Técnicas de Investigación Social

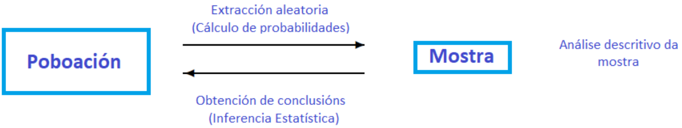
Grado en Relacións Laborais e Recursos Humanos

Curso 2019-2020

Tema 3. Mostraxe aleatoria estratificada

Alejandro Saavedra Nieves

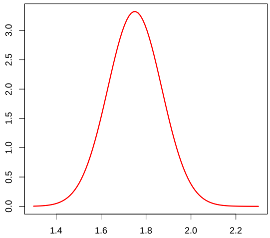
Variable: Horas empregadas en redes sociais (en media) polo estudantes de RRLL da UVigo



Estudantes Facultade RRLL

Modelo | Distribución normal

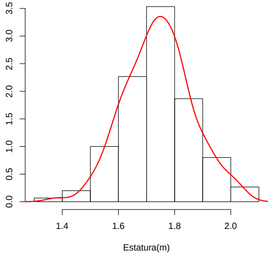
Media: $\mu = 1{,}75$; varianza: $\sigma^2 = 0{,}12^2$



Mostra estudantes RRLL, $n = 150$

$x_1 = 1.66$, $x_2 = 1.86$, $x_3 = 1.62$, ...

$\overline{X} = 1{,}74$, $s^2 = 0{,}119^2$



Algunhas cuestións naturais...

- Resulta de interés coñecer unha estimación do emprego medio diario dos alumnos de RRLL.
- En termos de productividade, poderiamos plantexarnos a necesidade de estudar o total de horas diarias que os estudantes de RRLL gastan nas redes sociais.
- Existe un problema de adicións ás redes sociais? Cal é a proporción de estudiantes que empregan máis de 2 horas diarias na súa xestión?

E se a poboacións se divide en subpoboacións heteroxéneas entre si?

Introdución

No que segue supoñemos a existencia dunha poboación finita de N individuos, sobre a que pode resultar de interese unha certa característica:

- a media dos seus valores,
- o total, en termos da suma dos valores,
- a proporción de individuos con certos valores da característica.

As metodoloxías de mostraxe permiten a estimación destes parámetros poboacionais, analizando o problema desde unha perspectiva estatística.

Notación.

- Os valores que esta variable toma para cada un dos individuos da poboación denotaranse por $X_1, \cdots, X_N$.
- Desa poboación extráese unha mostra con n observacións, $x_1, x_2, \ldots, x_n$.

Mostraxe estratificada

Mostraxe estratificada

Denomínase mostraxe estratificada a un método de selección dunha mostra de tamaño n, ó supoñer que:

- a poboación está dividida en grupos (ou estratos) homoxéneos e
- de cada un deles, extráese un subconxunto de observacións.

No caso no que a poboación sexa moi heteroxénea, con diferencias acusadas ente os individuos, úsase mostraxe estratificada co fin de aumentar a precisión con respecto á mostraxe aleatoria simple.

Por que usar mostraxe estratificada?

- A mostraxe estratificada pode aportar información máis precisa dalgunhas subpoblacións que varían bastante en tamaño e propiedades entre si, pero que son homoxéneas dentro de si.
- Os estratos deberían, no posible, estar constituido por unidades homoxéneas.
- O uso axeitado da mostraxe estratificada pode xerar ganancias na precisión, pois ó dividir unha poboación heteroxénea en estratos homoxéneos
- A mostraxe dentro dos estratos ten pouco erro debido precisamente á homoxeneidade.

Exemplo. Motivacións de tipo xeográfico xa que se requiren estimacións para certas áreas ou rexións.

