frac{1}{NA} \leq \frac{1}{A}$$

lo cual implica que

$$\frac{1}{NA}-\frac{1}{A} \leq 0$$

Pero el lado izquierdo de la expresión anterior es igual a $\frac{1}{N}-\frac{1}{n}$ según la expresión (2.41) por lo que

$$\frac{1}{N}-\frac{1}{n} \leq 0$$

$$\frac{1}{N} \leq \frac{1}{n}$$

De donde finalmente se obtiene que

$$n \leq N$$

Lo anterior sucede para toda n y N en los naturales

2.4 Caso con ley de distribución normal

Supóngase que se tienen $X_1, X_2, \ldots X_n$ variables aleatorias normales independientes e idénticamente distribuidas con media θ y varianza σ^2. La distribución de $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ es normal cuando las variables aleatorias X_i son normales (Canavos, 1991). Por otro lado la desviación estándar de $\hat{\theta}$ esta dada por la expresión (2.35) de modo que la variable aleatoria $\dfrac{\hat{\theta}-\theta}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$ se distribuye normal con media 0 y varianza 1 por lo que para una probabilidad $1-\alpha$ fija, se pueden encontrar 2 números reales denotados como $-Z_{\alpha/2}$ y $Z_{\alpha/2}$ tales que:

$$P\left(-Z_{\alpha/2} \le \frac{\hat{\theta}-\theta}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \le Z_{\alpha/2}\right) = 1-\alpha \qquad (2.42)$$

donde:

$-Z_{\alpha/2}$ representa un número real tal que, la probabilidad de que los valores de la variable aleatoria normal estándar sean menores que este valor es $\alpha/2$, y

$Z_{\alpha/2}$ representa un número real tal que, la probabilidad de que los valores de la variable aleatoria normal estándar sean mayores que este valor es $\alpha/2$

Se sabe que para cualquier variable aleatoria se cumple que:

$$P\left(\left|\hat{\theta}-\theta\right| \leq d\right)=1-\alpha$$

La expresión anterior se puede escribir como sigue

$$P\left(-d \leq \hat{\theta}-\theta \leq d\right)=1-\alpha$$

En donde $\hat{\theta}-\theta$ es una variable aleatoria que puede dividirse entre el número $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ y entonces se tiene que

$$P\left(\frac{-d}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \leq \frac{\hat{\theta}-\theta}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \leq \frac{d}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}\right)=1-\alpha \qquad (2.43)$$

Ahora considerando que en (2.42) y (2.43) se trata de la misma variable aleatoria y la misma probabilidad es posible igualar los intervalos de forma que

$$-Z_{\alpha/2}=\frac{-d}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \qquad (2.44)$$

$$Z_{\alpha/2}=\frac{d}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \qquad (2.45)$$

En donde al despejar a n de (2.44) o (2.45) se obtiene que[2]:

$$\boxed{n = \frac{\sigma^2\left(\frac{Z_{\alpha/2}}{d}\right)^2}{1+\frac{1}{N}\left[\sigma^2\left(\frac{Z_{\alpha/2}}{d}\right)^2-1\right]}} \qquad (2.46)$$

donde:

σ^2 es la varianza de cualquiera de las variables X_i

d es la desviación máxima del estimador y el valor real θ.

$Z_{\alpha/2}$ representa un número real tal que, la probabilidad de que los valores de la variable aleatoria normal estándar sean mayores que este valor es $\alpha/2$

n representa el tamaño de la muestra o bien el número de individuos que es necesario seleccionar de la población de estudio.

2.5 Caso en que las variables aleatorias tienen una distribución Bernoulli y Binomial

Al sustituir en (2.40) el valor de $\sigma^2 = PQ$ para el caso en donde cada una de las X_i tiene una distribución de tipo Bernoulli se obtiene que:

[2] El despeje es análogo al realizado en la sección anterior sólo hay que sustituir k por $Z_{\alpha/2}$

$$n=\frac{PQ\left(\frac{k}{d}\right)^{2}}{1+\frac{1}{N}\left[PQ\left(\frac{k}{d}\right)^{2}-1\right]} \tag{2.47}$$

donde:

PQ es la varianza de cualesquiera de las variables aleatorias $X_1, X_2, \cdots, X_n$.

d es la desviación máxima del estimador y el valor real θ.

k es un número real determinado con base en la probabilidad $1-\alpha$,

P representa la proporción o porcentaje de interés

Q es el complemento de P es decir $Q=(1-P)$

n representa el tamaño de la muestra o bien el número de individuos que es necesario seleccionar de la población de estudio y,

N representa el número de individuos de la población de estudio.

