Conceptos básicos

Todo empieza haciendo observaciones... Medimos la ley de cobre de una muestra de mineral extraída de un yacimiento. Es una observación. Podemos pensar que una observación es el encuentro de una variable (la ley de cobre del mineral) con una muestra determinada.

La ley de cobre es la cantidad de cobre que contiene una determinada cantidad de material mineralizado, generalmente expresada como un porcentaje de su masa. En Ingeniería de Minas, el término ley se utiliza para indicar la concentración de un elemento de interés económico en un material.

Ejemplo

Supongamos que extraemos una muestra de mineral de 100 kg y, mediante un análisis de laboratorio, determinamos que contiene 1,5 kg de cobre.

La ley de cobre se calcula como:

$$\text{Ley de cobre} = \frac{\text{Masa de cobre}}{\text{Masa total de la muestra}} \times 100$$

$$\text{Ley de cobre} = \frac{1,5}{100} \times 100 = \boxed{1,5\%}$$

Esto significa que el 1,5 % de la masa de la muestra corresponde a cobre.

¿Por qué es importante?

Porque permite caracterizar la concentración del mineral de interés y contribuye a evaluar un yacimiento.

Por ejemplo:

Una muestra con 0,5 % Cu contiene 5 kg de cobre por tonelada de material.

Una muestra con 1,0 % Cu contiene 10 kg de cobre por tonelada.

Una muestra con 2,0 % Cu contiene 20 kg de cobre por tonelada.

Una ley más alta significa mayor concentración de cobre, aunque no implica automáticamente que su extracción sea económicamente rentable, porque también intervienen los costos, la recuperación metalúrgica y otros factores.

Del mismo modo, medir el tiempo de operación de un camión minero es observar una variable (tiempo de operación) durante un período determinado (por ejemplo, un turno de trabajo).

Llamaremos **universo o población** al conjunto completo de elementos u observaciones de interés, definido de acuerdo con el problema que queremos estudiar.

Para evitar sorpresas desagradables hay que señalar que existe un margen de arbitrariedad en la definición de un universo, ya que bien podría ser:

- Todas las mediciones de ley de cobre realizadas en un sector del yacimiento durante una campaña de exploración.
- Todas las mediciones de ley de cobre realizadas en ese sector a lo largo de su explotación.

Ambas son alternativas posibles. En cualquier caso, conviene siempre aclarar, frente a un problema, de qué universo se está hablando.

Nota para los estudiantes de Ingeniería de Minas:

Muchas veces el universo incluye observaciones que están en el futuro, como las mediciones de producción que se realizarán durante los próximos meses. Eso hace que cualquier atributo del universo completo sea desconocido en el presente.

Ahora introduciremos otro concepto: Una **muestra** es un subconjunto de elementos u observaciones de esa población o universo, seleccionado para su estudio.

Finalmente nos queda considerar el concepto de **variable aleatoria**.

Podemos pensar en una variable aleatoria como en una fábrica de observaciones.

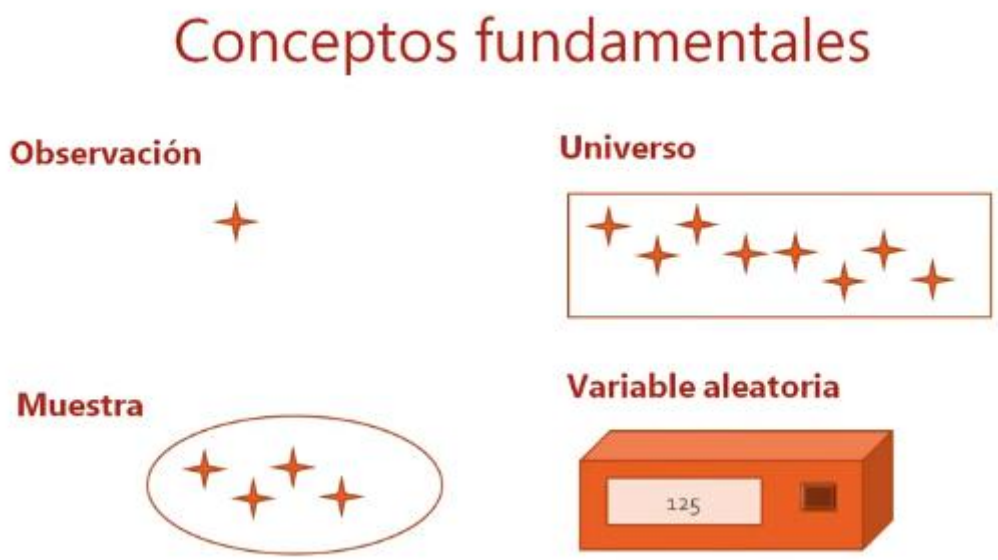
Otra forma de conceptualizar la variable aleatoria es como una caja opaca con un visor y un botón. Cada vez que se presiona el botón se obtiene un nuevo resultado en el visor (una nueva observación).



