CAPÍTULO 3.
METODOLOGÍA

3. METODOLOGÍA

En este capítulo tiene como objetivo el describir la manera en que se realizó este proyecto, explicar los pasos que se siguieron para poder demostrar la existencia de contrastes en estadística no paramétrica. Aquí podemos encontrar una simulación de lo que es el programa, el diagrama de flujo del mismo, la descripción del algoritmo y del programa mismo.

3.1 Prueba Propuesta

Esta prueba consiste en estimar los efectos principales y de las interacciones en un experimento factorial del tipo 2^k con una sola réplica por medio de contrastes estimados a través de estadística no paramétrica y de simulación, todo esto para saber cuáles de los factores son significativos y deben seguirse investigando.

Se tiene un experimento factorial del tipo 2^k, con k observaciones, una por cada tratamiento en el diseño factorial 2^k con una sola réplica. Los pasos a seguir son los siguientes:

1. Partiendo de la hipótesis de que el efecto de cada factor y sus interacciones son significativos, el primer paso a seguir es el de asignar rangos a las observaciones del experimento, cabe decir que las observaciones no se ordenan, sólo adquieren el número de posición que les corresponde sin perder su posición de acuerdo al tratamiento, en seguida se ilustra en la figura 3.1.1 la manera en que se asignan rangos a las observaciones, para ello se tomó los datos de la primera réplica de un experimento 2^3. (Ver Montgomery 1991)

	A	B	AB	C	AC	BC	ABC	Observaciones
(1)	-1	-1	1	-1	1	1	-1	-3
a	1	-1	-1	-1	-1	1	1	0
b	-1	1	-1	-1	1	-1	1	-1
ab	1	1	1	-1	-1	-1	-1	2
c	-1	-1	1	1	-1	-1	1	-1
ac	1	-1	-1	1	1	-1	-1	2
bc	-1	1	-1	1	-1	1	-1	1
abc	1	1	1	1	1	1	1	6

Observaciones	Rangos
-3	1
0	4
-1	2.5*
2	6.5*
-1	2.5*
2	6.5*
1	5
6	8

* Cuando se tienen dos números iguales entonces el rango que les asignado es el promedio de los rangos que les corresponden

Figura 3.1.1

2. Ahora que se tienen los rangos se calculan los contrastes de cada uno de los factores de la siguiente forma:

$$\boxed{Contraste^2 = \left(\frac{1}{2^{k-1}}\left[\sum R^+ - \sum R^-\right]\right)^2}$$

donde $\sum R^+$ denota la suma de rangos de los tratamientos con coeficientes positivos y $\sum R^-$ denota la suma la suma rangos de los tratamientos con coeficientes negativos

Fórmula 3.1

3. Ya que se tiene el contraste, obtenido de las observaciones originales, se comienza con la simulación. Esta consiste en generar números aleatorios utilizando una distribución Uniforme U (0,1), uno para cada observación, generar rangos igual que se hizo con las observaciones y calcular el contraste con la fórmula 3.1 que se utilizó para el cálculo de los contrastes antes de la simulación.

4. Cuando se han realizado las simulaciones necesarias para estimar el efecto de cada factor y sus interacciones entonces se procede a calcular el Valor P mediante la fórmula siguiente:

$$\boxed{Valor - P = \frac{1}{B}\sum_{i=1}^{B} I\left(A^{i^2} > A^2\right)}$$

Donde B es el número de simulaciones y A^{i^2} son los efectos estimados a través de simulación y A^2 es el resultado del contraste estimado con las observaciones originales.

Fórmula 3.2

5. Por último se calculan los intervalos de confianza para cada efecto calculado y Valor P correspondiente, para ello se utilizará la siguiente fórmula:

$$\boxed{\text{Intervalo de confianza} = \overline{x} \pm Z_{\alpha/2}\, s}$$

Fórmula 3.3

En el punto 3.2 de este capítulo se muestra un ejemplo de los pasos que se han explicado anteriormente.

