

# Formulario 6: Diseño Factorial

## Diseño de experimentos

Andrés Felipe Pico Zúñiga

### Problema 1

Se llevaron a cabo un experimento  $2^2$  con 3 réplicas para evaluar la capacidad de limpieza de una solución en pruebas de lavado, A: Cantidad de Bicarbonato de sodio [5 y 10 %] y B: Cantidad de celulosa de carboximetilo de sodio (2 y 4%). Si las sumas de las réplicas de cada corrida son las siguientes:

$$(1) = 21$$

$$a = 11$$

$$b = 23$$

$$ab = 29$$

El valor del efecto del factor AB es igual a:

### Solución

Se aplica la fórmula de diseño factorial para  $2^2$ :

$$AB = \frac{1}{2n} (ab + (1) - a - b)$$

$$AB = \frac{1}{2(3)} (29 + 21 - 11 - 23)$$

$$AB = 2.6666667$$

### Problema 2

Se llevaron a cabo un experimento  $2^2$  con 3 réplicas para evaluar la capacidad de limpieza de una solución en pruebas de lavado, A: Cantidad de Bicarbonato de sodio (5 y 10 %) y B: Cantidad de celulosa de carboximetilo de sodio (2 y 4%). Si las sumas de las réplicas de cada corrida son las siguientes:

$$(1) = 21$$

$$a = 11$$

$$b = 23$$

$$ab = 29$$

El valor de la varianza del factor AB es igual a:

## Solución

Se calcula el contraste AB:

$$\text{Contraste}_{AB} = ab + (1) - a - b$$

$$\text{Contraste}_{AB} = 29 + 21 - 11 - 23$$

$$\text{Contraste}_{AB} = 16$$

Se calcula la suma de cuadrados de la interacción AB:

$$SS_{AB} = \frac{(\text{Contraste}_{AB})^2}{n2^k}$$

$$SS_{AB} = \frac{(16)^2}{32^2}$$

$$SS_{AB} = 21.3333333$$

Se calcula la varianza o cuadrado medio de la interacción:

$$MS_{AB} = \frac{SS_{AB}}{\text{grados de libertad}}$$

$$MS_{AB} = \frac{21.3333333}{1}$$

$$MS_{AB} = 21.3333333$$

## Problema 3

Se llevaron a cabo un experimento  $2^2$  con 3 réplicas para evaluar la capacidad de limpieza de una solución en pruebas de lavado, A: Cantidad de Bicarbonato de sodio [5 y 10 %] y B: Cantidad de celulosa de carboximetilo de sodio [2 y 4%]. Si las sumas de las réplicas de cada corrida son las siguientes:

$$(1) = 21$$

$$a = 11$$

$$b = 23$$

$$ab = 29$$

Diga cuáles son las condiciones que permiten maximizar la variable de respuesta.

## Solución

Para determinar la combinación óptima, debemos comparar el resultado promedio de cada una de las cuatro condiciones del experimento. Los datos proporcionados son las sumas de 3 réplicas.

Cálculo de los promedios:

$$(1) \text{ Bicarbonato 5\% y Celulosa 2\%: } \frac{21}{3} = 7$$

$$(a) \text{ Bicarbonato 10\% y Celulosa 2\%: } \frac{11}{3} = 3.6666667$$

$$(b) \text{ Bicarbonato 5\% y Celulosa 4\%: } \frac{23}{3} = 7.6666667$$

(ab) Bicarbonato 10% y Celulosa 4%:  $\frac{28}{3} = 9.6666667$

El promedio más alto de limpieza es aproximadamente 9.67, que corresponde a la condición ab. Esta condición representa el nivel alto de ambos factores.

Ambas gráficas muestran que el punto más alto (máxima limpieza) se alcanza cuando el Bicarbonato está en su nivel alto (10) y la Celulosa también está en su nivel alto (4). Las líneas no paralelas en las gráficas indican que existe una interacción significativa entre los dos factores, lo que significa que el efecto de un factor depende del nivel del otro.

