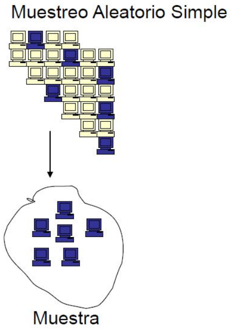
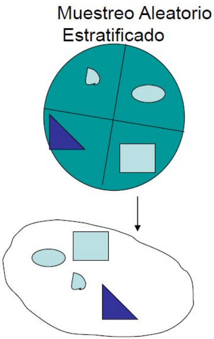
# Muestreo Estadístico-SEMANA-2

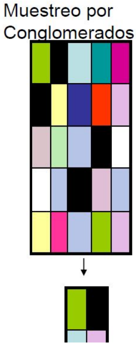
Raúl Alberto Pérez
raperez1@unal.edu.co
Profesor Asociado - Escuela de Estadística
Universidad Nacional de Colombia, Sede Medellín

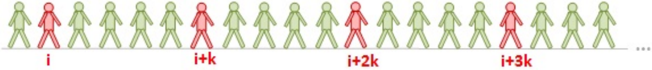
Semestre 2021-I

# Algunos tipos de Muestreo son los siguientes:



# Diseños Muestrales Básicos

Entre los distintos métodos de *muestreo probabilístico* para poblaciones finitas existen los considerados *métodos básicos* que al combinarse originan otros métodos llamados *polietápicos* los cuales son mas complejos de estudiar.

A continuación se hace una descripción general breve de cada uno de estos métodos.

# Muestreo Aleatorio Simple (MAS) sin reemplazo

## El MAS tiene los siguientes Supuestos:

1. La población es homogénea con respecto a la característica de interés.
2. Todas las muestras de tamaño $n$ tienen la misma probabilidad de ser seleccionada.
3. Se selecciona una muestra aleatoria de tamaño **n** de tal forma que cada muestra tenga probabilidad:

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

- La importancia de este diseño es que sirve como base para las definición de muchos otros.
- Es muy fácil de aplicar cuando se cuenta con un marco muestral completamente identificado.

# Muestreo Estratificado Aleatorio Simple:

**Supuestos**

1. Población heterogénea.
2. Se puede dividir la población en subpoblaciones homogéneas (**ESTRATOS**) internamente no traslapadas.
3. De cada subpoblación (**ESTRATO**) se selecciona una MAS de unidades de muestreo para conformar la muestra final.

# Muestreo Sistemático Aleatorio:

## Supuestos:

1. Útil para cuando hay población en movimiento. Proporciona más información por unidad de costo que la proporcionada por MAS en poblaciones que tienen patrones en la organización de elementos.
2. Se escoge un número aleatorio de los primeras k-unidades del marco, y luego las siguientes unidades se escogen igualmente distanciadas, se sigue el procedimiento hasta completar la muestra de n-unidades.
3. Las estimaciones se hacen en forma similar a como se hace en MAS.

**Usos:** En muestreos de Control de calidad industrial, investigaciones de mercado.

# Muestreo Por Conglomerados:

**Supuestos:**

1. La población esta conformada por subpoblaciones (**conglomerados**) pequeñas homogéneas internamente, disjuntas de unidades de muestreo.

2. Se selecciona una MAS de **conglomerados** y se muestrean todas las unidades de cada conglomerado seleccionado. Es decir, se seleccionan, antes que unidades elementales, grupos de unidades.

- Es recomendable cuando existe cercanía de algún tipo entre las unidades finales. La cercanía es generalmente de tipo físico o administrativo. Su principal ventaja radica principalmente en la disminución de los costos de recolección. Sin embargo la eficiencia relativa de sus estimadores puede ser muy baja si se le compara con otros diseños.

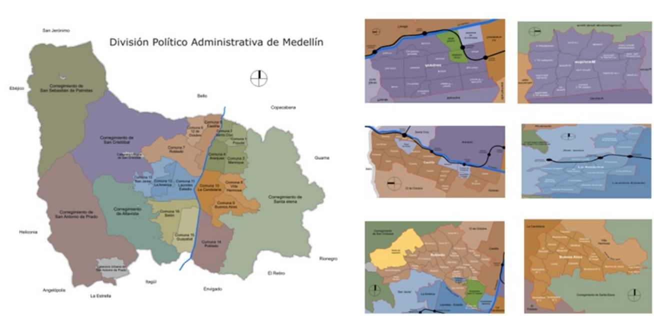
# Diseños Muestrales Avanzados:

Además de los diseños básicos mencionados anteriormente existen muchos otros, más complejos, que se usan en estudios de gran escala o cuando las características poblacionales a medir tienen ciertas condiciones especiales.

Una categoría general que podría abarcar a la mayoría de estos métodos, sería la del *Muestreo Multietápico o polietápico* que como su nombre lo indica consta de varias etapas en cada una de las cuales es necesario usar un diseño simple, ie. cualquier diseño que solamente comprenda una etapa de selección.

# Muestreo Multietápico

## Muestreo en Varias Etapas:



# Previo a la selección de la Muestra

## Muestreo Piloto

Constituye en la selección de una **muestra inicial** para:

1. La estimación la **varianza** de la variable de interés: $\sigma^2$.
2. Verificación de los cuestionarios o encuestas- **Validar las preguntas de la encuesta.**

- **NOTA:** Se recomienda que la Muestra PILOTO sea seleccionada con el mismo tipo de muestreo elegido.

# Previo a la selección de la Muestra

## Determinación del Tamaño de Muestra

Una vez elegido el **tipo de muestreo** a ser utilizado y los cálculos obtenidos del **muestreo piloto** se halla el número mínimo (**n**) de unidades de muestreo que estarán en la muestra.

El **tamaño de muestra** tiene dos componentes importantes que se fijan por el investigador antes de la selección de la muestra:

- **El nivel de Confianza**: $1 - \alpha$: Valor asociado a las estimaciones de los parámetros de inteés.

- **El límite en el error de estimación**, B: el margen de error que está dispuesto a cometer el investigador en la estimación del parámetro de interés:

$$P\left[|\theta - \hat{\theta}| < B\right] = 1 - \alpha$$

# Algunos Problemas Prácticos

**P1: Proporción de estudiantes** de primer semestre de la Facultad de Minas que utilizan la bicicleta para llegar a la Universidad.

### Definición de Términos

**Elemento:** Cada uno de los estudiantes matriculados en el primer semestre de la Facultad de Minas.

**Población:** Todos los estudiantes matriculados en el primer semestre de la Facultad de Minas.

**Unidad de Muestreo:** Cada una de las asignaturas introductorias de cada una de las carreras de la Facultad de Minas (FM).

**Marco:** Listado completo de todas asignaturas introductorias de cada una de las carreras de la FM.

# Algunos Problemas Prácticos

**P2: Proporción de Tiendas de Barrio** en Medellín con nombre Visible.

### Definición de Términos

**Elemento:** Tiendas de Barrio en Medellín.

**Población:** Todas las tiendas de Barrio en Medellín.

**Unidad de Muestreo:** Agrupaciones de manzanas.

**Marco:** Listado completo de todas las UM por cada uno de los Barrios

# Algunos Problemas Prácticos

**P3:** **Total de Viviendas de un Barrio** en Medellín que poseen mascotas con nombre Visible.

### Definición de Términos

**Elemento:** Cada una de las viviendas del Barrio.

**Población:** Todas las viviendas del Barrio en estudio.

**Unidad de Muestreo:** Cada una de las viviendas del Barrio.

**Marco:** Listado completo de todas las UM del Barrio bajo estudio.

# Algunos Problemas Prácticos

**P4:** **Monto promedio de la matrícula** de los estudiantes de la UNAL Sede Medellín que pertenecen a las selecciones deportivas.