3.2 Ejemplo de simulación del programa

Tenemos un diseño factorial 2^3, con los factores A, B, C a dos niveles, un experimento con una sola réplica y los datos son los siguientes:

A	B	C	OBSERVACIONES
-1	-1	-1	32.8
+1	-1	-1	27.4
-1	+1	-1	54.1
+1	+1	-1	16.6
-1	-1	+1	45.3
+1	-1	+1	83.4
-1	+1	+1	22.7
+1	+1	+1	17.4

Figura 3.2.1

Este ejemplo tiene el propósito de demostrar si el Efecto de C es significativo para el experimento o si no lo es. Por lo tanto tenemos la siguiente hipótesis:

H_0: El efecto de C es significativo (α=0.05)

H_1: El efecto de C no es significativo (α=0.05)

Se calculó el Efecto de C por medio de contrastes en estadística no paramétrica por medio de rangos.

A continuación se le asignan rangos a las observaciones, como sigue:

C	OBSERVACIONES	RANGOS
-1	32.8	5
-1	27.4	4
-1	54.1	7
-1	16.6	1
+1	45.3	6
+1	83.4	8
+1	22.7	3
+1	17.4	2

Figura 3.2.2

Ya establecidos los rangos, se procede a calcular el contraste de C, con la siguiente fórmula 3.1 entonces:

$$\text{Contraste de } C^2 = \left[\left[\left(6+8+3+2\right) - \left(5+4+7+1\right)\right]\right]^2 = 0.25$$

Ahora comienza la simulación

Primera Simulación

C	U (0,1)	RANGOS
-1	0.035	2
-1	0.824	7
-1	0.956	8
-1	0.011	1
+1	0.233	3
+1	0.379	4
+1	0.726	6
+1	0.708	5

Figura 3.2.3

$$\text{Contraste de } C^{I^2} = \left[\left[\left(2+7+8+1\right) - \left(3+4+6+5\right)\right]\right]^2 = 0$$

Segunda simulación

C	U (0,1)	RANGOS
-1	0.963	8
-1	0.687	7
-1	0.26	3
-1	0.527	6
+1	0.018	1
+1	0.306	4
+1	0.471	5
+1	0.156	2

Figura 3.2.4

$$\text{Contraste de } C^{II^2} = \left[\left[\left(1+4+5+2\right) - \left(8+7+3+6\right)\right]\right]^2 = 9$$

Tercera Simulación

C	U (0,1)	RANGOS
-1	0.363	2
-1	0.742	7
-1	0.761	8
-1	0.474	5
+1	0.24	1
+1	0.401	4
+1	0.395	3
+1	0.627	6

Figura 3.2.5

$$\text{Contraste de } C^{III^2} = \left[\left[(1+4+3+6) - (2+7+8+5)\right]\right]^2 = 4$$

Cuarta Simulación

C	U (0,1)	RANGOS
-1	0.809	8
-1	0.537	5
-1	0.091	1
-1	0.144	2
+1	0.326	4
+1	0.782	7
+1	0.629	6
+1	0.161	3

Figura 3.2.6

$$\text{Contraste de } C^{IV^2} = \left[\left[(4+7+6+3) - (8+5+1+2)\right]\right]^2 = 1$$

Quinta Simulación

C	U (0,1)	RANGOS
-1	0.449	3
-1	0.087	1
-1	0.765	5
-1	0.768	7
+1	0.496	4
+1	0.096	2
+1	0.948	8
+1	0.766	6

Figura 3.2.7

$$\text{Contraste de } C^{v2} = \left[\left[(4+2+8+6) - (3+1+5+7)\right]\right]^2 = 1$$

Concluidas las simulaciones, se calcula el Valor P con la fórmula 3.2 probar la hipótesis nula.

$$P - VALUE = \frac{1}{5}(4) = 0.8$$

Con esto podemos concluir que,

Como el Valor P es 0.8 > 0.05, no se rechaza H_0.

