

Técnicas de Muestreo con SPSS Y STATA

Escuela Nacional de Estadística e Informática



Centro Andino de Formación y Capacitación en Estadística

Descuento aprobado
Inicios Próximos

Cursos de SPSS®

Básico
Intermedio
Avanzado
Análisis de Datos con SPSS
Técnicas de Muestreo con SPSS y STATA
Análisis de Muestras Complejas con SPSS
Técnicas Estadísticas Predictivas con SPSS
Estadística Descriptiva e Inferencial con SPSS



NUESTRAS MODALIDADES



Se desarrollan en las modernas instalaciones de la ENEI



Se desarrollan mediante la plataforma Zoom de manera síncrona



Los cursos se realizan de manera síncrona y asíncrona

Próximos inicios para
este año

TODO

Virtual

**Presen
cial**

*Nuestros cursos virtuales y/o
presenciales de Estadística,
Economía, Informática,
Investigación y TI !!!*



MUESTREO ALEATORIO SIMPLE

Muestreo Aleatorio Simple

- Las encuestas por muestreo tienen que ver con muestras seleccionadas de poblaciones que contienen un número finito N de unidades.
- Si todas estas unidades pueden ser distinguidas una de otra, el número de muestras distintas de tamaño n que pueden ser seleccionadas de N unidades está dado por:

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

- Por ejemplo, si la población contiene cinco unidades identificadas por A, B, C, D y E , existen 10 diferentes muestras de tamaño 3, que son:

- | | | | | |
|---------|-------|-------|-------|-------|
| ■ ABC | ABD | ABE | ACD | ACE |
| ■ ADE | BCD | BCE | BDE | CDE |

Muestreo Aleatorio Simple

- Debe notarse que la misma letra no aparece dos veces en la misma muestra.
- El orden de los elementos en la muestra no tiene importancia, las seis muestras *ABC*, *ACB*, *BAC*, *BCA*, *CAB* y *CBA* son consideradas idénticas.
- En una encuesta por muestreo, se eligen ciertas propiedades que son de interés para el investigador, las cuales intentamos medir y registrar para cada unidad que se encuentre dentro de la muestra.
- Estas propiedades de las unidades son identificadas como *características* o más simplemente, *atributos*.

Muestreo Aleatorio Simple

- Una muestra aleatoria simple es seleccionada de tal manera que cada muestra posible del mismo tamaño tiene igual probabilidad de ser seleccionada de la población.
- Cada elemento de la población tiene la misma probabilidad de ser seleccionada.
- Necesita un marco muestral de lista.
- Utiliza tabla de números aleatorios o generación de números aleatorios por computadora.
- Eficiente sólo en poblaciones homogéneas.
- Aplicable en encuestas de pequeña escala.
- Forma parte de diseños de muestra más complejos.

Variantes del MAS

- *MAS con reemplazo – MAScr*

- Una unidad de la población puede ser seleccionada más de una vez.
- Las selecciones son independientes.
- Puede haber unidades repetidas en la muestra

- *MAS sin reemplazo – MASsr*

- Una unidad de la población sólo es seleccionada una vez.
- Las selecciones no son independientes.
- No hay unidades repetidas en la muestra.

Procedimiento de Selección

- 1° Preparar el marco muestral de lista numerando las unidades desde 1 hasta N .
- 2° Elegir un número aleatorio entre 1 y N , la unidad que le corresponde a dicho número forma parte de la muestra.
- 3° Continuar la selección hasta completar el tamaño de muestra n incluyendo las unidades repetidas (si es con reemplazo) y excluyendo las unidades repetidas (si es sin reemplazo)

Muestreo Aleatorio Simple - MAS cr

- 1° La unidad U_i se selecciona en cualquier selección con probabilidad $1/N$.
- 2° Las selecciones son independientes.
- 3° El número de muestras posibles es N^n
- 4° Cada muestra posible de tamaño n es seleccionada con probabilidad $\frac{1}{N^n}$
- 5° La fracción o tasa de muestreo es $f = \frac{n}{N}$
- 6° Si t_i es la variable aleatoria definida como el número de veces que aparece la unidad U_i de la población, entonces U_i es una variable aleatoria con distribución binomial de parámetros n y $1/N$ para todo $i = 1, 2, \dots, N$

$$E(t_i) = \frac{n}{N} \quad V(t_i) = \frac{n(N-1)}{N^2} \quad COV(t_i, t_j) = -\frac{n}{N^2}$$

