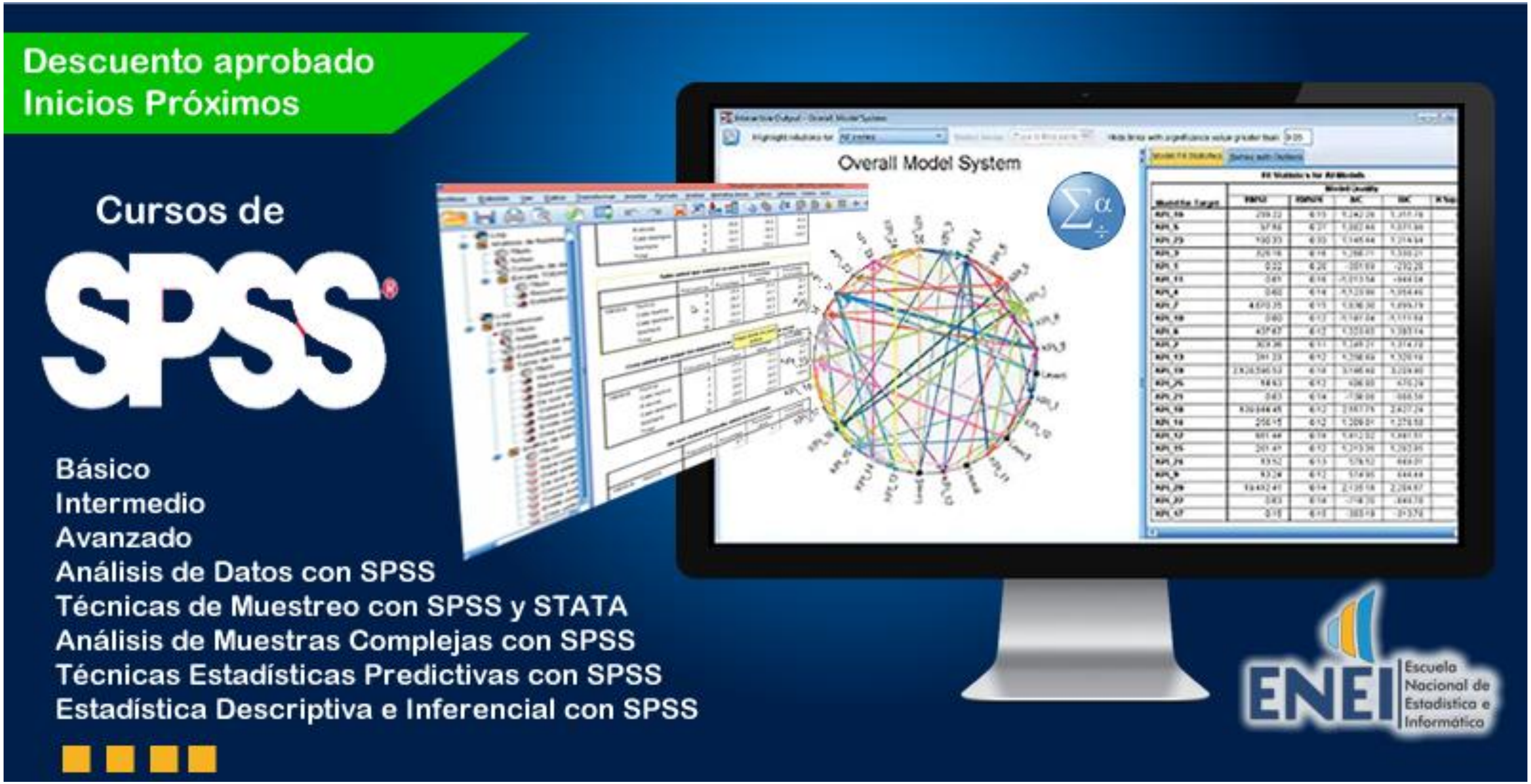
Técnicas de Muestreo con SPSS Y STATA

SPSS AN IBM COMPANY | stata | R | Microsoft | INEI Instituto Nacional de Estadística e Informática | C@mpus Virtual Tecnología Informática aplicada a la Educación

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
Overall Model System
ENEI
Escuela Nacional de Estadística e Informática

NUESTRAS MODALIDADES

Presencial

Se desarrollan en las modernas instalaciones de la ENEI

Online

Se desarrollan mediante la plataforma Zoom de manera síncrona

E-Learning

Los cursos se realizan de manera síncrona y asíncrona


Próximos inicios para este año
INEI
TODO
Virtual
Presen
cial
Nuestros cursos virtuales y/o presenciales de Estadística, Economía, Informática, Investigación y TI !!!
ENEI
Escuela Nacional de Estadística e Informática