Y (2.47) es la fórmula para el cálculo del tamaño de muestra para cuando las variables X_i con $i=1,\ldots,n$ tienen asociada una función de distribución Bernoulli

Para determinar el máximo en la igualdad (2.47) para una confianza y precisión fijas se expresa la ecuación como función de P y se deriva dicha función para encontrar los puntos críticos

Expresando a (2.47) como función de P y haciendo $A=\left(\frac{k}{d}\right)^{2}$ se tiene

$$n = f(p) = \frac{PQA}{1+\frac{1}{N}[PQA-1]} = \frac{[P(1-P)]A}{1+\frac{1}{N}[[P(1-p)]A-1]}$$

$$= \frac{[P-P^2]A}{1+\frac{1}{N}[(P-P^2)A-1]}$$

Se obtiene la derivada de la función y se iguala a cero para encontrar el valor crítico:

$$f'(P) = \frac{(A-2AP)\left(1+\frac{AP}{N}-\frac{AP^2}{N}-\frac{1}{N}\right)-\left(AP-AP^2\right)\left(\frac{A}{N}-\frac{2AP}{N}\right)}{\left(1+\frac{AP}{N}-\frac{AP^2}{N}-\frac{1}{N}\right)^2} = 0$$

Después se despeja a P de la expresión anterior:

$$(A-2AP)\left(1+\frac{AP}{N}-\frac{AP^2}{N}-\frac{1}{N}\right) = \left(AP-AP^2\right)\left(\frac{A}{N}-\frac{2AP}{N}\right)$$

$$A+\frac{A^2P}{N}-\frac{3A^2P^2}{N}-\frac{A}{N}-2AP+\frac{2A^2P^3}{N}+\frac{2AP}{N}=\frac{A^2P}{N}-\frac{3A^2P^2}{N}-\frac{2A^2P^3}{N}$$

$$A-\frac{A}{N}-2AP+\frac{2AP}{N}=0$$

$$1-\frac{1}{N}-2P+\frac{2P}{N}=0$$

$$1-\frac{1}{N}+P\left(\frac{2}{N}-2\right)=0$$

$$P\left(\frac{2}{N}-2\right)=\frac{1}{N}-1$$

$$P=\frac{\frac{1}{N}-1}{\left(\frac{2}{N}-2\right)}=\frac{\frac{1}{N}-1}{2\left(\frac{1}{N}-1\right)}=\frac{1}{2}$$

El siguiente paso consiste en obtener la segunda derivada y evaluarla en el valor crítico $P=\frac{1}{2}$:

Sea

$\omega=(A-2AP)$

$\lambda=\left(\frac{A}{N}-\frac{2AP}{N}\right)$

$\psi=\left(1+\frac{AP}{N}-\frac{AP^2}{N}-\frac{1}{N}\right)$

$\varphi=(AP-AP^2)$

$$f''(P)=\frac{\left\{[\omega\lambda+(-2A)\psi]-\left[\varphi\left(\frac{-2A}{N}\right)+\omega\lambda\right]\right\}\psi^2-[\omega\psi-\varphi\lambda]2\psi\lambda}{\psi^4}$$

De modo que

$$f''\left(\frac{1}{2}\right)=\frac{\left(-2A-\frac{A^2}{N}+\frac{1}{2}\frac{A^2}{N}+\frac{2A}{N}\right)-\left(\frac{-A^2}{N}+\frac{1}{2}\frac{A^2}{2\cdot N}\right)}{\left(1+\frac{1}{2}\frac{A}{N}-\frac{1}{4}\frac{A}{N}-\frac{1}{N}\right)^2}=\frac{\frac{2A}{N}-2A}{\left(1+\frac{1}{2}\frac{A}{N}-\frac{1}{4}\frac{A}{N}-\frac{1}{N}\right)^2}$$

$$=\frac{2A\left(\frac{1}{N}-1\right)}{\left(1+\frac{1}{2}\frac{A}{N}-\frac{1}{4}\frac{A}{N}-\frac{1}{N}\right)^2}<0$$

Es importante aclarar que el resultado anterior sucede cuando $N>1$ ya que no tiene sentido llevar a cabo un muestreo sobre una población de tamaño uno.