Este es un ejemplo de lo que el programa hará, calculará los contrastes de los factores del experimento utilizando estadística no paramétrica. En este ejemplo sólo se utilizaron 5 simulaciones pero el programa es capaz de llevar a cabo hasta 100,000 de ellas.

3.3 Descripción del algoritmo

En esta sección del capítulo se describe el algoritmo propuesto, para el que se ha escrito un programa en el que el investigador solamente tiene que introducir los datos de las observaciones obtenidos por medio de un experimento y el programa se encargará de llevar a cabo los cálculos y procedimientos necesarios. Cabe mencionar que el programa no le pide al investigador el estadístico $Z_{\alpha/2}$ porque ya lo tiene predefinido y este trabaja con un porcentaje del 95% de confiabilidad.

El algoritmo propuesto es el siguiente:

1. Se introducen los datos de las observaciones y la tabla de signos para los factores principales en dos archivos, entonces se compila y se corre el programa.
2. Se solicitan al usuario los siguientes datos:

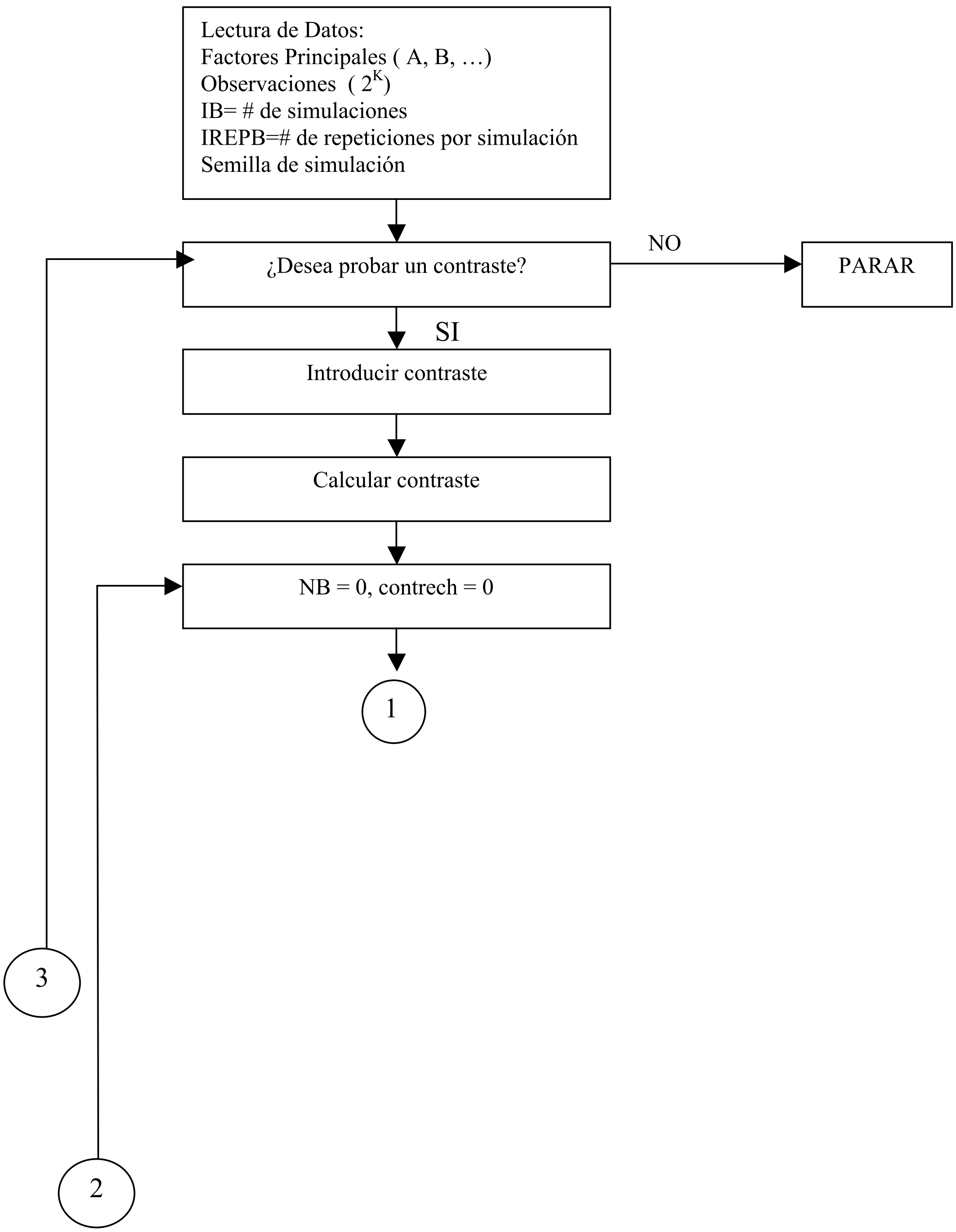
   - La semilla para la simulación
   - Número de simulaciones (IB)
   - Número de repeticiones por simulación (IREPB).

