



03061
2

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

UNIDAD ACADÉMICA DE LOS CICLOS PROFESIONAL Y DE POSGRADO
INSTITUTO DE INVESTIGACION EN MATEMÁTICAS APLICADAS
Y SISTEMAS

COMPARACION DE METODOS PARA
MODELAR LA VARIANZA EN
PROCESOS INDUSTRIALES

TESIS

Que para obtener el grado de
MAESTRA EN ESTADÍSTICA E
INVESTIGACION DE OPERACIONES

presenta

HORTENSIA MORENO MACIAS

Director de tesis: M. en C. Jorge Domínguez Domínguez

Ciudad Universitaria, Marzo de 2001



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*"No se puede pensar en terminar, porque apuntar a las estrellas, tanto en el sentido literal como en el figurado, es un problema que ocupará a varias generaciones; y así, por mucho que se avance, siempre se sentirá la emoción de estar apenas empezando"*¹

Deseo agradecer, con gran cariño:

A Jorge, por su amistad y su invaluable apoyo y dirección en la elaboración de este trabajo.

A todos mis profesores de la maestría, muy en especial al Dr. Ignacio Méndez, por su confianza y motivación.

A toda la Educación Pública y en particular a la maravillosa U.N.A.M por ofrecerme la oportunidad de ser mejor.

A Cristina Escobar por su orientación en los momentos más importantes de decisión.

¹ Fragmento de una carta enviada en 1932 por Robert Goddard, pionero de la investigación espacial, a H.G. Wells, escritor de Ciencia Ficción.

*Una dedicatoria a Luis, mi compañero de proyectos,
de triunfos y también de frustraciones.*

FE DE VIDA

*Desde mis ojos que miran
desde mis manos que tocan
desde mi lengua y mi boca,
doy fe de vida,
de vida por todas partes,
de vida por donde quiera,
de cualquier modo y manera
a pesar de los pesares.*

*Vida, vida de mi vida
vida que da vida, vida...
vida que apresa, vida que ríe
vida que besa.*

*vida, vida bienvenida
vida que convida, vida...
vida que duele, vida que mata,
vida que muere.*

*Vida, vida compañera,
vida por bandera
vida, vida, vida, vida*

*Desde los sueños pendientes
y los fracasos cumplidos
la memoria y el olvido
dan fe de vida.*

*En las luces y en la sombra
con su miel y su vinagre
a mi espejo miserable
viene a asomarse la vida*

Vida, vida de mi vida...

J.M. Serrat

Contenido

1	Efectos de dispersión	1
1.1	Antecedentes	1
1.2	Modelación de la variabilidad	3
1.3	Objetivos	3
1.4	Contenido de la tesis	4
2	Exploración de Efectos de Dispersión	7
2.1	El modelo clásico	7
2.2	Motivación	8
2.3	Ejemplos	15
2.3.1	Ejemplo: Presión arterial diastólica	15
2.3.2	Ejemplo: Árboles de cereza	28
2.3.3	Ejemplo: Resortes automotrices	32
2.3.4	Ejemplo: Circuitos integrados	47
2.3.5	Comentarios	54
3	Estudio de efectos de dispersión	55
3.1	Antecedentes	55
3.2	El modelo	57
3.3	Notación	62
4	Modelación de efectos de dispersión	65
4.1	Introducción.	65
4.2	Procedimientos de estimación	66
4.2.1	Regresión en dos etapas (2E)	67
4.2.2	Mínimos Cuadrados Generalizados (MCG)	70
4.2.3	Máxima Verosimilitud (MV)	75

4.2.4	Pseudoverosimilitud (PV)	79
4.2.5	Quasiverosimilitud (QV)	80
4.3	Aplicación de los métodos de estimación	85
5	Métodos Montecarlo	91
5.1	Introducción	91
5.2	Experimento I	93
5.2.1	El valor esperado.	93
5.2.2	La varianza	94
5.3	Experimento II	100
5.4	Algoritmos	101
5.4.1	Regresión en dos etapas.	101
5.4.2	Mínimos cuadrados generalizados	102
5.4.3	Máxima verosimilitud	103
5.4.4	Pseudoverosimilitud	103
5.4.5	Quasiverosimilitud	104
6	Presentación de resultados	107
6.1	Introducción	107
6.2	Experimento I	107
6.3	Experimento II	113
7	Conclusiones	119
	Referencias bibliográficas	123
	Apéndice A	127
A.1	Presión arterial diastólica	127
A.2	Árboles de cereza	128
A.3	Resortes automotrices	129
A.4	Circuitos integrados	129
	Apéndice B	131
B.1	Tablas de resultados del Experimento I	131
B.2	Tablas de resultados del Experimento II	134
	Apéndice C	137
C.1	Programas de simulación	137

Capítulo 1

Efectos de dispersión

1.1 Antecedentes

En todo proceso productivo intervienen diversos factores que pueden o no tener un efecto importante en las características del artículo terminado. En una situación ideal, los industriales desearían contar con un proceso en el que todos los factores estuviesen bajo control absoluto de manera tal que cada unidad producida fuese idéntica al prototipo establecido por el cliente. Una de las tareas importantes del control estadístico de la calidad es identificar y estimar los efectos de los factores que influyen, de manera sistemática, en el valor promedio de una característica para mantenerlos en los niveles adecuados y poder ofrecer al consumidor, un producto que cumpla ciertos requerimientos específicos.

Sin embargo, está claro que la temperatura ambiente, la precisión de los instrumentos de medición, los cambios físicos y químicos de la materia prima y el desgaste de la maquinaria son algunos ejemplos de factores que no son controlables y que hacen diferentes a los elementos al interior de las poblaciones. Estos factores son considerados aleatorios y son los que, tradicionalmente, asumen la responsabilidad de provocar la variabilidad en las características del producto, no obstante, existen factores de control que presentan, de manera simultánea, un efecto sistemático sobre la media y la varianza de las observaciones, es más, en algunos casos, un factor no refleja su efecto sobre el valor esperado pero sí tiene una influencia importante en la dispersión de la variable respuesta que se hace evidente en la calidad del producto.

El número de unidades que requiere un reproceso o que deben ser eliminadas por *no permanecer dentro de los límites de control* es mayor conforme la variabilidad del proceso aumenta, en consecuencia, el costo de la producción es más alto.

El reto, entonces, es lograr un proceso generador de productos que cumplan las especificaciones sobre la característica media pero con la mínima varianza para, en la medida de lo posible, lograr la eliminación de los defectos sistemáticos garantizando artículos de mayor calidad a menor precio. Al respecto, Montgomery (1993) comenta:

*"...Un esfuerzo de mejoramiento de la calidad que tenga éxito puede eliminar muchas de estas pérdidas y conducir a costos menores, mayor productividad, satisfacción creciente del cliente, aumento de la reputación comercial, mayor participación en el mercado y, a la larga, mayores rendimientos para la compañía."*¹

La importancia de la variabilidad en el control de la calidad es indiscutible, no obstante, los métodos estadísticos que se aplican de manera común están orientados a la estimación de los parámetros de un **modelo sólo para la media** en el que se supone que las observaciones provienen de distribuciones normales que pueden diferir en el valor promedio pero no en la varianza. Es decir, sólo se analizan los efectos de los factores sobre la media pero no se estudia su posible **influencia sistemática sobre la varianza**.

La suposición de igualdad de varianzas (homoscedasticidad), no es la única que soporta la teoría de los modelos de regresión clásica, la normalidad de los datos y la independencia de las observaciones también juegan un papel importante. El estudio de las propuestas alternativas ante la violación de cada uno de los supuestos puede resultar por demás extenso por lo que, en este trabajo, el interés se centra en el problema de la presencia de heteroscedasticidad sistemática bajo normalidad e independencia.

La estimación ponderada de los parámetros de los modelos de regresión y la transformación de los datos por medio de la función potencia de Box-Cox son alternativas a las que con mucha frecuencia se recurre para contrarrestar

¹Montgomery, Douglas *et al.* *Probabilidad y Estadística para ingeniería y administración*, p.666.

los efectos de las varianzas distintas, sin embargo, en el fondo el interés continúa apostado en el modelo para la media mientras la varianza se sigue considerando como un "factor de ruido" donde se conjugan todos aquellos factores no controlables y se soslayan posibles efectos sistemáticos.

A través de la revisión de las propuestas de varios autores, en este trabajo de tesis se busca proporcionar -principalmente a los profesionistas interesados en el control de la calidad industrial y, en general a las personas involucradas en el análisis de datos de todas las áreas del conocimiento (Economía, Medicina, Ecología, Educación, Psicología, por mencionar sólo algunas), en las que con frecuencia se encuentran los modelos de regresión con problemas de heteroscedasticidad- un camino diferente, una alternativa en la que la varianza y la media son igualmente importantes, motivo por el cual, ambos merecen un análisis detallado y cuidadoso para determinar los factores de los que reciben una influencia notable y, en su caso, plasmarlos en un modelo no sólo para la media sino también para la varianza.

1.2 Modelación de la variabilidad

La experiencia que se tiene en el estudio y modelado de la media es considerablemente más extensa que la que se puede adjudicar en el mismo sentido a la varianza. Los métodos de estimación de los parámetros de la media y las propiedades de sus estimadores son ampliamente conocidos y, actualmente, la literatura al respecto es por demás vasta. Está claro que no se puede decir lo mismo al referirse a la varianza, razón por la que este trabajo se ocupa de abordar el tema presentando una amplia variedad de alternativas de exploración; las propuestas de modelación que han planteado diferentes autores y algunos métodos de estimación de los parámetros.

Las propiedades de los estimadores han sido punto de discusión entre los estudiosos del tema por lo que, aquí se analizan a través de los resultados de la aplicación de métodos montecarlo.

1.3 Objetivos

A lo largo de este trabajo se pretende:

- Motivar la necesidad de identificar, ante un problema de heteroscedasticidad, los efectos sistemáticos que algunos factores pueden tener sobre la varianza.
- Resaltar la importancia de desarrollar un análisis exploratorio sobre los efectos de dispersión.
- Revisar diferentes métodos para estimar los parámetros de un modelo para la varianza en procesos industriales.
- Presentar los algoritmos de cada uno de los métodos de estimación de parámetros.
- Implementar los algoritmos en datos simulados.
- Estimar los efectos de los factores que tienen influencia sobre la media y hacer lo propio con los que afectan la varianza de un proceso a través de diversos ejemplos documentados en la literatura.
- Comparar, entre sí, los métodos presentados en términos de sesgo y varianza.
- Aplicar los diferentes métodos en el análisis de dos experimentos.

1.4 Contenido de la tesis

Para el logro de los objetivos, este material se ha organizado en siete capítulos de la siguiente manera:

La presente introducción forma el primero de ellos; el capítulo segundo está dedicado a sensibilizar al lector sobre la necesidad de realizar un análisis para la varianza similar al que se lleva a cabo en el caso de la media.

A través de cuatro ejemplos ya documentados (en algunos casos han sido objeto de largas discusiones por parte de varios investigadores) se desarrolla de manera amplia un análisis exploratorio que incluye desde diagramas de dispersión y cajas hasta el uso del coeficiente de correlación de Spearman

como prueba no paramétrica de asociación lineal entre variables. La gráfica de Daniel es una herramienta importante en la exploración de los efectos.

Es en este apartado donde se introducen algunas ideas sobre los modelos para la varianza a través de regresiones aplicadas a algunas transformaciones de los residuos.

En el tercer capítulo se hace una breve reseña histórica sobre el tema y se revisan algunas discusiones sobre las características de los modelos. Aquí se formaliza el problema y se da a conocer la notación empleada en el resto del material.

La descripción de cinco métodos de estimación propuestos por varios autores se encuentra en el capítulo 4. Después de proporcionar detalles de cada uno de los métodos, estos son ilustrados a través de su aplicación a los ejemplos del capítulo 2. Las primeras comparaciones entre las bondades de los diferentes procedimientos se remiten al cuadrado medio del error provocados por el modelo. Los cambios en esta estadística motivan a recalcular el error estándar y el nivel de significancia de los estimadores. Las gráficas de los residuos constituyen un elemento importante en este capítulo.

Con la idea de estudiar el comportamiento de los estimadores obtenidos por los métodos descritos en el capítulo 4, se recurre a la simulación como técnica de muestreo. Dicho recurso se aplica en dos experimentos: uno de ellos representa el caso en que la varianza está en función de las variables regresoras y el otro consiste en un diseño Box-Behnken. El detalle de los experimentos y los algoritmos empleados se encuentran en el capítulo número cinco.

El sexto apartado contiene el análisis de los resultados generados por las simulaciones. El sesgo y la varianza de los estimadores son el centro de atención a lo largo del capítulo.

Finalmente, en el capítulo siete se presentan las conclusiones, las referencias bibliográficas y, como parte de los apéndices, se proporcionan los conjuntos de datos aquí analizados; los programas de computo empleados y las tablas de resultados.

Capítulo 2

Exploración de Efectos de Dispersión

Para proponer un modelo del valor esperado de un conjunto de datos, la primera recomendación importante es hacer una exploración sobre el comportamiento de las observaciones con respecto a las variables regresoras. Una vez ajustado el modelo propuesto, es indispensable hacer una revisión del cumplimiento de los supuestos, misma que se realiza por medio de los residuos. Se tiene claro que la forma de "embudo" (figura 2.4) o de "moño" en una gráfica de dispersión entre residuos y valores ajustados (o variables regresoras) es una muestra clara de la diferencia de varianzas entre las poblaciones. En este apartado se busca resaltar la importancia de realizar también un análisis exploratorio sobre los residuos con la idea de identificar, si es el caso, los patrones de comportamiento sistemático de la dispersión.

2.1 El modelo clásico

A manera de establecer un vínculo entre el material de ésta tesis y el tema de regresión clásica, aquí se hace una breve descripción del modelo y sus supuestos.

Un modelo estadístico lineal es representado de manera común con la expresión

$$Y = \mu_i + \varepsilon, \quad (2.1)$$

donde:

- $\mu_i = \mu + X\beta$ es el *componente sistemático* que está formado por una media general (μ), cuyo valor constante depende de los factores comunes a todas las poblaciones de las que provienen los datos, y por el término $X\beta$ que representa las desviaciones a partir de la media general debidas a las características comunes (x_i) dentro de cada población pero que varían de una población a otra.
- ε es el *componente aleatorio* y representa a todos aquellos factores no controlables que hacen diferentes a los elementos al interior de las poblaciones. Se le conoce como el término del error.

En el diseño de experimentos las hipótesis que se plantean son con respecto a los efectos que los tratamientos tienen *sobre la media*. Para probar tales hipótesis se supone que los errores del modelo son variables aleatorias independientes con distribución normal, con media cero y con varianza constante σ^2 . Es decir, se supone que las observaciones, en cada tratamiento, forman una muestra aleatoria de una población normal y que se tienen tantas poblaciones (con igual varianza) como criterios de clasificación incluya el experimento.

2.2 Motivación

En esta sección se ilustra gráficamente y por medio de datos simulados, el caso homoscedástico y su representación en la tabla de análisis de varianza. De la misma manera, se usa un problema de heteroscedasticidad para motivar en el lector la necesidad de brindar a la varianza la oportunidad de jugar un papel relevante en el modelo.

Homoscedasticidad

La gráfica del efecto que los cinco niveles de un factor X tiene, de manera exclusiva, sobre la media de la variable respuesta de un experimento con cinco repeticiones por punto diseño es la que muestra la figura 2.1.

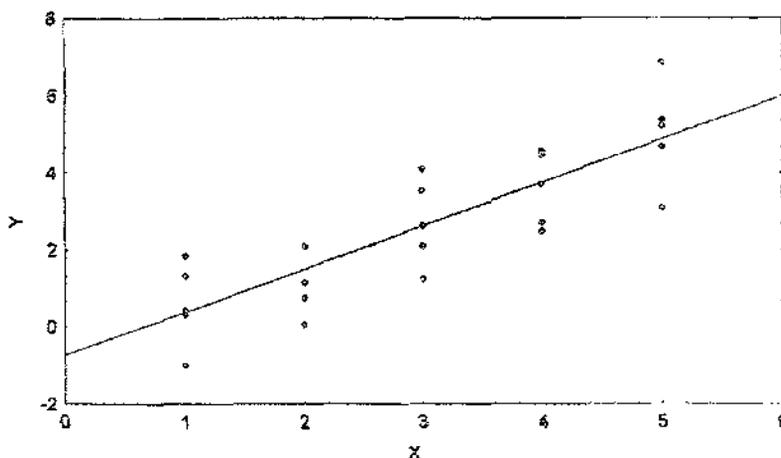


Figura 2.1: Efecto sobre la media

Cada uno de los conjuntos, de cinco repeticiones, forma una muestra aleatoria de una población normal con ciertos parámetros. Es decir, en la gráfica se pueden dibujar cinco distribuciones normales con igual varianza (exactly la misma forma) que se van desplazando suavemente conforme se modifica el valor de X (Figura 2.2).

El análisis de varianza, para este ejemplo, (tabla 2.1) confirma que el efecto del factor X (con cinco categorías) sobre la media es significativamente diferente de cero:

Fuente de variación	g.l.	SC	CM	F	p
X	4	63.67	15.92	13.32	0
Error	20	23.91	1.2		
Total	24	87.57			

En particular, en este caso, las medias de las poblaciones muestran un comportamiento sistemático ascendente que se puede describir a través de una función que depende de los niveles del factor, a saber:

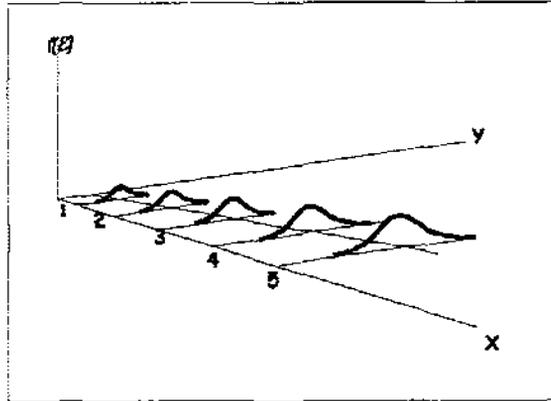


Figura 2.2: Homoscedasticidad

$$\hat{Y} = 2.6188 + \left\{ \begin{array}{l} -2.0318 \text{ si } x = 1 \\ -1.3992 \text{ si } x = 2 \\ 0.0859 \text{ si } x = 3 \\ 0.9494 \text{ si } x = 4 \\ 2.3955 \text{ si } x = 5 \end{array} \right\} \quad (2.2)$$

y los intervalos, al 95%, de confianza para las medias de cada tratamiento son los que ilustra la figura 2.3

Trat	N	Media	D. est	
1	5	0.587	1.098	(---*)
2	5	1.220	0.870	(---*)
3	5	2.705	1.127	(---*)
4	5	3.568	0.956	(---*)
5	5	5.014	1.352	(---*)
d. e.	común	1.093		(---*)

-2 0 2 4 6 8

Figura 2.3: Intervalos de confianza

En este sencillo ejemplo, cabe aclarar que dada la homoscedasticidad, los valores de los errores estándar de las medias son similares entre sí por lo

que no tiene mayor relevancia si se toma el valor promedio y los intervalos de confianza parecen ser todos del mismo tamaño. Lo importante, en estos casos, es observar el comportamiento de las medias.

Heteroscedasticidad

Por otro lado, el ejemplo gráfico que muestra el efecto que los cinco niveles de un factor X tiene, de manera exclusiva, sobre la **varianza** de la variable respuesta de un experimento con cinco repeticiones por punto diseño es el que contiene la figura 2.4.

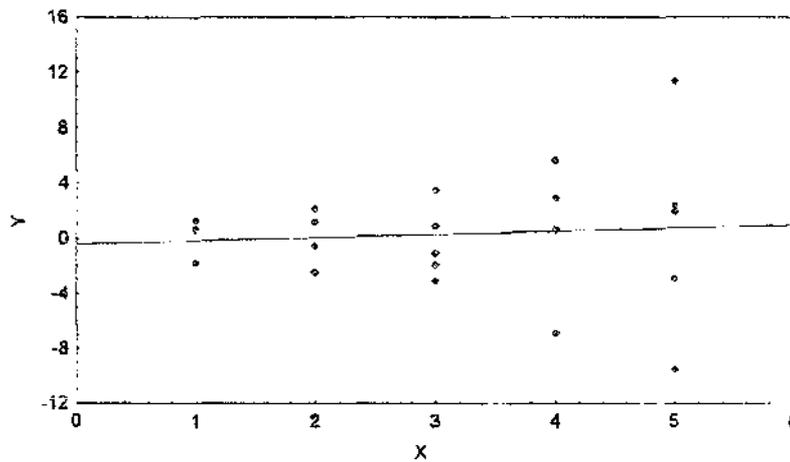


Figura 2.4: La varianza en función de X

El análisis de varianza, en la tabla 2.2, no rechaza la hipótesis de que el efecto del factor **sobre la media** es nulo.

Tabla 2.2. Análisis de varianza

Fuente	g.l	SC	CM	F	P
X	4	6.6	1.6	0.09	0.985
Error	20	376.1	18.8		
Total	24	382.7			

En este sentido, el factor X no debe formar parte de un modelo para la media. ¿Deben entonces ser ignorados sus niveles? ¿Basta con observar que, en este ejemplo, el nivel más bajo es el que genera menor varianza?

Los intervalos al 95% de confianza para la media, basados en la desviación estándar común se ilustran en la figura 2.5

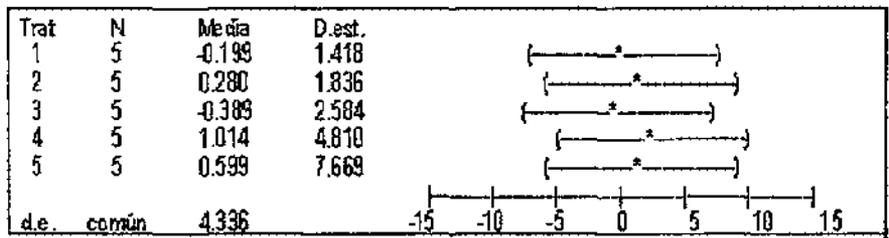


Figura 2.5: Intervalos de confianza

Si se observa la columna que contiene la desviación estándar de cada tratamiento, se puede entender que el error estándar común sea más grande que bajo condiciones de homoscedasticidad, además de esta manera, se está "forzando" a que las poblaciones parezcan igualmente dispersas y se pierde de vista el comportamiento de la varianza. Puede notarse que el intervalo en el nivel 1 es más grande de lo que realmente debería ser mientras sucede lo contrario en el último nivel.

En este sentido parece más adecuado presentar intervalos basados en su propio error estándar (figura 2.6).

En casos como el anterior, lo relevante es observar el comportamiento de la varianzas y no el de las medias. Nótese que la varianza muestra un comportamiento sistemático creciente acorde con los valores de los diferentes niveles del factor mientras las medias de las poblaciones no parecen ser significativamente diferentes.

La figura 2.7 muestra el caso en que el factor X tiene efecto tanto en la media como en la varianza de la variable respuesta.

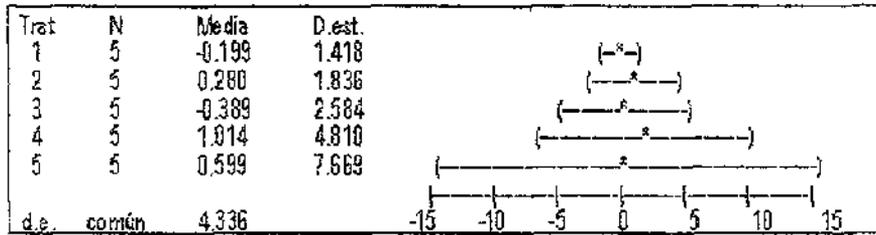


Figura 2.6: Intervalos de confianza

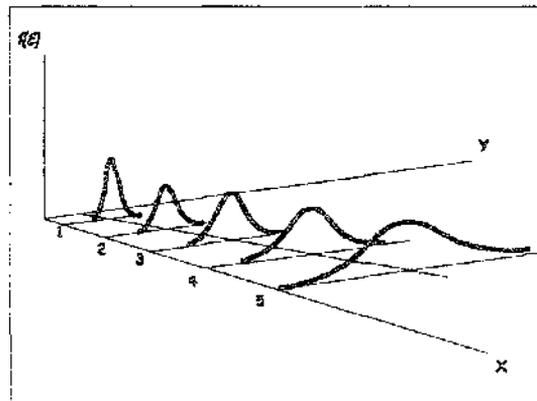


Figura 2.7: Heteroscedasticidad

El análisis de varianza (tabla 2.3) indica un efecto significativo del factor X sobre la media de las observaciones pero no refleja lo correspondiente a la varianza.

Tabla 2.3. Análisis de varianza

Fuente de variación	g.l	SC	CM	F	p
x	4	899.8	225	11.96	0
Error	20	376.1	18.8		
Total	24	1275.9			

Los intervalos de confianza al 95% podrían ser los que se encuentran en la figura 2.8

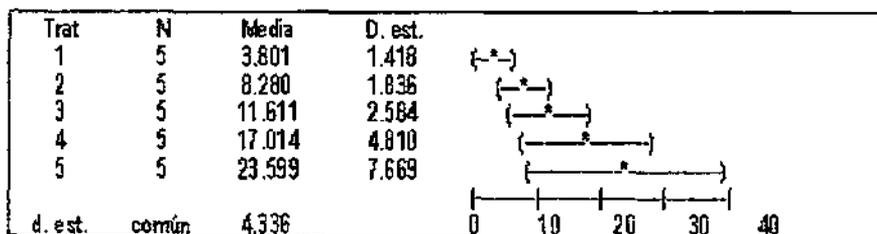


Figura 2.8: Intervalos de confianza

Si la influencia del factor X es sistemática tanto en la media como en la varianza, la pregunta es, ¿por qué debe representarse tal comportamiento en un modelo únicamente para la media?.

En la industria, la variabilidad de los procesos es un aspecto por demás importante y con frecuencia ésta presenta un **comportamiento sistemático** a través de los factores. Tal es la razón por la que surge, de la misma manera que en el modelo de medias, la necesidad de plantear una función de varianza, es decir, un modelo para la variabilidad.

En este sentido, Carroll y Ruppert expresan:

"... *Nosotros vemos la heterogeneidad de la varianza como un problema de regresión, es decir, cambios sistemáticos y suaves de la variabilidad conforme*

las predictoras son perturbadas. Visto de esta manera, hay mucha similitud con el modelado del vector media".¹

En la medida en que se logre identificar a aquellos factores que afectan significativamente el comportamiento de la varianza y la relación que pueda existir entre ellos y las variables regresoras, se tendrá control sobre el proceso y el producto a desarrollar podrá conservar características de calidad con mayor certeza.

2.3 Ejemplos

La falta de homoscedasticidad es frecuente en la aplicación del modelo de regresión a problemas de diversas áreas de estudio. En esta sección se hace referencia a cuatro casos que han sido publicados como ejemplos de tal situación. El primero de ellos pertenece al área de salud; el segundo puede asociarse con el estudio del medio ambiente, y los dos últimos pertenecen a procesos industriales.

Tres de los ejemplos, además de contener la solución propuesta por su respectivo autor, son sometidos a un análisis exploratorio para identificar aquellos factores que tienen efecto significativo sobre la media o la varianza de las observaciones.

2.3.1 Ejemplo: Presión arterial diastólica

Netter y Wasserman (1989) ofrecen un conjunto de datos sobre la presión arterial diastólica de un grupo de 54 mujeres con edades entre 20 y 60 años (ver tabla 1 en el apéndice A).

En el diagrama de dispersión (figura 2.9) se tiene claro que tanto la media como la varianza de las observaciones presentan un comportamiento creciente conforme la edad de las personas avanza.

A continuación se presentan tres soluciones al problema de modelar el conjunto de datos. En primera instancia se obtiene un modelo para la media basado en la estimación por mínimos cuadrados ordinarios; más adelante se

¹Carroll R. y Ruppert D. *Transformation and Weighting in Regression*. pag 10

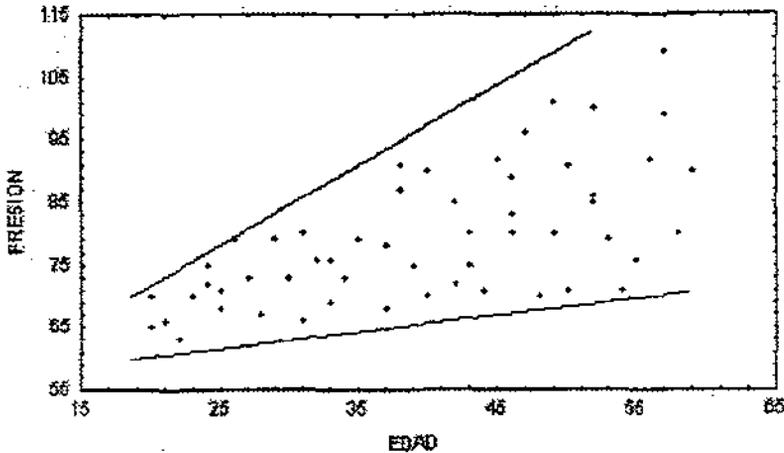


Figura 2.9: Presión contra edad

proporciona la solución que propusieron Netter y Wasserman mediante el uso de los mínimos cuadrados ponderados y finalmente se propone un par de modelos para la dispersión.

Regresión por Mínimos Cuadrados Ordinarios

Ignorando la falta de homogeneidad de varianzas, se aplica el método de mínimos cuadrados

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (2.3)$$

para estimar los parámetros del modelo de regresión lineal ordinario. $Y_{54 \times 1}$ es el vector respuesta (presión arterial de las 54 pacientes) y $X_{54 \times 2}$ es la matriz de covariables, de esta manera se tiene el modelo:

$$\hat{Y}_i = 56.157 + 0.580x_i \quad (2.4)$$

mismo que genera un Cuadrado Medio del Error (*CME*) de 66.35 y los residuos que aparecen en las figuras 2.10 y 2.11 contra los valores ajustados (\hat{Y}) y edad respectivamente.

