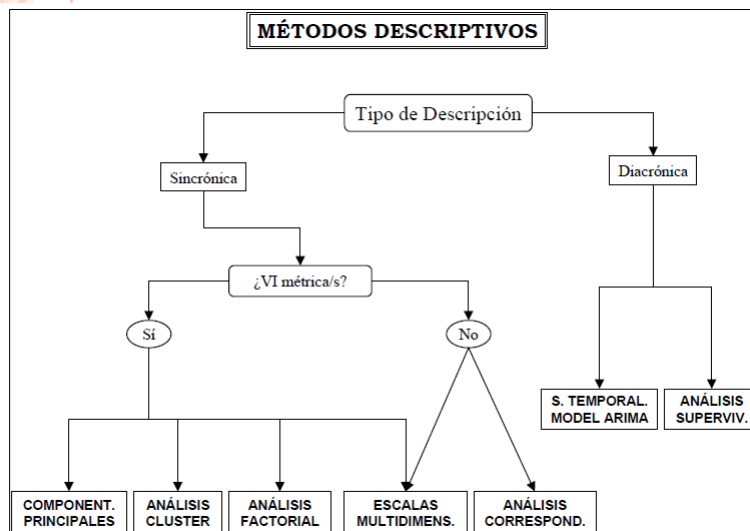


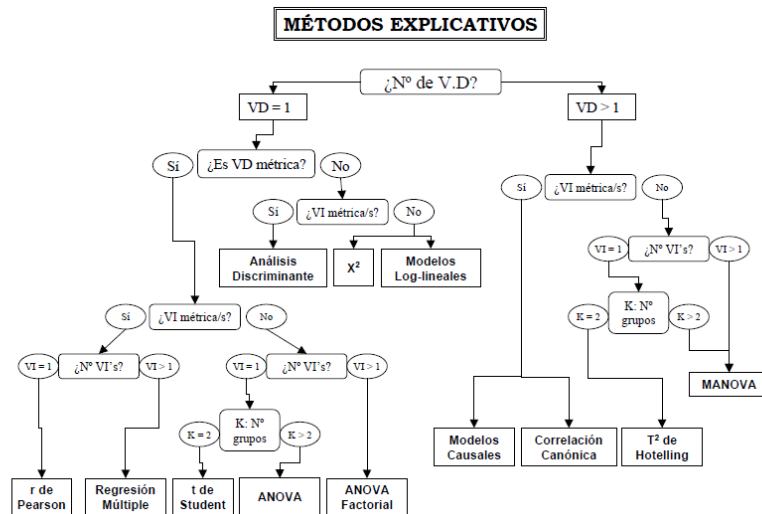
METODOLOGÍA ESTADÍSTICA SIMPLE

METODOLOGÍA ESTADÍSTICA (SIMPLE)

Sólo Afecta A Una o Dos Variables

- **MÉTODOS DESCRIPTIVOS** → Estadística Descriptiva:
Ofrece una Visión Superficial de las Características.
Frecuencias, Tablas de Contingencia o Exploración.
- **MÉTODOS EXPLICATIVOS** → Estadística Inferencial
Ofrece Visión Profunda de Características de Fenómeno.
ANOVA, t-test y Medias





METODOLOGÍA ESTADÍSTICA COMPLEJA

METODOLOGÍA ESTADÍSTICA (MÚLTIPLE O COMPLEJA)

Afecta A Una Multiplicidad de Variables: Métodos Multivariantes

• MÉTODOS JERÁRQUICOS

Estadística que divide las Variables en Dependientes (VD) e Independientes (VI)
Análisis de la Regresión Múltiple

• MÉTODOS NO JERÁRQUICOS

Estadística que NO se divide a las Variables en 2,
Sino que todas ellas tienen Idéntico Status:
Análisis Cluster, Factorial o Escalamiento.

ACLARACIÓN MÁS QUE JUSTIFICACIÓN

- La estadística nos ayuda a corroborar hipótesis dando un soporte matemático a observaciones realizadas. La estadística es la ciencia de la probabilidad y por ello **no es correcto realizar afirmaciones categóricas o negaciones rotundas, sino que estas afirmaciones o rechazos hay que enmarcarlos siempre en un nivel de significación**, que no es más que encuadrarlo dentro de un margen de error que nosotros mismos nos estamos fijando (generalmente entre el 1-5%).
- Lo primero que debe considerarse al realizar un experimento que posteriormente llevará un tratamiento estadístico es:

....

