



# Muestreo Aleatorio



Una vez planteado un experimento aleatorio, de esa población de infinitos resultados posibles tomo un conjunto finito de n resultados; a este conjunto finito se lo denomina **muestreo aleatorio**.

## Muestreo aleatorio Homogéneo

Posee las siguientes características:

Dado un conjunto de n resultados de una muestra (1) se tiene:

$$x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}$$

Si tenemos l muestras se tiene que:  $x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}$

$$x_1^{(j)}, x_2^{(j)}, \dots, x_n^{(j)}$$

$$x_1^{(\ell)}, x_2^{(\ell)}, \dots, x_n^{(\ell)}$$

Es decir los  $n$  resultados de una muestra cualquiera se pueden escribir:

$$\mathbf{x}^{(j)} = (x_1^{(j)}, x_2^{(j)}, \dots, x_n^{(j)})$$

La función de densidad de probabilidad del vector de muestra puede escribirse como:

$$g(\mathbf{x}) = g(x_1, x_2, \dots, x_n)$$

a) Cada elemento de la muestra es independiente:

$$g(\mathbf{x}) = g_1(x_1)g_2(x_2) \dots g_n(x_n)$$

b) La función de probabilidad de cada elemento son iguales entre si e igual a la de la población

$$g_1(x) = g_2(x) = \dots = g_n(x) = f(x)$$

# Estimadores de parámetros estadísticos

Si quiero encontrar  $\lambda$

$S = S(x_1, x_2, \dots, x_n)$  Defino como estimador del parámetro si:

a) Sin tendencia:  $E\{S(x_1, x_2, \dots, x_n)\} = \lambda$

b) Consistente:  $\lim_{n \rightarrow \infty} \sigma(S) = 0$

Estimaremos la media y varianza.....

Proponemos:

1) Estimador de la media:  $\tilde{\mu} = \frac{\sum_{i=1}^n x_i}{n}$  también llamado promedio simple o media aritmética

a) 
$$E[\tilde{\mu}] = E\left[\frac{\sum_{i=1}^n x_i}{n}\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} n \mu = \mu$$

↓

$$E[x_i] = \int_{-\infty}^{\infty} x_i g(x_i) dx_i = \int_{-\infty}^{\infty} x_i f(x_i) dx_i = \mu$$



b) Varianza del estimador de la media...

$$\begin{aligned} E[(\tilde{\mu} - \mu)^2] &= E\left[\left(\frac{\sum_{i=1}^n X_i}{n} - \mu\right)^2\right] = E\left[\left(\frac{\sum_{i=1}^n X_i}{n} - n\mu \frac{1}{n}\right)^2\right] = E\left[\left(\frac{\sum_{i=1}^n (X_i - \mu)}{n}\right)^2\right] \\ &= \frac{1}{n^2} E\left(\sum_{i=1}^n (x_i - \mu)^2 + \sum_{i=1}^n \sum_{j=1}^n (x_i - \mu)(x_j - \mu)\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n E(x_i - \mu)^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n E(x_i - \mu)(x_j - \mu) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 = \frac{1}{n} \sigma^2 \end{aligned}$$

$$E[(x_i - \mu)^2] = \int_{-\infty}^{\infty} (x_i - \mu)^2 g(x_i) dx_i = \int_{-\infty}^{\infty} (x_i - \mu)^2 f(x_i) dx_i = \sigma^2$$

$$\sigma_{\tilde{\mu}}^2 = \frac{1}{n} \sigma^2$$



Estimador de la varianza del estimador de la varianza...

Proponemos:

1) Estimador de la varianza muestral,  $\tilde{\sigma}^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \tilde{\mu})^2$

$$E[\tilde{\sigma}^2] = E\left[\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \tilde{\mu})^2\right] = \frac{1}{(n-1)} \sum_{i=1}^n E(x_i - \tilde{\mu})^2 =$$

-  $E(x_i - \tilde{\mu})^2 = E(x_i^2) + E(\tilde{\mu}^2) - 2 E(x_i \tilde{\mu})$

-  $E(x_i^2) = \sigma^2 + \mu^2$

-  $E(\tilde{\mu}^2) = \frac{\sigma^2}{n} + \mu^2$

-  $E(x_i \tilde{\mu}) = \frac{1}{n} E(x_i x_1 + x_i x_2 + \dots + x_i x_i + \dots + x_i x_n) = \frac{n-1}{n} \mu^2 + \frac{1}{n} \sigma^2 + \frac{1}{n} \mu^2$

