Organización y representación de datos

Hasta aquí hemos estudiado diferentes medidas estadísticas que permiten resumir las características de un conjunto de observaciones.

La media y la mediana permiten identificar valores representativos; la desviación estándar describe la dispersión de los datos; y los cuartiles, deciles y percentiles permiten reconocer posiciones dentro de un conjunto ordenado.

Estas herramientas forman parte de la **estadística descriptiva**, cuyo propósito es organizar, representar y resumir los datos disponibles para facilitar su comprensión e interpretación.

La estadística descriptiva puede aplicarse tanto a una población completa como a una muestra. Cuando trabajamos con una muestra, las medidas calculadas describen directamente los datos observados; para extender las conclusiones a toda la población necesitamos herramientas adicionales, propias de la estadística inferencial.

Sin embargo, resumir los datos mediante medidas numéricas no siempre es suficiente. También necesitamos conocer cómo se distribuyen las observaciones: qué valores aparecen con mayor frecuencia, cuáles son menos habituales y cómo varían dentro del conjunto.

Para ello utilizamos **tablas de frecuencias y representaciones gráficas**, que complementan las medidas estadísticas estudiadas.

Distribución de Frecuencias

Cuando contamos con un conjunto de observaciones, una primera tarea consiste en organizar los datos para facilitar su interpretación.

Una distribución de frecuencias muestra cuántas veces aparece cada valor o categoría, o cuántas observaciones pertenecen a determinados intervalos.

Para construirla, utilizamos principalmente tres tipos de frecuencias:

- **Frecuencia absoluta (f_i)**: cantidad de observaciones que corresponden a un determinado valor, categoría o intervalo.
- **Frecuencia relativa (h_i)**: proporción de observaciones que corresponde a ese valor, categoría o intervalo, respecto del total.
- **Frecuencia porcentual**: frecuencia relativa expresada como porcentaje.

La frecuencia relativa se calcula mediante:

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

Donde (f_i) representa la frecuencia absoluta y (n) la cantidad total de observaciones.

La frecuencia porcentual se obtiene multiplicando la frecuencia relativa por 100:

$$p_i = h_i \times 100$$

Ejemplo: clasificación de materiales en una operación minera

Supongamos que se analizan veinte muestras de material y se clasifican según su tipo:

Tipo de material	Frecuencia absoluta	Frecuencia relativa	Frecuencia porcentual
Mineral sulfurado	8	0,40	40 %
Mineral oxidado	7	0,35	35 %
Material estéril	5	0,25	25 %
Total	**20**	**1,00**	**100 %**

Por ejemplo, para el mineral sulfurado:

$$h_1 = \frac{f_1}{n} = \frac{8}{20} = 0{,}40$$

$$p_1 = h_1 \times 100 = 0{,}40 \times 100 = 40\%$$

Interpretación: el 40 % de las muestras analizadas corresponde a mineral sulfurado.

Esta tabla permite conocer rápidamente la composición del conjunto de observaciones según las categorías consideradas.

Representaciones gráficas

Las tablas de frecuencias organizan los datos numéricamente. Sin embargo, muchas veces resulta más sencillo reconocer patrones, diferencias y concentraciones mediante gráficos.

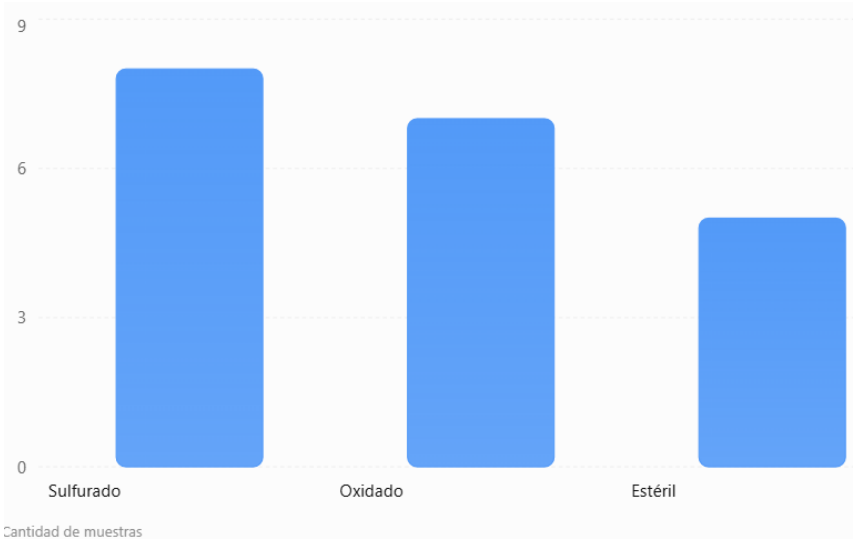
Existen diferentes representaciones gráficas, cuya elección depende del tipo de variable que estamos estudiando.