En ambos casos, la heteroscedasticidad es notoria. En el primero se observa que a \hat{Y} "pequeña" corresponden residuos menos dispersos que a \hat{Y}

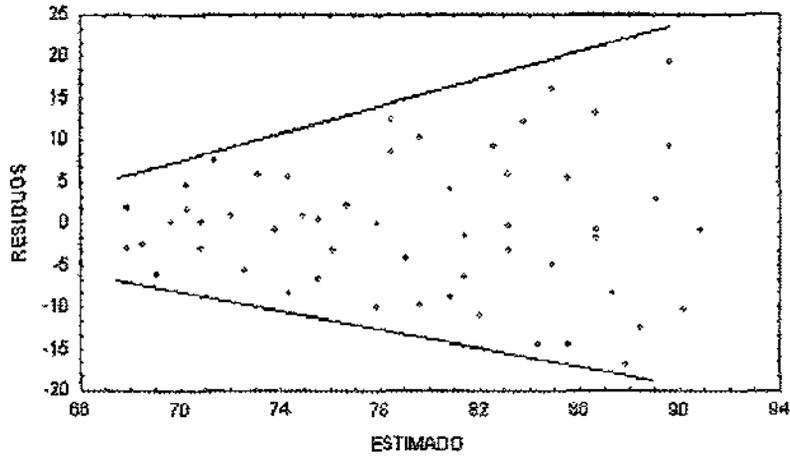


Figura 2.10: Valores ajustados contra residuos

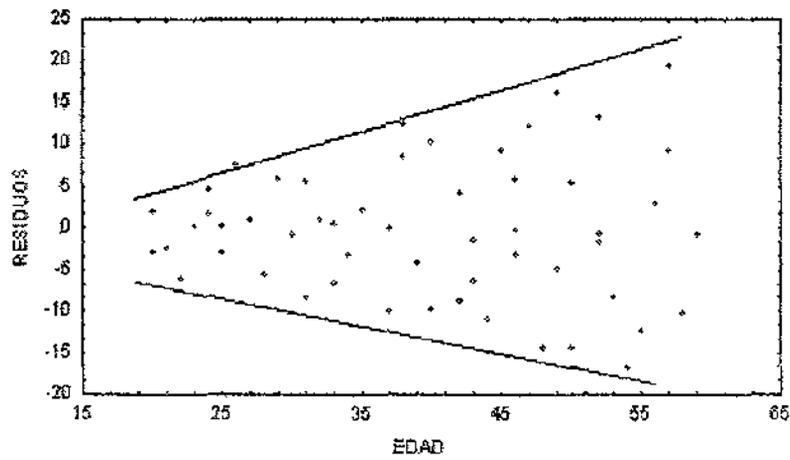


Figura 2.11: Edad contra residuos

”grande”, es decir, conforme la presión arterial diastólica es más alta, el modelo pierde precisión. En el segundo, la relación entre la dispersión de los residuos y la edad se manifiesta.

Regresión por Mínimos Cuadrados Ponderados

Como alternativa, Neter y Wasserman (1989), para estimar los parámetros del modelo, toman el camino de los mínimos cuadrados ponderados, esto es

$$\hat{\beta}_{mcp} = (X'WX)^{-1}X'WY \tag{2.5}$$

donde $W_{54 \times 54}$ es la matriz diagonal con el peso de cada observación.

Ante la falta de repeticiones para la estimación de la varianza, forman cuatro grupos de edad, uno por década, y obtienen la varianza muestral de los residuos de una regresión no ponderada de tal manera que el peso asignado a cada observación es igual al inverso de la varianza muestral del grupo al que ésta pertenece (tabla 2.4).

Tabla 2.4. Medias y varianzas muestrales de los residuos por grupo de edad.

Grupo (j)	Edad	n	media muestral	varianza muestral (vm)	peso = $\frac{1}{vm}$
1	[20-30)	13	0.448481	17.74260	0.0563615
2	[30-40)	13	0.085086	42.13678	0.0237322
3	[40-50)	15	0.201695	87.93657	0.0113718
4	[50-60)	13	-0.766291	124.14565	0.0080551

La representación gráfica de las estadísticas de los residuos, mediante cajas, está en la figura 2.12

Nótese que la media de los residuos, por grupo, es muy cercana al cero pero su dispersión va en aumento con la edad.

Así, el modelo para la media que proponen Neter y Wasserman es:

$$\hat{Y}_i = 56.0896 + 0.5896x_i \tag{2.6}$$

La gráfica de los valores ajustados contra los residuos ponderados ($\sqrt{peso_j} r_{ij}$) ahora muestra un comportamiento más homogéneo en cuanto a dispersión se refiere (figura 2.13) pero el modelo 2.6 no sólo no refleja la relación de

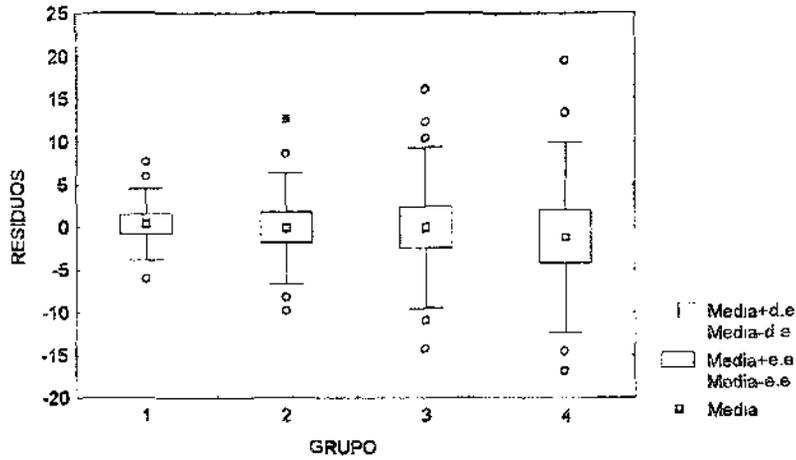


Figura 2.12: Residuos en cuatro grupos.

la varianza con la edad sino, además, las observaciones con mayor varianza tienen menor participación en él. El CME que se obtiene con este modelo

$$CME = \frac{\sum w_i(Y_i - \hat{Y}_i)^2}{n - p} \tag{2.7}$$

es menor que el que genera la regresión no ponderada, a saber

$$CME = 0.9641 \tag{2.8}$$

Para completar la idea de aplicar mínimos cuadrados ponderados a este problema, en seguida se presenta la misma idea de Neter y Wasserman pero ahora formando ocho grupos con los residuos del ajuste no ponderado. Las estadísticas de cada grupo de residuos están en la tabla 2.5 y la figura 2.14.

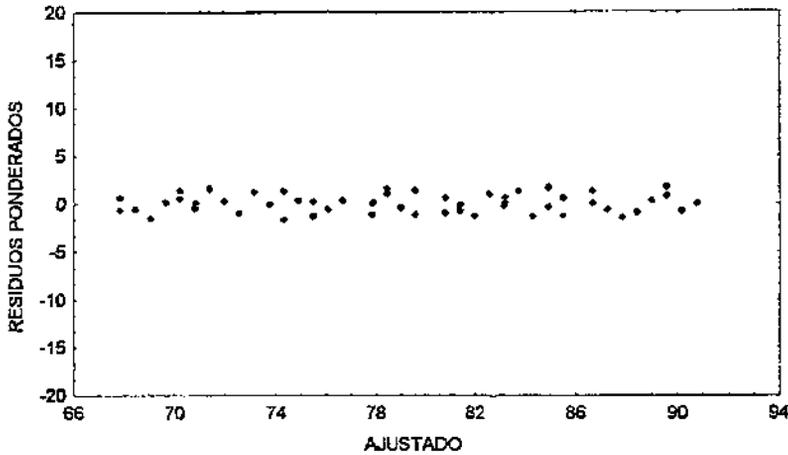


Figura 2.13: Residuos ponderados contra valores ajustados

Tabla 2.5 Medias y varianzas muestrales de los residuos por grupo de edad

Grupo	edad	n	media muestral	varianza muestral(vm)	peso($\frac{1}{vm}$)
1	[20-25)	7	-0.203	13.50	0.074
2	[25-30)	6	1.210	25.09	0.040
3	[30-35)	7	-1.430	22.77	0.044
4	[35-40)	6	1.855	66.82	0.015
5	[40-45)	7	-2.950	64.18	0.016
6	[45-50)	8	2.960	102.27	0.010
7	[50-55)	7	-2.950	116.10	0.009
8	[55-60)	6	1.780	144.16	0.007

El modelo ponderado es

$$\hat{Y}_i = 56.14 + 0.574x_i \quad (2.9)$$

con el que se obtiene un $CME = 0.9526$ y los residuos que se grafican en la figura 2.15.

Modelar la variabilidad

Con el método de estimación ponderada, lo que se busca es "corregir" el problema de heteroscedasticidad y recuperar, en lo posible, los supuestos del modelo clásico para proponer un modelo para la media. Sin embargo, se

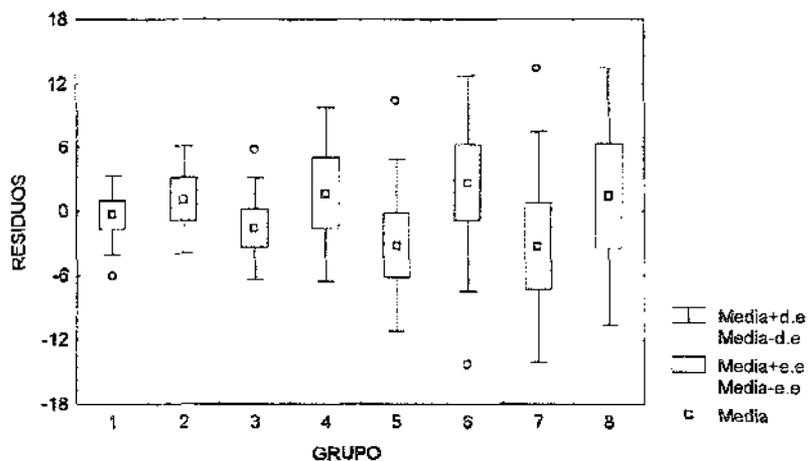


Figura 2.14: Residuos en ocho grupos.

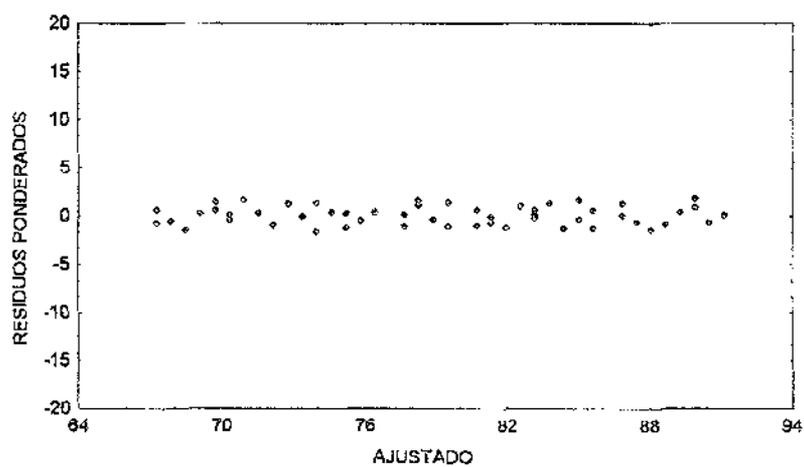


Figura 2.15: Residuos ponderados vs. ajustados (8 gpos)

han trazado caminos nuevos en los que se busca modelar más que corregir la variabilidad.

Con la intención de proponer un modelo para la varianza en términos de la variable regresora, a continuación, y siguiendo la idea de Carroll y Ruppert (1989), se hace una exploración gráfica sobre la relación entre la edad y algunas transformaciones de los residuos que arroja el modelo no ponderado.

Ante la premisa de que el valor absoluto de los residuos proporciona información sobre el comportamiento de la desviación estándar y los residuos al cuadrado hacen lo propio con la varianza, Carroll y Ruppert (1989) recomiendan el uso de la transformación logaritmo natural en ambos casos para obtener las correspondientes aproximaciones.

Transformación I

Es así como al aplicar el logaritmo natural al valor absoluto de los residuos del ajuste no ponderado, (modelo 2.4), se percibe en ellos un comportamiento creciente con respecto a la edad (figura 2.17), de hecho, la prueba no paramétrica basada en el coeficiente de correlación de Spearman rechaza la hipótesis de asociación lineal nula con $\rho = 0.4113$ y una significancia del 0.002. Esto es, se puede decir que la desviación estándar aumenta en función lineal con la edad.

Para identificar el patrón de crecimiento, la idea es armar conjuntos de observaciones y calcular el promedio de los residuos transformados en cada uno de ellos. De esta manera, se busca un modelo para las medias de los residuos en los grupos.

Si en este ejemplo se forman 8 grupos de edad, uno por lustro, y se calcula la media de los residuos transformados, en cada grupo, se tienen las estadísticas que muestra la tabla 2.6.

El comportamiento creciente de la desviación estándar (figura 2.17) es aún más claro y motiva la búsqueda de un modelo de la forma

$$\ln |r| = \alpha_0 + \alpha_1 x \quad (2.10)$$

por lo que la desviación estándar se expresa como

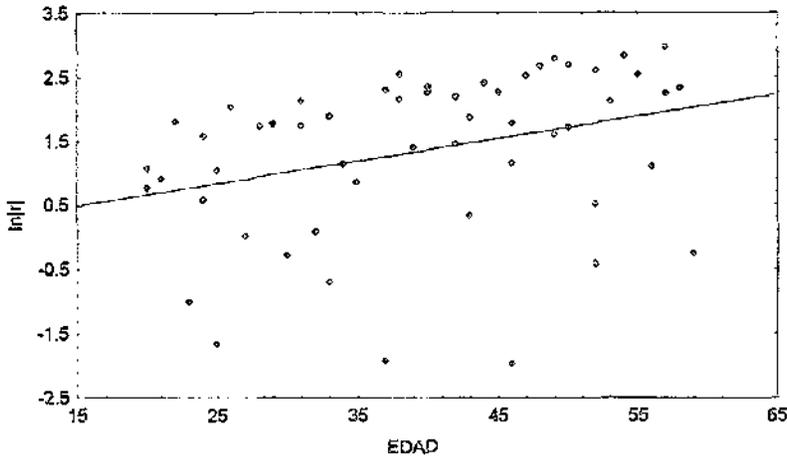


Figura 2.16: Residuos transformados contra edad

$$d.e.(y) = \exp(\alpha_0 + \alpha_1 x) \tag{2.11}$$

Tabla 2.6. Media de residuos transformados por grupo de edad

grupo de edad	n	promedio de residuos transformados
[20 - 25)	7	0.8583
[25 - 30)	6	0.9339
[30 - 35)	7	0.8676
[35 - 40)	6	1.3813
[40 - 45)	7	1.7876
[45 - 50)	8	1.6008
[50 - 55)	7	1.5761
[55 - 60)	6	1.7131

Es así como a través del valor absoluto de los residuos, y después de aplicar mínimos cuadrados ordinarios para estimar los parámetros de la recta que describe el comportamiento de la desviación estándar media, se tiene como modelo tentativo para la desviación estándar

$$d.e.(y) = \exp(0.203 + 0.029 * edad) \tag{2.12}$$

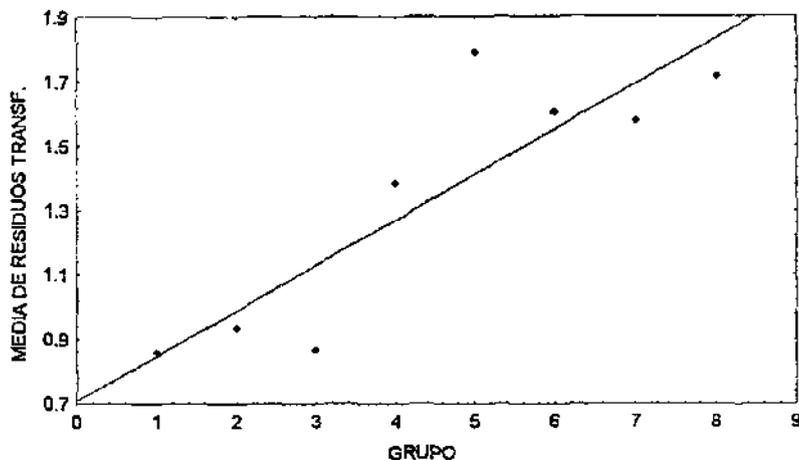


Figura 2.17: Medias de $\ln|r|$ por grupo

cuyos residuos no muestran un comportamiento sistemático con respecto a la variable regresora (figura 2.18).

Si el modelo para la desviación estándar es adecuado, resulta interesante aplicarlo en la estimación ponderada de los parámetros para el modelo de la media. Es decir, a través de la expresión 2.12 se obtiene la estimación del peso asignado por observación esto es,

$$w_i = \exp(0.203 + 0.029 * edad)^{-2} \quad (2.13)$$

De esta manera, al aplicar la expresión 2.5, el modelo ponderado para la media es

$$\hat{Y}_i = 55.9587 + 0.58578x_i \quad (2.14)$$

mismo que genera un $CME = 1.376$

Transformación II

Por otro lado, la relación entre el logaritmo natural del cuadrado de los residuos y la edad se observa en la figura 2.19. La relación creciente que se observó con anterioridad aquí se vuelve a manifestar.

Bajo el supuesto de normalidad en los errores, los residuos al cuadrado seguirían, aproximadamente, una distribución ji-cuadrada con un grado de

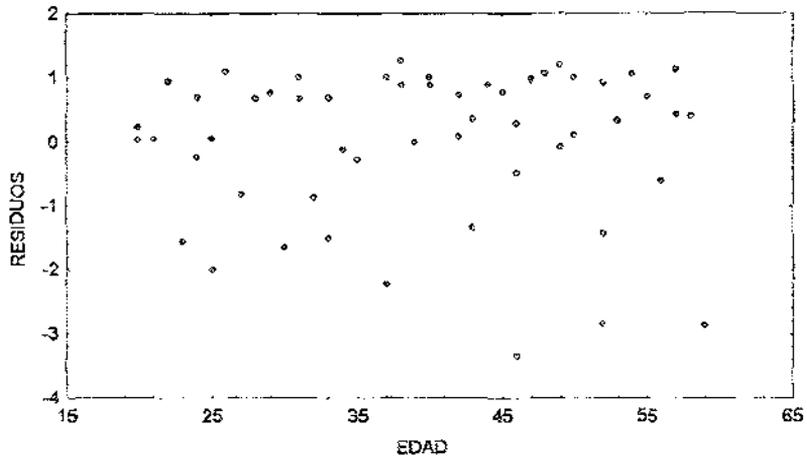


Figura 2.18: Residuos del modelo de desviación estándar

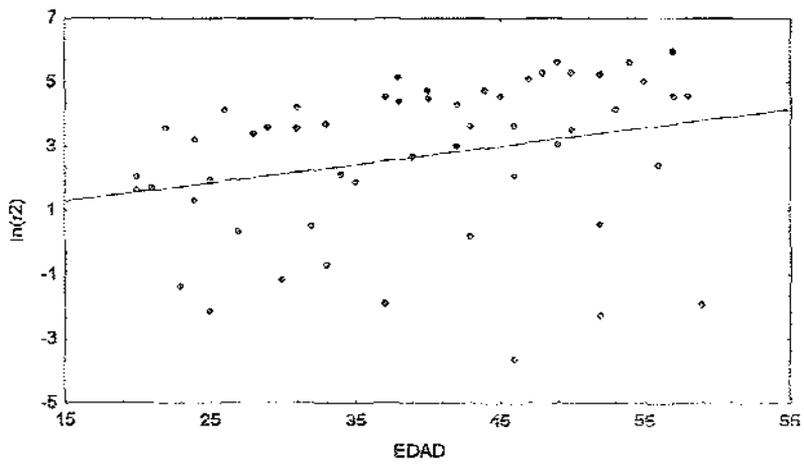


Figura 2.19: Edad vs. residuos al cuadrado

libertad por lo que, Carrrol y Ruppert (1989), proponen otra alternativa que consiste en observar el comportamiento de la raíz cúbica del cuadrado de los residuos no ponderados (transformación para conseguir normalidad) con respecto a la edad, (figura 2.20).y sin formar grupos, buscar una expresión que explique el comportamiento de ambas variables.

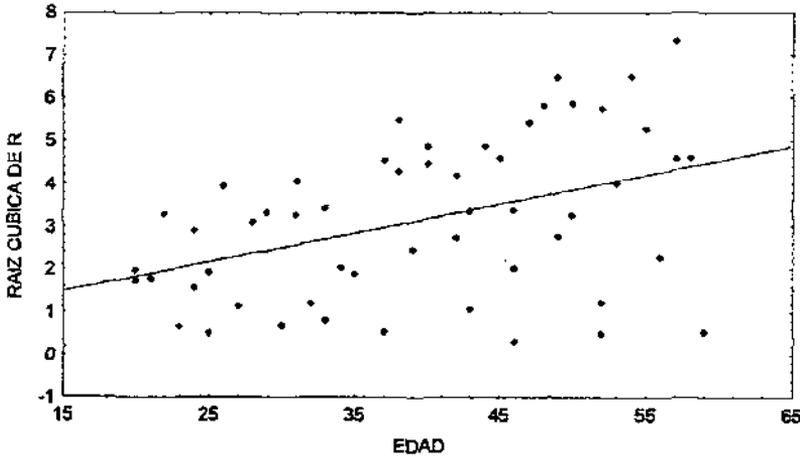


Figura 2.20: Raíz cúbica del cuadro de los residuos

En el ejemplo, un modelo que representa una relación lineal entre la varianza y la edad es

$$Var(y_i) = 0.46393 + 0.06763x_i \tag{2.15}$$

Los residuos del ajuste del modelo para la varianza aún muestran un comportamiento creciente con la edad de las pacientes (figura 2.21).

No obstante, si la expresión 2.15 se usa para la estimación de los pesos de las obsevaciones y se aplica una regresión ponderada para obtener los estimadores del modelo de la media, se tiene

$$\hat{Y}_i = 56.0868 + 0.5818x_i \tag{2.16}$$

con un $CME = 2.8606$

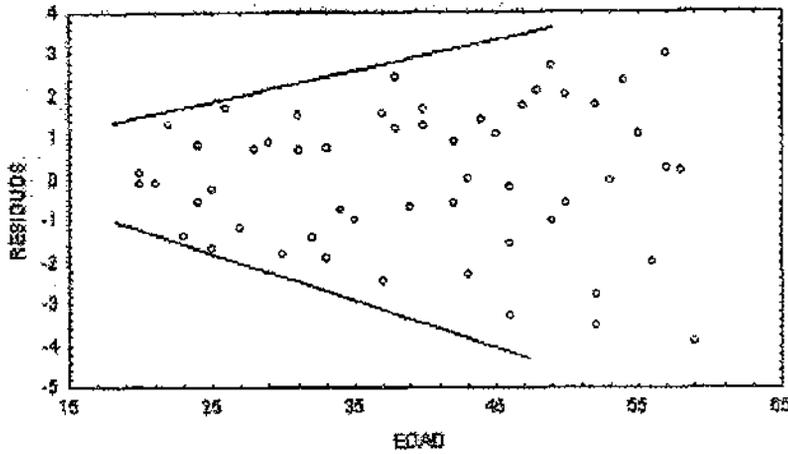


Figura 2.21: Residuos de la función varianza

En resumen, los diferentes métodos de estimación aplicados arrojan los siguientes resultados para los estimadores de los parámetros del modelo de la media (β) y de los residuos transformados (α).

Tabla 2.7 Estimaciones de los parámetros

Estimador	MCO	MCP ₄ ²	MCP ₈ ³	Transf. I	Transf. II
$\widehat{\beta}_0$	56.157	56.0890	56.1410	55.9587	56.0868
$\widehat{\beta}_1$	0.580	0.5896	0.5742	0.5858	0.5818
$\widehat{\alpha}_0$				0.2034	0.4639
$\widehat{\alpha}_1$				0.0288	0.0676
<i>CME</i>	66.350	0.9641	0.9526	1.3760	2.8606

Se observa que el método que minimiza el cuadrado medio del error es el de mínimos cuadrados ponderados a través de 8 grupos, sin embargo, el modelo así obtenido no refleja la relación que existe entre la varianza y la edad. Por medio de las transformaciones se busca explicar dicho comportamiento.

2.3.2 Ejemplo: Árboles de cereza

Este es un ejemplo famoso porque ha sido discutido y analizado por diversos autores como Aitkinson, Cook-Weisberg y Mc Cullaugh-Nelder. El conjunto de observaciones forma parte de las bases de datos que incluye el paquete estadístico MINITAB.

Los primeros en documentar este problema fueron Ryan, Joiner y Ryan (1976). Se tienen 31 observaciones sobre las dimensiones de los árboles de cereza negra. La idea es encontrar un modelo para el volumen de madera útil en los árboles en función de la altura y el diámetro de los mismos.

La figura. 2.22 contiene el diagrama de dispersión de los datos. Para identificar el problema de heteroscedasticidad por este medio, lo ideal sería observar campanas de Gauss bidimensionales con diferente escala de las que, se supone, provienen las observaciones.

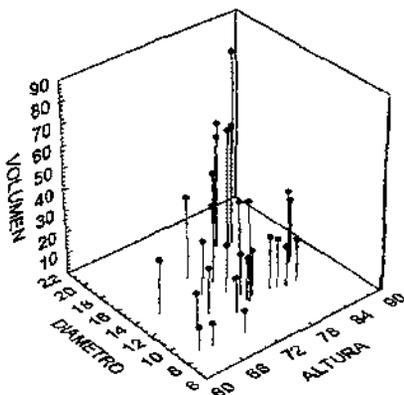


Figura 2.22: Diagrama de dispersión

Un modelo lineal ordinario que relaciona el volumen con la altura y el diámetro bajo los supuestos de homoscedasticidad, independencia y normalidad de los errores es:

$$\hat{Y} = -57.98766 + 4.7082x_1 + 0.3393x_2 \quad (2.17)$$

donde:

$$\begin{aligned} x_1 &= \text{diámetro} & y \\ x_2 &= \text{altura} \end{aligned}$$

Pero, la gráfica de los residuos contra los valores ajustados sugiere la inclusión de un término cuadrático en el modelo (figura 2.23)

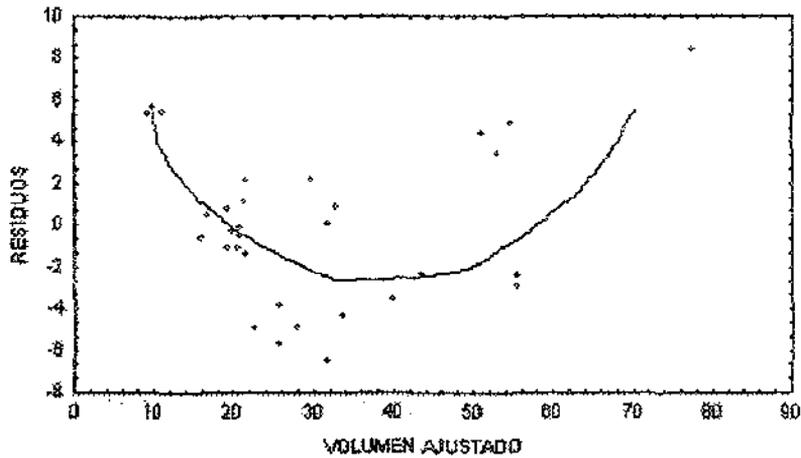


Figura 2.23: Valores ajustados vs. residuos

De hecho, la gráfica de los residuos contra el diámetro (figura 2.24) confirma un patrón cuadrático entre ambas variables.

Al ajustar, por mínimos cuadrados ordinarios, un modelo que incluye el término cuadrático para el diámetro, es decir, un modelo de la forma

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i}^2 + \epsilon_i \quad (2.18)$$

se tiene que los tres coeficientes son significativamente diferentes de cero y toman los siguientes valores para los estimadores de los parámetros:

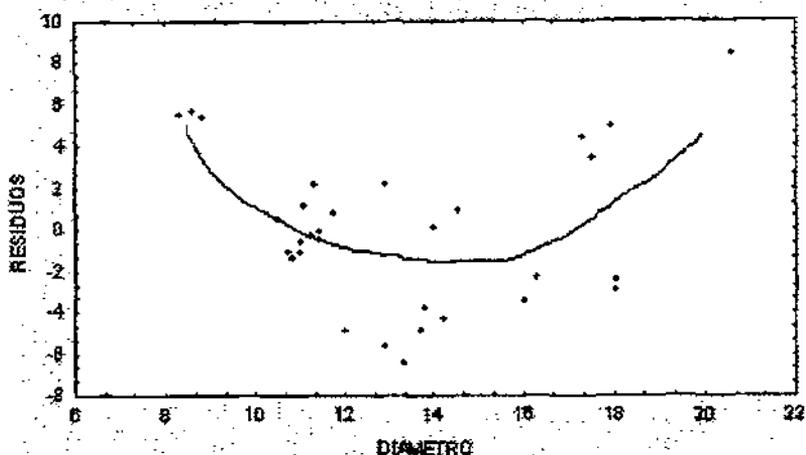


Figura 2.24: Residuos vs. diámetro

Tabla 2.8. Estimadores del modelo

Término	Estimador	Error estándar
Intercepto (β_0)	-9.9204	10.079
diámetro(β_1)	-2.8851	1.309
altura(β_2)	0.3763	0.088
diámetro ² (β_3)	0.2686	0.046

sin embargo, la forma de embudo en la gráfica de los residuos permanece, esto es, el problema de heteroscedasticidad persiste (figura 2.25)

Ante tal situación, Atkin (1987) sugiere la transformación de los datos por medio de la raíz cúbica para ajustar un modelo lineal sin interacciones obteniendo así los siguientes resultados:

Tabla 2.9. Estimadores del modelo para datos transformados

Término	Estimador	Error estándar
Intercepto (β_0)	-0.0853	0.1843
Diámetro (β_1)	0.1515	0.0056
Altura (β_2)	0.0145	0.0028

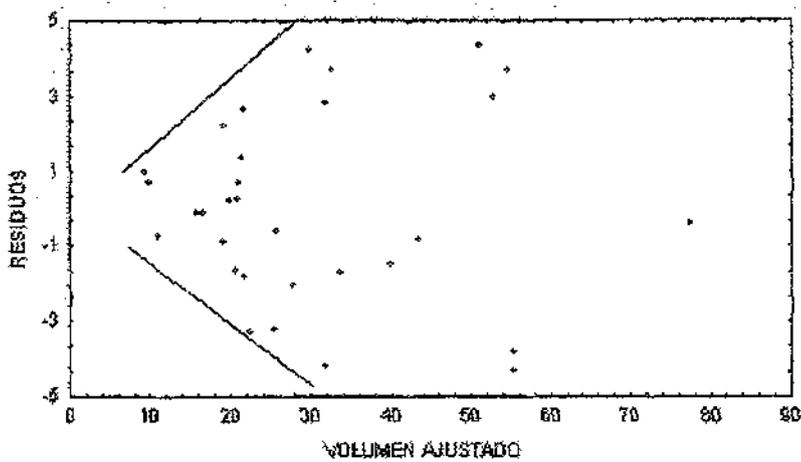


Figura 2.25: Residuos. Modelo con término cuadrático

por lo que el modelo propuesto es

$$\hat{Y}_i^{\frac{1}{3}} = -0.853 + 0.1515x_1 + .0145x_2$$

Aunque el análisis de los residuos (figura 2.26) no resalta algún problema con la dispersión de los datos, Cook y Weisberg (1983) insisten en resaltar la relación entre la varianza del volumen con el diámetro de los árboles y ponen en tela de juicio la homoscedasticidad.