Probabilidades de inclusión y peso muestral

7° La probabilidad de inclusión de primer orden es: $\pi_i = P(t_i \neq 0) = 1 - \left(1 - \frac{1}{N}\right)^n$

8° La probabilidad de inclusión de segundo orden es: $\pi_{ij} = P[(t_i \neq 0) \cap (t_j \neq 0)] = 1 - 2\left(1 - \frac{1}{N}\right)^n + \left(1 - \frac{2}{N}\right)^n$

9° El peso muestral o factor de expansión es:

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

- *Caso Sífilis*

Ejemplo 1:

Se realizó un estudio descriptivo que incluyó las características clínicas de la sífilis reciente en el área de salud de los centros de salud Comunitario de la Ciudad de la Habana en el periodo de 1990 a 2000. Se encontraron 850 fichas epidemiológicas en las que se obtuvieron datos en las siguientes variables que se encuentran en el archivo sifilis.sav

Seleccionar una muestra aleatoria simple con reemplazo de 200 fichas clínicas del archivo sífilis.sav En el archivo muestra indique los pesos muestrales, probabilidad de inclusión y fracción de muestreo.

Muestreo Aleatorio Simple - MAS sr

1° La unidad U_i se selecciona en cualquier selección con probabilidad $1/N$.

2° Las selecciones no son independientes.

3° El número de muestras posibles es $C_n^N = \frac{N!}{n!(N-n)!}$

4° Cada muestra posible de tamaño n es seleccionada con probabilidad $f = \frac{n}{N}$

5° La fracción o tasa de muestreo es :

6° Si t_i es la variable aleatoria definida como el número de veces que aparece la unidad U_i de la población, entonces t_i es una variable aleatoria con distribución Bernoulli de parámetros n/N para todo $i = 1, 2, \dots, N$

$$E(t_i) = \frac{n}{N} \quad V(t_i) = \frac{n(N-n)}{N^2} \quad COV(t_i, t_j) = -\frac{n}{N^2} \left(\frac{N-n}{N-1} \right)$$

$i = 1, 2, \dots, N$

Probabilidades de inclusión y peso muestral

7° La probabilidad de inclusión de primer orden es:

$$\pi_i = P(t_i = 1) = \frac{n}{N}$$

8° La probabilidad de inclusión de segundo orden es:

$$\pi_{ij} = P[(t_i = 1) \cap (t_j = 1)] = \frac{n}{N} \left(\frac{n-1}{N-1} \right)$$

9° El peso muestral o factor de expansión es:

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

Ejemplo 2

- Seleccionar una muestra aleatoria simple sin reemplazo de 200 fichas clínicas del archivo sífilis.sav.
- En el archivo muestra indique los pesos muestrales, probabilidad de inclusión y fracción de muestreo.

Procedimiento de estimación ~~de~~ μ medias

- Un estimador general para la media poblacional μ esta dado por

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^N y_i$$

- La varianza teórica del estimador anterior es

□

MAScr

$$V(\mu) = \frac{\sigma^2}{n} = \frac{(N-1)}{nN} S^2$$

MAS sr

$$V(\mu) = \left(\frac{N-n}{N-1} \right) \frac{\sigma^2}{n} = \frac{(1-f)}{n} S^2$$

- La varianza estimada del estimador anterior es

$$\hat{v}(\mu) = \frac{s^2}{n}$$

$$\hat{v}(\mu) = \frac{(1-f)}{n} s^2 \quad s^2 = \text{varianza muestral}$$

- EL MAS sr es superior al MAS cr

Procedimiento de estimación de totales

- Un estimador general para el total poblacional T esta dado por

$$T = N\mu = \frac{N}{n} \sum_{i=1}^n y_i$$

- La varianza teórica del estimador anterior es

$$\begin{array}{ll} \text{-- MAScr} & \text{MAS sr} \\ V(\mu) = \frac{N^2 \sigma^2}{n} = \frac{N(N-1)}{n} S^2 & V(\mu) = \left(\frac{N-n}{N-1} \right) \frac{N^2 \sigma^2}{n} = \frac{N^2 (1-f)}{n} S^2 \end{array}$$