Mostraxe estratificada: os fumadores en México

Plantexamos o seguinte exemplo como motivación da mostraxe estratificada:

Se queremos estudar o tanto por cento da poboación que fuma en México, a idade pode ser un bo criterio para estratificar (é dicir, pensamos que existen diferencias importantes no hábito de fumar dependendo da idade). Polo tanto, podemos definir 3 estratos: menores de 20 anos, de 20 a 44 anos e maiores de 44 años. É de esperar que ó dividir toda a poboación mexicana nestos 3 estratos non resulten grupos de igual tamaño.

Efectivamente, se miramos datos oficiais, obtemos:

- Estrato 1. Poboación Mexicana menor de 19 anos: 42,4 millones (41,0 %).
- Estrato 2. Población Mexicana de 20 a 44 años: 37,6 millones (36,3 %).
- Estrato 3. Población Mexicana maior de 44 años: 23,5 millones (22,7 %)

Mostraxe estratificada

Hipótese. Supoñemos a estratificación da poboación, é dicir, a subdivisión da poboación de N unidades en L subpoboacións, de tamaños $N_1, N_2, \ldots, N_L$, que non se superpoñen e que xuntas forman a totalidade da poboación.

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

Mostraxe estratificada

- Cada unha das subpoboacións nas que queda dividida a poboación recibe o nome de estratos.
- Para realizar a mostraxe, extráese unha mostra de cada un dos estratos, tomando como mostra final o conxunto destas submostras.
- Cada operación de mostraxe en cada estrato debe realizarse de maneira independente, podendo aplicar metodoloxías diferentes de mostraxe en cada un deles.
- Se se realiza mostraxe aleatoria simple en cada estrato, o procedemento coñécese como mostraxe estratificada aleatoria.

Proceso para realizar Mostraxe Estratificado

De maneira esquemática, os pasos a seguir son os seguintes:

1. Partimos dunha poboación con N individuos, que se asume dividida en L subgrupos disxuntos de tamaño $N_1, N_2, \ldots, N_L$.
2. Obtemos en cada estrato h unha mostra de tamaño n_h, de tal maneira que $\sum_{h=1}^{L} n_h = n$, sendo n o tamaño da mostra.
3. A mostra final da poboación será a formada por tódalas submostras obtidas en cada subpoboación.

Notación específica para a mostraxe estratificada

Notación. Consideremos o estrato h.

- N_h, número total de unidades no estrato h.
- n_h, número de unidades da mostra no estrato h.
- x_{ih}, valor obtido na unidade i do estrato h.
- $W_h = \frac{N_h}{N}$, ponderación do estrato h na poboación.
- $f_h = \frac{n_h}{N_h}$, fracción de mostraxe no estrato h.
- $\bar{X}_h$, media poboacional do estrato h.
- $\bar{x}_h$, media da mostra do estrato h.
- S_h^2, cuasivarianza poboacional do estrato h.

Estimación de θ

Supoñamos unha poboación $X_1, X_2, \ldots, X_N$ composta por N individuos, da que se extrae unha mostra aleatoria simple $x_1, x_2, \ldots, x_n$ de n observacións.

O obxectivo pasa por analizar algunha característica da poboación, θ, que é descoñecida e da que resulta de interese a súa estimación.

- Usaremos unha función obtida da mostra, o estimador $\hat{\theta}$, como aproximación de θ.
- Como é construido a partir da mostra, terá unha certa distribución.
- Debe ser fácil obter $E(\hat{\theta})$ e $V(\hat{\theta})$, que será aproximado por $\hat{V}(\hat{\theta})$.

Estimación de θ

Desde o punto de vista estatítico, cómpre controlar o erro na estimación, isto é,

$$P(|\theta - \hat{\theta}| < \varepsilon) \geq 1 - \alpha,$$

para:

- unha certa cota do erro cometido, ε, e
- un nivel de confianza (probabilidade) $1 - \alpha$.

Nesta dupla xogará un papel fundamental o tamaño da mostra seleccionado.

Equivalentemente, este problema tradúcese en determinar valores L_1 e L_2 (dependentes da mostra) tal que se satisfagan

$$P(\theta \in [L_1, L_2]) \geq 1 - \alpha.$$

Intervalos de confianza!