Gráfico de barras

El gráfico de barras permite representar las frecuencias de distintas categorías o valores discretos.

Cada barra corresponde a una categoría y su altura representa la frecuencia observada.

Por ejemplo, podemos representar gráficamente la cantidad de muestras de mineral sulfurado, mineral oxidado y material estéril del ejemplo anterior.



Interpretación: el gráfico permite identificar visualmente que el mineral sulfurado es la categoría más frecuente entre las veinte muestras analizadas.

En los gráficos de barras, las categorías son independientes entre sí y las barras suelen representarse separadas.

Histograma

Cuando estudiamos una variable cuantitativa continua, como la ley de cobre, el tiempo de carga de un camión o el tamaño de las partículas de mineral, puede resultar conveniente agrupar las observaciones en intervalos.

Un **histograma** representa gráficamente la distribución de frecuencias de una variable cuantitativa agrupada en intervalos.

A diferencia del gráfico de barras, los intervalos del histograma son contiguos, porque representan partes consecutivas de una escala numérica.

Ejemplo: distribución de leyes de cobre

Supongamos que se analizaron veinte muestras de mineral y se obtuvieron las siguientes leyes de cobre, expresadas en porcentaje:

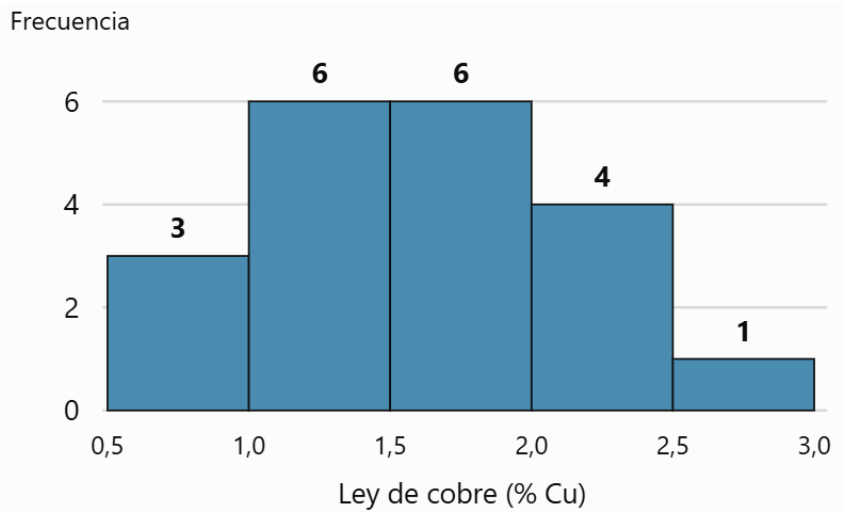
0,6; 0,8; 0,9; 1,0; 1,1; 1,2; 1,3; 1,4; 1,4; 1,5; 1,6; 1,6; 1,7; 1,8; 1,9; 2,0; 2,1; 2,2; 2,4; 2,7.

Para organizar los datos, establecemos intervalos de amplitud 0,5 puntos porcentuales.

Intervalo de ley (% Cu)	**Frecuencia absoluta**	**Frecuencia relativa**	**Porcentaje**
[0,5; 1,0)	3	0,15	15 %
[1,0; 1,5)	6	0,30	30 %
[1,5; 2,0)	6	0,30	30 %
[2,0; 2,5)	4	0,20	20 %
[2,5; 3,0)	1	0,05	5 %
Total	**20**	**1,00**	**100 %**

La notación [0,5; 1,0) indica que el intervalo incluye 0,5, pero no incluye 1,0. De esta manera, una observación ubicada exactamente en un límite pertenece a un único intervalo.

Al representar estas frecuencias mediante un histograma, podemos observar cómo se distribuyen las leyes de cobre.