En el caso en que las variables aleatorias tienen una distribución Binomial al sustituir en (2.40) el valor de $\sigma^2=mPQ$ se obtiene que:

$$\boxed{n=\frac{mPQ\left(Z_{\alpha/2}/d\right)^2}{1+\frac{1}{N}\left[mPQ\left(\frac{Z_{\alpha/2}}{d}\right)^2-1\right]}} \tag{2.48}$$

Y (2.48) es la fórmula para el cálculo del tamaño de muestra para cuando las variables X_i con $i=1,\dots,n$ tienen asociada cada una función de distribución Binomial

A continuación se obtiene el máximo para la expresión (2.48) como función de P

Para fines de notación se hace $A = m\left(\frac{k}{d}\right)^2$ entonces la expresión (2.48) queda

$$n = f(p) = \frac{PQA}{1 + \frac{1}{N}[PQA - 1]} = \frac{[P(1-P)]A}{1 + \frac{1}{N}[[P(1-p)]A - 1]}$$

$$= \frac{[P - P^2]A}{1 + \frac{1}{N}[(P - P^2)A - 1]}$$

Y se resuelve de forma análoga al caso en que las variables tienen una distribución Bernoulli

CAPITULO 3

APLICACIONES

3.1 Encuesta de salud reproductiva en población derechohabiente

El diseño muestral de la encuesta tiene como objetivo general proporcionar datos, que permitan determinar en la población derechohabiente del Instituto Mexicano del Seguro Social de régimen ordinario las características sociodemográficas y el estado actual de la fecundidad, los derechos sexuales y reproductivos, la detección y manejo de la infertilidad, la salud materna y la infantil, la planificación familiar, el conocimiento de enfermedades de transmisión sexual incluido el VIH/SIDA, el conocimiento y la práctica sobre la detección de los cánceres cérvico-uterino y de mama, patrones de violencia doméstica, la atención al climaterio y la menopausia así como la calidad de los servicios. Para dar respuesta a estos objetivos y con el fin de incorporar la perspectiva de género en la salud reproductiva, se definieron las siguientes poblaciones de estudio:

Población 1

- Mujeres que fueran derechohabientes del IMSS, y
- Que su edad se encontrara entre los 12 y los 54 años.

Población 2

- Hombres derechohabientes del IMSS y

- Que su edad estuviera comprendida entre los 12 y 59 años

Población 3

- Los hogares en los que residían los hombres y las mujeres de las poblaciones 1 y 2

3.1.1 Tamaño de la muestra

Datos oficiales del sistema de registro del IMSS indican que el total de la población de estudio 1 era en 1998 de aproximadamente 2 millones por lo que en este caso se puede considerar que la población es de tamaño infinito. La fórmula para el cálculo del tamaño de muestra es la siguiente:

$$n = \sigma^2 \left(\frac{Z_{\alpha/2}}{d} \right)^2 \qquad (3.1)$$

El cálculo del tamaño de muestra se hizo de tal forma que garantiza estimaciones confiables y precisas del promedio de hijos nacidos vivos y del porcentaje de uso de métodos anticonceptivos entre las mujeres de la población de estudio.

En los cálculos se consideraron estimaciones del área urbana, ya que las y los derechohabientes del IMSS del régimen ordinario, generalmente residen en este tipo de áreas.

Datos de la Encuesta Nacional de la Dinámica Demográfica de 1992 indican que el promedio de hijos nacidos vivos de las mujeres mexicanas en edad fértil del área urbana era de 2.259 y que su varianza era de 0.0004. El cuadro 1 contiene los cálculos del tamaño de muestra considerando diferentes confiabilidades, precisiones y el estimador de la varianza

CUADRO 1
TAMAÑO DE MUESTRA REQUERIDO PARA ESTIMAR EL PROMEDIO DE HIJOS NACIDOS VIVOS SEGÚN CONFIABILIDAD Y PRECISION

PRECISION	CONFIABILIDAD		
(Por diez mil)	90%	95%	99%
8	1681	2401	4160
10	1076	1537	2663
15	478	683	1183

