CENTRO DE FORMACIÓN TÉCNICA

LOTA ARAUCO

Documento de Trabajo Nº 13

DETERMINACIÓN DEL TAMAÑO DE UNA MUESTRA

Juan A. Campodonico Saieh

Lota Mayo 2005

INDICE

1.-Introducción.. 3

2.-Determinación del tamaño de la muestra.......................... 4

2.1. Cálculo del tamaño de la muestra para variables cualitativas.......... 4

2.2. Cálculo del tamaño de la muestra para variables cuantitativas........13

3.- Observaciones generales.. 15

4.- Bibliografía... 16

5.- Anexos... 17

INTRODUCCIÓN

En el diseño y desarrollo de diversos trabajos que deben preparar, presentar y/o defender los alumnos y egresados del CFT Lota Arauco, tales como: en informes de practicas laborales, elaboración de planes, proyectos o programas, trabajos de investigación aplicada y en sus trabajos de título, uno de los problemas recurrentes con que se encuentran, es la determinación **del tamaño** que deben tomar de una muestra estadística significativa, muestras que deben ser obtenidas de una determinada población o universo y que luego deben utilizar para calcular algunas variables claves de dichos trabajos.

En la bibliografía existente al respecto, este tema se desarrolla a veces, de forma **poco clara**, en otras de forma **no muy precisa** y en muchas ocasiones, en función de variables que son **muy difíciles** de calcular ó determinar.

Este esfuerzo, surgido de la experiencia de siete años de trabajos con alumnos del CFT Lota Arauco, pretende presentar una solución **práctica, sencilla y clara** que permita calcular el tamaño de una muestra estadística en función del tamaño de la población afectada por el proyecto, estudio, plan ó trabajo de título en curso en cualquier carrera del Centro.

El sólo objetivo de este trabajo es ofrecer a alumnos y docentes del Centro una metodología **clara, precisa y aplicada** al tema de la **determinación de una muestra** y que permita luego a los estudiantes tener argumentos suficientes para justificar ciertas extrapolaciones y conclusiones que generalmente se derivan del uso de este instrumento de estadística básica.

Determinación del Tamaño de la Muestra.

Una parte fundamental para realizar un estudio estadístico de cualquier tipo es utilizar las propiedades y características de una **muestra válida** que permitan luego hacer afirmaciones que afectan o inciden sobre el conjunto de una población, grupo ó universo en general.

Dos de estas características básicas tienen que ver principalmente con el **tamaño** de la muestra y con **la manera** de obtenerla.

A continuación desarrollaremos dos métodos que nos servirá para determinar la relación entre el tamaño de la muestra y el tamaño de la población.

A.- Método Nº 1. Cálculo del tamaño de la muestra para variables cualitativas.

Estas variables cualitativas pueden estar contenidas en instrumentos tales como encuestas, cuestionarios, en determinados fenómenos sociales, etc.

En este método para calcular el tamaño de una muestra hay que tomar en cuenta tres factores o conceptos claves, a saber **:**

a.- **El porcentaje de confianza** (**Z**) ó grado de confianza con el cual se quiere luego generalizar los resultados obtenidos.

El porcentaje de confianza es el porcentaje de seguridad que existe para generalizar los resultados obtenidos. Esto quiere decir que si se toma un valor de la muestra igual al del 100% de la población o grupo involucrado, ello equivale a decir o a afirmar que **no existe duda en los resultados obtenidos de la supuesta muestra. Pero este hecho** implicaría estudiar **toda la población considerada en el estudio**, lo que generalmente es imposible de realizar ó bien ello tiene un costo muy alto. Entonces se acepta para una muestra determinada un grado de confianza menor, comúnmente este es de un **95%** [1]

1 "Estadística" .Murray y Spiegel; 1998

b.- **El porcentaje de error E** que se pretende aceptar de la muestra equivale a tomar un riesgo al aceptar ó rechazar una hipótesis que puede ser falsa ó verdadera. Si el riesgo a tomar es de 0%, entonces la muestra debe ser del mismo tamaño de la población.

Como ya dijimos que esto es muy difícil de realizar, normalmente conviene tomar un cierto riesgo de equivocarse. En consecuencia el porcentaje de error es el error que se esta dispuesto a aceptar para estimar un determinado parámetro de la población o grupo estudiado. **Comúnmente se acepta como porcentaje de error entre el 4 % y el 6%.**[2]

c **.- La variabilidad** es la probabilidad de ocurrencia con éxito **(p)** ó la probabilidad de que no ocurra **(q)** la variable estudiada ya sea en alguna investigación anterior ó en un ensayo previo. Hay que considerar que **p** y **q** son complementarios, es decir, su suma es igual a 1.