- - Plantear la hipótesis de trabajo que se quiere demostrar.
- - Definir bien las variables a estudiar.
- - Cómo recoger y recopilar los datos (TIPOS DE MUESTREO).
- - Elección del método estadístico más apropiado para demostrar la hipótesis de trabajo de la mejor manera posible.

Es conveniente resaltar que el fin de los muestreos es extraer una muestra lo suficientemente representativa de una población para que **las conclusiones muestrales obtenidas puedan extrapolarse a nivel poblacional, de ahí que sea de suma importancia la minuciosa elección y preparación en la recogida de datos.**

TEORÍA DE MUESTREO

Una parte fundamental para realizar un estudio estadístico de cualquier tipo es obtener unos resultados confiables y que puedan ser aplicables. Como ya se comentó anteriormente, resulta casi imposible o impráctico llevar a cabo algunos estudios sobre toda una población, por lo que la solución es llevar a cabo el estudio basándose en un subconjunto de ésta denominada muestra.

Sin embargo, para que los estudios tengan la validez y confiabilidad buscada es necesario que tal subconjunto de datos, o muestra, posea algunas características específicas que permitan, al final, generalizar los resultados hacia la población en total. Esas características tienen que ver principalmente con el tamaño de la muestra y con la manera de obtenerla.

TEORÍA DE MUESTREO

- A la hora de determinar el tamaño que debe alcanzar una muestra hay que tomar en cuenta varios factores: el tipo de muestreo, el parámetro a estimar, el error muestral admisible, la varianza poblacional y el nivel de confianza. Por ello antes de presentar algunos casos sencillos de cálculo del tamaño muestral delimitemos estos factores.
- **PARAMETRO:** Son las medidas o datos que se obtienen sobre la población.
- **ESTADISTICO:** Los datos o medidas que se obtienen sobre una muestra y por lo tanto una estimación de los parámetros.

TEORÍA DE MUESTREO

- **ERROR MUESTRAL**, de estimación o standard: Es la diferencia entre un estadístico y su parámetro correspondiente. Es una medida de la variabilidad de las estimaciones de muestras repetidas en torno al valor de la población, nos da una noción clara de hasta dónde y con qué probabilidad una estimación basada en una muestra se aleja del valor que se hubiera obtenido por medio de un censo completo. Siempre se comete un error, pero la naturaleza de la investigación nos indicará hasta qué medida podemos cometerlo (los resultados se someten a error muestral e intervalos de confianza que varían muestra a muestra).
Varía según se calcule al principio o al final.

TEORÍA DE MUESTREO

- Un estadístico será más preciso en cuanto y tanto su error es más pequeño. Podríamos decir que es la desviación de la distribución muestral (por distribución muestral se entiende la distribución de frecuencias de los valores de un estadístico en infinitas muestras iguales) de un estadístico y su fiabilidad.
- **NIVEL DE CONFIANZA**: Probabilidad de que la estimación efectuada se ajuste a la realidad. Cualquier información que queremos recoger está distribuida según una ley de probabilidad (Gauss o Student), así llamamos nivel de confianza a la probabilidad de que el intervalo construido en torno a un estadístico capte el verdadero valor del parámetro.
- **VARIANZA POBLACIONAL**: Cuando una población es más homogénea la varianza es menor y el número de entrevistas necesarias para construir un modelo reducido del universo, o de la población, será más pequeño. Generalmente es un valor desconocido y hay que estimarlo a partir de datos de estudios previos.

TEORÍA DE MUESTREO

- A lo largo del curso se haremos uso de dos tipos de razonamiento: el **deductivo** y el **inductivo**. El primero está relacionado directamente con la teoría de probabilidad, que se abordó de manera elemental, y que a partir de las características de la población se obtienen las posibles características de una muestra. El segundo tipo de razonamiento se relaciona con la denominada **inferencia estadística**: utilizar las características de un subconjunto de la población (la muestra) para hacer afirmaciones (inferir) sobre la población en general. Éste será el caso de la parte final del curso.
- El muestro, como ya se mencionó, implica algo de incertidumbre que debe ser aceptada para poder realizar el trabajo, pues aparte de que estudiar una población resulta ser un trabajo en ocasiones demasiado grande, Wonnacott y Wonnacott ofrecen las siguientes razones extras:

TEORÍA DE MUESTREO

- **Recursos limitados.** Es decir, no existen los recursos humanos, materiales o económicos para realizar el estudio sobre el total de la población. Es como cuando se compra un aparato, un automóvil usado (por ejemplo), que se prueba unos minutos (el encendido, una carrerita, etc.) para ver si funciona correctamente y luego se adquiere, pero no se espera a probarlo toda la vida (encendiéndolo y apagándolo o, simplemente, dejándolo encendida) antes de realizar la adquisición.
- **Escasez.** Es el caso en que se dispone de una sola muestra. Por ejemplo, para el estudio paleontológico de los dinosaurios (el T. Rex por ejemplo) sería muy bueno contar con, al menos, muchos restos fósiles y así realizar tales investigaciones; sin embargo, se cuenta sólo con una docena de esqueletos fosilizados (casi todos incompletos) de esas criaturas en todo el mundo.

TEORÍA DE MUESTREO

- **Pruebas destructivas.** Es el caso en el que realizar el estudio sobre toda la población llevaría a la destrucción misma de la población. Por ejemplo, si se quisiese saber el conteo exacto de hemoglobina de una persona habría que extraerle **toda** la sangre.
- **El muestreo puede ser más exacto.** Esto es en el caso en el que el estudio sobre la población total puede causar errores por su tamaño o, en el caso de los censos, que sea necesario utilizar personal no lo suficientemente capacitado; mientras que, por otro lado, el estudio sobre una muestra podría ser realizada con menos personal pero más capacitado.



ALGUNAS CARACTERÍSTICAS PARA HACER INFERENCIAS

- Para calcular el tamaño de una muestra hay que tomar en cuenta tres factores:
 - El porcentaje de confianza con el cual se quiere generalizar los datos desde la muestra hacia la población total.
 - El porcentaje de error que se pretende aceptar al momento de hacer la generalización.
 - El nivel de variabilidad que se calcula para comprobar la hipótesis.
- La **confianza** o el **porcentaje de confianza** es el porcentaje de seguridad que existe para generalizar los resultados obtenidos. Esto quiere decir que un porcentaje del 100% equivale a decir que no existe ninguna duda para generalizar tales resultados, pero también implica estudiar a la totalidad de los casos de la población.



ALGUNAS CARACTERÍSTICAS PARA HACER INFERENCIAS

- Para evitar un costo muy alto para el estudio o debido a que en ocasiones llega a ser prácticamente imposible el estudio de todos los casos, entonces se busca un porcentaje de confianza menor. Comúnmente en las investigaciones sociales se busca un 95%.
- El **error** o **porcentaje de error** equivale a elegir una probabilidad de aceptar una hipótesis que sea falsa como si fuera verdadera, o la inversa: rechazar a hipótesis verdadera por considerarla falsa. Al igual que en el caso de la confianza, si se quiere eliminar el riesgo del error y considerarlo como 0%, entonces la muestra es del mismo tamaño que la población, por lo que conviene correr un cierto riesgo de equivocarse.
- Comúnmente se aceptan entre el 4% y el 6% como error, tomando en cuenta de que **no** son complementarios la confianza y el error.

ALGUNAS CARACTERÍSTICAS PARA HACER INFERENCIAS

- La **variabilidad** es la probabilidad (o porcentaje) con el que se aceptó y se rechazó la hipótesis que se quiere investigar en alguna investigación anterior o en un ensayo previo a la investigación actual. El porcentaje con que se aceptó tal hipótesis se denomina **variabilidad positiva** y se denota por p , y el porcentaje con el que se rechazó se la hipótesis es la **variabilidad negativa**, denotada por q .
- Hay que considerar que p y q son complementarios, es decir, que su suma es igual a la unidad: $p+q=1$. Además, cuando se habla de la máxima variabilidad, en el caso de no existir antecedentes sobre la investigación (no hay otras o no se pudo aplicar una prueba previa), entonces los valores de variabilidad es $p=q=0.5$.