Como se puede observar el tamaño de muestra aumenta en función de la confiabilidad al considerar la precisión fija, así como para una confiabilidad fija el tamaño de muestra aumenta en función de la precisión. A manera de ejemplo, según los datos del cuadro se requiere tomar una muestra de 683 mujeres que se estén en edad para procrear hijos es decir, en edad fértil, que en lo referente a su estado civil estén unidas y que sean derechohabientes del IMSS para que la estimación del promedio de hijos nacidos vivos tenga a lo más un error de 15 por diez mil y una confiabilidad del 95%

Con respecto a la prevalencia, las estimaciones según datos de la Encuesta Nacional de Planificación Familiar, indican que en México en 1995 el porcentaje de uso de métodos anticonceptivos en mujeres en edad fértil unidas del área urbana fue de 71.3%. Con base en este nivel de uso y utilizando también la fórmula (3.1), se calculó el tamaño de muestra según diferentes confiabilidades y precisiones. En el cuadro 2 se pueden observar los resultados

CUADRO 2
TAMAÑO DE MUESTRA REQUERIDO PARA ESTIMAR EL PORCENTAJE DE USO DE METODOS ANTICONCEPTIVOS SEGÚN CONFIABILIDAD Y PRECISION

PRECISION	CONFIABILIDAD		
%	90%	95%	99%
10	55	79	136
5	220	314	545
3	612	873	1513

Los cuadros tales como el cuadro 1 y 2 permiten decidir el número de elementos en la muestra con base en la confiabilidad, precisión y el costo en el muestreo de dichos elementos.

3.2 Conteo Rápido para las elecciones presidenciales del año 2000

El objetivo general del conteo rápido fue proporcionar información confiable y precisa acerca de las tendencias de las votaciones en mucho menos tiempo que el Instituto Federal Electoral (IFE). El objetivo especifico fue estimar el porcentaje de votos a favor de cada uno de los candidatos presidenciales.

La población de estudio fue el total de electores que había en el padrón electoral hasta el 2 de julio del año 2000. Datos oficiales del IFE indican que el tamaño de la población de estudio era de 45,266,098 electores.

Para alcanzar el objetivo se diseñó una muestra representativa de la población de estudio. El tamaño de muestra se calculó con base en el presupuesto asignado a la investigación. De acuerdo con ello se determinó muestrear sólo 1200 casillas de un total de 115,146. Con base en cifras oficiales del IFE se obtuvo que el promedio de electores por casilla es de 393 electores, dicha cifra se consideró como el número real de electores en cada una de las casillas; entonces el tamaño de muestra esperado para estimar el porcentaje de interés fue de 471,600 electores. Por lo que en este caso el problema fue determinar la confiabilidad y precisión asociadas con este tamaño de muestra.

3.2.1 Cálculo de la precisión

Dado que el objetivo es estimar un porcentaje, se consideró el caso en el que las variables aleatorias tienen una distribución Bernoulli; además como la población de estudio es tan grande, puede considerarse como infinita y como no es posible conocer el valor de la varianza real se utilizó la siguiente fórmula

$$n = \frac{1}{4}\left(\frac{Z_{\alpha/2}}{d}\right)^2$$

Sustituyendo $n = 471{,}600$ y despejado la precisión d se tiene que

$$d = \frac{1}{2}\frac{Z_{\alpha/2}}{\sqrt{471{,}600}}$$

Como se puede observar la precisión para un tamaño de muestra fijo está en función de la confiabilidad. Utilizando esta fórmula y diferentes confiabilidades se calcularon diferentes precisiones asociadas con este tamaño de muestra. En el siguiente cuadro se presentan los resultados

CUADRO 3

PRECISION REQUERIDA PARA ESTIMAR LAS TENDENCIAS DE LAS VOTACIONES PRESIDENCIALES DADO UN TAMAÑO DE MUESTRA FIJO

CONFIABILIDAD	VALOR NORMAL ASOCIADO $\left(Z_{\alpha/2}\right)$	PRECISION
90%	1.64	.0012
95%	1.96	.0014
99%	2.58	.0019

Como se puede observar en el cuadro, el estimador pierde precisión entre más confiable es, de tal manera que para una confiabilidad de 90% la precisión es de 12 por diez mil, mientras que para una confiabilidad de 99% la precisión es de 19 por diez mil, es decir, mientras más confiable es el estimador menos preciso es y viceversa.

