

Formulario de Estadística Descriptiva

1. Organización de Datos (Agrupación)

Formularios usados para convertir datos individuales en una tabla de intervalos.

- *Rango (R)*
 - *Fórmula:* $R = X_{\max} - X_{\min}$
 - *Cuándo se usa:* Siempre, al inicio del proceso de agrupar datos.
 - *Descripción:* Mide la diferencia entre el valor más alto y el más bajo de tus datos.
 - *Número de Intervalos (k)*
 - *Fórmula:* $k \approx \sqrt{n}$ (Regla sugerida)
 - *Cuándo se usa:* Para decidir en cuántos grupos (filas) dividirás tus datos.
 - *Descripción:* n es el número total de datos. El resultado se redondea a un entero (ej. $\sqrt{54} \approx 7.34 \rightarrow k = 7$).
 - *Amplitud (A)*
 - *Fórmula:* $A = R/k$
 - *Cuándo se usa:* Para definir el ancho de cada intervalo de clase.
 - *Descripción:* Se divide el Rango entre el número de intervalos. Es crucial redondear este valor hacia arriba para asegurar que el último intervalo cubra el X_{\max} .
 - *Marca de Clase (m_i)*
 - *Fórmula:* $m_i = (L_i + L_{i+1})/2$
 - *Cuándo se usa:* En datos agrupados, para calcular la media (\bar{x}) y la desviación estándar (s).
 - *Descripción:* Es el punto medio de un intervalo. Actúa como el valor representativo de todos los datos que cayeron en ese intervalo.
-

2. Medidas de Tendencia Central

- *Media (\bar{x}) - Datos No Agrupados*
 - *Fórmula:* $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
 - *Cuándo se usa:* Para datos individuales (ej. listas de números, como en el ejercicio 33 o 11).
 - *Descripción:* El promedio simple. Suma todos los valores y divide por la cantidad de datos.
- *Media (\bar{x}) - Datos Agrupados*
 - *Fórmula:* $\bar{x} = \frac{1}{n} \sum_{i=1}^k (m_i \cdot f_i)$

- *Cuándo se usa:* Cuando tienes una tabla de frecuencias con intervalos (ej. ejercicio 20).
- *Descripción:* Multiplica cada marca de clase (m_i) por su frecuencia (f_i), suma esos resultados y divide por el total de datos (n).
- *Mediana (Me) - Datos No Agrupados*
 - *Fórmula:* (Método) [386-392]
 - *Cuándo se usa:* Para datos individuales.
 - *Descripción:* Ordena los datos de menor a mayor.
 - * Si n es *impar*: La mediana es el valor en la posición $\frac{n+1}{2}$.
 - * Si n es *par*: La mediana es el promedio de los dos valores centrales (posiciones $\frac{n}{2}$ y $\frac{n+2}{2}$).
- *Mediana (Me) - Datos Agrupados*
 - *Fórmula:* $Me = L_{i-1} + \left(\frac{\frac{n}{2} - F_{i-1}}{f_i} \right) \cdot A$
 - *Cuándo se usa:* Cuando buscas la mediana en una tabla de intervalos.
 - *Descripción:* Es la fórmula del Percentil 50. L_{i-1} es el límite inferior de la clase mediana, F_{i-1} es la frecuencia acumulada anterior, y f_i es la frecuencia de la clase mediana.

