
CR 4. FUNDAMENTOS DE DISEÑO EXPERIMENTAL PARA LA MODELACIÓN Y SIMULACIÓN DE PROCESOS BIOLÓGICOS

Luis Miguel Mejía

Ingeniero agrónomo

Especialista en gestión para el desarrollo empresarial y docencia universitaria

Director del grupo de investigación para el desarrollo agroindustrial GIDA

Coordinador de investigaciones facultad de ingeniería universidad La Gran Colombia

Universidad La Gran Colombia seccional Armenia

miguelsejia99@gmail.com

RESUMEN: El Diseño de Experimentos ha sido una ramificación de la Estadística que se ha caracterizado por su fundamentación inductiva de procesos de diversas índoles, siendo los procesos biológicos propios de la agricultura y el sector agropecuario área de permanente estudio; pero con la aparición de la minería de datos (*Data Mining*) y la Inteligencia Analítica se han comenzado a explorar desde el diseño, diversas formas de modelar y posteriormente simular variables respuesta en función de variables explicatorias o fuentes de variación. No obstante, implica una serie de condiciones para llevar a buen término tales modelos y simuladores y en última instancia maximizar el potencial de “minar datos experimentales”; aspectos como las condiciones de control del trabajo experimental, la fiabilidad y validez de la modelación, la naturaleza propia de las variables, el método científico y la metodología de modelación y simulación como tal, la cual parte de la formulación del problema de investigación hasta desembocar en el análisis con base en la respectiva simulación.

DISEÑO EXPERIMENTAL EN EL CONTEXTO DE I+D

En los años recientes se observa una tendencia creciente hacia la creación y desarrollo de empresas productoras de bienes y servicios con un alto valor agregado de conocimientos, caracterizadas por su pequeño tamaño, estar fuertemente influenciadas por las funciones de Investigación y Desarrollo (I+D), por poseer altos márgenes de ganancia en sus productos, y por su potencial para sustituir importaciones y competir en los mercados externos, hechos que las hacen diferentes a las empresas tradicionales.

Estas empresas, conocidas como empresas de base tecnológica, se presentan principalmente en áreas tales como la informática, las comunicaciones, la mecánica de precisión, la biotecnología, la química fina, la electrónica, la instrumentación, la elaboración de nuevos materiales, entre otros, y en muchas ocasiones sus orígenes se encuentran en spin-offs de proyectos llevados a cabo por universidades y el sector privado, el cual ha efectuado inversiones en infraestructura para la investigación.

Sin embargo, tanto la investigación científica como tecnológica se desarrollan actualmente bajo procesos de gestión, sustentados en proyectos sólidamente estructurados, ya que la gestión de la Innovación Tecnológica se sustenta en la introducción comercial de nuevos productos o procesos, logrados a partir de la generación de conocimiento sobre los medios empleados y con un fuerte apoyo de la investigación experimental; algunos autores han definido la innovación tecnológica como lo afirmado por Waissbluth et, al (1986), que la innovación tecnológica es un Proceso que consiste en conjugar oportunidades técnicas con necesidades, y que conduce a la integración de un paquete tecnológico, cuyo objetivo es introducir o modificar productos o procesos en el sector productivo, con su consecuente comercialización.

Otra manera de definir la innovación tecnológica es la planteada por Nelson (1993): “Cambio que requiere un considerable grado de imaginación; constituye una ruptura relativamente profunda con las formas establecidas de hacer las cosas y con ello crea fundamentalmente nuevas capacidades, por lo cual no debe entenderse como un concepto técnico, sino de raíz económica y social”.

En el proceso de innovación o de cambio tecnológico existen tres momentos o estados fundamentales, tal como lo indica Camacho (1998) en la Tabla 1.

TABLA 1. ESTADOS FUNDAMENTALES DEL PROCESO DE INNOVACION.

{PRIVATE}ESTADO O MOMENTO	DEFINICION
La invención	Creación de una idea potencialmente generadora de beneficios comerciales pero no necesariamente realizada de forma concreta.
La innovación	Aplicación comercial de una idea; se trata de un hecho comercial y social que crea riqueza pero no conocimiento.
La difusión	Diseminación en la sociedad de la utilización de una innovación; es el estado en el cual se ve afectada la economía, obteniendo los beneficios de la innovación.

