

PRECISION Y EXACTITUD. Ejemplo

El siguiente ejemplo corresponde a una situación muy habitual: la comparación de conjuntos de mediciones diferentes (correspondientes a diferentes métodos, analistas, laboratorios,...) de la misma cantidad desconocida para ver, por ejemplo, cuál de ellos es mejor en el sentido de ser menos “propenso” a cometer errores.

Cuatro laboratorios farmacéuticos, que llamaremos A, B, C y D, realizan 5 determinaciones, cada uno de ellos, de una muestra de sangre cuya lectura real suponemos conocida e igual a 10.00 ml. Las determinaciones obtenidas por los cuatro laboratorios (en ml.) se recogen en una tabla, donde además se recogen algunos cálculos realizados con ellos.

Laboratorio	Determinaciones	Media \bar{x}	Rango r	Desv. s	CV
A	10.09	10.10	0.04	0.01581	0.00157
	10.10				
	10.08				
	10.11				
	10.12				
B	9.88	10.01	0.41	0.17176	0.01716
	10.02				
	10.14				
	10.21				
	9.80				
C	9.79	9.90	0.50	0.21048	0.02126
	10.19				
	9.78				
	10.05				
	9.69				
D	10.04	10.01	0.07	0.03317	0.00331
	10.02				
	9.98				
	9.97				
	10.04				

Si denotamos las determinaciones por x_1, \dots, x_5 , la media, que denotaremos por \bar{x} , es la estimación del valor verdadero de la lectura de la muestra que da el laboratorio correspondiente (y que debería aproximarse al valor

real 10.00, aunque puede no hacerlo), y no es más que la media aritmética de las observaciones (y tiene sus mismas unidades):

$$\bar{x} = \frac{x_1 + \cdots + x_5}{5} = \frac{\sum_{i=1}^5 x_i}{5}.$$

En la tabla, además de la media, se han calculado algunas medidas de la dispersión de las determinaciones para cada laboratorio. Como estimación de la *dispersión* de las determinaciones se ha calculado la **desviación**, s , que es la raíz cuadrada (positiva) de la **variancia**, y tiene las mismas unidades que las observaciones, ml. en este caso. Tanto la desviación como la variancia son estimaciones de la dispersión de las observaciones de los datos alrededor del valor central que es la media, que dependen de la diferencia entre las observaciones y este valor central, tomadas al cuadrado para evitar que diferencias positivas y negativas se compensen, y se definen de esta manera:

$$s^2 = \frac{(x_1 - \bar{x})^2 + \cdots + (x_5 - \bar{x})^2}{5 - 1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

es la variancia, y $s = \sqrt{s^2}$ la desviación, donde n es el número de observaciones, $n = 5$ en el ejemplo. Obsérvese que se divide por el número de observaciones menos una en vez de por el número de observaciones, lo que podría parecer más natural de entrada, porque de esta manera se estima mejor la dispersión. Aunque no entraremos en los detalles de en qué sentido es esta mejora, **siempre** dividiremos por $n - 1$ para calcular s^2 o s .

Nota: en la obtención de s^2 es más cómodo utilizar como expresión equivalente a la definición ésta otra,

