TEC | Tecnológico de Costa Rica

Instituto Tecnológico de Costa Rica

Maestría en Computación

MC6103: Diseño de Experimentos

Apuntes Semana 13

Fecha de entrega: 18 de Mayo de 2023

Profesor:

Ernesto Rivera Alvarado

Alumno:

Angel Phillips Ortega

Índice

1. Presentaciones de artículos de la Semana 14

1. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection presentado por Angel Phillips Ortega

2. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection presentado por Victoria Orozco

2. Experimentos Multifactoriales Fracciónales

Esta es una método exploratoria para llegar ejecutar experimentos más baratos en menor tiempo, a pesar que existe perdida de información en el camino.

2.1. Diseños Multifactoriales Fracciónales

Durante el curso, se ha realizo experimentos factoriales completos seleccionando los factores que se consideran pertinentes como parte del experimento y adicionalmente se ha podido realizar replicaciones.

Sin embargo, si este escenario no siempre en posible. Entonces se puede utilizar diseño fraccional cuando no se puede costear ni siquiera una replica del diseño factorial completo.

Para el diseño fraccional, el numero de ejecuciones estará dado por $N = 2^{k-p}$, donde k es el numero de factores del experimento y p el numero de reducciones en el experimento.

Por ejemplo, si se tienen 3 factores y la reducción del experimento es de 1, entonces $2^{3-1} = 4$. Esto se llama experimento $\frac{1}{2} * 2^k$, es decir, se tiene dos posible mitades para ejecutar las pruebas

Lo que buscan estos diseños es evaluar muchas combinaciones, sin la necesidad de aplicar un experimento grande.

Importante: El diseño factorial fraccional se realiza con factores de únicamente dos niveles, con esto se minimiza la cantidad de información que se pierde.

Adicionalmente el diseño factorial fraccional se utiliza para explorar un gran numero de factores bajo el supuesto de que solo algunos son importantes, esto para tratar de reducir el numero de experimentos a ejecutar.

Si tenemos un diseño experimental completo con los factores de la Tabla 1, habría ejecutar 72 experimentos solo para una replica.

Tabla 1: Factores del diseño experimental completo

Factor A	Factor B	Factor C	Factor D
1	1	1	1
2	2	2	2
	3	3	3
			4

A través de un experimento previo eliminando un factores, por ejemplo el **factor D**, se reducir el numero de experimentos a 18 por replica.

Entonces, los diseño multifactoriales son utilizados como experimentos preliminares o cuando la cantidad de muestra es limitada, se debe tener presente que se puede perder información con experimento completo. Sirve para identificar los factores vitales y distinguir estos de los factores triviales.

2.2. Aplicación de experimentos fraccionales

- Se aplican por limitante de presupuesto.
- Cuando no nos alcanza para repeticiones
- Ni siquiera el experimento completo.
- El ANOVA se puede aplicar para estos escenarios fraccionales, pero el potencial incremento del Error Tipo II. También al menos se requieren dos datos.

Un buen ejemplo de aplicación de experimento fraccional en el área de la química, ya que para las experimentaciones en esta área se utilizan reactivos de alto costo (millones de colones). Entonces para experimento que tengo 8 combinaciones, se puede dividir un diseño fraccional y mandar y ejecutar la muestra que tiene un alto costo. Si la muestra no es exitosa por contaminación u otros factores, el diseño fraccional permitió ahorrar el dinero de las otras combinaciones no aplicadas.

Escenarios para la aplicación del ANOVA:

- Ideal: repeticiones infinitas.
- Normal: según se requiera. Según la experiencia, 150 repeticiones para un solo escenario en pruebas médicas. También basado en la literatura se encuentran distintos valores de repeticiones 200, 100, 50, 38, 29.

- Típico en ciencias de la computación: 15 y 5 para la tesis.
- No ideal: 4-2 repeticiones.
- Si no hay otra opción: 1 repetición.
- Peor escenario: 1/2, 1/4, 1/8, etc.

A pesar, que son experimentos incompletos, no se debe perder el valor de estos, ya que se puede extraer información bastante útiles en las etapas previas de las investigaciones. Adicionalmente, son sumamente útiles en experimentos exploratorios. Estos experimentos fraccionales son una excelente herramienta para identificar los factores que realmente son importantes para nuestro experimento teniendo en cuenta la claridad del contexto.

