1885

UNAM

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO
ESCUELA NACIONAL DE INGENIEROS

ESTUDIO ESTADISTICO DEL TRANSITO DE VEHICULOS EN LAS CARRETERAS NACIONALES

TESIS

QUE PARA OBTENER EL TITULO DE

INGENIERO CIVIL

PRESENTA EL PASANTE

ADOLFO OCAMPO NAVARRO

MEXICO 1958

TESIS CON FALLA DE ORIGEN

A mis Padres.

A mis Maestros.

Al Departamento de muestreo de la Dirección General de Estadística.

El estudio del tránsito de vehículos en las carreteras tema de este trabajo, fué hecho a iniciativa de la Asesoría Técnica de la Sub-secretaría de Obras Públicas de la S. C. O. P., bajo el atinado cargo del Sr. Ing. Fernando Espinosa a cuya clara visión se debe el tratamiento estadístico al problema.

Se hace patente el mas sincero agradecimiento a la Srita. Ana María Flores Avila, jefe del Departamento de Muestreo de la Dirección General de Estadística por las innumerables facilidades otorgadas sin las cuales no hubiera sido posible efectuarse este trabajo. Así mismo, se agradece a los M. M. en C. C. del Departamento de Muestreo Srita. María Guadalupe Lomelí y Sr. Remigio Valdéz Gámez por sus valiosas contribuciones al presente trabajo.

CAPITULO I

CONSIDERACIONES GENERALES

SOBRE LA CAPACIDAD Y FUNCIONAMIENTO DE CAMINOS EN LA REPUBLICA MEXICANA

El buen proyecto de un camino no sólo estriba en el aspecto económico y técnico del buen trazo del camino, sino también habrá que atender al aspecto funcionamiento del camino.

A pesar de que las características de un camino y las condiciones impuestas por el tránsito, deben estar íntimamente relacionadas entre - si, su diseño y construcción se lleva a cabo sin tener en cuenta el servicio que es capaz de proporcionar a los usuarios. En la generalidad de los casos, los elementos de proyecto, tales como número y ancho de bandas de circulación, acotamientos, curvas verticales y horizontales, sobre elevaciones, pendientes, etc., se eligen de acuerdo con las condiciones del terreno, económicas u otras, más bien que para cumplir propósitos formulados de antemano para determinada clase y densidad de tránsito; de suerte que, una vez puesta en servicio una vía pública diseñada- y construida con estos lineamientos, suele resultar inadecuada.

Caso palpable lo tenemos en algunas de las carreteras Nacionales en las que a pesar de haber compaginado el factor económico con el factor técnico del trazo de curvas verticales y horizontales, al presente - resultan poco menos que obsoletas por lo que se refiere al número y calidad de vehículos que transitan por ellos.

Vistos pues los trastornos o retrasos que esto ocasiona la S.-C.O.P. se ha propuesto iniciar nuevos estudios tendientes a determinar - de una manera racional la ley de variación del tránsito carretero a fin de poder predecir mediante razonamientos e hipótesis con bases matemáticas el incremento de tránsito para un determinado período de tiempo.

Se dan casos en que el camino se proyecta para grandes capacidades que nunca se alcanzan en la práctica, mientras en otros, el grado - de saturación rebasa con mucho su capacidad, lo que supone una falta de plan adecuado para su realización.

Otra irregularidad se observa con el tránsito de vehículos de carga, que en algunos caminos es más pesado de lo que son capaces de resistir, lo cual ocasiona desperfectos de consideración.

Esto conduce a pensar en la necesidad de realizar estudios sobre el uso que puede tener un camino y establecer normas para su construcción.

En otros países, como por ejemplo Estados Unidos de Norteamérica, para construir un camino, se piensa primeramente en el servicio que va a prestar, esto es, se procura tener conocimiento del número y clase de vehículos que habrán de usar el camino durante períodos de tiempo determinados, ya sea por hora o por día, o ambos, y se pronostica el incremento que pueda haber en un número de años, elegido por lo regular desde el punto de vista económico.

Con este fin, practican aforos permanentemente en muchos puntos de su amplia red de caminos y obtienen datos tales como fluctuaciones de tránsito en las diversas horas del día y en los distintos días del año; porcentajes de cada uno de los tipos de vehículos que hay en uso, promedios de velocidades, y otros.

Por otra parte, tienen establecidas normas generales de proyecto basadas principalmente en la experiencia, que les permiten determinar las características geométricas, de trazo, alineamiento, etc., en función del tránsito de vehículos.

Conociendo las características de construcción de un camino y teniendo datos del tránsito de vehículos sobre el mismo, pueden determinar su capacidad real y así saber si es suficiente o insuficiente, o si está demasiado sobrado, y en este caso prever para qué fecha llegará a saturarse.

Así mismo, con el acopio de datos, se pueden proyectar nuevos caminos, más apegados a las necesidades de cada región, con lo que se obtienen obras más económicas y eficientes.

Consecuentemente, si nosotros deseamos conservar y mejorar correctamente nuestros caminos existentes, y si queremos construir eficientemente nuevas vías de comunicación, se hace necesario emprender la tarea de recabar datos precisos de los caminos que están actualmente en operación.

Ahora bien, estos datos no tendrán significación alguna si no contamos con especificaciones adecuadas para relacionarlos con ellas. Como guía pueden sernos de utilidad las que han sido implantadas en otros países, pero sin perder de vista el caracter y fisonomía particulares de México.

Es indudable que la tradición caminera y las condiciones económicas de algunos países les permiten fijar normas con las que obtienen grados de comodidad y calidad insuperables. En contraste, la situación económica de nuestro país, ni siquiera nos permite integrar al presente una red caminera que sea suficiente para nuestras necesidades más urgentes.- Entonces, es lógico pensar en construir el máximo posible de caminos, - aunque no sean de los llamados de primer orden.

Para conseguir este propósito se hace necesario sacrificar en - cierta proporción los requisitos de comodidad que dictan las especificaciones de otros países, sin perjuicio desde luego, de la estabilidad de los caminos, y no rebasando el límite de aquellas normas que garantizan - la seguridad de los usuarios.

Es por tanto que, si por ejemplo, adoptamos como guía las especificaciones de los Estados Unidos, deberemos emplear un criterio más liberal para su aplicación en la práctica, haciendo ajustes y concesiones - que armonicen con las condiciones apuntadas en los párrafos anteriores.-

Así por ejemplo, en los tramos de camino con mayor densidad de - tránsito, es probable que su capacidad, desde el punto de vista de las - especificaciones americanas, sea actualmente superada en un 200% o 300%. Sin embargo, se observa que las incomodidades y demoras en esos lugares no son de caracter insoportable. Mas adelante, con los aforos que se - practiquen en esos tramos, se podrá definir si la densidad de tránsito - va en ascenso y en ese caso, prever con oportunidad para cuándo sea intolerable el congestionamiento.

Respecto a los datos por recopilar, conviene clasificarlos en dos tipos, a saber:

I.- Geométricos, topográficos y climatéricos.

II.- Aforos de tránsito.

Teniendo los primeros, se puede conocer la capacidad de cada uno - de los caminos que se estudie y con los aforos se puede determinar el grado de saturación de los mismos. Si los aforos se establecen con caracter permanente, con el transcurso de los años se podrá determinar la Ley de - variación de tránsito.

Cabe aclarar que las condiciones climatéricas, tales como niebla,- lluvia, heladas, etc., de hecho influyen en la capacidad de un camino, pero no son determinantes para la "Capacidad de Diseño", toda vez que su - efecto retardador de la circulación de vehículos puede considerarse el - mismo, ya sea que se trate de un camino de dimensiones restringidas o de una supercarretera.

Sin embargo, en algunas regiones es importante conocer las condiciones climatéricas para definir ciertas características geométricas y para precisar el sistema de señales del camino, que si bien no remedian la situación desde el punto de vista de la capacidad del camino, si se hacen necesarias para su seguridad.

En efecto, si se piensa en hacer una ruta desde el punto de vista - exclusivo de su capacidad, puede resultar inadecuada, como es el caso de una región montañosa en que abunde la niebla durante una considerable parte del año. Si por ejemplo, se proyecta el camino con bandas de circulación angostas, sin tener en cuenta la reducción de visibilidad ocasionada por la niebla, el tramo en cuestión es peligroso y puede ser foco de continuos accidentes.

Para determinar la capacidad de un camino, será necesario contar - con los siguientes datos:

a).- Número y ancho de bandas de circulación.

b).- Obstáculos laterales existentes a lo largo del camino.

c).- Ancho de acotamientos.

d).- Alineamiento.

e).- Distancias de visibilidad.

Los aforos deberán suministrar las siguientes cifras:

a).- Número de vehículos por banda de circulación, que transitan - durante cada hora del día. Estos datos deben ser por un tiem po suficientemente grande para poder establecer buenas conclusiones y para juzgar sin mucho error de la frecuencia con que ocurren los períodos de máximo congestionamiento.

b).- Velocidades con que operan los vehículos.

c).- Porcentaje de cada uno de los tipos de vehículos que hacen uso de la carretera.

C A P I T U L O II

CONCEPTOS FUNDAMENTALES DE LA TEORIA DE PROBABILIDAD Y VARIABLE ALEATORIA, ESPERANZA MATEMATICA SESGO

CONCEPTOS FUNDAMENTALES DE LA TEORIA

DE PROBABILIDAD Y VARIABLE ALEATORIA

A continuación se expondrán las ideas básicas sobre Probabilidad y Variable Aleatoria, que son indispensables para el desarrollo de los elementos de Muestreo que se presentarán en el presente trabajo, como base al problema de tránsito en carreteras que se estudia.

Como la definición clásica del Concepto de Probabilidad adolece de ciertos defectos que hacen que no sea aplicable en múltiples problemas estadísticos, el punto de partida para una definición mas apropiada es la idea de Frecuencia Relativa.

Consideremos un experimento $\mathcal{E}$ en el que los resultados mas simples se representarán con: e_1, e_2, ..., etc., que serán llamados eventos elementales. Un conjunto de estos, constituyen un evento que indicaremos con A, B, ..., etc.

Si repetimos el experimento $\mathcal{E}$ bajo idénticas condiciones, N veces y si indicamos con N_A el número de veces en que se verifica el evento A, N_A nos representa la frecuencia con que se repite el evento y $\frac{N_A}{N}$ es llamada la frecuencia relativa con que se verifica el evento en las N repeticiones del evento. Se supone que dicha frecuencia relativa tiende a aproximarse a un cierto valor, a medida que N crece indefinidamente. En tal caso, a cada evento A asociado al experimento se le asocia un número real no negativo P(A) llamado su probabilidad, del cual se consideran como valores aproximados las frecuencias relativas de A, para un N suficientemente grande.

De las propiedades de las frecuencias relativas de los eventos y dado que no se precisa en que forma dichas frecuencias relativas tienden a aproximarse a las probabilidades correspondientes, se establecen como axiomas fundamentales de la Teoría de la Probabilidad las tres siguientes proposiciones:

1er. AXIOMA:

$$P(A) \text{ es } \geq 0.$$

2º. AXIOMA:

La probabilidad de un evento seguro, es decir, de un evento que necesariamente se verifica al efectuar el experimento, es uno.

3er. AXIOMA:

Si dos eventos A y B son mutuamente exclusivos, esto es, que no tienen eventos elementales en común, la $P(A+B) = P(A) + P(B)$, en donde $(A+B)$ representa al evento A o B.

De los tres axiomas anteriores se puede desarrollar la Teoría Elemental de Probabilidad, precisando incluso, a posteriori, la forma en que las frecuencias relativas tienden a las probabilidades.

(Esto se desprende de las Leyes de los Grandes Números:

Ley Débil de los Grandes Números:

$$\lim_{N \to \infty} P\left\{\left[\frac{N_A}{N} - p\right] \geq \varepsilon\right\} = 0 \quad \text{para todo} \quad \varepsilon \geq 0.$$

Ley Fuerte de los Grandes Números:

$$\lim_{N \to \infty} \frac{N_A}{N} = p \text{ con probabilidad 1.)}$$

En esta Teoría, el principio de Probabilidad compuesta:

$P(AB) = P(A) \times P(B)$ cuando A y B son independientes, se toma como una definición de independencia de eventos. En la fórmula anterior, (AB) representa el evento A y B, es decir, la verificación simultánea de ambos.

CONCEPTO DE VARIABLE CASUAL O ALEATORIA

Consideremos una función ξ que asocia a cada evento elemental e de un experimento E un valor real ξ (e). Representemos con $\{e | \xi (e) \leq x\}$ al evento que está constituido por todos los eventos elementales e para los cuales ξ (e) tiene un valor menor o igual que x. Si para todo número real x, dichos eventos tienen una probabilidad bien definida, entonces, ξ es una variable casual o variable aleatoria.

Así, por ejemplo, si $\mathcal{E}$ consiste en tirar un dado y observar la cara que queda hacia arriba, los eventos elementales son: 1, 2, 3, 4, 5 y 6.

Supongamos asociado a cada evento elemental la probabilidad de $\frac{1}{6}$. Si ξ es una función que asocia el valor 1 cuando queda par, y -1 cuando queda impar, la probabilidad $P\left\{e/\xi(e) < 0\right\} = \frac{1}{2}$ por ejemplo, es la probabilidad del evento constituido por los eventos elementales para los cuales $\xi(e)$ tiene un valor menor que cero. Dichos eventos elementales son: 1, 3 y 5 y la probabilidad de dicho evento es: $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$.

El símbolo $\left\{e/\xi(e) \leq x\right\}$ se representa en forma abreviada con $(\xi \leq x)$. La probabilidad $P(\xi \leq x)$ se representa con $F(x)$, que recibe el nombre de función de distribución de la variable aleatoria ξ.

Una variable aleatoria se clasifica de acuerdo con sus valores que puede adquirir, de las dos maneras siguientes:

1o. ξ toma sólo un número finito o infinito numerable de valores. En tal caso se dice que ξ es una variable aleatoria discreta.

Si ξ puede tomar los valores: $x_1, x_2, \ldots$ y la probabilidad

$$P(\xi = x_i) = p_i, \quad i = 1, 2, \ldots \tag{1}$$

entonces el sistema de valores (1) recibe el nombre de distribución de probabilidad de ξ.

Por supuesto que $\sum_i p_i = 1$.

No se ponen los límites de la suma, porque el límite superior puede ser finito o infinito. Se sobre entiende que la sumatoria es respecto a todos los valores que puede tomar i.

2o. ξ puede tomar valores reales cualesquiera.

Si la probabilidad $P\left\{\xi \in [a, b]\right\} = \int_a^b f(x)dx.$ donde $f(x) \geq 0$, entonces ξ es una variable aleatoria continua y $f(x)$ es su densidad de probabilidad.

En tal caso:

$$P(\xi \leq x) = \int_{-\infty}^{x} f(x)dx = F(x).$$

Tanto en el caso continuo como en el caso discreto, se tiene que:

$$\lim_{x \to -\infty} F(x) = 0 \qquad \text{y}$$

$$\lim_{x \to +\infty} F(x) = 1$$

CONCEPTO DE ESPERANZA MATEMATICA

Si ξ es una variable aleatoria, aceptemos que una función $\varphi(\xi)$ es a su vez variable aleatoria.

Definiremos el concepto de Esperanza Matemática de $\varphi(\xi)$ separadamente para el caso continuo y el caso discreto.

1er. Caso discreto:

Si ξ es una variable aleatoria discreta con distribución de probabilidad (1) $(P(\xi = x_i) = p_i \; ; \; i = 1, 2, \ldots)$, entonces la Esperanza Matemática de $\varphi(\xi)$ se define:

$$E\left[\varphi(\xi)\right] = \sum_i \varphi(x_i)p_i$$

o sea, la suma de los productos de cada valor que toma $\varphi(\xi)$ para cada valor x_i de ξ, por la probabilidad de que ξ tome dicho valor x_i.

El símbolo $E\left[\varphi(\xi)\right]$ se lee: "la esperanza matemática de fí - de kei".

COROLARIOS DE LA DEFINICION DE ESPERANZA MATEMATICA:

1.- $E(1) = E\left[(\xi)^0\right] = \sum_i p_i = 1$

2.- $E(c) = c$ $\quad$ (c = constante)

3.- $E\left[c\,\varphi(\xi)\right] = cE\left[\varphi(\xi)\right]$ $\quad$ (c = constante)

4.- $E\left[\varphi(\xi) + \psi(\xi)\right] = E\left[\varphi(\xi)\right] + E\left[\psi(\xi)\right]$

Algo mas sobre Esperanzas Matemáticas:

En particular, si $\varphi(\xi) = \xi$:

$$E(\xi) = m = \sum_i x_i p_i \qquad (2)$$

que es lo que se llama la media de la distribución de ξ, o la media de ξ.

Si n es entero no negativo, entonces:

$E(\xi)^n = \mu'_n$, recibe el nombre de momento de orden n respecto al origen.

En particular: $\mu_1' = m.$ (2')

La $E\left[(\xi - m)^n\right] = \mu_n$, y es el momento de orden n respecto — a la media.

$$E\left[(\xi - m)^2\right] = \sigma^2 \tag{3}$$

es la variancia de la distribución y $\sigma = \sqrt{\sigma^2}$ es la desviación estándar de la distribución.

Se puede ver que:

$$\mu_1 = E(\xi - m) = 0$$

$$\mu_2 = E(\xi - m)^2 = \mu_2' - m^2 \tag{4}$$

que es una relación muy útil.

2º. Caso continuo:

Si ξ es variable aleatoria continua con densidad de probabilidad $f(x)$:

$$E\left[\varphi(\xi)\right] = \int_{-\infty}^{\infty} \varphi(x)\, f(x)\, dx.$$

También, como en el caso discreto tendremos:

1.- $E(1) = 1$

2.- $E(c) = c$

3.- $E\left\{c\,\varphi(\xi)\right\} = cE\left\{\varphi(\xi)\right\}$

Y también:

$$\mu_n' = \int_{-\infty}^{\infty} \varphi(x)^n\, f(x)\, dx.$$

$\mu_1' = m$

$\mu_n = \int_{-\infty}^{\infty} \left\{\varphi(x) - m\right\}^n f(x)\, dx.$; en particular:

$\mu_1 = 0$

$\mu_2 = \sigma^2$

$\mu_2 = \mu_1' - m^2$

CONCEPTO DE SESGO

Hipótesis: Por simplicidad supondremos en el desarrollo de la teo ría que y_i es el verdadero valor de la i-ésima unidad. (Despreciaremos pues los errores de medida).

Definición: Se dice que un procedimiento de muestreo es insesgado si la media de la distribución de frecuencia de las estimaciones proporcionadas por la totalidad de las $\binom{N}{n}$ muestras posibles de tamaño n tomadas de una población N, es igual al valor de la característica que se está estimando de la población.

O sea, si la esperanza matemática de la estimación, es el valor del parámetro de la población.