-  $E(x_i - \tilde{\mu})^2 = \frac{n-1}{n} \sigma^2$

$$E(x_i x_1) = \iint_{-\infty}^{\infty} x_i x_1 f(x_i x_1) dx_i dx_1$$

$$\iint_{-\infty}^{\infty} x_i x_1 g_i(x_i) g_1(x_1) dx_i dx_1 = \mu^2$$

Finalmente:

$$E[\tilde{\sigma}^2] = \frac{1}{n-1} n \frac{n-1}{n} \sigma^2 = \sigma^2$$



- Estimación de la media:  $\tilde{\mu} = \frac{\sum_{i=1}^n X_i}{n}$  y estimación de la varianza del estimador de la media  $\tilde{\sigma}_{\tilde{\mu}}^2 = \frac{1}{n} \tilde{\sigma}^2$

- Estimación de la varianza:  $\tilde{\sigma}^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \tilde{\mu})^2$

### **Ejemplo:**

Tenemos 100 mediciones de resistencias que debieran medir 200 kOhmio

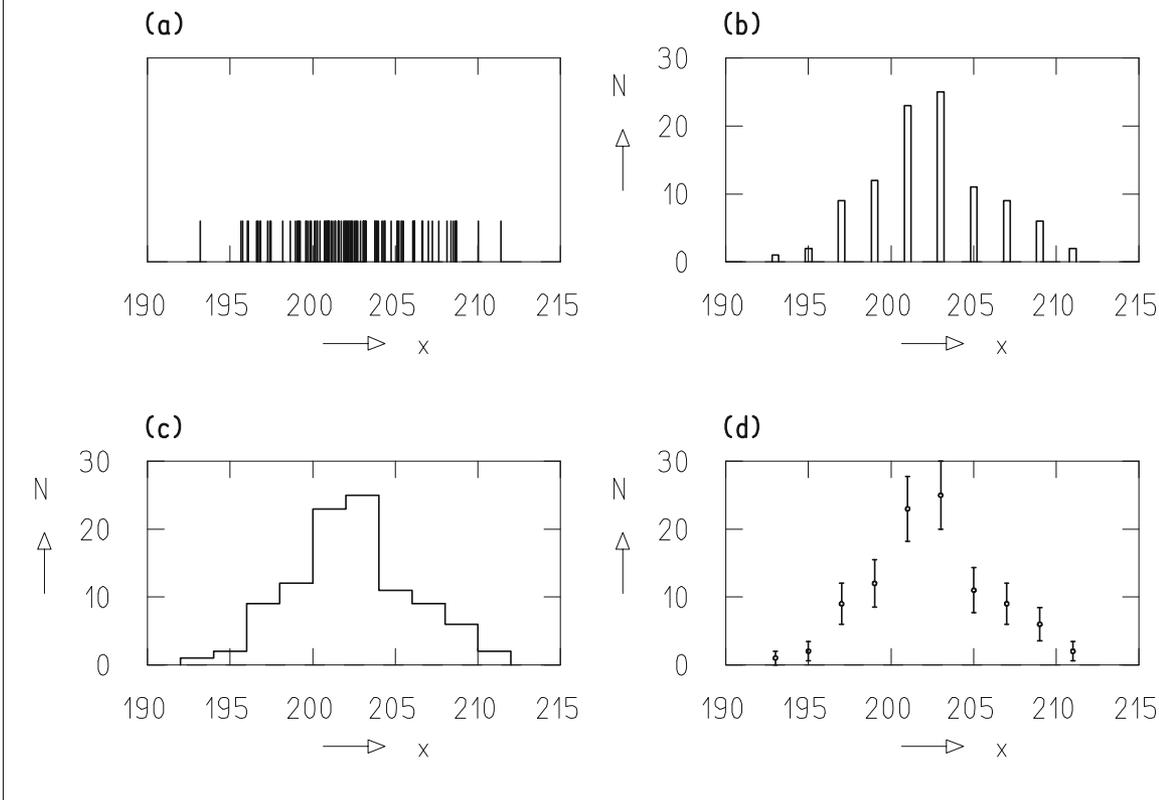
¿Cuál es el valor mas representativo dela resistencia?