Por su parte, Mc Cullaugh y Nelder (1989) proponen usar un modelo de la forma:

$$Vol = diámetro^2 * altura * c \quad (2.19)$$

con c constante.

Para que el modelo sea aditivo, usan la transformación logaritmo y buscan los estimadores de los coeficientes en la expresión

$$\ln(vol) = \beta_0 + \beta_1 * \ln(diámetro) + \beta_2 * \ln(altura) \quad (2.20)$$

y obtienen:

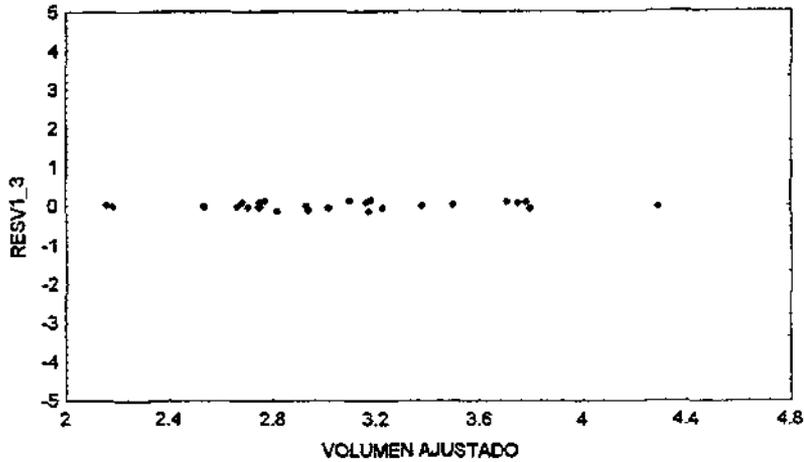


Figura 2.26: Residuos del modelo para los datos transformados

$$\widehat{\ln(Y)} = -6.6316 + 1.9826 * \ln(\text{diámetro}) + 1.1171 * \ln(\text{altura}) \quad (2.21)$$

de tal manera que los residuos no reflejan problemas con la homoscedasticidad (figura 2.27).

Tanto el modelo de Aitkin como el de Mc Cullagh y Nelder corresponden a una transformación de la variable respuesta, por lo que resulta interesante buscar un modelo para la media y la dispersión de las observaciones.

2.3.3 Ejemplo: Resortes automotrices

Este ejemplo se ha discutido en la literatura estadística para modelar efectos de dispersión, y con él se pretende, en este caso, resaltar la importancia de estudiar la varianza en procesos industriales. El problema originalmente fue planteado por Pignatello y Ramberg en 1985 pero Nair y Pregibon en 1988 y, posteriormente Mc Cullagh y Nelder en 1989 hicieron una propuesta de análisis.

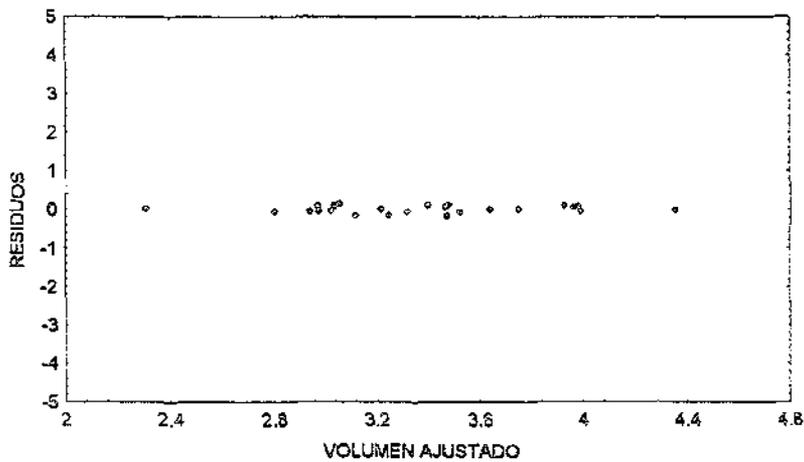


Figura 2.27: Residuos del modelo para $\ln(\text{volumen})$

Aquí se discute la solución que proporcionan Mc Cullagh y Nelder y, adicionalmente se hace un análisis descriptivo para evaluar el efecto de dispersión.

El problema consiste en establecer un sistema de calentamiento adecuado para la producción de muelles automotrices (resortes de hoja). Se busca que la longitud promedio de las muelles sin carga sea de 8 pulgadas con la menor varianza posible. Para tal efecto se diseñó un experimento 2^{5-1} con tres repeticiones y los siguientes factores:

- B: Temperatura del horno
- C: Tiempo de calentamiento
- D: Tiempo de transferencia
- E: Tiempo de inmersión
- O: Temperatura del aceite de templar.

La variable respuesta es la longitud del resorte sin carga.

Cabe aclarar que los diseños factoriales fraccionarios representan una alternativa cuando el número de ensayos necesarios para obtener una repetición completa es tal que el experimento resulta incosteable. En este sentido, diseñar un experimento factorial fraccionario con repeticiones parece contradictorio cuando se busca un modelo para la media sin embargo, en este

caso se trata de estimar varianzas por lo que se hace necesario contar con repeticiones.

Análisis inicial.

El análisis que, en su momento, desarrollaron Pignatiello y Ramberg (1985) sobre la dispersión dentro de las repeticiones, está basado en un modelo para los logaritmos de las varianzas muestrales de cada uno de los 16 tratamientos, sin embargo, Mc.Cullagh y Nelder (1989) consideran incorrecto este camino argumentando que a valores "pequeños" corresponden logaritmos "grandes" y pueden ser negativos.

Por su parte, éstos últimos desarrollaron un análisis que consiste en tomar las varianzas muestrales como la respuesta en un modelo lineal generalizado bajo el supuesto de que los errores son variables aleatorias con distribución gama, la función liga es el logaritmo y el factor de escala es igual a 1.

El resultado que obtuvieron fue que tanto la temperatura del horno (B) como el tiempo de calentamiento (C) tienen efecto significativo sobre la dispersión y los estimadores del modelo log-lineal correspondiente (B+C) son:

$$\hat{b} = 1.369 \pm 0.514 \quad (2.22)$$

$$\hat{c} = -1.092 \pm 0.514 \quad (2.23)$$

Con la idea de incluir en el análisis la información relativa a la media, los autores sustituyeron todas las observaciones por el valor promedio estimado correspondiente y repitieron el ejercicio anterior con lo que obtuvieron:

$$\hat{b} = -1.820 \pm 0.943 \quad (2.24)$$

$$\hat{c} = 4.785 \pm 0.943 \quad (2.25)$$

Mc.Cullagh y Nelder atribuyen la diferencia entre las estimaciones a que el efecto de ambos factores no es tan relevante.

Análisis alternativo

De acuerdo con Mc Cullagh y Nelder, durante el proceso productivo es difícil controlar la temperatura del aceite para templar por lo que se puede considerar un "factor de ruido". con tal antecedente, a continuación se realiza el análisis de los datos bajo dos esquemas: σ como factor de ruido y σ como factor de control.

σ como factor de ruido. Se dice que un factor es de "ruido" cuando no se puede controlar por completo en el ambiente del experimento. Es responsable del error experimental.

Si se considera σ como factor de ruido, se tienen 6 repeticiones por cada uno de los ocho tratamientos. En la figura 2.28 se observan los diagramas de caja correspondientes. En ellas se distingue la diferencia de varianzas, de hecho, la hipótesis de homoscedasticidad es rechazada, es decir, la muestra no proporciona evidencias suficientes para pensar que los datos provienen de distribuciones con la misma dispersión.

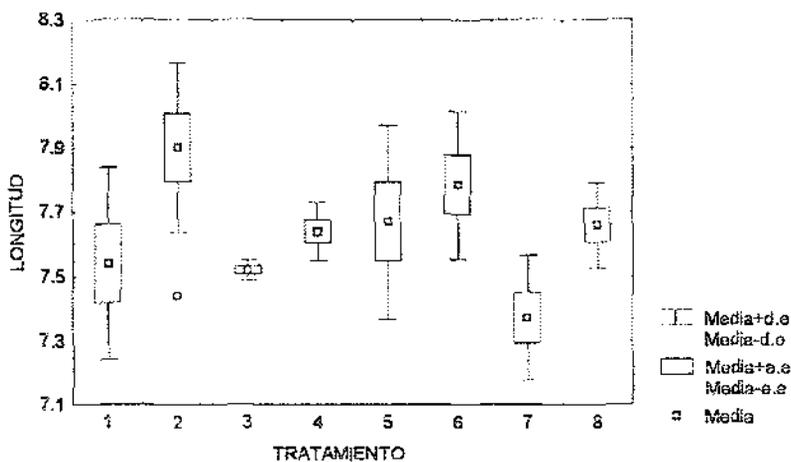


Figura 2.28: Cajas de 8 tratamientos

En este caso, el análisis de varianza y la gráfica de Daniel (figura 2.29) destacan el efecto que B y C tienen sobre la respuesta media sin que se presenten interacciones

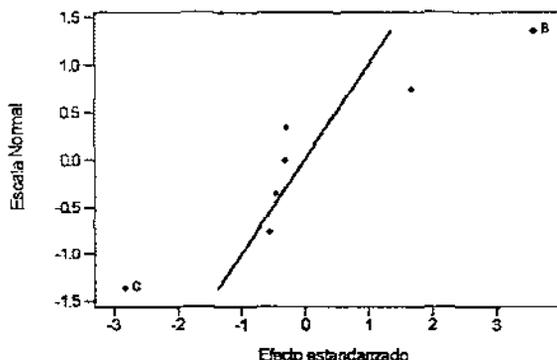


Figura 2.29: Efectos significativos para la media

La estimación, por mínimos cuadrados ordinarios, de los parámetros de un modelo para la media (tabla 2.10) sugiere, para mantener una longitud promedio cercana a las 8 pulgadas, mantener el tiempo de calentamiento (C) en su nivel bajo y la temperatura del horno (B) en el alto.

Tabla 2.10. Estimadores para el modelo de la media

Término	estimador	error estándar	valor p
Intercepto	7.63604	0.03047	<.0001
C[-1 - 1]	0.088125	0.03047	0.0059
B[-1 - 1]	-0.1106	0.03047	0.0007

El modelo es, entonces

$$\hat{Y} = 7.63604 + \left\{ \begin{array}{l} 0.088125 \text{ cuando } C^- \\ -0.088125 \text{ cuando } C^+ \end{array} \right\} + \left\{ \begin{array}{l} -0.1106 \text{ cuando } B^- \\ 0.1106 \text{ cuando } B^+ \end{array} \right\} \tag{2.26}$$

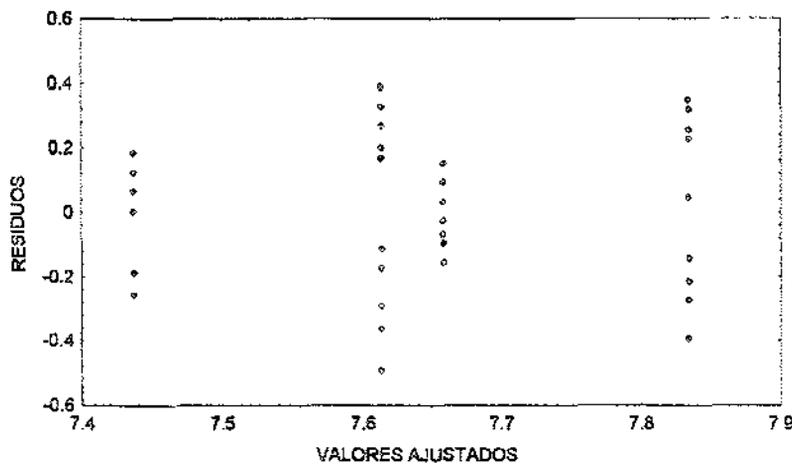


Figura 2.30: Valores ajustados vs. residuos

y los residuos, (figura 2.30) hacen evidente las diferentes dispersiones.

Dado que se tienen 6 repeticiones por punto diseño, cabe la oportunidad de calcular las varianzas muestrales para tomar sus logaritmos como la respuesta en un modelo para la función varianza

$$\ln(s_i^2) = X\alpha \quad (2.27)$$

la estimación por mínimos cuadrados ordinarios resalta al factor C como el único que tiene un efecto significativo en la dispersión (tabla 2.11).

Tabla 2.11. Estimadores para la función varianza

Término	estimador	error estándar	valor p
Intercepto	-3.689	0.4	<0.0001
$C[-1 - 1]$	-1.0901	0.4	0.0035

La forma del modelo es

$$\ln(s_i^2) = -3.689 - 1.0901 * c \quad (2.28)$$

esto es, para minimizar la varianza se debe tener C en su nivel alto, pero de la tabla 2.9 se sabe que C debe estar en bajo para tener una longitud promedio de 8 pulgadas.

Después de separar aquellos puntos donde B está en un nivel alto, independientemente del valor de C, se tienen 24 datos: 12 para C⁺ y 12 para C⁻, las estadísticas por grupo están en la tabla 2.12 y se ilustran con diagramas de caja en la figura 2.31

Tabla 2.12. Estadísticas por grupo

C	n	media	desv. est.
-1	12	7.84	0.245
1	12	7.65	0.1075

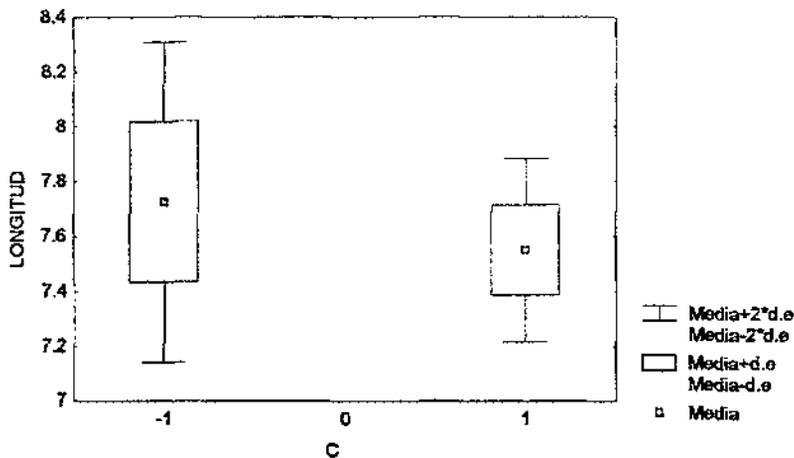


Figura 2.31: Cajas para los dos niveles del factor C

En este caso, las pruebas de Levene y Bartlett descartan la igualdad de varianzas y la prueba de Welch rechaza la hipótesis de igualdad de medias, esto es, bajo el supuesto de normalidad, la decisión a tomar está entre:

- mantener el tiempo de calentamiento (C) en su nivel bajo y ofrecer muelles con una longitud promedio entre 7.35 y 8.33 con el 95% de confianza.
- mantener el tiempo de calentamiento en su nivel alto y ofrecer muelles con una longitud promedio entre 7.435 y 7.865 con un 95% de confianza.

O como factor de control Por otro lado, si O es considerado un factor se tienen 16 tratamientos con tres repeticiones cada uno.

La figura 2.32 contiene los diagramas de caja para las observaciones de cada tratamiento. En ella se sugiere la existencia de un problema de heteroscedasticidad aunque es importante tomar en cuenta que con tres observaciones, la varianza puede ser poco estable.

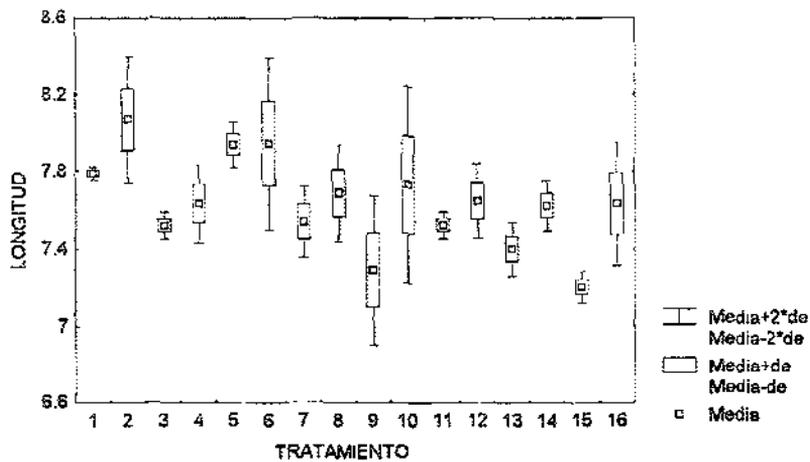


Figura 2.32: Diagramas de caja. Longitud de los resortes

MODELO I

Bajo el supuesto de homoscedasticidad y normalidad, un modelo inicial para la longitud media de los resortes debe incluir, tal como lo indica la gráfica de Daniel, figura 2.33, todos los efectos principales, excepto D, y las interacciones B*O y C*O.

Los estimadores para la longitud media, obtenidos por mínimos cuadrados ordinarios, están en la tabla 2.13

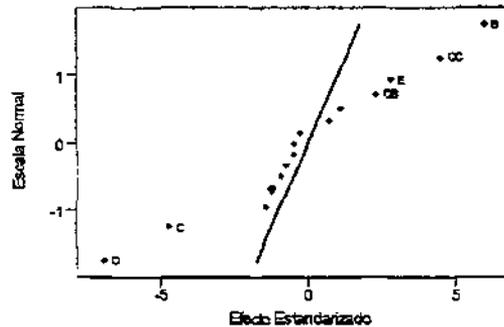


Figura 2.33: Efectos significativos para la media

Tabla 2.13. Estimadores para el modelo para la longitud media

Término	estimador	error estándar	valor p
Intercepto	7.6364	0.018227	<.0001
O[-1 - 1]	0.1298	0.018227	<.0001
C[-1 - 1]	0.088	0.018227	<.0001
O[-1 - 1]*C[-1 - 1]	0.0827	0.018227	<.0001
B[-1 - 1]	-0.1106	0.018227	<.0001
O[-1 - 1]*B[-1 - 1]	0.04229	0.018227	0.0254
E[-1 - 1]	-0.05188	0.018227	0.0069

Esto es, para conseguir muelles con longitud promedio de 8 pulgadas se recomienda mantener los factores O y C en los niveles bajos pero B y E en el nivel alto.

El análisis de los residuos del modelo no muestra ningún problema. En la figura 2.34 se puede ver que no se distingue un patrón entre los valores ajustados y los residuos. Además, en la gráfica de los residuos contra cada uno de los factores, tampoco hay evidencia de una diferencia notable en la dispersión de los datos, de hecho, la prueba de hipótesis basada en la estadística F^{*4} no rechaza la igualdad de varianzas en ningún caso. La tabla

⁴Bajo el supuesto de igualdad de varianzas entre los dos niveles del factor la estadística sigue una distribución normal estándar

2.14 muestra el valor de la estadística para los efectos principales

$$F^* = \ln \frac{s^2(i^+)}{s^2(i^-)} \quad (2.29)$$

donde: $s^2(i^+)$ es la varianza muestral de los residuos de un ajuste no ponderado en el nivel alto del factor i ; $s^2(i^-)$ es el valor correspondiente en el nivel bajo.

El cuadrado medio del error que se obtiene a través de este modelo es $CME = 0.015946$

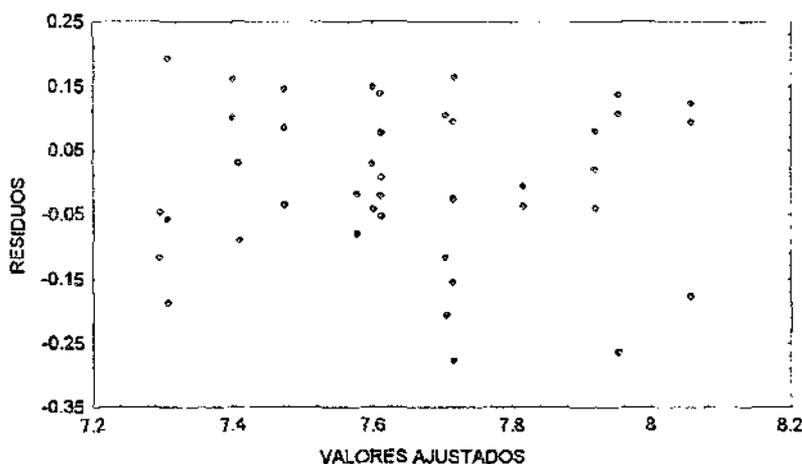


Figura 2.34: Valores ajustados vs. residuos

Tabla 2.14 Estadística de prueba para la igualdad de varianzas

	O	D	C	B	F
$s^2(i^+)$	0.1302	0.1104	0.1072	0.1383	0.1255
$s^2(i^-)$	0.1071	0.1257	0.1302	0.0963	0.1126
F^*	0.3906	-0.2595	-0.3887	0.7239	0.2169

No obstante, si se toma el logaritmo del valor absoluto de los residuos como variable respuesta en un modelo para la varianza, se observa que la

interacción entre la temperatura del aceite de templar y el tiempo de transferencia ($O * D$) tienen un comportamiento significativo en la dispersión de las observaciones; pero no así los efectos principales.

De cualquier manera, para proponer un modelo jerárquico, se toman en cuenta los resultados de la estimación de los parámetros de una función varianza después de aplicar mínimos cuadrados ordinarios

Tabla 2.15. Estimadores para la función varianza

Término	estimador	error estándar	valor p
Intercepto	-2.626	0.1221	<.0001
$O[-1 - 1]$	-0.1491	0.1221	0.2286
$D[-1 - 1]$	0.4647	0.1221	0.7053
$O[-1 - 1]*D[-1 - 1]$	-0.2487	0.1221	0.0477

La interacción ($O*D$) sugiere que la varianza mínima para la respuesta media se obtiene cuando ambos factores se encuentran en el nivel bajo. De esta manera, si se distinguen los 3 puntos que cumplen las condiciones para lograr una longitud media de 8 pulgadas (O^- , C^- , B^+ y E^+) con la menor varianza (O^- , D^-) y se separan del resto de las observaciones, se tienen las siguientes estadísticas

Tabla 2.16. estadísticas por grupo

n	media	desviación est.
3	8.07	0.16523
45	7.607	0.22907

es importante aclarar que, en este caso, la prueba de Levene no rechaza la igualdad de varianzas, por lo que se confirma la homoscedasticidad.

MODELO II

Una propuesta alternativa para estudiar el problema de los resortes automotrices consiste en tomar como variable respuesta, para el modelo de la

media, el promedio de las repeticiones por punto diseño y hacer lo propio con la varianza, esto es

$$\overline{E(Y)} = X\beta$$

y

$$\ln(s^2) = X\alpha$$

Tabla 2.17. Medias y varianzas muestrales

CORRIDA	B	C	D	E	O	\bar{Y}	s^2
1	-	-	-	-	-	7.79	0.0003
2	+	-	-	+	-	8.07	0.2730
3	-	+	-	+	-	7.52	0.0012
4	+	+	-	-	-	7.63	0.0104
5	-	-	+	+	-	7.94	0.0036
6	+	-	+	-	-	7.95	0.0496
7	-	+	+	-	-	7.54	0.0084
8	+	+	+	+	-	7.69	0.0156
9	-	-	-	-	+	7.29	0.0373
10	+	-	-	+	+	7.73	0.0645
11	-	+	-	+	+	7.52	0.0012
12	+	+	-	-	+	7.65	0.0092
13	-	-	+	+	+	7.40	0.0048
14	+	-	+	-	+	7.62	0.0042
15	-	+	+	-	+	7.20	0.0016
16	+	+	+	+	+	7.63	0.0254

La gráfica de Daniel (figura 2.35) resalta que los efectos principales significativos en un modelo para la media son O, B y C, además de la interacción O*C. En el caso de la desviación estándar, el efecto importante corresponde al factor B (figura 2.36).

En este punto resulta interesante observar el diagrama de dispersión de los efectos que los factores tienen sobre la media y la varianza. En la figura 2.37 se destaca lo importante que resultan B y C (en sentidos opuestos)

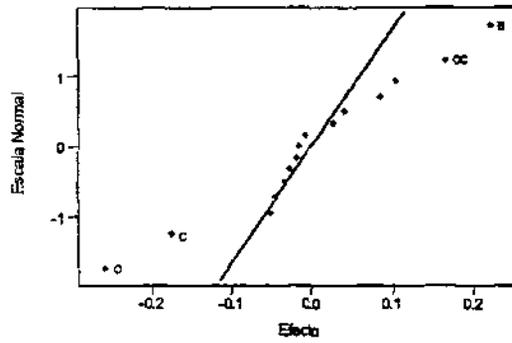


Figura 2.35: Efectos en la media

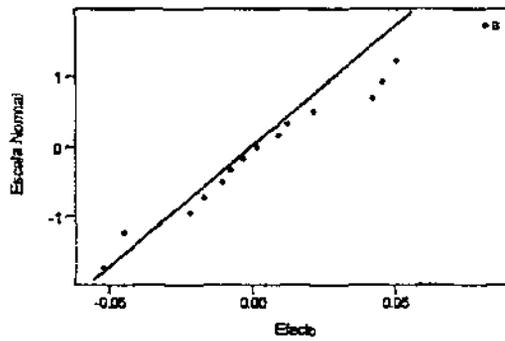


Figura 2.36: Efectos en la desviación estándar

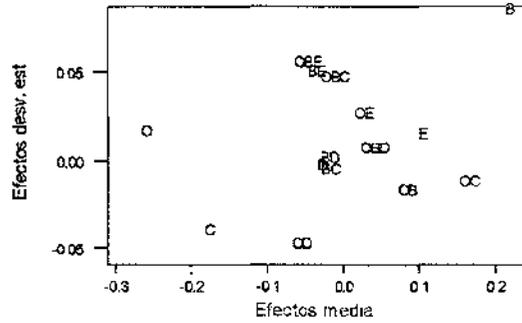


Figura 2.37: Efectos media vs. efectos desv. est.

sobre ambos parámetros. Como *O* permanece "cerca" del origen vertical pero "lejos" del horizontal, se interpreta que sólo tiene efecto significativo en la media. El caso de la interacción *OC* es el mismo.

Los estimadores, para el modelo de medias, obtenidos por mínimos cuadrados y a través de esta estrategia son los siguientes

Tabla 2.18 Estimadores para el modelo de la media

Término	estimador	error estándar	valor p
intercepto	7.636	0.0253	<.0001
<i>O</i> [-1-1]	0.130	0.0253	0.0003
<i>B</i> [-1-1]	-0.111	0.0253	0.0011
<i>C</i> [-1-1]	0.088	0.0253	0.0052
<i>O</i> [-1-1]* <i>C</i> [-1-1]	0.083	0.0253	0.0075

Es decir, para obtener hojas de longitud promedio de 8 pulgadas, se requiere que tanto la temperatura del aceite para templar como el tiempo de calentamiento estén en un nivel alto y la temperatura del horno en el nivel inferior. La gráfica de los residuos contra las medias ajustadas se encuentra en la figura 2.38. puede observarse que no se detecta ningún patrón de comportamiento. El $CME = 0.0103$.

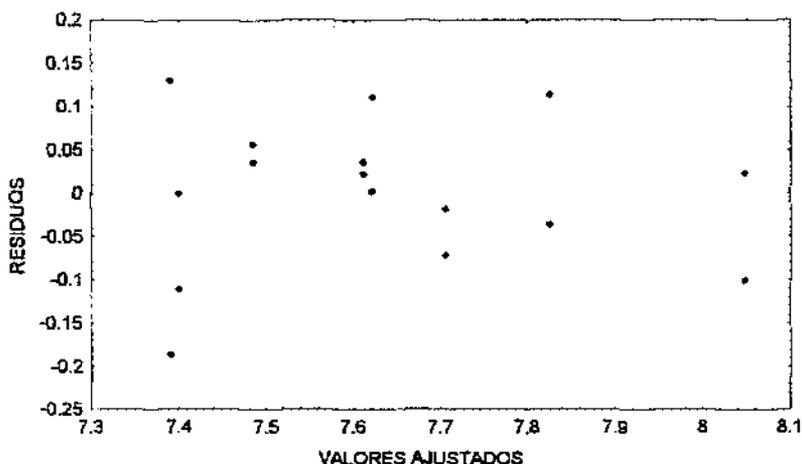


Figura 2.38: Residuos media vs. ajustados media

Por otro lado, los estimadores de los parámetros de un modelo para el logaritmo de las varianzas muestrales son los que se muestran en la tabla 2.19 mismos que resaltan que la temperatura del horno debe estar en su nivel bajo para minimizar la varianza. (El factor C no es significativo para este modelo). De hecho, si se observa la figura 2.39, se tiene que las cajas más cortas (representan datos menos dispersos) corresponden a los tratamientos 1,3,5,11 y 15. Todos ellos tiene en común el nivel bajo de B.

Tabla 2.19 Estimadores para el modelo de la media

Término	estimador	error estándar	valor p
intercepto	-4.931	0.3046	<.0001
B[-1-1]	-0.945	0.3046	0.0078

Tal condición implicaría tomar la decisión sobre si se debe dejar B en su nivel alto para acercarse a la longitud promedio deseada aunque el resultado sea más disperso o buscar mayor precisión con una longitud menor a las 8 pulgadas. Sin embargo, después de seleccionar los puntos en los que se cumple que O y C están en su nivel bajo (requerimiento para una longitud promedio de 8 pulgadas), se tienen las siguientes estadísticas:

Tabla 2.20. Estadísticas de los puntos de interés.

nivel	n	media	desv. est.
B^-	6	7.865	0.0912
B^+	6	8.008	0.1880

La prueba de homogeneidad de varianzas de Bartlett no rechaza tal hipótesis con un valor p de 0.1382 y, Levene tampoco la rechaza con una significancia del 10.48%. Además, la diferencia entre las medias no resulta significativa. Ante tal situación se propone seguir las recomendaciones del modelo de la media.