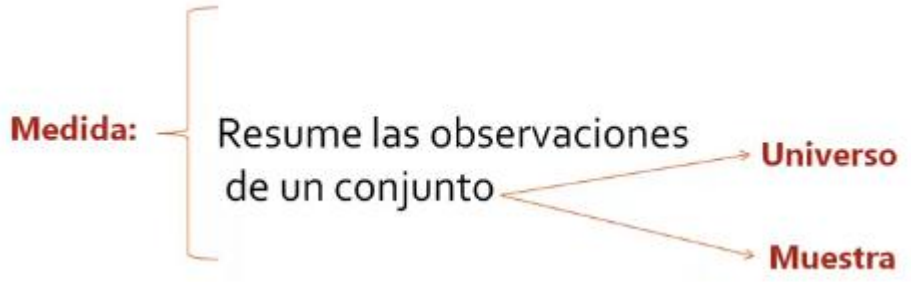
Por ejemplo, cada vez que analizamos una nueva muestra de mineral, obtenemos un valor de ley de cobre.

¿Y de dónde sale el resultado del visor? De un mecanismo cuyo comportamiento puede describirse mediante una distribución de probabilidad. Investigar ese comportamiento puede ser relevante para nosotros.

Formalmente, una variable aleatoria es un modelo matemático del resultado numérico de un fenómeno incierto.



Vamos a llamar una **medida** a cualquier resultado que calculamos sobre los datos de un conjunto para resumir nuestro conocimiento sobre dicho conjunto.



Seguramente ya conocen varias medidas como:

- Promedio
- Mediana
- Desviación estándar

Las medidas que resumen el conocimiento del universo las llamamos **medidas universales**:

- Promedio universal (habitualmente notado con la letra griega μ).
- Mediana universal.

- Desviación estándar universal (habitualmente notado con la letra griega σ).

A las medidas que resumen el conocimiento de una muestra las llamamos **medidas muestrales**:

- Promedio muestral (habitualmente notado con ($\bar{X}$).
- Mediana muestral.
- Desviación estándar muestral (habitualmente notado con la letra (S)).

	Muestra	Universo o población
Promedio	$\bar{x} = \frac{\sum x_i}{n}$	$\mu = \frac{\sum x_i}{N}$
Mediana	$M_m = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & n\ es\ impar \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & n\ es\ par \end{cases}$	$M_p = \begin{cases} x_{\left(\frac{N+1}{2}\right)}, & N\ es\ impar \\ \frac{x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)}}{2}, & N\ es\ par \end{cases}$
Desviación Estándar	$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$	$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$

Donde:

- *N* es el número de elementos del universo y *n* es el número de elementos de la muestra.
- Los (x_i) son los distintos elementos del conjunto.
- Importante: los paréntesis en $x_{(i)}$ indican una posición dentro del conjunto ordenado, no el número de identificación original de la observación

Estas medidas describen características diferentes de un mismo conjunto de datos, no deben interpretarse como fórmulas aisladas

Nota para los estudiantes de Ingeniería de Minas:

Una operación que naturalmente hacemos los humanos es reemplazar al universo por la repetición del promedio. Eso nos lleva a conclusiones incorrectas. Debemos estar en guardia contra esa tendencia.

