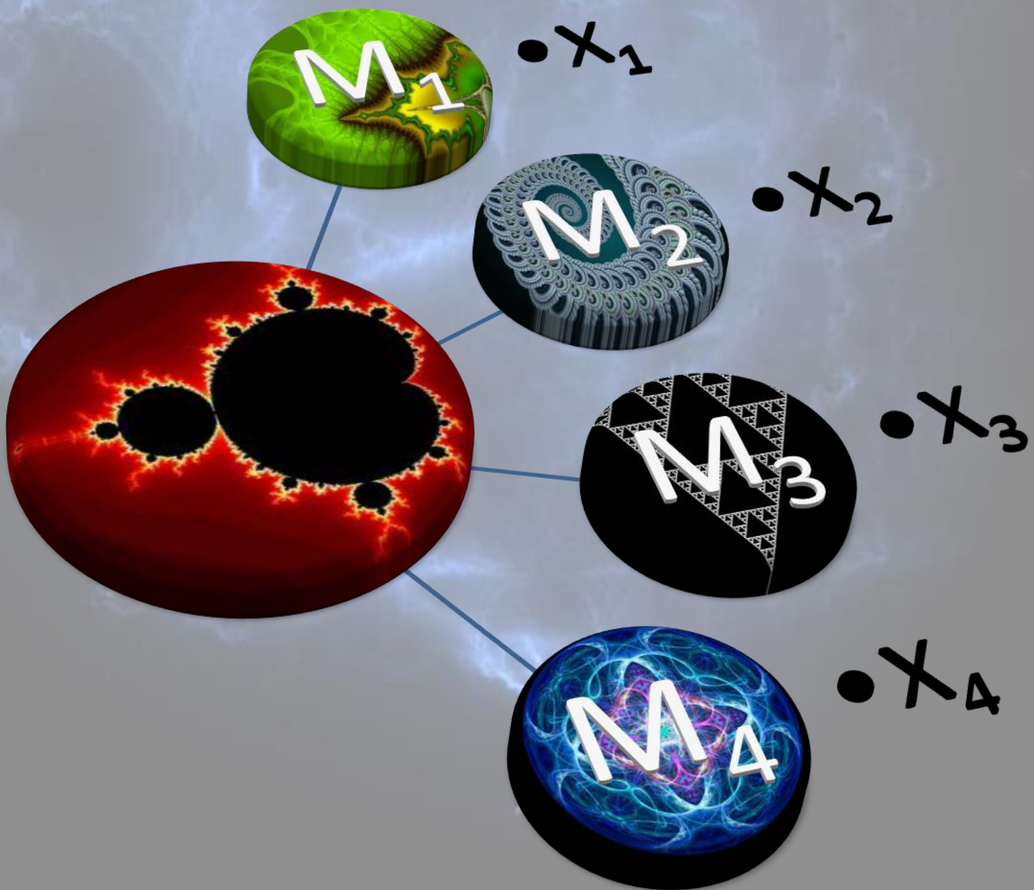
Teorema del límite central

Distribuciones muestrales

MAS obtenidas de una población *N*, son por naturaleza propia impredecibles. No esperamos que dos muestras aleatorias de tamaño *n*, tomadas de la misma población *N*, tenga la misma media muestral o que sean completamente parecidas. Puede esperarse que cualquier estadístico, como la media muestral, calculado a partir de las medias en una muestra aleatoria, cambie su valor de una muestra a otra, por ello, se requiere estudiar la ***distribución*** de todos los valores posibles de un estadístico. Tales distribuciones serán muy importantes en el estudio de la estadística inferencial, porque las inferencias sobre las poblaciones se harán usando estadísticas muestrales. Con el análisis de las distribuciones asociadas con los estadísticos muestrales, podremos juzgar la confiabilidad de un estadístico muestral como un instrumento para hacer inferencias sobre un parámetro poblacional desconocido. Como los valores de un estadístico, tal como *X*, varían de una muestra aleatoria a otra, se le puede considerar como una *variable aleatoria con su* correspondiente ***distribución de frecuencias***.