**MAGGIA**  
LABORATORIO



193.199	195.673	195.757	196.051	196.092
196.596	196.679	196.763	196.847	197.267
197.392	197.477	198.189	198.650	198.944
199.070	199.111	199.153	199.237	199.698
199.572	199.614	199.824	199.908	200.118
200.160	200.243	200.285	200.453	200.704
200.746	200.830	200.872	200.914	200.956
200.998	200.998	201.123	201.208	201.333
201.375	201.543	201.543	201.584	201.711
201.878	201.919	202.004	202.004	202.088
202.172	202.172	202.297	202.339	202.381
202.507	202.591	202.633	202.716	202.884
203.051	203.052	203.094	203.094	203.177
203.178	203.219	203.764	203.765	203.848
203.890	203.974	204.184	204.267	204.352
204.352	204.729	205.106	205.148	205.231
205.357	205.400	205.483	206.070	206.112
206.154	206.155	206.615	206.657	206.993
207.243	207.621	208.124	208.375	208.502
208.628	208.670	208.711	210.012	211.394



Estimar la media y la varianza del promedio simple y la varianza del dato

$$\tilde{\mu} = \frac{\sum_{i=1}^n X_i}{n} = 202.2167157 (=202.217)$$

$$\tilde{\sigma}^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \tilde{\mu})^2 = 12.5272919 \quad \text{o} \quad \tilde{\sigma} = 3.53939145 (=3.539)$$





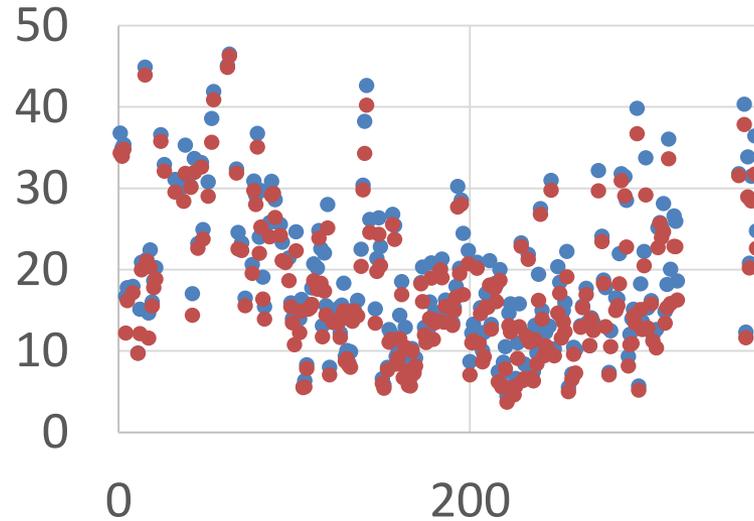
# MAGGIA

LABORATORIO

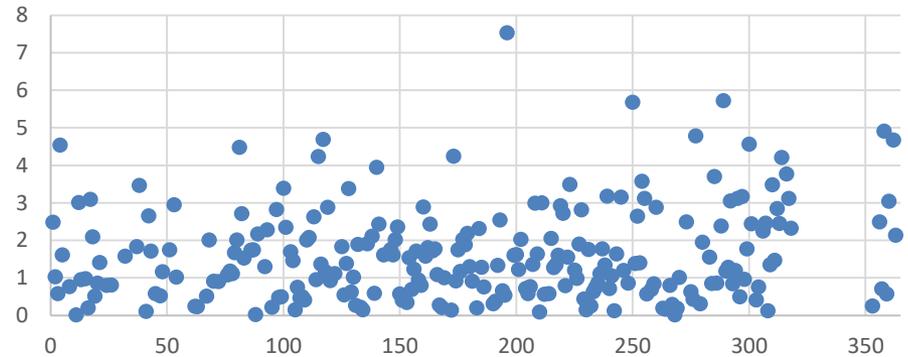
Se mide el vapor de agua integrado con dos instrumentos Radiosonda y GNSS