- La varianza estimada del estimador anterior es

$$\hat{v}(\mu) = \frac{N^2 s^2}{n} \qquad \hat{v}(\mu) = \frac{N^2 (1-f)}{n} s^2$$

Procedimiento de estimación de proporciones

- Un estimador general para la proporción poblacional P esta dado por

$$P = \frac{1}{n} \sum_{i=1}^n y_i = \frac{n_e}{n}$$

- La varianza teórica del estimador anterior es

– MAScr

$$V(P) = \frac{P(1-P)}{n}$$

MAS sr

$$V(P) = \left(\frac{N-n}{N-1} \right) \frac{P(1-P)}{n}$$

- La varianza estimada del estimador anterior es

$$v(P) = \frac{P(1-P)}{n-1}$$

$$v(P) = \frac{(1-f)}{n-1} P(1-P) \quad s^2 = \text{varianza muestral}$$

Ejemplo 3:

- Con la muestra aleatoria simple con reemplazo seleccionada en el ejemplo 1, determine las siguientes estimaciones con sus respectivos errores muestrales:
 - Promedio de días de curación de la enfermedad.
 - Promedio de días de curación de la enfermedad, según sexo.
 - Proporción de pacientes con diagnósticos SRAS.
 - Total de pacientes con diagnóstico SRAL
 - Total de pacientes con diagnóstico SRAL según raza.
 - Procedimiento con SPSS:
 - Analizar / Muestras complejas / Descriptivos
 - Analizar / Muestras complejas / Frecuencias

Ejemplo 4:

- Con la muestra aleatoria simple sin reemplazo seleccionada en el ejemplo 2, determine las siguientes estimaciones con sus respectivos errores muestrales:
 - Promedio de días de curación de la enfermedad.
 - Promedio de días de curación de la enfermedad, según sexo.
 - Proporción de pacientes con diagnósticos SRAS.
 - Total de pacientes con diagnóstico SRAL
 - Total de pacientes con diagnóstico SRAL según raza.
 - Procedimiento con SPSS:
 - Analizar / Muestras complejas / Descriptivos
 - Analizar / Muestras complejas / Frecuencias

Tamaño de muestra

- Tenemos tres expresiones técnicas sobre los tamaños muestrales. En este caso está el *tamaño esperado* de muestra, el *tamaño planeado* de muestra y el *tamaño efectivo*.
- El *tamaño esperado* resume solamente consideraciones estadísticas de acuerdo con las exigencias ex ante sobre precisión y confiabilidad.
- El *tamaño planeado* incluye la provisión necesaria para la *No Respuesta* lo que es de especial importancia cuando no se admiten sustituciones que siempre afectan al diseño muestral.
- Finalmente el *tamaño final* se refiere al número de entrevistas o cuestionarios que se han aceptado como válidas y que corresponden a la muestra.

- *Tamaño Esperado de Muestra (TEM):*

Es el tamaño muestral necesario para lograr estimaciones de valores poblacionales con especificaciones de confiabilidad y precisión.

- *Tamaño Planeado de Muestra (TPM)*

Es el tamaño muestral necesario para satisfacer el tamaño esperado de muestra considerando la *No Respuesta*.

$$TPM = TEM / \text{Tasa anticipada de respuesta}$$

- *Tamaño Final de Muestra (TFM)*

El número de unidades muestrales que finalmente son observadas (o entrevistadas) de acuerdo con las reglas y procedimientos de campo definidos para la encuesta.