ALGUNAS CARACTERÍSTICAS PARA HACER INFERENCIAS

- Una vez que se han determinado estos tres factores, entonces se puede calcular el tamaño de la muestra como a continuación se expone.
- Hablando de una población de alrededor de 10,000 casos, o mínimamente esa cantidad, podemos pensar en la manera de calcular el tamaño de la muestra a través de las siguientes fórmulas. Hay que mencionar que estas fórmulas se pueden aplicar de manera aceptable pensando en instrumentos que no incluyan preguntas abiertas y que sean un total de alrededor de 30.
- Vamos a presentar dos fórmulas, siendo la primera la que se aplica en el caso de que **no se conozca con precisión el tamaño de la población**, y es:

CALCULAR TAMAÑO DE LA MUESTRA ¿? TAMAÑO PRECISO DE LA POBLACIÓN

- Hay que tomar nota de que debido a que la variabilidad y el error se pueden expresar por medio de porcentajes, hay que convertir todos esos valores a proporciones en el caso necesario.

$$n = \frac{Z^2 p q}{E^2}$$

donde:

n es el tamaño de la muestra;

Z es el nivel de confianza;

p es la variabilidad positiva;

q es la variabilidad negativa;

E es la precisión o error.

Calcular tamaño de la muestra

- También hay que tomar en cuenta que el nivel de confianza no es ni un porcentaje, ni la proporción que le correspondería, a pesar de que se expresa en términos de porcentajes. El nivel de confianza se obtiene a partir de la distribución normal estándar, pues la proporción correspondiente al porcentaje de confianza es el área simétrica bajo la curva normal que se toma como la confianza, y la intención es buscar el valor Z de la variable aleatoria que corresponda a tal área.

POR EJEMPLO....

- Si se quiere un porcentaje de confianza del 95%, entonces hay que considerar la proporción correspondiente, que es 0.95. Lo que se buscaría en seguida es el valor Z para la variable aleatoria z tal que el área simétrica bajo la curva normal desde $-Z$ hasta Z sea igual a 0.95, es decir, $P(-Z < z < Z) = 0.95$.
- Utilizando las tablas, o la función DISTR.NORM.ESTAND.INV() del *Excel*, se puede calcular el valor de Z , que sería 1.96 (con una aproximación a dos decimales).
- Esto quiere decir que $P(-1.96 < z < 1.96) = 0.95$.

CALCULAR TAMAÑO DE LA MUESTRA iii TAMAÑO DE LA POBLACIÓN

La ventaja sobre la primera fórmula es que al conocer exactamente el tamaño de la población, el tamaño de la muestra resulta con mayor precisión y se pueden incluso ahorrarse recursos y tiempo para la aplicación y desarrollo de

$$n = \frac{Z^2 pq N}{NE^2 + Z^2 pq}$$

donde

n es el tamaño de la muestra;

Z es el nivel de confianza;

p es la variabilidad positiva;

q es la variabilidad negativa;

N es el tamaño de la población;

E es la precisión o el error.

POR EJEMPLO....

En el *Colegio de Bachilleres*, una institución de nivel medio superior, se desea realizar una investigación sobre los alumnos inscritos en primer y segundo años, para lo cual se aplicará un cuestionario de manera aleatoria a una muestra, pues los recursos económicos y el tiempo para procesar la información resultaría insuficiente en el caso de aplicársele a la población estudiantil completa.

....

- En primera instancia, suponiendo que no se conoce el tamaño exacto de la población, pero con la seguridad de que ésta se encuentra cerca a los diez millares, se aplicará la primera fórmula.
- Se considerará una confianza del 95%, un porcentaje de error del 5% y la máxima variabilidad por no existir antecedentes en la institución sobre la investigación y porque no se puede aplicar una prueba previa.
- Primero habrá que obtener el valor de Z de tal forma que la confianza sea del 95%, es decir, buscar un valor de Z tal que $P(-Z < z < Z) = 0.95$. Utilizando las tablas o las funciones de *Excel* se pueden obtener, o viendo (en este caso) el ejemplo anterior, resulta que $Z = 1.96$.

....

- En primera instancia, suponiendo que no se conoce el tamaño exacto de la población, pero con la seguridad de que ésta se encuentra cerca a los diez millares, se aplicará la primera fórmula.
- Se considerará una confianza del 95%, un porcentaje de error del 5% y la máxima variabilidad por no existir antecedentes en la institución sobre la investigación y porque no se puede aplicar una prueba previa.
- Primero habrá que obtener el valor de Z de tal forma que la confianza sea del 95%, es decir, buscar un valor de Z tal que $P(-Z < z < Z) = 0.95$. Utilizando las tablas o las funciones de *Excel* se pueden obtener, o viendo (en este caso) el ejemplo anterior, resulta que $Z = 1.96$.

