



Universidad Católica del Maule
Ingeniería en Estadística

Diseño factorial general: Experimentos de tres factores con análisis de resultados.

Felipe Neira Rojas & Angel Llanos Herrera

Profesora: Carolina Marchant Fuentes

Un informe de Diseño de Experimento
IES 413
Ingeniería en Estadística

3 de julio de 2025

Índice general

1. Introducción al Diseño de Experimentos	1
1.1. ¿Qué es el Diseño de Experimento?	1
1.2. Objetivo del Diseño de Experimento	1
1.3. Importancia del control, aleatorización y replicación	1
1.3.1. Control local	1
1.3.2. Aleatorización	2
1.3.3. Replicación	2
2. Diseño Factorial	3
2.1. ¿Qué es el Diseño Factorial?	3
2.2. Comparación con otros diseños (ej. unifactoriales)	3
2.3. Historia y evolución del diseño factorial	3
2.4. Ventajas y limitaciones del enfoque factorial	4
2.5. Aplicaciones comunes en la industria	5
3. Tipos de Diseños Factoriales	6
4. Diseño Factorial General: Aplicación para 3 factores	7
4.1. Definición del diseño	7
4.2. Modelo general del diseño	7
4.3. Hipótesis del modelo	8
4.4. Supuestos y Condición de Estimabilidad	9
4.4.1. Supuestos del modelo	9
4.4.2. Comprobación de Supuestos	9
4.4.3. Condición de Estimabilidad	10
4.5. Estimación de parámetros	10
4.5.1. Función de Máxima Verosimilitud y Log-Verosimilitud	10
4.5.2. Para σ^2	15
4.5.3. Resumen de estimaciones finales	17
4.5.4. Suma Cuadrática Total	17
4.5.5. Suma Cuadrática por Factor	17
4.5.6. Suma Cuadrática por Intervención de segundo orden	17
4.5.7. Suma Cuadrática por Intervención de tercer orden	18
4.5.8. Suma Cuadrática del Error	18
4.5.9. Cálculo del Cuadrado Medio	18
4.5.10. Cálculo del Estadístico F	18
4.6. Tabla ANOVA y Aplicación	19
4.7. Región de Rechazo	19
4.8. Comparaciones Múltiples	20

4.8.1. Prueba de Duncan	20
4.9. Imputación de valores faltantes	20
4.10. Utilidad del Diseño Factorial General	20
4.11. Ejercicio	21
5. Referencias	34

Capítulo 1

Introducción al Diseño de Experimentos

El diseño de experimento nace en el año 1926, en el cual Ronald A. Fisher publica el artículo "The arrangement of Field Experiments", donde, describió tres componentes principales en los diseños: control local, replicación y aleatorización. También, organizó y desarrolló el análisis de varianza para el diseño de experimento.

1.1. ¿Qué es el Diseño de Experimento?

El diseño de experimento es una estrategia estadística que permite planificar y analizar un experimento, para su posterior interpretación. Con el fin de evaluar el efecto de una o más variables (factores) sobre una variable de interés (variable de respuesta). También, permite cuantificar el impacto de esos factores, minimizando la variabilidad del experimento (error experimental) para llegar a conclusiones e interpretaciones válidas.

1.2. Objetivo del Diseño de Experimento

El objetivo principal del diseño de experimento es evaluar y cuantificar el efecto de uno o más factores sobre una variable de interés, a través de un proceso con análisis controlado de un conjunto de observaciones.

1.3. Importancia del control, aleatorización y replicación

La importancia de estos componentes principales definidos para el diseño de experimento son:

1.3.1. Control local

Estrategia para reducir la variabilidad del experimento (error experimental), mediante condiciones constantes, diseño estructurado y la aplicación de técnicas de ser necesario (ej. bloques) para mejorar la precisión del experimento.

1.3.2. Aleatorización

Asignación aleatoria de los tratamientos a las unidades experimentales (UE). Proporcionando que las estimaciones sobre la varianza del error experimental sean válidas. Esto, elimina sesgo en la selección, distribuyendo equitativamente los efectos no controlables en las unidades experimentales.

Esta aleatorización permite aplicar modelos estadísticos válidos, ya que puede garantizar la independencia de los errores.

1.3.3. Replicación

Aplicación del mismo tratamiento en varias unidades experimentales de manera independiente. Es un medio par estimar la varianza del error experimental. Su importancia recae en que permiten tener una mayor precisión en la estimación. Debido a que en cada tratamiento aplicado sobre una unidad experimental existe una variabilidad.

Capítulo 2

Diseño Factorial

2.1. ¿Qué es el Diseño Factorial?

Un diseño factorial es un tipo de diseño experimental, el cual considera el estudio de dos o más factores al mismo tiempo, considerando también las combinaciones posibles entre sus niveles. Donde, cada tratamiento en la interacción corresponde a una combinación.

2.2. Comparación con otros diseños (ej. unifactoriales)

Debido a lo explicado anteriormente. El diseño factorial no solo se estudian los efectos individuales, sino que también la influencia de la interacción entre niveles. Al comparar, un diseño unifactorial (ej. ANOVA ONE WAY) existe solo un factor de estudio; como el efecto de 3 diferentes fertilizantes. Y un diseño factorial (ej. 2 factores) existen 2 o más factores de estudio; como el efecto de 3 diferentes fertilizantes dado 2 diferentes tipos de riego. Permitiendo así detectar las interacciones entre factores (¿Cuál es el efecto del fertilizante A en el tipo de suelo 1?) en el diseño factorial.

Aquí, el número de tratamientos estaría dado por:

$$\text{número de tratamientos} = \prod_{i=1}^{n_F} n_{fi}$$

Donde,

n_F , número de factores.

n_{fi} , número de niveles en el factor i.

En el ejemplo con 3 niveles el factor fertilizantes y 2 niveles en el factor tipo de suelo:

$$\text{número de tratamientos} = \prod_{i=1}^{n_F=2} n_{fi} = 3 \cdot 2 = 6 \text{ tratamientos}$$

2.3. Historia y evolución del diseño factorial

Su origen recae en la agricultura, donde gracias al trabajo de Ronald A. Fisher el diseño factorial nace como respuesta al problema de evaluar múltiples factores que afectan a los cultivos con datos confiables. En ese contacto, donde se ve enfrentado a la variabilidad natural de los cultivos. Fisher tuvo que desarrollar un método para controlar aquella variación, permitiendo identificar efectos válidos entre tratamientos agrícolas. Con ello, publica un artículo "The

Arrangement of Field Experiments.^{en} 1926, donde, explica principios fundamentales; aleatorización, replicación y control local. También, define el análisis de varianza (ANOVA) como técnica estadística para comparar tratamientos y establece el diseño factorial completo como un método para evaluar múltiples factores a la vez.

Más tarde, entre 1930 y 1950 se desarrollan métodos para el análisis y simplificación de diseños factoriales, como por ejemplo los diseños fraccionados. Al igual que se consolidan modelos como: Diseño completamente aleatorizado, Diseño en bloques, Diseño factorial completo y fraccionado. Convirtiéndose en modelos estándar en la investigación agrícola, biológica y química.

Hasta llegar a la actualidad, que desde los 2000 el diseño factorial se ha expandido hacia áreas como las ciencias sociales, para el estudio de múltiples tratamientos o políticas, en áreas de finanzas, para comportamientos del consumidor dado múltiples factores.

Entonces, el diseño factorial se ha convertido en una herramienta aplicable a múltiples áreas y sectores, desde la experimentación agrícola hasta el estudio del comportamiento de las personas. Su fortaleza radica en su capacidad para analizar interacciones entre factores.

2.4. Ventajas y limitaciones del enfoque factorial

Ventajas

- El enfoque factorial permite evaluar múltiples factores de manera simultánea, teniendo una mayor eficiencia que al evaluar estos factores por separado en experimentos diferentes.
- Además de evaluar cada factor por sí solo, permite evaluar interacciones entre estos, indicando si el efecto de un factor depende del nivel de otro factor, generando un análisis más robusto.
- Mayor representatividad de la realidad, ya que en la mayoría de los casos de estudio intervienen múltiples factores de manera simultánea.
- Con base en el punto anterior, mientras mayor sea la cantidad de factores estudiados, se explicará una mayor variabilidad del diseño.

Limitaciones

- A mayor cantidad de factores y niveles estudiados, mayor será la cantidad de observaciones que se analizará, lo que puede generar altos costos.
- Una gran cantidad de niveles y factores puede generar complejidad en la interpretación de las interacciones.
- Proceso costoso en tiempo y dinero al evaluar la mayor cantidad de factores posible.

2.5. Aplicaciones comunes en la industria

El enfoque factorial puede utilizarse para realizar diseños experimentales en variadas industrias, como por ejemplo:

Medicina

Se puede utilizar para evaluar el efecto de un medicamento o tratamiento con base en distintos factores, como la dosis administrada, la frecuencia de aplicación y el tipo de paciente (por ejemplo edad o condición de salud), permitiendo analizar tanto los efectos individuales de cada factor como las posibles interacciones entre ellos.

Manufacturera

En controles de calidad de diversos productos, podemos evaluar esto según maquinaria utilizada, fábrica, tipo de empleado a cargo (ya sea evaluar edad, sexo o puesto de trabajo que posee), para conocer qué factor puede estar afectando y tomar medidas correctoras para la mejora continua de la empresa.

Alimenticia

En la industria alimenticia puede evaluarse el peso de los alimentos empaquetados con base en la maquinaria, tiempos de cocción, proporción de los ingredientes, etc.

Agricultura

Casos comunes analizados en agricultura es como afectan factores como temperaturas, uso de fertilizantes, tipo de suelo, cantidad de regadío o de luz en el crecimiento o desarrollo de una plantación.

