## VI.6.3 REPLICADO:

**Aquí suponga que se usaron los 5 diseños replicado: 5 muestras de 80 personas fueron seleccionadas de la población; de cada una de las 125 zonas registradas**

| Replica | Nº entrevistas | T.V. encendidas | P |
|---|---|---|---|
| 1 | 80 | 59 | 74 % |
| 2 | 80 | 57 | 71 % |
| 3 | 80 | 61 | 76 % |
| 4 | 80 | 53 | 66 % |
| 5 | 80 | 62 | 78 % |
| Total | 400 | 292 | 73 % |

$$\sigma_{\bar{x}} = \left|\frac{0.78 - 0.66}{5}\right| \sqrt{\frac{5(125-5)}{125 * 4}} = 0.026 \text{ ó } 2.6\%$$

El intervalo es 73 % ± 2 (2.6 %) ó entre 67.8 % y 78.2 %.

Vemos que el menor error estándar se obtiene en el muestreo estratificado, razón por la que siempre se recomienda usarlo.

## VI.6.4 TAMAÑO DE LA MUESTRA

Por su importancia derivada de los ejemplos anteriores, veamos de nuevo como se obtiene el tamaño de la muestra (n) a partir de las fórmulas del error estándar, en este caso de una proporción.

Se toma una muestra para estimar entre otras cosas, la proporción de familias viendo T.V. en la tarde entre semana.

Se desea que ese estimador esté entre el 5 % del porcentaje actual con 95% de seguridad.

N = 10,000
$S^2$= para un porcentaje = p*q
p = .5 por seguridad, es decir, trabajando con variancia máxima.

$\sigma_p$ debe ser tal que $2\sigma_p$ incluyan el 95% de los estimadores de p, luego $2\sigma_p$ = 0.05 de aquí qué $\sigma_p$=0.025

De $\sigma_p = \sqrt{p * q \frac{N-n}{N * n}}$ ; $\sigma_p^2 = \frac{pqN - npq}{Nn}$ tenemos $n(N\sigma_p^2 + pq) = Npq$

$n = \frac{Npq}{pq + N\sigma_p^2}$ entonces

$$n = \frac{(0.25)(10,000)}{(0.25) + 10,000(0.025)^2} = 385 \text{ familias}$$

Vemos que el tamaño apropiado seria de 385 familias y no 400 para hacer la investigación.

De manera similar, se puede obtener los tamaños de muestra para cada uno de los modelos muestrales bajo estudio.

Resumiendo: en forma didáctica y sencilla hemos expuesto las características de los principales diseños de muestreo de mayor uso en economía. En este sentido, como una de sus aplicaciones es en la elaboración de ENCUESTAS, a continuación presentamos la relación de actividades que deben de efectuarse para hacer una encuesta.

Para afianzar el conocimiento decidí explicar algunas de las actividades, como son las siguientes:

**i) DISEÑO DEL CUESTIONARIO**

El diseño de un cuestionario, involucra la consideración de un sin número de aspectos diferentes, de los cuales quizá los más importantes son:
- Los objetivos del estudio;
- La forma que debe tener;
- Si contendrá preguntas abiertas - codificación previa o posterior de las preguntas;
- La forma como se harán las preguntas;
- La organización e instrucción del cuestionario; etc.

Lo que también indudablemente determina su diseño es el tipo de datos que se desean obtener; el método usado para obtenerlos y en última instancia el uso de los resultados. Adicionalmente, podría señalarse que el diseño depende fuertemente de los antecedentes y experiencias del investigador, el tipo de entrevistadores disponibles, costo y tiempo.

Así, basándose en los formatos de la tabulación del guión de información, los rangos probables de variación tomados de las experiencias anteriores - si las hay - y las posibles respuestas, el cuestionario debe diseñarse en forma simple, fácil de seguir y si es posible atractiva.

Lo último es particularmente importante en el caso de los cuestionarios que se envían por correo, donde la decisión de los miembros de la muestra, sobre llenarlo o no, depende de la impresión que tengan sobre la apariencia del cuestionario. Al respecto, se aconseja recabar la información a través de entrevistas directas, ya que el enumerador puede inmediatamente captar los datos en forma precisa o corregirlos cuando el caso lo amerite.

El formato del cuestionario puede tener entradas múltiples o una sola; puede ser de preguntas cerradas o abiertas; las respuestas pueden estar precodificadas o no; cuando las preguntas son abiertas, las respuestas se codificarán con base en un INSTRUCTIVO DE CODIFICACIÓN.

**ii) TRABAJO DE CAMPO**

Es conveniente mencionar que existen diversos métodos para la recolección de datos, de los cuales los principales son:

a) La selección de la muestra a partir de la información de los archivos de la empresa. Así, una muestra puede ser escogida sin mayor problema y al mismo tiempo los datos pueden ser obtenidos con alto grado de confianza a un costo relativamente bajo. Además de que la muestra puede mantenerse continuamente sin representar mayores cargos o esfuerzos extraordinarios;

b) Métodos de observación: La recolección de los datos por observación, es otro instrumento que indirectamente capta la información. Como la información interna, este método no requiere contacto directo con los elementos de la muestra. Estos métodos se utilizan observadores humanos y/o mecánicos, prefiriendo los primeros en casos donde haya que distinguir; por ejemplo: los adultos de los niños, o las personas por sexo.

c) Entrevistas telefónicas: cuando se puede aplicar este método resulta altamente eficiente en la recolección directa de la información. Lo anterior, se debe a que la población virtualmente esta contenida en un directorio y la selección de la muestra, se convierte en una actividad de rutina. Las entrevistas son de lo más económico -excepto cuando hayan que hacerse bastante llamadas de larga distancia- y los datos se obtienes rápidamente. Sin embargo, como los demás métodos, también tiene sus limitaciones. Obviamente no es aplicable si las entrevistas comprenden cuestiones visuales - publicidad, pruebas de interpretación, etc. - . A la vez, información altamente personal se obtiene con menos éxito por teléfono que -por ejemplo.- a través de una entrevista personal.

d) Entrevistas personales: dentro de las formas directa de obtener los datos, este método es sin lugar a dudas el más popular, por referirse a una conversación directa " frente a frente " entre un miembro de la muestra y el entrevistador. Como resultado, se puede obtener una gran variedad de información con este método, el cual es flexible en varios sentidos. Por ejemplo, los datos pueden ser registrados en grabadora o en cuestionarios.

La construcción de los cuestionario es un arte en si; requiere numerosas precauciones para evitar respuestas sesgadas.

Desde el punto de vista de la obtención de los datos, puede decirse que existen dos formas de entrevistar: En un extremo se haya la entrevista altamente estructurada, en la cual se prepara un cuestionario formal y las preguntas se hacen bajo instrucciones precisas y el entrevistador mantiene un orden estricto para su contestación.

Esta forma, se usa generalmente para obtener una variedad de información diferente acerca de una materia, siguiendo algún orden particular. Esta forma en cierto modo, evita que la información recabada refleje sesgos debidos a juicios personales de los enumeradores.

En el otro extremo esta la entrevista carente de formalidad para la cual no se requiere un cuestionario, basta una lista de preguntas generales o temas relacionados con la información que se busca.

Dentro de estos extremos, existen varias combinaciones. El enumerador puede usar un cuestionario estructurado, pero se le permite hacer las preguntas como el quiera.

Como podrá intuirse, el enumerador es la piedra angular de una entrevista, indistintamente de la forma que adopten para entrevistar o cual sea la unidad de muestreo. Si está debidamente entrenado (a), no solamente entrevistará a un mayor número de personas, sino que los datos serán más confiables.

Parece que los mejores enumeradores son personas entre los 25 y 50 años, que tienen una evidente disposición, son inteligentes, poseen cierta cultura, son flexibles y precisos en sus hábitos de trabajo.

Indudablemente que la experiencia es útil, pero si se proporciona un buen entrenamiento puede no ser necesaria. En ciertos tipos de nuevas encuestas, la experiencia puede ser una limitante, ya que se requiere que el enumerador siga procedimientos contrarios a los acostumbrados en el pasado.

Por lo que se refiere a la organización y control del trabajo de campo, como las demás etapas requiere una programación de tiempos y actividades para asignar al personal correspondiente. Dentro de los aspectos básicos esta la fijación de las rutas de trabajo, el plan de trabajo o forma de entrevistar y la supervisión -sobre todo- cuando el grupo de trabajo es numeroso o la captación de los datos presentan dificultades.

**iii) CRITICA DE CUESTIONARIOS**

Los cuestionarios, codificados o no previamente, llegan a la oficina con el orden y presentación de las respuestas dadas por los enumeradores. En algunas ocasiones el trabajo se realiza de acuerdo a las instrucciones establecidas y enseguida pasa al departamento de captura, para ser procesado inmediatamente. Sin embargo, en la mayoría de los casos se requiere una crítica o revisión cuidadosa ya que:

a) Pueden traer las respuestas ilegibles;
b) El orden en que aparecen las respuestas no es el indicado;
c) Se contradicen unas respuestas con otras al compararse entre si;
d) Existen preguntas que vienen en blanco y debían haberse contestado en alguna u otra forma etc.
e) Se requiere preparar los cuestionarios para la codificación de las respuestas; y
f) Se desea verificar la autenticidad de los datos y preliminarmente comprobar ciertas hipótesis establecidas en la programación inicial de actividades, etc.

Tal que en esta etapa la información debe quedar depurada y ordenada hasta donde sea posible para su posterior transformación y vaciada en formatos previamente diseñados. En algunos casos se acostumbra usar la computadora -filtrado electrónico- para realizar esta etapa.

**iv) CODIFICACIÓN Y PROCESAMIENTO DE DATOS**

Una vez que los datos han sido obtenidos y revisados, deben ser procesados para hacer posible un análisis del fenómeno estudiado. Es generalmente aceptado que esta actividad es un tanto tediosa, pero también que es crítica para asegurar exactitud en los resultados.

Una tabulación hecha sin cuidado puede viciar una buena planeación y el método de obtención de los datos. Así mismo, los peligros de los sesgos a un se presentan en los procesos de preparación, clasificación y tabulación.

Esta etapa esta fuertemente ligada a la anterior, ya que, por ejemplo, la preparación consiste en la inspección de cuestionarios o cualquier otra forma usada para captar los datos, su exactitud, si están completos o no, la inspección de trabajo de campo, arreglos o eliminación de respuestas por su inconsistencia o desconfianza la clasificación o estandarización de los datos en base comunes y sobre todo su preparación para ser tabulados.

**v) CLASIFICACIÓN.**

Es el arreglo de los datos en clases o categorías para ser manipulados de acuerdo con la verificación de la hipótesis de trabajo.

**vi) TABULACIÓN.**

La tabulación es la etapa que sucede inmediatamente después a la crítica de cuestionarios y es un conjunto de procedimientos que se adoptan para la recopilación o vaciado de los datos en cuadros. Estos últimos comprenden las diferentes relaciones que se establecen entre las variables comprendidas en el estudio, así, habrá cuadros de una sola entrada, doble entrada, etc.

Los datos pueden ser tabulados manualmente o mecánicamente. La tabulación manual se aconseja cuando las encuestas son pequeñas, existen problemas de presupuesto o no hay ninguna posibilidad de procesar los datos electrónicamente. Por el contrario, cuando la encuesta es grande, la tabulación manual, además de tardada acarrea el riesgo de cálculos erróneos por lo voluminoso de la información, aconsejándose el uso de las computadoras. Por ello, será necesario que la información sea capturada y se diseñan los programas que calcularán los datos de acuerdo con instrucciones específicas.

**vii) EVALUACIÓN ESTADÍSTICA DE RESULTADOS**

El análisis de los datos recabados con la muestra, incluye indicaciones del valor hasta el cual las estimaciones derivadas de la muestra pueden desviarse de los valores verdaderos de la población. Esta evaluación debe comprender datos sobre la precisión de los estimadores, sobre todo si la selección ha sido probabilística, así como consideraciones sobre algunos sesgos en la operación de reconocimiento que tienda a distorsionar el valor de los estimadores.