Por ejemplo, si la ley media de cobre de un conjunto de muestras es del 1,5 %, no significa que todas las muestras del sector mineralizado tengan exactamente esa concentración.

Ejemplo de aplicación de las medidas estadísticas

Supongamos que, durante una campaña de exploración minera, se analizaron cinco muestras de mineral y se obtuvieron las siguientes leyes de cobre:

Muestra	M1	M2	M3	M4	M5
Ley de cobre (% Cu)	0,8	1,0	1,2	1,5	1,9

Consideraremos estas cinco observaciones como una muestra de un sector mineralizado. Los valores son ficticios y se utilizan exclusivamente con fines didácticos.

Ejemplo 1. Media aritmética

La media aritmética es una medida de tendencia central que se obtiene sumando todos los valores observados y dividiendo el resultado por la cantidad de observaciones.

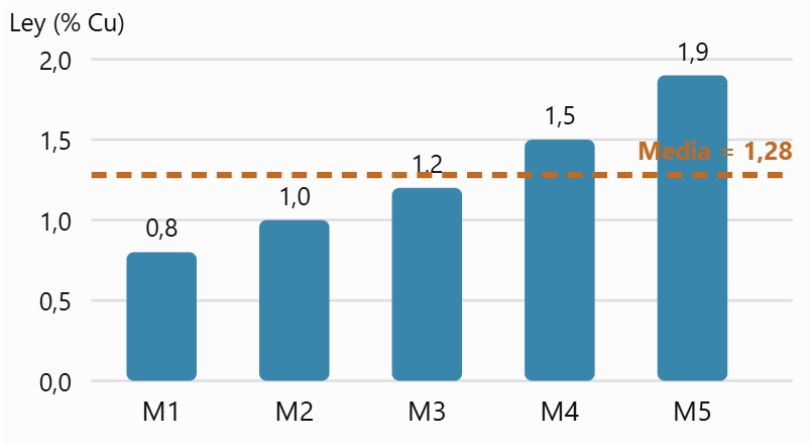
Para nuestros datos:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{0{,}8 + 1{,}0 + 1{,}2 + 1{,}5 + 1{,}9}{5}$$

$$\bar{x} = 1{,}28\ \%Cu$$

Esto significa que la ley media aritmética de las cinco muestras analizadas es del 1,28 % Cu.

La media resume el conjunto mediante un único valor, pero no significa que todas las muestras presenten esa concentración. Tampoco permite afirmar, por sí sola, que la ley media de todo el sector mineralizado sea del 1,28 %.



Las barras en la figura anterior representan las cinco observaciones. La línea discontinua indica su media aritmética.

Ejemplo 2. Mediana

La mediana es una medida de tendencia central que divide un conjunto de datos ordenados en dos partes, de modo que aproximadamente la mitad de las observaciones queda a cada lado.

Primero ordenamos los valores de menor a mayor. En este caso ya están ordenados:

$$0{,}8;\ 1{,}0;\ \ 1{,}2;\ 1{,}5;\ 1{,}9$$

Como tenemos cinco observaciones ($n = 5$), la cantidad es impar. La posición central se obtiene mediante:

$$M_m = x_{\left(\frac{n+1}{2}\right)}$$

$$\frac{n+1}{2} = \frac{5+1}{2} = 3$$

Por lo tanto, el valor es aquel que corresponde a la tercera posición, es decir:

$$M_m = x_3 = 1{,}2\%Cu$$

La mediana es 1,2 % Cu, porque ocupa la tercera posición del conjunto ordenado. Hay dos observaciones menores y dos mayores.



La observación central se destaca en la figura en color naranja.