## Definición de Términos

**Elemento:** Cada uno de los estudiantes que integran las selecciones deportivas de la Unal Sede Medellín.

**Población:** Todos los estudiantes que integran las selecciones deportivas de la UNAL Sede Medellín.

**Unidad de Muestreo:** Cada una de las selecciones deportivas de la Unal sede Medellín.

**Marco:** Listado completo de todas las selecciones deportivas de la Unal sede Medellín

# Instructivo Preinforme TRABAJO PRÁCTICO

## Instructivo Preinforme TRABAJO PRÁCTICO

- **Equipos:** Máximo cuatro integrantes.
- **Fecha Entrega:** **Marzo 12 de 2021**

**Objetivo:**
Cada uno de los equipos debe hacer una propuesta de un estudio de muestreo por encuestas que sea de interés (temática llamativa que les interese estudiar a todos los integrantes del equipo) y que sea factible de ser realizada durante todo el semestre.

**Etapa 1:** **Preinforme.**
En esta etapa, para el problema propuesto por cada una de los grupos, se debe enviar por email (raperez1@unal.edu.co), por escrito en computador(Portada, Tabla de contenido, números de página, bibliografía), y de forma detallada:

# Instructivo Preinforme TRABAJO PRÁCTICO

## Instructivo Preinforme TRABAJO PRÁCTICO

a) **Motivación** del problema a resolver (antecedentes)

b) Identificar:
   - **Parámetro** (s) a estimar.
   - **Elemento**.
   - **Población**.
   - **Marco Muestral**.

c) **Tipo de Muestreo** propuesto para la estimación del(de los) parámetro(s) de interés (explicar la forma como se usaría dicho muestreo para la obtención de la información).

# Métodos de recolección de datos

1. Encuestas personales
2. *Encuestas por teléfono*

- **NOTA:** Con frecuencia se puede encontrar información más objetiva a través de **información directa**, que al realizar una *entrevista* o el diligenciamiento de un ***cuestionario enviado por correo***.

# Reglas generales para la construcción de un cuestionario

## Elaboración del Cuestionario

1. **Ordene las preguntas** que contrarresten problemas potenciales, si es necesario **explique el significado** de cada una de las preguntas.

2. Considere **preguntas cerradas**, de preferencia, las preguntas abiertas son necesarias sólo en caso de tener que dar alguna justificación.

3. Recuerde ser cauteloso en la **redacción** de las preguntas, las preguntas deben ser directas pero no generar indisposición por parte del encuestado.

# Reglas generales para la construcción de un cuestionario

## Planeación de una encuesta

1. Establezca los objetivos.
2. Defina la población objetivo.
3. Defina el marco muestral.
4. Diseñe el muestreo.
5. Indique el método de recolección de la información.
6. Instrumento de medición.

# Reglas generales para la construcción de un cuestionario

## Planeación de una encuesta

7. Seleccione y capacite los investigadores de campo.
8. Realice una prueba piloto.
9. Organice el trabajo de campo.
10. Organice el manejo de datos.
11. Analice los datos.
12. Realice las inferencias adecuadas a partir de la información obtenida.

# Muestreo Aleatorio Simple (M.A.S)

## TEMÁTICA A DESARROLLAR:

1. ¿Cómo se selecciona la muestra?
2. Estimaciones de los parámetros de interés: $\mu$, $\tau$, $p$
3. Cálculo de la Varianza de los Estimadores

## Estudio de un CASO

**EJEMPLO**

Recuerde la Aplicación 1: **Estimar la Proporción de estudiantes de recién ingreso de la Facultad de Minas que utilizan bicicleta para llegar a la Universidad.**

Suponga que un investigador encuesta aleatoriamente **25** estudiantes de la población **OBJETIVO**, que tiene un total de **739** estudiantes.

# ESTUDIO DE UN CASO

Adicional a la pregunta de interés, $Z_i$, se registra el **valor de la matrícula** (VM: en miles de pesos). Los datos se muestran a continuación:

| **Est.** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **VM** | 834 | 576 | 930 | 282 | 258 | 1005 | 930 | 870 | 663 | 504 | 564 | 636 | 276 |
| $Z_i$ | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

| **Est.** | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **VM** | 168 | 426 | 111 | 558 | 663 | 657 | 915 | 1815 | 1230 | 2580 | 1278 | 2700 |
| $Z_i$ | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |

$Y_i$: **Valor de la matrícula** del $i$-ésimo estudiante en la muestra, VM\ $i = 1, 2, \ldots, n = 25.$\

$$Z_i = \begin{cases} 1 & \text{Estudiante llega en bicicleta a la Unal} \\ 0 & \text{Estudiante NO llega en bicicleta a la Unal} \end{cases}$$

# PREGUNTAS POR RESOLVER

1. Identificar: **ELEMENTO, POBLACIÓN, UNIDAD DE MUESTREO, MARCO**
2. Indicar como seleccionó la muestra el investigador para garantizar que sea una Muestra Aleatoria Simple.
3. Estime el valor de la Matrícula promedio y el Número total de estudiantes de recién ingreso de la Facultad de Minas que llegan en Bicicleta a la UNAL.
4. Estime la proporción de estudiantes de la Facultad de Minas de recién ingreso que llegan en bicicleta a la UNAL.

# Muestreo Aleatorio Simple (M.A.S)

Es un procedimiento estadístico a partir del cual se selecciona una MUESTRA de tamaño $n$ de una población de $N$ unidades garantizando que cada muestra posible de tamaño $n$ tenga la misma probabilidad de ser seleccionada.

La muestra obtenida se denomina muestra aleatoria simple (MAS).

Lo anterior conlleva a que todas las unidades de la población también tengan la misma probabilidad de ser seleccionadas, ie. de formar parte de la muestra (es un muestreo equiprobable).

# MUESTREO ALEATORIO SIMPLE

## NOTAS SOBRE EL M.A.S

1. En la práctica una MAS es seleccionada unidad por unidad.
2. El muestreo se realiza sin reemplazo, es decir las Unidades de Muestreo no se repiten en la Muestra Aleatoria que se seleccione.
3. El MAS se utiliza cuando:
   - La población es HOMOGÉNEA con respecto a la característica de interés
   - Las estimaciones de interés se refieren a toda la POBLACIÓN objeto de estudio y no a SUBGRUPOS O SUBPOBLACIONES de la POBLACIÓN completa.
4. Preguntas a resolver:

   **a.** ¿Cómo seleccionar la MUESTRA?

   **b.** ¿Cómo estimar los parámetros de interés?

# Muestreo Aleatorio Simple

## Selección de la MAS

Las Unidades muestrales se enumeran de 1 a N, se generan n números aleatorios por algún mecanismo de aleatorización:

**Mecanismos de Selección MAS**

1. En el programa Estadístico R:
   Usar la función sample (rótulos, n). Por ejemplo:
   $muestra <- sample(1:739, 25); sort(muestra)$
   Lo anterior proporciona los códigos de las **Unidades de Muestreo** a ser incluidas en la MAS de tamaño $n = 25$, de una población con $N = 739$ UM.
   Se identifican en la población a que unidades corresponden y se muestrean dichas unidades.
2. En EXCEL: En la celda: $= ALEATORIO.ENTRE(1; N)$, genera un número aleatorio entre 1 y N
3. En la Calculadora se identifica la tecla que genera números aleatorios en $(0, 1)$, sea $a$ dicho número, luego el código a ser seleccionado será:

$$||\mathbf{a} * \mathbf{N} + 1||,$$

$||x||$: Denota parte entera de $x$.