Dentro de los sesgos puede considerarse las "no respuestas", cobertura, influencia de los enumeradores sobre la unidad de muestre entrevistado y lo que anoten en el cuestionario, una codificación de respuestas inadecuada, etc.

Por lo que se refiere a la precisión esta se refiere al error de muestreo de un estimador: mientras más pequeño sea el error, mejor será la precisión. El error de muestreo se mide con la fórmula del error estándar, la cual varía de acuerdo con el tipo de estimador - media, mediana, razón, etc. y con el diseño muestral.

La exposición de las fórmulas de los errores estándar se presentan en la sección de los métodos de muestreo, donde se deducen de las varianzas de los estimadores - media, total, etc.

**viii) DISEÑO DE LOS FORMATOS DE TABULACIÓN :**

Los requerimientos de información y las relaciones significativamente importantes, deben exhibirse en estos formatos con claridad y sencillez, dado que con el éxito que esto se logre, la solución del problema será más convincente y fácil. Deben definirse aquí los títulos de todos los cuadros.

**ix) DISEÑO DEL CUESTIONARIO E INSTRUCTIVO :**

Basándose en los formatos de tabulación, del guión de información, de los rangos probables de variación, de las experiencias anteriores y de las posibles respuestas de las preguntas, hágase el diseño de un cuestionario precodificado, procurando y verificando que no se omita ningún concepto, que el llenado del cuestionario, sea lo más sencillo y rápido posible, que el encadenamiento de las preguntas sea el más adecuado, que algunas preguntas sirvan para comprobar las respuestas de otras, etc. Un cuestionario precidificado asigna en cada pregunta un conjunto de claves numéricas, correspondiendo en forma biunívoca, en el conjunto de las posibles respuestas, esta claves se anotan cifra por cifra, en las posiciones -en cuadrícula- que se hayan designado para el caso.

**x) INVESTIGACIÓN SOBRE FUENTES DE INFORMACIÓN**

Un marco muestral es un conjunto de listas o de mapas, o una combinación de estos elementos, de tal manera, que todas las unidades de interés estén contenidas y que al seleccionar las muestra se pueda determinar la probabilidad de su inclusión, asimismo en el momento de levantar la encuesta, la identificación de cada unidad en la muestra sea posible hacerla sin ninguna ambigüedad.

Para obtener un marco muestral puede recurrirse a ciertas instituciones y recopilar además , datos para: calcular el tamaño de la muestra , confrontar y complementar los resultados de la encuesta, determinar aproximadamente algunos rangos de variación, etc., si es que en los antecedentes -archivos propios- no se tienen.

**xi) PRUEBA DEL CUESTIONARIO Y AJUSTES FINALES.**

Con objeto de determinar cuáles ajustes deben hacerse al cuestionario para poder lograr los objetivos en forma satisfactoria, es necesario realizar algunas entrevistas en el campo de estudio, llenar los cuestionarios correspondientes y evaluar los resultados a este nivel.

**xii) FORMULACIÓN DEL GUIÓN DE INFORMACIÓN**

Partiendo de un examen del problema, se recomienda hacer una relación de todas las variables, cuyos valores puedan ser significativamente relevantes, en la resolución del problema.

**xiii) OBTENCIÓN DE INFORMACIÓN COMPLEMENTARIA**

Prepárese todo el material que sea necesario, como oficios debidamente dirigidos y firmados, formas para captar información, etc. Los métodos de muestreo tienen por objeto indicar el número de unidades que deben incluirse en la muestra, dependiendo de la forma como éstas se seleccionen, del nivel de confianza que se requiera, del error de muestreo permisible y del fondo disponible para la realización de la encuesta.

**xiv) LEVANTAMIENTO DE LA ENCUESTA**

El trabajo de los enumeradores debe hacerse exactamente con las unidades de última etapa, determinadas en la selección de la muestra y si ello no fuera posible por deficiencias en el marco muestral, resuélvase el problema con apego a las instrucciones precisas que se hayan hecho para estos pasos. Al hacerse las preguntas, téngase cuidado de que las respuestas sean correctas y veraces, considerando los rangos aproximados para los valores que puedan tomar las variables involucradas en el estudio.

**xv) SUPERVISIÓN DEL LEVANTAMIENTO DE LA ENCUESTA**

Es conveniente utilizar una forma de reporte, en la cual el supervisor anote cómo se desarrolla el levantamiento de la encuesta, esto es, registrar el material recibido y entregado, folio de los cuestionarios entregados a su grupo, casos de no respuesta y especificación de la resolución tomada, folio de los cuestionarios que fue necesario aclarar, número diario de cuestionarios entregados y de errores por enumerador, porcentaje del avance total del trabajo -llenado de cuestionarios-, día y hora para cada reporte a oficinas centrales, números de cuestionarios efectivamente llenados al terminar la encuesta y registro de los demás documentos recogidos, calificación final de los enumeradores, etc.

**xvi) ADMINISTRACIÓN DEL LEVANTAMIENTO DE LA ENCUESTA:**

Se refiere a todas las actividades, como:

-Autorización de gastos y obtención de fondos junto con las directrices administrativas para su uso;
-Acuse de lo recibido a oficinas centrales;
-Pago del trabajo de campo;
-Observación del sistema de envíos;
-Tiempos transcurridos entre envío y recepción;
-Condición de llegada del material;
-Retroalimentación de las experiencias de la fase inicial y ajuste en donde ello sea necesario;
-Registro de aquéllos procedimientos -o personas- que no funcionaron para referencias futuras y para obtener de ello una experiencia;
-Terminación de obligaciones con el personal eventual; etc.

**xvii) CRÍTICA DE LOS CUESTIONARIOS Y DETERMINACIÓN DEL TAMAÑO EFECTIVO DE LA MUESTRA:**

Esto es, hacer un filtrado de todos los errores que no hayan sido detectados por los supervisores, así como también verificar y concentrar el número total de cuestionarios encomendados a cada supervisor, para obtener el tamaño efectivo de la muestra.

**xviii) ANÁLISIS Y DETERMINACIÓN DE LOS ESTÁNDARES DE TRABAJO:**

Basandose en el trabajo realizado, al probar el cuestionario y en experiencias anteriores, determínese el número de cuestionarios por individuos y por día como: cargo de trabajo, número de visitas antes de declarar la no respuesta, mínimo de los rangos de variación para algunas variables, etc.

**xix) DETERMINACIÓN DETALLADA DE TODOS LOS CONTROLES ADMINISTRATIVOS Y DE TRABAJO**

Elaboración de todos los mecanismos relativos al procedimiento de pago, métodos de retiro de fondos, de adquisición de materiales, de órdenes de trabajo, autorizaciones necesarias, procedimientos para pago de impuestos, definición de obligaciones por ambas partes, procedimientos de envíos, etc.

Redacción de formas para: para reportes del avance del trabajo, registro de personal - con número de clave del registro federal de causantes -, credenciales de identificación, contratos, etc.

Establecimiento de los requisitos del personal de campo: escolaridad, disponibilidad de tiempo, sexo, número de ello en las diversas plazas y zonas de la encuesta, preferencia de contratar maestros, trabajadores sociales, enfermeras, etc.

**xx) ENTRENAMIENTO DE SUPERVISORES:**

En todas las actividades, tanto administrativas, como de trabajo de campo.

**xxi) PROCEDIMIENTOS PARA EL VIAJE DE SUPERVISORES:**

Habilitación de supervisores de campo con visticos, vehículos, boletos, credenciales, cartas de presentación debidamente dirigidas y firmadas etc.

**xxii) APROVISIONAMIENTO DE MATERIALES PARA TRABAJO:**

Como cuestionarios, instructivos, fotografías, planos, cintas métricas, etiquetas, ligas, requisiciones para vehículos, gasolina, papelería, tabletas para escribir en el campo, lápices, gomas, plumones, tarjetas con direcciones importantes y teléfonos, instrucciones para casos de emergencias, formas para control y registro del material, para el avance del trabajo, etc.

**xxiii) ORGANIZACIÓN DE LOS MATERIALES DE TRABAJO:**

En paquetes para las diversas zonas y unidades; ¿a quién van dirigidos? controles de entradas y salida de documentos.

**xxiv) ANÁLISIS ESTADÍSTICO DE LOS RESULTADOS:**

Partiendo de una evaluación de la no respuesta y la no cobertura, los intervalos de confianza y los coeficientes de variación se puede obtener un criterio, acerca de la validez de las estimaciones hechas a partir de la muestra, de tal manera que el confrontar algunos resultados de la encuesta, con cifras de otras fuentes, se podrá determinar el origen de las posibles discrepancias

significativas, como coberturas distintas, error por no respuesta, error de muestreo, etc.
Este análisis debe de tenerse en cuenta para el reporte final.

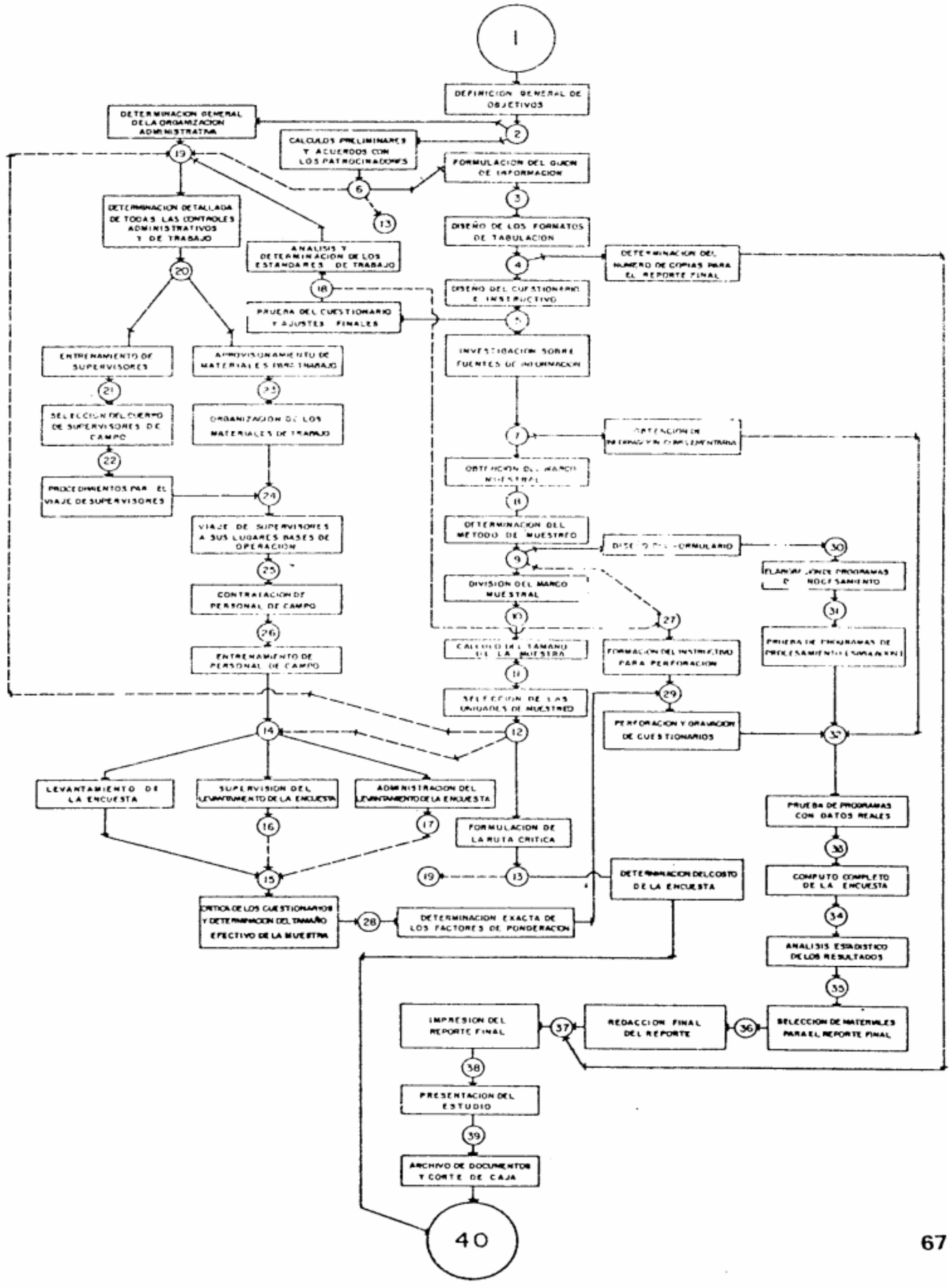
**xxv) ARCHIVO DE DOCUMENTOS**

Envío del reporte a la biblioteca para su conservación y clasificación. Destrucción de cuestionarios y otros materiales no utilizables. Control sobre estas operaciones. Felicitaciones al personal involucrado. Cierre de libros y disposición final de fondos. Terminación final de la encuesta.