¿Qué sucede si la cantidad de observaciones es par? Supongamos que agregamos una sexta muestra cuya ley es 2,0 % Cu. Los valores ordenados serían:

$$0{,}8;\ 1{,}0;\ \ 1{,}2;\ 1{,}5;\ 1{,}9;\ 2{,}0$$

Como $n = 6$, no existe una única observación central. La mediana se calcula promediando los valores de las posiciones 3 y 4:

$$M_m = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

$$\frac{x_{(3)} + x_{(4)}}{2} = \frac{1{,}2 + 1{,}5}{2}$$

$$M_m = 1{,}35\%Cu$$

Así, la mediana puede coincidir con un valor observado o encontrarse entre dos valores del conjunto.

Ejemplo 3. Desviación estándar

La desviación estándar es una medida de dispersión que permite describir cuánto varían los datos alrededor de su media aritmética.

Retomamos las cinco observaciones originales, cuya media es 1,28 % Cu.

Para calcular la desviación estándar muestral corregida utilizamos:

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

Primero calculamos las diferencias entre cada observación y la media, y luego elevamos esas diferencias al cuadrado.

Ley (% Cu)	Diferencia respecto de la media	Diferencia al cuadrado
0,8	−0,48	0,2304
1,0	−0,28	0,0784
1,2	−0,08	0,0064
1,5	+0,22	0,0484
1,9	+0,62	0,3844
Total		0,7480

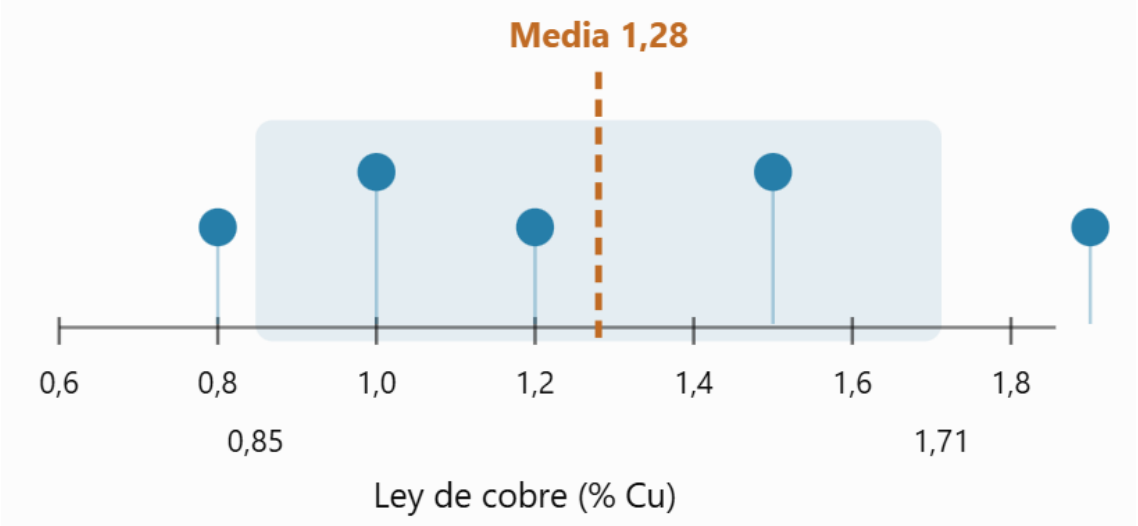
Sustituimos en la fórmula:

$$s = \sqrt{\frac{0{,}7480}{5-1}} = \sqrt{0{,}1870}$$

$$s \approx 0{,}4324\ \%Cu$$

La desviación estándar muestral corregida es aproximadamente 0,43 puntos porcentuales de ley de cobre.

Este valor expresa la dispersión de las mediciones alrededor de su media. No significa que todas las observaciones se encuentren necesariamente a esa distancia de la media.



Los puntos azules representan las mediciones; la línea naranja indica la media. La franja sombreada representa el intervalo que cubre la media ± una desviación estándar muestral corregida, aproximadamente de 0,85 % a 1,71 % Cu. No implica que todas las observaciones deban quedar dentro de ese intervalo.