La formula general que relaciona el tamaño de la muestra **(n)** con el tamaño de la población **(N)** es la siguiente [3] **:**

$$\boxed{n = \frac{Z^2 (p)(q) N}{E^2 N + Z^2 (p)(q)} \qquad (1)}$$

Donde :

n : es el tamaño de la muestra.
Z : es el grado de confianza.
p : es la variabilidad.
q : 1-p
N : es el tamaño de la población.
E : el error muestral.

2 www.uaq.mx./matematicas/estadistica/xu5.com
3 “ www.uaq.mx./matematicas/estadistica/xu5.com

d. Análisis de las variables de la ecuación (1) :

El valor de **Z** se determina en función del porcentaje de confianza que se le otorga a la muestra y el se obtiene de la **tabla de probabilidades Normal**[4], Algunos de estos valores son los siguientes**:**

Porcentaje de confianza	92 %	95 %	98 %
Valor de Z	1,75	1,96	2,33

La variabilidad **p** de acuerdo a la bibliografía existente esta variable se puede estimar de acuerdo a datos ó experiencias históricas en alguna prueba previa**.**

Como esta variable no es conocida ó difícil de calcular entonces determinaremos, usando la ecuación **(1),** para qué valor de **p,** el valor de la muestra **n** es máximo.

Al respecto se debe tener en cuenta la siguiente tabla de probabilidades y gráfico Nº1, para diferentes valores de **N**.

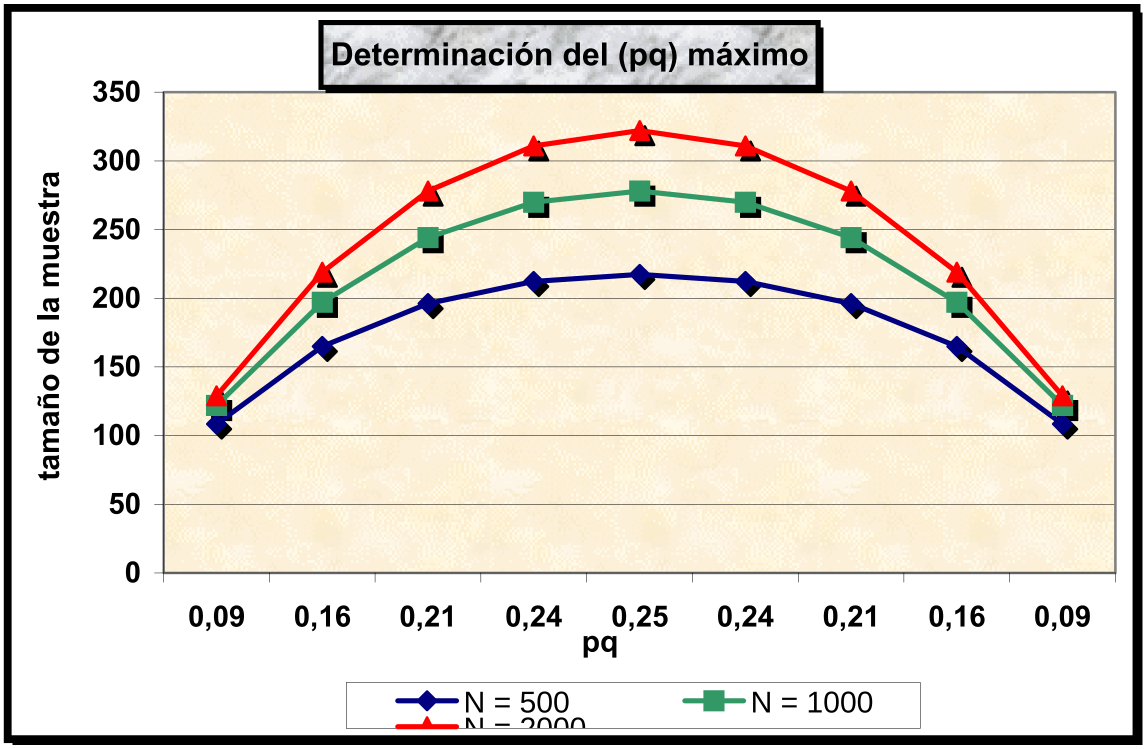
Tabla Nº 1.

p	**pq**	**n (N =500)**	**n (N=1000)**	**n (N=2000)**
0,1	0,09	108,4	122	129
0,2	0,16	164,9	197	219
0,3	0,21	196,2	244	278
0,4	0,24	212,3	270	311
0,5	0,25	217,3	278	322
0,6	0,24	212,3	270	311
0,7	0,21	196,2	244	278
0,8	0,16	164,9	197	219
0,9	0,09	108,4	122	129

4. **"**Estadística para Administración y Negocios". M.L.Berenson

A partir de esta tabla se puede construir el siguiente gráfico**:**

Gráfico Nº 1



De acuerdo a lo observado en este gráfico Nº1 el valor con el cual se obtiene el **n** máximo corresponde a un **pq** = 0,25.