DE ESTA MANERA SE REALIZA LA SUSTITUCIÓN Y SE OBTIENE:

$$n = \frac{(1.96^2)(0.5)(0.5)}{0.05^2} = \frac{(3.8416)(0.25)}{0.0025} = \frac{0.9604}{0.0025} = 384.16$$

Esto quiere decir que el tamaño de la muestra es de 385 alumnos.

Supongamos ahora que sí se conoce el tamaño de la población estudiantil y es de 9,408, entonces se aplicará la siguiente fórmula para obtener el tamaño de la muestra:

$$n = \frac{(1.96^2)(0.5)(0.5)(9408)}{(9408)(0.05^2) + (1.96^2)(0.5)(0.5)} = \frac{9035.4432}{24.4804} = 369.088...$$

Con lo que se tiene una cota mínima de 370 alumnos para la muestra y así poder realizar la investigación sin más costo del necesario, pero con la seguridad de que las condiciones aceptadas para la generalización (confiabilidad, variabilidad y error) se mantienen.

MUESTREO NO PROBABILÍSTICO

- Los elementos de la muestra son seleccionados por procedimientos al azar ó con probabilidades conocidas de selección. Por lo tanto es imposible determinar el grado de representatividad de la muestra.
- Dentro de los tipos de muestreo no Probabilístico, podemos mencionar los siguientes:
- **Muestreo por Juicio, Selección Experta o Selección Intencional:**
- El investigador toma la muestra seleccionando los elementos que a él le parecen representativos o típicos de la población, por lo que depende del criterio del investigador.

MUESTREO NO PROBABILÍSTICO

- **Muestreo casual o fortuito:**
- Se usa en los casos en no es posible seleccionar los elementos, y deben sacarse conclusiones con los elementos que esten disponibles. Por ejemplo: en el caso de voluntarios para pruebas de medicamentos de enfermedades como el corazón, cáncer, etc.
- **Muestreo de cuota:**
- Se utiliza en estudios de opinión de mercado. Los enumeradores, reciben instrucciones de obtener cuotas especificas a partir de las cuales se constituye una muestra relativamente proporcional a la población.
- **Muestreo de poblaciones móviles:**
- Este tipo de muestreo utiliza métodos de captura, marca y recaptura. Se utiliza mucho en el estudio de migración de poblaciones de animales y otras características.

MUESTREOS PROBABILÍSTICOS, ALEATORIOS O ESTOCÁSTICOS

- Las técnicas de **muestreo probabilístico** son aquellas en las que se determina al azar los individuos que constituirán la muestra. Estas técnicas nos sirven cuando se desean generalizar los resultados que se obtienen a partir de la muestra hacia toda la población. Lo anterior se dice dado que se supone que el proceso aleatorio permitirá la obtención de una muestra **representativa** de la población.
- Los muestreos probabilísticos pueden ser **con** o **sin** reemplazo.
- Los **muestreos con reemplazo** son aquellos en los que una vez que ha sido seleccionado un individuo (y estudiado) se le toma en cuenta nuevamente al elegir el siguiente individuo a ser estudiado. En este caso cada una de las observaciones permanece independiente de las demás, pero con poblaciones pequeñas (un grupo de escuela de 30 alumnos, por ejemplo) tal procedimiento debe ser considerado ante la posibilidad de repetir observaciones. En el caso de poblaciones grandes no importa tal proceder, pues no afecta sustancialmente una repetición a las frecuencias relativas.

MUESTREOS PROBABILÍSTICOS, ALEATORIOS O ESTOCÁSTICOS

- Los **muestreos sin reemplazo** son los que una vez que se ha tomado en cuenta un individuo para formar parte de la muestra, no se le vuelve a tomar en cuenta nuevamente. En este caso, y hablando específicamente para el caso de poblaciones pequeñas, las observaciones son dependientes entre sí, pues al no tomar en cuenta nuevamente el individuo se altera la probabilidad para la selección de otro individuo de la población. Para el caso de las poblaciones grandes (por ejemplo la población de un país) dicha probabilidad para la selección de un individuo se mantiene prácticamente igual, por lo que se puede decir que existe independencia en las observaciones.
- Las técnicas de muestreo probabilístico que mencionaremos serán básicamente tres: el aleatorio simple, el aleatorio estratificado y el sistemático.