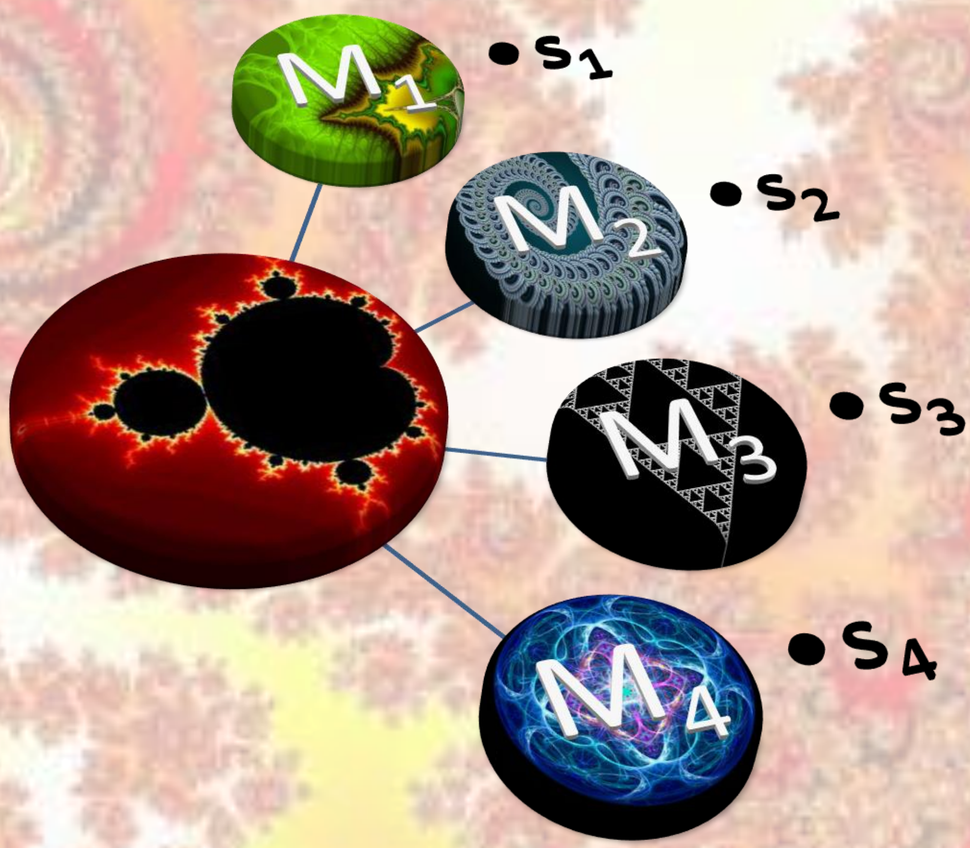
Definición

La distribución de frecuencia de un estadístico muestral se denomina ***distribución muestral.***
En general, la ***distribución muestral*** de un estadístico es la de todos sus valores posibles calculados a partir de muestras del mismo tamaño.



Definición

Supongamos que elegimos muestras aleatorias de tamaño n, de una población de tamaño *N*, calculamos la deviación estándar de cada muestra. La colección de todas estas desviaciones estándar muestrales se llamará **distribución muestral de la desviación estándar.**



Ejemplo

Se eligen muestras ordenadas de tamaño 2, con reemplazo, de la población de valores 0, 2, 4 y 6. Encontrar:
μ, la media poblacional.
σ, la desviación estándar poblacional.
μ_x, la media de la distribución muestral de medias.
σ_x, la desviación estándar de la distribución muestral de medias.
Además, graficar las frecuencias para la población y para la distribución muestral de medias.

Solución

μ la media poblacional.

$$\mu = \frac{1}{4}\sum_{i=1}^{4} x_i$$

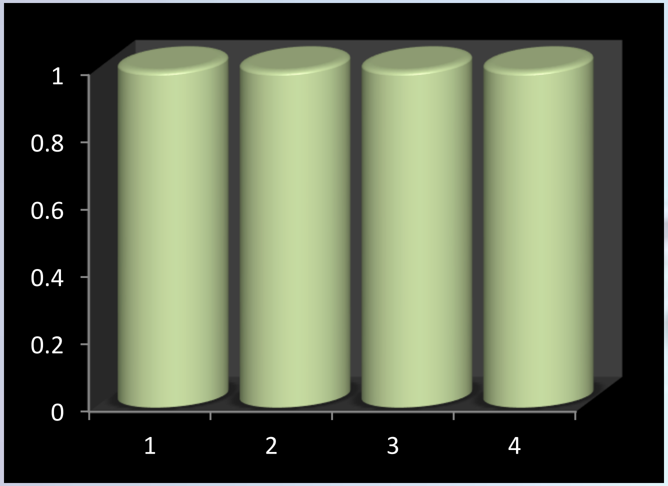
$$= \frac{1}{4}(0+2+4+6) = \frac{12}{4} = 3$$