El mismo autor (Camacho, 1998) define 6 fases que conforman el referido proceso de innovación tecnológica como se indica en la Tabla 2.

TABLA 2. FASES DEL PROCESO DE INNOVACION.

{PRIVATE}FASE	DEFINICION
Idea	Base del proceso de innovación; para generarla es necesaria la información y para implementarla se requiere la decisión de los responsables de la empresa

	y de su financiación.
Investigación	Estudio original y planificado que se emprende con la finalidad de obtener conocimientos nuevos.
Desarrollo tecnológico	Ensayo y elaboración de una aplicación potencial a un modelo o a una serie de especificaciones que demuestren la practicabilidad física de un nuevo proceso o producto.
Elaboración de prototipo	Se persigue conocer la practicabilidad económica y física de utilizar realmente un modelo o unas especificaciones.
Producción	Estructuración y montaje de nuevos medios de producción, seguido del ensayo y modificación de los mismos hasta que resulten posibles las operaciones a ritmo normal. En esta fase la normalización, la homologación y la garantía de la calidad tienen una importancia fundamental.
Comercialización	Puesta a disposición de los consumidores del nuevo producto, a través de determinados canales de distribución y puntos de venta.

La innovación tecnológica, apoyada en investigación es entendida como la conversión de conocimiento tecnológico en nuevos productos o procesos para su introducción en el mercado, en función de las necesidades sentidas de nichos específicos, lo cual es una actividad fundamentalmente empresarial.

Sin embargo, al aplicar criterios de investigación científica y tecnológica la adaptación no suele ser drástica, dados los métodos de investigación, como son aquellos de índole inductivo, sobre los cuales se fundamenta el diseño experimental, aunque con ciertas reglas:

1. Integrar la investigación en el entorno biológico y las reglas del diseño experimental, con base en las repeticiones, la aleatorización, efectos fijos, aleatorios.
2. Identificar las suposiciones y tratarlas como hipótesis.
3. Medir y reportar bajo monitoreo científico los efectos del manejo de los datos tomados y vistos como experimentales.
4. Aspectos ligados íntimamente a la naturaleza del diseño, la modelación y la simulación como son:
 - Cuestionamiento
 - Número de Repeticiones

- Azar
- Control
- Diseño, en función de los objetivos.
- Eficiencia
- Exactitud y Precisión
- Naturaleza de los datos en función de los supuestos estadísticos en el análisis de varianza.
- Selección adecuada de las variables y análisis de los resultados.

LA REGRESIÓN LINEAL COMO HERRAMIENTA PARA LA OBTENCIÓN DE MODELOS

La regresión lineal simple se basa en la construcción de modelos en los cuales una variable dependiente Y está en función de una variable explicatoria X , es decir,

$$Y = F_{(x)} + \varepsilon$$

Y : Variable dependiente, respuesta, explicada o predecida.

X : Variable independiente, control, estímulo, predictora, explicatoria.

El modelo base para regresión lineal simple es entonces, de la siguiente forma:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Donde:

β_0 : Intercepto (corte con el eje x).

β_1 : Pendiente del modelo.

ε_i : Error experimental, en este recae todo lo que el investigador no es capaz de explicar en el modelo, son los factores no controlados.

Propiedades:

- $E(y_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$
- $V(y_i) = V(\beta_0 + \beta_1 x_i) = \sigma^2$
- $\varepsilon_i \approx N(0, \sigma^2)$
- $COV(\varepsilon_i, \varepsilon_j) = 0 \quad \forall_{i \neq j}$
- $V(\varepsilon_i) = \sigma^2$

- $E\left(\frac{y}{x}\right) = \hat{\beta}_0 + \hat{\beta}_1 x$
- $V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$
- $V(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$

Estimación de los parámetros del modelo de regresión lineal simple:

Para $\hat{\beta}_1$:
$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n}}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}$$

Para $\hat{\beta}_0$:
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Cuando se estiman los betas, el modelo es de la siguiente forma:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

El análisis de varianza para el modelo de regresión lineal simple es de la siguiente forma:

- ANAVA

• FV	• GL	• SC	• CM	• Fcalc	• Ftabla
Modelo	1	$SC_{Total} - SC_{Error}$	SC_{Modelo}	$\frac{CM_{Modelo}}{CM_{Error}}$	$f_{\alpha}(GL_{Modelo}, GL_{Error})$
Error Experimental	n-2	$\sum_{i=1}^n e_i^2$	$\frac{\sum_{i=1}^n e_i^2}{n-2}$		
Total	n-1	$\sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}$			

Si $f_{calc} < f_{tabla(0,05,0,01)}$: El modelo no es significativo.