MUESTREO ALEATORIO ESTRATIFICADO

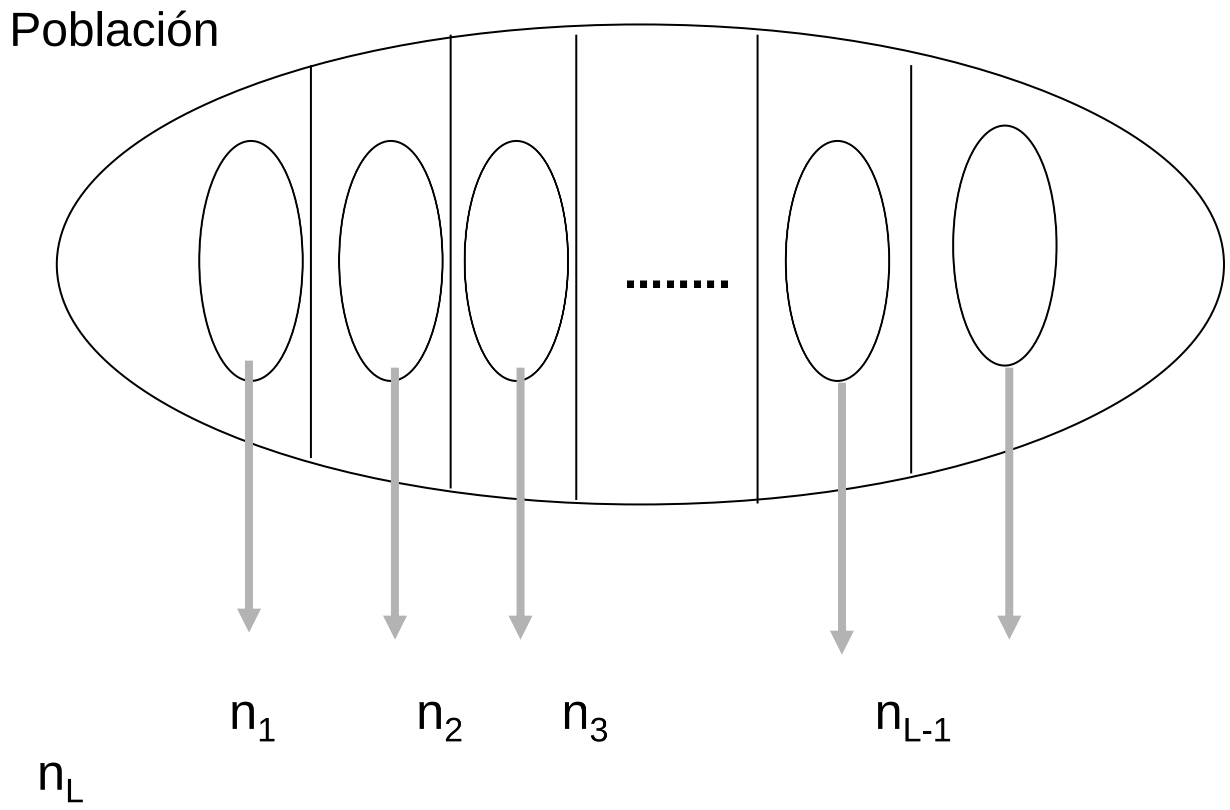
Muestreo Aleatorio Estratificado

- En muchas ocasiones es conveniente dividir a la población en grupos o estratos para mejorar la eficiencia del muestreo o bien obtener resultados desagregados por dominios de estudio.
- La población de estudio, formada por *N* unidades, se divide en *L* estratos, los cuales constituyen una población, es decir, no se solapan y la unión de todos ellos es el total.
- La muestra estratificada se obtiene seleccionando n_h unidades de cada uno de los *L* estratos de forma independiente en cada estrato.
- Los estratos, para mejorar la eficiencia del diseño, se forman en función de variables altamente correlacionadas con las variables en estudio, tales como nivel socioeconómico, tamaño de la localidad, giro de empresas, etc.

