cano pina

Tema 1
Organización del plan de muestreo

Contenidos

Extras

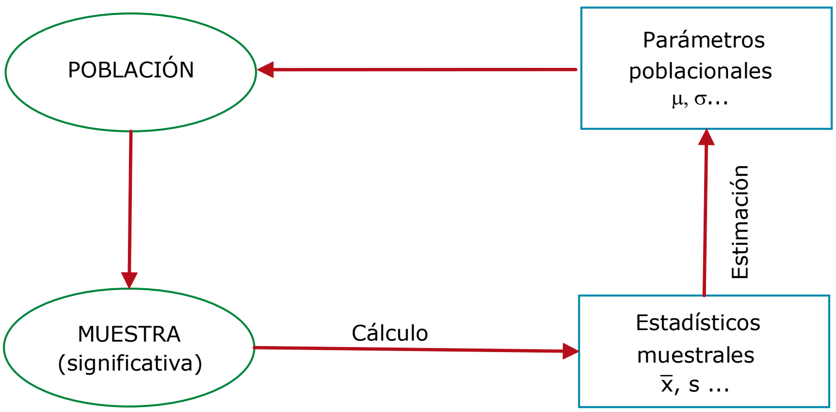
1.1 Importancia del muestreo

El muestreo es el primer paso de un análisis químico, constituyendo la fuente de error más importante, incluido el análisis.

Aunque existen métodos que intentan determinar los componentes del material sin necesidad de muestrear, la mayoría de los análisis todavía se hacen mediante el análisis en laboratorio de una parte del material (muestra).

Una población es el grupo de todos los elementos que se proponen para obtener una medida característica. Ante la imposibilidad de estudiar toda la población se selecciona un subgrupo de elementos representativos de la población que constituye lo que se denomina muestra.

Uno de los objetivos del análisis estadístico consiste en estimar los parámetros de la población a partir de los valores estadísticos calculados de los datos proporcionados por la muestra, según se muestra en la figura:



Como no resulta práctico analizar todo el lote de producción, existen dos formas de estimar los parámetros de la población, a partir de una muestra:

- **Estimación puntual.** Una estimación puntual de la media poblacional μ es la media muestral $\bar{x}$ que varía de una muestra a otra, teniendo como inconveniente no poder conocer cómo de buena es la aproximación entre la estimación ($\bar{x}$) y el parámetro de población (μ).

Ejemplo 1

Si se dice que el contenido en hierro de una muestra es del 0,23%, se está dando una estimación puntual.

- **Estimación por intervalo de confianza.** Permite estimar el parámetro desconocido (μ), proporcionando el error en nuestra estimación. El término confianza se utiliza para indicar que se confía en la exactitud o precisión de la medida.

Ejemplo 2

Si se dice que el contenido en hierro es del 0,23 ± 0,05%, se está diciendo que el porcentaje de hierro en la muestra estará comprendido entre el 0,18 y 0,28%, siendo este caso una estimación por intervalo.

Existen dos maneras de estimar los parámetros poblacionales:

- Analizando toda la población, lo que se ve claramente imposible cuando se trata de un gran número de elementos.

- Estimando el rango dentro del cual el valor del parámetro de la población sea más probable. El nivel de confianza del intervalo lo podemos fijar y se suele trabajar con el 95%, aunque a veces también con el 99% o el 90%.

Ejemplo 3

Una interpretación del nivel de confianza es la probabilidad a largo plazo. Por ejemplo, si una estimación de una media tiene una probabilidad del 95% quiere decir que si el muestreo se realiza una y otra vez, a largo plazo la estimación de la media se dará en el 95% de los casos.

Cuanto mayor sea el intervalo de confianza menor será la precisión de la medida y cuanto menos sean las muestras, menor es el nivel de confianza que puede ser asignado a un intervalo particular.

Análogamente, para pocas medidas se utilizarán intervalos de confianza más grandes para un nivel de confianza dado.

Los resultados analíticos obtenidos de n muestras se pueden distribuir con una desviación estándar:

$s = \frac{\sigma}{\sqrt{n}}$, donde σ es la desviación típica y es la raíz cuadrada de la diferencia entre las observaciones individuales y la media poblacional $\sigma = \sqrt{\frac{\sum(x_i - \mu)}{n}}$, considerándose s como una estimación de σ cuando el nº de datos es pequeño.

En general σ y μ no son conocidas, pero s puede ser una estimación de σ y la media de los datos obtenidos en el muestreo puede usarse como una estimación de μ.

Si se toman varias muestras y se analiza su contenido medio, el intervalo de confianza para la media se estima, como:

$\mu = \bar{x} \pm \frac{t \cdot s}{\sqrt{n}}$, donde s es una estimación de σ y t un parámetro derivado de la distribución de Student, que depende del número de grados de libertad (n - 1), o el número de desviaciones independientes que se utilizan en calcular s, restándole 1 porque la última desviación se puede calcular matemáticamente, ya que la suma de todas las desviaciones debe ser 0.

Para determinar el valor de t si se dispone de una tabla como la del Apéndice 7 y si lo que se desea es:

a. Buscar el valor de una cola, se entra por la columna encabezada el valor de α en el área en una cola.

b. Buscar el valor de dos colas, hay que entrar por la columna encabezada por α en el área en dos colas, siendo α el nivel de significación, empleándose habitualmente α = 0,05 (nivel de confianza del 95%).

Ejemplo 4

La determinación del grado alcohólico de 10 botellas de vino ha dado los siguientes resultados en % volumen etanol (grado).

Muestra Nº	Grado
1	12,8
2	12,3
3	11,9
4	12,2
5	11,8
6	12,2
7	11,9
8	12,0
9	11,8
10	12,1

Debemos calcular el contenido medio y sus límites de confianza para los niveles de significación:

a) $\alpha = 0,1$

b) $\alpha = 0,05$

c) $\alpha = 0,01$

Primero, determinamos la media y la desviación estándar, según se muestra en la tabla:

Muestra	Grado	$x - \bar{x}$	$(x - \bar{x})^2$
1	12,8	0,70	0,490
2	12,3	0,20	0,040
3	11,9	-0,20	0,040
4	12,2	0,10	0,010
5	11,8	-0,30	0,090
6	12,2	0,10	0,010
7	11,9	-0,20	0,040
8	12,0	-0,10	0,010
9	11,8	-0,30	0,090
10	12,1	0,00	0,000
Suma =	121	0	0,820
Media	$\bar{x} = \frac{121}{10} = 12,10$	Desviación estándar	$s = \sqrt{\frac{0,820}{10-1}} = 0,302$

a) Un nivel de significación de $\alpha = 0,1$, indica un nivel de confianza del (1 - 0,1) · 100 = 90%.

Al ser una muestra pequeña (n<30), debemos aplicar el estadístico t y como es un intervalo de confianza, debemos buscar el valor en el área de dos colas, entrando por la columna encabezada por área de dos colas con $\alpha = 0,1$

Para determinar el valor de t, entramos por la columna 0,1 y con 9 (10-1) grados de libertad nos da un valor de t = 1,833.

Sustituyendo, queda $\mu = \bar{x} \pm t \cdot \left(\frac{s}{\sqrt{n}}\right) = 12,10 \pm 1,833 \cdot \left(\frac{0,302}{\sqrt{10}}\right) = 12,10 \pm 0,175$, siendo el resultado promedio para un nivel de confianza del 90% de 12,10 ± 0,18.