Si $f_{calc} \geq f_{tabla(0,05)}$:El modelo es significativo y se denota al lado del valor de F calculado con un asterisco (*), indicando que el modelo se explica con un 95% de confiabilidad.

Si $f_{calc} \geq f_{tabla(0,01)}$:El modelo es altamente significativo y se denota al lado del valor de F calculado con dos asteriscos (**), indicando que el modelo se explica con un 99% de confiabilidad.

- **COEFICIENTE DE DETERMINACIÓN (R^2)**

El coeficiente de determinación R^2 , explica la variación de y_i , atribuible a los factores adscritos al modelo, indica el porcentaje de explicación del modelo en torno a y_i , posee la siguiente fórmula:

$$R^2 = \frac{SC_{Modelo}}{SC_{Total}}$$

Propiedades:

- Oscila entre 0 y 1, así: $0 \leq R^2 \leq 1$
- R^2 cercano a 1 explica casi toda la variación de y_i , pero no es siempre el mejor modelo.

- **RESIDUALES (e_i)**

Los residuales son la divergencia existente entre el valor y_i real y el estimado, entendiéndose como un error atribuido en cada observación y es de la siguiente forma:

$$e_i = y_i - \hat{y}_i \quad \text{donde} \quad \hat{\epsilon}_i = e_i$$

Propiedades:

- $\sum_{i=1}^n e_i = 0$
- $\sum_{i=1}^n e_i x_i = 0$
- $\sum_{i=1}^n e_i y_i = 0$

los residuales juegan un papel importante en el modelo de regresión lineal simple, ya que con base en ellos se validan los supuestos adscritos al mismo.

Ejemplo de Regresión lineal simple:

Un investigador desea saber el comportamiento del rendimiento (gramos) de una sustancia sometida bajo cierto proceso biotecnológico en función de la cantidad (mililitros) de un sustrato aplicado; tomó en forma aleatoria 10 muestras a las cuales les midió la cantidad de sustancia aplicada y el rendimiento respectivos encontrando:

Xi	Yi
3.2	16.5
3.1	14.3
3.5	17
4	13.2
4.2	15.5
3.8	18.3
3.7	16.9
4.1	14.6

1. Se calcula la media aritmética tanto para Xi como para Yi:

$$\bar{x} = \frac{3.2 + 3.1 + \dots + 4.1}{8} = 3.7$$

e

$$\bar{y} = \frac{16.5 + 14.3 + \dots + 14.6}{8} = 15.7875$$

2. Seguidamente se estima β_1 :

$$\hat{\beta}_1 = \frac{(3.2 * 16.5) + (3.1 * 14.3) + \dots + (4.1 * 14.6) - \frac{(3.2 + 3.1 + \dots + 4.1) * (16.5 + 14.3 + \dots + 14.6)}{8}}{3.2^2 + 3.1^2 + \dots + 4.1^2 - \frac{(3.2 + 3.1 + \dots + 4.1)^2}{8}}$$

$$\rightarrow \hat{\beta}_1 = -0.7327586207$$

3. Seguidamente se estima β_0 :

$$\hat{\beta}_0 = 15.7875 - (-0.7327586207 * 3.7) = 18.4987069$$

Entonces la ecuación obtenida es la siguiente:

$$\hat{y}_i = 18.4987069 - 0.7327586207 * x_i$$

4. Posteriormente se hallan los valores estimados para Y_i con base en la siguiente operación:

X_i	Y_i	$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 * x_i$	\hat{y}_i
3.2	16.5	$\hat{y}_i = 18.4987069 - 0.7327586207 * 3.2$	16.15387931
3.1	14.3	$\hat{y}_i = 18.4987069 - 0.7327586207 * 3.1$	16.22715518
3.5	17	$\hat{y}_i = 18.4987069 - 0.7327586207 * 3.5$	15.93405173
4	13.2	$\hat{y}_i = 18.4987069 - 0.7327586207 * 4$	15.56767242
4.2	15.5	$\hat{y}_i = 18.4987069 - 0.7327586207 * 4.2$	15.42112069
3.8	18.3	$\hat{y}_i = 18.4987069 - 0.7327586207 * 3.8$	15.71422414
3.7	16.9	$\hat{y}_i = 18.4987069 - 0.7327586207 * 3.7$	15.7875
4.1	14.6	$\hat{y}_i = 18.4987069 - 0.7327586207 * 4.1$	15.49439656