La siguiente tabla muestra una síntesis del resultado de estos tres ejemplos

Medida	**Resultado**	**¿Qué describe?**
Media aritmética	1,28 % Cu	El valor promedio de las observaciones
Mediana	1,20 % Cu	El valor central del conjunto ordenado
Desviación estándar muestral corregida	0,43 puntos porcentuales	La dispersión de los valores alrededor de la media

Las tres medidas se calculan sobre los mismos datos, pero aportan información diferente.

La media y la mediana describen la tendencia central del conjunto, mientras que la desviación estándar describe su dispersión.

Por ello, ninguna de estas medidas, considerada de manera aislada, permite caracterizar completamente el comportamiento de las observaciones.

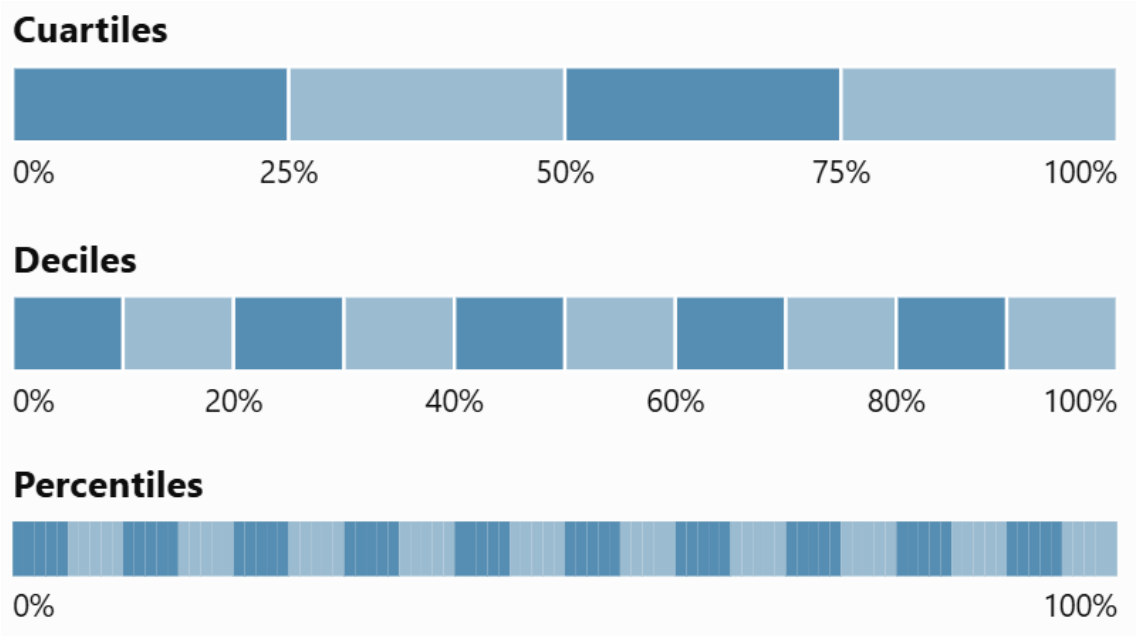
Las medidas de posición

Los cuartiles, deciles y percentiles permiten identificar posiciones dentro de un conjunto de datos ordenados. A diferencia de la media, que resume los valores mediante un promedio, estas medidas indican aproximadamente qué proporción de observaciones se encuentra por debajo de determinados valores.

Para calcularlas, primero debemos ordenar los datos de menor a mayor.

- **Cuartiles**. El primer cuartil deja aproximadamente el 25 % de los datos por debajo de su valor y el 75 % por encima. El segundo cuartil coincide con la mediana. El tercer cuartil deja aproximadamente el 75 % de los valores por debajo y el 25 % por encima.
- **Deciles**. Análogamente a los cuartiles, el primer decil deja aproximadamente el 10 % de los datos por debajo de su valor y el 90 % por encima. Así va avanzando hasta que el noveno decil deja aproximadamente el 90 % de los datos por debajo y el 10 % por encima.
- **Percentiles**. Son conceptualmente equivalentes a los deciles, pero, por ejemplo, el primer percentil deja aproximadamente el 1 % de los datos por debajo y el 99 % por encima.