En este caso, las mayores frecuencias se encuentran entre 1,0 % y 2,0 % Cu, mientras que las leyes superiores a 2,5 % Cu son menos frecuentes.

El histograma permite reconocer estas características de manera visual, complementando la información proporcionada por la media, la mediana y la desviación estándar. Las barras son contiguas porque representan intervalos consecutivos.

¿Qué ocurre cuando aumenta la cantidad de observaciones?

Otra cuestión importante es la cantidad de datos disponibles.

Con pocas observaciones, un histograma puede presentar una forma irregular y resultar difícil identificar patrones.

Cuando disponemos de una mayor cantidad de observaciones representativas del mismo fenómeno, podemos construir distribuciones de frecuencias más detalladas y, en determinadas condiciones, reconocer patrones con mayor claridad.

Sin embargo, disponer de más datos no garantiza por sí solo una representación adecuada: también importa cómo se obtuvieron las observaciones y si representan el fenómeno que queremos estudiar.

En Ingeniería de Minas, esto resulta especialmente relevante porque las muestras deben permitir caracterizar adecuadamente el sector del yacimiento o el proceso que se está analizando.

Frecuencia relativa y una primera aproximación a la probabilidad

Hasta aquí hemos utilizado las frecuencias relativas para describir qué proporción de observaciones corresponde a determinados valores o categorías.

Pero también podemos preguntarnos qué sucederá en una nueva observación.

Por ejemplo, si en veinte muestras de mineral encontramos cuatro con leyes superiores o iguales al 2,0 % Cu, la frecuencia relativa observada es:

$$h = \frac{4}{20} = 0{,}20$$

Esto significa que el 20 % de las muestras analizadas presenta una ley de cobre mayor o igual al 2,0 %.

¿Podemos afirmar entonces que existe exactamente un 20 % de probabilidad de encontrar esa ley en cualquier punto del yacimiento?

No necesariamente.

La frecuencia relativa describe lo que ocurrió en las muestras observadas. Para utilizarla como estimación de una probabilidad necesitamos considerar cómo fueron seleccionadas las muestras y qué población queremos estudiar.

Una idea fundamental de la probabilidad es que, cuando repetimos numerosas veces un experimento aleatorio bajo condiciones comparables e independientes,

la frecuencia relativa de un determinado resultado tiende a aproximarse a su probabilidad.

Por ejemplo, al lanzar repetidamente un dado equilibrado, la proporción de veces que aparece el número cinco tiende a aproximarse a:

$$P(5) = \frac{1}{6} \approx 0{,}167$$

Es decir, aproximadamente el 16,7 % de los lanzamientos.

Esta relación permite conectar las frecuencias observadas con el concepto de probabilidad y comprender por qué resulta importante disponer de datos adecuados para estudiar fenómenos variables.

En el ámbito minero, las observaciones pueden utilizarse para estudiar la frecuencia de determinados eventos o características, pero la interpretación probabilística requiere considerar las condiciones de muestreo y el fenómeno analizado.

Síntesis: de los datos a la información estadística

El análisis descriptivo comienza con un conjunto de observaciones.

Estas observaciones pueden organizarse mediante tablas de frecuencias, representarse gráficamente y resumirse utilizando medidas estadísticas.

Cada herramienta aporta información diferente:

Herramienta	**¿Qué permite conocer?**
Tabla de frecuencias	Cuántas observaciones corresponden a cada valor, categoría o intervalo.
Gráfico de barras	Comparar visualmente frecuencias de categorías o valores discretos.
Histograma	Reconocer cómo se distribuyen los valores de una variable cuantitativa.
Media y mediana	Identificar valores representativos del conjunto.
Desviación estándar	Describir la dispersión de las observaciones alrededor de la media.

Cuartiles, deciles y percentiles	Identificar posiciones dentro de la distribución ordenada.

En conjunto, estas herramientas permiten transformar datos individuales en información organizada y comprensible.

La estadística descriptiva no reemplaza el conocimiento técnico del fenómeno estudiado, sino que proporciona herramientas para analizarlo e interpretarlo de manera fundamentada.

En Ingeniería de Minas, esta información puede contribuir a la caracterización de materiales, el análisis de procesos y la toma de decisiones técnicas, siempre considerando las condiciones en que se obtuvieron los datos y las limitaciones de las conclusiones.