El porcentaje de error **E** se tomará, como se indicó anteriormente, en un rango de 4% a 6%.

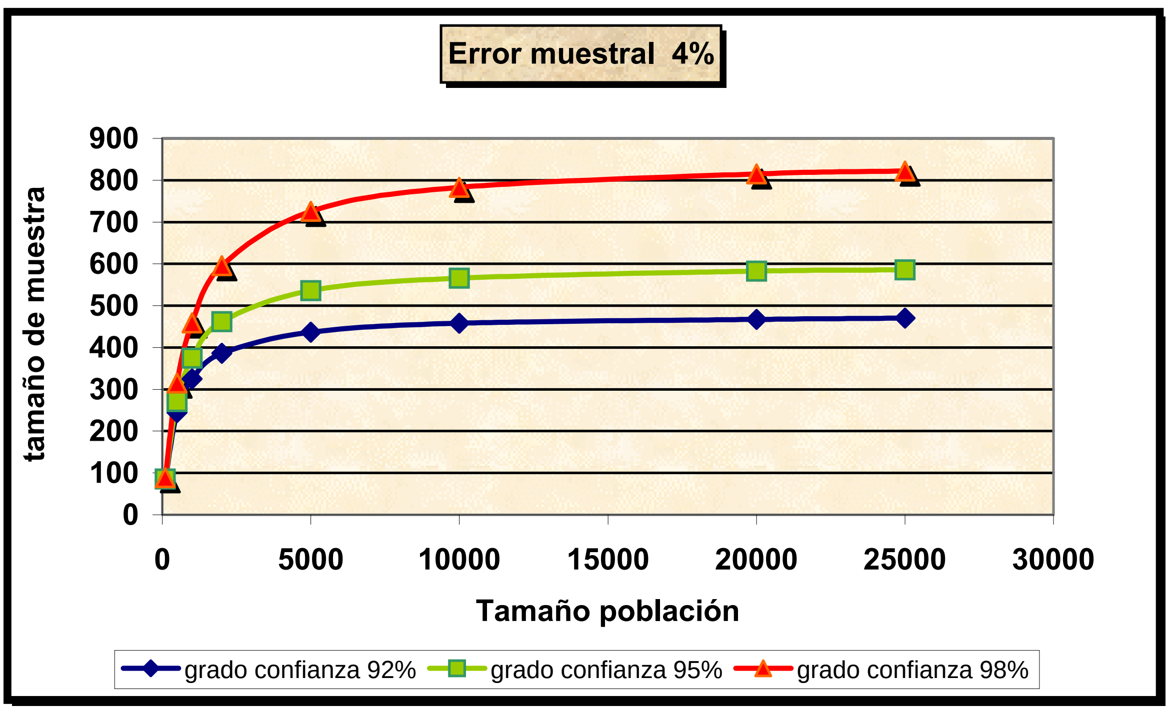
e. Análisis de la ecuación (1).-

Si en la ecuación (1) reemplazamos los valores determinados para un *grado de confianza* de 92 % ,95 % y 98% con un *error* del 4% y un *p* de 0,5 para diferentes valores de **N** obtenemos la tabla Nº 2 y el gráfico Nº 2 (ver anexo 1) **:**

Tabla Nº 2

N	n (Z = 92%)	n (Z= 95%)	n (Z= 98%)
100	83	86	89
500	245	273	315
1000	324	375	459
2000	386	462	596
5000	437	536	727
10000	457	566	783
20000	467	583	815
25000	470	586	822