MUESTREO ALEATORIO SIMPLE

- Podemos aquí mencionar que para el caso de que se estuviese estudiando una proporción dentro de la población (una elección de candidato, la aceptación o rechazo de una propuesta en una comunidad, la presencia o ausencia de una característica hereditaria), y en el caso de un muestreo aleatorio simple, la estimación que se puede hacer de la proporción buscada a partir de la proporción hallada en la muestra se obtiene mediante la construcción de un **intervalo de confianza**:

$$p = P \pm \text{tolerancia de la muestra}$$

- Donde p es la proporción buscada en la población y P es la proporción presente en la muestra.
- Por otro lado, la **tolerancia de la muestra** está relacionada directamente con el nivel de **confianza** y se obtiene a partir de la distribución normal al igual que como se obtuvo para el cálculo del **tamaño de las muestras** con Z para obtener la fórmula:

$$\pi = P \pm Z \sqrt{\frac{P(1-P)}{n}}$$



TAMAÑO DE MUESTRA PARA ESTIMAR LA MEDIA CON M.S.A.

$$n = \frac{N Z_{a/2}^2 S^2}{N d^2 + Z_{a/2}^2 S^2}$$

- de donde:
- n = tamaño de la muestra.
- N = tamaño de la población.
- $Z_{a/2}$ = variable estandarizada de distribución normal.
- S^2 = varianza de la muestra.
- d = precisión del muestreo.
- a = Nivel de significancia.

Generalmente es necesario hacer un premuestreo de 30 elementos, con el objetivo de hacer una primera estimación de S^2 .

Ejemplo:

En un lote de frascos para medicina, con una población de 8000 unidades, se desea estimar la media de la capacidad en centímetros cúbicos de los mismos.

A través de un premuestreo de tamaño 35 se ha estimado que la desviación estándar es de 2 centímetros cúbicos. Si queremos tener una precisión 0.25 cms³, y un nivel de significancia del 5%. ¿De que tamaño debe de ser la muestra ?

MUESTREO ALEATORIO SIMPLE

DATOS:

$S = 2 \text{ cms}^3$; $N = 8000$; $d = 0.25 \text{ cms}^3$; $a = 0.05$ (5%)

$Z_{a/2} = 1.96$

$$n = \frac{N Z_{a/2}^2 S^2}{N d^2 + Z_{a/2}^2 S^2} = \frac{8000(1.96)^2(2)^2}{8000(0.25)^2 + (1.96)^2(2)^2} = 238$$

frascos

Solo faltaría muestrear 203 frascos, pues los datos de los 35 frascos del premuestreo siguen siendo válidos.

TAMAÑO DE MUESTRA PARA ESTIMAR PROPORCIONES CON M.S.A.

En bastantes ocasiones, la variable bajo estudio es de tipo binomial, en ese caso para calcular el tamaño de muestra bajo el muestreo simple aleatorio, se haría de la siguiente manera:

$$n = \frac{N Z_{a/2}^2 pq}{Nd^2 + Z_{a/2}^2 pq}$$

de donde:

- p = probabilidad de éxito.
- q = probabilidad de fracaso.
- d = precisión expresada en porcentaje.
- en este caso para la estimación de la varianza, tenemos dos opciones:
- a) hacer un premuestreo.
- b) asumir varianza máxima.

...EJEMPLO

En un estudio, se desea determinar en que proporción los niños de una región toman incaparina en el desayuno. Si se sabe que existen 1,500 niños y deseamos tener una precisión del 10 por ciento, con un nivel de significancia del 5% . De que tamaño debe de ser la muestra?.

DATOS:

N = 1500 ; d = 10 % = 0.1 ; a = 5 %; $Z_{a/2} = 1.96$
 $p = 0.5$ y $q = 0.5$ (asumiendo varianza máxima).

$$n = \frac{N Z_{a/2}^2 pq}{d^2 + Z_{a/2}^2 pq} = \frac{1500 (1.96)^2(0.5)(0.5)}{1500(0.1)^2 + (1.96)^2(0.5)(0.5)} = 91$$

Se deben de muestrear 91 niños.