Resumen

Las diferentes propuestas de análisis han proporcionado resultados distintos. Para el caso de la media, los factores B y C tienen presencia de manera consistente en todos los modelos; aunque queda en tela de juicio la influencia de D y de algunas interacciones. La situación de los parámetros de la función varianza es diferente pues cada propuesta identifica diferentes protagonistas para el modelo.

2.3.4 Ejemplo: Circuitos integrados

Otro autor que se ha preocupado por el estudio de los efectos de dispersión en los procesos industriales es John Grego. En 1993 publicó un artículo en el que manifestó su interés por aplicar los modelos lineales generalizados en la optimización simultánea de la media y la varianza. En tal documento planteó un experimento para la producción de circuitos integrados. Con un diseño 2_{iv}^{4-1} buscó determinar el efecto de cuatro factores sobre la fuerza adherente de los circuitos montados sobre una base de metal.

Los factores son

Tabla 2.21. Factores en el experimento con circuitos

Factor	Descripción	-	+
A	Tipo de adhesivo	D2A	H-1-E
B	Material conductor	cobre	níquel
C	Tiempo	90min	120 min
D	Material del depósito	estaño	plata

Análisis inicial

En su documento, John M. Grego (1993) propuso una pauta para el análisis de estos datos a través de dos métodos:

- Una regresión "stepwise" ponderada para la respuesta media de las observaciones y, por separado, aplicó la misma técnica tomando la varianza muestral como variable respuesta.
- Usó los 40 datos y un modelo de la forma:

$$Y = X\beta + e \quad (2.30)$$

$$Var(e_i) = \exp(z_i\alpha) \quad (2.31)$$

estimó de manera simultánea, por máxima verosimilitud aplicada a un modelo lineal generalizado con errores que se distribuyen gamma, usando el logaritmo como función liga y un factor de escala igual a 0.5.

En ambos casos, los factores A, C y D resultaron significativos para la respuesta media y, A y D para la varianza. Sin embargo, el investigador pone en duda la normalidad asintótica de los estimadores que se obtuvieron por máxima verosimilitud simultánea por lo que no confía en sus errores estándar.

Análisis alternativo

Para elaborar una propuesta sobre la forma de los modelos de la media y la varianza, se empieza por explorar el comportamiento de las observaciones:

La media y varianza muestrales para cada uno de los ocho tratamientos son:

Tabla 2.22. Media y varianza por tratamiento.

Trat	A	B	C	D	\bar{Y}	s^2
1	-	-	-	-	73.48	2.452
2	-	-	+	+	87.06	0.5030
3	-	+	-	+	81.58	0.6470
4	-	+	+	-	79.38	1.9820
5	+	-	-	+	83.88	4.2325
6	+	-	+	-	79.54	8.5625
7	+	+	-	-	75.60	26.705
8	-	+	+	+	90.32	3.9774

Los diagramas de caja por tratamiento, figura ?? muestran las diferentes medias y varianzas muestrales. Mientras la media más alta se logra con el tratamiento 8, la menor varianza corresponde a los tratamientos 2 y 3. Cabe notar que la segunda mitad de la gráfica contiene las cajas más largas, y corresponden a los tratamientos que incluyen al factor A^+ , por lo que se puede pensar que el adhesivo tipo H-1-E provoca datos más dispersos que el D2A, es decir, este factor tiene cierta influencia sobre la varianza de los datos.

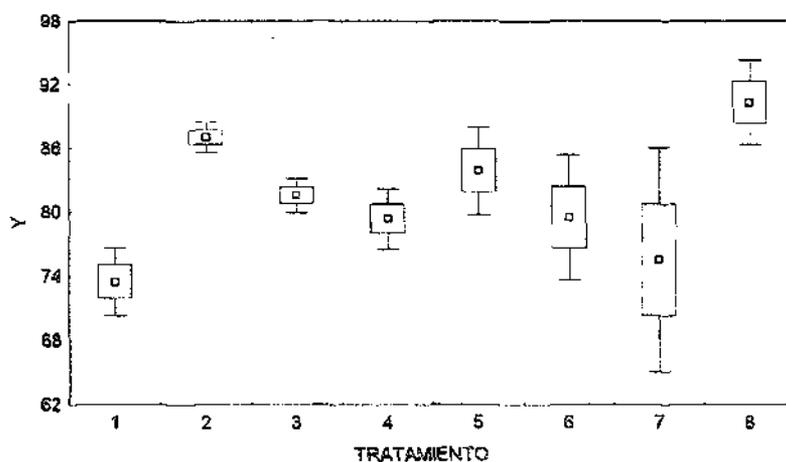


Figura 2.39: Diagramas de caja

Por otro lado, los diagramas de caja por factor, figura 2.40 no reflejan, de manera contundente, la relación del tipo de adhesivo con la varianza de las

observaciones de hecho, aparentemente la influencia del material del depósito (D) puede ser más importante en este sentido y en el comportamiento de la media.

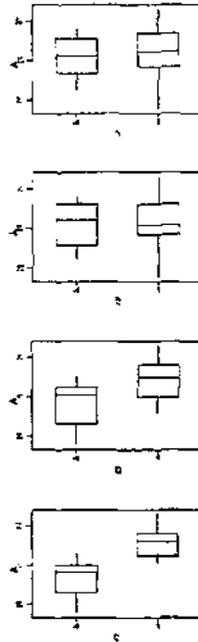


Figura 2.40: Diagrama de cajas por factor

Aplicando las pruebas de Levene, Bartlett y Hartley, se rechaza la hipótesis de igualdad de varianzas.

Si se busca un modelo lineal ordinario para la media, la gráfica de Daniel propone únicamente los efectos principales A, C y D como significativos (figura 2.41).

La estimación por mínimos cuadrados ordinarios arroja los resultados de la tabla 2.23

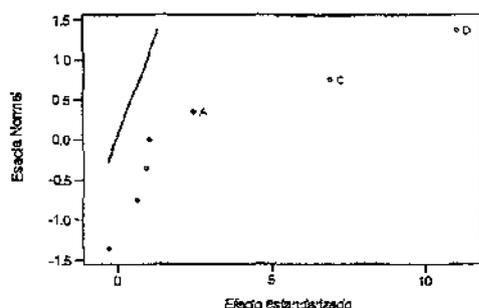


Figura 2.41: Efectos significativos

Tabla 2.23. Estimadores para el modelo de la media

Término	Estimador	Error estándar	valor p
Intercepto	81.350	0.386	<.0001
A[-1-1]	-0.975	0.386	0.0161
C[-1-1]	-2.725	0.386	<.0001
D[-1-1]	-4.360	0.386	<.0001

pero los residuos del modelo

$$Y = \left\{ \begin{array}{l} -0.975 \text{ cuando } A^- \\ 0.975 \text{ cuando } A^+ \end{array} \right\} + \left\{ \begin{array}{l} -2.725 \text{ cuando } C^- \\ 2.725 \text{ cuando } C^+ \end{array} \right\} + \left\{ \begin{array}{l} -4.36 \text{ cuando } D^- \\ 4.36 \text{ cuando } D^+ \end{array} \right\} \quad (2.32)$$

ratifican la falta de homogeneidad de varianzas (figura 2.42) y generan un $CME = 5.961$.

MODELO I

Para analizar la dispersión de las observaciones se propone usar el logaritmo natural de los residuos, del modelo no ponderado, al cuadrado para estimar los parámetros de una función varianza de la forma 2.31. El resultado, que se muestra en la siguiente tabla, coincide con el análisis de Grego al resaltar la influencia significativa de los factores A y D en sus niveles bajo y alto, respectivamente, para minimizar la varianza.

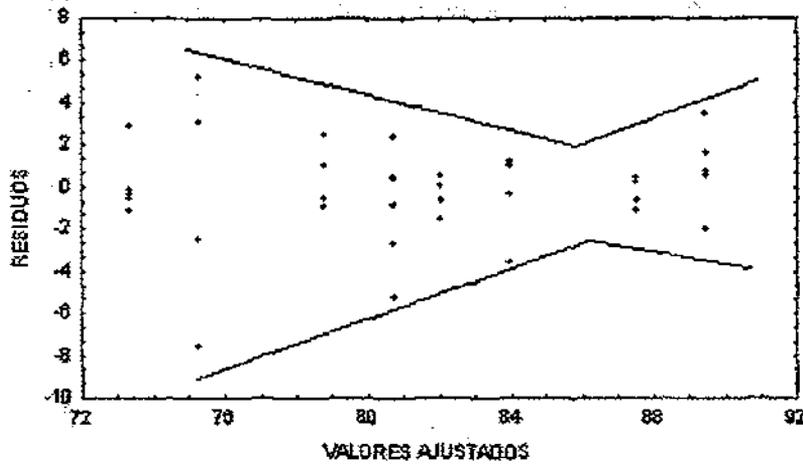


Figura 2.42: Residuos vs. valores ajustados

Tabla 2.24. estimadores para la función varianza

Término	Estimador	Error estándar	Valor p
Intercepto	0.1505	0.2696	0.5801
A[-1 -1]	-1.04	0.2696	0.0004
D[-1 -1]	0.515	0.2696	0.0639

MODELO II

Si se consideran las medias y varianzas, por tratamiento, como las variables respuesta para la búsqueda de los modelos respectivos y se grafican los efectos que los factores tienen sobre ambos parámetros (figura 2.43) se observa con claridad que se confirma lo antes expuesto, esto es: los factores A y D tienen influencia sobre la media y sobre la varianza; mientras C únicamente participa en el comportamiento de la fuerza adherente promedio.

Los resultados de la estimación de los parámetros (tabla 2.25) para la función varianza son similares a los ya obtenidos.

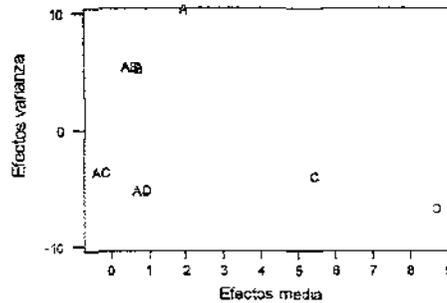


Figura 2.43: Efectos para la media y la varianza

Tabla 2.25. Estimadores para la función varianza

término	estimador	error estándar	valor p
intercepto	2.088	0.2993	0.0009
A[-1 - 1]	-0.967	0.2993	0.0232
D[-1 - 1]	0.697	0.2993	0.0673

En resumen, de cualquier manera, se puede concluir que se debe tener especial atención en los factores A y D si se busca minimizar la varianza. Para lograrlo se necesitan sus niveles bajo y alto respectivamente mientras, la máxima fuerza adherente promedio se logra con los niveles altos de A, C y D.

Analizando los grupos formados para A^+ y A^- en los niveles altos de C y D se tienen las estadísticas de la tabla 2.26 y las pruebas de Levene, Bartlett y Hartley no rechazan la igualdad de varianzas al 95% de confianza por lo que, lo recomendable es aplicar adhesivo tipo H-1-E durante 120 minutos en un depósito de plata sin importar si el material conductor empleado es el níquel o el cobre.

Tabla 2.26. estadísticas por grupo

A	n	media	desviación est.
-1	5	87.06	0.709
1	5	90.32	1.994

2.3.5 Comentarios

A lo largo de este capítulo se exploró, por medio de diversas estrategias, el comportamiento de la dispersión de las observaciones. Cuando la heteroscedasticidad ha sido por demás clara, como es el caso del ejemplo de Grego, las conclusiones a las que se llega por diferentes caminos coinciden. Sin embargo, ante una situación un tanto polémica como el problema de los resortes automotrices, el análisis exploratorio puede abrir diferentes alternativas de solución.

Capítulo 3

Estudio de efectos de dispersión

La inquietud por modelar la varianza surgió hace poco más de cincuenta años. De esa época a la fecha se han documentado una serie de alternativas que se han visto reforzadas con el gran desarrollo cibernético que permite desarrollar cálculos bastante engorrosos a una velocidad increíble. En este capítulo se presenta una breve cronología sobre la formación de una idea final: estimar los parámetros de un modelo para la varianza de las observaciones.

3.1 Antecedentes

Los trabajos de Bartlett, a mediados del siglo XX, fueron los primeros que motivaron el estudio de la relación entre la media y la varianza. Un método empírico que propuso fue graficar el logaritmo de la varianza dentro de cada celda de información contra el logaritmo de la media en la celda correspondiente y observar su comportamiento. El mismo autor sugirió el uso de una transformación de los datos para minimizar alguna medida de heteroscedasticidad sin alterar la independencia de las observaciones y a la vez acercarse más a la distribución normal recuperando así los supuestos básicos del modelo de regresión. Esta idea se avanza con las transformaciones de Box-Cox.

A partir de aquí empezaron a surgir una serie de propuestas para modelar la varianza, por ejemplo, Hildreth y Houck en 1968; Goldfeld y Quandt en 1972 y Amemiya en 1977 estudiaron las siguientes expresiones:

$$\text{Var}(y_i) = \theta_0 + \theta_1 x_i \quad (3.1)$$

función lineal de las predictoras

$$\text{Var}(y_i) = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 \quad (3.2)$$

función cuadrática de las predictoras

$$\text{Var}(y_i) = \theta_0 + \theta_1 (\mu_i(\beta))^{\theta_2} \quad (3.3)$$

función que depende de la media

$$\text{Var}(y_i) = \sigma^2 (\exp(\theta_0 + \theta_1 x_i)) \quad (3.4)$$

función no lineal en los parámetros

Es importante aclarar que a diferencia de los modelos de medias, las funciones de varianza tienen una restricción en cuanto a valores ajustados se refiere, es decir, mientras los promedios pueden tomar valores en el eje real, las varianzas deben ser siempre positivas, por lo que otra posibilidad es tomar en cuenta modelos lineales en las predictoras para el logaritmo de las varianzas, por ejemplo:

$$\ln(\sigma^2) = \theta_0 + \theta_1 x_i \quad (3.5)$$

$$\ln(\sigma^2) = \theta_0 + \theta_1 \mu(\beta) \quad (3.6)$$

$$\ln(\sigma^2) = \theta_0 + \theta_1 \ln(\mu(\beta)) \quad (3.7)$$

Esta propuesta es la única que garantiza que las varianzas estimadas serán positivas.

3.2 El modelo

Entre los autores que han estudiado este tipo de modelos están Box y Hill en 1974; Harvey en 1976 y la pareja Carroll - Ruppert en 1982.

La diversidad de propuestas para los modelos de varianza puede ser tan amplia como la variedad de modelos para la media, por lo que, igual que en el caso de la media, es importante tener elementos para seleccionar el mejor.

Davidian y Carroll (1987), generalizan las expresiones para la varianza y la media a través de un modelo de regresión heteroscedástico que consta de dos partes: la primera representa al valor esperado de las observaciones como una función de las variables regresoras $\{x_i\}$ y de un vector de parámetros $\beta_{(p \times 1)}$; la segunda corresponde a la función varianza que queda en términos de una función de las covariables $\{z_i\}$ y de los parámetros $\beta_{(p \times 1)}$ y $\alpha_{(r \times 1)}$. Así, el modelo es:

$$E(y_i) = \mu_i = f(x_i, \beta) \quad (3.8)$$

$$Var(y_i) = \sigma^2 g^2(\beta, z_i, \alpha) \quad (3.9)$$

donde :

- $\{x_i\}$ y $\{z_i\}$ son vectores conocidos, ($\{z_i\}$ puede ser un subconjunto propio de $\{x_i\}$);
- f y g son funciones con forma conocida y
- σ es un parámetro de escala desconocido.

Una característica importante de este modelo es que no se requiere hacer ningún supuesto sobre la distribución de Y , los supuestos sólo se hacen con respecto a los dos primeros momentos de los datos.

Para hacer explícita una posible relación entre la dispersión y la media, Carroll y Ruppert (1988) presentan el mismo modelo pero con una pequeña modificación en la función varianza, quedando la siguiente expresión:

$$E(y_i) = \mu_i = f(x_i, \beta) \quad (3.10)$$

$$Var(y_i) = \sigma^2 g^2(\mu_i(\beta), z_i, \alpha) \quad (3.11)$$

Por otro lado, Nelder y Pregibon (1987) ofrecen un enfoque basado en modelos lineales generalizados para analizar efectos de localización y dispersión de los datos en el caso en que la varianza depende de la media por lo que puede considerarse un caso particular de la propuesta anterior. El modelo que proponen es:

$$E(y_i) = \mu_i = f(x_i, \beta) \quad (3.12)$$

$$Var(y_i) = \phi_i V(\mu_i) \quad (3.13)$$

$$f(\mu_i) = x_i \beta \quad (3.14)$$

$$g(\phi_i) = z_i \alpha \quad (3.15)$$

donde:

- f y g son funciones liga conocidas,
- $V(\mu)$ es una función varianza, positiva, también conocida.
- ϕ es el parámetro de dispersión.

La selección de la función varianza determina la interpretación de ϕ , esto es, si $V(\mu) = 1$, ϕ es la varianza de la respuesta; si $V(\mu) = \mu^2$, ϕ es el cuadrado del coeficiente de variación de la respuesta. En los modelos lineales generalizados más simples, el parámetro de dispersión ϕ es una constante, usualmente desconocida, pero en los casos en que Y_i es el promedio de m_i repeticiones puede ser apropiado suponer que ϕ_i es proporcional a $w_i = \frac{1}{m_i}$ ("pesos" conocidos).

La idea de la transformación de observaciones es retomada por Logothetis (1990) y plantea que si existe una relación funcional entre la media y la varianza dada por $\sigma_y = f(\mu_y)$, entonces, para obtener datos en los que la dispersión σ_T sea igual a una constante c , se requiere de una transformación de la forma:

$$T(\mu) = \int \frac{c}{f(\mu)} d\mu \quad (3.16)$$

En el caso en que la varianza sea una potencia de la media, por ejemplo,

$$\sigma_y = f(\mu) = \alpha\mu^\theta \quad (3.17)$$

la transformación es equivalente a la propuesta de la función potencia de Box-Cox:

$$T(\mu) = \begin{cases} C_1\mu^{1-\theta} & \text{si } \theta \neq 1 \\ C_1 \ln \mu & \text{si } \theta = 0 \end{cases} \quad (3.18)$$

y una manera de determinar si la relación funcional es adecuada, es a través de la estimación por mínimos cuadrados de θ en la regresión que involucra la desviación estándar y la media muestrales :

$$\ln(s_i) = \ln \alpha + \theta \ln(\bar{X}_i) + \varepsilon_i \quad (3.19)$$

Logothetis (1990) afirma que, en muchos casos, el resultado de la estimación de θ en el modelo anterior es igual al valor que se obtiene para λ al aplicar la técnica de la transformación Box-Cox y aclara que cuando los resultados no son iguales, se debe a que, en el modelo, se requiere de aumentar términos de un orden mayor o algunas interacciones.

Desde este punto de vista, Logothetis concluye que resolver la expresión 3.19 es un método suficiente para determinar el tipo de transformación para los datos.

Sin embargo, Engel (1992) cuestiona la propuesta de Logothetis argumentando la desventaja que tienen los métodos basados en transformaciones de datos en el sentido de que se está proponiendo un modelo lineal para alguna función estabilizadora de varianza no interpretable. Pero además, Engel afirma que el estimador por mínimos cuadrados para θ puede ser inconsistente cuando el número de repeticiones aumenta ($m \rightarrow \infty$) y que debido a que el intercepto en 3.19 es constante, el error estándar de α tiende a ser grande.

De esta manera, Engel retoma la experiencia tanto de Carroll-Ruppert como de McCullagh-Nelder y de Logothetis para proponer un modelo de la forma:

$$E(y_i) = \mu_i = x_i\beta, \quad (3.20)$$

$$Var(y_i) = \phi_i V(\mu_i, \alpha), \quad (3.21)$$

$$\ln \phi_i = z_i\gamma \quad (3.22)$$

donde β, α, ϕ_i y γ son parámetros.

La idea de Engel es hacer una separación explícita de los factores en aquéllos que tienen efecto sobre la media (factores control), los que afectan a la varianza (factores señal) y de los llamados factores de costo.

Es en este sentido que la expresión para la varianza en este modelo resulta ser un caso especial de la que usaron Davidian y Carroll, pues mientras ϕ_i es un parámetro que no depende de la media, $V(\mu_i, \alpha)$ es un término que está en función de ella (además de α).

Por otro lado, $Var(y_i) = \phi_i V(\mu_i, \alpha)$ es una extensión de la expresión usada por Nelder y Pegibon al hacer depender la función varianza de un parámetro α . Nótese, además, que las funciones liga f y g son la identidad y el logaritmo natural respectivamente.

A partir de lo anterior, el estudio que se hace en esta tesis sobre los modelos de regresión heteroscedásticos queda definido formalmente de la siguiente manera:

Se tiene un vector variable de respuesta, Y , que describe una característica del producto. Las observaciones y_1, y_2, \dots, y_n son n variables aleatorias independientes que provienen de poblaciones normales con diferentes medias y varianzas, es decir, tanto su valor promedio como su variabilidad se ven afectadas por la influencia de covariables.

Las covariables $\{x_i\}$, llamadas factores de control, pueden tener efecto tanto en la media como en la varianza. Por otro lado, se tienen factores no controlados $\{z_i\}$, denominados de ruido, que afectan la dispersión de las respuestas.

De esta manera, si $\varphi^2(\alpha)$ representa el modelo para la varianza de Y , la respuesta tiene la siguiente expresión:

$$Y = \mu(\beta) + \varphi(\alpha) \varepsilon \quad (3.23)$$

donde:

- ε es un vector de variables aleatorias,
- β y α son los parámetros para la media y la función varianza respectivamente.

3.3 Notación

En particular, el interés se centra en un modelo lineal para la media y uno multiplicativo para la varianza. De manera específica, las n respuestas Y_i se suponen normalmente distribuidas con media y varianza dadas por las siguientes expresiones:

$$E(Y) = X\beta + e \quad (3.24)$$

$$Var(Y) = \exp(X\alpha) \quad (3.25)$$

donde:

- e es un vector de variables aleatorias que siguen una distribución normal:

$$e \sim N(0, \exp(X\alpha)) \quad (3.26)$$

En términos de la notación usada por Carroll y Ruppert (1988), se tiene que

$$E(y_i) = \mu_i = f(x_i, \beta) = X\beta \quad (3.27)$$

$$Var(y_i) = \sigma^2 \exp(Z\alpha) \quad (3.28)$$

donde la relación entre Z y X puede ser:

- $X = Z$, es decir los factores que tiene influencia significativa sobre la media son los mismos que afectan la varianza de las observaciones.
- $X \neq Z$, cuando los factores que tienen efectos significativos sobre la media son diferentes a los que explican el comportamiento de la varianza
- X y Z coinciden sólo en algunos componentes cuando existen factores que repercuten tanto en el valor esperado como en la dispersión de la variable respuesta.

Los vectores β_{mx1} y α_{px1} , son los parámetros del modelo de medias y de la función varianza respectivamente. Además, σ^2 se conoce como el parámetro estructural y es fijo pero desconocido.

De esta manera, la estimación de parámetros se convierte en un problema de estimación del vector

$$\delta_{(m+p+1)x1} = \begin{pmatrix} \beta \\ \alpha \\ \sigma^2 \end{pmatrix} \quad (3.29)$$

Capítulo 4

Modelación de efectos de dispersión

4.1 Introducción.

Los estimadores de los parámetros de un modelo de regresión clásico se obtienen a través de métodos como máxima verosimilitud o mínimos cuadrados. Se sabe que en el primero es necesario conocer la distribución de los datos (o por lo menos hacer un supuesto al respecto con algún fundamento importante) y que en el caso de que esta distribución sea la normal, los estimadores que se obtienen por ambos métodos coinciden.

Dado que las funciones varianza también dependen de parámetros, ha sido necesario establecer métodos para su estimación. Los procedimientos que aquí se presentan son:

- Regresión en dos etapas
- Mínimos cuadrados generalizados
- Máxima verosimilitud
- Pseudoverosimilitud
- Quasiverosimilitud

Todos ellos tienen una base en los métodos conocidos para la media y, con excepción del primero de la lista, se trata de métodos que funcionan de manera iterativa.

Para la descripción de los procedimientos, se usa la notación de Carroll y Ruppert (1988), es decir, se considera el modelo de la forma:

$$E(y_i) = f(x_i, \beta) = \mu_i$$

$$Var(y_i) = \sigma^2 g^2(\mu_i(\beta), z_i, \alpha)$$

donde:

$$g^2(\mu_i(\beta), z_i, \alpha) = \exp(X\alpha)$$

Cada procedimiento es ejemplificado con detalle a través de su aplicación en la estimación de los parámetros del modelo para la presión arterial diastólica de las 54 mujeres (ejemplo 2.1).

En este capítulo se empieza a hacer algunas comparaciones entre métodos a través del *CME* y al final se presentan los resultados de la estimación de los parámetros en los problemas de los resortes automotrices y los circuitos integrados.

4.2 Procedimientos de estimación

En el capítulo 2 se hizo un análisis exploratorio del comportamiento de la dispersión de los datos y se propuso, por separado, un modelo para la media y para la varianza. Aquí se presentan métodos de estimación que permiten reflejar el efecto de la varianza en el modelo de la media.

4.2.1 Regresión en dos etapas (2E)

Harvey (1976) propone la estimación de los parámetros de la función varianza siguiendo el principio del método de mínimos cuadrados ordinarios¹ (MCO).

Los cuadrados de los residuos que se obtienen de un modelo de regresión ordinaria para la media, son considerados como un reflejo de la estructura de la variabilidad de las observaciones por lo que, en la segunda etapa, se buscan los estimadores de un modelo de regresión lineal donde éstos juegan un papel relevante.

Los pasos a seguir son:

1. Obtener estimadores por MCO para la media , esto es:

$$\hat{\beta} = (X'X)^{-1}(X'Y) \quad (4.1)$$

2. Obtener estimadores por MCO para la función varianza, usando el logaritmo de los residuos al cuadrado como la variable respuesta en el modelo lineal

$$\ln(r_i^2) = Z_i\alpha + w_i \quad (4.2)$$

de tal manera que

$$\hat{\alpha} = (Z'Z)^{-1}Z'\ln(r_i^2) \quad (4.3)$$

Una de las ventajas más reconocidas del método de mínimos cuadrados es que no se requiere ningún supuesto distribucional para las observaciones, sin embargo, la transformación que se hace de los residuos, en la segunda

¹El método de mínimos cuadrados ordinarios se debe al matemático alemán Carl Friedrich Gauss(1777-1855)

etapa, genera valores negativos para residuos menores a 1. Para garantizar la no negatividad de la función varianza, Harvey hace

$$\widehat{Var}(y_i) = \widehat{\sigma}_i^2 = \exp(Z\widehat{\alpha}). \quad (4.4)$$

Es importante notar que bajo el supuesto de normalidad de las observaciones, los residuos, $r_i = y_i - \widehat{y}_i$, deben seguir una distribución aproximadamente normal y si la expresión $\exp(Z\alpha)$ realmente representa la parte sistemática de la variabilidad, la variable de la forma

$$w_i = \ln\left(\frac{r_i^2}{\sigma_i^2}\right) \quad (4.5)$$

converge en distribución a una variable aleatoria w_i^* que se distribuye como el logaritmo de una χ^2 con un grado de libertad cuyos primeros dos momentos están dados por

$$E(w_i^*) = \psi\left(\frac{1}{2}\right) - \ln\left(\frac{1}{2}\right) = -1.2704 \quad (4.6)$$

$$Var(w_i^*) = \psi^{(1)}\left(\frac{1}{2}\right) = 4.9348 \quad (4.7)$$

donde: $\psi(s) = \frac{d \ln(\Gamma(s))}{ds}$ es la función digamma y $\psi^{(1)}(s)$ denota la primer derivada de $\psi(s)$

De esta manera, Harvey concluye que

$$Var(\widehat{\alpha}) = 4.9348 \left[\sum z_i z_i' \right]^{-1} \quad (4.8)$$

y que $\widehat{\alpha}$ es un estimador asintóticamente eficiente.

Ejemplo

En el problema de la presión arterial diastólica de las 54 mujeres con edades entre 20 y 60 años, como resultado de la estimación por el método de regresión en dos etapas, se tiene

$$\beta = \begin{bmatrix} 56.1569 \\ 0.580 \end{bmatrix}$$

y

$$\alpha = \begin{bmatrix} 0.4068 \\ 0.0577 \end{bmatrix}$$

de tal manera que el modelo es

$$\hat{Y}_i = 56.1569 + 0.5800x_i + \sqrt{\exp(0.4068 + 0.05767x_i)} * \varepsilon_i \quad (4.9)$$

con

$$\varepsilon \sim N(0, 1) \quad (4.10)$$

Nótese que, la diferencia entre una estimación ordinaria y la regresión en dos etapas está en la existencia del segundo paso y que a partir del análisis exploratorio que se llevó a cabo a este conjunto de datos en el capítulo 2, se ha usado la edad de las personas como una variable explicativa tanto en el modelo para la media como en la función varianza.

Las estadísticas de los residuos del modelo de la función varianza son

$$\bar{w} = 0 \quad (4.11)$$

$$var(w) = 5.3998 \quad (4.12)$$

La figura 4.1 contiene los residuos del modelo para la varianza que muestran un comportamiento visiblemente más disperso para los residuos negativos que para los positivos como una clara consecuencia de la falta de ajuste para los residuos pequeños (menores que 1) del modelo para la media.

De esta manera, se tiene un argumento importante para pensar que la función varianza no representa la variabilidad intrínseca de las observaciones. No obstante, si se calcula la suma de cuadrados del error a través de residuos de la forma

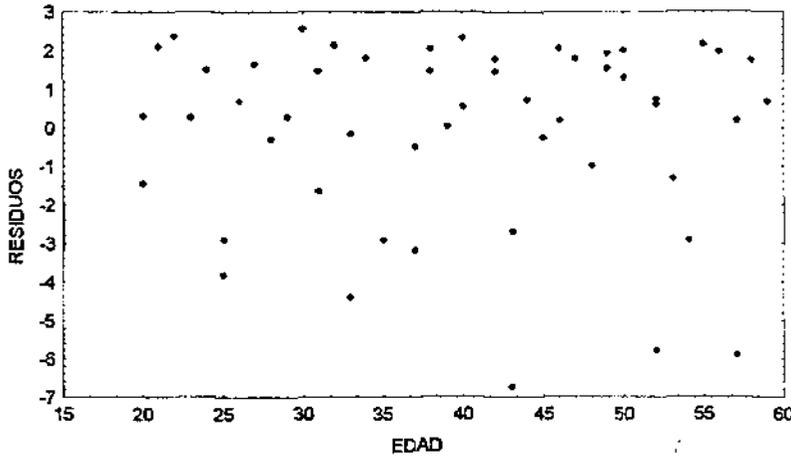


Figura 4.1: Residuos función varianza

$$SCE = \sum \left(\frac{y_i - \hat{y}_i}{\sqrt{\exp(0.4068 + 0.0577x_i)}} \right)^2 \quad (4.13)$$

se tiene $SCE = 1152.154$ por lo que

$$CME = \frac{SCE}{n - p} = \frac{1152.124}{52} = 22.156 \quad (4.14)$$

El valor del cuadrado medio del error que se obtiene con este modelo será comparado con el que resulte de aplicar otros métodos.