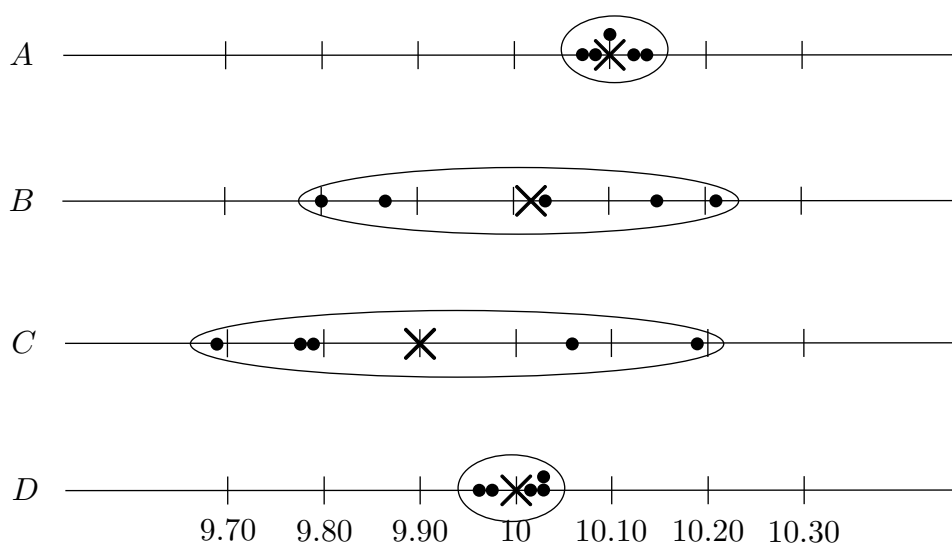
$$s^2 \left(= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \right) = \frac{\sum_{i=1}^n x_i^2 - n(\bar{x})^2}{n - 1}.$$

Otras medidas de la variabilidad de los datos son el **rango**, r , definido como el máximo menos el mínimo de las observaciones (es una medida muy *naïf*) o el **coeficiente de variación**, **CV**, que no depende de las unidades utilizadas y que está definido por

$$CV = \frac{s}{|\bar{x}|}$$

donde $|\cdot|$ indica *valor absoluto*. Al no depender de las unidades, esta medida presenta la ventaja de permitir comparar la variabilidad de grupos de observaciones dados en diferentes unidades, o de diferentes órdenes de magnitud. El coeficiente de variación se conoce también por las siglas **DER** (desviación estándar relativa) y **RSD** (del inglés, *ratio standard deviation*). Lo hemos calculado, junto con la media, el rango y la desviación, para los datos de los cuatro laboratorios, redondeando al quinto decimal.

Además de utilizar estas medidas para comparar los laboratorios, podemos también hacer una representación gráfica como la siguiente (este tipo de gráfico se conoce como *diagrama de puntos* o “*dot plot*”, en inglés, por razones obvias).



En ella se representan mediante puntos las cinco observaciones de cada laboratorio y su media se indica mediante un aspa. Hemos rodeado las cinco observaciones para dar una idea gráfica de la dispersión de las mismas. Naturalmente, las informaciones aportadas por el cálculo de la media, el rango, la desviación y el coeficiente de variación, por una parte, y el gráfico por otra, se complementan.

Así, vemos que los resultados del laboratorio A cumplen que todos están muy próximos (CV pequeño), lo que indica una alta **precisión**. Sin embargo, son demasiado altos (respecto del valor de referencia, que era 10.00), por lo que son **inexactos**.

Podemos apreciar que hay dos tipos de errores diferentes en los resultados: los **aleatorios**, que provocan *dispersión* de los resultados individuales alrededor de su media (en el caso de los resultados del laboratorio A, los

errores aleatorios son pequeños y obtenemos resultados con poca dispersión alrededor de su valor medio, esto es, con gran precisión). Los errores **sistemáticos**, por su parte, provocan que los resultados sean erróneos en el mismo sentido (para este laboratorio, todos por exceso), luego afectan a la exactitud de los resultados, que es la proximidad de la media al valor verdadero; para los resultados del laboratorio A los errores sistemáticos son importantes, lo que provoca que los resultados sean inexactos. Observamos que, efectivamente, la estimación del valor verdadero de la lectura que de la muestra obtiene el laboratorio A dista bastante del valor real 10.00 debido a los errores sistemáticos, que provocan un sesgo en las determinaciones.

El laboratorio B ha obtenido resultados exactos, ya que su media es 10.01, muy próxima al valor real 10.00, pero imprecisos, pues su CV es grande (por comparación con A). El C muestra unos resultados inexactos e imprecisos. Por último, el laboratorio D muestra resultados exactos y precisos.

Podemos resumir la información que hemos obtenido sobre los cuatro laboratorios en la siguiente tabla:

	Precisión	
Exactitud	Alta	Baja
Alta	D	B
Baja	A	C