5. Con base en la diferencia entre el valor observado de Y_i y su estimado se calculan los residuales por cada uno:

y_i	\hat{y}_i	$e_i = y_i - \hat{y}_i$
16.5	16.15387931	0.346120686
14.3	16.22715518	-1.92715518
17	15.93405173	1.065948272

13.2	15.56767242	-2.36767242
15.5	15.42112069	0.078879307
18.3	15.71422414	2.585775859
16.9	15.7875	1.112499997
14.6	15.49439656	-0.89439656
$\sum e_i =$		0

6. Después de calcular los residuales cada uno de ellos se eleva al cuadrado y se suman para obtener así la suma de cuadrado del error experimental:

e_i	e_i^2
0.346120686	0.119799529
-1.92715518	3.713927072
1.065948272	1.13624572
-2.36767242	5.605872675
0.078879307	0.006221945
2.585775859	6.686236791
1.112499997	1.237656242
-0.89439656	0.799945198
$\sum e_i^2 =$	19.30590517

7. Seguidamente se calcula la suma de cuadrado total con base en los valores de Y_i , de la siguiente manera.

$$SC_{Total} = 16.5^2 + 14.3^2 + \dots + 14.6^2 - \frac{(16.5 + 14.3 + \dots + 14.6)^2}{8} = 19.92875$$

8. Se calcula la suma de cuadrado del modelo, restándole a la suma de cuadrado total la suma de cuadrado del error experimental ($\sum e_i^2$):

$$SC_{Modelo} = 19.92875 - 19.30590517 = 0.62284483$$

9. Se construye el respectivo Análisis de Varianza (ANAVA):

ANAVA

FV	GL	SC	CM	Fcalculado	Ftabla 5% 1%
Modelo	1	0.62284483	0.62284483	$\frac{0.62284483}{3.218} = 0.194$	(1,6)= 5.99 13.75
Error Experimental	7- 1=6	19.30590517	$\frac{19.30590517}{6} = 3.218$		
Total	7	19.92875			

Interpretación: Dado que el valor Fcalculado es menor que ambos valores F de la tabla, indica que el modelo no es significativo; es decir, que el rendimiento en gramos del producto sometido a proceso biotecnológico NO está en función de la cantidad en mililitros del sustrato.

Coeficiente de Determinación: $R^2 = \frac{0.62284483}{19.92875} = 0.0313$ Indica que el modelo es explicado por

la cantidad de sustrato en un 3,13%, ratificando lo dicho en la interpretación anterior.

LOS DISEÑOS FACTORIALES COMO HERRAMIENTAS DE MODELACIÓN Y SIMULACIÓN

Este diseño experimental se puede amoldar dentro de cualquier tipo de diseño experimental visto con antelación, tal como bloques al azar, cuadrado latino, completamente aleatorizado, entre otros.

El arreglo factorial se origina cuando se dispone de varios factores a varios niveles cada uno, donde los factores se denotan A, B, C y los niveles se representan con letras minúsculas, números o caracteres alfa-numéricos.

Los factores pueden ser de carácter simétrico o asimétrico, el primero se origina cuando el número de niveles por factor es igual, el segundo cuando es diferente el número de niveles por factor.

Simétrico:

A	B	C
<hr/>		
Ao	Bo	Co
A1	B1	C1

Asimétrico:

A	B	C	D
<hr/>			
Ao	Bo	Co	Do
A1	B1	C1	D1
	B2		D2

Los factores en un arreglo pueden ser dispuestos en forma numérica, letras, binaria y siempre son definidos por el investigador, de antemano, acudiendo a su experiencia sobre unidad experimental o su inquietud por encontrar a través de factores, niveles óptimos de un proceso.

Los niveles son dobles, triples o múltiples y los factores se clasifican en:

Unifactorial: Sólo existe un factor a diferentes niveles.

Bifactorial: Se dispone de 2 factores en un arreglo.

