

# Taller # 1 de Muestreo. 2021-I

## Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín

1. Enumere todas las posibles muestras aleatorias simples de tamaño  $n = 2$  que se pueden seleccionar de una población cuyos valores son: 0, 3, 5, 8, 12.
  - (a) Calcule  $\sigma^2$  de la población y  $Var[\bar{Y}]$ .
  - (b) Calcule  $S^2$  de cada muestra y verifique numéricamente que:

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

- (c) Verifique numéricamente que  $Var[\bar{Y}] = \frac{\sigma^2}{n} \frac{N-n}{N-1}$
2. Decida cual de los siguientes procedimientos de muestreo generan una MAS de 10 estudiantes a partir de un grupo de 60 estudiantes. En caso negativo justifique.
    - (a) Seleccionar los primeros diez estudiantes del listado de estudiantes inscritos en el grupo.
    - (b) Seleccionar un dígito de forma aleatoria y elegir aquellos 10 estudiantes cuyos documentos terminan en ese dígito.
    - (c) Si el grupo tiene seis filas de sillas con 10 sillas en cada fila, seleccionar una fila de forma aleatoria y muestrear todos los estudiantes de esa fila.
    - (d) Si el grupo se compone de 30 hombres y 30 mujeres, se asigna a los hombres los números del 1 al 30 y a las mujeres los números del 31 al 60. Luego, utilizar un generador de números aleatorios para seleccionar diez números del 1 al 60. Seleccionar los estudiantes asignados a esos números en la muestra.
    - (e) Si el grupo se compone de 30 hombres y 30 mujeres, asigne a los hombres los números del 1 al 30 y a las mujeres los números del 31 al 60. Luego, utilice un generador de números aleatorios para generar cinco números del 1 al 30 y otros cinco números aleatorios del 31 al 60. Seleccione los estudiantes asignados a esos números en la muestra.
    - (f) Elija de forma aleatoria una letra del alfabeto y seleccione aquellos estudiantes cuyo apellido comience por esa letra. Si ningún apellido comienza por esa letra, elija de forma aleatoria otra letra del alfabeto.
  3. Suponga que se desea conocer el porcentaje de hogares donde viven niños menores de 13 años en el barrio donde Ud. vive. Cada día de la semana, desde las nueve de la mañana hasta las cinco de la tarde, los encuestadores llaman a los hogares de la muestra. Cada vez que localizan a una persona en los hogares de la muestra, les preguntan: ¿Viven niños menores de 13 años en su hogar?. Finalmente desechan los hogares con los que no es posible ponerse en contacto.
    - (a) Cree que la estimación de hogares de su barrio donde viven niños menores de 13 años sea probablemente demasiado baja, demasiado alta o aproximadamente correcta?
    - (b) ¿Cómo ayuda este ejemplo a explicar por qué los encuestadores realizan las llamadas probablemente en horas de la noche?