El siguiente gráfico muestra la relación entre estas medidas de posición



Los cuartiles dividen las posiciones de los datos ordenados en cuatro partes; los deciles, en diez; y los percentiles, en cien.

Ejemplo 4. Cálculo de cuartiles

Supongamos que se analizaron once muestras de mineral y se obtuvieron las siguientes leyes de cobre, ordenadas de menor a mayor:

Muestra	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
Ley de cobre (% Cu)	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2	1,3	1,4	1,5

Los cuartiles permiten identificar tres valores que dividen el conjunto ordenado en cuatro partes.

Para este ejemplo utilizaremos el procedimiento de posiciones $k(n+1)/4$, con k=1,2,3.

Primer cuartil (Q_1)

$$P_{Q_1} = \frac{1(11+1)}{4} = 3$$

El primer cuartil corresponde al valor que ocupa la tercera posición:

$$Q_1 = 0{,}7\%\ Cu$$

Segundo cuartil (Q_2)

$$P_{Q_2} = \frac{2(11+1)}{4} = 6$$

El sexto valor es de 1,0 % Cu:

$$Q_2 = 1{,}0\%\ Cu$$

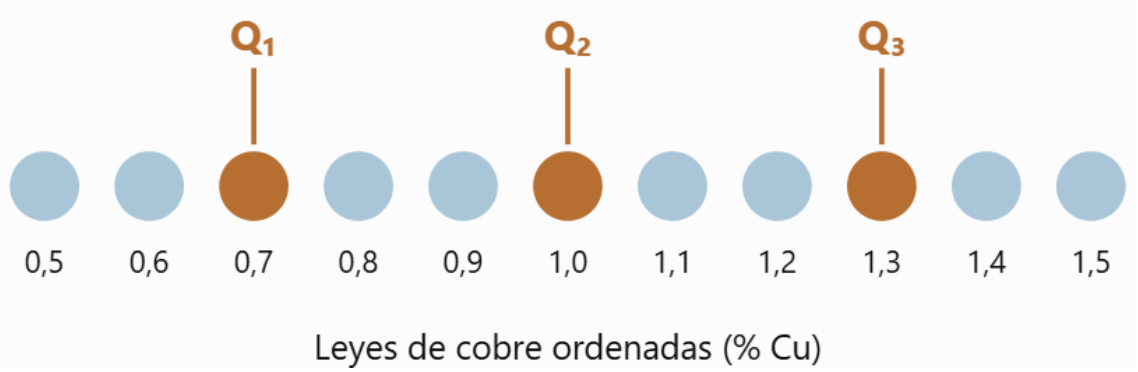
El segundo cuartil coincide con la mediana

Tercer cuartil (Q_3)

$$P_{Q_3} = \frac{3(11+1)}{4} = 9$$

El noveno valor es de 1,3 % Cu:

$$Q_3 = 1{,}3\%\ Cu$$



Los cuartiles se encuentran en las posiciones 3, 6 y 9. Las observaciones extremas también forman parte del conjunto.

Interpretación: aproximadamente el 25 % de las observaciones se encuentra por debajo de Q_1, el 50 % por debajo de Q_2 y el 75 % por debajo de Q_3.

Ejemplo 5. Cálculo de deciles

Supongamos ahora que se registraron los tiempos de carga de nueve camiones mineros, expresados en minutos y ordenados de menor a mayor:

$$4{,}2;\ 4{,}5;\ 4{,}8;\ 5{,}0;\ 5{,}2;\ 5{,}5;\ 5{,}8;\ 6{,}0;\ 6{,}4$$

Los deciles permiten dividir las posiciones de los datos ordenados en diez partes.

Para ilustrar el procedimiento, calcularemos el cuarto decil (D_4), que corresponde aproximadamente al 40 %.