Muestreo Aleatorio Estratificado

- Si la selección en cada estrato es aleatoria simple, el muestreo se denomina ***Muestreo Aleatorio Estratificado (MAE)***.
- Su principal objetivo es mejorar la precisión de las estimaciones reduciendo los errores de muestreo. Minimiza la varianza de los estimadores mediante la creación de estratos lo más homogéneos posible entre sus elementos y lo más heterogéneo entre estratos.
- Es eficiente en poblaciones heterogéneas.
- Reduce el costo del muestreo al reducir los tamaños de muestra sin perder precisión.
- Forma parte de los diseños de muestras complejas.
- Administrativamente el muestreo estratificado facilita la designación de supervisiones y equipos de campo que controlen y ejecuten la encuesta de cada región o estrato.

Muestreo Aleatorio Estratificado



$$n = n_1 + n_2 + n_3 + \ldots\ldots\ldots + n_L$$

Muestreo Aleatorio Estratificado

Estrato	Valores de la población			
1	Y_{11}	Y_{12}	....	Y_{1N1}
2	Y_{21}	Y_{22}	....	Y_{2N2}
......	.....	.....	....	......
L	Y_{L1}	Y_{L2}	....	Y_{LNL}

$$N = \sum_{h=1}^{L} N_h$$

Estrato	Valores de la muestra			
1	y_{11}	y_{12}	....	y_{1n1}
2	y_{21}	y_{22}	....	y_{2n2}
......	.....	.....	....	......
L	y_{L1}	y_{L2}	....	y_{LnL}

$$n = \sum_{h=1}^{L} n_h$$

Supuestos del muestreo estratificado

- ***HOMOGENEIDAD:***

  Entre elementos de un mismo estrato

- ***HETEROGENEIDAD:***

  Entre estratos

- ***INDEPENDENCIA:***

  Entre estratos al seleccionar la muestra

El número de estratos:

- No se debe pensar que aumentando notablemente el número de estratos se obtienen altos beneficios. En la práctica el aumento mas allá de 6 estratos produce pocas ganancias en la reducción de las varianzas.
- Definido por criterio del investigador
- Calculado por fórmula teórica:

$$L = \frac{2nC_n}{C_e}$$

donde,

C_n : costo por unidad de muestra

C_e : costo por estratificación

Formación de los estratos

- ***Método de Dalenius***

  Cuando se tiene una variable cuantitativa de estratificación, se puede determinar los límites de los intervalos para cada estrato.

- ***Método Cluster***

  Cuando se tienen varias variables de estratificación, se puede formar estratos de elementos similares en base a las medidas de distancia entre elementos.

- ***A criterio del investigador***

  Cuando se forman estratos por dominios geográficos, por dominios temáticos, etc, que favorecen el análisis de la muestra.

Método de Dalenius

- Dalenius (1957), diseñó un método para determinar los mejores límies para estratos cuando se dispone de datos correspondientes a una variable cuantitativa para toda la población.
- Los resultados son muy buenos cuando la variable de estratificación está altamente correlacionada con la variable de interés.
- Este método tiende a minimizar la varianza del estimador.
- Se requiere de manera preliminar contar con una gran cantidad de estratos estrechos (intervalos de clase)

Método de Dalenius

Ejemplo:

- En un estudio de múltiples propósitos se necesita seleccionar una muestra de 400 abonados de Lima Metropolitana. Una de las variables más importantes es el gasto en **tráfico telefónico.**

- Se decidió por utilizar L=5 estratos.

$$\frac{3533.11}{5} = 706.62$$

- Luego, obtenemos los límites:

706.62x2=1413.24

706.62x3=2119.86

706.62x4=2826.48

706.62x5=3533.11

Gasto en tráfico Telefónico (S/.)	Número de Abonados (f_i)	$\sqrt{f_i}$	Acum $\sqrt{f_i}$
0 – 50	120000	346.41	346.41
50 – 100	220000	469.04	815.45
100 – 150	80000	282.84	1098.29
150 – 200	50920	225.65	1323.94
200 – 250	20300	142.48	1466.42
250 – 300	10120	100.59	1567.01
300 – 350	10000	100	1667.01
350 – 400	9500	97.47	1764.48
400 – 450	9280	96.33	1860.81
450 – 500	8620	92.84	1953.65
500 – 550	7950	89.16	2042.81
550 – 600	7230	85.03	2127.84
600 – 650	7020	83.78	2211.62
650 – 700	6320	79.49	2291.11
700 – 750	6000	77.45	2368.56
750 – 800	5980	77.33	2445.89
800 – 850	5000	70.71	2516.60
850 – 900	4880	69.86	2586.46
900 – 950	4300	65.57	2652.03
950 – 1000	3900	62.45	2714.48
1000 – 1050	3560	59.67	2774.15
1050 – 1100	3110	55.77	2829.92
1100 – 1150	2910	53.94	2883.86
1150 – 1200	2490	49.89	2933.75
1200 – 1250	2100	45.83	2979.58
1250 – 1300	2000	44.72	3024.30
1300 – 1350	1980	44.49	3068.79
1350 –1400	1830	42.78	3111.57
1400 – 1450	1800	42.43	3154.00
1450 – 1500	1790	42.31	3196.31
1500 – 1550	1750	41.83	3238.14
1550 – 1600	1710	41.35	3279.49
1600 – 1650	1630	40.37	3319.86
1650 – 1700	1500	38.73	3358.59
1700 – 1750	1320	36.33	3394.92
1750 – 1800	1280	35.78	3430.70
1800 – 1850	1000	31.62	3462.32
1850 – 1900	820	28.64	3490.96
1900 – 1950	410	20.25	3511.21
1950 – 2000	220	14.83	3526.04
2000 a más	50	7.07	3533.11
TOTAL	632580		