# Algunas propiedades del MAS

1. La probabilidad de que la observación $U_i$, para $i = 1, 2, \dots, N$ aparezca en cualesquier selección de la muestra es:

$$P_i = 1/N$$

2. La probabilidad de que la observación $U_i$ aparezca en una extracción específica dado que la observación $U_j$ aparece en cualesquier otra extracción es:

$$P_{i|j} = 1/(N-1)$$

3. La probabilidad de que las observaciones $U_i$ y $U_j$ aparezcan en dos extracciones específicas es:

$$P_{ij} = \frac{1}{N}\frac{1}{N-1}$$

**EJEMPLO:**

Para ilustrar las propiedades anteriores en este tipo de muestreo considere el siguiente ejemplo de 15-muestras de tamaño $n = 2$ de una población de tamaño $N = 6$, con sus resúmenes estadísticos dados en la tabla presentada a continuación de los datos poblacionales.

| Observaciones | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $Y_5$ | $Y_6$ | $\mu$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|
| Valores | 670 | 720 | 650 | 720 | 560 | 700 | 670 | 3066.67 |

A continuación de la tabla resumen también se tienen: la media de las medias muestrales, la media de las varianzas-muestrales y las medias de las desviaciones estándar.

| Muestras | Unid | Obs | Medias=$\overline{X}$ | Varianza=$S^2$ | S:D |
|---|---|---|---|---|---|
| 1 | 1,2 | 670,720 | 695 | 1250 | 35.36 |
| 2 | 1,3 | 670,650 | 660 | 200 | 14.14 |
| 3 | 1,4 | 670,720 | 695 | 1250 | 35.36 |
| 4 | 1,5 | 670,560 | 615 | 6050 | 77.28 |
| 5 | 1,6 | 670,700 | 685 | 450 | 21.21 |
| 6 | 2,3 | 720,650 | 685 | 2450 | 49.50 |
| 7 | 2,4 | 720,720 | 720 | 0 | 0 |
| 8 | 2,5 | 720,560 | 640 | 12800 | 113.14 |
| 9 | 2,6 | 720,700 | 710 | 200 | 14.14 |
| 10 | 3,4 | 650,720 | 685 | 2450 | 49.50 |
| 11 | 3,5 | 650,560 | 605 | 4050 | 63.64 |
| 12 | 3,6 | 650,700 | 675 | 1250 | 35.36 |
| 13 | 4,5 | 720,560 | 640 | 12800 | 113.14 |
| 14 | 4,6 | 720,700 | 710 | 200 | 14.14 |
| 15 | 5,6 | 560,700 | 630 | 9800 | 99.00 |
| Valor esperado= $E[\hat{\theta}]$ | | | 670 | | 49.03 |

## Distribución Muestral de $\overline{X}$:

| $\overline{X}$ | 695 | 660 | 615 | 685 | 720 | 640 | 710 | 605 | 675 | 630 | *Total* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $P[\overline{X}=\overline{x}]$ | $\frac{2}{15}$ | $\frac{1}{15}$ | $\frac{1}{15}$ | $\frac{3}{15}$ | $\frac{1}{15}$ | $\frac{2}{15}$ | $\frac{2}{15}$ | $\frac{1}{15}$ | $\frac{1}{15}$ | $\frac{1}{15}$ | 1 |

Ahora,

$$E[\overline{X}] = \sum_{1}^{10} \overline{x}_i P[\overline{x}_i] = 695(\frac{2}{15}) + 660(\frac{1}{15}) + \cdots + 630(\frac{1}{15}) = 670 = \mu$$

$$Var[\overline{X}] = E(\overline{X} - E[\overline{X}])^2 = \sum_{i=1}^{10} (\overline{x}_i - 670)^2 P[\overline{x}_i]$$

$$= (695-670)^2(\frac{2}{15}) + (660-670)^2(\frac{1}{15}) + \cdots + (630-670)^2(\frac{1}{15})$$

$$= 1226.67 = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n} = \left(\frac{6-2}{6-1}\right)\frac{3066.67}{2}$$

Ejercicio: Hallar la distribución muestral de $S^2$ y verificar que:
$E[S^2] = (\frac{N}{N-1})\sigma^2$

Para los datos presentados en la tabla anterior se tiene que:

1. Existen
$$\binom{N}{n} = \binom{6}{2} = 15$$
muestras posibles de tamaño $n = 2$, **y cada unidad poblacional** $U_i$ $i = 1, 2, \ldots, N = 6$-aparecen en:
$$\binom{1}{1}\binom{N-1}{n-1} = \binom{5}{1} = 5$$
muestras.

Por lo tanto la probabilidad de que una unidad poblacional aparezca en la muestra es:

$$\frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N} = \frac{2}{6} = \frac{1}{3} = \frac{5}{15}$$

2. **Cada par de las $N$ unidades poblacionales** aparecen en:

$$\binom{2}{2}\binom{N-2}{n-2} = \binom{4}{0} = 1$$

de las muestras y por lo tanto tienen probabilidad de que aparezcan en las muestras de:

$$\frac{\binom{2}{2}\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)} = \frac{2(2-1)}{6(6-1)} = \frac{2}{30} = \frac{1}{15}$$

3. Similarmente resulta para el caso de más combinaciones de unidades poblacionales.

Cada terna de las $N$ unidades poblacionales aparecen en:

$$\binom{3}{3}\binom{N-3}{n-3}$$

de las muestras.

# Más Acerca del Muestreo Aleatorio Simple

En este tipo de muestreo ninguna unidad de la población puede estar representada en la muestra mas de una vez.

La probabilidad de seleccionar una muestra específica de tamaño $n$ está dada por:

$$\frac{1}{\binom{N}{n}} = \frac{1}{\frac{N!}{n!(N-n)!}} = \frac{n!(N-n)!}{N!}$$

pues el número total de muestras posibles de tamaño $n$ que pueden ser seleccionadas de un conjunto (o población) de tamaño $N$ es:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}.$$

**La probabilidad anterior se obtiene como sigue**:

En la primera extracción, la probabilidad de que se seleccione una de estas unidades es: $\frac{n}{N}$; en la segunda extracción, la probabilidad de que se seleccione una de las restantes $(n-1)$-unidades es: $\frac{n-1}{N-1}$ y así sucesivamente, de donde la probabilidad de que se extraigan las $n$-unidades específicas es,

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

como se ve a continuación.

$$\frac{n}{N} \times \frac{n-1}{N-1} \times \frac{n-2}{N-2} \cdots \times \frac{n-(n-2)}{N-(n-2)} \times \frac{n-(n-1)}{N-(n-1)}$$

$$= \frac{n}{N} \times \frac{n-1}{N-1} \times \frac{n-2}{N-2} \cdots \times \frac{2}{N-(n-2)} \times \frac{1}{N-(n-1)}$$

$$= \frac{n!}{\frac{N!}{(N-n)!}} = \frac{n!(N-n)!}{N!} = \frac{1}{\binom{N}{n}}$$