2.3. Diseño Fraccional

Para tres factores dada A, B, y C, estos se van a ordenar de acuerdo al impacto que estos factores generan en la variable de respuesta. La clasificación del impacto de los factores en la variable de respuesta se basa en conocimiento previo. Por ejemplo en la **Tarea 2**, la arquitectura era el factor de más impacto en la variable de respuesta.

El planteamiento del diseño fraccional se realiza como un conteo binario como en la Tabla 2

Tabla 2: Conteo binario para tres factores

A	B	C
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

Importante: La representación de los 0's y 1's no necesariamente significa presente o no, podría ser la escénica entre dos valores o niveles determinados del factor, por ejemplo: para factor Arquitectura, $0 \rightarrow$ GPU, $1 \rightarrow$ CPU.

Formalmente para diseño fraccional de un experimento con 3 factores en 2^k se visualización como en la Tabla 3

Tabla 3: Diseño fraccional para tres factores

A	B	C
-	-	-
-	-	+
-	+	-
-	+	+
+	-	-
+	-	+
+	+	-
+	+	+

Dado esto ultimo, se dice que para las pruebas solo se pueden hacer 4 experimentos por cuestión de presupuesto. Entonces, ¿como se deben escoger los experimentos?

Importante: Estrategias de escogencia de experimentos a conveniencia como escogencia a dedo o mitad y mitad no son apropiadas, pues esto ocasiona solapamiento y se pierde información.

La estrategia más apropiada es tomar el factor menos significativo **C** y expresar en función de los otros factores más significativos $\mathbf{C} = \mathbf{AB}$, por lo tanto se obtiene la escogencia en la Tabla 4

Tabla 4: Estrategia de escogencia de los 4 experimentos

A	B	C
-	-	+
-	+	-
+	-	-
+	+	+

Esta estrategia de escogencia se consideran todos las combinaciones de los factores más significativos **AB** y esto disminuye la perdida de información.

Si las combinaciones dadas por esta estrategia no son las que se desea explorar por razones practicas si no la otra mitad, se plantea $\mathbf{C} = \mathbf{-AB}$, entonces esto resulta en la Tabla 5

Tabla 5: Estrategia negada de escogencia de los 4 experimentos

A	B	C
-	-	-
-	+	+
+	-	+
+	+	-

El orden de los factores es importante, ya que se consideran todas las combinaciones de factores más significativos **A** y **B**, y no así para el menos significativos **C**, esto basado en el conocimiento del impacto de los factores en la variable de respuesta.

La Tabla 6 se extiende la matriz incluyendo la intersección de modelo lineal **I** y las interacciones entre los factores.

Tabla 6: Matriz completa de las combinaciones e interacciones de los factores

Ejecución	I	A	B	C	AB	AC	BC	ABC
1	+	-	-	-	+	+	+	-
2	+	-	-	+	+	-	-	+
3	+	-	+	-	-	+	-	+
4	+	-	+	+	-	-	+	-
5	+	+	-	-	-	-	+	+
6	+	+	-	+	-	+	-	-
7	+	+	+	-	+	-	-	-
8	+	+	+	+	+	+	+	+

Previamente se realizó la escogencia de los experimentos considerando los signos de los factores más significativos **A** y **B**, pero este mismo resultado se puede obtener considerando los signos de la interacción de factores de mayor nivel **ABC**. En la Tabla 7 se observa la selección de los experimentos según el signo, recordando estos experimentos se ejecutan aleatoriamente.

Tabla 7: Matriz completa de las combinaciones e interacciones de los factores resaltando las escogencias por combinaciones o interacciones entre los factores

Ejecucion	I	A	B	C	AB	AC	BC	ABC
1	+	-	-	-	+	+	+	-
2	+	-	-	+	+	-	-	+
3	+	-	+	-	-	+	-	+
4	+	-	+	+	-	-	+	-
5	+	+	-	-	-	-	+	+
6	+	+	-	+	-	+	-	-
7	+	+	+	-	+	-	-	-
8	+	+	+	+	+	+	+	+

Importante: Recordar que en ANOVA se consideran los factores y sus interacciones, por lo que en experimento escogidos, un factor tiene las mismas combinaciones de signos que las interacciones de los otros factores. Entonces, puede existir incertidumbre entre la significacia del factor y las interacciones de los otros factores, este efecto se conoce como **solapamiento** o **aliasing**.