3.3 Encuesta sobre la fecundidad reciente en mujeres indígenas

La fecundidad reciente se define como el número de hijos nacidos vivos que tiene una mujer un año antes de la fecha de la encuesta, por lo que el objetivo de la encuesta fue medir el promedio de hijos nacidos vivos en las mujeres en edad fértil.

3.3.1 Tamaño de la muestra

Para calcular el tamaño de muestra es importante tener claro que un nacido vivo, es uno de los dos posibles resultados del embarazo y que en el transcurso de un año una mujer sólo puede tener 2 hijos nacidos vivos. En términos probabilísticos la situación anterior se modela considerando al resultado del embarazo como una variable aleatoria de Bernoulli con los 2 posibles resultados: nacido vivo o bien nacido muerto y el número de nacidos vivos un año antes de la encuesta es la suma de las 2 variables aleatorias de Bernoulli, es decir:

$$Y = X_1 + X_2$$

donde

Y representa el número de hijos nacidos vivos que tiene una mujer de la población de estudio un año antes de la encuesta,

X_1 es la variable aleatoria de Benroulli asociada con el primer embarazo, y

X_2 es la variable aleatoria de Bernoulli asociada con el segundo embarazo

En este contexto la variable Y tiene una distribución Binomial y dado que la población de estudio es lo suficientemente grande para considerarse como de tamaño infinito se consideró la siguiente fórmula para el cálculo del tamaño de muestra.

$$n = 2PQ\left(\frac{Z_{\alpha/2}}{d}\right)^2$$

Como no se conocía el valor de P ni se tenían estudios previos de este tipo se calculó el tamaño de muestra utilizando $P = 1/2$ así como diferentes confiabilidades y precisiones. En el siguiente cuadro se pueden observar los resultados:

CUADRO 4
TAMAÑO DE MUESTRA REQUERIDO PARA ESTIMAR LA FECUNDIDAD RECIENTE SEGÚN CONFIABILIDAD Y PRECISION

PRECISION	CONFIABILIDAD		
%	90%	95%	99%
10	134	192	333
5	538	768	1331
3	1494	2134	3698

De acuerdo con los datos del cuadro, se requieren 768 mujeres en edad fértil para que el estimador de la fecundidad reciente tenga una confiabilidad de 95% y una precisión de 5%.

3.4 Encuesta sobre estudiantes de nivel medio superior

El diseño muestral de la encuesta tuvo como objetivo general proporcionar datos que permitieran conocer las causas por las que los estudiantes de nivel medio superior de la región oriente del estado de Morelos no ingresan a la universidad de Cuautla y como objetivo específico estimar el porcentaje de estudiantes que desean ingresar a la misma universidad. Para dar respuesta a estos objetivos se definió la siguiente población de estudio:

- Estudiantes de nivel medio superior del estado de Morelos
- Que estudian actualmente en alguno de los planteles de las siguientes modalidades: CBTA, COBAEM, CETIS, CBTIS, PREFECO, CONALEP, UAM, PARTICULAR y CECYT, y
- Que egresarán en el año 2001

Datos de la Secretaría de Educación Pública del estado de Morelos indican que actualmente existen de manera oficial 3,321 alumnos distribuidos en 29 planteles clasificados en alguna de las modalidades antes mencionadas. Por lo que, el tamaño de la población de estudio si la tasa de no-egreso es cero es precisamente 3,321 estudiantes.

Actualmente existen 350 estudiantes en la universidad Cuautla. Si se supone que este año ingresará este mismo número y que egresarán de la población de estudio definida anteriormente entonces una aproximación al valor del porcentaje de estudiantes que desean ingresar a la universidad de Cuautla es de 10.5% que es el porcentaje que representa 350 de 3,321.