- **Teorema Central do Límite.** IC para $\hat{\theta}$ con un nivel de confianza $1-\alpha$ (con $n > 30$):

$$IC_\theta = \left(\hat{\theta} - z_{\alpha/2}\sqrt{V(\hat{\theta})}, \hat{\theta} + z_{\alpha/2}\sqrt{V(\hat{\theta})}\right)$$

sendo z_α o punto que deixa unha probabilidade á súa dereita igual a α.

Si $Z \sim N(0,1)$, $P(Z \geq z_\alpha) = \alpha$. Entonces, $z_{0,1} = 1{,}281$, $z_{0,05} = 1{,}649$ y $z_{0,025} = 1{,}960$.

Polo xeral, $V(\hat{\theta})$ é descoñecido.

- **Teorema Central do Límite.** IC para $\hat{\theta}$ con un nivel de confianza $1-\alpha$ (con $n > 30$):

$$IC_\theta = \left(\hat{\theta} - z_{\alpha/2}\sqrt{\hat{V}(\hat{\theta})}, \hat{\theta} + z_{\alpha/2}\sqrt{\hat{V}(\hat{\theta})}\right)$$

sendo z_α o punto que deixa unha probabilidade á súa dereita igual a α.

Si $Z \sim N(0,1)$, $P(Z \geq z_\alpha) = \alpha$. Entonces, $z_{0,1} = 1{,}281$, $z_{0,05} = 1{,}649$ y $z_{0,025} = 1{,}960$.

A estimación da media baixo mostraxe estratificada

A estimación da media poboacional

Neste caso o parámetro descoñecido é a media poboacional:

$$\bar{X} = \sum_{h=1}^{L} W_h \bar{X}_h,$$

con $\bar{X}_h$ a media poboacional do estrato h. O estimador máis utilizado na estimación da media poboacional é:

$$\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h,$$

con x_h a estimación da media da mostra do estrato h.

- Trátase dun estimador insesgado da media poboacional na mostraxe estratificada.

A varianza do estimador da media

Tras certas operacións, a varianza do estimador para a media baixo mostraxe estratificada é:

$$V(\bar{x}_{st}) = \sum_{h=1}^{L} W_h^2 (1 - f_h) \frac{S_h^2}{n_h},$$

onde

- f_h é a fracción de mostra seleccionada no estrato h, e
- S_h^2 denota a varianza poboacional do estrato h.

Dado que σ_h^2 e S_h^2 son descoñecidas, utilizamos como aproximación a cuasivarianza muestral do estrato h

$$S_{c,h}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_i - \bar{x}_h)^2.$$

Desta maneira, un estimador insesgado da varianza do estimador é

$$\hat{V}(\bar{x}_{st}) = \sum_{h=1}^{L} W_h^2 (1 - f_h) \frac{S_{c,h}^2}{n_h}$$

A estimación do total baixo mostraxe estratificada

A estimación do total poboacional

Neste caso o parámetro descoñecido é o total poboacional:

$$X = \sum_{h=1}^{L} X_h.$$

O estimador máis utilizado na estimación do total poboacional é:

$$\hat{X}_{st} = \sum_{h=1}^{L} N_h \bar{x}_h,$$

sendo $\hat{X}_h = N_h \bar{x}_h$ o estimador do total do estrato h.

- Trátase dun estimador insesgado do total poboacional.

A varianza do estimador do total

Tras certas operacións, a varianza do estimador para o total baixo mostraxe estratificada é:

$$V(\hat{X}_{st}) = \sum_{h=1}^{L} N_h^2 (1 - f_h) \frac{S_h^2}{n_h},$$

onde

- f_h é a fracción de mostra seleccionada no estrato h, e
- S_h^2 denota a varianza poboacional do estrato h.

