

Tarea dos de Diseño de Experimentos

Sofía Cuartas García (C.C. 1.000.874.937) – scuartasg@unal.edu.co
Simón Cuartas Rendón (C.C. 1.037.670.103) – scuartasr@unal.edu.co

Septiembre 7, 2022

Punto uno

- **Enunciado.** Se desea investigar el efecto del pH en el crecimiento de cierto microorganismo en un medio específico. Para ello se realiza un experimento, teniendo como punto de partida la misma cantidad de microorganismos. Se hacen cuatro repeticiones y se obtienen los siguientes resultados:

Table 1: Crecimiento porcentual promedio según el nivel de pH

Nivel de pH	Crecimiento porcentual promedio
1	80
2	105
3	75

¿Estos datos son evidencia suficiente para afirmar que los niveles de pH donde se logra menor y mayor crecimiento son el tres y el dos respectivamente? Explique y liste las consideraciones que se deben tener en cuenta para que las conclusiones obtenidas sean válidas.

Consideraciones básicas del experimento

A continuación se explicarán los elementos básicos relacionados al diseño de este experimento.

- **Unidad experimental.** Cada uno de los microorganismos que fueron considerados para la realización del experimento.
- **Factor.** Nivel de pH .
- **Niveles.** Esta variable fue categorizada para estudiar el efecto que tienen tres valores de pH sobre la variable de respuesta: $pH = 1, 2, 3$.
- **Variable respuesta.** Crecimiento de cierta especie de microorganismo.
- **Garantía de la aleatorización del experimento.** Para poder garantizar que se aleatorice el experimento, se pueden aislar cada uno de los microorganismos con los que se desarrollará el experimento, de tal forma que hayan tres grupos con la misma cantidad de microorganismos, de tal suerte que el experimento sea balanceado. Luego, se puede asignar de forma aleatoria el nivel de pH que tendrá el medio en el que habita cada grupo con ayuda de R , definiendo los grupos ‘uno’, ‘dos’ y ‘tres’, se ejecuta un comando del tipo `sample(1:3, size = 3, replace = FALSE`, y el primer elemento se le asigna al primer grupo y de forma análoga para los otros dos. Luego, se estudian los microorganismos de forma aleatoria ejecutando un comando similar para los n microorganismos que hayan sido considerados para la realización del experimento.

¿Se puede concluir sobre el crecimiento de los microorganismos?

En primer lugar es importante tener presente que no se puede asumir un acercamiento determinístico a este problema para considerar si alguno de los tres tratamientos es mayor o menor que los otros a partir del valor de la media muestral reportada en la tabla 1.

Entonces, una primera aproximación para poder hacer juicios basados en la evidencia estadística es mediante un análisis de varianza con la tabla ANOVA para este experimento de un solo factor, la cual permite determinar si la media de cada uno de los tratamientos considerados es igual o no desde un punto de vista estadístico. Así pues, si Y_{ij} es el crecimiento promedio porcentual de los microorganismos de acuerdo con el i -ésimo nivel de pH en la j -ésima repetición, entonces tal modelo es:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim NID(0, \sigma^2), \quad i = 1, 2, 3, \quad j = 1, 2, 3, 4$$

donde μ es la media global, por lo que es un parámetro común a todos los tratamientos, τ_i es un parámetro único del i -ésimo tratamiento y ε_{ij} es el error aleatorio relacionado a cada repetición para cada tratamiento.

Sin embargo, como el mismo nombre lo dice, este método requiere de la varianza para poder llevar a cabo el análisis, pero en este caso solo conocemos el valor de las medias para cada tratamiento (los \bar{y}_i , $i = 1, 2, 3$), lo cual impide estudiar si quiera si existe al menos una media que sea diferente (mayor o menor) a las otras. Esto quiere decir que, deberíamos conocer los siguientes valores: $y_{11}, \dots, y_{14}, y_{21}, \dots, y_{24}, y_{31}, \dots, y_{34}$, así como los residuales para poder calcular la suma de los cuadrados totales y la suma de los cuadrados asociados a los tratamientos, necesarios para poder obtener los valores que se reportan en una tabla ANOVA.

Y a propósito de los residuales, un asunto muy importante es que deben ser estudiados los supuestos del modelo ANOVA, y es que los errores aleatorios son independientes e idénticamente distribuidos siguiendo una normal con media nula y homocedásticos. Para poder chequear estos supuestos es necesario conocer los residuales del experimento.

Asimismo, y ahora con el objetivo de determinar si algún tratamiento en particular tiene resultados en promedio mayores o menores a los de algún otro tratamiento (o varios de ellos), se podría considerar cualquiera de los testes que comparan las medias de los tratamientos de un experimento, tales como el test de Tukey o el de Duncan, pero semejante al caso de la ANOVA, es necesario conocer los valores obtenidos en cada una de las cuatro repeticiones para los cuatro tratamientos considerados, de tal forma que se puedan calcular las varianzas para obtener los estadísticos de prueba de estos testes.

En conclusión, no se tienen elementos suficientes para sugerir que alguno de los tratamientos considerados derivan en un crecimiento porcentual promedio significativamente mayor o menor que los otros dos desde el punto de vista estadístico.