3. Medidas de Posición (Cuantiles)

- *Percentiles (P_k) - Datos No Agrupados*
 - *Fórmula (Método):* [419-431]
 - *Cuándo se usa:* Para calcular Q_1 , Q_3 , Deciles, etc., en datos individuales (ej. ejercicio 33, 40).
 - *Descripción:*
 - Calcula la posición: $\frac{nk}{100}$.
 - Si el resultado es *decimal* (ej. 2.25): Se usa la regla $r > 0$. El valor es el de la posición $j + 1$ (ej. x_3).
 - Si el resultado es *entero* (ej. 42): Se usa la regla $r = 0$. El valor es el promedio de las posiciones j y $j + 1$ (ej. $(x_{42} + x_{43})/2$).
- *Percentiles (P_k) - Datos Agrupados (Fórmula Maestra)*
 - *Fórmula:* $P_k = L_{i-1} + \left(\frac{\frac{nk}{100} - F_{i-1}}{f_i} \right) \cdot A$
 - *Cuándo se usa:* Para calcular cualquier cuantil en una tabla de intervalos (ej. 15, 18, 20).
 - *Descripción:* Es la fórmula de interpolación.
 - * Para Q_1 (Cuartil 1), usa $k = 25$.
 - * Para Q_3 (Cuartil 3), usa $k = 75$.
 - * Para D_7 (Decil 7), usa $k = 70$.

4. Medidas de Dispersión

- *Varianza Muestral (s^2) - Datos No Agrupados*
 - *Fórmula:* $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ [681]
 - *Cuándo se usa:* Para datos individuales (ej. ejercicio 33).
 - *Descripción:* Mide la dispersión promedio al cuadrado. Se divide por $(n - 1)$ por ser una muestra.
 - *Varianza Muestral (s^2) - Datos Agrupados*
 - *Fórmula:* $s^2 = \frac{1}{n-1} \sum_{i=1}^k (m_i - \bar{x})^2 \cdot f_i$ [742]
 - *Cuándo se usa:* Para datos en intervalos (ej. 20).
 - *Descripción:* Idéntica a la no agrupada, pero usa la marca de clase (m_i) y pondera por la frecuencia (f_i).
 - *Desviación Estándar (s)*
 - *Fórmula:* $s = \sqrt{s^2}$ [670]
 - *Cuándo se usa:* Siempre.
 - *Descripción:* Es la raíz cuadrada de la varianza. Regresa la medida de dispersión a las unidades originales (ej. "mm" en lugar de "mm²").
-

5. Diagrama de Caja y Datos Atípicos

- *Rango Intercuartil (RIQ)*
 - *Fórmula:* $RIQ = Q_3 - Q_1$
 - *Cuándo se usa:* Para construir los límites del diagrama de caja.
 - *Descripción:* Mide la dispersión del 50% central de los datos.
 - *Límites para Datos Atípicos (L_i, L_s)*
 - *Fórmula (L_i):* $L_i = Q_1 - 1.5 \cdot RIQ$
 - *Fórmula (L_s):* $L_s = Q_3 + 1.5 \cdot RIQ$
 - *Cuándo se usa:* Para definir las "cercas" e identificar datos atípicos (ej. 33).
 - *Descripción:* *No son los bigotes.* Son umbrales invisibles. Cualquier dato fuera de $[L_i, L_s]$ es un dato atípico. Los bigotes se dibujan hasta el último dato no atípico.
-

6. Aplicaciones (Interpolación y Extrapolación)

- *Interpolación (para Conteo N_{pos})*
 - *Fórmula:* $N_{pos} = F_{i-1} + f_i \cdot \left(\frac{P_k - L_{i-1}}{A} \right)$

- *Cuándo se usa:* Cuando tienes un valor (P_k) y quieres saber cuántos datos hay por debajo de él (ej. 18, 20).
- *Descripción:* Es la fórmula del percentil P_k "al revés", despejada para encontrar la frecuencia acumulada (N_{pos}). F_{i-1} es la acumulación de los intervalos completos y la fracción es la porción del intervalo parcial.
- *Extrapolación (Estimación Poblacional)*
 - *Fórmula:* Estimación = $\left(\frac{f_i}{n}\right) \cdot N_{\text{población}}$
 - *Cuándo se usa:* Cuando tienes una muestra (n) y quieres estimar un conteo en la población ($N_{\text{población}}$) (ej. 10, 18, 20).
 - *Descripción:* Se asume que la muestra es representativa. La frecuencia relativa ($\frac{f_i}{n}$) de la muestra se aplica al total de la población.