Por otra parte y considerando que la economía y los negocios son diversos y complejos, decidí incluir para los especialistas una relación adicional de 10 modelos de muestreo, que complementan los anteriores y brindan al lector una gama de alternativas para seleccionar el método apropiado para la investigación especifica que pretenda hacer.

## VI.6.5 RED GENERAL DE ACTIVIDADES EN UNA ENCUESTA DE MUESTREO

3.1.8 RED GENERAL DE ACTIVIDADES EN UNA ENCUESTA DE MUESTREO




67

SINTESIS DE LAS CARACTERISTICAS DE ALGUNOS MODELOS DE MUESTREO

| DENOMINACION DEL MODELO DE MUESTREO | MEDIA POBLACIONAL | ESTIMADORES | | VARIANCIA DEL ESTIMADOR DE LA MEDIA POBLACIONAL | TAMAÑO DE LA MUESTRA |
|---|---|---|---|---|---|
| | | MEDIA POBLACIONAL | VARIANCIA DEL ESTIMADOR DE LA MEDIA POBLACIONAL | | |
| Muestreo monoetápico, equiprobable y sin reemplazo (muestreo irrestricto aleatorio). | $\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$ | $\hat{\mu} = \hat{X} = \frac{1}{n}\sum_{i=1}^{n} \hat{X}_i$ | $\text{Est. } V(\hat{\mu}) = \frac{N-n}{Nn} s^2$ | $V(\hat{\mu}) = \frac{N-n}{Nn} S^2$<br>$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{X}_i-\mu)^2$ | $n = \frac{\frac{S^2 t^2_{(\alpha,\infty)}}{\varepsilon^2 \mu^2}}{1+\frac{1}{N}\frac{S^2 t^2_{(\alpha,\infty)}}{\varepsilon^2\mu^2}}$ |
| Muestreo bietápico, equiprobable y sin reemplazo. | $\mu = \frac{1}{M_0}\sum_{i=1}^{N} M_i \mu_i$<br>$\mu_i = \frac{1}{M_i}\sum_{j=1}^{M_i} X_{ij}$<br>$M_0 = \sum_{i=1}^{N} M_i$<br>$\bar{M} = \frac{M_0}{N}$ | $\hat{\mu} = \frac{1}{\sum_{i=1}^{n} M_i}\sum_{i=1}^{n} M_i \hat{X}_i$<br>$\hat{X}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} X_{ij}$ | $\text{Est. } V(\hat{\mu}) = (\frac{1}{n}-\frac{1}{N}) s^2$<br>$+\frac{1}{Nn}\sum_{i=1}^{n}(\frac{M_i}{\bar{M}})^2(\frac{1}{m_i}-\frac{1}{M_i}) s_i^2$<br>$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\frac{M_i}{\bar{M}})^2(\hat{X}_i-\hat{X})^2$<br>$s_i^2 = \frac{1}{m_i-1}\sum_{j=1}^{m_i}(X_{ij}-\hat{X}_i)^2$ | $V(\hat{\mu}) = (\frac{1}{n}-\frac{1}{N}) S^2 + \frac{1}{Nn}\sum_{i=1}^{N}(\frac{M_i}{\bar{M}})^2$<br>$(\frac{1}{m_i}-\frac{1}{M_i}) S_i^2$<br>$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\frac{M_i}{\bar{M}})^2(\mu_i-\mu)^2$<br>$S_i^2 = \frac{1}{M_i-1}\sum_{j=1}^{M_i}(X_{ij}-\mu_i)^2$ | $n = a(S^2 + b(m-1))$<br>$a = \frac{N t^2_{(\alpha,\infty)}}{\varepsilon^2\mu^2 N + S^2 t^2_{(\alpha,\infty)}}$<br>$b = \frac{1}{N\bar{M}^2}\sum_{i=1}^{N} M_i S_i^2$<br>$m_i = \frac{M_i}{m}$<br>$1 \le m \le (\frac{N}{ab}+1-\frac{S^2}{b})$ |
| Muestreo monoetápico, equiprobable, sin reemplazo y estratificado. | $\mu = \frac{1}{N_0}\sum_{i=1}^{K} N_i \mu_i$<br>$\mu_i = \frac{1}{N_i}\sum_{j=1}^{N_i} X_{ij}$<br>$N_0 = \sum_{i=1}^{K} N_i$ | $\hat{\mu} = \frac{1}{N_0}\sum_{i=1}^{K} N_i \hat{X}_i$<br>$\hat{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$ | $\text{Est. } V(\hat{\mu}) = \sum_{i=1}^{K}(\frac{1}{n_i}-\frac{1}{N_i})(\frac{N_i}{N_0})^2 s_i^2$<br>$s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij}-\hat{X}_i)^2$ | $V(\hat{\mu}) = \sum_{i=1}^{K}(\frac{N_i}{N_0})^2(\frac{1}{n_i}-\frac{1}{N_i}) S_i^2$<br>$S_i^2 = \frac{1}{N_i-1}\sum_{j=1}^{N_i}(X_{ij}-\mu_i)^2$ | $n_i = \frac{\frac{S_i^2 t^2_{(\alpha,\infty)}}{(\frac{N_0}{N_i})^2\frac{\varepsilon^2}{K}\mu^2}}{1+\frac{1}{N_i}\frac{S_i^2 t^2_{(\alpha,\infty)}}{(\frac{N_0}{N_i})^2\frac{\varepsilon^2}{K}\mu^2}}$ |
| Muestreo bietápico, equiprobable, sin reemplazo y estratificado. | $\mu = \frac{1}{M_0}\sum_{i=1}^{K} M_{0i}\mu_i$<br>$\mu_i = \frac{1}{M_{0i}}\sum_{j=1}^{N_i} M_{ij}\mu_{ij}$<br>$\mu_{ij} = \frac{1}{M_{ij}}\sum_{k=1}^{M_{ij}} X_{ijk}$<br>$M_0 = \sum_{i=1}^{K} M_{0i}$ $\bar{M}_i = \frac{M_{0i}}{N_i}$<br>$M_{0i} = \sum_{j=1}^{N_i} M_{ij}$ | $\hat{\mu} = \frac{1}{M_0}\sum_{i=1}^{K} M_{0i}\hat{X}_i$<br>$\hat{X}_i = \frac{1}{n_i\bar{M}_i}\sum_{j=1}^{n_i} M_{ij}\hat{X}_{ij}$<br>$\hat{X}_{ij} = \frac{1}{m_{ij}}\sum_{k=1}^{m_{ij}} X_{ijk}$<br>$\hat{M}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} M_{ij}$ | $\text{Est. } V(\hat{\mu}) = \sum_{i=1}^{K}(\frac{M_{0i}}{M_0})^2[(\frac{1}{n_i}-\frac{1}{N_i})s_i^2$<br>$+\frac{1}{N_i n_i}\sum_{j=1}^{n_i}(\frac{M_{ij}}{\bar{M}_i})^2(\frac{1}{m_{ij}}-\frac{1}{M_{ij}})s_{ij}^2]$<br>$s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(\frac{M_{ij}}{\bar{M}_i})^2(\hat{X}_{ij}-\hat{X}_i)^2$<br>$s_{ij}^2 = \frac{1}{m_{ij}-1}\sum_{k=1}^{m_{ij}}(X_{ijk}-\hat{X}_{ij})^2$ | $V(\hat{\mu}) = \sum_{i=1}^{K}(\frac{M_{0i}}{M_0})^2[(\frac{1}{n_i}-\frac{1}{N_i})S_i^2+\frac{1}{N_i n_i}\sum_{j=1}^{N_i}$<br>$\frac{M_{ij}^2}{\bar{M}_i^2}(\frac{1}{m_{ij}}-\frac{1}{M_{ij}})S_{ij}^2]$<br>$S_{ij}^2 = \frac{1}{M_{ij}-1}\sum_{k=1}^{M_{ij}}(X_{ijk}-\mu_{ij})$<br>$S_i^2 = \frac{1}{N_i-1}\sum_{j=1}^{N_i}(\frac{M_{ij}}{\bar{M}_i}\mu_{ij}-\mu_i)^2$ | $n_i = a_i[S_i^2 + b_i(m_i-1)]$<br>$a_i = \frac{K N_i M_{0i}^2 t^2_{(\alpha,\infty)}}{N_i M_0^2\varepsilon^2\mu^2 + K M_{0i}^2 S_i^2 t^2_{(\alpha,\infty)}}$<br>$b_i = \frac{1}{N_i\bar{M}_i^2}\sum_{j=1}^{N_i} M_{ij} S_{ij}^2$<br>$m_{ij} = \frac{1}{m_i} M_{ij}$<br>$1 \le m_i \le \frac{N_i}{a_i b_i}+1-\frac{S_i^2}{b_i}$ |
| Muestreo monoetápico, equiprobable y con reemplazo (muestreo irrestricto aleatorio con reemplazo). | $\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$ | $\hat{\mu} = \hat{X} = \frac{1}{n}\sum_{i=1}^{n}\hat{X}_i$ | $\text{Est. } V(\hat{\mu}) = \frac{1}{n} s^2$<br>$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\hat{X})^2$ | $V(\hat{\mu}) = \frac{N-1}{Nn} S^2$ | $n = \frac{(N-1)S^2 t^2_{(\alpha,\infty)}}{N\varepsilon^2\mu^2}$ |
| Muestreo bietápico, equiprobable y con reemplazo. | $\mu = \frac{1}{M_0}\sum_{i=1}^{N} M_i\mu_i$<br>$\mu_i = \frac{1}{M_i}\sum_{j=1}^{M_i} X_{ij}$<br>$M_0 = \sum_{i=1}^{N} M_i$<br>$\bar{M} = \frac{M_0}{N}$ | $\hat{\mu} = \frac{N}{nM_0}\sum_{i=1}^{n} M_i\hat{X}_i$<br>$\hat{X}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} X_{ij}$ | $\text{Est. } V(\hat{\mu}) = \frac{s^2}{nM_0^2}$<br>$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(NM_i\hat{X}_i-\frac{N}{n}\sum_{i=1}^{n}M_i\hat{X}_i)^2$ | $V(\hat{\mu}) = \frac{N-1}{NnM_0^2}S^2 + \frac{1}{NnM_0^2}\sum_{i=1}^{N}\frac{M_i-1}{M_i m_i}S_i^2$<br>$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(NM_i\mu_i-M_0\mu)^2$<br>$S_i^2 = \frac{1}{M_i-1}\sum_{j=1}^{M_i}(NM_iX_{ij}-NM_i\mu_i)^2$ | $n = a[(N-1)S^2 + m\,b]$<br>$a = \frac{t^2_{(\alpha,\infty)}}{\varepsilon^2\mu^2 N M_0^2}$<br>$b = \sum_{i=1}^{N}\frac{M_i-1}{M_i}S_i^2$<br>$m_i = \frac{M_i}{m}$<br>$1 \le m \le \frac{N-a(N-1)S^2}{ab}$ |
| Muestreo monoetápico, equiprobable con reemplazo y estratificado. | $\mu = \frac{1}{N_0}\sum_{i=1}^{K} N_i\mu_i$ $N_0 = \sum_{i=1}^{K} N_i$<br>$\mu_i = \frac{1}{N_i}\sum_{j=1}^{N_i} X_{ij}$ | $\hat{\mu} = \frac{1}{N_0}\sum_{i=1}^{K} N_i\hat{X}_i$<br>$\hat{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$ | $\text{Est. } V(\hat{\mu}) = \sum_{i=1}^{K}(\frac{N_i}{N_0})^2\frac{s_i^2}{n_i}$<br>$s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij}-\hat{X}_i)^2$ | $V(\hat{\mu}) = \sum_{i=1}^{K}(\frac{N_i}{N_0})^2(\frac{1}{n_i}-\frac{1}{N_i n_i})S_i^2$<br>$S_i^2 = \frac{1}{N_i-1}\sum_{j=1}^{N_i}(X_{ij}-\mu_i)^2$ | $n_i = (\frac{N_i}{N_0})^2\frac{K(N_i-1)S_i^2 t^2_{(\alpha,\infty)}}{N_i\varepsilon^2\mu^2}$ |
| Muestreo bietápico, equiprobable, con reemplazo y estratificado. | $\mu = \frac{1}{M_0}\sum_{i=1}^{K} M_{0i}\mu_i$<br>$\mu_i = \frac{1}{M_{0i}}\sum_{j=1}^{N_i} M_{ij}\mu_{ij}$<br>$\mu_{ij} = \frac{1}{M_{ij}}\sum_{k=1}^{M_{ij}} X_{ijk}$<br>$M_0 = \sum_{i=1}^{K} M_{0i}$<br>$M_{0i} = N_i\bar{M}_i$<br>$\bar{M}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} M_{ij}$ | $\hat{\mu} = \frac{1}{M_0}\sum_{i=1}^{K} M_{0i}\hat{X}_i$<br>$\hat{X}_i = \frac{1}{n_i\hat{M}_i}\sum_{j=1}^{n_i}\hat{Y}_{ij}$<br>$\hat{X}_{ij} = \frac{1}{m_{ij}}\sum_{k=1}^{m_{ij}}\hat{X}_{ijk}$<br>$\hat{M}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} M_{ij}$ | $\text{Est. } V(\hat{\mu}) = \sum_{i=1}^{K}(\frac{M_{0i}}{M_0})^2\frac{s_i^2}{n_i}$<br>$s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}[\frac{1}{\hat{M}_i}(M_{ij}\hat{X}_{ij}-M_i\hat{X}_i)]^2$ | $V(\hat{\mu}) = \sum_{i=1}^{K}(\frac{M_{0i}}{M_0})^2\frac{1}{n_i}[\frac{1}{N_i}\sum_{j=1}^{N_i}(\frac{M_{ij}}{\bar{M}_i})^2$<br>$\frac{M_{ij}-1}{M_{ij}m_{ij}}S_{ij}^2+\frac{N_i-1}{N_i}S_i^2]$<br>$S_i^2 = \frac{1}{N_i-1}\sum_{j=1}^{N_i}[\frac{M_{ij}}{\bar{M}_i}(\mu_{ij}-\mu_i)]^2$<br>$S_{ij}^2 = \frac{1}{M_{ij}-1}\sum_{k=1}^{M_{ij}}(X_{ijk}-\mu_{ij})^2$ | $n_i = a_i[(N_i-1)S_i^2+b_i m_i]$<br>$a_i = \frac{t^2_{(\alpha,\infty)}}{\frac{\varepsilon^2}{K(\frac{M_{0i}}{M_0})^2}\mu^2 N_i}$<br>$b_i = \frac{1}{\bar{M}_i^2}\sum_{j=1}^{N_i}(M_{ij}-1)S_{ij}^2$<br>$m_{ij} = \frac{M_{ij}}{m_i}$<br>$1 \le m_i \le \frac{N_i}{a_i b_i}-\frac{N_i-1}{b_i}S_i^2$ |
| Muestreo monoetápico con probabilidades variables de selección y con reemplazo. | $\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$<br>$\sum_{i=1}^{N} P_i = 1$ | $\hat{\mu} = \frac{1}{Nn}\sum_{i=1}^{n}\frac{X_i}{P_i}$ | $\text{Est. } V(\hat{\mu}) = \frac{s^2}{n}$<br>$s^2 = \frac{1}{N^2(n-1)}\sum_{i=1}^{n}(\frac{\hat{X}_i}{P_i}-N\hat{\mu})^2$ | $V(\hat{\mu}) = \frac{1}{n}\sigma^2$<br>$\sigma^2 = \frac{1}{N^2}\sum_{i=1}^{N} P_i(\frac{X_i}{P_i}-N\mu)^2$ | $n = \frac{\sigma^2 t^2_{(\alpha,\infty)}}{\varepsilon^2\mu^2}$ |
| Muestreo bietápico con probabilidades variables de selección en cada etapa con reemplazo. | $\mu = \frac{1}{M_0}\sum_{i=1}^{N} M_i\mu_i$<br>$\mu_i = \frac{1}{M_{ij}}\sum_{j=1}^{M_{ij}} X_{ij}$<br>$M_0 = \sum_{i=1}^{N} M_i$<br>$\bar{M} = \frac{M_0}{N}$ | $\hat{\mu} = \frac{1}{M_0 n}\sum_{i=1}^{n}\hat{X}_i$<br>$\hat{Y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i}\frac{X_{ij}}{P_{ij}}$ | $\text{Est. } V(\hat{\mu})\ \frac{s^2}{n}$<br>$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{X}_i-M_0\hat{\mu})^2$ | $V(\hat{\mu}) = \frac{1}{M_0^2 n}[\sum_{i=1}^{N}\frac{P_i}{m_i}\sigma_i^2+\sigma^2]$<br>$\sigma^2 = \sum_{i=1}^{N} P_i(\frac{M_i\mu_i}{P_i}-M_0\mu)^2$<br>$\sigma_i^2 = \sum_{j=1}^{M_i}\frac{P_{ij}}{P_i}(\frac{X_{ij}}{P_{ij}}-\frac{M_i\mu_i}{P_i})^2$ | $n = a(m\,b+\sigma^2)$<br>$a = \frac{t^2_{(\alpha,\infty)}}{M_0^2\varepsilon^2\mu^2}$<br>$b = \sum_{i=1}^{N}\frac{P_i}{M_i}\sigma_i^2$<br>$m_i = \frac{M_i}{m}$<br>$1 \le m \le \frac{N}{ab}-\frac{\sigma^2}{b}$ |