Multifactorial: Es el tipo de arreglo más importante dentro de la estadística experimental, ya que éste permite interacciones de diferentes órdenes como dobles, triples, cuádruples y analizar efectos cruzados, simples, entre otros.

Este arreglo permite también, optimizar la superficie de respuesta y analizar operaciones evolutivas (EVOP) dentro de optimización de procesos.

Los tratamientos se originan al interaccionar los niveles (En este caso el diagrama de árbol es sumamente útil).

Ejemplo:

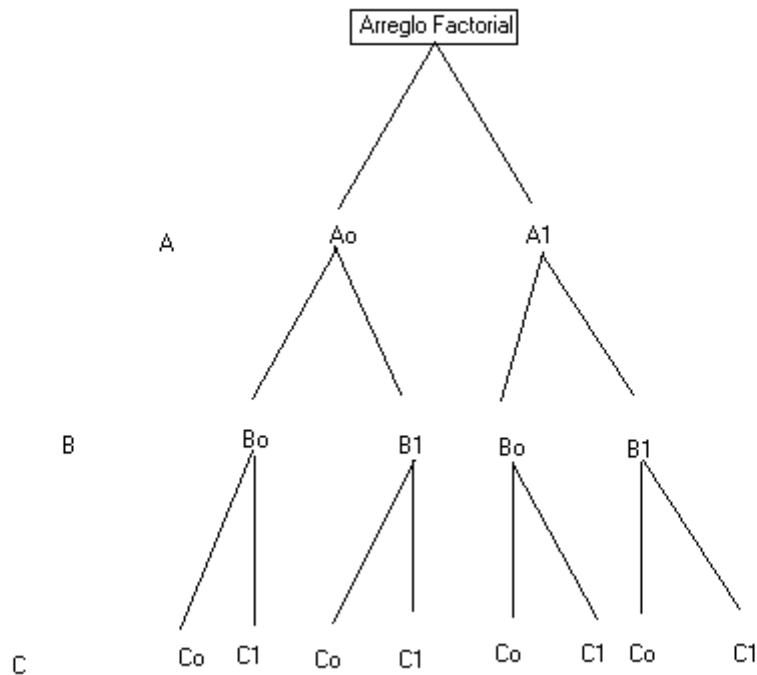
A	B	C
<hr/>		

Ao Bo Co

A1 B1 C1

Cada factor tiene 2 niveles a (Ao-A1, Bo-B1 y Co-C1), lo cual quiere decir que es un arreglo 2^3 , porque:

$2 \times 2 \times 2 = 2^3 = 8$ Tratamientos creados y establecidos así:



Tratamientos:

Ao Bo Co

Ao Bo C1

Ao B1 Co

Ao B1 C1

A1 Bo Co

A1 Bo C1

A1 B1 Co

A1 B1 C1

Los tipos de arreglos factoriales más comunes son $2^2, 2^3, 3^2, 3^3$ siendo estos de índole simétrico.

Cuando es de índole asimétrico es:

A	B	C	D
Ao	Bo	Co	Do
A1	B1	C1	D1
	B2		D2
			D3

$$2 \times 3 \times 2 \times 4 = 2^2 \times 3 \times 4 = 48 \text{ Tratamientos}$$

Ventajas:

- El arreglo factorial se aplica cuando el investigador desea obtener los niveles más adecuados para un cultivo, proceso industrial o agroindustrial, biotecnológico, entre otros.
- Este arreglo permite interacciones entre factores; además, se pueden conocer los efectos simples, principales, cruzados.
- El arreglo factorial minimiza el error experimental, cuando se encuadra el arreglo en un diseño adecuado.
- Este arreglo permite la estimación de datos faltantes.
- Es de vital importancia cuando se desea estimar el efecto del medio ambiente en el tratamiento, o a través del tiempo o el efecto del tratamiento en diferentes medios.
- Sirve para la evaluación de variedades, híbridos y cultivares.

La notación del arreglo factorial se realiza por los métodos: Tradicional (visto con anterioridad), numérica (Tanto niveles como factores se denotan con números) y de tercer tipo o α -numérica (Combina los 2 anteriores).