3.4.1 Tamaño de la muestra

Un primer cálculo del tamaño de la muestra se hizo bajo el supuesto de que la población de estudio es tan grande que se puede considerar como de tamaño infinito. Bajo este supuesto y utilizando el valor de $P = 10.5\%$ se calculó el tamaño de muestra según diferentes precisiones y confiabilidades. En el siguiente cuadro se presentan los resultados:

CUADRO 5
TAMAÑO DE MUESTRA REQUERIDO PARA ESTIMAR EL PORCENTAJE DE INGRESO SEGÚN CONFIABILIDAD Y PRECISION

PRECISION	C O N F I A B I L I D A D		
%	90%	95%	99%
10	25	36	62
5	101	144	246
1	2528	3610	6159

Como se puede observar para confiabiliades de 95% o del 99% y una precisión del 1% el tamaño de muestra es mayor que la población de estudio por lo que en este caso se utilizó la siguiente formula:

$$n = \frac{PQ\left(\frac{k}{d}\right)^2}{1 + \frac{1}{N}\left[PQ\left(\frac{k}{d}\right)^2 - 1\right]}$$

Calculando nuevamente los tamaños de muestra ajustados con el tamaño de la población de estudio y usando los mismos valores de precisión y confiabiliad se tienen los siguientes resultados:

CUADRO 6
TAMAÑO DE MUESTRA REQUERIDO PARA ESTIMAR EL PORCENTAJE DE INGRESO SEGÚN CONFIABILIDAD Y PRECISION

PRECISION	CONFIABILIDAD		
%	90%	95%	99%
10	91	178	733
5	338	613	1763
1	2454	2822	3208

Es importante observar que esta fórmula arroja tamaños de muestra que son menores a la población de estudio en cualquier caso.

3.5 Encuesta de la calidad de los servicios

El objetivo general de la encuesta es medir la calidad de los servicios que otorga la Aseguradora Hidalgo a las personas que demandan uno de los servicios que están incluidos en la póliza del seguro de gastos médicos mayores.

Para medir este objetivo es necesario tener presente que en la medición de la calidad de cualquier servicio médico, se tienen que considerar necesariamente al prestador de los servicios, al usuario de los servicios y a las unidades médicas en donde se presta el servicio. La compañía otorga el servicio a través de una red de 100 hospitales privados que están asociados y distribuidos en la ciudad de México. Lo anterior permitió definir las siguientes poblaciones de estudio:

- Los hospitales privados de la ciudad de México que pertenecen a la red contratada por la Aseguradora Hidalgo

- Los médicos contratados por estos hospitales que otorgan los servicios que están cubiertos por el seguro de gastos médicos mayores

- Las enfermeras que apoyan a los médicos en la prestación de los servicios cubiertos por el seguro de gastos médicos mayores, y

- Los usuarios que han sido atendidos en alguno de estos hospitales por alguno de los servicios que cubre el seguro de gastos médicos mayores.

A continuación se presentan los objetivos específicos en este caso:

Estimar el porcentaje de hospitales que tienen instalaciones adecuadas para la prestación de los servicios

Estimar el porcentaje de prestadores de servicios (médicos y enfermeras) que otorgan servicios de buena calidad

Estimar el porcentaje de usuarios de los servicios que opinan que el prestador de los servicios les da un servicio de buena calidad

3.5.1 Tamaño de muestra

El cálculo del tamaño de muestra de hospitales se calculó con base en la confiabilidad y precisión que se desea tengan los estimadores, mientras que la muestra de prestadores de servicios y de usuarios se calculó con base en los números esperados por hospital. Dado que no se tenían estudios previos similares a éste, para el cálculo del tamaño de muestra de hospitales se utilizó la siguiente fórmula para poblaciones finitas pequeñas.

$$n = \frac{PQ\left(\frac{k}{d}\right)^2}{1+\frac{1}{N}\left[PQ\left(\frac{k}{d}\right)^2-1\right]}$$

Los resultados se pueden observar en el siguiente cuadro.

CUADRO 7
TAMAÑO DE MUESTRA PARA EVALUAR LA CALIDAD DE LOS SERVICIOS
SEGUN PRECISION Y CONFIABILIDAD

PRESICION	CONFIABILIDAD		
	90	95	99
10	72	83	96
5	91	95	99
1	100	100	100

Según los datos del cuadro, para que el porcentaje de hospitales que tienen instalaciones adecuadas para la prestación de los servicios tenga una confiabilidad de 95% y una precisión de 5% se necesitan 95 hospitales. Si la confiabilidad baja a 90%, así como la precisión a 10% el tamaño de muestra baja a 72 hospitales. Es importante notar según la fórmula, que si se desea una precisión del 1% se requiere medir a los 100 hospitales, sin embargo, es importante dejar claro que en este caso al tomar a los 100 hospitales, se tiene un censo y no hay error en las estimaciones ni confiabilidad alguna.