En la Tabla 7, el factor **A** y la interacción **BC** comparten signo en las ejecuciones 2 y 3, entonces si el ANOVA reporta que el factor **A** es significativo, esto podría ser debido a la interacción **BC**.

Para diseños fraccionales que incluyen interacción de mayor nivel (mayor numero de letras de factores **ABCDE**) es poco probable que tenga significacia, entonces si estas interacciones de mayor se reportan como significativa, podría ser debido a la factor con el mismo signo.

2.3.1. Solapamiento

El **aliasing** o **solapamiento** en diseño experimental fraccional es un fenómeno que ocurre cuando los efectos de variables confundidas se combinan o se confunden entre sí, lo que dificulta la identificación y la interpretación precisa de los efectos individuales. Esto puede llevar a conclusiones incorrectas o poco claras sobre los efectos reales de las variables en estudio.

Para minimizar el efecto del solapamiento en el diseño, se crea un solapamiento entre la intersección del modelo lineal **I** y la interacción de mayor nivel **ABC** para la Tabla 8.

Tabla 8: Matriz reducida con las combinaciones e interacciones de los factores resaltando las escogencias por combinaciones o interacciones entre los factores

A	B	C	AB	AC	BC	ABC	I
-	-	+	+	-	-	+	+
-	+	-	-	+	-	+	+
+	-	-	-	-	+	+	+
+	+	+	+	+	+	+	+

La implicación del solapamiento es la perdida de información debido a la incertidumbre de factores y las interacciones con los mismo signos en la fila. Esto es conocido como **factores confundidos**

Observación 1

La definición del factor menos significativo **C** entre términos los otros factores más significativo **A** y **B** se denomina como factor *generador* **C** = **AB**. También este factor generado se puede expresar como **C** = **-AB** y esto permite explorar las combinaciones contrarias como muestra en la Tabla 9.

Tabla 9: Matriz con las combinaciones e interacciones de los factores con la escogencia de los experimentos para los factores generadores **C = AB** y **C = -AB**

Ejecución	I	A	B	C	AB	AC	BC	ABC
1	+	-	-	-	+	+	+	-
2	+	-	-	+	+	-	-	+
3	+	-	+	-	-	+	-	+
4	+	-	+	+	-	-	+	-
5	+	+	-	-	-	-	+	+
6	+	+	-	+	-	+	-	-
7	+	+	+	-	+	-	-	-
8	+	+	+	+	+	+	+	+

2.4. Generadores y relaciones de definición

Para k factores, se escribe los primero **k** - 1 factores. Dados los factores **A**, **B**, **C**, y **D**. Se tiene en la Tabla 10.

Tabla 10: Combinaciones de k - 1 factores

A	B	C
-	-	-
-	-	+
-	+	-
-	+	+
+	-	-
+	-	+
+	+	-
+	+	+

Se generar la combinaciones para el factor **D** en términos de los otros factores **A**, **B** y **C** como en la Tabla 11.

Tabla 11: Combinaciones de k factores

A	B	C	D
-	-	-	-
-	-	+	+
-	+	-	+
-	+	+	-
+	-	-	+
+	-	+	-
+	+	-	-
+	+	+	+

2.4.1. Intersección

Anteriormente se incluyo la intersección con el modelo lineal en la Tabla 6. con signo positivo para todas las ejecuciones, ya que la intersección siempre forma parte del modelo lineal. Se muestra en la siguiente ecuación del modelo lineal.

$$V.R = I + \beta_1 A + \beta_2 B + \beta_3 C + \beta_4 AB + \beta_5 AC + \beta_6 BC + \beta_7 ABC$$

2.4.2. Relación definición

Como se menciono en la sección 2.3.1, la intersección se igualada con la interacción de mayor nivel. Este método se conoce como relación de definición y permite definir las fracciones de experimentos

Observación 2

Opciones para escoger las mitades de los experimentos.

1. Factor generador **C** = **AB**
2. Signo de la interacción de mayor nivel **ABC**
3. Intersección igualada a la interacción de de mayor nivel **I** = **ABC**

Todos las opciones permiten llegar al mismo resultado.

2.5. Relación entre la intersección y el aliasing

Para un diseño de 4 factores y presupuesto para 8 experimentos, se tienen estas condiciones.