Procedimiento de Selección

1° Preparar el marco muestral tal que contenga la variable que identifica el estrato al que pertenece cada unidad del marco

2° Seleccionar la muestra aleatoria (simple con o sin reemplazo, sistemática, etc) de forma independiente en cada estrato

3° La muestra estratificada es la unión de todas las muestras obtenidas de cada estrato

Muestreo Aleatorio Estratificado

MASsr en cada estrato

- En el estrato h, las selecciones no son independientes
- El número de muestras posibles en el estrato *h* es: $C_{n_h}^{N_h} = \dfrac{N_h!}{n_h!(N_h - n_h)!}$
- Cada muestra posible de tamaño n_h es seleccionada con probabilidad $1 / C_{n_h}^{N_h}$
- La fracción o tasa de muestreo en cada estrato es: $f_h = \dfrac{n_h}{N_h}$
- La probabilidad de inclusión de primer orden en cada estrato es: $\pi_{hi} = \dfrac{n_h}{N_h}$
- El peso muestral o factor de expansión en cada estrato es: $\omega_{hi} = \dfrac{N_h}{n_h}$

Procedimiento de Estimación

- El estimador del total poblacional esta dado por:

$$\hat{T} = \sum_{h=1}^{L} \hat{T}_h$$

- Debido a la independencia, la varianza es:

$$V(\hat{T}) = \sum_{h=1}^{L} V(\hat{T}_h)$$

- El estimador de la varianza es:

$$\hat{V}(\hat{T}) = \sum_{h=1}^{L} \hat{V}(\hat{T}_h)$$

Procedimiento de estimación de medias

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

$$\hat{\mu} = \frac{\hat{T}}{N} = \frac{1}{N}\sum_{h=1}^{L}\hat{T} = \frac{1}{N}\sum_{h=1}^{L}N_h\hat{\mu}_h = \sum_{h=1}^{L}W_h\hat{\mu}_h$$

- La varianza teórica del estimador anterior es:

$$V(\hat{\mu}) = \sum_{h=1}^{L}W^2{}_hV(\hat{\mu}_h)$$

- La varianza estimada del estimador anterior es:

$$\hat{V}(\hat{\mu}) = \sum_{h=1}^{L}W^2{}_h\hat{V}(\hat{\mu}_h)$$

Procedimiento de estimación de proporciones

- Un estimador general para la proporción poblacional P está dado por:

$$\hat{P} = \frac{1}{N}\sum_{h=1}^{L} N_h \hat{P}_h = \sum_{h=1}^{L} W_h \hat{P}_h$$

- La varianza teórica del estimador anterior es:

$$V(\hat{P}) = \sum_{h=1}^{L} W^2{}_h V(\hat{P}_h)$$

- La varianza estimada del estimador anterior es:

$$\hat{V}(\hat{P}) = \sum_{h=1}^{L} W^2{}_h \hat{V}(\hat{P}_h)$$

Tamaño de muestra

- ¿Cuál es el tamaño de muestra n? y ¿cómo afijar el tamaño de muestra a cada estrato?
- Existen muchas maneras de dividir el tamaño de muestra total entre los estratos. (afijación de la muestra).
- Cada división diferente puede originar una precisión diferente para el estimador.
- ¿Qué factores influyen en el mejor esquema de afijación?
- La variabilidad de las observaciones dentro de cada estrato.
- El número total de elementos de cada estrato.
- El costo de obtener una observación de cada estrato.