4.2.2 Mínimos Cuadrados Generalizados (MCG)

Ante la falta de homogeneidad de varianzas se tiene que algunas de las observaciones son "menos confiables" que otras en el sentido de que si la variable respuesta y_1 proviene de una población más dispersa que otra variable y_2 , entonces, y_1 proporciona más incertidumbre sobre los resultados del proceso que la ha generado en comparación con y_2 . De esta manera, la idea del método de mínimos cuadrados generalizados sigue el mismo principio que la

regresión ponderada, esto es, a cada observación se le asigna un peso que es inversamente proporcional a la varianza de la población a la que pertenece.

Si las varianzas son desconocidas, éstas pueden ser estimadas a través de las varianzas muestrales en cada punto de diseño, sin embargo, es sabido que tal estimador puede ser muy inestable si el número de repeticiones por punto es pequeño, además, por este medio no se logra identificar una estructura para las varianzas.

La alternativa que plantean Carroll y Ruppert (1983) es estimar la varianza por medio de una regresión y, por consiguiente, sin necesidad de tener repeticiones.

El método de mínimos cuadrados generalizados parte de una regresión ordinaria para obtener un estimador inicial de β (vector parámetro del modelo para la media) y, los residuos generados son usados para estimar el vector parámetro (α) de una función varianza propuesta. El proceso es iterativo y se busca minimizar la suma de cuadrados de los residuos ponderados, esto es:

$$\sum w_i [y_i - f(x_i, \beta)]^2 \quad (4.15)$$

donde:

$$w_i = \frac{1}{g^2(\mu(\beta), z_i, \alpha)}$$

representa el peso de la observación i y

$$f(x_i, \beta)$$

es el valor esperado de y_i .

No se requieren supuestos sobre la distribución de los datos pero sí es necesario establecer una relación entre la media y la varianza, es decir, se debe proponer una expresión para la función varianza.

El algoritmo que proponen Carroll y Ruppert (1988) es el siguiente:

1.- Empezar con un estimador preliminar $\hat{\beta}_*$ (puede ser el resultado de una estimación no ponderada)

2.- Estimar α a través de la solución a una ecuación de la forma :

$$\sum H(y_i, z_i, \mu_i(\hat{\beta}_*), \alpha) = 0$$

3.- Calcular los pesos estimados:

$$\hat{w}_i = \frac{1}{g^2(\mu_i(\hat{\beta}_*), z_i, \hat{\alpha})}$$

4.- Obtener un estimador por mínimos cuadrados ponderados para β y actualizar el valor de $\hat{\beta}_*$.

5.- Repetir los pasos 2, 3 y 4 hasta que los cambios en los valores de los estimadores en dos iteraciones sucesivas sean tan pequeños como se desee o hasta acumular un cierto número de iteraciones para prevenir aquellos casos en los que el proceso no converge.

Cuando el número de iteraciones tiende a infinito, los resultados del algoritmo coinciden con la solución de las ecuaciones

$$\sum f_{\beta}(x_i, \beta) [y_i - f(x_i, \beta)] / g^2(\mu_i(\beta), z_i, \alpha) = 0 \quad (4.16)$$

$$\sum H(y_i, z_i, \mu_i(\beta), \alpha) = 0 \quad (4.17)$$

donde $f_{\beta}(x_i, \beta)$ es la derivada de f con respecto a β .

En cuanto a la distribución asintótica de $\hat{\beta}_*$ se puede mencionar el siguiente

Teorema: Para cualquier estimador inicial $N^{\frac{1}{2}}$ - consistente $\hat{\beta}_*$ y cualquier número de iteraciones, el estimador por mínimos cuadrados generalizados $\hat{\beta}_{mcg}$ se distribuye, asintóticamente, como una normal con parámetros β y $\sigma^2 / \sum \frac{f_{\beta}(x_i, \beta) f_{\beta}(x_i, \beta)^t}{g^2(\mu_i(\beta), z_i, \alpha)}$, esto es

$$\hat{\beta}_{mcg} \sim N\left(\beta, \sigma^2 / \sum \frac{f_{\beta}(x_i, \beta) f_{\beta}(x_i, \beta)^t}{g^2(\mu_i(\beta), z_i, \alpha)}\right) \quad (4.18)$$

Por otro lado, Goldfeld y Quandt (1972), por medio de simulación Monte Carlo, y Williams (1975) por medios analíticos y también empíricos concluyen que, en muestras pequeñas, la eficiencia de la estimación de β depende de la eficiencia con que se estima α . Posteriormente, Toyooka (1982) y Rothenberg (1984) explican que para datos normalmente distribuidos y cuya varianza depende de las variables predictoras pero no de la media, se tiene que:

$$Var \left[\sqrt{N}(\hat{\beta}_* - \beta) \right] \simeq \sigma^2 S_G^{-1} + N^{-1} V(\hat{\alpha})$$

donde:

$$S_G = N^{-1} \sum \frac{f_{\beta}(x_i, \beta) f_{\beta}(x_i, \beta)^t}{g^2(\mu_i(\beta), z_i, \alpha)}$$

y $V(\hat{\alpha})$ es una función creciente de la varianza asintótica de $\hat{\alpha}$.

De esta manera, al menos para datos normales, entre mejor estimador sea $\hat{\alpha}$ mejor será también $\hat{\beta}$.

Ejemplo

En los datos de la presión sanguínea de las 54 mujeres se tiene un estimador por mínimos cuadrados ordinarios $\hat{\beta}_* = [56.1569, 0.58]$ como valor inicial. Esto es,

$$f(x_i, \beta) = X\beta$$

por lo que

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

Tomando en cuenta la propuesta de Harvey, en el segundo paso se usan los residuos (\widehat{u}_i) del ajuste preliminar y se estima α por mínimos cuadrados ordinarios a través de

$$\ln(\widehat{u}_i^2) = x_i \widehat{\alpha}$$

De esta manera, en el paso 3, se toma

$$g^2(\mu(\beta), z_i, \alpha) = \exp(x_i \alpha)$$

por lo que los pesos son de la forma:

$$\widehat{w}_i = \frac{1}{\exp(x_i \widehat{\theta})} \quad (4.19)$$

y se usan para estimar β ponderado quien ahora toma el lugar del estimador preliminar para una nueva estimación de α .

Después de 615 iteraciones (para obtener diferencias menores a 0.00001 entre dos estimaciones sucesivas), se tiene:

$$\widehat{\beta}'_* = [56.113, 0.5812] \quad (4.20)$$

$$\widehat{\alpha}'_* = [0.8989, 0.0289] \quad (4.21)$$

Por lo que el modelo propuesto es de la forma:

$$\widehat{y}_i = 56.113 + 0.5812x_i \quad (4.22)$$

$$\widehat{Var}(y_i) = \exp(0.8989 + 0.0289x_i) \quad (4.23)$$

y

$$r_i^2(\beta, \alpha) = \frac{(Y_i - 56.113 - 0.5812x_i)^2}{\exp(0.8989 + 0.0289x_i)} \quad (4.24)$$

mismos que generan una $SCE = 377.35$, esto es,

$$CME = \frac{377.35}{52} = 7.256 \quad (4.25)$$

Aunque el CME es menor al que se obtuvo por el método de regresión en dos etapas, la representación gráfica de estos residuos contra la edad (figura 4.2) aún muestra que la varianza de las observaciones es mayor conforme la edad de las mujeres de la muestra aumenta.

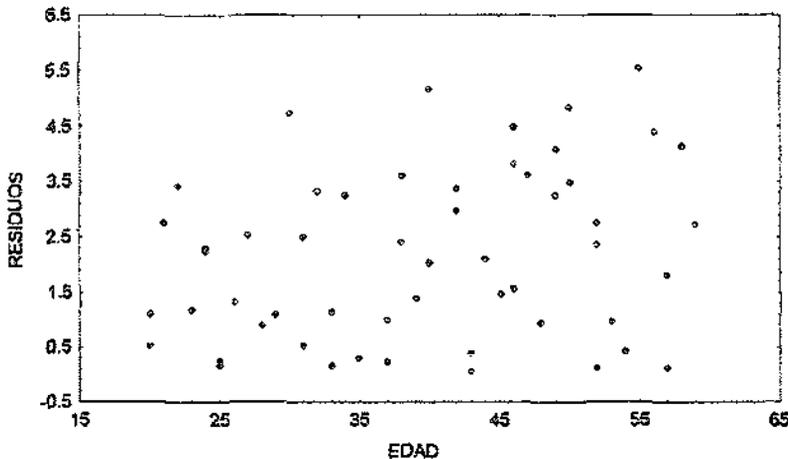


Figura 4.2: Residuos Mínimos Cuadrados Generalizados

4.2.3 Máxima Verosimilitud (MV)

En este método se parte del supuesto de que la forma de la distribución de los datos es conocida pero es necesario estimar sus parámetros. La estimación se lleva a cabo de manera simultánea para β y θ .

En el caso de la distribución normal se tiene que la función de densidad es de la forma:

$$f(y_i; \mu_i, \alpha_i) = \frac{1}{\sqrt{2\pi}\sigma g(\mu_i(\beta), z_i, \alpha)} \exp \frac{-1}{2} \left(\frac{y_i - \mu_i(\beta)}{\sigma g(\mu_i(\beta), z_i, \alpha)} \right)^2 \quad (4.26)$$

de aquí que la expresión para la función de verosimilitud es:

$$L(\theta) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \left(\frac{1}{\prod \sigma g(\mu_i(\beta), z_i, \alpha)} \right) \exp \left[\frac{-1}{2} \sum_1^n \left(\frac{y_i - \mu_i(\beta)}{\sigma g(\mu_i(\beta), z_i, \alpha)} \right)^2 \right]$$

con

$$\theta = (\beta_1, \beta_2, \dots, \beta_k, \alpha_1, \alpha_2, \dots, \alpha_m)$$

y la log-verosimilitud está dada por :

$$\ln L(\theta) = \frac{-n}{2} \ln(2\pi) - \frac{1}{2} \sum_1^n \ln(\sigma^2 g^2(\mu_i(\beta), z_i, \alpha)) - \frac{1}{2} \sum_1^n \left(\frac{y_i - \mu_i(\beta)}{\sigma g(\mu_i(\beta), z_i, \alpha)} \right)^2 \quad (4.27)$$

Aplicando el método de Newton-Raphson para maximizar, se tiene

$$U(\theta) = [U_1(\theta), U_2(\theta), \dots, U_{m+k}(\theta)] \quad (4.28)$$

es el vector *score* en θ , donde

$$U_i(\theta) = \frac{\partial \ln L(\theta)}{\partial \theta_i} = 0 \quad (4.29)$$

para $i = 1, 2, \dots, m+k$ son las ecuaciones de verosimilitud.

Sea θ_0 un valor inicial de $\hat{\theta}$. Cada $U_i(\theta)$ se expande en una serie de Taylor alrededor de θ_0 , esto es

$$U(\theta) \doteq U(\theta_0) + G(\theta_0)(\theta - \theta_0)$$

donde $G(\theta)$ es la matriz con entradas

$$G_{ij}(\theta) = \frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \quad (4.30)$$

como $\hat{\theta}$ debe cumplir que $U(\hat{\theta}) = 0$, se tiene que

$$\hat{\theta} \doteq \theta_0 - G(\theta_0)^{-1} U(\theta_0) \quad (4.31)$$

o equivalentemente, en términos de la matriz de información de Fisher,

$$\hat{\theta} \doteq \theta_0 + I(\theta_0)^{-1} U(\theta_0) \quad (4.32)$$

Los valores iniciales θ_0 pueden ser los estimadores obtenidos por el método de mínimos cuadrados ordinarios.

El estimador máximo verosímil de σ^2 es

$$\hat{\sigma}^2 = \sum \frac{(y_i - x_i \beta)^2}{n(g^2(\mu_i(\beta), z_i, \alpha))} = \frac{SCE}{n} \quad (4.33)$$

Para corregir el sesgo, se tiene la siguiente expresión:

$$\hat{\sigma}^2 = \sum \frac{(y_i - x_i \beta)^2}{(n-p)(g^2(\mu_i(\beta), z_i, \alpha))} = \frac{SCE}{n-p} \quad (4.34)$$

Como propiedades de los estimadores máximo verosímiles se tiene que son funciones de estadísticas suficientes, son consistentes y asintóticamente normales.

La estimación de parámetros por máxima verosimilitud es más eficiente que la de mínimos cuadrados cuando la distribución de los datos es normal y si las funciones de la media y la varianza están correctamente especificadas.

Ejemplo

En el ejemplo de la presión sanguínea del grupo de mujeres, el resultado de la estimación por máxima verosimilitud es

$$\hat{\beta}_0 = 55.0428$$

$$\hat{\beta}_1 = 0.61157$$

$$\hat{\alpha}_0 = 1.02009$$

$$\hat{\alpha}_1 = 0.07249$$

La gráfica de los residuos estandarizados está en la figura 4.3

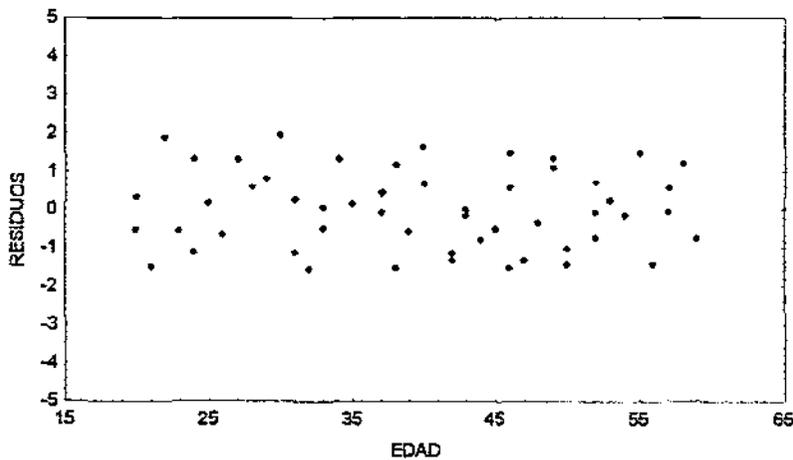


Figura 4.3: Residuos máximo verosímiles

y el cuadrado medio del error es $CME = 1.03225$.

Además de que el CME es menor aún que el generado por la estimación por mínimos cuadrados generalizados, el diagrama de dispersión de los residuos contra la edad, no muestra patrón alguno de comportamiento por lo que se puede pensar que ya no se tiene ningún efecto sistemático sin modelar.

4.2.4 Pseudoverosimilitud (PV)

Dado que en el método de mínimos cuadrados generalizados se tiene que la eficiencia en la estimación de β depende de la eficiencia en la estimación de α , la idea, entonces, es buscar el mejor estimador posible para α .

Se sabe que bajo un supuesto distribucional correctamente sustentado, el estimador máximo verosímil es el más eficiente, por lo que Carroll y Rupert (1988) proponen el método de estimación por máxima pseudoverosimilitud que consiste en la combinación de dos procesos de estimación: uno para obtener un valor inicial de $\hat{\beta}$ y el otro para estimar α y σ por máxima verosimilitud normal, suponiendo β conocido. El nombre de pseudoverosimilitud se debe a que los estimadores así obtenidos no coinciden con los valores respectivos que se obtienen por máxima verosimilitud simultánea a menos que el estimador inicial de $\hat{\beta}$ se haya obtenido también por máxima verosimilitud y la varianza no esté en función de la media.

Este es un método en el que el supuesto distribucional no es tan importante como el que establece una relación entre la media y la varianza.

Con lo que se tiene que

$$E(y_i) = \mu_i(\beta) = f(x_i, \beta) \quad (4.35)$$

$$Var(y_i) = \sigma^2 g^2(\mu_i(\beta), z_i, \alpha) \quad (4.36)$$

son las expresiones para la media y para la varianza en función de la media de las observaciones, respectivamente. Si se supone $\beta = \hat{\beta}_n$ conocida, la función a maximizar para α es

$$l_{PL}(\beta, \alpha, \sigma) = -n \ln \sigma - \sum \ln [g \{ \mu_i(\beta), z_i, \alpha \}] - (2\sigma^2)^{-1} \sum \frac{\{y_i - f(x_i, \beta)\}^2}{g^2 \{ \mu_i(\beta), z_i, \alpha \}} \quad (4.37)$$

El proceso es iterativo y se puede empezar con una estimación por mínimos cuadrados ordinarios para β .

El estimador insesgado para σ^2 tiene la misma expresión que el obtenido por máxima verosimilitud simultánea.

$$\hat{\sigma}^2 = \sum \frac{(y_i - x_i\beta)^2}{(n-p)(g^2(\mu_i(\beta), z_i, \alpha))} = \frac{SCE}{n-p} \quad (4.38)$$

Ejemplo

En el ejemplo de la presión sanguínea del grupo de mujeres, el resultado de la estimación por pseudoverosimilitud es

$$\begin{aligned} \hat{\beta}_0 &= 55.756098 \\ \hat{\beta}_1 &= 0.588986 \\ \hat{\alpha}_0 &= 1.04294 \\ \hat{\alpha}_1 &= 0.0719185 \end{aligned}$$

La gráfica de los residuos estandarizados está en la figura 4.4

El cuadrado medio del error es igual al que se obtiene en la estimación por máxima verosimilitud y los diagramas de dispersión, en ambos casos, son prácticamente iguales.

4.2.5 Quasiverosimilitud (QV)

A diferencia de los procedimientos anteriores, en este método sólo se estiman los parámetros de la media porque se parte del supuesto de que la función varianza es conocida.

En los modelos lineales generalizados, un criterio de ajuste está dado por la *devianza* :

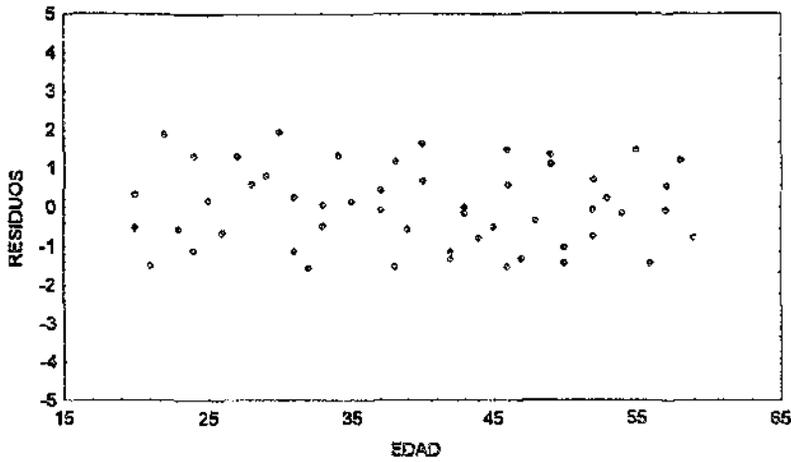


Figura 4.4: Residuos pseudoverosímiles

$$D = -2 \sum \int_y^\mu \frac{y-u}{V(u)} du \quad (4.39)$$

A partir de esta medida de discrepancia entre las observaciones y los valores esperados, Wedderburn (1974) define la función quasi-log-verosimilitud, para una sola observación de la siguiente manera:

Definición: Suponiendo que se tiene un conjunto de n observaciones independientes y_i ($i = 1, 2, \dots, n$) tales que $E(y_i) = \mu_i$ y $Var(y_i) = V(\mu_i)$ donde V es alguna función conocida y μ_i es alguna función conocida del vector de parámetros β . Entonces, para cada observación se define la función de quasiverosimilitud por la relación

$$\frac{\partial q(y_i, \mu_i)}{\partial \mu_i} = \frac{y_i - \mu_i}{V(\mu_i)} \quad (4.40)$$

es decir,

$$q(y_i, \mu_i) = \int \frac{y_i - u_i}{V(u_i)} du_i + f(y_i) \quad (4.41)$$

donde f es cualquier función.

Wedderburn (1974) demuestra que la función de quasiverosimilitud que propone tiene propiedades similares a las de la función de log-verosimilitud y que ambas funciones son iguales en el caso de la familia exponencial con un sólo parámetro. De hecho, a partir de la quasi-log-verosimilitud se obtiene una función de densidad que está dada por

$$f_q = \frac{\exp(q)}{\int \exp(q) dy} dy \quad (4.42)$$

Para relajar el supuesto de que la función varianza es conocida, Wedderburn incluye, en la expresión, una constante de proporcionalidad (σ^2) que es desconocida y la identifica como un parámetro de dispersión. De esta manera, la función de quasiverosimilitud está dada por

$$q(y_i, \mu_i) = \int_{\mu_i}^{\mu_i} \frac{y_i - u_i}{\sigma^2 V(u_i)} du_i \quad (4.43)$$

Para obtener los estimadores máximo quasiverosímiles se empieza por determinar las ecuaciones de quasiverosimilitud al derivar $q(y, \mu)$ con respecto a β e igualar a cero, esto es

$$\frac{\partial q(y, \mu)}{\partial \beta} = \frac{\partial}{\partial \beta} \int_y^{\mu} \frac{y - u}{\sigma^2 V(u)} du = D'V^{-1} \left(\frac{y - \mu}{\sigma^2} \right) = U(\beta) \quad (4.44)$$

donde:

$$D_{ir} = \frac{\partial \mu_i}{\partial \beta_r} \quad (4.45)$$

para $i = 1, 2, \dots, n$; $r = 1, 2, \dots, p$

Obteniendo así el equivalente a la función *score* en el caso de la función de verosimilitud por lo que, en este caso, se llama función de *quasi-score*.

Aplicando el método de Newton-Raphson para maximizar la función de quasiverosimilitud y empezando con un valor arbitrario $\hat{\beta}$ para el vector de parámetros β , se tiene

$$\widehat{\beta}_{i+1} = \widehat{\beta}_i + (\widehat{D}'\widehat{V}^{-1}\widehat{D})^{-1}\widehat{D}'\widehat{V}^{-1}(y - \mu) \quad (4.46)$$

Nótese que, al igual que en el caso de la verosimilitud, la estimación de β no depende de σ^2 y que si el modelo para la media es una función lineal de la matriz diseño, $D = X$ por lo que el segundo sumando de 4.46 es una estimación por mínimos cuadrados ponderados con los residuos como respuesta.

Por otro lado, la matriz de covarianza de $U(\beta)$ está dada por

$$-E \left[\frac{\partial U(\beta)}{\partial \beta} \right] = -E \left[\frac{\partial^2 q(y, \mu)}{\partial \beta^2} \right] = E \left[\frac{\partial q(y, \mu)}{\partial \beta} \right]^2 = E \left[D'V^{-1} \left(\frac{y - \mu}{\sigma^2} \right) \right]^2 \quad (4.47)$$

de aquí que

$$-E \left[\frac{\partial U(\beta)}{\partial \beta} \right] = \frac{D'V^{-1}D}{\sigma^2} \quad (4.48)$$

por lo que la matriz de covarianza asintótica de $\widehat{\beta}$ es

$$Cov(\widehat{\beta}) \simeq \sigma^2(D'V^{-1}D)^{-1} \quad (4.49)$$

El estimador insesgado quasiverosímil del parámetro de dispersión está dado por la expresión

$$\widehat{\sigma}^2 = \sum \frac{(y_i - \widehat{\mu}_i)^2}{(n - p)V(\widehat{\mu}_i)} \quad (4.50)$$

De acuerdo con Wedderburn, $\widehat{\beta}$ quasiverosímil tiene las mismas propiedades que el estimador máximo-verosímil, en particular, se tiene que es insesgado y asintóticamente normal, esto es

$$\widehat{\beta}_q \sim N(\beta, \sigma^2(D'V^{-1}D)^{-1}) \quad (4.51)$$

Ejemplo

El resultado de aplicar la expresión 4.32 a los datos de la presión sanguínea es

$$\hat{\beta} = \begin{bmatrix} 55.26233 \\ 0.60411 \end{bmatrix} \quad (4.52)$$

y los residuos se grafican en la figura 4.5:

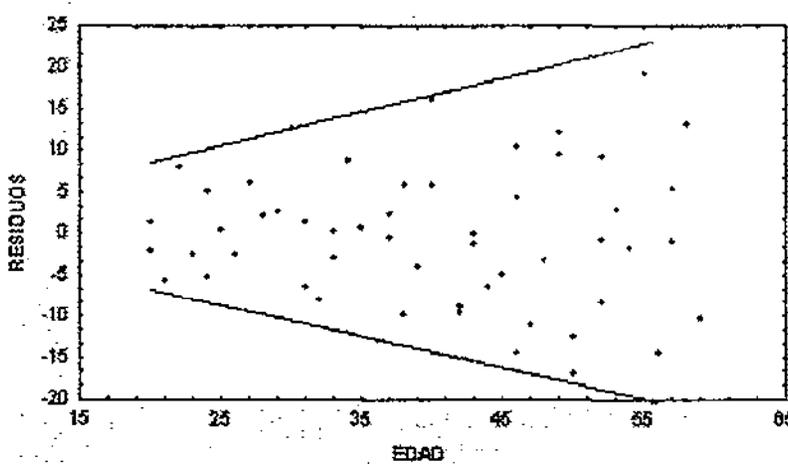


Figura 4.5: Residuos Quasiverosímiles

Nótese la dependencia entre la varianza de las observaciones y la edad. El valor del cuadrado medio del error es mayor que en los dos métodos anteriores $CME = 3.2147$

Comparaciones

La siguiente tabla resume los resultados de las estimaciones por los diferentes métodos para el ejemplo de la presión arterial sanguínea de un grupo de 54 mujeres.

.Tabla 4.1 Resultados de las estimaciones.

Estim	MCO	MCP ₄ ²	MCP ₈	MCP ₂	2E	MCG	MV	PV	QV
$\hat{\beta}_0$	56.16	56.09	56.14	56.01	56.16	56.11	55.04	55.76	55.26
$\hat{\beta}_1$	0.58	0.59	0.57	0.59	0.58	0.58	0.61	0.59	0.60
$\hat{\alpha}_0$					0.41	0.90	1.02	1.04	
$\hat{\alpha}_1$					0.06	0.03	0.07	0.07	
CME	66.35	0.96	0.95	1.00	22.16	7.26	1.03	1.03	3.21

La partición en ocho grupos es la que ha minimizado el cuadrado medio del error, sin embargo, además de la subjetividad que acompaña la decisión del número de grupos a formar, el modelo que se obtiene por mínimos cuadrados ponderados no refleja la estructura de la varianza.

De los modelos iterativos, máxima verosimilitud y pseudoverosimilitud son los que provocan un ECM menor y los modelos para la varianza son semejantes entre sí.

4.3 Aplicación de los métodos de estimación

Los diferentes métodos fueron utilizados para estimar los parámetros de los modelos correspondientes a los ejemplos de Pignatello y Grego ejemplificados en el capítulo 2.

Resortes automotrices

Para el conjunto de datos sobre la longitud de los resortes (Pignatello) se ha tomado como referencia el análisis exploratorio que considera O como factor de control y bajo el subtítulo de MODELO II (página 42). Ya se había observado, en el diagrama de Daniel, que en el modelo de la media los efectos principales O, B y C son significativos así como la interacción OC, mientras en el caso de la varianza resulta significativo el factor B.

Los estimadores para el modelo de la media de la forma

$$E(Y_{ijkl}) = \mu + \tau_i + \gamma_j + \delta_k + \tau\delta_{ij} \quad (4.53)$$

donde μ es el efecto medio general y τ, γ y δ son los efectos de los factores, en ese orden, O, B, C con la interacción $\tau\delta$, obtenidos a través de los métodos aquí documentados son por demás similares entre sí y con la estimación por mínimos cuadrados ordinarios aplicados en el capítulo 2. (Tabla.4.2).

Tabla 4.2 Estimadores para el modelo de la media de los resortes automotrices.

Término	2E	MCG	MV	PV	QV	MCO
Intercepto (μ)	7.636	7.636	7.636	7.636	7.636	7.636
O[-1-1] ($\hat{\tau}$)	0.130	0.130	0.138	0.130	0.130	0.128
B[-1-1] ($\hat{\gamma}$)	-0.111	-0.111	-0.110	-0.111	-0.111	-0.111
C[-1-1] ($\hat{\delta}$)	0.088	0.088	0.086	0.088	0.089	0.088
OC[-1-1] ($\hat{\tau\delta}$)	0.083	0.083	0.084	0.083	0.082	0.083

Mientras los resultados en un modelo para la varianza de la forma

$$Var(Y_{ijkl}) = \exp(\psi + \gamma_i) \quad (4.54)$$

donde ψ es el intercepto y γ representa el efecto del factor B en la dispersión de las observaciones, son

Tabla 4.3 Estimadores para la función varianza. Resortes automotrices

Término	2E	MCG	MV	PV	MCO
Intercepto (ψ)	-5.175	-5.175	-4.028	-4.025	-4.93
B[-1-1] ($\hat{\gamma}$)	-0.052	-0.052	0.203	0.161	-0.945

Nótese que los valores de los estimadores para la función varianza se alejan más entre sí con respecto a lo que sucede en el caso de la media.

El *CME*, en cada caso es:

Tabla 4.4 Cuadrado medio del error

2E	MCG	MV	PV	QV	MCO
0.020	0.020	1.116	1.116	3.605	0.0103

Si se toma el CME como criterio para seleccionar un método de estimación, resulta que una regresión ordinaria es la mejor opción. De aquí que puede concluirse que, en efecto, no se tiene un problema de heteroscedasticidad en el experimento de Pignatello.