4. Se desea estimar el promedio de Departamentos de Colombia que han visitado las personas que viven en Colombia. Formule una metodología apropiada para realizar dicha estimación: Indique: Elemento, población, Unidad de Muestreo, Marco Muestral, justifique el Tipo de muestreo que utilizaría para hacer este estudio.
5. Suponga que desea estimar la altura promedio de los estudiantes de los cursos de Estadística I, II y III, de su Universidad. Compare la ventaja de estos dos métodos de muestreo.
  - (a) Obtiene un listado de todos los estudiantes inscritos en los cursos de Estadística I, II y III, realiza un muestreo aleatorio simple de esos estudiantes, y a los estudiantes seleccionados les pregunta su altura.
  - (b) Obtiene, usando el SIA, número de estudiantes matriculados en cada grupo de Estadística I, II y III. Genera  $n$  números aleatorios y determina cuantos estudiantes se deben muestrear en los respectivos grupos.
6. Una m.a.s de  $n = 100$  contadores de agua es controlada dentro de una comunidad para estimar el promedio de consumo de agua diario por casa, durante un periodo seco. Si se supone que hay  $N = 10000$  casas dentro de la comunidad, indique: Elemento, población, Unidad de Muestreo, Marco Muestral, justifique el Tipo de muestreo que se utilizó para hacer este estudio. Diga como seleccionaría la MAS de los 100 contadores. Una vez realizado el estudio se halló que la media y la varianza muestral fueron  $\bar{y} = 12.5$  y  $S^2 = 1.252$ . Estime  $\mu$ , el promedio de consumo diario verdadero, y establezca el límite para el error de estimación. Además, estime el número total de galones de agua  $\tau$ , usados diariamente durante el periodo de sequía. Establezca un límite para el error de estimación. Halle los respectivos I. C. del 95% e interprételos.
7. El departamento de caza y pesca de una ciudad está interesado en la dirección de sus programas futuros de caza. Para mantener un potencial mayor de caza futura, el departamento desea determinar la proporción de cazadores que buscan cualquier tipo de ave de caza. Se obtuvo una m.a.s de  $n = 1000$  de los  $N = 99000$  cazadores con permiso. Indique: Elemento, población, Unidad de Muestreo, Marco Muestral, justifique el Tipo de muestreo que se utilizó para hacer este estudio. Diga cómo seleccionaría la MAS de los 1000 cazadores con permiso.  
 Suponga que 430 indicaron que cazaron aves. Estime la proporción  $P$  de cazadores con permiso que buscan aves de caza. Establezca un límite para el error de estimación y halle el respectivo I. C. del 90%.  
 Cuál deberá ser el tamaño de muestra necesario para estimar la proporción de cazadores de aves de caza  $P$ , con un límite para el error de estimación de magnitud  $B = 0.02$ ? y un nivel de confianza del 90%.
8. (Selección Múltiple) Se realizó una encuesta con el fin de determinar la preferencias de los adultos en los servicios de telefonía celular. Los resultados de la encuesta mostraron que el 73% de las personas pedían servicio de correo electrónico, con un margen de error de más/menos el 4%, con un nivel de confianza del 95%. ¿Qué significa la frase de más/menos el 4%?
  - (a) Que el 4% de la población encuestada puede cambiar de opinión entre el periodo desde que se realizó la encuesta y el periodo en que se publicó la misma.
  - (b) Que existe un 4% de oportunidad de que el porcentaje verdadero de adultos que desean el servicio de correo electrónico no se encuentre en el I.C del 69 al 77%.
  - (c) Que sólo se encuentra el 4% de la población.

- (d) Que sería improbable obtener la proporción de la muestra observada del 73%, a menos que el porcentaje actual de todos los adultos que deseen el servicio de correo electrónico se encuentre entre el 69 y 77%.
- (e) Que la probabilidad de que la proporción muestral se encuentre en el I.C es de 0.04.
- (f) Que se espera que la proporción de personas que piden el servicio de correo electrónico se encuentre entre 69% y el 77% con una confianza del 95%.
9. Un auditor muestrea aleatoriamente 20 cuentas pendientes de cobro de las 500 de una determinada empresa. El auditor enumera la cantidad de cada cuenta y comprueba si los documentos anexos coinciden con los procedimientos llevados a cabo. Indique: Elemento, población, Unidad de Muestreo, Marco Muestral, justifique el Tipo de muestreo que se utilizó para hacer este estudio. Diga cómo seleccionaría la MAS de las 20 cuentas pendientes de cobro.

Los datos obtenidos se muestran en la siguiente tabla. Las cantidades están en dólares. Y=SI, N=NO.

| Cuenta | Cantidad | Conformidad | Cuenta | Cantidad | Conformidad |
|--------|----------|-------------|--------|----------|-------------|
| 1      | 278      | Y           | 11     | 188      | N           |
| 2      | 192      | Y           | 12     | 212      | N           |
| 3      | 310      | Y           | 13     | 92       | Y           |
| 4      | 94       | N           | 14     | 56       | Y           |
| 5      | 86       | Y           | 15     | 142      | Y           |
| 6      | 335      | Y           | 16     | 37       | Y           |
| 7      | 310      | N           | 17     | 186      | N           |
| 8      | 290      | Y           | 18     | 221      | Y           |
| 9      | 221      | Y           | 19     | 219      | N           |
| 10     | 168      | Y           | 20     | 305      | Y           |