Gráfico Nº 2

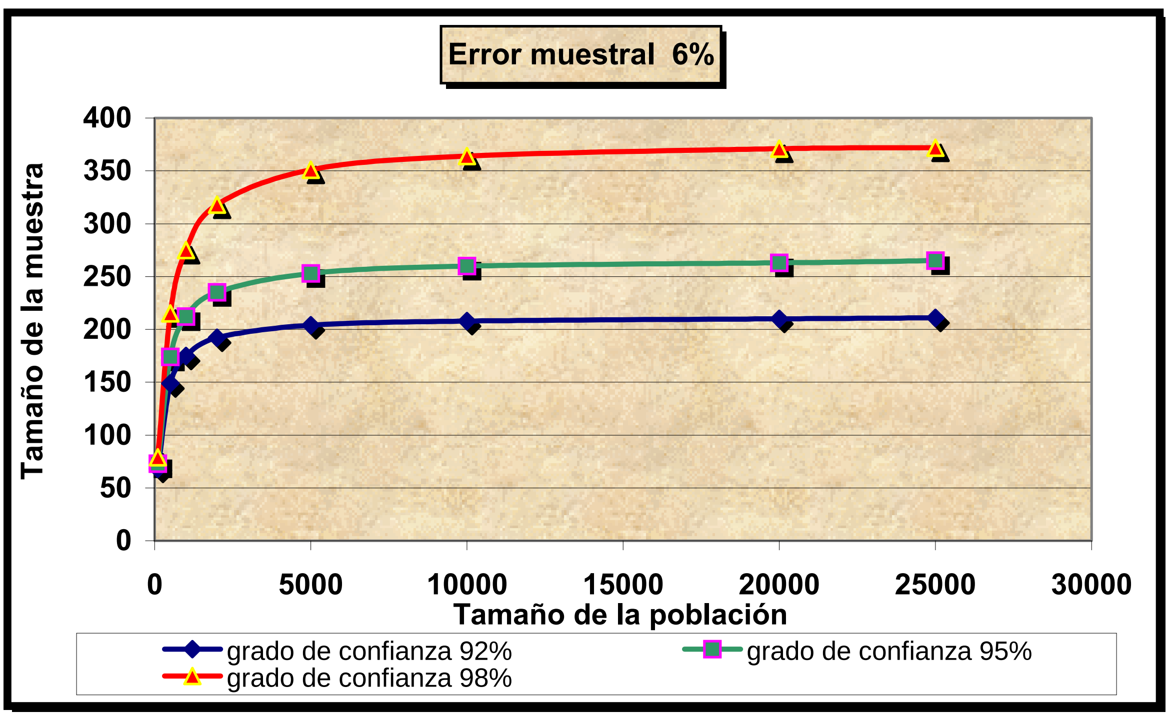


Realizando el mismo análisis anterior con los mismos grados de confianza pero con un error del 6 % obtenemos la Tabla Nº 3 y el gráfico Nº 3 **:**

Tabla Nº 3.

N	n (Z = 92%)	n (Z= 95%)	n (Z= 98%)
100	83	73	79
500	245	174	215
1000	324	212	275
2000	386	235	318
5000	437	253	351
10000	457	260	364
20000	467	263	371
25000	470	265	372

Gráfico Nº3



Conclusiones.-

De todo lo analizado anteriormente podemos concluir que en la ecuación (**1)** podemos reemplazar el producto **pq** por 0,25 y tomar un **E** (error muestral) del 5% y nos queda una ecuación más simple y de más fácil aplicación **:**

$$n = \frac{0{,}25\, Z^2\, N}{0{,}0025\, N + 0{,}25\, Z^2} \qquad (2)$$

Ejemplos de cálculo de esta ecuación se muestran en el anexo Nº 1

B.- Método Nº 2. Cálculo del tamaño de la muestra para variables cuantitativas.

En este caso, la ecuación para determinar el tamaño de la muestra **n** es [5] **:**

$$n = \frac{Z^2 \sigma^2}{E^2} \qquad (3)$$

Donde **:**

n : es el tamaño de la muestra.
Z : es el grado de confianza.
E : el error muestral.
σ : la desviación estándar

Los valores de **Z** y **E** se determinan como se explicó en la primera parte: el valor de **Z se obtiene** de la **tabla de probabilidades Normal** para un determinado grado de confianza y un **E** del 5 %.

Para determinar el valor de **σ** se debe conocer el valor de la desviación estándar de la variable y en otros casos pueden estar disponibles datos pasados (históricos) de la población que se pueden extrapolar para estimar la desviación estándar actual.