Para el cálculo y construcción de la ANAVA se hará con base en el arreglo 2^3 en este documento, así:

1. Cálculo de la suma de cuadrado total:

$$SC_{Total} = \sum y_i^2 - \frac{(y_i)^2}{n}$$

2. Cálculo para cada factor (A, B, C, respectivamente, cada uno con 2 niveles).

Para A:

$$SC_A = \frac{\sum A_i^2}{r} - FC$$

Para B:

$$SC_B = \frac{\sum B_i^2}{r} - FC$$

Para C:

$$SC_C = \frac{\sum C_i^2}{r} - FC$$

3. Son calculadas las interacciones de primer orden:

• **A x B:**

$$SC_{AxB} = \frac{\sum A_0B_0}{r} + \frac{\sum A_0B_1}{r} + \frac{\sum A_1B_0}{r} + \frac{\sum A_1B_1}{r} - FC - SC_A - SC_B$$

• **A x C:**

$$SC_{AxC} = \frac{\sum A_0C_0}{r} + \frac{\sum A_0C_1}{r} + \frac{\sum A_1C_0}{r} + \frac{\sum A_1C_1}{r} - FC - SC_A - SC_C$$

• **B x C:**

$$SC_{BxC} = \frac{\sum B_0C_0}{r} + \frac{\sum B_0C_1}{r} + \frac{\sum B_1C_0}{r} + \frac{\sum B_1C_1}{r} - FC - SC_B - SC_C$$

4. Se calcula la interacción de segundo orden (A x B x C):

$$SC_{AxBxC} = \frac{\sum A_0B_0C_0}{r} + \frac{\sum A_0B_0C_1}{r} + \frac{\sum A_0B_1C_0}{r} + \frac{\sum A_0B_1C_1}{r} + \frac{\sum A_1B_0C_0}{r} + \frac{\sum A_1B_0C_1}{r} + \frac{\sum A_1B_1C_0}{r} + \frac{\sum A_1B_1C_1}{r} - FC - SC_A - SC_B - SC_C - SC_{AxB} - SC_{AxC} - SC_{BxC}$$

5. Se halla la suma de cuadrado del tratamiento:

$$SC_{Tto} = SC_A + SC_B + SC_C + SC_{AxB} + SC_{AxC} + SC_{BxC} + SC_{AxBxC}$$

6. Calcula la suma de cuadrado del error experimental:

$$SC_{ErrorExp} = SC_{Total} - SC_{Tto}$$

Nota: Las réplicas (r) para cada suma de cuadrado depende del número de datos de cada, exclusivamente.

• ANAVA

FV	GL	SC	CM	f _{Calculado}	F _{Tabla, 5%,1%}
Tratamientos	$GL_A + GL_B + GL_C + GL_{AxB} + GL_{AxC} + GL_{BxC} + GL_{AxBxC}$	$SC_{Tto} (*)$	$\frac{SC_{Tratamientos}}{GL_{Tratamientos}}$	$\frac{CM_{Tratamientos}}{CM_{ErrorExp}}$	(GL_{Tto}, Gl_{Error})
A	NºNiveles(A)-1	$SC_A (*)$	$\frac{SC_A}{GL_A}$	$\frac{CM_A}{CM_{ErrorExp}}$	(GL_A, Gl_{Error})
B	NºNiveles(B)-1	$SC_B (*)$	$\frac{SC_B}{GL_B}$	$\frac{CM_B}{CM_{ErrorExp}}$	(GL_B, Gl_{Error})
C	NºNiveles(C)-1	$SC_C (*)$	$\frac{SC_C}{GL_C}$	$\frac{CM_C}{CM_{ErrorExp}}$	(GL_C, Gl_{Error})
AxB	$GL_A * GL_B$	$SC_{AxB} (*)$	$\frac{SC_{AxB}}{GL_{AxB}}$	$\frac{CM_{AxB}}{CM_{ErrorExp}}$	(GL_{AxB}, Gl_{Error})
AxC	$GL_A * GL_C$	$SC_{AxC} (*)$	$\frac{SC_{AxC}}{GL_{AxC}}$	$\frac{CM_{AxC}}{CM_{ErrorExp}}$	(GL_{AxC}, Gl_{Error})

