



CIMAT

Centro de Investigación en Matemáticas, A.C.

**Detección de Efectos Activos Bajo
Condiciones de Datos Anómalos y
No-Normalidad**

TESIS

que para obtener el Grado de

Doctor en Ciencias

con Orientación en

Probabilidad y Estadística

Presenta:

Román de la Vara Salazar

Directores de Tesis:

Dr. Víctor Manuel Armando Aguirre Torres

Dr. Rogelio Ramos Quiroga

Guanajuato, Gto., 29 de Septiembre de 2006

AGRADECIMIENTOS

A Rosalinda, mi esposa, por su apoyo permanente.

A mis padres y hermanos que siempre me han animado a seguir adelante.

A los doctores Víctor Aguirre Torres y Rogelio Ramos Quiroga por la dirección de este trabajo de tesis.

A los doctores Andrés Christen, Jorge Domínguez y Ernesto Barrios por ser sinodales, y por sus observaciones y comentarios que ayudaron a mejorar sensiblemente el trabajo de tesis.

Al CIMAT por darme el tiempo necesario para culminar este proyecto académico.

Al Dr. Víctor Pérez-Abreu por alentar mi entrada al programa de doctorado del CIMAT.

Al director del CIMAT, Dr. José Carlos Gómez Larrañaga, por su apoyo en momentos críticos.

A la maestra Stephanie Dunbar por hacerme hablar en inglés.

A mis compadres, los doctores Armando Domínguez y Lupita Russell por su soporte estadístico y anímico.

A Humberto Gutiérrez Pulido por compartir parte de esta aventura.

A mis colegas y/o amigos del CIMAT: Enrique, Ramón, Mirabal, Solis, Fina, Moreles y Johan por su apoyo moral.

Contenido

1	Prefacio	4
2	Preliminares de Factoriales No-Replicados y Contaminación por Datos Anómalos	12
2.1	Introducción a los Experimentos Factoriales	12
2.2	Modelo Estadístico en el Análisis de los Diseños Factoriales No-Replicados	17
2.3	Impacto de las Observaciones Anómalas	19
2.4	Reestimado una observación sospechosa	21
2.5	El caso de dos datos anómalos	23
2.6	Reestimando dos observaciones sospechosas	26
3	Métodos Formales Existentes para Analizar Factoriales con Posibles Datos Anómalos	28
3.1	Método Bayesiano	28
3.2	Método de Rangos	36
3.3	Sobre la Comparación de los Métodos	40
4	Nuevas Propuestas para el Análisis de Factoriales No-Replicados con Posibles Datos Anómalos	44
4.1	Método de rangos modificados	46
4.2	Método de ajuste de espacios entre efectos	48
4.3	Método de datos anómalos reestimados	51
4.3.1	Cuando hay dos datos anómalos	56
5	Métodos de Detección de Efectos Activos Basados en Técnicas de Regresión Robusta	60
5.1	Estimadores de Regresión Robusta	60
5.2	Procedimiento lmRobMM de SPLUS	65
5.3	Método Basado en Técnicas de Regresión Robusta, Variante I	66
5.3.1	Sobre el Primer Paso del Método Robusto I	68

5.4	Método Basado en Técnicas de Regresión Robusta, Variante II	70
6	Ejemplos	73
6.1	Ejemplo 1	74
6.2	Ejemplo 2	77
6.3	Ejemplo 3	81
7	Estudio de Simulación	86
7.1	Calibración de los métodos	87
7.2	Escenarios estudiados y resultados obtenidos	88
7.3	Efecto de la Eficiencia y del Número de Términos a Fijar en el Modelo	92
8	Detección de Efectos Activos en Experimentos Factoriales con Respuesta No Normal	95
8.1	Famila Exponencial de Distribuciones	97
8.2	Estructura de los Modelos Lineales Generalizados	99
8.3	Inferencia tipo Wald e inferencia basada en verosimilitud	100
8.4	Elementos de Estadística Bayesiana	101
8.5	Selección de Bayesiana de Modelos	104
8.6	Caso Normal	105
9	Detección de Efectos Activos en Experimentos Factoriales con Respuesta Poisson	108
9.1	Poisson con liga identidad	108
9.2	Poisson con liga log	110
9.3	Poisson con liga raíz cuadrada	112
9.4	Ejemplo, Regresión de Poisson	113
10	Detección de Efectos Activos en Experimentos Factoriales con Respuesta Binomial	121
10.1	Binomial con liga logística	121
10.2	Ejemplo, Regresión Logística	123
11	Detección de Efectos Activos en Experimentos Factoriales con Respuesta Gama	125
11.1	Regresión gama, liga inversa	126
11.2	Regresión gama, liga log	127
11.3	Uso de la transformación $\log(y)$	128
11.4	Ejemplo	129
12	Conclusiones	136
13	Apéndice: Demostración de la Fórmula 3.3	141

14 Apéndice: Programas	149
14.1 Detección de Efectos Activos en Experimentos Factoriales No-Replicados Contemplando la Posibilidad de Datos Anómalos	149
14.1.1 Método de Rangos	150
14.1.2 Método de Rangos Modificados	152
14.1.3 Método de Ajuste de Espacios Entre Efectos	153
14.1.4 Método de Estimación de Datos Extremos (Anómalos)	154
14.1.5 Método Basado en Técnicas de Regresión Robusta I	157
14.1.6 Método Basado en Técnicas de Regresión Robusta II	159
14.1.7 Método Bayesiano de Box y Meyer que Admite la Posibilidad de Datos Anómalos	160
14.2 Detección Bayesiana de Efectos Activos Cuando la Respuesta es No- Normal	168
14.2.1 Regresión de Poisson, Liga Identidad	169
14.2.2 Regresión de Poisson, Liga Log	174
14.2.3 Regresión de Poisson, Liga Raíz Cuadrada	175
14.3 Regresión Logística	176
14.4 Modelo Gama	177

Capítulo 1

Prefacio

Frecuentemente no es posible correr repeticiones en cada combinación de un experimento factorial completo, e incluso es común que sólo se corra una réplica de una fracción de éste. En estas situaciones, aunque la respuesta sea normal y no haya datos atípicos o aberrantes, la detección de los efectos activos o significativos se complica al no contar con repeticiones naturales que permitan obtener un estimador independiente de la varianza del error. No es posible en este caso aplicar la prueba F estándar para detectar dichos efectos.

El primer método que se propuso para el análisis de factoriales no replicados y todavía uno de los más utilizados es el método de Daniel (1959), que consiste en graficar los efectos en papel de probabilidad normal y considerar activos los efectos que no se alinean e inertes los que se alinean. La línea de referencia es la que señalan los puntos correspondientes a los efectos pequeños centrales, ya que esa tendencia es la esperada si los efectos poblacionales fueran nulos. El problema es que en muchos experimentos no es fácil decidir visualmente si un efecto está lo suficientemente alejado de dicha línea como para concluir que es significativo. Esta subjetividad del método de Daniel y el hecho de no contar con un estimador natural de la varianza del error son las razones principales por las que en los últimos 25 años se hayan propuesto por diferentes autores no menos de 30 métodos que buscan ser objetivos al momento

decidir cuáles efectos están activos en el experimento. Una revisión de la mayoría de los métodos propuestos hasta ahora se puede ver en el trabajo de Hamada y Balakrishnan (H&B, 1998). Se puede afirmar que la mayoría de los métodos explotan de alguna manera el “principio de escasez” o “principio de Pareto” de efectos (Wu y Hamada, 2000, pág. 112), que dice que de antemano se espera que solo algunos (digamos entre 20% y 30%) de los efectos estarán activos, y apoyados en este principio proponen alguna estrategia que genera un punto o puntos de referencia contra los que se evalúa la significancia de los efectos.

Cuando se cumplen los supuestos tradicionales de normalidad, varianza constante e independencia de los residuos y no hay presencia de observaciones atípicas o datos aberrantes, varios de los 24 métodos estudiados por H&B tienen un desempeño adecuado para detectar efectos activos. Sin embargo, en la práctica no es improbable la ocurrencia de al menos una observación atípica, de las cuales Daniel (1959) afirma que en su experiencia suceden con una probabilidad de 0.01 a 0.1 en este tipo de experimentos. Y en los experimentos factoriales sin réplicas es más difícil percatarse de ello al no contar con la referencia que representan el resto de las repeticiones en cada tratamiento, pero es un hecho que una sola observación anómala puede afectar sensiblemente el desempeño de la mayoría de los métodos propuestos, que no contemplan dicha posibilidad. El impacto de las observaciones anómalas en 11 de los métodos se muestra a través de un ejemplo en el trabajo de Aguirre-Torres (1998). Del análisis de dicho ejemplo se concluye que solo los métodos que contemplan la posibilidad de observaciones anómalas tienen buen desempeño en la detección de los efectos activos.

Una forma de manejar un experimento factorial no replicado que tiene un dato anómalo, una vez que éste es identificado, fue propuesta por Box (1991), quien parte de la observación de Daniel (1976) en el sentido de que el dato anómalo causa una separación o espaciamiento en la escala de los efectos graficados en el papel de probabilidad normal. Este desfase surge porque el dato extremo típicamente entra sumando en $(2^{k-1} - 1)$ de los contrastes y restando en 2^{k-1} de ellos. Y es precisamente este

desfase el que a su vez provoca que los métodos formales que no contemplan la posibilidad de datos discrepantes se vean fuertemente afectados. Una solución dada por Box (1991) en ese caso consiste en estimar el valor "correcto" del dato anómalo a partir de la separación observada en los 8 efectos centrales; se sustituye la observación correspondiente y se vuelve a obtener el gráfico de Daniel. Otra estrategia sugerida por Box (1990) es considerar las observaciones anómalas como datos faltantes que se reestiman a partir de los datos "buenos", suponiendo nulos algunos de los contrastes que corresponden a interacciones de alto orden.

Dos métodos formales que admiten la factibilidad de observaciones discrepantes son el método Bayesiano de Box y Meyer (1987) y el método basado en la transformación de rangos de Aguirre-Torres (1993). Una versión mejorada de este último método se encuentra en el artículo de Aguirre-Torres y Pérez-Trejo (AT&PT,2001), donde a los efectos estimados con la transformación de rangos se les aplica una variante del método de Benski (1989) para dar lugar a un método híbrido (combinación de dos métodos o pruebas) que tiene control sobre el error tipo I aún bajo condiciones de contaminación. El método de Box y Meyer (1987) calcula iterativamente las probabilidades posteriores de que los efectos están activos y de que los datos sean anómalos. Con las probabilidades posteriores altas se decide el resultado del análisis.

De lo anterior se puede afirmar que cuando hay presencia de datos anómalos en el experimento habría varias maneras de afrontar el problema: 1. Reestimado la observación anómala que se declara faltante como lo hace Box (1990), 2. Estimando los efectos de manera robusta, como lo hace AT&PT (2001) o usando técnicas de regresión robusta, 3. Ajustando los espacios o boquetes que generarían los datos atípicos entre los efectos usuales ordenados (no robustos), y 4. Con métodos Bayesianos como en Box y Meyer (1987).

En una primera vertiente este trabajo de tesis contribuye con métodos nuevos para identificar efectos activos en experimentos factoriales no replicados con posible contaminación por datos anómalos explotando las diferentes maneras de enfocar

el problema mencionadas en el párrafo anterior. En una segunda vertiente se propone también un método Bayesiano para detectar efectos activos en experimentos factoriales no replicados cuando la respuesta es Poisson, binomial o gama. En este contexto de modelos lineales generalizados se utiliza selección Bayesiana de modelos para identificar los efectos activos (Box y Meyer, 1993).

Para clarificar lo que se estudia en esta tesis, y también lo que no se estudia, se presenta el diagrama de la Figura 1.1. En el trabajo siempre se hace referencia a experimentos factoriales no-replicados, en los cuales se puede suponer que la respuesta sigue una distribución normal o no-normal (Poisson, binomial o gama). Tanto en el caso normal como no-normal puede haber o no datos atípicos o aberrantes. En esta tesis se estudian los casos normal con posibles datos atípicos y no-normal sin datos atípicos, que son las ramas del diagrama que terminan con recuadro. No se incluyen en la tesis los casos normal sin datos atípicos y tampoco el caso no-normal con datos atípicos. El caso normal sin datos atípicos ha sido ampliamente estudiado en la literatura (ver por ejemplo H&B, 1998). El segundo caso (no-normal con datos atípicos) rebasa el alcance del presente trabajo.

Las nuevas propuestas de métodos que se hacen en esta tesis para el caso normal con datos atípicos se concretan al enfoque frecuentista, y en todos se aplica el método de Bensi (1989) en el mismo sentido en que lo utilizan AT&PT (2001). Se decidió utilizar el método de Bensi (1989) en principio porque había mostrado buen desempeño en el método propuesto por AT&PT (2001), pero sin duda que es posible encontrar nuevos métodos robustos a datos atípicos a partir de combinar ideas de otros métodos como los descritos por H&B (1998). La investigación y desarrollo de esas nuevas propuestas queda por el momento como una línea de investigación futura. El método propuesto para el caso no-normal sin datos atípicos se concentra en el enfoque Bayesiano, buscando extender el método de Box y Meyer (1987) a los modelos lineales generalizados más usuales.

Mediante estudios de simulación se comparan los diferentes métodos propuestos

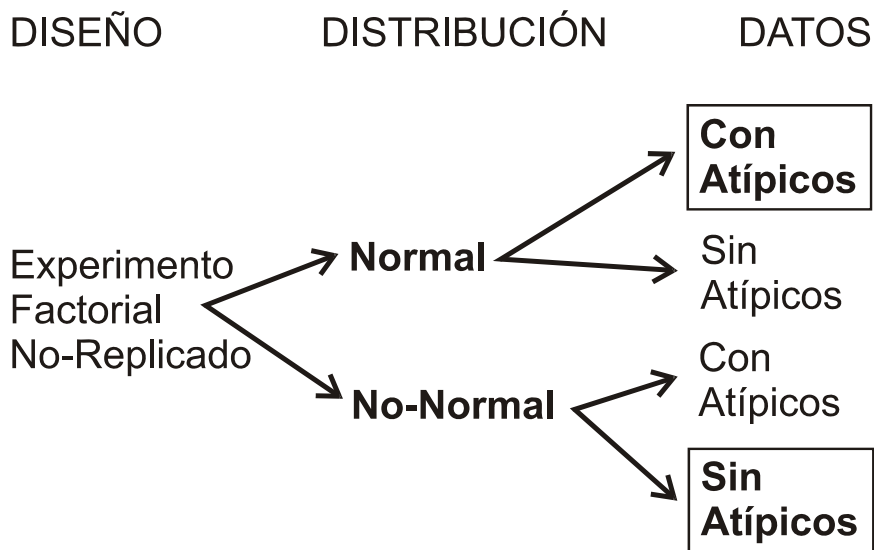


Figura 1.1: Qué se estudia en la Tesis

con el método basado en la transformación de rangos de AT&PT (2001) considerando los mismos tres escenarios de contaminación que ellos consideran en su artículo. Los diferentes métodos propuestos arrojan resultados globalmente mejores que el método de rangos.

Se utilizan técnicas de simulación Monte Carlo para comparar el desempeño de los métodos propuestos dado que no se cuenta con estadísticos que sigan alguna distribución específica conocida bajo la hipótesis de que sólo hay efectos nulos. Aún en los métodos directos en el caso normal que son aquellos donde se propone un pseudoerror estándar para estandarizar los efectos (H&B, 1998) típicamente no se conoce la distribución de referencia exacta ni aproximada del estadístico resultante que pueda utilizarse para obtener la curva de potencia del método. En los métodos que consideran la posibilidad de datos anómalos, que son los que se estudian y proponen en este trabajo de tesis, es todavía más complicado pensar en una comparación analítica de los métodos, puesto que todos los métodos estudiados son híbridos, esto es, se aplican dos pruebas estadísticas y la segunda se aplica sólo si la primera es rechazada,

y no se tiene siquiera un estadístico que sea función de la magnitud de los efectos.

La tesis está organizada como sigue: en el Capítulo 2 se presentan algunos elementos preliminares de factoriales no replicados y su análisis, y se muestra el impacto que pueden tener los datos anómalos cuando están presentes en el experimento. El impacto de los datos anómalos se muestra básicamente usando en gráfico de efectos en papel de probabilidad. El impacto de los datos aberrantes sobre el desempeño de algunos de los métodos existentes que no contemplan este tipo de datos se puede ver en AT&PT (2001). En el Capítulo 3 se revisan los métodos hasta ahora propuestos que admiten la posibilidad de datos discrepantes. Estos son el método Bayesiano de Box y Meyer (1987) y el método que se basa en la transformación de rangos propuesto por Aguirre-Torres, (1993) y mejorado por AT&PT (2001).

En los Capítulos 4 y 5 se describen los métodos nuevos que se plantean en la tesis para analizar factoriales no-replicados con posibles datos anómalos. La idea fundamental es lograr robustez a la hora de afirmar cuáles efectos están activos. La robustez se puede lograr de diferentes maneras y en diferentes momentos: 1. Una es trabajando directamente sobre los datos, detectando y corrigiendo los datos anómalos aplicando por ejemplo la idea de Box (1990). 2. Otra es lograr la robustez al momento de estimar los efectos, y 3. La robustez se puede lograr incluso hasta después de haber estimado los efectos de la manera usual, buscando corregir el impacto que se sabe producen los datos anómalos en los efectos ordenados cuando ese tipo de datos están presentes. En resumen, se puede lograr robustez trabajando antes, durante y después de la estimación de los efectos, en las etapas que se muestran en la Figura 1.2.

De los métodos nuevos que se proponen en el Capítulo 4 uno se enfoca a corregir los datos anómalos desde antes de estimar los efectos, trabajando en la etapa *datos* de la Figura (1.2). Otro de ellos se enfoca a lograr la robustez en la etapa de *estimación* y un tercero se enfoca (en la etapa de *análisis*) a tratar de corregir el impacto negativo que tienen los datos anómalos sobre la estimación usual no robusta de los efectos.



Figura 1.2: Etapas donde lograr robustez

Los métodos nuevos que se describen en el Capítulo 5 se enfocan a lograr robustez en la etapa de *estimación* utilizando técnicas de regresión robusta en ese paso. En el Capítulo 6 se analizan tres conjuntos de datos experimentales utilizando seis métodos diferentes que contemplan la posibilidad de datos anómalos: los dos que ya existían y cuatro métodos nuevos que se proponen en este trabajo. Además, como referencia se incluye el método de Bensi (1989) para el caso normal sin datos atípicos. Los tres conjuntos de datos experimentales son diferentes entre sí: el primero no tiene datos atípicos, de manera que se esperaría que ninguno de los siete métodos expuestos detecte efectos activos. El segundo ejemplo contiene un dato atípico, y se esperaría por ejemplo que el método de Bensi (1989) no se desempeñe bien al no contemplar la posibilidad de ese tipo de datos. El tercer ejemplo tiene también un dato atípico, pero que no es un dato extremo, y otra vez se esperaría que los métodos robustos tiendan a tener un mejor desempeño. De la comparación vía ejemplos del Capítulo 6 se desprende que algunos de los métodos nuevos que se proponen tienen un buen desempeño al compararse con los dos métodos existentes.

En el Capítulo 7 se describe un estudio de simulación para comparar el desempeño de los seis métodos mencionados, excluyendo el método Bayesiano, utilizando para ello un modelo con tres efectos activos de magnitudes diferentes con tres diferentes niveles de contaminación, que son los mismos escenarios estudiados por AT&PT (2001). En un primer paso los métodos se calibran en un escenario no contaminado y sin efectos activos para que todos tengan el mismo error global. No se incluye el método Bayesiano por lo tardado que resultan sus cálculos.

En el Capítulo 8 comienza la segunda vertiente de este trabajo que tiene que ver

con la rama respuesta no-normal sin datos atípicos de la Figura 1.1, misma que se aborda desde el enfoque Bayesiano. En este capítulo se describen los preliminares para esta vertiente como son la familia exponencial de distribuciones, los modelos lineales generalizados, un resumen de la inferencia que en ellos se utiliza y algunos elementos de Estadística Bayesiana, entre ellos la Selección Bayesiana de Modelos. Luego se describe el método Bayesiano de Box y Meyer (1986) el cual se pretende extender en los siguientes capítulos a situaciones experimentales donde la respuesta es no-normal. En el Capítulo 9 se estudia el caso Poisson, en el Capítulo 10 el caso binomial y en el Capítulo 11 el caso gama. Una característica distintiva de la detección de efectos activos en modelos lineales generalizados es la imposibilidad de obtener expresiones cerradas para las probabilidades posteriores de interés. Esto lleva a la necesidad de aproximar dichas probabilidades mediante técnicas numéricas, que en nuestro caso se refieren a simulación Quasi-Monte Carlo, para calcular las integrales múltiples que surgen.

Las conclusiones se dan en en Capítulo 12 donde se resumen las aportaciones principales de este trabajo de tesis en las dos vertientes consideradas: la detección de efectos activos en el caso de respuesta normal con posibles datos anómalos y la detección de efectos activos cuando la respuesta es no-normal sin contemplar la existencia de datos atípicos. En ambas vertientes se propone metodología nueva para detectar los efectos activos en el contexto de experimentos factoriales no-replicados. También se mencionan algunas posibilidades de trabajo futuro por desarrollar en las vertientes mencionadas. Finalmente, se incluye un Apéndice con los detalles del método Bayesiano y otro con los programas utilizados realizados en el lenguaje S-PLUS.

Capítulo 2

Preliminares de Factoriales

No-Replicados y Contaminación por Datos Anómalos

En este capítulo se muestra mediante un ejemplo el impacto que pueden tener uno o dos datos anómalos en los resultados del análisis de un experimento factorial no-replicado. Se muestra en las Secciones 2.3 y 2.4 cómo se puede manejar de acuerdo con Box (1990, 1991) el análisis del experimento una vez que se han identificado uno o dos datos atípicos. Si bien estas estrategias sugeridas por Box no son métodos formales, una de ellas se utiliza en el Capítulo 4 como parte de uno de los métodos nuevos que se desarrollan en este trabajo de tesis. Antes damos una breve introducción a experimentos factoriales donde se definen algunos conceptos que se utilizan a lo largo del trabajo.

2.1 Introducción a los Experimentos Factoriales

Uno de los aspectos importantes en el diseño de experimentos es precisamente decidir cuáles pruebas experimentales deben hacerse para poder dar respuesta a las preguntas

planteadas por el experimentador: más específicamente, deben elegirse los factores o variables de proceso cuyo impacto en la(s) respuesta(s) quiere estudiarse, cuántos niveles se probarán en cada factor y cuántas combinaciones de éstos serán probadas en el experimento. Estos aspectos determinan la llamada matriz de diseño o simplemente "el diseño" que contiene todos los puntos experimentales.

Diseños Factoriales. En un experimento factorial completo se prueban todas las posibles combinaciones que pueden formarse con los niveles de los factores. Evidentemente el número de combinaciones crece rápidamente con el número de factores y con el número de niveles, por lo que frecuentemente no es posible correr todas las combinaciones sino sólo una parte de ellas, dando lugar a los diseños factoriales fraccionados. Por ejemplo, el factorial completo 2^k estudia k factores cada uno probado en 2 niveles, lo que da un total de 2^k puntos experimentales. El factorial fraccionado 2^{k-p} estudia k factores utilizando una fracción $1/2^p$ del factorial completo 2^k .

En este trabajo los métodos estudiados y propuestos se ilustran principalmente con el experimento factorial 2^4 cuya matriz de diseño se muestra en la Tabla 2.1. Para denotar los valores de la variable de respuesta se utiliza la notación de Yates en la que las letras presentes en cada combinación indican que los correspondientes factores se encuentran en su nivel alto; por el contrario las letras ausentes señalan los factores que se encuentran en su nivel bajo. Esta notación permite no solo denotar el valor observado de la respuesta sino identificar fácilmente en cuál combinación fue obtenido. Por ejemplo, (1) denota la observación obtenida en la combinación $(-1, -1, -1, -1)$ puesto que las cuatro letras están ausentes; si hubiera réplicas o repeticiones de todo el arreglo, (1) sería la suma de las observaciones en el punto $(-1, -1, -1, -1)$.

Contrastes y Efectos. El efecto del factor A es la diferencia entre las medias de la respuesta en los niveles alto y bajo de este factor. Esta diferencia de medias o efecto se puede calcular fácilmente obteniendo primero el *contraste* A el cual resulta de multiplicar la columna de signos o de más-menos unos que correspondiente a este factor en la matriz de diseño por los datos o vector de observaciones dado por la

columna en la notación Yates. El contraste dividido por una constante apropiada (2^{k-1}) se convierte en la diferencia de medias que llamamos efecto. Por ejemplo, en el factorial 2^4 no-replicado el contraste A y el efecto A están dados por

Tabla 2.1. Matriz de diseño del factorial 2^4

#	A	B	C	D	Y <small>(Notación de Yates)</small>
1	-1	-1	-1	-1	(1)
2	1	-1	-1	-1	a
3	-1	1	-1	-1	b
4	1	1	-1	-1	ab
5	-1	-1	1	-1	c
6	1	-1	1	-1	ac
7	-1	1	1	-1	bc
8	1	1	1	-1	abd
9	-1	-1	-1	1	d
10	1	-1	-1	1	ad
11	-1	1	-1	1	bd
12	1	1	-1	1	abd
13	-1	-1	1	1	cd
14	1	-1	1	1	acd
15	-1	1	1	1	bcd
16	1	1	1	1	$abcd$

$$\begin{aligned} \text{Contraste } A &= -(1) + a - b + ab - c + ac - bc + abc - d \\ &\quad + ad - bd + abd - cd + acd - bcd + abcd \end{aligned} \quad (2.1)$$

$$\text{Efecto } A = (\text{Contraste } A)/8 = \mu_{A+} - \mu_{A-},$$

Note que todos los términos de la notación de Yates que "entran sumando" al *Contraste* A tienen la letra a presente; y por el contrario los que no la tienen "entran restando".

Con un diseño factorial completo 2^k se pueden estimar y estudiar los $2^k - 1$ efectos posibles que son: k efectos principales (efectos individuales), $\binom{k}{2}$ interacciones dobles, $\binom{k}{3}$ interacciones triples, y así hasta, $\binom{k}{k} = 1$ interacción de los k factores. Por ejemplo, con el factorial 2^4 dado en la Tabla 2.1 se pueden estudiar los 15 efectos

$$A, B, C, D, AB, AC, AD, BC, BD, CD, ABC, ABD, ACD, BCD, ABCD,$$

donde cada uno es una diferencia de medias. Para calcular el contraste de una interacción primero se obtiene su correspondiente columna de signos dada por el producto de las columnas de los factores que la conforman.

En los factoriales 2^k (en dos niveles) los efectos estimados no son otra cosa que los tradicionales coeficientes del modelo de regresión multiplicados por 2 (ver Sección 2.2); o lo que es lo mismo, los coeficientes de regresión son los efectos divididos entre 2: Sea la matriz \mathbf{X} la matriz que contiene las columnas de los 15 efectos más una primera columna de unos. Entonces, los coeficientes de regresión se estiman como

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}\mathbf{X})^{-1} \mathbf{X}\mathbf{y},$$

donde \mathbf{y} es el vector de observaciones. Como la matriz \mathbf{X} es ortogonal de dimensión $n \times n$ donde $n = 2^k$ es el total de observaciones, se tiene que $(\mathbf{X}\mathbf{X})^{-1}$ es una matriz diagonal, con un $1/2^k$ en cada entrada ($diag(1/2^k)$). Entonces

$$\widehat{\boldsymbol{\beta}} = \frac{1}{2} \left(\frac{1}{2^{k-1}} \mathbf{X}\mathbf{y} \right) = \begin{bmatrix} \widehat{\mu} \\ \frac{1}{2} \widehat{Efecto A} \\ \vdots \\ \frac{1}{2} \widehat{Efecto ABCD} \end{bmatrix} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_{15} \end{bmatrix}. \quad (2.2)$$

Factoriales Fraccionados. Para más de 4 factores ($k > 4$) es frecuente que no se pueda o no convenga correr el factorial completo sino sólo una fracción de éste. Al correr una fracción necesariamente se pierde información en relación a los efectos: habrá efectos que no podrán estimarse y los efectos estimados tienen al menos un efecto alias (que es el mismo efecto pero con otro nombre). La fracción a correr se selecciona de manera que se minimize la información que se pierde en relación a los efectos de interés. Dicha selección se basa en el *principio de jerarquía* de los efectos (Wu y Hamada, 2000, pág. 112) que dice que son más importantes *a priori* los efectos principales, seguidos por las interacciones dobles, luego las triples, etc., hasta la interacción de mayor orden. Así a la hora de fraccionar se sacrifican primero los efectos de menor jerarquía que son las interacciones de mayor orden.

Por ejemplo, en el diseño factorial fraccionado 2^{5-1} (notación que indica la fracción a la mitad $(1/2)2^5$) la fracción se elige con base en los signos del contraste de la interacción de mayor orden $ABCDE$ con lo cual se pierde el mínimo de información relativa a los efectos importantes *a priori* que son los 5 efectos principales A, B, C, D y E , y las 10 interacciones dobles $AB, AC, AD, AE, BC, BD, BE, CD, CE$ y DE . La interacción de mayor orden $ABCDE$ usada para generar la fracción no se puede estimar al quedar confundida con la media, y los efectos restantes tienen la siguiente estructura de alias

$$\begin{aligned} &A + BCDE; \quad B + ACDE; \quad C + ABDE; \quad D + ABCE; \quad E + ABCD \\ &AB + CDE; \quad AC + BDE; \quad AD + BCE; \quad AE + BCD; \quad BC + ADE \\ &BD + ACE; \quad BE + ACD; \quad CD + ABE; \quad CE + ABD; \quad DE + ABC, \end{aligned}$$

que se obtiene de multiplicar (módulo 2) cada efecto por el generador de la fracción, que en este caso son los signos positivos del contraste de la interacción de mayor orden $(+ABCDE)$. Se observa en la estructura de alias que los efectos potencialmente importantes son alias de interacciones de alto orden que típicamente no son importantes, de manera que en este caso prácticamente no se ha perdido nada de la información relevante. Este grado de confusión entre los efectos se llama *resolución V* en la cuál los efectos relevantes están "limpiamente estimados", al ser alias de efectos que se pueden ignorar *a priori* como lo son las interacciones triples en adelante.

Otras fracciones que tienen 16 corridas experimentales y que estudian más de 5 factores son los diseños factoriales fraccionados $2^{6-2}, 2^{7-3}, \dots, 2^{15-11}$, a las cuales se les pueden aplicar de manera casi inmediata los métodos que se proponen en esta tesis. Entre 6 y 8 factores las fracciones son de resolución IV en la cual los efectos principales están limpiamente estimados pero hay interacciones dobles que son alias entre ellas; Entre 9 y 15 factores se tiene resolución III lo que significa que hay efectos principales que son alias de al menos una interacción doble. Note que a menor resolución es menos clara la información que se tiene sobre los efectos potencialmente

importantes. La fracción con 15 factores es el caso saturado que implica que todos los grados de libertad del experimento se usan en los 15 efectos principales y se tendrían que ignorar todas las interacciones dobles al momento de interpretar los efectos.

Al interpretar el efecto observado de un grupo de efectos alias éste se atribuye al efecto de mayor jerarquía del grupo, y cuando los efectos del grupo son de la misma jerarquía (típicamente un grupo de interacciones dobles) se aplica el *principio de herencia* (Wu y Hamada, 2000, pág. 112) que recomienda atribuir el efecto observado a la interacción que involucra al menos uno de los factores significativos a nivel de efecto principal. Esto es, el efecto de interacción "hereda" el efecto de los factores que la componen.

2.2 Modelo Estadístico en el Análisis de los Diseños Factoriales No-Replicados

El modelo estadístico usual en el análisis de experimentos factoriales no-replicados es el tradicional modelo lineal

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.3)$$

donde $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ es el vector de observaciones y $n = 2^k$ es el número de corridas experimentales; $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_m)'$ son los parámetros desconocidos, que como vimos representan los efectos que se pueden estudiar ($m = 2^k - 1$); \mathbf{X} es la matriz $2^k \times 2^k$ que contiene las columnas de todos los posibles contrastes más una primera columna de unos, y $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ es el vector de errores. Se asume que los errores cumplen los supuestos de normalidad, varianza constante e independencia, esto es, ε_i , $i = 1, 2, \dots, n$ son variables aleatorias independientes con media cero y varianza común σ^2 ($\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$).

Como se sabe, la varianza de $\hat{\boldsymbol{\beta}}$ está dada por $\sigma^2(\mathbf{X}\mathbf{X})^{-1}$, que en el caso del diseño

factorial 2^k o fracciones 2^{k-p} la varianza de cada parámetro se reduce a

$$var(\widehat{\beta}_i) = (1/2^k)\sigma^2 \quad (2.4)$$

Por otra parte, la varianza de $\widehat{Efecto\ i}$, es $var(\widehat{Efecto\ i}) = var[(\widehat{Contraste\ i})/2^{k-1}]$, que es igual a

$$var(\widehat{Efecto\ i}) = \frac{2^k\sigma^2}{2^{2k-2}} = \frac{\sigma^2}{2^{k-2}} = 4var(\widehat{\beta}_i) \quad (2.5)$$

dado que $Efecto\ i = 2\beta_i$. Suponiendo normalidad de las observaciones y asumiendo que se cuenta con un estimador de la varianza residual σ^2 se prueba la significancia de los coeficientes individuales β_i ($H_0 : \beta_i = 0$ vs. $H_A : \beta_i \neq 0$) usando el hecho de que

$$\widehat{\beta}_i / \sqrt{var(\widehat{\beta}_i)} \sim T \text{ de Student con } \nu \text{ g.l. del error,} \quad (2.6)$$

y previamente verificando la significancia del modelo con una prueba F (ver Montgomery y Peck, (1992) pág.135). Cuando hay repeticiones en el experimento la varianza residual se estima con el cuadrado medio del error definido por

$$\widehat{\sigma}^2 = CM_E = \frac{(\mathbf{y} - \mathbf{X}\widehat{\beta})(\mathbf{y} - \mathbf{X}\widehat{\beta})}{n - p}, \quad (2.7)$$

donde p es el número de parámetros estimados en el modelo y \mathbf{y} es el vector de observaciones.

El problema es que en un diseño factorial completo o fraccionado no-replicado, en el que se realiza sólo una corrida experimental en cada combinación de la matriz de diseño, no se cuenta con repeticiones naturales que permitan obtener un estimador independiente de la varianza del error, y por lo tanto, no se tiene una referencia contra la cual verificar la significancia de los efectos como se hace con la prueba T o con la prueba F cuando sí hay réplicas (repeticiones de todo el arreglo experimental). El estimador de la varianza del error dado en la ecuación (2.7) es igual a cero en un experimento factorial no-replicado si se estudian todos los efectos posibles.

Bajo el supuesto de que sólo un porcentaje menor de los efectos es activo (principio de escasez de efectos), se han propuesto varios métodos en el caso normal sin datos

anómalos en los cuales se propone un estimador del error estándar (pseuderror) con el que se estandarizan los efectos para concluir sobre su significancia al compararlos con algún valor crítico. Este es caso de los métodos "directos" mencionados por Hamada y Balakrishnan (1998). Entre estos métodos se encuentran varios de los más utilizados en la práctica, después del método de Daniel (1959), como son el método de Lenth (1989) y los métodos que combinan sumas de cuadrados de los efectos más pequeños (ver por ejemplo Berk y Picard (1991)).

2.3 Impacto de las Observaciones Anómalas

Varios de los métodos discutidos por H&B (1998) tienen un buen desempeño para detectar los efectos activos en un experimento factorial no-replicado, pero su desempeño se puede ver seriamente afectado cuando se tiene la presencia de al menos 1 observación anómala (ver Aguirre-Torres y Pérez-Trejo, 2001). Esta situación viene a complicar un poco más el análisis del factorial no-replicado al hacer necesario el desarrollo de métodos de análisis robustos a la posible presencia de datos anómalos. También en los factoriales con réplicas puede haber datos anómalos, pero al menos se cuenta con la referencia que representan el resto de las observaciones en el mismo punto experimental, cosa que no se tiene en un experimento sin réplicas.

Tabla 2.2. Experimento factorial 2^4 no-replicado

#	A	B	C	D	Y	#	A	B	C	D	Y
1	-	-	-	-	(1) = 47.46	9	-	-	-	+	$d = 46.76$
2	+	-	-	-	$a = 49.62$	10	+	-	-	+	$ad = 48.56$
3	-	+	-	-	$b = 43.13$	11	-	+	-	+	$bd = 44.83$
4	+	+	-	-	$ab = 46.31$	12	+	+	-	+	$abd = 44.45$
5	-	-	+	-	$c = 51.47$	13	-	-	+	+	$cd = 59.15$
6	+	-	+	-	$ac = 48.49$	14	+	-	+	+	$acd = 51.33$
7	-	+	+	-	$bc = 49.34$	15	-	+	+	+	$bcd = 47.02$
8	+	+	+	-	$abc = 46.10$	16	+	+	+	+	$abcd = 47.90$

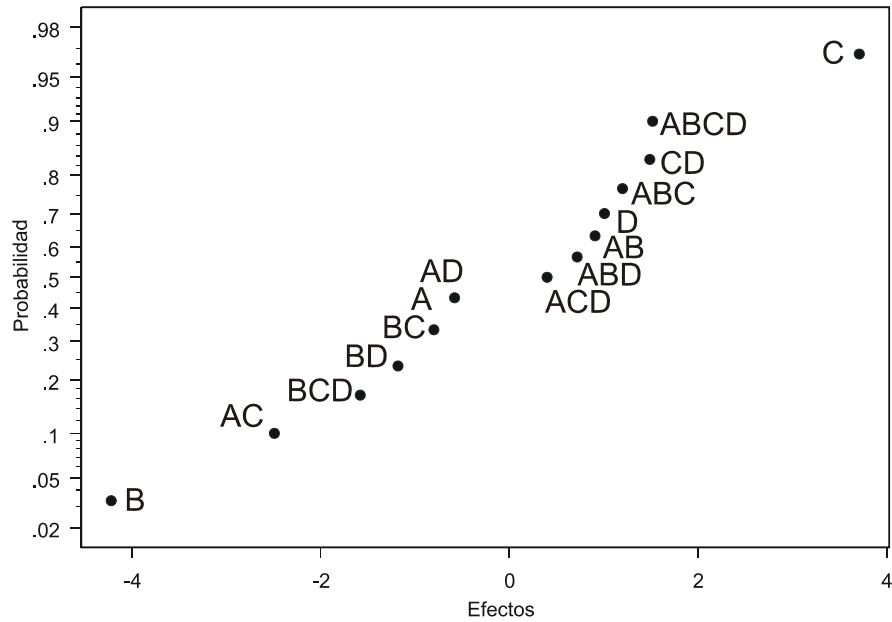


Figura 2.1: Efectos en papel de probabilidad normal

Consideremos por ejemplo el experimento factorial 2^4 no-replicado dado por Box (1991), que se muestra en la Tabla (2.2). Dada su magnitud relativa al resto de los datos, es posible que la observación en el tratamiento cd sea un dato anómalo.

En este experimento se tienen 4 efectos principales, 6 interacciones dobles, 4 interacciones triples y 1 interacción cuádruple, dando un total de 15 efectos, cuyos valores estimados a partir estos datos son:

$A = -0.8$	$AC = -2.49$	$ABC = 1.20$
$B = -4.22$	$AD = -0.58$	$ABD = 0.72$
$C = 3.71$	$BC = -0.8$	$ACD = 0.40$
$D = 1.01$	$BD = -1.18$	$BCD = -1.58$
$AB = 0.91$	$CD = 1.49$	$ABCD = 1.52$

La gráfica de efectos en papel de probabilidad normal se muestra en la Figura 2.1.

A reserva de hacer una prueba formal, se observa en la Figura (2.1) que los efectos B y C están potencialmente activos. Otro aspecto notorio en la figura es el espacio horizontal que se observa entre los efectos a la altura del cero, que se puede deber a la

presencia de un dato anómalo y que enmascara el impacto visual de los efectos potencialmente activos. Este espacio se genera porque un dato anómalo entra sumando en 7 de los contrastes y entra restando en 8 de ellos (ver tratamiento cd en la Tabla 2.3). En particular, en este ejemplo el dato más grande y sospechoso de ser anómalo es $cd = 59.15$, el cual entra restando en los contrastes $A, B, AC, AD, BC, BC, ACD$ y BCD y sumando en los contrastes C, D, AB, CD, ABC, ABD y $ABCD$. Nótese que no necesariamente quedan los efectos repartidos a cada lado del hueco como indica el signo del dato cd a lo largo de los contrastes, como sucede en este caso con el efecto ACD , que aparece a la derecha del espacio generado por el dato sospechoso aunque éste resta en su contraste.

2.4 Reestimado una observación sospechosa

A partir de los 8 efectos centrales. Suponiendo que los ocho efectos centrales, los primeros cuatro a cada lado del hueco, son iguales a cero, se puede estimar el desfase provocado por la observación sospechosa o anómala (Box, 1991). Así, tenemos que los ocho efectos centrales sin signo son

$$\begin{aligned} -A &= 0.80, & -AD &= 0.58, & -BC &= 0.80, & -BD &= 1.18 \\ +D &= 1.01, & +AB &= 0.91, & +ABD &= 0.72, & +ACD &= 0.40, \end{aligned}$$

de cuyo promedio se obtiene un estimador del desfase debido al dato anómalo. Este desfase es $d = 0.8$ que al multiplicarlo por 2^{k-1} (para pasar de la escala de las medias a la escala de los contrastes) estima el desfase en la escala original que tiene el dato anómalo ($0.8 \times 8 = 6.4$). De manera que el dato anómalo se puede "corregir" restándole esta cantidad, con lo que la observación cd corregida es

$$cd_{(corregida)} = 59.15 - 6.4 = 52.75.$$

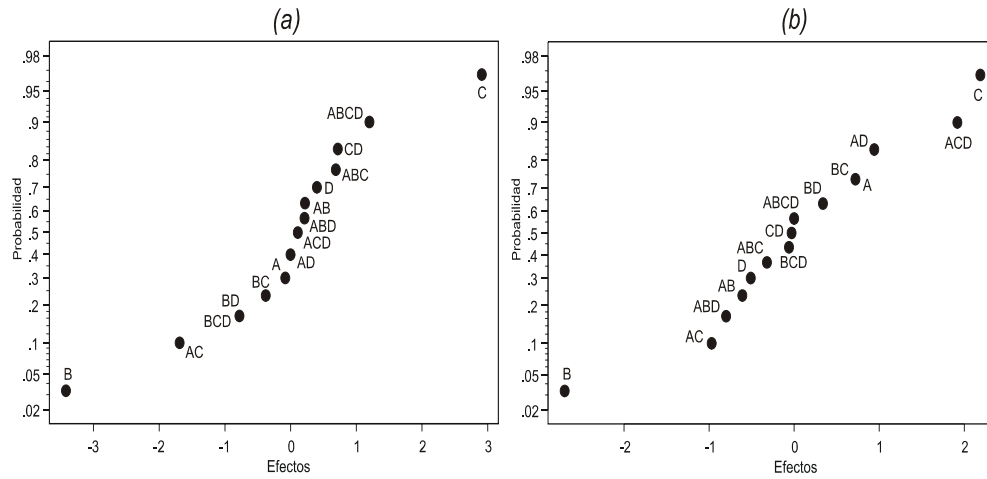


Figura 2.2: Efectos en papel de probabilidad normal con dato anómalo corregido

En la Figura 2.2(a) se muestra otra vez el gráfico de efectos en papel de probabilidad normal obtenidos a partir de los datos de la Tabla (2.2) con la observación corregida $cd = 52.75$. Se observa que el espacio en los efectos interiores ha desaparecido, dando la gráfica una mejor apreciación del potencial efecto de los efectos B y C .

A partir del contraste ABCD. Otra manera de manejar la situación de un dato anómalo es considerarlo dato faltante y estimarlo a partir de las observaciones restantes y de suponer el contraste $ABCD$ igual a cero (ver Box, 1990). Haciendo esto, tenemos que

$$\begin{aligned}
 ABCD &= (1) - a - b + ab - c + ac + bc - abc - d \\
 &\quad + ac + bd - abd + cd - acd - bcd + abcd \\
 0 &= 47.46 - 49.62 - 43.13 + 46.31 - 51.47 + 48.49 + 49.34 - 46.1 \\
 &\quad - 46.76 + 48.56 + 44.83 - 44.45 + x - 51.33 - 47.02 + 47.90
 \end{aligned}$$

de donde se obtiene que

$$x = cd = 46.99.$$

Se incorpora este valor en el vector de observaciones y se procede con la estimación de los efectos. El gráfico de éstos en papel de probabilidad normal se muestra en la Figura 2.2(b). Se aprecia que difiere del caso anterior (Figura 2.2a), en principio debido a que en este ejemplo particular la interacción cuádruple ($ABCD$) no estaba precisamente cercana a cero, por lo que el supuesto de nulidad de dicha interacción viene a trastocar las estimaciones de todos los efectos. Note que la mayoría de los efectos centrales pierden el orden original (dado en Figura 2.1) que sí preserva la estrategia anterior, pero se siguen separando en los extremos, aunque con menor claridad, los efectos B y C como candidatos a estar activos.

2.5 El caso de dos datos anómalos

A diferencia de la estrategia que promedia los ocho efectos centrales sin signo para estimar el desfase causado por un dato anómalo, la estrategia de estimar la observación anómala a partir de igualar a cero el contraste de mayor orden se puede generalizar fácilmente a dos o más datos anómalos suponiendo nulas algunas de las interacciones triples. En particular con dos datos anómalos se supone nula la interacción de mayor orden ($ABCD$) y una de las interacciones triples, una de las que permita definir un sistema de dos ecuaciones con dos incógnitas, cuyas soluciones son los estimados de los datos anómalos que se supusieron perdidos. Consideremos a continuación este caso, para lo cual es conveniente recordar aquí la Tabla 2.3 de signos para el diseño factorial completo 2^4 que se muestra en la página siguiente.

Lo primero por notar es que la multiplicación módulo 2 de los dos datos anómalos declarados perdidos expresados en la notación de Yates da cierta información sobre cómo entran los datos extremos en los contrastes de los efectos: Si el resultado de la multiplicación es una palabra de 1 letra entonces todos los efectos que no contienen dicha letra tendrán sumados o restados simultáneamente en sus contrastes ambos datos anómalos, lo que puede sesgarlos bastante si ambos datos anómalos caen en el

mismo lado (ambos son grandes o ambos son pequeños). Por ejemplo, supongamos que los datos anómalos corresponden a (1) y a , entonces $a \times (1) = a$, de manera que todos los efectos que no contienen la letra a tendrán a ambos datos sumando o restando simultáneamente en su contraste. Estos efectos y signos se encierran en cuadritos en los primeros dos renglones de la Tabla 2.3. Específicamente son los efectos y signos

Tabla 2.3. Contrastes del factorial 2^4

<i>Trat.</i>	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
(1)	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+
a	+	-	-	-	-	-	-	+	+	+	+	+	+	-	-
<i>b</i>	-	+	-	-	-	+	+	-	-	+	+	+	-	+	-
<i>ab</i>	+	+	-	-	+	-	-	-	-	+	-	-	+	+	+
<i>c</i>	-	-	+	-	+	-	+	-	+	-	+	-	+	+	-
<i>ac</i>	+	-	+	-	-	+	-	-	+	-	-	+	-	+	+
<i>bc</i>	-	+	+	-	-	-	+	+	-	-	-	+	+	-	+
<i>abc</i>	+	+	+	-	+	+	-	+	-	-	+	-	-	-	-
<i>d</i>	-	-	-	+	+	+	-	+	-	-	-	+	+	+	-
<i>ad</i>	+	-	-	+	-	-	+	+	-	-	+	-	-	+	+
<i>bd</i>	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+
<i>abd</i>	+	+	-	+	+	-	+	-	+	-	-	+	-	-	-
<i>cd</i>	-	-	+	+	+	-	-	-	-	+	+	+	-	-	+
<i>acd</i>	+	-	+	+	-	+	+	-	-	+	-	-	+	-	-
<i>bcd</i>	-	+	+	+	-	-	-	+	+	+	-	-	-	+	-
<i>abcd</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

$$-B, -C, -D, +BC, +BD, +CD \text{ y } -BCD.$$

Se entiende que en el resto de los contrastes (efectos) uno de los datos extremos entra sumando y el otro restando, situación en la que se nulifican uno al otro si los dos datos extremos caen en mismo lado y tienen aproximadamente la misma magnitud.

Si el resultado de la multiplicación módulo 2 son dos letras, los efectos que se pueden ver más afectados por dos datos extremos o anómalos en el mismo lado son tanto los efectos que contienen ambas letras como los que no contienen ninguna de esas letras. Por ejemplo, considere que los datos extremos son las observaciones ac y

bc , y que ambos caen al mismo lado. Como $ac \times bc = ab$, entonces los efectos en los que ambos datos entran sumando o restando simultáneamente son

$$+C, -D, -AB, -CD, -ABC, +ABD \text{ y } +ABCD,$$

que son los efectos que tienen o no tienen ambas letras A y B (verificar en Tabla 2.3).