---

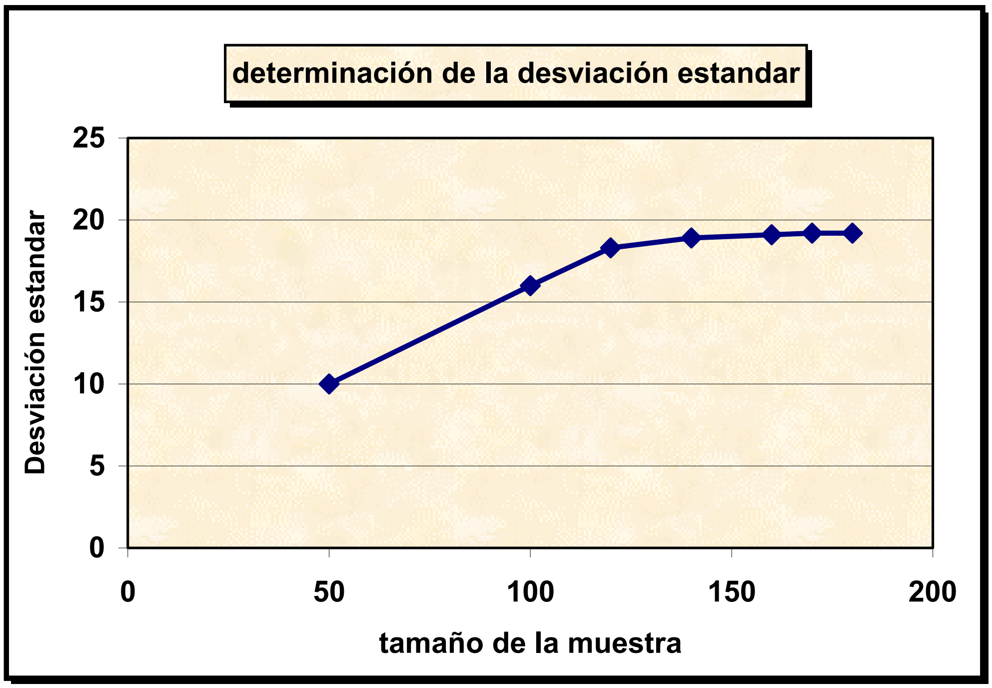
5 Estadística" .Murray y Spiegel

Si la **σ** no se conoce, que es lo más frecuente, para su determinación se usa el sistema *iterativo* utilizando el siguiente procedimiento[6] :

• primeramente se toma una pequeña muestra, $\mathbf{n_0}$, que se llama **muestra piloto,** con ella se evalúa la **σ** y con este valor se calcula un **n** en la ecuación (3).El nuevo valor , $\mathbf{n_1}$ obtenido nuevamente con ese se extrae una segunda muestra de ese tamaño de la población y se determina una nueva **σ** a esa muestra y se aplica de nuevo la ecuación (3), obteniendo un nuevo $\mathbf{n_2}$ como muestra para la siguiente iteración.

• se plantea que a medida que aumenta **n** el valor de la **σ** tiende a estabilizarse alrededor de cierto valor por lo que llegará el momento en que se encuentre el tamaño de la muestra conveniente. (ver gráfico Nº5)

Gráfico Nº 5.-



6 www.ilustrados.com

Observaciones generales

1.- Para poblaciones mayores de 5000 unidades se observa en el gráfico Nº 2 y 3 que la variación del tamaño de la muestra, en función del tamaño de la población, **es poco significativa.**

2.- Para poblaciones mayores de 1000 unidades se observa que tiene gran importancia en la determinación del tamaño de la muestra el **grado de confianza** que se acepta.

No ocurre lo mismo para poblaciones menores de 1000 en el cual el **grado de confianza no es tan significativo.**

3.- Para poblaciones menores de 1000 observamos que a medida que aumenta el tamaño de la población (**N**) *aumenta* el tamaño de la muestra a medida que *disminuye* el valor de **E** ,con diferencias entre un 15% hasta un 40% entre ambos valores. Pero como el rango estimado para el error es pequeño (4% a 6%) podemos considerar un error estándar intermedio (**E** = 5 %) y este no va a afectar mayormente el cálculo de la muestra.

Del análisis efectuado para determinar el tamaño de una muestra de una población, que se expresa en su forma más simple en la ecuación (2) , es necesario recordar que el cálculo del tamaño de la muestra **n** se puede ajustar para determinar un tamaño optimo modificando el error muestral **E** y/ó el grado de confianza **Z** dentro de rangos aceptables que en el desarrollo de las ecuaciones se deben indicar.

Bibliografía.-

" Estadística" Murray y Spiegel, 2º edición. 1998.-

" Estadística para Administración y Negocios" M.L. Berenson & Levine.1991.-

www.uaq.mx./matematicas/estadistica/xu5.com

www.ilustrados.com

www.psico.uniovi. es

A N E X O S

ANEXO Nº 1.-

Los ejemplos que a continuación se presentan fueron obtenidos de Trabajos de Tesis de alumnos del CFT Lota-Arauco

$$n = \frac{Z^2 (p)(q) N}{E^2 N + Z^2 (p)(q)} \qquad (1)$$

Ej.1 : Determinar el número mínimo de personas a encuestar de una población de 420 personas.[1]
Si tomamos un grado de confianza del 95% que de acuerdo a la *tabla normal* le corresponde un valor de **Z** = 1,96 ,un error muestral **E** del 6% **.**Reemplazando estos valores en la ecuación (1) para un **pq** = 0,25 obtenemos **:**

$$n = \frac{(1,96)^2 (0,25) (420)}{(0,06)^2 (420) + (1,96)^2 (0,25)}$$

**n = ** 163 personas encuestadas

1 " Estudio sobre la participación ciudadana en la comuna de Lota 1996-2000" Pág.52-57. Erika Alarcón M. y Roxana Mátamala V. Enero 2001.