Si z_i es la estimación del parámetro a partir de la muestra Q_i -- $\left\{ i = 1, 2, \ldots \binom{N}{n} \right\}$; p_i la probabilidad de selección de z_i y θ_i el valor del parámetro de la población, el procedimiento de muestreo es insesgado si:

$$\sum_i p_i z_i = \theta_i \quad \text{que es equivalente a:}$$

$$E(z_i) = \theta_i \tag{5}$$

Si

$$\sum_i p_i z_i \neq \theta_i , \quad \text{a la diferencia } B = \sum_i p_i z_i - \theta_i \tag{6}$$

se le llama Sesgo.

C A P I T U L O   III

TEORIA DEL MUESTREO PROBABILISTICO

I.- DEFINICION DE MUESTREO PROBABILISTICO.
II.- MUESTREO IRRESTRICTO ALEATORIO.
III.- MUESTREO ESTRATIFICADO.
IV.- MUESTREO SISTEMATICO.

M U E S T R E O

Hasta hace pocos años al problema de la selección de una muestra - representativa de una población se le concedía muy poca atención.

El diagnóstico de un laboratorio acerca del estado de salud de un individuo es fácil hacerse a partir de unas cuantas gotas de sangre. Este procedimiento se basa en la hipótesis de que la sangre circulante es homogénea, de tal suerte que, una gota es completamente semejante a otra y - con las mismas características.

Hipótesis de este tipo son frecuentemente correctas y las consecuencias obtenidas a partir de muestras en esos casos dan buenos resultados; pero cuando la población en estudio no es homogénea el método según el -- cual se obtenga la muestra presenta a menudo serios problemas y el estudio de la técnica que asegure una muestra digna de confianza llega a ser importante.

Desarrollaremos pues algunas ideas de la Teoría del Muestreo a fin de seleccionar después una muestra trabajable matemáticamente.

Señalemos algunas ventajas de las investigaciones por muestreo:

i) Reducción de costos.

ii) Mayor velocidad.

iii) Mayor campo de acción y aplicabilidad.

iv) Mayor acuciocidad.

MUESTREO PROBABILISTICO

Para poder desarrollar una teoría matemática un muestreo debe cumplir los siguientes requisitos:

i) Debe poderse definir el conjunto de distintas muestras Q_1, Q_2, ..., Q_n en una población determinada. Es decir, se podrá precisar qué - unidades de muestreo pertenecen a cada muestra Q_i.

ii) A toda muestra Q_i se le asociará una determinada probabilidad - de selección π_i.

iii) Q_i será seleccionada de acuerdo con su probabilidad de selección π_i asociada.

iv) El método de cálculo de la estimación a partir de la muestra debe dar una estimación única para cada muestra Q_i.

A un muestreo que cumpla con los requisitos de (i a i v) se le llama Muestreo Probabilístico. Para cualquier procedimiento de muestreo que cumpla con la definición estamos en posibilidad de calcular la distribución de frecuencia de las estimaciones que se genera si aplicamos repetidamente el procedimiento a una misma población ya que sabremos cuán frecuentemente se presentará una determinada muestra Q_i y podremos calcular exactamente la estimación a partir de los elementos de Q_i.

MUESTREO IRRESTRICTO ALEATORIO

Es un método de seleccionar muestras de n unidades de una población de N elementos, de tal suerte que, cada una de las $\genfrac{}{}{0pt}{}{N}{n}$ posibles muestras, tengan la misma probabilidad de selección.

En la práctica, una muestra en un MIA* se forma seleccionando elemento por elemento. Los elementos de la población se numeran de 1 a N y con ayuda de una tabla de números aleatorios son seleccionados los n elementos que constituye la muestra. Un procedimiento así efectuado dará igual probabilidad de selección a todas las $\binom{N}{n}$ posibles muestras. Esto es fácil de verificar.

Una propiedad del MIA es que la probabilidad de seleccionar una unidad dada de la población en cualquier selección, es igual a la probabilidad de seleccionarla en la primera selección; en efecto, suponer que: n sea el número de unidades que van a formar la muestra.

La probabilidad de que la unidad especificada sea seleccionada en la r-ésima selección será el producto de la probabilidad de no ser seleccionada en ninguna de las (r-1) selecciones anteriores y la probabilidad de ser seleccionada en la r-ésima selección.

La probabilidad de no ser seleccionada en las primeras (r-1)-ésimas selecciones es:

$$\frac{N-1}{N} \times \frac{N-2}{N-1} \times \frac{N-3}{N-2} \times \ldots \times \frac{N-\quad+1}{N-r+2} = \frac{N-r+1}{N}$$

La probabilidad de ser escogida en la r-ésima selección es:

$$\frac{1}{N-r+1}$$

∴ la probabilidad de seleccionar la unidad especificada en la r-ésima selección será:

$$\frac{N-r+1}{N} \times \frac{1}{N-r+1} = \frac{1}{N}$$ que es precisamente la probabilidad de seleccionar la unidad especificada en la primera selección.

Puesto que la unidad especificada puede quedar incluida en la muestra en cualquiera de las n selecciones, la probabilidad de que quede incluida dentro de la muestra será la suma de las probabilidades de que sea escogida en las n selecciones, o sea: * $P(\xi \in \mathfrak{F}) = \frac{n}{N}$ y puesto que este razonamiento puede hacerse para cualquiera de las N unidades de la población, se infiere que cada una de las unidades de la población tiene la misma probabilidad de ser incluida en la muestra.

Aunque a veces se ha pretendido dar esta propiedad como definición de MIA, no es tal, puesto que existen otros procedimientos que no dan la misma probabilidad de selección en la primera selección a cada unidad de la población y sin embargo la probabilidad de que cualquier unidad especificada esté incluida en la muestra es $\frac{n}{N}$.

Demostremos pues que el procedimiento descrito en párrafos anteriores para seleccionar una muestra cumple con la definición del MIA de dar igual probabilidad de selección a cualquiera de las $\binom{N}{n}$ posibles muestras.

Siendo $\binom{N}{n}$ el número de las posibles muestras, por definición el MIA implica que cada una de estas muestras tenga por probabilidad:

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

En efecto, la probabilidad de escoger una unidad cualquiera en la primera selección es por definición: $\frac{1}{N}$; habiendo seleccionado una, la probabilidad de escoger cualquiera de las unidades restantes es: $\frac{1}{N-1}$; la probabilidad de escoger una de las restantes en la tercera selección es: $\frac{1}{N-2}$ y así sucesivamente. Por lo tanto, la probabilidad de seleccionar

---

* donde ($\xi \in \mathfrak{F}$) nos simboliza: unidad ξ dentro de la muestra .

n unidades dadas cualesquiera en sucesión en un orden especificado es:

$$\frac{1}{N} \times \frac{1}{N-1} \times \frac{1}{N-2} \times \dots \times \frac{1}{N-n+1}$$

y puesto que el orden en que se escogen las unidades no es importante, - la probabilidad de cualesquiera n unidades dadas de formar la muestra está dada por:

$$\frac{n!}{N\cdot(N-1)(N-2)\ \dots\ (N-n+1)} = \frac{1}{\binom{N}{n}}$$ que nos exige la definición.

La palabra aleatorio se refiere al "procedimiento" de seleccionar la muestra mas bien que a la muestra escogida en particular; de aquí que, cualquier muestra así escogida puede ser una muestra irrestricta aleatoria, por muy poco representativa que pueda resultar.

Es conveniente aclarar que la selección de unidades puede hacerse "con reemplazo" o "sin reemplazo"; es fácil hacer ver -no se demostrará- que las propiedades anteriores se cumplen en cualquiera de los dos casos. Sin embargo, a menos que se mencione lo contrario, supondremos que el — muestreo se hace sin reemplazo de las unidades.

Estimadores y sus propiedades.

Sea:

N - el número de unidades de muestreo en la población.

y - la característica en consideración.

y_i - el valor de la característica para la i-ésima unidad.

$\bar{Y}_N$ - el valor medio de la característica de la población dado - por

$$\bar{Y}_N = \frac{\sum_{i=1}^{N} y_i}{N} \qquad (7)$$

S^2 - el cuadrado medio para la población dado por:

$$S^2 = \frac{\sum_{i=1}^{N} (y_i-\bar{Y}_N)^2}{N-1} = \frac{\sum_{i=1}^{N} y_i^2 - N\bar{Y}_N^2}{N-1} \qquad (8)$$

$V(y)$ - la variancia de una muestra tomada de la población dada - por:

$$V(y) = \frac{\sum_{i=1}^{N} (y_i-\bar{Y}_N)^2}{N} = \frac{N-1}{N} S^2 \qquad (9)$$

n - el tamaño de la muestra, o sea, el número de unidades en la muestra.

$\bar{y}_n$ - la media de la muestra dada por:

$$\bar{y}_n = \frac{\sum_{i=1}^{n} y_i}{n} \tag{10}$$

s^2 - el cuadrado medio de la muestra dado por:

$$s^2 = \frac{\sum_{i=1}^{n} (y_i-\bar{Y}_n)^2}{n - 1} \quad \frac{\sum_{i=1}^{n} y_i^2-n\bar{Y}_n^2}{n - 1} \tag{11}$$

Estimaciones insesgadas de los valores de la población

Es evidente que una estimación variará de muestra a muestra, pues to que las estimaciones son funciones de las unidades incluidas en la muestra; sin embargo, los promedios de las medias de todas las muestras y los cuadrados medios de todas las muestras, son iguales a los valores de la media y del cuadrado medio de la población respectivamente.

A las estimaciones que cumplan con la propiedad anterior, se les llama estimaciones insesgadas de los valores de la población. Algebráicamente, esto se expresa:

$$E(\bar{y}_n) = \frac{\sum^{\binom{N}{n}} \bar{y}_n}{\binom{N}{n}} = \bar{Y}_N \qquad y \tag{12}$$

$$E(s^2) = \frac{\sum^{\binom{N}{n}} s^2}{\binom{N}{n}} = S^2 \tag{13}$$

donde el símbolo E indica la esperanza matemática.

Demostraremos que un MIA proporciona estimaciones insesgadas de la media de la población y el cuadrado medio de la muestra proporciona una estimación insesgada del cuadrado medio de la población.

Variancia de Muestreo, Error Estándar y Error Cuadrático Medio.

Hemos visto que una estimada por muestreo no es exactamente el va lor de la característica de la población y mas aún, varía de estimada a estimada dependiendo de la muestra. A la diferencia entre la estimada de muestreo y el valor para la población se le llama ERROR DE MUESTREO de la estimada. Asentado esto, un requisito que se le exige al método de muestreo es que además de dar una estimación única del valor para la población, dé una medida del error de muestreo para la estimada. Nótese sin embargo que el verdadero valor del error de muestreo no puede llegarse a conocer puesto que no conocemos el verdadero valor para la población; en virtud de esto, únicamente obtenemos una medida de la magnitud media del error de muestreo de la estimada en todas las muestras posibles. Un simple promedio de los verdaderos errores de muestreo de todas las muestras posibles, será cero -como es fácil comprobarse- en el caso de estimadas insesgadas. Para subsanar esto a la magnitud media de los cuadrados de los errores de muestreo en todas las muestras se le han denominado VARIANCIA DE MUESTREO de la estimada y a su raíz cuadrada ERROR MEDIO DE MUESTREO o ERROR ESTANDAR.

Generalizando para estimadas sesgadas la Variancia de Muestreo se define como la media aritmética de los cuadrados de las diferencias entre la estimada de la muestra y la esperanza matemática de la estimada sobre todas las muestras. A su raíz cuadrada se la llama Error Estándar.

La media aritmética de los cuadrados de las diferencias entre la estimada de la muestra y el valor de la población se llama Error Cuadrático Medio.

La definición de variancia se traduce al álgebra como:

$$\sigma^2 = E\left\{x_i - E(x_i)\right\}^2 \qquad (14)$$

<u>Esperanzas Matemáticas de los estimadores de la población en un MIA</u>

a).- E.M. de la media de la muestra:

Escribamos:

$$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (15)$$

tomando esperanzas:

$$E(\bar{y}_n) = E\left\{\frac{1}{n}\sum_{i=1}^{n} y_i\right\} \tag{16}$$

donde y_i representa el valor de la i-ésima unidad de la población y la sumatoria se toma para todas las n unidades de la muestra. Numerando seriadamente las unidades en la muestra como 1, 2, ..., r, ... n, podemos escribir a (16) como:

$$E(\bar{y}_n) = E\left\{\frac{1}{n}\sum_{r=1}^{n} y'_r\right\} \tag{17}$$

donde y'_r representa ahora el valor de la unidad incluida en la muestra en la r-ésima selección; o sea:

$$E(\bar{y}_n) = \frac{1}{n}\sum_{r=1}^{n} E(y'_r) \tag{18}$$

pero por definición:

$$E(y'_r) = \sum_{i=1}^{N} p_{ir}\, y_i$$

siendo p_{ir} la probabilidad de selección una unidad y_i especificada, en la selección r-ésima. Demostramos anteriormente que esta probabilidad es $\frac{1}{N}$

$$\therefore E(y'_r) = \frac{1}{N}\sum_{i=1}^{N} y_i, \text{ pero } \frac{1}{N}\sum_{i=1}^{N} p_i = \bar{Y}_N.$$

$$\therefore E(y'_r) = \bar{Y}_N \tag{19}$$

(r = 1, e, ... , n); substituyendo (19) en (18):

$$E(\bar{y}_n) = \frac{1}{n}\sum_{r=1}^{n} \bar{Y}_N \text{ finalmente}$$

$$E(\bar{y}_n) = \bar{Y}_N \tag{20}$$

que nos dice que la media de la muestra es una estimación insesgada de la -

media de la población.

Otra manera más sugestiva de obtener este resultado es escribir - la media de la muestra en la forma:

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\left\{\sum_{i=1}^{N} \alpha_i\, y_i\right\} \tag{21}$$

donde α_i es una variable casual tal que:

$\alpha_i = 1$ si y_i está en la muestra y

$\alpha_i = 0$ si y_i no está en la muestra

Tomando esperanzas en (21)

$$E\left\{\frac{1}{n}\sum_{i=1}^{n} y_i\right\} = \frac{1}{n}\left[\sum_{i=1}^{N}\left\{E(\alpha_i)\right\} y_i\right] \tag{22}$$

ahora, por definición:

$E(\alpha_i) = 1 \times$ probabilidad de que y_i se incluya en la muestra; o sea:

$$E(\alpha_i) = 1 \times \frac{n}{N} \tag{23}$$

substituyendo (23) en (22):

$$E\left\{\frac{1}{n}\sum_{i=1}^{n} y_i\right\} = \frac{1}{n}\sum_{i=1}^{N}\frac{n}{N}\, y_i = \frac{1}{n}\,\frac{n}{N}\sum_{i=1}^{N} y_i;\ \text{pero}\ \frac{1}{N}\sum_{i=1}^{N} y_i = \bar{Y}_N$$

$$\therefore\ E\left\{\frac{1}{n}\sum_{i=1}^{n} y_i\right\} = \bar{Y}_N \tag{24}$$

<u>Esperanza matemática de la variancia de la muestra</u>.

Recordemos que:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} y_i^2 - n\bar{\bar{y}}_n^2\right\} \tag{25}$$

escribimos pues:

$$E(s^2) = E\left\{\frac{1}{n-1}\left(\sum_{i=1}^{n} y_i^2 - n\bar{\bar{y}}_n^2\right)\right\} \tag{26}$$

o sea:

$$= \frac{1}{n-1}\left\{E\sum_{i=1}^{n} y_i^2 - nE(\bar{y}_n^2)\right\} \qquad (27)$$

Obtengamos separadamente

$E\sum_{i=1}^{n} y_i^2$ y $E(\bar{y}_n^2)$; obtengamos primero $E\sum_{i=1}^{n} y_i^2$, para esto hagamos:

$$E\left\{\frac{1}{n}\sum_{i=1}^{n} y_i^2\right\} = E\left\{\frac{1}{n}\sum_{r=1}^{n} y_r'^2\right\} \qquad (28)$$

en la que y_r' representa el valor de la i-ésima unidad incluida en la r-ésima selección de la muestra.

Tomando esperanzas del 2do. miembro:

$$E\left\{\frac{1}{n}\sum_{r=1}^{n} y_r'^2\right\} = \frac{1}{n}\sum_{r=1}^{n} E(y_r'^2) = \frac{1}{n}\, nE(y_r'^2) \qquad (29)$$

pero por definición:

$$E(y_r'^2) = \frac{1}{N}\sum_{i=1}^{N} y_i^2 \qquad (30)$$

sumando y restando $\bar{Y}_N^2$ en el segundo miembro de (30)

$$E(y_r'^2) = \bar{Y}_N^2 + \frac{1}{N}\sum_{i=1}^{N}(y_i^2 - \bar{Y}_N^2) \qquad (31)$$

pero por la ec:

$$S^2 = \frac{\sum_{i=1}^{N} y_i^2 - N\bar{Y}_N^2}{N-1}$$

tenemos:

$$\frac{1}{N}\sum_{i=1}^{N}(y_i^2 - \bar{Y}_N^2) = \frac{N-1}{N} S^2 \qquad (32)$$

substituyendo (32) en (31)

$$E(y_r'^2) = \bar{Y}_N^2 + (1 - \frac{1}{N})S^2 \qquad (33)$$

substituyendo (33) en (29)

$$E\left\{\frac{1}{n}\sum_{r=1}^{n} y_r'^2\right\} = \bar{Y}_N^2 + (1 - \frac{1}{N})S^2 \qquad (34)$$

y reemplazando este valor en (28)

$$E\left\{\frac{1}{n}\sum_{i=1}^{n} y_i^2\right\} = \bar{Y}_N^2 + \left(1 - \frac{1}{N}\right)S^2 \tag{35}$$

de donde concluimos que:

$$E\sum_{i=1}^{n} y_i^2 = n\,\bar{Y}_N^2 + \left(n - \frac{n}{N}\right)S^2 \tag{36}$$

Para obtener $E(\bar{y}_n^2)$ escribamos:

$$E(\bar{y}_n^2) = E\left\{\frac{1}{n}\sum_{i=1}^{n} y_i\right\}^2 \tag{37}$$

$$= \frac{1}{n^2}\; E\left\{\sum_{i=1}^{n} y_i^2 + \sum_{i\neq j=1}^{n} y_i\, y_j\right\}$$

$$= \frac{1}{n^2}\left\{E\sum_{i=1}^{n} y_i^2 + E\sum_{i\neq j=1}^{n} y_i\, y_j\right\} \tag{38}$$

$E\sum_{i=1}^{n} y_i^2$ la hemos obtenido en la ec. (36); obtengamos $E\sum_{i\neq j=1}^{n} y_i y_j$ para ello hagamos:

$$\frac{1}{n(n-1)}\; E\left\{\sum_{i\neq j=1}^{n} y_i\, y_j\right\} = \frac{1}{n(n-1)}\, E\left\{\sum_{r\neq s=1}^{n} y'_r\, y'_s\right\} \tag{39}$$

donde y'_r y y'_s nos representan los valores de las unidades y_i y y_j seleccionadas en la r-ésima y s-ésima selección respectivamente; tomando esperanzas del segundo miembro de (39):

$$\frac{1}{n(n-1)}\, E\left\{\sum_{i\neq j=1}^{n} y_i\, y_j\right\} = \frac{1}{n(n-1)}\left\{\sum_{r\neq s=1}^{n} E\,(y'_r\, y'_s)\right\} \tag{40}$$

pero por definición:

$$E(y'_r\, y'_s) = \sum_{i\neq j=1}^{N} P_{ir}\cdot P_{js/i}\cdot y_i\, y_j \tag{41}$$

P_{ir} es la probabilidad de seleccionar y_i en la r-ésima selección

y

$P_{js/i}$ la probabilidad de selección de y_j en la s-ésima selección-una vez elegida y_i.