Si el resultado de la multiplicación módulo 2 son tres letras, los efectos que se ven más afectados por dos datos extremos son los efectos que no contienen ninguna letra y los efectos que contienen dos de las tres letras. Por ejemplo, considere que los datos extremos o anómalos son las observaciones abd y cd . Como $abd \times cd = abc$, entonces los efectos en los que ambos datos entran sumando o restando son

$$+D, +AB, -AC, -BC, +ABD, -ACD \text{ y } -BCD.$$

Finalmente, si el resultado de la multiplicación son 4 letras, como es el caso por ejemplo si los datos extremos caen en ab y cd , ($ab \times cd = abcd$) entonces los efectos que tendrán sumando o restando a ambos datos atípicos en sus contrastes son los que tienen 2 ó 4 de las letras, esto es,

$$+AB, -AC, -AD, -BC, -BD, +CD \text{ y } +ABCD.$$

Por ejemplo, consideremos otra vez los datos de la Tabla 2.2 agregando un segundo dato extremo en la observación c . Específicamente, sea $c = 61.47$ en lugar de $c = 51.47$. La gráfica de efectos en papel de probabilidad normal se muestra en la Figura 2.3. De acuerdo a lo discutido arriba los efectos que no tienen la letra d son los que se verán doblemente afectados por los datos extremos ya que ambos entran sumando o restando en sus contrastes. En la figura se observa que precisamente los efectos sin la letra d tienden a caer en los puntos extremos del gráfico, mientras que los efectos que tienen la letra d aparecen agrupados más cerca del centro. Se observa también la aparición de otros huecos o espaciamentos en el sentido horizontal entre los efectos debido al impacto de los dos datos extremos, dando ahora la impresión visual de que

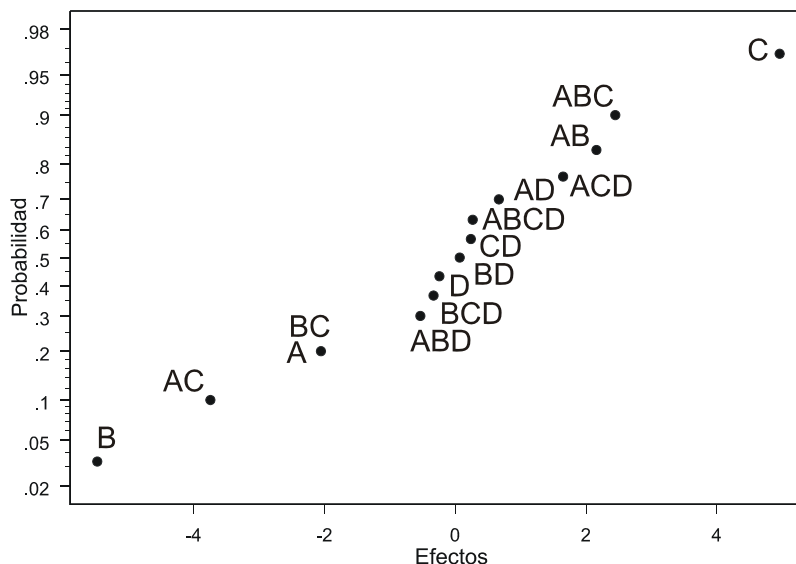


Figura 2.3: Efectos en papel de probabilidad normal con impacto de dos datos extremos

cuando menos el efecto AC está también activo. También es notorio el cambio de posición que sufrieron los efectos A y BC respecto a las figuras anteriores (2.1 y 2.2), donde se ubicaban claramente más al centro.

2.6 Reestimando dos observaciones sospechosas

Suponiendo las dos observaciones extremas como datos perdidos, éstas se pueden reestimar a partir de suponer nulos dos contrastes de interacciones de alto orden. Naturalmente, uno de los contrastes sería, en el caso de un factorial 2^4 , la interacción cuádruple $ABCD$ y el segundo alguna de las interacciones triples elegida esta última de manera que el sistema de ecuaciones resultante tenga solución.

Aplicando este procedimiento a la situación descrita arriba con datos extremos $c = 61.47$ y $cd = 59.15$, se obtiene el gráfico de Daniel de la Figura 2.4, que busca corregir la situación provocada por los datos extremos en la Figura 2.3. Efectivamente, de alguna manera se corrige un poco la situación anterior al observarse con mayor

potencial de estar activos a los efectos B y C . Los detalles de cómo decidir cuáles datos son anómalos, así como su reestimación al declararlos observaciones perdidas se discuten en el método de la Sección 4.3.

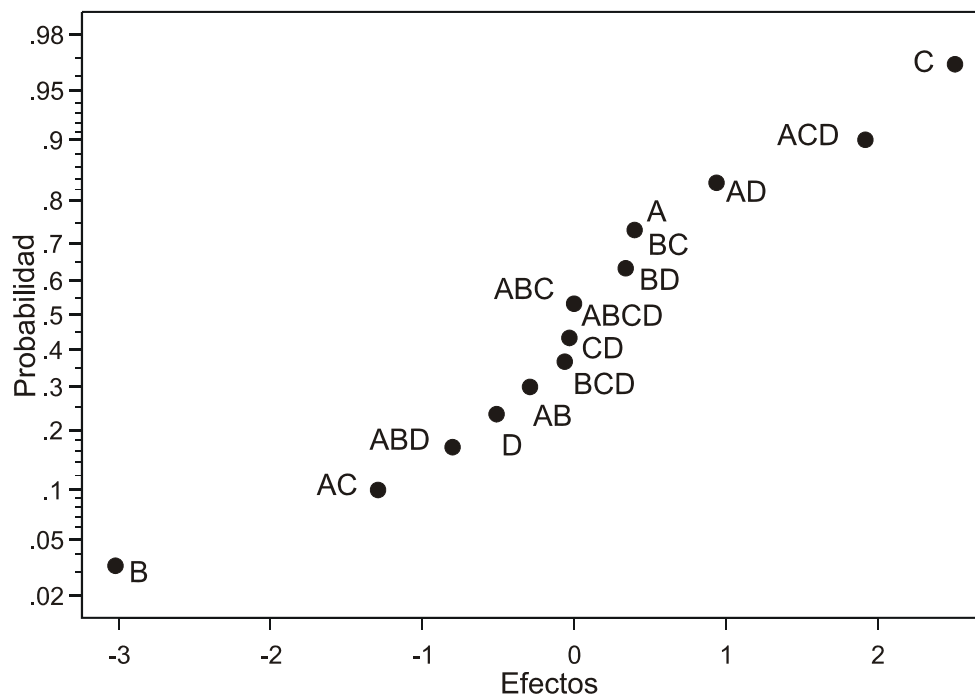


Figura 2.4. Efectos en papel de probabilidad normal con datos extremos corregidos

Capítulo 3

Métodos Formales Existentes para Analizar Factoriales con Posibles Datos Anómalos

En la literatura se han propuesto hasta ahora básicamente dos métodos para el análisis formal de diseños factoriales no-replicados y que contemplan la posibilidad de datos anómalos: uno de ellos es el método Bayesiano de Box y Meyer (1987) y el otro es el método de rangos de Aguirre-Torres y Pérez-Trejo (2003). En este capítulo se describen brevemente estos métodos y se ilustran con los datos de la Tabla 2.2. Se discuten algunos conceptos importantes para la comparación de los métodos.

3.1 Método Bayesiano

Box y Meyer (1986) proponen un método Bayesiano para analizar factoriales no-replicados sin contemplar la posibilidad de datos anómalos. Box y Meyer (1987) extienden el método Bayesiano para incluir la posibilidad de observaciones anómalas que pueden influir en las probabilidades posteriores de que los efectos sean activos. En este caso los conjuntos de interés $a_{(r_1, r_2)}$; $r_1 = 0, 1, 2, \dots, 15$; $r_2 = 0, 1, 2, \dots, 16$, son

todas las posibles combinaciones de efectos activos y de observaciones anómalas, que en el caso por ejemplo de un factorial 2^4 , son tantos como $2^{15} \times 2^{16}$, que es una cantidad enorme como para hacer los cálculos de manera exhaustiva. Los mismos autores proponen un procedimiento iterativo para aproximar las probabilidades posteriores tanto de que los efectos sean activos como de que los datos sean anómalos.

Sea la matriz \mathbf{X} formada por las columnas de todos los posibles efectos más una primera columna de unos. Sea $a_{(r_1, r_2)}$ el evento definido por un conjunto particular de r_1 efectos activos y un conjunto particular de r_2 observaciones anómalas. Esto es, r_1 representa las columnas de la matriz \mathbf{X} que corresponden a los efectos activos y r_2 son los renglones de la matriz \mathbf{X} y del vector \mathbf{y} que corresponden a los datos anómalos. Sea $M_i = M_{(r_1, r_2)}$ el modelo con parámetro β_i asociado con el evento $a_{(r_1, r_2)}$, que de acuerdo a esto tiene r_1 términos independientes. Notamos que la probabilidad posterior de cada $a_{(r_1, r_2)}$ es

$$p(a_{(r_1, r_2)} | \mathbf{y}) \propto p(M_{(r_1, r_2)}) p(\mathbf{y} | a_{(r_1, r_2)}) \quad (3.1)$$

donde $p(M_{(r_1, r_2)})$ es la probabilidad *a priori* del modelo y

$$p(\mathbf{y} | a_{(r_1, r_2)}) = \int_{\Theta} p(\mathbf{y} | a_{(r_1, r_2)}, \boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i | a_{(r_1, r_2)}) d\boldsymbol{\theta}_i, \quad (3.2)$$

donde $p(\mathbf{y} | a_{(r_1, r_2)}, \boldsymbol{\theta}_i)$ es la densidad del vector de observaciones dado $a_{(r_1, r_2)}$, y $p(\boldsymbol{\theta}_i | a_{(r_1, r_2)})$ es la distribución *a priori* de los parámetros donde $\boldsymbol{\theta}_i = (\beta_0, \beta_1, \dots, \beta_{r_1}, \sigma)$.

Para permitir la presencia de observaciones anómalas se supone que los errores del modelo lineal asociados con tales observaciones tienen una varianza inflada $k^2\sigma^2$ ($k > 1$), esto es, el error sigue la distribución

$$\varepsilon \sim (1 - \alpha_2)N(0, \sigma^2) + \alpha_2N(0, k^2\sigma^2),$$

donde α_2 es la probabilidad *a priori* de dato anómalo el cual proviene de la distribución $N(0, k^2\sigma^2)$. Sean \mathbf{X}_s y \mathbf{X}_r las submatrices de \mathbf{X}_{r_1} que corresponden a los \mathbf{y}_s

datos normales y \mathbf{y}_r datos anómalos. La distribución del vector de observaciones, dado $a_{(r_1, r_2)}$, está dada por

$$p(\mathbf{y} | a_{(r_1, r_2)}, \boldsymbol{\beta}_i, \sigma) = \sigma^{-n} k^{-r} \exp \left\{ -\frac{(\mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta}_i)' (\mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta}_i)}{2\sigma^2} - \frac{(\mathbf{y}_r - \mathbf{X}_r \boldsymbol{\beta}_i)' (\mathbf{y}_r - \mathbf{X}_r \boldsymbol{\beta}_i)}{2k^2 \sigma^2} \right\}.$$

Sean las distribuciones *a priori*

$$p(\boldsymbol{\beta}_i | a_{(r_1, r_2)}, \sigma) = \gamma^{-r_1} \sigma^{-r_1} \exp \left\{ -\frac{\boldsymbol{\beta}_i' \boldsymbol{\Gamma}_{r_1} \boldsymbol{\beta}_i}{2\sigma^2} \right\}$$

para los parámetros $\boldsymbol{\beta}_i = (\beta_1, \dots, \beta_{r_1})$, donde

$$\boldsymbol{\Gamma}_{r_1} = \frac{1}{\gamma^2} \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{r_1} \end{bmatrix},$$

y la distribución no informativa (ver Box y Tiao, 1973) dada por

$$p(\beta_0, \sigma | a_{(r_1, r_2)}) = p(\beta_0, \sigma) \propto \frac{1}{\sigma}$$

para los parámetros β_0 y σ . Se puede mostrar que sustituyendo estas distribuciones en la fórmula 3.2, resolviendo la integral y sustituyendo el resultado en la ecuación 3.1, así como la distribución *a priori* del modelo, (ver detalles en Apéndice) se llega a que la probabilidad posterior de que un conjunto particular de r_1 efectos esté activo y de que al mismo tiempo las observaciones r_2 sean anómalas (evento $a_{(r_1, r_2)}$), dados los datos \mathbf{y} , está dada por

$$p(a_{(r_1, r_2)} | \mathbf{y}) \propto \left(\frac{\alpha_1}{1 - \alpha_1} \right)^{r_1} \left(\frac{\alpha_2}{1 - \alpha_2} \right)^{r_2} \gamma^{-r_1} k^{-r_2} \quad (3.3)$$

$$\times \frac{|\mathbf{X}_{(0)} \mathbf{X}_{(0)}|^{1/2}}{|\boldsymbol{\Gamma}_{r_1} + \mathbf{X}_{(r_1)} \mathbf{X}_{(r_1)} - \varphi \mathbf{X}_{(r_1, r_2)} \mathbf{X}_{(r_1, r_2)}|^{1/2}}$$

$$\times \left(\frac{\mathbf{S}(\hat{\tau}_{(r_1, r_2)}) + \hat{\boldsymbol{\tau}}_{(r_1, r_2)} \boldsymbol{\Gamma}_{(r_1)} \hat{\boldsymbol{\tau}}_{(r_1, r_2)}}{\mathbf{S}(\hat{\tau}_{(0)})} \right),$$

donde \mathbf{y} es el vector de observaciones y $\mathbf{X}_{(r_1, r_2)}$ es la matriz de columnas y renglones de \mathbf{X} que corresponden a los r_1 efectos activos y r_2 observaciones anómalas, $\mathbf{X}_{(r_1)}$ contiene las columnas de \mathbf{X} que corresponden a los r_1 efectos activos, α_1 y α_2 son las probabilidades *a priori* de que un efecto esté activo y de que un dato sea anómalo, respectivamente; γ y k son los factores de inflamiento de la varianza debida a efecto activo y a dato anómalo, respectivamente; y además

$$\begin{aligned}\varphi &= 1 - 1/k^2, \\ \hat{\boldsymbol{\tau}}_{(r_1, r_2)} &= (\boldsymbol{\Gamma}_{r_1} + \mathbf{X}_{(r_1)}\mathbf{X}_{(r_1)} - \varphi\mathbf{X}_{(r_1, r_2)}\mathbf{X}_{(r_1, r_2)})^{-1} (\mathbf{X}_{(r_1)}\hat{\mathbf{y}} - \varphi\mathbf{X}_{(r_1, r_2)}\hat{\mathbf{y}}_{(r_2)}) \\ \mathbf{S}(\hat{\boldsymbol{\tau}}_{(r_1, r_2)}) &= (\mathbf{y} - \mathbf{X}_{(r_1)}\hat{\boldsymbol{\tau}}_{(r_1, r_2)})'(\mathbf{y} - \mathbf{X}_{(r_1)}\hat{\boldsymbol{\tau}}_{(r_1, r_2)}) - \\ &\quad \varphi(\mathbf{y}_{(r_2)} - \mathbf{X}_{(r_1, r_2)}\hat{\boldsymbol{\tau}}_{(r_1, r_2)})'(\mathbf{y}_{(r_2)} - \mathbf{X}_{(r_1, r_2)}\hat{\boldsymbol{\tau}}_{(r_1, r_2)})\end{aligned}$$

donde $\mathbf{y}_{(r_2)}$ son las supuestas observaciones anómalas. Ver Apéndice para mayores detalles de la demostración de la fórmula (3.3). Los autores sugieren los siguientes valores de la información *a priori*: $\alpha_1 = 0.2$, $\alpha_2 = 0.05$, $\gamma = 2.5$ y $k = 5$. Los valores de α_1 y γ han sido sugeridos por (Box y Meyer, 1986) con base en el análisis de varios experimentos de la literatura, a partir de contabilizar la proporción de efectos declarados activos y el cociente entre el cuadrado medio de un efecto activo en relación al cuadrado medio de un efecto inerte, respectivamente. La probabilidad *a priori* de dato anómalo ($\alpha_2 = 0.05$) se ha propuesto (Box y Tiao, 1968), porque la experiencia indica que a lo más se esperarían 1 ó 2 datos anómalos cada 20 observaciones en este tipo de experimentos (ver también Daniel, 1959). El valor del parámetro $k = 5$ se toma a partir de considerar qué tanto debe la varianza residual inflarse para permitir la presencia de datos anómalos. Por otra parte se han hecho estudios de sensibilidad (Meyer, 1987) para estudiar cómo cambian las conclusiones al variar los valores de estas probabilidades *a priori* y se ha encontrado que las conclusiones son robustas a cambios moderados en los parámetros. De cualquier manera se recomienda probar otras combinaciones de los parámetros en un ejemplo específico para comprobar la estabilidad de la solución obtenida.

La probabilidad posterior de que el efecto i sea activo se obtiene sumando las probabilidades posteriores de todos los eventos que contienen al efecto i , esto es,

$$p_i = \sum_{(r_1, r_2): i \text{ activo}} p(a_{(r_1, r_2)} | \mathbf{y}) \quad (3.4)$$

y de manera análoga la probabilidad posterior de que el dato y_j sea anómalo se obtendría como

$$q_j = \sum_{(r_1, r_2): y_j \text{ anómalo}} p(a_{(r_1, r_2)} | \mathbf{y}). \quad (3.5)$$

Dada la imposibilidad de calcular directamente estas sumas, el siguiente procedimiento iterativo aproxima las probabilidades deseadas: se empieza por suponer que no hay observaciones anómalas, esto es, en la fórmula (3.2) se toma $r_2 = 0$, $\mathbf{X}_{(r_1, r_2)} = \mathbf{0}_{(r_1)}$ y $\mathbf{y}_{(r_2)} = \mathbf{0}$, de manera que ésta se reduce al caso donde no se suponen observaciones anómalas (Box y Meyer, 1986). En un segundo paso y con los r_1 efectos que hayan resultado más importantes, digamos con probabilidades posteriores de 0.5 o más, se supone temporalmente el modelo correspondiente y se calculan las probabilidades posteriores de que los datos sean anómalos; así, en este segundo paso, r_1 y las columnas de las matrices $\mathbf{X}_{(r_1)}$ y $\mathbf{X}_{(r_1, r_2)}$ se mantienen fijas. Luego, en un tercer paso y suponiendo como datos anómalos los r_2 que tengan probabilidades posteriores altas (> 0.5) en el paso anterior, se vuelven a calcular las probabilidades posteriores de que los efectos estén activos. En este tercer paso r_2 vuelve a ser fijo al igual que los renglones de la matriz $\mathbf{X}_{(r_1, r_2)}$ y de $\mathbf{y}_{(r_2)}$. Y así sucesivamente. Este procedimiento generalmente converge en una o dos iteraciones, es decir, en la segunda o tercera vez que se calculan las probabilidades posteriores de efecto activo. Los programas para los métodos Bayesianos que incluyen o no la posibilidad de datos anómalos se pueden ver en De la Vara, R. (2004) y en el Apéndice.

Ejemplo con un dato anómalo. Para ilustrar el método Bayesiano se analizan los datos de la Tabla (2.2). Se comienza por calcular las probabilidades de efecto activo suponiendo que no hay datos anómalos. Los resultados obtenidos se muestran en la

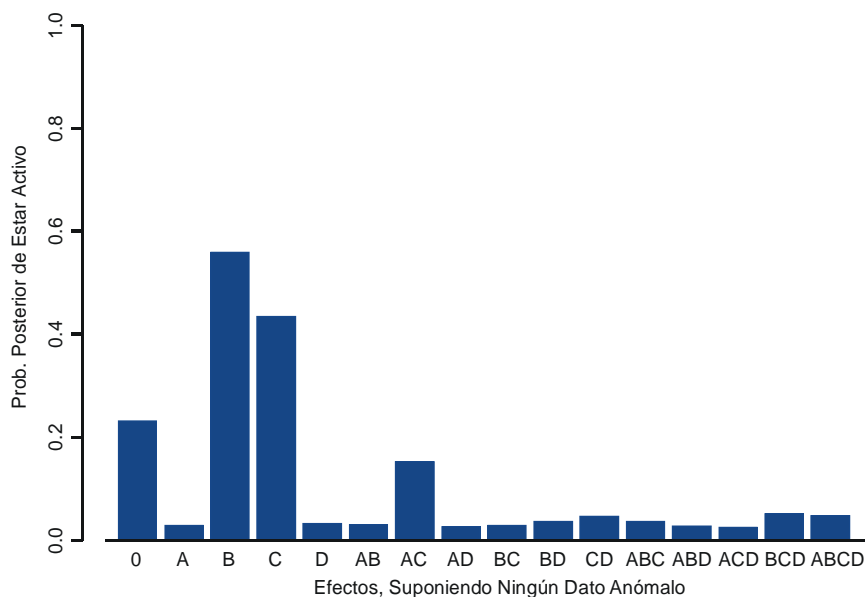


Figura 3.1: Probabilidad posteriores de efecto activo, suponiendo que no hay datos anómalos.

Figura 3.1, donde se observan relativamente importantes los efectos B y C , pero sus probabilidades posteriores de estar activos son apenas de alrededor de 0.5. En tercera instancia aparece el modelo constante con probabilidad posterior un poco arriba de 0.2.

Tomando por el momento como activos los efectos B y C , se obtiene la probabilidad posterior de que cada dato sea anómalo. En la Figura 3.2 se destaca claramente la alta probabilidad de ser anómala que tiene la la observación 13 (ver Tabla 2.2).

Considerando la observación 13 como anómala se recalculan las probabilidades posteriores de efecto activo, y los resultados se muestran en la Figura 3.3. Se aprecia que las probabilidades posteriores de que los efectos B y C sean activos son cercanas a 0.9, y aparecen también con probabilidades posteriores superiores a 0.5 los efectos AC y ACD .

Como último paso de este ejemplo, se supone ahora que los cuatro términos B , C ,

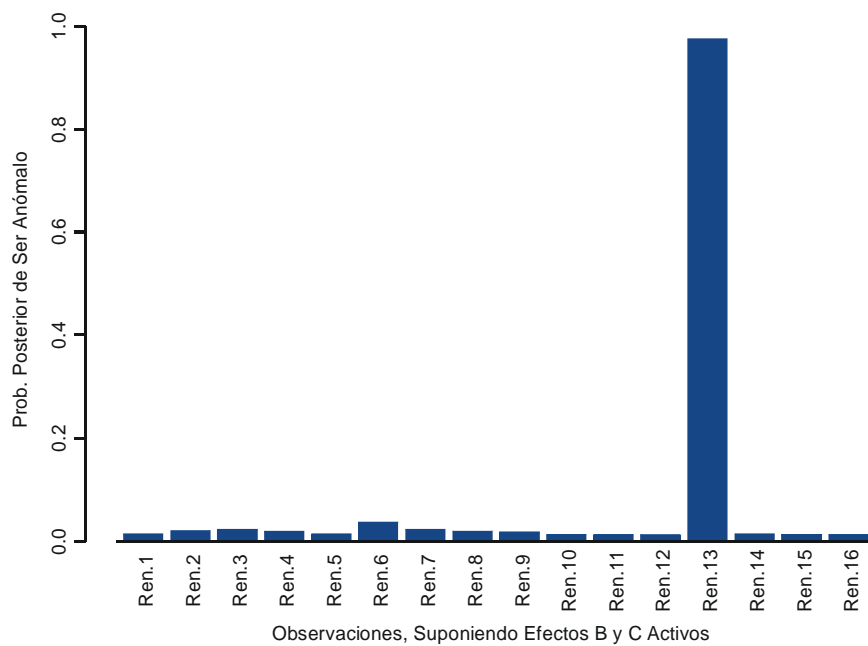


Figura 3.2: Probabilidad posterior de dato anómalo, suponiendo efectos B y C activos.

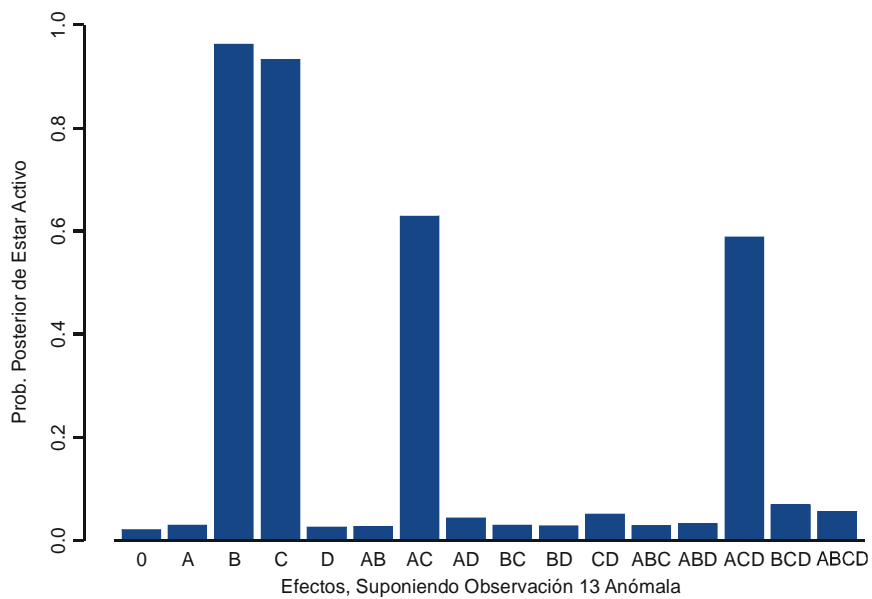


Figura 3.3: Probabilidades posteriores de efecto activo, suponiendo dato 13 anómalo.

AC y ACD están activos y se calculan otra vez las probabilidades posteriores de dato anómalo. El resultado, es parecido al de la Figura 3.2, esto es, se confirma que sólo la observación 13 es anómala, con lo cual se concluye precisamente éso, y que los efectos B , C , AC y ACD están activos.

Es importante hacer notar que este método (de Box y Meyer, 1987) tal como fue propuesto no ofrece ningún control del tamaño del error tipo I (ver Sección 3.3), y que muy probablemente está detectando más efectos activos de los que en realidad hay, tal como ocurre con el método Bayesiano que no contempla la posibilidad de datos anómalos propuesto por los mismos autores (Box y Meyer, 1986). Los estudios hechos a este último método muestran que declarar activos los efectos con probabilidades posteriores mayores a 0.5 da por resultado un exceso de detección de efectos activos, y por ende, un error tipo I grande.

Cabe aclarar que el error tipo I no es una cantidad Bayesiana y es por eso que el método de Box y Meyer (1987) no tendría en principio porqué contemplarlo. Sin embargo, el error tipo I del método Bayesiano sí puede evaluarse, al menos por simulación Monte Carlo, como de hecho lo han hecho autores como AT&PT (2001) y Chen (2003) para el método de Box y Meyer (1986).

Ejemplo con dos datos anómalos. Para ilustrar el método Bayesiano cuando hay 2 datos anómalos consideramos otra vez los datos de la Tabla 2.2 pero con el dato en la combinación $(-, -, +, -)$ modificado a $c = 61.47$ en lugar del $c = 51.47$ original. Los resultados obtenidos se muestran en la Figura (3.4). Se observa que el método detecta bastante bien los dos datos anómalos, pero no detecta con la misma claridad del análisis con un dato anómalo (Figura 3.3) los efectos activos. Pero sin duda mejora la detección en relación al análisis que ignora *a priori* la existencia de los datos discrepantes.

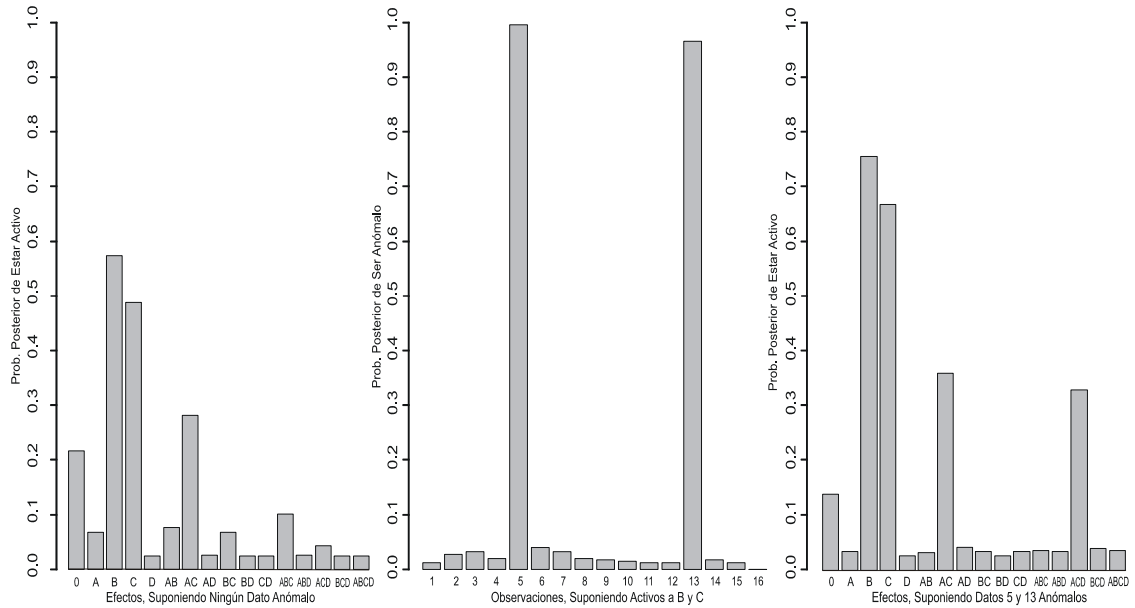


Figura 3.4: Método Bayesiano, con Dos Datos Anómalos

3.2 Método de Rangos

Aguirre-Torres (1993) introduce un método basado en rangos para la detección de efectos activos en experimentos factoriales no replicados que pudieran tener la presencia de anomalías en los datos. El rango de una observación es la posición que ésta ocupa en el vector de datos ordenados. En caso de empates en las observaciones, se asigna a ellas el rango promedio de los rangos que habrían tenido si no hubiera empates. Una vez obtenido el vector de rangos $R(\mathbf{Y})$ los efectos estimados basados en rangos, T_i^R ($i = 1, 2, \dots, 2^k - 1$), se obtienen de acuerdo a la fórmula usual

$$\begin{bmatrix} T_0^R \\ T_1^R/2 \\ \vdots \\ T_{2^k-1}^R/2 \end{bmatrix} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t R(\mathbf{Y}).$$

El método de Aguirre-Torres (1993) consiste básicamente en calcular y graficar en papel de probabilidad normal este vector de efectos basados en los rangos de las ob-

servaciones originales. El gráfico de probabilidad resultante se compara con el gráfico de los mismos efectos calculados en la escala original. En caso de discrepancias en los análisis, éstas son indicio de la presencia de anomalías en los datos, y entonces el análisis basado en la transformación de rangos es probablemente el más apropiado (Montgomery, 2001, pág.118). Esto es porque el análisis de la transformación de rangos es más resistente a la falta de normalidad o a la presencia de datos anómalos, y en todo caso cuando esa discrepancia entre los análisis ocurre, el experimentador también puede buscar alguna transformación a normalidad o determinar cuáles datos son atípicos para considerar su impacto en el análisis original. Note que el análisis propuesto por Aguirre-Torres (1993) es fundamentalmente subjetivo al basarse en el análisis visual de gráficas de probabilidad normal de los efectos.

Para darle objetividad al método de rangos, Aguirre-Torres y Pérez-Trejo (2001) proponen aplicar una variante del método de Benski (1989) al vector de rangos que resulta del vector de observaciones, logrando con ello tener control sobre el error tipo I. Benski (1989) propone un método para analizar experimentos factoriales en el caso normal, es decir, sin contemplar la posibilidad de datos anómalos. Básicamente este método aplica una prueba de normalidad tipo Shapiro-Wilk (Shapiro y Francia, 1972) al vector de efectos, que en caso de resultar significativa indica la presencia de efectos activos, y en ese caso se procede a identificarlos mediante una prueba para detectar datos anómalos.

Específicamente, en el método de rangos Aguirre-Torres y Pérez-Trejo (2001), a los efectos basados en rangos y ordenados dados por $T_{(i)}^R$ ($i = 1, 2, \dots, 2^k - 1$) se les aplica una prueba de normalidad basada en el estadístico de prueba dado por

$$W' = \frac{\left(\sum_{i=1}^{2^k-1} m_i T_{(i)}^R\right)^2}{\sum_{i=1}^{2^k-1} m_i^2 \sum_{i=1}^{2^k-1} \left(T_{(i)}^R - \bar{T}\right)^2}, \quad (3.6)$$

con $m_i = \Phi^{-1}(p_i)$, donde Φ^{-1} es el inverso de la distribución normal y $p_i = (i - \alpha) / (2^k - 2\alpha + 1)$ son probabilidades repartidas en el intervalo $[0,1]$, donde los valores de

α se definen como (Benski, 1989)

$$\alpha_i = \begin{cases} 0.275499 + 0.072884 \times [\ln(2^k - 1)]^{0.41148}, & \text{si } 1 < i < 2^k - 1 \\ 0.205146 + 0.1314965 \times [\ln(2^k - 1)]^{0.226701}, & \text{si } i = 1 \text{ ó } i = 2^k - 1 \end{cases}$$

El p-valor P_1 de la prueba de normalidad se obtiene de

$$P_1 = \exp(C), \text{ para } P_1 > 0.005,$$

donde

$$\begin{aligned} C &= \frac{(W' - A)/B + 0.0486128}{0.02760309} - \ln(100) \\ A &= 1.031918 - 0.183573 \times [0.1(2^k - 1)]^{-0.5447402} \\ B &= -0.5084706 + 2.076782 \times [0.1(2^k - 1)]^{-0.4905993}. \end{aligned}$$

En caso de rechazar la prueba de normalidad, lo que ocurre cuando $P_1 < 0.033$, se concluye que al menos un efecto está activo y se procede a identificar a los efectos activos mediante la prueba de dispersión de cuartos, que consiste en declarar activos los efectos que caigan fuera del intervalo $[-2d_F, 2d_F]$, donde d_F es el rango intercuartílico estimado de los efectos. La significancia predefinida de 0.033 se determina mediante simulación Monte Carlo (ver Sección 7.1).

Ejemplo. La gráfica de efectos en papel de probabilidad normal utilizando el método de rangos se muestra en la Figura 3.5. Se observa una separación clara de los efectos B y C , y no tan clara de AC . En cuanto a los resultados analíticos del método se encuentra que $W' = 0.87$, p -valor = 0.044 y $2d_F = 3.25$. Los efectos estimados son

A	B	C	D	AB	AC	AD	BC
0.00	-6.25	4.75	0.00	-0.75	-3.25	1.00	0.50
BD	CD	ABC	ABD	ACD	BCD	$ABCD$	
-0.25	1.25	1.00	0.25	2.25	-1.00	1.50	

De aquí que, de acuerdo a las magnitudes de los efectos y con $\alpha = 0.05$, se concluye que B , C y AC están activos. Esto es, los efectos B y C rechazan claramente la prueba

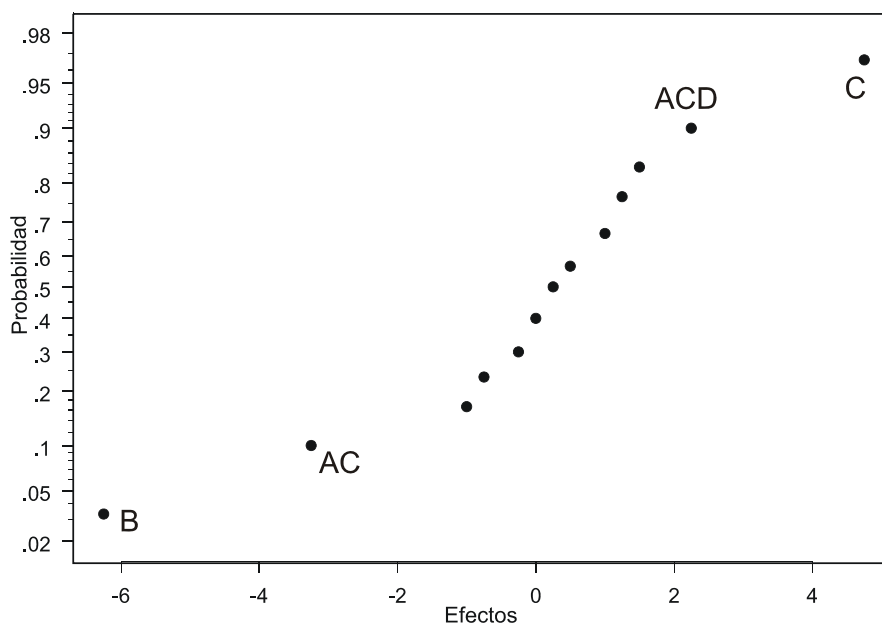


Figura 3.5: Efectos en papel de probabilidad normal, método de rangos

de datos anómalos y el efecto AC coincide con el valor crítico. El p -valor = 0.044 es mucho más pequeño que el que resultaría de analizar directamente los datos sin la transformación de rangos (ver Tabla 6.4 del Capítulo 6). Sin embargo, usando estrictamente el valor calibrado $\alpha = 0.033$, (Sección 3.3) se afirmarí­a que ningún efecto está activo, puesto que se aceptaría la normalidad de los efectos.

Si agregamos el segundo dato anómalo $c = 61.47$ a los datos de la Tabla 2.2 y re-hacemos el análisis con el método de rangos, se sigue observando una clara separación de los efectos B y C como activos e incluso el efecto AC se aprecia ahora un poco más claro (Figura 3.6). En cuando a los resultados analíticos, se tiene que $W' = 0.87$, p -valor = 0.033 y $2dF = 3.25$. El p -valor prácticamente coincide con el parámetro $\alpha = 0.033$, lo cual confirma que el método de rangos está haciendo en este ejemplo incluso un mejor trabajo con los dos datos anómalos en relación al análisis previo con un solo dato anómalo. De acuerdo a las magnitudes de los efectos estimados,

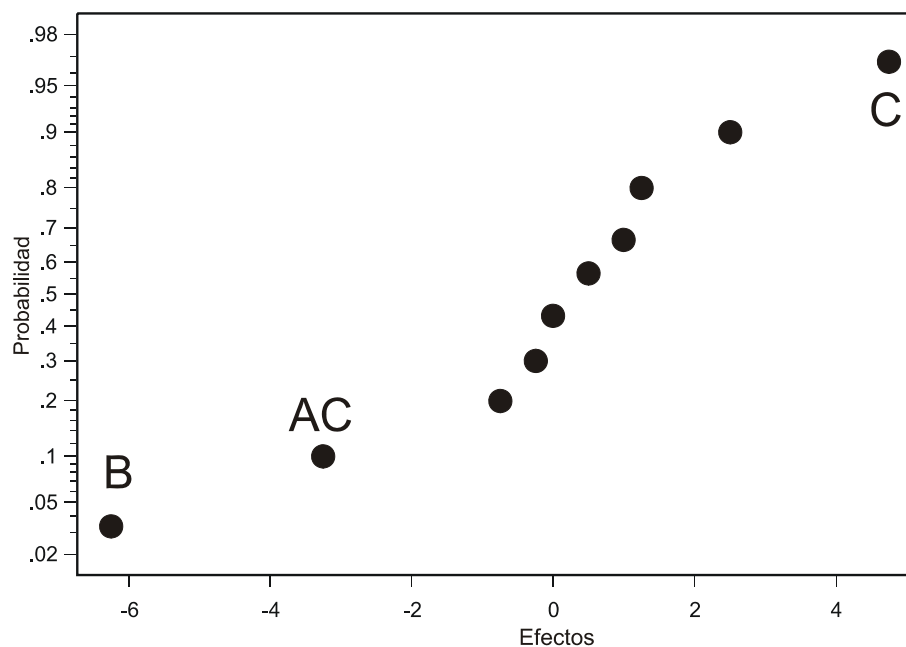


Figura 3.6: Efec. en Papel Normal, Método de Rangos, caso con Dos Datos Anómalos

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>AB</i>	<i>AC</i>	<i>AD</i>	<i>BC</i>
0.00	-6.25	4.75	-0.25	-0.75	-3.25	1.25	0.50
<i>BD</i>	<i>CD</i>	<i>ABC</i>	<i>ABD</i>	<i>ACD</i>	<i>BCD</i>	<i>ABCD</i>	
0.00	1.00	1.00	0.00	2.50	-0.75	1.25	

se declararían activos en la prueba de dispersión de cuartos los tres efectos mencionados.

3.3 Sobre la Comparación de los Métodos

Para poder hacer una comparación justa de los métodos para analizar factoriales no-replicados, tanto en el caso que no considera la posibilidad de datos anómalos como el caso que si la considera, se requiere que el desempeño sea igual cuando todos los efectos son inertes. Difícilmente se cumple esto con los métodos tal cual fueron propuestos. Sin embargo, en ocasiones los métodos se pueden calibrar o ajustar para que su desempeño sea igual cuando no hay efectos activos. Dos medidas de error

comúnmente usadas para este propósito son el error global EER (experimentwise error rate) y el error individual IER (individual error rate). Suponiendo que no hay efectos activos, el EER es el porcentaje de experimentos en la simulación donde ocurre al menos una falsa detección y el IER es el porcentaje de efectos inertes declarados activos considerando el total de efectos estimados en la simulación.

En el caso que no considera la posibilidad de datos anómalos se han reportado varios estudios donde se compara el desempeño de diferentes conjuntos de métodos: Hamada & Balakrishnan (1998) comparan 24 métodos controlando el error individual en 0.044. Haciendo esto, el error global EER varía entre los métodos comparados desde 0.1 hasta 0.44. Aguirre-Torres y Pérez-Trejo (2001) comparan 11 de los métodos bajo condiciones de contaminación considerando 3 efectos activos, y previamente fijando el EER en 0.05. Chen, Y. (2003) compara 13 métodos, calibrando los métodos a un error global (EER) de 0.05. Se desprende de estos estudios que no existe un método que se pueda afirmar que es el mejor en todos los escenarios posibles: el desempeño de un método depende de aspectos tales como el número de efectos activos considerados, de los tamaños de dichos efectos, de los valores en que controla el EER o el IER y de los criterios para medir su desempeño.

El método Bayesiano que considera la posibilidad de datos anómalos resulta difícil de calibrar dado el tiempo que consume cada análisis (media hora), pero hay evidencia para afirmar que las probabilidades críticas de $P = Q = 0.5$ sugeridas por Box y Meyer (1987) dan lugar a un exceso de detecciones falsas, y que en realidad dichas probabilidades deberían de ser mucho más altas para tener por ejemplo un $EER = 5\%$. La evidencia se fundamenta en lo que sucede con el método Bayesiano de Box y Meyer (1986) el cual trabaja (ver Aguirre-Torres y Pérez-Trejo, 2001) con un EER de alrededor de 27%, con lo que la probabilidad crítica ($P = 0.5$) en ese método debe incrementarse hasta $P = 0.89$ para tener un $EER = 5\%$.

El método de rangos trabaja con un $EER = 5\%$ si se toma $\alpha = 0.033$ en lugar de $\alpha = 0.05$ en la prueba de normalidad lo cual se determinó con un estudio de simulación

basada en 50000 experimentos: Se generaron 50000 experimentos factoriales 2^4 sin considerar efectos activos, es decir, los 16 datos de cada experimento se obtienen de la misma distribución $N(\mu, \sigma^2)$, por lo que el valor esperado de cada efecto es igual a cero. Para cada conjunto de datos se aplica el método de rangos y se guarda el p-valor de la prueba de normalidad de los efectos, y también se guarda el hecho de que la prueba de datos anómalos detecte al menos un efecto activo. Finalmente se elige como el valor α calibrado aquel donde solo el 5% de los experimentos detectaron al menos un efecto activo. Cabe agregar que si se aplicara sólo la prueba de datos anómalos esta detectaría efectos activos en alrededor del 30% de los experimentos.

Por lo anterior, no es obvio comparar estos dos métodos dada la falta de una calibración adecuada del método Bayesiano. En el ejemplo particular discutido en este capítulo, es aparente que el método Bayesiano detecta más efectos activos: pero posiblemente ajustando adecuadamente la probabilidad crítica, solo detectaría los efectos B y C . Por su parte, el método de rangos no detectaría ningún efecto activo si estrictamente se utilizara $\alpha = 0.033$, dado que el p – *valor* estimado resulta ser 0.044; hay alguna evidencia importante aunque no suficiente de que los efectos B y C estarían activos.

Una vez calibrados los métodos se puede proceder a comparar su desempeño, por ejemplo en términos de su potencia para detectar uno o varios efectos activos de magnitudes diferentes o iguales. La potencia con la que se detecta un efecto activo se define como el porcentaje de las veces que se declara activo en la simulación. Otra medida importante para comparar métodos, que toma en cuenta tanto los efectos activos declarados activos como los efectos inactivos erróneamente declarados activos, es la figura de mérito Q introducida por Benski y Cabau (1995). Esta se define como

$$Q = \frac{n^+}{N^+} \left(1 - \frac{n^-}{N^-} \right),$$

donde N^+ es el total de efectos activos, n^+ es el número de efectos activos declarados como tal, N^- es el total de efectos inertes y n^- es el número de efectos inertes

erróneamente juzgados como activos. Otros criterios de evaluación y comparación se pueden ver en Chen (2003) págs. 67-70. En los Capítulos 5 y 6 se utilizan la potencia y la figura de mérito Q para comparar los métodos, primero usando ejemplos específicos y luego mediante un estudio de simulación.

La razón fundamental de utilizar técnicas de simulación Monte Carlo para comparar el desempeño de los métodos propuestos en lugar de utilizar técnicas analíticas es que con estos métodos no se cuenta con estadísticos que sigan alguna distribución específica conocida bajo la hipótesis de que sólo hay efectos nulos. Aún en los métodos directos en el caso normal que son aquellos donde se propone un pseudoerror para estandarizar los efectos (H&B, 1998), típicamente no se conoce la distribución de referencia exacta ni aproximada del estadístico resultante que pueda utilizarse para obtener la curva de potencia del método. En los métodos que consideran la posibilidad de datos anómalos, que son los de interés en este trabajo de tesis, es todavía más intratable el problema de encontrar alguna distribución de referencia.

Capítulo 4

Nuevas Propuestas para el Análisis de Factoriales No-Replicados con Posibles Datos Anómalos

En este capítulo se presentan algunos métodos nuevos que superan el desempeño del método de AT&PT (2001), partiendo del argumento de que al aplicar la transformación de rangos a la respuesta se pierde información relevante sobre la significancia de los efectos. Es claro, por ejemplo, que el método de rangos no considera las distancias relativas entre las observaciones del experimento, lo que causa que las distancias relativas entre las magnitudes de los efectos no se estimen adecuadamente. Esto es, si bien la transformación de rangos elimina el impacto de los datos extremos o anómalos, por otro lado causa distorsiones en la estimación de los efectos, mismas que acaban reflejándose en la potencia del método.

Para detectar los efectos activos en todas las propuestas de métodos que se hacen se explota fundamentalmente la variante del Benski (1989) que utilizan AT&PT (2001), esto es, una prueba de normalidad seguida de una prueba para datos anómalos: la primera nos dice cuando hay al menos un efecto activo, y cuando los hay, la segunda nos dice cuáles específicamente son los efectos activos. En cada método

propuesto se busca llegar a la prueba de normalidad con efectos a los que se haya eliminado, de alguna manera, el impacto de los posibles datos anómalos. Esto se logra fundamentalmente de tres maneras diferentes en los cinco métodos propuestos: 1. Detectando y corrigiendo las posibles observaciones atípicas, 2. Estimando de manera robusta los efectos y 3. Corrigiendo el impacto de los posibles datos anómalos en los efectos usuales no robustos. Esto es, la robustez se puede lograr antes, durante y después de la estimación de los efectos.

Los cinco métodos nuevos que se proponen entre este capítulo y el siguiente son:

1. Rangos modificados.
2. Ajuste de espacios entre efectos ordenados
3. Observaciones anómalas reestimadas
4. Técnicas de regresión robusta (variante I)
5. Técnicas de regresión robusta (variante II)

Como su nombre lo dice el método de Rangos Modificados cambia los rangos interiores a valores que recuperan las distancias interiores entre las observaciones originales, de manera que los rangos modificados son una especie de combinación de rangos y estadísticas de orden reescaladas. El método de Ajuste de Espacios entre Efectos Ordenados busca corregir el posible impacto de los datos atípicos en los efectos usuales no robustos. El método de Observaciones Anómalas Reestimadas incorpora alguna de las ideas de Box (1990) comentadas en el Capítulo 2, con el propósito de reestimar los datos anómalos como si fueran datos faltantes. Estos primeros tres métodos se describen en el presente capítulo. Los métodos basados en Técnicas de Regresión Robusta se exponen en el Capítulo 5.

En la descripción de los métodos básicamente nos enfocamos a los pasos de cada método hasta antes de aplicar la prueba de normalidad de la variante del método de Bensi (1989), y se ilustran esos primeros pasos con los datos de la Tabla (2.2). Es decir, no es objetivo de este capítulo presentar los análisis de significancia ni hacer comparaciones de los métodos. Los análisis de significancia de este y otros

ejemplos comparando los métodos se presentan hasta el Capítulo 6. La comparación vía simulación Monte Carlo se presenta en el Capítulo 7.