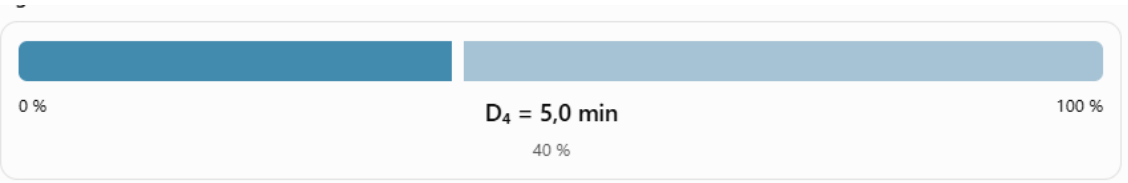
Utilizaremos la posición:

$$P_{D_k} = \frac{k(n+1)}{10}$$

Como *n*=9 y *k*=4

$$P_{D_4} = \frac{4(9+1)}{10} = 4$$

Esto indica que el cuarto decil corresponde a un tiempo de carga de 5,0 minutos. En términos de posición, se encuentra aproximadamente en el 40 % de la distribución ordenada. La siguiente figura expresa una representación conceptual de la posición del cuarto decil. Los porcentajes se refieren a la distribución ordenada, no a una proporción exacta de los nueve datos.



Ejemplo 6. Cálculo de percentiles

Supongamos que una planta de procesamiento registra los tamaños de partícula de nueve muestras, expresados en milímetros y ordenados de menor a mayor:

$$2{,}0;\ 2{,}2;\ 2{,}4;\ 2{,}6;\ 2{,}8;\ 3{,}0;\ 3{,}2;\ 3{,}4;\ 3{,}6$$

Los percentiles permiten dividir las posiciones de los datos ordenados en cien partes.

Calculemos el percentil 80 (P_{80}).

Utilizaremos:

$$L_{P_k} = \frac{k(n+1)}{100}$$

Para *n*=9 y *k*=80

$$L_{P_{80}} = \frac{80(9+1)}{100} = 8$$

El octavo valor es 3,4 mm.

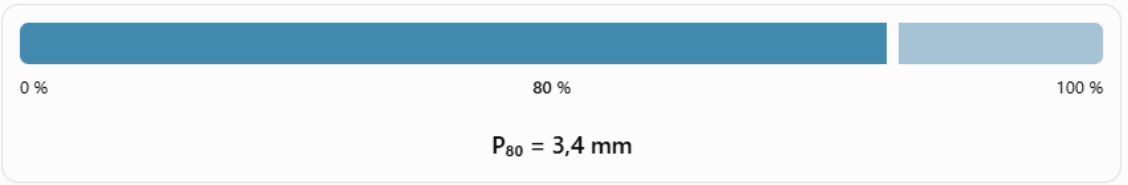
$$P_{80} = 3{,}4\ mm$$

Esto significa que el percentil 80 corresponde a un tamaño de partícula de 3,4 mm.

En el análisis granulométrico, este tipo de medida resulta especialmente útil para caracterizar la distribución de tamaños del material. Por ejemplo, el parámetro D_{80} suele indicar el tamaño de partícula por debajo del cual se encuentra aproximadamente el 80 % del material, considerando la distribución acumulada en masa.

Una distinción importante: el percentil 80 calculado sobre nueve mediciones individuales no equivale necesariamente al D_{80} granulométrico de un lote de mineral, porque este último se determina a partir de la distribución de tamaños ponderada por la masa del material.

La siguiente figura expresa que el percentil 80 identifica una posición de referencia en la distribución ordenada. La representación es conceptual y no una curva granulométrica real.



Nota sobre los procedimientos de cálculo: existen diferentes convenciones para determinar cuartiles, deciles y percentiles, especialmente cuando las posiciones calculadas no son números enteros. En los ejemplos anteriores seleccionamos cantidades de datos que permiten obtener posiciones enteras y aplicar directamente el procedimiento. Cuando la posición no es entera, algunos métodos utilizan interpolación entre observaciones consecutivas. Por esta razón, distintos programas estadísticos pueden producir resultados ligeramente diferentes para un mismo conjunto de datos.

En síntesis, estas medidas permiten identificar valores asociados a determinadas posiciones de una distribución. Complementan a las medidas de tendencia central

y dispersión, aportando información sobre cómo se encuentran distribuidas las observaciones.