- (a) Estime la cantidad total de las cuentas por cobrar para las 500 cuentas de la empresa y establezca un límite para el error de estimación y el respectivo I. C. del 95%. ¿Cree que el valor promedio de las cuentas por cobrar de la empresa sobrepasa los U\$250? ¿Por qué?
- (b) Estime la proporción de las cuentas de la empresa que no cumplen los procedimientos indicados. Establezca un límite para el error de estimación. ¿Cree que la proporción de cuentas que cumplen los procedimientos indicados supere el 80%? ¿Por qué?
10. En un estudio para evaluar las actitudes de los empresarios respecto al impacto que tenían los anuncios de sus servicios se enviaron 200 cuestionarios a empresarios de un listado de 1400 empresarios. Se devolvieron un total de 82 cuestionarios correctamente diligenciados. El resumen de los datos de una pregunta se muestra en la siguiente tabla:

| probabilidad de que realice anuncios en el futuro (en %). |                      |                                       |
|---|----------------------|---------------------------------------|
|   | Todas las respuestas | Los que tenían anuncios anteriormente |
|   | ( $n = 82$ )         | ( $n = 46$ )                          |
| Seguro  | 22                   | 35                                    |
| Muy Probable  | 4                    | 5                                     |
| Algo Probable   | 19                   | 35                                    |
| Alrededor del 50-50                                       | 18                   | 15                                    |
| Algo Improbable   | 6                    | 10                                    |
| Muy improbable  | 12                   | 0                                     |
| Completamente Nada  | 15                   | 0                                     |
| Sin respuesta   | 4                    | 0                                     |
|   | 100%                 | 100%                                  |

- (a) Estime la proporción de la población que, con toda seguridad, pondrá anuncios en el futuro y reporte el respectivo I. C.
  - (b) Estime la proporción de la población que, al menos, señala que tiene un 50-50 de probabilidad de anunciarse en el futuro y reporte el respectivo I. C.
  - (c) Entre los que habían puesto anuncios con anterioridad, estime la proporción de la población que responde que es algo improbable que se vuelva a anunciar y reporte el respectivo I. C.
  - (d) Entre los que habían puesto anuncios con anterioridad, estime la proporción de la población, que al menos, señala un 50-50 de probabilidad de anunciarse en el futuro y reporte el respectivo I. C.
11. Una empresa de investigación de mercadotecnia desea estimar la proporción de posibles clientes que prefieren un labial seleccionando aleatoriamente 100 mujeres que se acercaron de forma espontánea a un punto de venta de un centro comercial. De las 100 mujeres muestreadas, 65 indicaron una preferencia por la marca A.
- (a) ¿Cómo estimaría la proporción real de mujeres que prefieren la marca A?, reporte el límite en el error de estimación? Discuta alrededor del diseño de muestreo utilizado, se puede considerar una MAS?, en caso negativo, como haría Ud. el estudio para que realmente las 100 mujeres si sean consideradas una M.A.S.?
  - (b) ¿Cuál es la población objetivo de este estudio?
  - (c) ¿La empresa de investigación de marketing seleccionó una muestra aleatoria simple?
  - (d) ¿Qué problemas adicionales ve con este tipo de muestreo?
12. El auditor de una compañía está interesado en estimar el número total de comprobantes de viaje que fueron archivados incorrectamente. En una muestra aleatoria simple de  $n = 50$  comprobantes tomados de un grupo de  $N = 250$ , 20 fueron archivados incorrectamente. Estime el número total de comprobantes archivados incorrectamente de los  $N = 250$  y establezca un límite para el error de estimación. Adicionalmente: Indique: Elemento, población, Unidad de Muestreo, Marco Muestral, justifique el Tipo de muestreo que se utilizó para hacer este estudio.
13. En un estudio sociológico, realizado en una pequeña ciudad, se hicieron llamadas telefónicas para estimar la proporción de hogares donde habita al menos una persona mayor de 65 años de edad. La ciudad tiene 621 hogares, según el directorio telefónico más reciente. Se seleccionó un m.a.s de  $n = 60$  hogares del directorio. Al terminar la investigación de campo de los 60 hogares muestreados, en 11 de ellos habitaba al menos una persona mayor de 65 años de edad.

Estime la proporción poblacional  $P$  y establezca un límite para el error de estimación y calcule el respectivo I. C.

Cuál deberá ser el tamaño de muestra necesario para estimar a  $P$ , con un límite para el error de estimación de 0.08? Suponga que la proporción verdadera es  $P \approx 0.2$ .