Recordemos que $P_{ir} = \frac{1}{N}$ y con un razonamiento análogo al hecho para obtener P_{ir}, concluimos que:

$P_{js/i} = \frac{1}{N-1}$; substituyendo estos valores en (41):

$$E(y_r' \, y_s') = \frac{1}{N(N-1)} \sum_{i \neq j = 1}^{N} y_i \, y_j \tag{42}$$

valor que substituyendo en (40) nos da:

$$\frac{1}{n(n-1)} E \left\{ \sum_{i \neq j = 1}^{n} y_i y_j \right\} = \frac{1}{N(N-1)} \sum_{i \neq j = 1}^{N} y_i y_j \tag{43}$$

Otra manera de obtener este resultado es escribiendo:

$$\frac{1}{n(n-1)} E \left\{ \sum_{i \neq j = 1}^{n} y_i \, y_j \right\} = \frac{1}{n(n-1)} \left[\sum_{i \neq j = 1}^{N} E(\alpha_i \, \alpha_j) y_i y_j \right] \tag{44}$$

en la que α_i y α_j son variables casuales del tipo antes mencionado,- la sumatoria $\sum\limits_{i \neq j}^{n}$ se extiende sobre los n(n-1) términos del producto - $y_i \, y_j$ en la muestra y $E(\alpha_i \, \alpha_j)$ la probabilidad de incluir y_i y y_j en la muestra. Obtengamos pues: $E(\alpha_i \, \alpha_j)$; escribamos:

$$E(\alpha_i \, \alpha_j) = E(\alpha_i) \cdot E(\alpha_j / \alpha_i) \tag{45}$$

siendo

$E(\alpha_j / \alpha_i) = 1 \times$ probabilidad de incluir y_j en una muestra de -- (n-1) elementos una vez seleccionado y_i de una población de (N-1).

Razonando análogamente que para $E(\alpha_i)$ tenemos:

$$E(\alpha_j / \alpha_i) = \frac{n-1}{N-1} \tag{46}$$

y

$$\therefore \; E(\alpha_i \, \alpha_j) = \frac{n}{N} \frac{n-1}{N-1} \tag{47}$$

Valor que substituyendo en (44) da:

$$\frac{1}{n(n-1)} E\left\{\sum_{i\neq j=1}^{n} y_i\, y_j\right\} = \frac{1}{N(N-1)} \sum_{i\neq j=1}^{N} y_i y_j \tag{48}$$

valor que coincide con (43).

Trabajando con

$\sum_{i\neq j=1}^{N} y_i\, y_j$; es evidente que podemos escribir:

$$\sum_{i\neq j=1}^{N} y_i y_j = \left(\sum_{i=1}^{N} y_i\right)^2 - \sum_{i=1}^{N} y_i^2 \quad \text{que es equivalente a:}$$

$$= N^2\, \bar{Y}_N^2 - N\bar{Y}_N^2 - (N-1)S^2$$

que substituyendo en (48) y simplificando:

$$\frac{1}{n(n-1)} E\left\{\sum_{i\neq j=1}^{n} y_i y_j\right\} = \bar{Y}_N^2 - \frac{S^2}{N} \tag{49}$$

$$\therefore E\left\{\sum_{i\neq j=1}^{n} y_i y_j\right\} = n(n-1)\left\{\bar{Y}_N^2 - \frac{S^2}{N}\right\} \tag{50}$$

Estamos pues en condiciones de valuar $E(\bar{y}_n^2)$ substituyendo a (36) y (50) en (38):

$$E(\bar{y}_n^2) = \frac{1}{n^2}\left[\left\{n\bar{Y}_N^2 + (n - \frac{n}{N})S^2\right\} + n(n-1)\left\{\bar{Y}_N^2 - \frac{S^2}{N}\right\}\right]$$

operando:

$$= \frac{1}{n^2}\left[n\bar{Y}_N^2 + (n - \frac{n}{N})S^2 + n(n-1)\bar{Y}_N^2 - \frac{n(n-1)}{N}\, S^2\right]$$

simplificando:

$$E(\bar{y}_n^2) = \bar{Y}_N^2 + \left(\frac{1}{n} - \frac{1}{N}\right) S^2 \tag{51}$$

Hemos obtenido pues $E\sum_{i=1}^{n} y_i^2$ y $E(\bar{y}_n^2)$ mediante los valores (36) y (51), finalmente reemplazandolos en la ec. (27):

$$E(s^2) = \frac{1}{n-1}\left\{n\bar{Y}_N^2 + (n - \frac{n}{N})S^2 - n\bar{Y}_N^2 - (1 - \frac{n}{N})S^2\right\}$$

Operando:

$$E(S^2) = \frac{1}{n-1}\left\{ n\bar{Y}_N^2 + nS^2 - \frac{n}{N}\ S^2 - n\bar{Y}_N^2 - S^2 + \frac{n}{N} S^2 \right\}$$

Simplificando:

$$E(s^2) = S^2 \qquad (52)$$

que nos dice que s^2 es una estimación insesgada de S^2.

VARIANCIA DE MUESTREO DE LA MEDIA

Es obvio que la media de una muestra es función de conjunto, en consecuencia variará de muestra a muestra; midamos pues, esa variación.

Sea $V(\bar{y}_n)$ la variancia de muestreo de la media. Entonces, por definición:

$$V(\bar{y}_n) = E\left[\left\{\bar{y}_n - E(\bar{y}_n)\right\}^2\right] \qquad (53)$$

desarrollando el cuadrado:

$$V(\bar{y}_n) = E\left[\bar{y}_n^2 + E(\bar{y}_n)^2 - 2\bar{y}_n\ E(\bar{y}_n)\right]$$ y tomando esperanzas:

$$V(\bar{y}_n) = E(\bar{y}_n^2) + \left\{E(\bar{y}_n)\right\}^2 - 2\ E(\bar{y}_n)\ E(\bar{y}_n)$$

$$V(\bar{y}_n) = E(\bar{y}_n^2) - \left\{E(\bar{y}_n)\right\}^2 \qquad (54)$$

y recordando los valores de $E(\bar{y}_n^2)$ y $E(\bar{y}_n)$ dados por (51) y (20) respectivamente y simplificando:

$$V(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2$$ que podemos escribir

$$V(\bar{y}_n) = \frac{N-n}{N} \cdot \frac{S^2}{n} \qquad (55)$$

Al factor $\frac{N-n}{N}$ se le llama "corrección por probabilidad finita" - o "multiplicador finito," debido al tamaño finito N de la población. Estudiémoslo un poco:

Si $N \to \infty$ o bien es muy grande comparado con n:

$$\frac{N-n}{N} \xrightarrow[N\to\infty]{} 1$$

y podremos considerar a la variancia de la media como la de una media obte nida de una población infinita o de una población en la que se ha hecho un muestreo con reemplazo.

En la práctica, basta con que $\frac{n}{N} \leq .05$ para que podamos despreciar el multiplicador finito en vista de que $\frac{N-n}{N} \doteq 1$ quedando la expresión de la variancia de la media como:

$$V(\bar{y}_n) = \frac{S^2}{n} \tag{56}$$

Por otra parte, el valor de S^2 generalmente no lo conocemos, únicamente tendremos una estimación s^2 de él; en consecuencia solamente podremos tener una estimación de la variancia de la media dada por la expresión:

$$V(\bar{y}_n) = \frac{N-n}{N} \cdot \frac{s^2}{n} \tag{57}$$

y de acuerdo con la definición dada de error estandar podremos escribir:

$$\widehat{E.S.}(\bar{y}_n) = \sqrt{\frac{N-n}{N}} \cdot \frac{s}{n} \tag{58}$$

o en los casos en que $\frac{n}{N} \leq .05$:

$$\hat{V}(\bar{y}_n) = \frac{s^2}{n} \tag{59}$$

y

$$\widehat{E.S.}(\bar{y}_n) = \frac{s}{\sqrt{n}} \tag{60}$$

LIMITES DE CONFIANZA

Habíamos dicho que una estimación $\bar{y}_n$ de $\bar{Y}_N$ es una función de conjunto y en general, $\bar{y}_n$ será diferente de $\bar{Y}_N$, es decir, estará afectada de un error (que puede ser positivo o negativo). Una medida de este error nos la da el error estándar mediante el cual podemos formar un criterio que de idea de la frecuencia con que se presentan errores de determinada magnitud. En efecto, la experiencia indica que aproximadamente el 68% de las veces la magnitud del error es menor que el valor del error estándar y que en un 95% de los casos la magnitud del error será menor del duplo del error estándar para poblaciones con estimadores normalmente distribuídos. (Se demuestra que aun para poblaciones con leyes de --

distribución diferentes a la normal, la media de las muestras puede aceptarse como normalmente distribuida).

Si el tamaño de la muestra n no es muy pequeño ($n \geqq 35$), N suficientemente grande ($N \geqq 95n$) y el estimador es insesgado, la frecuencia con que el error excederá en determinada proporción del error estandar de la estimación será aproximadamente igual a la probabilidad de la integral normal, con lo cual estamos en situación de localizar el verdadero valor (desconocido) de la característica de la población dentro de ciertos límites con una probabilidad determinada.

Hablando de la media, (normalmente distribuida), podemos esperar que:

$$\left| \bar{y}_n - \bar{Y}_N \right| \leq \sqrt{\frac{N-n}{Nn}}\, S$$ aproximadamente en 68 de 100 ocasiones y

$$\left| \bar{y}_n - \bar{Y}_N \right| \leq 2\sqrt{\frac{N-n}{Nn}}\, S$$ aproximadamente en 95 de 100 ocasiones.

En general, podemos esperar que la desigualdad:

$$\bar{y}_n - t_{(\alpha)}\sqrt{\frac{N-n}{Nn}}\, S \leq \bar{Y}_N \leq \bar{y}_n + t_{(\alpha)}\sqrt{\frac{N-n}{Nn}}\, S \qquad (61)$$

en la que: $t_{(\alpha)}$ es el valor de la variable normal para el valor $1 - \frac{\alpha}{2}$ en la curva de la integral de probabilidad normal, se mantenga en promedio con una probabilidad de $1 - \alpha$.

A los valores $\bar{y}_n - t_{(\alpha)}\sqrt{\frac{N-n}{Nn}}\, S$ y $\bar{y}_n + t_{(\alpha)}\sqrt{\frac{N-n}{Nn}}\, S$ se les llama "límites de confianza"; la diferencia entre ellos es el "intervalo de confianza"; $1 - \alpha$ es la probabilidad con que la desigualdad se verifica y α es el "nivel de significación".

En (61) estamos suponiendo conocido el valor de S; sin embargo, en la mayoría de los casos esto no sucede, sino que únicamente tenemos una estimación s a partir de la muestra. Se puede demostrar que $\dfrac{\bar{y}_n - \bar{Y}_N}{\sqrt{\frac{N-n}{N}} \cdot \frac{s}{\sqrt{n}}}$ sigue la distribución t de Student con n-1 grados de libertad. Llamando $t_{\alpha,(n-1)}$ el valor de t correspondiente a un nivel de significación α para (n-1) grados de libertad podemos escribir:

$$\bar{y}_n - t_{\alpha,(n-1)}\sqrt{\frac{N-n}{N}}\,\frac{s}{\sqrt{n}} \leq \bar{Y}_N \leq \bar{y}_n + t_{\alpha,(n-1)}\sqrt{\frac{N-n}{N}}\,\frac{s}{\sqrt{n}} \qquad (62)$$

Para poblaciones infinitas o bien en aquellas en las que

$\frac{N-n}{N} \doteq 1$ (61) y (62) quedan:

$$P\left\{\bar{y}_n - t_\alpha \frac{s}{\sqrt{n}} \leq \bar{Y}_N \leq \bar{y}_n + t_\alpha \frac{s}{\sqrt{n}}\right\} = 1 - \alpha \qquad (63)$$

y

$$P\left\{\bar{y}_n - t_{\alpha,(n-1)} \frac{s}{\sqrt{n}} \leq \bar{Y}_N \leq \bar{y}_n + t_{\alpha,(n-1)} \frac{s}{\sqrt{n}}\right\} = 1 - \alpha. \qquad (64)$$

que son las expresiones usuales con las que se hace una estimación por intervalo de la media de una población.

<u>Tamaño de la muestra para una precisión especificada.</u>

Uno de los principales problemas en la teoría del muestreo, es precisamente elegir el tamaño de la muestra. Hemos escrito que una muestra debe darnos una medida de su precisión y también que la precisión de la muestra es justamente función de su tamaño. La precisión pues, la especificaremos en función del error permisible para la estimación y del grado de confianza asignado para que la estimación esté dentro del intervalo permisible de error. Planteado esto, si $\epsilon \bar{Y}_N$ es el error permisible para la estimación y $1 - \alpha$ el grado de confianza, debe cumplirse:

$$P\left\{\left|\bar{Y}_N - \bar{y}_n\right| \geq \epsilon \bar{Y}_N\right\} = 1 - \alpha \qquad (65)$$

Recordando la expresión (61):

$$\bar{y}_n - t_\alpha \sqrt{\frac{N-n}{Nn}}\, S \leq \bar{Y}_N \leq \bar{y}_n + t_\alpha \sqrt{\frac{N-n}{Nn}}\, S \qquad (61)$$

podemos escribir:

$$\bar{Y}_N = \bar{y}_n + t_\alpha \sqrt{\frac{N-n}{Nn}}\, S \quad \text{o bien:} \quad \bar{Y}_N - \bar{y}_n = t_\alpha \sqrt{\frac{N-n}{Nn}}\, S \qquad (66)$$

pero

$$\left|\bar{Y}_N - \bar{y}_n\right| = \left|\bar{y}_n - \bar{Y}_N\right| = \epsilon \bar{Y}_N \qquad (67)$$

substituyendo (67) en (66):

$\epsilon \bar{Y}_N = t_\alpha \sqrt{\frac{N-n}{Nn}}\, S$ elevando al cuadrado a ambos miembros y operando:

$$\frac{\epsilon^{2}\bar{Y}_N^2}{t_\alpha^2 \; S^2} = \frac{N-n}{Nn} \quad \text{o sea:} \quad \frac{\epsilon^{2}\bar{Y}_N^2}{t_\alpha^2 \; S^2} = \frac{1}{n} - \frac{1}{N} \quad \text{o mejor:}$$

$$\frac{\epsilon^{2}\bar{Y}_N^2}{t_\alpha^2 \; S^2} + \frac{1}{N} = \frac{1}{n} \quad \text{finalmente } n = \frac{1}{\dfrac{\epsilon^{2}\bar{Y}_N^2}{t_\alpha^2 \; S^2} + \dfrac{1}{N}} \tag{68}$$

expresión que es equivalente a:

$$n = \frac{\dfrac{t_\alpha^2}{\epsilon^2} \; \dfrac{S^2}{\bar{Y}_N^2}}{1 + \dfrac{1}{N} \; \dfrac{t_\alpha^2}{\epsilon^2} \; \dfrac{S^2}{\bar{Y}_N^2}} \tag{69}$$

Al cociente $\frac{S}{\bar{Y}_N}$ se le conoce como "coeficiente de variación" de la población y nos da idea de la homogeneidad de la población.

Obsérvese en (69) que el valor de n depende pues del coeficiente de variación, valor que se desconoce por no conocerse S y $\bar{Y}_N$; sin embargo, puede estimarse burdamente y en consecuencia (69) también será una estimación burda del tamaño de la muestra.

Una mejor estimación de n sería razonando a partir de la expresión (62) lo que nos conduciría a:

$$n = \frac{\dfrac{t_{\alpha,(n-1)}^2}{\epsilon^2} \; \dfrac{S^2}{\bar{Y}_N^2}}{1 + \dfrac{1}{N} \; \dfrac{t^2_{\alpha,(n-1)}}{\epsilon^2} \; \dfrac{S^2}{\bar{Y}_N^2}} \tag{70}$$

en la que como objeción se presenta el desconocimiento de $t_{\alpha,(n-1)}$ que depende precisamente de n que andamos buscando.

Una manera pues de estimar n es usar (69) con la que tendría una n', con esta n' obtener $t_{\alpha,(n'-1)}$ que nos da una idea de $t_{\alpha,(n-1)}$ y calcular n con $t_{\alpha,(n'-1)}$ aplicando (70). Como t_α es menor que $t_{\alpha,(n-1)}$ (69) nos da una subestimación n' de n y $\therefore t_{\alpha,(n'-1)} < t_{\alpha,(n-1)}$; la corrección obvia que se ocurre es aumentar n en la relación

$$\frac{t^2_{\alpha,(n'-1)}}{t_\alpha^2}$$

pero para los fines de la práctica esta corrección es despreciable en la mayoría de los casos a menos que n sea pequeña.

Vistas las anteriores objeciones la manera práctica de calcular n es seleccionar una "muestra piloto", digamos de n elementos, que nos dará una estimación s_1^2 de S^2; entonces la muestra adicional para estimar el valor de la población con la precisión especificada (suponiendo que N es suficientemente grande), será $n-n_1$ donde:

$$n = \frac{t^2_{\alpha,(n_1-1)}}{\epsilon^2} \cdot \frac{s_1^2}{\bar{Y}_N^2} \tag{71}$$

Se demuestra que la n así obtenida satisface la condición

$$P\left\{\left|\bar{y}_n - \bar{Y}_N\right| \geq \epsilon \bar{Y}_N\right\} = 1 - \alpha$$

y en promedio, da un intervalo de confianza más preciso que cuando S es desconocido y se supone a priori.

MUESTREO ESTRATIFICADO

La precisión de una estimación por muestreo depende principalmente de dos factores: i).- el tamaño de la muestra ii).- la heterogeneidad de los elementos de la población. Consecuentemente, un método de aumentar la precisión de la estimación sin aumentar el tamaño de la muestra, es idear un procedimiento en el que se disminuya la influencia de la heterogeneidad de la población. A procedimiento tal es al que se le ha llamado procedimiento de Muestreo Estratificado.

El tal procedimiento consiste en clasificar de acuerdo con su heterogeneidad a los elementos de la población en k clases o grupos, a cada uno de los cuales se les conoce como Estratos.

La clasificación hecha debe garantizar la representación en la muestra de todos los estratos.