- Mitades: **D** = **ABC** o **D** = **ABC**
- Relación de definición: **I** = **ABCD**

Importante: La relación de definición permite facilita identificar el solapamiento de cada factor con respecto a los demás factores, por ejemplo **A** = **BCD**, **B** = **ACD**, etc. También se puede aplicar para solapamiento de dos niveles como **AB** = **CD**, **AC** = **BD**, etc.

2.6. Análisis del diseño fraccional y factores confundidos

Utilizando la matriz del diseño fraccional de la Tabla 6 para un experimento exploratorio donde se sospecha que el factor menos significativo es **C** y por ello **C** se define en términos de **A** y **B** como en la Tabla 3. Sin embargo, después de ejecutar el ANOVA del diseño fraccional, resulta que el factor **C** es muy significativo, esto indica que el factor **C** requiere ser explorado a profundidad mediante la evaluación de la otra mitad **C** = **-AB**, si el factor **C** también resultado significativo. Se podría indicar que se debe hacer un replanteamiento del orden de los factores.

2.6.1. Supuesto de aleatoriedad en diseños fraccionales

En diseños fraccionales no se puede cumplir estrictamente el supuesto de aleatoriedad, pues no se tiene el espectro completo de los experimentos, esto genera un error en el ANOVA, pero es pequeño.

2.7. Propiedades del diseño fraccional

2.7.1. Principio de escasez de efectos

- Solo unos cuantos factores son importantes en un modelo.
- Un sistema es dominado por factores principales e interacciones de dos niveles.

- Ocasionalmente se observan interacciones de tres niveles.
- Mas de tres niveles es extraño encontrarlas.
- El impacto en la variable de respuesta esta dada por el efecto de los factores principales e interacciones de dos niveles. En términos de solapamiento para interacciones de dos niveles es probable que se de por la factor principal.
- El objetivo es identificar los factores claves y las interacciones que causan mayor efecto en la variable de respuesta.
- Es clave para determinar que ignorar en un diseño factorial.

2.7.2. Principio de proyección

Para tres factores **A**, **B**, y **C**, se determina que **A** no es significativo y esto ayudar a proyectar un diseño completo para factores **B**, y **C**. Estos permite estimar los términos del modelo con **B**, y **C** del diseño completo.

2.8. Resultados del diseño fraccional

Se puede ejecutar experimentos secuenciales, esto permite continuar con la ejecución de otra fracción y si la primera no brinda la información requerida. También se puede seleccionar la fracción del experimento con base al conocimiento previo, suele ser una practica recomendada. Finalmente la información brindada por estos experimentos debe ser complementada con el conocimiento previo.

3. Diseño fraccionales en R

La biblioteca que se requiere la diseño fraccionales en R:

```
> if(!require(FrF2)){install.packages("FrF2")}
```

Para un diseño fraccional de 5 factores y 8 experimentos, se utiliza el siguiente código:

```
> dsg <- FrF2(nfactors = 5, nruns = 8)
> summary(dsg)
```

Donde `nfactors` es numero de factores y `nruns` es numero de ejecuciones. R devolverá el diseño con la resolución más alta posible y la mínima cantidad de solapamiento.

```
Call:
FrF2(nfactors = 5, nruns = 8)

Experimental design of type  FrF2
```

```
8  runs

Factor settings (scale ends):
   A  B  C  D  E
1 -1 -1 -1 -1 -1
2  1  1  1  1  1

Design generating information:
$legend
[1] A=A B=B C=C D=D E=E

$generators
[1] D=AB E=AC

Alias structure:
$main
[1] A=BD=CE B=AD    C=AE    D=AB    E=AC

$fi2
[1] BC=DE BE=CD


The design itself:
   A  B  C  D  E
1 -1  1  1 -1 -1
2  1  1  1  1  1
3 -1  1 -1 -1  1
4  1 -1 -1 -1 -1
5  1  1 -1  1 -1
6 -1 -1 -1  1  1
7 -1 -1  1  1 -1
8  1 -1  1 -1  1
class=design, type= FrF2
```

La información brindada:

- Numero de ejecuciones.
- `Design Generators`: valor de los signos para **D** y **E** y con quienes se solapan.
- Los factores principales están solapados con interacciones de dos factores.
- Las interacciones de dos factores solapadas con otras interacciones de dos factores.