## Problema de aplicación 2

Una empresa electrónica quiere minimizar el número de piezas defectuosas generadas en el proceso conocido como “soldadora de ola”. Los factores y niveles son A: velocidad de conveyor (4 y 7 pies/min), B: temperatura de precalentado (80 y 120 grados), C: temperatura de soldadura (470 y 500 grados), debido a que el proceso es muy rápido se decide probar en cada condición de prueba 25 tarjetas, la variable de respuesta es el promedio de las 25 tarjetas, los resultados obtenidos son:

(1) = <39, 45, 42>

a = <110, 110, 104>

b = <23, 27, 30>

ab = <37, 39, 35>

c = <42, 44, 43>

ac = <146, 162, 153>

bc = <31, 35, 39>

abc = <42, 48, 38>

### Tabla ANOVA

```
import numpy as np
from scipy import stats
import pandas as pd

# --- 1. Entrada de Datos ---
# Datos brutos de las réplicas para cada combinación del experimento
data = {
    '(1)': [39, 45, 42],
    'a': [110, 110, 104],
    'b': [23, 27, 30],
    'ab': [37, 39, 35],
    'c': [42, 44, 43],
    'ac': [146, 162, 153],
    'bc': [31, 35, 39],
    'abc': [42, 48, 38]
}

# --- 2. Parámetros del Experimento ---
k = 3 # Número de factores
n = 3 # Número de réplicas por combinación
N = n * (2**k) # Número total de observaciones
```

```

# --- 3. Cálculos para el ANOVA ---

# Suma de los totales para cada combinación en el orden estándar de Yates
order = ['(1)', 'a', 'b', 'ab', 'c', 'ac', 'bc', 'abc']
totals = np.array([sum(data[key]) for key in order])

# Coeficientes de contraste (+1/-1) para cada efecto
contrasts_def = {
    'A': [-1, 1, -1, 1, -1, 1, -1, 1],
    'B': [-1, -1, 1, 1, -1, -1, 1, 1],
    'C': [-1, -1, -1, -1, 1, 1, 1, 1],
    'AB': [1, -1, -1, 1, 1, -1, -1, 1],
    'AC': [1, -1, 1, -1, -1, 1, -1, 1],
    'BC': [1, 1, -1, -1, -1, -1, 1, 1],
    'ABC': [-1, 1, 1, -1, 1, -1, -1, 1]
}

# Diccionario para almacenar los resultados
anova_results = {}

# Calcular la Suma de Cuadrados (SS) para cada efecto
for effect, coeffs in contrasts_def.items():
    contrast_value = np.sum(np.array(coeffs) * totals)
    ss_effect = (contrast_value**2) / N
    anova_results[effect] = {'SS': ss_effect}

# Calcular la Suma Total de Cuadrados (SST)
all_data = np.array([obs for run in data.values() for obs in run])
grand_total = np.sum(all_data)
sst = np.sum(all_data**2) - (grand_total**2 / N)

# Calcular la Suma de Cuadrados del Error (SSE)
ss_effects_total = sum(res['SS'] for res in anova_results.values())
sse = sst - ss_effects_total

# --- 4. Construcción de la Tabla ANOVA ---

# Grados de libertad
df_effect = 1
df_error = (2**k) * (n - 1)
df_total = N - 1

# Calcular MS, F y Valor p para cada efecto
ms_error = sse / df_error
for effect in anova_results:
    res = anova_results[effect]
    res['df'] = df_effect
    res['MS'] = res['SS'] / df_effect
    res['F'] = res['MS'] / ms_error
    res['P-value'] = stats.f.sf(res['F'], df_effect, df_error)

# Añadir las filas de Error y Total
anova_results['Error'] = {'SS': sse, 'df': df_error, 'MS': ms_error, 'F': '', 'P-value': ''}

```

```
anova_results['Total'] = {'SS': sst, 'df': df_total, 'MS': '', 'F': '', 'P-value': ''}

# Convertir a un DataFrame de pandas para una bonita impresión
df_anova = pd.DataFrame.from_dict(anova_results, orient='index')

# Formatear la salida para que coincida con la tabla
df_anova['F'] = df_anova['F'].apply(lambda x: f"{x:.2f}" if isinstance(x, (int, float)) else x)
df_anova['P-value'] = df_anova['P-value'].apply(lambda x: f"{x:.4f}" if isinstance(x, (int, float)) else x)
df_anova['P-value'] = df_anova['P-value'].str.replace('0.0000', '<0.0001', regex=False)

print("Tabla de Análisis de Varianza (ANOVA)")

## Tabla de Análisis de Varianza (ANOVA)
print(df_anova)
```