σ, la desviación estándar poblacional.

$$\sigma^2 = \frac{1}{4}\sum_{i=1}^{4}(x_i-\mu)^2$$

$$= \frac{1}{4}\left(\sqrt{(0-3)^2+(2-3)^2+(4-3)^2+(6-3)^2}\right) = 5$$

DFMM



La media de la distribución muestral de la media se obtiene entonces de la siguiente forma:

$$\mu_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} \overline{x_i}$$

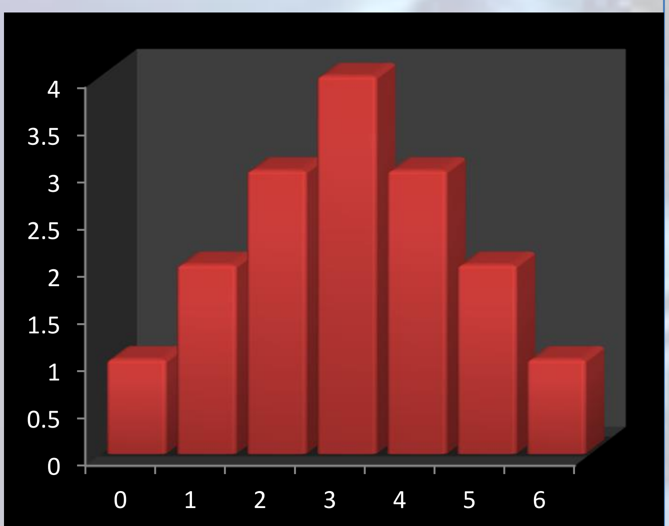
$$= \frac{0+1+2+3+1+2+3+4+2+3+4+5+3+4+5+6}{16}$$

$$= \frac{0(1)+1(2)+2(3)+3(4)+4(3)+5(2)+6(1)}{16}$$

$$= 0(1/16)+1(2/16)+2(3/16)+3(4/16)+4(3/16)+5(2/16)+6(1/16)$$

$$= 0P(\overline{X}=0)+1(\overline{X}=1)+2P(\overline{X}=2)+\cdots+6P(\overline{X}=6)$$

$$= \sum_{i=1}^{n} \bar{x}_i P(\overline{X}=\bar{x}_i) = 3.0$$



La media de la distribución muestral de la varianza se obtiene entonces de la siguiente forma:

$$S_{\bar{x}}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu_{\bar{x}}\right)^2$$

$$= \frac{(0-3)^2+(1-3)^2+(2-3)^2+\cdots(6-3)^2}{16}$$

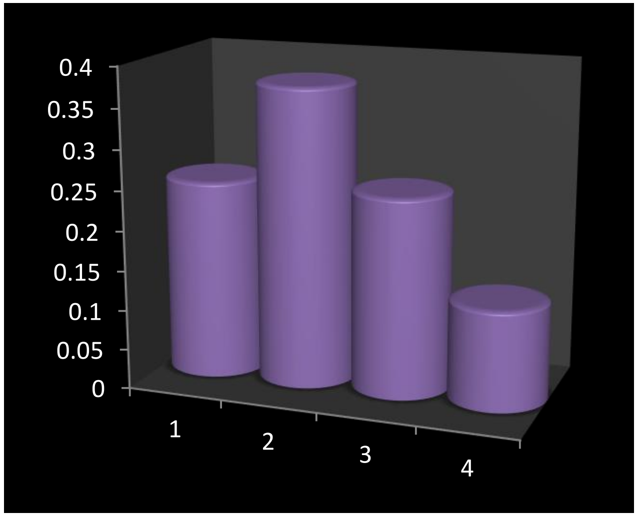
$$= \frac{(0-3)^2+(1-3)^2(2)+(2-3)^2(3)+\cdots+(6-3)^2}{16}$$

$$= (0-3)^2\,\tfrac{1}{16}+(1-3)^2(\tfrac{2}{16})+(2-3)^2(\tfrac{3}{16})+\cdots+(6-3)^2\,\tfrac{1}{16}$$

$$= (0-3)^2P(\overline{X}=0)+(1-3)^2P(\overline{X}=1)+(2-3)^2P(\overline{X}=2)+\cdots+(6-3)^2P(\overline{X}=6)$$

$$= \sum_{i=1}^{n}\left(x_i-\mu_{\bar{x}}\right)^2 P(\overline{X}=\bar{x}_i) = 2.5$$