Dado que σ_h^2 e S_h^2 son descoñecidas, utilizamos como aproximación a cuasivarianza muestral do estrato h

$$S_{c,h}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_i - \bar{x}_h)^2.$$

Desta maneira, un estimador insesgado da varianza do estimador é

$$\hat{V}(\hat{X}_{st}) = \sum_{h=1}^{L} N_h^2 (1 - f_h) \frac{S_{c,h}^2}{n_h}$$

A estimación da proporción baixo mostraxe estratificada

A estimación da proporción poboacional

Neste caso o parámetro descoñecido é a proporción poboacional:

$$P = \sum_{h=1}^{L} W_h P_h.$$

Denótase por P_h a proporción de individuos que presenta a característica no estrato h e $Q_h = 1 - P_h$, a proporción contraria.

O estimador máis utilizado na estimación do total poboacional é:

$$p_{st} = \sum_{h=1}^{L} W_h p_h,$$

con p_h a proporción estimada no estrato h.

- Trátase dun estimador insesgado da proporción poboacional.

A varianza do estimador da proporción

Tras certas operacións, a varianza do estimador para o total baixo MASsr é:

$$V(p_{st}) = \sum_{h=1}^{L} W_h^2 \frac{N_h - n_h}{N_h - 1} \frac{P_h Q_h}{n_h},$$

onde

- P_h é a proporción de individuos que presenta a característica no estrato h, e
- $Q_h = 1 - P_h$, a proporción contraria.

Dado que P_h e Q_h son descoñecidos, un estimador insesgado da varianza do estimador é

$$\hat{V}(p_{st}) = \sum_{h=1}^{L} W_h^2 (1 - f_h) \frac{p_h q_h}{n_h - 1},$$

sendo $q_h = 1 - p_h$.

Exemplo: mostraxe estratificada

As granxas dunha certa rexión divídense en catro categorías dacordo coa súa superficie. O número de granxas en cada categoría é 72, 37, 50 e 11. Un estudo para estimar o total de vacas produtoras de leite na rexión produce unha mostra estratificada de 28 granxas. O total de vacas produtoras de leite nestas 28 granxas vén dado na seguinte táboa:

Categoría	Total de vacas
Categoría I	61, 47, 44, 70, 28, 39, 51, 52, 101, 49, 54, 71
Categoría II	160, 148, 89, 139, 142, 93
Categoría III	26, 19, 21, 34, 28, 15, 20, 24
Categoría IV	17, 11

Estimar o total de vacas produtoras de leite.

Exemplo: mostraxe estratificada

Datos:

h	Estrato 1	Estrato 2	Estrato 3	Estrato 4
N_h	72	37	50	11
n_h	12	6	8	2
f_h	12/72	6/37	8/50	2/11

Resultados:

h	Estrato 1	Estrato 2	Estrato 3	Estrato 4
$\bar{x}_h$	55.58	128.5	23.375	14
$\hat{S}^2_{c,h}$	350.99	8970.1	35.411	18

Cal é a estimación do total?

A estimación dos intervalos de confianza baixo mostraxe estratificada

IC con nivel de confianza $1-\alpha$

- **Media.**

$$IC_{\bar{X}} = \left(\bar{x}_{st} - z_{\alpha/2}\sqrt{\hat{V}(\bar{x}_{st})}, \bar{x}_{st} + z_{\alpha/2}\sqrt{\hat{V}(\bar{x}_{st})}\right)$$

- **Total.**

$$IC_{X} = \left(\hat{X}_{st} - z_{\alpha/2}\sqrt{\hat{V}(\hat{X}_{st})}, \hat{X}_{st} + z_{\alpha/2}\sqrt{\hat{V}(\hat{X}_{st})}\right)$$

- **Proporción.**

$$IC_{P} = \left(p_{st} - z_{\alpha/2}\sqrt{\hat{V}(p_{st})}, p_{st} + z_{\alpha/2}\sqrt{\hat{V}(p_{st})}\right)$$

sendo z_{α} o punto que deixa unha probabilidade á súa dereita igual a α.

Mostraxe aleatoria simple: determinación do tamaño da mostra

Afixación da mostra

Supoñamos unha poboación $X_1, X_2, \ldots, X_N$ composta por N individuos.