Capítulo 3

Tipos de Diseños Factoriales

Existen distintos tipos de diseños factoriales, los cuales se elegirán en función del caso a estudiar, la cantidad de factores a evaluar, el número de niveles considerados, la presencia de bloques para controlar la variabilidad, o la necesidad de recurrir a diseños cuando los recursos no permiten ejecutar todas las combinaciones posibles, ya sea por temas de costos o tiempo. Algunos tipos de diseños factoriales que se pueden utilizar son los siguientes.

Diseño Factorial General

Considera estudiar todos los factores que el experimento desee, tomando en cuenta todos los niveles de cada factor, donde se buscará estudiar al factor por sí solo y la interacción entre sus niveles.

Diseño Factorial 2^k

Este diseño estudia k cantidad de factores, los cuales poseen solamente 2 niveles por factor. Al igual que el Diseño Factorial General desee estudiar al factor por sí solo y la interacción entre sus niveles. Este diseño se puede utilizar en presencia de niveles dicotómicos.

Formación de bloques

Se desea controlar la variabilidad mediante el análisis de bloques aleatorios.

Diseños factoriales fraccionados de dos niveles

Se utiliza cuando la cantidad de factores es muy alto, en dónde solo se considera una fracción del diseño 2^k , en dónde se reduce la cantidad de interacciones analizadas.

Diseños factoriales fraccionados de tres niveles

Extensión del diseño 2^k en dónde cada factor tiene 3 niveles, en dónde nuevamente se considera solamente una fracción del diseño a analizar.

Diseños factoriales con niveles mixtos

Se considera este diseño cuando se desean analizar diversos factores con una cantidad de niveles distintos. Para este modelo se debe considerar la cantidad de combinaciones para que se estudien todas las interacciones posibles.

Capítulo 4

Diseño Factorial General: Aplicación para 3 factores

4.1. Definición del diseño

El Diseño Factorial General (DFG) se utiliza para analizar el efecto de dos o más factores sobre una variable de respuesta. A diferencia de los diseños unifactoriales, este enfoque permite evaluar no solo los efectos principales de cada factor por separado, sino también las interacciones que puedan existir entre ellos. Cada factor puede tener un número distinto de niveles, siempre que cuente con al menos dos niveles. Así, en un diseño factorial general con tres factores, se consideran: el factor A con a niveles, el factor B con b niveles y el factor C con c niveles, cumpliéndose que $r, s, t \geq 2$.

El análisis ANOVA Factorial de DFG busca:

- **Analizar los efectos principales de cada factor:** Permite determinar si cada factor, considerado de forma individual, ejerce una influencia significativa sobre la variable de respuesta.
- **Detectar interacciones entre factores:** Consiste en identificar si el efecto de un factor varía en función del nivel de otro, lo que evidencia una dependencia entre factores, sugiriendo que no actúan de manera independiente.

4.2. Modelo general del diseño

Se plantea un Modelo ANOVA factorial para la variable de respuesta y_{ijkl} , considerando 3 factores:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

- y_{ijkl} : Variable de respuesta correspondiente al nivel i del factor A, nivel j del factor B, nivel k del factor C, en la réplica l
- μ : Media global
- α_i : Efecto del nivel i del factor A, con $i = 1, 2, \dots, r$, donde $(a \geq 2)$
- β_j : Efecto del nivel j del factor B, con $j = 1, 2, \dots, s$, donde $(b \geq 2)$

- γ_k : Efecto del nivel k del factor C, con $k = 1, 2, \dots, t$, donde $(c \geq 2)$
- $(\alpha\beta)_{ij}$: Efecto de la interacción entre A y B
- $(\alpha\gamma)_{ik}$: Efecto de la interacción entre A y C
- $(\beta\gamma)_{jk}$: Efecto de la interacción entre B y C
- $(\alpha\beta\gamma)_{ijk}$: Efecto de la interacción entre A, B y C
- ε_{ijkl} : Error aleatorio en el nivel i del factor A, nivel j del factor B, nivel k del factor C, en la observación l .

4.3. Hipótesis del modelo

Para el planteamiento del modelo ANOVA en Diseño Factorial General, se evalúan múltiples hipótesis, las cuales buscan conocer si los factores influyen en los modelos por si solos, evaluando además todas sus posibles combinaciones. Por lo tanto, la cantidad de hipótesis planteadas dependerá de la cantidad de factores a estudiar. Para el caso de DFG de 3 factores las hipótesis serían las siguientes:

- **Por factor:**

- Factor A:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a = 0$$

$$H_1 : \text{Al menos un } \alpha_i \neq 0$$

- Factor B:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_b = 0$$

$$H_1 : \text{Al menos un } \beta_j \neq 0$$

- Factor C:

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_c = 0$$

$$H_1 : \text{Al menos un } \gamma_k \neq 0$$

- **Interacciones entre 2 factores:**

- Factores A y B:

$$H_0 : (\alpha\beta)_{ij} = 0 \quad \forall i, j$$

$$H_1 : \text{Existe al menos un } (\alpha\beta)_{ij} \neq 0$$

- Factores A y C:

$$H_0 : (\alpha\gamma)_{ik} = 0 \quad \forall i, k$$

$$H_1 : \text{Existe al menos un } (\alpha\gamma)_{ik} \neq 0$$

- Factores B y C:

$$H_0 : (\beta\gamma)_{jk} = 0 \quad \forall j, k$$

$$H_1 : \text{Existe al menos un } (\beta\gamma)_{jk} \neq 0$$

■ **Interacción entre tres factores:**

$$H_0 : (\alpha\beta\gamma)_{ijk} = 0 \quad \forall i, j, k$$

$$H_1 : \text{Existe al menos un } (\alpha\beta\gamma)_{ijk} \neq 0$$

Es importante destacar que a , b y c son la cantidad de niveles de los factores A, B y C respectivamente. Además, considerar que los subíndices i , j , k pueden tomar los siguientes valores:

$$i = 1, 2, \dots, a$$

$$j = 1, 2, \dots, b$$

$$k = 1, 2, \dots, c$$

4.4. Supuestos y Condición de Estimabilidad

4.4.1. Supuestos del modelo

El análisis de DFG se realiza bajo los siguientes supuestos:

$$\varepsilon_{ijkl} \sim \mathcal{N}(0, \sigma^2), \quad \forall i, j, k, l$$

- **Normalidad de los Residuos:** Los errores de cada tratamiento siguen una distribución normal con media 0 y varianza σ^2 constante.
- **Homocedasticidad:** La varianza de los residuos es constante.

1Nota: Los errores deben ser independientes, condición que se cumple si se respeta la estructura del Diseño Factorial General, el cual exige una asignación aleatoria de los tratamientos a las unidades experimentales.

4.4.2. Comprobación de Supuestos

Para la comprobación de los supuestos del modelo ANOVA, podemos utilizar los siguientes test:

Supuesto	Test	Estadístico de prueba	Hipótesis
Normalidad	Shapiro–Wilk	W	H_0 : los residuos provienen de una distribución normal. H_1 : los residuos no provienen de una distribución normal.
	Lilliefors	D	H_0 : los residuos provienen de una distribución normal. H_1 : los residuos no provienen de una distribución normal.
Homogeneidad de varianzas	Bartlett	χ^2	H_0 : todas las varianzas poblacionales son iguales (homocedasticidad). H_1 : al menos una varianza poblacional difiere (heterocedasticidad).

Table 4.1: Pruebas estadísticas para verificar supuestos del modelo ANOVA

4.4.3. Condición de Estimabilidad

Para que los parámetros del modelo factorial general sean estimables de forma única mediante mínimos cuadrados, deben cumplirse las siguientes condiciones de estimabilidad:

Tipo de efecto	Condición de Estimabilidad
Efectos principales	$\sum_{i=1}^a \hat{\alpha}_i = 0$ $\sum_{j=1}^b \hat{\beta}_j = 0$ $\sum_{k=1}^c \hat{\gamma}_k = 0$
Interacciones dobles	$\sum_{i=1}^a \sum_{j=1}^b (\widehat{\alpha\beta})_{ij} = 0$ $\sum_{i=1}^a \sum_{k=1}^c (\widehat{\alpha\gamma})_{ik} = 0$ $\sum_{j=1}^b \sum_{k=1}^c (\widehat{\beta\gamma})_{jk} = 0$
Interacción triple	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (\widehat{\alpha\beta\gamma})_{ijk} = 0$

Table 4.2: Condiciones de estimabilidad en el modelo factorial general de 3 factores

4.5. Estimación de parámetros

4.5.1. Función de Máxima Verosimilitud y Log-Verosimilitud

La estimación de parámetros se llevará a cabo mediante el método de máxima verosimilitud considerando nuestro vector de parámetros como:

$$\theta = (\mu, \alpha_i, \beta_j, \gamma_k, (\alpha\beta)_{ij}, (\alpha\gamma)_{ik}, (\beta\gamma)_{jk}, (\alpha\beta\gamma)_{ijk}, \sigma^2)^\top, \quad i = 1, \dots, a; \quad j = 1, \dots, b; \quad k = 1, \dots, c$$

$$\varepsilon_{ijkl} \sim \mathcal{N}(0, \sigma^2) \Rightarrow y_{ijkl} \sim \mathcal{N}(\mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}, \sigma^2)$$

Entonces, la función de verosimilitud para:

$$\theta = (\mu, \alpha_i, \beta_j, \gamma_k, (\alpha\beta)_{ij}, (\alpha\gamma)_{ik}, (\beta\gamma)_{jk}, (\alpha\beta\gamma)_{ijk}, \sigma^2)^\top$$

es:

$$L(\theta) = \prod_{i=1}^a \prod_{j=1}^b \prod_{k=1}^c \prod_{l=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_{ijkl} - \mu_{ijk})^2}{2\sigma^2}\right)$$

donde:

$$\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$$

Por lo tanto, se obtiene lo siguiente:

$$L(\theta) = \prod_{i=1}^a \prod_{j=1}^b \prod_{k=1}^c \prod_{l=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(y_{ijkl} - \mu - \alpha_i - \beta_j - \gamma_k - (\alpha\beta)_{ij} - (\alpha\gamma)_{ik} - (\beta\gamma)_{jk} - (\alpha\beta\gamma)_{ijk})^2}{2\sigma^2} \right)$$

A esto aplicamos función de log-verosimilitud, quedando:

$$\begin{aligned} \ell(\theta) &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{(y_{ijkl} - \mu - \alpha_i - \beta_j - \gamma_k - (\alpha\beta)_{ij} - (\alpha\gamma)_{ik} - (\beta\gamma)_{jk} - (\alpha\beta\gamma)_{ijk})^2}{2\sigma^2} \right] \\ &= -\frac{abcn}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \alpha_i - \beta_j - \gamma_k - (\alpha\beta)_{ij} - (\alpha\gamma)_{ik} - (\beta\gamma)_{jk} - (\alpha\beta\gamma)_{ijk})^2 \end{aligned}$$

Ya calculada la función de log-verosimilitud, se deriva esta función dependiendo del parámetro que se desea estimar. Es importante tener en cuenta las condiciones de estimabilidad definidas anteriormente.