Un Muestreo Estratificado presupone desde luego el conocimiento de los Estratos, es decir, el tamaño de cada estrato y una lista del total de estratos con sus correspondientes unidades a la que llamaremos Marco.. El Marco es indispensable para la selección de la muestra.

La clasificación de las unidades dentro de cada estrato puede hacer se siguiendo diferentes criterios, no necesariamente deben formarse estratos con unidades físicamente contiguas dentro de la población.

ESTIMADA DE LA MEDIA DE LA POBLACION Y SU VARIANCIA

Sea N_i el tamaño de i-ésimo estrato y n_i el tamaño de la muestra seleccionada en él, de manera que:

$$\sum_{i=1}^{k} N_i = N \qquad y \qquad \sum_{i=1}^{k} n_i = n$$

La media por estimar de la población se define como:

$$\bar{Y}_N = \frac{1}{N} \sum_{i=1}^{k} N_i \bar{Y}_{Ni} = \sum_{i=1}^{k} p_i \bar{Y}_{Ni} \tag{72}$$

donde $p_i = \frac{N_i}{N}$; teniendo en cuenta que n_i es una muestra aleatoria de N_i - (a), podemos escribir:

$$\hat{\bar{Y}}_N = \sum_{i=1}^{k} p_i \bar{y}_{ni}$$ que será la estimada de la media de la población.

Obsérvese que esta media es una estimación ponderada con el tamaño de los estratos como ponderaciones podremos designarla pues con $\bar{y}_\omega$, siendo ω el símbolo que adoptemos para indicar ponderación.

$$\bar{y}_\omega = \sum_{i=1}^{k} p_i \bar{y}_{ni} \tag{73}$$

Demostremos que $\bar{y}_\omega$ es una estimación insesgada de la media de la población.

En el capítulo anterior se vió que:

$$E(\bar{y}_n) = \bar{Y}_N$$

que para este caso podremos escribir:

$$E(\bar{y}_{ni}) = \bar{Y}_{Ni}$$

según lo dicho en el párrafo (a); en consecuencia:

$$E(\bar{y}_\omega) = E\left\{\sum_{i=1}^{k} p_i \bar{y}_{ni}\right\} = \sum_{i=1}^{k} p_i E(\bar{y}_{ni}) = \sum_{i=1}^{k} p_i \bar{Y}_{Ni} = \bar{Y}_N \tag{74}$$

Obtengamos la variancia de muestreo de $\bar{y}_\omega$; por definición:

$$V(\bar{y}_\omega) = E(\bar{y}_\omega - \bar{Y}_N)^2$$

$$= E(\sum_{i=1}^{k} p_i \bar{y}_{ni} - \sum_{i=1}^{k} p_i \bar{Y}_{Ni})^2$$

$$= E\left\{\sum_{i=1}^{k} p_i (\bar{y}_{ni} - \bar{Y}_{Ni})\right\}^2$$

$$V(\bar{y}_\omega) = E\left\{\sum_{i=1}^{k} p_i^2 (\bar{y}_{ni} - \bar{Y}_{Ni})^2 + \sum_{i\neq j=1}^{k} p_i p_j (\bar{y}_{ni} - \bar{Y}_{Ni})(\bar{y}_{nj} - \bar{Y}_{Nj})\right\}$$

$$= \sum_{i=1}^{k} p_i^2 E(\bar{y}_{ni} - \bar{Y}_{Ni})^2 + \sum_{i\neq j=1}^{k} p_i p_j E\left\{(\bar{y}_{ni} - \bar{Y}_{Ni})(\bar{y}_{nj} - \bar{Y}_{Nj})\right\} \quad (75)$$

y recordando (a) y el resultado (55) del capítulo anterior:

$$E(\bar{y}_{ni} - \bar{Y}_{Ni})^2 = (\frac{1}{n_i} - \frac{1}{N_i}) S_i^2 \quad (76)$$

en la que S_i^2 es el cuadrado medio de la población en el i-ésimo estrato según la expresión:

$$S_i^2 = \frac{1}{N_i - 1} \sum_{i=1}^{N_i} (y_{ij} - \bar{Y}_{Ni})^2 \quad (77)$$

Volvamos a (75) y analicemos

$$\sum_{i\neq j=1}^{k} p_i p_j E\left\{(\bar{y}_{ni} - \bar{Y}_{Ni})(\bar{y}_{nu} - \bar{Y}_{Nj})\right\} ;$$

puesto que la selección de las muestras en el i-ésimo y j-ésimo estratos se hacen independientemente, concluimos que:

$$E\left\{(\bar{y}_{ni} - \bar{Y}_{Ni})(\bar{y}_{nj} - \bar{Y}_{Nj})\right\} = 0$$ con lo cual (75) queda:

$$V(\bar{y}_\omega)_{St.} \doteq \sum_{i=1}^{k} (\frac{1}{n_i} - \frac{1}{N_i}) p_i^2 S_i^2 \quad (78)$$

en la que se ve:la variancia en el muestreo estratificado depende de las - variancias dentro de cada estrato, lo que nos da un criterio para reducir la variancia en un muestreo estratificado al lograr reducir la variancia - dentro de cada estrato mientras más homogéneos se hagan los estratos.

El subíndice St. de $V(\bar{y}_\omega)_{St.}$ indica que se trata de un muestreo - estratificado.

SELECCION DE LOS TAMAÑOS DE LA MUESTRA EN DIFERENTES ESTRATOS

En la expresión (78) del párrafo anterior vemos que la variancia - de la media estimada es función de los tamaños de la muestra n_i dentro de los estratos.

Hay dos criterios a seguir para fijar el tamaño de las n_i, a saber:

I.- Selección de las n_i para obtener una estimación de $V(\bar{y}_\omega)_{St}$ con una precisión dada a costo mínimo y

II.- Para un costo prefijado seleccionar las n_i para hacer mínima - la variancia.

En el primero de los casos, nos fijamos la variancia dependiendo el costo de ella, y en el segundo nos fijamos costo, y supeditamos la varian cia al costo.

Sea C_i el costo por unidad muestreada en el estrato i-ésimo. El - costo total del muestreo estará dado por:

$$C = \sum_{i=1}^{k} c_i \, n_i \qquad (79)$$

Si c_i fuera el mismo de estrato a estrato, digamos c, el costo to- tal de la investigación será:

$$C = cn \qquad (80)$$

Para determinar los valores óptimos de n_i cuando la función del cos to es (79), consideremos la función:

$$\phi = V(\bar{y}_\omega) + \mu C \quad \text{donde } \mu \text{ es una constante.}$$

Substituyendo en ϕ los valores de $V(\bar{y}_\omega)$ y C, tendremos:

$$V(\bar{y}_\omega) + \mu C = \sum_{i=1}^{k} \left(\frac{1}{n_i} - \frac{1}{N_i}\right) p_i^2 \, S_i^2 + \mu \left\{ \sum_{i=1}^{k} c_i \, n_i \right\} = I$$

$$= \sum_{i=1}^{k} \left\{ \frac{p_i \, S_i^2}{n_i} - \frac{p_i S_i^2}{N_i} + \mu c_i \, n_i \right\} \quad \text{y sumando y}$$

restando $2p_i\, S_i \sqrt{\mu c_i}$ dentro del paréntesis y arreglando:

$$I = \sum_{i=1}^{k} \left\{ \left(\frac{p_i^2\, S_i^2}{n_i} + \mu c_i\, n_i - 2p_i\, S_i \sqrt{\mu c_i} \right) + 2p_i S_i \sqrt{\mu c_i} - \frac{1}{N_i}\, p_i^2 S_i^2 \right\}$$

o bien:

$$I = \sum_{i=1}^{k} \left(\frac{p_i\, S_i}{\sqrt{n_i}} - \sqrt{\mu c_i n_i} \right)^2 + 2 \sum_{i=1}^{k} p_i\, S_i \sqrt{\mu c_i} - \sum_{i=1}^{k} \frac{1}{N_i}\, p_i^2\, S_i^2$$

$$= \sum_{i=1}^{k} \left(\frac{p_i\, S_i}{\sqrt{n_i}} - \sqrt{\mu c_i n_i} \right)^2 + \text{términos independientes de } n_i \quad (81)$$

Claramente $V(\bar{y}_\omega)$ es mínima para c fijo, o el costo de una investigación es mínimo para un valor fijo de $V(\bar{y}_\omega)$, cuando cada uno de los términos cuadráticos en el segundo miembro de (81) es cero, o en otras palabras, cuando:

$$n_i \propto \frac{p_i\, S_i}{\sqrt{c_i}} \qquad (i = 1, 2, \ldots, k) \qquad (82)$$

La ecuación (82) nos hace ver:

i).- Mientras mayor sea el tamaño del estrato, mayor debe ser el tamaño de la muestra por seleccionar de allí;

ii).- mientras mayor sea la variabilidad dentro de un estrato, mayor debe ser el tamaño de la muestra para ese estrato; y

iii).- mientras más barato sea el costo dentro de un estrato, mayor debe ser la muestra en ese estrato.

AFIJACION DE NEYMAN

Determinemos el valor de las n_i. Evaluemos pues $\frac{1}{\sqrt{\mu}}$ de manera de satisfacer la condición de costo fijo o variancia fija. En el primer caso substituyendo en (79) a (82)

$$C_o = \sum_{i=1}^{k} \frac{p_i\, S_i}{\sqrt{\mu c_i}}\; c_i$$

en donde C_o es la cantidad presupuesta para el costo de la investigación con la cual se desea estimar la media con máxima precisión.

Cuando $c_i = c$, (88) se reduce a:

$$n_i = p_i\, S_i \frac{\sum_{i=1}^{k} p_i\, S_i}{V_o + \frac{1}{N}\sum_{i=1}^{k} p_i\, S_i^2} \qquad (89)$$

de tal suerte que la muestra ~~más~~ requerida para estimar la media con varian cia fija V_o está dada por:

$$n = \frac{\left(\sum_{i=1}^{k} p_i\, S_i\right)^2}{V_o + \frac{1}{N}\sum_{i=1}^{k} p_i\, S_i^2} \qquad (90)$$

TAMAÑO DE LA MUESTRA PARA ESTIMAR LA MEDIA CON UNA VARIANCIA DADA BAJO a) AFIJACION OPTIMA O DE NEYMAN Y b) AFIJACION PROPORCIONAL.

Hemos visto que para n_i - s arbitrarias, la variancia de la media - está dada por:

$$V(\bar{y}_\omega) = \sum_{i=1}^{k} \left(\frac{1}{n_i} - \frac{1}{N_i}\right) p_i^2 \quad S_i^2 \qquad (91)$$

bajo la afijación de Neyman tenemos:

$$n_i = \frac{np_i\, S_i}{\sum_{i=1}^{k} p_i\, S_i} \qquad (92)$$

substituyendo n_i de (92) en (91), tenemos:

$$V(\bar{y}_\omega)_N = \sum_{i=1}^{k} \left(\frac{\sum_{i=1}^{k} p_i\, S_i}{n\, p_i\, S_i} - \frac{1}{N_i} \right) p_i^2\, S_i^2$$

$$= \frac{1}{n}\left(\sum_{i=1}^{k} p_i\, S_i\right)^2 - \frac{1}{N}\sum_{i=1}^{k} p_i\, S_i^2 \qquad (93)$$

Cuando $c_i = c$, (88) se reduce a:

$$n_i = p_i\ S_i\ \frac{\sum_{i=1}^{k} p_i\ S_i}{V_o + \frac{1}{N}\sum_{i=1}^{k} p_i\ S_i^2} \tag{89}$$

de tal suerte que la muestra ~~mínima~~ requerida para estimar la media con varianza fija V_o está dada por:

$$n = \frac{\left(\sum_{i=1}^{k} p_i\ S_i\right)^2}{V_o + \frac{1}{N}\sum_{i=1}^{k} p_i\ S_i^2} \tag{90}$$

TAMAÑO DE LA MUESTRA PARA ESTIMAR LA MEDIA CON UNA VARIANCIA DADA BAJO a) AFIJACION OPTIMA O DE NEYMAN Y b) AFIJACION PROPORCIONAL.

Hemos visto que para n_i - s arbitrarias, la variancia de la media - está dada por:

$$V(\bar{y}_\omega) = \sum_{i=1}^{k} \left(\frac{1}{n_i} - \frac{1}{N_i}\right) p_i^2\ \ S_i^2 \tag{91}$$

bajo la afijación de Neyman tenemos:

$$n_i = \frac{np_i\ S_i}{\sum_{i=1}^{k} p_i\ S_i} \tag{92}$$

substituyendo n_i de (92) en (91), tenemos:

$$V(\bar{y}_\omega)_N = \sum_{i=1}^{k}\left(\frac{\sum_{i=1}^{k} p_i\ S_i}{n\ p_i\ S_i} - \frac{1}{N_i}\right) p_i^2\ S_i^2$$

$$= \frac{1}{n}\left(\sum_{i=1}^{k} p_i\ S_i\right)^2 - \frac{1}{N}\ \sum_{i=1}^{k} p_i\ S_i^2 \tag{93}$$

El subíndice de $V(\bar{y}_{\omega})_N$ nos indica que la variancia de la media es según la afijación de Neyman.

Para la afijación proporcional de la muestra entre los estratos:

$$n_i = np_i \tag{94}$$

que substituida en (91) da:

$$V(\bar{y}_{\omega})_P = \sum_{i=1}^{k}\left(\frac{1}{np_i} - \frac{1}{Np_i}\right) p_i^2\, S_i^2 = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{i=1}^{k} p_i\, S_i^2$$

$$= \frac{N-n}{Nn}\sum_{i=1}^{k} p_i\, S_i^2 \tag{95}$$

en la que el subíndice P indica que la variancia esta bajo el sistema de afijación proporcional.

Volviendo a la ecuación (93) podremos despejar n:

$$n = \frac{\left(\sum_{i=1}^{k} p_i\, S_i\right)^2}{V_o + \frac{1}{N}\sum_{i=1}^{k} p_i\, S_i^2} \tag{96}$$

que nos da el tamaño de la muestra estratificada bajo la afijación de Neyman con una variancia V_o dada.

Análogamente, de la ecuación (95) podremos despejar n:

$$n = \frac{\sum_{i=1}^{k} p_i\, S_i^2}{V_o + \frac{1}{N}\sum_{i=1}^{k} p_i\, S_i^2} \tag{97}$$

que nos da el tamaño de la muestra estratificada bajo la afijación proporcional con una variancia V_o dada.

COMPARACION DEL MUESTREO ESTRATIFICADO CON EL MUESTREO IRRESTRICTO ALEATORIO NO ESTRATIFICADO

Compararemos el M.E. con el MIA. bajo las distintas afijaciones:

a).- Afijación proporcional.

Habíamos visto que: $V(\bar{y}_{\omega})_P = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{i=1}^{k} p_i\, S_i^2$ (98)

Escribamos la variancia de la media en el MIA como:

$$V(\bar{y}_n)_{MIA} = \left(\frac{1}{n} - \frac{1}{N}\right) S^2 \tag{99}$$

Expresemos S^2 en términos de S_i^2 ; para esto, dividamos la suma de cuadrados total en la población en 2 partes:

i).- dentro de los estratos

ii).- entre los estratos:

$$\sum_{i=1}^{k} \sum_{j=1}^{N_i} (y_{ij} - \bar{y}_N)^2 \equiv \sum_{i=1}^{k} \sum_{j=1}^{N_i} (y_{ij} - \bar{y}_{Ni} + \bar{y}_{Ni} - \bar{y}_N)^2$$

$$\equiv \sum_{i=1}^{k} \sum_{j=1}^{N_i} (y_{ij} - \bar{y}_{Ni})^2 + \sum_{i=1}^{k} N_i \, (\bar{y}_{Ni} - \bar{y}_N)^2$$

que es equivalente a:

$$(N-1)S^2 = \sum_{i=1}^{k} (N_i - 1)S_i^2 + \sum_{i=1}^{k} N_i \, (\bar{y}_{Ni} - \bar{y}_N)^2 \tag{100}$$

para simplificar la expresión, aceptemos que:

$$\frac{N_i - 1}{N_i} \doteq 1 \qquad \text{y} \qquad \frac{N-1}{N} \doteq 1$$

suponiendo que N_i y N son suficientemente grandes para permitir estas aproximaciones.

Dividiendo en N y haciendo uso de estas aproximaciones:

$$S^2 \doteq \sum_{i=1}^{k} p_i \, S_i^2 + \sum_{i=1}^{k} p_i \, (\bar{y}_{Ni} - \bar{y}_N)^2$$

$$\therefore \sum_{i=1}^{k} p_i \, S_i^2 \doteq S^2 - \sum_{i=1}^{k} p_i \, (\bar{y}_{Ni} - \bar{y}_N)^2 \tag{101}$$

Substituyendo (101) en (98)

$$V(\bar{y}_\omega)_p \doteq \left(\frac{1}{n} - \frac{1}{N}\right) \left\{ S^2 - \sum_{i=1}^{k} p_i \, (\bar{y}_{Ni} - \bar{y}_N)^2 \right\} \tag{102}$$

restando (102) de (99):

$$V(\bar{y}_N)_{MIA} - V(\bar{y}_\omega)_P \doteq \frac{N-n}{Nn} \sum_{i=1}^{k} p_i \, (\bar{y}_{Ni} - \bar{y}_N)^2 \tag{103}$$

que nos dice que: mientras mas difieran los estratos en sus medias, mayor será la ganancia en precisión debida al muestreo proporcional sobre el muestreo no estratificado irrestricto aleatorio.

b).- afijación de Neyman:

Expresemos primero la variancia de la media bajo afijación de Neyman en una forma que permita la comparación con la variancia en muestreo no estratificado irrestricto aleatorio.