Baixo os supostos da mostraxe estratificada, a poboación atópase dividida en en L subpoboacións, de tamaños N_1, N_2, ..., N_L, que non se superpoñen e que xuntas forman a totalidade da poboación:

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

O obxectivo consiste en extraer unha mostra aleatoria simple $x_1, x_2, \ldots, x_n$ de n observacións do conxunto da poboación.

Cantos individuos selecciono en cada estrato?

Cantos individuos selecciono en cada estrato?

Chámase afixación da mosta ó reparto ou distribución do tamaño da mostra n entre os diferentes estratos. É dicir, temos que determinar os valores n_h, con $h = 1, 2, \ldots, L$ que verifiquen

$$n_1 + n_2 + \cdots + n_L = n$$

Tipos de afixacións

Entre outras:

- **Uniforme**
- **Proporcional**
- **Varianza mínima**
- **Óptima**

Afixación uniforme

Este tipo de reparto asigna o mesmo número de unidades da mostra a cada un dos estratos.

Se n é o tamaño da mostra e L o número de estratos nos que queda dividia a poboación, tomarase de cada unha das subpoboacións h unha submostra de tamaño:

$$n_h = \frac{n}{L}.$$

O tamaño de mostra é igual para todos os estratos ($\frac{n}{L}$).

Algúns inconvenientes

- Este tipo de afixación dá a mesma importancia a tódolos estratos, en canto o tamaño da mosta, co que os estratos de menor tamaño se verán favorecidos, fronte ós grandes en canto á precisión.
- Só é conveniente en poboacións con estratos de tamaño similar.

Mostraxe estratificada: os fumadores en México

No caso de querer analizar a porcentaxe de fumadores en México, e dacordo cos datos oficiais, tiñamos:

- Estrato 1. Poboación Mexicana menor de 19 anos: 42,4 millones (41,0 %).
- Estrato 2. Población Mexicana de 20 a 44 años: 37,6 millones (36,3 %).
- Estrato 3. Población Mexicana maior de 44 años: 23,5 millones (22,7 %)

Se queremos extraer unha mostra de tamaño $n = 1000$, dado que temos L estratos, o tamaño de mostra que extraemos de cada estrato e que proporciona a afixación uniforme é

$$n_1 = n_2 = n_3 = \frac{1000}{3} \approx 333.$$

Afixación uniforme: tamaño da mostra

Fixado un certo erro $\varepsilon > 0$, os tamaños de mostra requerido son, para os parámetros baixo estudo, os seguintes:

- **Media.**

$$n = \frac{L(\sum_{h=1}^{L} W_h^2 S_h^2)}{\varepsilon^2 + \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h}}$$

- **Total.**

$$n = \frac{L(\sum_{h=1}^{L} N_h^2 S_h^2)}{\varepsilon^2 + \sum_{h=1}^{L} N_h^2 S_h^2}$$

- **Proporción.**

$$n = \frac{L(\sum_{h=1}^{L} \frac{W_h^2 N_h P_h Q_h}{N_h - 1})}{\varepsilon^2 + \sum_{h=1}^{L} \frac{W_h^2 P_h Q_h}{N_h - 1}}$$

Afixación proporcional

A afixación proporcional consiste en asignar a cada subpoboación un número de unidades da mostra proporcional ó seu tamaño. As n unidades de mostra distribúense proporcionalmente ós tamaños dos estratos expresados en números de unidades que pertencen a ese estrato.

Desa forma, se o tamaño de cada submostra é proporcional ó tamaño do estrato, entonces existe unha constante $k > 0$ tal que

$$kn_h = N_h, \text{ para } h = 1, 2, \ldots, L.$$

Desa maneira, para coñecer o tamaño global de mostra cómpre coñecer o valor desa constante k. Debemos ter en conta que

$$kn = N$$

Afixación proporcional

Na práctica, k denota a porcentaxe de individuos que extraen da poboación (a fracción de mostraxe na poboación).