Ej. 2 : El tamaño de la población **N** es de 8.531 personas [2] ,como no se menciona el grado de confianza ni el error muestral **E** ,asumiremos los valores de un **E** = 6% con un **Z** = 1,96 (95%). Reemplazando estos datos en la ecuación (1) :

$$\mathbf{n} = \frac{(0{,}25)(1{,}96)^2\ (8.531)}{(0{,}06)^2\ (8.531) + (0{,}25)(1{,}96)^2}$$

**n = ** 259 personas encuestadas.

---

2 "Estudio para la creación de un sistema de recaudaciones en el consultorio rural Isabel Jiménez R. de Tirua" Pág. 25-27. Susana Antilao G.y Ginette Saavedra P. Octubre 2000.

Ej.: Se quiere determinar la *nota promedio de ingreso* y su *ingreso per capita* de los alumnos de la sección diurna del CFT Lota Arauco inscritos en el 1er semestre 2005.[3] **.**

Tomaremos un **E** = 6% con un grado de confianza del 95% (**Z** = 1,96).de una población total de 593 alumnos.Como no se tiene referencia ni datos históricos de la desviación estándar (σ) a usar en la ecuación (3) se calculará la σ usando el método iterativo a partir de una *muestra piloto*

$$\mathbf{n} = \frac{Z^2 \ \sigma^2}{E^2} \qquad \mathbf{(3)}$$

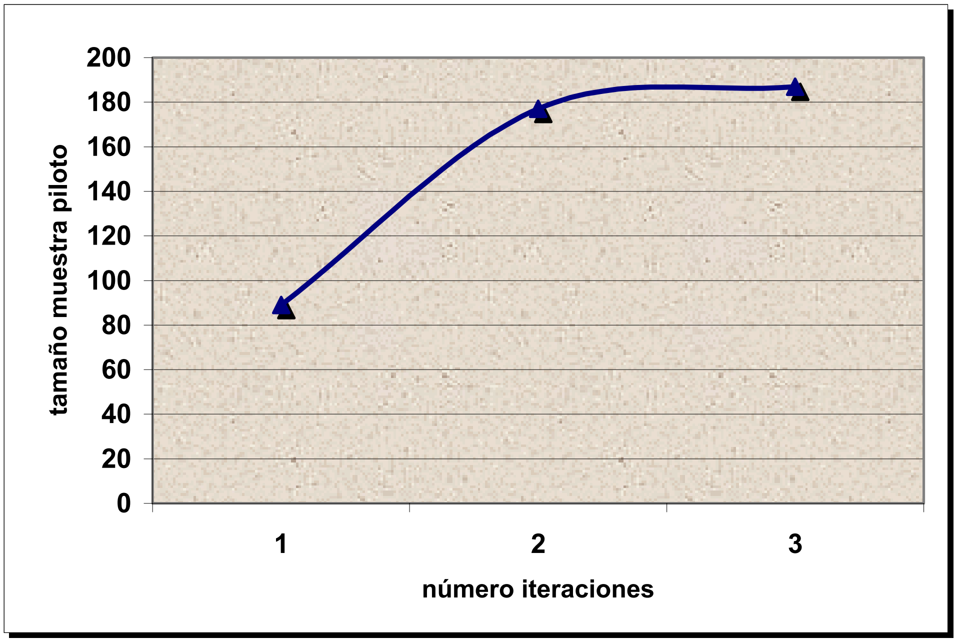
Tomaremos una muestra piloto pequeña $\mathbf{n}_0$ = 89 alumnos (equivalente al 15% de la población)[4] elegidos al azar de las diferentes carreras.Conocidas las notas de ingresos de 89 alumnos se calculo la desviación estándar, σ = 0,407.Con este valor de σ se determina un nuevo valor de **n** y así sucesivamente (ver tabla A y gráfico A)

Tabla A

Muestra piloto	Desviación estándar
89	0,407
177	0,419
187	0,415

3 Datos proporcionados por el área de Comunicaciones
4 ver anexo nº 2

Gráfico A.-



Como se observa en la Tabla de datos y su correspondiente gráfico el valor de la desviación estándar se hace constante para una muestra de 187 individuos. Conocido el tamaño mínimo (187 alumnos) se eligen al azar esa cantidad de alumnos y se calcula las variables de ***nota de ingreso*** y su ***ingreso per capita***. De la información proporcionada por el área de Comunicaciones obtenemos :

Promedio de notas = 5,56

Promedio de ingreso/per capita = $ 49.175.

