

FORMULARIO

Muestreo

POBLACIONES INFINITAS (N ≥ 100,000)	POBLACIONES FINITAS (N < 100,000 es decir 99,999)
$n_0 = \frac{Z^2 * P * Q}{E^2}$ <p style="text-align: center; background-color: yellow;">Si el ERROR MUESTRAL se da como porcentaje</p> $n_0 = \frac{Z^2 * \sigma^2}{E^2}$ <p style="text-align: center; background-color: yellow;">Si el ERROR MUESTRAL se da como cantidad</p>	$n_0 = \frac{N * Z^2 * P * Q}{(N-1)E^2 + (Z^2 * P * Q)}$ <p style="text-align: center; background-color: yellow;">Si el ERROR MUESTRAL se da como porcentaje</p> $n_0 = \frac{N * Z^2 * \sigma^2}{(N-1)E^2 + (Z^2 * \sigma^2)}$ <p style="text-align: center; background-color: yellow;">Si el ERROR MUESTRAL se da como cantidad</p>
<p>Dónde:</p> <p>n₀ = Tamaño mínimo de muestra N = Tamaño de la población Z = Nivel de confianza en valores Z, p.ej. 95% ---- Z= 1.96 P = probabilidad de éxito o preferencia, en valor relativo, p.ej. 60% ---- 0.6 Q = probabilidad de fracaso o No preferencia, en valor relativo, p.ej. 40 % ---- 0.4 E = error muestral, en valor relativo, p.ej. 3% de error máximo ----- 0.03 σ^2 = <i>varianza poblacional</i></p>	

Cuartiles y Percentiles

CUARTILES Y PERCENTILES	
Dato suelto	Dato agrupado
$\text{Ubicación de } Q_1 = \frac{(n+1)}{4}$	$Q_1 = L_i + \left(\frac{\frac{n}{4} - fa}{f} \right) A$
$\text{Ubicación de } Q_3 = \frac{3(n+1)}{4}$	$\text{Ubicación} = \frac{n}{4}$
$\text{Ubicación de } P_k = \frac{k(n+1)}{100}$	$Q_3 = L_i + \left(\frac{\frac{3n}{4} - fa}{f} \right) A$
<p>Dónde: P = Percentil n = tamaño de la muestra k = percentil deseado</p>	$\text{Ubicación} = \frac{3n}{4}$
<p>Busque el número correspondiente a la ubicación, determine la distancia proporcional según el cálculo al siguiente número y súmelas.</p>	$P_k = L_i + \left(\frac{\frac{kn}{100} - fa}{f} \right) A$
	$\text{Ubicación} = \frac{kn}{100}$ <p>Si sale decimal redondee L_i = lim inferior REAL fa = frec acum antes de llegar</p>

Agrupación de datos

Rango (R) = X_{máx.} - X_{mín.} (No redondear)

Criterio de Sturges, No. Intervalos = 1 + (3.322 log n) (Redondear de forma entera convencional)

Amplitud (A ó k) = $\frac{\text{Rango}}{\text{No. intervalos (ya redondeado)}}$ (Redondear hacia arriba, según la complejidad del problema)

Por ejemplo, A=3.12255555, si los datos manejados por el problema son:

Enteros, el redondeo queda en 4 (A=3.~~12255555~~, el 3 se lleva a 4)

Si tienen 1 decimal, el redondeo queda en 3.2 (A=3.12~~255555~~,

Si tienen 2 decimales, el redondeo queda en 3.13 (A=3.12~~255555~~,

Si tienen 3 decimales, el redondeo queda en 3.123 (A=3.122~~55555~~,

Nota: Todo esto se hace con el fin de ser práctico y que el valor máximo siempre sea contabilizado en el último intervalo y este no quede fuera, sin importar nos pasemos un poco.