MUESTREO ALEATORIO ESTRATIFICADO

- Una muestra aleatoria estratificada es la obtenida mediante la separación de los elementos de la población en grupos que no presenten traslapes, llamados estratos y la selección posterior de una muestra irrestrictamente aleatoria simple en cada estrato.
- En resumen, los motivos principales para utilizar un muestreo aleatorio estratificado son los siguientes:
- a) La estratificación puede producir un error de estimación más pequeño que el que generaría un m.s.a. del mismo tamaño. Este resultado es particularmente cierto si las mediciones dentro de los estratos son homogéneas.
- b) El costo por observación en la encuesta puede ser reducido mediante la estratificación de los elementos de la población en grupos convenientes.

MUESTREO ALEATORIO ESTRATIFICADO

- c) Se pueden obtener estimaciones de parámetros poblacionales para subgrupos de la población. Los subgrupos deben de ser entonces estratos identificables.
- Lo anterior debe de tomarse en cuenta cuando se está planeando estratificar o no una población o decidiendo en que forma se definirán los estratos.

TAMAÑO DE MUESTRA PARA ESTIMAR LA MEDIA CON M.A.E.

de donde:

N_i = tamaño del i ésimo estrato.

N = tamaño de la población.

S_i^2 = varianza del i ésimo estrato.

w_i = importancia o peso del i ésimo estrato.

$$n = \frac{\sum N_i^2 S_i^2 / w_i}{N^2 D + \sum N_i S_i^2}$$

$$D = \frac{B^2}{4} \quad B = \text{precisión.}$$

EJEMPLO

- En un Ingenio, desea hacer una estimación del promedio de grados Brix con que llega la caña a la fabrica.
- Para el efecto desea realizar un muestreo aleatorio estratificado, puesto que la caña puede provenir de tres tipos de proveedores.
- proveedor tipo A (estrato 1) la caña proviene de lotes de la misma finca.
- proveedor tipo B (estrato 2) la caña proviene de fincas de particulares en donde el ingenio ha prestado servicios.
- proveedor tipo C (estrato 3) la caña proviene de fincas de particulares en donde el ingenio no ha tenido ningún servicio.
- De estudios anteriores, se conoce el tamaño y desviación estándar de cada estrato y además se desea tener una precisión de un grado brix en el estudio. De que tamaño debe de ser la muestra total y de cada estrato?.

DATOS:

ESTRATO	N_i	S_i	w_i^*
1	558	3.5	$558/998 = 0.56$
2	190	5.4	$190/998 = 0.19$
3	250	6.2	$250/998 = 0.25$

$$N = \sum N_i = 998$$

* con distribución proporcional.

$$n = \frac{\sum N_i^2 S_i^2 / w_i}{N^2 D + \sum N_i S_i^2}$$

...Ejemplo

$$\sum N_i^2 S_i^2 / w_i = N_1^2 S_1^2 / w_1 + N_2^2 S_2^2 / w_2 + N_3^2 S_3^2 / w_3$$

$$\sum N_i^2 S_i^2 / w_i = (558)^2(3.5)^2/0.56 + (190)^2(5.4)^2/0.19 + (250)^2(6.2)^2/0.25 = 6,811,087.5 + 5,540,400 + 9,610,000$$

$$\sum N_i^2 S_i^2 / w_i = 21,961,87.5$$

$$\sum N_i S_i^2 = N_1 S_1^2 + N_2 S_2^2 + N_3 S_3^2$$

$$\sum N_i S_i^2 = 558(3.5)^2 + (190)(5.4)^2 + (250)(6.2)^2$$

$$\sum N_i S_i^2 = 6835.5 + 5540.4 + 9610 = 21,985.9$$

...Ejemplo

$$D = \frac{1^2}{4} = 0.25$$

$$N^2D = (998)^2(0.25) = 249,001$$

$$n = \frac{\sum N_i^2 S_i^2 / w_i}{N^2D + \sum N_i^2 S_i^2} = \frac{21,961,487.5}{249,001 + 21,985} = 81$$

como se utilizó distribución proporcional, a cada estrato le tocaría el siguiente tamaño de muestra:

$$n_1 = 81(558/998) = 45 ; \quad n_2 = 81(190/998) = 15$$

$$n_3 = 81(250/998) = 20.$$

Gracias por su atención