- Desta forma, $f = \frac{1}{k} = n/N$ é a fracción de mostraxe para a poboación.
- Para o estrato h,

$$f_h = \frac{n_h}{N_h} = \frac{n_h}{kn_h} = \frac{1}{k} = f, \text{ para todo } h = 1, 2, \ldots, L.$$

  Isto é, a fracción de mostraxe é o valor inverso da constante de proporcionalidade.

- Ademais, a ponderación de cada estrato h é

$$W_h = \frac{N_h}{N} = \frac{n_h k}{nk} = \frac{n_h}{n}, \text{ para todo } h = 1, 2, \ldots, L.$$

Mostraxe estratificada: os fumadores en México

No caso de querer analizar a porcentaxe de fumadores en México, e dacordo cos datos oficiais, tiñamos:

- Estrato 1. Poboación Mexicana menor de 19 anos: 42,4 millones (41,0 %).
- Estrato 2. Población Mexicana de 20 a 44 años: 37,6 millones (36,3 %).
- Estrato 3. Población Mexicana maior de 44 años: 23,5 millones (22,7 %)

Se queremos extraer unha mostra de tamaño $n = 1000$, a fracción de mostraxe é $f = \frac{n}{N} = \frac{1000}{(42{,}4+37{,}6+23{,}5)\cdot 10^6} = 9{,}661 \cdot 10^{-6}$.

O tamaño de mostra que extraemos de cada estrato e que proporciona a afixación proporcional é

$$n_1 = f_1 \times N_1 = f \times N_1 = 9{,}661 \cdot 10^{-6} \times 42400000 \approx 410.$$

$$n_2 = f_2 \times N_2 = f \times N_2 = 9{,}661 \cdot 10^{-6} \times 37600000 \approx 363.$$

$$n_3 = f_3 \times N_2 = f \times N_3 = 9{,}661 \cdot 10^{-6} \times 23500000 \approx 227.$$

Afixación proporcional: tamaño da mostra

Fixado un certo erro $\varepsilon > 0$, o tamaño de mostra requerido son, para os parámetros baixo estudo, os seguintes:

- **Media.**

$$n = \frac{L(\sum_{h=1}^{L} W_h^2 S_h^2)}{\varepsilon^2 + \frac{1}{N}\sum_{h=1}^{L} W_h^2 S_h^2}$$

- **Total.**

$$n = \frac{N(\sum_{h=1}^{L} N_h S_h^2)}{\varepsilon^2 + \sum_{h=1}^{L} N_h S_h^2}$$

- **Proporción.**

$$n = \frac{\sum_{h=1}^{L} \frac{W_h N_h P_h Q_h}{N_h - 1}}{\varepsilon^2 + \frac{1}{N}\sum_{h=1}^{L} \frac{W_h N_h P_h Q_h}{N_h - 1}}$$

Exemplo: mostraxe estratificada

As granxas dunha certa rexión divídense en catro categorías dacordo coa súa superficie. O número de granxas en cada categoría é 72, 37, 50 e 11. Un estudo para estimar o total de vacas produtoras de leite na rexión produce unha mostra estratificada de 28 granxas. O total de vacas produtoras de leite nestas 28 granxas vén dado na seguinte táboa:

Categoría	Total de vacas
Categoría I	61, 47, 44, 70, 28, 39, 51, 52, 101, 49, 54, 71
Categoría II	160, 148, 89, 139, 142, 93
Categoría III	26, 19, 21, 34, 28, 15, 20, 24
Categoría IV	17, 11

Calcula o tamaño de mostra para cada un dos estratos mediante afixación uniforme e proporcional na estimación do total.

Afixación de mínima varianza

Este tipo de afixación, tamén coñecida como afixación de Neyman, consiste en determinar os valores de n_h de forma que para un tamaño de mostra fixo e igual a n, a varianza dos estimadores sexa mínima.