14. Un investigador está interesado en estimar el número total de árboles marcados (árboles más grandes que cierto tamaño específico) en una población de  $N = 1500$  acres. Esta información se utiliza para estimar el volumen total de madera aserrada para los árboles en la plantación. Se seleccionó una m.a.s de  $n = 100$  parcelas de un acre y se examinó cada parcela en relación con el número de árboles marcados. El promedio muestral para las  $n = 100$  parcelas de 1 acre fue de  $\bar{y} = 25.2$  con una varianza muestral de  $S^2 = 136$ . Estime el número total de árboles marcados en la plantación. Establezca un límite para el error de estimación y calcule el respectivo I. C. Además: Indique: Elemento, población, Unidad de Muestreo, Marco Muestral, justifique el Tipo de muestreo que se utilizó para hacer este estudio.

¿Cuál deberá ser el tamaño de muestra necesario para estimar a  $\tau$ , el número total de árboles en la plantación, con un límite para el error de estimación de magnitud  $B = 1500$ ?

15. Los guardas de un parque estaban interesados en la proporción de campistas que consideraban un determinado lugar como espacio adecuado de acampada. Decidieron realizar una m.a.s de  $n = 30$  a partir de los primeros  $N = 300$  campistas que visitaron el camping. Sea  $y_i = 0$ -si el director del  $i$ -ésimo grupo de campistas muestreado no cree que el espacio para acampar sea adecuado y  $y_i = 1$  si cree que sí lo es ( $i = 1, 2, \dots, 30$ ). Utilice los datos de la siguiente tabla para estimar a  $P$ -la proporción de campistas que tienen en cuenta el lugar adecuado de camping y calcule el respectivo I. C. e interprete.

| Campista Encuestado | Respuesta, $y_i$ |
|---------------------|------------------|
| 1                   | 1                |
| 2                   | 0                |
| 3                   | 1                |
| $\vdots$            | $\vdots$         |
| 29                  | 1                |
| 30                  | 1                |

$$\sum_{i=1}^{30} y_i = 25$$

¿Cuál deberá ser el tamaño de muestra necesario para estimar a  $P$ , con un límite para el error de estimación de magnitud  $B = 0.05$ ?

16. Un psicólogo desea estimar el tiempo de reacción promedio para un estímulo entre 200 pacientes de un hospital especializado en trastornos nerviosos. Se seleccionó una m.a.s de  $n = 20$  pacientes y se midieron sus tiempos de reacción obteniendo los siguientes resultados:

$$\bar{y} = 2.1 \text{ segundos} \quad S = 0.4 \text{ segundos.}$$

Estime la media poblacional  $\mu$  y establezca un límite para el error de estimación y el respectivo I. C.

¿Cuál deberá ser el tamaño de muestra necesario para estimar a  $\mu$ , con un límite para el error de estimación de 1 segundo? Use 1 segundo como una aproximación de la desviación estándar poblacional.

17. Seleccione una encuesta cualesquiera realizada por los medios de comunicación escritos (revistas, periódicos) e identifíquela (medio, fecha, etc.). Lea detenidamente la información y analice los resultados. ¿Qué problemas detecta en la información suministrada? ¿Falta o sobra algo en los datos proporcionados? ¿Tiene alguna sugerencia para mejorar el estudio?
18. Se desea estimar la proporción de estudiantes de la Universidad Nacional ( $N = 40000$ ) que practican regularmente un deporte (al menos una vez a la semana). ¿Cuál debería ser el número de estudiantes muestreados aleatoriamente para estar 95% confiados de que el porcentaje estimado no variará en más de dos puntos (2%) del verdadero porcentaje?
19. Considere que la siguiente información corresponde a una muestra aleatoria piloto de una comunidad de 10000 personas.

| Nro. | 1 | 2   | 3 | 4   | 5   | 6  | 7   | 8   | 9   | 10 | 11  | 12  | 13 | 14  | 15  |
|------|---|-----|---|-----|-----|----|-----|-----|-----|----|-----|-----|----|-----|-----|
| G    | f | m   | f | m   | f   | m  | f   | f   | f   | m  | f   | f   | m  | m   | m   |
| EC   | c | s   | c | s   | c   | c  | c   | o   | o   | s  | o   | s   | s  | s   | c   |
| I    | 2 | 2.5 | 4 | 3.8 | 7.2 | 10 | 5.6 | 4.9 | 3.3 | 4  | 3.5 | 5.7 | 10 | 8.1 | 4.4 |