Tamaño Esperado de Muestra

- Bajo la teoría de normalidad asintótica, para muestras grandes:

$$P(|\theta - \theta| \leq E) = \gamma \qquad E = Z\sqrt{V(\theta)}$$

- Las fórmulas de tamaño esperado de muestra en MASr son:

- $$TEM = \frac{N\sigma^2}{(N-1)D + \sigma^2} \longrightarrow \text{Para estimar medias y totales}$$

$$TEM = \frac{NP(1-P)}{(N-1)D + P(1-P)} \longrightarrow$$

- $$D = \left(\frac{E}{Z}\right)^2 \quad \text{Para estimar proporciones}$$

Tamaño Planeado de Muestra:

- La fórmula de tamaño planeado de muestra que se genera es:

$$TPM = \frac{TEM}{TAR}$$

- *TAR= Tasa anticipado de respuesta*
- Por ejemplo, si se espera que un 10% de los entrevistados no contestarán la encuesta (*no respuesta*) entonces $TAR = 0,9$, entonces:

$$TPM = \frac{TEM}{0,9}$$

Ejemplo 5:

- Se desea determinar el tamaño de una muestra de fichas clínicas de pacientes con sífilis reciente en los centros de Salud Comunitario de la Ciudad de la Habana utilizando MASsr. El propósito principal del muestreo es estimar el promedio de días de curación de la enfermedad, con un error de no más de 4 días y una confianza del 95%.

Ejemplo 6:

- Se desea determinar el tamaño de una muestra de fichas clínicas de pacientes con sífilis reciente en los centros de Salud Comunitario de la Ciudad de la Habana utilizando MASsr. El propósito principal del muestreo es estimar el porcentaje de pacientes con diagnóstico de sífilis SRAS, con un error absoluto de no más del 2% y una confianza del 95%.

Tamaño de muestra en mediciones cualitativas:

Margen de Error (%)	P (%)							
	1	5	10	20	25	35	45	50
0,1	39 600	190 000	360 000	640 000	750 000	910 000	990 000	1 000 000
0,5	1 584	7 600	14 400	25 600	30 000	36 400	39 600	40 000
1,0	396	1 900	3 600	6 400	7 500	9 100	9 900	10 000
2,0	99	475	900	1 600	1,875	2 275	2 475	2 500
3,0	44	211	400	711	833	1 011	1 100	1 100
4,0	25	119	225	400	469	569	619	625
5,0	16	76	144	256	300	364	396	400
10,0	4	19	36	64	75	91	99	100



IMPROVEMENT

Al terminar todo curso en la Escuela del INEI recibirás un correo con los datos de acceso para poder descargar tu certificado Digital a nombre del INEI.

ESCUELA INEI
Certificados
Digitales



Descarga el certificado Digital del curso



INSTITUTO
NACIONAL DE
ESTADÍSTICA E
INFORMÁTICA

ESCUELA NACIONAL DE ESTADÍSTICA E INFORMÁTICA



ESCUELA
NACIONAL DE
ESTADÍSTICA E
INFORMÁTICA

CERTIFICADOS Y CONSTANCIAS DIGITALES

Ingresar todos los datos solicitados

Datos de Identidad

TIPO DE DOCUMENTO DE IDENTIDAD: (*)

Seleccione

Nº DE DOCUMENTO DE IDENTIDAD: (*)

SELECCIONE UN TIPO DE DOCUMENTO

(*) Ingrese el código de alumno a validar y seleccione el documento a descargar

CÓDIGO DE ALUMNO: (*)

INGRESE CÓDIGO ACCESO

DOCUMENTO A DESCARGAR: (*)

Seleccione

☐ He leído y acepto los términos y condiciones de uso

Por favor, leer el siguiente aviso: El presente certificado es un documento digital emitido por el INEI, el cual no tiene validez jurídica si no es emitido por el INEI. El código de acceso también está en el documento impreso.

<https://sistemas.inei.gob.pe/WebCerEnei/>

Contactos de la Escuela del INEI



Nuestros teléfonos

☎ 433-3127

☎ 997-567-428

☎ 991-686-020

Nuestros correos

✉ cursos@inei.gob.pe

✉ enei@inei.gob.pe

Horario de atención: Lunes a Viernes de 9 a.m. a 6 p.m.