La figura 4.6 contiene la gráfica de los residuos estandarizados del ajuste máximo verosímil

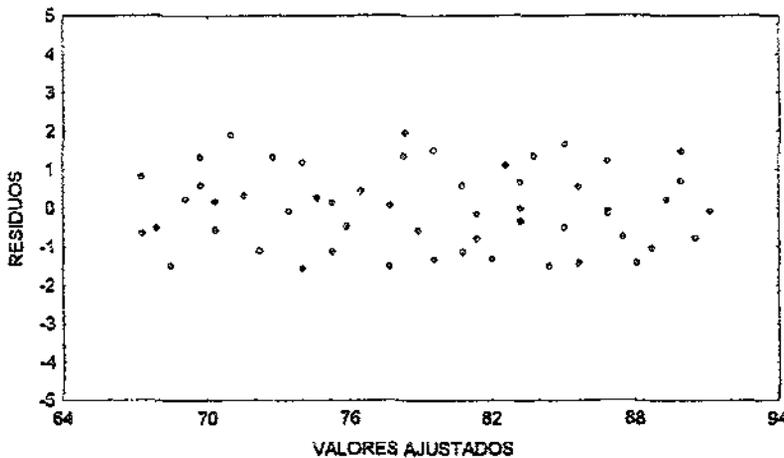


Figura 4.6: Residuos estandarizados vs. ajustados

y en ella se observa un comportamiento aleatorio de los residuos con al menos el 95% de las observaciones dentro de la franja horizontal de -2 a 2 de tal manera que se puede pensar que el modelo es adecuado.

Circuitos integrados

En el caso del ejemplo de Grego y, de acuerdo con el análisis exploratorio del capítulo 2, el modelo que se ajusta para la media tiene la forma

$$E(Y_{ijk}) = \mu + \tau_i + \gamma_j + \delta_k \quad (4.55)$$

donde μ es el efecto medio general y τ, γ, δ son los efectos de los factores A, C y D respectivamente.

Los resultados de las estimaciones son los siguientes:

Tabla 4.5 Estimadores de los parámetros del modelo de la media. Fuerza adherente

Término	2E	MCG	MV	PV	QV	MCO
Intercepto (μ)	81.350	81.350	81.550	81.350	81.501	81.350
A[-1-1] (τ)	-0.975	-0.975	-1.230	-0.975	-1.172	-0.975
C[-1-1] (γ)	-2.725	-2.725	-2.800	-2.725	-2.806	-2.725
D[-1-1] (δ)	-4.360	-4.360	-4.034	-4.360	-4.037	-4.360

en tanto un modelo para la varianza de la forma

$$Var(Y_{ij}) = \exp(\psi + \tau_i + \delta_j) \quad (4.56)$$

tiene los siguientes estimadores

Tabla 4.6 Estimadores de los parámetros de la función varianza

Término	2E	MCG	MV	PV	MCO
Intercepto (ψ)	0.151	0.151	0.976	1.067	0.151
A[-1-1] (τ)	-1.040	-1.040	-1.059	-0.948	-1.040
D[-1-1] (δ)	0.515	0.515	0.730	0.639	0.515

De nueva cuenta se tiene que hay mayor discrepancia entre las estimaciones para la varianza que para la media. Los cuadrados medios del error se reportan en la tabla 4.7

Tabla 4.7 Cuadrado medio del error

2E	MCG	MV	PV	QV	MCO
5.961	6.167	1.111	1.111	2.591	5.961

De esta manera, se tiene que las estimaciones por máxima y por pseudo-verosimilitud son las que minimizan tal estadística.

Dado que al ajustar un modelo para la media que de manera implícita considera el problema de heteroscedasticidad y que como consecuencia se ha modificado el CME y por tanto el error estándar de los estimadores, cabe hacerse la pregunta si tales estimadores siguen siendo significativos.

Suponiendo normalidad asintótica en los estimadores máximo verosímiles y, considerando CME=1.111 se ha recalculado el error estándar y se ha usado la estadística *t-student* para probar la hipótesis nula de que los estimadores son iguales a cero. (Tabla 4.8).

Tabla 4.8 Significancia de Estimadores

	Término	Estimador	error estándar	valor p
Media				
	Intercepto	81.55	0.1667	<.0001
	A[-1-1]	-1.23	0.1667	<.0001
	C[-1-1]	-2.8	0.1667	<.0001
	D[-1-1]	-4.034	0.1667	<.0001
Varianza				
	Intercepto	0.976	0.1667	<.0001
	A[-1-1]	-1.059	0.1667	<.0001
	D[-1-1]	-0.7259	0.1667	<.0001

La gráfica de los residuos estandarizados contra los valores ajustados por el método de máxima verosimilitud se encuentra en la figura 4.7 y en ella no se observa alguna evidencia para descartar el modelo.

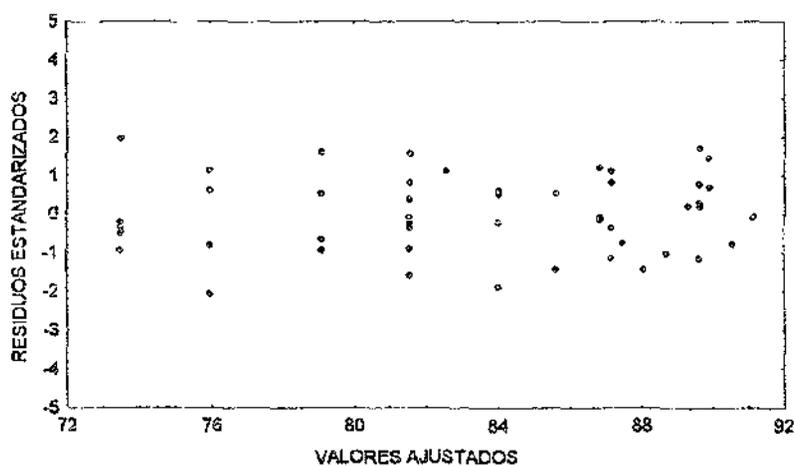


Figura 4.7: Residuos estandarizados

Capítulo 5

Métodos Montecarlo

5.1 Introducción

Para estudiar, empíricamente, el comportamiento estadístico de algún fenómeno aleatorio se requiere de la observación de un conjunto de datos que conforman una muestra de tal situación. En la actualidad, gracias al desarrollo de importantes teorías, en ocasiones es posible desarrollar métodos analíticos para deducir o, en su caso, inferir características o propiedades relevantes de cierto proceso. Sin embargo, es común encontrarse con que el aspecto matemático implicado en un estudio teórico es bastante complicado por lo se requiere de un análisis empírico.

La observación directa del desarrollo de todo un proceso o de la variable respuesta de un experimento puede ser muy costosa en términos de tiempo y recursos humanos y financieros. La alternativa, en estos casos, es la simulación de las circunstancias en las que se desarrolla el proceso o experimento y obtener la muestra por el método de Monte Carlo.

El punto de partida es un modelo matemático que se "activa" utilizando números aleatorios que siguen cierta distribución.

La idea es modificar el modelo tanto como sea necesario, es decir, los experimentos se llevan a cabo en el modelo en lugar de hacerlo en el proceso real. Al introducir una distribución de probabilidades y la generación de números aleatorios se obtienen muestras del proceso.

Es así como, con la ayuda de una computadora, se pueden construir grandes cantidades de realizaciones numéricas del proceso de manera rápida y económica. Una colección de estas proporcionará una imagen de las características del proceso en su totalidad.

Con la finalidad de examinar y comparar las propiedades de los estimadores que se obtienen a través de los diferentes métodos descritos en el capítulo anterior y así tener más elementos para tomar una decisión sobre el método a elegir, se desarrollaron las simulaciones de dos experimentos, que tienen en común las siguientes características:

La forma general de las observaciones es

$$Y_i = \beta_0 + \beta_1 x_i + \sqrt{\exp\left(\frac{\alpha_0 + \alpha_1 x_i}{2}\right)} e_i \quad (5.1)$$

con

$$e_i \sim N(0, 1) \quad (5.2)$$

de manera que

$$E(Y_i) = \beta_0 + \beta_1 x_i = X\beta \quad (5.3)$$

y

$$\text{Var}(Y_i) = \exp(\alpha_0 + \alpha_1 x_i) = \exp(X\alpha) \quad (5.4)$$

y que en las siguientes secciones se describen.

5.2 Experimento I

En el primer experimento se simularon muestras aleatorias con distribución normal que fueron sometidas a tres distintas situaciones de varianza.

Se parte de la simulación de un experimento con 7 tratamientos y $n = 3, 5$ y 8 repeticiones para cada caso, es decir, se hicieron nueve grupos de simulaciones¹ por cada método de estimación.

Los diferentes tratamientos están en función de una sola variable regresora con 7 puntos igualmente espaciados en el intervalo $[2.3, 3.68]^2$, por lo que la matriz diseño (X) consta de $7 * n$ renglones y 2 columnas:

$$X_{7n \times 2} = \begin{bmatrix} 1 & 2.3 \\ 1 & 2.53 \\ 1 & 2.76 \\ 1 & 2.99 \\ 1 & 3.22 \\ 1 & 3.45 \\ 1 & 3.68 \\ & \dots \\ 1 & 3.45 \\ 1 & 3.68 \end{bmatrix}$$

5.2.1 El valor esperado.

La media de las observaciones sigue el modelo

$$E(Y_i) = \beta_0 + \beta_1 x_i = X\beta$$

¹Cada grupo consta de mil simulaciones. A partir de un modelo con $\alpha_1=0$ (homoscedasticidad), se ha verificado que el rechazo de las Hipótesis que hacen referencia al valor de los parámetros ocurre con probabilidad menor a 0.1

²Valores arbitrarios

con

$$\beta = \begin{bmatrix} 1.5 \\ 0.6 \end{bmatrix}$$

los valores para los parámetros fueron asignados de manera arbitraria debido a que el centro de atención está en la dispersión de los datos.

5.2.2 La varianza

Con respecto a la varianza se tiene que las tres diferentes condiciones a tratar son:

1. La hipótesis de homoscedasticidad no es rechazada
2. La decisión de rechazar homoscedasticidad es dudosa
3. Se rechaza de manera contundente la hipótesis de igualdad de varianzas.

Una de las pruebas más sencillas de homogeneidad de varianzas es la de Hartley:

El cociente de las varianzas muestrales mínima y máxima es la estadística de prueba. En términos de la función varianza (5.4), la estadística de prueba H toma la forma:

$$H = \frac{\exp(\alpha_0 + \alpha_1 x_{\max})}{\exp(\alpha_0 + \alpha_1 x_{\min})}$$

de aquí que

$$H = \exp(\alpha_1 x_{\max} - \alpha_1 x_{\min})$$

$$\iff \ln(H) = \alpha_1 (x_{\max} - x_{\min})$$

$$\iff \alpha_1 = \frac{\ln(H)}{(x_{\max} - x_{\min})}$$

esto es, para los modelos de la forma 5.4 y para la prueba de Hartley, α_1 es el parámetro más importante dentro de la función varianza por lo que se ha considerado α_0 con un valor fijo, arbitrario, igual a $\frac{1}{2}$ y α_1 se calcula en cada caso:

Homoscedasticidad

En esta ocasión, los valores de α son "tan pequeños" que no se puede considerar una diferencia significativa entre las varianzas de las distintas poblaciones.

Dado que, para un experimento con 7 tratamientos y 8 repeticiones en cada uno ellos, se tiene que, al 95% de confianza, la hipótesis nula de igualdad de varianzas no se rechaza cuando $H < 11.8$, se tiene

$$\alpha_1 < \frac{\ln(11.8)}{(3.68 - 2.3)} = 1.7885 \quad (5.5)$$

por lo que se propuso $\alpha_1 = 1.2$ de manera que se esperarían, respectivamente, los siguientes valores máximo y mínimo para la varianza de las observaciones:

$$\sigma_{\max}^2 = \exp(0.5 + 1.2x_{\max}) = 136.46$$

y

$$\sigma_{\min}^2 = \exp(0.5 + 1.2x_{\min}) = 26.05$$

$$\implies H = \frac{136.46}{26.05} = 5.24$$

\therefore no se rechaza la hipótesis de igualdad de varianzas.

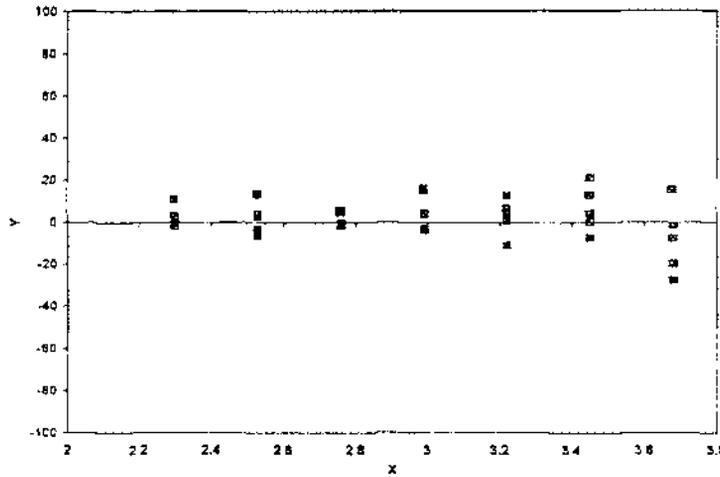


Figura 5.1: Homoscedasticidad

Ejemplo

La figura 5.1 muestra uno de los conjuntos de observaciones generadas por el procedimiento antes descrito para $n = 5$.

Las varianzas muestrales, para los datos del ejemplo, son:

Tabla 5.1. Varianzas muestrales

x_i	s_i^2
2.30	36.72
2.53	58.06
2.76	18.40
2.99	49.56
3.22	69.72
3.45	83.91
3.68	120.34

por lo que $H = \frac{120.34}{18.40} = 6.54 < 11.8$ es decir, no se rechaza la igualdad de varianzas.

Cabe mencionar que este resultado es consistente con el que se obtiene con otras pruebas de homogeneidad de varianzas:

Tabla 5.2. Pruebas de homoscedasticidad		
Prueba	Valor de la estadística	Valor p
Levene	1.0246	0.4207
Bartlett	1.0942	0.3629

Homoscedasticidad dudosa

En este caso α toma valores que delatan una discreta heteroscedasticidad.

Siguiendo el procedimiento 5.5 se propuso $\alpha_1 = 1.79$ de manera que se esperaría tener un valor $H = 11.82$.

Si se aplica esta modificación al mismo conjunto de datos del ejemplo anterior, se tiene la figura 5.2

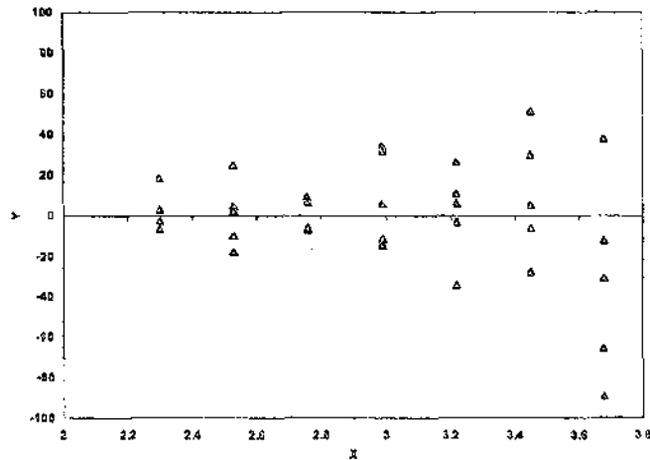


Figura 5.2: Homoscedasticidad dudosa

Las varianzas muestrales ahora son:

Tabla 5.3 varianzas muestrales

x_i	s_i^2
2.30	142.32
2.53	296.18
2.76	81.96
2.99	288.93
3.22	465.60
3.45	736.31
3.68	922.27

esto es, $H = \frac{922.27}{81.96} = 11.25$. El valor de la estadística se acerca al punto crítico pero aún así, para este conjunto de observaciones, la hipótesis de igualdad de varianzas no se rechaza con un nivel de significancia del 5%.

Los resultados de otras pruebas aplicadas con respecto a la homoscedasticidad son:

Prueba	Valor de la estadística	Valor p
Levene	1.927	0.0951
Bartlett	2.182	0.0416

Es decir, Levene no rechaza la igualdad de varianzas al 5% de significancia pero sí lo hace al 10%, mientras Bartlett descarta la homoscedasticidad con una significancia del 5%.

Heteroscedasticidad

Aquí se trata la situación en que *la heterogeneidad de varianzas es evidente*.

Con $\alpha_1 = 2.55$, se esperaría tener $H = 33.75$, de tal manera que el problema de heteroscedasticidad sería indudablemente importante.

Continuando con el ejemplo anterior, el conjunto de observaciones se vería como lo muestra la figura 5.3

Las varianzas muestrales son:

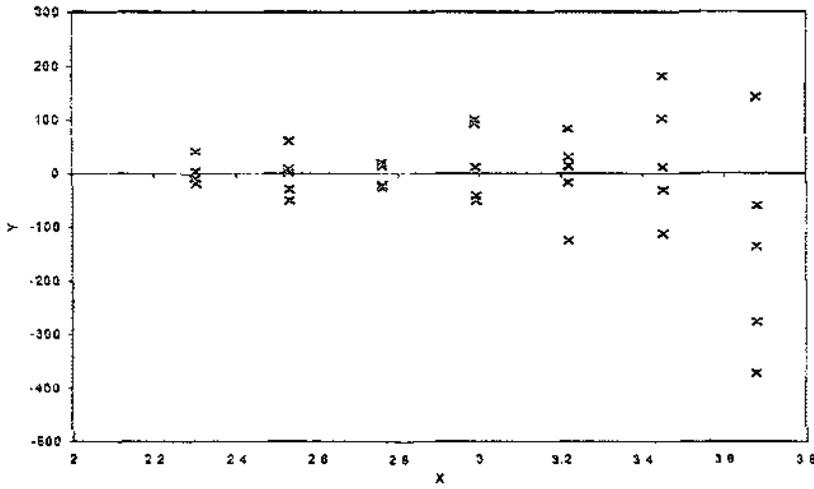


Figura 5.3: Heteroscedasticidad

Tabla 5.4. varianzas muestrales

x_i	s_i^2
2.30	818.53
2.53	2411.79
2.76	560.74
2.99	2803.7
3.22	5380.22
3.45	12069.22
3.68	12687.77

por lo que $H = \frac{12687.77}{560.74} = 22.63$ rechazando la hipótesis de homoscedasticidad.

Las pruebas de Levene y Bartlett llevan a la misma conclusión.

Tabla 5.5. Pruebas de homoscedasticidad

Prueba	Valor de la estadística	Valor p
Levene	3.3044	0.0082
Bartlett	4.1578	0.0003

Nótese la diferencia de escala entre las tres gráficas anteriores. Para resaltar las distintas varianzas, la figura 5.4 contiene de manera simultánea a los tres grupos

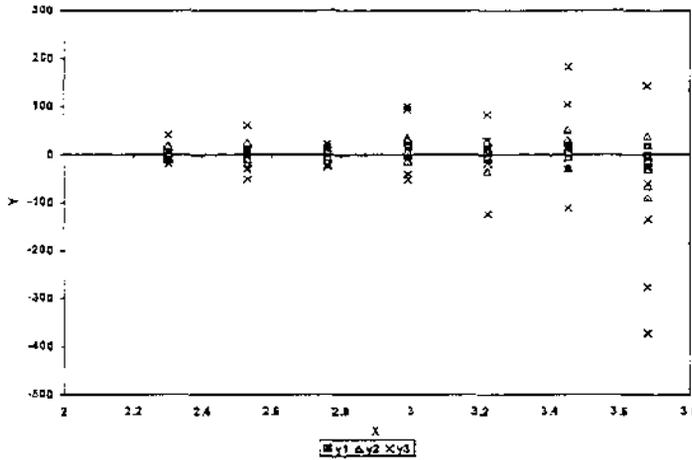


Figura 5.4: Tres grupos de datos con diferente varianza

5.3 Experimento II

En este caso se simuló un diseño de tres niveles que se utiliza para ajustar superficies de respuesta y que lleva el nombre de sus autores: Box-Behnken.

El diseño consta de tres variables. Dos de ellas tienen efecto sobre el valor esperado, incluyendo una interacción, y la tercera sobre la varianza, esto es

$$E(Y) = 6.5 + 0.5x_1 + 0.95x_2 - 0.8x_1x_2 \quad (5.6)$$

$$Var(Y) = \exp(2.4 + 0.8x_3) \quad (5.7)$$

A diferencia del experimento I, en este caso no se varían las condiciones de homoscedasticidad pero sí se toman en cuenta los tamaños de muestra, a saber, $n = 1, 3$ y 5 .

5.4 Algoritmos

Una vez generados los conjuntos de observaciones, se aplicaron los diferentes métodos siguiendo los algoritmos que a continuación se detallan.

5.4.1 Regresión en dos etapas.

1. Estimar β por mínimos cuadrados ordinarios

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

2. Calcular el vector de residuos:

$$r = Y - X\hat{\beta}$$

3. Estimar α por mínimos cuadrados ordinarios tomando como respuesta el logaritmo de los residuos:

$$\hat{\alpha} = (X'X)^{-1}(X'\ln(r)).$$

4. Calcular $\hat{\sigma}^2$

$$\hat{\sigma}^2 = CME = \sum \frac{(y_i - x_i\hat{\beta})^2}{(n-p)\exp(x_i\hat{\alpha})} = \frac{SCE}{n-p}$$

5 Fin.

5.4.2 Mínimos cuadrados generalizados

1. Estimar β por mínimos cuadrados ordinarios

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

2. Calcular el vector de residuos:

$$r = Y - X\hat{\beta}$$

3. Estimar α por mínimos cuadrados ordinarios tomando como respuesta el logaritmo de los residuos:

$$\hat{\alpha} = (X'X)^{-1}(X'\ln(r)).$$

4. Construir la matriz de varianza estimada. Es una matriz diagonal cuyos componentes son

$$\hat{v}_{ii} = \exp(\hat{\alpha}_1 + \hat{\alpha}_2 x_i)$$

5. Estimar β por mínimos cuadrados ponderados:

$$\hat{\beta}_{pon} = (X'V^{-1}X)^{-1}(X'V^{-1}Y)$$

6. Calcular los residuos de la estimación ponderada:

$$r_{pon} = Y - X\hat{\beta}_{pon}$$

7. Actualizar el valor de $\hat{\alpha}$ usando los residuos del punto anterior

$$\hat{\alpha} = (X'X)^{-1}(X'\ln(r_{pon})).$$

8. Calcular la diferencia entre dos estimaciones sucesivas:

$$i \quad \left| \hat{\beta} - \hat{\beta}_{pon} \right| \leq 0.00001 \quad \text{y} \quad \left| \hat{\alpha} - \hat{\alpha}_{pon} \right| \leq 0.00001 \quad ?$$

Si, entonces calcular $\hat{\sigma}^2 = CME = \sum \frac{(y_i - x_i \hat{\beta})^2}{(n-p) \exp(x_i \hat{\alpha})} = \frac{SCE}{n-p}$
detener el proceso.

No, entonces

$$\hat{\beta} = \hat{\beta}_{pon} \quad \text{y} \quad \hat{\alpha} = \hat{\alpha}_{pon}$$

9. Regresar al paso 5 para obtener nuevos estimadores

5.4.3 Máxima verosimilitud

1. Obtener un estimador inicial para β por mínimos cuadrados ordinarios

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

2. Calcular el vector de residuos:

$$r = Y - X\hat{\beta}$$

3. Obtener un estimador inicial para α por mínimos cuadrados ordinarios tomando como respuesta el logaritmo de los residuos:

$$\hat{\alpha} = (X'X)^{-1}(X' \ln(r)).$$

4. Encontrar el valor de $\hat{\beta}_{pv}$ y de $\hat{\alpha}_{pv}$ que simultáneamente maximizan la función de log-verosimilitud esto es, maximizar:

$$\ln(L) = \frac{-1}{2} \left[n\alpha_0 + \alpha_1 \sum x_i + \sum \frac{(Y_i - \beta_0 - \beta_1 x_i)^2}{\sigma^2 \exp(\alpha_0 + \alpha_1 x_i)} \right]$$

5. Calcular

$$\hat{\sigma}^2 = CME = \sum \frac{(y_i - x_i \hat{\beta})^2}{(n - p) \exp(x_i \hat{\alpha})} = \frac{SCE}{n - p}$$

6. Fin

5.4.4 Pseudoverosimilitud

1. Obtener un estimador inicial para β por mínimos cuadrados ordinarios

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

2. Calcular el vector de residuos:

$$r = Y - X\hat{\beta}$$

3. Obtener un estimador inicial para α por mínimos cuadrados ordinarios tomando como respuesta el logaritmo de los residuos:

$$\hat{\alpha} = (X'X)^{-1}(X' \ln(r)).$$

4. Encontrar el valor de $\hat{\beta}_{pv}$ que maximiza la función de log-verosimilitud normal suponiendo α conocido, esto es, maximizar:

$$\ln(L) = \frac{-1}{2} \left[n\alpha_0 + \alpha_1 \sum x_i + \sum \frac{(Y_i - \beta_0 - \beta_1 x_i)^2}{\sigma^2 \exp(\alpha_0 + \alpha_1 x_i)} \right] \quad (5.8)$$

5. Encontrar el valor de $\hat{\alpha}_{pv}$ que maximiza la función de log-verosimilitud normal ahora suponiendo β conocido.
6. ¿ $|\hat{\beta} - \hat{\beta}_{pv}| \leq 0.00001$ y $|\hat{\alpha} - \hat{\alpha}_{pv}| \leq 0.00001$?

Si, entonces calcular

$$\hat{\sigma}^2 = CME = \sum \frac{(y_i - x_i \hat{\beta})^2}{(n-p) \exp(x_i \hat{\alpha})} = \frac{SCE}{n-p} \text{ y detener el proceso.}$$

No, entonces regresar al paso 4 para obtener nuevos estimadores

5.4.5 Quasiverosimilitud

1. Estimar β por mínimos cuadrados ordinarios

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

2. Calcular el vector de residuos:

$$r = Y - X\hat{\beta}$$

3. Estimar α por mínimos cuadrados ordinarios tomando como respuesta el logaritmo de los residuos:

$$\hat{\alpha} = (X'X)^{-1}(X' \ln(r)).$$

4. Construir la matriz de varianza estimada. Es una matriz diagonal cuyos componentes son

$$\hat{v}_{ii} = \exp(\hat{\alpha}_1 + \hat{\alpha}_2 x_i)$$

5. Estimar β por mínimos cuadrados ponderados, tomando los residuos como respuesta:

$$\widehat{\beta}_{pom} = (X'V^{-1}X)^{-1}(X'V^{-1}r)$$

6. Calcular $\widehat{\beta}_q = \widehat{\beta} + \widehat{\beta}_{pom}$

7. ¿ $|\widehat{\beta} - \widehat{\beta}_q| \leq 0.00001$?

Si, entonces calcular

$$\widehat{\sigma}^2 = CME = \sum \frac{(y_i - x_i\widehat{\beta})^2}{(n-p) \exp(x_i\widehat{\alpha})} = \frac{SCE}{n-p}$$

detener el proceso

8. No, entonces,

$$\widehat{\beta} = \widehat{\beta}_q \quad \text{regresar al paso 2}$$

Capítulo 6

Presentación de resultados

6.1 Introducción

La comparación de los métodos de estimación se establece a partir de las propiedades distribucionales de los estimadores obtenidos por los diferentes métodos. Es por esto que en este apartado se presentan el Sesgo, las desviaciones estándar (Error Estándar) y el Error Cuadrático Medio (ECM) de los valores logrados en cada caso.

La presentación se hace por medio de una serie de gráficas que pretenden ilustrar y resaltar algunas características importantes. Las tablas en las que se puede observar de manera directa el valor de un estimador específico, ante una situación particular y para un método en especial se encuentran en el apéndice B.

6.2 Simulación I

El modelo que se usó para la simulación tiene la forma:

$$Y_i = 1.5 + 0.6x_i + \exp\left(\frac{0.5 + 1.2x_i}{2}\right)e_i \quad e_i \sim N(0, 1)$$

donde: $\beta_0 = 1.5$; $\beta_1 = 0.6$; $\alpha_0 = 0.5$ y $\alpha_1 = 1.2, 1.79$ y 2.55

Esto es, el valor de α_1 va cambiando para modificar las condiciones de homoscedasticidad. En cada una de ellas se tomaron en cuenta situaciones

en las que se tienen 3, 5 y 8 repeticiones por punto diseño, de esta manera por cada método se realizaron 9 grupos de simulaciones.

Después de mil simulaciones se armaron vectores de igual tamaño para cada estimador y por cada uno de los métodos estudiados. Se observó que todas las distribuciones obtenidas son simétricas y, en su mayoría, no rechazan la hipótesis de normalidad.

En las figuras 6.1 a 6.4 se representa el sesgo de cada estimador conforme se alteraron las condiciones de heteroscedasticidad y para los diferentes tamaños de muestra. Cada figura está formada por tres gráficas (los tres tipos de experimentos con diferente número de repeticiones). En el eje horizontal se representan los valores de α_1 , es decir, las distintas circunstancias de homoscedasticidad. En todas ellas es claro que conforme α_1 aumenta, los estimadores tienden a alejarse del valor del parámetro, es decir, muestran mayor sesgo pero conforme se tiene una muestra más grande éste disminuye, es decir, ante un problema de heteroscedasticidad, lo recomendable es contar con muestras grandes (nada nuevo).

Puede observarse también que el estimador máximo verosímil es el más sesgado cuando $n = 3$ pero el menos cuando $n = 8$.

Resulta interesante resaltar que cualquiera de los métodos de estimación para α_1 , subestima el valor del parámetro y que esta situación se acentúa conforme la heteroscedasticidad es más evidente (figura 6.4).

Las figuras 6.5 a 6.8 representan el error estándar de cada estimador ante las diferentes situaciones de homoscedasticidad y para los tamaños de muestra 3,5 y 8. Tal como era de esperarse, se observa que, conforme aumenta el número de repeticiones, el error estándar de los estimadores es menor y resulta interesante destacar que en el caso de β_0 y β_1 se forman dos grupos: los métodos que generan estimadores más dispersos son los de regresión en dos etapas y pseudoverosimilitud mientras los estimadores con menor error estándar son los que se obtienen por MCG, Máxima verosimilitud y Quasiverosimilitud.

En el caso de α_0 y α_1 también se distinguen dos grupos: Máxima verosimilitud y pseudoverosimilitud conducen a estimadores con menor error estándar.