EZE	Día	GNSS	Radiosonda	Dif	Dif (-3)
87576	1	36.78	34.3	2.48	2.48
87576	2	34.94	33.91	1.03	1.03
87576	3	35.38	34.8	0.58	0.58
87576	4	16.72	12.18	4.54	4.54
87576	5	17.77	16.16	1.61	1.61
87576	8	17.92	17.15	0.77	0.77
87576	11	9.72	9.71	0.01	0.01
87576	12	15.11	12.1	3.01	3.01
87576	13	20.89	19.94	0.95	0.95
87576	15	44.88	43.9	0.98	0.98
87576	16	21.26	21.06	0.2	0.2
87576	17	14.66	11.57	3.09	3.09
87576	18	22.42	20.33	2.09	2.09
87576	19	16.02	15.51	0.51	0.51
87576	20	18.62	17.76	0.86	0.86
87576	21	20.24	18.83	1.41	1.41
87576	24	36.58	35.78	0.8	0.8
87576	26	32.92	32.11	0.81	0.81
87576	32	31.09	29.51	1.58	1.58
87576	37	30.2	28.37	1.83	1.83
87576	38	35.29	31.83	3.46	3.46
87576	41	30.21	30.1	0.11	0.11
87576	42	17.03	14.38	2.65	2.65
87576	43	33.64	31.93	1.71	1.71
87576	45	23.2	22.62	0.58	0.58
87576	47	33.14	32.62	0.52	0.52
87576	48	24.92	23.76	1.16	1.16

...



• GNSS • Radiosonda



GNSS - Radiosonda			
IWV(GNSS)	IWV(R)(Rad.)	$\sigma^2, \sigma, 3\sigma$	$\sigma^2, \sigma, 3\sigma$
17.0165902	15.3607012	1.12498619	1.10102027
		1.23574662	1.12089698
		4.83222605	4.46371122



# Muestreo aleatorio inhomogeneo

Sean  $t$  subpoblaciones que definen todos los casos que representan la población (todos los casos posibles):

$$G_1, G_2, \dots, G_t$$

Cuyas densidades de probabilidad son:

$$f_1(x), f_2(x), \dots, f_t(x)$$

La función de distribución de c/u, se puede escribir como:

$$F_i(x) = \int_{-\infty}^x f_i(x) dx = P(\mathbf{X} < x | \mathbf{X} \in G_i)$$

Teniendo en cuenta que:

$$F(x) = \sum_{i=1}^t P(\mathbf{X} \in G_i) F_i(x)$$

Entonces puedo escribir:

$$F(x) = P(\mathbf{X} < x | \mathbf{X} \in G) = \sum_{i=1}^t P(\mathbf{X} < x | \mathbf{X} \in G_i) P(\mathbf{X} \in G_i)$$

Y su f.d.p:

$$f(x) = \sum_{i=1}^t P(\mathbf{X} \in G_i) f_i(x)$$

## Estimación de la media y varianza



De la definición teórica de media, inferimos la estimación de la media, que llamaremos promedio pesado:

$$\mu = E[x] = \int_{-\infty}^{\infty} x f(x) dx = \sum_{i=1}^t p_i \int_{-\infty}^{\infty} x f_i(x) dx = \sum_{i=1}^t p_i \mu_i$$

El estimador de la media lo podemos escribir, en función de la estimación de media de cada subpoblación:

$$\mu = \sum_{i=1}^t p_i \mu_i \quad \rightarrow \quad \tilde{\mu} = \sum_{i=1}^t p_i \tilde{\mu}_i \quad \text{promedio pesado}$$