Hagamos uso de la identidad:

$$\sum_{i=1}^{k} p_i \, (s_i - \bar{S}_\omega)^2 \equiv \sum_{i=1}^{k} p_i \, S_i^2 - \Big(\sum_{i=1}^{k} p_i \, S_i\Big)^2 \tag{104}$$

donde

$$\bar{S}_\omega = \sum_{i=1}^{k} p_i \, S_i$$

Substituyendo a

$$\left(\sum_{i=1}^{k} p_i \, S_i\right)^2$$

según (104) en la expresión para la variancia de la media bajo afijación de Neyman dada en (93), obtenemos:

$$V(\bar{y}_\omega)_N = \frac{1}{N}\left[\sum_{i=1}^{k} p_i \, S_i^2 - \sum_{i=1}^{k} p_i (s_i - \bar{S}_\omega)^2\right] - \frac{1}{N} \sum_{i=1}^{k} p_i \, S_i^2$$

$$= \frac{N-n}{Nn} \sum_{i=1}^{k} p_i \, S_i^2 - \frac{1}{n} \sum_{i=1}^{k} p_i (S_i - \bar{S}_\omega)^2 \tag{105}$$

El primer término del segundo miembro de la ecuación (105) representa la variancia bajo la afijación proporcional.

$$V(\bar{y}_{\omega})_P - V(\bar{y}_{\omega})_N = \frac{1}{n} \sum_{i=1}^{k} p_i \, (s_i - \bar{S}_{\omega})^2 \qquad (106)$$

Vemos que mientras mayores sean las diferencias entre las desviaciones estándar de los estratos, mayor será la ganancia en precisión de la afijación óptima sobre la afijación proporcional. Más aún, substituyendo a $V(\bar{y}_{\omega})_P$ según (102), tenemos:

$$V(\bar{y}_{\omega})_N \doteq \frac{N-n}{N_n}\left\{S^2 - \sum_{i=1}^{k} p_i(\bar{y}_{Ni} - \bar{y}_N)^2 - \frac{N}{N-n}\sum_{i=1}^{k} p_i(s_i - \bar{S}_{\omega})^2\right\} \quad (107)$$

puesto que el primer término del segundo miembro de (107) es la variancia de la media de un muestreo no estratificado de n, podemos escribir:

$$V(\bar{y}_n)_{MIA} - V(\bar{y}_{\omega})_N \doteq \frac{N-n}{Nn}\left\{\sum_{i=1}^{k} p_i(\bar{y}_{Ni} - \bar{y}_N)^2 + \frac{N}{N-n}\sum_{i=1}^{k} p_i(s_i-\bar{S}_{\omega})^2\right\} (108)$$

La ecuación (108) muestra que la ganancia en precisión de la afijación de Neyman sobre el muestreo no estratificado irrestricto aleatorio radica en dos factores:

i).- diferencias entre las medias de los estratos y

ii).- diferencias entre las desviaciones estandar de los estratos.

O sea, para una estratificación eficiente, la población debe dividirse de manera que las diferencias entre las medias y las desviaciones estándar de los estratos sean tan grandes como sea posible.

La mejor manera de hacer esto en la práctica es agrupar las unidades parecidas de la población.

A veces las unidades contiguas geográficamente, son mas semejantes que las que están separadas.

Otras veces, se forman los estratos basándose en algún carácter correlacionado; por ejemplo, el tamaño de una granja se sabe que está correlacionado con un cierto número de caracteres de la granja, como: el área de cultivos, número de ganado, etc.

EVALUACION A PARTIR DE LA MUESTRA DE LA GANANCIA EN PRECISIÓN DEBIDA A LA LA ESTRATIFICACIÓN

Hemos comparado la precisión del muestreo estratificado aleatorio con el muestreo irrestricto aleatorio no estratificado, suponiendo que los valores de las medias y las desviaciones estándar de la población de los estratos se conocían. Sin embargo, generalmente no será este el caso; sino que únicamente dispondremos de una muestra estratificada y el problema consistirá en estimar la ganancia en precisión debida a la estratificación.

Supongamos que $n_1, n_2, \ldots, n_k$ representan a la muestra estratificada. La variancia de la media de la muestra es entonces:

$$V(\bar{y}_\omega)_s = \sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right) p_i^2\, S_i^2$$

y su estimada está dada por:

$$\text{Est. } V(\bar{y}_\omega)_s = \sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right) p_i\, s_i^2 \tag{109}$$

donde s_i^2 es el cuadrado medio en la muestra tomada del i-ésimo estrato. Si hubiéramos seleccionado la muestra total según un muestreo irrestricto aleatorio sin estratificación, la variancia de la media de la muestra sería:

$$V(\bar{y}_n)_{MIA} = \frac{N-n}{N} \cdot \frac{S^2}{n} \tag{110}$$

Sin embargo, su estimada no puede obtenerse substituyendo el cuadrado medio para la muestra total en lugar de S^2, porque, aunque dentro de cada estrato la muestra se selecciona por el método de muestreo irrestricto aleatorio, la muestra total no puede considerarse que ha sido seleccionada de esa manera para la población total.

El problema es, por lo tanto, estimar s^2, dados:

$\bar{y}_{n1}, \bar{y}_{n2}, \ldots, \bar{y}_{nk}$ y $s_1^2, s_2^2, \ldots s_k^2$.

Tenemos de (100):

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{k} (N_i - 1)S_i^2 + \frac{N}{N-1}\sum_{i=1}^{k} p_i(\bar{y}_{Ni} - \bar{y}_N)^2 \tag{111}$$

Puesto que s_i^2 proporciona una estimada insesgada de S_i^2, el problema de estimar S^2 se reduce a la estimación de:

$$\sum_{i=1}^{k} p_i(\bar{y}_{ni} - \bar{y}_N)^2 \tag{112}$$

a partir de la muestra.

Sea

$$\bar{y}_{ni} = \bar{y}_{Ni} + \varepsilon_i \tag{113}$$

donde $E(\varepsilon_i) = 0$, y $E(\varepsilon_i^2) = \frac{N_i - n_i}{N_i\ n_i} S_i^2$

Elevando al cuadrado ambos miembros de (113) tenemos:

$$\bar{y}_{ni}^2 = \bar{y}_{Ni}^2 + \varepsilon_i^2 + 2\bar{y}_{Ni}\varepsilon_i \tag{114}$$

tomando esperanzas matemáticas tenemos:

$$E(\bar{y}_{ni}^2) = \bar{y}_{Ni}^2 + \frac{N_i - n_i}{N_i\ n_i} S_i^2 \tag{115}$$

de donde:

$$E\left(\sum_{i=1}^{k} p_i\ \bar{y}_{ni}^2\right) = \sum_{i=1}^{k} p_i\ \bar{y}_{Ni}^2 + \sum_{i=1}^{k} \frac{N_i - n_i}{N_i\ n_i}\ p_i\ S_i^2 \tag{116}$$

Se infiere que:

$$\text{Est.}\left(\sum_{i=1}^{k} p_i\ \bar{y}_{Ni}^2\right) = \sum_{i=1}^{k} p_i\ \bar{y}_{ni}^2 - \sum_{i=1}^{k} \frac{N_i - n_i}{N_i\ n_i'}\ p_i\ s_i^2 \tag{117}$$

De (113) tenemos:

$$\sum_{i=1}^{k} p_i\ \bar{y}_{ni} = \sum_{i=1}^{k} p_i\ \bar{y}_{Ni} + \sum_{i=1}^{k} p_i\varepsilon_i$$

o bien:

$$\bar{y}_{\omega} = \bar{y}_N + \sum_{i=1}^{k} p_i\varepsilon_i \tag{118}$$

Elevando al cuadrado mabos miembros, obtenemos:

$$\bar{y}_{\omega}^2 = \bar{y}_N^2 + \sum_{i=1}^{k} p_i^2\varepsilon_i^2 + \sum_{i\neq j=1}^{k} p_i p_j \varepsilon_i \varepsilon_j + 2\bar{y}_N \sum_{i=1}^{k} p_i\varepsilon_i \tag{119}$$

Tomando esperanzas matemáticas y notando que $E(\varepsilon_i \varepsilon_j) = 0$ ya que - las muestras de los diferentes estratos se han seleccionado independientemente, obtenemos:

$$E(\bar{y}^2_{\omega}) = \bar{y}^2_N + \sum_{i=1}^{k} \frac{N_i - n_i}{N_i\, n_i} p_i^2 S_i^2 \qquad (120)$$

De donde

$$\text{Est.}(\bar{y}^2_N) = \bar{y}^2_{\omega} - \sum_{i=1}^{k} \frac{N_i - n_i}{N_i\, n_i} p_i^2 s_i^2 \qquad (121)$$

Restando (121) de (117) tenemos:

$$\text{Est.} \sum_{i=1}^{k} p_i(\bar{y}_{Ni} - \bar{y}_N)^2 = \sum_{i=1}^{k} p_i(\bar{y}_{ni} - \bar{y}_{\omega})^2 - \sum_{i=1}^{k} p_i(1-p_i)\frac{N_i - n_i}{N_i\, n_i} s_i^2 \qquad (122)$$

De (111) y (122) obtenemos:

$$\text{Est. } S^2 = \frac{1}{N-1} \sum_{i=1}^{k} (N_i-1)s_i^2 + \frac{N}{N-1}\left\{\sum_{i=1}^{k} p_i(\bar{y}_{ni}-\bar{y}_{\omega})^2 - \sum_{i=1}^{k} p_i(1-p_i) \frac{N_i - n_i}{N_i\, n_i} s_i^2\right\}$$

que puede escribirse como:

$$\text{Est.}S^2 = \frac{N}{N-1}\sum_{i=1}^{k} (p_i - \tfrac{1}{N})s_i^2 + \frac{N}{N-1}\left\{\sum_{i=1}^{k} p_i(\bar{y}_{ni}-\bar{y}_{\omega})^2 - \sum_{i=1}^{k} p_i(1-p_i)\frac{s_i^2}{n_i} + \frac{1}{N}\sum_{i=1}^{k} (1-p_i)s_i^2\right\}$$

simplificando:

$$\text{Est. } S^2 = \sum_{i=1}^{k} p_i\, s_i^2 + \frac{N}{N-1}\left\{\sum_{i=1}^{k} p_i(\bar{y}_{ni} - \bar{y}_{\omega})^2 - \sum_{i=1}^{k} p_i(1 - p_i) \frac{s_i^2}{n_i}\right\} \qquad (123)$$

substituyendo en (110) tenemos:

$$\text{Est.}V(\bar{y}_n)_{MIA} = \frac{N-n}{Nn} \sum_{i=1}^{k} p_i s_i^2 + \frac{N-n}{n(N-1)}\left\{\sum_{i=1}^{k} p_i(\bar{y}_{ni} - \bar{y}_{\omega})^2 - \sum_{i=1}^{k} p_i(1-p_i) \frac{s_i^2}{n_i}\right\} \qquad (124)$$

De aquí, el cambio en la variancia debido a la estratificación se - estima por la diferencia entre (124) y (109), o sea:

$$\text{Est.}\left\{V(\bar{y}_n)_{MIA} - V(\bar{y}_{\omega})_s\right\} = \frac{1}{n} \sum_{i=1}^{k} p_i s_i^2 - \sum_{i=1}^{k} \frac{p_i s_i^2}{n_i} + \frac{N - n}{n(N-1)}\left\{\sum_{i=1}^{k} p_i(\bar{y}_{ni}-\bar{y}_{\omega})^2 - \sum_{i=1}^{k} p_i(1-p_i) \frac{s_i^2}{n_i}\right\}$$

Tomando

$\frac{N}{N-1} \doteq$ tendremos:

$$\text{Est.}\left\{V(\bar{y}_n)_{MIA}-V(\bar{y}_\omega)_s\right\} \doteq \frac{1}{n}\sum_{i=1}^{k} p_i s_i^2 - \sum_{i=1}^{k}\frac{p_i s_i^2}{n_i} + \frac{N-n}{Nn}\left\{\sum_{i=1}^{k} p_i(\bar{y}_{ni}-\bar{y}_\omega)^2 - \sum_{i=1}^{k} p_i(1-p_i)\frac{s_i^2}{n_i}\right\} \quad (126)$$

La razón de (125) o (126) a (109) expresada como un porcentaje, da la estimada de la ganancia en eficiencia debida a la estratificación.

Estos resultados toman una forma particularmente sencilla en el caso de muestreo proporcional para el cual $\bar{y}_\omega = \bar{y}_n$.

La ecuación (109) se convierte a:

$$\text{Est. } V(\bar{y}_\omega)_p = \frac{N-n}{Nn}\sum_{i=1}^{k} p_i\, s_i^2 \quad (127)$$

y los primeros dos términos de (126) desaparecen. La reducción neta en varianza, debida a la estratificación está dada por los últimos términos de (126).

Al substituir $\frac{n_i}{n}$ por p_i, (126) queda:

$$\text{Est.} V(\bar{y}_n)_{MIA} - \text{Est.} V(\bar{y}_\omega)_p = \frac{N-n}{N_n}\left\{\frac{1}{n}\sum_{i=1}^{k} n_i(\bar{y}_{ni}-\bar{y}_\omega)^2 - \frac{1}{n}\sum_{i=1}^{k}\left(1-\frac{n_i}{n}\right)s_i^2\right\} \quad (128)$$

Si se supone que el multiplicador finito es uno, y S_i^2 es constante, digamos: $S_i^2 = S_\omega^2$ $(i = 1, 2, \ldots k)$, la mejor estimación insesgada de éste último se obtiene combinando la suma de cuadrados dentro de los estratos para la muestra. Podemos escribir:

$$\text{Est. } S_\omega^2 = \frac{1}{n-k}\sum_{i=1}^{k}\sum_{j}^{n_i}(y_{ij}-\bar{y}_{ni})^2 = s_\omega^2 \quad (129)$$

y puesto que:

$$E(s_\omega^2) = \frac{1}{n-k}E\left\{\sum_{i=1}^{k}\sum_{j}^{J_i}(y_{ij}-\bar{y}_{ni})^2\right\} = \frac{1}{n-k}E\sum_{i=1}^{k} E\sum_{j}^{n_i}(y_{ij}-\bar{y}_{ni})^2$$

$$= \frac{1}{n-k}E\sum_{i=1}^{k}(n_i-1)S_i^2 = S_\omega^2$$

Sea:

$$\sum_{i=1}^{k} n_i(\bar{y}_{ni}-\bar{y}_\omega)^2 = (k-1)\,\bar{n}\, s_b^2 \quad (130)$$

donde $\bar{n} = \frac{n}{k}$

Substituyendo (129) y (130) en (124), tenemos:

$$\text{Est. } V(\bar{y}_n)_{MIA} \doteq \frac{1}{n^2}\left\{(n-k+1)s_\omega^2 + (k-1)\bar{n}\ s_b^2\right\} \qquad (131)$$

Las cantidades s_ω^2 y s_b^2 se llaman los cuadrados medios dentro y entre los estratos respectivamente, y se calculan mas fácilmente con una tabla de análisis de variancia:

Causa de variación	G. L.	Suma de cuadrados	Cuadrado medio
Entre estratos	$k - 1$	$\sum_{i=1}^{k} n_i(\bar{y}_{ni} - \bar{y}_n)^2$	$\bar{n}\ s_b^2$
Dentro de estratos	$n - k$	$\sum_{i=1}^{k}\sum_{j}^{n_i} (y_{ij} - \bar{y}_{ni})^2$	s_ω^2
Total	$n - 1$	$\sum_{i=1}^{k}\sum_{j}^{n_i} (y_{ij} - \bar{y}_n)^2$	s^2

También de (127):

$$\text{Est. } V(\bar{y}_\omega)_p \doteq \frac{s_\omega^2}{n} \qquad (132)$$

Una estimada de la reducción en variancia está dada restando (132) de (131) o directamente de (128)

$$\text{Est.}V(\bar{y}_n)_{MIA} - \text{Est.}V(\bar{y}_\omega)_p = \frac{k-1}{n^2}\left\{\bar{n}\ s_b^2 - s_\omega^2\right\} \qquad (133)$$

La razón de (133) a (132) da la ganancia en precisión relativa, debida a la estratificación y es:

$$\frac{k-1}{n}\left\{\frac{\bar{n}\ s_b^2}{s_\omega^2} - 1\right\} \qquad (134)$$

La eficiencia de la estratificación a veces se calcula directamente comparando el cuadrado medio total s^2 con s_ω^2, y la ganancia en precisión relativa está dada por:

$$\frac{s^2}{s_w^2} - 1 = \frac{(n-k)s_w^2 + (k-1)\bar{n}\ s_b^2}{(n-1)s_w^2} - 1$$

$$\frac{s^2}{s_w^2} - = \frac{k-1}{n-1}\left\{\frac{\bar{n}s_b^2}{s^2} - 1\right\} \qquad (135)$$

La ganancia en precisión estimada de esta manera es $\frac{n}{n-1}$ veces el valor dado por (134), lo que no es probable que sea de una diferencia material en - muestras grandes, supuesto que la muestra está repartida en proporción a los - tamaños de los diferentes estratos. Cuando la muestra no está afijada así, - ni (134) ni (135) es probable que sean satisfactorias. La expresión exacta - dada por la razón de (125) a (109) deberá usarse en ese caso.

MUESTREO SISTEMATICO

Hemos visto en los capítulos anteriores métodos de muestreo en los que las unidades de la muestra se seleccionaron aleatoriamente. Consideremos ahora un método de muestreo en el que solamente la primera unidad se selecciona - aleatoriamente, y el resto se selecciona automáticamente de acuerdo con un - plan predeterminado. El método se conoce como MUESTREO SISTEMATICO.

El plan seguido al seleccionar una muestra sistemática consiste en espaciar regularmente las unidades. Así, supongamos que una población consiste de N unidades numeradas seriadamente de la N. Supongamos además que N puede - expresarse como el producto de 2 enteros: k y n, de manera que $N = kn$. Se toma el número aleatorio menor o igual a k, digamos i; de i en adelante se se-- lecciona la unidad con el número de serie correspondiente a cada unidad k-ésima de la población.

Evidentemente, la muestra contendrá las n unidades:

$i;\ i + k;\ i + 2k;\ \ldots\ ;\ i + (n-1)k$, y se conoce como una MUESTRA SISTEMATICA.

Supondremos que $N = nk$ donde n es el tamaño de la muestra, y k un entero positivo. En la práctica N puede no ser expresable de esta menra, y los resultados presentados en este capítulo pueden no ser estrictamente aplicables.- Sin embargo, los errores que se cometen son de poca importancia a menos que n sea pequeña.

Haremos un cuadro esquemático de la población, dividiéndola en k columnas de n renglones (elementos) cada una.

Columna / Renglón	1	2	i	k
1	1	2	i	k
2	1 + k	2 + k	i + k	2k
3	1 + 2k	2 + 2k	i + 2k	3k
j	1 + (j-1)k	2 + (j-1)k	i + (j-1)k	jk
n	1 + (n-1)k	2 + (n-1)k	i + (n-1)k	nk

LA MEDIA DE LA MUESTRA Y SU VARIANCIA

Sea:

y_{ij} el valor de la unidad que lleva el número seriado $i + (j-1)k$ en la población ($i=1,2,\ldots k$; $j=1,2,\ldots,n$).

$\bar{y}_i$ la media de la muestra.

$$\bar{y}_i = \frac{1}{n} \sum_{j=1}^{n} y_{ij} \qquad (136)$$

$\bar{Y}$ la media de la población, expresada por:

$$\bar{y} = \frac{1}{nk} \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij} \qquad (137)$$

Puesto que la probabilidad de seleccionar la columna i-ésima es $\frac{1}{k}$, - se concluye que:

$$E(\bar{y}_i) = \frac{1}{k} \sum_{i=1}^{k} \bar{y}_i = \bar{y} \qquad (138)$$

que nos dice que una muestra sistemática proporciona una estimada insesgada de la media de la población.

La variancia está dada por:

$$V(\bar{y}_i)_s = E\left\{(\bar{y}_i - \bar{y})^2\right\}$$

$$= \frac{1}{k} \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2 \qquad (139)$$

$$= \frac{k-1}{k} S_c^2 \qquad (140)$$

donde S_c^2 es el cuadrado medio entre las medias de las columnas, el índice c indica una columna.