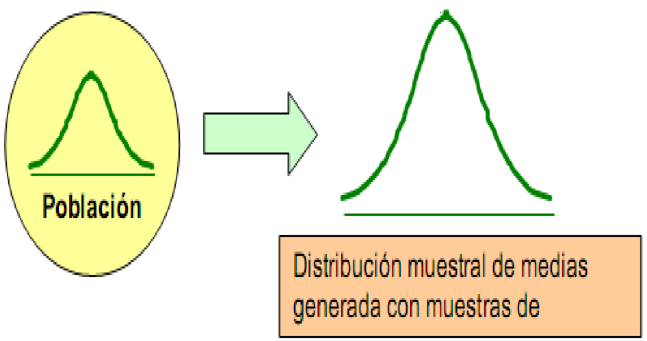
Deducción



$$\mu = \mu_{\overline{X}}$$

$$\sigma^2 = n S_{\overline{X}}^2$$

$$\sigma_X = \frac{\sigma}{\sqrt{n}}$$



Para cualquier variable aleatoria, la distribución muestral de medias tiene una media o valor esperado, una varianza y una desviación estándar, se ha mostrado para dos ejemplos distintos que la distribución muestral de medias tiene una media igual a la media poblacional.

Podemos ver que una distribución muestral se genera extrayendo todas las posibles muestras del mismo tamaño de la población y calculándoles a éstas su estadístico.

Si la población de la que se extraen las muestras es normal, la distribución muestral de medias será normal sin importar el tamaño de la muestra.

parámetros poblacionales								Estadisticos muestrales																
X	fx	P(X)	μ	$(xi-\mu)^2$	σ^2	DSTP	$_NP_n$	x_1,x_2	$x_1+x_2/2$	X_m	$P(X_1,X_2)$	X_m	fx_m	$(x_i-\mu)^2$	s_i^2	fs2	$P(S_i^2)$	$(x_i-\mu)^2P(s_i^2)$	fx_m/M	$P(x_m)$	$x_iP(x_i)$	μ	S_i^2
0	1	0.25	3	9	5	2.24	12	0,0	0	0	0.0625	0	1	9	0	4	0.25	0	0.0625	0.0625	0.000	3	2.50
2	1	0.25		1			6	0,2	1	1	0.0625	1	2	4	1	6	0.375	0.375	0.1250	0.1250	0.125		
4	1	0.25		1				0,4	2	2	0.0625	2	3	1	4	4	0.25	1	0.1875	0.1875	0.375		
6	1	0.25		9				0,6	3	3	0.0625	3	4	0	9	2	0.125	1.125	0.2500	0.2500	0.750		
		1.00						2,0	1	4	0.0625	4	3	4		16	1	2.5	0.1875	0.1875	0.750		
								2,2	2	5	0.0625	5	2	1					0.1250	0.1250	0.625		
								2,4	3	6	0.0625	6	1	0					0.0625	0.0625	0.375		
								2,6	4		0.0625			1					1.0000	1.000			
								4,0	2		0.0625			1									
								4,2	3		0.0625			0									
								4,4	4		0.0625			1									
								4,6	5		0.0625			4									
								6,0	3		0.0625			0									
								6,2	4		0.0625			1									
								6,4	5		0.0625			4									
								6,6	6		0.0625			9									

Si la población de donde se extraen las muestras no es normal, entonces el tamaño de la muestra debe ser mayor o igual a 30, para que la distribución muestral tenga una forma acampanada. Mientras mayor sea el tamaño de la muestra, más cerca estará la distribución muestral de ser normal. Para muchos propósitos, la aproximación normal se considera buena si se cumple n=30. La forma de la distribución muestral de medias sea aproximadamente normal, aún en casos donde la población original es bimodal, es realmente notable.