Derivadas respecto a cada parámetro

Para la estimación de los parámetros $\theta_1 = [\mu, \alpha, \beta, \gamma, (\alpha\beta), (\alpha\gamma), (\beta\gamma), (\alpha\beta\gamma)]$, se aplica la derivada:

considerando $a * b * c * n = N$

$$-\frac{\partial N \log(2\pi\sigma^2)}{\partial \theta_1} - \frac{\partial}{\partial \theta_1} \frac{1}{2\sigma^2} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \alpha_i - \beta_j - \gamma_k - (\alpha\beta)_{ij} - (\alpha\gamma)_{ik} - (\beta\gamma)_{jk} - (\alpha\beta\gamma)_{ijk})^2$$

Al ser $-\frac{\partial N \log(2\pi\sigma^2)}{\partial \theta_1}$ una derivación de constante, por lo que se considera como valor 0.

Por lo tanto, se deriva la siguiente expresión e iguala a 0

$$\frac{\partial}{\partial \theta_1} \frac{1}{2\sigma^2} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \alpha_i - \beta_j - \gamma_k - (\alpha\beta)_{ij} - (\alpha\gamma)_{ik} - (\beta\gamma)_{jk} - (\alpha\beta\gamma)_{ijk})^2 = 0$$

$$\frac{1}{\sigma^2} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \alpha_i - \beta_j - \gamma_k - (\alpha\beta)_{ij} - (\alpha\gamma)_{ik} - (\beta\gamma)_{jk} - (\alpha\beta\gamma)_{ijk}) = 0$$

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \alpha_i - \beta_j - \gamma_k - (\alpha\beta)_{ij} - (\alpha\gamma)_{ik} - (\beta\gamma)_{jk} - (\alpha\beta\gamma)_{ijk}) = 0$$

Por lo tanto, ahora se considera el cálculo para cada parámetro, con esto se ven aplicadas las siguientes condiciones de estimabilidad:

$$\begin{aligned}\sum_{i=1}^a \hat{\alpha}_i &= 0, \sum_{j=1}^b \hat{\beta}_j = 0, \sum_{k=1}^c \hat{\gamma}_k = 0 \\ \sum_{i=1}^a \sum_{j=1}^b (\hat{\alpha\beta})_{ij} &= 0, \sum_{i=1}^a \sum_{k=1}^c (\hat{\alpha\gamma})_{ik} = 0, \sum_{j=1}^b \sum_{k=1}^c (\hat{\beta\gamma})_{jk} = 0 \\ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (\hat{\alpha\beta\gamma})_{ijk} &= 0\end{aligned}$$

Considerando la siguiente expresión:

$$\sum_{i,j,k,l}^N y_{ijkl} - \sum_{i,j,k,l}^N \mu - \sum_{i,j,k,l}^N \alpha_i - \sum_{i,j,k,l}^N \beta_j - \sum_{i,j,k,l}^N \gamma_k - \sum_{i,j,k,l}^N (\alpha\beta)_{ij} - \sum_{i,j,k,l}^N (\alpha\gamma)_{ik} - \sum_{i,j,k,l}^N (\beta\gamma)_{jk} - \sum_{i,j,k,l}^N (\alpha\beta\gamma)_{ijk} = 0$$

Para μ

$$\begin{aligned}\sum_{i,j,k,l}^N y_{ijkl} - \sum_{i,j,k,l}^N \mu &= 0 \\ \sum_{i,j,k,l}^N \mu &= \sum_{i,j,k,l}^N y_{ijkl} \\ N\mu &= \sum_{i,j,k,l}^N y_{ijkl} \\ \hat{\mu} &= \frac{\sum_{i,j,k,l}^N y_{ijkl}}{N} = \bar{y}_{....}\end{aligned}$$

Para α_i

Se considera la misma función utilizada, teniendo en cuenta que la sumatoria del índice evaluado, en este caso i se mantiene fijo para la estimación puntual, caso que ocurrirá para la estimación de los demás factores o interacciones.

$$\begin{aligned}\sum_{j,k,l}^{bcn} y_{ijkl} - \sum_{j,k,l}^{bcn} \mu - \sum_{j,k,l}^{bcn} \alpha_i &= 0 \\ \sum_{j,k,l}^{bcn} y_{ijkl} - bcn \cdot \mu - bcn \cdot \alpha_i &= 0 \\ \sum_{j,k,l}^{bcn} y_{ijkl} - bcn \cdot \mu &= bcn \cdot \alpha_i \\ \hat{\alpha}_i &= \frac{\sum_{j,k,l}^{bcn} y_{ijkl}}{bcn} - \hat{\mu} = \bar{y}_{i...} - \bar{y}_{....}\end{aligned}$$

Para β_j

$$\begin{aligned}\sum_{i,k,l}^{acn} y_{ijkl} - \sum_{i,k,l}^{acn} \mu - \sum_{i,k,l}^{acn} \beta_j &= 0 \\ \sum_{i,k,l}^{acn} y_{ijkl} - acn \cdot \mu - acn \cdot \beta_j &= 0 \\ \sum_{i,k,l}^{acn} y_{ijkl} - acn \cdot \mu &= acn \cdot \beta_j \\ \hat{\beta}_j &= \frac{\sum_{i,k,l}^{acn} y_{ijkl}}{acn} - \hat{\mu} = \bar{y}_{\cdot j \cdot \cdot} - \bar{y}_{\cdot \cdot \cdot \cdot}\end{aligned}$$

Para γ_k

$$\begin{aligned}\sum_{i,j,l}^{abn} y_{ijkl} - \sum_{i,j,l}^{abn} \mu - \sum_{i,j,l}^{abn} \gamma_k &= 0 \\ \sum_{i,j,l}^{abn} y_{ijkl} - abn \cdot \mu - abn \cdot \gamma_k &= 0 \\ \sum_{i,j,l}^{abn} y_{ijkl} - abn \cdot \mu &= abn \cdot \gamma_k \\ \hat{\gamma}_k &= \frac{\sum_{i,j,l}^{abn} y_{ijkl}}{abn} - \hat{\mu} = \bar{y}_{\cdot \cdot k \cdot} - \bar{y}_{\cdot \cdot \cdot \cdot}\end{aligned}$$

Para $(\alpha\beta)_{ij}$

$$\begin{aligned}\sum_{k,l}^{cn} y_{ijkl} - \sum_{k,l}^{cn} \mu - \sum_{k,l}^{cn} \alpha_i - \sum_{k,l}^{cn} \beta_j - \sum_{k,l}^{cn} (\alpha\beta)_{ij} &= 0 \\ cn \cdot (\alpha\beta)_{ij} &= \sum_{k,l}^{cn} y_{ijkl} - cn \cdot \mu - cn \cdot \alpha_i - cn \cdot \beta_j \\ (\hat{\alpha}\hat{\beta})_{ij} &= \frac{\sum_{k,l}^{cn} y_{ijkl}}{cn} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j \\ (\hat{\alpha}\hat{\beta})_{ij} &= \frac{\sum_{k,l}^{cn} y_{ijkl}}{cn} - \bar{y}_{\cdot \cdot \cdot \cdot} - (\bar{y}_{i \cdot \cdot \cdot} - \bar{y}_{\cdot \cdot \cdot \cdot}) - (\bar{y}_{\cdot j \cdot \cdot} - \bar{y}_{\cdot \cdot \cdot \cdot}) \\ (\hat{\alpha}\hat{\beta})_{ij} &= \bar{y}_{ij \cdot \cdot} - \bar{y}_{i \cdot \cdot \cdot} - \bar{y}_{\cdot j \cdot \cdot} + \bar{y}_{\cdot \cdot \cdot \cdot}\end{aligned}$$

Para $(\alpha\gamma)_{ik}$

$$\begin{aligned} \sum_{j,l}^{bn} y_{ijkl} - \sum_{j,l}^{bn} \mu - \sum_{j,l}^{bn} \alpha_i - \sum_{j,l}^{bn} \gamma_k - \sum_{j,l}^{bn} (\alpha\gamma)_{ik} &= 0 \\ bn \cdot (\alpha\gamma)_{ik} &= \sum_{j,l}^{bn} y_{ijkl} - bn \cdot \mu - bn \cdot \alpha_i - bn \cdot \gamma_k \\ (\hat{\alpha\gamma})_{ik} &= \frac{\sum_{j,l}^{bn} y_{ijkl}}{bn} - \hat{\mu} - \hat{\alpha}_i - \hat{\gamma}_k \\ (\hat{\alpha\gamma})_{ik} &= \frac{\sum_{j,l}^{bn} y_{ijkl}}{bn} - \bar{y}_{....} - (\bar{y}_{i...} - \bar{y}_{....}) - (\bar{y}_{..k} - \bar{y}_{....}) \\ (\hat{\alpha\gamma})_{ik} &= \bar{y}_{i..k} - \bar{y}_{i...} - \bar{y}_{..k} + \bar{y}_{....} \end{aligned}$$