## VI.6.6 PRÁCTICA VIII

Una población consta de los dígitos: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a partir de la cual obtenga:

a) El número y composición de las muestras de tamaño n = 2 que pueda surgir de esa población, con reemplazo y sin reemplazo.
b) Considerando el número de muestras obtenidas sin reemplazo,
Obtenga $\mu\bar{x}$ y $\sigma_{\bar{x}}$; y para la población obtenga μ y σ.
c) Analice, compare interprete la relación que hay entre μ, $\mu\bar{x}$, σ, $\sigma_{\bar{x}}$ tanto para "variables" como para proporciones (atributos).
d) Obtenga $\overline{X}_i$ y *Si* para cada muestra obtenida sin reemplazo.
e) Compare e interprete los valores de los parámetros μ y σ y de los estadísticos $\overline{X}_i$ y Si, si es que existe.
f) Usando la tabla de números aleatorios seleccione aleatoria y sistemáticamente una muestra. Para la muestra seleccionada, calcule su $\bar{x}$, $S_i$ correspondiente. Con α = 0.05, determine e interprete los limites de confianza dentro de las cuales se halla $\mu\bar{x}$.
g) Con e = 0.05 (error permitido) y α = 0.05 determine el tamaño de la muestra adecuada, aplicando la fórmula del muestreo simple aleatorio.
h) Relación entre $\sigma_{\bar{x}}$, e.

SOLUCIÓN DE LA PRACTICA VIII

A) Con $X_i$ : 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.
tenemos N = 10, luego con n = 2 se obtiene, sin reemplazo

$\begin{bmatrix} N \\ n \end{bmatrix} = \dfrac{N!}{(N-n)!\,n!} = \dfrac{10}{(10-2)!2!} = 45$ muestras de tamaño dos y que constituyen la nueva distribución de muestreo, que son:

| Muestra | Composición de la muestra | P(muestra) | Media Muestral | $Si = \sqrt{\frac{\sum\left(xi - \bar{x}\right)^2}{n}}$ |
|---|---|---|---|---|
| 1 | 0 , 1 | 1 ÷ 45 | 0.5 | 0.5 |
| 2 | 0 , 2 | 1 ÷ 45 | 1.0 | 1.0 |
| 3 | 0 , 3 | 1 ÷ 45 | 1.5 | 1.5 |
| 4 | 0 , 4 | 1 ÷ 45 | 2.0 | 2.0 |
| 5 | 0 , 5 | 1 ÷ 45 | 2.5 | 2.5 |
| 6 | 0 , 6 | 1 ÷ 45 | 3.0 | 3.0 |
| 7 | 0 , 7 | 1 ÷ 45 | 3.5 | 3.5 |
| 8 | 0 , 8 | 1 ÷ 45 | 4.0 | 4.0 |
| 9 | 0 , 9 | 1 ÷ 45 | 4.5 | 4.5 |
| 10 | 1 , 2 | 1 ÷ 45 | 1.5 | 0.5 |
| 11 | 1 , 3 | 1 ÷ 45 | 2.0 | 1.0 |
| 12 | 1 , 4 | 1 ÷ 45 | 2.5 | 1.5 |
| 13 | 1 , 5 | 1 ÷ 45 | 3.0 | 2.0 |
| 14 | 1 , 6 | 1 ÷ 45 | 3.5 | 2.5 |
| 15 | 1 , 7 | 1 ÷ 45 | 4.0 | 3.0 |
| 16 | 1 , 8 | 1 ÷ 45 | 4.5 | 3.5 |
| 17 | 1 , 9 | 1 ÷ 45 | 5.0 | 4.0 |
| 18 | 2 , 3 | 1 ÷ 45 | 2.5 | 0.5 |
| 19 | 2 , 4 | 1 ÷ 45 | 3.0 | 1.0 |
| 20 | 2 , 5 | 1 ÷ 45 | 3.5 | 1.5 |
| 21 | 2 , 6 | 1 ÷ 45 | 4.0 | 2.0 |
| 22 | 2 , 7 | 1 ÷ 45 | 2.5 | 2.5 |
| 23 | 2 , 8 | 1 ÷ 45 | 5.0 | . |
| 24 | 2 , 9 | 1 ÷ 45 | 5.5 | . |
| 25 | 3 , 4 | 1 ÷ 45 | 3.5 | . |
| 26 | 3 , 5 | 1 ÷ 45 | 4.0 | . |
| 27 | 3 , 6 | 1 ÷ 45 | 4.5 | . |
| 28 | 3 , 7 | 1 ÷ 45 | 5.0 | . |
| 29 | 3 , 8 | 1 ÷ 45 | 5.5 | . |
| 30 | 3 , 9 | 1 ÷ 45 | 6.0 | . |
| 31 | 4 , 5 | 1 ÷ 45 | 4.5 | . |
| 32 | 4 , 6 | 1 ÷ 45 | 5.0 | . |
| 33 | 4 , 7 | 1 ÷ 45 | 5.5 | . |
| 34 | 4 , 8 | 1 ÷ 45 | 6.0 | . |
| 35 | 4 , 9 | 1 ÷ 45 | 6.5 | . |
| 36 | 5 , 6 | 1 ÷ 45 | 5.5 | . |
| 37 | 5 , 7 | 1 ÷ 45 | 6.0 | . |
| 38 | 5 , 8 | 1 ÷ 45 | 6.5 | . |

| | | | | |
|---|---|---|---|---|
| 39 | 5 , 9 | 1 ÷ 45 | 7.0 | . |
| 40 | 6 , 7 | 1 ÷ 45 | 6.5 | . |
| 41 | 6 , 8 | 1 ÷ 45 | 7.0 | . |
| 42 | 6 , 9 | 1 ÷ 45 | 7.5 | . |
| 43 | 7 , 8 | 1 ÷ 45 | 7.5 | . |
| 44 | 7 , 9 | 1 ÷ 45 | 8.0 | . |
| 45 | 8 , 9 | 1 ÷ 45 | 8.5 | . |
| | | 45 ÷ 45 | 202.5 | |

