

Distribución de Frecuencias para Datos Agrupados en Intervalos de Clases

Fundamentos Teóricos, Metodología de Construcción y Aplicaciones Prácticas

Institución: IUTEPI | Asignatura: Estadística Descriptiva | Unidad Curricular: Análisis e Interpretación de Datos

1. Introducción y Justificación Teórica

En el estudio de la estadística descriptiva, cuando trabajamos con conjuntos de datos cuantitativos masivos (donde el número de observaciones N es grande o la cantidad de valores distintos es muy alta), el análisis individualizado de cada dato se vuelve impracticable y carente de claridad visual. Para resolver este inconveniente, se recurre a la técnica de **agrupación de datos en intervalos de clases**.

Definición fundamental: Consiste en organizar un conjunto de datos numéricos dividiéndolos en subintervalos o clases contiguas y mutuamente excluyentes, permitiendo resumir la información mediante frecuencias que muestran cómo se distribuyen los valores a lo largo de la escala de medición.

Esta metodología facilita la identificación de patrones de comportamiento, tendencias centrales, dispersión y la posterior construcción de representaciones gráficas avanzadas (como histogramas, polígonos de frecuencia y ojivas).

2. Conceptos Clave y Terminología Técnica

Para construir una distribución de frecuencias para datos agrupados de manera rigurosa, es indispensable dominar los siguientes elementos y parámetros estadísticos:

- **Rango (R):** Es la medida de dispersión más simple. Representa la diferencia aritmética entre el valor máximo ($X_{\{max\}}$) y el valor mínimo ($X_{\{min\}}$) del conjunto de datos. Su fórmula es:

$$R = X_{\{max\}} - X_{\{min\}}$$

- **Número de Clases (k):** Indica cuántos intervalos se van a formar en la tabla. Un número muy pequeño de clases oculta detalles importantes; uno muy excesivo fragmenta demasiado la información. El método estándar universalmente aceptado es la **Regla de Sturges**:

$$k = 1 + 3.322 \log(N)$$

Donde N es el número total de observaciones. El resultado decimal obtenido se redondea al número entero más próximo.

- **Amplitud o Anchura del Intervalo (C):** Es la longitud de cada intervalo de clase. Se calcula dividiendo el rango entre el número de clases:

$$C = \frac{R}{k}$$

Si el cociente resulta exacto, puede utilizarse directamente; si tiene decimales, por norma práctica se redondea por exceso al entero inmediato superior para garantizar que todos los datos queden plenamente cubiertos dentro de los límites.

- **Límites de Clase (L_i y L_s):** Son los valores extremos que delimitan cada intervalo. Se definen el *Límite Inferior* (L_i) y el *Límite Superior* (L_s). Generalmente se trabaja con intervalos semiabiertos de la forma $[L_i - L_s)$, donde el límite inferior está incluido y el superior no.
- **Marca de Clase (x_i):** Es el punto medio de cada intervalo de clase y representa a todos los valores comprendidos en él. Se calcula mediante la semisuma de sus límites:

$$x_i = \frac{L_i + L_s}{2}$$

3. Tipos de Frecuencias en Datos Agrupados

Una vez definidos los intervalos de clase y sus marcas, se calculan las distintas columnas de frecuencias que componen la tabla estadística:

- **Frecuencia Absoluta (f_i):** Número exacto de observaciones (datos) que caen dentro de un determinado intervalo de clase. La suma de todas las frecuencias absolutas debe ser igual al tamaño total de la muestra ($\sum f_i = N$).
- **Frecuencia Absoluta Acumulada (F_i):** Suma sucesiva de las frecuencias absolutas desde la primera clase hasta la clase en consideración ($F_i = F_{i-1} + f_i$). Representa cuántos datos son menores o iguales al límite superior de dicha clase.
- **Frecuencia Relativa (h_i):** Cociente entre la frecuencia absoluta de una clase y el tamaño total de la muestra:

$$h_i = \frac{f_i}{N}$$

La suma de todas las frecuencias relativas es siempre igual a 1 ($\sum h_i = 1$).

- **Frecuencia Porcentual (%):** Es la frecuencia relativa expresada en porcentaje multiplicándola por 100 ($\%_i = h_i \times 100$). Su sumatoria total es 100%.

4. Metodología Paso a Paso para la Construcción de la Tabla

El proceso operativo para llevar un conjunto bruto de datos desordenados a una tabla de distribución de frecuencias completa en el ámbito del IUTEPI comprende los siguientes pasos secuenciales:

1. **Recopilación y Ordenamiento:** Obtener los datos estadísticos de la muestra u población bajo estudio y ordenarlos de menor a mayor para identificar visualmente $X_{\{min\}}$ y $X_{\{max\}}$.
2. **Cálculo del Rango (R):** Sustituir los valores extremos en la fórmula $R = X_{\{max\}} - X_{\{min\}}$.
3. **Determinación del número de clases (k):** Aplicar la regla de Sturges $k = 1 + 3.322 \log(N)$ y redondear al entero conveniente.
4. **Cálculo de la amplitud (C):** Dividir R/k y ajustar por exceso si fuese necesario.
5. **Formación de los Intervalos:** Comenzar el primer intervalo con $L_1 = X_{\{min\}}$ y sumar C sucesivamente para hallar los límites inferiores y superiores de cada clase subsiguiente.
6. **Conteo y Registro de Frecuencias:** Proceder al conteo riguroso de los valores originales que pertenecen a cada intervalo para asentar las frecuencias absolutas (f_i).
7. **Cálculo de Marcas de Clase y Acumuladas:** Completar las columnas de marcas de clase (x_i), frecuencias acumuladas (F_i), relativas (h_i) y porcentuales.

5. Ejemplo Práctico Integral

Planteamiento del problema: Se registraron las calificaciones obtenidas por 30 estudiantes en la evaluación final de estadística en el IUTEPI. Los datos ordenados son los siguientes:

32, 35, 40, 42, 45, 48, 50, 52, 53, 55, 58, 60, 62, 63, 65, 66, 68, 70, 72, 74, 75, 78, 80, 82, 85, 88, 90, 92, 95, 98

Desarrollo de parámetros:

- Tamaño de muestra: $N = 30$
- Valor mínimo: $X_{\{min\}} = 32$ | Valor máximo: $X_{\{max\}} = 98$
- Rango: $R = 98 - 32 = 66$
- Número de clases (Sturges): $k = 1 + 3.322 \log(30) = 1 + 3.322(1.4771) = 5.90$ *pprox 6* clases.
- Amplitud: $C = \frac{66}{6} = 11$ unidades por intervalo.

Tabla de Distribución de Frecuencias Resultante:

Clase (i)	Intervalos	Marca de Clase (x_i)	Absoluta (f_i)	Acumulada (F_i)	Relativa (h_i)	Porcentual (%)
1	[32 - 43)	37.5	4	4	0.133	13.3%
2	[43 - 54)	48.5	5	9	0.167	16.7%
3	[54 - 65)	59.5	6	15	0.200	20.0%
4	[65 - 76)	70.5	6	21	0.200	20.0%
5	[76 - 87)	81.5	4	25	0.133	13.3%
6	[87 - 98]	92.5	5	30	0.167	16.7%
Totales:			30	—	1.00	100%

Interpretación analítica: A partir de la tabla construida, es posible observar con gran precisión que el 20% de los estudiantes se concentra en los intervalos centrales de calificaciones [54-65) y [65-76), evidenciando el rendimiento promedio del grupo evaluado en la institución.