

# Histograma y Tabla de Frecuencias

Al analizar un conjunto de datos, lo fundamental es conocer su tendencia central y su dispersión. A continuación, veremos dos herramientas que nos permiten visualizar ambos elementos y a la vez mostrarnos cómo se distribuyen los datos dentro de su rango de variación: **el *histograma* y la *tabla de frecuencias*.**

Un *histograma* es una representación gráfica, en forma de barras, que muestra la distribución de un grupo de datos o una variable, cuyos datos se clasifican por su magnitud en cierto número de grupos o clases, y cada clase se representa con una barra, cuya longitud es proporcional a la frecuencia de los valores representados. En la gráfica, el eje horizontal está integrado por una escala numérica para mostrar la magnitud de los datos; y en el eje vertical se representan las frecuencias.

Generalmente, el histograma se obtiene a partir de la *tabla de frecuencias*. Para lograr obtener la tabla de frecuencia, lo primero que haremos es dividir el rango de variación de los datos en cierta cantidad de intervalos que abarquen todo el rango, y al final se determina cuántos datos caen en cada intervalo. Se recomienda que el número de intervalos o clases sea de 5 a 15.

Para elegir un valor entre este rango, podemos utilizar varios métodos, uno de ellos menciona que el número de clases debe ser aproximadamente igual a la raíz cuadrada del número de datos.

Existe otro método llamado la regla de *Sturges*, el cual nos indica que el número de clases se obtiene mediante la siguiente fórmula:

$$1 + 3.3 * \log_{10} (\text{número total de datos})$$

En la tabla 2.2 tenemos la tabla de frecuencias para los 125 datos del grosor de los discos de la tabla del ejemplo 2.1. Podemos ver que al aplicar la regla de Sturges obtenemos lo

# Histograma y Tabla de Frecuencias

siguiente:  $(1 + 3.3 * \log_{10}(125) = 7.9)$ , por lo cual se decidió formar ocho clases; la primera clase representa a los datos con magnitud entre 1.10 y 1.12, y la última clase es para los datos entre 1.24 y 1.26.

En el intervalo de la primera clase hay tres datos que corresponden a 2.4% del total; la clase 5 es la de mayor frecuencia e indica que entre 1.18 y 1.20 hay 39 datos (31.2%).

**TABLA 2.1** Datos para el grosor de los discos, ejemplo 2.1.

1.15	1.20	1.17	1.16	1.16	1.15	1.17	1.20	1.16	1.19	1.17	1.13	1.15	1.20	1.18	1.17	1.16
1.20	1.17	1.17	1.20	1.14	1.19	1.13	1.19	1.16	1.18	1.16	1.17	1.15	1.21	1.15	1.20	1.18
1.17	1.17	1.13	1.16	1.16	1.17	1.20	1.18	1.15	1.13	1.20	1.17	1.19	1.23	1.20	1.24	1.17
1.17	1.17	1.17	1.18	1.24	1.16	1.18	1.16	1.22	1.23	1.22	1.19	1.13	1.15	1.15	1.22	1.19
1.18	1.19	1.17	1.16	1.17	1.18	1.19	1.23	1.19	1.16	1.19	1.20	1.17	1.13	1.22	1.19	1.21
1.20	1.19	1.17	1.19	1.22	1.19	1.18	1.11	1.19	1.19	1.17	1.19	1.17	1.20	1.16	1.19	1.20
1.20	1.17	1.25	1.16	1.16	1.20	1.20	1.16	1.18	1.21	1.20	1.22	1.19	1.14	1.19	1.17	1.20
1.16	1.15	1.20	1.12	1.11	1.18											

**TABLA 2.2** Tabla de frecuencia para el grosor de los discos.

CLASE	GROSOR DE DISCOS, X	MARCAS PARA CONTEO	FRECUENCIA	FRECUENCIA PORCENTUAL
1	$1.10 < x \leq 1.12$	///	3	2.4
2	$1.12 < x \leq 1.14$	///// ///	8	6.4
3	$1.14 < x \leq 1.16$	///// ///// ///// ///// /////	26	20.8
4	$1.16 < x \leq 1.18$	///// ///// ///// ///// ///// ///// /////	34	27.2
5	$1.18 < x \leq 1.20$	///// ///// ///// ///// ///// ///// ///// /////	39	31.2
6	$1.20 < x \leq 1.22$	///// ///	9	7.2
7	$1.22 < x \leq 1.24$	/////	5	4.0
8	$1.24 < x \leq 1.26$	/	1	0.8

# Histograma y Tabla de Frecuencias

Una vez que tenemos la tabla de frecuencia veremos cómo queda el histograma, en la figura 2.1 se muestra el histograma correspondiente, en el cual tenemos como eje vertical a la frecuencia (la cual también puede ser relativa o porcentual). Podemos observar que la tendencia central de los datos se ubica alrededor de 1.18, no se observan datos raros o atípicos y la distribución de los datos tiene una forma similar a una campana.

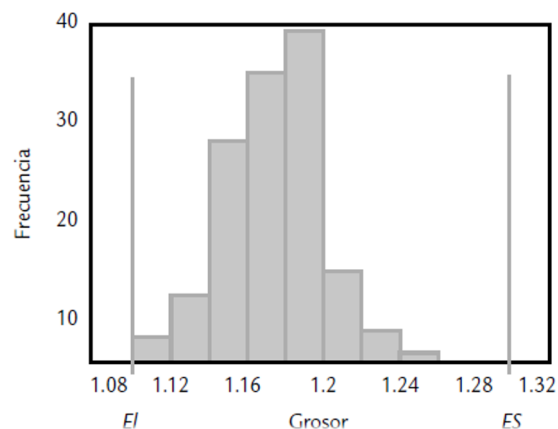


Figura 2.1. Histograma para grosor de discos

Si en este histograma agregáramos las especificaciones (1.10 y 1.30) para el grosor del disco, podríamos ver que la variación de los datos (amplitud del histograma) es un poco menor que las especificaciones. Pero, debido a que el grosor óptimo es de 1.20, podemos observar que está moderadamente descentrado a la izquierda, como ya se había visto cuando se calculó la media. Aparte, el grosor de los discos no es satisfactorio, ya que la orilla izquierda del histograma debería estar alejada de la especificación inferior ( $E_l=1.10$ ), lo cual no ocurre. Vale la pena mencionar que, aunque no hay ningún dato por debajo de la  $E_l$ , no debemos olvidar que el estudio se hace a partir de una muestra, por lo tanto, si se continúa tomando datos, es casi seguro que se encontrarán mediciones fuera, como lo sugiere la prolongación de la cola izquierda de la curva imaginaria que suaviza al

# Histograma y Tabla de Frecuencias

histograma. De acuerdo con lo anterior, la primera acción que se habría de tomar para mejorar la capacidad del proceso de inyección de discos es mejorar su centrado.

Con este ejemplo podemos entender que el histograma ayuda a ver la tendencia central de los datos, facilita el entendimiento de la variabilidad y favorece el pensamiento estadístico, ya que con un solo vistazo podemos obtener una idea acerca de la capacidad de un proceso, se evitan tomar decisiones solo apoyándose en la media y se detectan datos raros y formas especiales de la distribución de los datos. En el siguiente tema, veremos la interpretación del histograma con mayor profundidad.

## REFERENCIA:

Gutiérrez, H. y De la Vara, R. (2009). Control estadístico de calidad y seis sigma. Recuperado de: <https://www.uv.mx/personal/ermeneses/files/2018/05/6-control-estadistico-de-la-calidad-y-seis-sigma-gutierrez-2da.pdf>