	SS	df	MS	F	P-value
A	14210.666667	1	14210.666667	789.48	<0.0001
B	15810.666667	1	15810.666667	878.37	<0.0001
C	1380.166667	1	1380.166667	76.68	<0.0001
AB	9440.666667	1	9440.666667	524.48	<0.0001
AC	661.500000	1	661.5	36.75	<0.0001
BC	400.166667	1	400.166667	22.23	0.0002
ABC	840.166667	1	840.166667	46.68	<0.0001
Error	288.000000	16	18.0		
Total	43032.000000	23			

## Planteamiento 3

Se está estudiando la dureza de una tableta, medida en Kg fuerza. Se Piensa que depende de A: tipo de proceso (Vía húmeda: 1, vía seca : 2, granulado: 3) y B: concentración de carbón activado (1% :1, 2% : 2). Se efectúa un experimento que consiste en tomar 2 réplicas de la dureza para cada combinación (a,b) de los niveles de los factores.

Los resultados del experimento son los siguientes:

(1,1) = 218

(1,2) = 187

(2,1) = 202

(2,2) = 204

(3,1) = 175

(3,2) = 217

(1,1) = 205

(1,2) = 201

(2,1) = 220

(2,2) = 233

(3,1) = 179

(3,2) = 301

## Tabla ANOVA

```
import pandas as pd
import statsmodels.api as sm
from statsmodels.formula.api import ols

# --- 1. Crear el DataFrame con los datos ---
# Organizar los datos en un formato adecuado para el análisis
data = {
    'dureza': [218, 187, 202, 204, 175, 217, 205, 201, 220, 233, 179, 301],
    'proceso': ['P1', 'P1', 'P2', 'P2', 'P3', 'P3', 'P1', 'P1', 'P2', 'P2', 'P3', 'P3'],
    'concentracion': ['C1', 'C2', 'C1', 'C2', 'C1', 'C2', 'C1', 'C2', 'C1', 'C2', 'C1', 'C2']
}
df = pd.DataFrame(data)

# Convertir las variables categóricas a tipo 'category'
df['proceso'] = pd.Categorical(df['proceso'])
df['concentracion'] = pd.Categorical(df['concentracion'])

# --- 2. Realizar el Análisis de Varianza (ANOVA) ---
# Definir el modelo: la dureza depende del proceso, la concentración y su interacción
# C(variable) se usa para tratar las variables como categóricas
model = ols('dureza ~ C(proceso) + C(concentracion) + C(proceso):C(concentracion)', data=df).fit()

# Generar la tabla ANOVA a partir del modelo ajustado
anova_table = sm.stats.anova_lm(model, typ=2)

# --- 3. Imprimir la Tabla ---
# Renombrar las columnas para mayor claridad en español
anova_table.index.name = 'Fuente de Variación'
anova_table.rename(columns={
    'sum_sq': 'SS',
    'df': 'gl',
    'F': 'F',
    'PR(>F)': 'p'
}, inplace=True)

# Calcular y añadir la columna de Cuadrado Medio (MS)
anova_table['MS'] = anova_table['SS'] / anova_table['gl']

# Reordenar las columnas para el formato estándar
anova_table = anova_table[['SS', 'gl', 'MS', 'F', 'p']]

print("Tabla de Análisis de Varianza (ANOVA)")

## Tabla de Análisis de Varianza (ANOVA)
print(anova_table)
```

##	SS	gl	MS	F	p
## Fuente de Variación					
## C(proceso)	516.166667	2.0	258.083333	0.360033	0.711760
## C(concentracion)	1728.000000	1.0	1728.000000	2.410602	0.171504
## C(proceso):C(concentracion)	5358.500000	2.0	2679.250000	3.737619	0.088276
## Residual	4301.000000	6.0	716.833333	NaN	NaN