Para $(\beta\gamma)_{jk}$

$$\begin{aligned} \sum_{i,l}^{an} y_{ijkl} - \sum_{i,l}^{an} \mu - \sum_{i,l}^{an} \beta_j - \sum_{i,l}^{an} \gamma_k - \sum_{i,l}^{an} (\beta\gamma)_{jk} &= 0 \\ an \cdot (\beta\gamma)_{jk} &= \sum_{i,l}^{an} y_{ijkl} - an \cdot \mu - an \cdot \beta_j - an \cdot \gamma_k \\ (\hat{\beta\gamma})_{jk} &= \frac{\sum_{i,l}^{an} y_{ijkl}}{an} - \hat{\mu} - \hat{\beta}_j - \hat{\gamma}_k \\ (\hat{\beta\gamma})_{jk} &= \frac{\sum_{i,l}^{an} y_{ijkl}}{an} - \bar{y}_{....} - (\bar{y}_{j..} - \bar{y}_{....}) - (\bar{y}_{..k} - \bar{y}_{....}) \\ (\hat{\beta\gamma})_{jk} &= \bar{y}_{.jk} - \bar{y}_{j..} - \bar{y}_{..k} + \bar{y}_{....} \end{aligned}$$

Para $(\alpha\beta\gamma)_{ijk}$

$$\begin{aligned} \sum_l^n y_{ijkl} - \sum_l^n \mu - \sum_l^n \alpha_i - \sum_l^n \beta_j - \sum_l^n \gamma_k - \sum_l^n (\alpha\beta)_{ij} - \sum_l^n (\alpha\gamma)_{ik} - \sum_l^n (\beta\gamma)_{jk} &= 0 \\ \sum_l^n y_{ijkl} - \sum_l^n \mu - \sum_l^n \alpha_i - \sum_l^n \beta_j - \sum_l^n \gamma_k - \sum_l^n (\alpha\beta)_{ij} - \sum_l^n (\alpha\gamma)_{ik} - \sum_l^n (\beta\gamma)_{jk} - \sum_l^n (\alpha\beta\gamma)_{ijk} &= 0 \\ \sum_l^n (\alpha\beta\gamma)_{ijk} &= \sum_l^n y_{ijkl} - \sum_l^n \mu - \sum_l^n \alpha_i - \sum_l^n \beta_j - \sum_l^n \gamma_k - \sum_l^n (\alpha\beta)_{ij} - \sum_l^n (\alpha\gamma)_{ik} - \sum_l^n (\beta\gamma)_{jk} \end{aligned}$$

$$n(\alpha\beta\gamma)_{ijk} = \sum_l^n y_{ijkl} - n\mu - n\alpha_i - n\beta_j - n\gamma_k - n(\alpha\beta)_{ij} - n(\alpha\gamma)_{ik} - n(\beta\gamma)_{jk}$$

$$(\alpha\beta\gamma)_{ijk} = \frac{\sum_l^n y_{ijkl}}{n} - \mu - \alpha_i - \beta_j - \gamma_k - (\alpha\beta)_{ij} - (\alpha\gamma)_{ik} - (\beta\gamma)_{jk}$$

$$(\hat{\alpha}\hat{\beta}\hat{\gamma})_{ij} = \bar{y}_{ijk.} - \bar{y}_{....} - (\bar{y}_{i...} - \bar{y}_{....}) - (\bar{y}_{.j..} - \bar{y}_{....}) - (\bar{y}_{..k.} - \bar{y}_{....}) - (\bar{y}_{ij..} - \bar{y}_{i...} - \bar{y}_{.j..} + \bar{y}_{....})$$

$$- (\bar{y}_{i.k.} - \bar{y}_{i...} - \bar{y}_{.k..} + \bar{y}_{....}) - (\bar{y}_{.jk.} - \bar{y}_{.j..} - \bar{y}_{..k.} + \bar{y}_{....})$$

$$(\hat{\alpha}\hat{\beta}\hat{\gamma})_{ij} = \bar{y}_{ijk.} - \bar{y}_{ij..} - \bar{y}_{i.k.} - \bar{y}_{.jk.} + \bar{y}_{i...} + \bar{y}_{.j..} + \bar{y}_{..k.} - \bar{y}_{....}$$

4.5.2. Para σ^2

$$\ell(\sigma^2) = -\frac{abcn}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n \left(y_{ijkl} - \mu - \alpha_i - \beta_j - \gamma_k - (\alpha\beta)_{ij} - (\alpha\gamma)_{ik} - (\beta\gamma)_{jk} - (\alpha\beta\gamma)_{ijk} \right)^2.$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{abcn}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n \left(y_{ijkl} - \mu - \alpha_i - \beta_j - \gamma_k - (\alpha\beta)_{ij} - (\alpha\gamma)_{ik} - (\beta\gamma)_{jk} - (\alpha\beta\gamma)_{ijk} \right)^2 = 0.$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{abcn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n \left(y_{ijkl} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k - \widehat{(\alpha\beta)}_{ij} - \widehat{(\alpha\gamma)}_{ik} - \widehat{(\beta\gamma)}_{jk} - \widehat{(\alpha\beta\gamma)}_{ijk} \right)^2$$

Error (residuo) y Suma de Cuadrados del Error

$$e_{ijkl} = y_{ijkl} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k - \widehat{(\alpha\beta)}_{ij} - \widehat{(\alpha\gamma)}_{ik} - \widehat{(\beta\gamma)}_{jk} - \widehat{(\alpha\beta\gamma)}_{ijk},$$

$$i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, c; l = 1, \dots, n.$$

$$SCE = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n e_{ijkl}^2$$

Grados de libertad del error

Número total de observaciones: $N = abcn$.

Número de parámetros ajustados (p):

$$p = 1 + (a-1) + (b-1) + (c-1) + (a-1)(b-1) + (a-1)(c-1) + (b-1)(c-1) + (a-1)(b-1)(c-1).$$

$$gl_E = N - p$$

CME y estimador de σ^2

$$\text{CME} = \frac{\text{SCE}}{\text{gl}_E} = \frac{1}{abcn - p} \sum_{i,j,k,l} e_{ijkl}^2.$$

Modelo balanceado (efectos fijos):

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}, \quad \varepsilon_{ijkl} \sim \mathcal{N}(0, \sigma^2),$$

con las restricciones de centrado $\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = \dots = 0$.

$$\hat{\mu} = \bar{y}_{....} = \frac{1}{abcn} \sum_{i,j,k,l} y_{ijkl}, \quad \hat{\sigma}_{\text{MLE}}^2 = \frac{1}{abcn} \sum_{i,j,k,l} \hat{e}_{ijkl}^2, \quad N = abcn.$$

1. Insesgadez de $\hat{\mu}$

$$\begin{aligned} E[\hat{\mu}] &= \frac{1}{N} \sum_{i,j,k,l} E[y_{ijkl}] \\ &= \frac{1}{N} \sum_{i,j,k,l} (\mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}) \\ &= \mu + \underbrace{\frac{1}{N} \sum_i \alpha_i \sum_{j,k,l} 1}_{=0} + \underbrace{\frac{1}{N} \sum_j \beta_j \sum_{i,k,l} 1}_{=0} + \dots \\ &= \mu. \end{aligned}$$

$$E[\hat{\mu}] = \mu \implies \hat{\mu} \text{ es insesgado}$$

Insesgadez de la varianza

Consideramos a CME como estimador de la varianza para comprobar su insesgadez

$$\hat{\sigma}^2 = \frac{\text{SCE}}{\text{gl}_E} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n e_{ijkl}^2}{N - p}, \quad N = abcn$$

$$\frac{\text{SCE}}{\sigma^2} = \frac{\sum_{i,j,k,l} e_{ijkl}^2}{\sigma^2} \sim \chi_{\text{gl}_E}^2 \quad (\text{distribución ji-cuadrada con } N - p \text{ grados de libertad}).$$

Por la propiedad básica de la ji-cuadrada:

$$E[\chi_\nu^2] = \nu \implies E\left[\frac{\text{SCE}}{\sigma^2}\right] = N - p.$$

Por lo tanto, calculamos la esperanza del Cuadrado Medio del Error (CME)

$$E[\hat{\sigma}_{\text{CME}}^2] = E\left[\frac{\text{SCE}}{N-p}\right] = \frac{1}{N-p} E[\text{SCE}] = \frac{1}{N-p} [(N-p)\sigma^2] = \sigma^2.$$

$$E[\hat{\sigma}_{\text{CME}}^2] = \sigma^2 \implies \hat{\sigma}_{\text{CME}}^2 \text{ es insesgado.}$$

4.5.3. Resumen de estimaciones finales

Por lo tanto, de la estimación por máxima verosimilitud se obtienen los siguientes valores a estudiar para el análisis de varianzas:

$$\hat{\mu} = \bar{y}_{...}$$

$$\hat{\alpha}_i = \bar{y}_{i...} - \bar{y}_{...}, \quad \hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...}, \quad \hat{\gamma}_k = \bar{y}_{..k} - \bar{y}_{...}$$

$$\widehat{(\alpha\beta)}_{ij} = \bar{y}_{ij..} - \bar{y}_{i...} - \bar{y}_{.j.} + \bar{y}_{...}$$

$$\widehat{(\alpha\gamma)}_{ik} = \bar{y}_{i.k.} - \bar{y}_{i...} - \bar{y}_{..k} + \bar{y}_{...}$$

$$\widehat{(\beta\gamma)}_{jk} = \bar{y}_{.jk.} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...}$$

$$\widehat{(\alpha\beta\gamma)}_{ijk} = \bar{y}_{ijk.} - \bar{y}_{ij..} - \bar{y}_{i.k.} - \bar{y}_{.jk.} + \bar{y}_{i...} + \bar{y}_{.j.} + \bar{y}_{..k} - \bar{y}_{...}$$