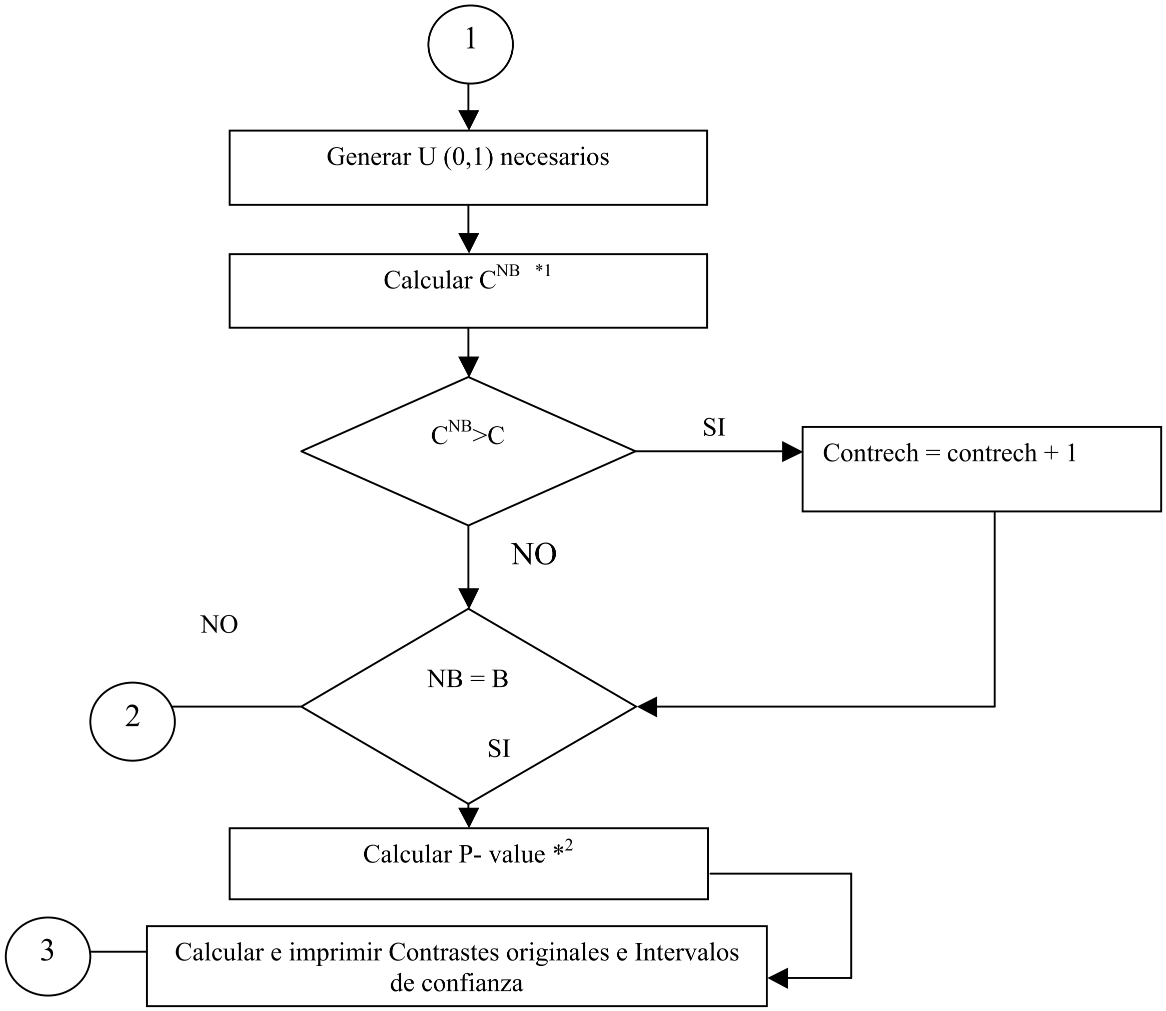
3. Se le asignan rangos a las observaciones originales.
4. Se genera la tabla de signos para cada factor y se calculan los contrastes para cada factor e interacciones.
5. Empieza la simulación (NB=0). El contador de simulaciones está en cero (NB=0)
6. Se obtienen números aleatorios para las nuevas observaciones U (0,1)
7. Se asignan rangos a estas observaciones simuladas y se calculan los contrastes simulados para cada factor e interacciones.
8. Se incrementa en 1 el contador de simulaciones (IB=IB+1).