Generación de la distribución de muestras, con reemplazo; $N^n = 10^2 = 100$

| $n_i$ | Composición | $n_i$ | Composición | $n_i$ | Composición | $n_i$ | Composición |
|---|---|---|---|---|---|---|---|
| 1 | 0 , 0 | 26 | 2 , 5 | 51 | 5 , 0 | 76 | 7 , 5 |
| 2 | 0 , 1 | 27 | 2 , 6 | 52 | 5 , 1 | 77 | 7 , 6 |
| 3 | 0 , 2 | 28 | 2 , 7 | 53 | 5 , 2 | 78 | 7 , 7 |
| 4 | 0 , 3 | 29 | 2 , 8 | 54 | 5 , 3 | 79 | 7 , 8 |
| 5 | 0 , 4 | 30 | 2 , 9 | 55 | 5 , 4 | 80 | 7 , 9 |
| 6 | 0 , 5 | 31 | 3 , 0 | 56 | 5 , 5 | 81 | 8 , 0 |
| 7 | 0 , 6 | 32 | 3 , 1 | 57 | 5 , 6 | 82 | 8 , 1 |
| 8 | 0 , 7 | 33 | 3 , 2 | 58 | 5 , 7 | 83 | 8 , 2 |
| 9 | 0 , 8 | 34 | 3 , 3 | 59 | 5 , 8 | 84 | 8 , 3 |
| 10 | 0 , 9 | 35 | 3 , 4 | 60 | 5 , 9 | 85 | 8 , 4 |
| 11 | 1 , 0 | 36 | 3 , 5 | 61 | 6 , 0 | 86 | 8 , 5 |
| 12 | 1 , 1 | 37 | 3 , 6 | 62 | 6 , 1 | 87 | 8 , 6 |
| 13 | 1 , 2 | 38 | 3 , 7 | 63 | 6 , 2 | 88 | 8 , 7 |
| 14 | 1 , 3 | 39 | 3 , 8 | 64 | 6 , 3 | 89 | 8 , 8 |
| 15 | 1 , 4 | 40 | 3 , 9 | 65 | 6 , 4 | 90 | 8 , 9 |
| 16 | 1 , 5 | 41 | 4 , 0 | 66 | 6 , 5 | 91 | 9 , 0 |
| 17 | 1 , 6 | 42 | 4 , 1 | 67 | 6 , 6 | 92 | 9 , 1 |
| 18 | 1 , 7 | 43 | 4 , 2 | 68 | 6 , 7 | 93 | 9 , 2 |
| 19 | 1 , 8 | 44 | 4 , 3 | 69 | 6 , 8 | 94 | 9 , 3 |
| 20 | 1 , 9 | 45 | 4 , 4 | 70 | 6 , 9 | 95 | 9 , 4 |
| 21 | 2 , 0 | 46 | 4 , 5 | 71 | 7 , 0 | 96 | 9 , 5 |
| 22 | 2 , 1 | 47 | 4 , 6 | 72 | 7 , 1 | 97 | 9 , 6 |
| 23 | 2 , 2 | 48 | 4 , 7 | 73 | 7 , 2 | 98 | 9 , 7 |
| 24 | 2 , 3 | 49 | 4 , 8 | 74 | 7 , 3 | 99 | 9 , 8 |
| 25 | 2 , 4 | 50 | 4 , 9 | 75 | 7 , 4 | 100 | 9 , 9 |

B) Cálculo de los parámetros de la población.

Valores fijos.

$\mu = 45/10 = 4.5$ .

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} = \sqrt{\frac{82.50}{10}} = \sigma = 2.87$$

Calculo de $\sigma$:

| $X_i$ | $X_i\text{-}\mu$ | $(X_i\text{-}\mu)^2$ |
|---|---|---|
| 0 | -4.5 | 20.25 |
| 1 | -3.5 | 12.25 |
| 2 | -2.5 | 6.25 |
| 3 | -1.5 | 2.25 |
| 4 | -0.5 | 0.25 |
| 5 | 0.5 | 0.25 |
| 6 | 1.5 | 2.25 |
| 7 | 2.5 | 6.25 |
| 8 | 3.5 | 12.25 |
| 9 | 4.5 | 20.25 |
| | 0 | 82.5 |

Ahora, calculando.

$$\mu\bar{x} = E(\bar{X}_i) = \frac{\sum \bar{X}_i}{n} = \frac{202.5}{45} = 4.5 = \mu\bar{x}$$

NOTA :número de muestras = 45.

$$\sigma\bar{x} = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}} = \frac{2.87}{\sqrt{2}} * \sqrt{\frac{10-2}{10-1}} = \frac{2.87}{1.41}\sqrt{0.89}$$

$$\sigma\bar{x} = 2.03(0.94) = 1.92$$

C) Vemos que $\mu\bar{x} = \mu = 4.5$ mientras que $\sigma_{\bar{x}} = 1.92$ y que $\sigma = 2.87$

En el caso de "VARIABLES", lo mismo sucede en el caso de las "PROPORCIONES", es decir, $P = \Pi$ donde :

P: Proporción de la distribución de muestras.

$\Pi$: Parámetro poblacional.

Sp: Error estándar de la distribución de proporciones.

$\sigma$p: Parámetro denominado desviación estándar de la población.

Sp diferente de σp.

D) Obtenga $\overline{X}_i$ y si para cada muestra obtenida sin reemplazo.
Solución:
i = 1, 2,...,44, 45.
Calculando como ilustración S1 y S45, porque el método de calculo es el mismo, tenemos:

$$S_1 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \qquad S_{45} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

$$S_1 = \sqrt{\frac{\sum\left[(0-0.5)^2 + (1-0.5)^2\right]}{2}} \qquad S_{45} = \sqrt{\frac{\sum\left[(8-8.5)^2 + (9-8.5)^2\right]}{2}}$$

$$S_1 = \sqrt{\frac{\sum[0.25 + 0.25]}{2}} \qquad S_{45} = \sqrt{\frac{\sum[0.25 + 0.25]}{2}}$$

$$S_1 = \sqrt{\frac{0.50}{2}} \qquad S_{45} = \sqrt{\frac{0.50}{2}}$$

$$S_1 = \sqrt{0.25} = 0.5 \qquad S_{45} = \sqrt{0.25} = 0.5$$

E) Al comparar los valores de μ, σ con $\overline{X}_i$, Si, vemos que el valor de los parámetros es FIJO, mientras que el de las "estadísticas" es variable puesto que esta en función de la composición de cada muestra.
F) La selección aleatoria determinó la obtención de la muestra compuesta por los dígitos 0 y 8, puesto que la tabla de números aleatorios, trabajando horizontalmente, determinó que se tomara la muestra número 08 del marco muestral que está compuesto por 45 muestras disponibles y obtenidas en un muestreo sin remplazo.

| MARCO MUESTRAL | |
|---|---|
| Número de muestra | Composición de la muestra |
| 1 | 0 , 1 |
| 2 | 0 , 2 |
| 3 | 0 , 3 |
| 4 | 0 , 4 |
| 5 | 0 , 5 |
| 6 | 0 , 6 |
| 7 | 0 , 7 |

| | | |
|---|---|---|
| 8 | 0 , | 8 |
| . | . | |
| . | . | |
| . | . | |
| . | . | |
| . | . | |
| 45 | 8 , | 9 |

Así, a partir de la selección aleatoria que determinó la muestra compuesta por los dígitos 8 y 0 vamos a determinar los límites de confianza con: $\alpha = 0.05$; $Z_\alpha = \pm\ 1.96$.

$\overline{X} = 8 + 0 / 2 = 4$

$$S = \sqrt{\frac{\sum\left[(8-4)^2 + (0-4)^2\right]}{2}} = \sqrt{16} = 4 = 4$$

Sabemos que los límites de confianza se determinan con :

$$\overline{X} \pm Z_\alpha \sigma\overline{x} \text{ donde } \sigma\overline{x} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = 1.92 \text{ del inciso B).}$$

Luego sustituyendo tendremos:

4 - (1.96) (1.92) límite inferior del intervalo de confianza = 0.2368

4 + (1.96)(1.92) límite superior del intervalo de confianza = 7.7632.

Interpretación: hay una probabilidad del 95% de que el valor de $\mu$ se halle en el intervalo de 0.2368 a 7.7632, lo cual es cierto puesto que $\mu\overline{x} = 4.5$.

G) Con $e = 0.05$; $\alpha = 0.05$ tenemos $Z_\alpha = \pm\ 1.96$; $\sigma = 2.87$ luego:

$$n = \frac{Z_\alpha^2 \sigma_x^2 N}{e^2 N - e^2 + Z_\alpha^2 \sigma_x^2} = \frac{(3.84)(3.68)(10)}{(0.25)-(0.0025)+(14.15)} = \frac{141.7}{14.17} = 10$$

Ahora bien usando $n = \dfrac{N}{1+Ne^2} = \dfrac{10}{1+10(0.05)^2} = \dfrac{10}{1+0.025}$

$$n = \frac{10}{1.025} = 9.7 \cong 10$$

Observaciones:

Con las dos fórmulas se obtiene el mismos resultado. Ello indica que el tamaño de la muestra debe ser el del universo. Esto es así, no debe sorprendernos porque el universo es tan pequeño que la muestra debe ser igual a 10 para que sea representativa.

h) La relación entre $\sigma_{\bar{x}}$ y e.

$$\sigma\bar{x} = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}} = \frac{2.87}{\sqrt{2}} * \sqrt{\frac{10-2}{10-1}} = 1.92$$

e = $Z_\alpha \sigma\bar{x}$ = 1.96 (1.92) = 3.7632

i).- La determinación de e se hace a partir de los límites de confianza.

Así, de $\overline{X} \pm Z_\alpha \sigma\bar{x}$ tenemos que $e = Z_\alpha \sigma\bar{x}$ = 1.96 (1.92) = 3.7632

comparación $\sigma_{\bar{x}}$ = 1.92 y e = 3.7632, luego el error estándar, es menor que el error de muestreo o error permitido.

Pero si el error estándar ( 1.92 ) lo usamos en términos de probabilidad para cuantificar el error de muestreo $|\bar{x} - \mu_x|$, entonces recordemos que **idealmente** èste último debe ser menor o igual que e = error permitido= $Z_\alpha \ \sigma_{\bar{x}}$

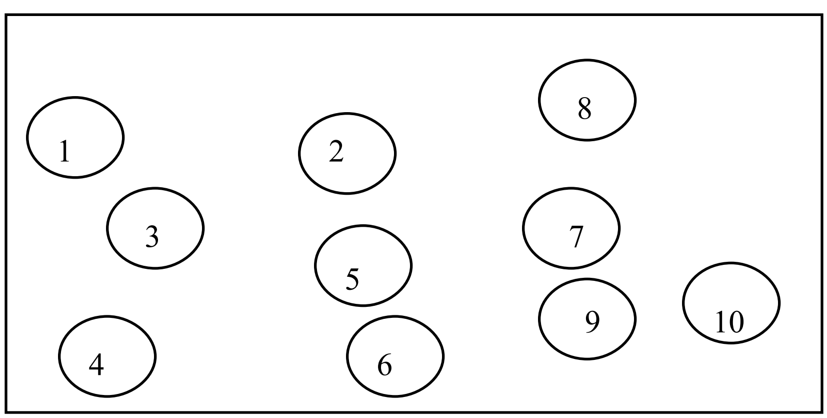
Del inciso b), sabemos que $\mu_x = 4.5$ y del inciso f) sabemos que $\bar{x} = 4.0$ luego el error de muestreo = $|4-4.5| = |0.5| \leq 3.7637 =$ error permitido. Es bueno el resultado.