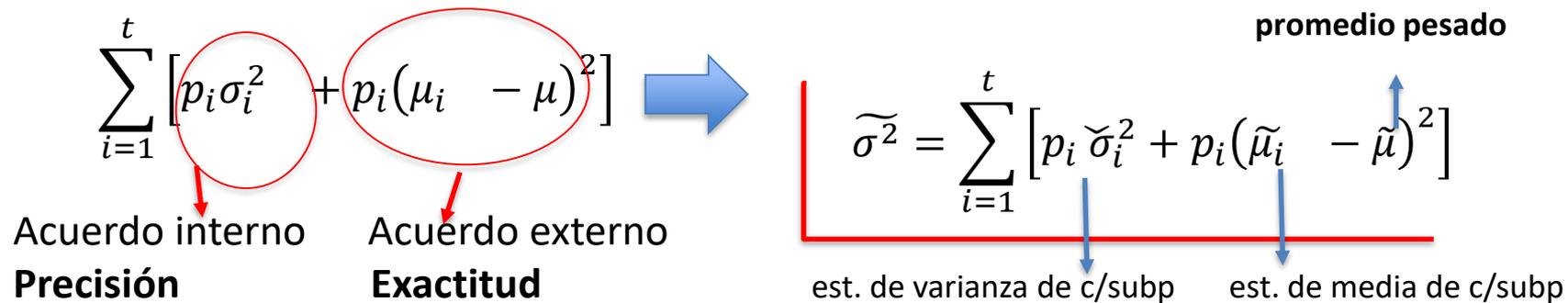




De la definición teórica de varianza, inferimos la estimación de la varianza (del dato)

$$\begin{aligned}
 E[(x - \mu)^2] &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \sum_{i=1}^t p_i \int_{-\infty}^{\infty} (x - \mu)^2 f_i(x) dx = \\
 &= \sum_{i=1}^t p_i \int_{-\infty}^{\infty} (x - \mu_i + \mu_i - \mu)^2 f_i(x) dx = \\
 &= \sum_{i=1}^t p_i \int_{-\infty}^{\infty} \left[ (x - \mu_i)^2 + (\mu_i - \mu)^2 + (x - \mu_i)(\mu_i - \mu) \right] f_i(x) dx = \\
 &= \sum_{i=1}^t p_i \int_{-\infty}^{\infty} \left[ (x - \mu_i)^2 + (\mu_i - \mu)^2 \right] f_i(x) dx = \sum_{i=1}^t \left[ p_i \sigma_i^2 + p_i (\mu_i - \mu)^2 \right]
 \end{aligned}$$

El **estimador de la varianza** lo podemos escribir, en función de la estimación de varianza de cada subpoblación:



De la regla de propagación podemos inferir la varianza del promedio pesado....

$$\tilde{\mu} = \sum_{i=1}^t p_i \tilde{\mu}_i$$

Nota: utilice la regla de propagación de errores y tome a la matriz A como un vector de pi, y el vector de x son las medias estimadas.

$$\vec{\tilde{\mu}} = \begin{pmatrix} \tilde{\mu}_1 \\ \vdots \\ \tilde{\mu}_t \end{pmatrix}, C_{\vec{\tilde{\mu}}} = \begin{bmatrix} \tilde{\sigma}_{\tilde{\mu}_1}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \tilde{\sigma}_{\tilde{\mu}_t}^2 \end{bmatrix}, A^T = \begin{bmatrix} p_1 \\ \vdots \\ p_t \end{bmatrix}$$

$$\tilde{\sigma}_{\tilde{\mu}}^2 = A C_{\vec{\tilde{\mu}}} A^T = \sum_{i=1}^t p_i^2 \tilde{\sigma}_{\tilde{\mu}_i}^2 = \sum_{i=1}^t p_i^2 \frac{1}{n_i} \tilde{\sigma}_i^2$$

Numero de elementos en la muestra de la subpoblación

Cual será el valor de  $p_i = P(X \text{ pertenezca a } G_i)$ ???

- Para responder esta pregunta debemos asumir una función de densidad de probabilidad o distribución de probabilidad que gobierne a la variable X...

Si la f. d. p. es gaussiana entonces:

$$p_i = \frac{1/\sigma_{\tilde{\mu}_i}^2}{\sum_{j=1}^t 1/\sigma_{\tilde{\mu}_j}^2}$$

Nota: Rosenfield (1967) propone un estimador de promedio pesado teniendo en cuenta el acuerdo externo.

$$\tilde{\sigma}_{\tilde{\mu}}^2 = \frac{1}{t-1} \sum_{i=1}^t p_i (\tilde{\mu}_i - \tilde{\mu})^2 \text{ (Acuerdo externo)}$$



# Método de máxima verosimilitud

Para poder hacer uso de este método tenemos en cuenta las siguientes características:

- Función de densidad de probabilidad que rige las muestras  $f(\vec{\mathbf{x}}, \vec{\lambda})$ , donde  $\vec{\lambda}$  es el vector de parámetros que definen a la f.d.p. Es decir:  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  y  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_p)$
- Las N muestras son:  $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(j)}, \dots, \mathbf{x}^{(N)}$

La probabilidad asociada a una muestra será:

$$dP^{(j)} = f(\mathbf{x}^{(j)}; \boldsymbol{\lambda}) d\mathbf{x} \quad .$$

Si las N muestras son linealmente independientes:

$$dP = \prod_{j=1}^N f(\mathbf{x}^{(j)}; \boldsymbol{\lambda}) d\mathbf{x}$$

La función de probabilidad se define como:

$$L = \prod_{j=1}^N f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

# Método de máxima verosimilitud (o probabilidad)

$$\ell = \ln L = \sum_{j=1}^N \ln f(\mathbf{x}^{(j)}; \lambda).$$

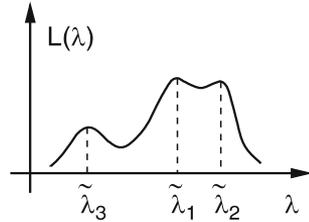
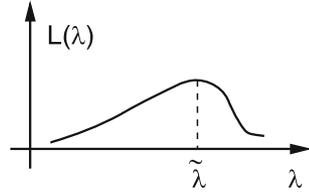
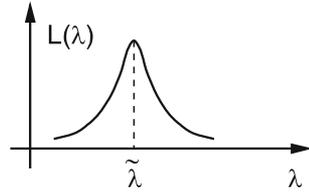
Función logarítmica: MAXIMIZAR!!!

$$\ell' = d\ell/d\lambda = 0.$$

Supongamos  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$ ,  $p=1$

$$\ell' = \sum_{j=1}^N \frac{d}{d\lambda} \ln f(\mathbf{x}^{(j)}; \lambda) = \sum_{j=1}^N \frac{f'}{f} = \sum_{j=1}^N \varphi(\mathbf{x}^{(j)}; \lambda),$$

$$\varphi(\mathbf{x}^{(j)}; \lambda) = \left( \frac{d}{d\lambda} f(\mathbf{x}^{(j)}; \lambda) \right) / f(\mathbf{x}^{(j)}; \lambda)$$



En general:

$$\frac{\partial \ell}{\partial \lambda_i} = 0, \quad i = 1, 2, \dots, p.$$

Volvemos a el valor de  $p_i$  !!!!!

Asumimos una f.d.p gaussiana, cada muestra tiene varianza conocida y todas tienen el mismo valor medio:

$$f(\mathbf{x}^{(j)}; \lambda) dx = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(\mathbf{x}^{(j)} - \lambda)^2}{2\sigma_j^2}\right) dx$$

$$L = \prod_{j=1}^N \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(\mathbf{x}^{(j)} - \lambda)^2}{2\sigma_j^2}\right) \longrightarrow \ell = -\frac{1}{2} \sum_{j=1}^N \frac{(\mathbf{x}^{(j)} - \lambda)^2}{\sigma_j^2} + \text{const.}$$

$$\frac{d\ell}{d\lambda} = \sum_{j=1}^N \frac{\mathbf{x}^{(j)} - \lambda}{\sigma_j^2} = 0.$$

$$\tilde{\lambda} = \frac{\sum_{j=1}^N \frac{\mathbf{x}^{(j)}}{\sigma_j^2}}{\sum_{j=1}^N \frac{1}{\sigma_j^2}} \longrightarrow p_j = \frac{1/\sigma_j^2}{\sum_{j=1}^N 1/\sigma_j^2}$$





Volvemos a el valor de  $p_i$  !!!!!

$$\frac{d\ell}{d\lambda} = \sum_{j=1}^N \frac{\mathbf{x}^{(j)} - \lambda}{\sigma_j^2} = 0.$$

$$\tilde{\lambda} = \frac{\sum_{j=1}^N \mathbf{x}^{(j)}}{\sum_{j=1}^N \frac{1}{\sigma_j^2}} \quad \Longrightarrow \quad p_j = \frac{1/\sigma_j^2}{\sum_{j=1}^t 1/\sigma_j^2}$$