Pues,

$$\begin{aligned} \frac{N!}{(N-n)!} &= \frac{N(N-1)\ldots[N-(n-2)][N-(n-1)](N-n)!}{(N-n)!} \\ &= N(N-1)\ldots[N-(n-2)][N-(n-1)] \end{aligned}$$

Similarmente, la probabilidad de que una unidad particular de la población esté presente en la muestra es:

$$\frac{\text{Número de formas posibles de que la unidad caiga en la muestra}}{\text{Número total de muestras posibles}}$$

$$= \frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{\frac{(N-1)!}{(n-1)!(N-1-(n-1))!}}{\frac{N!}{n!(N-n)!}}$$

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

$$= \frac{n}{N}$$

# Estimación de los parámetros en M.A.S

## Ejemplos donde es Improtante Estimar la Media Poblacional

| Área | Variable de Interés | Media Poblacional |
|---|---|---|
| Agricultura | Producción anula de maíz en una finca de colombia en el año 2001 | Producción media anual de maíz por finca en colombia en el año 2001 |
| Economía | Gasto mensual en alimentación de una familia colombiana en el año 2001 | Gasto promedio mensual en alimentación por familia en colombia en el año 2001 |
| Transporte | Número diario de pasajeros transportados por un vehículo de servicio público en Medellín en el año 2001 | Promedio diario de pasajeros transportados por un vehículo de servicio público en Medellín en el año 2001 |

# Estimación de la Media Poblacional $\mu$ en M.A.S

En el M.A.S sin reemplazo, la media muestral definida como:

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

es un estimador insesgado de la media poblacional $\mu$.

$$\text{es decir :} \quad E[\bar{y}] = \mu$$

Además, se puede mostrar que:

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

donde $\left(\frac{N-n}{N-1}\right)$ se conoce como el **factor de corrección**, fc.

# Estimación de la Media Poblacional $\mu$ en M.A.S

El valor de $\bar{Y}$ por si sólo no revela con exactitud el valor de $\mu$, es necesario hallar un Intervalo de Confianza para $\mu$.

Para ello se requiere determinar el límite estimado del Error de Estimación.

# Estimación de la Media Poblacional $\mu$ en M.A.S

Las expresiones anteriores se pueden verificar con un ejemplo numérico sencillo.

Suponga que su población está conformada por cuatro, 4, **Unidades de Muestreo**, cuyos valores son conocidos:

$$\{1, 4, 7, 10\}.$$

Y se seleccionan muestras de tamaño **dos**, 2.

# Estimación de la Media Poblacional $\mu$ en M.A.S

Se halla la distribución muestral de $\bar{Y}$ para todas las muestras posibles de tamaño **dos**:

| muestra | | $\bar{y}$ | $P[\bar{Y} = \bar{y}]$ | $s^2$ |
|---|---|---|---|---|
| 1 | $\{1, 4\}$ | 2.5 | $\frac{1}{6}$ | ...... |
| 2 | $\{1, 7\}$ | 4.0 | $\frac{1}{6}$ | ...... |
| 3 | $\{1, 10\}$ | 5.5 | $\frac{1}{6}$ | ...... |
| 4 | $\{4, 7\}$ | 5.5 | $\frac{1}{6}$ | ...... |
| 5 | $\{4, 10\}$ | 7.0 | $\frac{1}{6}$ | ...... |
| 6 | $\{7, 10\}$ | 8.5 | $\frac{1}{6}$ | ...... |

# Estimación de la Media Poblacional $\mu$ en M.A.S

La distribuión de mustral de $\bar{y}$ s:

| $\bar{Y}_i$ | 2.5 | 4.0 | 5.5 | 5.5 | 7.0 | 8.5 |
|---|---|---|---|---|---|---|
| $P[\bar{Y}_i]$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

$$
\begin{aligned}
E[\bar{Y}] &= \sum_{i=1}^{6} \bar{y}_i P\left[\bar{Y} = \bar{y}_i\right] \\
&= 2.5\frac{1}{6} + 4\frac{1}{6} + 5.5\frac{1}{6} + 5.5\frac{1}{6} + 7\frac{1}{6} + 8.5\frac{1}{6} \\
&= \frac{1}{6}\{2.5 + 4.0 + 5.5 + 5.5 + 7.0 + 8.5\} = \frac{11}{2}
\end{aligned}
$$

Por otro lado :

$$
\begin{aligned}
\mu &= \sum_{i=1}^{4} y_i P[Y = y_i] = \frac{1}{4}[1 + 4 + 7 + 10] = \frac{1}{4} * 22 = \frac{11}{2} \\
E[\bar{Y}] &= \mu
\end{aligned}
$$

# Estimación de la Media Poblacional $\mu$ en M.A.S

Además,

$$
\begin{aligned}
\sigma^2 &= \sum_{i=1}^{4} (y_i - \mu)^2 \, P[Y = y_i] \\
&= \frac{1}{4}\left[\left(1 - \frac{11}{2}\right)^2 + \left(4 - \frac{11}{2}\right)^2 + \left(7 - \frac{11}{2}\right)^2 + \left(10 - \frac{11}{2}\right)^2\right] \\
&= \frac{1}{4}\left[\left(\frac{9}{2}\right)^2 + \left(\frac{3}{2}\right)^2 + \left(\frac{3}{2}\right)^2 + \left(\frac{9}{2}\right)^2\right] \\
&= \frac{1}{4}\frac{1}{4}[81 + 9 + 9 + 81] = \frac{180}{16} = \frac{90}{8} \\
&= \frac{45}{4}
\end{aligned}
$$

# Estimación de la Media Poblacional $\mu$ en M.A.S

Por otro lado se halla la Varianza de $\bar{Y}$:

$$\begin{aligned}
V(\bar{Y}) &= \sum_{i=1}^{6}\left(\bar{y}_i - \frac{11}{2}\right)^2 * \frac{1}{6} \\
&= \frac{1}{6}\left[\left(\frac{5}{2} - \frac{11}{2}\right)^2 + \left(4.0 - \frac{11}{2}\right)^2 + \left(5.5 - \frac{11}{2}\right)^2 + \left(5.5 - \frac{11}{2}\right)^2\right. \\
&= \left. + \left(7 - \frac{11}{2}\right)^2 + \left(8.5 - \frac{11}{2}\right)^2\right] \\
&= \frac{1}{6}\left[9 + \frac{9}{4} + \frac{9}{4} + 9\right] = \frac{1}{6}\left[18 + \frac{18}{4}\right] = 3 + \frac{3}{4} \\
&= \frac{15}{4}
\end{aligned}$$

Y:

$$\frac{\sigma^2}{n}\frac{N-n}{N-1} = \frac{45}{4} * \frac{1}{2} * \frac{4-2}{4-1} = \frac{45}{4} * \frac{1}{3} = \frac{15}{4} = V(\bar{Y}).$$

# Estimación de la Media Poblacional $\mu$ en M.A.S

Además, como $\sigma^2$, en general, es **desconocido**, se necesita un **estimador insesgado** de éste, con el fin de encontrar un estimador insesgado de $V(\bar{Y})$.