## VI.6.7 PRÁCTICA IX

### 1.- APLICACIONES DEL MUESTREO SIMPLE ALEATORIO.

Referencias:

El plano de la colonia del Valle del Distrito Federal es el siguiente:



Con : n = 2
$t_\alpha = 2$
Obtenga:
a) El número de manzanas en la colonia del Valle.
hay N = 10 manzanas.

b) La fracción del muestreo.
Fracción de muestreo = F = n / N = 2 / 10 = 0.2

c) Seleccione con la tabla de números aleatorios las dos manzanas que integren la muestra, indica como son y como le hizo.
Digamos que fueron las manzanas 2 y 7, que cayeron en la muestra mediante el manejo ya conocido de la tabla de números aleatorios.

d) Suponiendo que:
La primer manzana tiene 40 familias.
La segunda manzana tiene 36 familias.
Cálcule la media y desviación estándar de la muestra.
Puesto que:
Manzana 2 tiene 40 familias.
Manzana 7 tiene 36 familias.

$\bar{X} = 40 + 36 / 2 = 76 / 2 = 38$ familias.

$$S = \sqrt{\frac{(40-38)^2 + (36-38)^2}{2}} = \sqrt{\frac{(2)^2 + (-2)^2}{2}} = \sqrt{\frac{8}{2}} = 2$$

e) Determine el total de familias en la colonia del Valle.
El total de familias se estima por :
$\hat{Y} = N\overline{X}$ ; $\hat{Y} = 10(38);$
$\hat{Y} = 380$ familias.

f) Determine e interprete los límites de confianza del total de familias.
Sabemos que:

$$N\bar{x} - \frac{tNS}{\sqrt{n}}\sqrt{1-F} \leq \hat{Y} \leq N\bar{x} + \frac{tNS}{\sqrt{n}}\sqrt{1-0.2}$$

Sustituyendo

$$380 - \frac{2(10)2}{1.41}\sqrt{1-0.2} \leq \hat{Y} \leq 380 + \frac{2(10)2}{1.41}\sqrt{1-0.2}$$

$$380 - \frac{2(10)2}{1.41}(0.89) \leq \hat{Y} \leq 380 + \frac{2(10)2}{1.41}(0.89)$$

380 - 28.37 (0.89) | $\hat{Y}$ | 380 + 28.37 (0.89)

380 - 25.2 | $\hat{Y}$ | 380 + 25.2

354.8 | $\hat{Y}$ | 405.2

Interpretación:

El total estimado de familias Y, se halla entre 355 y 405 familias con una probabilidad o seguridad del 95.45%

g) Determine el número de habitantes en la colonia del Valle tomando en cuenta que 5 es el promedio de personas por familia.

El total de habitantes en la colonia del Valle es:

Total de habitantes = 380 (5) = 1900 personas.

## 2.- APLICACIÓN DEL MUESTREO ALEATORIO ESTRATIFICADO:

Referencias: El canal 22 de televisión ha sido puesto en venta y la empresa "escorpión" que esta interesado en adquirirlo disidió hacer una encuesta para conocer el número de horas que el público ve televisión y de ahí saber cuántos hogares (mediante entrevistas) ven el canal 22.

La empresa escorpión puede producir estimaciones por separado es decir, puede estratificar para estimar el número promedio de horas que se ve televisión en cada estrato, ya que, la información disponible revela que hay tres estratos que componen el universo o población con:

| POBLACIÓN | MUESTRA | ESTRATO HOGARES |
|---|---|---|

| N | n | |
|---|---|---|
| $N_1$= 180 hogares | $n_1$= 15 hogares | 1 |
| $N_2$= 70 hogares | $n_2$= 4 hogares | 2 |
| $N_3$= 100 hogares | $n_3$= 5 hogares | 3 |
| Total 350 | Total 24 | |

Mediante la cual se realizan las entrevistas, con los siguientes resultados:
Tiempo que se ve televisión, en horas por semana.

| Estrato 1 | Estrato 2 | Estrato 3 |
|---|---|---|
| 30, 27, 40<br>45, 26, 35<br>33, 29, 37<br>34, 25, 41<br>43, 32, 31 | 4, 49<br>25, 30 | 9, 20<br>11, 34<br>24, |

Con esos datos sustituya y obtenga:
a) $\overline{X}_i$, Si con i=1, 2, 3.

| Estrato 1 | Estrato 2 | Estrato 3 |
|---|---|---|
| $n_1 = 15$ | $n_2 = 4$ | $n_3 = 5$ |
| $\bar{x}_1 = 34$ | $\bar{x}_2 = 27$ | $\bar{x}_3 = 20$ |
| $S_1 = 6$ | $S_2 = 16$ | $S_3 = 9.1$ |
| $N_1 = 180$ | $N_2 = 70$ | $N_3 = 100$ |

ESTRATO 1

$$\overline{X}_1 = \frac{30+27+40+45+26+35+33+29+37+34+25+41+43+32+31}{15}$$

$$\overline{X}_1 = \frac{508}{15} = 33.87$$

33.87 equivalente a 34

$$S_1 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{546}{15}} = \sqrt{36.4} = 6$$

| $\bar{X}_i$ | $x_i - \bar{X}_1$ | | | | | $(X_i - \bar{X}_1)^2$ |
|---|---|---|---|---|---|---|
| 30 | 30 | - | 34 | = | -4 | 16 |
| 27 | 27 | - | 34 | = | -7 | 49 |
| 40 | 40 | - | 34 | = | 6 | 36 |
| 45 | 45 | - | 34 | = | 11 | 121 |
| 26 | 26 | - | 34 | = | -8 | 64 |
| 35 | 35 | - | 34 | = | 1 | 1 |
| 33 | 33 | - | 34 | = | -1 | 1 |
| 29 | 29 | - | 34 | = | -5 | 25 |
| 37 | 37 | - | 34 | = | 3 | 9 |
| 34 | 34 | - | 34 | = | 0 | 0 |
| 25 | 25 | - | 34 | = | -9 | 81 |
| 41 | 41 | - | 34 | = | 7 | 49 |
| 43 | 43 | - | 34 | = | 9 | 81 |
| 32 | 32 | - | 34 | = | -2 | 4 |
| 31 | 31 | - | 34 | = | -3 | 9 |
| | | | | | | 546 |

ESTRATO 2

$$\bar{X}_2 = \frac{4+49+25+30}{4} = \frac{108}{4} = 27$$

$$S_2 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{1026}{4}} = \sqrt{256.5} = 16.0$$

| $\bar{X}_i$ | $x_i - \bar{X}_2$ | | | | | $(X_i - \bar{X}_2)^2$ |
|---|---|---|---|---|---|---|
| 4 | 4 | - | 27 | = | -23 | 529 |
| 49 | 49 | - | 27 | = | -22 | 484 |
| 25 | 25 | - | 27 | = | -2 | 4 |
| 30 | 30 | - | 27 | = | 3 | 9 |

1026

ESTRATO 3

$$\bar{X}_3 = \frac{9+20+11+34+24}{5} = \frac{98}{5} = 19.6 \text{ equivale a } 20$$

$$S_3 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{414}{5}} = \sqrt{82.8} = 9.1$$

| $\overline{X}_1$ | $x_i - \overline{X}_3$ | $(X_1 - \overline{X}_3)^2$ |
|---|---|---|
| 9 | 9 - 20 = -11 | 121 |
| 20 | 20 - 20 = 0 | 0 |
| 11 | 11 - 20 = -9 | 81 |
| 34 | 34 - 20 = 14 | 196 |
| 24 | 24 - 20 = 4 | 16 |
| | | 414 |

b) Usando la información anterior, estime el tiempo promedio que se ve televisión en horas por semana, para todos los hogares, en la población constituida por todos, sabiendo que esta medida poblacional, que no es muestral, se calcula con la fórmula:

$$\overline{\overline{X}} \text{ De todos los estratos } = \frac{1}{N}\sum_{i=1}^{3} N_i \overline{X}_i = \frac{1}{N}\left[N_1\overline{X}_1 + N_2\overline{X}_2 + N_3\overline{X}_3\right]$$

$$\overline{\overline{X}} = \frac{1}{350}\left[180(34) + 70(27) + 100(20)\right]; \overline{\overline{X}} = \frac{1}{350}\left[6120 + 1890 + 2000\right]$$

$$\overline{\overline{X}} = \frac{10{,}010}{350} = 28.6 \text{ equivalente a 29 horas}$$

c) Obtenga la varianza de $\overline{\overline{X}}$ con la fórmula:

$$V(\overline{\overline{X}}) = \frac{1}{N^2}\left[N_1\left(\frac{N_1 - n_1}{N_1}\right)\left(\frac{S^2}{n_1}\right) + N_2\left(\frac{N_2 - n_2}{N_2}\right)\left(\frac{S_2^2}{n_2}\right) + N_3\left(\frac{N_3 - n_3}{N_3}\right)\left(\frac{S_3^3}{n_3}\right)\right]$$

$$V(\overline{\overline{X}}) = \frac{1}{(350)^2}\left[180\left(\frac{180-15}{180}\right)\left(\frac{36.4}{15}\right) + 70\left(\frac{70-4}{70}\right)\left(\frac{256.5}{4}\right) + 100\left(\frac{100-5}{100}\right)\left(\frac{82.8}{5}\right)\right]$$

$$V(\overline{\overline{X}}) = \frac{1}{122500}\left[402.41 + 4219.75 + 15.732\right]$$

$$V(\overline{X}) = \frac{6195.36}{122500} = 0.0506$$

d) Obtenga los limites de confianza con:

$\overline{\overline{X}} \pm 2\sqrt{V(\overline{\overline{X}})}$ ; Interprete conociendo que.

$2\sqrt{V(\overline{\overline{X}})}$ = e = error de estimación permitido

$\overline{\overline{X}} \pm 2\sqrt{V(\overline{\overline{X}})}$ $\qquad 29 \pm 2\sqrt{0.0506}$

29 ± 2 (0.22) ; 29 ± 0.44; luego e = 0.44

Limite inferior = 28.56
Limite superior = 29.44
Interpretación:

Usando el muestreo aleatorio estratificado hemos estimado que el número promedio de horas por semana que se ve televisión en todos los hogares es de 29 horas, el error de estimación permitido es de 0.44 horas, con una probabilidad de 95.45% .

Calificación:

| | |
|---|---|
| Solución del caso No.1 ; | 34 puntos |
| Solución del caso No.2 ; | 66 puntos |
| total: | 100 puntos |

# VII. ESTIMACIÓN DE PARÁMETROS [9]

## VII.1 DEFINICIÓN

Estimación, es el proceso de usar un "estadístico muestral" para estimar el parámetro desconocido.

Un parámetro se puede estimar de dos maneras:

1.- Estimación de un punto: Es la selección de un número único que se utiliza para representar o estimar al parámetro.

Ejemplo: el precio de la leche es de $30.00 por litro en el D.F.

2.- Estimación de un intervalo: Es la estimación de un recorrido o rango dentro del cual se espera que está, contenido el parámetro.

Ejemplo: El precio de la leche está entre los $28.00 y los $35.00 por litro.

En otras palabras, cuando no se conocen los parámetros de la población, se pueden estimar recurriendo a muestras que permitan calcular intervalos dentro de los cuáles pueden estar contenido el valor de los parámetros.

Estos intervalos se llaman intervalos de confianza y sus extremos se llaman limites de confianza.

