

FORMULARIO

Muestreo

POBLACIONES INFINITAS (n ≥ 100,000)	POBLACIONES FINITAS (n < 100,000 es decir 99,000)
$n_0 = \frac{Z^2 * P * Q}{E^2}$ <p style="text-align: center; background-color: yellow;">Si se dan proporciones</p> $n_0 = \frac{Z^2 * \sigma^2}{E^2}$ <p style="text-align: center; background-color: yellow;">Si se dan cantidades</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 se dan proporciones</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 se dan cantidades</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</p>	

Cuartiles y Percentiles

<i>Datos Suetos</i>		<i>Datos agrupados</i>
<i>Cálculo rápido</i>	<i>Cálculo exacto</i>	
$P_k = \frac{n * k}{100}$	$Q_1 = \frac{(n + 1)}{4}$ $Q_3 = \frac{3(n + 1)}{4}$	$Q_1 = L_i + \left(\frac{\frac{n}{4} - fa}{f} \right) A$ $Q_3 = L_i + \left(\frac{\frac{3n}{4} - fa}{f} \right) A$
<p>Dónde:</p> <p>P_k = percentil buscado n = tamaño de la muestra k = percentil</p> <p>Regla de redondeo Si el resultado es: Decimal---- Redondear de forma normal, 0.49 baja, 0.50 sube Entero-----Sumar 0.5</p>		

Agrupación de datos

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

Criterio de Stürges, No. Intervalos = 1 + (3.322 log n) Redondear al entero inmediato, si se pasa por muy poco pruebe con esos intervalos, p.ej

No.interv = 6.01 quizás dejar en 6)

Rango

Amplitud (A ó k) = $\frac{\text{Rango}}{\text{No.intervalos(ya redondeado)}}$ (Redondear hacia arriba, según la complejidad del problema, si sale exacta y es la necesaria según sus decimales dejar tal cual, sino seguir lo siguiente:

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=~~3.12255555~~,

Si tienen 2 decimales, el redondeo queda en 3.13 (A=~~3.12255555~~,

Si tienen 3 decimales, el redondeo queda en 3.123 (A=~~3.12255555~~,

Nota: Todo esto se hace con el fin de 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 noes (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: Li = Límite inferior REAL donde está la mediana fa= frecuencia acumulada ANTES de llegar a la mediana A= Amplitud n = tamaño de la muestra f = 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: Li = Límite inferior REAL donde está la moda A= Amplitud f modal= frecuencia de la moda f anterior= frecuencia antes de la moda f 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 las haya.</p>	
Varianza muestral			
<p>Fórmula Original</p> $s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$	<p>Fórmula Alternativa</p> $s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1}$	<p>Fórmula Original</p> $s^2 = \frac{\sum(x - \bar{x})^2 f}{n - 1}$	<p>Fórmula alternativa</p> $s^2 = \frac{\sum x^2 f - \frac{(\sum xf)^2}{n}}{n - 1}$
Desviación estándar muestral			
$s = \sqrt{s^2}$		$s = \sqrt{s^2}$	
Curtosis ó Apuntalamiento			
<p>$A_p = \frac{\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_j - \bar{x}}{s} \right)^4}{\frac{3(n-1)^2}{(n-2)(n-3)}}$</p>	<p>Ap >0 Leptocurtica Ap <0 Platicurtica Ap =0 Mesocurtica</p>	<p>$A_p = \frac{\sum(x - \bar{x})^4 f}{n \sigma^4}$</p> <p>$\sigma^4 = \left(\frac{\sum(x - \bar{x})^2 f}{n} \right)^2$</p>	<p>Ap >3 Leptocurtica Ap <3 Platicurtica Ap =3 Mesocurtica</p>
Asimetría			
<p>$A_s = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_j - \bar{x}}{s} \right)^3$</p>	<p>As >0 Asimetría "+" As <0 Asimetría "-" As = 0 Simétrica</p>	<p>Pearson $A_s = \frac{3(\bar{x} - \tilde{x})}{s}$ (media, mediana, desv est muestral)</p> <p>Bowley $A_s = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$</p> <p>Fischer $A_s = \frac{\sum(x - \bar{x})^3 f}{n \sigma^3}$</p>	<p>As >0 Asimetría "+" As <0 Asimetría "-" As = 0 Simétrica</p>
Coefficiente de Varianza			
<p>c. v = $\frac{s}{\bar{x}} \times 100$</p> <p>Desviación Estándar, media</p>		<p>c. v = $\frac{s}{\bar{x}} \times 100$</p> <p>Desviación Estándar, media</p>	
		<p>0 ≤ C.V < 20 % Muy confiable 20 ≤ C.V < 30 % Confiable 30 ≤ C.V < 40 % Relat. Confiable 40 % ↑ No confiable</p>	