BxC	$GL_B * GL_C$	$SC_{BxC} (*)$	$\frac{SC_{BxC}}{GL_{BxC}}$	$\frac{CM_{BxC}}{CM_{ErrorExp}}$	(GL_{BxC}, GI_{Error})
AxBxC	$GL_A * GL_B * GL_C$	$SC_{AxBxC} (*)$	$\frac{SC_{AxBxC}}{GL_{AxBxC}}$	$\frac{CM_{AxBxC}}{CM_{ErrorExp}}$	(GL_{AxBxC}, GI_{Error})
Error Exp.	$GL_{Total} - GL_{Tratamientos}$	$SC_{Total} - SC_{Tratamientos}$	$\frac{SC_{ErrorExp}}{GL_{ErrorExp}}$		
Total	n-1	$\sum y_{ijs}^2 - \frac{(\sum y_{ijs})^2}{n}$			

(*) Cálculos definidos antes de la construcción de la ANAVA.

Si $f_{calc} < f_{tabla(0.05,0.01)}$: El modelo no es significativo.

Si $f_{calc} \geq f_{tabla(0.05)}$: El modelo es significativo y se denota al lado del valor de F calculado con un asterisco (*), indicando que el modelo se explica con un 95% de confiabilidad.

Si $f_{calc} \geq f_{tabla(0.01)}$: El modelo es altamente significativo y se denota al lado del valor de F calculado con dos asteriscos (**), indicando que el modelo se explica con un 99% de confiabilidad.

Ejercicio de Diseño experimental en Arreglo Factorial:

Un investigador dedicado a la investigación en Biotecnología Vegetal *In vitro* está realizando un protocolo para la propagación masiva de cierta especie de Heliconia; dicho protocolo lo basó en función de 2 niveles de concentración de auxinas, 2 de citoquininas y 2 de azúcar no refinada.

El diseño a establecer es del tipo arreglo factorial 2^3 (2x2x2), planteándose 8 tratamientos provenientes de la combinación de los factores anteriormente mencionados.

La variable medida fue tasa de Supervivencia de Embriones, dado el balance de las sustancias obteniéndose lo siguiente:

Tratamiento			Repeticiones					Total
Auxina	Citoquinina	Azúcar	1	2	3	4	5	

1	1	1	16.3	16.6	15.9	16.7	16.9	82.4
1	1	2	15.4	15.5	15.6	11.5	12.3	70.3
1	2	1	14.8	14.3	12.1	11.8	13.5	66.5
1	2	2	12.2	13.3	14.5	16.1	15.4	71.5
2	1	1	13.3	14.5	15.6	15.4	14.8	73.6
2	1	2	14.1	16.4	16.1	15.9	15.8	78.3
2	2	1	17.3	18.4	18.8	17.4	16.3	88.2
2	2	2	12.1	10.1	11.8	12	12.1	58.1

1. Se calcula la suma de cuadrado total:

$$SC_{Total} = 16.3^2 + 16.6^2 + \dots + 12.1^2 - \frac{(16.3 + 16.6 + \dots + 12.1)^2}{40}$$

$$\rightarrow SC_{Total} = 8840.59 - 8670.08 = 170.51$$

2. Se calculan las sumas de cuadrados para cada factor, individualmente:

Para Auxina:

$$SC.A = \frac{(82.4 + 70.3 + 66.5 + 71.5)^2 + (73.6 + 78.3 + 88.2 + 58.1)^2}{20} - 8670.08 = 1.4063$$

Para Citoquinina:

$$SC.B = \frac{(82.4 + 70.3 + 73.6 + 78.3)^2 + (66.5 + 71.5 + 88.2 + 58.1)^2}{20} - 8670.08 = 10.302$$

Para Azúcar:

$$SC.C = \frac{(82.4 + 66.5 + 73.6 + 88.2)^2 + (70.3 + 71.5 + 78.3 + 58.1)^2}{20} - 8670.08 = 26.406$$

3. Se calculan las sumas de cuadrados para las interacciones simples (entre pares de factores):

Auxina x Citoquinina:

$$SC.AxB = \frac{(82.4 + 70.3)^2 + (66.5 + 71.5)^2 + (73.6 + 78.3)^2 + (88.2 + 58.1)^2}{10} - 8670.08 - 1.4063 - 10.302 = 2.07$$

Auxina x Azúcar:

$$SC.AxC = \frac{(82.4 + 66.5)^2 + (70.3 + 71.5)^2 + (73.6 + 88.2)^2 + (78.3 + 58.1)^2}{10} - 8670.08 - 1.4063 - 26.406 = 8.37$$