4.5.4. Suma Cuadrática Total

$$SC_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n y_{ijkl}^2 - \frac{y_{....}^2}{abcn}$$

4.5.5. Suma Cuadrática por Factor

$$SC_A = \frac{1}{bcn} \sum_{i=1}^a y_{i...}^2 - \frac{y_{....}^2}{abcn}$$

$$SC_B = \frac{1}{acn} \sum_{j=1}^b y_{.j.}^2 - \frac{y_{....}^2}{abcn}$$

$$SC_C = \frac{1}{abn} \sum_{k=1}^c y_{..k.}^2 - \frac{y_{....}^2}{abcn}$$

4.5.6. Suma Cuadrática por Intervención de segundo orden

$$SC_{AB} = \left(\frac{1}{cn} \sum_{i=1}^a \sum_{j=1}^b y_{ij..}^2 - \frac{y_{....}^2}{abcn} \right) - SC_A - SC_B$$

$$SC_{AC} = \left(\frac{1}{bn} \sum_{i=1}^a \sum_{k=1}^c y_{i.k.}^2 - \frac{y_{....}^2}{abcn} \right) - SC_A - SC_C$$

$$SC_{BC} = \left(\frac{1}{an} \sum_{j=1}^b \sum_{k=1}^c y_{.jk.}^2 - \frac{y_{....}^2}{abcn} \right) - SC_B - SC_C$$

4.5.7. Suma Cuadrática por Intervención de tercer orden

$$SC_{ABC} = \left(\frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c y_{ijk}^2 - \frac{y_{\dots}^2}{abcn} \right) - SC_A - SC_B - SC_C - SC_{AB} - SC_{AC} - SC_{BC}$$

4.5.8. Suma Cuadrática del Error

$$SC_E = SC_T - \left(\frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c y_{ijk}^2 - \frac{y_{\dots}^2}{abcn} \right)$$

4.5.9. Cálculo del Cuadrado Medio

Por otra parte el cálculo del Cuadrado Medio corresponderá a:

$$\frac{SC_{tratamiento}}{g \cdot l_{tratamiento}}$$

4.5.10. Cálculo del Estadístico F

Por último el cálculo del estadístico F se realiza de la siguiente formula.

$$F_0 = \frac{CM_{tratamiento}}{CM_{error}}$$

el cual sigue la siguiente distribución:

$$F_0 \sim F_{(1-\phi), (y), (u)}$$

en dónde:

ϕ : corresponde al nivel de significancia.

y : Grados de libertad del caso a evaluar. (Si se desea evaluar un factor corresponde a sus grados de libertad, en cambio si desea evaluar la interacción corresponde a la multiplicación de los grados de libertad de los factores).

u : Grados de libertad del error.

4.6. Tabla ANOVA y Aplicación

Fuente de variación	GL	Suma de cuadrados (SC)	Cuadrado medio (CM)	Estadístico F	Valor p
Factor A	$a - 1$	SC_A	$\frac{SC_A}{a - 1}$	$\frac{CM_A}{CM_E}$	p_A
Factor B	$b - 1$	SC_B	$\frac{SC_B}{b - 1}$	$\frac{CM_B}{CM_E}$	p_B
Factor C	$c - 1$	SC_C	$\frac{SC_C}{c - 1}$	$\frac{CM_C}{CM_E}$	p_C
Interacción A \times B	$(a-1)(b-1)$	SC_{AB}	$\frac{SC_{AB}}{(a-1)(b-1)}$	$\frac{CM_{AB}}{CM_E}$	p_{AB}
Interacción A \times C	$(a-1)(c-1)$	SC_{AC}	$\frac{SC_{AC}}{(a-1)(c-1)}$	$\frac{CM_{AC}}{CM_E}$	p_{AC}
Interacción B \times C	$(b-1)(c-1)$	SC_{BC}	$\frac{SC_{BC}}{(b-1)(c-1)}$	$\frac{CM_{BC}}{CM_E}$	p_{BC}
Interacción A \times B \times C	$(a-1)(b-1)(c-1)$	SC_{ABC}	$\frac{SC_{ABC}}{(a-1)(b-1)(c-1)}$	$\frac{CM_{ABC}}{CM_E}$	p_{ABC}
Error	$abc(n-1)$	SC_E	$\frac{SC_E}{abc(n-1)}$		
Total	$abcn - 1$	SC_T			

Table 4.3: Tabla ANOVA para un Diseño Factorial General con tres factores

4.7. Región de Rechazo

Se rechaza H_0 bajo el criterio de la región crítica si el estadístico calculado es mayor al estadístico F a evaluar.

$$F_{\text{calculado}} > F_{(1-\phi), (y), (u)}$$

En caso de desear evaluar el valor-p, el cálculos es el siguiente

$$\text{valor} - p = 1 - P(F_{\text{calculado}} < F)$$

en dónde rechazamos si:

$$\text{valor} - p < \phi$$

4.8. Comparaciones Múltiples

4.8.1. Prueba de Duncan

El método de comparación múltiple de Duncan se aplica **solo después** de que la ANOVA indique diferencias significativas. El procedimiento esencial es:

$$S_{\bar{y}} = \sqrt{\frac{CME}{n}}$$

Los rangos críticos de Duncan calculado corresponde a $S_{\bar{y}} \cdot q_{\phi(y,w)}$, donde y representa $y=[2,...,a]$ siendo a la mayor cantidad de niveles del factor y y w siendo los grados de libertad del error.

4.9. Imputación de valores faltantes

Para un diseño factorial de 3 factores A, B, C con efectos fijos, n réplicas por celda y un único valor faltante $y_{i_0 j_0 k_0 l_0}$, se utilizan las siguientes fórmulas:

$$Y_{VF} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n y_{ijkl},$$

$$\alpha_{VF} = \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n y_{ij_0 k l}, \quad \beta_{VF} = \sum_{i=1}^a \sum_{k=1}^c \sum_{l=1}^n y_{i j_0 k l}, \quad \gamma_{VF} = \sum_{i=1}^a \sum_{j=1}^b \sum_{l=1}^n y_{i j k_0 l},$$

$$(\alpha\beta)_{VF} = \sum_{k=1}^c \sum_{l=1}^n y_{i_0 j_0 k l}, \quad (\alpha\gamma)_{VF} = \sum_{j=1}^b \sum_{l=1}^n y_{i_0 j k_0 l}, \quad (\beta\gamma)_{VF} = \sum_{i=1}^a \sum_{l=1}^n y_{i j_0 k_0 l}.$$

La ecuación final para el valor faltante es:

$$(a-1)(b-1)(c-1) n y_{i_0 j_0 k_0 l_0} = Y_{VF} - a \alpha_{VF} - b \beta_{VF} - c \gamma_{VF} + ab (\alpha\beta)_{VF} + ac (\alpha\gamma)_{VF} + bc (\beta\gamma)_{VF}. \quad (4.1)$$

Por lo tanto, la fórmula de imputación general para n réplicas es:

$$\hat{y}_{i_0 j_0 k_0 l_0} = \frac{Y - a \alpha_{VF} - b \beta_{VF} - c \gamma_{VF} + ab (\alpha\beta)_{VF} + ac (\alpha\gamma)_{VF} + bc (\beta\gamma)_{VF}}{(a-1)(b-1)(c-1) n}. \quad (4.2)$$

Es importante considerar que al imputar un valor faltante, los grados de libertad se reducen a un valor menos, siendo el modelo ANOVA un análisis menos robusto en comparación a análisis de experimentos sin datos faltantes.

4.10. Utilidad del Diseño Factorial General

El diseño factorial general es una herramienta fundamental en la investigación experimental, ya que permite analizar de forma simultánea múltiples factores independientemente

y también sus interacciones. Gracias a esta capacidad, el modelo proporciona una visión integral sobre la relación entre variables independientes y la variable de respuesta, ayudando a optimizar procesos, mejorar productos y reducir costos.

En particular, la utilidad del DFG se expresa en los siguientes aspectos:

- **Identificación de factores relevantes:** posibilita determinar cuáles factores tienen un impacto estadísticamente significativo sobre la variable de interés, priorizando acciones de mejora y reduciendo el riesgo de decisiones subjetivas y además aportando información valiosa para un posible análisis predictivo.
- **Detección de interacciones:** hace evidente cómo el efecto de un factor depende de los niveles de otros factores, fenómeno crucial en sistemas industriales y de servicios donde los factores difícilmente actúan de forma independiente.
- **Eficiencia en costos:** al analizar múltiples factores en un solo estudio, se logra reducir el número total de ensayos necesarios respecto a diseños unifactoriales.
- **Potencial:** al considerar interacciones y efectos principales dentro de un solo modelo, se eleva la capacidad para detectar efectos reales con mayor precisión.
- **Versátil:** se adapta tanto a experimentos de laboratorio controlados como a condiciones reales de planta o de terreno, siendo aplicable a estudios piloto, control de calidad, desarrollo de nuevos procesos y validaciones reglamentarias.

Aplicaciones generales en la industria

El diseño factorial general tiene amplias aplicaciones industriales, entre ellas:

- Optimización de parámetros en procesos de manufactura (por ejemplo, temperatura, presión, velocidad de corte).
- Desarrollo de nuevos productos alimenticios, farmacéuticos o químicos.
- Mejora de la eficiencia energética y reducción de desperdicios en líneas de producción.
- Control de calidad, identificando condiciones óptimas de operación.
- Ajustes de formulaciones o ingredientes en la industria o alimentaria.
- Evaluación de la interacción de factores agronómicos (dosis de fertilizantes, densidad de siembra, riego).
- Validación de procesos de esterilización y sanitización en la industria de la salud.
- Diseño para minimizar la variabilidad de procesos ante cambios en materia prima o condiciones ambientales.

4.11. Ejercicio

La empresa Agrícola del Maule desea evaluar el crecimiento de un nuevo cultivo de *Maravillas*, midiendo su crecimiento mensual en milímetros como variable de respuesta. Para este estudio se han considerado tres factores experimentales:

- **Dosis de fertilizante**, con tres niveles: alta (A1), media (A2) y baja (A3).