Es importante enfatizar que aunque utilizamos principalmente el experimento factorial 2^4 para ilustrar los métodos nuevos, éstos se pueden aplicar con modificaciones menores a cualquier miembro de la familia de experimentos factoriales con 16 corridas que van desde el factorial completo 2^4 hasta la fracción saturada 2^{15-11} (ver Sección 2.1). Con modificaciones "mayores" se podrían aplicar a otros tamaños de experimentos factoriales no-replicados.

4.1 Método de rangos modificados

La idea de este método surge de observar que el análisis de los rangos, si bien protege de posibles datos anómalos, puede causar distorsiones en la estimación de los efectos: por ejemplo, un efecto que es claramente importante en la escala original se ve menos claro en la escala de los rangos, como se aprecia en la Figura (4.1), haciendo coincidir para fines de la representación la escala de los rangos con la escala original arbitraria. Cuando un efecto es muy claro en la escala original todos los datos que restan en su contraste se separan notoriamente de los datos que suman, como se observa en los puntos de la parte superior de la figura. Sin embargo, al considerar la transformación de rangos, éstos pasan a quedar equidistantes entre ellos como se aprecia en la parte intermedia de la Figura (4.1), de manera que el efecto X que se calcula con los rangos resulta más pequeño o menos claro que el original. Pero si se ajustaran los rangos interiores de manera proporcional a las distancias originales es posible atenuar ese impacto negativo en la estimación del efecto X, lográndose entonces un efecto más parecido al original, como se aprecia en los puntos de la escala inferior de la Figura (4.1).

El método de rangos modificados consiste básicamente en recuperar las distancias relativas entre observaciones interiores. Esto es, se asignan los rangos como antes,

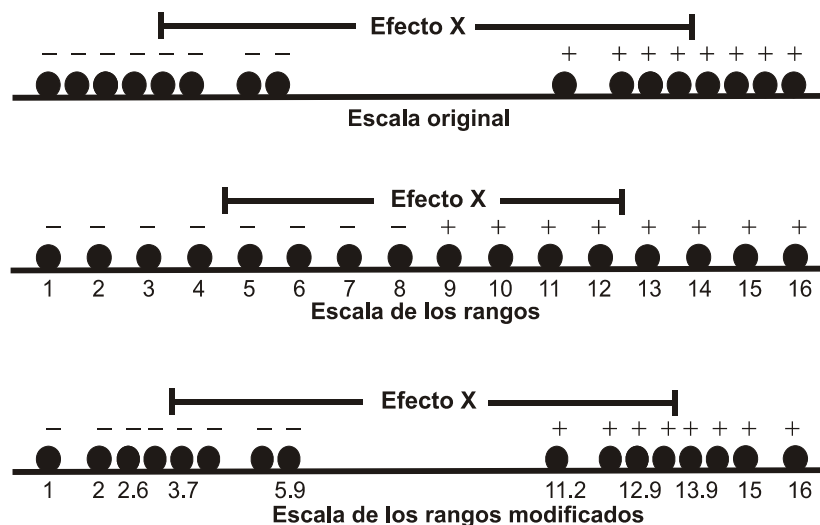


Figura 4.1: Escala original, escalas de rangos y rangos modificados

que por ejemplo, en un factorial con 16 corridas son los números del 1 al 16. Los rangos interiores del 3 al 14 se ajustan de acuerdo a las distancias relativas entre las observaciones originales que dichos rangos representan. Sean los rangos R_1, R_2, \dots, R_{16} que se asignan a las observaciones ordenadas $Y_{(1)}, Y_{(2)}, \dots, Y_{(16)}$. Las dos observaciones más chicas y las dos observaciones más grandes toman sus rangos usuales, pero a las observaciones intermedias se les asignan los "rangos modificados"

$$R_i = R_{i-1} + \left(\frac{Y_{(i)} - Y_{(i-1)}}{Y_{(15)} - Y_{(2)}} \right) \times 13; \quad i = 3, 4, \dots, 14.,$$

recuperando de esta manera las distancias relativas intermedias originales. Por ejemplo, para los datos de la Tabla 2.2 se llega a los rangos modificados de la Tabla 4.1.

En lugar de los rangos usuales se propone analizar los rangos modificados dados en la tabla. Es claro que este ajuste solo puede proteger adecuadamente de un dato anómalo en cada lado, puesto que si se observan dos datos anómalos en el mismo lado los rangos modificados R_{14} y R_{15} estarán alejados como lo están las observaciones originales ordenadas $Y_{(14)}$ y $Y_{(16)}$. La protección de dos datos anómalos en cada lado

se daría fijando tres rangos en cada extremo, sin embargo al reducirse la cantidad de rangos interiores por ajustar, el método acaba perdiendo la potencia ganada. Es posible que con experimentos factoriales con más de 16 corridas se puedan fijar tres rangos y tener todavía mejores resultados.

Tabla 4.1. Rangos Modificados

<i>datos</i>	<i>datos ordenados</i>	<i>rangos</i>	<i>rangos ajustados</i>
(1) = 47.46	43.13	1	1
<i>a</i> = 49.62	44.45	2	2
<i>b</i> = 43.13	44.83	3	2.70
<i>ab</i> = 46.31	46.10	4	5.05
<i>c</i> = 51.47	46.31	5	5.44
<i>ac</i> = 48.49	46.76	6	6.28
<i>bc</i> = 49.34	47.02	7	6.76
<i>abc</i> = 46.10	47.46	8	7.57
<i>d</i> = 46.76	47.90	9	8.39
<i>ac</i> = 48.56	48.49	10	9.48
<i>bd</i> = 44.83	48.56	11	9.61
<i>abd</i> = 44.45	49.34	12	11.05
cd = 59.15	49.62	13	11.57
<i>acd</i> = 51.33	51.33	14	14.74
<i>bcd</i> = 47.02	51.47	15	15
<i>abcd</i> = 47.90	59.15	16	16

4.2 Método de ajuste de espacios entre efectos

Como se menciona en el Capítulo 2, cuando se tienen observaciones anómalas éstas causan huecos o espacios entre los efectos estimados, como se aprecia en los gráficos en papel de probabilidad normal de las figuras 2.1 y 2.3. La idea de este método es ajustar los 3 huecos más grandes de los 6 huecos centrales que ocurren entre los efectos estandarizados y ordenados, antes de proseguir con la prueba de normalidad. Específicamente este método consiste en lo siguiente:

1. Primero se estiman los efectos tradicionales normales T_i .
2. Los efectos se ordenan y se obtienen los ocho huecos centrales $T_{(i)} - T_{(i-1)}$, $i = 5, 6, \dots, 12$.

3. Con estos huecos centrales se obtiene un estimador de la desviación estándar residual, dado por la mediana de los huecos dividida por 0.07. Esto es,

$$\hat{\sigma} = \text{mediana}(\text{hue cos})/0.07$$

donde la constante 0.07 se determinó por simulación. Con la fórmula $\sqrt{CM_E/2^{k-2}}$ se obtiene un estimador del error estándar de los efectos, donde el cuadrado medio del error $\sqrt{CM_E}$ es precisamente el estimador mencionado de la desviación estándar residual basado en los 8 espacios centrales entre los efectos ordenados.

4. Se procede a estandarizar los efectos, dividiéndolos entre su error estándar,

$$\frac{T_i}{\hat{\sigma}/2}; \quad i = 1, 2, \dots, 15.$$

5. Se ordenan los efectos estandarizados, que denotamos igual que los originales.

6. Se estiman los seis huecos centrales entre efectos estandarizados

$$T_{(i)} - T_{(i-1)}; \quad i = 6, \dots, 11.$$

7. Se identifican los tres espacios más grandes.

8. Se ajustan esos tres espacios mayores a valores pequeños. En particular se proponen 0.023, 0.015 y 0.01 para los correspondientes tres huecos en orden descendente. La idea de estos valores pequeños es básicamente cerrar los huecos potencialmente generados por datos anómalos. Los tamaños relativos entre los valores son aproximadamente los mismos que ocurren en cuantiles inferiores (abajo de 10%) de las distribuciones simuladas de dichos huecos. Cada ajuste consiste en sumar una constante a todos los efectos a la izquierda del hueco que se está ajustando, y restar esa misma cantidad a todos los efectos que caen a la derecha de dicho hueco.

9. Se obtiene el vector de efectos actualizado que considera los espacios ajustados. Evidentemente son efectos más compactados que los efectos normales.

10. Se procede de la manera usual con el método de Benski modificado.

Una consecuencia de este ajuste drástico que se aplica siempre es que el vector de efectos se aleja de la normalidad aún cuando no haya efectos activos, por lo tanto

se encuentra que el p-valor calibrado a utilizar en la prueba de normalidad es 0.001, para poder tener un EER final de 5% (Capítulo 7). El gráfico de efectos en papel de probabilidad normal debe verse con reservas ya que puede estar enviando un mensaje visual de efectos activos que no son tal.

Ilustramos los primeros nueve pasos del procedimiento usando los datos de la Tabla 2.2. En las columnas 1 y 2 se muestran los datos y efectos tradicionales ordenados, respectivamente. Los huecos o espacios entre efectos se muestran en la tercera columna, y se resaltan los 8 huecos centrales que se utilizan para estimar el error estándar de los efectos como

$$\text{Error estándar de } T_i = \frac{\hat{\sigma}}{2} = \frac{\text{mediana}(\text{huecos})}{2 \times 0.07}$$

donde la constante que divide se encuentra por simulación, esto es, haciendo los cálculos para 50000 experimentos con error estándar conocido se determina la constante que logra igualdad en promedio al error estándar dado. Se estandarizan los efectos en la columna 4. En la columna 5 se calculan los huecos entre efectos estandarizados y se resaltan los tres huecos más grandes de los 6 huecos centrales. Estos tres huecos se ajustan y se recalculan los efectos ajustados que se muestran en la última columna. Por ejemplo, el hueco 0.67 se ajusta a 0.023 sumando la cantidad $(0.67 - 0.023)/2$ a los efectos a la izquierda del hueco y restando esa misma cantidad a los efectos a la derecha. De manera similar se hacen los ajustes del hueco 0.22 a 0.015 y de 0.15 a 0.01. Los tres ajustes entran sumando al efecto estandarizado $A = -2.88$ de tal forma que este se transforma en

$$A_{ajustado} = -2.88 + (0.67 - 0.023)/2 + (0.22 - 0.015)/2 + (0.15 - 0.01)/2 = -2.38.$$

Tabla 4.2. Método de espacios ajustados

<i>datos</i>	Efectos ordenados	Huecos entre efectos	Efectos estandarizados	Espacios entre efectos estan.	Efectos ajustados
(1) = 47.46	–	–	–	–	–
<i>a</i> = 49.62	<i>B</i> = –4.22	–	–2.88	–	–2.38
<i>b</i> = 43.13	<i>AC</i> = –2.49	1.73	–1.70	1.18	–1.20
<i>ab</i> = 46.31	<i>BCD</i> = –1.58	0.91	–1.08	0.62	–0.58
<i>c</i> = 51.47	<i>BD</i> = –1.18	0.40	–0.80	0.27	–0.31
<i>ac</i> = 48.49	<i>A</i> = –0.80	0.38	–0.54	0.26	–0.05
<i>bc</i> = 49.34	<i>BC</i> = –0.80	0.00	–0.54	0.00	–0.05
<i>abc</i> = 46.10	<i>AD</i> = –0.58	0.22	–0.39	0.15	–0.04
<i>d</i> = 46.76	<i>ACD</i> = 0.40	0.98	0.27	0.67	–0.02
<i>ac</i> = 48.56	<i>ABD</i> = 0.72	0.32	0.49	0.22	–0.00
<i>bd</i> = 44.83	<i>AB</i> = 0.91	0.19	0.62	0.13	0.12
<i>abd</i> = 44.45	<i>D</i> = 1.01	0.10	0.69	0.07	0.19
cd = 59.15	<i>ABC</i> = 1.20	0.19	0.82	0.13	0.32
<i>acd</i> = 51.33	<i>CD</i> = 1.49	0.29	1.02	0.20	0.52
<i>bcd</i> = 47.02	<i>ABCD</i> = 1.52	0.03	1.04	0.02	0.54
<i>abcd</i> = 47.90	<i>C</i> = 3.71	2.19	2.53	1.49	2.04

4.3 Método de datos anómalos reestimados

Este método consiste en reestimar los datos anómalos como si fueran observaciones perdidas, como lo sugiere Box (1990). El método considera sólo los casos de uno o dos datos anómalos. La reestimación de los datos anormales como "datos perdidos", una vez identificados, se obtiene de suponer el contraste del efecto de mayor orden igual a cero para el dato más alejado, y para el segundo dato extremo se considera nula una de las interacciones del mayor orden posible. En el caso del factorial 2^4 se trabajaría con la interacción de mayor orden *ABCD* y con alguna de las interacciones triples (*ABC*, *ABD*, *ACD* y *BCD*). En el caso de un dato anómalo se resuelve una ecuación cuya solución es el dato reestimado y en el caso de dos datos atípicos se resuelve un sistema de ecuaciones con dos incógnitas que corresponden a las estimaciones de las dos observaciones extremas.

Este método resulta interesante dada la combinación que se hace de técnicas de naturaleza un tanto diferente como son la regresión L1, análisis de cúmulos (cluster

analysis) jerárquico con encadenamiento simple, estimación de datos perdidos, prueba de normalidad y prueba para detección de datos anómalos. La idea de utilizar análisis de cúmulos para la detección de datos anómalos se tomó de Sebert *et al.* (1997). Específicamente, el método consiste de los siguientes pasos:

1. Primero se obtienen estimadores robustos de los efectos para un modelo que considera hasta las interacciones dobles, para lo cual se utiliza regresión L1. Se sabe (Rousseeuw y Leroy, 1987, pág. 10) que esta regresión es robusta a valores extremos de la respuesta.
2. Usando los valores estimados y los residuos se hace un análisis de cúmulos con el método jerárquico de encadenamiento simple para identificar los datos que se agrupan al final, que corresponden a datos anómalos potenciales.
3. En cada paso del análisis de cúmulos se tienen las distancias a las cuales se van agrupando los datos. En este proceso se debe tomar en cuenta que no es lo mismo que los dos datos potencialmente anómalos se agrupen al final de manera separada a datos ya agrupados previamente, a que se agrupen primero entre ellos, ya que en este último caso la distancia de agrupación a considerar ocurre necesariamente en un paso posterior, justo cuando ambos datos se agrupan a datos de grupos previos. Un ejemplo de esto se muestra en la Sección 4.3.1, en particular en la Tabla (4.6). Allí las dos observaciones anómalas se agrupan entre ellas en el paso 12, pero su distancia relevante de agrupación al resto de los datos ocurre hasta el paso 15.
4. Usando las distancias de agrupación mencionadas en el paso anterior, se obtiene un umbral a partir del cual poder considerar anómalo un dato. Esto es, hay una distancia mínima de agrupación que indica cuando el dato se puede considerar una observación anómala.
5. Una vez identificado el o los datos anómalos, se procede a reestimarlos utilizando en este propósito la interacción de mayor orden y una de las interacciones triples, si fueran dos los datos anómalos por reestimar. La interacción triple se elige considerando que el sistema de ecuaciones resultante tenga solución.

6. Los datos encontrados anómalos se sustituyen por los datos reestimados.
7. Con el vector de datos corregido se procede con el método de Benski de la manera usual.

Consideremos una vez más el ejemplo de la Tabla 2.2 para mostrar los 6 primeros pasos del método de reestimación de datos extremos. Lo primero es obtener el modelo de regresión L1 que considera hasta interacciones dobles, que en este caso está dado por

$$\hat{Y} = 48.23 - 0.36x_1 - 1.94x_2 + 2.24x_3 + 0.27x_4 + 0.25x_1x_2 - 1.60x_1x_3 - 0.09x_1x_4 + 0.17x_2x_3 - 0.20x_2x_4 + 0.91x_3x_4.$$

donde las x_i son los factores de control en el experimento en la escala codificada (-1=nivel bajo,+1=nivel alto). Con este modelo se obtienen el vector de valores ajustados y de residuos, que se muestran en las columnas 2, 3, 5 y 6 de la Tabla 4.3, respectivamente. Vemos en esta tabla que la observación $cd = 59.15$, considerada por simple inspección la más sospechosa de ser anómala, tiene apenas el segundo residual más grande; el más grande corresponde a la observación $bcd = 47.02$, que no parece ser extrema. Sin embargo, al calcular la distancia euclidiana entre las observaciones (ver Tabla 4.4) considerando las dos variables (\hat{Y}_i y e_i) se encuentra que la observación 13 (cd) es la que tiene las mayores distancias al resto de los datos, lo que la pone otra vez como la más sospechosa de ser dato anómalo.

Tabla 4.3. Datos observados, estimados y residuales

<i>Datos</i> (Y_i)	\hat{Y}_i	e_i	<i>Datos</i> (Y_i)	\hat{Y}_i	e_i
(1) = 47.46	47.46	0.00	$d = 46.76$	46.76	0.00
$a = 49.62$	49.62	0.00	$ac = 48.56$	48.56	0.00
$b = 43.13$	43.13	0.00	$bd = 44.83$	41.63	3.20
$ab = 46.31$	46.31	0.00	$abd = 44.45$	44.45	0.00
$c = 51.47$	52.99	-1.52	$cd = 59.15$	55.95	3.20
$ac = 48.49$	48.73	-0.24	$acd = 51.33$	51.33	0.00
$bc = 49.34$	49.34	0.00	$bcd = 47.02$	51.5	-4.48
$abc = 46.10$	46.10	0.00	$abcd = 47.90$	47.9	0.00

Tabla 4.4. Matriz de distancias considerando valores estimados y residuales

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	2.16	4.33	1.15	5.73	1.29	1.88	1.36	0.70	1.10	6.65	3.01	9.07	3.87	6.03	0.44
2	0.00	6.49	3.31	3.70	0.92	0.28	3.52	2.86	1.06	8.61	5.17	7.09	1.71	4.86	1.72
3		0.00	3.18	9.97	5.60	6.21	2.97	3.63	5.43	3.53	1.32	13.21	8.20	9.49	4.77
4			0.00	6.85	2.43	3.03	0.21	0.45	2.25	5.67	1.86	10.16	5.02	6.85	1.59
5				0.00	4.45	3.95	7.05	6.41	4.68	12.30	8.67	5.57	2.25	3.31	5.31
6					0.00	0.65	2.64	1.98	0.29	7.89	4.28	8.00	2.61	5.06	0.86
7						0.00	3.24	2.58	0.78	8.35	4.89	7.34	1.99	4.97	1.44
8							0.00	0.66	2.46	5.49	1.65	10.35	5.23	7.01	1.80
9								0.00	1.80	6.04	2.31	9.73	4.57	6.52	1.14
10									0.00	7.63	4.11	8.05	2.77	5.36	0.66
11										0.00	4.26	14.32	10.21	12.50	7.04
12											0.00	11.93	6.88	8.35	3.45
13												0.00	5.62	8.87	8.66
14													0.00	4.48	3.43
15														0.00	5.75
16															0.00

Con base en la matriz de distancias euclidianas entre los puntos (\hat{Y}_i, e_i) se procede a realizar análisis de cúmulos utilizando el método de encadenamiento simple, que consiste en definir la distancia entre dos grupos como la distancia entre sus miembros más cercanos (Kaufman y Rousseeuw, 1993). Así, en un primer paso se agrupan los datos 8 y 4 a una distancia de 0.21 y en el último paso se agrupa la observación 13 a una distancia de 5.57, que es su distancia al dato 5 (ver dendograma de la Figura 4.2 y Tabla 4.5).

Con las distancias de agrupación dadas en la Tabla 4.5 se procede a determinar la posible presencia de datos anómalos: para ello lo primero es aclarar que en la columna "*Se agrupan*" los signos negativos indican un dato que se agrupa por primera vez, mientras que los valores positivos marcan un grupo previo formado hasta ese paso. Por ejemplo, en el paso 15 de la Tabla 4.5 se agrupa por primera vez la observación 13 al grupo conformado hasta el paso 14; y en el paso 14 se había unido por primera vez la observación 11 al grupo conformado hasta el paso 13, y así sucesivamente. De aquí que las dos observaciones más raras son la 13 y la 11, pero falta ver que sean anómalas. Con este propósito se considera el vector de distancias hasta el último paso

donde se haya agrupado por primera vez una observación, o las dos observaciones más raras al resto de los datos, que en este ejemplo es el paso 15.

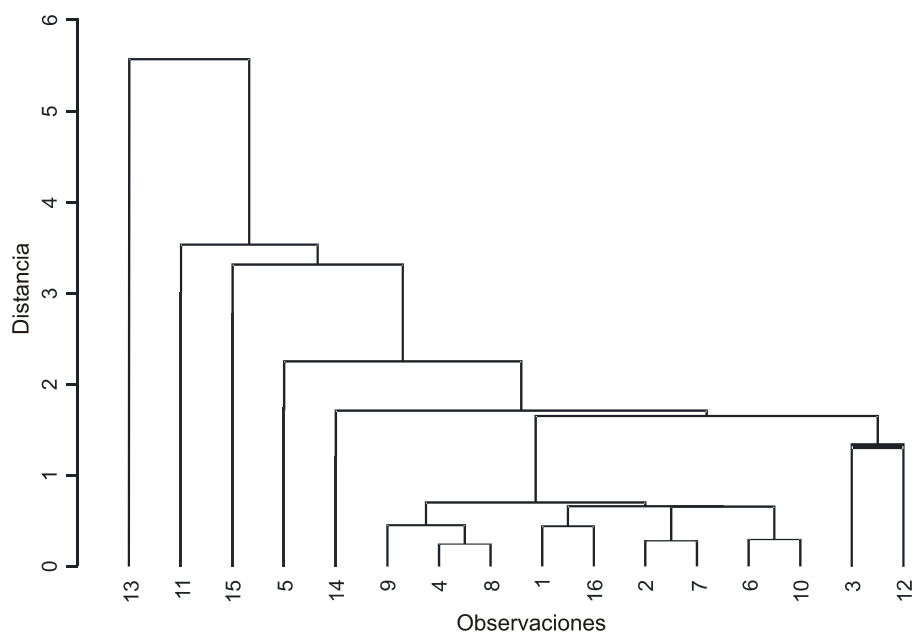


Figura 4.2: Dendrograma, método de encadenamiento simple

Tabla 4.5. Pasos de agrupación

<i>Paso</i>	<i>Se agrupan</i>	<i>Distancia</i>	<i>Paso</i>	<i>Se agrupan</i>	<i>Distancia</i>
1	-4, -8	0.21	9	3, -12	1.32
2	-2, -7	0.28	10	8, 9	1.65
3	-6, -10	0.29	11	-14, 10	1.71
4	-1, -16	0.44	12	-5, 11	2.25
5	-9, 1	0.45	13	-15, 12	3.31
6	2, 3	0.65	14	-11, 13	3.53
7	4, 6	0.66	15	-13, 14	5.57
8	5, 7	0.70	—	—	—

Sea el vector *DISTAN* con las 15 distancias de agrupación de la Tabla 4.5; sean Q_{1D} y Q_{3D} el primero y tercer cuartil de las distancias y sea $RI = Q_{3D} - Q_{1D}$ el rango intercuartílico. La distancia mínima significativa para declarar anómala una observación se define como

$$DMS = Q_{3D} + 2.2RI,$$

donde el valor 2.2 se determinó por simulación. En el ejemplo, se tiene que $Q_{1D} = 0.445$, $Q_{3D} = 1.98$ y $RI = 1.535$, por lo que la distancia mínima significativa resulta ser

$$DMS = 1.98 + 2.2 \times (1.98 - 0.445) = 5.36,$$

que lleva a considerar anómala la observación 13 puesto que su distancia de agrupación es mayor a este límite ($5.57 > DMS$), mientras que la observación 11 no lo es, dado que su distancia de agrupación es menor que DMS . Finalmente la observación 13 se declara faltante y se reestima a partir del contraste correspondiente a la interacción de mayor orden. Para ello se iguala a cero el $Contraste_{ABCD(-13)}$ sin la observación 13 y se despeja el valor de dicho dato, y se obtiene que $Y_{13} = 46.99$. Sustituyendo este valor en el vector de datos originales, se sigue luego con el procedimiento de Benski en la variante de AT&PT (2001) para detectar efectos activos, mismo que no detecta nada en este caso. El análisis completo se puede ver en Capítulo 6.

4.3.1 Cuando hay dos datos anómalos

Cuando se detectan dos datos anómalos, queda el problema de reestimarlos suponiendo nulos el contraste de mayor orden ($ABCD$) y una de las interacciones triples (ABC , ABD , ACD , BCD). Y lo primero es seleccionar una interacción triple adecuada, de manera que el sistema de ecuaciones resultante tenga solución. Un procedimiento para seleccionar dicha interacción es el siguiente: usando multiplicación módulo 2 se obtiene el producto de las dos observaciones anómalas expresadas en notación de Yates. Si el resultado es una letra, se multiplica (módulo 2) ésta por la interacción de mayor orden para obtener la interacción triple deseada. Si el resultado son dos letras, nos deshacemos de una de ellas, sistemáticamente de la segunda en orden alfabético, y la letra que queda se multiplica por la interacción cuádruple para obtener una interacción triple apropiada. Si el resultado son tres letras, se eliminan la segunda y la tercera, y la restante se multiplica por $ABCD$, y resulta la interacción triple ade-

cuada. Finalmente, si el resultado son cuatro letras, cualquiera de las interacciones triples funcionaría, pero siendo consistentes se utiliza $A \times ABCD = BCD$.

Por ejemplo, consideremos el caso donde las observaciones atípicas caen en c y cd . El producto de éstas es $c \times cd = d$. Se multiplica ahora por $ABCD$, y se obtiene $D \times ABCD = ABC$ como la interacción triple apropiada. Entonces para reestimar los dos datos anómalos se usan los contrastes $ABCD$ y ABC . Sean $ABCD_{(-c,-cd)}$ y $ABC_{(-c,-cd)}$ los valores de los contrastes correspondientes sin considerar los datos c y cd . Con esta información se plantea el sistema de ecuaciones

$$\begin{aligned} Y_c - Y_{cd} &= ABCD_{(-c,-cd)} \\ -Y_c - Y_{cd} &= ABC_{(-c,-cd)}, \end{aligned}$$

cuya solución son los valores reestimados para los dos datos anómalos, esto es,

$$\begin{aligned} \hat{Y}_{cd} &= -(ABCD_{(-c,-cd)} + ABC_{(-c,-cd)})/2 \\ \hat{Y}_c &= (ABCD_{(-c,-cd)} - ABC_{(-c,-cd)})/2. \end{aligned}$$

Para ilustrar el procedimiento con dos observaciones anómalas se analizan otra vez los datos de la Tabla 2.2, que tiene en la observación 13 (cd) un dato anómalo de acuerdo al análisis de la sección previa. Para completar dos datos candidatos a ser anómalos se modifica la observación $c = 51.47$ a $c = 61.47$, como se hizo en la Sección 2.1 para mostrar el impacto de dos datos anómalos en el gráfico de probabilidad normal de los efectos. Las distancias a las que se van agrupando los datos se muestran en la Tabla (4.6). El dendograma resultante se muestra en la Figura 4.3. En él se aprecia que las observaciones 5 y 13 (c y cd) se agrupan al último al resto de los datos a una distancia de 5.62, lo que indica que son observaciones anómalas puesto que la distancia crítica es $DMS = 4.86$. Los contrastes $ABC_{(-c,-cd)}$ y $ABCD_{(-c,-cd)}$ considerando las dos observaciones anómalas perdidas valen

$$ABC_{(-c,-cd)} = -101.02; \quad ABCD_{(-c,-cd)} = 4.48,$$

y el sistema de ecuaciones queda como

$$\begin{aligned} Y_c - Y_{cd} &= 4.48 \\ -Y_c - Y_{cd} &= -101.02, \end{aligned}$$

cuya solución es

$$\begin{aligned} \hat{Y}_{cd} &= -(4.48 - 101.02)/2 = 48.27 \\ \hat{Y}_c &= (4.48 + 101.02)/2 = 52.75. \end{aligned}$$

Con estos valores se sustituyen las observaciones originales correspondientes antes de proseguir con la prueba de normalidad del método de Benski. Aunque estrictamente no se alcanza a rechazar la prueba de normalidad, el *p* – *valor* que resulta es incluso menor en este caso que el obtenido cuando se corrigió un solo dato anómalo, detectado en la observación 13 (*cd*). Los gráficos de efectos en papel de probabilidad normal considerando uno o dos datos anómalos se muestran en la Figura 4.4.

Tabla 4.6. Pasos de agrupación, caso con dos datos anómalos

<i>Paso</i>	<i>Se agrupan</i>	<i>Distancia</i>	<i>Paso</i>	<i>Se agrupan</i>	<i>Distancia</i>
1	-4, -8	0.21	9	-11, 7	1.32
2	-2, -7	0.28	10	-3, 9	1.65
3	-1, -16	0.29	11	-14, 10	1.71
4	-9, 1	0.44	12	-5, -13	2.25
5	-10, 3	0.45	13	-12, 11	3.31
6	4, 5	0.65	14	8, 13	3.53
7	2, 6	0.66	15	12, 14	5.57
8	-6, -15	0.70	–	–	–

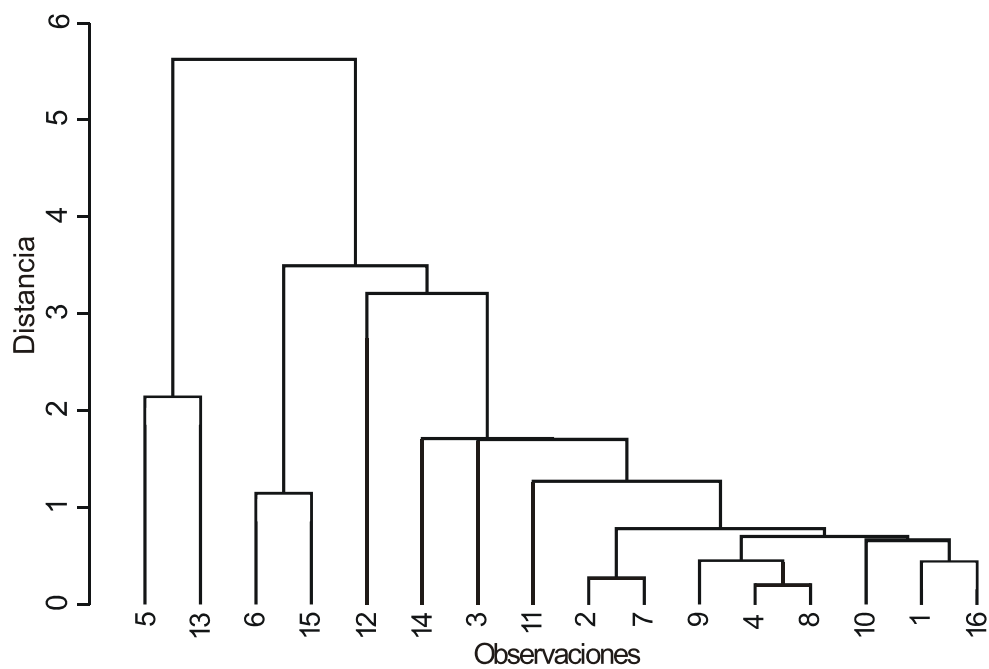


Figura 4.3: Dendrograma, considerando dos datos anómalos

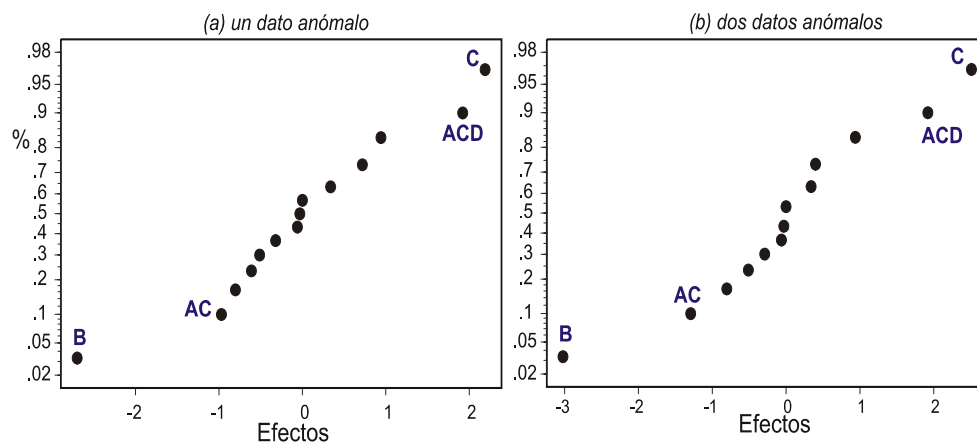


Figura 4.4: Efectos en papel de probabilidad normal, considerando uno y dos datos anómalos

Capítulo 5

Métodos de Detección de Efectos Activos Basados en Técnicas de Regresión Robusta

En este capítulo se presentan un par de métodos nuevos para detectar efectos activos basados ambos en técnicas de regresión robusta. Los distinguimos como variantes I y II (Secciones 5.3 y 5.4, respectivamente). La diferencia entre ellos es básicamente la manera en que se selecciona un modelo inicial robusto a partir del cual, aplicando una técnica de regresión también robusta, se obtienen de manera iterativa estimadores de todos los efectos o coeficientes.

5.1 Estimadores de Regresión Robusta

Es deseable que un estimador de regresión robusta tenga alto punto de rompimiento y buenas propiedades asintóticas. Se llama punto de rompimiento la fracción mínima de puntos contaminados que requiere la muestra para que el sesgo del estimador pueda ser infinito. El máximo punto de rompimiento que puede llegar a tener un estimador es $1/2$, en cuyo caso puede soportar una contaminación por datos anómalos

hasta el 50% de los datos (ver Rousseeuw y Leroy, 1987). En cuanto a las propiedades asintóticas se requiere que el estimador sea consistente y eficiente, es decir, que cuando $n \rightarrow \infty$, el estimador $\hat{\theta}$ converge en probabilidad a θ y que la varianza (asintótica) del estimador sea comparable a la del estimador de mínimos cuadrados cuando los errores son normales.

Como se sabe, el estimador ordinario de mínimos cuadrados resulta de minimizar la suma del cuadrado de los residuos dada por

$$\sum_{i=1}^n r_i^2,$$

donde $r_i = Y_i - \hat{Y}_i$. Esto es, se minimiza la función $\rho(r) = r^2$ de los residuos. Este estimador es el más eficiente cuando la distribución de los residuos es normal, pero no es robusto. Su punto de rompimiento es $1/n$, lo que significa que un solo dato anómalo puede tener un gran efecto en el estimador (Rousseeuw y Leroy, 1987).

Dada la falta de robustez del estimador de mínimos cuadrados ordinarios, se han propuesto muchas alternativas de funciones de los residuos a minimizar buscando que sean insensibles a observaciones atípicas. Es así como Huber (1973) propone el uso de *M-estimadores* que se basan precisamente en utilizar otras funciones de los residuos, de manera que los estimadores resultan de minimizar

$$\sum_{i=1}^n \rho(r_i)$$

donde ρ es una función simétrica ($\rho(t) = \rho(-t)$ para todo t) con un mínimo único en cero. Note que si $\rho(r; \theta) = -\log f(r; \theta)$ se tiene el estimador de máxima verosimilitud, de allí el nombre de M-estimador. Derivando esta función respecto a los coeficientes de regresión ($\hat{\beta}_i$'s) se obtiene el sistema de ecuaciones

$$\sum_{i=1}^n \psi(r_i) \mathbf{x}_i = \mathbf{0}$$

cuya solución es un estimador de regresión. Sin embargo, esta solución no es “equivariante” con respecto a cambios de escala en la respuesta “Y”, por lo que los residuos se

estandarizan dividiendo por algún estimador de σ que se estima simultáneamente al resolver el sistema de ecuaciones

$$\sum_{i=1}^n \psi(r_i/\hat{\sigma})\mathbf{x}_i = \mathbf{0}.$$

Desgraciadamente los M-estimadores siguen teniendo el problema de punto de rompimiento $1/n$ para datos muy alejados en \mathbf{x}_i . De aquí que se propusieran generalizaciones de estos estimadores (estimadores GM), que buscan limitar la influencia de los puntos \mathbf{x}_i mediante ponderaciones. Pero el punto de rompimiento de estas generalizaciones nunca pasó de cierto valor (35%) y tiene la indeseable propiedad de que decrece en función de p (el número de coeficientes de regresión, ver Rousseeuw y Leroy, 1987).

Posteriormente en la década de los ochentas aparecieron varios estimadores de regresión con alto punto de rompimiento. Uno de ellos, relacionado con el M-estimador, es el *S-estimador* propuesto por Rosseeauw y Yohai (1984) el cual minimiza la dispersión de los residuos

$$\min_{\theta} S(r_1(\theta), r_2(\theta), \dots, r_n(\theta))$$

definida ésta como la solución de

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{r_i}{S}\right) = b, \quad (5.1)$$

donde b se toma como $E_{\phi}(\rho)$ donde ϕ es la normal estándar. La función ρ satisface las siguientes condiciones:

C1.- Es simétrica, continua y $\rho(0) = 0$.

C2.- Existe una constante $c > 0$ tal que ρ es estrictamente creciente en el intervalo $[0, c]$ y constante sobre $[c, \infty)$. Esto implica que $\rho' = \psi$ es igual a cero desde un cierto valor en adelante. A las funciones ψ con esta propiedad se les llama “re-descendentes”.

Una tercera condición,

C3.- $\frac{b}{\rho(c)} = \frac{1}{2}$, es necesaria para probar el punto de rompimiento de 50% que tiene este estimador (ver lema 4 y teorema 8, pág. 136-139 de Rousseeuw y Leroy, 1987).

Por ejemplo para la función ρ definida por

$$\rho_c(u) = \begin{cases} \frac{u^6}{6c^4} - \frac{u^4}{2c^2} + \frac{u^2}{2} & \text{si } |u| < c \\ c^2/6 & \text{si } |u| \geq c, \end{cases}$$

cuya derivada es la función doble cuadrada re-descendente dada por

$$\psi_c(u) = \begin{cases} u\left(\left(\frac{u}{c}\right)^2 - 1\right)^2 & \text{si } |u| < c \\ 0 & \text{si } |u| \geq c. \end{cases}$$

Definiendo $b = E_\phi(\rho)$ la condición C3 se cumple tomando $c = 1.547$. Específicamente, para este valor de la constante c se tiene que $b = 0.1995$. Asimismo, por el teorema 8 de la página 139 de Rosseeuw y Leroy (1987), el estimador resultante con esta función tiene un punto de rompimiento de 50% y una eficiencia de 28.7% (ver Tabla 19 de la página 142 de Rosseeuw y Leroy, 1987). En dicha tabla se observa que el incrementar la eficiencia del S-estimador lleva a una fuerte reducción de su punto de rompimiento. Por ejemplo, para una eficiencia de 96.6% se tendría un punto de rompimiento de 10%, con constantes $c = 5.182$ y $b = 0.4475$.

Buscando desarrollar estimadores con alto punto de rompimiento y alta eficiencia, Yohai (1987) propone los *MM-estimadores*, que se definen en las siguientes tres etapas:

1. Se toma un estimador inicial θ_0 con alto punto de rompimiento, posiblemente de 0.5.
2. Se obtienen los residuos r_i ($i = 1, 2, \dots, n$) y se calcula el estimador de escala $s_n = s(r(\theta_0))$ definido por

$$(1/n) \sum_{i=1}^n \rho_0(r_i/s) = b$$

usando una función ρ_0 que satisface las siguientes condiciones: es simétrica, continua, no decreciente en $|r|$ y $\rho_0(0) = 0$; En esta relación b es una constante tal que $b/a = 0.5$, donde $a = \max \rho_0(r)$. Esta relación entre b y a implica que el estimador de escala tenga un punto de rompimiento de 0.5 (por condición C3 de los S-estimadores).

3. Se toma otra función ρ_1 satisfaciendo las mismas condiciones y tal que

$$\begin{aligned}\rho_1(r) &\leq \rho_0(r), \\ \sup \rho_1(r) &= \sup \rho_0(r) = a.\end{aligned}$$

Sea $\psi_1 = \rho_1'$. Entonces el MM-estimador θ_1 se define como cualquier solución de

$$\sum_{i=1}^n \psi_1(r_i(\theta)/s_n) \mathbf{x}_i = 0$$

la cual verifica

$$S(\theta_1) \leq S(\theta_0),$$

donde

$$S(\theta) = \sum_{i=1}^n \rho_1(r_i(\theta)/s_n),$$

donde $\rho_1(0/0)$ se define como 0. Yohai (1987) establece la consistencia (teorema 3.2) y la normalidad asintótica (teorema 4.1) de los MM-estimadores suponiendo que las $x'_i s$ son aleatorias. Una manera de escoger a ρ_0 y ρ_1 de manera que satisfagan las condiciones mencionadas es como sigue: sea ρ una función que satisface las condiciones C1 a C3 y sea $0 < c_0 < c_1$. Defina $\rho_0(r) = \rho(r/c_0)$ y $\rho_1(r) = \rho(r/c_1)$. El valor de c_0 se escoge de manera que $b/a = 0.5$ y el valor de c_1 determina la eficiencia asintótica del estimador. Esto último se desprende del teorema 4.1 en el que se observa que la varianza asintótica de los MM-estimadores depende solo de la constante c_1 que define la función ρ_1 . Otro hecho interesante es que la eficiencia asintótica es independiente de la distribución de las $x'_i s$.

Ahora sí estamos en posibilidades de describir el procedimiento de MM-estimación que realiza la función `lmRobMM` de `SPLUS`, que en palabras del propio Yohai *et al.* (1991), no es más que un MM-estimador inicializado por un S-estimador.

5.2 Procedimiento ImRobMM de SPLUS

Este procedimiento permite realizar MM-estimación (Yohai *et al.*, 1991) usando la función doble cuadrada definida como

$$\psi_c(u) = \begin{cases} \frac{u}{c} \left[\left(\frac{u}{c} \right)^2 - 1 \right]^2 & \text{si } |u| \leq c \\ 0 & \text{si } |u| > c, \end{cases}$$

donde c es una constante de afinación, que permite lograr el punto de rompimiento y la eficiencia deseada. La integral de esta función dividida por $c/6$ es

$$\rho_c(u) = \begin{cases} \left(\frac{u}{c} \right)^6 - 3 \left(\frac{u}{c} \right)^4 + 3 \left(\frac{u}{c} \right)^2 & \text{si } |u| < c \\ 1 & \text{si } |u| \geq c. \end{cases}$$

Para cualquier θ , sea $s_c(\theta)$ la solución de

$$\frac{1}{n} \rho_c \left(\frac{y_i - \theta^t x_i}{s} \right) = 0.5. \quad (5.2)$$

donde s es un estimador de escala que se obtiene de la siguiente manera. Las etapas para obtener el MM-estimador usando esta función ρ son las siguientes:

1. Yohai *et al.* (1991) propone obtener el estimador inicial $\tilde{\theta}_0$ como sigue: se resuelve esta última igualdad en el conjunto de los estimadores θ_i 's que resultan de ajustar subconjuntos de p puntos de la matriz X para los cuales se tiene un ajuste de mínimos cuadrados perfecto. Esto es, se calculan los $\binom{n}{p}$ estimadores θ_i 's con los cuales se resuelve la igualdad anterior usando $c_0 = 1.548$, y se identifica el estimador $\tilde{\theta}_0$ para el cual s es mínimo. Este valor de la constante c_0 asegura un punto de rompimiento de 50% del estimador de escala resultante. Por ejemplo, al estimar modelos con 7 coeficientes de regresión en el factorial 2^4 , se tienen $\binom{n=16}{p=7} = 11440$ subconjuntos de 7 renglones. Se implementa un muestreo tal que la probabilidad de que algún subconjunto contenga solo puntos “buenos” sea alta.

2. Tomando como valor inicial este estimador, se vuelve a resolver la ecuación (5.2) para obtener el mínimo local que corresponde al estimador $\hat{\theta}_0$.

3. Finalmente, después de sustituir el estimador $\hat{\theta}_0$, se calcula el estimador robusto $\hat{\theta}_1$ resolviendo otra vez la igualdad anterior usando ahora la constante $c_1 = 4.687$. Este valor de la constante corresponde a una eficiencia del 95% bajo el modelo normal. Para una eficiencia de 99.3% el valor de la constante es $c_1 = 7.695$.

5.3 Método Basado en Técnicas de Regresión Robusta, Variante I

Al ser los factoriales no replicados experimentos con pocas observaciones que pueden contener datos anómalos no se puede hacer MM-estimación (aplicar la función `lmRobMM`) directamente al modelo saturado, sino que es necesario estimar los efectos de forma iterativa, previamente fijando de manera robusta una parte del modelo. Específicamente, el método que proponemos basado en técnicas de regresión robusta (variante I) consiste de tres pasos, a saber:

1a. Primero se obtiene un estimador inicial robusto de los efectos principales e interacciones dobles aplicando mínimos cuadrados ordinarios a subconjuntos aleatorios de 13 renglones a la matriz \mathbf{X} (ver Subsección 5.3.1). Este muestreo de subconjuntos de 13 renglones de la matriz \mathbf{X} se repite 40 veces y se preserva aquel subconjunto en el cual la suma de cuadrados del error (SC_E) es mínima. Por ejemplo, en el factorial 2^4 no replicado se tienen 16 observaciones, 4 efectos principales, 6 interacciones dobles y 5 interacciones de alto orden. Se estiman con mínimos cuadrados ordinarios los 4 efectos principales y 6 interacciones dobles usando cada vez subconjuntos de 13 renglones aleatoriamente seleccionados de la matriz \mathbf{X} , y se preserva la estimación que minimiza la suma de cuadrados del error.

1b. En el caso de un experimento factorial fraccionado se hace muestreo tanto de columnas como de renglones: en cada paso se seleccionan al azar 8 columnas de la matriz \mathbf{X} de las que corresponden a contrastes de efectos, y se muestrean 40 conjuntos

de 13 renglones de la submatriz resultante. A cada conjunto de 13 renglones y 8 columnas se les aplica mínimos cuadrados ordinarios, guardando cada vez el modelo que resulta con menor suma de cuadrados del error. El muestreo de columnas se repite también 40 veces, es decir, se ajustan 1600 modelos. Con los 4 ó 5 términos mayores del modelo ganador se sigue con el paso 2.

La idea del muestreo anterior es tratar de evitar los renglones que corresponden a observaciones atípicas, eliminando de esta manera su impacto en los estimadores iniciales de esos 10 efectos. En otras palabras, estos estimadores iniciales de los efectos principales e interacciones dobles son en alguna medida robustos a observaciones atípicas, y por el momento no se tienen estimadores de las interacciones de alto orden. El muestreo de renglones se repite sólo 40 veces por razones de costo computacional, pero en realidad sigue siendo alta la probabilidad de no detectar el subconjunto de 13 renglones que no tiene datos anómalos: son $\binom{16}{13} = 560$ maneras de seleccionar 13 renglones de 16, de manera que la probabilidad de no seleccionar en cada intento los 13 renglones correctos, suponiendo tres datos anómalos, es $559/560$. Luego, la probabilidad de no seleccionar los 13 renglones "buenos" en los 40 intentos independientes es $(559/560)^{40} = 0.93$. Pero la probabilidad de no seleccionar en los 40 intentos al menos 12 de los 13 renglones buenos es 0.05, de manera que el muestreo si logra alguna protección con alta probabilidad respecto a dos de los tres datos anómalos.

2. Con los 5 o 4 efectos más grandes obtenidos en el paso anterior se decide la parte fija del modelo a utilizar con el procedimiento lmRobMM y los términos faltantes (10 términos) se obtienen iterativamente utilizando las 16 observaciones del experimento. Debe utilizarse la función de pesos doble cuadrada y una eficiencia de 99.3% en el procedimiento lmRobMM, para llegar a resultados interesantes. Por ejemplo, para el factorial 2^4 se fija el modelo correspondiente a los 5 términos mayores del paso anterior y un sexto término se va estimando iterativamente hasta completar los 15 efectos posibles.

3. Se aplica el método de Benski (1989) a los efectos robustos obtenidos en el paso

anterior en la misma forma en que lo aplica Aguirre-Torres y Pérez-Trejo (2001) a los efectos estimados con la transformación de rangos.

5.3.1 Sobre el Primer Paso del Método Robusto I

El primer paso del método robusto variante I, consistente en muestrear 40 veces 13 renglones de la matriz X , no es más que un mecanismo para decidir el modelo a fijar en el procedimiento de estimación del segundo paso. Se puede verificar por simulación que fijar el modelo con base en los estimadores de los efectos más grandes que arroja el método de mínimos cuadrados ordinarios sobre la matriz \mathbf{X} completa no funciona, puesto que éstos están afectados por la contaminación presente; y a la postre este hecho afecta sensiblemente el procedimiento de estimación robusta de Yohai *et al.* (1991) que aplica la función `lmRobMM` de `SPLUS`, al utilizarla de manera iterativa para estimar los efectos faltantes.