Coa aplicación de técnicas de optimización, o número de unidades que se deberían extraer do estrato h é:

$$n_h = n\frac{N_h S_h}{\sum_{j=1}^{L} N_j S_j}, \text{ para todo } h = 1, 2, \ldots, L,$$

e onde S_h é a cuasidesviación típica correspondente ó estrato h.

Afixación de mínima varianza

Algúns comentarios sobre a afixación de mínima varianza

- Dado que os valores de n_h son proporcionais ós produtos $N_h S_h$, se $S_h = S$ para todo estrato h, a afixación de mínima varianza coincide coa proporcional.

  Se $S_h = S$, temos que

$$n_h = n \frac{N_h S_h}{\sum_{j=1}^{L} N_j S_j} = \frac{n N_h}{N} = f N_h, \text{ con } f = \frac{n}{N}, \text{ e para todo } h = 1, 2, \ldots, L.$$

- Este tipo de afixación só mellora os resultados se hai grandes diferenzas na variabilidade dos estratos.

- No caso de variabilidade homoxénea entre estatos, a afixación proporcional é preferible.

Afixación de mínima varianza: tamaño da mostra

Fixado un certo erro $\varepsilon > 0$, o tamaño de mostra requerido son, para os parámetros baixo estudo, os seguintes:

- **Media.**

$$n = \frac{(\sum_{h=1}^{L} W_h S_h)^2}{\varepsilon^2 + \frac{1}{N}\sum_{h=1}^{L} W_h S_h^2}$$

- **Total.**

$$n = \frac{(\sum_{h=1}^{L} N_h S_h)^2}{\varepsilon^2 + \frac{1}{N}\sum_{h=1}^{L} N_h S_h^2}$$

- **Proporción.** Os tamaños de mostra para a proporción calcúlanse substituindo S_h^2 por $\frac{N_h}{N_h - 1} P_h Q_h$ sobre as fórmulas para a media.

Afixación óptima

A afixación óptima permite determinar os valores de n_h de forma que, para un coste fixo X e para un tamaño de mostra fixo e igual a n, a varianza dos estimadores sexa mínima.

O coste fixo C é a suma dos costes derivados da selección das unidades da mostra dos estratos. É dicir, se C_h é o coste de extraer unha unidade no estrato h, o coste de extraer n_h unidades é $C_h n_h$. Para os L estratos o coste total é a suma dos costes dos estratos.

Coa aplicación de técnicas de optimización, o número de unidades que se deberían extraer do estrato h é:

$$n_h = n \frac{N_h S_h / \sqrt{C_h}}{\sum_{j=1}^{L} N_j S_j / \sqrt{C_j}}, \text{ para todo } h = 1, 2, \ldots, L,$$

e onde S_h é a cuasidesviación típica correspondente ó estrato h.

Algúns comentarios sobre a afixación óptima

- Dado que os valores de n_h son proporcionais ós produtos $N_h S_h / \sqrt{C_h}$,
- Se $C_h = K$ para todo estrato h, a afixación óptima coincide coa de mínima varianza.
- Se, ademais, $S_h = S$ para todo estrato h, a afixación óptima coincide coa proporcional.

Afixación de mínima varianza: tamaño da mostra

Fixado un certo erro $\varepsilon > 0$, o tamaño de mostra requerido son, para os parámetros baixo estudo, os seguintes:

- **Media.**

$$n = \frac{(\sum_{h=1}^{L} W_h S_h/\sqrt{C_h})(\sum_{h=1}^{L} W_h S_h \sqrt{C_h})}{\varepsilon^2 + \frac{1}{N}\sum_{h=1}^{L} W_h S_h^2}$$

- **Total.**

$$n = \frac{(\sum_{h=1}^{L} N_h S_h/\sqrt{C_h})(\sum_{h=1}^{L} N_h S_h \sqrt{C_h})}{\varepsilon^2 + \frac{1}{N}\sum_{h=1}^{L} N_h S_h^2}$$

- **Proporción.** Os tamaños de mostra para a proporción calcúlanse substituindo S_h^2 por $\frac{N_h}{N_h-1} P_h Q_h$ sobre as fórmulas para a media.