ANEXO Nº 2

ESTADÍSTICA

Juan A. Campodonico S.

En la mayoría de los textos de estadística en el calculo del tamaño muestral,se usa la siguiente formula **:**

$$n = \frac{Z^2\ \sigma^2}{E^2}$$

El problema mayor que tiene la determinación de una muestra estadística deriva de la estimación de la desviación estándar **σ.**

Para su estimación se plantean dos alternativas , a saber, **:**

a) estimar esta variable de datos históricos anteriores para determinar el tamaño de la muestra que necesitamos.Pero aparece la siguiente pregunta ¿cómo se obtuvo ese dato?,en algunos casos mediante una investigación anterior.¿Y si la investigación anterior no usó una muestra probabilística?,si fue así implica un margen de error desconocido.

   Pero aquí no terminan las dificultades, ya que si aceptamos una **σ** de un estudio anterior ,tenemos que precisar a que variable en estudio le corresponde la **σ** ya que la muestra sería " representativa" sólo con relación a esa variable y no a otras.

b) A través de una muestra piloto.En este caso se plantea tomar una pequeña muestra de la población y en esta calcular la correspondiente desviación estándar **σ** y mediante iteración ("ensayo y error") se determina el tamaño de la muestra.
   La pregunta que surge es cual será el tamaño de la muestra piloto ó que porcentaje de la población tendrá esta muestra piloto?

   Para resolver este problema se decidió determinar cual sería el tamaño optimo de la muestra piloto.
   Para esto se tomaron diferentes porcentaje de población **N** (alumnos del 1 er semestre 2005 inscritos en el CFT Lota Arauco) y se determino en cada caso la **σ** [1] sabiendo que la muestra que tenga una **σ** mínima correspondería a la muestra piloto optima.

---

1 Datos proporcionados por área Comunicaciones.

El desarrollo se muestra en la Tabla 1 y los correspondientes gráficos nº 1,2 y 3.

Tabla Nº 1

		N = 406	N = 593	N = 999
muestra nº	**Porcentaje de N (%)**	**Desviación estándar**	**Desviación estándar**	**Desviación estándar**
1	2	518×10^{-3}	499×10^{-3}	501×10^{-3}
2	5	500×10^{-3}	476×10^{-3}	489×10^{-3}
3	8	467×10^{-3}	419×10^{-3}	462×10^{-3}
4	10	473×10^{-3}	394×10^{-3}	441×10^{-3}
5	12	470×10^{-3}	358×10^{-3}	432×10^{-3}
6	15	465×10^{-3}	407×10^{-3}	439×10^{-3}
7	18	479×10^{-3}	392×10^{-3}	458×10^{-3}
8	20	499×10^{-3}	400×10^{-3}	450×10^{-3}