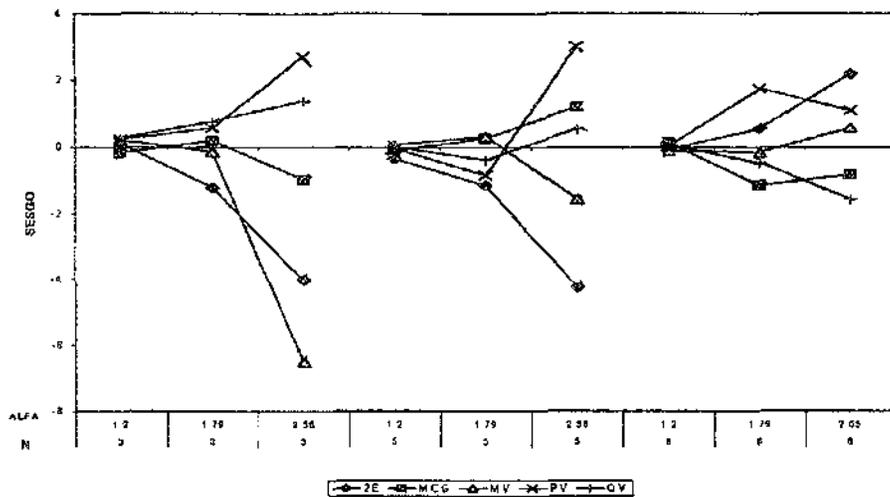


Figura 6.1: Sesgo de B0

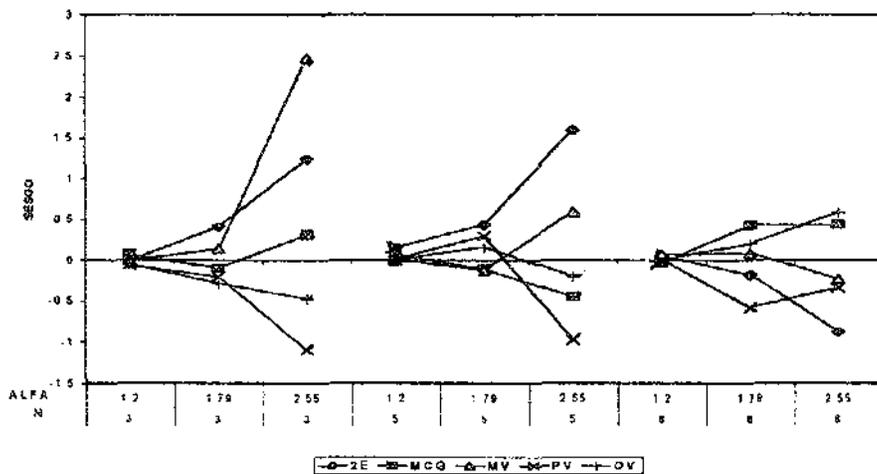


Figura 6.2: Sesgo de B1

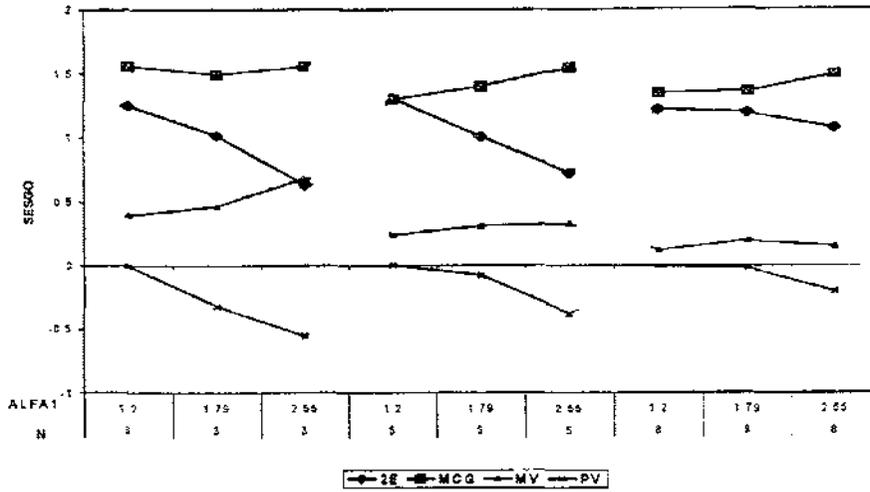


Figura 6.3: Sesgo de Alfa0

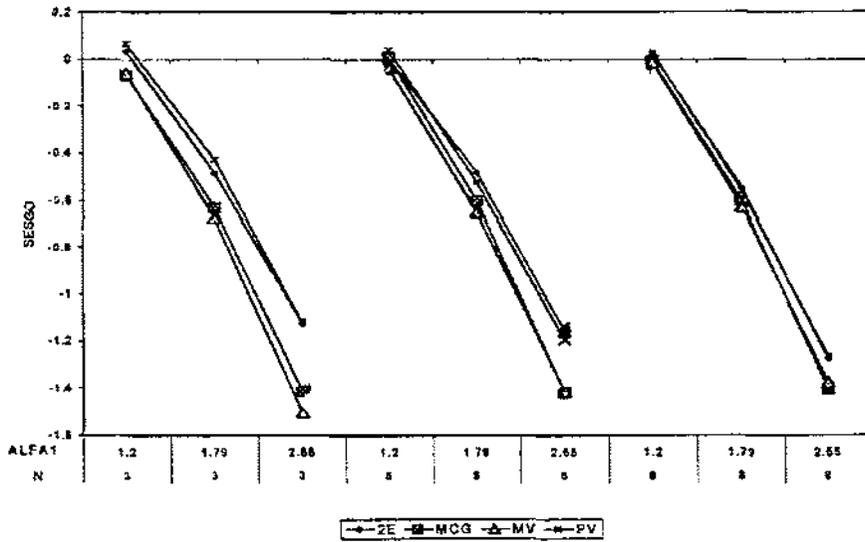


Figura 6.4: Sesgo de Alfa1

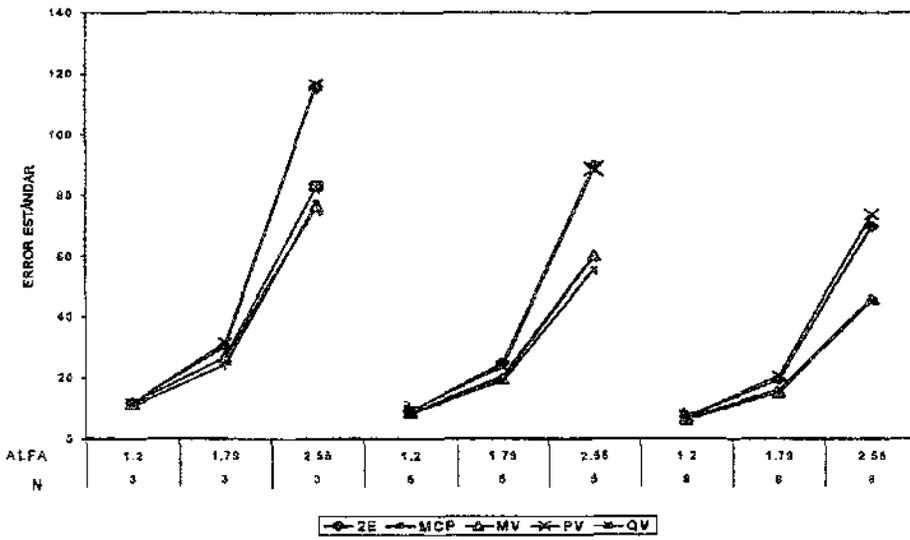


Figura 6.5: Error estándar de B0

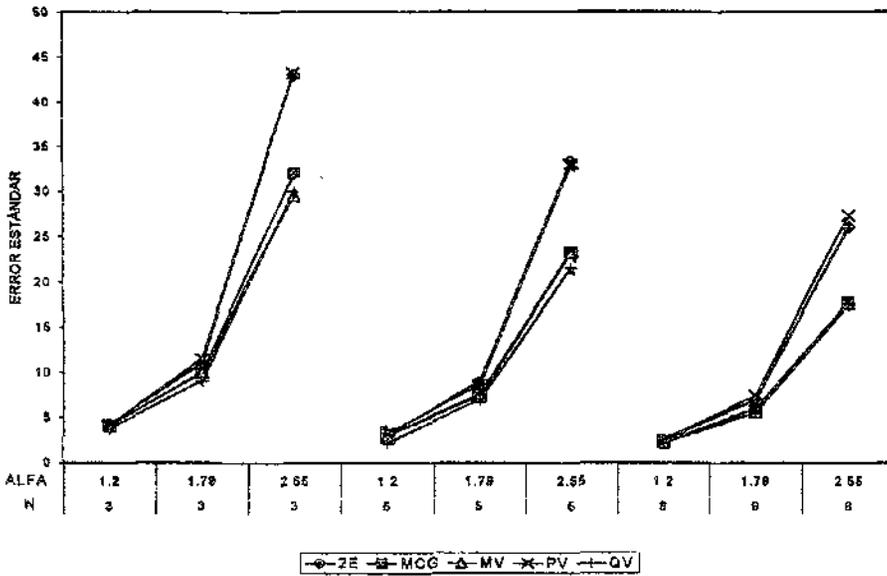


Figura 6.6: Error estándar de B1

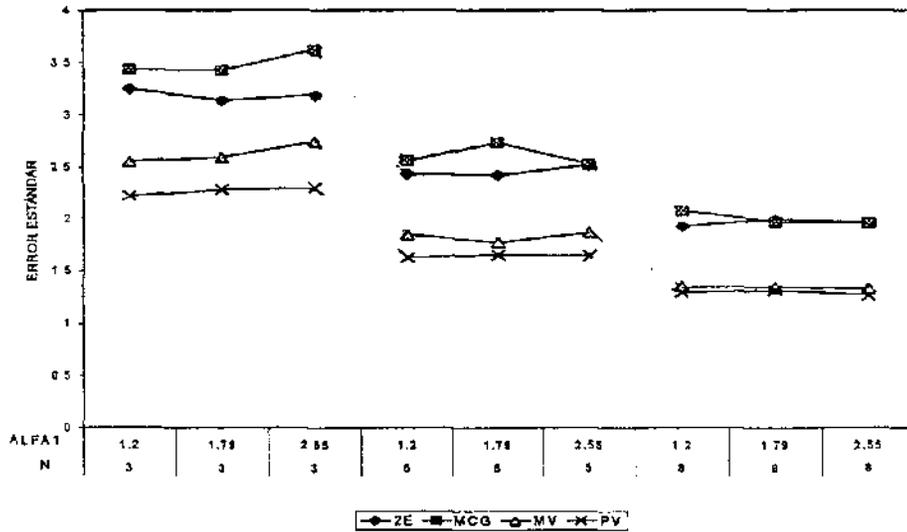


Figura 6.7: Error estándar de Alfa0

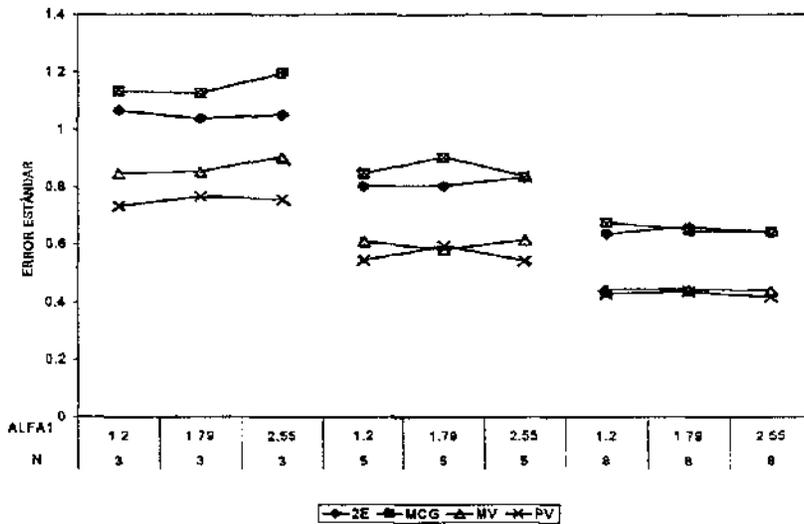


Figura 6.8: Error estándar de Alfa1

Las gráficas del ECM (figuras 6.9 a 6.12) reflejan la situación observada en el sesgo y el error estándar, razón por la que, en el caso de β_0 y β_1 , la estimación obtenida por Pseudoverosimilitud y Regresión en dos etapas genera un ECM más grande mientras que para α_0 Regresión en dos etapas y mínimos cuadrados generalizados se encuentran en el mismo caso. Para α_1 el ECM de la estimación por MCG es el más grande y por Pseudoverosimilitud es el menor aunque conforme el tamaño de la muestra aumenta, los valores generados por los cuatro métodos tienden a acercarse.

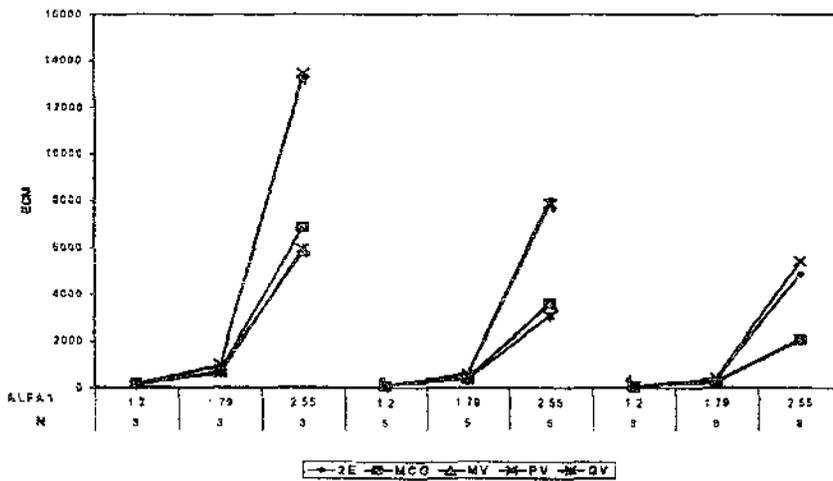


Figura 6.9: ECM B0

6.3 Experimento II

El diseño Box-Benhken usado en este experimento es

$$E(Y_i) = 6.5 + 0.5x_1 + 0.95x_2 - 0.8x_1x_2 \tag{6.1}$$

$$V(Y_i) = \exp(2.4 + 0.8x_3) \tag{6.2}$$

En este caso, la condición experimental está únicamente en el número de repeticiones por punto diseño (1,3 y 5). El valor de la varianza no se modifica a lo largo de las simulaciones.

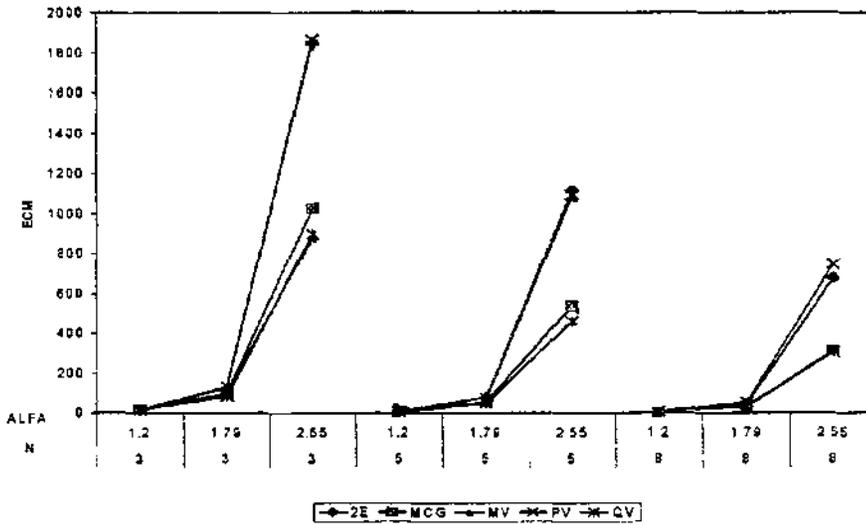


Figura 6.10: ECM B1

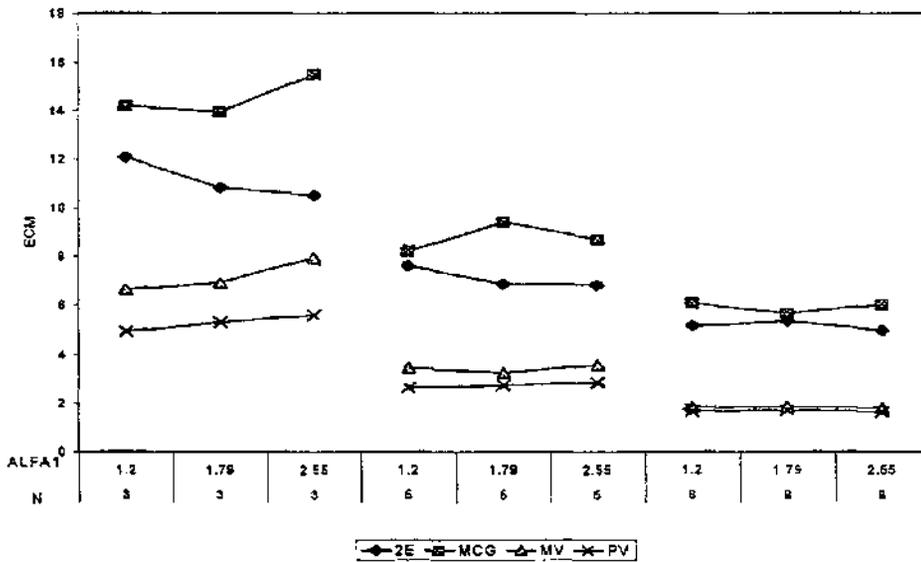


Figura 6.11: ECM Alfa0

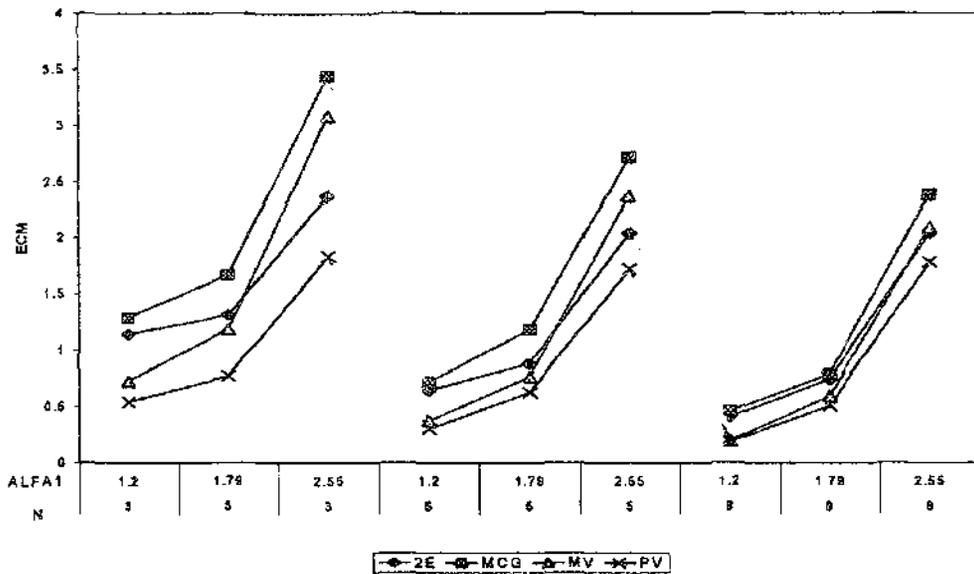


Figura 6.12: ECM Alfa1

En las figuras 6.13 y 6.14 se muestran, respectivamente, el sesgo y el error estándar de los estimadores del modelo 6.1. En ellas se observa que ambas estadísticas de los estimadores de β (parámetro de la media) son por demás similares entre sí (estimadores insesgados con menos varianza conforme aumenta el tamaño de la muestra).

En el caso de la estimación de α_0 , los métodos de mínimos cuadrados generalizados y regresión en dos etapas proporcionan estimadores con mayor sesgo y más dispersos que los que se obtienen por máxima y pseudo verosimilitud. Cabe notar que el error estándar del vector α es menor que correspondiente a β .

En el caso de α_1 se tiene que su estimador tiende a ser insesgado y el error estándar disminuye cuando la muestra es más grande. Si se usa máxima verosimilitud cuando no se tienen repeticiones, el estimador tiene mayor varianza y subestima el valor del parámetro, aunque con cinco repeticiones por punto diseño, el estimador tiende a ser insesgado y con menor varianza.

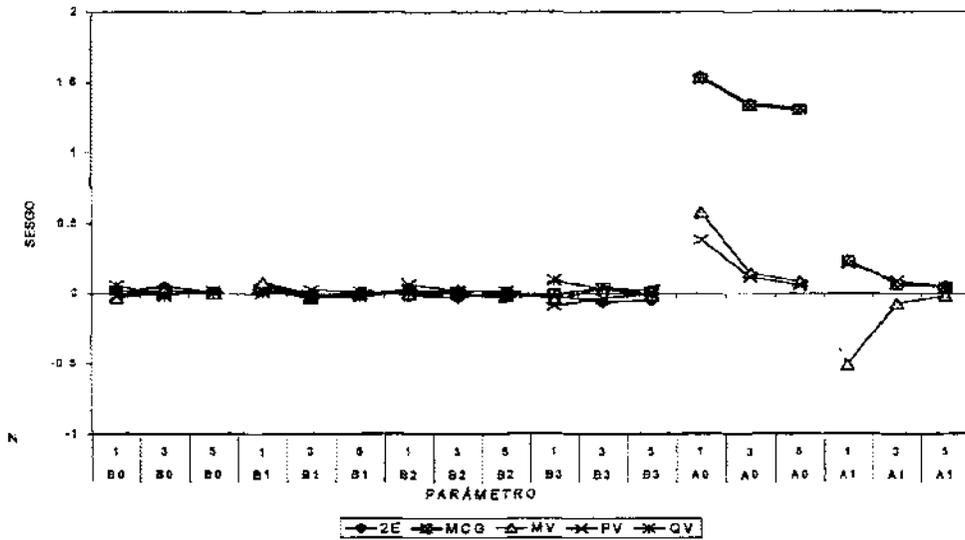


Figura 6.13: Sesgo de los estimadores

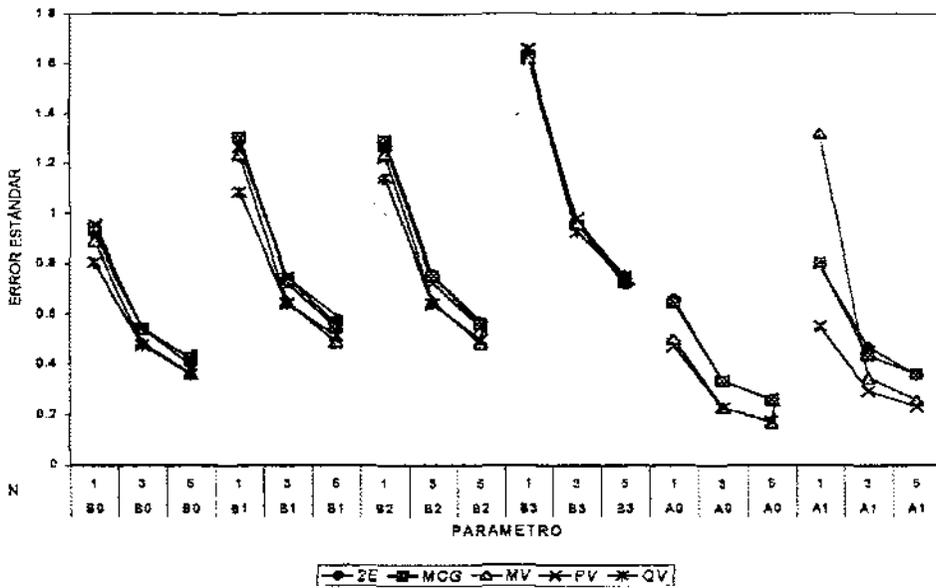


Figura 6.14: Error estándar de los estimadores

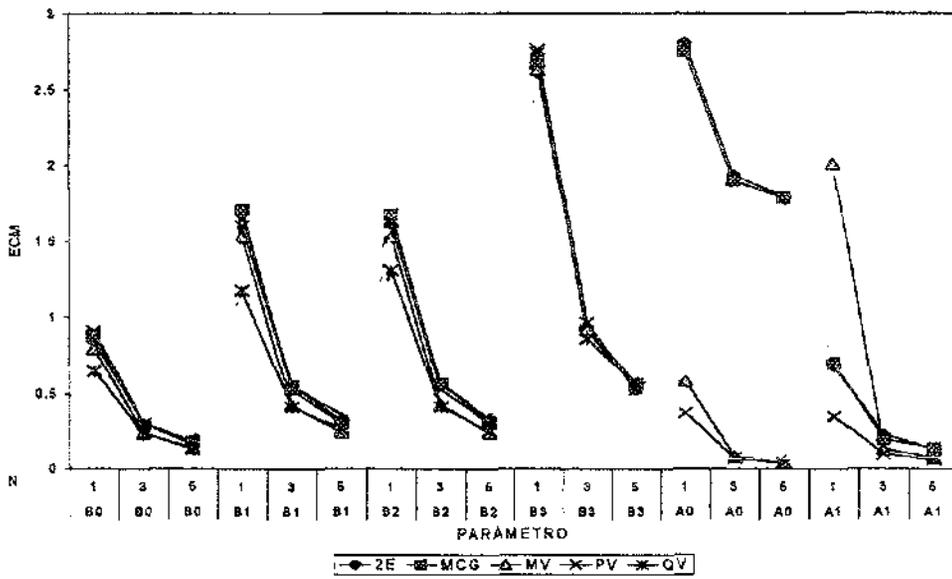


Figura 6.15: ECM de los estimadores

Capítulo 7

Conclusiones

A lo largo de este trabajo de tesis se ha desarrollado la idea de explorar, analizar y modelar la varianza más que corregirla.

El análisis exploratorio aplicado a los efectos de dispersión cumple una tarea tan importante como la que se le otorga cuando se busca modelar la media. Los ejemplos utilizados mostraron que cuando el problema de heteroscedasticidad es muy claro, las estrategias empleadas coinciden en delatar a los mismos factores como responsables, pero si es un tanto ambiguo (ejemplo de Pignatiello), pueden tenerse propuestas encontradas que de cualquier manera es importante tomar en cuenta.

Con el desarrollo de los ejemplos famosos (algunos), se ha puesto de manifiesto que la dispersión de la variable respuesta no debe ser atribuible, de manera automática y exclusiva, a factores no controlables. Después de tener algún indicio de heteroscedasticidad es necesario indagar su naturaleza para valorar la posibilidad de controlarla.

En cuanto a las propiedades de los estimadores obtenidos por los diferentes métodos aquí revisados, se observó que la distribución de los estimadores tiende a la normal de manera asintótica, por lo que es necesario tomar otros puntos de comparación. En este caso se han usado el sesgo, el error estándar y el Error Cuadrático Medio de tal manera que se concluye lo siguiente:

1. Cuando la varianza está en función de una covariable que también tiene efecto sobre la media:

- Los estimadores para el modelo de la **media** (β_0 y β_1) tienen un comportamiento esencialmente diferente a los correspondientes en la función varianza, en el primer caso:
 - Conforme la heteroscedasticidad es más severa, aumenta el sesgo y el error estándar.
 - Conforme aumenta el número de repeticiones por punto diseño, el sesgo y el error estándar disminuyen.
 - Cuando se cumple el supuesto de homoscedasticidad, los estimadores obtenidos por los diferentes métodos son por demás similares entre sí en términos del error cuadrático medio (ECM).
 - Si se tiene un problema de heteroscedasticidad y pocas repeticiones, mínimos cuadrados generalizados y quasiverosimilitud generan estimadores con menor ECM pero conforme aumenta el número de repeticiones, el máximo verosímil es el mejor. (bajo un supuesto de normalidad)
- Por otro lado, los estimadores de los parámetros de la **función varianza**:
 - El intercepto obtenido por pseudoverosimilitud o máxima verosimilitud es el mejor (menor sesgo, menor varianza) y, en la medida que aumenta el número de repeticiones, tales propiedades tienden a mejorar.
 - Para α_1 se observa que conforme la heteroscedasticidad es más fuerte, el estimador es más sesgado sin importar el método ni el número de repeticiones pero por pseudo y máxima verosimilitud se obtienen valores de $\hat{\alpha}_1$ con menor varianza.

2 En el diseño Box-Behnken se observa que:

- El sesgo y varianza de los estimadores de la media son prácticamente iguales independientemente del método.
- La diferencia está en la estimación de los parámetros de la función varianza, si no se cuenta con repeticiones, pseudoverosimilitud es mejor pero máximo verosimilitud tiene mejores propiedades conforme aumenta el número de observaciones por punto diseño.

En general, por supuesto, que conforme se agudiza un problema de heterogeneidad de varianzas, es preferible contar con gran número de observaciones para poder analizar de manera representativa tal comportamiento. A mayor número de observaciones, menor incertidumbre en los resultados. En este mismo sentido y bajo el supuesto de normalidad, el estimador máximo verosímil sigue mostrando su relevancia en la tarea estadística de inferencia.

Las propiedades aquí analizadas corresponden a situaciones muy particulares de heteroscedasticidad, será importante estudiar el comportamiento de las estimaciones ante una gran variedad de situaciones, por ejemplo, el caso en que la varianza es función de una potencia de la media; cuando aumenta el número de covariables en la función varianza sin que éstas participen en la media; cuando se trata de modelos no lineales o cuando la distribución de las observaciones es diferente a la normal.

Referencias bibliográficas

- Aitkin, M.(1987). "Modeling variance heterogeneity in normal regression using GLIM". *Appl. Statistics* Vol 36 No 3 pp.332-339. Letters to the editor.
- Amemiya, T.(1977) "A note on heteroscedastic model". *Journal of econometrics* No. 6 pp. 365-370.
- Box, George. (1994)"Statistics and quality improvement".*Journal of Royal Statistical Society, serie "A"* pp 209-229
- Box, G., Meyer, D. (1986) "Dispersion effect from fractional designs". *Technometrics* Vol 28 No.1 pp.19-27.
- Carroll,R.J., Ruppert, D. (1982) "Robust estimation in heteroscedastic linear models". *Annals of Statistics*. Vol 10 No.2 pp.429-441
- Carroll,R.J., Ruppert, D. (1988) "Transformation and weighting in regression". Chapman and Hall, U.S.A.
- Cook, D, Weisberg S.(1983)."Diagnostics for heteroscedasticity in regression. *Biometrika*. Vol 70 No1 pp 1-10.
- Davidian, M., Carroll, R.J. (1977)."Variance functions estimation". *JASA* Vol 82 No. 4 pp.1079-1091.
- Davidian, M., Carroll, R.J. (1988). " A note on extended quasilielihood". *Journal of the Royal Statistical Society "B"*. Vol 50 No.1 pp.74-82
- Engel, J. (1992). "Modelling variation in industrial experiments". *Appl Statistics*. Vol 41 No 3 pp. 579-593.