En las técnicas de regresión robusta se suele suponer que n es mucho más grande que p , cosa que no ocurre en los factoriales no-replicados. Por ejemplo, en el factorial 2^4 sin réplicas se tienen $n = 16$ observaciones y son $p = 15$ efectos por estimar. Puesto que no es posible utilizar la función `lmRobMM` para estimar de una sola vez todos los efectos de interés, es necesaria la estimación iterativa de los efectos restantes a partir de un modelo base robustamente elegido. Si el modelo base se viera afectado por los datos atípicos, como ocurre con los estimadores de mínimos cuadrados usuales, se afecta el procedimiento `lmRobMM`, dado que se basa en una función de pesos de los residuos obtenidos de dicho modelo; residuos que resultan sesgados debido al modelo sesgado utilizado.

Para obtener estimadores robustos desde este primer paso nos fijamos solamente en los efectos potencialmente importantes de acuerdo al principio de jerarquía de efectos (Wu y Hamada, 2000, pág. 112), que son los efectos principales y las interacciones dobles. Así por ejemplo, en esta primera estimación robusta no se estiman para el factorial 2^4 las cuatro interacciones triples y la interacción cuádruple. Estos 5 grados

de libertad se utilizan para poder aplicar el método de mínimos cuadrados a un subconjunto de renglones de la matriz \mathbf{X} conformada por una primera columna de 1's más 10 columnas de los efectos principales e interacciones dobles. Posiblemente lo ideal sería aplicar el método de mínimos cuadrados a todas las posibles combinaciones de 13 renglones de los 16 que tiene la matriz \mathbf{X} , que son un total de $\binom{16}{13} = 560$, para seleccionar aquella donde el efecto de los datos atípicos sea mínimo, medido este con la suma de cuadrados del error. La combinación de renglones seleccionada tiende a ser aquella que no incluye a los datos atípicos presentes en el experimento.

El problema es que resulta computacionalmente caro tanto la selección de todas las posibilidades como el ajustar mínimos cuadrados tantas veces. De aquí que se proponga probar una muestra aleatoria de tales combinaciones. Se observó que con un muestreo de tamaño 40 o más se logra ubicar una de tales combinaciones con menos impacto de los puntos aberrantes. Con los más grandes estimadores de mínimos cuadrados basados en esta submatriz \mathbf{X} se fija el modelo a ser utilizado en el paso 2. La razón de utilizar 13 renglones de la matriz \mathbf{X} descrita arriba es porque con 12 renglones comienzan a aparecer combinaciones de renglones en las cuales la matriz $X'X$ no es invertible. Con 13 renglones nunca ocurre que las columnas de la matriz resultante sean linealmente dependientes.

Consideremos la selección de 12 renglones y 11 columnas de la matriz \mathbf{X} de la página siguiente. Se puede ver que la columna C es combinación lineal de otras columnas. Específicamente ocurre que $C = D + I - A$.

Entonces no se puede trabajar con 12 o menos renglones porque comienzan a suceder dependencias como la que acabamos de mostrar, que impiden el cálculo de la inversa de $X'X$ y botan el programa, mientras que con 13 renglones la matriz $X'X$ es siempre no singular, y además 13 renglones permiten una protección hasta de 3 puntos aberrantes.

renglón	A	B	C	D	E	F	G	H	I	J	K
16	1	1	1	1	1	1	1	1	1	1	1
9	1	-1	-1	-1	1	1	1	-1	1	-1	-1
15	1	-1	1	1	1	-1	-1	-1	1	1	1
1	1	-1	-1	-1	-1	1	1	1	1	1	1
2	1	1	-1	-1	-1	-1	-1	-1	1	1	1
8	1	1	1	1	-1	1	1	-1	1	-1	-1
7	1	-1	1	1	-1	-1	-1	1	1	-1	-1
10	1	1	-1	-1	1	-1	-1	1	1	-1	-1
14	1	1	-1	1	1	-1	1	1	-1	-1	1
13	1	-1	-1	1	1	1	-1	-1	-1	-1	1
5	1	-1	-1	1	-1	1	-1	1	-1	1	-1
6	1	1	-1	1	-1	-1	1	-1	-1	1	-1

5.4 Método Basado en Técnicas de Regresión Robusta, Variante II

El paso 1 de la variante 1 presentada en las secciones previas consiste en muestrear 40 veces 13 renglones de la matriz \mathbf{X} buscando detectar aquel subconjunto que no contenga datos anómalos. Luego, aplicando mínimos cuadrados al subconjunto seleccionado se eligen los 4 o 5 términos a fijar en el paso 2. Una desventaja clara de este procedimiento es el costo computacional, que debería de incrementarse si se quiere un mejor trabajo de selección del subconjunto de 13 renglones deseado: suponiendo la existencia de 3 datos anómalos, la probabilidad de seleccionar con este muestreo el conjunto de 13 renglones que corresponden a las observaciones no atípicas es apenas de 0.07, mientras que la probabilidad de seleccionar un conjunto de 13 renglones con al menos 12 sin datos anómalos es de 0.95.

Otra desventaja del método basado en muestreo es que los resultados del análisis pueden variar de una realización a otra, debido a la selección de un conjunto diferente de 13 renglones de la matriz X . Por lo mismo, es probable entonces que un análisis específico no sea el mejor resultado robusto que se pueda tener.

Una alternativa de selección robusta de la parte fija del modelo la provee la regre-

sión L1, que consiste en minimizar la suma de los valores absolutos de los residuos

$$\sum_{i=1}^n |r_i|.$$

La idea surge de que la regresión L1 es robusta a observaciones discrepantes en la respuesta (Y) (Rousseuw y Leroy, 2003, pág. 10); y aunque no es robusta a datos anómalos en las \mathbf{x}'_i s, en el contexto de diseño de experimentos en que nos encontramos, esto último no es un problema. De aquí que se decida utilizar esta regresión en el paso 1 para decidir la parte fija del modelo en el paso 2.

El paso 1 para la variante 2 del método basado en técnicas de regresión robusta queda entonces como sigue: se aplica regresión L1 al modelo de 10 términos conformados por los efectos principales e interacciones dobles. Se ordenan los efectos absolutos resultantes y con los 4 o 5 mayores se decide el modelo a utilizar en el paso 2.

Como se menciona aquí arriba, el utilizar la regresión L1 en lugar del muestreo de renglones en el paso 1 ofrece ventajas tanto de costo computacional como de consistencia de los resultados en cada repetición del análisis. Además de que en los estudios de simulación del Capítulo 7 la variante 2 del método robusto ofrece resultados un poco mejores en lo que a potencia para detectar efectos activos se refiere.

La regresión L1 como criterio para seleccionar la parte fija del modelo que se utilizará en el paso 2 funciona porque ésta es robusta a datos discrepantes en la variable de respuesta.

Del paso 1 se sale con estimadores robustos iniciales de los efectos principales e interacciones dobles, de los cuales se eligen 4 o 5 para la parte fija del modelo a utilizar en la función `lmRobMM` de `SPLUS`. Se utiliza este procedimiento 10 veces hasta obtener estimadores robustos de los efectos no incluidos en la parte fija del modelo, ajustando cada vez un sexto término al modelo inicial. Al estimar el último efecto, que corresponde al sexto mayor del paso 1, se leen también los estimadores

obtenidos para la parte fija del modelo y de esta manera se tienen los 15 efectos estimados robustamente con la función `lmRobMM`.

Capítulo 6

Ejemplos

En este capítulo se presentan tres ejemplos de la literatura que se resuelven usando todos los métodos expuestos en este trabajo, a saber: 1. Método de rangos, 2. Método de rangos modificados, 3. Método de ajuste de espacios entre efectos, 4. Método de datos anómalos reestimados, 5. Método basado en técnicas de regresión robusta, variante II y 6. Método Bayesiano. El primero y el último se han propuesto en la literatura por Aguirre-Torres y Pérez-Trejo (2001) y Box y Meyer (1987), respectivamente. Los cuatro métodos intermedios son nuevas propuestas. No se presentan los resultados del método basado en técnicas robustas variante I porque son muy similares a los de la variante II.

En cuanto a los ejemplos o experimentos que se analizan a continuación conviene mencionar que el Ejemplo 1 consiste de datos normales, sin datos atípicos. El Ejemplo 2 es el conjunto de datos con el que hemos venido ilustrando los métodos y tiene una observación discrepante, y finalmente, el Ejemplo 3 que es un experimento factorial fraccionado 2^{5-1} también contiene un dato atípico, aunque no extremo.

6.1 Ejemplo 1

Este corresponde al ejemplo 1 de AT&PT (2001) y al ejemplo IV de Box y Meyer (1986). En la Tabla 6.1 se muestran los datos observados, los efectos estimados en la escala original, efectos con la transformación de rangos, efectos con rangos modificados, efectos con espacios ajustados, efectos de datos anómalos reestimados y los efectos estimados con técnicas de regresión robusta. En la en la Figura 6.1 se muestran los gráficos de efectos en papel normal para los seis tipos de efectos. Cinco de los gráficos resultan visualmente similares y la excepción es el método de espacios ajustados, en el cual se refleja su alejamiento de la normalidad aun bajo H_0 . Para los seis tipos de efectos el análisis formal confirma la ausencia de efectos activos, según se observa en el valor del estadístico W' y su p-valor observado reportados en la Tabla 6.2. La definición del estadístico W' se encuentra en la ecuación 3.6.

Tabla 6.1. Datos y Efectos Estimados, Ejemplo 1

Nombre	Datos	Efectos (res-orig)	Efectos (rangos)	Efectos (ran.mod)	Efectos (esp.aj)	Efectos (ext.per)	Efectos (reg-rob)
(1)	0.08						
A	0.04	-0.19	-3.62	-3.71	-1.46	-0.19	-0.097
B	0.53	-0.02	0.12	-0.45	-0.22	-0.02	-0.008
AB	0.43	0.00	-0.25	0.08	-0.16	0.00	-0.000
C	0.31	-0.08	-0.75	-1.48	-0.38	-0.08	-0.038
AC	0.09	0.03	0.87	0.74	-0.01	0.03	0.020
BC	0.12	-0.07	-1.37	-1.20	-0.29	-0.07	-0.033
ABC	0.36	0.15	2.00	2.69	1.06	0.14	0.074
D	0.79	0.27	4.62	5.26	2.23	0.27	0.135
AD	0.68	-0.16	-2.75	-3.08	-1.18	-0.16	-0.080
BD	0.73	-0.25	-5.00	-4.84	-2.02	-0.25	-0.125
ABD	0.08	-0.10	-1.62	-1.86	-0.62	-0.10	-0.050
CD	0.77	-0.03	-0.37	-0.60	-0.27	-0.03	-0.013
ACD	0.38	-0.01	0.00	-0.07	-0.21	-0.01	-0.000
BCD	0.49	0.12	2.75	2.44	0.83	0.12	0.061
ABCD	0.23	0.02	0.37	0.30	-0.15	0.02	0.007

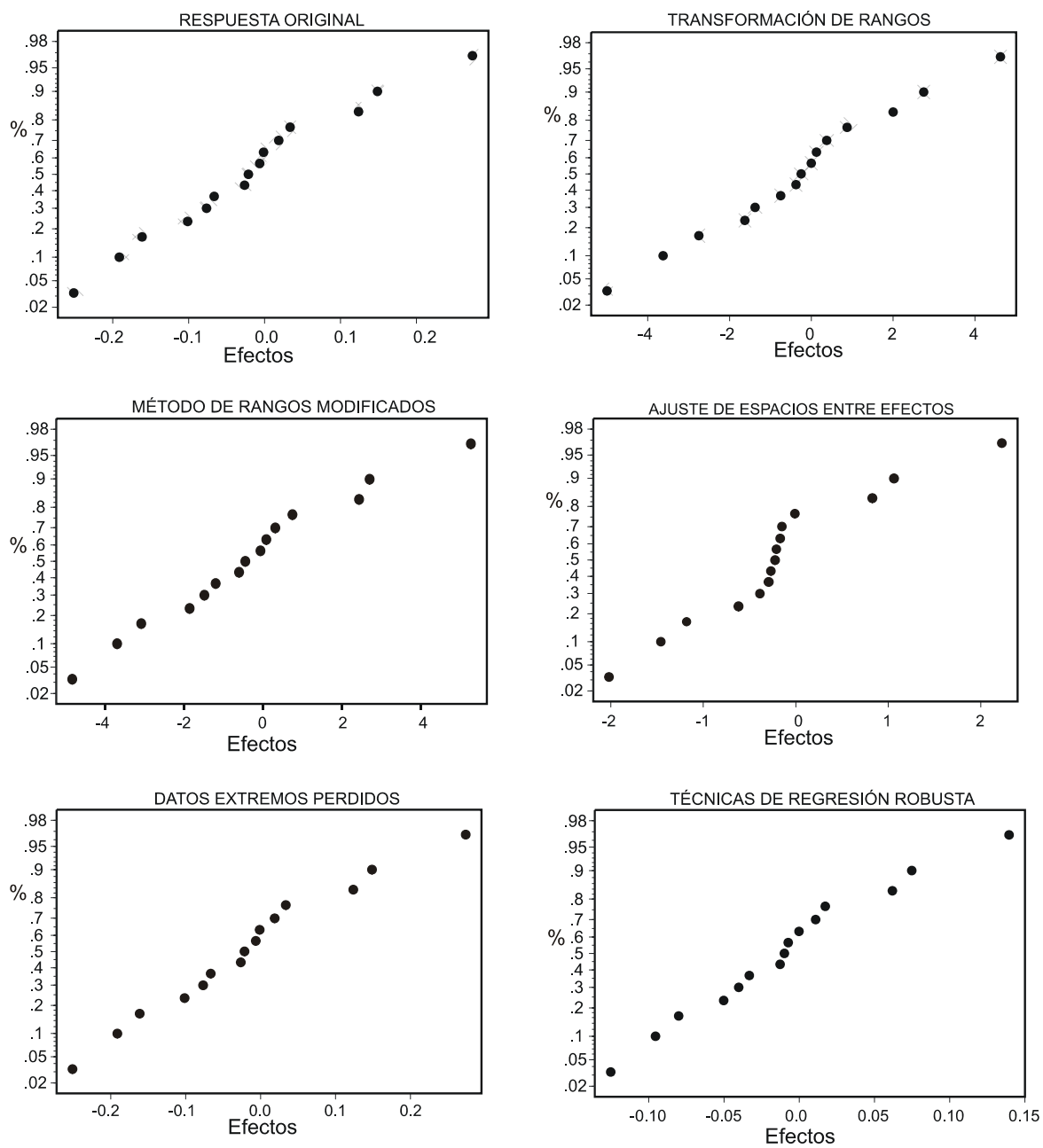


Figura 6.1: Efectos en papel de probabilidad normal, Ejemplo 1

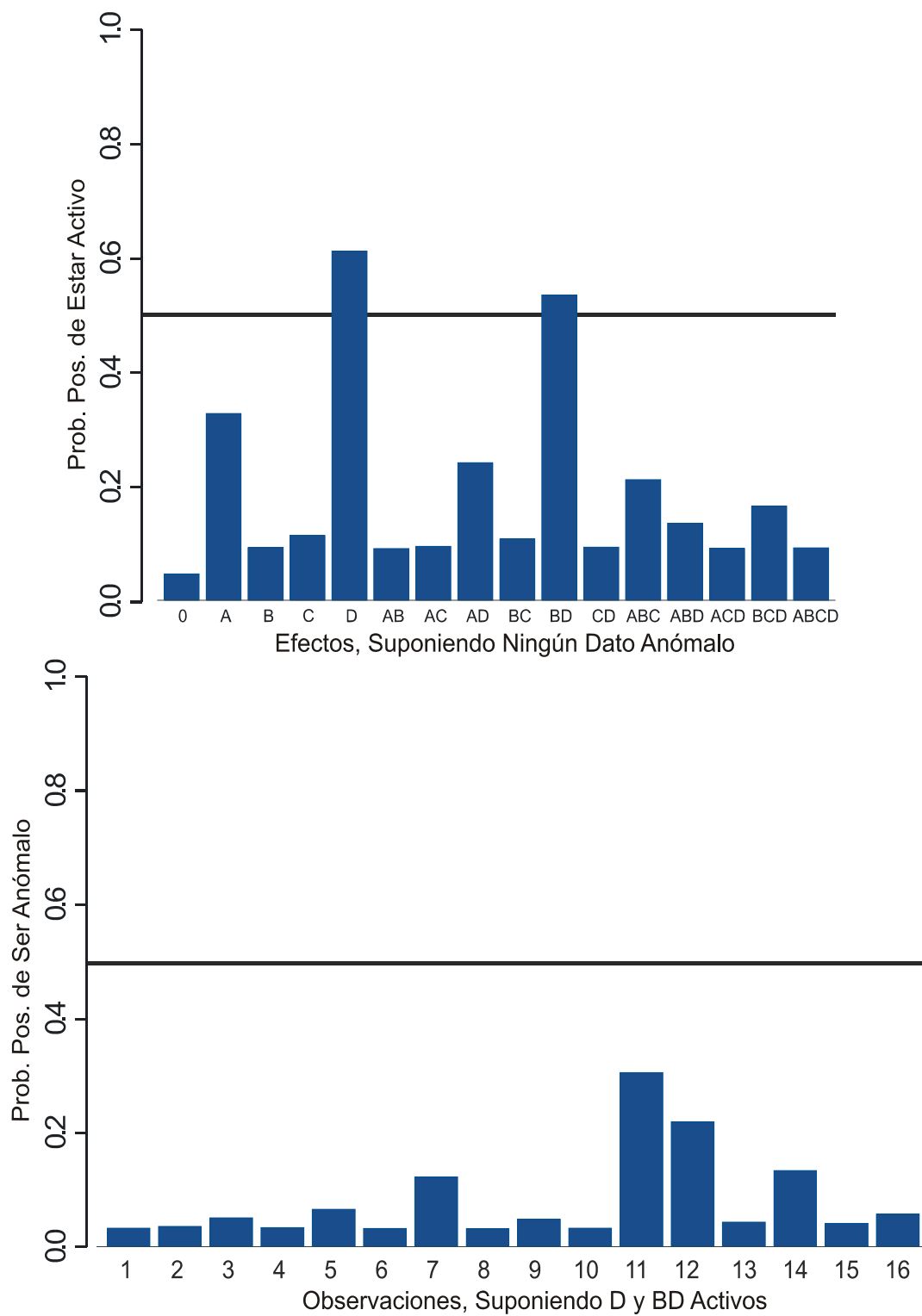


Figura 6.2: Método Bayesiano, Ejemplo 1

Tabla 6.2. Comparación analítica, Ejemplo 1

Método	W'	P-valor	FL	FU	dF	$\pm 2dF$
Respuesta original	0.9687	0.7435	-0.089	0.026	0.115	± 0.23
Rangos original	0.9763	0.9377	-1.500	0.625	2.125	± 4.25
Rangos modificados	0.9697	0.7666	-1.672	0.525	2.197	± 4.39
Ajuste de espacios entre efectos	0.9065	0.1128	-0.502	-0.081	0.421	± 0.84
Estimación de datos extremos	0.9687	0.7435	-0.089	0.026	0.115	± 0.23
Técnicas de regresión robusta	0.9675	0.7183	-0.045	0.014	0.0593	± 0.12

En la Figura 6.2 se muestran las gráficas de las probabilidades posteriores de que los efectos estén activos y de que los datos sean anómalos, que corresponden a las dos primeras iteraciones del método de Box y Meyer (1987). Se observa que los efectos D y BD tienen probabilidades posteriores arriba de $P=0.5$, que indicaría que están activos. Sin embargo, esta probabilidad $P=0.5$ favorece la detección de efectos que en realidad no están activos, de manera que si se utiliza el valor $P=0.89$ reportado por AT&PT (2001) se concluye que ningún efecto está activo y ningún dato es anómalo. Esto último lleva a la misma conclusión que arrojaron los métodos de la Tabla 6.2.

6.2 Ejemplo 2

Este ejemplo corresponde al ejemplo 2 de AT&PT (2001). Los datos, efectos estimados en la escala original, efectos usando la transformación de rangos, efectos con rangos modificados, efectos con espacios ajustados, efectos con datos anómalos reestimados y los efectos estimados con técnicas de regresión robusta se muestran en la Tabla 6.3. Los efectos de mayor magnitud en las diferentes escalas resultan ser B , C y AC . Varias de las técnicas consideradas detectan que el cuarto efecto en magnitud es ACD , mismo que en el análisis de Box y Meyer (1987) tiene la cuarta probabilidad posterior mayor. Las excepciones a esto son el método en la escala original y el método de espacios ajustados, con valores pequeños para ese efecto. En los gráficos de Daniel dados en la Figura 6.3 se observa el impacto de un punto aberrante sobre los efectos en la escala original, que causa en esta escala no se detecte ningún efecto

activo. En casi todas las gráficas se observan separados tres efectos de la línea de normalidad, que corresponden a los efectos B , C y AC .

Tabla 6.3. Datos y Efectos Estimados, Ejemplo 2

Nombre	Datos	Efectos (res.orig)	Efectos (rangos)	Efectos (ran.mod)	Efectos (esp.aj)	Efectos (ext.per)	Efectos (reg.rob)
(1)	47.46						
A	49.62	-0.80	0.0	-0.01	-0.051	0.72	-0.358
B	43.13	-4.22	-6.25	-5.98	-2.39	-2.70	-1.974
AB	46.31	0.91	-0.75	-0.15	0.13	-0.61	0.356
C	51.47	3.71	4.75	5.04	2.04	2.19	1.810
AC	48.49	-2.49	-3.25	-2.78	-1.20	-0.97	-1.195
BC	49.34	-0.80	0.50	-0.01	-0.05	0.72	0.030
ABC	46.10	1.20	1.00	0.75	0.32	-0.32	0.503
D	46.76	1.01	0.00	0.04	0.19	-0.51	0.453
AD	48.56	-0.58	1.00	0.76	-0.04	0.94	-0.243
BD	44.83	-1.18	-0.25	-0.71	-0.31	0.34	-0.478
ABD	44.45	0.72	0.25	-0.14	-0.00	-0.80	0.227
CD	59.15	1.49	1.25	1.29	0.52	-0.03	0.702
ACD	51.33	0.40	2.25	2.21	-0.02	1.92	0.719
BCD	47.02	-1.58	-1.00	-1.09	-0.58	-0.06	-0.383
ABCD	47.90	1.52	1.50	0.98	0.54	0.00	0.681

En el análisis formal usando $\alpha = 0.045$ (nominal) sólo el método basado en rangos modificados es capaz de detectar los efectos activos B , C y AC (Tabla 6.4). El método de rangos también detecta estos efectos con este valor nominal del parámetro α , pero si se utilizara el valor calibrado de $\alpha = 0.033$, tampoco este método detectaría nada. Los métodos en la escala original y con estimación robusta no detectan ningún efecto activo al arrojar p-valores de 0.47 y 0.25, respectivamente.

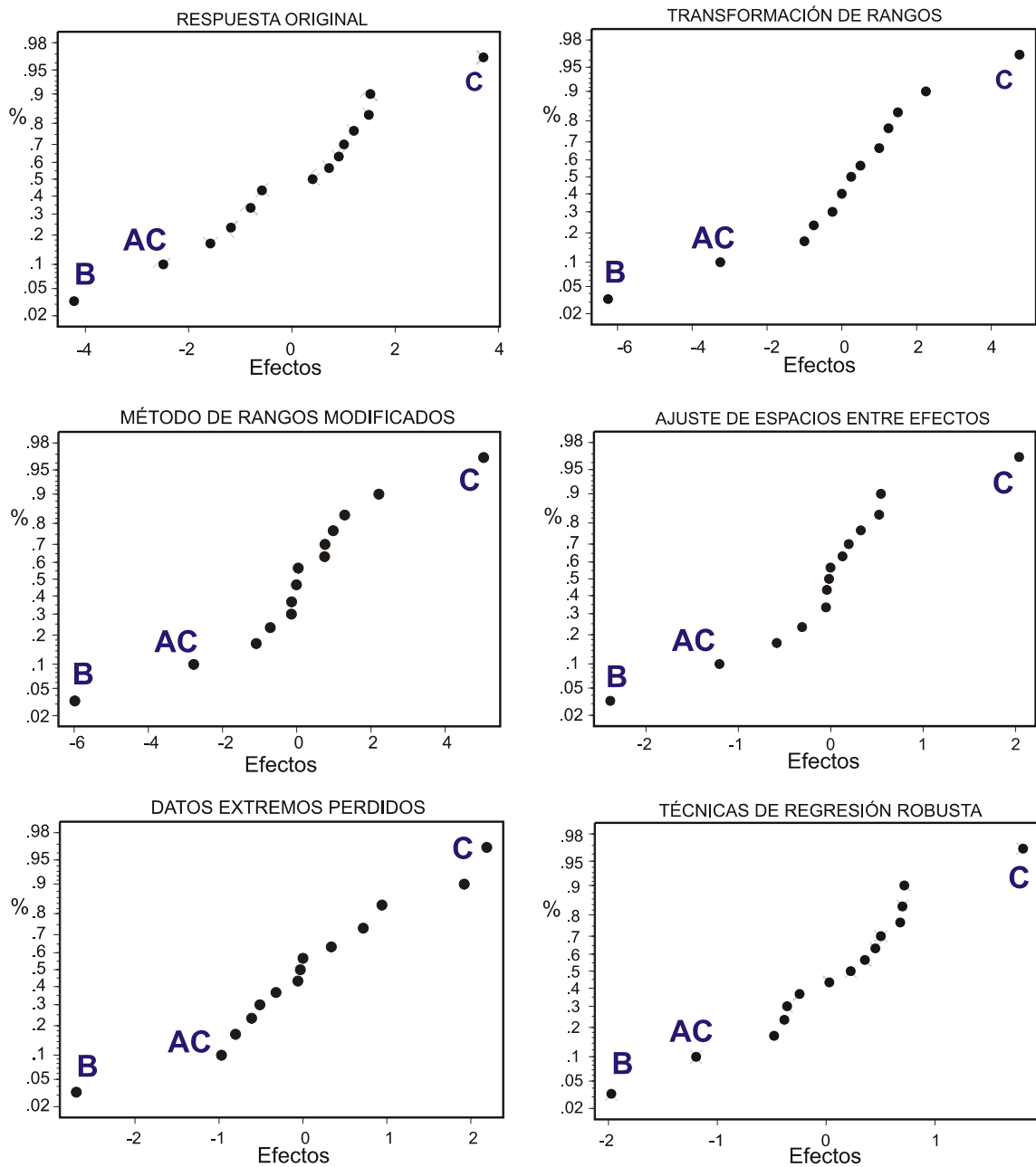


Figura 6.3: Efectos en papel de probabilidad normal, Ejemplo 2

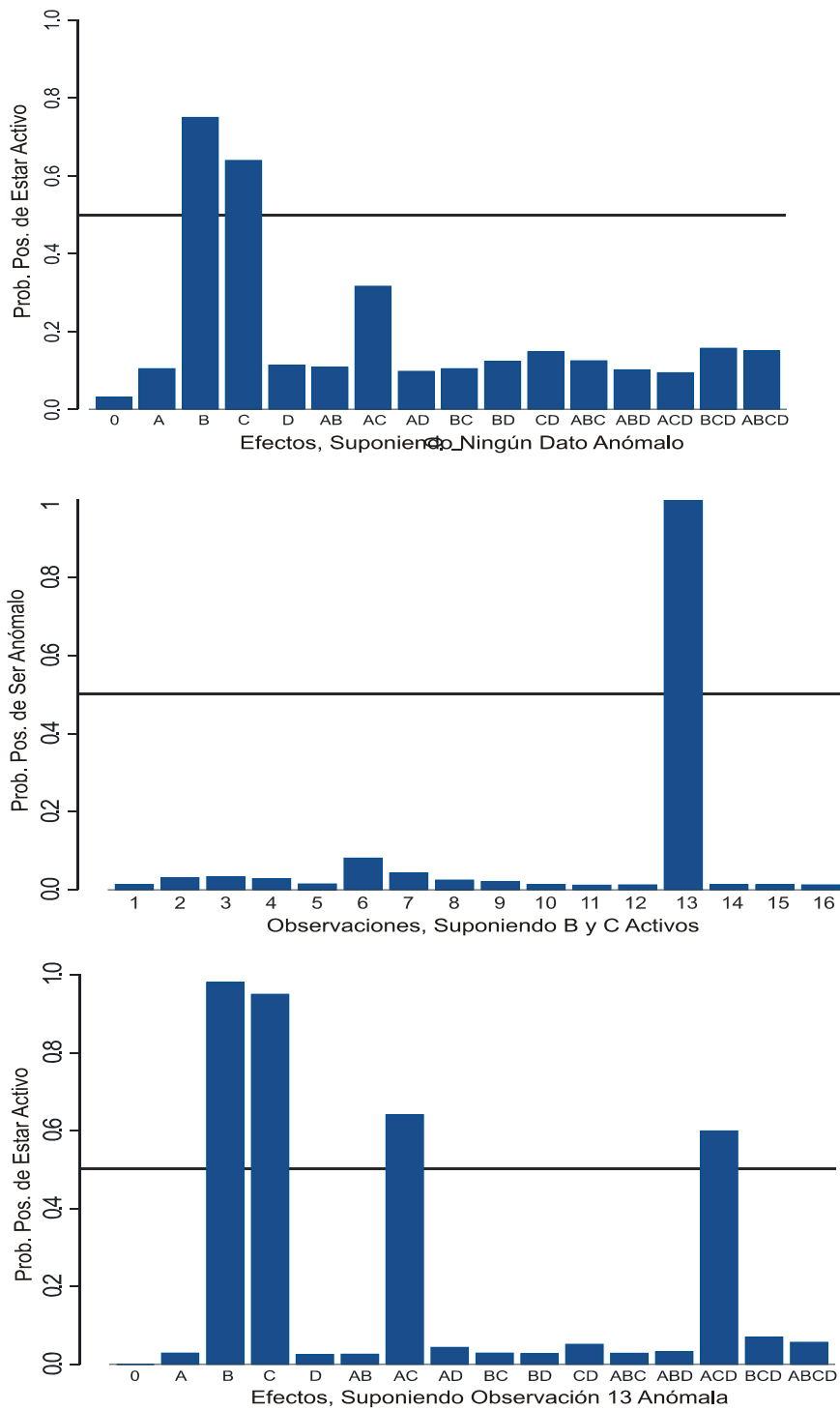


Figura 6.4: Análisis Bayesiano, Ejemplo 2

Tabla 6.4. Comparación analítica, Ejemplo 2

Método	W	P-valor	FL	FU	dF	$\pm 2dF$
Respuesta original	0.9536	0.4703	-0.99	1.105	2.095	± 4.19
Rangos original	0.8757	0.0443	-0.50	1.125	1.625	± 3.25
Rangos Modificados	0.8653	0.0322	-0.430	0.870	1.301	± 2.60
Ajuste de espacios	0.8422	0.0160	-0.181	0.259	0.441	± 0.88
Estimación de datos extremos	0.9440	0.3523	-0.56	0.72	1.28	± 2.56
Técnicas de regresión robusta	0.9675	0.2540	-0.045	0.014	0.059	± 0.12

En la Figura 6.4 se muestran los resultados del método Bayesiano. Suponiendo que no hay datos anómalos se obtienen los efectos B y C con posteriores superiores a 0.5. Suponiendo por el momento este modelo como válido, se obtiene una alta probabilidad posterior de que la observación 13 es anómala. Finalmente, suponiendo que la observación 13 es anómala, se concluye más claramente que B y C están activos, además de que ahora también los efectos AC y ACD tienen posteriores superiores a 0.5. En conclusión, cuando menos B y C están activos y la observación 13 es anómala.

6.3 Ejemplo 3

Este ejemplo, que es un diseño factorial fraccionado 2^{5-1} , fue publicado por Shari Kraber en el Stat-Teaser de diciembre de 1999. Los datos, y los efectos estimados con seis métodos se muestran en la Tabla 6.5. Los correspondientes gráficos de Daniel para los seis tipos de efectos se muestran en la Figura 6.5. En los cuatro gráficos de arriba no se aprecian efectos que puedan estar activos, pero tanto en la gráfica del método de estimación de datos extremos y del método basado en técnicas de regresión robusta, si es notorio el alejamiento de la linealidad de los efectos B , D y DB que parecen estar activos.

De acuerdo al análisis hecho por Kraber (1999) la primera observación (0.14) es un punto aberrante, hecho que se puede ver con los residuos estandarizados considerando tentativamente un modelo con los efectos mayores. Kraber rehace el análisis sin considerar dicho dato y concluye que los efectos activos son B , D y DB , los mismos

que detectan dos de los métodos robustos aquí propuestos (Tabla 6.6). Note que los p-valores de los primeros cuatro métodos son claramente mayores que el p-valor de los dos últimos métodos. Por cierto que el análisis de estimación de datos extremos también detecta como dato anómalo la primera observación.

Tabla 6.5. Datos y Efectos Estimados, Ejemplo 3

Nombre	Datos	Efectos (resp.orig)	Efectos (rangos)	Efectos (ran.mod)	Efectos (esp.ajus)	Efectos (dat.ext.per)	Efectos (reg.robust)
(1)	0.14						
A	0.98	0.045	-0.250	0.58	0.10	-0.12	-0.034
B	0.36	-0.195	-0.750	-3.02	-0.42	-0.36	-0.153
C	0.42	0.050	0.125	0.83	0.11	-0.12	-0.031
D	1.00	-0.285	-3.750	-4.41	-0.73	-0.45	-0.199
E	0.90	-0.005	2.375	-0.15	0.09	0.00	0.053
AB	0.28	-0.090	-1.000	-1.26	-0.05	-0.08	0.009
AC	0.14	-0.125	-1.875	-1.99	-0.17	-0.12	-0.010
AD	0.22	-0.120	-2.250	-1.74	-0.15	-0.11	-0.011
AE	0.26	0.170	3.625	2.67	0.33	0.00	0.029
BC	0.38	-0.115	-2.125	-1.81	-0.14	-0.11	-0.006
BD	0.12	0.260	2.750	4.03	0.64	0.26	0.187
BE	0.30	0.160	3.375	2.49	0.29	-0.01	0.030
CD	0.06	-0.055	0.125	-0.93	0.07	-0.05	0.042
CE	0.22	0.115	1.750	1.66	0.14	-0.05	0.028
DE	0.38	0.180	3.875	2.82	0.36	0.01	0.006

El análisis Bayesiano se muestra en la Figura 6.6. Tomando como modelo tentativo al compuesto por los términos D y BD , se encuentra que la primera observación tiene una probabilidad posterior relativamente alta de ser anómala. Considerando esta posibilidad, se encuentra que los efectos B , D y BD están activos dada su alta probabilidad posterior. Esta conclusión coincide con la obtenida de los dos últimos métodos de la Tabla 6.6 y difiere de la conclusión que arrojan los primeros cuatro métodos.

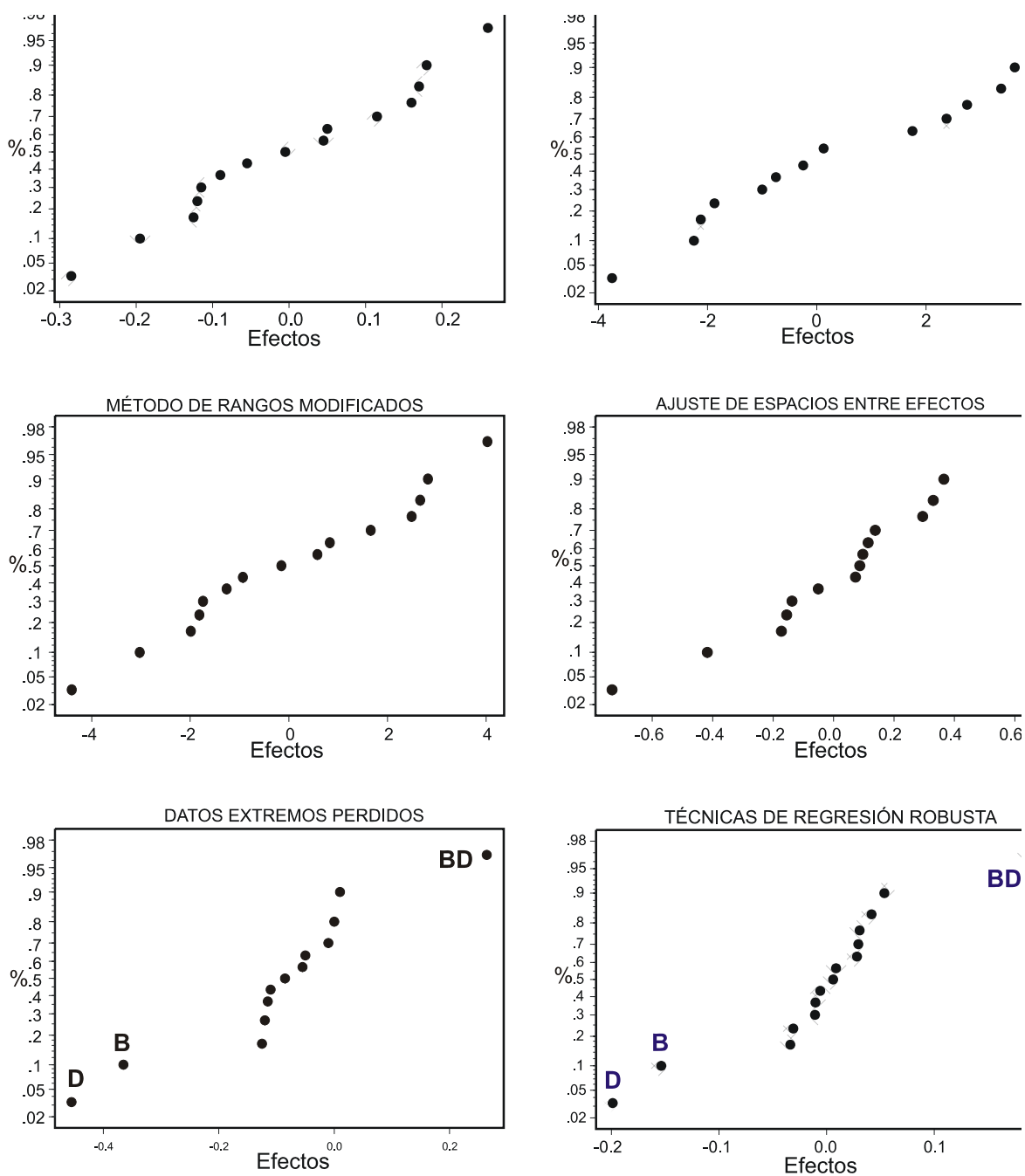


Figura 6.5: Efectos en papel de probabilidad normal, Ejemplo 3

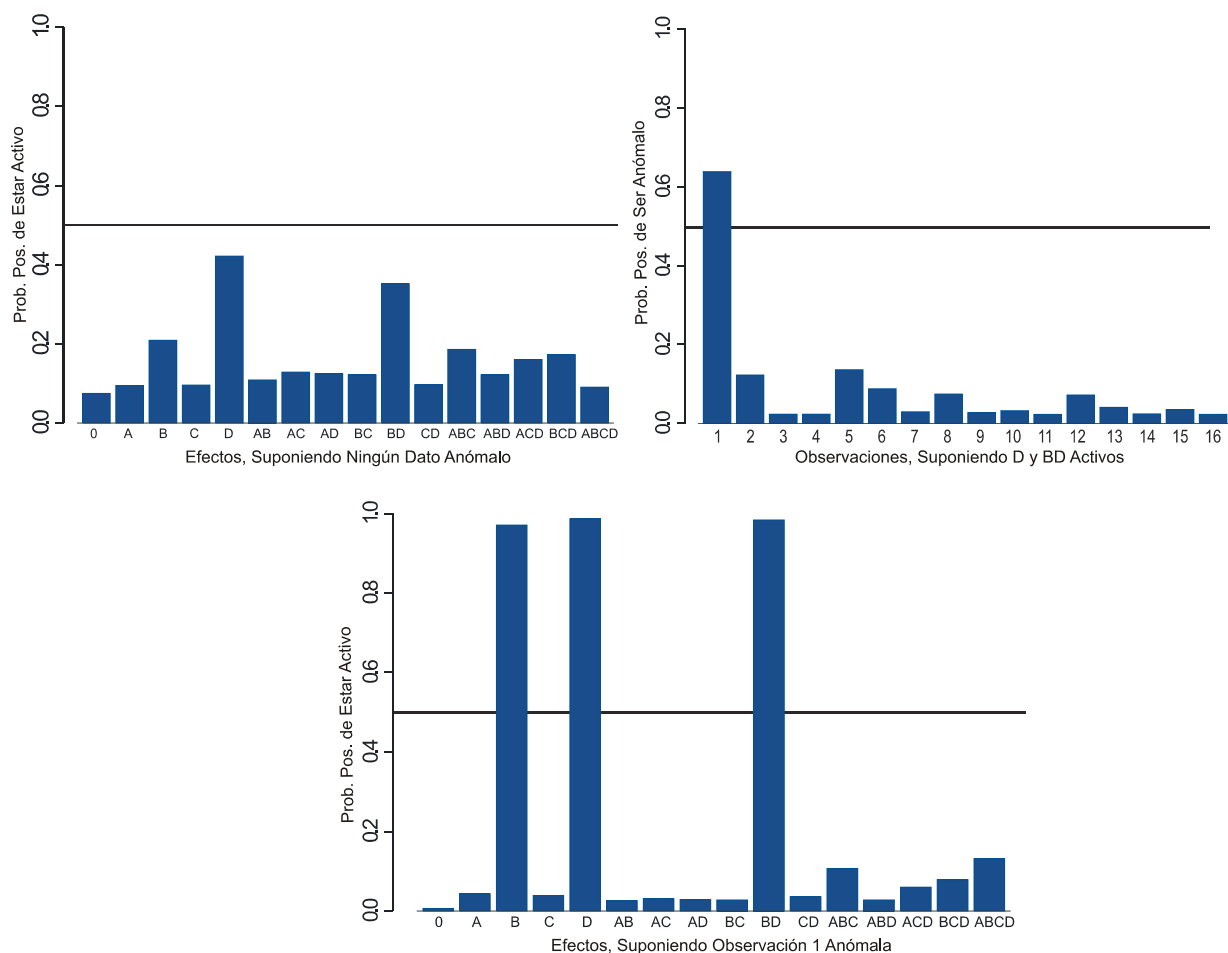


Figura 6.6: Análisis Bayesiano, Ejemplo 3

Tabla 6.6. Comparación analítica, Ejemplo 3

Método	W	P-valor	FL	FU	dF	$\pm 2dF$
Respuesta original	0.9755	0.9148	-0.117	0.137	0.255	± 0.51
Rangos original	0.9586	0.5473	-1.437	2.562	4.000	± 8.00
Rangos modificados	0.9786	1.000	-1.776	2.078	3.854	± 7.71
Ajuste de espacios entre efectos	0.9510	0.4347	-0.146	0.216	0.363	± 0.72
Estimación de datos extremos	0.8448	0.0173	-0.120	-0.005	0.115	± 0.23
Técnicas de regresión robusta	0.8494	0.0199	-0.021	0.030	0.051	± 0.10

Del análisis de estos ejemplos se desprende que tanto los métodos existentes como los métodos nuevos propuestos tienden a detectar mejor los efectos los efectos activos bajo la presencia de datos anómalos que el método que no contempla esa posibilidad,

que aparece en las tablas como Respuesta original y que corresponde a una variante del método de Benski para el caso normal sin datos atípicos. En los dos ejemplos con con datos atípicos el método de Benski no detectó los efectos activos.

En el Ejemplo 2 con un dato atípico extremo, el método de rangos modificados es el que tiene el mejor desempeño seguido de cerca por el método de rangos y por el método Bayesiano. En el Ejemplo 3 con un dato atípico no extremo se destacan dos de los métodos nuevos propuestos junto con el método Bayesiano. En este ejemplo se aprecia cómo el método de rangos no tiene la habilidad de proteger contra datos anómalos internos. Una comparación más exhaustiva de los métodos propuestos y del método de rangos, mediante datos simulados normales sin y con contaminación por datos anómalos, se presenta en el siguiente capítulo.

Capítulo 7

Estudio de Simulación

En este capítulo se calibran y se comparan mediante simulación los métodos propuestos entre ellos y con el método basado en rangos de AT&PT (2001). Para la comparación se considera un escenario con tres efectos activos de diferentes magnitudes y tres niveles de contaminación. Uno de los criterios de comparación es la potencia que demuestra cada método para detectar cada efecto activo, entendida ésta como la proporción de simulaciones en las que el efecto es declarado correctamente activo. Otro criterio es la estadística de mérito Q que toma en cuenta tanto los efectos activos correctamente declarados activos como los efectos inertes incorrectamente declarados activos (ver última sección del Capítulo 3). Específicamente, los métodos a comparar son:

1. Método de rangos, de AT&PT (2001)
2. Método de rangos modificados
3. Método de ajuste de espacios entre efectos
4. Método de reestimación de datos extremos
5. Método de regresión robusta (variante I)
6. Método de regresión robusta (variante II)

7.1 Calibración de los métodos

Para poder hacer una comparación justa de los métodos es preciso calibrarlos para que, bajo la hipótesis de ningún efecto activo, detecten falsamente la misma proporción de experimentos con al menos un efecto activo. Es decir, para cada método se determinó mediante simulación el valor del parámetro α a utilizar en la prueba de normalidad, de manera que trabaje con un error global de 5% ($EER = 5\%$). Por ser los casos más demandantes computacionalmente hablando, los métodos basados en técnicas de regresión robusta se calibraron simulando 12000 experimentos 2^4 no-replicados sin efectos activos, y se encontró que la variante (I) que utiliza muestreo de renglones de la matriz X en el paso 1 debe utilizar un valor de $\alpha = 0.034$, para tener un $EER = 5\%$. Este método considera una eficiencia de 99% en el procedimiento lmRobMM de SPLUS. Por su parte la variante (II) que utiliza regresión L1 requiere un valor $\alpha = 0.047$ con una eficiencia de 99.3% en el procedimiento de SPLUS.

Los métodos 1-4 se calibraron usando 40000 réplicas del experimento. De esta manera se determinó que el método de rangos original debe utilizar $\alpha = 0.033$ para que trabaje con un $EER = 5\%$, en lugar del 0.028 que AT&PT (2001) encuentran con 600 réplicas. El método de rangos modificados requiere $\alpha = 0.045$, y el método de anómalos reestimados trabaja con $\alpha = 0.033$. Finalmente, el método de espacios ajustados requiere un valor de $\alpha = 0.001$, que es bastante pequeño puesto que al ajustar los espacios entre los efectos se trastoca sensiblemente la normalidad de los mismos, aun bajo condiciones de ausencia de efectos activos. No queda pues justificado con este método la aplicación de la prueba de normalidad. Los métodos con los correspondientes valores de su parámetro α se resumen en la tabla de la siguiente página. Cabe aclarar que los métodos son invariantes a transformaciones de escala, es decir, se obtienen los mismos valores de α si la calibración se realiza con base en y_i/a en lugar de y_i usando cualquier escalar $a > 0$.

Método	Valor α	Experimentos Simulados
1. Mét. de rangos	0.033	40000
2. Mét. de rangos modificados	0.045	40000
3. Mét. de espacios ajustados	0.001	40000
4. Mét. de datos anómalos reestimados	0.033	40000
5. Mét. de técnicas de regresión robusta I	0.034	12000
6. Mét. de técnicas de regresión robusta II	0.047	12000

También conviene recordar que los valores de α calibrados dados en esta tabla se aplican solamente a experimentos factoriales con 16 corridas (2^4 , 2^{5-1} , 2^{6-2} , etc.). Para experimentos factoriales no-replicados con otros tamaños, como pueden ser 8 ó 32 corridas, habría que volver a obtener nuevos valores α calibrados, además de afinar algunos otros aspectos particulares de cada uno de los métodos propuestos.

7.2 Escenarios estudiados y resultados obtenidos

En el estudio se utilizan los mismos efectos activos y los mismos escenarios de contaminación estudiados por AT&PT (2001). Esto es, se consideran los tres efectos A , AB y C y en cada réplica del experimento se generan 16 observaciones con el modelo

$$Y_i = \alpha_1 A + \alpha_2 AB + \alpha_3 C + \varepsilon_i \quad ; \quad i = 1, 2, \dots, 16, \quad (7.1)$$

donde la distribución del término error es la mezcla de normales

$$\varepsilon_i \sim (1 - \beta)N(0, 1) + \beta N(0, K^2).$$

La segunda distribución $[N(0, K^2)]$ representa la contaminación, la cual ocurre en cada observación con probabilidad β . De manera que los diferentes niveles de contaminación se pueden denotar con el par (β, K) . En particular, los escenarios que se utilizan en este estudio son: $(0, 0)$, $(0.05, 5)$ y $(0.10, 10)$. El par $(0, 0)$ es el caso normal sin contaminación y el caso más contaminado es el par $(0.10, 10)$ que nos dice que cada dato tiene una probabilidad de 0.10 de provenir de la distribución normal

con desviación estándar inflada 10. Note que en cada réplica del experimento en el par (β, K) se esperarían 16β datos provenientes de la segunda distribución, número que oscila de acuerdo a la varianza $16\beta(1 - \beta)$. Por ejemplo, para el par de mayor contaminación (0.10,10) se esperaría en cada réplica del experimento observar de 1 a 3 datos de la segunda distribución.