- **Tipo de riego**, con dos niveles: goteo (B1) y aspersión (B2).
- **Cobertura de suelo**, con dos niveles: con mulch (C1) y sin mulch (C2).

A continuación, se presentan los valores correspondientes a cada combinación de tratamientos para su análisis:

C	Rép	A1						A2						A3					
		B1			B2			B1			B2			B1			B2		
		1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
C1	1	60	75	75	67	73	73	62	68	65	71	80	80	76	71	75	75	75	75
	2	86	70	70	67	68	68	76	65	65	72	80	80	70	68	73	75	75	77
C2	1	55	53	53	52	52	57	44	44	45	60	60	60	52	51	50	56	55	57
	2	55	55	55	52	54	54	48	48	45	67	67	65	52	48	54	59	50	55

Table 4.4: Datos del ejercicio factorial con 3 réplicas

1. Unidad experimental, Factores y Niveles

- **Unidad Experimental:** Cultivo de Maravilla de una parcela.
- **Factores, niveles y observaciones:**

Símbolo	Factor	Niveles	Descripción de niveles
A	Dosis de fertilizante	A1, A2, A3	A1 = Alta, A2 = Media, A3 = Baja
B	Tipo de riego	B1, B2	B1 = Goteo, B2 = Aspersión
C	Cobertura de suelo	C1, C2	C1 = Con mulch, C2 = Sin mulch

Table 4.5: Factores y niveles del experimento

- **Número de combinaciones de tratamiento:** $3 \times 2 \times 2 = 12$.
- **Réplicas por combinación:** $n = 6$ (independientes).
- **Observaciones totales:** $N = 12 \times 6 = 72$.

2. Descripción de Factores e Interacciones

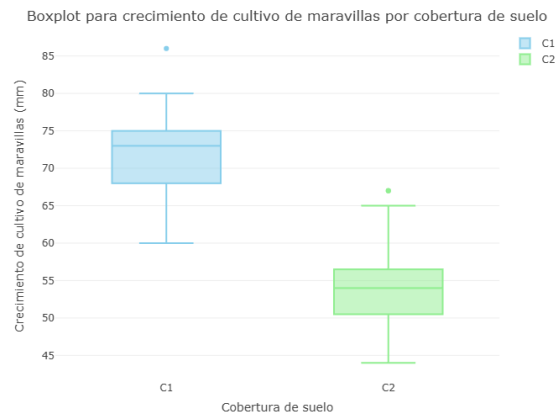


Figure 4.1: Boxplot para crecimiento de cultivo de maravillas por cobertura de suelo

En este caso, para crecimiento de cultivo de maravillas por cobertura de suelo se logra notar una clara diferencia. Donde por cobertura con mulch el crecimiento de los cultivos de maravillas se encuentran entre 60 mm y 86 mm. En cambio, por cobertura sin mulch entre 44 mm y 67 mm. El 100 % de las observaciones de crecimiento con suelo con mulch es superior al menos al 75 % de las observaciones de crecimiento con suelo sin mulch. Ambas coberturas de suelo mantienen un dato atípico hacia crecimientos más elevados, en suelo con mulch sería el máximo en 86 mm, y en suelo sin mulch el máximo en 67 mm.

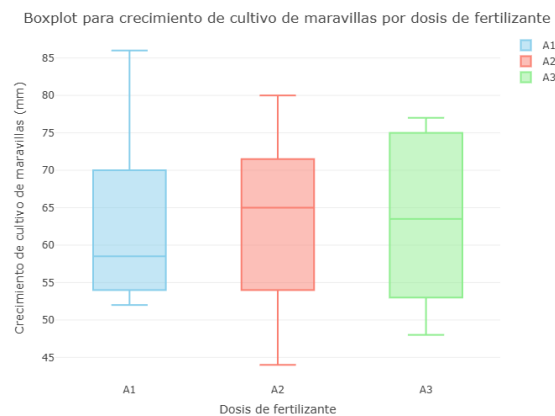


Figure 4.2: Boxplot para crecimiento de cultivo de maravillas por dosis de fertilizante

Nos podemos dar cuenta que la mediana del nivel alto en dosis de fertilizante (58,5 mm) es la menor entre las dosis de fertilizante medio (65 mm) y bajo (63,5 mm). Pudiendo sugerir que un exceso en la dosis de fertilizante puede ser perjudicial para el crecimiento de los cultivos de maravilla. Sin embargo, el crecimiento de cultivos de maravilla alcanza el máximo en la dosis de fertilizante alto (86 mm) aunque valores altos de crecimiento con dosis alta son menos comunes.

- El 50 % de las observaciones de crecimiento de cultivo de maravillas con dosis de fertilizante alta se encuentra entre 54 mm y 70 mm.

- El 50 % de las observaciones de crecimiento de cultivo de maravillas con dosis de fertilizante media se encuentra entre 54 mm y 71,5 mm.
- El 50 % de las observaciones de crecimiento de cultivo de maravillas con dosis de fertilizante baja se encuentra entre 53 mm y 75 mm.

Es por esto, que se propone que los crecimientos de cultivo de maravilla máximos en dosis de fertilizante alta son menos usuales.

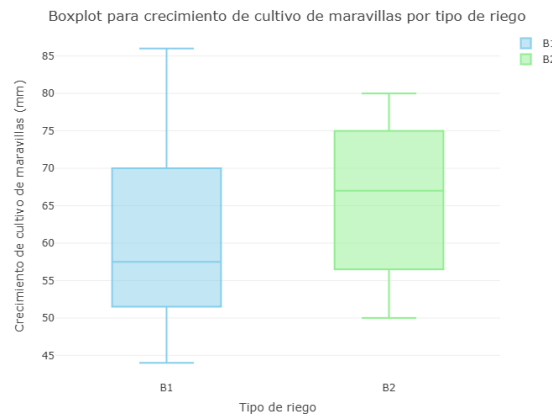


Figure 4.3: Boxplot para crecimiento de cultivo de maravillas por tipo de riego

Ahora, viendo el crecimiento de los cultivos de maravilla por tipo de riego, podemos notar que riego por goteo mantiene crecimientos entre 44 mm y 86 mm en los cultivos de maravilla. En cambio, riego por aspersión mantiene entre 50 mm y 80 mm en los cultivos. La mediana en tipo de riego por goteo es mucho menor (57,5 mm) que tipo de riego por aspersión (67 mm). Donde, cerca del 75 % de crecimientos en riego por aspersión son mayores al 50 % del crecimiento de cultivo de maravilla en riego por goteo.

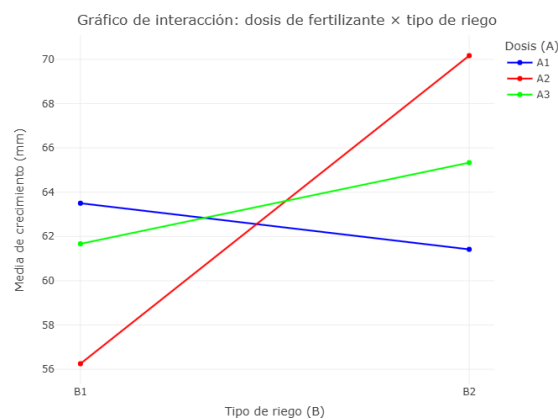


Figure 4.4: Gráfico de interacción para dosis de fertilizante y tipo de riego

En cuanto a las interacciones entre factores, viendo el tipo de riego con la dosis de fertilizante, podemos ver la posible existencia de interacción al haber cruces en el promedio de crecimiento de cultivo de maravillas de las interacciones.

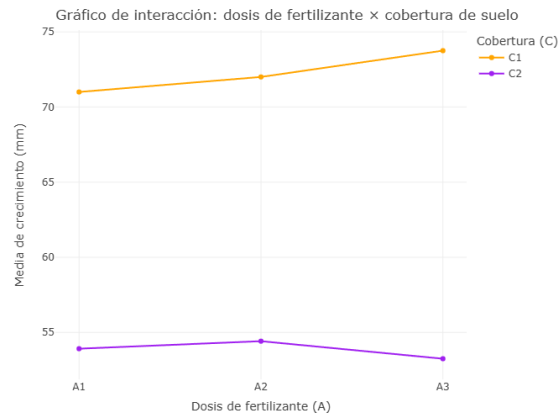


Figure 4.5: Gráfico de interacción para dosis de fertilizante y cobertura de suelo

Ahora, viendo la dosis de fertilizante por cobertura de suelo, tampoco se logra notar una posible interacción entre ellos. Sin embargo, al igual que el caso anterior, sí se logra notar una posible diferencia entre coberturas de suelo.

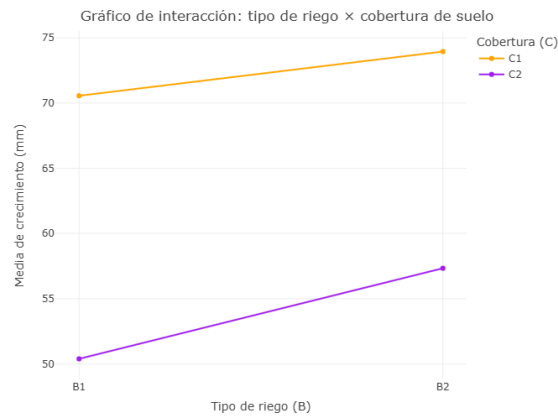


Figure 4.6: Gráfico de interacción para tipo de riego y cobertura de suelo

En el caso de interacciones entre tipo de riego y cobertura de suelo no se logra notar una posible interacción entre ellos. Sin embargo, sí se puede notar la posible diferencia entre coberturas de suelo en el crecimiento promedio de cultivo de maravillas.