ESTIMACION DE LA VARIANCIA

Puesto que una muestra sistemática es una muestra aleatoria de un solo conglomerado (un solo elemento dentro de cada conglomerado), no se puede formar ninguna estimada de la variancia a partir de la muestra. Esto constituye una gran desventaja del método, que por otra parte, ofrece grandes - beneficios.

En la práctica se usan ciertas aproximaciones para calcular la varian cia. Una de éstas, consiste en tratar la muestra sistemática como si fuera una aleatoria de n unidades, y calcular la variancia usando la fórmula:

$$\left\{\frac{1}{n} - \frac{1}{nk}\right\} S^2_{\omega c} \tag{141}$$

donde $S^2_{\omega c}$ es el cuadrado medio entre unidades dentro de la muestra sistemá tica seleccionada, es decir, una columna del diagrama (1). Evidentemente (141) no proporciona una estimación insesgada de (140).

Otra aproximación usada a veces para calcular la variancia a partir de la muestra es:

$$\left\{\frac{1}{n} - \frac{1}{nk}\right\} \frac{\sum_{j=1}^{n-1} (y_{ij} - y_{ij+1})^2}{2(n-1)} \tag{142}$$

que tampoco da una estimación insesgada de la variancia (no se justificarán estas dos últimas afirmaciones.)

C A P I T U L O IV

DETERMINACION DE LAS ESTACIONES DE AFORO

DETERMINACION DE LOS DIAS DE AFORO

DETERMINACION DE LAS HORAS DE AFORO

El Departamento de Planeación Vial de la S.C.O.P. por medio de los - Departamentos de Conservación y Cooperación de la Dirección Nacional de Caminos proporcionó los datos de los aforos existentes hasta el año de 1954.

Las cifras qué aparecen a continuación de los nombres de los tramos - carreteros, corresponden al promedio diario trimestral de vehículos (sin — atender a su calidad) que transitaron por ese tramo, en el trimestre de máximo tráfico en el año de 1954.

Se escogieron los datos de 1954 por haber sido los mas recientes y - completos existentes en el Departamento de Planeación Vial.

Préviamente se compararon estos datos con los de años anteriores y - se observó que la intensidad de tráfico en dichos tramos salvo su magnitud - que iba en aumento de año a año conservaba (con excepción de 2 o 3 tramos)- la misma ley de variación respecto a los trimestres del año, es decir, que el trimestre de máxima intensidad de tráfico se conservaba el mismo para cada tramo. De aquí surgió la idea de hacer una clasificación de los tramos de acuerdo con el trimestre de máxima intensidad de tráfico.

Se agruparon bajo los número del I al IV a los tramos según los trimestres del año, correspondiendo el I a aquellos en los que el tráfico máximo se presento en el primer trimestre, el II a los que el máximo fue en el segundo trimestre y así sucesivamente.

Una segunda clasificación se hizo atendiendo únicamente a la intensidad de tráfico durante los trimestres del año, sin importar en que trimestre se había presentado el máximo.

Y por último la tercera estratificación se hizo con la combinación - de las dos clasificaciones anteriores, es decir, se formaron grupos tomando en cuenta la intensidad de tránsito en el trimestre de máximo tráfico, y el trimestre en que se presentó ese máximo, es decir, en cada grupo están los tramos semejantes no únicamente respecto al número de vehículos por trimestre que circularon, sino también si ese número de vehículos circuló en el - primero, segundo, tercero o cuarto trimestres.

Las estratificaciones se presentan a continuación:

ESTRATIFICACION A

(Por trimestres únicamente)

ESTRATO I	Número de Vehículos	ESTRATO II	Número de Vehículos	ESTRATO III	Número de Vehículos	ESTRATO IV	Número de Vehículos
Toluca-Zitácuaro	1 677	Toluca-México	10 342	Nogales-Hermosillo	623	Tijuana-Mexicali	1 767
Zitácuaro-Cd.Hidalgo	941	Guadalajara-Jocotepec	603	Cd.Hidalgo-Morelia	975	Zamora-Zahuayo	2 152
Durango-Sombrerete	781	Zahuayo-Tizapán	971	Sta.Bárbara-Chalco	2 362	S.Juan del Río-Toluca	2 007
Fresnillo-Zacatecas	1 115	Guadalajara-Tepatitlán	834	S.Juan del Río-Querétaro	1 632	Cd.Juárez-Chihuahua	3 292
Allende-Sabinas	978	Llano Grande-El Salto	1 029	Piedras Niegras-Allende	2 063	Durango-Llano Grande	558
Sabinas-Monclova	978	Huichapan-Ixmiquilpan	766	Saltillo-Camaleón	880	S.Juan del Río-Huichapan	760
Antiguo Morelos-Cd. del Maíz	288	Zacatepec-Perote	771	Camaleón-Monclova	773	Villagrán-Cd.Victoria	590
Jacala-Tamazunchale	574	Guadalupe-Venta de Carpio	3 325	Nuevo Laredo-Sabinas Hidalgo	551	Zacatepec-Zaragoza	517
Tamazunchale-Valles	699	Venta de Carpio-Colonia	3 057	Sabinas Hidalgo-Monterrey	699	México-Los Reyes	8 268
México-Cuernavaca	1 527			Cd.Victoria-Mante	729	Los Reyes-Sta.Bárbara	6 077
Cuernavaca-Taxco(V)	611			Valles-Tampico	336	S.Hipólito-Zacatepec	727
Perote-Jalapa	1 745			Mante-Tampico	426	Córdoba-La Capilla	2 016
				Mante-Cd.Victoria	739	La Capilla-Paso del Toro	883
				Zimapán-Jacala	761	Paso del Toro-Veracruz	925
				Ixmiquilpan-Zimapán	820	Puebla-S.Hipólito	2 464
				Insurgentes-Ecatepec	3 382	México-Cuernavaca(N)	2 092
				Papantla-Poza Rica	1 143		
				Chalco-Cuautla	1 885		
				Colonia-Pachuca	1 567		

E S T R A T I F I C A C I O N   B

(Por intervalos únicamente)

De 1 a 500	Número de Vehículos	De 501 a 750	Número de Vehículos	De 751 a 1000	Número de Vehículos	De 1001 a 1500	Número de Vehículos	De 1501 a 2000	Número de Vehículos
Antiguo Morelos-C.Maíz	288	Jacala-Tamazunchale	574	Zitácuaro-C.Hidalgo	941	Fresnillo-Zacatecas	1 115	Toluca-Zitácuaro	1 677
Valles-Tampico	336	Tamazunchale-Valles	699	Durango-Sombrerete	781	Llano Grande-El Salto	1 029	México-Cuernavaca	1 527
Mante-Tampico	426	Cuernavaca-Taxco	611	Allende-Sabinas	978	Papantla-Poza Rica	1 148	Perote-Jalapa	1 745
		Guadalajara-Jocotepec	603	Sabinas-Monclova	978			S.Juan del Río-Querétaro	1 632
		Nogales-Hermosillo	623	Zahuayo-Tizapan	971			Chalco-Cuautla	1 885
		Nvo.Laredo-Sabinas Higaldo	551	Guadalajara-Tepatitlán	834			Colonia-Pachuca	1 567
		Sabinas-Hgo.-Monterrey	699	Huichapan-Ixmiquilpan	766			Tijuana-Mexicali	1 767
		C.Victoria-Mante	729	Zacatepec-Perote	771				
		Mante-C.Victoria	739	C.Hidalgo-Morelia	975				
		Dgo.-Llano Grande	558	Saltillo-Camaleón	880				
		Villagrán-C.Victoria	590	Camaleón-Monclova	773				
		Zacatepec-Zaragoza	517	Zimapán-Jacala	761				
		S.Hipólito-Zacatepec	727	Ixmiquilpan-Zimapán	820				
				S.Juan del Río-Huichapan	760				
				P.del Toro-Veracruz	925				
				La Capilla-P.del Toro	883				

De 2001 a 2500	Número de Vehículos	De 3001 a 3500	Número de Vehículos	De 6001 a 6500	Número de Vehículos	De 8001 a 8500	Número de Vehículos	De 10 001 a 10 500	Número de Vehículos
S.Bárbara-Chalco	2 362	[illegible].-Venta de Carpio	3 325	Los Reyes-S.Bárbara	6 077	México-Los Reyes	8 268	Toluca-México	10 342
P.Negras-Allende	2 063	V.de Carpio-Colonia	3 057						
Zamora-Zahuayo	2 152	Insurgentes-Ecatepec	3 382						
S.J.del Río-Toluca	2 007	C.Juárez-Chihuahua	3 292						
Córdoba-La Capilla	2 016								
Puebla-S.Hipólito	2 464								
México-Cuernavaca (N)	2 092								

E S T R A T I F I C A C I O N   C

(Por intervalos y por trimestres)

Primer trimestre		Segundo trimestre		Tercer trimestre	
Estrato 1	Número de Vehículos	Estrato 4	Número de Vehículos	Estrato 7	Número de Vehículos
Zitácuaro-Cd.Hidalgo	941	Toluca-México	10 342	Nogales-Hermosillo	623
Durango-Sombrerete	781			Cd.Hidalgo-Morelia	975
Allende-Sabinas	978			Saltillo-Camaleón	880
		Estrato 5			
Sabinas-Monclova	978	Guadalajara-Jocotepec	603	Camaleón-Monclova	773
Tamazunchale-Valles	699	Zahuayo-Tizapan	971	Sabinas Hgo.-Monterrey	699
Cuernavaca-Taxco (V)	611	Guadalajara-Tepatitlán	834	C.Victoria-Mante	729
Estrato 2					
Toluca-Zitácuaro	1 677	Llano Grande-El Salto	1 029	Ixmiquilpan-Zimapan	820
Fresnillo-Zacatecas	1 115	Huichapan-Ixmiquilpan	766	Zimapán-Jacala	761
México-Cuernavaca	1 527	Zacatepec-Perote	771	Mante-C.Victoria	739
Estrato 3		Estrato 6		Estrato 8	
Antiguo Morelos-C. Maíz	288	Guadalupe-Venta de Carpio	3 325	Sta.Bárbara-Chalco	2 362
Jacala-Tamazunchale	574	Venta de Carpio-Colonia	3 057	P.Negras-Allende	2 063
				Cuautla-Chalco	1 835

Tercer trimestre		Cuarto trimestre		Cuarto trimestre	
Estrato 9	Número de Vehículos	Estrato 12	Número de Vehículos	Estrato 14	Número de Vehículos
S.Juan del Río-Querétaro	1 632	Tijuana-Mexicali	1 767	C.Juárez-Chihuahua	3 292
Papantla-Poza - Rica	1 148	Zamora-Zahuayo	2 152		
Colonia-Pachuca	1 567	S.Juan del Río-Toluca	2 007		
Estrato 10				Estrato 15	
Nvo.Laredo-Sabinas Hgo.	551	Córdoba-La Capilla	2 016	México-Los Reyes	8 268
Valles-Tampico	336	Puebla-S.Hipólito	2 464		
Mante-Tampico	426	México-Cuernavaca (N)	2 092		

Tercer trimestre		Cuarto trimestre		Cuarto trimestre	
Estrato 11	Número de Vehículos	Estrato 13	Número de Vehículos	Estrato 16	Número de Vehículos
Insurgentes-Ecatepec	3 382	Durango-Llano Grande	558	Los Reyes-S.Bárbara	6 077
		S.Juan del Río-Huichapan	760		
		Villagrán-C.Victoria	590		
		Zacatepec-Zaragoza	517		
		S.Hipólito-Zacatepec	727		
		La Capilla-P.del Toro	883		
		P.del Toro-Veracruz	925		

EFICIENCIA DE LA ESTRATIFICACION A

Empleando la expresión (135):

$$\frac{k-1}{N}\left\{\frac{S_E^2}{S_D^2}-1\right\}$$

Calculemos:

$$S_E^2 = \frac{1}{3}\left\{11\ 828\ 616 + 52\ 311\ 467 + 26\ 293\ 011 + 76\ 978\ 689 - \frac{(91\ 058)^2}{56}\right\} =$$

$$= \frac{1}{3}(167\ 441\ 783 - 19\ 348\ 223) = \frac{19\ 348\ 223}{3} = 6\ 449\ 408$$

$$S_D^2 = \frac{1}{52}(181\ 129\ 372 - 19\ 348\ 223) = \frac{161\ 781\ 149}{52} = 3\ 111\ 176$$

$$\eta_A = \frac{3}{56}\left(\frac{6\ 449\ 408}{3\ 111\ 176} - 1\right) = \frac{3}{56}\ (2.073 - 1) = .054 \times 1.073 = .058$$

o en porciento:

5.8% más eficiente que el MIA

EFICIENCIA DE LA ESTRATIFICACION B

$$S_E^2 = \frac{180\ 579\ 091}{9} = 20\ 064\ 343$$

$$S_D^2 = \frac{148\ 613\ 841}{46} = 3\ 230\ 736$$

$$\eta_B = \frac{9}{56} \quad \frac{20\ 064\ 343}{3\ 230\ 736} - 1 \quad = \frac{9}{56}\,(6.210 - 1) = .161 \times 5.210 = .839$$

o en porciento

83.9% más eficiente que el MIA.

EFICIENCIA DE LA ESTRATIFICACION C

$$S_E^2 = \frac{179\ 899\ 363}{15} = 12\ 993\ 291$$

$$\eta_C = \frac{15}{56}\left(\frac{12\ 993\ 291}{3\ 132\ 339} - 1\right) = \frac{15}{56}\,(4.213 - 1) \doteq .268 \times 3.213 = .861$$

o en porciento:

86.1% más eficiente que el MIA.

Comparando:

η_A, η_B y η_C. escogemos la estratificación C por ser la

la más eficiente: $\eta_C \doteq 86.1\%$

CALCULO DEL TAMAÑO DE LA MUESTRA

Fijaremos el tamaño de la muestra bajo una "afijación óptima" o - "afijación de Neyman"; para esto, utilizaremos la expresión (96)

$$n = \frac{\left(\sum_{i=1}^{k} p_i\, S_i\right)^2}{V_0 + \frac{1}{N}\sum_{i=1}^{k} p_i\, S_i^2}$$

en la que:

n = tamaño de la muestra

$p_i = \frac{N_i}{N}$

S_i^2 = variancia del estrato i-ésimo

V_0 = variancia con que se desea estimar la media = .05 $\bar{Y}_N$

n = tamaño de la muestra

Calculemos primeramente las variancias de los estratos de C.

$$S_1^2 = \frac{1}{5}\left\{4\ 270\ 332 - \frac{(4\ 988)^2}{6}\right\} = 2\ \ 728$$

$$\therefore\ S_1 = 157.247$$

$$S_2^2 = \frac{1}{3}\left\{ 9\ 432\ 308 - \frac{(6\ 064)^2}{4} \right\} = 79\ 761$$

$$\therefore \quad S_2 = 282.420$$

$$S_3^2 = 412\ 420 - \frac{(862)^2}{2} = 40\ 898$$

$$\therefore \quad S_3 = 202.233$$

$$S_5^2 = \frac{1}{5}\left\{ 4\ 242\ 044 - \frac{(4\ 974)^2}{6} \right\} = 23\ 720$$

$$\therefore \quad S_5 = 154.013$$

$$S_6^2 = 20\ 400\ 874 - \frac{(6\ 382)^2}{2} = 35\ 912$$

$$\therefore \quad S_6 = 189.505$$

$$S_7^2 = 5\ 528\ 367 - \frac{(6\ 999)^2}{9} = 10\ 685$$

$$\therefore \quad S_7 = 103.368$$

$$S_8^2 = \frac{1}{2}\left\{ 18\ 158\ 238\ \frac{(7\ 310)^2}{3} \right\} = 173\ 103$$

$$\therefore \quad S_8 = 416.056$$

$$S_9^2 = \frac{1}{2}\left\{ 6\ 436\ 817 - \frac{(4\ 347)^2}{3} \right\} = 69\ 007$$

$$\therefore \quad S_9 = 262.692$$

$$S_{10}^2 = \frac{1}{2}\left\{ 597\ 973 - \frac{(1\ 313)^2}{3} \right\} = 11\ 658$$

$$\therefore \quad S_{10} = 107.972$$

$$S_{12}^2 = \frac{1}{5}\left\{ 26\ 293\ 458 - \frac{(12\ 498)^2}{6} \right\} = 52\ 025$$

$$\therefore \quad S_{12} = 228.090$$

$$S_{13}^2 = \frac{1}{6}\left\{ 3\ 668\ 196 - \frac{(4\ 960)^2}{7} \right\} = 25\ 614$$

$$\therefore \quad S_{13} = 160.044$$

Las cantidades $p_i = \frac{N_i}{N}$ serán:

$p_1 = \frac{6}{56} = .107$ $\qquad$ $p_2 = \frac{4}{56} = .071$

$p_3 = \frac{2}{56} = .036$ $\qquad$ $p_4 = \frac{1}{56} = .018$

$p_5 = \frac{6}{56} = .107$ $\qquad$ $p_6 = \frac{2}{56} = .036$

$p_7 = \frac{9}{56} = .161$ $\qquad$ $p_8 = \frac{3}{56} = .054$

$p_9 = \frac{3}{56} = .054$ $\qquad$ $p_{10} = \frac{3}{56} = .054$

$p_{11} = \frac{1}{56} = .018$ $\qquad$ $p_{12} = \frac{6}{56} = .107$

$p_{13} = \frac{7}{56} = .125$ $\qquad$ $p_{14} = \frac{1}{56} = .018$

$p_{15} = \frac{1}{56} = .018$ $\qquad$ $p_{16} = \frac{1}{56} = .018$

En el párrafo correspondiente al intervalo de confianza para la media habríamos obtenido:

$$P\left\{\bar{y} - t\,\frac{s}{\sqrt{n}} \leq \bar{Y} \leq \bar{y} + t\,\frac{s}{\sqrt{n}}\right\} = 1 - \alpha$$

que podemos simplificar a:

$$\bar{y} \pm t\sqrt{Vo}$$

en la que Vo nos representa la variancia de la media de la muestra.

Hagamos:

$$d = t\sqrt{Vo}$$

que será la desviación respecto de la media de la población de la media de la muestra. Fijémonos arbitrariamente una desviación tolerable de 5%:

$$d = .05\,\bar{Y} \qquad \therefore\ Vo = \frac{d^2}{t^2} = \frac{(.05 \times 1\,626)^2}{1.96^2} = \frac{6\,609.69}{3.8416}$$

$$Vo = 1\ 720.557$$

$$\left(\sum_{i=1}^{k} p_i\ S_i\right)^2 = (170.995)^2 = 29\ 239.290$$

$$\frac{1}{N} \sum_{i=1}^{k} p_i\ S_i^2 = \frac{1}{56}\ 37\ 804.309 = 675.077$$

$$\therefore \quad n = \frac{29\ 239.290}{1\ 720.557 + 675.077} = \frac{29\ 239.290}{2\ 395.634} = 12.205$$

$$n = 12$$

Es claro que en aquellos estratos que constan de un solo tramo carece de sentido hablar de Variancia dentro del Estrato ya que para que pueda aplicarse la definición de Variancia se necesita un mínimo de dos observaciones; sin embargo, si se tomaron en cuenta para calcular la Eficiencia de las Estratificaciones en las que interviene la Variancia entre Estratos.'