Una vez calibrados los métodos se probaron los tres escenarios mencionados considerando tres efectos activos (A, C, AB) de diferente magnitud. Se les asignaron los valores $A = 2\sigma_{error}$, $AB = \sigma_{error}$ y $C = 4\sigma_{error}$, que en términos de los coeficientes del modelo dado arriba (ecuación 1) son $\alpha_1 = 1$, $\alpha_2 = 0.5$ y $\alpha_3 = 2$, respectivamente; Se toma $\sigma_{error} = 1$ que es la desviación estándar del error experimental cuando no hay observaciones anómalas. Una vez obtenidos los datos en cada réplica se calculan los 15 efectos del factorial 2^4 y con cada método se determina cuáles de los efectos activos y cuántos de los efectos inertes se declaran activos. En cada escenario se realizaron 3000 réplicas del experimento con los métodos basados en regresión robusta y 20000 réplicas con el resto de los métodos. Los resultados obtenidos en los tres escenarios se muestran en las tablas 7.1-7.3.

Tabla 7.1. Potencia y estadístico QG en el par $(\beta = 0, K = 0)$

Método	A	C	AB	n⁺	n⁻	QG
1. Rangos	60.27	94.20	10.30	82387	2819	54.66
2. Rangos modificados	67.05	96.64	12.07	17576	585	58.30
3. Espacios ajustados	88.74	91.24	49.81	45959	23680	69.04
4. Datos anómalos reestimados	65.42	92.03	13.49	17094	786	56.61
5. Reg. robusta (variante I)	64.72	95.10	12.62	6898	268	57.16
6. Reg. robusta (variante II)	65.85	95.12	12.30	6931	264	57.44

Tabla 7.2. Potencia y estadístico QG en el par $(\beta = 0.05, K = 5)$

Método	A	C	AB	n⁺	n⁻	QG
1. Rangos	44.22	78.47	7.34	65020	2742	43.15
2. Rangos modificados	49.62	81.43	8.32	13937	586	46.23
3. Espacios ajustados	67.65	74.53	36.20	35678	21706	54.08
4. Datos anómalos reestimados	51.55	79.79	11.35	14269	739	47.27
5. Reg. robusta (variante I)	51.67	83.25	10.72	5826	286	48.61
6. Reg. robusta (variante II)	54.55	85.3	10.60	6018	247	49.89

Tabla 7.3. Potencia y estadístico QG en el par ($\beta = 0.10, K = 10$)

Método	A	C	AB	n^+	n^-	QG
1. Rangos	25.41	54.60	4.31	42158	2680	27.98
2. Rangos modificados	29.72	55.26	4.84	8982	606	29.79
3. Espacios ajustados	35.69	45.16	18.91	19952	15909	31.05
4. Datos anómalos reestimados	38.81	62.67	7.77	10925	794	36.17
5. Reg. robusta (variante I)	39.02	65.65	7.80	4499	385	37.19
6. Reg. robusta (variante II)	42.00	69.57	8.45	4801	366	39.70

Las potencias de todos los métodos propuestos y en todos los escenarios son sistemáticamente mayores que la del método de rangos, con excepción del escenario sin contaminación en el efecto más grande (C) (Tabla 7.1), donde el método de rangos resulta mejor que el método de espacios ajustados y que el método de datos anómalos reestimados. Otra excepción es el método de espacios ajustados cuya potencia para detectar el efecto más grande se mantiene siempre resagada respecto a todos los métodos restantes, resago que empeora con la contaminación; pero su potencia con los efectos pequeños es de llamar la atención, aún en el escenario más contaminado.

Los métodos de reestimación de datos extremos y de técnicas de regresión robusta mantienen un desempeño consistentemente bueno a lo largo de todos los escenarios y su potencia respecto al método de rangos se incrementa en la medida que aumenta la contaminación. Por su parte el método de rangos modificados comienza a perder su poca ventaja respecto al método de rangos al incrementarse la contaminación.

La estadística de mérito QG es sistemáticamente mejor con todos los métodos propuestos y en todos los escenarios, en relación al método de rangos. En la Figura 7.1 se hace una representación gráfica de las tablas 7.1-7.3, efecto por efecto con el mismo tamaño de la escala vertical para una mejor comparación relativa. La más punteada corresponde al método de regresión robusta variante II (método 6) y la línea sólida es el método de rangos (método 1). Note que el método de rangos tiene la menor potencia global (QG) y con los efectos A y AB , logrando buen desempeño sólo con el efecto mayor (C) cuando no hay contaminación. Es claro en la figura el desempeño atípico del método de ajuste de espacios entre efectos, en particular

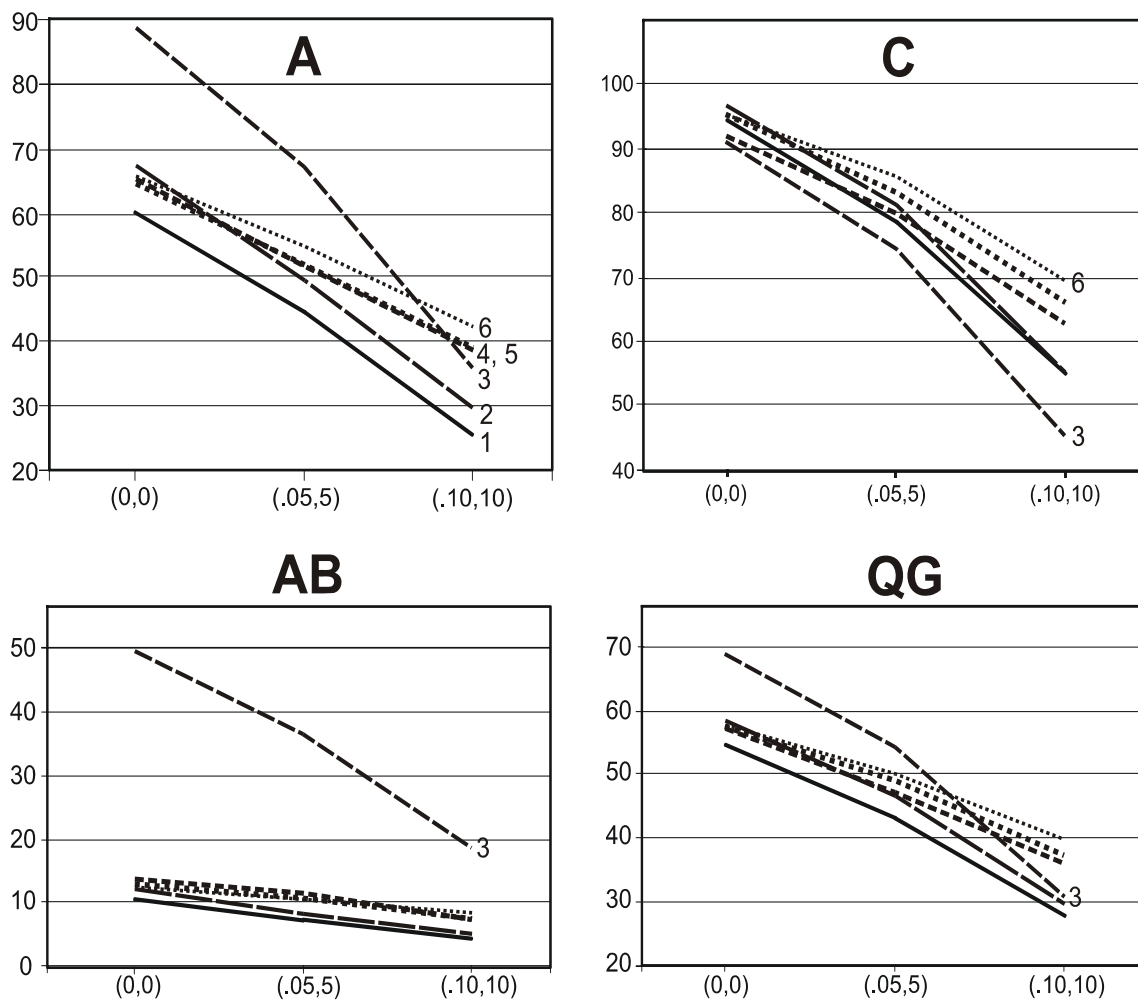


Figura 7.1: Potencias por Efecto y Estadística Global de Mérito

con los efectos más pequeños *AB* y *A*. Los métodos basados en técnicas de regresión robusta (líneas de punto más fino) tienen el mejor desempeño global.

7.3 Efecto de la Eficiencia y del Número de Términos a Fijar en el Modelo

El número de términos que se fijan en el paso 2, así como la eficiencia solicitada al procedimiento de MM-estimación, influyen de manera importante en el desempeño de los métodos basados en regresión robusta propuestos (variantes I y II). La eficiencia mide el desempeño del estimador robusto bajo el modelo normal sin contaminación. Decidimos probar 2 tamaños de modelo y 3 eficiencias. En cada una de las 6 combinaciones se calibró el procedimiento al EER deseado de 5%, y luego se probaron los mismos tres escenarios mencionados arriba. Los resultados obtenidos se resumen en la Tabla 7.4, para el caso con paso 1 basado en el muestreo de renglones como estrategia para seleccionar los renglones buenos (variante I).

Aunque las eficiencias de 0.991, 0.993 y 0.995 no fueron lo suficientemente espaciadas para detectar mejor su efecto, sí se aprecia que a mayor eficiencia se tiene mayor potencia para detectar los efectos activos en el escenario sin contaminación ($\beta = 0, K = 0$), mientras al mismo tiempo se pierde potencia para detectar los efectos de interés en el escenario más contaminado. Este comportamiento de la eficiencia es el esperado de acuerdo a la teoría de MM-estimación descrita en el Capítulo 5. Se observaron los mejores desempeños con eficiencias de 0.993 utilizando 4 ó 5 términos en la parte fija del modelo, con cualquiera de las dos variantes del método.

En cuanto a los modelos, se tiene mejor desempeño si en el paso 2 del método propuesto se fijan 4 términos en lugar de 5 términos. Note en la Tabla 7.4 que las potencias con 4 términos en la parte fija del modelo tienden a ser ligeramente mejores que con 5 términos. Esto ocurre en principio porque sólo hay 3 efectos activos por detectar, además que el método robusto trabaja mejor al tener que estimar menos parámetros con las 16 observaciones del experimento. Posiblemente con más efectos activos resulte mejor el modelo con 5 términos. .

Tabla 7.4. Potencias considerando Tres eficiencias, Dos modelos y Tres escenarios de contaminación

térm. fijos	Escenario	Eficie.	A	C	AB	n ⁺	n ⁻	QG	α
5	$(\beta = 0, K = 0)$	0.991	62.55	92.97	12.60	6725	269	55.73	0.038
5	$(\beta = 0, K = 0)$	0.993	64.12	94.30	12.97	6856	313	56.76	0.043
5	$(\beta = 0, K = 0)$	0.995	65.20	94.92	12.50	6905	312	57.17	0.048
5	$(\beta = 0.05, K = 5)$	0.991	50.32	81.82	9.40	5662	336	46.85	0.038
5	$(\beta = 0.05, K = 5)$	0.993	50.95	81.42	10.72	5724	300	47.40	0.043
5	$(\beta = 0.05, K = 5)$	0.995	51.85	82.72	9.35	5757	291	47.68	0.048
5	$(\beta = 0.1, K = 10)$	0.991	37.72	64.45	8.17	4414	496	36.40	0.038
5	$(\beta = 0.1, K = 10)$	0.993	38.12	65.15	8.02	4452	448	36.75	0.043
5	$(\beta = 0.1, K = 10)$	0.995	36.85	64.42	6.82	4324	443	35.70	0.048
4	$(\beta = 0, K = 0)$	0.991	66.12	94.65	12.90	6947	270	57.57	0.042
4	$(\beta = 0, K = 0)$	0.993	64.72	95.10	12.62	6898	268	57.16	0.047
4	$(\beta = 0, K = 0)$	0.995	66.02	96.10	13.60	7029	267	58.25	0.048
4	$(\beta = 0.05, K = 5)$	0.991	52.47	83.07	10.60	5846	303	48.41	0.042
4	$(\beta = 0.05, K = 5)$	0.993	51.67	83.25	10.72	5826	286	48.61	0.047
4	$(\beta = 0.05, K = 5)$	0.995	51.92	83.20	8.97	5764	279	47.75	0.048
4	$(\beta = 0.1, K = 10)$	0.991	37.05	63.35	6.90	4292	394	35.47	0.042
4	$(\beta = 0.1, K = 10)$	0.993	39.02	65.65	7.80	4499	385	37.19	0.047
4	$(\beta = 0.1, K = 10)$	0.995	37.15	64.45	7.22	4353	357	35.70	0.048

Se realiza un estudio de simulación similar al de la Tabla 7.4, pero ahora utilizando el paso 1 basado en regresión L1 (variante II), y los resultados se muestran en la Tabla 7.5. El efecto de la eficiencia es ahora más claro puesto que se espaciaron más los niveles de prueba utilizados: se observa que a mayor eficiencia, se tiene mayor potencia en la situación sin contaminación y menos potencia en los escenarios con contaminación. Se tienen mejores potencias fijando 4 términos en el paso 2. Otra vez, la eficiencia de 0.993 resulta la de mayor potencia para los dos modelos.

Tabla 7.5. Potencias con Paso1 basado en Regresión L1

térn. fijos	Escenario	Eficie.	A	C	AB	n ⁺	n ⁻	QG	α
5	$(\beta = 0, K = 0)$	0.987	64.55	92.7	13.05	6812	325	56.38	0.034
5	$(\beta = 0, K = 0)$	0.993	65.47	94.45	12.82	6910	294	57.23	0.038
5	$(\beta = 0, K = 0)$	0.999	67.07	96.52	12.50	7044	224	58.42	0.054
5	$(\beta = 0.05, K = 5)$	0.987	52.3	81.54	11.00	7242	369	47.98	0.034
5	$(\beta = 0.05, K = 5)$	0.993	51.45	81.52	10.45	5737	275	47.53	0.038
5	$(\beta = 0.05, K = 5)$	0.999	49.6	81.25	9.45	5612	234	46.54	0.054
5	$(\beta = 0.1, K = 10)$	0.987	39.72	67.95	7.52	4608	478	38.02	0.034
5	$(\beta = 0.1, K = 10)$	0.993	40.77	67.60	7.95	4653	507	38.36	0.038
5	$(\beta = 0.1, K = 10)$	0.999	32.87	58.80	6.20	3915	328	32.40	0.054
4	$(\beta = 0, K = 0)$	0.987	64.05	93.82	12.55	6817	273	56.48	0.036
4	$(\beta = 0, K = 0)$	0.993	65.85	95.12	12.30	6931	264	57.44	0.044
4	$(\beta = 0, K = 0)$	0.999	68.02	97.05	12.42	7100	251	58.86	0.060
4	$(\beta = 0.05, K = 5)$	0.987	52.47	83.12	10.35	5838	250	48.39	0.036
4	$(\beta = 0.05, K = 5)$	0.993	54.55	85.3	10.60	6018	247	49.89	0.044
4	$(\beta = 0.05, K = 5)$	0.999	51.22	83.82	8.82	5755	236	47.72	0.060
4	$(\beta = 0.1, K = 10)$	0.987	42.95	69.00	7.97	4797	470	39.58	0.036
4	$(\beta = 0.1, K = 10)$	0.993	42.00	69.57	8.45	4801	366	39.70	0.044
4	$(\beta = 0.1, K = 10)$	0.999	33.85	60.62	6.60	4043	241	33.52	0.060

Capítulo 8

Detección de Efectos Activos en Experimentos Factoriales con Respuesta No Normal

Muchos experimentos que se hacen en la práctica tienen respuesta no normal. Este es el caso, por ejemplo, cuando la respuesta es un conteo de defectos, o de defectuosos (ver por ejemplo Lewis y Montgomery (2001), Hamada y Nelder (1997)). En este capítulo se presenta una metodología de selección Bayesiana de modelos en el contexto de modelos lineales generalizados para detectar efectos activos cuando la respuesta sigue distribuciones diferentes a la normal, como son las distribuciones Poisson, binomial y gama.

Buscando subsanar la subjetividad del método de Daniel para detectar efectos activos en el caso normal, Box y Meyer (1993) propuso un método Bayesiano para calcular las probabilidades *a posteriori* de todos los modelos posibles que se pueden formar con los factores considerados en el experimento, y de aquí, marginalizando, obtiene las probabilidades *a posteriori* de que cada efecto sea activo. Todo esto lo realiza para el caso normal, que es un caso particular de los llamados modelos lineales generalizados. Cabe decir que el método Bayesiano no se limita al caso no replicado

sino que se podría aplicar sin problemas a experimentos con réplicas.

En el análisis clásico de los modelos lineales generalizados la significancia de los términos se verifica con una prueba tipo Wald (ver Sección 8.3) que se basa en la normalidad asintótica de los estimadores de máxima verosimilitud. Cada coeficiente dividido entre su error estándar es asintóticamente normal estándar bajo H_0 , de manera que al elevar dicho cociente al cuadrado es asintóticamente ji-cuadrada con un grado de libertad (χ_1^2). Sin embargo, no se sabe el desempeño que tienen estas pruebas en el caso no replicado, ni se han propuesto hasta ahora métodos especiales para GLM's con factoriales no replicados, sino que se usan las mismas pruebas que en el caso replicado. Una recomendación que se da (ver Myers *et al.* (2002) página 270) es graficar los coeficientes divididos por sus errores estándar en papel de probabilidad normal, pero debe tenerse cuidado si la correlación entre los coeficientes es grande. Una alternativa es utilizar una función liga que estabilice la varianza, y que da lugar a coeficientes no correlacionados.

Cabe aclarar que algunos autores sí han trabajado la selección Bayesiana de variables en modelos lineales generalizados, aunque considerando situaciones más generales que la detección de efectos activos en diseños factoriales no replicados, ver por ejemplo Wang y Chen (2004). Un tema relacionado es la ponderación Bayesiana de modelos (Hoeting, *et. al.*, 1999; Clyde, M. 1999), que como su nombre lo dice considera los m modelos con probabilidades posteriores mayores para hacer predicciones de la variable de interés, incorporando de esta manera la incertidumbre de utilizar un solo modelo. Este no es el problema que interesa en esta tesis, sin embargo, la ponderación de modelos requiere necesariamente calcular las probabilidades posteriores de los modelos, mismas que se pueden utilizar también para hacer selección de variables, y específicamente para detectar efectos activos en factoriales no-replicados en el contexto de modelos lineales generalizados.

En este capítulo se presentan los fundamentos teóricos necesarios para explorar la posibilidad de utilizar métodos Bayesianos para la selección de efectos activos en

modelos lineales generalizados, de manera similar a como lo hace Box y Meyer (1993) en el caso normal. Comenzamos con una breve introducción al los GLM's, luego se describe el método de Box y Meyer, para posteriormente, en los capítulos 9, 10 y 11 discutir los casos de regresión de Poisson con diferentes ligas, regresión logística y regresión gama, respectivamente. En cada caso se presenta un ejemplo de la literatura que se analiza con los métodos Bayesianos propuestos y los resultados se comparan con los obtenidos por el método frecuentista.

La solución propuesta para los casos Poisson, binomial y gama se basa en la integración numérica usando técnicas de Quasi Monte Carlo (Krykova, 2003; Niederreiter, 1992) de la función predictiva *a priori* resultante. Se tiene que resolver una integral por cada modelo, lo que implica bastante costo computacional, dado que en un factorial 2^4 se tienen $2^{15} = 32768$ modelos posibles.

8.1 Familia Exponencial de Distribuciones

Sean y_1, y_2, \dots, y_n observaciones independientes con medias $\mu_1, \mu_2, \dots, \mu_n$, donde y_i tiene una distribución de la familia exponencial, lo que significa que su distribución se puede escribir en la forma

$$f(y_i | \theta_i, \phi) = \exp \{r(\phi)[y_i \theta_i - b(\theta_i)] + c(y_i, \phi)\}, \quad (i = 1, 2, \dots, n) \quad (8.1)$$

donde ϕ es el parámetro de escala y θ_i es el parámetro natural de localización (Myers y Montgomery, 1997). La parte sistemática del modelo involucra los factores o regresores x_1, x_2, \dots, x_k a través del predictor lineal $\eta = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k = \mathbf{x}'\boldsymbol{\beta}$. El modelo expresa la relación entre las medias y el predictor lineal mediante la función liga $\eta_i = g(\mu_i)$. La función liga g es una función monótona y diferenciable, que cuando es igual a θ_i se dice que es la liga canónica. La varianza σ_i^2 ($i = 1, 2, \dots, n$) es función de la media μ_i , excepto en el caso normal donde la varianza es constante.

Por ejemplo, en el caso normal

$$\begin{aligned} f(y_i | \mu_i, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\{-(y_i - \mu_i)^2/2\sigma^2\} \\ &= \exp\left\{\frac{1}{\sigma^2}[y_i\mu_i - \mu_i^2/2] - \frac{1}{2}[y_i^2/\sigma^2 + \log(2\pi\sigma^2)]\right\}, \quad (i = 1, 2, \dots, n), \end{aligned}$$

de manera que $\theta_i = \mu_i$, $b(\theta_i) = \theta_i^2/2$, $r(\phi) = 1/\sigma^2$ y $c(y_i, \phi) = -\frac{1}{2}[y_i^2/\sigma^2 + \log(2\pi\sigma^2)]$. Note que la liga canónica es la función identidad, es decir, $\eta_i = g(\mu_i) = \mu_i$ y que el parámetro de escala es $\phi = \sigma^2$.

En el caso Poisson

$$\begin{aligned} f(y_i | \mu_i) &= \mu_i^{y_i} \exp\{-\mu_i\}/y_i! \\ &= \exp\{y_i \log(\mu_i) - \mu_i - \log(y_i!)\}, \quad (i = 1, 2, \dots, n), \end{aligned}$$

de modo que $\theta_i = \log(\mu_i)$, $b(\theta_i) = e^{\theta_i}$, $r(\phi) = 1$ y $c(y_i, \phi) = -\log(y_i!)$. En este caso, si usamos la liga canónica, la función liga es $g(\mu_i) = \log(\mu_i)$.

En el caso binomial tenemos que

$$\begin{aligned} f(y_i | \mu_i) &= \binom{n_i}{y_i} \left(\frac{\mu_i}{n_i}\right)^{y_i} \left(1 - \frac{\mu_i}{n_i}\right)^{n_i - y_i} \\ &= \exp\left[y_i \log\left(\frac{\mu_i/n_i}{1 - \mu_i/n_i}\right) + n_i \log\left(\frac{e^{-\theta_i}}{1 + e^{-\theta_i}}\right) + \log\binom{n_i}{y_i}\right] \end{aligned}$$

de tal forma que $\theta_i = \log\left(\frac{\mu_i/n_i}{1 - \mu_i/n_i}\right)$, $b(\theta_i) = -n_i \log\left(\frac{e^{-\theta_i}}{1 + e^{-\theta_i}}\right)$, $r(\phi) = 1$ y $c(y_i, \phi) = \log\binom{n_i}{y_i}$.

Otro miembro relevante de la familia exponencial es la distribución gama, como se muestra a continuación.

$$\begin{aligned} f(y_i | \lambda_i, r) &= \frac{1}{\Gamma(r)} \left(\frac{1}{\lambda_i}\right)^r e^{-y_i/\lambda_i} y_i^{r-1}; \quad y_i \geq 0, \quad r > 0, \quad \lambda_i > 0 \\ &= \exp(-\log(\Gamma(r)) - r \log(\lambda_i) - y_i/\lambda_i + (r-1) \log(y_i)) \\ &= \exp\left(\frac{(-y_i/\lambda_i r + \log(1/\lambda_i r))}{1/r} + r \log r + (r-1) \log(y_i) - \log(\Gamma(r))\right) \end{aligned}$$

donde r es parámetro de escala y λ_i es parámetro de localización. De esta última expresión se desprende que $\theta_i = -1/\lambda_i r$, $b(\theta_i) = -\log(-\theta_i)$, $r(\phi) = r$ y $c(y_i, \phi) =$

$r \log r + (r - 1) \log(y_i) - \log(\Gamma(r))$. También se observa que la liga canónica es la función inversa $g(\mu_i) = 1/\mu_i$.

8.2 Estructura de los Modelos Lineales Generalizados

En esta sección se indica de manera más específica la estructura que tienen los modelos lineales generalizados (ver Myers *et al.* 2002 o McCullagh y Nelder, 1989) y es la siguiente:

1. Se tienen y_1, y_2, \dots, y_n observaciones independientes con medias $\mu_1, \mu_2, \dots, \mu_n$, respectivamente.
2. La observación y_i tiene una distribución que es miembro de la familia exponencial.
3. La parte sistemática del modelo involucra las variables regresoras o factores x_1, x_2, \dots, x_k .
4. El modelo se construye alrededor del predictor lineal $\eta = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k = \mathbf{x}'\boldsymbol{\beta}$.
5. El modelo expresa la relación entre las medias y el predictor lineal mediante la función liga $\eta_i = g(\mu_i)$; $i = 1, 2, \dots, n$.
6. La función liga es monótona y diferenciable.
7. La varianza σ_i^2 ($i = 1, 2, \dots, n$) es una función de la media μ_i .

En resumen, un modelo lineal generalizado en nuestro caso tiene tres componentes básicos, a saber: 1. La distribución de la variable de respuesta y , que es de la familia exponencial, 2. El predictor lineal η que involucra a las variables regresoras y 3. La función liga g que relaciona al predictor lineal con la media de la variable de respuesta. Sean y_i ($i = 1, 2, \dots, n$) los valores de la respuesta, entonces el modelo lineal generalizado se puede escribir como

$$g(\mu_i) = g[E(y_i)] = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k = \mathbf{x}'\boldsymbol{\beta},$$

donde y_i sigue una distribución de la familia exponencial con media μ_i . Veamos algunos ejemplos concretos.

En el caso normal el modelo lineal generalizado, que es simplemente el modelo de regresión múltiple, se puede escribir como

$$\mu_i = E(y_i) = \mathbf{x}'\boldsymbol{\beta}$$

donde y_i sigue una distribución normal con media μ_i . En este caso $g(\mu_i) = \mu_i$ es la liga identidad, que es la liga canónica para este caso.

En el caso Poisson se tiene que

$$g(\mu_i) = \log(\mu_i) = \log[E(y_i)] = \mathbf{x}'\boldsymbol{\beta}$$

donde y_i tiene la distribución de Poisson con media μ_i y se usa la liga canónica $g(\mu_i) = \log(\mu_i)$.

En el caso Binomial se tiene que

$$g(\mu_i) = \log\left(\frac{\mu_i/n_i}{1 - \mu_i/n_i}\right) = \log\left(\frac{E(y_i)/n_i}{1 - E(y_i)/n_i}\right) = \mathbf{x}'\boldsymbol{\beta}$$

donde y_i sigue la distribución de binomial con media $\mu_i = n_i p_i$, usando su liga canónica $g(\mu_i) = \log\left(\frac{\mu_i/n_i}{1 - \mu_i/n_i}\right)$.

8.3 Inferencia tipo Wald e inferencia basada en verosimilitud

La inferencia tipo Wald se basa fundamentalmente en la aplicación de las propiedades asintóticas de los estimadores de máxima verosimilitud, en particular su normalidad asintótica. Usando esta propiedad se puede probar la hipótesis

$$H_0 : \beta_i = 0$$

para cada parámetro en el predictor lineal ($i = 0, 1, \dots, k$). Bajo H_0 , el estadístico de prueba

$$\left(\frac{b_i}{\widehat{\sigma}_{b_i}} \right)^2$$

sigue una distribución χ_1^2 para muestras grandes. El error estándar del estimador $\widehat{\sigma}_{b_i}$ es el i -ésimo elemento de la diagonal de la matriz de información inversa. Del hecho de que el estadístico $b_i/\widehat{\sigma}_{b_i}$ se distribuye asintóticamente $N(0, 1)$ se deduce que un intervalo al $100(1 - \alpha)\%$ de confianza para el parámetro β_i es

$$b_i - z_{\alpha/2}\widehat{\sigma}_{b_i} \leq \beta_i \leq b_i + z_{\alpha/2}\widehat{\sigma}_{b_i}.$$

La inferencia basada en verosimilitud utiliza la diferencia en log verosimilitud de un modelo reducido contra un modelo completo para probar hipótesis sobre suconjuntos de parámetros. Si se denota la verosimilitud por L , se tiene que, asintóticamente,

$$-2 \ln \left[\frac{L(\text{reducido})}{L(\text{completo})} \right] \sim \chi_\nu^2$$

donde ν es la diferencia en el número de parámetros entre los modelos completo y reducido.

Los usuarios de estas técnicas suelen no tomar en cuenta que el diseño experimental por analizar es no-replicado y que no necesariamente aplica esta inferencia basada en resultados asintóticos. El uso de estas pruebas sin mayor análisis o sin el apoyo de métodos como los que en esta segunda vertiente del trabajo de tesis proponemos, puede llevar por ejemplo a una detección excesiva de efectos activos que en realidad están inertes.

8.4 Elementos de Estadística Bayesiana

Sea una función de densidad $f(x|\boldsymbol{\theta})$ propuesta para los datos observados D , donde $\boldsymbol{\theta}$ es un vector de parámetros desconocidos. Suponiendo $\boldsymbol{\theta}$ aleatorio, sea $\pi(\boldsymbol{\theta})$ su distribución *a priori* que refleja el conocimiento, previo a la obtención de los datos, que

tiene el experimentador o usuario sobre el parámetro $\boldsymbol{\theta}$. La función de verosimilitud es la distribución conjunta de los datos condicionada a un valor del parámetro y se denota por $L(\boldsymbol{\theta}|D)$. El teorema de Bayes dice que la distribución posterior de los parámetros dados los datos, $\pi(\boldsymbol{\theta}|D)$, está dada por

$$\pi(\boldsymbol{\theta}|D) = \frac{L(\boldsymbol{\theta}|D) \pi(\boldsymbol{\theta})}{\int_{\Theta} L(\boldsymbol{\theta}|D) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

donde Θ es el espacio de los parámetros; el denominador $f(D) = \int_{\Theta} L(\boldsymbol{\theta}|D) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}$ es la distribución marginal conjunta de los datos, y es a su vez la constante normalizadora de $\pi(\boldsymbol{\theta}|D)$, y no depende de $\boldsymbol{\theta}$. Es claro de esta última relación que la distribución posterior $\pi(\boldsymbol{\theta}|D)$ es proporcional a la verosimilitud multiplicada por la distribución *a priori*,

$$\pi(\boldsymbol{\theta}|D) \propto L(\boldsymbol{\theta}|D) \pi(\boldsymbol{\theta}),$$

de donde se aprecia que $\pi(\boldsymbol{\theta}|D)$ se puede interpretar como una actualización de la distribución *a priori* $\pi(\boldsymbol{\theta})$ una vez observados los datos.

Densidad Predictiva. Supongamos que y es un valor futuro de la respuesta de interés con distribución $f(y|\boldsymbol{\theta})$. Se puede demostrar (ver Bernardo y Smith, 1994, pág. 243) que la distribución posterior de y dados los datos observados está dada por

$$f(y|D) = \int_{\Theta} f(y|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|D) d\boldsymbol{\theta}$$

si los datos son independientes dado $\boldsymbol{\theta}$. Esta función se conoce también como distribución posterior predictiva de la observación futura y . En algunos problemas es necesario obtener la distribución predictiva *a priori* que se define por

$$f(y) = \int_{\Theta} f(y|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (8.2)$$

donde aparece la distribución *a priori* en lugar de la distribución posterior de los parámetros.

Selección de Modelos. Supongamos que se tienen dos modelos M_1 y M_2 en competencia con parámetros $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}_2$, respectivamente, y se quieren comparar a la luz de

los datos observados D . Por el teorema de Bayes, la probabilidad posterior de que el modelo M_1 sea el correcto, dados los datos, es

$$p(M_1 | D) = \frac{p(D | M_1) p(M_1)}{p(D | M_1) p(M_1) + p(D | M_2) p(M_2)} \quad (8.3)$$

donde $p(D | M_i)$ es la probabilidad marginal de los datos dado el modelo M_i ; $p(M_i)$ es la probabilidad *a priori* del modelo M_i , y por construcción $p(M_1 | D) + p(M_2 | D) = 1$. La probabilidad $p(D | M_1)$ se obtiene de la siguiente integral respecto a θ_1 :

$$p(D | M_1) = \int_{\Theta_1} p(D | \theta_1, M_1) p(\theta_1 | M_1) d\theta_1 \quad (8.4)$$

donde $p(D | \theta_1, M_1)$ es la función de verosimilitud del modelo M_1 y $p(\theta_1 | M_1)$ es la distribución *a priori* de los parámetros. Note que $p(D | M_1)$ es un caso de distribución predictiva *a priori* como la definida en la ecuación (8.2). El qué tanto los datos apoyan el modelo 2 se mide con el momio posterior del modelo M_2 contra el modelo M_1 , esto es, el cociente de las probabilidades posteriores

$$\frac{p(M_2 | D)}{p(M_1 | D)} = \left[\frac{p(D | M_2)}{p(D | M_1)} \right] \frac{p(M_2)}{p(M_1)}$$

donde el término entre corchetes, que es el cociente de las distribuciones predictivas *a priori*, se conoce como *factor de Bayes* (B_{21}). Cuando $B_{21} > 1$ los datos favorecen al modelo M_2 ; si $B_{21} < 1$ los datos apoyan al modelo M_1 (Raftery, 1995).

La obtención de la distribución predictiva *a priori* dada en la ecuación (8.4) puede ser un problema de integración multidimensional e intratable. De aquí que se hayan propuesto varias maneras de aproximar la integral de forma analítica y también numéricamente (ver Evans y Swartz, 1995). En el caso de selección de modelos lineales generalizados que planteamos en las siguientes secciones es típico que la integral en cuestión resulta intratable (excepto en el caso normal). En este trabajo de tesis se aproxima su valor para cada modelo utilizando integración Quasi-Monte Carlo (Papageorgiou y Traub, 1997). Ver ecuación (8.8) en la siguiente sección.

8.5 Selección de Bayesiana de Modelos

Consideremos un experimento factorial 2^k sin réplicas, donde $n = 2^k$ es el total de observaciones, y con el cual se quiere investigar la posible significancia de los $2^k - 1 = n - 1$ efectos. Sea el conjunto de posibles modelos $M_0, M_1, M_2, \dots, M_{m=2^n-1}$ que se pueden formar con los $n - 1$ efectos, más el modelo sin efectos activos (M_0). Cada modelo tiene asociado un vector de parámetros $\boldsymbol{\theta}_i^t = (\beta_0, \beta_1, \dots, \beta_{t_i})$. Sea \mathbf{y} el vector de observaciones. Si la distribución de \mathbf{y} dado el modelo se denota por $f(\mathbf{y} | M_i, \boldsymbol{\theta}_i)$, la probabilidad *a priori* de M_i por $p(M_i)$ y la densidad *a priori* de $\boldsymbol{\theta}_i$ es $f(\boldsymbol{\theta}_i | M_i)$, entonces la densidad predictiva *a priori* de \mathbf{y} , dado el modelo M_i , está dada por

$$f(\mathbf{y} | M_i) = \int_{\Theta_i} f(\mathbf{y} | M_i, \boldsymbol{\theta}_i) f(\boldsymbol{\theta}_i | M_i) d\boldsymbol{\theta}_i \quad (8.5)$$

donde Θ_i es el conjunto de posibles valores de $\boldsymbol{\theta}_i$. La probabilidad posterior del modelo M_i , dados los datos \mathbf{y} , está dada por

$$p(M_i | \mathbf{y}) = \frac{p(M_i) f(\mathbf{y} | M_i)}{\sum_{i=0}^m p(M_i) f(\mathbf{y} | M_i)}. \quad (8.6)$$

Note que esta fórmula no es más que una generalización de la fórmula dada en la ecuación (8.2) para el caso de m modelos.

Para calcular la probabilidad posterior de que un efecto individual T_j está activo se acumulan las probabilidades posteriores de todos los modelos en los que aparece ese efecto, esto es

$$P_j = \sum_{M_i: T_j \text{ es activo}} p(M_i | \mathbf{y}). \quad (8.7)$$

Para aplicar este procedimiento se requiere el cálculo de la densidad predictiva *a priori* $f(\mathbf{y} | M_i)$, lo que implica resolver una integral múltiple en el espacio de parámetros, lo que muchas veces solo se puede hacer por métodos numéricos. Generalmente los métodos numéricos usados para resolver este tipo de integrales se refieren a técnicas de simulación Monte Carlo. En este trabajo usamos Quasi-Monte Carlo para

aproximar dicha integral en los modelos estudiados, lo que implica utilizar secuencias de baja discrepancia para evaluar el integrando en lugar de números pseudo aleatorios (Niederreiter, 1992). Esto significa que se aproxima la integral como el valor esperado de una función de $\boldsymbol{\theta}_i$. Se obtienen N valores de $\boldsymbol{\theta}_i$ a partir de su distribución *a priori*, se evalúan en la función de $\boldsymbol{\theta}_i$ de interés y se promedia, es decir,

$$\widehat{E}[f(\mathbf{y} | M_i, \boldsymbol{\theta}_i)] = \frac{1}{N} \sum_{j=1}^N f(\mathbf{y} | M_i, \boldsymbol{\theta}_i). \quad (8.8)$$

El problema es que esta integración se tendría que hacer para cada modelo, que por ejemplo en el caso de 15 efectos de interés se tienen $2^{15} = 32768$ modelos posibles. Se puede ver que es posible obtener resultados razonables considerando sólo los modelos "pequeños", que de cumplirse el principio de escasez de efectos serían los modelos más factibles.

8.6 Caso Normal

En el caso normal sí es posible llegar a una expresión cerrada de la predictiva *a priori* $f(\mathbf{y} | M_i)$, como lo muestra Meyer y Box (1992). En este caso la densidad de \mathbf{y} dado el modelo M_i está dada por

$$f(\mathbf{y} | M_i, \sigma, \boldsymbol{\theta}_i) \propto \frac{1}{\sigma^n} \exp \left[-(\mathbf{y} - \mathbf{X}_i \boldsymbol{\theta}_i)^t (\mathbf{y} - \mathbf{X}_i \boldsymbol{\theta}_i) / 2\sigma^2 \right].$$

Las densidades *a priori* de los coeficientes $\beta_1, \beta_2, \dots, \beta_{t_i}$ se toman como normales independientes con media cero y varianza $\gamma^2 \sigma^2$, donde el parámetro conocido γ^2 captura la magnitud del efecto relativa al ruido experimental. El hecho de que sean normales con media cero hace que de antemano no se tenga preferencia por algún signo del coeficiente o efecto. De manera que

$$f(\boldsymbol{\theta}_i | M_i, \sigma) \propto \frac{1}{\gamma^{t_i} \sigma^{t_i}} \exp \left(-\boldsymbol{\theta}_i^t \Gamma_i \boldsymbol{\theta}_i / 2\sigma^2 \right),$$

donde

$$\Gamma_i = \frac{1}{\gamma^2} \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{t_i} \end{pmatrix}.$$

Para el parámetro β_0 , que es la primera entrada del vector $\boldsymbol{\theta}_i$, así como para σ , se elige la *a priori* no informativa dada por

$$f(\beta_0, \sigma) \propto \frac{1}{\sigma}.$$

(Box y Tiao, 1973, págs. 49-51). Integrando respecto a los parámetros, la densidad predictiva *a priori* resulta de resolver la integral múltiple

$$\begin{aligned} f(\mathbf{y} | M_i) &= \int_0^\infty \int_{-\infty}^\infty \dots \int_{-\infty}^\infty \frac{1}{\sigma^n} \exp [-(\mathbf{y} - \mathbf{X}_i \boldsymbol{\theta}_i)^t (\mathbf{y} - \mathbf{X}_i \boldsymbol{\theta}_i) / 2\sigma^2] \frac{1}{\gamma^{t_i} \sigma^{t_i}} \times \\ &\quad \exp (-\boldsymbol{\theta}_i^t \Gamma_i \boldsymbol{\theta}_i / 2\sigma^2) \frac{1}{\sigma} d\beta_1 \dots d\beta_{t_i} d\sigma \\ &= \int_0^\infty \int_{-\infty}^\infty \dots \int_{-\infty}^\infty \frac{1}{\gamma^{t_i} \sigma^{n+1+t_i}} \exp [-((\mathbf{y} - \mathbf{X}_i \boldsymbol{\theta}_i)^t (\mathbf{y} - \mathbf{X}_i \boldsymbol{\theta}_i) + \boldsymbol{\theta}_i^t \Gamma_i \boldsymbol{\theta}_i) / 2\sigma^2] \\ &\quad \times \frac{1}{\sigma} d\beta_1 \dots d\beta_{t_i} d\sigma. \end{aligned}$$

Meyer y Box (1992) resuelven esta integral y llegan a que

$$f(\mathbf{y} | M_i) \propto \gamma^{-t_i} |\Gamma_i + \mathbf{X}_i^t \mathbf{X}_i|^{-1/2} \left(S(\hat{\boldsymbol{\theta}}_i) + \hat{\boldsymbol{\theta}}_i^t \Gamma_i \hat{\boldsymbol{\theta}}_i \right)^{-(n-1)/2}$$

donde

$$\begin{aligned} \hat{\boldsymbol{\theta}}_i &= (\Gamma_i + \mathbf{X}_i^t \mathbf{X}_i)^{-1} \mathbf{X}_i^t \mathbf{y} \\ S(\hat{\boldsymbol{\theta}}_i) &= \left(y - \mathbf{X}_i \hat{\boldsymbol{\theta}}_i \right)^t \left(y - \mathbf{X}_i \hat{\boldsymbol{\theta}}_i \right) \end{aligned}$$

Una vez calculada la integral se obtiene la posterior del modelo como

$$p(M_i | \mathbf{y}) = C \times p(M_i) f(\mathbf{y} | M_i)$$

donde C es la constante de normalización, que hace que las probabilidades sumen la unidad. La distribución del número de términos en el modelo se considera binomial con parámetro α , de modo que la probabilidad *a priori* del modelo es

$$p(M_i) \propto (\alpha / (1 - \alpha))^{t_i}.$$

Equivalentemente y para fines de cómputo es mejor calcular

$$p(M_i | \mathbf{y}) = C^* \frac{p(M_i) f(\mathbf{y} | M_i)}{p(M_0) f(\mathbf{y} | M_0)}$$

donde M_0 es el modelo constante (sin efectos activos). El utilizar esta última expresión evita en alguna medida la obtención de números demasiado grandes que rebasen la capacidad de cálculo del programa utilizado. Y es equivalente a la expresión anterior porque al dividir por la constante $p(M_0) f(\mathbf{y} | M_0)$ no cambia el resultado, puesto que al final todo se escala para que sume uno, de manera que esta contante es absorbida por C^* .

Esta última expresión está dada por

$$p(M_i | \mathbf{y}) = C^* \left(\frac{\alpha}{1 - \alpha} \right)^{t_i} \gamma^{-t_i} \frac{|\mathbf{X}_0^t \mathbf{X}_0|^{1/2}}{|\mathbf{\Gamma}_i + \mathbf{X}_i^t \mathbf{X}_i|^{1/2}} \left(\frac{S(\hat{\boldsymbol{\theta}}_i) + \hat{\boldsymbol{\theta}}_i^t \mathbf{\Gamma}_i \hat{\boldsymbol{\theta}}_i}{S(\hat{\boldsymbol{\theta}}_0)} \right)^{-(n-1)/2}. \quad (8.9)$$

Calculando $p(M_i | \mathbf{y})$ para los $m = 2^{n-1}$ modelos posibles, las probabilidades posteriores de los efectos se obtienen sumando las posteriores de todos los modelos que los contienen.

Esta última expresión puede interpretarse como el producto de dos factores: el primer factor

$$\left(\frac{\alpha}{1 - \alpha} \right)^{t_i} \gamma^{-t_i} \frac{|\mathbf{X}_0^t \mathbf{X}_0|^{1/2}}{|\mathbf{\Gamma}_i + \mathbf{X}_i^t \mathbf{X}_i|^{1/2}}$$

penaliza el incremento de variables en el modelo. A mayor cantidad de variables el valor de este factor es menor. El segundo factor

$$\left(\frac{S(\hat{\boldsymbol{\theta}}_i) + \hat{\boldsymbol{\theta}}_i^t \mathbf{\Gamma}_i \hat{\boldsymbol{\theta}}_i}{S(\hat{\boldsymbol{\theta}}_0)} \right)^{-(n-1)/2}$$

mide la bondad de ajuste del modelo. Mientras más pequeña es la suma de cuadrados $S(\hat{\boldsymbol{\theta}}_i)$ el valor de este factor es más grande. De esta manera, el producto de los dos factores que conforman la probabilidad posterior $p(M_i | \mathbf{y})$ tiende a favorecer los modelos con pocas variables (principio de escasez de efectos) que se ajustan adecuadamente a los datos.

Capítulo 9

Detección de Efectos Activos en Experimentos Factoriales con Respuesta Poisson

9.1 Poisson con liga identidad

La expresión dada por (8.9) es la forma cerrada de las probabilidades posteriores de los modelos en el caso normal. Pero como veremos, en el caso de Poisson no es posible llegar a una forma cerrada para la predictiva $f(\mathbf{y} | M_i)$ y será necesario resolver la integral resultante por métodos numéricos. Sea el vector de observaciones dado por $\mathbf{y}^t = [y_1, y_2, \dots, y_n]$, donde y_j es el número de defectos o eventos observados en el tratamiento j . Consideremos al modelo M_i con vector de parámetros $\boldsymbol{\theta}_i^t = [\beta_0, \beta_1, \dots, \beta_{i_i}]$ y la función liga identidad dada por $g(\mu_j) = \mu_j = \mathbf{x}_j^t \boldsymbol{\beta}$. La distribución de la observación y_j , dado el modelo M_i , está dada por

$$f(y_j | M_i, \boldsymbol{\theta}_i) = \frac{1}{y_j!} e^{-\mathbf{x}_j^t \boldsymbol{\theta}_i} (\mathbf{x}_j^t \boldsymbol{\theta}_i)^{y_j}.$$

Entonces la verosimilitud de los n datos es

$$f(\mathbf{y} | M_i, \boldsymbol{\theta}_i) = \prod_{j=1}^n \frac{1}{y_j!} e^{-\mathbf{x}_j^t \boldsymbol{\theta}_i} (\mathbf{x}_j^t \boldsymbol{\theta}_i)^{y_j}.$$

Considerando que las columnas de la matriz de diseño son ceros y unos, con excepción de la primera columna de unos, esta verosimilitud se reduce a

$$f(\mathbf{y} | M_i, \boldsymbol{\theta}_i) = \frac{e^{-n\beta_0}}{\prod_{j=1}^n y_j!} \prod_{j=1}^n (\mathbf{x}_j^t \boldsymbol{\theta}_i)^{y_j}.$$

Para los coeficientes del modelo M_i se propone una distribución *a priori* normal con media cero y varianza $\gamma^2 \sigma^2$, donde γ es un parámetro que se supone conocido y que permite capturar la magnitud de los efectos significativos relativa a los efectos inertes. Esto es, el hiperparámetro σ^2 tiene que ver con la variación de los coeficientes cuando no hay efectos activos y se elige de acuerdo a la información proporcionada por el experimentador. Mientras que el parámetro γ^2 tiene que ver con la variación adicional que tienen los coeficientes que están activos respecto a los inactivos.

Típicamente en el caso del modelo de Poisson el experimentador puede decir antes del experimento cuántos eventos espera observar en promedio en el experimento, y de aquí se deduce un valor de referencia σ^2 , usando la aproximación normal de la distribución de Poisson, y sabiendo que la varianza del efecto está dada por $\widehat{CM}_E / 2^{k-2}$. Note que en este caso no necesariamente el parámetro σ^2 se refiere a variación residual, sino que es un valor de referencia a partir del cual se propone un valor $\gamma^2 \sigma^2$ para la varianza de los β_i .