3. Modelo ANOVA

En base al caso anterior, deseamos estudiar si existe diferencias entre los niveles de los distintos factores mencionados anteriormente y sus interacciones, por lo que planteamos un modelo ANOVA de 3 factores

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl},$$

$$i = 1, 2, 3 \quad (A1, A2, A3),$$

$$j = 1, 2 \quad (B1, B2),$$

$$k = 1, 2 \quad (C1, C2),$$

$$l = 1, \dots, 6 \quad (\text{réplicas})$$

En donde,

- y_{ijkl} : Crecimiento en milímetros del l -ésimo cultivo de maravilla bajo la i -ésima dosis de fertilizante, con el j -ésimo tipo de riego, utilizando la k -ésima cobertura de suelo.
- μ : media global del crecimiento de maravillas.
- α_i : efecto de la i -ésima dosis de fertilizante.
- β_j : efecto del j -ésimo tipo de riego.
- γ_k : efecto de la k -ésima cobertura de suelo.
- $(\alpha\beta)_{ij}$: interacción dosis de fertilizante i , bajo el j -ésimo tipo de riego.
- $(\alpha\gamma)_{ik}$: interacción dosis de fertilizante i , bajo la k -ésima cobertura.
- $(\beta\gamma)_{jk}$: interacción riego j , bajo la k -ésima cobertura.
- $(\alpha\beta\gamma)_{ijk}$: interacción de la i -ésima dosis de fertilizante, bajo el tipo de riego j , con la cobertura de suelo k .
- ε_{ijkl} : Error experimental en milímetros del l -ésimo cultivo de maravilla bajo la i -ésima dosis de fertilizante, con el j -ésimo tipo de riego, utilizando la k -ésima cobertura de suelo.

4. Supuestos del modelo

Se realiza el experimento bajo los siguientes supuestos:

$$\varepsilon_{ijkl} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

1. **Independencia**: las unidades experimentales han sido aleatorizadas.
2. **Normalidad**: los errores ε_{ijkl} siguen una distribución normal.
3. **Homocedasticidad**: varianza constante $\text{Var}(\varepsilon_{ijkl}) = \sigma^2$ para todas las observaciones.

5. Condiciones de estimabilidad (restricciones de suma cero)

Por otra parte, se consideran las siguientes condiciones de estimabilidad para todo valor de i, j, k .

$$\begin{aligned}\sum_{i=1}^a \alpha_i &= 0, \quad \sum_{j=1}^a \beta_j = 0, \quad \sum_{k=1}^a \gamma_k = 0, \\ \sum_{i=1}^a \sum_{j=1}^b (\alpha\beta)_{ij} &= 0 \quad \forall i, \\ \sum_{i=1}^a \sum_{k=1}^c (\alpha\gamma)_{ik} &= 0, \quad \sum_{j=1}^b \sum_{k=1}^c (\beta\gamma)_{jk} = 0, \\ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (\alpha\beta\gamma)_{ijk} &= 0.\end{aligned}$$

6. Hipótesis a contrastar

Efecto	Hipótesis nula (H_0) vs. alternativa (H_1)
Dosis de fertilizante (A)	$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ vs. $H_1 : \text{Para al menos un } i (i = 1, 2, 3) : \alpha_i \neq 0$
Tipo de riego (B)	$H_0 : \beta_1 = \beta_2 = 0$ vs. $H_1 : \text{Para al menos un } j (j = 1, 2) : \beta_j \neq 0$
Cobertura de suelo (C)	$H_0 : \gamma_1 = \gamma_2 = 0$ vs. $H_1 : \text{Para al menos un } k (k = 1, 2) : \gamma_k \neq 0$
Interacción $A \times B$	$H_0 : (\alpha\beta)_{ij} = 0 \quad \forall i, j$ vs. $H_1 : \text{Para al menos un } i, j : (\alpha\beta)_{ij} \neq 0$
Interacción $A \times C$	$H_0 : (\alpha\gamma)_{ik} = 0 \quad \forall i, k$ vs. $H_1 : \text{Para al menos un } i, k : (\alpha\gamma)_{ik} \neq 0$
Interacción $B \times C$	$H_0 : (\beta\gamma)_{jk} = 0 \quad \forall j, k$ vs. $H_1 : \text{Para al menos un } j, k : (\beta\gamma)_{jk} \neq 0$
Interacción triple $A \times B \times C$	$H_0 : (\alpha\beta\gamma)_{ijk} = 0 \quad \forall i, j, k$ vs. $H_1 : \text{Para al menos un } i, j, k : (\alpha\beta\gamma)_{ijk} \neq 0$

7. Estadístico y Región de Rechazo

Se evaluará este experimento con un nivel de significación de un 5 %

$$\phi = 0,05$$

Utilizando el siguiente estadístico calculado:

$$F_{\text{calculado}} = \frac{CM_{\text{tratamiento}}}{CM_{\text{error}}} \sim F_{(1-\phi), (y), (u)}$$

en donde:

y : Pertenecce a los grados de libertad del tratamiento o interacción a evaluar.

u : Pertenecce a los grados de libertad del error

En dónde rechazamos si

$$\text{valor-p} < \phi$$

$$F_{\text{calculado}} > F_{(1-\phi), (y), (u)}$$

8. Tabla ANOVA

Consideramos la siguiente tabla que contiene la suma de observaciones por factor o combinación entre factores para facilitar el cálculo y entendimiento.

	A1		A2		A3		Total
C	B1	B2	B1	B2	B1	B2	
C1	436	416	401	463	433	452	2 601
C2	326	321	274	379	307	332	1 939
Total	762	737	675	842	740	784	4 540

Table 4.6: Suma de réplicas por combinación factorial $3 \times 2 \times 2$

Por lo tanto, es importante considerar los siguientes valores:

$$a = 3, b = 2, c = 2, n = 6, N = abc n = 72$$

$$y_{....} = 4540$$

$$y_{....}^2 = 20,611,600$$

$$\frac{y_{....}^2}{N} = 286,272,22$$

$$\sum_{i,j,k,l}^N y_{ijkl}^2 = 294,612$$

$$SCT = 294,612 - 286,272,22 = 8,339,78$$

$$SCA = 286,286,0833 - 286,272,22 = 13,863$$

$$SCB = 286,752,72 - 286,272,22 = 480,5022$$

$$SCC = 292,358,944 - 286,272,22 = 6,086,7244$$

$$SCAB = 287,554,83 - 286,272,22 - 13,863 - 480,5022 = 788,25$$

$$SCAC = 292,413,6667 - 286,272,22 - 13,863 - 6,086,7244 = 40,86$$

$$SCBC = 292,896,33 - 286,272,22 - 480,5022 - 6,086,7244 = 56,89$$

$$SCABC = 293,770,33 - 286,272,22 - 13,863 - 480,5022 - 6,086,7244 - 788,25 - 40,86 - 56,89 = 31,02$$

$$SCE = 8,428,78 - 13,863 - 480,5022 - 6,086,7244 - 788,25 - 40,86 - 56,89 - 31,02 = 841,67$$

Por lo tanto, nos quedaría la siguiente tabla ANOVA

Fuente de variación	GL	Suma de cuadrados (SC)	Cuadrado medio (CM)	Estadístico F	Valor p
Factor A	2	13,863	6,9315	0,494	0,6126
Factor B	1	480,5022	480,5022	34,25	2,16e-07*
Factor C	1	6,086,7244	6,086,7244	433,84	0*
Interacción A × B	2	788,25	394,125	28,09	2,45e-09*
Interacción A × C	2	40,86	20,43	1,46	0,2413
Interacción B × C	1	56,89	56,89	4,05	0,0485*
Interacción A × B × C	2	31,02	15,51	1,11	0,3377
Error	60	841,67	14,03		
Total	71	8,339,78			

Table 4.7: Tabla ANOVA para crecimiento de cultivos de maravillas según tipo de riego, dosis de fertilizante y cobertura de suelo *(diferencias significativas $\phi = 0,05$)

Además se comparan los estadísticos F calculados con los valores $F_{(0,95),(1),(60)} = 4,0012$ y $F_{(0,95),(2),(60)} = 3,1504$, obteniendo las siguientes comparativas:

- **Para factor A:** $F_{(0,95),(2),(60)} = 3,1504 > 0,494$ **No se rechaza H_0**
- **Para factor B:** $F_{(0,95),(1),(60)} = 4,0012 < 34,25$ **Se rechaza H_0**
- **Para factor C:** $F_{(0,95),(1),(60)} = 4,0012 < 433,84$ **Se rechaza H_0**
- **Para interacción AB:** $F_{(0,95),(2),(60)} = 3,1504 < 28,09$ **Se rechaza H_0**
- **Para interacción AC:** $F_{(0,95),(2),(60)} = 3,1504 > 1,46$ **No se rechaza H_0**
- **Para interacción BC:** $F_{(0,95),(1),(60)} = 4,0012 < 4,05$ **Se rechaza H_0**
- **Para interacción ABC:** $F_{(0,95),(2),(60)} = 3,1504 > 1,11$ **No se rechaza H_0**

9. Conclusiones Generales

A modo general, con un nivel de significancia de $\phi = 0,05$, se concluye lo siguiente:

- No se rechaza H_0 para el análisis de los niveles del Factor A, indicando que no suficiente evidencia muestral para concluir que hay diferencias significativas en promedio entre las dosis de fertilizante aplicado al cultivo de maravillas.