Otra manera para crear un diseño fraccional en R:

```
> dsg <- FrF2(nfactors = 5, resolution = 4)
> summary(dsg)
```

Donde `nfactors` es numero de factores y `resolution` es numero de factores principales. R devolverá el diseño:

```
Call:
FrF2(nfactors = 5, resolution = 4)

Experimental design of type  FrF2
16  runs

Factor settings (scale ends):
   A  B  C  D  E
1 -1 -1 -1 -1 -1
2  1  1  1  1  1

Design generating information:
$legend
[1] A=A B=B C=C D=D E=E

$generators
[1] E=ABCD


Alias structure:
[[1]]
[1] no aliasing among main effects and 2fis


The design itself:
    A  B  C  D  E
1  -1  1 -1 -1 -1
2   1  1  1  1  1
3   1 -1 -1 -1 -1
4  -1 -1  1  1  1
5   1  1 -1 -1  1
6   1 -1  1  1 -1
7  -1  1  1  1 -1
8   1 -1  1 -1  1
9   1  1 -1  1 -1
10 -1 -1 -1 -1  1
11  1 -1 -1  1  1
12 -1  1 -1  1  1
13 -1  1  1 -1  1
14 -1 -1 -1  1 -1
15  1  1  1 -1 -1
16 -1 -1  1 -1 -1
class=design, type= FrF2
```

Importante: R solo muestra las interacciones entre factores principales e interacciones de dos niveles debido a la propiedad de escasez de efectos. Para interacciones niveles superiores no se analizan.

4. Tabla del diseño fraccional

En la Figura 1 se muestra la tabla de diseño fraccional, que permite seleccionar de manera óptima las fracciones.

Un ejemplo: Para 3 factores y 4 experimentos, se ejecuta $\frac{1}{2} * 2^k$ y la funcion generadora es $\pm\mathbf{C} = \mathbf{AB}$. Para 3 factores y 8 experimentos, la tabla indica que se esta ejecutando el experimento completo *full*. Para 3 factores y 16 experimentos, la tabla indica que se ejecuta dos veces o una replica.

Number of factors, k

increasing cost

Number of runs	3	4	5	6	7	8	9
4	2^{3-1}_{III} ±C=AB						
8	2^3 *full*	2^{4-1}_{IV} ±D=ABC	2^{5-2}_{III} ±D=AB ±E=AC	2^{6-3}_{III} ±D=AB ±E=AC ±F=BC	2^{7-4}_{III} ±D=AB ±E=AC ±F=BC ±G=ABC		
16	2^3 *twice*	2^4 *full*	2^{5-1}_{V} ±E=ABCD	2^{6-2}_{IV} ±E=ABC ±F=ABD	2^{7-3}_{IV} ±E=ABC ±F=ABD ±G=ACD	2^{8-4}_{IV} ±E=ABC ±F=ABD ±G=ACD ±H=BCD	2^{9-5}_{III}
32	2^3 *4 times*	2^4 *twice*	2^5 *full*	2^{6-1}_{VI} ±F=ABCDE	2^{7-2}_{IV} ±F=ABC ±G=ABDE	2^{8-3}_{IV} ±F=ABC ±G=ABD ±H=ACDE	2^{9-4}_{IV}
64	2^3 *8 times*	2^4 *4 times*	2^5 *twice*	2^6 *full*	2^{7-1}_{VII} ±G=ABCDEF	2^{8-2}_{V} ±G=ABCD ±H=ABEF	2^{9-3}_{IV}

increasing information about additional factors → lower resolution greater aliasing

Figura 1: Tabla del diseño fraccional

5. Aclaración del proyecto 2

Debido a limitaciones practicas para la ejecuciones para los 80 escenarios y las 5 replicas de cada escenarios en los diferentes sistemas operativos.

Se propone los descrito en la Tabla 12, donde se aleatorizan las ejecuciones en cada sistema operativos y adicionalmente se aleatoriza el orden de los sistemas operativos, esto nos permite ecualizar el error introducido por no cumplir el principio de aleatoriedad a través de replicaciones.

Tabla 12: Propuesta para aleatorizar las ejecuciones en los diferentes sistemas operativos

OS 1	OS 2	OS 3	OS 4	OS 5
16	16	16	16	16