El criterio a seguir es pues, al calcular el tamaño de la muestra abstraerse momentáneamente de ellos trabajando con el resto de la población incluyendo al final necesariamente dichos estratos en la muestra.

AFIJACION DE LA MUESTRA EN LOS DIFERENTES ESTRATOS SEGUN LA AFIJACION DE NEYMAN

La afijación de la muestra en los diferentes estratos para satisfacer la condición de estimar la media con la precisión especificada habíamos visto que se hace según la expresión:

$$n_i = n \frac{p_i\ S_i}{\sum_{i=1}^{k} p_i\ S_i}$$

$p_1\ S_1 = 16.825$	$p_5\ S_5 = 16.479$	$p_8\ S_8 = 22.467$
$p_2\ S_2 = 20.052$	$p_6\ S_6 = 6.822$	$p_9\ S_9 = 14.185$
$p_3\ S_3 = 7.280$	$p_7\ S_7 = 16.642$	$p_{10} S_{10} = 5.830$

$$p_{12}S_{12} = 24.406 \qquad p_{13}S_{13} = 20.006$$

$$n_1 = 12.205\ \frac{16.825}{170.995} = 1.201 \qquad n_7 = 12.205\ \frac{16.642}{170.995} = 1.188$$

$$n_2 = 12.205\ \frac{20.052}{170.995} = 1.431 \qquad n_8 = 12.205\ \frac{22.467}{170.995} = 1.604$$

$$n_3 = 12.205\ \frac{7.280}{170.995} = .520 \qquad n_9 = 12.205\ \frac{14.185}{170.995} = 1.012$$

$$n_5 = 12.205\ \frac{16.479}{170.995} = 1.176 \qquad n_{10} = 12.205\ \frac{5.830}{170.995} = .416$$

$$n_6 = 12.205\ \frac{6.822}{170.995} = .487 \qquad n_{12} = 12.205\ \frac{24.406}{170.995} = 1.742$$

$$n_{13} = 12.205\ \frac{20.006}{170.995} = 1.428$$

Teniendo en cuenta la aclaración hecha respecto a los estratos de un solo tramo, los tamaños de las muestras en cada estrato serán:

$n_1 = 1$	$n_4 = 1$	$n_7 = 1$
$n_2 = 1$	$n_5 = 1$	$n_8 = 2$
$n_3 = 1$	$n_6 = 1$	$n_9 = 1$
$n_{10} = 1$	$n_{13} = 1$	$n_{16} = 1$
$n_{11} = 1$	$n_{14} = 1$	
$n_{12} = 2$	$n_{15} = 1$	

y el tamaño de la muestra total será: $n = 18$.

SELECCION DE LAS ESTACIONES DE AFORO

Numerando progresivamente las estaciones dentro de cada estrato, - con la ayuda de una tabla de números aleatorios se hará la selección buscando en cualquier columna y cualquier renglón de la tabla los números - comprendidos del uno al mayor número del tamaño de cada estrato; hasta - completar los requeridos por el tamaño de la muestra.

Así, para el primer estrato tendremos:

Número de orden	Nombre de la Estación
1	Zitácuaro-C. Hidalgo
2	Durango-Sombrerete
3	Allende-Sabinas
4	Sabinas-Monclova
5	Tamazunchale-Valles
6	Cuernavaca-Taxco (V)

$n_1 = 1$

Números de orden de las estaciones:

1, 2, 3, 4, 5, 6.

Empezando en la 5a. columna renglón sexto, el primer número comprendido entre 1 y 6 es 5.

O sea, la estación determinada aleatoriamente será la 5a. que corresponde a Tamazunchale-Valles.

Procediendo análogamente con el resto de los estratos tendremos:

Estrato Número	Tamaño de la muestra	Estaciones seleccionadas	Tipo
1	1	Tamazunchale-Valles	I
2	1	Fresnillo-Zacatecas	I
3	1	Antiguo-Morelos-C.del Maíz	I
4	1	Toluca-México	II
5	1	Huichapan-Hixmiquilpan	II
6	1	Venta de Carpio-Colonia	II
7	1	Camaleón-Monclova	III
8	2	Piedras Negras-Allende- Cuautla-Chalco	III
9	1	Papantla-Poza Rica	III

TABLA DE NUMEROS ALEATORIOS

2017	4228	2317	5966	3861	0210	8610	5155	9252	4425
7449	0449	0304	1033	5370	1154	4863	9460	9449	5738
9470	4931	3867	2342	2965	4088	7871	3718	4864	0657
2215	7815	6984	3252	3254	1512	5402	0137	3837	1293
9329	1218	2730	3055	9187	5057	5851	4936	1253	9640
4504	7797	3614	9945	5295	6985	0383	5187	8556	2237
4491	9949	8939	9460	4849	0677	6472	5926	0851	2557
1623	9102	1996	4759	8965	2784	3092	6337	2624	2366
0450	6504	6565	8242	7051	5504	6147	8883	9934	8237
3270	1772	0361	6626	2471	2277	8833	1778	0892	7349
0364	5907	4295	8139	0641	2081	9234	5190	3908	2142
6249	0090	6786	9348	3183	1907	6768	4903	2747	5203
6100	9586	9836	1403	4988	5107	3340	0686	3376	6857
8903	9049	2874	2104	0996	6045	2203	5280	0179	3381
0172	3385	9240	6007	0671	8927	1429	5524	8579	3196
2796	4979	3434	3222	6053	9117	3326	4470	9314	9970
4305	7448	1055	3525	2428	2022	3566	6634	2635	9123
4974	2725	9726	3394	4223	0128	5958	9269	0366	7382
7026	2243	8906	1985	1812	4765	6563	5607	9785	5679
4067	7796	4339	7693	0879	2218	5455	9375	9726	9077
4872	8746	7573	0011	2707	0520	3085	2221	0467	1913
9597	9862	1727	3142	6471	4622	3275	1932	2099	9485
2799	5731	7040	4655	4612	2432	3674	6920	7210	9593
0579	5837	8533	7518	8871	2344	5428	0048	9623	6645
5585	6342	0079	9122	2901	4139	5140	3665	2611	7832
5728	9625	6836	2472	0385	4924	0569	6426	0819	9121
8556	9478	3259	5182	8643	7384	4560	8957	0687	0815
4010	6009	0588	7844	6313	5825	3711	1847	7562	5221
9455	6948	9080	7780	2689	8744	2374	5620	2019	2652
1163	7777	2320	3362	6219	2903	9415	5637	1409	4716
6400	2604	5455	3857	9462	6840	2604	2425	0361	0120
5094	1323	7841	6058	1060	8846	3021	4598	7096	3689
6698	3796	4413	4505	3459	7585	4897	2719	1785	4851
6691	4283	6877	9091	6090	7962	5766	7228	0870	9603
3356	1218	0207	1980	2129	3945	9042	5884	8543	9567
5249	4016	7240	7305	5090	0204	9824	0530	2725	2088
7496	9399	7830	7947	9692	4558	4037	8976	8441	7468
5026	5430	0188	6957	5445	6988	2321	0569	9344	0532
4946	6189	3379	9684	2834	1935	2873	3959	5634	9707
1965	1344	7839	7388	6203	3600	2596	8676	6790	2168
6417	4767	8759	8140	7261	1400	2828	5586	2338	1615
1843	9737	6897	5656	5795	0188	1189	4807	4260	1192
6558	6087	5109	9661	1553	6681	6688	4475	3701	2888
7990	3100	9114	8565	3175	4315	4593	6478	3453	8802
0723	0015	5905	1609	9442	2040	6376	6567	3411	9410
9008	1424	0151	3546	3032	3319	0014	1928	4051	9269
5382	6202	2102	3413	4103	1285	6530	0097	5630	1548
9817	2615	0450	7625	2033	5484	3931	2333	5964	9627
0891	1244	8240	3062	4550	6454	6517	8925	5944	9995
3721	4677	8487	6739	0554	9737	3341	1174	9050	2962

Estrato Número	Tamaño de la muestra	Estaciones seleccionadas	Tipo
10	1	Valles-Tampico	III
11	1	Insurgentes-Ecatepec	III
12	2	Zamora-Zahuayo Tijuana-Mexicali	IV
13	1	Paso del Toro-Veracruz	IV
14	1	C.Juárez-Chihuahua	IV
15	1	México-Los Reyes	IV
16	1	Los Reyes-Sta.Bárbara	IV

Los días en los que habrá de efectuarse el aforo se determinaron mediante un muestreo aleatorio tomando en consideración los trimestres de máximo tránsito, sin por esto descuidar los otros meses del año en los que también se fijaron aleatoriamente días en que habrá de aforarse.

En el trimestre de máximo tránsito se fijaron aleatoriamente 3 días por semana dentro de ese trimestre. En los meses fuera del trimestre de máxima intensidad se determinaron únicamente 5 días por mes que se consideraron suficientes para detectar alguna variación importante en el transcurso de ese mes.

El primer trimestre de 1956 tiene 13 semanas completas.

El segundo trimestre de 1956 tiene 13 semanas completas.

El tercer trimestre de 1956 tiene 13 semanas completas mas 1 día.

El cuarto trimestre de 1956 tiene 13 semanas completas mas 1 día.

Total de días aforados al año:

$$\left(\frac{3 \text{ días}}{\text{semana}} \times 13 \text{ semanas} = 39 \text{ días}\right) + \left(\frac{5 \text{ días}}{\text{mes}} \times 9 \text{ meses} = 45 \text{ días}\right) = 84 \text{ días}$$

Los calendarios se dan a continuación:

Se hizo un solo calendario para cada uno de los 4 tipos de estaciones de aforo según la Estratificación A.

La selección de Gs días se hace con la ayuda de una tabla de números aleatorios.

Hagamos la siguiente correspondencia de los días de la semana con números:

D	L	M	M	J	V	S
1	2	3	4	5	6	7

que nos permitirá seleccionarlos aleatoriamente vgr: si la tabla nos da como número aleatorio el 4, querrá decir que el día de la semana será - el miércoles.

Con este artificio es posible formar un calendario.

Los calendarios ya hechos para cada tipo de estación según la estratificación A se dan a continuación:

Calendario de los días de aforo para estaciones del tipo I

Número de la semana	Números aleatorios	Días de la semana
1	6,7,4	miércoles,viernes,sábado
2	3,6,4	martes,miércoles,viernes
3	1,7,3	domingo,martes,sábado
4	6,3,2	lunes,martes,viernes
5	6,5,2	lunes,jueves,viernes
6	5,6,4	miércoles,jueves,viernes
7	1,7,5	domingo,jueves,sábado
8	1,3,7	domingo,martes,sábado
9	4,2,6	lunes,miércoles,viernes
10	3,2,4	lunes,martes,miércoles
11	4,5,1	domingo,miércoles,jueves
12	4,5,3	martes,miércoles,jueves
13	6,5,1	domingo,jueves,viernes

Mes del año	Día del mes
Abril	6,11,14,23,27
Mayo	5,10,19,30,31
Junio	4, 6,16,17,21
Julio	2,11,16,21,31
Agosto	3,14,19,26,29
Septiembre	9,11,19,23,25
Octubre	2,10,11,22,24
Noviembre	5,8 ,10,18,20
Diciembre	12,16,18,24,25

Calendario de los días de aforo para estaciones del tipo II

Número de la semana	Días de la semana
1	domingo, jueves, sábado
2	lunes, martes, viernes
3	domingo, martes, jueves
4	lunes, jueves, sábado
5	martes, jueves, sábado
6	lunes, miércoles, sábado
7	domingo, jueves, sábado
8	domingo, lunes, jueves
9	lunes, viernes, sábado
10	miércoles, viernes, sábado
11	lunes, jueves, sábado
12	domingo, miércoles, jueves
13	miércoles, jueves, sábado

Mes del año	Día del mes
Enero	8,18,20,21,27
Febrero	7, 8,11,18,21
Marzo	2, 4, 6,26,30
Julio	3,11,21,29,31
Agosto	5,14,25,28,29
Septiembre	5, 8,10,18,20
Octubre	6, 7,19,20,27
Noviembre	3,11,12,21,29
Diciembre	5, 8,12,18,21

Calendario de los días de aforo para estaciones del tipo III

Número de la semana	Días de la semana
1	Domingo,miércoles,sábado
2	Miércoles,jueves,viernes
3	Jueves,viernes,sábado
4	Lunes,miércoles,sábado
5	Lunes,martes,jueves
6	Martes,viernes,sábado
7	Domingo,martes,miércoles
8	Domingo,miércoles,sábado
9	Lunes,jueves,viernes
10	Martes,miércoles,jueves
11	Lunes,jueves,sábado
12	Lunes,martes,viernes
13	Lunes,jueves,sábado

Mes del año	Días del mes
Enero	2,16,20,26,30
Febrero	3,11,12,21,29
Marzo	4, 6,16,17,21
Abril	4,16,23,24,27
Mayo	5,14,25,28,29
Junio	5, 8,10,18,20
Octubre	4, 6,16,17,21
Noviembre	2, 8,11,23,25
Diciembre	4, 7,11,18,28

Calendario de los días de aforo para estaciones del tipo IV

Número de la semana	Días de la semana
1	lunes,martes,viernes
2	lunes,miércoles,jueves
3	domingo,miércoles,jueves
4	martes,viernes,sábado
5	domingo,martes,viernes
6	martes,miércoles,jueves
7	domingo,lunes,miércoles
8	lunes,miércoles,jueves
9	domingo,martes,sábado
10	martes,viernes,sábado
11	domingo,lunes,martes
12	lunes,martes,viernes
13	domingo,martes,jueves

Mes del año	Días del mes
Enero	6,11,23,27,31
Febrero	2,11,14,19,21
Marzo	3,10,11,22,24
Abril	2,14,17,24,26
Mayo	10,15,19,27,29
Junio	1, 5, 8,15,19
Julio	5, 8,10,18,20
Agosto	10,15,18,24,29
Septiembre	2, 6,12,17,25

DETERMINACION DEL HORARIO DE AFORO

Las horas para aforar se determinaron según un Muestreo Sistemático, tomando como período un intervalo de 3 horas dentro de cada turno de 8 horas que puede considerarse de sensibilidad necesaria para poder conocer la tendencia o variación horaria del tráfico.

El adoptar este tipo de muestreo tiene como ventajas la simplicidad de operación, que el número máximo de horas por aforar dentro de cada turno es únicamente 3, lo que permite hacer el conteo de vehículos por conglomerado, es decir, las 3 horas seguidas si el personal empleado es eventual (contratado únicamente para este objeto) ó bien contar los vehículos precisamente en las horas señaladas si el personal es de planta y por lo tanto debe permanecer las 8 horas del turno en su puesto. Evidentemente esta segunda alternativa es preferible.

Si el conteo se hace automáticamente es obvio que el aforo durante las 24 horas del día es óptimo.

El horario determinado es único para todos los tramos de la muestra sin distinción de clasificación a que pertenezcan aunque sí se precisa una hora específica para cada día de aforo.

El horario es el siguiente:

Día	H O R A S								
	Turno 1			Turno 2			Turno 3		
1	2	5	8	11	14	-	17	20	23
2	-	3	6	9	12	15	18	21	24
3	-	3	6	9	12	15	18	21	24
4	2	5	8	11	14	-	17	20	23
5	1	4	7	10	13	16	19	22	-
6	-	3	6	9	12	15	18	21	24
7	2	5	8	11	14	-	17	20	23
8	1	4	7	10	13	16	19	22	-
9	-	3	6	9	12	15	18	21	24
10	2	5	8	11	14	-	17	20	23
11	1	4	7	10	13	16	19	22	-
12	-	3	6	9	12	15	18	21	24
13	1	4	7	10	13	16	19	22	-
14	1	4	7	10	13	16	19	22	-
15	1	4	7	10	13	16	19	22	-

Día	H O R A S								
	Turno 1			Turno 2			Turno 3		
16	1	4	7	10	13	16	19	22	-
17	2	5	8	11	14	-	17	20	23
18	1	4	7	10	13	16	19	22	-
19	2	5	8	11	14	-	17	20	23
20	-	3	6	9	12	15	18	21	24
21	-	3	6	9	12	15	18	21	24
22	1	4	7	10	13	16	19	22	-
23	-	3	6	9	12	15	18	21	24
24	1	4	7	10	13	16	19	22	-
25	1	4	7	10	13	16	19	22	-
26	1	4	7	10	13	16	19	22	-
27	-	3	6	9	12	15	18	21	24
28	-	3	6	9	12	15	18	21	24
29	-	3	6	9	12	15	18	21	24
30	1	4	7	10	13	16	19	22	-
31	2	5	8	11	14	-	17	20	23
32	1	4	7	10	13	16	19	22	-
33	-	3	6	9	12	15	18	21	24
34	-	3	6	9	12	15	18	21	24
35	2	5	8	11	14	-	17	20	23
36	2	5	8	11	14	-	17	20	23
37	-	3	6	9	12	15	18	21	24
38	-	3	6	9	12	15	18	21	24
39	2	5	8	11	14	-	17	20	23
40	1	4	7	10	13	16	19	22	-
41	2	5	8	11	14	-	17	20	23
42	1	4	7	10	13	16	14	22	-
43	1	4	7	10	13	16	19	22	-
44	2	5	8	11	14	-	17	20	23
45	2	5	8	11	14	-	17	20	23
46	2	5	8	11	14	-	17	20	23
47	1	4	7	10	13	16	19	22	-
48	-	3							
49	1	4	7	10	13	16	19	22	-
50	-	3	6	9	12	15	18	21	24
51	-	3	6	9	12	15	18	21	24
52	2	5	8	11	14	-	17	20	23
53	1	4	7	10	13	16	19	22	-
54	2	5	8	11	14	-	17	20	23
55	2	5	8	11	14	-	17	20	23
56	2	5	8	11	14	-	17	20	23
57	-	3	6	9	12	15	18	21	24
58	2	5	8	11	14	-	17	20	23
59	2	5	8	11	14	-	17	20	23
60	1	4	7	10	13	16	19	22	-

Día	H O R A S								
	Turno 1			Turno 2			Turno 3		
61	2	5	8	11	14	-	17	20	23
62	2	5	8	11	14	-	17	20	23
63	2	5	8	11	14	-	17	20	23
64	1	4	7	10	13	16	19	22	-
65	-	3	6	9	12	15	18	21	24
66	1	4	7	10	13	16	19	22	-
67	2	5	8	11	14	-	17	20	23
68	-	3	6	9	12	15	18	21	24
69	-	3	6	9	12	15	18	21	24
70	1	4	7	10	13	16	19	22	-
71	-	3	6	9	12	15	18	21	24
72	1	4	7	10	13	16	18	22	-
73	-	3	6	9	12	15	18	21	24
74	-	3	6	9	12	15	18	21	24
75	2	5	8	11	14	-	17	20	23
76	2	5	8	11	14	-	17	20	23
77	2	5	8	11	14	-	17	20	23
78	1	4	7	10	13	16	19	22	-
79	1	4	7	10	13	16	19	22	-
80	2	5	8	11	14	-	17	20	23
81	1	4	7	10	13	16	19	22	-
82	2	5	8	11	14	-	17	20	23
83	1	4	7	10	13	16	19	22	-
84	1	4	7	10	13	16	19	22	-

CAPITULO V

CONCLUSIONES GENERALES

La determinación así hecha de los tramos carreteros por aforar, - garantiza una muestra representativa de la población (con un 5% tolerado en la desviación de la media - Vo = .05 $\bar{Y}$ -) permitiendo obtener conclusiones que se pueden generalizar para todos los tramos dentro de un mismo estrato. Es decir, la ley de variación del tránsito de vehículos hecha para el tramo aforado, puede suponerse la misma para todos los tra-- mos dentro del mismo estrato a que pertenece ese tramo aforado.