Nota para los estudiantes de *Ingeniería* de Minas:

La mediana suele destacarse porque es más resistente a los valores extremos (*outliers*, en la literatura en inglés). Esto significa que, cuando una observación presenta un valor considerablemente mayor o menor que las demás, la mediana generalmente se ve menos afectada que la media aritmética.

En el análisis de datos mineros, es importante reconocer que **un valor extremo no es necesariamente un error**. Puede deberse a una medición incorrecta, pero también puede reflejar una característica real del material o del proceso estudiado.

Por ejemplo, si una muestra de mineral presenta una ley de cobre considerablemente mayor que las demás, no deberíamos descartarla únicamente por ser diferente. Primero debemos investigar si corresponde a un error de medición o si refleja una característica real del sector estudiado.

Ejemplo: ¿qué sucede con una ley de cobre inusualmente elevada?

Supongamos que, durante una campaña de exploración, se obtienen las siguientes leyes de cobre en cinco muestras de mineral:

Muestra	M1	M2	M3	M4	M5
Ley de cobre (% Cu)	1,0	1,1	1,2	1,3	4,0

Observamos que cuatro muestras presentan concentraciones relativamente próximas, mientras que la quinta tiene una ley considerablemente mayor.

Para comprender cómo influye esta observación, comparemos las medidas estadísticas calculadas con las cinco muestras y las obtenidas si excluyéramos la quinta.

Medida	**Cinco muestras**	**Sin la quinta muestra**
Media aritmética	1,72 % Cu	1,15 % Cu
Mediana	1,20 % Cu	1,15 % Cu
Desviación estándar muestral corregida	1,28 puntos porcentuales	0,13 puntos porcentuales

¿Qué podemos observar?

La media aritmética cambia considerablemente al excluir la quinta muestra: pasa de 1,72 % a 1,15 % Cu. En cambio, la mediana presenta una modificación mucho menor, pasando de 1,20 % a 1,15 % Cu.

Esto permite comprender por qué la mediana se considera una medida más resistente a los valores extremos.

También observamos que la desviación estándar disminuye notablemente cuando se elimina la quinta observación. Sin embargo, esta reducción no significa que hayamos conseguido que el fenómeno estudiado sea menos variable: simplemente hemos modificado el conjunto de datos utilizado para describirlo.

¿Por qué hablamos de una reducción artificial de la variabilidad?

Existe cierta tendencia a descartar observaciones que se alejan considerablemente de las demás. Sin embargo, **eliminar un dato extremo verdadero sin una justificación técnica puede conducir a conclusiones incorrectas**.

Cuando excluimos una observación extrema, las medidas de dispersión calculadas pueden disminuir. No obstante, esto no significa necesariamente que la variabilidad real del fenómeno sea menor.

Si el valor eliminado corresponde a una observación válida, situaciones semejantes podrían volver a presentarse en otras zonas del yacimiento o durante futuras operaciones.

Por ello, excluir datos extremos sin justificación puede conducir a una descripción incompleta de la variabilidad y afectar las estimaciones de incertidumbre que se realicen posteriormente.

En Ingeniería de Minas, esto podría repercutir en decisiones relacionadas con la exploración, la caracterización de yacimientos, la planificación de la extracción y el control de los procesos productivos.

En consecuencia, antes de descartar una observación extrema debemos investigar su origen, verificar su validez y justificar cualquier decisión de exclusión.

El objetivo del análisis estadístico no consiste en obtener datos que se ajusten a nuestras expectativas, sino en comprender adecuadamente el comportamiento del fenómeno estudiado.

El estudio de probabilidades y estadística se suele dividir en dos grandes partes:

- **Estadística descriptiva:** Dispongo de un conjunto de datos, correspondiente a una muestra o población, y quiero organizarlo, representarlo y resumir sus características.

- **Inferencia estadística:** Dispongo de datos de una muestra y quiero obtener conclusiones sobre una población más amplia, considerando las condiciones y la incertidumbre del procedimiento.