El cuadrado medio se deduce de la media de defectos que el experimentador cree observará en el experimento: considerando que la media declarada representa en sí misma un promedio de defectos que interesa eliminar, se toma esta media como el valor del contraste de referencia y entonces $\widehat{CM}_E = (\lambda)^2 / 2^k$. No olvidar además que los coeficientes son la mitad del efecto (Sección 2.1), de aquí que la varianza de referencia se toma como

$$var(\beta_i) = (\lambda)^2 / 2^{2k}.$$

El utilizar la distribución *a priori* normal con media cero para los coeficientes permite la posibilidad de efectos negativos y positivos sin preferir algún signo. Del parámetro β_0 , al ser un conteo de eventos, se sabe que es no negativo por lo que se le asigna una distribución *a priori* gama (también puede considerarse una uniforme, o una normal que no tome valores negativos) con media igual a λ , y desviación estándar obtenida de la varianza de referencia, que se deduce de la media de defectos declarada por el usuario. De acuerdo a esto, tenemos que la predictiva *a priori* es

$$f(\mathbf{y} | M_i) = \int_{\Theta_i} \frac{e^{-n\beta_0}}{\prod_{j=1}^n y_j!} \prod_{j=1}^n (\mathbf{x}_j^t \boldsymbol{\theta}_i)^{y_j} \frac{1}{(2\pi)^{t_i/2} (\gamma\sigma)^{t_i}} \times \\ e^{-(1/2\gamma^2\sigma^2) \sum_{k=1}^{t_i} \beta_k^2} \frac{1}{\Gamma(\delta)\varphi^\delta} \beta_0^{\delta-1} e^{-\beta_0/\varphi} d\boldsymbol{\theta}_i$$

donde los hiperparámetros σ^2 , γ , δ y φ se elucidan de acuerdo con la información dada por el usuario y a lo discutido en el párrafo anterior. La información *a priori* relativa a los t_i coeficientes β_i es el producto de t_i normales con la misma media y varianza. Por ejemplo, para el modelo sin efectos activos (M_0), la distribución predictiva *a priori* se reduce a

$$f(\mathbf{y} | M_0) = \int_0^\infty \frac{1}{\prod_{j=1}^n y_j!} e^{-n\beta_0} (\beta_0)^{\sum_{j=1}^n y_j} \frac{1}{\Gamma(\delta)\varphi^\delta} \beta_0^{\delta-1} e^{\beta_0/\varphi} d\beta_0 \\ = \frac{1}{\prod_{j=1}^n y_j!} \int_0^\infty e^{-n\beta_0} (\beta_0)^{\sum_{j=1}^n y_j} \frac{1}{\Gamma(\delta)\varphi^\delta} \beta_0^{\delta-1} e^{\beta_0/\varphi} d\beta_0.$$

9.2 Poisson con liga log

La liga log implica que la relación entre las variables explicativas y la media de la distribución en el j -ésimo punto experimental está dada por la función

$$g(\mu_j) = \log(\mu_j) = \mathbf{x}'\boldsymbol{\beta}.$$

Sea otra vez el vector de observaciones dado por $\mathbf{y}^t = [y_1, y_2, \dots, y_n]$. Consideremos el modelo M_i con vector de parámetros $\boldsymbol{\theta}_i^t = [\beta_0, \beta_1, \dots, \beta_{t_i}]$. La distribución de la observación y_j , dado el modelo M_i , en este caso está dada por

$$f(y_j | M_i, \boldsymbol{\theta}_i) = \frac{1}{y_j!} \exp\left(-e^{\mathbf{x}_j^t \boldsymbol{\theta}_i}\right) \left(\exp(\mathbf{x}_j^t \boldsymbol{\theta}_i)\right)^{y_j}.$$

Entonces la verosimilitud considerando los n datos es

$$\begin{aligned} f(\mathbf{y} | M_i, \boldsymbol{\theta}_i) &= \prod_{j=1}^n \frac{1}{y_j!} \exp\left(-e^{\mathbf{x}_j^t \boldsymbol{\theta}_i}\right) \left(\exp(\mathbf{x}_j^t \boldsymbol{\theta}_i)\right)^{y_j}. \\ &= \prod_{j=1}^n \frac{1}{y_j!} \exp\left(-e^{\mathbf{x}_j^t \boldsymbol{\theta}_i} + y_j \mathbf{x}_j^t \boldsymbol{\theta}_i\right) \\ &= \frac{1}{\prod y_j!} \exp\left(\sum_{j=1}^n \left(-e^{\mathbf{x}_j^t \boldsymbol{\theta}_i} + y_j \mathbf{x}_j^t \boldsymbol{\theta}_i\right)\right) \end{aligned}$$

Considerano otra vez distribuciones *a priori* normales con media cero para los parámetros β'_i s y gama para el parámetro β_0 , la distribución predictiva *a priori* resulta de resolver la integral

$$\begin{aligned} f(\mathbf{y} | M_i) &= \int_{\Theta_i} \frac{1}{\prod y_j!} \exp\left(\sum_{j=1}^n \left(-e^{\mathbf{x}_j^t \boldsymbol{\theta}_i} + y_j \mathbf{x}_j^t \boldsymbol{\theta}_i\right)\right) \times \\ &\quad \frac{1}{(2\pi)^{\frac{t_i}{2}} (\gamma\sigma)^{t_i}} e^{-(1/2\gamma^2\sigma^2) \sum_{k=1}^{t_i} \beta_k^2} \frac{1}{\Gamma(\delta)\varphi^\delta} \beta_0^{\delta-1} e^{\beta_0/\varphi} d\boldsymbol{\theta}_i. \end{aligned}$$

El valor del hiperparámetro σ se obtiene de la información *a priori* dada por el experimentador y el hiperparámetro γ se selecciona de acuerdo a la varianza adicional considerada para los efectos activos. La distribución *a priori* para β_0 se toma como una gama con media acorde a la información dada por el usuario (ver ejemplo más adelante).

9.3 Poisson con liga raíz cuadrada

En el caso de la liga raíz cuadrada, la relación entre el predictor lineal y la media de la distribución de Poisson está dada por la función

$$g(\mu_j) = (\mu_j)^{1/2} = \mathbf{x}'\boldsymbol{\beta}.$$

Usando esta función liga, la distribución de la observación y_j , dado el modelo M_i , está dada por

$$f(y_j | M_i, \boldsymbol{\theta}_i) = \frac{1}{y_j!} e^{-(\mathbf{x}_j^t \boldsymbol{\theta}_i)^2} (\mathbf{x}_j^t \boldsymbol{\theta}_i)^{2y_j}.$$

Entonces la verosimilitud de los n datos es

$$\begin{aligned} f(\mathbf{y}_j | M_i, \boldsymbol{\theta}_i) &= \prod_{j=1}^n \frac{1}{y_j!} e^{-(\mathbf{x}_j^t \boldsymbol{\theta}_i)^2} (\mathbf{x}_j^t \boldsymbol{\theta}_i)^{2y_j} \\ &= \frac{1}{\prod_{j=1}^n y_j!} e^{-\sum_{j=1}^n (\mathbf{x}_j^t \boldsymbol{\theta}_i)^2} \prod_{j=1}^n (\mathbf{x}_j^t \boldsymbol{\theta}_i)^{2y_j}. \end{aligned}$$

Otra vez, considerando como distribución *a priori* de los coeficientes del modelo M_i una distribución normal con media cero y varianza $\gamma^2 \sigma^2$, y una distribución gama para el parámetro β_0 , se obtiene la predictiva *a priori*

$$\begin{aligned} f(\mathbf{y} | M_i) &= \int_{\Theta_i} \frac{1}{\prod_{j=1}^n y_j!} e^{-\sum_{j=1}^n (\mathbf{x}_j^t \boldsymbol{\theta}_i)^2} \prod_{j=1}^n (\mathbf{x}_j^t \boldsymbol{\theta}_i)^{2y_j} \frac{1}{(2\pi)^{t_i/2} (\gamma\sigma)^{t_i}} \times \\ &\quad e^{-(1/2\gamma^2\sigma^2) \sum_{k=1}^{t_i} \beta_k^2} \frac{1}{\Gamma(\delta)\varphi^\delta} \beta_0^{\delta-1} e^{-\beta_0/\varphi} d\boldsymbol{\theta}_i \end{aligned}$$

Por ejemplo, para el modelo sin efectos activos (M_0), tenemos que

$$f(\mathbf{y} | M_0) = \int_0^\infty \frac{1}{\prod_{j=1}^n y_j!} e^{-16\beta_0^2 \sum_{j=1}^n y_j} \frac{1}{\Gamma(\delta)\varphi^\delta} \beta_0^{\delta-1} e^{\beta_0/\varphi} d\beta_0$$

9.4 Ejemplo, Regresión de Poisson

Este ejemplo corresponde al Ejemplo 7.5 de Myers *et al.* (2002 pág. 272) y también es el ejemplo 1 de Myers y Montgomery (1997). El experimento es un factorial 2^{9-5} de resolución III en el que se estudia el posible efecto de 9 factores en el número de defectos observados en el acabado de cierta parrilla para automóvil. Los datos obtenidos, así como cierta información acerca de la estructura de alias se muestra en la Tabla 9.1. La respuesta c son los defectos observados en cada combinación experimental y la última columna es una transformación raíz cuadrada modificada de Freeman y Tukey.

Análisis Frecuentista Si se analiza la respuesta transformada usando papel de probabilidad normal se encuentra que los efectos D , F y $BG + CJ + EH$ están activos. Myers y Montgomery (1997) analizan los datos usando un modelo de Poisson con liga \log y encuentran estos mismos efectos activos usando un análisis tipo Wald. Myers *et al.* (2001) rehacen el análisis usando ligas \log y raíz cuadrada. Usando papel de probabilidad semi-normal concluyen con la liga \log que están activos D , F y $BC + DF + GJ$ y con la liga raíz cuadrada llegan a que están activos D , F y $BG + CJ + EH$, siendo esta última conclusión la misma obtenida por Myers y Montgomery (1997), e igual a la que resulta de analizar la respuesta transformada usando un modelo normal.

Tabla 9.1. Ejemplo 1. Matriz de diseño, Defectos observados y Estructura de Alias

#	A	B	C	D	E	F	G	H	J	c	$(\sqrt{c} + \sqrt{c+1})/2$
1	-1	-1	-1	-1	1	-1	1	-1	1	56	7.52
2	1	-1	-1	-1	1	-1	-1	1	-1	17	4.18
3	-1	1	-1	-1	-1	1	1	-1	-1	2	1.57
4	1	1	-1	-1	-1	1	-1	1	1	4	2.12
5	-1	-1	1	-1	1	1	-1	1	1	3	1.87
6	1	-1	1	-1	1	1	1	-1	-1	4	2.12
7	-1	1	1	-1	-1	-1	-1	1	-1	50	7.12
8	1	1	1	-1	-1	-1	1	-1	1	2	1.57
9	-1	-1	-1	1	-1	1	1	1	1	1	1.21
10	1	-1	-1	1	-1	1	-1	-1	-1	0	0.50
11	-1	1	-1	1	1	-1	1	1	-1	3	1.87
12	1	1	-1	1	1	-1	-1	-1	1	12	3.54
13	-1	-1	1	1	-1	-1	-1	-1	1	3	1.87
14	1	-1	1	1	-1	-1	1	1	-1	4	2.12
15	-1	1	1	1	1	1	-1	-1	-1	0	0.50
16	1	1	1	1	1	1	1	1	1	0	0.50
$l_1 = A + BJ + CG$ $l_2 = B + AJ + DE$ $l_3 = C + AG + EF$ $l_4 = D + BE + GH$ $l_5 = E + BD + CF$ $l_6 = F + CE + HJ$ $l_7 = G + AC + DH$ $l_8 = H + DG + FJ$										$l_9 = J + AB + FH$ $l_{10} = AD + CH + EG$ $l_{11} = AE + FG + JD$ $l_{12} = AF + BH + EG$ $l_{13} = AH + BF + CD$ $l_{14} = BC + DF + GJ$ $l_{15} = BG + CJ + EH$	

El análisis que proponen Myers y Montgomery (1997) es el siguiente: primero ajustan el modelo sin intercepto y realizan pruebas ji-cuadradas para los 15 coeficientes del modelo. La salida de SPLUS para dicho análisis se muestra en la Tabla 9.2. Estos valores T elevados al cuadrado son los estadísticos ji-cuadrada reportados en dicho artículo, que detectan como significativos los efectos D , F y DF , aplicando el principio de herencia. Sin embargo, al rehacer el análisis para estos tres efectos incluyendo el intercepto se encuentra a DF no significativo, y posiblemente algún otro efecto ausente en este modelo reducido sea significativo. Para identificarlo, excluyen DF del modelo y se estima el modelo con los demás efectos incluyendo el intercepto y los resultados obtenidos se muestran en la Tabla 9.3. Concluyen de esta tabla que

los efectos significativos son D , F y BG , sin aclarar cómo es que eligen la interacción BG del grupo de alias $BG + CJ + EH$.

Tabla 9.2. Modelo sin intercepto usando liga log

Efecto.	Coef.	Error Std	T-valor
A	-0.570	4.062	-0.140
B	-0.558	4.062	-0.137
C	-0.644	4.062	-0.158
D	-2.003	0.173	-11.542
E	-0.225	4.062	-0.055
F	-2.191	0.158	-13.879
G	0.239	4.062	0.0590
H	0.544	4.062	0.134
J	0.428	4.062	0.105
$AD + CH + EG$	-0.141	4.062	-0.035
$AE + FG + JD$	0.630	4.062	0.155
$AF + BH + EG$	-0.228	4.062	-0.056
$AH + BF + CD$	-0.413	4.062	-0.102
$BC + DF + GJ$	-1.329	0.191	-6.964
$BG + CJ + EH$	-0.902	4.061	-0.222

Tabla 9.3. Modelo con intercepto usando liga log

<i>Efecto.</i>	<i>Coef.</i>	<i>ErrorStd</i>	<i>T - valor</i>
<i>Intercepto</i>	1.1929	0.178	6.677
A	-0.151	0.198	-0.762
B	-0.162	0.202	-0.801
C	-0.247	0.202	-1.224
D	-0.762	0.148	-5.151
E	0.159	0.203	0.785
F	-0.970	0.162	-5.975
G	-0.145	0.203	-0.715
H	0.148	0.202	0.731
J	0.063	0.205	0.307
$AD + CH + EG$	0.231	0.205	1.126
$AE + FG + JD$	0.233	0.202	1.155
$AF + BH + EG$	0.145	0.204	0.710
$AH + BF + CD$	-0.049	0.205	-0.239
$BG + CJ + EH$	-0.524	0.205	-2.561

Myers *et al.* 2002 analizan otra vez estos datos y utilizan gráficos de Daniel (coeficientes absolutos en papel normal) para discernir cuáles efectos están activos.

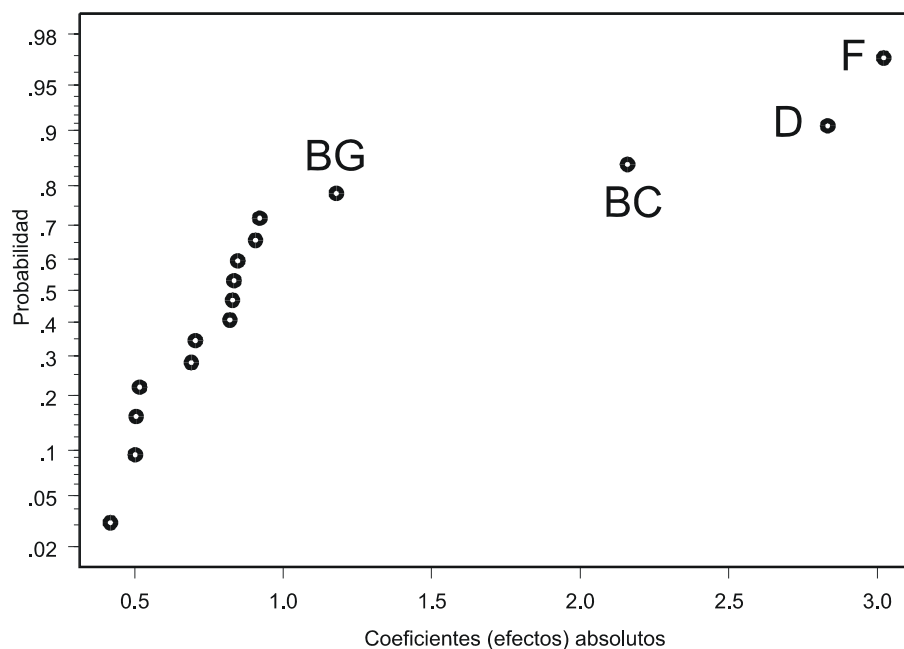


Figura 9.1: Coeficientes absolutos en papel de probabilidad normal, Poisson liga log

Encuentran que con la liga *log* los efectos activos parecen ser *F*, *D* y *BC*, y con la liga raíz cuadrada detectan a *F*, *D* y *BG* (ver Figuras 9.1 y 9.2). Concluyen que este último análisis es el correcto basados en el análisis de los residuos de ambos modelos.

Análisis Bayesiano Se analizan los datos del ejemplo considerando los modelos Poisson con ligas identidad, raíz cuadrada y log. Se supone que antes del experimento el usuario esperaba observar un promedio de 10 defectos por tratamiento. De aquí que para el caso de la liga identidad se toman como distribuciones *a priori*

$$\beta_0 \sim \text{gama}(256, 25.6)$$

$$\beta_i \sim \text{normal}(0, 0.39); (i = 1, 2, \dots, n - 1)$$

siendo esta última ($N(\text{media} = 0, \text{var} = 0.39)$) la distribución de referencia de β_i cuando no hay efectos activos, a partir de la cual se aplica el hiperparámetro de inflamiento de la varianza γ . Note que $(\lambda)^2/2^{2k} = (10^2/2^8) = 0.39$, que es también la

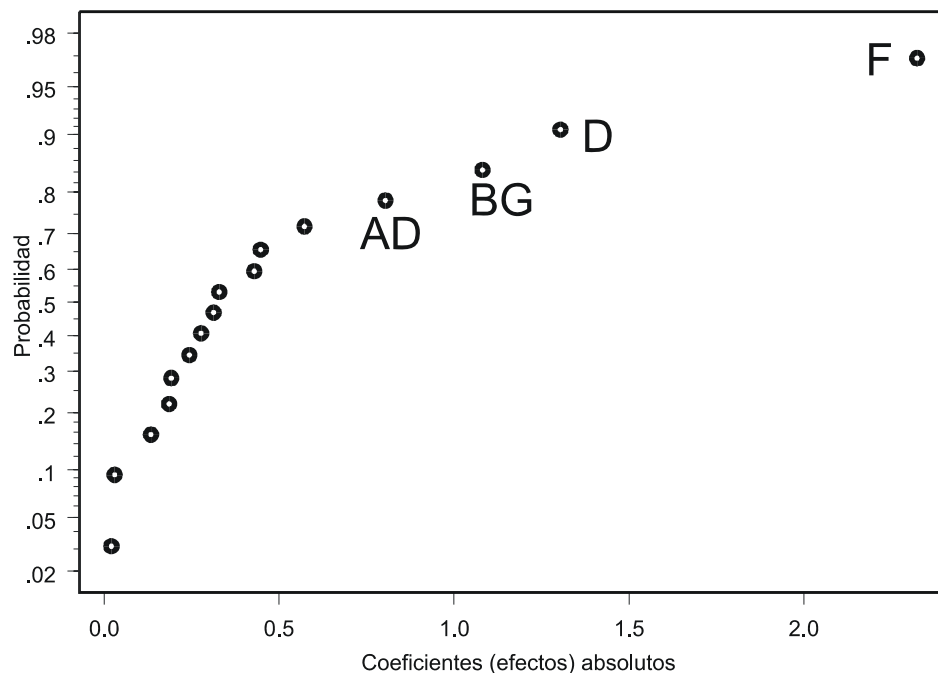


Figura 9.2: Coeficientes absolutos en papel de probabilidad normal, Poisson liga raíz varianza de la distribución *a priori* para β_0 . Esto es, la distribución *a priori* gama para β_0 tiene media 10 y varianza 0.39.

En el caso Poisson con liga log cuando no hay efectos activos se tiene que $\mu_i = e^{\beta_0}$, de aquí que se considere la distribución *a priori* $\beta_0 \sim \text{gama}(256, 111)$ y la distribución *a priori* de referencia para los β_i en el modelo es $\beta_i \sim \text{normal}(0, 0.02)$. Note que la distribución *gama*(256, 111) tiene aproximadamente media $\log(\lambda)$ y varianza $(\log \lambda)^2 / 256$, considerando $\lambda = 10$.

Finalmente, para el caso de la liga raíz cuadrada, se toma $\beta_0 \sim \text{gama}(255, 80)$ y a los β_i en el modelo se les asigna una distribución *a priori* *normal*(0, 0.039). En este caso la media de la distribución *gama* es cercana a $\sqrt{10}$ y su varianza es aproximadamente $(\sqrt{10})^2 / 256$.

Todos los análisis están basados en muestras de tamaño 200 para aproximar cada integral y se consideran modelos solo hasta con cuatro términos (1940 modelos). En todos los casos la información *a priori* de β_0 se basa en la supuesta media de 10 defec-

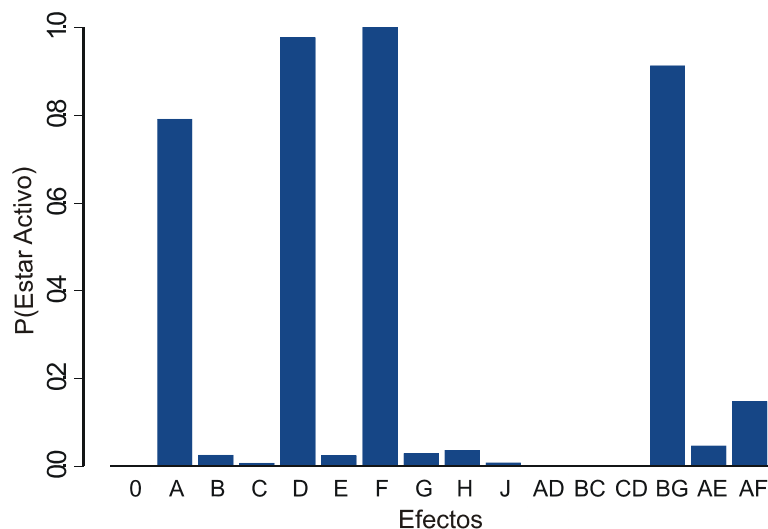


Figura 9.3: Probabilidades posteriores de estar activos, Poisson liga identidad

tos declarada por el usuario y considerando también la función liga correspondiente. Las varianzas de las distribuciones *a priori* de β_i mencionadas arriba son para un parámetro de inflamamiento de la varianza unitario ($\gamma = 1$). Funcionan razonablemente bien valores de γ en el rango $1 \leq \gamma \leq 2.5$, dadas las magnitudes de las varianzas de referencia de las que se parte, es decir, el rango puede cambiar si se definen de otra manera las varianzas de referencia.

Las probabilidades posteriores de los efectos considerando el modelo de Poisson con liga identidad se muestran en la Figura 9.3. Se detectan claramente como activos los efectos F , D y BG . Adicionalmente, aunque con menor probabilidad posterior destaca el efecto A .

En la Figura 9.4 se muestran las probabilidades posteriores de efectos que arroja el modelo Poisson liga log. Detecta claramente como significativos los efectos F , D y BG . También con esta liga destaca el efecto A .

Finalmente, en la Figura 9.5 se presentan resultados obtenidos con el modelo Poisson liga raíz cuadrada. También en este caso se logran detectar los efectos F , D y BG . Cabe señalar que con las tres ligas se detectan fundamentalmente los mismos

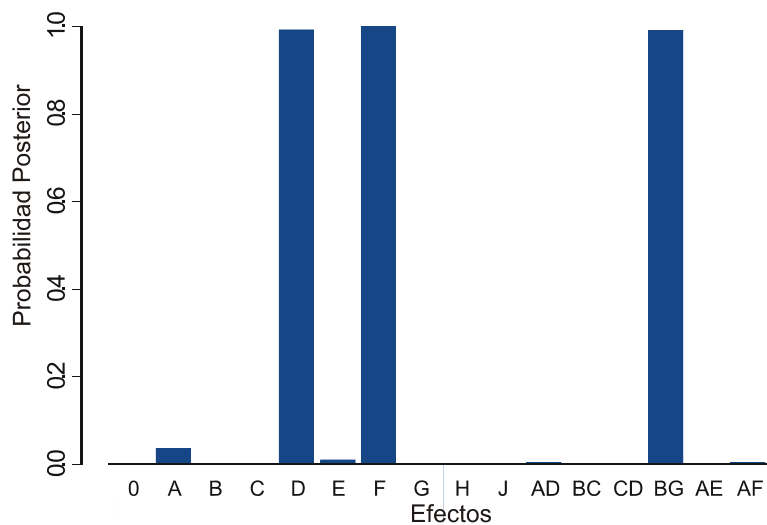


Figura 9.4: Probabilidades posteriores de estar activos, Poisson liga log

tres efectos más importantes: D , F y BG , y que son los mismos efectos detectados por Myers *et al.* (2002) usando técnicas frecuentistas.

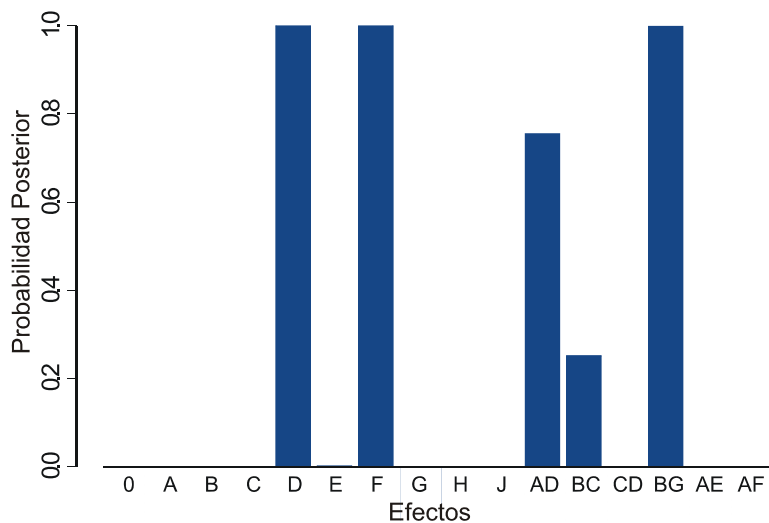


Figura 9.5. Probabilidades posteriores de estar activos, Poisson liga raíz

Una vez detectados los efectos activos se puede utilizar MCMC (Markov Chain Monte Carlo, ver por ejemplo, Robert, C. P. y Casella, G., 1999) para estimar los efectos o coeficientes del modelo sugeridos por este análisis, o bien se pueden utilizar las técnicas frecuentistas usuales (máxima verosimilitud).

En este ejemplo ocurre que las diferentes ligas utilizadas en el modelo Poisson pudieron detectar básicamente los mismos efectos activos, pero puede haber casos donde las discrepancias entre las diferentes ligas sean mayores. El problema de cuál función liga utilizar ya ha sido abordado en literatura, como un problema de selección Bayesiana de modelos usando factores de Bayes, donde se busca la función liga que más apoyo reciba de los datos observados (Czado y Raftery, 2005).

Capítulo 10

Detección de Efectos Activos en Experimentos Factoriales con Respuesta Binomial

10.1 Binomial con liga logística

Consideremos un experimento con respuesta binaria, donde y_1, y_2, \dots, y_n son los defectuosos o "éxitos" observados en muestras de tamaño n_j ($j = 1, 2, \dots, n$), procesadas en cada tratamiento. Sean $p_j = p(\mathbf{x}_j) = y_j/n_j$ la proporción observada en el tratamiento \mathbf{x}_j . El modelo de regresión logística se puede escribir como

$$p(\mathbf{x}_j) = \frac{1}{1 + e^{-\mathbf{x}_j^t \boldsymbol{\theta}}} = \frac{e^{\mathbf{x}_j^t \boldsymbol{\theta}}}{e^{\mathbf{x}_j^t \boldsymbol{\theta}} + 1}$$

donde $\boldsymbol{\theta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_k)$ es el vector de parámetros del predictor lineal. Alternativamente, el modelo se puede escribir de manera linealizada como

$$\ln \left[\frac{p(\mathbf{x}_j)}{1 - p(\mathbf{x}_j)} \right] = \mathbf{x}_j^t \boldsymbol{\theta}.$$

La distribución de y_j dado el modelo M_i con parámetro θ_i está dada por

$$\begin{aligned} f(y_j | M_i, \theta_i) &= \binom{n_j}{y_j} (p_i)^{y_j} (1 - p_i)^{n_j - y_j} = \binom{n_j}{y_j} \left(\frac{p_i}{1 - p_i} \right)^{y_j} (1 - p_i)^{n_j} \\ &= \binom{n_j}{y_j} \left(e^{\mathbf{x}_j^t \theta_i} \right)^{y_j} \left(\frac{1}{1 + e^{\mathbf{x}_j^t \theta_i}} \right)^{n_j}. \end{aligned}$$

La función de verosimilitud considerando las n observaciones está dada por

$$\begin{aligned} f(\mathbf{y} | M_i, \theta_i) &= \prod_{j=1}^n \binom{n_j}{y_j} \left(e^{\mathbf{x}_j^t \theta_i} \right)^{y_j} \left(\frac{1}{1 + e^{\mathbf{x}_j^t \theta_i}} \right)^{n_j} \\ &\propto \prod_{j=1}^n \left(e^{\mathbf{x}_j^t \theta_i} \right)^{y_j} \left(\frac{1}{1 + e^{\mathbf{x}_j^t \theta_i}} \right)^{n_j} \end{aligned}$$

Considerando como distribución *a priori* de los coeficientes del modelo M_i una distribución normal con media cero y varianza $\gamma^2 \sigma^2$, y una distribución normal para el parámetro β_0 , se obtiene la predictiva *a priori*

$$\begin{aligned} f(\mathbf{y} | M_i) &= \int_{\Theta_i} \prod_{j=1}^n \left(e^{\mathbf{x}_j^t \theta_i} \right)^{y_j} \left(\frac{1}{1 + e^{\mathbf{x}_j^t \theta_i}} \right)^{n_j} \frac{1}{(2\pi)^{t_i/2} (\gamma\sigma)^{t_i}} \times \\ &\quad e^{-(1/2\gamma^2\sigma^2) \sum_{k=1}^{t_i} \beta_k^2} \frac{1}{(2\pi)^{1/2}\sigma} e^{-(1/2\sigma^2)(\beta_0 - \mu)^2} d\theta_i, \end{aligned}$$

usando integración Quasi Monte Carlo, previamente elucidando los parámetros como se describe a continuación.

Información *a priori* En un experimento con respuesta binaria, generalmente el investigador puede anticipar el porcentaje de resultados favorables que espera observar en el estudio. Sea esta proporción p_0 . Con esta proporción se deducen los parámetros para las distribuciones *a priori* de los coeficientes β_0 y β_i ($i = 1, 2, \dots, 2^k - 1$) de la siguiente manera: Considerando la transformación logit calculamos $\ln(p_0/(1 - p_0))$, y tomamos este valor como el contraste de referencia y al mismo tiempo como el valor esperado *a priori* de β_0 . Esto es, este valor determina la media de la distribución normal que se asigna al parámetro β_0 , pero falta determinar su variabilidad. Se propone una distribución normal dado que en regresión logística el parámetro β_0 puede tomar

valores positivos y negativos, al estar en la escala de $\ln(p_0/(1 - p_0))$. La variabilidad de referencia *a priori* se determina considerando el valor de la transformación logit de p_0 como el valor de un contraste de referencia, cuya varianza estaría dada por

$$\frac{(\ln(p_0/(1 - p_0)))^2}{2^{2k}}.$$

Entonces, para el parámetro β_0 se propone una distribución normal con media $\ln(p_0/(1 - p_0))$ y varianza $(\ln(p_0/(1 - p_0)))^2/2^{2k}$ y para los β_i 's se propone una distribución normal con media cero con varianza inflada a partir de esta varianza de referencia $(\ln(p_0/(1 - p_0)))^2/2^{2k}$ de los efectos inactivos. Ver siguiente ejemplo.

10.2 Ejemplo, Regresión Logística

Este ejemplo (Myers, *et al.* 2002, pág. 116) es sobre la supervivencia de espermatozoides en un banco de esperma, sujetos al efecto de tres factores: citrato de sodio (A), glicerol (B) y tiempo de equilibrio (C). La respuesta es el número de muestras que sobreviven, donde esto significa que la muestra tiene habilidad para fecundar. Se utilizaron 50 muestras en cada punto de un diseño factorial 2^3 y se registró el número de las que sobrevivieron. El interés del estudio es conocer como afectan estos factores a las muestras de espermatozoides, considerando todos los efectos posibles. Los datos se muestran en la Tabla 10.1

A (Citrato de Sodio)	B (Glicerol)	C (Tiempo de Equilibrio)	Y (muestras sobrevivientes)
-1	-1	-1	34
1	-1	-1	20
-1	1	-1	8
1	1	-1	21
-1	-1	1	30
1	-1	1	20
-1	1	1	10
1	1	1	25

En el análisis frecuentista los autores encuentran, usando inferencia tipo Wald, que los efectos activos son B (glicerol) y AB (interacción de citrato de sodio con glicerol).

Inspeccionando los datos se puede notar que cuando los factores A y B se encuentran en los mismos niveles ($(-, -)$ y $(+, +)$) ocurre la mayor sobrevivencia de las muestras, y por el contrario, cuando se encuentran en niveles opuestos la sobrevivencia es menor.

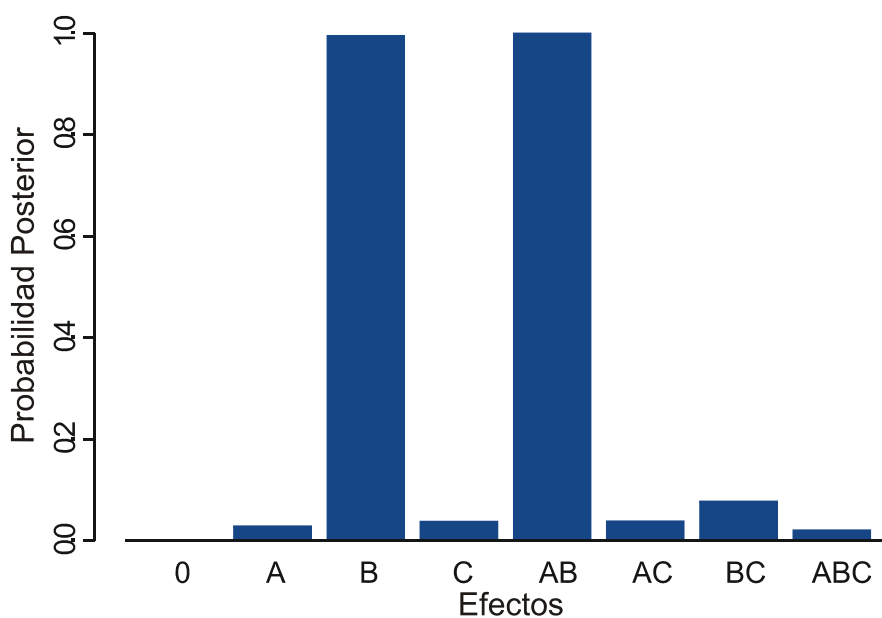
En este caso la información *a priori* se dedujo de la siguiente manera. Pensando que el experimentador espera una supervivencia de $p_0 = 0.32$, encontramos que las distribuciones *a priori* son las siguientes:

$$\beta_0 \sim \text{Normal}(\text{media} = \ln(.32/.68) = -0.75, \text{varianza} = 0.0088)$$

$$\beta_i \sim \text{Normal}(\text{media} = 0, \text{varianza} = \gamma^2 0.0088),$$

donde γ es el parámetro de inflamiento de la varianza para los efectos activos. La Figura (10.1) resulta de un inflamiento de la varianza de $\gamma = 5$ y $\alpha = 0.2$ como probabilidad *a priori* de efecto activo.

El análisis Bayesiano se resume en el gráfico de la Figura (10.1) en la cual se destacan los efectos B y AB con probabilidades posteriores de estar activos prácticamente unitarias, mientras el resto de los efectos tienen probabilidades posteriores muy pequeñas. Este resultado coincide con el análisis frecuentista.



Probabilidades posteriores de estar activo, binomial liga logit

Capítulo 11

Detección de Efectos Activos en Experimentos Factoriales con Respuesta Gama

Consideremos un experimento con respuesta continua no negativa donde la varianza no es constante, sino que es proporcional al cuadrado de la media. Este es el caso donde puede aplicarse la regresión gama en el contexto de modelos lineales generalizados, misma que asume la distribución gama en cada punto experimental. Otra alternativa para analizar este tipo de experimento es utilizar la transformación logaritmo natural ($\ln(y)$) de la respuesta, la cual estabiliza la varianza, y se analiza la respuesta transformada como el caso normal (Box y Meyer, 1993). En este caso implícitamente se estaría suponiendo una distribución lognormal en cada punto del experimento. Ambas soluciones Bayesianas pueden dar resultados equiparables dado que las distribuciones gama y lognormal son prácticamente indistinguibles cuando se tienen pocos datos (Hamada y Nelder, 1997).

11.1 Regresión gama, liga inversa

Sea el vector de observaciones dado por $\mathbf{y}^t = [y_1, y_2, \dots, y_n]$, donde y_j es el valor de la respuesta obtenido en el tratamiento j . Sea el modelo M_i con vector de parámetros $\boldsymbol{\theta}_i^t = [\beta_0, \beta_1, \dots, \beta_{t_i}]$ y la función liga inversa dada por $g(\mu_j) = 1/\mu_j = \mathbf{x}_j^t \boldsymbol{\beta}$ que es la liga canónica para la distribución gama, que en términos del parámetro λ_j de esta distribución se puede escribir como $g(\lambda_j) = 1/\lambda_j = \mathbf{x}_j^t \boldsymbol{\beta}$, puesto que el parámetro r no depende de la media. La distribución de la observación y_j , dado el modelo M_i , está dada por

$$f(y_j | M_i, \boldsymbol{\theta}_i, r) = \frac{1}{\Gamma(r)} (\mathbf{x}_j^t \boldsymbol{\theta}_i)^r e^{-\mathbf{y}_j \mathbf{x}_j^t \boldsymbol{\theta}_i} \mathbf{y}_j^{r-1}.$$

La función de verosimilitud expresada para los n datos es

$$f(\mathbf{y}_j | M_i, \boldsymbol{\theta}_i, r) = \left(\frac{1}{\Gamma(r)} \right)^n e^{-\sum \mathbf{y}_j \mathbf{x}_j^t \boldsymbol{\theta}_i} \prod_{j=1}^n (\mathbf{x}_j^t \boldsymbol{\theta}_i)^r \mathbf{y}_j^{r-1}.$$

Considerando como distribución *a priori* de los coeficientes del modelo M_i una distribución normal con media cero y varianza $\gamma^2 \sigma^2$, una distribución gama para el parámetro β_0 y una distribución uniforme para el parámetro r se obtiene la predictiva *a priori*

$$f(\mathbf{y} | M_i) = \int_{\Theta_i} \left(\frac{1}{\Gamma(r)} \right)^n e^{-\sum \mathbf{y}_j \mathbf{x}_j^t \boldsymbol{\theta}_i} \prod_{j=1}^n (\mathbf{x}_j^t \boldsymbol{\theta}_i)^r \mathbf{y}_j^{r-1} \frac{1}{(2\pi)^{t_i/2} (\gamma\sigma)^{t_i}} \times \\ e^{-(1/2\gamma^2\sigma^2) \sum_{k=1}^{t_i} (\beta_k^2 \in M_i)} \frac{1}{\Gamma(\delta)\varphi^\delta} \beta_0^{\delta-1} e^{-\beta_0/\varphi} \frac{1}{(\tau_2 - \tau_1)} d\boldsymbol{\theta}_i dr$$

donde los valores de los hiperparámetros γ , σ , δ y φ se elucidan a partir de la información proporcionada por el usuario antes de correr el experimento. En principio se pueden utilizar también una distribución normal para β_0 y una distribución gama para el parámetro r .

11.2 Regresión gama, liga log

Una liga no canónica muy recomendada en la regresión gama es la liga log, que evita estimar valores negativos de la respuesta. Con esta liga se tiene el modelo

$$\ln(\lambda_j) = \mathbf{x}'\boldsymbol{\beta}$$

con lo que $\lambda_j = \exp(\mathbf{x}'\boldsymbol{\beta})$ y la función de verosimilitud es de la forma

$$f(y_j | M_i, \boldsymbol{\theta}_i, r) = \frac{1}{\Gamma(r)} \left(e^{-\mathbf{x}_j^t \boldsymbol{\theta}_i} \right)^r e^{-y_j e^{-\mathbf{x}_j^t \boldsymbol{\theta}_i}} y_j^{r-1}.$$

La función de verosimilitud expresada para los n datos es

$$f(\mathbf{y}_j | M_i, \boldsymbol{\theta}_i, r) = \left(\frac{1}{\Gamma(r)} \right)^n e^{-\sum y_j e^{-\mathbf{x}_j^t \boldsymbol{\theta}_i}} \prod_{j=1}^n \left(e^{-\mathbf{x}_j^t \boldsymbol{\theta}_i} \right)^r y_j^{r-1}.$$

Considerando como distribución *a priori* de los coeficientes del modelo M_i una distribución normal con media cero y varianza $\gamma^2 \sigma^2$, una distribución gama para el parámetro β_0 y una distribución uniforme para el parámetro r se obtiene la predictiva *a priori*

$$f(\mathbf{y} | M_i) = \int_{\Theta_i} \left(\frac{1}{\Gamma(r)} \right)^n e^{-\sum y_j e^{-\mathbf{x}_j^t \boldsymbol{\theta}_i}} \prod_{j=1}^n \left(e^{-\mathbf{x}_j^t \boldsymbol{\theta}_i} \right)^r y_j^{r-1} \frac{1}{(2\pi)^{t_i/2} (\gamma\sigma)^{t_i}} \times \\ e^{-(1/2\gamma^2\sigma^2) \sum_{k=1}^{t_i} \beta_k^2} \frac{1}{\Gamma(\delta)\varphi^\delta} \beta_0^{\delta-1} e^{-\beta_0/\varphi} \frac{1}{(\tau_2 - \tau_1)} d\boldsymbol{\theta}_i dr$$

donde los valores de los hiperparámetros γ , σ , δ y φ se elucidan a partir de la información proporcionada por el usuario antes de correr el experimento. En cuanto a los parámetros τ_1 y τ_2 es muy difícil elucidarlos a partir de la información proporcionada por usuario, y entonces el parámetro de escala lo suponemos "conocido", en el sentido de que se ajusta al valor más grande posible cuidando que las integrales no exploten a infinito.

Dado que la distribución gama tiene dos parámetros la elucidación para especificar las distribuciones *a priori* es más complicada que en los casos anteriores, dado que

no basta una simple media para proponer valores adecuados de los hiperparámetros, sino que se requiere que el experimentador proporcione alguna información sobre la variabilidad de y como pueden ser los datos mayor y menor que esperaríamos observar en el experimento o bien la variación observada en la carta de control en el proceso actual. Otro problema es que aún con una elucidación apropiada de los hiperparámetros es posible que el resultado de la integral en cada punto sea cero o infinito, puesto que es demasiado sensible a los valores de los hiperparámetros. De aquí que no se puedan usar distribuciones *a priori* con soporte muy amplio dado que no se obtienen resultados.

11.3 Uso de la transformación $\log(y)$

Como se sabe, el caso de la regresión gama la transformación $\log(y)$ estabiliza la varianza de la respuesta. Si llamamos σ^2 a la varianza residual, la estimación por mínimos cuadrados arroja $(\mathbf{XX})^{-1}\sigma^2$ como la matriz de varianzas y covarianzas de los estimadores, donde σ^2 es el cuadrado del coeficiente de variación. Por otra parte, si se utiliza la liga log en un modelo lineal generalizado con distribución gama también se obtiene $(\mathbf{XX})^{-1}\sigma^2$ como la matriz de varianzas y covarianzas de los coeficientes estimados, donde σ^2 resulta ser el cuadrado del coeficiente de variación (ver Myers *et al.* (2002), pág. 171). De aquí que usando la transformación logaritmo de la respuesta con la regresión normal se puedan obtener frecuentemente resultados equiparables a la regresión gama en la detección de los efectos (coeficientes) activos. En particular, en el enfoque Bayesiano se puede utilizar el método de Box y Meyer (1993) con la respuesta transformada como lo hacemos en el siguiente ejemplo.

11.4 Ejemplo

Consideremos los datos del Ejemplo 1 de Lewis *et al.* (2001) que consisten de un experimento factorial 2^4 no replicado donde la respuesta estudiada es la tasa de avance de una taladradora. Los datos originales y la transformación logaritmo de éstos se muestran en la Tabla (11.1).

Tabla 11.1. Experimento factorial 2^4

#	A	B	C	D	y	$\ln(y)$
1	-1	-1	-1	-1	1.68	0.519
2	1	-1	-1	-1	1.98	0.683
3	-1	1	-1	-1	3.28	1.188
4	1	1	-1	-1	3.44	1.235
5	-1	-1	1	-1	4.98	1.605
6	1	-1	1	-1	5.70	1.740
7	-1	1	1	-1	9.97	2.299
8	1	1	1	-1	9.07	2.205
9	-1	-1	-1	1	2.07	0.727
10	1	-1	-1	1	2.44	0.892
11	-1	1	-1	1	4.09	1.408
12	1	1	-1	1	4.53	1.511
13	-1	-1	1	1	7.77	2.050
14	1	-1	1	1	9.43	2.244
15	-1	1	1	1	11.75	2.464
16	1	1	1	1	16.30	2.791

En el análisis frecuentista de estos datos usando la transformación logaritmo se encontraron tres efectos activos, a saber: B , C y D (ver Montgomery 2001, pág. 259). Y usando la regresión gama con función liga log se detectan también como efectos activos a B , C y D (ver Lewis *et al.* 2001).

Análisis Bayesiano Aquí se presenta el resultado de analizar estos datos usando el método Bayesiano con el modelo gama liga log y también utilizando la respuesta transformada con el método Bayesiano de Box y Meyer (1993), caso normal. Los resultados obtenidos se resumen en las figuras 11.1 y 11.2, respectivamente. En la Figura 11.1 se observan cuatro efectos activos que son A , B , C y D , donde el

efecto A no se había detectado significativo con los métodos frecuentistas, aunque es el siguiente efecto en magnitud usando la transformación logaritmo o la transformación de rangos de Aguirre-Torres (1993). En la Figura 11.2, que corresponde a la transformación logaritmo caso normal Bayesiano, sólo se detectan claramente los efectos activos B , C y D , y el efecto A no alcanza a superar la probabilidad posterior de 0.5 para declararse activo.

La elucidación de los hiperparámetros se propone de la siguiente manera: suponemos que el experimentador proporciona la media, el valor más pequeño y el valor más grande de la respuesta que esperaría observar en el experimento. Entonces el valor global de λ se propone como $(dato\ menor)/(dato\ mayor)$. Considerando que la función liga es $\lambda_i = \exp(\beta_0)$ en el caso del modelo nulo, se propone para β_0 una distribución *a priori* normal con media $\ln(\lambda)$ y varianza $(\ln(\lambda)/2)^2 / 256$. Esta varianza es la misma que se propone para los coeficientes β_i . En resumen, las distribuciones *a priori* para β_0 y β_i quedan como

$$\begin{aligned}\beta_0 &\sim Normal(\ln(\lambda), (\ln(\lambda)/2)^2 / 256) \\ \beta_i &\sim Normal(0, (\ln(\lambda)/2)^2 / 256)\end{aligned}$$

mientras que r lo suponemos conocido, o mejor dicho se utiliza como parámetro de ajuste que permite llevar al resultado de la integral del modelo nulo (que tiene sólo el término β_0) a un valor grande, cuidando que las integrales de los demás modelos no vayan a explotar a infinito. Todos los modelos se referencian al modelo nulo lo que evita que la suma de las integrales necesarias para obtener las posteriores de los efectos no se disparen a infinito. Específicamente, en este ejemplo las distribuciones *a priori* consideradas son $\beta_0 \sim Normal(-2, 0.0042)$ y $\beta_i \sim Normal(0, 0.0042)$.

Cabe agregar que se utiliza la función liga $\lambda_i = \exp(\mathbf{x}'\boldsymbol{\beta})$, probabilidad *a priori* de efecto activo $\alpha = 0.2$ y $\gamma = 1.25$ (factor de inflamamiento de la desviación estándar de referencia). Se aproximaron las integrales usando 400 simulaciones Quasi Monte-Carlo (Krykova, 2003) con secuencias de Halton usando parámetro de escala $r = 1$

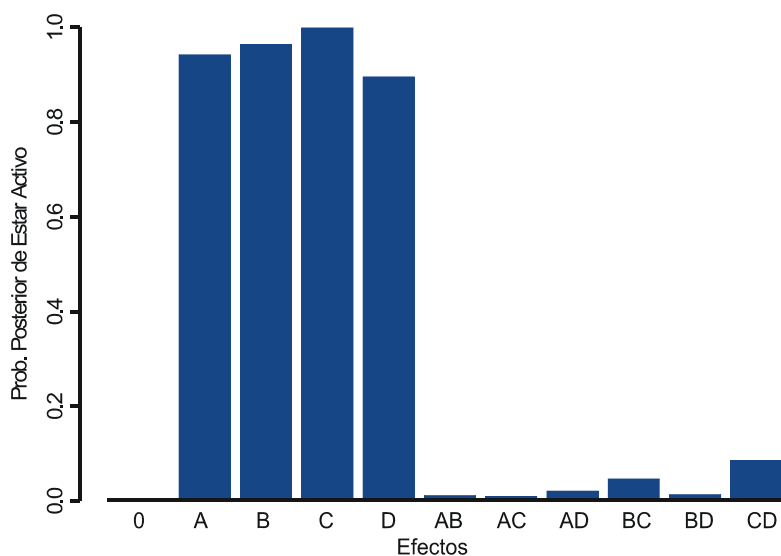


Figura 11.1: Regresión gama, liga log

con todos los posibles modelos que se pueden construir hasta con cuatro términos, sin considerar las interacciones de alto orden (triples en adelante).

En cuanto al análisis de la transformación logaritmo de la Figura (11.2), éste resulta de aplicar el método de Box y Meyer (1993) usando $\alpha = 0.2$ y $\gamma = 2.5$ (factor de inflamiento de la desviación estándar del error). Este método se trabaja con todos los posibles modelos considerando hasta la interacción cuádruple.