Para nuestro caso que tenemos un vector de estimador de medias, con sus desviaciones estandard

$$\tilde{\mu} = \sum_{i=1}^t p_i \tilde{\mu}_i \quad \text{y} \quad C_{\tilde{\mu}} = \begin{bmatrix} \widetilde{\sigma_{\mu_1}^2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \widetilde{\sigma_{\mu_t}^2} \end{bmatrix} \quad \Longrightarrow \quad p_i = \frac{1/\sigma_{\mu_i}^2}{\sum_{j=1}^t 1/\sigma_{\mu_j}^2}$$



Ejemplo de poblaciones fraccionadas:

Hemos medido 6 instrumentos distintos la intensidad de corriente en un circuito, en cada oportunidad el registro fue de 10 veces con cada instrumento. Obteniendo el siguiente resultado.

Instru mento	Ampere	
1	0.63	0.01
2	0.641	0.011
3	0.675	0.03
4	0.7	0.051
5	0.642	0.011
6	0.631	0.021

**Calcule:**

- El valor representativo de la corriente del circuito, teniendo en cuenta que las seis medidas fueran hechas con el mismo instrumento. Estime su error.
- El valor representativo de la corriente del circuito, teniendo en cuenta los diferentes instrumentos de donde provienen. Estime su error.

### Ejemplo de poblaciones fraccionadas:

Hemos medido 6 instrumentos distintos la intensidad de corriente en un circuito, en cada oportunidad el registro fue de 10 veces con cada instrumento. Obteniendo el siguiente resultado.

Instru mento	Ampere	
1	0.63	0.01
2	0.641	0.011
3	0.675	0.03
4	0.7	0.051
5	0.642	0.011
6	0.631	0.021

### Calcule:

- El **valor representativo** de la corriente del circuito, teniendo en cuenta que las seis medidas fueran hechas con el mismo instrumento. Estime **su error**.
- Rta: Promedio simple

$$\tilde{\mu} = \frac{\sum_{i=1}^n X_i}{n} \quad \longrightarrow \quad \mathbf{0.653}$$

$$\tilde{\sigma}^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \tilde{\mu})^2 \quad \longrightarrow \quad \tilde{\sigma}_{\tilde{\mu}}^2 = \frac{1}{n} \tilde{\sigma}^2$$

$$\tilde{\sigma} = \mathbf{0.028} \qquad \qquad \tilde{\sigma}_{\tilde{\mu}} = \mathbf{0.011}$$

### Ejemplo de poblaciones fraccionadas:

Hemos medido 6 instrumentos distintos la intensidad de corriente en un circuito, en cada oportunidad el registro fue de 10 veces con cada instrumento. Obteniendo el siguiente resultado.

Instru mento	Ampere	
1	0.63	0.01
2	0.641	0.011
3	0.675	0.03
4	0.7	0.051
5	0.642	0.011
6	0.631	0.021

### Calcule:

- El valor representativo de la corriente del circuito, teniendo en cuenta los diferentes instrumentos de donde provienen. Estime su error.

$$p_i = \frac{1/\sigma_{\mu_i}^2}{\sum_{j=1}^t 1/\sigma_{\mu_j}^2}$$

$$\tilde{\mu} = \sum_{i=1}^t p_i \tilde{\mu}_i = 0.639$$

$$\tilde{\sigma}^2 = \sum_{i=1}^t [p_i \tilde{\sigma}_i^2 + p_i (\tilde{\mu}_i - \tilde{\mu})^2] \rightarrow \tilde{\sigma} = 0.046$$

$$\tilde{\sigma}_{\tilde{\mu}}^2 = \sum_{i=1}^t p_i^2 \tilde{\sigma}_{\tilde{\mu}_i}^2 = \rightarrow \tilde{\sigma}_{\tilde{\mu}} = 0.006$$

$$\tilde{\sigma}_{\tilde{\mu}}^2 = \frac{1}{t-1} \sum_{i=1}^t p_i (\tilde{\mu}_i - \tilde{\mu})^2 \text{ (Acuerdo externo)} \rightarrow \tilde{\sigma}_{\tilde{\mu}}^{AE} = 0.005$$