Citoquinina x Azúcar:

$$SC.BxC = \frac{(82.4 + 73.6)^2 + (70.3 + 78.3)^2 + (66.5 + 88.2)^2 + (71.5 + 58.1)^2}{10} - 8670.08 - 10.302 - 26.406 = 7.83$$

4. Se calcula la suma de cuadrados para la interacción de segundo orden; es decir, la interacción de los 3 factores bajo investigación:

$$SC.AxBxC = \frac{82.4^2 + 70.3^2 + 66.5^2 + 71.5^2 + 73.6^2 + 78.3^2 + 88.2^2 + 58.1^2}{5} - 8670.08 - 1.4063 - 10.302 - 26.406 - 2.07 - 8.37 - 7.83 = 67.34$$

5. Se halla la suma de cuadrado del tratamiento que es la sumatoria de las anteriores sumas de cuadrados, salvo la suma de cuadrado total:

$$SC.Trata = 1.4063 + 10.302 + 26.406 + 2.07 + 8.37 + 7.83 + 67.34 = 123.724$$

6. Se halla la suma de cuadrado del error experimental, a través de la diferencia entre la suma de cuadrado total y la suma de cuadrado del tratamiento:

$$SC.Error.Exp = 170.51 - 123.724 = 46.79$$

7. Se construye el Análisis de Varianza (ANAVA):

- ANAVA

FV	GL	SC	CM	F calculado	F tabla 5% 1%
----	----	----	----	-------------	---------------

Tratamiento	7	123.724	$\frac{123.724}{7} = 17.7$	$\frac{17.7}{1.46} = 12.12^{**}$	(7,32)=2.33 3.30
A	1	1.4063	1.4063	$\frac{1.4063}{1.46} = 0.96$	(1,32)=4.17 7.56
B	1	10.302	10.302	$\frac{10.302}{1.46} = 7.06^*$	(1,32)=4.17 7.56
C	1	26.406	26.406	$\frac{26.406}{1.46} = 18.09^{**}$	(1,32)=4.17 7.56
AxB	1	2.07	2.07	$\frac{2.07}{1.46} = 1.42$	(1,32)=4.17 7.56
AxC	1	8.37	8.37	$\frac{8.37}{1.46} = 5.73^*$	(1,32)=4.17 7.56
BxC	1	7.83	7.83	$\frac{7.83}{1.46} = 5.36^*$	(1,32)=4.17 7.56
AxBxC	1	67.34	67.34	$\frac{67.34}{1.46} = 46.12^{**}$	(1,32)=4.17 7.56
Error Experimental	39-7=32	46.79	$\frac{46.79}{32} = 1.46$		
Total	39	170.51			

Interpretación: Existe un efecto altamente significativo para las fuentes de variación “Tratamiento” y para el factor “Azúcar” y para la interacción de segundo orden; mientras que existe efecto significativo para la fuente de variación “citoquinina” y las interacciones “Auxina x Azúcar” y “Citoquinina x Azúcar”.

Referencias

1. BUENO, Eduardo. La Gestión del Conocimiento, Nuevos Perfiles Profesionales. 1999.

2. CAMACHO, J. Incubadoras o Viveros de Empresas de Base Tecnológica: La reciente experiencia europea como referencia para las actuales y futuras iniciativas latinoamericanas. XII Congreso Latinoamericano sobre Espíritu Empresarial. 1998.
3. HILLER, L.G. INTRODUCCIÓN A LA INVESTIGACION DE OPERACIONES. Mc GRAW HILL, 1995
4. NELSON, R. (ed.) National Innovation Systems. Oxford University Press. Oxford. 1993.
5. MONTGOMERY, Douglas. DESIGN AND ANALYSIS OF EXPERIMENTS. 5TH EDITION. JOHN WILEY AND SONS INC, 2001
6. WAISSBLUTH, M. ET AL. Administración de Proyectos de Innovación Tecnológica. Centro para la Innovación Tecnológica UNAM y Ediciones Gernika. México. 1986.
7. WINSTON, WAYNE. INVESTIGACION DE OPERACIONES, APLICACIONES Y ALGORITMOS, EDICION, EDITORIAL THOMPSON
8. WRIGHT, Paul. Introducción a la Ingeniería: 2ª Edición, México Limusa.Wiley, 2004