Para ello se usa el hecho que:

$$E\left[S^2\right] = \frac{N}{N-1}\sigma^2 \quad ,$$

de donde:

$$\widehat{\sigma}^{\mathbf{2}} = \frac{N-1}{N}\mathbf{S^2}.$$

# Estimación de la Media Poblacional $\mu$ en M.A.S

Por tanto:

$$\hat{V}(\bar{Y}) = \frac{\hat{\sigma}^2}{n}\frac{N-n}{N-1} = \frac{N-1}{N}S^2\frac{1}{n}\frac{N-n}{N-1},$$

$$\hat{V}(\bar{Y}) = \frac{S^2}{n}\left(\frac{N-n}{N}\right)$$

donde $\frac{N-n}{N}$ se conoce como el factor de corrección para poblaciones finitas. Note que si $N >> n$ entonces:

$$\frac{N-n}{N} -->_{N-->\infty} 1,$$

y en estos casos este factor se puede omitir.

# Estimación de la Media Poblacional $\mu$ en M.A.S

Además, si $n > 30$, el límite en el error de estimación, estimado, estará dado por:

$$B = Z_{\frac{\alpha}{2}} \sqrt{\hat{V}(\bar{Y})} = Z_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n} \frac{N-n}{N}}$$

$$\text{y si} \quad n \leq 30: \quad B = t_{\frac{\alpha}{2}, n-1} \sqrt{\hat{V}(\bar{Y})} = t_{\frac{\alpha}{2}, n-1} \sqrt{\frac{S^2}{n} \frac{N-n}{N}}$$

**I. C. para $\mu$**

Un I. C. para $\mu$ del $(1-\alpha) \times 100\%$ está dado por:

| $n > 30$ | $n \leq 30$ |
| --- | --- |
| $\bar{Y} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n} \frac{N-n}{N}}$ | $\bar{Y} \pm t_{\frac{\alpha}{2}, n-1} \sqrt{\frac{S^2}{n} \frac{N-n}{N}}$ |

# Estimación del Total Poblacional en M.A.S

## Ejemplos donde es Improtante Estimar el total Poblacional

| Área | Variable de Interés | Total Poblacional |
|---|---|---|
| Agricultura | Producción anula de café en una finca colombiana en el año 2001 | Producción total anual de café en Colombia en el año 2001 |
| Economía | Ventas mensuales en un almacén de cadena en Medellín el año 2001 | Total de ventas mensuales en almacenes de cadena en Medellín en el año 2001 |
| Transporte | Número diario de pasajeros transportados por un vehículo de servicio público en Medellín en el año 2001 | Total diario de pasajeros transportados por los vehículos de servicio público en Medellín en el año 2001 |

# Estimación del Total Poblacional en M.A.S

El total Poblacional está relacionado con $\mu$ por:

$$\tau = N\mu,$$

luego:

$$\hat{\tau} = N\hat{\mu} = N\bar{Y},$$

$$\begin{aligned} V(\hat{\tau}) &= N^2 V(\hat{\mu}) = N^2 V(\bar{Y}) = N^2 \frac{\sigma^2}{n} \frac{N-n}{N-1} \\ \hat{V}(\hat{\tau}) &= N^2 \frac{S^2}{n} \frac{N-n}{N} \end{aligned}$$

# Estimación del Total Poblacional en M.A.S

## Un I. C. para $\tau$

Un I. C. para $\tau$ del $(1-\alpha)\times 100\%$ está dado por:

| $n > 30$ | $n \leq 30$ |
|---|---|
| $\hat{\tau} \pm Z_{\frac{\alpha}{2}} \sqrt{N^2 \frac{S^2}{n} \frac{N-n}{N}}$ | $\hat{\tau} \pm t_{\frac{\alpha}{2}, n-1} \sqrt{N^2 \frac{S^2}{n} \frac{N-n}{N}}$ |

# Estimación de la Proporción y Total Poblacional

En algunos casos se busca conocer las perspectivas o preferencias que las personas tienen por un determinado candidato o por un determinado producto, ie. se desea conocer la cantidad de elementos de la población que pueden clasificarse en un grupo específico.

La anterior implica la estimación de proporción poblacional del atributo deseado o la estimación del número total de elementos en la población que poseen dicho atributo.

# Estimación de la Proporción y Total Poblacional

## Ejemplos donde es Improtante Estimar la proporción y el Total Poblacional

| Área | Variable de Interés | Total o Proporción Poblacional |
|---|---|---|
| Agricultura | Cultivo de café en una finca Colombiana en el año 2001 | Proporción (Total, de fincas cafeteras en Colombia en al año 2001) |
| Economía | Ventas anuales superiores a 100-millones de peso en un almacen de Medellín en el año 2001 | Proporción (Total) de almacenes de Medellín con ventas anuales superiores a 100-millones de peso en el año 2001 |
| Mercadeo | Aceptación de un candidato a la presidencia de Colombia para el periodo 2016-2020 | Proporción (Total) de personas a favor de un candidato a la presidencia de Colombia para el periodo 2016-2020 |

# Estimación de la Proporción y Total Poblacional

Considere:

$$Y_i = \begin{cases} 1, & \text{si la unidad } i \text{ tiene el atributo} \\ 0, & \text{si la unidad } i \text{ NO tiene el atributo} \end{cases}$$

Considere la proporción POBLACIONAL $p$: Proporción de unidades en la población bajo estudio que posee un atributo de interés.

Y sea $\{y_1, y_2, \ldots, y_n\}$ los valores obtenidos a partir de una muestra aleatoria de tamaño $n$, entonces un **estimador insesgado** para $p$ será:

$$\widehat{\mathbf{p}} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\text{Nro de unidades que tienen el atributo}}{\text{Nro de unidades en la muestra}} = \bar{y} = \frac{a}{n}.$$

Y un estimador de la varianza de este estimador está dada por:

$$\widehat{\mathbf{V}}\left(\widehat{\mathbf{p}}\right) = \frac{\widehat{\mathbf{p}}(1-\widehat{\mathbf{p}})}{n-1}\left(\frac{N-n}{N}\right)$$

**EJEMPLO:**

Considérese una población de $N = 28$ colegios de una cierta ciudad. Se desea seleccionar una m.a.s sin reemplazo de tamaño $n = 8$-colegios con el fin de estimar el tamaño promedio de los colegios $\mu$, el número total de estudiantes en la ciudad $\tau$, la proporción $P$ y el número total $A$ de colegios privados en la ciudad.

Una selección aleatoria sin reemplazo de 8 colegios llevó a la selección de los colegios identificados con los números:
26, 27, 22, 14, 02, 08, 06, y 16.