---

| Nro. | 16  | 17  | 18 | 19 | 20  | 21  | 22  | 23  | 24  | 25   | 26  | 27  | 28  | 29 | 30 |
|------|-----|-----|----|----|-----|-----|-----|-----|-----|------|-----|-----|-----|----|----|
| G    | f   | m   | f  | f  | m   | m   | f   | m   | f   | m    | f   | m   | f   | m  | f  |
| EC   | o   | s   | o  | c  | s   | c   | s   | o   | c   | s    | o   | s   | o   | c  | s  |
| I    | 6.6 | 7.3 | 8  | 9  | 3.9 | 7.1 | 4.9 | 2.3 | 3.9 | 11.1 | 7.3 | 6.5 | 5.8 | 4  | 3  |

Estimar el ingreso promedio de la comunidad con un error absoluto de estimación no mayor a 0.2 salarios mínimos por persona y estimar el total de ingresos de la comunidad con un error absoluto de estimación no mayor a 2000 salarios mínimos en total. En ambos casos considere una confiabilidad mínima del 95% y utilice la alternativa modificada de Stein.

¿Cuál deberá ser el tamaño de muestra necesario para estimar a  $\mu$ , el número promedio de salarios por persona, con un límite para el error de estimación de magnitud 0.1 salario?

¿Cuál deberá ser el tamaño de muestra necesario para estimar a  $\tau$ , el total de ingresos percibidos por la comunidad (o el número total de salarios mínimos percibidos por la comunidad), con un límite para el error de estimación de magnitud 1000 salario?

20. Una muestra aleatoria simple sin reemplazo de 56 personas fue seleccionada de una población de 1000 trabajadores de una fábrica. Se midieron las variables Ingreso mensual (I) y Género (G) (H-Hombre, M-Mujer). La información obtenida es la siguiente:

| Nro. | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| I    | 800  | 960  | 670  | 688  | 1025 | 2346 | 780  | 1126 | 1780 | 2135 | 1446 | 2245 | 1760 | 2287 |
| G    | M    | H    | H    | M    | M    | H    | M    | H    | H    | H    | H    | H    | M    | H    |
| Nro. | 15   | 16   | 17   | 18   | 19   | 20   | 21   | 22   | 23   | 24   | 25   | 26   | 27   | 28   |
| I    | 686  | 997  | 1335 | 1567 | 1456 | 1234 | 2678 | 1456 | 1388 | 1785 | 1653 | 2121 | 880  | 984  |
| G    | H    | M    | H    | M    | M    | H    | H    | H    | M    | H    | M    | H    | M    | H    |
| Nro. | 29   | 30   | 31   | 32   | 33   | 34   | 35   | 36   | 37   | 38   | 39   | 40   | 41   | 42   |
| I    | 1256 | 946  | 2000 | 2037 | 3111 | 1042 | 1564 | 1222 | 1768 | 1984 | 2348 | 876  | 890  | 1452 |
| G    | H    | M    | M    | M    | H    | M    | H    | M    | H    | H    | H    | M    | H    | H    |
| Nro. | 43   | 44   | 45   | 46   | 47   | 48   | 49   | 50   | 51   | 52   | 53   | 54   | 55   | 56   |
| I    | 1678 | 1326 | 1843 | 880  | 760  | 1146 | 1680 | 2880 | 1890 | 1033 | 2668 | 3345 | 2156 | 1880 |
| G    | H    | M    | H    | H    | M    | M    | M    | H    | M    | H    | M    | H    | M    | M    |

- Estimar el ingreso promedio de los trabajadores. Calcule un I.C del 95% para dicho ingreso promedio.
- Estimar el ingreso total de todos los trabajadores de dicha empresa. Calcule un I.C del 95% para dicho ingreso total.
- Estimar la proporción y el número total de mujeres en la empresa.
- Calcule un I.C del 95% tanto para la proporción como para el número total de mujeres en la empresa.
- ¿Qué puedes decir acerca de la validez de la aproximación normal en este caso? Justifique su respuesta.
- ¿Cómo estimaría el ingreso promedio y el total de las mujeres para toda la empresa si no se conociera el número total de ellas?
- ¿Cuál de las dos subpoblaciones (hombres y mujeres) considera que es más homogénea con respecto a los ingresos? Justifique su respuesta.
- Estimar el ingreso promedio de cada una de las subpoblaciones (hombres y mujeres).