CONCLUSIONES

- Existen distintos casos para el cálculo del tamaño de muestra, los cuales en general se clasifican según si la población es muy grande o pequeña. En cada uno de estos casos los supuestos teóricos utilizados son diferentes.
- El uso de las expresiones para el cálculo del tamaño de muestra sin considerar los supuestos teóricos a partir de las cuales se obtuvieron puede llevar consigo errores en los resultados obtenidos.
- Cada una de las distintas expresiones para el cálculo del tamaño de muestra indican que éste es una función de la confiabilidad, la precisión y de la varianza de las variables aleatorias asociadas con los elementos de la población de estudio.
- La precisión y la confiabilidad las establece el investigador, sin embargo, la varianza es una característica propia de la población de estudio de interés, de tal forma que para poder calcular el tamaño de muestra se tiene que tener un conocimiento del valor de esta o aproximaciones mediante estudios previos o bien pruebas piloto. De lo contrario el tamaño de muestra no se puede calcular salvo en casos muy particulares.
- Si las variables asociadas con la población de estudio tienen una distribución Bernoulli o Binomial, entonces es factible encontrar el valor máximo posible que puede tomar la varianza y con dicho valor proceder al cálculo del tamaño de muestra.
- En ocasiones el tamaño de muestra se calcula con base en el tiempo y/o el presupuesto que se tienen para llevar a cabo la

encuesta. En este caso, dado el tamaño de muestra, se puede calcular la confiabilidad y/o la precisión asociados con este tamaño de muestra.

- La expresión para el cálculo del tamaño de muestra en el caso de poblaciones pequeñas asegura que la n muestral es menor o igual que la N poblacional.
- El papel del actuario dentro de la sociedad actual debe ser valorado no sólo en el campo de la administración de riesgos, ya que su preparación, le permite aplicar los conocimientos teóricos adquiridos durante su formación en cualquier área del estudio. En este sentido, el área de trabajo del actuario es muy grande, ya que muchos conocimientos de la carrera de actuaría pueden ser aplicados a cualquier campo de trabajo. En particular en el uso correcto de la teoría del muestreo para la obtención de información valiosa en el análisis y la toma de decisiones.

BIBLIOGRAFÍA

AMIDEM: Academia mexicana de investigación en demografía medica, AC., (1987), **Glosario de terminología en población**, Impreso en México, 80p.

Azorin F.,(1967), **Curso de muestreo y aplicaciones**, EditorialTolle, Lege Aguilar, Impreso en Madrid: Aguilar, 375 p.

Bartlett L, (1985), **El Universo y el Doctor Einstein**, Fondo de Cultura Económica, Impreso en México, Trad. Carlos Imaz, 103 p.

Bhat B.R., (1981), **Modern Probability Theory: An Introduction Text Book**, John Wiley and Sons, New York, Chichester, Brisbane, Toronto, Printed in India At Peorl Ofifset Press, New Delhi, 256 p.

Cardenas, H. y coautores, (1973), **Algebra superior**, Editorial Trillas, Impreso en México, 322 p.

Cochram W. G. (1993), **Técnicas de Muestreo,** Ed. CECSA, Impreso en México, 513 p.

Gnedenko B.V., (1968), **The Theory of Probability**, Chelsea Publishing Company New York, N.Y., Printed in U.S.A., 529 p.

INEGI, (1999), **Encuesta Nacional de la Dinámica Demográfica,** Impreso en México, 634p.

Kish L., (1982), **Muestreo de Encuestas**, Editorial Trillas, Impreso en México, Trad. Ricardo V. Cruz López, 736 p.

Leithold L., (1982), **El cálculo con geometría analítica**, Editorial Harla, Impreso en México 1319 p

Mendez I., (1981), **Modelos Estadísticos Lineales**, Consejo Nacional de Ciencia y Tecnología, Impreso en México, 139 p.

Mood Graybill B., (1985), **Introduction to theory of statistics**, Editorial Mc Graw Hill, Printed in Singapore, 654 p.

Parzen E., (1987), **Teoría Moderna de probabilidades y sus aplicaciones**, Editorial Limusa, Impreso en México, Trad. Edmundo Berumen Torres, 509 p.