La información correspondiente a estas 8 unidades aparece en la siguiente tabla:

| Nro. de Orden | Nro. Aleatorio | Tipo de Colegio | Nro. de Estud |
|---|---|---|---|
| 1 | 26 | O | 1455 |
| 2 | 27 | 1 | 326 |
| 3 | 22 | 1 | 265 |
| 4 | 14 | O | 440 |
| 5 | 02 | 1 | 360 |
| 6 | 08 | 1 | 446 |
| 7 | 06 | 1 | 540 |
| 8 | 16 | O | 412 |

Las estimaciones con las correspondientes estimaciones de sus errores estándar son las siguientes:

**Para la media $\mu$:**

$$\hat{\mu} = \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{8}\sum_{i=1}^{8} y_i = \frac{4244}{8} = 530.50 \approx 530\text{est}/\text{Coleg}$$

La varianza muestral es:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{1}{8-1}\sum_{i=1}^{8}(y_i - 530.50)^2 = 146506.29$$

La varianza-estimada del estimador de la media poblacional, es decir, la varianza de $\overline{y}$ es:

$$var[\overline{y}] = \left(1 - \frac{n}{N}\right)\frac{S^2}{n} = \left(1 - \frac{8}{28}\right)\frac{146506.29}{8} = 13080.92$$

y el error-estándar-estimado del estimador es:

$$e.e[\overline{y}] = \sqrt{var[\overline{y}]} = \sqrt{13080.92} = 114.37, \quad \text{est/coleg}$$

$$B = L.E.E = 2 \times e.e[\overline{y}] = 2 \times 114.37 = 228.74$$

**Para el total poblacional $\tau$:**

$$\hat{\tau} = N\overline{y} = 28(530.50) = 14854 \quad \text{estudiantes}$$

La varianza-estimada del estimador del total poblacional, es decir, la varianza de $\hat{\tau} = N\overline{y}$ es:

$$var[\hat{\tau}] = var[N\overline{y}] = N^2$$

$$var[\overline{y}] = 28^2(13080.92) = 10255441.28$$

y el error-estándar-estimado del estimador es:

$$e.e[\hat{\tau}] = e.e[N\overline{y}] = Ne.e[\overline{y}] = 28(114.37) = 3202.36, \quad \text{est/Coleg}$$

$$B = L.E.E = 2 \times e.e[\hat{\tau}] = 2 \times 3202.36 = 6404.72$$

**Para la Proporción poblacional $P$:**

$$p = \frac{a}{n} = \frac{5}{8} = 0.625(62.5\%) \quad \text{Coleg. privados}$$

La varianza-estimada del estimador de la proporción poblacional de colegios privados, es decir, la varianza de $p = \frac{a}{n}$ es:

$$var[p] = \left(\frac{N-n}{N}\right)\frac{pq}{n-1} = \left(\frac{28-8}{28}\right)\frac{(0.625)(0.375)}{8-1} = 0.023916$$

y el error-estándar-estimado del estimador es:

$$e.e[p] = \sqrt{\left(\frac{N-n}{N}\right)\frac{pq}{n-1}} = \sqrt{0.023916} = 0.1546$$

$$B = L.E.E = 2 \times e.e[p] = 2 \times 0.1546 = 0.3092$$

**Para el número total de coleg. privados en la población $A$:**
La estimación puntual es:

$$\hat{A} = Np = 28(0.625) = 17.5 \approx 17, \quad \text{Coleg. privados}$$

# I.C para la Proporción y Total Poblacional

Para $n$ lo suficientemente grande, el límite en el Error de Estimación estimado estará dado por

$$\mathbf{B} = Z_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{\mathbf{p}}(1-\widehat{\mathbf{p}})}{n-1}\left(\frac{N-n}{N}\right)},$$

en caso contrario se utilizaría como cuantiles aproximados los de $t_{\alpha/2,n-1}$ dependiendo de la magnitud de $\widehat{\mathbf{p}}$ y $n$ (**Ver Tabla**) o los cuantiles exactos o aproximados calculados a partir de una distribución Hipergeométrica o Binomial, respectivamente.

**Tabla de n y p apropiados para usar aproximación a la Normal**

| p | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0.05 | $<$0.05 |
|---|---|---|---|---|---|---|---|
| n | 30 | 50 | 80 | 200 | 600 | 1400 | $>$1400 |

# I.C. para proporción pobl. $P$, con Distribución Binomial

Se resuelve el siguiente sistema de ecuaciones para $p$:

$$P[X \geq a; p] = \sum_{x=a}^{n} \binom{n}{x} p^x (1-p)^{n-x} = \frac{\alpha}{2}$$

y

$$P[X \leq a; p] = \sum_{x=0}^{a} \binom{n}{x} p^x (1-p)^{n-x} = \frac{\alpha}{2}$$

Se toman los valores de $p$-que producen probabilidades tan cercanas como sea posible a $\alpha/2$.

El valor de $p$-que resuelve la primera ecuación es el límite inferior del I.C para $P$ y el valor de $p$ que resuelve la segunda ecuación es el límite superior del I.C para $P$. (Ver Ejemplo).

## Continuación del ejemplo anterior

### Para la Proporción poblacional $P$:

1. **Usando Aproximación Normal**

$$p = \frac{a}{n} = \frac{5}{8} = 0.625(62.5\%) \quad \text{Coleg. privados.}$$

La varianza-estimada del estimador de la proporción poblacional de colegios privados, es decir, la varianza de $p = \frac{a}{n}$ es:

$$var[p] = \left(\frac{N-n}{N}\right)\frac{pq}{n-1} = \left(\frac{28-8}{28}\right)\frac{(0.625)(0.375)}{8-1} = 0.023916,$$

y el error-estándar-estimado del estimador es:

$$e.e[p] = \sqrt{\left(\frac{N-n}{N}\right)}\sqrt{\frac{pq}{n-1}} = \sqrt{\left(\frac{28-8}{28}\right)}\sqrt{\frac{(0.625)(0.375)}{8-1}} = 0.1546$$

Un inter. de confianza del 90% para la proporción poblacional $P$ es:

$$p \pm t_{\alpha/2;n-1} e.e[p], \quad \text{es decir:}$$

$$0.625 \pm (1.895)(0.1546) \quad \Longleftrightarrow \quad [0.332\ ,\ 0.918], \quad \text{\% de col. Privados.}$$

Este intervalo, aunque parece ser apropiado a primera vista, no lo es, ya que al construir el I.C para el número total de colegios privados $A$ se obtiene:

$$N[0.332\ ,\ 0.918] = [9\ ,\ 26], \quad \text{Coelgios Privados.}$$

Este resultado está dentro de los límites teóricos establecidos para $A$. Sin embargo, el límite superior (26) no es consistente puesto que en la muestra ya se encontraron 3-colegios oficiales, lo que necesariamente implica que el máximo número posible para $A$ no puede ser mayor a: 28-3=25.

La pobre estimación tanto de $P$ como de $A$ se deben a que el tamaño de la población y el tamaño de la muestra son muy pequeños, y por tanto, la aproximación normal no es válida.

Algunas veces los límites estimados para la proporción poblacional, utilizando incorrectamente la aproximación normal, pueden dar valores negativos o mayores que 1, lo cual no tiene sentido.

En estos casos se debe recurrir a la aproximación Binomial, mediante la cual se obtienen directamente los límites para $P$ y de ahí deducirse los límites para $A$, multiplicando el I.C de $P$ por $N$.

## 2 Usando Aproximación Binomial

El valor de $p$ que sirve de solución a la siguiente ecuación proporciona el límite inferior para $P$:

$$\sum_{x=5}^{8} \binom{8}{x} p^x (1-p)^{8-x} = 0.025$$

Tomando una aproximación a tres cifras-decimales, se encuentra que:

$$\sum_{x=5}^{8} \binom{8}{x} (0.2448)^x (1-0.2448)^{8-x} = 0.0249726$$

$$\sum_{x=5}^{8} \binom{8}{x} (0.2449)^x (1-0.2449)^{8-x} = 0.02501595$$

de donde se deduce que el límite inferior para $P$ es: 0.2448.