- Se rechaza H_0 para el análisis de los niveles del Factor B, indicando que si hay diferencias significativas en promedio entre el tipo de riego utilizado para el cultivo de maravillas.
- Se rechaza H_0 para el análisis de los niveles del Factor C, indicando que si hay diferencias significativas en promedio entre la cobertura de suelo aplicada al cultivo de maravillas.
- Se rechaza H_0 para el análisis de interacciones entre los niveles A y B, indicando que si hay diferencias significativas en promedio entre las dosis de fertilizantes utilizadas, en conjunto con el tipo de riego aplicado para el cultivo de maravillas.
- No se rechaza H_0 para el análisis de interacciones entre los niveles A y C, indicando que no hay evidencia muestral suficiente para concluir que hay diferencias significativas en promedio entre el tipo de riego, en conjunto con la cobertura del suelo utilizada para el cultivo de maravillas.
- Se rechaza H_0 para el análisis de interacciones entre los niveles B y C, indicando que si hay diferencias significativas en promedio las dosis de fertilizantes utilizadas, en conjunto con el tipo de cobertura de suelo que tuvo el cultivo de maravillas.
- No se rechaza H_0 para el análisis de interacciones entre los niveles A, B y C, indicando que no hay evidencia muestral suficiente para concluir que hay diferencias significativas en promedio para los niveles de la dosis de fertilizante, tipo de riego, en conjunto con la cobertura del suelo utilizada para el cultivo de maravillas.

10. Comparaciones Múltiples

Medias

$$\bar{y}_{B_1} = \frac{2177}{36} = 60,47, \quad \bar{y}_{B_2} = \frac{2363}{36} = 65,64.$$

Diferencia observada

$$D = |\bar{y}_{B_2} - \bar{y}_{B_1}| = 5,17.$$

Diferencia crítica Para $r = 2$, $\nu_e = 60$ y $n = 36$:

$$s_B = \sqrt{\frac{14,03}{36}} = 0,625, \quad d_2 = 2,828 \times 0,625 = 1,77.$$

Conclusión

$$5,17 > 1,77 \implies \bar{y}_{B_2} > \bar{y}_{B_1} \quad (\alpha = 0,05).$$

Con un nivel de significación de 0.05 rechazamos la hipótesis nula. Es decir, existe suficiente evidencia muestral a favor de que el tipo de riego por goteo es menor a tipo de goteo por aspersión en cuanto al crecimiento promedio de los cultivos de maravilla.

2. Factor C (2 niveles)

Medias

$$\bar{y}_{C_1} = \frac{2601}{36} = 72,25, \quad \bar{y}_{C_2} = \frac{1939}{36} = 53,86.$$

Diferencia observada

$$D = 18,39.$$

Diferencia crítica Para $r = 2$, $\nu_e = 60$ y $n = 36$:

$$s_C = \sqrt{\frac{14,03}{36}} = 0,625, \quad d_2 = 2,828 \times 0,625 = 1,77.$$

Conclusión

$$18,39 > 1,77 \implies \bar{y}_{C_1} > \bar{y}_{C_2} \quad (\alpha = 0,05).$$

Con un nivel de significación de 0.05 rechazamos la hipótesis nula. Es decir, existe suficiente evidencia muestral a favor de que la cobertura de suelo con mulch es mayor a la cobertura de suelo sin mulch en cuanto al crecimiento promedio de los cultivos de maravilla.

Interacción $A \times B$ (6 niveles)

Medias

Tratamiento	A_2B_1	A_1B_2	A_3B_1	A_1B_1	A_3B_2	A_2B_2
Σ	675	737	740	762	784	842
$\bar{y} = \Sigma/12$	56,25	61,42	61,67	63,50	65,33	70,17

(12 observaciones por célula)

Valores críticos Para $\nu_e = 60$ y $n = 12$:

r	$q_{r,60;0,05}$	$d_r = q_{r,60;0,05} \times 1,081$
2	2,828	3,06
3	2,980	3,22
4	3,080	3,33
5	3,140	3,39
6	3,200	3,46

$$s_{AB} = \sqrt{\frac{14,03}{12}} = 1,081,$$

Comparaciones completas

$$\begin{aligned}
 r = 2 : & \quad |\bar{y}_{A_1B_2} - \bar{y}_{A_2B_1}| = 5,17 > 3,06 (*), \\
 & \quad |\bar{y}_{A_3B_1} - \bar{y}_{A_1B_2}| = 0,25 < 3,06 (ns), \\
 & \quad |\bar{y}_{A_1B_1} - \bar{y}_{A_3B_1}| = 1,83 < 3,06 (ns), \\
 & \quad |\bar{y}_{A_3B_2} - \bar{y}_{A_1B_1}| = 1,83 < 3,06 (ns), \\
 & \quad |\bar{y}_{A_2B_2} - \bar{y}_{A_3B_2}| = 4,84 > 3,06 (*); \\
 r = 3 : & \quad |\bar{y}_{A_3B_1} - \bar{y}_{A_2B_1}| = 5,42 > 3,22 (*), \\
 & \quad |\bar{y}_{A_1B_1} - \bar{y}_{A_1B_2}| = 2,08 < 3,22 (ns), \\
 & \quad |\bar{y}_{A_3B_2} - \bar{y}_{A_3B_1}| = 3,66 > 3,22 (*), \\
 & \quad |\bar{y}_{A_2B_2} - \bar{y}_{A_1B_1}| = 6,67 > 3,22 (*); \\
 r = 4 : & \quad |\bar{y}_{A_1B_1} - \bar{y}_{A_2B_1}| = 7,25 > 3,33 (*), \\
 & \quad |\bar{y}_{A_3B_2} - \bar{y}_{A_1B_2}| = 3,91 > 3,33 (*), \\
 & \quad |\bar{y}_{A_2B_2} - \bar{y}_{A_3B_1}| = 8,50 > 3,33 (*); \\
 r = 5 : & \quad |\bar{y}_{A_3B_2} - \bar{y}_{A_2B_1}| = 9,08 > 3,39 (*), \\
 & \quad |\bar{y}_{A_2B_2} - \bar{y}_{A_1B_2}| = 8,75 > 3,39 (*); \\
 r = 6 : & \quad |\bar{y}_{A_2B_2} - \bar{y}_{A_2B_1}| = 13,92 > 3,46 (*).
 \end{aligned}$$

Conclusiones Con un nivel de significación del 0.05 se rechaza la hipótesis nula. Es decir, existe suficiente evidencia muestral a favor de que en promedio las siguientes interacciones entregan un mayor o menor efecto sobre el crecimiento de los cultivos de maravilla:

- $A_1 : B_1 > A_2 : B_1$
- $A_1 : B_1 < A_2 : B_2$
- $A_1 : B_2 > A_2 : B_1$
- $A_1 : B_2 < A_2 : B_2$
- $A_1 : B_2 < A_3 : B_2$
- $A_2 : B_1 < A_2 : B_2$
- $A_2 : B_1 < A_3 : B_1$
- $A_2 : B_1 < A_3 : B_2$
- $A_2 : B_2 > A_3 : B_1$
- $A_2 : B_2 > A_3 : B_2$
- $A_3 : B_1 < A_3 : B_2$

En conclusión, la interacción mayor en este caso fue dosis de fertilizante media (A_2) y tipo de riego por aspersión (B_2). Esta combinación ofreció el mayor rendimiento promedio de crecimiento de cultivos de maravilla en comparación con sus otras interacciones. El menor rendimiento correspondió a dosis de fertilizante media (A_2) y tipo de riego por goteo (B_1).

Interacción $B \times C$ (4 niveles)

Medias

Tratamiento	$B_1 C_2$	$B_2 C_2$	$B_1 C_1$	$B_2 C_1$
Σ	907	1032	1270	1331
$\bar{y} = \Sigma/18$	50,39	57,33	70,56	73,94

(18 observaciones por célula)

Valores críticos Para $\nu_e = 60$ y $n = 18$:

$$s_{BC} = \sqrt{\frac{14,03}{18}} = 0,883,$$

r	$q_{r,60;0,05}$	$d_r = q_{r,60;0,05} \times 0,883$
2	2,828	2,50
3	2,980	2,63
4	3,080	2,72

Comparaciones completas

$$\begin{aligned}
 r = 2 : \quad & |\bar{y}_{B_2 C_2} - \bar{y}_{B_1 C_2}| = 6,94 > 2,50 (*), \\
 & |\bar{y}_{B_1 C_1} - \bar{y}_{B_2 C_2}| = 13,23 > 2,50 (*), \\
 & |\bar{y}_{B_2 C_1} - \bar{y}_{B_1 C_1}| = 3,38 > 2,50 (*); \\
 r = 3 : \quad & |\bar{y}_{B_1 C_1} - \bar{y}_{B_1 C_2}| = 20,17 > 2,63 (*), \\
 & |\bar{y}_{B_2 C_1} - \bar{y}_{B_2 C_2}| = 16,61 > 2,63 (*); \\
 r = 4 : \quad & |\bar{y}_{B_2 C_1} - \bar{y}_{B_1 C_2}| = 23,55 > 2,72 (*).
 \end{aligned}$$

Conclusión

- $B_1 C_1 > B_1 C_2$
- $B_1 C_1 > B_2 C_2$
- $B_1 C_2 < B_2 C_1$
- $B_1 C_2 < B_2 C_2$
- $B_2 C_1 > B_2 C_2$

Si ordenamos estas interacciones de mayor a menor en cuanto a crecimiento promedio de cultivo de maravillas, la combinación que ofrece el mayor rendimiento promedio es *riego por aspersión (B2) con mulch (C1)*, seguida en orden por:

1. Riego por goteo (B1) con cobertura de suelo con mulch (C1).
2. Riego por aspersión (B2) con cobertura de suelo sin mulch (C2).
3. Riego por goteo (B1) con cobertura de suelo sin mulch (C2), la combinación de menor crecimiento promedio.

11. Aplicación del experimento en Rstudio

El desarrollo completo de este ejemplo de aplicación en R se puede encontrar en el siguiente enlace asociado a un repositorio en GitHub https://github.com/angellds3/ANOVA_Disenho_Factorial_Tres_Factores o también escaneando el siguiente código QR.



Capítulo 5

Referencias

- Yates, F. (1937). *The design and analysis of factorial experiments* (Technical Communication No. 35). Imperial Bureau of Soil Science / Her Majesty's Stationery Office.
- Montgomery, D. C. (2007). *Diseño y análisis de experimentos* (2.^a ed.). Limusa Wiley.