Secretaría de Salud, (1994), **Norma Oficial Mexicana de los Servicios de Planificación Familiar**, Impreso en México, 64p.

Scheaffer R. y coautores, (1993), **Elementos de Muestreo**, Grupo Editorial Iberoamérica, Impreso en México, Trad. Dr. Gilberto Rendón S., Dr. José R. Gómez A., 321 p.

Stewart J.,(1991), **Calculus**, Brooks/Cole Publishing Company, California

Thomas/Finney,.(1987), **Cálculo con geometría analítica**,
Editorial Addison-Wesley Iberoaméricana,
Wilmington, Delaware, E.U.A.,
Trad.Manuel López Mateos, 1181 p.

GLOSARIO

Salud reproductiva: Es el estado de completo bienestar físico, mental y social y no solamente la ausencia de enfermedad durante el proceso de reproducción, así como del ejercicio de la sexualidad. En otras palabras la salud reproductiva es la capacidad de los individuos y de las parejas para disfrutar de una vida sexual y reproductiva satisfactoria, saludable y sin riesgos, con la absoluta libertad de decidir de una manera responsable y bien informada sobre el número y el espaciamiento de sus hijos.

Encuesta demográfica: Es una metodología para recolectar información confiable, precisa y representativa de una población de estudio.

Población derechohabiente del régimen ordinario del IMSS: Se refiere al total de personas afiliadas que tienen derecho a los servicios que proporciona el Instituto Mexicano del Seguro Social.

Características demográficas: Son características que permiten medir datos relacionados con la fecundidad, mortalidad, morbilidad y migración.

Características sociales: Características de la persona que permiten medir el desarrollo de una sociedad, por ejemplo: escolaridad, ocupación equidad de genero y derechos reproductivos (derecho a embarazarse, tener hijos, etc.) entre otras.

Características sociodemográficas: Es la suma de las características sociales y demográficas y su relación entre sí.

Fecundidad: Es la procreación efectiva (hijos nacidos vivos) de un individuo, pareja o población.

Planificación Familiar: Es el derecho de toda persona a decidir de manera libre, responsable e informada, sobre el número y el espaciamiento de sus hijos y a obtener la información especializada y los servicios idóneos.

El ejercicio de este derecho es independiente del género, la edad y el estado social o legal de las personas.

Población: Aunque en la terminología estadística la palabra población designa cualquier conjunto de unidades distintas (como sinónimo de universo), es usada comúnmente para designar el conjunto de personas que componen un pueblo o nación, es decir, el conjunto de habitantes de un cierto territorio.

Hogar: Unidad doméstica formada por una o más personas unidas o no por lazos de parentesco, que residen habitualmente en la misma vivienda y se sostienen de un gasto común para la alimentación, es decir, que comparten un mismo gasto para la comida.

Edad Fértil: Se refiere al la edad en que la mujer es capaz de procrear hijos.

Nacido vivo: Es la expulsión o extracción completa del cuerpo de la madre, independientemente de la duración del embarazo, de un producto de la concepción que, después de esta separación, respira o manifiesta cualquier otro signo de vida, tal como palpitación del corazón, pulsación del cordón umbilical o contracción efectiva del algún músculo sometido a la acción de la voluntad, haya o no haya sido cortado el cordón umbilical y esté o no adherido a la placenta.

Mujeres unidas: En los estudios demográficos se hace referencia al estado civil de las mujeres en edad fértil, clasificándolas en diferentes grupos. Así, se habla frecuentemente de las mujeres unidas o también actualmente unidas, para referirse a las casadas más las que viven o vivían en unión libre o consensual en un momento dado.

Nivel de instrucción: Último grado o año de estudio aprobado por la población de 5 años y más años de edad en alguno de los siguientes niveles del Sistema Educativo Nacional:

Nivel Básico. Comprende la educación primaria.

Nivel Medio Básico. Comprende la educación secundaria en sus diferentes modalidades y las carreras técnicas con primaria terminada.

Nivel medio superior. Tiene como antecedente inmediato la secundaria, abarca la educación preparatoria general y los bachilleratos técnicos, carreras técnicas terminales y la normal básica

Nivel superior. Tiene como antecedente inmediato la preparatoria o el bachillerato, abarca licenciatura, normal, carreras técnicas y posgrado.