Similarmente, para el límite superior, se debe cumplir que:

$$\sum_{x=0}^{5}\binom{n}{x}p^x(1-p)^{n-x}=0.025,$$

nuevamente, tomando tres cifras decimales significativas, se encuentra que:

$$\sum_{x=0}^{5}\binom{n}{x}(0.9147)^x(1-0.9147)^{n-x}=0.02505209,$$

$$\sum_{x=0}^{5}\binom{n}{x}(0.9148)^x(1-0.9148)^{n-x}=0.02497389,$$

de donde, el límite superior para $P$ es: 0.9148.

Luego el I.C para $P$ es:

$$[0.2449\ ,\ 0.9148].$$

**Para el número total de coleg. privados en la población $A$:**

La estimación puntual es:

$$\hat{A} = Np = 28(0.625) = 17.5 \approx 17, \quad \text{Coleg. privados.}$$

Un I.C aproximado para $A$ sería:

$$N[0.2449 \ , \ 0.9148] \quad \Longleftrightarrow \quad [7 \ , \ 25].$$

Los últimos dos I.C (ie. el de $P$ y el de $A$) son aceptables y consistentes y tienen un nivel de confiabilidad aproximado de:

$$100(1-\alpha_1-\alpha_2) = 100(1-0.04983074-0.04987037)\% = 90.03\%.$$

# Usando R Aproximación Binomial

Para el caso del ejemplo de la estimación de la proporción de colegios privdos, se tiene que:

$$\mathbf{N} = 28, \quad \mathbf{n} = 8, \quad \mathbf{a} = 5$$

## Programa en R

Aproximacion con la Binomial:

```
Ls.conf.binom.prop<-function(pp,a,alfa,n)
{ pbinom(a,size=n,prob= pp)-alfa*0.5}
Li.conf.binom.prop<- function(pp,a,alf,n)
{1-pbinom(a-1, size=n,prob=pp)-alf/2}
a=5; n=8; N=28; alf=0.05
```

Se definen los intervalos de búsqueda:

```
intervalols<- seq(0.42,1.00,0.0001)
intervaloli<-seq(0,0.42,0.0001)
```

```
ys<-apply(matrix(intervalols,ncol=1),1,Ls.conf.binom.prop,a,alf,n)

yi<-apply(matrix(intervaloli,ncol=1),1,Li.conf.binom.prop,a,alf,n)

par(mfrow=c(1,2))

plot(intervalols,ys, type=‘l’,las=1)

abline(h=0, col=‘blue’)

plot(intervaloli,yi,type=‘l’,las=1)

abline(h=0, col=‘blue’)

ind1<-which((abs(ys)<0.0001))

print(matrix(c(intervalols[ind1],ys[ind1]),ncol=2)) pp Error comparado con
alfa/2=0.025
```

```
     [,1]          [,2]

[1,] 0.9147 5.208621e-05

[2,] 0.9148 -2.611441e-05
```

```
ind2<-which(abs(yi)<0.0001)
print(matrix(c(intervaloli[ind2],yi[ind2]),ncol=2))
```

pp Error comparado con alfa/2=0.025

```
     [,1]          [,2]
[1,] 0.2447 -7.067317e-05
[2,] 0.2448 -2.738980e-05
[3,] 0.2449 1.594713e-05
[4,] 0.2450 5.933766e-05
```

# I.C. para proporción pobl. $P$, con Distribución Hiperg.

Al usar la distribución Hipergeométrica, con $n$ pequeño, se halla primero el I. C. del $(1-\alpha)\times 100\%$ para $A$: Total de unidades en la población que poseen el Atributo de interés.

Denote por $a$: Número de unidades en la muestra que posee el atributo de interés. Se resuelve el siguiente sistema de ecuaciones para $A$:

$$P[X \geq a; p] = \sum_{x=a}^{n} \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}} = \frac{\alpha}{2}$$

y

$$P[X \leq a; p] = \sum_{x=0}^{a} \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}} = \frac{\alpha}{2}$$

# I.C. para proporción pobl. $P$, con Distribución Hiperg.

Se toman los valores de $A$-que producen probabilidades tan cercanas como sea posible a $\alpha/2$.

El valor de $A$, $A_i$, que resuelve la primera ecuación es el límite inferior del I.C para $A$. El valor de $A$, $A_s$, que resuelve la segunda ecuación es el límite superior del I.C para $A$.

Luego el Intervalo de Confianza para la **Proporción de interés**, del $(1-\alpha) \times 100\%$ estará dado por:

$$\left( \frac{A_i}{N}, \quad \frac{A_s}{N} \right)$$

# Usando R Aproximación Hipergeométrica

## Programa en R

```
Ls.conf.hiper.prop <-function(A,a,N,n,alfa)

{ phyper(a, A, N-A, n)-alfa*0.5}

Li.conf.hiper.prop<- function(A,a,N,n,alf)

{ 1-phyper(a-1, A, N-A, n)-alf/2}

a=5; n=8; N=28; alf=0.05
```

Se definen los intervalos de búsqueda:

```
intervalols<- seq(10,28,1)

intervaloli<-seq(2,20,1)
```

```
ys<-apply(matrix(intervalols,ncol=1),1,Ls.conf.hiper.prop,a,N,n,alf)
yi<-apply(matrix(intervaloli,ncol=1),1,Li.conf.hiper.prop,a,N,n,alf)
par(mfrow=c(1,2))
plot(intervalols,ys, type=‘l’,las=1)
abline(h=0, col=‘blue’)
plot(intervaloli,yi,type=‘l’,las=1)
abline(h=0, col=‘blue’)
ind1<-which((abs(ys)<0.01))
print(matrix(c(intervalols[ind1],ys[ind1]),ncol=2))
```

Asup Error comparado con alfa/2=0.025

```
[1,] 25 -0.007905983
```

ind2<-which(abs(yi)<0.01)

print(matrix(c(intervaloli[ind2],yi[ind2]),ncol=2))

Ainf Error comparado con alfa/2=0.025

[1,] 8 -0.002696699

IC para Número total de colegios privados a un nivel del 95%

(8, 25)

IC para el porcentaje de coelgios privados a un nivel del 95%

propi<-8/N; props<-25/N

[propi,props]=[0.2857143, 0.8928571]

# Estimación de la Proporción Poblacional en MAS

## CASO DE ESTUDIO

Retomemos el caso de Estudio:
Recuerde la Aplicación 1: Estimar la Proporción de estudiantes de recién ingreso de la Facultad de Minas que utilizan bicicleta para llegar a la Universidad.

Suponga que un investigador encuesta aleatoriamente 25 estudiantes de la población OBJETIVO, que tiene un total de 739 estudiantes.

Adicional a la pregunta de interés se registra el valor de la matrícula (VM: en miles de pesos).