Conocidos los datos de aforo y puesto que la selección de la muestra dentro de cada estrato se ha hecho mediante un muestreo irrestricto-aleatorio, es fácil, obtener mediante la expresión (64) un intervalo de confianza para la media de cada estrato, aceptándose pues que el tránsito medio diario trimestral en esos tramos queda determinado entre esos - intervalos de confianza (con una probabilidad de 95% de certeza -para - = .05, t = 1.96-).

La estimación del tránsito medio diario en el trimestre de máxima intensidad se obtiene a partir de la estimación del tránsito medio horario dado por:

$$\bar{y}_{..} = \frac{1}{nk} \sum_{k=1}^{k} \sum_{j=1}^{n} y_{ij}$$

en la que:

$k = 8$ (horas por turno)

$n = 3$ (número de horas aforadas por turno)

que se multiplicaría por su correspondiente factor de ponderación 24, correspondiendo al número de horas del día.

O sea:

estimación del tránsito diario: $\hat{T}_D = 24\ \bar{y}_{..}$

Evidentemente el tránsito diario máximo en cada estrato estará dado por el extremo superior:

$$\hat{T}_{MD} = \hat{T}_D + 1.96 \frac{s}{\sqrt{n}} \qquad (143)$$

del intervalo, con una probabilidad de 95%.

La determinación de $\hat{T}_{MD}$ puede hacerse para el total de vehícu los que transitaron por cada estación de aforo, o bien puede hacerse — usando expresiones análogas a la (143) para cada tipo de vehículo, conocido el porcentaje de cada uno de ellos reportados por cada estación de aforo.

Puesto que los datos reportados son una Muestra Aleatoria del total de vehículos que transitaron, los porcentajes observados del número de vehículos atendiendo a su calidad, pueden también considerarse como estimaciones insesgadas de los porcentajes de totales que transitaron; es decir, que si los reportes indican un porcentaje p_1 respecto a $\bar{y}_{..}$ de automóviles; p_2 para camiones de 2 ejes; p_3 para camiones de 3 ejes; — etc., los porcentajes p_1, p_2, p_3, ... etc., son estimaciones insesgadas de P_1, P_2, P_3, ... etc., que serán los porcentajes respecto a $\hat{T}_D$ o a $\hat{T}_{MD}$ (según se considere) de cada uno de los tipos de vehículos correspondientes.

O sea:

$$p_k = \frac{\text{Número de vehículos de la clase k reportados}}{\text{Número total de vehículos reportados}} = \hat{P}_k$$

en la que:

$$\hat{P}_k = \text{estimación insesgada de } P_k$$

$$P_k = \frac{\text{Número de vehículos de la clase k que transitaron}}{\text{Número total de vehículos que transitaron}}$$

Conocidos los porcentajes $\hat{P}_k$, la estimación del número de vehículos de la clase k que transitaron es inmediata y con estos resultados se podrán proyectar las carreteras atendiendo no solamente a sus aspectos técnico y económico, sino también a su aspecto funcional para el cual deben estar adecuados.

La reiteración sistemática año tras año de los aforos de vehículos en los tramos incluidos en la muestra, nos pondrá en situación de un mejor conocimiento de la población de tramos, pudiéndose incluso variar la estratificación hecha, buscando una mayor eficiencia.

Las estaciones de aforo seleccionadas en la muestra, pueden variar de año a año o bien conservarse las mismas. El cambio en la muestra de -

las estaciones de aforo el lícito siempre y cuando la nueva muestra seleccionada se haga mediante un Muestreo Probabilístico siguiendo los lineamientos asentados en este estudio.

Lógicamente con el transcurso de los años el mejor conocimiento de la población redundará en una mayor estimación de las medias de los estratos.

Conocida una estimación de la media de un estrato que pueda considerarse consistente, mediante una Prueba de Hipótesis con la expresión:

$$\frac{\mu - \bar{x}}{\frac{s}{\sqrt{n}}}$$

en la que:

μ - media de un estrato considerara consistente.

$\bar{x}$ - media de un tramo por comparar

s - variancia del estrato

n - número de tramos del estrato,

podrá determinarse si hay variación significativa (positiva o negativa) comparándose con $t_{n-1, 1-\alpha}$ en el volumen de tránsito de vehículos para cada tramo por estudiar (No se demostrará). El valor de $t_{n-1,1-\alpha}$ se ve en una tabla de probabilidad para la distribución t. de Student con n-1 grados de libertad y un nivel (de significación α).

Las operaciones numéricas de este trabajo fueron hechas en la S.C.O.P. donde se recibieron los reportes de todas las estaciones de aforo. Se determinaron los grados de saturación para cada tramo. Los grados de saturación se obtuvieron en función del ancho de la corona del camino, grado de las curvas en cada tramo, pendientes máximas etc., mediante expresiones tomadas de la experiencia de otros países.

Se anexa el plano proporcionado por la S.C.O.P. en el que se muestran los tramos con los grados de saturación obtenidos.

Algunos de los tramos que resultaron sobresaturados están ya actualmente en proceso de ampliación.

1.- CAPACIDAD.- Representa el tráfico medio diario anual, expresado en automóviles, que pueden pasar en una sección en ambos sentidos sin que, en más de 30 horas al año los conductores de los vehículos se vean privados, por razones del tráfico, de una libertad razonable para elegir la velocidad a que deseen viajar.

Los factores fundamentales que definen la capacidad del camino - son: ancho de corona, ancho de carpeta, tipo de terreno y tanto por ciento de la longitud total del tramo en el que la distancia de visibilidad - es menor o igual a 500 m (1500 pies) de acuerdo con el criterio que aparece en el libro de Arthur G. Bruce y John Clarkeson "Highway Design -- and Construction" Págs. 101 a 103.

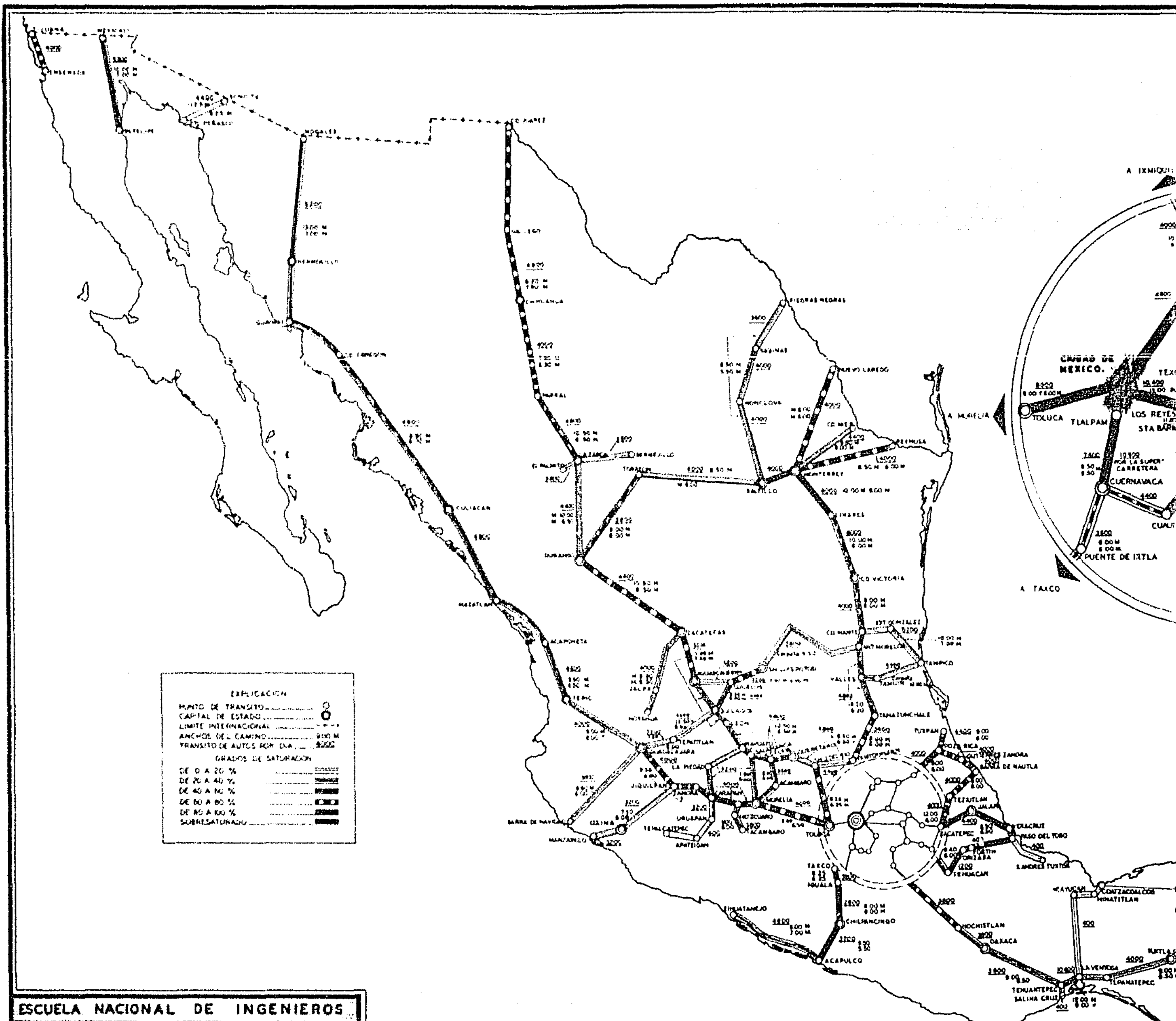
2.- TRAFICO DIARIO MEDIO ANUAL.- Se llama tráfico diario medio - anual en una sección del camino, al número medio de vehículos que pasan - por dicha sección diariamente, en ambos sentidos, considerando un período de un año para valuar el citado promedio.

Los tráficos diarios medios anuales que se indican están expresados en automóviles, teniendo en cuenta la naturaleza del terreno, las características de los vehículos registrados en los aforos y los factores - de conversión señalados en el libro de Arthur G. Bruce y John Clarkeson - "Highway Design and Construction" Pág. 103.

3.- GRADO DE SATURACION.- El término grado de saturación expresa - el porcentaje de la capacidad del camino que representa el tráfico diario medio anual aforado.

4.- AFOROS.- Los aforos utilizados fueron proporcionados por el - Depto. de Planeación Vial quien a su vez los obtuvo de los Departamentos - de Conservación y Cooperación de la Dirección Nacional de Caminos.

En este plano aparecen asciurados longitudinalmente aquellos tra-- mos de los que se recibieron los aforos correspondientes al año de 1956,- así como los datos necesarios para estimar su capacidad; estos tramos pertenecen en general, a los caminos más importantes del país arrojan una - longitud total de 13,015 Km. En aquellos tramos en los que no se dispo-- nía de los datos necesarios para estimar su capacidad o su tráfico diario medio anual, se fijaron éstos por comparación con otros importantes en - los que sí se habían estimado estos valores. Estos tramos no aparecen - asciurados longitudinalmente.



ESCUELA NACIONAL DE INGENIEROS

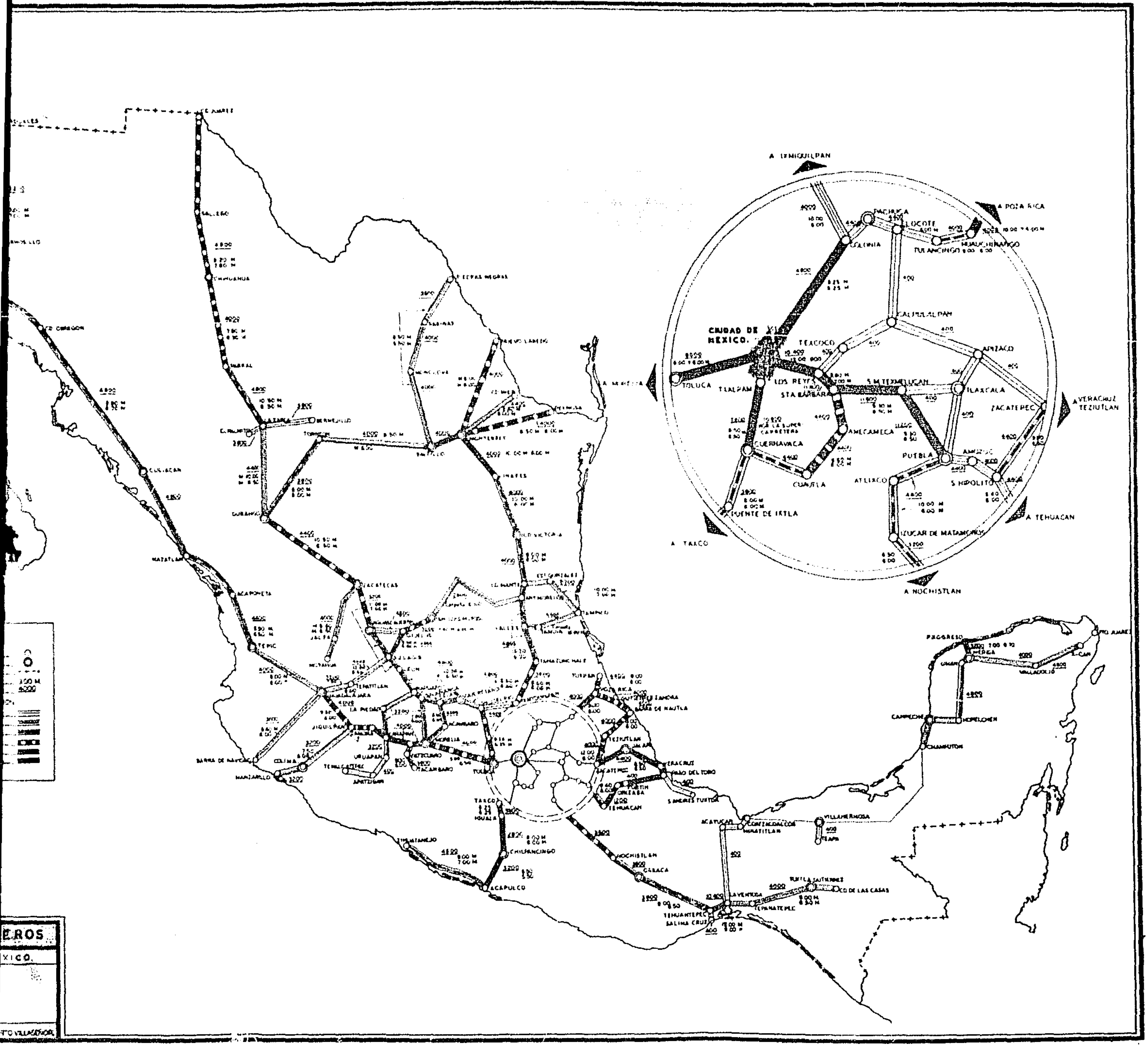
UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO.

GRADOS DE SATURACION DE LAS PRINCIPALES CARRETERAS NACIONALES

TESIS PROFESIONAL

ADOLFO OCAMPO NAVARRO.

FECHA: DIC. 4 DE [illegible] | ESCALA | DIBUJO: ALBERTO VILLASEÑOR



A IXMIQUILPAN
A POZA RICA
PACHUCA
COLONIA
TULANCINGO
HUAUCHINANGO
CALPULALPAN
CIUDAD DE MEXICO
TEXCOCO
APIZACO
TOLUCA
TLALPAM
LOS REYES
STA BARBARA
S.M. TEXMELUCAN
TLAXCALA
ZACATEPEC
A VERACRUZ TEZIUTLAN
AMECAMECA
CUERNAVACA
PUEBLA
S. HIPOLITO
CUAUTLA
ATLIXCO
PUENTE DE IXTLA
A TEHUACAN
IZUCAR DE MATAMOROS
A TAXCO
A NOCHISTLAN
CD. OBREGON
CULIACAN
MAZATLAN
ACAPONETA
TEPIC
CHIHUAHUA
PARRAL
DURANGO
ZACATECAS
TORREON
SALTILLO
MONTERREY
PIEDRAS NEGRAS
SABINAS
MONCLOVA
NUEVO LAREDO
LINARES
CD. VICTORIA
TAMPICO
VALLES
TAMAZUNCHALE
GUADALAJARA
TEPATITLAN
LA PIEDAD
JIQUILPAN
MORELIA
ACAMBARO
URUAPAN
APATZINGAN
COLIMA
MANZANILLO
BARRA DE NAVIDAD
TAXCO
IGUALA
CHILPANCINGO
ACAPULCO
ZIHUATANEJO
TUXPAN
POZA RICA
NAUTLA
TEZIUTLAN
JALAPA
VERACRUZ
CORDOBA
ORIZABA
TEHUACAN
NOCHISTLAN
OAXACA
TEHUANTEPEC
SALINA CRUZ
LA VENTOSA
TUXTLA GUTIERREZ
CD DE LAS CASAS
ACAYUCAN
COATZACOALCOS
MINATITLAN
VILLAHERMOSA
TEAPA
CAMPECHE
CHAMPOTON
HOPELCHEN
MERIDA
PROGRESO
VALLADOLID
PTO JUAREZ
EROS
XICO

BIBLIOGRAFIA

1.- "On Errors in Surveys" — Deming W.E. (1944)

2.- "Some Theory of Sampling" — John Wiley and Sons, Ltd.

3.- "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling - and the Method of Purposive — Selection" — Neyman, J. (1934)

4.- "On the Mathematical Expectation of the Moments of Frequency Distributions in the Case of Correlated Observations" — Tschuprow, A.A.(1923)

5.- "Contribution to the Theory of the Representative Method" — Sukhatme, P.V. (1935)

6.- "On Stratification and Optimun - Allocations" — Evans, W.D. (1951)

7.- "Teoria de Encuestas por Muestreo" — Pandurang V. Sukhatme (1956) (Fondo de Cultura Económica).