Tamaño de muestra: Afijaciones

- Afijación Óptima (costo, varianza y tamaño del estrato).
- Afijación de Neyman (varianza y tamaño del estrato).
- Afijación Proporcional (tamaño del estrato)
- Afijación uniforme (igual en cada estrato)
- Afijación Proporcional Valoral (Total X del estrato)
- Afijación óptimo relativo (proporcional al coeficiente de variación del estrato).
- Afijación desproporcional (no proporcional)

Tamaño de muestra: Afijación de Neyman

- Consiste en determinar los valores de n_h tal que para un tamaño de muestra n, la varianza del estimador sea mínima.
- También es llamada afijación de varianza mínima.
- Con *MASsr* en cada estrato:

$$V(\hat{\mu}) = \sum_{h=1}^{L} W^2{}_h V(\hat{\mu}_h) = \sum_{h=1}^{L} W^2{}_h \frac{(1-f_h)}{n_h} S^2{}_h = \sum_{h=1}^{L} \frac{W^2{}_h S^2{}_h}{n_h} - \frac{1}{N}\sum_{h=1}^{L} W_h S^2{}_h = \sum_{h=1}^{L} \frac{W^2{}_h S^2{}_h}{n_h} - \frac{1}{N}\sum_{h=1}^{L} W_h$$

- Se obtienen los valores óptimos de n_h que minimiza la función:

$$\Phi = V(\hat{\mu}) + \gamma\left[\sum_{h=1}^{L} n_h - n\right] = \sum_{h=1}^{L} \frac{W^2{}_h S^2{}_h}{n_h} - \frac{1}{N}\sum_{h=1}^{L} W_h S^2{}_h + \gamma\left[\sum_{h=1}^{L} n_h - n\right]$$

Tamaño de muestra . Afijación de Neyman:

- El valor óptimo de n_h resulta:

$$n_h = n\left[\frac{W_h S_h}{\sum W_h S_h}\right]$$

- Con este n_h se obtiene la varianza mínima del estimador:

$$V_{Min} = \frac{1}{n}\left[\sum_{h=1}^{L} W_h S_h\right]^2 - \frac{1}{N}\sum_{h=1}^{L} W_h \boldsymbol{S^2}_{\boldsymbol{h}}$$

- El tamaño de muestra n para un margen de error E con un nivel de confianza es:

$$n = \frac{\left(\sum W_h S_h\right)^2}{D + \frac{1}{N}\sum W_h \boldsymbol{S^2}_{\boldsymbol{h}}}$$

Tamaño de muestra. Afijación de Neyman

- Cuando se estiman proporciones o prevalencias se utilizan las fórmulas:

$$\boldsymbol{n_h = n\left[\frac{W_h\sqrt{P_h(1-P_h)}}{\sum W_h\sqrt{P_h(1-P_h)}}\right]}$$

$$\boldsymbol{n = \frac{\left(\sum W_h\sqrt{P_h(1-P_h)}\right)^2}{D + \frac{1}{N}\sum W_h P_h(1-P_h)}}$$

Tamaño de muestra. Afijación proporcional

- Consiste en repartir el tamaño de muestra n en forma proporcional al tamaño de los estratos de la población. Es decir:

$$\boldsymbol{n_h = n\left(\frac{N_h}{N}\right) = nW_h}$$

- Con ***MASsr*** en cada estrato, el n_h proporcional genera una varianza del estimador dado por:

$$\boldsymbol{V_{Prop} = \left(\frac{1}{n} - \frac{1}{N}\right)\left(\sum_{h=1}^{L} W_h {S^2}_h\right)^2}$$

Tamaño de muestra. Afijación proporcional

- El tamaño de muestra n para un margen de error E con un nivel de confianza es:

$$n = \frac{\sum \boldsymbol{W_h S^2{}_h}}{D + \frac{1}{N}\sum \boldsymbol{W_h S^2{}_h}}$$

- La afijación de Neyman es similar a la afijación proporcional, cuando las varianzas en los estratos son iguales. Por lo tanto, la afijación proporcional es conveniente cuando las varianzas son casi iguales en todos los estratos.

Tamaño de muestra. Afijación proporcional

- Cuando se estiman proporciones o prevalencias se utilizan las fórmulas:

$$\boldsymbol{n_h = n\left(\frac{N_h}{N}\right) = nW_h}$$

$$\boldsymbol{n = \frac{\sum W_h P_h(1 - P_h)}{D + \frac{1}{N}\sum W_h P_h(1 - P_h)}}$$

Descarga el certificado Digital del curso

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

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

Contactos de la Escuela del INEI

INEI INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA

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.**

ENEI ESCUELA NACIONAL DE ESTADÍSTICA E INFORMÁTICA