Medidas de tendencia central y de dispersión

Datos sueltos		Datos agrupados	
Promedio o Media aritmética			
$\bar{x} = \frac{\sum x}{n}$		$\bar{x} = \frac{\sum xf}{n}$	
Mediana			
<p>\tilde{X} = Por observación, es el valor a la mitad de la distribución de datos una vez ordenados de menor a mayor.</p> <p style="text-align: center;">$Ubicación = \frac{n+1}{2}$</p> <p>Si son datos pares (promedie bordes) Si son datos nones (saldrá directo)</p>		<p style="text-align: center;">$\tilde{x} = L_i + \left(\frac{\frac{n}{2} - fa}{f}\right) A$</p> <p style="text-align: right;">$Ubicación = \frac{n}{2}$</p> <p>Dónde: <i>L_i</i> = Límite inferior REAL donde está la mediana <i>fa</i>= frecuencia acumulada ANTES de llegar a la mediana <i>A</i>= Amplitud <i>n</i> = tamaño de la muestra <i>f</i> = frecuencia de la mediana</p> <p style="text-align: right;"><i>NOTA: La ubicación de la mediana se selecciona en base a fa</i></p>	
Moda			
<p>\hat{X} = Por observación, la frecuencia más alta.</p> <p>NOTA: Puede haber varias modas, tienen que tener el mismo valor numérico y ser las altas.</p> <p>AMODAL= Sin moda UNIMODAL= 1 moda BIMODAL= 2 MODAS MULTIMODAL= 3 ó más modas</p>		<p style="text-align: center;">$\hat{x} = L_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) A$</p> <p>Dónde: <i>L_i</i> = Límite inferior REAL donde está la moda <i>A</i>= Amplitud <i>f</i> modal= frecuencia de la moda <i>f</i> anterior= frecuencia antes de la moda <i>f</i> superior= frecuencia después de la moda $\Delta_1 = f \text{ modal} - f \text{ anterior}$ $\Delta_2 = f \text{ modal} - f \text{ superior}$</p> <p>NOTA: El intervalo que contiene la moda(s) se ubica por tener la frecuencia "f" más alta. Puede haber varias modas, y habrá que repetir esta fórmula "n" veces la haya.</p>	
Varianza muestral			
Fórmula Original $s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$	Fórmula Alternativa $s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1}$	Fórmula Original $s^2 = \frac{\sum(x - \bar{x})^2 f}{n - 1}$	Fórmula alternativa $s^2 = \frac{\sum x^2 f - \frac{(\sum xf)^2}{n}}{n - 1}$
Desviación estándar muestral			
$s = \sqrt{s^2}$		$s = \sqrt{s^2}$	
Curtosis ó Apuntalamiento, llamada g2			
$Ap = \frac{\left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_j - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}}{1}$	Ap > 0 Leptocurtica Ap < 0 Platicurtica AP = 0 Mesocurtica	$A_p = \frac{\sum(x - \bar{x})^4 f}{n \sigma^4}$ $\sigma^4 = \left(\frac{\sum(x - \bar{x})^2 f}{n} \right)^2$	Ap > 3 Leptocurtica Ap < 3 Platicurtica Ap = 3 Mesocurtica
Asimetría, llamada g1			
$As = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_j - \bar{x}}{s} \right)^3$	As > 0 Asimetría "+" As < 0 Asimetría "-" As = 0 Simétrica	Pearson $A_s = \frac{3(\bar{x} - \tilde{x})}{s}$ (media, mediana, desv est muestral) Bowley $A_s = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$ Fischer $A_s = \frac{\sum(x - \bar{x})^3 f}{n \sigma^3}$	As > 0 Asimetría "+" As < 0 Asimetría "-" As = 0 Simétrica
Coficiente de Variabilidad			
$c.v = \frac{s}{\bar{x}} \times 100$ Desviación Estándar, media	0 ≤ C.V < 20 % Muy confiable 20 ≤ C.V < 30 % Confiable 30 ≤ C.V < 40 % Relat. Confiable 40% ↑ No confiable	$c.v = \frac{s}{\bar{x}} \times 100$ Desviación Estándar, media	0 ≤ C.V < 20 % Muy confiable 20 ≤ C.V < 30 % Confiable 30 ≤ C.V < 40 % Relat. Confiable 40% ↑ No confiable