Como se menciona arriba, hay una discrepancia importante en cuanto al efecto A en las figuras 11.1 y 11.2. Lo que sucede es que el efecto A está, en cuanto a su tamaño, en una zona donde eventualmente con alguna selección de los parámetros (factor de inflamiento de la varianza de referencia y parámetro de escala) del método Bayesiano se podría concluir que está activo. Si se inspeccionan los datos de la Tabla 11.1 se observa en combinaciones vecinas una tendencia de menor tasa de avance en el nivel bajo del factor A respecto al nivel alto del mismo factor, con una única excepción, por lo que podría ser razonable que resultara activo. Otro aspecto determinante en el método Bayesiano propuesto es el número de simulaciones Quasi

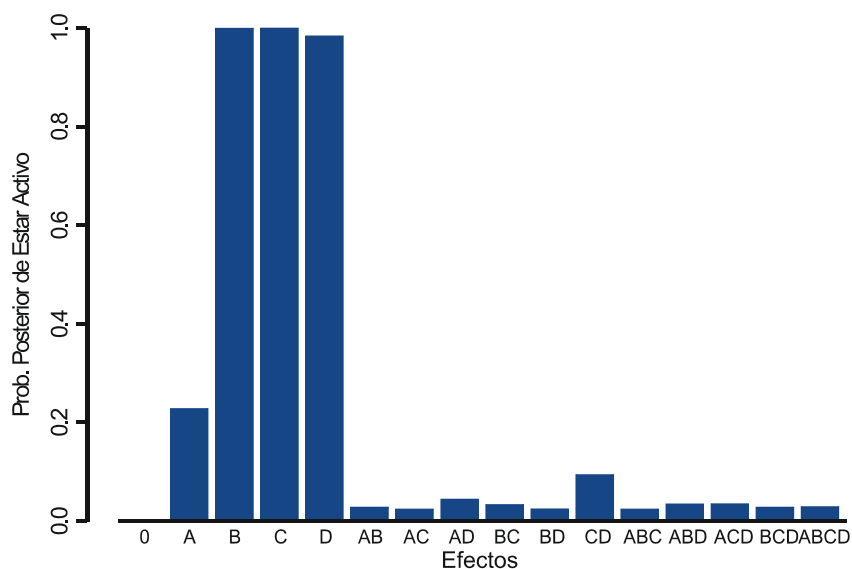


Figura 11.2: Método de Box y Meyer (1993) aplicado a $\log(y)$

Monte-Carlo utilizadas para calcular las integrales. Se puede mostrar que con una "mejor estimación" de las integrales se tiende a detectar activos solamente los efectos B , C y D con lo que la conclusión sería la misma con ambos métodos.

Para aclarar lo dicho al final del párrafo inmediato anterior se decide hacer un pequeño estudio de estabilidad de la solución Bayesiana al mover los parámetros: factor de inflamiento de la desviación estándar (γ), parámetro de escala de la distribución gama (r) y número de simulaciones Quasi Monte-Carlo usando secuencias de Halton. Se probaron cuatro valores del factor de inflamiento (1.0, 2.5, 5.0, 7.5), dos niveles del parámetro de escala (0.05, 40) y tres números de simulaciones (200, 600, 1000). Los valores del parámetro de escala se escogen básicamente cuidando que las integrales no se vayan todas a cero ($r = 0.05$) ni que se disparen a infinito ($r = 40$). Este último ($r = 40$) es más congruente con el hecho de que la media de la distribución gama es $\mu = r\lambda$, y es el valor que muestra un mejor desempeño. Los resultados de las $4 \times 3 \times 2$ combinaciones de parámetros se muestran en las figuras 11.3 y 11.4. Se concluye que el parámetro de escala $r = 40$ da soluciones más estables que el parámetro de escala

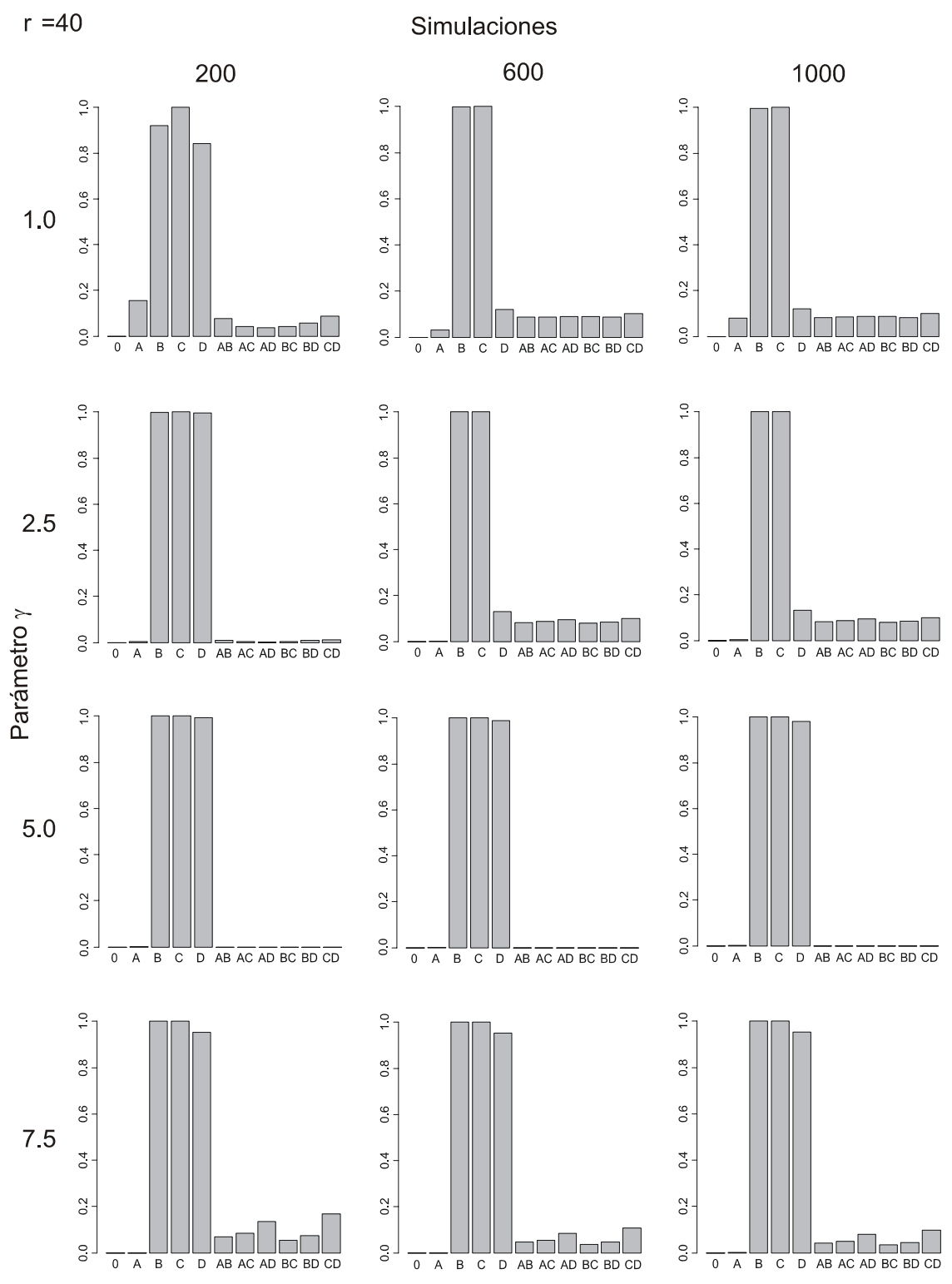


Figura 11.3: Soluciones con $r = 40$

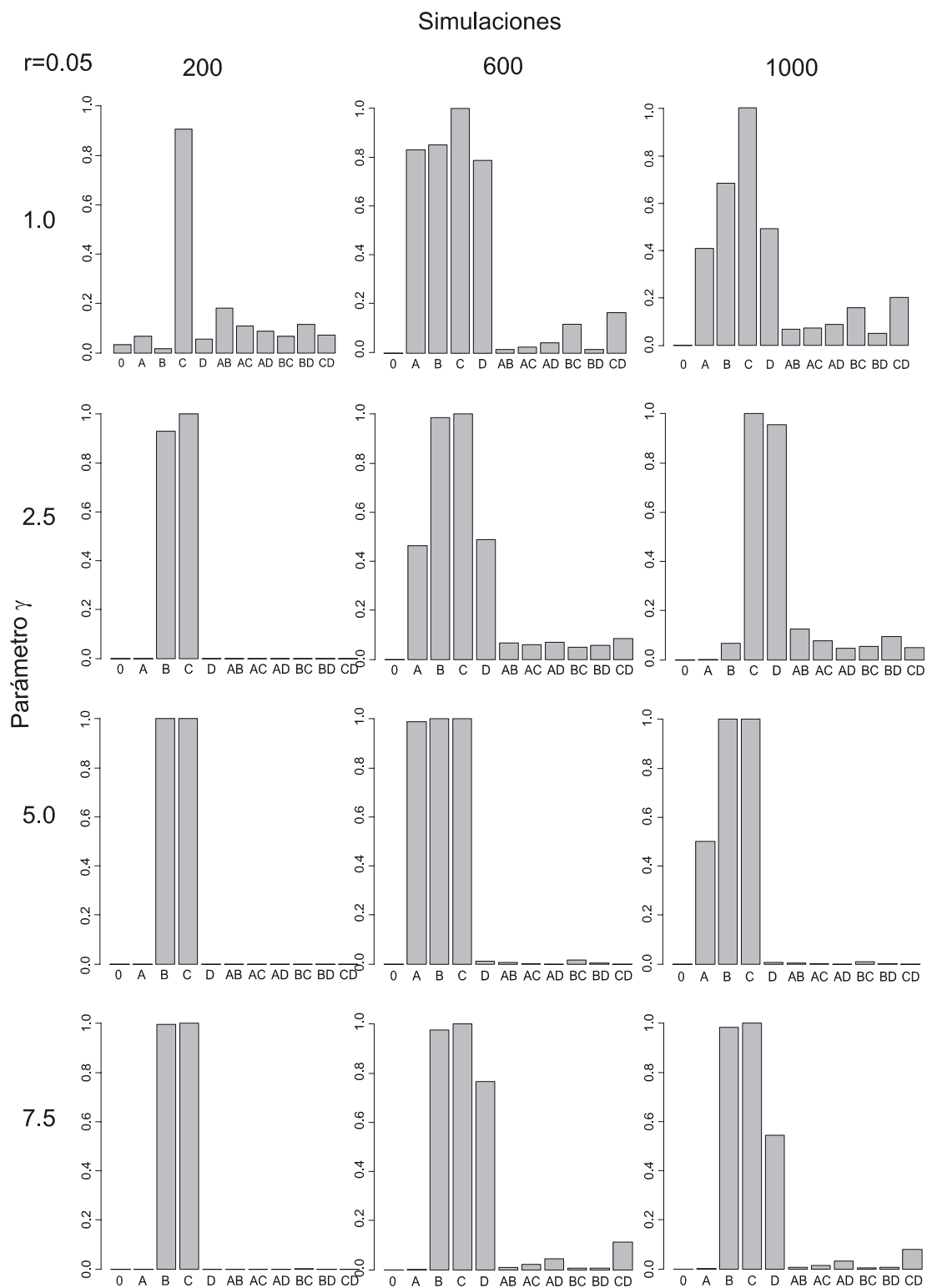


Figura 11.4: Soluciones con $r = 0.05$

$r = 0.05$, y que en el primer caso se detectan los efectos activos B , C y D cuando el factor de inflamamiento es mayor o igual que 5. Note en la Figura 11.3 que las soluciones con 600 y 1000 simulaciones son básicamente iguales.

En la Figura 11.4 se observa menos consistencia que no acaba de componerse ni con más simulaciones ni incrementando el parámetro de inflamamiento. En resumen, se concluye de ambas figuras que con una selección adecuada de los parámetros se llega a que los efectos activos son B , C y D , los mismos que detecta el análisis frecuentista. El parámetro de escala r se recomienda lo más grande posible evitando que las integrales exploten a infinito. Realizar al menos 600 simulaciones con factor de inflamamiento de la desviación estándar $\gamma \geq 5$. Evidentemente habría que analizar varios ejemplos para evaluar la pertinencia estas recomendaciones.

Capítulo 12

Conclusiones

En este trabajo de tesis se estudian y proponen métodos para el análisis de experimentos factoriales no-replicados en dos vertientes: la primera es cuando la respuesta normal puede estar contaminada por datos anómalos (Capítulos 1, 2,..., 7) y la segunda cuando la respuesta, debido a su naturaleza, no sigue una distribución normal pero no se consideran datos atípicos (Capítulos 8, 9, 10 y 11). Todos los métodos se enfocan fundamentalmente a la detección de los efectos activos en este tipo de experimentos.

Para la primer vertiente o problema de analizar experimentos con posible contaminación por datos atípicos ya existían los métodos propuestos por Box y Meyer (1987) y Aguirre Torres y Pérez Trejo (2001), el primero basado en técnicas Bayesianas y el segundo basado en el análisis de la transformación de rangos de los datos usando una variante del método de Bensi (1989).

En este trabajo de tesis se proponen varios métodos nuevos que dan resultados mejores que el método de rangos, al menos en los escenarios estudiados. El método Bayesiano no se incluyó en el estudio de simulación porque su aplicación a cada conjunto de datos es demasiado tardada. Los métodos propuestos a los que nos referimos, así como los métodos existentes, se resumen en el cuadro sinóptico de la Tabla 12.1. En la segunda columna de esta tabla se hace una breve descripción de

cada método; en la tercera se da el valor crítico utilizado en la prueba de normalidad el cual se obtuvo por simulación considerando un error global $EEER = 5\%$. Cuando se rechaza la normalidad se utiliza la prueba de dispersión de cuartos para decidir cuáles efectos están activos. Finalmente, en la última columna se señala si los métodos pueden o no identificar a los datos atípicos.

Tabla 12.1. Cuadro Sinóptico de Métodos Estudiados y Propuestos

Método	Descripción Breve	Valor α Crítico	Det. Atí.
1. Método de Rangos de AT&PT (2001)	Analiza la transformación de rangos de la respuesta	0.033	No
2. Método de Rangos Modificados	Rangos interiores con ajuste proporcional a los espacios entre los efectos ordenados	0.045	No
3. Método de Espacios Ajustados	Ajusta los efectos ordenados cerrando los huecos debidos a posibles datos anómalos	0.001	No
4. Mét. de Datos Anómalos Reestimados	Identifica y reestima los datos anómalos como si fueran observaciones faltantes	0.033	Sí
5. Método de Regresión Robusta I	Usa MM-estimación de forma iterativa a partir de un modelo selec. por muestreo	0.034	No
6. Método de Regresión Robusta II	Usa MM-estimación de forma iterativa a partir de un modelo sel. por regresión L1	0.047	No
7. Método Bayesiano de Box y Meyer (1987)	Calcula las probabilidades posteriores de efecto activo y de dato anómalo	P=0.89	Sí

Del estudio de simulación del Capítulo 7 se desprende que el método 6 basado en técnicas de regresión robusta es de los más destacados para detectar los efectos activos en los diferentes escenarios considerados. En particular, este método llega casi a duplicar la potencia de detección de los efectos pequeños considerados en la simulación en relación al método de rangos (Tabla 7.3) y en el escenario más contaminado. Otro método destacado es el método de datos anómalos reestimados (método 4) que permite detectar los datos anómalos y los corrige antes de estimar y pasar a detectar los efectos activos. Esto es interesante porque una vez detectados y corregidos los datos anómalos se podría utilizar prácticamente cualquiera de los métodos mencionados por Hamada y Balakrishnan (1998), con lo que se abren bastantes las posibilidades para encontrar nuevos métodos robustos.

En esta primera vertiente, que tiene que ver con la contaminación por datos anómalos, habría varias líneas que valdría la pena investigar en el futuro entre las que se encuentran:

1. Calibrar el método de Box y Meyer (1987) y luego compararlo de manera justa con los mejores métodos frecuentistas propuestos.
2. Explorar otros métodos híbridos. Por ejemplo, se podría utilizar alguna otra prueba en lugar de la prueba de normalidad que utiliza Bensi (1989), y para decidir cuáles efectos están activos también existirían otras posibilidades.
3. Ampliar los escenarios donde los métodos son comparados. Se sabe por estudios publicados (H&B, (1998) y Chen (2003)) sobre los métodos en el caso normal sin datos atípicos, que su desempeño depende del número de efectos activos que se consideran en la simulación, así como el tamaño de dichos efectos, o si todos son iguales o diferentes. También puede ser interesante considerar otros tamaños del error global EER o buscar controlar el error individual IER .
4. Investigar métodos robustos para la situación donde no necesariamente aplica el principio de escasez de efectos.
5. Utilizar otras medidas de desempeño para comparar los métodos (ver por ejemplo Chen, 2003).
6. Investigar estrategias secuenciales, que son aquellos en los que se va decidiendo la significancia efecto por efecto.
7. Afinar y calibrar los métodos nuevos más prometedores encontrados en este trabajo para que puedan aplicarse sin problemas a diferentes experimentos factoriales con 8 o más de 16 corridas.

En la segunda vertiente de esta tesis se exploró la factibilidad de utilizar Selección Bayesiana de modelos en el contexto de modelos lineales generalizados para detectar efectos activos en experimentos factoriales con respuesta no normal. Se estudiaron los casos Poisson, Binomial y Gama, además del caso Normal propuesto por Box y Meyer

(1993). En el caso Poisson se consideraron diferentes ligas, mismas que mostraron consistencia en el ejemplo analizado, en el cual todas las ligas detectaron los mismos efectos activos.

En general los métodos Bayesianos propuestos muestran bastante potencial para la detección de efectos activos en experimentos factoriales no replicados con respuesta no normal, y son una alternativa viable a los métodos frecuentistas como la inferencia tipo Wald y la inferencia basada en verosimilitud, ya que en experimentos no-replicados o fraccionados el tamaño de muestra es necesariamente pequeño. El problema con la inferencia usual frecuentista es que está basada en resultados asintóticos y hay evidencia que permite afirmar que en factoriales no replicados se tiende a detectar un exceso de efectos activos en la situación donde solo hay efectos inertes.

Para cada modelo considerado se propone un método para incorporar la información *a priori* sobre los coeficientes del predictor lineal. Por ejemplo, en cada caso se propone una distribución gama para el parámetro β_0 , cuyos hiperparámetros se deducen a partir de información mínima que puede proporcionar el usuario antes de correr el experimento. Para los parámetros β_i se propone como distribución *a priori* la normal con media cero y varianza "inflada" a partir de una varianza base atribuida al parámetro β_0 .

Una clara limitante del método es que la integración con Monte Carlo (o Quasi-Montecarlo) directo es muy tardada, y más por el hecho de la gran cantidad de modelos posibles, que se incrementan geométricamente cuando aumenta el número de efectos (factores) a estudiar. Recordemos que se debe resolver una integral por cada modelo. En la medida de que las computadoras sean más rápidas será más atractiva su aplicación, pero se pueden buscar formas más eficientes de realizar dicha simulación. Por ejemplo, buscar solamente en los modelos hasta cierta cantidad de términos o trabajar solamente con los modelos jerárquicos (aquellos donde dado un efecto de interacción, se incluyen también en el modelo los efectos más simples que componen dicha interacción).

Si bien los ejemplos considerados arrojaron resultados interesantes, habría varios aspectos por explorar en el futuro, entre los que se encuentran:

1. Buscar maneras más rápidas de hacer los cálculos, posiblemente mediante alguna aproximación tipo BIC (Criterio de Información Bayesiana), que de paso evite el problema de integración numérica. O bien, buscar en subconjuntos de modelos más reducidos, como pueden ser los modelos jerárquicos.

2. Probar los métodos en escenarios simulados, no solo con datos reales, de manera que se pueda estudiar su desempeño en situaciones conocidas.

3. Proponer criterios para seleccionar la función liga más apropiada en un modelo determinado.

Capítulo 13

Apéndice: Demostración de la Fórmula 3.3

En su método Bayesiano que contempla la posibilidad de datos anómalos, Box y Meyer (1987) reportan sin demostración que la probabilidad posterior de que un conjunto particular de r_1 efectos esté activo y de que al mismo tiempo las observaciones r_2 sean anómalas (evento $a_{(r_1, r_2)}$) está dada por

$$\begin{aligned}
 p(a_{(r_1, r_2)} | \mathbf{y}) &\propto \left(\frac{\alpha_1}{1 - \alpha_1}\right)^{r_1} \left(\frac{\alpha_2}{1 - \alpha_2}\right)^{r_2} \gamma^{-r_1} k^{-r_2} & (A1) \\
 &\times \frac{|\mathbf{X}_{(0)} \mathbf{X}_{(0)}|^{1/2}}{|\mathbf{\Gamma}_{r_1} + \mathbf{X}_{(r_1)} \mathbf{X}_{(r_1)} - \varphi \mathbf{X}_{(r_1, r_2)} \mathbf{X}_{(r_1, r_2)}|^{1/2}} \\
 &\times \left(\frac{\mathbf{S}(\hat{\tau}_{(r_1, r_2)}) + \hat{\tau}_{(r_1, r_2)} \mathbf{\Gamma}_{(r_1)} \hat{\tau}_{(r_1, r_2)}}{\mathbf{S}(\hat{\tau}_{(0)})}\right),
 \end{aligned}$$

(ecuación 3.3 del Capítulo 3) donde \mathbf{y} es el vector de observaciones y $\mathbf{X}_{(r_1, r_2)}$ es la matriz de columnas y renglones de \mathbf{X} que corresponden a los r_1 efectos activos y r_2 observaciones anómalas, α_1 y α_2 son las probabilidades *a priori* de que un efecto esté activo y de que un dato sea anómalo, respectivamente; γ y k son los factores de inflamamiento de la varianza debida a efecto activo y a dato anómalo, respectivamente;

y además

$$\begin{aligned}\varphi &= 1 - 1/k^2, & (A2) \\ \widehat{\boldsymbol{\tau}}_{(r_1, r_2)} &= (\boldsymbol{\Gamma}_{r_1} + \mathbf{X}_{(r_1)}\mathbf{X}_{(r_1)} - \varphi\mathbf{X}_{(r_1, r_2)}\mathbf{X}_{(r_1, r_2)})^{-1} (\mathbf{X}_{(r_1)}\dot{\mathbf{y}} - \varphi\mathbf{X}_{(r_1, r_2)}\dot{\mathbf{y}}_{(r_2)}) \\ \mathbf{S}(\widehat{\boldsymbol{\tau}}_{(r_1, r_2)}) &= (\mathbf{y} - \mathbf{X}_{(r_1)}\widehat{\boldsymbol{\tau}}_{(r_1, r_2)})'(\mathbf{y} - \mathbf{X}_{(r_1)}\widehat{\boldsymbol{\tau}}_{(r_1, r_2)}) - \\ &\quad \varphi (\mathbf{y}_{(r_2)} - \mathbf{X}_{(r_1, r_2)}\widehat{\boldsymbol{\tau}}_{(r_1, r_2)})'(\mathbf{y}_{(r_2)} - \mathbf{X}_{(r_1, r_2)}\widehat{\boldsymbol{\tau}}_{(r_1, r_2)})\end{aligned}$$

donde $\mathbf{y}_{(r_2)}$ son las supuestas observaciones anómalas. La matriz \mathbf{X} de dimensión $n \times n$ esta formada por las columnas que corresponden a los $n - 1$ contrastes posibles con los cuales se estiman los efectos de la manera usual (ver Capítulo 1), más una primera columna de unos.

A continuación damos la demostración de esta fórmula que es muy similar a la del método Bayesiano sin datos anómalos dada por Meyer y Box (1992). Agregando a ésta las ideas de Box y Tiao (1968) se obtiene el resultado. Sea $a_{(r_1, r_2)}$ el evento definido por un conjunto particular de r_1 efectos activos y un conjunto particular de r_2 observaciones anómalas. Esto es, r_1 representa las columnas de la matriz \mathbf{X} que corresponden a los efectos activos y r_2 son los renglones de la matriz \mathbf{X} y del vector \mathbf{y} que corresponden a los datos anómalos. Sea $M_i = M_{(r_1, r_2)}$ el modelo con parámetro $\boldsymbol{\beta}_i$ asociado con el evento $a_{(r_1, r_2)}$, que de acuerdo a esto tiene r_1 términos independientes. El problema es encontrar la probabilidad posterior del evento $a_{(r_1, r_2)}$ dada por

$$p(a_{(r_1, r_2)} | \mathbf{y}) = \frac{p(M_{(r_1, r_2)})p(\mathbf{y} | a_{(r_1, r_2)})}{\sum_{r_1} \sum_{r_2} p(M_{(r_1, r_2)})p(\mathbf{y} | a_{(r_1, r_2)})} \quad (A3)$$

donde $p(\mathbf{y} | a_{(r_1, r_2)})$ es la densidad predictiva de \mathbf{y} dado el evento $a_{(r_1, r_2)}$, la cual resulta de resolver la integral

$$p(\mathbf{y} | a_{(r_1, r_2)}) = \int_{\Theta} p(\mathbf{y} | a_{(r_1, r_2)}, \boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i | a_{(r_1, r_2)}) d\boldsymbol{\theta}_i, \quad (A4)$$

donde Θ es el espacio de los parámetros. El vector de parámetros está dado por $\boldsymbol{\theta}_i = (\beta_0, \beta_1, \beta_2, \dots, \beta_{r_1}, \sigma)$, que son los coeficientes de regresión y el parámetro σ .

La parte difícil del problema es resolver esta última integral, considerando que cuando ocurre el evento $a_{(r_1, r_2)}$ se ajusta el modelo lineal

$$\mathbf{y} = \mathbf{X}_{r_1} \boldsymbol{\beta}_i + \boldsymbol{\varepsilon} \quad (\text{A5})$$

donde $\boldsymbol{\beta}_i = (\beta_0, \beta_1, \beta_2, \dots, \beta_{r_1})$, \mathbf{X}_{r_1} contiene las columnas de \mathbf{X} que corresponden a los r_1 efectos activos y $\boldsymbol{\varepsilon}$ es el vector de errores. Para permitir la posibilidad de observaciones anómalas suponemos que los errores que corresponden a las observaciones anómalas provienen de una distribución normal con varianza inflada $k^2 \sigma^2$ ($k > 1$) y éstas ocurren con una probabilidad pequeña α_2 . Esto es, suponemos que cada error ε sigue la distribución mezcla de normales

$$\varepsilon \sim (1 - \alpha_2)N(0, \sigma^2) + \alpha_2 N(0, k^2 \sigma^2).$$

Considerando que se tienen n observaciones en total de las cuales $r = r_2$ son datos anómalos y $s = (n - r)$ son observaciones normales, y denotando por \mathbf{X}_s y \mathbf{X}_r las submatrices renglones de la matriz \mathbf{X}_{r_1} que corresponden a las \mathbf{y}_s observaciones normales y \mathbf{y}_r observaciones anómalas, respectivamente, se tiene que la densidad de \mathbf{y} dado el evento de interés y el modelo asociado está dada por

$$p(\mathbf{y} | a_{(r_1, r_2)}, \boldsymbol{\beta}_i, \sigma) \propto \sigma^{-n} k^{-r} \exp \left\{ -\frac{(\mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta}_i)' (\mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta}_i)}{2\sigma^2} - \frac{(\mathbf{y}_r - \mathbf{X}_r \boldsymbol{\beta}_i)' (\mathbf{y}_r - \mathbf{X}_r \boldsymbol{\beta}_i)}{2k^2 \sigma^2} \right\}. \quad (\text{A6})$$

Se supone *a priori* que los parámetros β_i ($1 \leq i \leq r_1$) se distribuyen de manera independiente normal con media cero y varianza inflada $\gamma^2 \sigma^2$, esto es,

$$p(\boldsymbol{\beta}_i | a_{(r_1, r_2)}, \sigma) \propto \gamma^{-r_1} \sigma^{-r_1} \exp \left\{ -\frac{\boldsymbol{\beta}_i \boldsymbol{\Gamma}_{r_1} \boldsymbol{\beta}_i}{2\sigma^2} \right\} \quad (\text{A7})$$

donde

$$\boldsymbol{\Gamma}_{r_1} = \frac{1}{\gamma^2} \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{r_1} \end{bmatrix}.$$

La media igual a cero de esta distribución *a priori* refleja la ignorancia del signo de cada efecto particular, y la magnitud del efecto relativa al ruido experimental se captura a través del parámetro γ . Se considera la distribución *a priori* no informativa (ver Box y Tiao, 1973) para los parámetros β_0 y σ dada por

$$p(\beta_0, \sigma | a_{(r_1, r_2)}) = p(\beta_0, \sigma) \propto \frac{1}{\sigma} \quad (\text{A8})$$

Una vez definidas la verosimilitud y las distribuciones *a priori* procedemos a calcular la predictiva de interés dada en (A4), comenzando por integrar respecto al vector de parámetros β_i .

$$\begin{aligned} p(\mathbf{y} | a_{(r_1, r_2)}, \sigma) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\mathbf{y} | a_{(r_1, r_2)}, \beta_i, \sigma) p(\beta_i | a_{(r_1, r_2)}, \sigma) d\beta_i \quad (\text{A9}) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \gamma^{-r_1} k^{-r_2} \sigma^{-(n+r_1)} \exp \left\{ - \left(\frac{1}{2\sigma^2} \right) [(\mathbf{y}_s - \mathbf{X}_s \beta_i)' (\mathbf{y}_s - \mathbf{X}_s \beta_i) \right. \\ &\quad \left. + \left(\frac{1}{k^2} \right) (\mathbf{y}_r - \mathbf{X}_r \beta_i)' (\mathbf{y}_r - \mathbf{X}_r \beta_i) + \beta_i \Gamma_{r_1} \beta_i] \right\} \end{aligned}$$

Utilizando la identidad

$$\mathbf{X}_s \mathbf{X}_s + \frac{1}{k^2} \mathbf{X}_r \mathbf{X}_r = \mathbf{X}_{r_1} \mathbf{X}_{r_1} - \varphi \mathbf{X}_{r_1, r_2} \mathbf{X}_{r_1, r_2}$$

donde la matriz \mathbf{X}_{r_1, r_2} contiene las columnas que corresponden a los r_1 efectos activos y los r_2 de las observaciones anómalas, la integral anterior (A9) se puede escribir como

$$\begin{aligned} p(\mathbf{y} | a_{(r_1, r_2)}, \sigma) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \gamma^{-r_1} k^{-r_2} k^{-r_2} \sigma^{-(n+r_1)} \exp \left\{ - \left(\frac{1}{2\sigma^2} \right) [(\mathbf{y} - \mathbf{X}_{r_1} \beta_i)' (\text{A10}) \right. \\ &\quad \left. (\mathbf{y} - \mathbf{X}_{r_1} \beta_i) - \varphi (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \beta_i)' (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \beta_i) + \beta_i \Gamma_{r_1} \beta_i] \right\} \end{aligned}$$

Sea

$$S(\beta_i) = (\mathbf{y} - \mathbf{X}_{r_1} \beta_i)' (\mathbf{y} - \mathbf{X}_{r_1} \beta_i) - \varphi (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \beta_i)' (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \beta_i) + \beta_i \Gamma_{r_1} \beta_i$$

que se puede escribir como

$$S(\boldsymbol{\beta}_i) = S(\widehat{\boldsymbol{\beta}}_i) + \left(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i\right)' [\boldsymbol{\Gamma}_{r_1} + \mathbf{X}_{r_1} \mathbf{X}_{r_1}' - \varphi \mathbf{X}_{r_1 r_2} \mathbf{X}_{r_1 r_2}'] \left(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i\right)$$

donde

$$\widehat{\boldsymbol{\beta}}_i = (\boldsymbol{\Gamma}_{r_1} + \mathbf{X}_{r_1} \mathbf{X}_{r_1}' - \varphi \mathbf{X}_{r_1 r_2} \mathbf{X}_{r_1 r_2}')^{-1} (\mathbf{X}_{r_1} \dot{\mathbf{y}} - \mathbf{X}_{r_1 r_2} \dot{\mathbf{y}}_{r_2}).$$

Reacomodando términos y multiplicando y dividiendo por

$$|\boldsymbol{\Gamma}_{r_1} + \mathbf{X}_{r_1} \mathbf{X}_{r_1}' - \varphi \mathbf{X}_{r_1 r_2} \mathbf{X}_{r_1 r_2}'|^{1/2}$$

se busca completar la integral de una distribución normal multivariada, lo que lleva a la expresión

$$\begin{aligned} p(\mathbf{y} | a_{(r_1, r_2)}, \sigma) &= \gamma^{-r_1} k^{-r_2} \sigma^{-(n-1)} |\boldsymbol{\Gamma}_{r_1} + \mathbf{X}_{r_1} \mathbf{X}_{r_1}' - \varphi \mathbf{X}_{r_1 r_2} \mathbf{X}_{r_1 r_2}'|^{-1/2} \quad (\text{A11}) \\ &\times \exp \left\{ - \left(\frac{1}{2\sigma^2} \right) \left[(\mathbf{y} - \mathbf{X}_{r_1} \widehat{\boldsymbol{\beta}}_i)' (\mathbf{y} - \mathbf{X}_{r_1} \widehat{\boldsymbol{\beta}}_i) \right. \right. \\ &\quad \left. \left. - \varphi (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \widehat{\boldsymbol{\beta}}_i)' (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \widehat{\boldsymbol{\beta}}_i) + \widehat{\boldsymbol{\beta}}_i' \boldsymbol{\Gamma}_{r_1} \widehat{\boldsymbol{\beta}}_i \right] \right\} \\ &\times \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |\boldsymbol{\Gamma}_{r_1} + \mathbf{X}_{r_1} \mathbf{X}_{r_1}' - \varphi \mathbf{X}_{r_1 r_2} \mathbf{X}_{r_1 r_2}'|^{1/2} \\ &\times \exp \left(\frac{-1}{2\sigma^2} (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)' [\boldsymbol{\Gamma}_{r_1} + \mathbf{X}_{r_1} \mathbf{X}_{r_1}' \right. \\ &\quad \left. - \varphi \mathbf{X}_{r_1 r_2} \mathbf{X}_{r_1 r_2}'] (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i) \right) d\boldsymbol{\beta}_i, \end{aligned}$$

donde el valor de la integral es uno, al ser la integral en todo el soporte de una distribución normal multivariada, quedando solamente lo que está afuera de la integral.

De esta forma

$$\begin{aligned} p(\mathbf{y} | a_{(r_1, r_2)}, \sigma) &\propto \gamma^{-r_1} \sigma^{-(n-1)} |\boldsymbol{\Gamma}_{r_1} + \mathbf{X}_{r_1} \mathbf{X}_{r_1}' - \varphi \mathbf{X}_{r_1 r_2} \mathbf{X}_{r_1 r_2}'|^{-1/2} \quad (\text{A12}) \\ &\times \exp - \left\{ \left(\frac{1}{2\sigma^2} \right) \left[(\mathbf{y} - \mathbf{X}_{r_1} \widehat{\boldsymbol{\beta}}_i)' (\mathbf{y} - \mathbf{X}_{r_1} \widehat{\boldsymbol{\beta}}_i) \right. \right. \\ &\quad \left. \left. - \varphi (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \widehat{\boldsymbol{\beta}}_i)' (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \widehat{\boldsymbol{\beta}}_i) + \widehat{\boldsymbol{\beta}}_i' \boldsymbol{\Gamma}_{r_1} \widehat{\boldsymbol{\beta}}_i \right] \right\}, \end{aligned}$$

pero falta por integrar respecto a σ , es decir,

$$\begin{aligned}
 p(\mathbf{y} | a_{(r_1, r_2)}) &= k^{-r_2} \gamma^{-r_1} |\mathbf{\Gamma}_{r_1} + \mathbf{X}_{r_1} \mathbf{X}_{r_1}' - \varphi \mathbf{X}_{r_1 r_2} \mathbf{X}_{r_1 r_2}'|^{-1/2} \\
 &\times \int_0^\infty \sigma^{-(n-1)-1} \exp - \left\{ \left(\frac{1}{2\sigma^2} \right) \left[(\mathbf{y} - \mathbf{X}_{r_1} \hat{\boldsymbol{\beta}}_i)' (\mathbf{y} - \mathbf{X}_{r_1} \hat{\boldsymbol{\beta}}_i) \right. \right. \\
 &\left. \left. - \varphi (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \hat{\boldsymbol{\beta}}_i)' (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i' \mathbf{\Gamma}_{r_1} \hat{\boldsymbol{\beta}}_i \right] \right\} d\sigma.
 \end{aligned} \tag{A13}$$

Se define la transformación

$$\begin{aligned}
 u &= \left(\frac{1}{2\sigma^2} \right) \left[(\mathbf{y} - \mathbf{X}_{r_1} \hat{\boldsymbol{\beta}}_i)' (\mathbf{y} - \mathbf{X}_{r_1} \hat{\boldsymbol{\beta}}_i) \right. \\
 &\left. - \varphi (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \hat{\boldsymbol{\beta}}_i)' (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i' \mathbf{\Gamma}_{r_1} \hat{\boldsymbol{\beta}}_i \right] \\
 &= S(\hat{\boldsymbol{\beta}}) / 2\sigma^2
 \end{aligned}$$

de manera que

$$\sigma = [S(\hat{\boldsymbol{\beta}})]^{1/2} (2u)^{-1/2}$$

y el Jacobiano de la transformación es

$$d\sigma = - [S(\hat{\boldsymbol{\beta}})]^{1/2} 2^{-1/2} u^{-3/2} du.$$

Sustituyendo en la integral (A13) se tiene que

$$\begin{aligned}
 p(\mathbf{y} | a_{(r_1, r_2)}) &\propto \gamma^{-r_1} k^{-r_2} |\mathbf{\Gamma}_{r_1} + \mathbf{X}_{r_1} \mathbf{X}_{r_1}' - \varphi \mathbf{X}_{r_1 r_2} \mathbf{X}_{r_1 r_2}'|^{-1/2} \\
 &\times \int_0^\infty \left([S(\hat{\boldsymbol{\beta}})]^{1/2} (2u) \right)^{-n/2} [S(\hat{\boldsymbol{\beta}})]^{1/2} 2^{-1/2} u^{-3/2} \exp \{-u\} du \\
 &= \gamma^{-r_1} |\mathbf{\Gamma}_{r_1} + \mathbf{X}_{r_1} \mathbf{X}_{r_1}' - \varphi \mathbf{X}_{r_1 r_2} \mathbf{X}_{r_1 r_2}'|^{-1/2} [S(\hat{\boldsymbol{\beta}})]^{-(n-1)/2} \\
 &\times \int_0^\infty (u)^{\frac{n-1}{2}-1} \exp \{-u\} du.
 \end{aligned} \tag{A14}$$

La integral en términos de u es la integral de una función gamma, que es una constante,

de manera que

$$\begin{aligned}
p(\mathbf{y} | a_{(r_1, r_2)}) &\propto \gamma^{-r_1} k^{-r_2} |\mathbf{\Gamma}_{r_1} + \mathbf{X}_{r_1} \mathbf{X}_{r_1} - \varphi \mathbf{X}_{r_1 r_2} \mathbf{X}_{r_1 r_2}|^{-1/2} \\
&\times \left[(\mathbf{y} - \mathbf{X}_{r_1} \widehat{\boldsymbol{\beta}}_i)' (\mathbf{y} - \mathbf{X}_{r_1} \widehat{\boldsymbol{\beta}}_i) \right. \\
&\left. - \varphi (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \widehat{\boldsymbol{\beta}}_i)' (\mathbf{y}_{r_2} - \mathbf{X}_{r_1 r_2} \widehat{\boldsymbol{\beta}}_i) + \widehat{\boldsymbol{\beta}}_i' \mathbf{\Gamma}_{r_1} \right]^{-\frac{(n-1)}{2}}
\end{aligned} \tag{A15}$$

Por otra parte, la predictiva para el modelo nulo con ninguna observación anómala sería

$$\begin{aligned}
p(\mathbf{y} | a_{(0)}) &\propto |\mathbf{X}_0 \mathbf{X}_0|^{-1/2} \times (\mathbf{y} - \mathbf{X}_0 \widehat{\boldsymbol{\beta}}_0)' (\mathbf{y} - \mathbf{X}_0 \widehat{\boldsymbol{\beta}}_0) \\
&= |\mathbf{X}_0 \mathbf{X}_0|^{-1/2} \times S(\widehat{\boldsymbol{\beta}}_0)
\end{aligned} \tag{A16}$$

donde \mathbf{X}_0 es el vector de unos de dimensión 16 (primera columna de \mathbf{X}) y $S(\widehat{\boldsymbol{\beta}}_0) = (\mathbf{y} - \bar{\mathbf{y}})' (\mathbf{y} - \bar{\mathbf{y}})$ con $\bar{\mathbf{y}}$ el vector con 16 entradas iguales a la media \bar{y} de las observaciones. De la ecuación (A3) tenemos que la probabilidad posterior del evento $a_{(r_1, r_2)}$ está dada por

$$p(a_{(r_1, r_2)} | \mathbf{y}) = C \times p(M_{(r_1, r_2)}) p(\mathbf{y} | a_{(r_1, r_2)}),$$

donde C es la constante de normalización y $p(M_{(r_1, r_2)})$ es la probabilidad *a priori* del modelo con r_1 efectos activos y de que, al mismo tiempo, ocurran r_2 datos anómalos. Esta probabilidad *a priori* está dada por

$$p(M_{(r_1, r_2)}) = \left(\frac{\alpha_1}{1 - \alpha_1} \right)^{r_1} \times \left(\frac{\alpha_2}{1 - \alpha_2} \right)^{r_2}.$$

donde α_1 es la probabilidad *a priori* de que un efecto particular esté activo y α_2 es la probabilidad *a priori* de que un dato se anómalo.

Expresando la predictiva $p(\mathbf{y} | a_{(r_1, r_2)})$ en relación a la predictiva del modelo nulo y sin datos anómalos $p(\mathbf{y} | a_{(0)})$, la probabilidad posterior deseada se puede expresar

como

$$\begin{aligned}
 p(a_{(r_1, r_2)} | \mathbf{y}) &= C \left(\frac{\alpha_1}{1 - \alpha_1} \right)^{r_1} \left(\frac{\alpha_2}{1 - \alpha_2} \right)^{r_2} \gamma^{-r_1} k^{-r_2} \\
 &\times \frac{|\mathbf{X}_{(0)} \mathbf{X}_{(0)}|^{1/2}}{|\mathbf{\Gamma}_{r_1} + \mathbf{X}_{(r_1)} \mathbf{X}_{(r_1)} - \varphi \mathbf{X}_{(r_1, r_2)} \mathbf{X}_{(r_1, r_2)}|^{1/2}} \\
 &\times \left(\frac{\mathbf{S}(\hat{\tau}_{(r_1, r_2)}) + \hat{\tau}_{(r_1, r_2)} \mathbf{\Gamma}_{(r_1)} \hat{\tau}_{(r_1, r_2)}}{\mathbf{S}(\hat{\tau}_{(0)})} \right),
 \end{aligned} \tag{A17}$$

donde C es una nueva constante de normalización. Esta es la expresión a la que se quería llegar (ecuación 3.3), donde los términos que la componen se definen en la ecuación (A2)..

Capítulo 14

Apéndice: Programas

14.1 Detección de Efectos Activos en Experimentos Factoriales No-Replicados Contemplando la Posibilidad de Datos Anómalos

En esta sección se presentan los programas desarrollados en SPLUS para aplicar, tanto los métodos existentes como los métodos propuestos en este trabajo de tesis, en el análisis de factoriales no-replicados contemplando la posibilidad de observaciones anómalas en el experimento.

Con excepción de método Bayesiano de Box y Meyer (1987), en todos los métodos programados se explota la idea de Aguirre-Torres y Pérez-Trejo (2001) de aplicar una variante del método híbrido de Benski (1989) al vector de efectos estimados de manera robusta. Así, comenzamos enlistando el método de rangos de AT&PT (2001) en el cual se señala lo específico del método entre los comentarios *comienza-método* y *termina-método*, en el entendido de que el resto del programa es lo mismo para todos los métodos. Por ello, sólo se enlistará la parte entre dichos comentarios en los métodos subsecuentes.