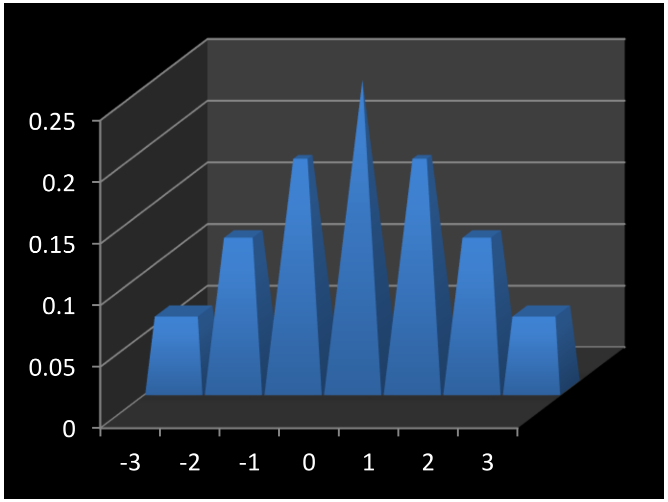
Error Muestral

Cualquier tipo de medida implica algún error en si misma. Si utilizamos la media para medir, y/o estimar, la media poblacional μ, entonces la media muestral, como medida, presenta algún error.

Se denomina el error muestral si la media muestral X puede expresarse como la suma de dos cantidades, la media poblacional μ y el error muestral, denotado por *e*.

$$e = X - \mu$$

Error Muestral



x1,x2	x1+x2/2	e= X-μ	μ_e	$e-\mu_e$	σ_e	dste	$P(X_1,X_2)$	e	fe	P(e)	P(e)	$(e-\mu_e)^2$	$(e-\mu_e)^2P(e)$	σ_e^2	σ
0,0	0	-3.000	0.000	9.000	2.50	1.581	0.0625	-3	1	0.0625	0.0625	9	0.5625	2.5	1.58
0,2	1	-2.000		4.000			0.0625	-2	2	0.125	0.125	4	0.5		
0,4	2	-1.000		1.000			0.0625	-1	3	0.1875	0.1875	1	0.1875		
0,6	3	0.000		0.000			0.063	0	4	0.25	0.250	0	0		
2,0	1	-2.000		4.000			0.0625	1	3	0.1875	0.1875	1	0.1875		
2,2	2	-1.000		1.000			0.0625	2	2	0.125	0.125	4	0.5		
2,4	3	0.000		0.000			0.0625	3	1	0.0625	0.0625	9	0.5625		
2,6	4	1.000		1.000			0.0625								
4,0	2	-1.000		1.000			0.0625								
4,2	3	0.000		0.000			0.0625								
4,4	4	1.000		1.000			0.0625								
4,6	5	2.000		4.000			0.0625								
6,0	3	0.000		0.000			0.0625								
6,2	4	1.000		1.000			0.0625								
6,4	5	2.000		4.000			0.0625								
6,6	6	3.000		9.000			0.0625								

Conclusión

La desviación estándar de la distribución muestral de un estadístico se conoce como error estándar del estadístico. Para el ejercicio anterior el error estándar de la media denotado por σ_e, es 1.58. Con esto se puede muestra que si para una población se eligen muestras de tamaño n con reemplazo, entonces el
error estándar de la media es igual a la desviación estándar de la distribución de los errores muestrales.

$$\sigma_X = \sigma_e$$

Cuando las muestras se toman de una población pequeña y sin reemplazo, se puede usar la formula siguiente para encontrar σ_x donde σ es la desviación estándar de la población de donde se toman las muestras, n es el tamaño de la muestra y N el tamaño de la población.
Como regla de cálculo, si el muestreo se hace sin reemplazo y el tamaño de la población es al menos 20 veces el tamaño de la muestra (N≥20), entonces se puede usar la fórmula.

$$\sigma_X = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$\sqrt{\frac{N-n}{N-1}}$$

Al factor se le denomina factor de corrección para una población finita.

Ejemplo

Suponga que la tabla siguiente muestra la antigüedad en años en el trabajo de tres maestros universitarios de matemáticas:

Maestro	Antigüedad
A	6
B	4
C	2

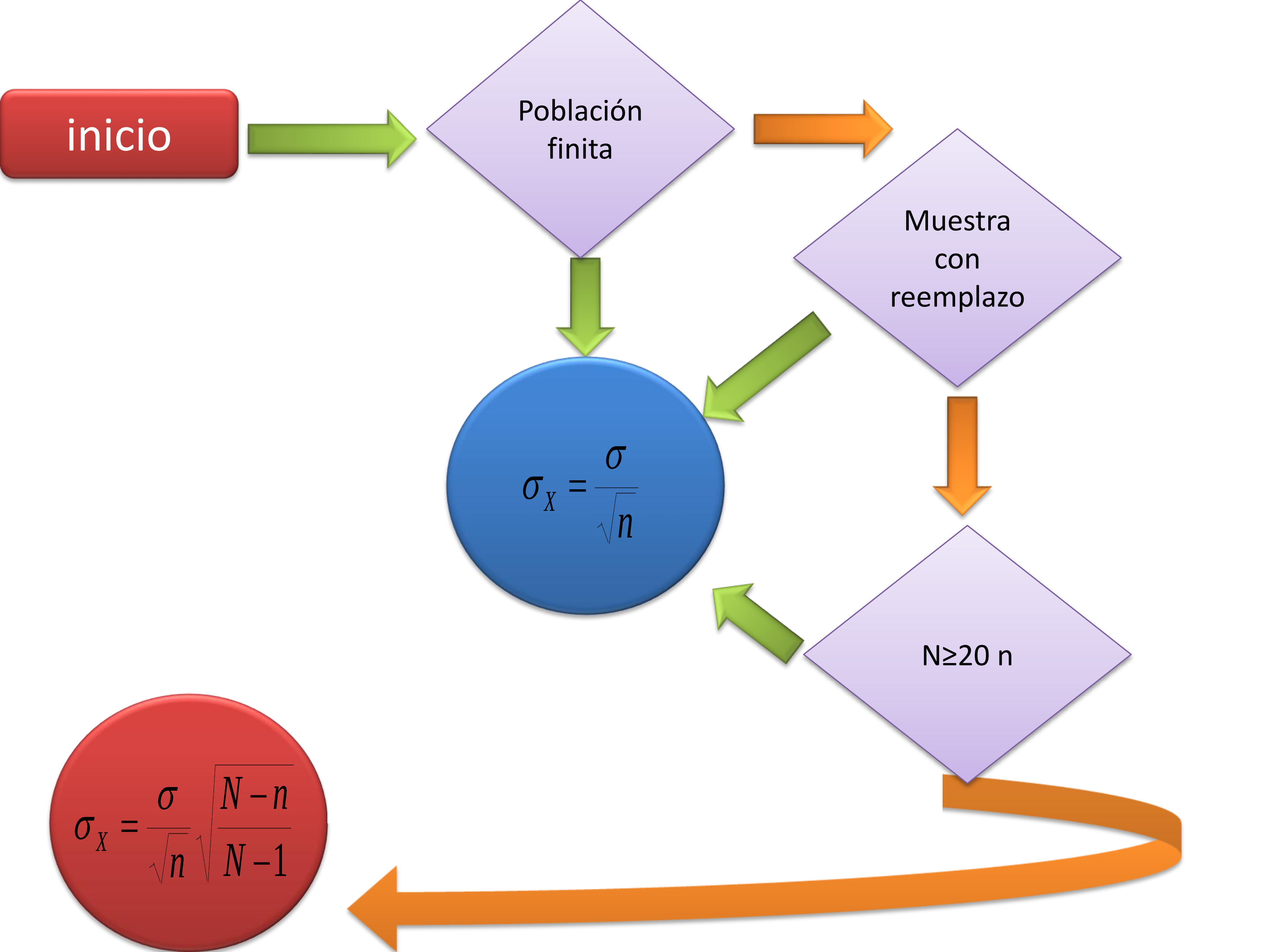
Suponga además que se seleccionan muestras aleatorias de tamaño 2 sin reemplazo. Calcule la antigüedad media para cada muestra, la media de la distribución muestral y el error estándar, o la desviación estándar de la distribución muestral.

Solución

Se pueden tener ${}_3C_2 = 3$ muestras posibles de tamaño 2, calcularemos sus respectivas medias muestrales.

población						muestra					
maestro	antigüedad	μ	σ_i^2	σ^2	σ	muestra 2	X_i	μ_x	σ_{Xi}^2	σ_x^2	σ_x
A	**6**	**4**	**4**	**2.66667**	**1.63**	**A,B**	**5**	**4**	**1**	**0.667**	**0.816**
B	**4**		**0**			**A,C**	**4**		**0**		
C	**2**		**4**			**B,C**	**3**		**1**	teorema LC	
											1.155
										Corrección LC	
											0.816

$$\sigma_X = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$


inicio
Población finita
Muestra con reemplazo
N≥20 n
$\sigma_X = \frac{\sigma}{\sqrt{n}}$
$\sigma_X = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$