El grado de confianza, de el parámetro que, esté, contenido en el intervalo, se determina por el número de errores estándar a los cuales les corresponde un área bajo la curva que se denomina "coeficiente de confianza" ($\xi$ épsilon ). Al riesgo que el valor estimado de $\mu$ no se encuentra dentro del intervalo de confianza construido alrededor de la media de la muestra, se le llama nivel de significación ($\alpha$ alfa ) y es el área o probabilidad complementaria del coeficiente de confianza, sí, numéricamente $\xi = 1 - \alpha$ ó $E + \alpha = 1$ = ÁREA bajo la curva.

De esta manera, el intervalo de confianza se determina con:

Limite de confianza $\overline{X} \pm Z\alpha\sigma_{\bar{x}}$ . . . . . . . . . . (1).

Donde $Z_\alpha$ = Valor especifico de Z en la tabla, asociado con determinado coeficiente de confianza.

$\alpha$ = nivel de significación.

$\overline{X}$ = Media muestral.

$\sigma_{\bar{x}} = \sigma / \sqrt{n}$

Luego, para calcular el intervalo de confianza que contenga a $\mu$, es necesario conocer $\overline{X}$, n y $\sigma$.

De (1) tenemos $\overline{X} - Z_\alpha\sigma_{\bar{x}}$ límite inferior del intervalo.

$\overline{X} + Z_\alpha\sigma_{\bar{x}}$ límite superior del intervalo

| Z$\alpha$ | 1.0 | 1.96 | 2.0 | 3.0 |
|---|---|---|---|---|

| $\xi$ | 0.68 | 0.95 | 0.955 | 0.997 |
|---|---|---|---|---|
| $\alpha$ | 0.32 | 0.05 | 0.045 | 0.003 |

Ejemplo 1:

Se desea estimar el ingreso medio de los trabajadores de la compañía NESTLE, con el fin de estudiar las condiciones de trabajo de los trabajadores y en su caso pedir la revisión del contrato. Para ello, se selecciona una muestra aleatoria de 49 trabajadores cuyo ingreso medio resultó ser de $ 5 500.00/mes.

Estudios previos, realizados por la Facultad de Economía -UNAM- en esta empresa señala que la $\sigma$ del universo es de $ 700.00/mes.

Con un nivel de significación de 5%, estimar: el ingreso medio de los trabajadores;

n = 49

$\sigma$ = 700

$\overline{X}$ = $ 5 500.00

$\alpha$ = 5%

$Z\alpha = \pm 1.96$

$\xi$ = 95%

$\overline{X} \pm Z_{\alpha}\sigma_{\bar{x}}$

5 500 ± 1.96 (100)

5 500 ± 196

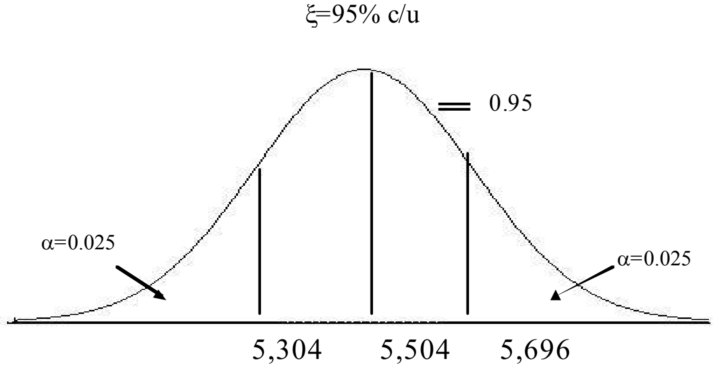
Limite de confianza = 5 500 ± 196

Intervalo de confianza = 5 304 a 5 696

donde el limite inferior = $ 5,304.00

donde el limite superior = $ 5,696.00

INTERPRETACIÓN : El ingreso medio de los trabajadores de la Cía. NESTLE, se halla entre los $ 5,304.00 y los $ 5,696.00 con una probabilidad del 95%. Gráficamente tendremos:



$$\bar{x} - Z_\alpha \sigma_{\bar{x}} \qquad \bar{x} \qquad \bar{x} + Z_\alpha \sigma_{\bar{x}}$$

Ejemplo 2.
Se desea estimar el gasto medio mensual en libros, del universo de estudiantes de la Universidad de Aguascalientes, con un nivel de significación del 5% y una muestra de 100 estudiantes seleccionados al azar, cuyo gasto medio mensual es de $288.000. La experiencia señala que la población tiene una desviación estándar de $ 20.000.

Como los limites de confianza = $\overline{X} \pm Z_\alpha \sigma_{\bar{x}}$

$\overline{X}$ = $ 288.00

n = 100 estudiantes

$\sigma$ = $ 20.00

$\alpha$ = 5%

Luego $Z_\alpha = \pm 1.96; y; \sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{20}{\sqrt{100}} = \dfrac{20}{10} = 2$

Por lo tanto limites de confianza = $\overline{X} \pm Z_\alpha \sigma_{\bar{x}}$

= $288 ± (1.96)(2)

= $288 ± 3.93

Intervalo de confianza = $284.08 a $ 291.92

Limite superior = $291.92

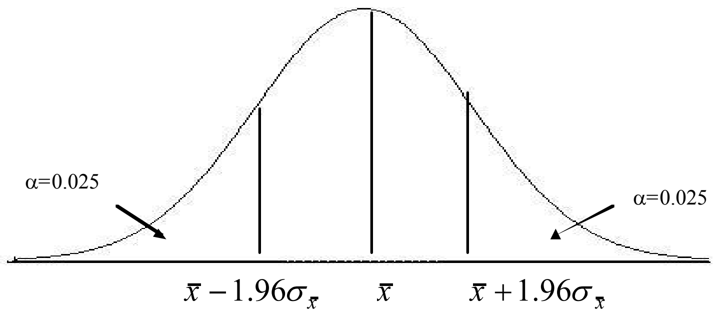
Limite inferior = $284.08

INTERPRETACIÓN :

Se estima que el gasto medio mensual en libros del universo constituidos por estudiantes, se halla entre $284.08 y los $291.92.

Gráficamente tendremos:

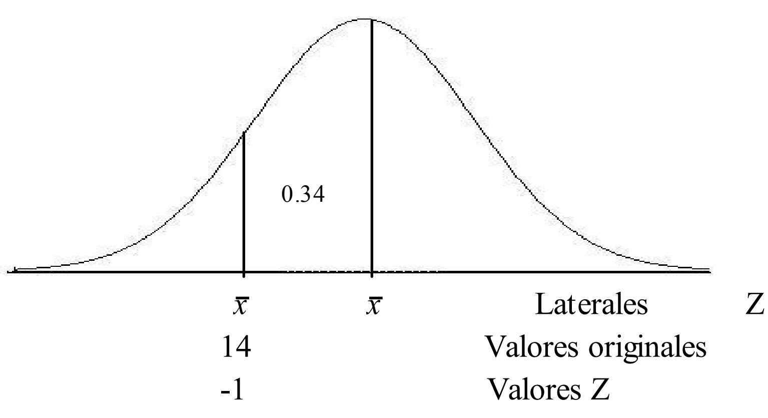
0.95 = ξ




## VII.2 DESIGUALDAD DE TCHEBYCHEFF

Antecedentes: Hemos visto en las distribuciones teóricas que con Z = x-μ/σ es posible conocer cierta porción o área de una distribución en el rango o intervalo x - μ.

Ejemplo, si los precios son:
μ = \$15.00 / Kg.
x = \$14.00 / Kg.
σ = \$1.00
Z = x - μ / σ = 14 -15 / 1 = -1 por lo tanto Z = -1
Gráficamente:




Donde Z = Número de desviaciones estándar de μ y cuya interpretación es: el 34% de los precios está entre los \$14 y \$15 (x -μ), o 0.34 es la probabilidad de que los precios estén entre \$14 y \$15. Para ello, se supone que los precios se distribuyen normalmente.

Cuando no se conoce la forma o características de la distribución, pero se conocen $\mu$ y $\sigma$, entonces con la desigualdad de TCHEBYCHEFF, cuya fórmula es:

$$P\left(|x-\mu| \geq K\sigma\right) \leq \frac{1}{K^2}$$

Se puede calcular la masa o porción de la distribución la interpretación es: la probabilidad de que un valor aleatorio ( $x_i$ ) de la distribución esté a un distancia igual o mayor de K desviaciones estándar de la media, cuando mucho es: $\frac{1}{K^2}$

En otra palabras, si K = 2, la probabilidad de que cualquier $x_i$ se encuentre a una distancia de 2 o más desviaciones estándar de la media, cuando más es:

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

De acuerdo con lo anterior en todos los casos la porción de la distribución situada a dos o más desviaciones estándar, nunca podrá exceder al 25% del total, independientemente de la forma de la distribución.

Otra interpretación sería que el 75% es la porción mínima de la distribución que se halla en la distancia comprendida dentro de $2\sigma$ de la media. En este caso, la fórmula sería: $1-\frac{1}{K^2} = K\sigma$ de $\mu$.

Comparando lo anterior con la distribución normal, se recordará que en la distancia de 2 $\sigma$ de $\mu$, se halla el 95.5% de la distribución, porción mucho mayor que los limites minimos dados por la desigualdad de TCHEBYCHEFF, situación atribuible a que dispone de menos información que la normal.

Sin embargo, la utilidad de TCHEBYCHEFF radica en que es aplicable a cualquier tipo de distribución.

Porcentaje del área de la ddistribución dentro de $k\sigma$ de $\mu$

| $Z = K = \frac{x-\mu}{\sigma}$ | Distribución Normal $\frac{1}{1-K^2}$ | Porcentaje mínimo dentro de $K\sigma$ |
|---|---|---|
| 1 | 68.27 % | 0.00 |
| 2 | 95.45 % | 75.00 |
| 3 | 99.73 % | 88.89 |
| 4 | 99.99 % | 93.75 |

Ejemplo:

Si conocemos la distribución del ingreso familiar en Pochutla, Oaxaca, tal que el ingreso medio mensual por familia es de \$10,000.00 con una desviación estándar de \$3,000.00 y deseamos conocer el rango que incluya cuando mucho al 50% de los ingresos familiares. Con la desigualdad de TCHEBYCHEFF podemos calcular el rango solicitado:

$$1 - \frac{1}{K^2} = 0.5$$

$K^2 = 2$

$K = 1.41$

Luego el rango será: \$10,000.00 ± 1.41σ

\$10,000.00 ± 1.41(3,000.00)

\$10,000.00 ± 4,330.00

Rango: de \$5,770.00 a \$14,330.00

si supiéramos que la distribución es normal, el rango se calculara así:

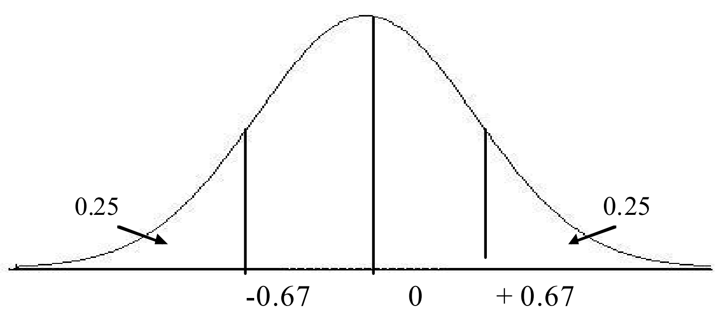
Luego: \$10,000.00 ± 0.67σ

\$10,000.00 ± 0.67(3,000)

\$10,000.00 ± \$2,010.00

Rango de: \$7,990.00 a \$12,010.00

Gráficamente:




## VII.3 ESTIMADORES INSESGADOS:

Son aquéllos cuya esperanza matemática es igual al parámetro poblacional.