Grafico Nº 1

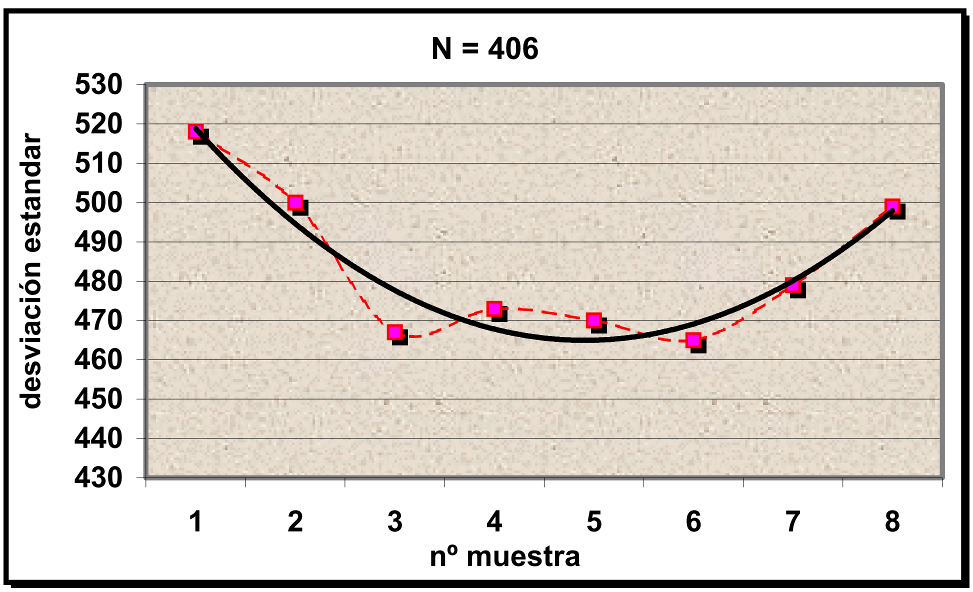


Grafico Nº 2

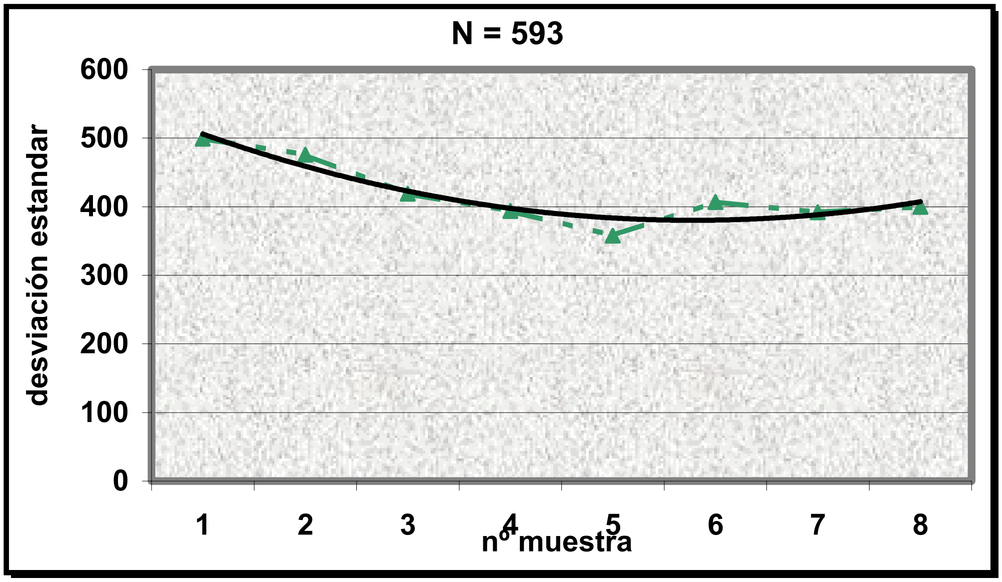
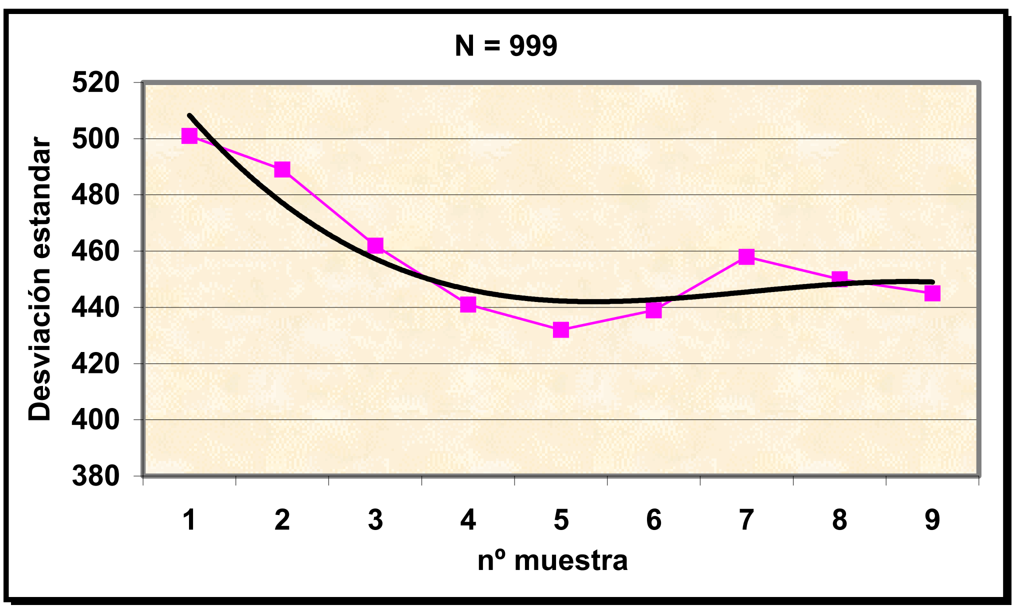


Grafico Nº 3



De los gráficos nº 1,2 y 3 se observa que en las muestras nº 5 equivalente a una muestra piloto entre el 12% y el 15% de la población esta el tamaño optimo a tomar de la muestra piloto.