# Estimación de la Proporción Poblacional en MAS

Los datos se muestran a continuación:

| Est. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **VM** | 834 | 576 | 930 | 282 | 258 | 1005 | 930 | 870 | 663 | 504 | 564 | 636 | 276 |
| **Cumple** | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

| Est. | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **VM** | 168 | 426 | 111 | 558 | 663 | 657 | 915 | 1815 | 1230 | 2580 | 1278 | 2700 |
| **Cumple** | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |

# Estimación de la Proporción Poblacional en MAS

Sea $Y_i$: Valor de la matrícula del $i$-ésimo estudiante en la muestra, $i = 1, 2, \ldots, n = 25$.

$$Z_i = \begin{cases} 1 & \text{Estudiante llega en bicicleta a la Unal} \\ 0 & \text{Estudiante NO llega en bicicleta a la Unal} \end{cases}$$

# RESPUESTA DE LAS PREGUNTAS

1. Identificar:**ELEMENTO, POBLACIÓN, UNIDAD DE MUESTREO, MARCO**.

- ELEMENTO: Cada uno de los estudiantes de recién ingreso matriculados en las carreras de la Fac. MINAS.
- POBLACIÓN: Todos los estudiantes matriculados en el primer semestre de la Facultad de MINAS.
- UNIDAD DE MUESTREO: Cada uno de los estudiantes de recién ingreso matriculados en las carreras de la Fac. MINAS.
- MARCO: Listado Completo de las asignaturas de Introducción a cada Ingeniería incluyendo en número de Est. matriculados en cada asignatura

# RESPUESTA DE LAS PREGUNTAS

2. Indicar como seleccionó la muestra el investigador para garantizar que sea una Muestra Aleatoria Simple.

El investigador debió:

- Enumerar todos los estudiantes de 1 hasta 739.
- Con la función *sample* del programa R, se generan $n = 25$ números aleatorios:
- Números aleatorios generados:

13 54 107 143 151 157 190 242 257 283 293 297 337

383 429 432 576 581 640 662 666 671 678 715 720

# RESPUESTA DE LAS PREGUNTAS

- Se agrupa la cantidad de estudiantes matriculados por cada asignatura de forma acumulada, como aparece en las tablas siguientes y se identifican los números aleatorios en los respectivos intervalos: **LISTADO DE ASIGNATURAS:**

| Asignatura | Est. Mat. | Int. Ac. | Nro Aleat. | Est. a Muest. |
|---|---|---|---|---|
| Int. Ing Civil | 90 | 1 90 | 13 54 | 2 |
| Int. Ing Admin. | 50 | 91 140 | 107 | 1 |
| Int. Ing Petróleos | 46 | 141 186 | 143 151 157 | 3 |
| Int. Ing Control | 63 | 187 249 | 190 242 | 2 |
| Int. Ing Minas | 48 | 250 297 | 257 283 293 297 | 4 |
| Int. Ing Sistemas | 77 | 298 374 | 337 | 1 |
| Int. Ing Eléctrica | 66 | 375 440 | 383 429 432 | 3 |
| Int. Ing Geológica | 63 | 441 503 | | 0 |
| Int. Ing Industrial | 46 | 504 549 | | 0 |
| Int. Ing Química | 68 | 550 617 | 576 581 | 2 |
| Int. Ing Ambiental | 50 | 618 667 | 640 662 666 | 3 |
| Int. Ing Mecánica | 72 | 668 739 | 671 678 715 720 | 4 |

La última columna de la tabla anterior proporciona el número de estudiantes a encuestar, de forma ALEATORIA, en cada asignatura.

# RESPUESTA DE LAS PREGUNTAS

3. Estime el valor de la Matrícula promedio y el Número total de estudiantes de recién ingreso de la Facultad de Minas que llegan en Bicicleta a la UNAL, con su respectivo I. C del 95%

$\mu$: Valor de la matrícula promedio de todos los estudiantes de la Facultad de Minas del primer semestre.

**Estimador de $\mu$**

**Estimador puntual de $\mu$:**

$$\hat{\mu} = \bar{y} = \frac{1}{25}\sum_{i=1}^{25} y_i = \frac{1}{25}\left(834 + 576 + \ldots + 2700\right)\times 1000 = 857,16\times 1000 = \$857.160$$

y la varianza muestral:
$s^2 = \frac{1}{n-1}\sum_{i=1}^{25}\left(y_i - \bar{y}\right)^2 = 434293,89 \times 1000^2$

# RESPUESTA DE LAS PREGUNTAS

**Límite en el error de estimación:**

$$\begin{aligned} B &= t_{\alpha/2,n-1}\sqrt{\frac{s^2}{n}\frac{N-n}{N}} \\ &= t_{\frac{0.05}{2},24}\sqrt{\frac{434293,89}{25}\frac{739-25}{739}} \times 1000 \\ &= 2,063899 * \sqrt{16784,07781} \times 1000 = \$267385,0877 \end{aligned}$$

I. C del 95% para la matrícula promedio:

$$\bar{y} \pm B \qquad \$857160 \pm \$267385,0877,$$

$$(589774.91, 1124545.088)$$

# RESPUESTA DE LAS PREGUNTAS

### INTERPRETACIÓN DEL INTERVALO

Con una confianza del 95% el valor de la matrícula promedio de los estudiantes de recién ingreso en los programas de la Facultad de Minas se encuentra entre \$ \$ 589774.91\$ y \$ \$1124545.088\$.

### Intervalo de Confianza para el Total Poblacional

Para hallar el I. C para el Total de Ingreso por matrícula de dichos estudiantes basta con multiplicar los extremos del Intervalo de confianza anterior por $N = 739$,
$(589774.91 * 739, 1124545.088 * 739)$, es decir, que con una confianza del 95% el ingreso total por matrículas para esta población se halla entre: \$435843660.2 y \$831038819.8

# RESPUESTA DE LAS PREGUNTAS

4. Estime $p$: Proporción de estudiantes de la Facultad de Minas de recién ingreso que llegan en bicicleta a la UNAL.

**Proporción**

$$\hat{p} = \frac{\text{total de estudiantes que llegan en bicicleta de la muestra}}{\text{total de estudiantes muestreados}} = \frac{7}{25} = 0.28$$

y su límite en el error de estimación:

$$\begin{aligned} B &= t_{\alpha/2, n-1}\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}\frac{N-n}{N}} = t_{\frac{0.05}{2},24}\sqrt{\frac{0.28(1-0.28)}{25-1}\frac{739-25}{739}} \\ &= 2,063899 * \sqrt{0,008115832} = 0.1859 \end{aligned}$$

# RESPUESTA DE LAS PREGUNTAS

**Interpretación**

Luego un Intervalo de confianza del 95% para la redproporción de estudiantes de primer semestre de la Facultad de Minas que llegan en bicicleta a la Universidad está dado por:

$$(\hat{p}-B, \quad \hat{p}+B) = (0.28-0.1859, 0.28+0.1859) = (0.094, 0.4659)$$

Con una confianza del 95% la redproporción de estudiantes de

primer semestre de la Facultad de Minas que llegan en bicicleta a la Universidad se encuentra entre 0.094 y 0.4659. $(9.4\%, 46.59\%)$

# RESPUESTA DE LAS PREGUNTAS

## Intervalo de Confianza para el Total

Con el I. C. anterior también se puede hallar un I. C. para $A$ : Número total de estudiantes de primer semestre que llegan a la UNAL en bicicleta.
En este caso: $A = Np$, y su estimador está dado por:

$$\hat{A} = N\hat{p} = 739 * 0.28 = 206.92 \approx 207, \quad \text{y su LEE:}$$

$$B_A = NB_p = 739 * 0.1859 = 137.40$$

y el I. C del 95% está dado por:

$$\begin{aligned}(\hat{A} - B_A, \hat{A} + B_A) &= (206.92 - 137.40, 206.92 + 137.40) \\ &= (69.51, 344.32) \approx (70, 344)\end{aligned}$$

I. C del 95%, usando la distribución Hipergeométrica, está dado por:

$$(91, 363)$$