$E(\overline{X}) = \mu$ pero $E(S^2) \neq \sigma^2$

Demostrar si la varianza muestral es estimador insesgado de $\sigma^2$;

con $x_i = x - \bar{x}$

| Muestras | Media Muestral | Varianza Sesgada | Varianza Insesgada |
|---|---|---|---|

| $X_i$ | | | | | $\bar{X}_i$ | $S^2 = \frac{\sum x_i^2}{n}$ | | | $\hat{S}^2 = \frac{\sum x_i^2}{n-1}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 , | 2 , | 3 , | 3 , | | 2.25 | 2.75 | ÷ | 4 | 2.75 | ÷ | 3 |
| 1 , | 2 , | 3 , | 4 , | | 2.50 | 5.00 | ÷ | 4 | 5.00 | ÷ | 3 |
| 1 , | 2 , | 3 , | 5 , | | 2.75 | 8.75 | ÷ | 4 | 8.75 | ÷ | 3 |
| 1 , | 2 , | 3 , | 4 , | | 2.50 | 5.00 | ÷ | 4 | 5.00 | ÷ | 3 |
| 1 , | 2 , | 3 , | 5 , | | 2.75 | 8.75 | ÷ | 4 | 8.75 | ÷ | 3 |
| 1 , | 2 , | 4 , | 5 , | | 3.00 | 10.00 | ÷ | 4 | 10.00 | ÷ | 3 |
| 1 , | 3 , | 3 , | 4 , | | 2.75 | 4.75 | ÷ | 4 | 4.75 | ÷ | 3 |
| 1 , | 3 , | 3 , | 5 , | | 3.00 | 8.00 | ÷ | 4 | 8.00 | ÷ | 3 |
| 1 , | 3 , | 4 , | 5 , | | 3.25 | 8.75 | ÷ | 4 | 8.75 | ÷ | 3 |
| 1 , | 3 , | 4 , | 5 , | | 3.25 | 8.75 | ÷ | 4 | 8.75 | ÷ | 3 |
| 2 , | 3 , | 3 , | 4 , | | 3.00 | 2.00 | ÷ | 4 | 2.00 | ÷ | 3 |
| 2 , | 3 , | 3 , | 5 , | | 3.25 | 4.75 | ÷ | 4 | 4.75 | ÷ | 3 |
| 2 , | 3 , | 4 , | 5 , | | 3.50 | 5.00 | ÷ | 4 | 5.00 | ÷ | 3 |
| 2 , | 3 , | 4 , | 5 , | | 3.50 | 5.00 | ÷ | 4 | 5.00 | ÷ | 3 |
| 3 , | 3 , | 4 , | 5 , | | 3.75 | 2.75 | ÷ | 4 | 2.75 | ÷ | 3 |
| Total | 15 | | | | 45.00 | 90/4=22.5 | | | 90/3=30 | | |
| E($\bar{X}$) | | | | | 45/45=3 | 22.5/15=1.5 | | | 30/15=2 | | |
| Parámetro | | | | | μ=3 | $\sigma^2$=1.67 | | | $\hat{\sigma}^2 = 10/5 = 2$ | | |

Calculo de las variancias de cada una de las muestras (primera muestra).

| $X_i$ | $(X_i - \bar{X})$ | $(X_i - \bar{X})^2$ |
|---|---|---|
| 1 | -1.25 | 1.5625 |
| 2 | -0.25 | 0.0625 |
| 3 | 0.75 | 0.5625 |
| 3 | 0.75 | 0.5625 |
| 9 | 0 | 2.75 |

$\bar{x}_i$ = 9/4 = 2.25

$\sigma^2$ = 2.750/4 = 0.6825

1ª CONCLUSIÓN

Se dice que $\bar{X}$ es un estimador insesgado de μ porque su esperanza matemática es igual al valor del parámetro, cuyo universo es:

$E(\bar{X}) = \mu$

| Trabajador | Salario/hrs (x) | X = X - μ | X = (X - μ)$^2$ |
|---|---|---|---|
| A | 1 | -2 | 4 |
| B | 2 | -1 | 1 |
| C | 3 | 0 | 0 |
| D | 3 | 0 | 0 |
| E | 4 | 1 | 1 |
| F | 5 | 2 | 4 |
| | 18 | 0 | 10 |

Así: $\mu$=18/6=3; $\sigma^2$=10/6=1.67 y $\hat{\sigma}^2 = 10/5 = 2$.

2ª CONCLUSIÓN

$S^2$ no es un estimador insesgado de $\sigma^2$ porque su esperanza matemática es diferente del valor del parámetro poblacional.

$E(S^2) \neq \sigma^2$

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n}; \sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

$\mathrm{E}\left(S^2\right) = 1.5$ pero $\sigma^2$ =1.67; luego $\mathrm{E}\left(S^2\right) \neq \sigma^2$

3a. CONCLUSIÓN

$\hat{S}^2$ es un estimador insesgado de $\hat{\sigma}^2$ porque $E(\hat{S}^2) = \hat{\sigma}^2$ donde

$$\hat{S}^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}; \hat{\sigma}^2 = \frac{\sum (x_i - \mu)^2}{N-1}$$

Ahora bien $\hat{S}^2$ también se puede obtener de:

$$\hat{S}^2 = S^2 \frac{n}{n-1}; \hat{\sigma}^2 = \sigma^2 \frac{N}{N-1} = \hat{\sigma}^2; \frac{\sum (x_i - \bar{x})^2}{n} * \frac{n}{n-1} = \hat{S}^2 = \frac{(x_i - \bar{x})^2}{n-1}$$

EJEMPLO # 2 :

Con N = 1,2,3 ; $\mu$ = 2; $\sigma$ = 0.67

Si tomamos $\binom{3}{2} = 3$ muestras de tamaño 2.

| Media | | Varianza | |
|---|---|---|---|
| Muestra | Muestral | Sesgada | Insesgada |
| 1 , 2 | 1.50 | 0.50 ÷ 2 | 0.50 ÷ 1 |
| 1 , 3 | 2.00 | 2.00 ÷ 2 | 2.00 ÷ 1 |
| 2 , 3 | 2.50 | 0.50 ÷ 2 | 0.50 ÷ 1 |
| 3 | 6 | 3/2 | 3/1 |

E( $\overline{X}$ ) 6/3 = 2 ; E($S^2$)=1.5/3 = 0.50 3/3 =1

Como $\sigma^2 = 0.67 \neq E(S^2) = 0.50$

En cambio, si $\hat{\sigma}^2 = \dfrac{\sum (x_i - \mu)^2}{N-1} = \dfrac{2}{2} = 1$

Luego $E(\hat{S}^2) = 1 = \hat{\sigma}^2$

**<u>P COMO ESTIMADOR INSESGADO DE P</u>:**

p es un estimador insesgado de P ya que E (p) = P

Ejemplo : Partiendo de una población binaria donde tenemos A, B, C, personas que fuman con valor 1 y X, Y, Z personas que no fuman con valor cero.

| Personas | X |
|---|---|
| A | 1 |
| B | 1 |
| C | 1 |
| X | 0 |
| Y | 0 |
| Z | 0 |

μ= P = 3/6 = 0.5 = 50%; $\sigma = \sqrt{PQ} = 0.5$ Donde μ y σ son los parámetros de la población.

Si ahora deseamos conocer la proporción de fumadores en muestras de n = 4; la media de las proporciones y su error estándar, tendremos en primer lugar que hay:

$\binom{6}{4} = 15$ muestras de cuatro personas en un muestreo sin reemplazo.

## VII.3.1 CALCULO DE LAS PROPORCIONES MUESTRALES

| Personas en la muestra, n=4 | Proporción de fumadores en la muestra | Probabilidad de c/u de las 15 muestras |
|---|---|---|
| 1 .- A , X , Y , Z , | 1 ÷ 4 = 0.25 | 1 ÷ 15 |
| 2 .- B , X , Y , Z , | 1 ÷ 4 = 0.25 | 1 ÷ 15 |
| 3 .- C , X , Y , Z , | 1 ÷ 4 = 0.25 | 1 ÷ 15 |
| 4 .- A , B , X , Y , | 2 ÷ 4 = 0.50 | 1 ÷ 15 |
| 5 .- A , B , X , Z , | 2 ÷ 4 = 0.50 | 1 ÷ 15 |
| 6 .- A , B , Y , Z , | 2 ÷ 4 = 0.50 | 1 ÷ 15 |
| 7 .- A , C , X , Z , | 2 ÷ 4 = 0.50 | 1 ÷ 15 |
| 8 .- A , C , Y , Z , | 2 ÷ 4 = 0.50 | 1 ÷ 15 |
| 9 .- B , C , X , Y , | 2 ÷ 4 = 0.50 | 1 ÷ 15 |
| 10 .- B , C , Y , Z , | 2 ÷ 4 = 0.50 | 1 ÷ 15 |
| 11 .- A , B , C , X , | 3 ÷ 4 = 0.75 | 1 ÷ 15 |
| 12 .- A , B , C , Y , | 3 ÷ 4 = 0.75 | 1 ÷ 15 |
| 13 .- A , B , C , Z , | 3 ÷ 4 = 0.75 | 1 ÷ 15 |
| 14 .- A , C , X , Y , | 2 ÷ 4 = 0.50 | 1 ÷ 15 |
| 15 .- B , C , Y , Z , | 2 ÷ 4 = 0.50 | 1 ÷ 15 |
| 15 | 7.5 | 15/15 |

luego, la media de las proporciones denotada por $\overline{p} = \dfrac{7.50}{15} = 0.50 = \mathrm{P}$

que significa que $\overline{p} = \mathrm{E(p)} = \mathrm{P}$.

El error estándar de la proporción es :

$$\sigma_p = \sqrt{\frac{PQ}{n}\frac{N-n}{N-1}} = \sqrt{\frac{0.5(0.5)}{4}\frac{6-4}{6-1}} = \sqrt{0.025} = 0.158$$

$\sigma_p = 0.158$

Ahora bien :

1.- $S^2$ no es un estimador insesgado de $\sigma^2$, ya que $E(\lozenge^2) \neq \sigma^2$ donde

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n} ; \sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

2.- En cambio, si tomamos $\hat{S}^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$; tendremos:

E($\hat{S}^2$)=$\hat{\sigma}^2 = \dfrac{\sum (x_i - \mu)^2}{N-1}$

3.- Cuando N $\to \infty$ y no conocemos a $\sigma^2$, entonces $\hat{S}^2$ se hace un estimador insesgado de $\sigma^2$ porque si:

$\hat{S}^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$; y $\hat{\sigma}^2 = \dfrac{\sum (x_i - \mu)^2}{N-1}$ cuando N $\to \infty$, tenemos:

$\hat{S}^2 = \sigma^2$ porque $\dfrac{N}{N-1} = 1$

o sea, que $\hat{S}^2$ estima $\sigma^2$ cuando N $\to \infty$ y $\hat{\sigma}\bar{x}$ estima $\hat{S}\bar{x}$; es decir, cuando no se conozca $\sigma$ usaremos $\hat{S}_{\bar{x}} = \dfrac{\hat{S}}{\sqrt{n}}$ en lugar de: $\sigma\bar{x} = \dfrac{\sigma}{\sqrt{n}}$ en $\bar{x} \pm Z_\alpha \sigma_{\bar{x}}$; por lo que la expresión de los limites de confianza se convierte en $\bar{x} \pm Z_\alpha \hat{S}_{\bar{x}}$.

4.- También podemos decir que cuando N > 100, $\hat{S}^2 = S^2$, tal que $S\bar{x} = \dfrac{S}{\sqrt{n}}$

### Estimadores eficientes, suficientes y consistentes.[13]

**VII.4 ESTIMADOR EFICIENTE:** Es aquel que tiene la variancia mínima.

**VII.5 ESTIMADOR SUFICIENTE ($\bar{X}$):** Es el que utiliza toda la información que posee la muestra sobre el parámetro que se estima.

**VII.6 ESTIMADOR CONSISTENTE ($\bar{X}$):** Es el que se aproxima al parámetro ($\mu$) que se va a estimar, al aumentar la muestra. Con literales $\bar{X} \to \mu$, cuando n $\to$ N .