9. El contador de rechazos para la obtención del Valor P está en cero (Contrech=0)
10. Se comparan los contrastes simulados con los originales y se incrementa el contador de rechazos (Contrech=contrech+1)
11. Se calcula el Valor P cuando se termina las repeticiones de la simulación, es decir cuando (RB=B)
12. Se calculan las varianzas de cada contraste.
13. Y por último se calcula el intervalo de confianza correspondiente a cada contraste
14. Se imprimen los contrastes originales e intervalos de confianza para cada factor e interacciones.
15. Parar

3.4 Diagrama de Flujo

El siguiente diagrama de flujo representa gráficamente el algoritmo computacional para el cálculo de contrastes en estadística no paramétrica, el cual es la base para escribir el programa de simulación en cualquier lenguaje.





*[1] Después de generar U (0,1), se crean los rangos y se calcula el contraste de los efectos.

*[2] $\text{Valor P} = \frac{\text{contrech}}{B}$

3.4 Descripción del programa de simulación

El programa diseñado para este algoritmo tiene tres versiones y fueron escritos en el lenguaje de programación Fortran y denominados como fabiola1.exe, fabiola2.exe, fabiola3.exe. A continuación se describe como trabajan dichos programas.
En general, el propósito del programa es el de generar los contrastes para cada uno de los factores y sus interacciones y saber cuales de ellos son significativos para seguir estudiándolos. El programa también proporciona al usuario el intervalo de confianza mediante la repetición de las simulaciones, determinadas por el experimentador, y con la ayuda del valor P previamente calculado. Esto da una mayor claridad acerca de la precisión de la prueba.

Este programa tiene una capacidad de correr hasta un diseño factorial 2^8 , esto es 256 observaciones pero puede ser modificado según las necesidades del investigador. Como ya se mencionó antes, hasta el momento se tienen tres versiones disponibles del programa, la de los diseños 2^3, 2^4 y 2^5 pero es posible modificarlo fácilmente. Los aspectos que se pueden modificar son la matriz de signos de los factores principales, las columnas de las interacciones de los factores en la matriz de factores y los valores numéricos de las observaciones. La forma de hacer dichas modificaciones se encuentra en el Apéndice II. También cabe mencionar que el programa no imprime los contrastes originales y los intervalos de confianza para los valores P en el orden natural de Fisher, si no que primero imprime factores principales, después interacciones dobles, interacciones triples, así hasta llegar a imprimir la interacción de los k factores. El program le especifica al usuario el orden para una fácil comprensión de los resultados de éste.