14.1.1 Método de Rangos

```

# METODO DE RANGOS (Aguirre-Torres y Pérez-Trejo, 2001)
# PRIMERO DECLARE LOS DATOS:
Y1 <- c(47.46,49.62,43.13,46.31,51.47,48.49,49.34,46.10,
        46.76,48.56,44.83,44.45,59.15,51.33,47.02,47.90)
# MATRIZ DE SIGNOS DE UN FACTORIAL 2^4
X0 <- c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)
X1 <- c(-1, 1,-1, 1,-1, 1,-1, 1,-1, 1,-1, 1,-1, 1,-1, 1)
X2 <- c(-1,-1, 1, 1,-1,-1, 1, 1,-1,-1, 1, 1,-1,-1, 1, 1)
X3 <- c(-1,-1,-1,-1, 1, 1, 1, 1,-1,-1,-1,-1, 1, 1, 1, 1)
X4 <- c(-1,-1,-1,-1,-1,-1,-1,-1, 1, 1, 1, 1, 1, 1, 1, 1)
X1X2 <- X1*X2; X1X3 <- X1*X3; X1X4 <- X1*X4; X2X3 <- X2*X3
X2X4 <- X2*X4; X3X4 <- X3*X4; X1X2X3 <- X1*X2*X3
X1X2X4 <- X1*X2*X4; X1X3X4 <- X1*X3*X4; X2X3X4 <- X2*X3*X4
X1X2X3X4 <- X1*X2*X3*X4
X <-cbind(X1,X2,X3,X4,X1X2,X1X3,X1X4,X2X3,X2X4,X3X4,X1X2X3, X1X2X4,
X1X3X4,X2X3X4,X1X2X3X4)
# "EFAC" ES EL VECTOR DE RESULTADOS: 0 ES EFECTO INERTE, 1 ES ACTIVO
EFX1 <- 0; EFX2 <- 0; EFX3 <- 0; EFX4 <- 0;
EFX1X2 <- 0; EFX1X3 <- 0; EFX1X4 <- 0; EFX2X3 <- 0;
EFX2X4 <- 0; EFX3X4 <- 0; EFX1X2X3 <- 0; EFX1X2X4 <- 0;
EFX1X3X4 <- 0; EFX2X3X4 <- 0; EFX1X2X3X4 <- 0;
EFAC <- cbind(EFX1, EFX2, EFX3, EFX4, EFX1X2, EFX1X3, EFX1X4, EFX2X3,
EFX2X4,EFX3X4, EFX1X2X3, EFX1X2X4, EFX1X3X4, EFX2X3X4, EFX1X2X3X4)
# -----COMIENZA MÉTODO -----#
NOMBRE <- "MÉTODO DE RANGOS"; PC <- 0.033
# TRANSFORMACIÓN DE RANGOS Y EFECTOS ESTIMADOS
Y2 <- rank(Y1); CON2 <- t(X)%*%Y2; EF2 <- CON2/8

```



```

#-----TERMINA MÉTODO -----#
# POSICIONES DE GRAFICACIÓN PARA LA PRUEBA DE NORMALIDAD
n <- 15; alfa <- rep(0,n); p <- rep(0,n)
for (i in 2:n-1)
  {
    alfa[i] <- 0.275499 + 0.072884*(log(n))^0.41148
    alfa[1] <- 0.205146 + 0.1314965*(log(n))^0.226701
    alfa[n] <- alfa[1]
  }
for (j in 1:n)
  {
    p[j] <- (j-alfa[j])/(n-2*alfa[j]+1)
    m <- qnorm(p)
  }
# ESTADÍSTICO W
EFORD <- sort(EF2)
NUMERADOR <- (sum(m*EFORD))^2
DENOMINA <- (sum(m*m))*(n-1)*var(EFORD)
W <- NUMERADOR/DENOMINA
# SIGNIFICANCIA
A <- 1.031918 - 0.183573*(0.1*n)^-0.5447402
B <- -0.5084706 + 2.076782*(0.1*n)^-0.4905993
C <- ((W-A)/B + 0.0486128 )/0.02760309-log(100)
PR <- exp(C)
# PRUEBA DE DISPERSIÓN DE LOS CUARTOS
FL <- (EFORD[4]+EFORD[5])/2
FU <- (EFORD[11]+EFORD[12])/2
DF <- FU-FL

```

```

EFACT <- EFORD[abs(EFORD)>2*DF]
EFACTR <- length(EFACT)
# DETECCIÓN DE EFECTOS ACTIVOS
if (PR < PC && EFACTR > 0)
  {for (i in 1:15) if (abs(EF2[i])>2*DF) EFAC[i] <- 1};
# EFECTOS, EFECTOS ACTIVOS Y EFECTOS EN PAPEL NORMAL
EFECTOS <- EF2; EFAC; probplot(EFECTOS);
mtext(outer=F, NOMBRE, line=.5, side=3);
#-----FIN DEL PROGRAMA-----#

```

14.1.2 Método de Rangos Modificados

```

# ----- COMIENZA MÉTODO-----#
NOMBRE <- "MÉTODO DE rangos modificados"; PC <- 0.045
Y1STR <- rep(1,16); Y11 <- rep(1,16)
Y1S <- sort(Y1); DISOR <- Y1S[15]-Y1S[2]; posY1 <- order(Y1)
Y1STR[16] <- 16; Y1STR[15] <- 15; Y1STR[1] <- 1; Y1STR[2] <- 2;
  for (i in 3:15)
    {
      Y1STR[i] <- Y1STR[i-1]+((Y1S[i]- Y1S[i-1])/DISOR)*13
    }
  for (i in 1:16)
    {Y11[posY1[i]]<-Y1STR[i]}
# EFECTOS ESTIMADOS
Y2 <- Y11; CON2 <- t(X)%*%Y2; EF2 <- CON2/8
# -----TERMINA MÉTODO-----#

```

14.1.3 Método de Ajuste de Espacios Entre Efectos

```

# ----- COMIENZA MÉTODO-----#
NOMBRE <- "MÉTODO DE AJUSTE DE ESPACIOS ENTRE EFECTOS"; PC <- 0.001
# CALCULO DE LOS EFECTOS Y DE LOS OCHO HUECOS CENTRALES
CON2 <- t(X)%*%Y1
EF2 <- CON2/8; EFOR1 <- sort(EF2); posi <- order(EF2)
GAPS <- rep(0,8)
  for (j in 1:8) {GAPS[j] <- (EFOR1[j+4]-EFOR1[j+3])}
MD <- median(GAPS)/0.07
# SEIS HUECOS CENTRALES DE EFECTOS ESTANDARIZADOS
EF2 <- EF2/(MD/2) # EFECTOS ESTANDARIZADOS
EFORD <- sort(EF2)
GAPS <- rep(0,6)
  for (j in 1:6) {GAPS[j] <- (EFORD[j+5]-EFORD[j+4])}
posGap <- order(GAPS)
GAP6 <-GAPS[posGap[6]]; GAP5 <-GAPS[posGap[5]];
GAP4 <-GAPS[posGap[4]];
# AJUSTE DE LOS TRES MAYORES GAPS DE LOS SEIS INTERIORES
for (j in 1:15){if (j < posGap[6]+ 5) EFOR1[j]<-EFORD[j]+
  (GAP6 - 0.023)/2 else EFOR1[j]<-EFORD[j]- (GAP6 -0.023)/2}
for (j in 1:15){if (j< posGap[5]+5) EFOR1[j]<-EFORD[j]+
  (GAP5-0.015)/2 else EFOR1[j]<-EFORD[j]- (GAP5-0.015)/2}
for (j in 1:15){if (j< posGap[4]+5) EFOR1[j]<-EFORD[j]+
  (GAP4-0.01)/2 else EFOR1[j]<-EFORD[j]- (GAP4-0.01)/2}
  for (i in 1:15)
    {EF2[posi[i]]<-EFORD[i]}
# -----TERMINA MÉTODO-----#

```

14.1.4 Método de Estimación de Datos Extremos (Anómalos)

```

#----- COMIENZA MÉTODO-----#
NOMBRE <- "MÉTODO DE DATOS EXTREMOS REESTIMADOS"; PC <- 0.033
# AJUSTE DE LA REGRESIÓN L1
REGL1 <- l1fit(XR,Y1, intercept=T, print=F)
RESIDUOS <- REGL1$res; YGORRO <- rep (0,16)
for (i in 1:16) {YGORRO[i] <- XRO[i,]*%REGL1$coef}
YGRES <- cbind(YGORRO,RESIDUOS)
# CONSTRUCCIÓN DE LA MATRIZ DE DISTANCIAS
distancias <- dist(YGRES,metric ="euclidean"); MD <-matrix(0,16,16);
  for (i in 1:16)
  {
    for (j in 1:16)
    {
      if (i<j)
      MD[i,j]<- distancias[16*(i-1) - i*(i-1)/2 + j-i]
    }
  }
for (i in 1:16)
{
  for (j in 1:16)
  {
    MD[j,i]<- MD[i,j]
  }
}
# ANÁLISIS DE CLUSTER

```

```

CLUST <- hclust(distancias, method = "connected")
# DETERMINACIÓN DE LOS CANDIDATOS A OUTLIERS, SU POSICIÓN Y DISTANCIA
A LA CUAL SE AGRUPAN
pasos <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15)
COMB <- cbind(pasos,CLUST$merge)
COMBS <- select.rows(COMB, rows=[(COMB[,2]<0 |COMB[,3]<0)])
nc <- dim(COMBS)[1]; COMBR1 <- COMBS[nc,];COMBR2 <- COMBS[nc-1,]
nega1 <- rbind(COMBR2,COMBR1)
PASO1 <- nega1[1,1]; PASO2 <- nega1[2,1]
nega1[1,1] <- nega1[1,2]; nega1[2,1] <- nega1[2,2]
# MANEJANDO LA UNIÓN DE DOS DATOS EXTREMOS QUE APARECEN POR PRIMERA
VEZ, LOS CUALES SE AGRUPAN AL RESTO DE LOS DATOS A UNA DISTANCIA MAYOR
COMBSC <- select.rows(COMB, rows=[(COMB[,1]>PASO2)])
DIFER1 <- sum(COMBSC==PASO1);
nr <- dim(COMBSC)[1]
if (DIFER1==0) ren1 <- 0
if (DIFER1==1) {ren1 <- ((which(COMBSC==PASO1))%nr);
if (ren1==0) ren1 <- nr}
DIFER2 <- sum(COMBSC==PASO2);
if (DIFER2==0) ren2 <- 0
if (DIFER2==1) {ren2 <- ((which(COMBSC==PASO2))%nr);
if (ren2==0) ren2 <- nr}
if ((nega1[1,3] < 0) && (DIFER1 ==1) && (COMBSC[ren1,1] > PASO2))
    {nega1[1,1] = nega1[1,3]; nega1[2,1] = nega1[1,2];
PASO1 <- COMBSC[ren1,1]; PASO2 <- COMBSC[ren1,1]}
if (nega1[2,3] < 0) {nega1[1,1] = nega1[2,3]; nega1[2,1] = nega1[2,2];
PASO1 <- COMBSC[ren2,1]; PASO2 <- COMBSC[ren2,1]}
# DISTANCIA CRÍTICA

```

```

DISTAN <- select.rows(CLUST$height, 1:PASO2)
LSD <- quantile(DISTAN,c(.75))+2.2*diff(quantile(DISTAN,c(.25, .75)))
# INICIA PROCEDIMIENTO PARA ESTIMARLOS COMO OBSERVACIONES PERDIDAS
LETRA1 <- XY[,-nega1[1,1]]*XY[,-nega1[2,1]]
LETRAS <- cbind(X0,X0,X0,X2,X0,X3,X3,X2X3,X0,X4,X4,
                X2X4,X4,X3X4,X3X4,X2X3X4)
COL <- 0
  for (i in 1:16)
  {
    VALOR <- LETRA1%*%XY[,i]
    if (VALOR==16) COL <- i
  }
LETRA2 <- XY[,COL]*LETRAS[,COL]
LETRA3 <- LETRA2*ABCD # CONTIENE EL SEGUNDO CONTRASTE A UTILIZAR
  if (CLUST$height[PASO1] > LSD) {Y1[-nega1[1,1]]=0;
  Y1[-nega1[2,1]]=0; Y11 <- Y1;
TOTAL1 <- -sum(ABCD*Y11); TOTAL2 <- -sum(LETRA3*Y11)}
if (CLUST$height[PASO1] > LSD) {DENOM <-rbind(c(ABCD[-nega1[1,1]],
ABCD[-nega1[2,1]]),c(LETRA3[-nega1[1,1]],LETRA3[-nega1[2,1]]));
NY1<-rbind(c(TOTAL1,ABCD[-nega1[2,1]]),c(TOTAL2,LETRA3[-nega1[2,1]]));
NY2<-rbind(c(ABCD[-nega1[1,1]],TOTAL1),c(LETRA3[-nega1[1,1]],TOTAL2))}
if (CLUST$height[PASO1] > LSD) {Y1[-nega1[1,1]]=det(NY1)/det(DENOM);
Y1[-nega1[2,1]]=det(NY2)/det(DENOM)}
if ((CLUST$height[PASO2] > LSD)&(CLUST$height[PASO1] < LSD))
  {Y1[-nega1[2,1]]=0; Y11 <- Y1; TOTAL1 <- -sum(ABCD*Y11);
  Y1[-nega1[2,1]] <- sign(ABCD[-nega1[2,1]])*TOTAL1}
Y2 <- Y1
# CALCULO DE EFECTOS PARA LA RESPUESTA CORREGIDA Y2

```

```

CON2 <- t(X)%*%Y2; EF2 <- CON2/8; EFORD <- sort(EF2)
# ----- TERMINA MÉTODO -----#

```

14.1.5 Método Basado en Técnicas de Regresión Robusta I

```

#----- COMIENZA MÉTODO-----#
NOMBRE <- "MÉT.BASADO EN TÉCNICAS DE REGRESIÓN ROBUSTA I"; PC<-0.034
# SELECCION ROBUSTA DE LA PARTE FIJA DEL MODELO
N <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16)
X <- cbind(X0,X1,X2,X3,X4,X1X2,X1X3,X1X4,X2X3,X2X4,X3X4)
XI <- cbind(X1X2X3,X1X2X4,X1X3X4,X2X3X4,X1X2X3X4)
BETA0 <- solve(t(X)%*%X)%*%t(X)%*%Y1; SCR1 <-sum(abs(Y1-X)%*%BETA0));
RENG1 <- c(1,2,3,4,5,6,7,8,9,10,11,12,13)
for (j in 1:40)
{
RENG <- sample(N, size=13)
XX<-rbind(X[RENG[1],],X[RENG[2],],X[RENG[3],],X[RENG[4],],
X[RENG[5],],X[RENG[6],],X[RENG[7],],X[RENG[8],], X[RENG[9],],
X[RENG[10],],X[RENG[11],],X[RENG[12],],X[RENG[13],])
YY1 <- rbind(Y1[RENG[1]],Y1[RENG[2]],Y1[RENG[3]],Y1[RENG[4]],
Y1[RENG[5]],Y1[RENG[6]],Y1[RENG[7]],Y1[RENG[8]], Y1[RENG[9]],
Y1[RENG[10]],Y1[RENG[11]],Y1[RENG[12]],Y1[RENG[13]])
BETA0 <- solve(t(XX)%*%XX)%*%t(XX)%*%YY1
SCR <- sum(abs(YY1-XX)%*%BETA0))
if (SCR < SCR1) {BETA01<-BETA0; SCR1<-SCR; RENG1 <- RENG}
}
BETA01 <- BETA01[-1]
# MÉTODO ROBUSTO DE ESTIMACIÓN lmRobMM

```

```

BETAR <- rep(0,15); BETABS <- abs(BETA01); BS <- order(BETABS)
for (k in 1:6)
{
XXX<-cbind(X[,BS[10]+1],X[,BS[9]+1],X[,BS[8]+1],X[,BS[7]+1],X[,BS[k]+1])
ROBUS<-lmRobMM(Y1~XXX,robust.control=lmRobMM.robust.control(tlo=0.0001,
tua=1.5e-06, mxr=50,mxf=50, mxs=50,tl=1e-1000,estim="Auto", seed=1313,
level=0.01, efficiency=0.993, sampling="Random",
weight=c("Bisquare","Bisquare")))
BETAR[11-k] <- ROBUS$coef[6]
}
  BETAR[1] <- ROBUS$coef[2]
  BETAR[2] <- ROBUS$coef[3]
  BETAR[3] <- ROBUS$coef[4]
  BETAR[4] <- ROBUS$coef[5]
for (k in 1:5)
{
XXX<-cbind(X[,BS[10]+1],X[,BS[9]+1],X[,BS[8]+1],X[,BS[7]+1],XI[,k])
ROBUS<-lmRobMM(Y1~XXX,robust.control=lmRobMM.robust.control(tlo=0.0001,
tua=1.5e-06, mxr=50,mxf=50, mxs=50,tl=1e-1000,estim="Auto", seed=1313,
level=0.01, efficiency=0.993,sampling="Random",
weight=c("Bisquare","Bisquare")))
  BETAR[10+k] <- ROBUS$coef[6]
}
for (i in 1:10) {BETABS[BS[11-i]]<-BETAR[i]}
for (i in 1:10) {BETAR[i]<-BETABS[i]}
EF2 <- BETAR; EFORD <- sort(EF2)
#-----TERMINA MÉTODO-----#

```


14.1.6 Método Basado en Técnicas de Regresión Robusta II

```

#-----COMIENZA MÉTODO-----#
NOMBRE <- "TÉCNICAS DE REGRESIÓN ROBUSTA II"; PC <- 0.047
# SELECCION ROBUSTA DE LA PARTE FIJA DEL MODELO
XB <- cbind(X1,X2,X3,X4,X1X2,X1X3,X1X4,X2X3,X2X4,X3X4)
XI <- cbind(X1X2X3,X1X2X4,X1X3X4,X2X3X4,X1X2X3X4)
REGL1 <- l1fit(XB,Y1, intercept=T, print=F)
BETA01<-REGL1$coef; BETA01 <- select.rows(BETA01, 2:11)
# METODO ROBUSTO DE ESTIMACION lmRobMM
BETAR <- rep(0,15)
BETABS <- abs(BETA01); BS <- order(BETABS)
for (k in 1:6)
{
XXX<-cbind(XB[,BS[10]],XB[,BS[9]],XB[,BS[8]],XB[,BS[7]],XB[,BS[k]])
ROBUS<-lmRobMM(Y1~XXX,robust.control=lmRobMM.robust.control(tlo=0.0001,
tua=1.5e-06, mxr=50,mxf=50, mxs=50,tl=1e-1000,estim="Auto", seed=1313,
level=0.01, efficiency=0.993, sampling="Random",
weight=c("Bisquare","Bisquare")))
  BETAR[11-k] <- ROBUS$coef[6]
}
BETAR[1] <- ROBUS$coef[2]
BETAR[2] <- ROBUS$coef[3]
BETAR[3] <- ROBUS$coef[4]
BETAR[4] <- ROBUS$coef[5]
for (k in 1:5)
{
XXX <- cbind(XB[,BS[10]],XB[,BS[9]],XB[,BS[8]],XB[,BS[7]],XI[,k])
ROBUS<-lmRobMM(Y1~XXX,robust.control=lmRobMM.robust.control(tlo=0.0001,

```

```

tua=1.5e-06, mxr=50,mxf=50, mxs=50,t1=1e-1000,estim="Auto", seed=1313,
level=0.01, efficiency=0.993, sampling="Random",
weight=c("Bisquare","Bisquare")))
  BETAR[10+k] <- ROBUS$coef[6]
}
for (i in 1:10){BETABS[BS[11-i]]<-BETAR[i]}
for (i in 1:10){BETAR[i]<-BETABS[i]}
EF2 <- BETAR; EFORD <- sort(EF2);
#-----TERMINA MÉTODO-----#

```

14.1.7 Método Bayesiano de Box y Meyer que Admite la Posibilidad de Datos Anómalos

El procedimiento iterativo propuesto por Box y Meyer (1987) comienza por suponer en un primer paso la ausencia de observaciones anómalas. En un segundo paso y con los efectos que hayan resultado más importantes, digamos con probabilidades posteriores de 0.5 o más, se supone temporalmente el modelo correspondiente y se calculan las probabilidades posteriores de que los datos sean anómalos. Luego, en un tercer paso y suponiendo como datos anómalos los que tengan probabilidades posteriores altas (> 0.5) en el paso anterior, se vuelven a calcular las probabilidades posteriores de que los efectos estén activos, y así sucesivamente. Este procedimiento generalmente converge en una o dos iteraciones.

A continuación se presenta el programa paso a paso, y es también como se debe correr el programa, puesto que la salida de un paso es la entrada del siguiente. Incluso se recomienda correr de manera separada cada uno de los bloques donde se obtienen y ordenan los conjuntos potencia de los 15 efectos y 16 observaciones.

Comienza Programa

```
# PASO 1: SUPONIENDO QUE NO HAY DATOS ANÓMALOS SE CALCULAN
```

```

LAS PROBABILIDADES POSTERIORES DE QUE LOS EFECTOS SEAN
ACTIVOS.

# VECTOR Y VALOR DONDE SE GUARDAN LA PROBABILIDAD POSTERIOR
DE CADA EFECTO Y SU ACUMULADO
nn1<-15; ppEF1 <- rep(0,nn1+1); ppT1 <- 0

# INFORMACIÓN A PRIORI TOMADA DE BOX Y MEYER (1987)
alf1 <- 0.2; alf2 <- 0.05; g <- 2.5; h <- 5; fi <- 1-1/h^2
re <- 0 # SE SUPONE QUE NO HAY DATOS ANORMALES

# SE OBTIENE EL CONJUNTO POTENCIA DE LOS nn1=15 EFECTOS
powerSet <- function(x)
{
  K <- NULL
  for(m in x)
    K <- rbind(cbind(K, F), cbind(K, T))
  apply(K, 1, function(x, s) s[x], s = x)
}

xx1 <- powerSet(1:nn1)
xx1 <- xx1[-1]

# SE ORDENA EL CONJUNTO POTENCIA DE LOS nn1=15 EFECTOS
numdat1 <- length(xx1); ps <- rep(0,numdat1)
for (i in 1:numdat1) {ps[i] <- length(xx1[[i]])}
xx1 <- xx1[sort.list(ps)]

# DATOS DE EJEMPLO 1
y <- c(47.46,49.62,43.13,46.31,51.47,48.49,49.34,46.10,
  46.76,48.56,44.83,44.45,59.15,51.33,47.02,47.90)

# EFECTOS Y SU MATRIZ DE SIGNOS
X00 <- c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)

```

```

A <- c(-1, 1,-1, 1,-1, 1,-1, 1,-1, 1,-1, 1,-1,1)
B <- c(-1,-1, 1, 1,-1,-1, 1, 1,-1,-1, 1, 1,-1,-1, 1,1)
C <- c(-1,-1,-1,-1, 1, 1, 1, 1,-1,-1,-1,-1, 1, 1, 1,1)
D <- c(-1,-1,-1,-1,-1,-1,-1,-1, 1, 1, 1, 1, 1, 1,1)
E <- A*B; G <- A*C; H <- A*D; I <- B*C; J <- B*D; K <- C*D
L <- A*B*C; M <- A*B*D; N <- A*C*D; O <- B*C*D; P<-A*B*C*D
Cr <- cbind(A,B,C,D,E,G,H,I,J,K,L,M,N,O,P)

# PROBABILIDADES POSTERIORES PARA MODELO SIN TÉRMINOS
ppEF1[1] <- 1; ppT1 <- 1

# PROBABILIDADES POSTERIORES PARA MODELOS CON HASTA 7
EFECTOS (16383 MODELOS)
for(i in 1:16383)
  {
    ps <- xx1[[i]]; te <- length(ps)
    Xr1 <- cbind(X00,Cr[,ps]); Xr1r2 <- rep(0,te+1);
    yr2 <- 0; gg <- matrix(0,te+1,te+1)
    for (ii in 1:te) {gg[ii+1,ii+1] <- 1}; gg1 <- (1/g^2)*gg
    tr1r2 <- solve(gg1+t(Xr1)%*%Xr1-fi*t(t(Xr1r2))%*%Xr1r2)
      %*%(t(Xr1)%*%y-fi*t(t(Xr1r2))%*%yr2)
    Sr1r2 <- t(y-Xr1%*%tr1r2)%*%(y-Xr1%*%tr1r2)
      -fi*t(yr2-Xr1r2%*%tr1r2)%*%(yr2-Xr1r2%*%tr1r2)
    pp1 <- ((alf1*g^-1)/(1-alf1))^te*((alf2*h^-1)/(1-alf2))^re
    pp2 <- 4/det(gg1+t(Xr1)%*%Xr1-fi*t(t(Xr1r2))%*%Xr1r2)^0.5
    pp3 <- ((Sr1r2+t(tr1r2)%*%gg1%*%tr1r2)
      /t(y-mean(y))%*%(y-mean(y)))^-7.5
    ppEF1[ps+1] <- ppEF1[ps+1] + pp1*pp2*pp3
    ppT1 <- ppT1+ pp1*pp2*pp3
  }

```

```

# NORMALIZACIÓN Y GRAFICACIÓN DE LAS PROBABILIDADES
POSTERIORES
posterior <- ppEF1/ppT1
names <- c("0","A", "B","C","D","AB", "AC","AD","BC","BD",
          "CD", "ABC", "ABD", "ACD", "BCD","ABCD")
barplot(posterior,names=names,xlab="Efectos, Sin Suponer
Datos Anómalos", ylab="Prob. Posterior de Estar Activo")
### TERMINA PASO 1 ###

```

Del gráfico resultante de este primer paso se toman temporalmente como importantes los efectos con probabilidades posteriores altas, digamos mayores a 0.5, sin que esto implique que esta regla se deba aplicar de manera estricta. Con este modelo se entra al paso 2.

```

# PASO 2: CALCULANDO LAS POSTERIORES DE QUE LOS DATOS SEAN
ANÓMALOS, SUPONIENDO POR EL MOMENTO EL MODELO IDENTIFICADO EN
EL PASO ANTERIOR.

# CONSTRUCCIÓN DE MATRICES Xr1 y gg PARA EL MODELO
SELECCIONADO EN PASO 1
Xr1 <- cbind(X00,Cr[,2],Cr[,3])
gg <- matrix(0, 3, 3); gg[2,2] <- 1; gg[3,3] <- 1;
te <- 2 # DOS TÉRMINOS ACTIVOS SE DETECTARON EN EL PASO 1

# CALCULANDO EL CONJUNTO POTENCIA DE LOS nn2=16 DATOS
powerSet <- function(x)
{
  K <- NULL
  for(m in x)
    K <- rbind(cbind(K, F), cbind(K, T))
  apply(K, 1, function(x, s) s[x], s = x)
}

```

```

}

nn2 <- 16
xx2 <- powerSet(1:nn2)
xx2 <- xx2[-1]

# SE ORDENA EL CONJUNTO POTENCIA DE LOS nn2=16 DATOS
numdat2 <- length(xx1); ps <- rep(0,numdat2)
for (i in 1:numdat2) {ps[i] <- length(xx2[[i]])}
xx2 <- xx2[sort.list(ps)]

# CONTADORES Y ACUMULADORES
pYY <- rep(0,nn2) # PROBABILIDADES POSTERIORES DE QUE
CADA DATO SEA ANÓMALO
ppyy <- 0 # ACUMULA TODAS LAS PROBABILIDADES POSTERIORES

# INICIA CÁLCULO DE PROBABILIDADES POSTERIORES USANDO 14892
MODELOS, QUE SON LOS CONJUNTOS CON A LO MÁS 6 DATOS ANÓMALOS
for(i in 1:14892)
{
  ps <- xx2[[i]]
  re <- length(ps)
  if (re == 1)
  {
    Xr1r2 <- Xr1[ps,]
    yr2 <- as.matrix(y[ps])
    tr1r2 <- solve(gg+t(Xr1)%*%Xr1-fi*t(t(Xr1r2))%*%Xr1r2)
      %*%(t(Xr1)%*%y-fi*t(t(Xr1r2))%*%yr2)
    Sr1r2 <- t(y-Xr1)%*%tr1r2)%*%(y-Xr1)%*%tr1r2)
      -fi*t(yr2-Xr1r2)%*%tr1r2)%*%(yr2-Xr1r2)%*%tr1r2)
  }
  pp1 <- ((alf1*g^-1)/(1-alf1))^te*((alf2*h^-1)/(1-alf2))^re
  pp2 <- 4.0/sqrt(det(gg+t(Xr1)%*%Xr1-fi*t(t(Xr1r2))%*%Xr1r2))
}

```

```

pp3 <- ((Sr1r2+t(tr1r2)%*%gg%*%tr1r2)/
        t(y-mean(y))%*%(y-mean(y)))^(-7.5)
    }
    else
    {
        Xr1r2 <- Xr1[ps,]
        yr2 <- as.matrix(y[ps])
        tr1r2 <- solve(gg+t(Xr1)%*%Xr1-fi*t(Xr1r2)%*%Xr1r2)
            %*%(t(Xr1)%*%y-fi*t(Xr1r2)%*%yr2)
        Sr1r2 <- t(y-Xr1)%*%tr1r2)%*%(y-Xr1)%*%tr1r2)
            -fi*t(yr2-Xr1r2)%*%tr1r2)%*%(yr2-Xr1r2)%*%tr1r2)

pp1 <- ((alf1*g^-1)/(1-alf1))^te*((alf2*h^-1)/(1-alf2))^re
pp2 <- 4.0/sqrt(det(gg+t(Xr1)%*%Xr1-fi*t(Xr1r2)%*%Xr1r2))
pp3 <- ((Sr1r2+t(tr1r2)%*%gg%*%tr1r2)/
        t(y-mean(y))%*%(y-mean(y)))^(-7.5)
    }

pYY[ps] <- pYY[ps] + pp1*pp2*pp3
ppyy <- ppyy + pp1*pp2*pp3
}

# NORMALIZACIÓN Y GRAFICACIÓN DE PROBABILIDADES POSTERIORES
posterior <- pYY/ppyy
names <- paste("Ren",1:nn2,sep=".")
barplot(posterior,names=names,srt=90,xlab="Observaciones,
Suponiendo Modelo B y C", ylab="Prob. Posterior de Ser
Anómalo")
### TERMINA PASO 2 ###

```

Las observaciones que hayan resultado anómalas en el segundo paso se suponen como tal y se vuelven a calcular las probabilidades posteriores de que los efectos están activos. Esto se hace en el tercer paso.

```

# PASO 3: SE VUELVEN A CALCULAR LAS PROBABILIDADES
POSTERIORES DE QUE LOS EFECTOS ESTÁN ACTIVOS SUPONIENDO
LAS OBSERVACIONES ANÓMALAS DETECTADAS EN EL PASO ANTERIOR.

# SE DEFINEN LAS OBSERVACIONES ANÓMALAS ENCONTRADAS EN EL
PASO 2, EN ESTE CASO ES SOLO LA OBSERVACIÓN 13
YA1 <- 13; re <- 1

# SE DEFINE EL VECTOR Y EL ACUMULADO QUE GUARDA LAS
PROBABILIDADES POSTERIORES SIN NORMALIZAR
ppEF3 <- rep(0,nn1+1)
ppT3 <- 0

# PROBABILIDADES POSTERIORES PARA MODELO SIN TÉRMINOS
(MODELO CONSTANTE)

te <- 0

Xr1 <- cbind(X00); Xr1r2 <- Xr1[YA1,] ; yr2 <- y[YA1]
gg <- matrix(0,1,1); gg1 <- (1/g^2)*gg

tr1r2 <- solve(gg1+t(Xr1)%*%Xr1-fi*t(t(Xr1r2))
              %*%Xr1r2)%*(t(Xr1)%*%y-fi*t(t(Xr1r2))%*%yr2)
Sr1r2 <- t(y-Xr1)%*%tr1r2)%*(y-Xr1)%*%tr1r2
        -fi*t(yr2-Xr1r2)%*%tr1r2)%*(yr2-Xr1r2)%*%tr1r2)
pp1 <- ((alf1*g^-1)/(1-alf1))^te*((alf2*h^-1)/(1-alf2))^re
pp2 <- 4/det(gg1+t(Xr1)%*%Xr1-fi*t(t(Xr1r2))%*%Xr1r2)^0.5
pp3 <- ((Sr1r2+t(tr1r2)%*%gg1)%*%tr1r2)/
        t(y-mean(y))%*(y-mean(y)))^-7.5

```



```

ppEF3[1] <- pp1*pp2*pp3
ppT3 <- pp1*pp2*pp3

# PROBABILIDADES POSTERIORES PARA MODELOS HASTA CON 7
TÉRMINOS (16383 MODELOS)
for(i in 1:16383)
{
  ps <- xx1[[i]]
  te <- length(ps)

  Xr1 <- cbind(X00,Cr[,ps]); Xr1r2 <- Xr1[YA1,] ;
  yr2 <- y[YA1]; gg <- matrix(0,te+1,te+1)
  for (ii in 1:te) {gg[ii+1,ii+1] <- 1}
  gg1 <- (1/g^2)*gg

  tr1r2 <- solve(gg1+t(Xr1)%*%Xr1-fi*t(t(Xr1r2))%*%Xr1r2)
  %*%(t(Xr1)%*%y-fi*t(t(Xr1r2))%*%yr2)
  Sr1r2 <- t(y-Xr1)%*%tr1r2)%*%(y-Xr1)%*%tr1r2)
  -fi*t(yr2-Xr1r2)%*%tr1r2)%*%(yr2-Xr1r2)%*%tr1r2)
  pp1 <- ((alf1*g^-1)/(1-alf1))^te*((alf2*h^-1)/(1-alf2))^re
  pp2 <- 4/det(gg1+t(Xr1)%*%Xr1-fi*t(t(Xr1r2))%*%Xr1r2)^0.5
  pp3 <- ((Sr1r2+t(tr1r2)%*%gg1)%*%tr1r2)/
  t(y-mean(y))%*%(y-mean(y)))^-7.5
  ppEF3[ps+1] <- ppEF3[ps+1] + pp1*pp2*pp3
  ppT3 <- ppT3 + pp1*pp2*pp3
}

# NORMALIZACIÓN Y GRAFICACIÓN DE LAS PROBABILIDADES POSTERIORES
posterior3 <- ppEF3/ppT3
names <- c("0", "A", "B", "C", "D", "AB", "AC", "AD", "BC",

```

```

"BD", "CD", "ABC", "ABD", "ACD", "BCD", "ABCD")
barplot(posterior3,names=names,xlab="Efectos, Suponiendo
Observación 13 Anómala", ylab="Prob. Posterior de Estar
Activo")
### TERMINA PASO 3 ###

```

Lo que procede a continuación es suponer el modelo que haya resultado importante y volver a calcular las probabilidades posteriores de que los datos sean anómalos (Paso 2), declarando las matrices $X_{(r_1)}$ y $gg = \Gamma_{r_1}$ al comienzo de este paso como

```

Xr1 <- cbind(X00,Cr[,2],Cr[,3],Cr[,6],Cr[,13])
gg <- matrix(0, 5, 5);
gg[2,2] <- 1; gg[3,3] <- 1; gg[4,4] <- 1; gg[5,5] <- 1
te <- 4 # CUATRO TÉRMINOS ACTIVOS SE DETECTARON EN EL PASO 1

```

en el caso de un modelo tentativo con cuatro términos.

14.2 Detección Bayesiana de Efectos Activos Cuando la Respuesta es No-Normal

En esta sección se presentan los programas desarrollados en SPLUS para en análisis de experimentos factoriales cuando la respuesta es no-normal. Básicamente estos programas resuelven o aproximan vía simulación Monte Carlo las integrales múltiples que definen la distribución posterior de cada modelo considerado. Finalmente, marginalizando, se obtienen las probabilidades posteriores de que los efectos (coeficientes del modelo) sean activos.

Se obtienen números aleatorios (nd_{ij}) de las distribuciones *a priori* de los parámetros, a partir de números pseudo-aleatorios o secuencias de baja discrepancia (h_{ij}). Se recomiendan éstos últimos para resolver problemas de integración en varias variables, como es el caso.

14.2.1 Regresión de Poisson, Liga Identidad

```

# REGRESION DE POISSON BAYESIANA
# RESULTADOS DEL PROGRAMA: PROBABILIDADES POSTERIORES DE EFECTOS
# remove(ls())
# MUESTREANDO LAS DISTRIBUCIONES A PRIORI
  ud1 <- qgamma(h11,256, rate=25.6)
  g <- 1.25
  nd22 <- g*qnorm(h22); nd32 <- g*qnorm(h32); nd33 <- g*qnorm(h33)
  nd42 <- g*qnorm(h42); nd43 <- g*qnorm(h43); nd44 <- g*qnorm(h44)
  nd52 <- g*qnorm(h52); nd53 <- g*qnorm(h53); nd54 <- g*qnorm(h54)
  nd55 <- g*qnorm(h55)
# POSIBLES MODELOS QUE SE PUEDEN CONSTRUIR CON 15 TÉRMINOS
  powerSet <- function(x)
  {
    K <- NULL
    for(m in x)
      K <- rbind(cbind(K, F), cbind(K, T))
    apply(K, 1, function(x, s) s[x], s = x)
  }
  nn <- 15
  xx <- powerSet(1:nn)
  xx <- xx[-1]
  numdat <- length(xx); ti <- rep(0,numdat)
  for (i in 1:numdat) {ti[i] <- length(xx[[i]])}
  xx <- xx[sort.list(ti)]
  numdat <- 120
  for (i in 1:numdat) {ti[i] <- length(xx[[i]])}
# EXPERIMENTO Y DATOS

```

```

n <- 15
X0 <- c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)
X1 <- c(-1, 1,-1, 1,-1, 1,-1, 1,-1, 1,-1, 1,-1, 1,-1, 1)
X2 <- c(-1,-1, 1, 1,-1,-1, 1, 1,-1,-1, 1, 1,-1,-1, 1, 1)
X3 <- c(-1,-1,-1,-1, 1, 1, 1, 1,-1,-1,-1,-1, 1, 1, 1, 1)
X4 <- c(-1,-1,-1,-1,-1,-1,-1,-1, 1, 1, 1, 1, 1, 1, 1, 1)
X5 <- X2*X4; X6 <- X2*X3*X4; X7 <- X1*X3
X8 <- X1*X3*X4; X9 <- X1*X2; X10 <- X1*X4
X11 <- X2*X3; X12 <- X3*X4; X13 <- X1*X2*X3
X14 <- X1*X2*X4; X15 <- X1*X2*X3*X4
X <- cbind(X1,X2,X3,X4,X5,X6,X7,X8,X9,X10,X11,X12,X13,X14,X15)
y <- c(56,17,2,4,3,4,50,2,1,0,3,12,3,4,0,0)
# ACUMULADORES Y PROBABILIDAD A PRIORI DE ESTAR ACTIVO
pEF <- rep(0,nn+1)
ppT <- 0; alfa <- 0.2
# PROBABILIDAD POSTERIOR PARA EL MODELO SIN TÉRMINOS
gk <- 0;
for (k in 1:200)
{
  b0 <- ud1[k]; xtb <- rep(0,16);
  for (j in 1:16)
  {
    xtb[j] <- exp(-b0)*(b0)^y[j];
  }
  gk <- gk + prod(xtb)
}
ppT <- (gk/200); pEF[1] <- gk/200
ppb0 <- ppT; ppT <- 1; pEF[1] <- 1

```

```

# PROBABILIDAD POSTERIOR DE MODELOS CON un TÉRMINO
# 1-15
MUESTRA <- 200;
for(i in 1:15)
{
  ps <- xx[[i]]
  gk <- 0
  for (k in 1:MUESTRA)
  {
    b0 <- ud1[k]; bk <- nd22[k];
    BI <- c(b0,bk); XI <- cbind(X0,X[,ps])
    xtb <- rep(0,16);
    for (j in 1:16)
    {
      xtb[j] <- exp(-b0)*(XI[j,]%*%BI)^y[j];
    }
    gk <- gk + (prod(xtb))
  }
  pp <- ((alfa/(1-alfa))^ti[i])*gk/(MUESTRA*ppb0)
  pEF[ps+1] <- pEF[ps+1] + pp
  ppT <- ppT + pp
  print(i)
}

# PROBABILIDAD POSTERIOR DE MODELOS CON dos TÉRMINOS
# 16-120
for(i in 16:120)
{
  ps <- xx[[i]]

```

```

gk <- 0
for (k in 1:MUESTRA)
{
  b0 <- ud1[k]; bk <- c(nd32[k],nd33[k]);
  BI <- c(b0,bk); XI <- cbind(X0,X[,ps])
  xtb <- rep(0,16);
  for (j in 1:16)
  {
    xtb[j] <- exp(-b0)*(XI[j,]%*%BI)^y[j];
  }
  gk <- gk + (prod(xtb))
}
pp <- ((alfa/(1-alfa))^ti[i])*gk/(MUESTRA*ppb0)
pEF[ps+1] <- pEF[ps+1] + pp
ppT <- ppT + pp
print(i)
}

# PROBABILIDAD POSTERIOR DE MODELOS CON tres TÉRMINOS
# 121-575
for(i in 421:575)
{
  ps <- xx[[i]]
  gk <- 0
  for (k in 1:MUESTRA)
  {
    b0 <- ud1[k]; bk <- c(nd42[k],nd43[k],nd44[k]);
    BI <- c(b0,bk); XI <- cbind(X0,X[,ps])
    xtb <- rep(0,16);

```

```

    for (j in 1:16)
    {
      xtb[j] <- exp(-b0)*(XI[j,]%*%BI)^y[j];
    }
    gk <- gk + (prod(xtb))
  }
  pp <- ((alfa/(1-alfa))^ti[i])*gk/(MUESTRA*ppb0)
  pEF[ps+1] <- pEF[ps+1] + pp
  ppT <- ppT + pp
  print(i)
}

# PROBABILIDAD POSTERIOR DE MODELOS CON cuatro TÉRMINOS
# 576-1940
for(i in 1276:1375)
{
  ps <- xx[[i]]
  gk <- 0
  for (k in 1:MUESTRA)
  {
    b0 <- ud1[k]; bk <- c(nd52[k],nd53[k],nd54[k],nd55[k]);
    BI <- c(b0,bk); XI <- cbind(X0,X[,ps])
    xtb <- rep(0,16);
    for (j in 1:16)
    {
      xtb[j] <- exp(-b0)*(XI[j,]%*%BI)^y[j];
    }
    gk <- gk + (prod(xtb))
  }
}

```

```

pp <- ((alfa/(1-alfa))^ti[i])*gk/(MUESTRA*ppb0)
pEF[ps+1] <- pEF[ps+1] + pp
ppT <- ppT + pp
print(i)
}

# PROBABILIDADES POSTERIORES DE EFECTOS
names <- c("O","A", "B","C","D","E", "F","G","H","J",
          "AD","BC","CD","BG", "AE","AF")
barplot(pEF/ppT, names=names)
#-----TERMINA PROGRAMA-----#

```

14.2.2 Regresión de Poisson, Liga Log

En el programa anterior solo se cambia la expresión de la verosimilitud por la que corresponde a esta liga. Basta entonces que enlistemos uno de los ciclos, por ejemplo para calcular las probabilidades posteriores de los modelos con cuatro términos.

```

# PROBABILIDAD POSTERIOR DE MODELOS CON cuatro TÉRMINOS
# 576-1940
for(i in 576:1940)
{
  ps <- xx[[i]]
  gk <- 0
  for (k in 1:MUESTRA)
  {
    b0 <- ud1[k];
    bk <- c(nd52[k],nd53[k],nd54[k],nd55[k]);
    BI <- c(b0,bk); XI <- cbind(X0,X[,ps])
    xtb <- rep(0,16);
    for (j in 1:16)

```



```

    {
        xtb[j] <- -exp(XI[j,]%*%BI)+ y[j]*(XI[j,]%*%BI)
    }
    gk <- gk + exp(sum(xtb))
}
pp <- ((alfa/(1-alfa))^ti[i])*gk/(MUESTRA*ppb0)
pEF[ps+1] <- pEF[ps+1] + pp
ppT <- ppT + pp
print(i)
}

```

14.2.3 Regresión de Poisson, Liga Raíz Cuadrada

En este caso las probabilidades posteriores para los modelos con cuatro términos se obtienen en el siguiente extracto del programa.

```

# PROBABILIDAD POSTERIOR DE MODELOS CON cuatro TÉRMINOS
for(i in 576:1940)
{
    ps <- xx[[i]]
    gk <- 0
    for (k in 1:MUESTRA)
    {
        b0 <- ud1[k];
        bk <- c(nd52[k],nd53[k],nd54[k],nd55[k]);
        BI <- c(b0,bk); XI <- cbind(X0,X[,ps])
        xtb <- rep(0,16);
        for (j in 1:16)
        {
            xtb[j] <- exp(-(XI[j,]%*%BI)^2)*((XI[j,]%*%BI))^(2*y[j]);

```

```

    }
    gk <- gk + prod(xtb)
  }
  pp <- ((alfa/(1-alfa))^ti[i])*gk/(MUESTRA*ppb0)
  pEF[ps+1] <- pEF[ps+1] + pp
  ppT <- ppT + pp
  print(i)
}

```

14.3 Regresión Logística

El ciclo para estimar las probabilidades posteriores de modelos con cuatro términos es el siguiente:

```

# PROBABILIDAD POSTERIOR DE MODELOS CON cuatro TÉRMINOS
for(i in 64:98)
{
  ps <- xx[[i]]
  gk <- 0
  for (k in 1:MUESTRA)
  {
    b0 <- ud1[k];
    bk <- c(nd52[k],nd53[k],nd54[k],nd55[k]);
    BI <- c(b0,bk); XI <- cbind(X0,X[,ps])
    xtb <- rep(0,8);
    for (j in 1:8)
    {
      xtb[j]<-((exp(XI[j,]%*%BI))^y[j])*(1/(1+exp(XI[j,]%*%BI)))^50;
    }
  }
}

```

```

    gk <- gk + prod(xtb)
  }
  pp <- ((alfa/(1-alfa))^ti[i])*gk/(MUESTRA*ppb0)
  pEF[ps+1] <- pEF[ps+1] + pp
  ppT <- ppT + pp
  print(i)
}

```

14.4 Modelo Gama

Se enlista el ciclo aproximar las probabilidades posteriores de modelos con cuatro términos. Cabe recordar que el valor del hiperparámetro r se ajusta lo más grande posible cuidando que las integrales no se disparen a infinito. Basta generalmente con verificar esto en el modelo nulo que solo tiene la constante β_0 .

```

# PROBABILIDAD POSTERIOR DE MODELOS CON cuatro TÉRMINOS
for(i in 176:385)
{
  ps <- xx[[i]]
  gk <- 0
  for (k in 1:MUESTRA)
  {
    b0 <- ud1[k];
    bk <- c(nd52[k],nd53[k],nd54[k],nd55[k]);
    BI <- c(b0,bk); XI <- cbind(X0,X[,ps])
    xtb <- rep(0,16);
    for (j in 1:16)
    {
      xtb[j] <- (-r*XI[j,]%*%BI)-y[j]*exp(-XI[j,]%*%BI)
    }
  }
}

```

```
    }  
    gk <- gk + exp(sum(xtb))  
  }  
  pp <- ((alfa/(1-alfa))^ti[i])*gk/(MUESTRA*ppb0)  
  pEF[ps+1] <- pEF[ps+1] + pp  
  ppT <- ppT + pp  
  print(i)  
}
```

Referencias

Aguirre-Torres V. y Pérez-Trejo M. E. (2001). Outliers and the use of the rank transformation to detect active effects in unreplicated 2^f experiments. *Communications in Statistics* 30, pp. 637-663.

Aguirre-Torres V. (1993). A simple analysis of unreplicated factorials with possible abnormalities. *Journal of Quality Technology* 25, pp. 183-187.

Aguirre-Torres V. (1998). Impact of outliers on methods for detecting active effects in unreplicated experiments. *Technical Report I-98-22*, Centro de Investigación en Matemáticas, Guanajuato, Gto., 34 pp.

Benski, H. C. y Cabau, E. (1995). Unreplicated experimental designs in reliability growth programs. *IEEE Trans. Reliability* 44, 199-205.

Benski, H. C. (1989). Use of a normality test to identify significant effects in factorial designs. *Journal of Quality Technology* 21, pp. 174-178.

Berk, K. y R. Picard (1991). Significance tests for saturated orthogonal arrays. *Journal of Quality Technology* 23, pp. 79-89.

Bernardo, J. M. y Smith, A. F. M. (1994). *Bayesian Theory*. Wiley, New York.

Box, G. E. P. y Meyer, R. (1986). An analysis for unreplicated fractional factorials. *Technometrics* 28, pp. 11-18.

Box, G. E. P. y Meyer R. (1987). Analysis of unreplicated factorials allowing for possible faulty observations. In *Design, Data and Analysis*. Mallow, C. Ed. Wiley, New York.

Box, G. E. P. y Meyer, R. (1993). Finding the active factors in fractionated screening experiments. *Journal of Quality Technology* 25, pp. 94-104.

Box, G. E. P. y Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Wiley, New York.

Box, G. E. P. (1990). A simple way to deal with missing observations from designed experiments. *Quality Engineering* 3, pp. 249-254.

Box, G. E. P. (1991). Finding bad values in factorial designs. *Quality Engineering* 3, pp. 405-410

Chen, Y. (2003). On the analysis of unreplicated factorial designs. PhD Thesis, Dortmund, Germany.

Clyde, M. (1999). Bayesian model averaging and model search strategies (with discussion). In *Bayesian Statistics 6*. J. M. Bernardo, A. P. Dawid, J. O. Berger and A. F. M. Smith Eds. Oxford University Press, pp. 157-185.

Czado, C. y Raftery, A. E. (2005). Choosing the link function and accounting for link uncertainty in generalized linear models using Bayes factors. *Statistical Papers*, to appear.

Daniel, C. (1959). Use of half normal plots in interpreting factorial two-level experiments. *Technometrics*, 1, 311-341.

Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation*. John Wiley, New York.

De la Vara, R. (2004). Probabilidades posteriores para identificar efectos activos en experimentos factoriales sin y con posibles datos anómalos. *Comunicación Técnica del CIMAT* No. I-04-09/18-09-2004.

Evans, M. y Swartz, (1995). Methods for approximating integrals with special emphasis on Bayesian integration problems. *Statistical Science* 10, 254-272.

- Hamada, M. y Nelder, J. A. (1997). Generalized linear models for quality-improvement experiments. *Journal of Quality Technology* 29, pp. 292-303.
- Hamada M. y N. Balakrishnan (1998). Analyzing unreplicated factorial experiments: a review with some new proposals. *Statistica Sinica* 8, pp. 1-41.
- Hoeting, J. A., Madigan, D., Raftery, A. E. y Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical Science* 14, 382-417.
- Huber, P. J. (1973). Robust Regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics* 1, pp. 799-821.
- Kaufman, L. y Rousseeuw (1993). Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York.
- Kraber, S. (1999). A case to test your metal (results). *Stat-Teaser*. News from Stat-Ease, Inc. December.
- Krykova, I. (2003). Evaluating of Path-Dependent Securities with Low Discrepancy Methods. Master Thesis, Worcester, Massachusetts.
- Lenth, R. V. (1989). Quick and easy analysis of unreplicated factorial experiments. *Technometrics* 31, pp. 469-473.
- Lewis, S. L. y Montgomery, D. C. (2001). Examples of designed experiments with nonnormal responses. *Journal of Quality Technology* 33, pp. 265-278.
- McCullagh, P. y Nelder, J. A. (1989). Generalized Linear Models, 2nd ed., Chapman and Hall, New York.
- Meyer, R. y Box, G. E. P. (1992). Finding the active factors in fractionated screening experiments. *CQPI Report* No. 80, april 1992. University of Wisconsin, Madison.

- Meyer, R. (1987). Further details of an analysis for unreplicated fractional factorials. *CQPI Report* No. 22, february 1987. University of Wisconsin, Madison.
- Montgomery, D. C. and E. A. Peck (1992). *Introduction to Linear Regression Analysis*. Second Edition. Wiley, New York.
- Montgomery, D. C. (2001). *Design and Analysis of Experiments*. 5th Edition, Wiley, New York.
- Myers, R. H. y Montgomery, D. C. (1997). A tutorial on generalized linear models. *Journal of Quality Technology* 29, pp. 274-291.
- Myers, R. H., Montgomery, D. C. y G. Vining (2002). *Generalized Linear Models*. Wiley, New York.
- Niederreiter, H. (1992). Random Numbers Generation and Quasi-Monte Carlo Methods. *CBMS-NSF Regional Conference Series in Applied Mathematics*, Vol. 63, SIAM, Philadelphia.
- Papageorgiou, A. y Traub, J. F. (1997). Faster evaluation of multidimensional integrals. *Computers in Physics*, november, pp. 574-578.
- Kaufman, L. y P. J. Rousseeuw (1993). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.
- Raftery, A. E. (1995). Bayesian model selection in social research (with discussion) *Sociological Methodology* 25, pp. 111-196.
- Robert, C. P. y Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer Verlag, New York.
- Rousseeuw, P. J. y A. M. Leroy (1987). *Robust Regression and Outlier Detection*. Wiley, New York.

Rousseeuw, P. J. y Yohai, V. J. (1984). Robust Regression by means of S-estimators. In *Robust and Nonlinear Time Series*, edited by J. Franke, W. Hardle and D. Martin, Lecture Notes in Statistics 26, pp. 256-272. Springer Verlag, New York.

Sebert, D. M., Montgomery, D. C. y D. A. Rollier (1997). A clustering algorithm for identifying multiple outliers in linear regression. *Computational Statistics and Data Analysis* 27, pp. 461-484.

Shapiro, S. S. y R.S. Francia (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association* 67, pp. 215-216.

SPLUS 6.1 for Windows, Professional Edition, Release 1, 2002.

Wang, X. y Chen, M. (2004). Approximate predictive densities and fully bayes variable selection in generalized linear models. *Technical Report SMU-TR-322*, Southern Methodist University, Dallas, Tx.

Wu, C. F. J. y M. Hamada (2000). Experiments: Planning, Analysis, and Parameter Design Optimization. Wiley, New York.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics* 15, 20, 642-656.

Yohai, V. J, Stahel, W. A. y Zamar, R. H. (1991). A procedure for robust estimation and inference in linear regression. In *Directions in Robust Statistics and Diagnostics, Part II*. Eds. Stahel, W. A. y Weisber, S. Springer Verlag, New York.