- Firth, D. (1987) "On the efficiency of quasi-likelihood estimation". *Biometrika*, Vol 74 No.2 pp.233-245
- Gauss System versión 3.11 (1990-1995). Aptech systems .Inc. Wash- ington.
- Grego, J. (1993) "Generalized Linear Models and process variation". *Journal of Quality Technology*. Vol 25 No. 4
- Gong, G., Samaniego, F. (1981). "Pseudo maximum likelihood estima- tion: Theory and Applications". *Annals of Statistics*, Vol 9 No. 4 pp. 881-869
- Harvey, A.C. (1976). "Estimating regression models with multiplicative heteroscedasticity" *Econometrica*, Vol 44 No. 3 pp. 461-465
- Harville D.(1977) "Maximum likelihood approaches to variance com- ponents estimation and to related problems". *JASA* Vol 72 No. 358
- Heiler, S. (1981). *Robust estimates in linear regression, a simulation approach*. Computational Statistics. Berlín
- Jobson, J., Fuller, W. (1980). "Least Squares Estimation when the co- variance matrix and the parameter vector are functionally related". *JASA* Vol 75 No. 369 pp.176-181
- Kackar, R (1985)"Off-line quality control, parameter design, and the Taguchi method". *Journal of the Quality Technology*". Vol 17 No.4 pp.176-204
- Logothetis N. (1990). "Box-Cox transformations and the Taguchi me- thods". *Appl. Statistics* Vol 39 No. 1 pp-31-48
- McCullagh, P., Nelder, J.S. (1988). *Generalized Linear Models*. 2nd. ed. Chapman and Hall. Great Britain
- Mc.Cullagh, P. (1983)"Quasi-likelihood functions" *The Annals of Statis- tics*. Vol.11 No.1 pp.59-67
- Nelder, J.A., Pegibon, D. (1987). "An extended quasiliikelihood func- tions". *Biometrika*, Vol 74 No. 2 pp.221-232

- Nelder, J.A., Lee, Y. (1992) "Likelihood, quasilielihood and pseudolikelihood: some comparisons". J.R. Statist. Soc. B. Vol 54, No.1, pp.273-284
- Nelder, J.A. (1992) "Pseudolikelihood and Quasi-likelihood". Vol 41 No 3 pp.595-600
- Neter J., et al.(1989). Applied Linear Statistical Models. Irwin Illinois.
- Wedderburn, R.W.M (1974). "Quasilielihood functions, generalized linear models and the Gauss-Newton method". Biometrika Vol 61, No.3 pp. 439-447

Apéndice A

Conjuntos de datos empleados en el desarrollo de ésta tesis.

A.1 Presión arterial diastólica

<i>i</i>	EDAD	PRESION	<i>i</i>	EDAD	PRESION	<i>i</i>	EDAD	PRESION
1	27	73	19	37	78	37	42	85
2	21	66	20	38	87	38	44	71
3	22	63	21	33	76	39	46	80
4	26	79	22	35	79	40	47	96
5	25	68	23	30	73	41	45	92
6	28	67	24	37	68	42	55	76
7	24	75	25	31	80	43	54	71
8	25	71	26	39	75	44	57	99
9	23	70	27	46	89	45	52	86
10	20	65	28	49	101	46	53	79
11	29	79	29	40	70	47	56	92
12	24	72	30	42	72	48	52	85
13	20	70	31	43	80	49	57	109
14	38	91	32	46	83	50	50	71
15	32	76	33	43	75	51	59	90
16	33	69	34	49	80	52	50	91
17	31	66	35	40	90	53	52	100
18	34	73	36	48	70	54	58	80

Fuente: Neter- Wasserman (1989) página 421.

A.2 Árboles de cereza

i	DIAM	ALTURA	VOL	i	DIAM	ALTURA	VOL
1	8.3	70	10.3	17	8.8	63	10.2
2	10.7	81	18.8	18	11	66	15.6
3	11.1	80	22.6	19	11.3	79	24.2
4	11.4	76	21.4	20	12	75	19.1
5	12.9	85	33.8	21	13.7	71	25.7
6	14	78	34.5	22	14.5	74	36.3
7	16.3	77	42.6	23	17.5	82	55.7
8	18	80	51.5	24	20.6	87	77
9	8.6	65	10.3	25	10.5	72	16.4
10	10.8	83	19.7	26	11	75	18.2
11	11.2	75	19.9	27	11.4	76	21
12	11.7	69	21.3	28	12.9	74	22.2
13	13.3	86	27.4	29	13.8	64	24.9
14	14.2	80	31.7	30	16	72	38.3
15	17.3	81	55.4	31	17.9	80	58.3
16	18	80	51				

Fuente: Ryan, Joiner y Ryan (1976)

A.3 Resortes automotrices

CORRIDA	B	C	D	E	O	1	2	3
1	-	-	-	-	-	7.78	7.78	7.81
2	+	-	-	+	-	8.15	8.18	7.88
3	-	+	-	+	-	7.50	7.56	7.50
4	+	+	-	-	-	7.59	7.56	7.75
5	-	-	+	+	-	7.94	8.00	7.88
6	+	-	+	-	-	7.69	8.09	8.06
7	-	+	+	-	-	7.56	7.62	7.44
8	+	+	+	+	-	7.56	7.81	7.69
9	-	-	-	-	+	7.50	7.25	7.12
10	+	-	-	+	+	7.88	7.88	7.44
11	-	+	-	+	+	7.50	7.56	7.50
12	+	+	-	-	+	7.63	7.75	7.56
13	-	-	+	+	+	7.32	7.44	7.44
14	+	-	+	-	+	7.56	7.69	7.62
15	-	+	+	-	+	7.18	7.18	7.25
16	+	+	+	+	+	7.81	7.50	7.59

Fuente: Mc Cullagh y Nelder (1983). página 366

A.4 Circuitos integrados

CORRIDA	A	B	C	D	1	2	3	4	5
1	D2A	Cu	90	Tin	73.0	73.2	72.8	72.2	76.2
2	D2A	Cu	120	Ag	87.7	86.4	86.9	87.9	86.4
3	D2A	Ni	90	Ag	80.5	81.4	82.6	81.3	82.1
4	D2A	Ni	120	Tin	79.8	77.8	81.3	79.8	78.4
5	H-1-E	Cu	90	Ag	85.2	85.0	80.4	85.2	83.6
6	H-1-E	Cu	120	Tin	78.0	75.5	83.1	81.2	79.9
7	H-1-E	Ni	90	Tin	78.4	72.8	80.5	78.4	67.9
8	H-1-E	Ni	120	Ag	90.2	87.4	92.9	90.0	91.9

Fuente: Grego, John (1993)

Apéndice B

B.1 Tablas de resultados del Experimento I

SESGO

Estimador	n	ALFA	2E	MCP	MV	PV	GV
<u>S 0</u>							
3	1.2	0.062	-0.195	0.169	0.208	0.241	
	1.79	-1.233	0.174	-0.168	0.543	0.757	
	2.55	-4.025	-0.984	-6.51	2.694	1.366	
5	1.2	-0.361	-0.121	0.044	-0.045	-0.007	
	1.79	-1.169	0.25	0.302	-0.859	-0.398	
	2.55	-4.257	1.188	-1.582	2.999	0.524	
8	1.2	-0.103	0.113	-0.133	0.009	0.043	
	1.79	0.531	-1.176	-0.19	1.728	-0.497	
	2.55	2.16	-0.845	0.566	1.084	-1.588	
<u>S 1</u>							
3	1.2	-0.021	0.061	-0.014	-0.069	-0.056	
	1.79	0.403	-0.106	0.133	-0.213	-0.295	
	2.55	1.234	0.304	2.474	-1.166	-0.481	
5	1.2	0.148	0.063	-0.002	0.006	-0.004	
	1.79	0.425	-0.116	-0.127	0.291	0.153	
	2.55	1.593	-0.453	0.587	-0.983	-0.207	
8	1.2	0.043	-0.045	0.085	-0.001	-0.004	
	1.79	-0.191	0.419	0.077	-0.589	0.196	
	2.55	-0.884	0.432	-0.241	-0.348	0.58	
<u>Alfa 0</u>							
3	1.2	1.251	1.559	0.394	-0.01		
	1.79	1.017	1.49	0.462	-0.322		
	2.55	0.634	1.551	0.684	-0.554		
5	1.2	1.306	1.297	0.236	-0.001		
	1.79	1.004	1.396	0.305	-0.081		
	2.55	0.725	1.54	0.323	-0.383		
8	1.2	1.216	1.342	0.117	-0.008		
	1.79	1.188	1.353	0.191	-0.024		
	2.55	1.069	1.487	0.149	-0.208		
<u>Alfa 1</u>							
3	1.2	0.033	-0.068	-0.061	0.064		
	1.79	-0.487	-0.633	-0.678	-0.429		
	2.55	-1.126	-1.418	-1.507	-1.124		
5	1.2	0.003	0.008	-0.038	0.039		
	1.79	-0.485	-0.603	-0.652	-0.524		
	2.55	-1.157	-1.421	-1.416	-1.156		
8	1.2	0.024	-0.008	-0.014	0.028		
	1.79	-0.55	-0.605	-0.629	-0.562		
	2.55	-1.278	-1.405	-1.377	-1.287		

ERROR ESTÁNDAR

ESTIMADOR	n	alfa1	2E	MCP	MV	PV	QV
<u>B0</u>							
3	1.2		11.679	11.347	11.121	11.542	10.377
	1.79		30.175	26.626	26.635	31.22	24.181
	2.55		115.251	82.769	76.108	115.991	77.118
5	1.2		8.786	8.466	8.349	9.065	8.157
	1.79		24.409	20.084	19.46	23.362	18.667
	2.55		89.58	60.108	59.966	88.708	55.481
8	1.2		7.243	6.778	6.445	7.268	6.385
	1.79		19.142	15.931	14.905	20.351	14.785
	2.55		69.737	45.782	45.58	73.644	45.299
<u>B1</u>							
3	1.2		4.182	4.086	4.039	4.135	3.769
	1.79		11.007	9.881	9.928	11.475	8.987
	2.55		43.002	32.012	29.584	43.163	29.959
5	1.2		3.148	3.06	3.003	3.224	2.194
	1.79		8.935	7.497	7.315	8.583	6.993
	2.55		33.397	23.273	23.195	33.029	21.5
8	1.2		2.557	2.42	2.317	2.582	2.288
	1.79		6.992	5.948	5.542	7.405	5.547
	2.55		26.007	17.738	17.607	27.313	17.518
<u>Alfa0</u>							
3	1.2		3.239	3.428	2.544	2.216	
	1.79		3.131	3.424	2.587	2.279	
	2.55		3.181	3.616	2.738	2.297	
5	1.2		2.429	2.556	1.841	1.627	
	1.79		2.419	2.73	1.77	1.651	
	2.55		2.512	2.519	1.861	1.639	
8	1.2		1.918	2.066	1.347	1.292	
	1.79		1.987	1.963	1.342	1.314	
	2.55		1.956	1.951	1.33	1.265	
<u>Alfa1</u>							
3	1.2		1.067	1.133	0.847	0.731	
	1.79		1.038	1.127	0.853	0.766	
	2.55		1.052	1.196	0.905	0.755	
5	1.2		0.802	0.847	0.609	0.542	
	1.79		0.804	0.905	0.581	0.594	
	2.55		0.834	0.837	0.615	0.54	
8	1.2		0.636	0.676	0.441	0.427	
	1.79		0.661	0.649	0.443	0.435	
	2.55		0.641	0.642	0.438	0.416	

ECM

<u>Estimador</u>	<u>n</u>	<u>ALFA</u>	<u>2E</u>	<u>MCP</u>	<u>MV</u>	<u>PV</u>	<u>QV</u>
<u>B 0</u>							
3	1 2		136 40	128 79	123 71	133 26	107 74
	1 79		912.05	708 97	709 45	974.98	585 29
	2 55		13298 98	6851 68	5834 81	13461.17	5949 05
5	1 2		77.32	71 69	69 71	82.18	66 54
	1 79		597.17	403 43	378 78	545.52	348 82
	2 55		8042 70	3614.38	3596.42	7678.10	3078 42
8	1 2		52 47	45.95	41.56	52 82	40 77
	1 79		366.70	256 18	222 20	417.15	218.84
	2 55		4867 91	2098.71	2077.86	5424 61	2054 52
<u>B 1</u>							
3	1 2		17 49	16 70	16 31	17 10	14 21
	1 79		121.32	97 65	98 58	131.72	80 85
	2 55		1850.69	1024 86	881 33	1864.27	887 77
5	1 2		9.93	9 37	9 02	10.39	4.81
	1 79		80.01	56.22	53.53	73 75	48 93
	2 55		1117.90	541 84	538 35	1091 88	462 29
8	1 2		6 54	5.86	5 37	6.67	5 23
	1 79		48 92	35.55	30 72	55.18	30 81
	2 55		677 15	314 82	310 06	746 12	307 22
<u>Alfa 0</u>							
3	1 2		12 06	14.18	6 63	4.91	
	1 79		10 84	13.94	6 91	5.30	
	2 55		10.52	15 48	7 96	5.58	
5	1 2		7 61	8 21	3 44	2.85	
	1 79		6 86	9 40	3 23	2.73	
	2 55		6.84	8 72	3 57	2.83	
8	1 2		5 16	6 07	1 83	1.67	
	1 79		5 36	5.68	1 84	1.73	
	2 55		4.97	6.02	1.79	1 64	
<u>Alfa 1</u>							
3	1 2		1.14	1 29	0.72	0.54	
	1 79		1.31	1 67	1 19	0 77	
	2 55		2 37	3 44	3 09	1 83	
5	1 2		0 64	0 72	0 37	0 30	
	1 79		0 88	1 18	0 76	0 63	
	2 55		2 03	2 72	2 38	1 72	
8	1 2		0 41	0 46	0 19	0 18	
	1 79		0 74	0 79	0 59	0 51	
	2 55		2 04	2 39	2 09	1 78	

B.2 Tablas de resultados del Experimento II

SESGO

<i>Estimador</i>	<i>n</i>	<i>2E</i>	<i>MCG</i>	<i>MV</i>	<i>PV</i>	<i>QV</i>
B0						
	1	0.010	0.018	-0.035	-0.018	0.053
	3	0.044	-0.007	0.018	-0.011	-0.020
	5	0.008	0.001	-0.001	0.002	0.016
B1						
	1	0.059	0.037	0.078	0.016	0.006
	3	-0.025	-0.039	-0.011	0.019	-0.023
	5	0.007	-0.013	-0.004	0.009	-0.026
B2						
	1	-0.013	0.025	-0.007	0.063	0.013
	3	-0.030	-0.003	0.021	0.015	0.000
	5	0.011	-0.031	0.016	0.013	-0.023
B3						
	1	-0.018	-0.010	-0.048	-0.088	0.090
	3	-0.062	0.038	0.029	-0.040	0.030
	5	-0.052	0.013	-0.023	-0.005	-0.010
A0						
	1	1.540	1.528	0.576	0.380	
	3	1.350	1.335	0.147	0.116	
	5	1.313	1.309	0.086	0.064	
A1						
	1	0.230	0.230	-0.512	0.204	
	3	0.062	0.061	-0.076	0.086	
	5	0.050	0.035	-0.023	0.043	

ERROR ESTÁNDAR

Estimador	n	2E	MCG	MV	PV	QV
B 0						
	1	0.917	0.942	0.889	0.953	0.803
	3	0.542	0.537	0.483	0.545	0.474
	5	0.396	0.425	0.36	0.42	0.358
B 1						
	1	1.293	1.305	1.236	1.264	1.084
	3	0.744	0.722	0.647	0.744	0.64
	5	0.56	0.545	0.486	0.593	0.512
B 2						
	1	1.267	1.293	1.236	1.274	1.143
	3	0.761	0.75	0.648	0.722	0.64
	5	0.572	0.556	0.481	0.546	0.495
B 3						
	1	1.629	1.632	1.621	1.66	1.653
	3	0.961	0.953	0.958	0.979	0.924
	5	0.721	0.735	0.753	0.723	0.746
A 0						
	1	0.859	0.845	0.498	0.469	
	3	0.333	0.335	0.229	0.229	
	5	0.263	0.259	0.171	0.174	
A 1						
	1	0.792	0.802	1.322	0.55	
	3	0.467	0.432	0.345	0.291	
	5	0.352	0.358	0.256	0.231	

ECM

Estimador	n	2E	MCG	MV	PV	QV
B 0						
	1	0.841	0.888	0.792	0.909	0.648
	3	0.296	0.286	0.234	0.297	0.225
	5	0.157	0.181	0.130	0.176	0.128
B 1						
	1	1.675	1.704	1.534	1.598	1.175
	3	0.554	0.523	0.419	0.554	0.410
	5	0.314	0.297	0.236	0.352	0.263
B 2						
	1	1.605	1.672	1.528	1.627	1.307
	3	0.565	0.563	0.420	0.522	0.410
	5	0.327	0.310	0.232	0.298	0.246
B 3						
	1	2.654	2.664	2.630	2.763	2.741
	3	0.927	0.910	0.919	0.960	0.855
	5	0.523	0.540	0.568	0.523	0.561
A 0						
	1	2.606	2.751	0.580	0.364	
	3	1.933	1.894	0.074	0.066	
	5	1.793	1.781	0.037	0.034	
A 1						
	1	0.680	0.696	2.010	0.344	
	3	0.222	0.190	0.125	0.092	
	5	0.126	0.129	0.066	0.055	

Apéndice C

C.1 Programas de simulación

Regresión en dos etapas

```
/* METODO DE REGRESION EN DOS ETAPAS */
/*****/
library maxlik;
#include maxlik.ext;
maxset;
let TV=1; /*tamaño del vector*/
m=1;
n=5; /* tamaño de muestra*/
ni=7; /* número de tratamientos */
TM=7*n; /*numero total de observaciones: n repeti-
ciones por 7 tratamientos*/
let beta=1.5 0.6;
let alfa=0.5 3.5;
MC=zeros(4,1); /* b0, b1, t0,t1 por minimos cuadrados*/
ESTIMA=ZEROS(TV,5); /*MATRIZ DE RESULTADOS */
/* MATRIZ DISEÑO */
proc matdis(tt);
local x1,i; i=1;
x1=zeros(tt,2); x1[.,1]=ones(tt,1);
do while i<8;
x1[i,2]=2.3+0.23*(i-1);
i=i+1;
enddo;
retp(x1); endp;
{x1}=matdis(ni);
```

```

x=x1; j=1;
do while j<n;
  x=x|x1;
  j=j+1;
endo;
DO WHILE M<TV+1;
  /* VECTOR DE OBSERVACIONES */
proc observ(tt);
  local y1,e,e1;
  e=rndn(tt,1);
  e1=e.*exp((alfa[1]+alfa[2]*x[.,2])/2);
  y1=beta[1]+beta[2]*x[.,2]+e1;
  retp(y1);
endp;
{y}=observ(tm);
/* ESTIMACION POR MINIMOS CUADRADOS (regresion en dos etapas */
proc (3)=mincuad(x1, y1);
  local mcb1,r1,yr,mca1,s1,v;
  mcb1=inv(x1'x1)*x1'y1; /*estimacion de beta*/
  r1=y1-x1*mcb1; /*residuos del modelo de medias*/
  yr=ln(r1^2);
  mca1=inv(x1'x1)*x1'yr; /*estimacion de teta*/
  v=exp(x1*mca1);
  s1=((r1./v)*r1)/(tm-2); /*suma de cuadrados de los residuos media*/
retp(mcb1,mca1,s1); endp;
{mcb,mca,cme}=mincuad(x,y);
ESTIMA[M,1]=MCB[1]; ESTIMA[M,2]=MCB[2]; ESTIMA[M,3]=MCA[1];
ESTIMA[M,4]=MCA[2]; ESTIMA[M,5]=cme;
M=M+1;
ENDO;
output file=c:\gauss\mod13e3.dat on;
print estima;

```

Mínimos cuadrados generalizados

```

/* METODO DE MINIMOS CUADRADOS GENERALIZADOS*/
/*****/
library maxlik;

```

```

#include maxlik.ext;
maxset;
let TV=1000;          /*tamaño del vector*/
m=1;
n=5;                 /* tamaño de muestra*/
ni=7;                /* número de tratamientos */
TM=ni*n;             /*número total de observaciones: n repeticiones
por 7 tratamientos*/
let beta=1.5 0.6;
let alfa=0.5 3.5;
ESTIMA=ZEROS(TV,6); /*MATRIZ DE RESULTADOS*/
/* MATRIZ DISEÑO */
proc matdis(tt);
local x1,i;
i=1;
x1=zeros(tt,2);
x1[.,i]=ones(tt,1);
do while i<8;
    x1[i,2]=2.3+0.23*(i-1);
    i=i+1;
end;
retp(x1); endp;
{x1}=matdis(ni);
x=x1; j=1;
do while j<n;
    x=x|x1;
    j=j+1;
end;
DO WHILE M<TV+1;
/* VECTOR DE OBSERVACIONES */
proc observ(tt);
local y1,e,e1;
e=rndn(tt,1);
e1=e.*exp((alfa[1]+alfa[2]*x[.,2])/2);
y1=beta[1]+beta[2]*x[.,2]+e1;
retp(y1); endp;
{y}=observ(tm);
/* ESTIMACION POR MINIMOS CUADRADOS */

```

```

proc (2)= mincuad(x1, y1);
  local mcb1,yr,mca1,s1,r1;
  mcb1=inv(x1'x1)*x1'y1;          /*estimación de beta*/
  r1=y1-x1*mcb1;                 /*residuos del modelo de medias*/
  yr=ln(r1^2);
  mca1=inv(x1'x1)*x1'yr;         /*estimación de teta*/
  retp(mcb1,mca1);
endp;
{mcb,mca}=mincuad(x,y);
/* PONDERACIONES */
proc (2)= pondera(x1,mca1,tt);
  local var,w1,bpon1,id;
  var=exp(x*mca1);
  id=eye(tt);
  w1=diagrv(id,(1/var));
  bpon1=inv(x1'w1*x1)*(x1'w1*y);
  retp(bpon1,w1);
endp;
/*ESTIMACION POR MINIMOS CUADRADOS PONDERADOS*/
difb=1; difa=1; amcp=mca; bmcp=mcb; i=1;
do until abs(difb)<1e-05 and abs(difa)<1e-05 or i>1000;
  {bpon,w1}=pondera(x,amcp,tm);
  res=y-x*bpon;
  yr=ln(res^2);
  apon=inv(x'x)*(x'yr);
  difb=bmcp-bpon; difa=amcp-apon;
  bmcp=bpon; amcp=apon; i=i+1;
endo;
var=exp(x*apon);
cme=((res./var)**res)/(tm-2);
estima[m,1]=bpon[1]; estima[m,2]=bpon[2]; estima[m,3]=apon[1];
estima[m,4]=apon[2]; estima[m,5]=cme; estima[m,6]=i;
M=M+1;
ENDO;
output file=mod13cp5.dat on;

print estima;

```

Máxima Verosimilitud

/* PROGRAMA DE SIMULACION PARA ESTIMAR LOS PARAMETROS DE UN MODELO

```

CON HETROSCEDASTICIDAD*/
library maxlik;
#include maxlik.ext;
maxset;
let TV=1000;          /*tamaño del vector*/
m=1;
n=8;                 /* tamaño de muestra*/
ni=7;                /* número de tratamientos */
TM=ni*n;             /*número total de observaciones : n repeticiones por
7 tratamientos*/
let beta=1.5 0.6;
let alfa=0.5 3.5;
VEROSI=ZEROS(TV,5);    /*MATRIZ DE RESULTADOS*/
MC=ZEROS(4,1);
/* MATRIZ DISEÑO */
proc matdis(tt);
  local x1,j;
  i=1;
  x1=zeros(tt,2);
  x1[.,1]=ones(tt,1);
  do while i<8;
    x1[i,2]=2.3+0.23*(i-1);
    i=i+1;
  endo;
  retp(x1);
endp;
{x1}=matdis(ni);
x=x1; j=i;
do while j<n;
  x=x|x1;
  j=j+1;
endo;
DO WHILE M<TV+1;
/* VECTOR DE OBSERVACIONES */

```

```

proc observ(tt);
  local y1,e,e1;
  e=rndn(tt,1);
  e1=e.*exp((alfa[1]+alfa[2]*x[.,2])/2);
  y1=beta[1]+beta[2]*x[.,2]+e1;
  retp(y1);
endp;
{y}=observ(tm);
output file=mod13v8.dat on;
/* ESTIMACION POR MINIMOS CUADRADOS (regresión en dos etapas */
proc (2)= mincuad(x1, y1);
  local mcb1,r1,yr,mca1,s1;
  mcb1=inv(x1'x1)*x1'y1; /*estimacion de beta*/
  r1=y1-x1*mcb1; /*residuos del modelo de medias*/
  yr=ln(r1^2);
  mca1=inv(x1'x1)*x1'yr; /*estimacion de teta*/
retp(mcb1,mca1);
endp;
{mcb,mca}=mincuad(x,y);
mc[1,1]=mcb[1]; mc[2,1]=mcb[2];
mc[3,1]=mca[1]; mc[4,1]=mca[2];
/* DEFINICION DE LA FUNCION DE VEROSIMILITUD */
proc normal(b,x1);
  local b0,b1,t0,t1,pp,fv,v;
  b0=b[1]; b1=b[2]; t0=b[3]; t1=b[4];
  v=exp(x*(b[3]|b[4]));
  pp=x[.,1]*(x*(b[3]|b[4]));
  fv=((-0.5)*(pp+(y-(x*(b[1]|b[2])))*((y-(x*(b[1]|b[2])))./v)));
retp(fv);
endp;
/* ESTIMACION POR MAXIMA VEROSIMILITUD */
vin=mc;
proc (3)= maxver(x1);
  local mvb1, mvb,mva, var, fmin, g,r2,s2, cov,retcode;
  {mvb1,fmin,g, cov,retcode}=maxlik(x1,0,&normal,vin);
  mvb=mvb1[1]|mvb1[2];
  mva=mvb1[3]|mvb1[4];
  r2=y-x1*mvb;

```

```

    var=exp(x1*mva);
    s2=((r2./var)*r2)/(tm-2);
    retp(mvb,mva,s2);
endp;
{mvb,mva,CME}=maxver(x);
VEROSI[M,1]=mvb[1];VEROSI[M,2]=mvb[2]; VEROSI[M,3]=MVA[1];
VEROSI[M,4]=MVA[2]; VEROSI[M,5]=CME;
M=M+1;
ENDO;
print verosi;

```

Pseudoverosimilitud

/* PROGRAMA DE SIMULACION PARA ESTIMAR LOS PARAMETROS DE UN MODELO

CON VARIANZA NO CONSTANTE

```

library maxlik;
#include maxlik.ext;
maxset;
let TV=i000;           /*tamaño del vector*/
m=1;
n=5;                  /* tamaño de muestra*/
ni=8;                 /* número de tratamientos */
TM=ni*n;              /*número total de observaciones : n repeticiones
por 7 tratamientos*/
let beta=1.5 0.6;
let alfa=0.5 3.5;
PSEUDO=ZEROS(TV,6);   /*MATRIZ DE RESULTADOS*/
/* MATRIZ DISEÑO */
proc matdis(tt);
    local x1,i;
    i=1;
    x1=zeros(tt,2);
    x1[:,1]=ones(tt,1);
    do while i<8;
        x1[i,2]=2.3+0.23*(i-1);
        i=i+1;
    endo;

```

```

retp(x1);
endp;
{x1}=matdis(ni);
x=x1; j=1;
do while j<n;
    x=x|x1;
    j=j+1;
endo;
DO WHILE M<TV+1;
/* VECTOR DE OBSERVACIONES */
proc observ(tt);
    local y1,e,e1;
    e=rndn(tt,1);
    e1=e.*exp((alfa[1]+alfa[2]*x[.,2])/2);
    y1=beta[1]+beta[2]*x[.,2]+e1;
retp(y1);
endp;
{y}=observ(tm);
output file=mod13p8.dat on;
/* ESTIMACION POR MINIMOS CUADRADOS (regresión en dos etapas)
*/
proc (3)= mincuad(x1, y1);
    local mcb1,r1,yr,mca1,s1;
    mcb1=inv(x1'x1)*x1'y1; /*estimacion de beta*/
    r1=y1-x1*mcb1; /*residuos del modelo de medias*/
    yr=ln(r1^2);
    mca1=inv(x1'x1)*x1'yr; /*estimacion de teta*/
retp(mcb1,mca1,r1);
endp;
{mcb,mca,res}=mincuad(x,y);
/*FUNCION DE VEROSIMILITUD NORMAL PARA ALFA*/
proc normala(a,x1);
    local v;
    v=exp(x1*(a[1]|a[2]));
    retp((-0.5)*(x1[.,1]'*(x1*(a[1]|a[2]))+(res./v)**res));
endp;
/*ESTIMACION DE ALFA POR MAXIMA VEROSIMILITUD*/
vin1=mca; /* valores iniciales*/

```

```

proc maxvert(x1);
    local mval,fmin,g,cov,retcode;
    {mval,fmin,g,cov,retcode}=maxlik(x1,0,&normala,vin1);
retp(mval);
endp;
/* FUNCION DE VEROSIMILITUD NORMAL PARA BETA*/
proc normalb(b,x1);
    local v;
    v=exp(x*mca);
retp((-0.5)*((y-(x1*(b[1]|b[2])))*(y-(x1*(b[1]|b[2]))))./v));
endp;
/*ESTIMACION DE BETA POR MAXIMA VEROSIMILITUD */
vin2=mcb;
proc maxverb(x1);
    local mvb1,fmin,g,cov,retcode;
    {mvb1,fmin,g,cov,retcode}=maxlik(x1,0,&normalb,vin2);
retp(mvb1);
endp;
/* ESTIMACION POR PSEUDOVEROSIMILITUD */
/* SE ESTIMA BETA Y ALFA POR VEROSIMILITUD POR SEPARADO*/
difb=1; difa=1; iq=1;
do until abs(difb)<1e-05 and abs(difa)<1e-05 or iq>1000;
    {aq}=maxvert(x);
    {bq}=maxverb(x);
    res=y-x*bq;
    difb=vin2-bq; difa=vin1-aq;
    vin2=bq; mct=aq; vin1=aq; iq=iq+1;
enddo;
vari=exp(x*aq);
CME=((res./vari)*res)/(tm-2);
pseudo[m,1]=bq[1]; pseudo[m,2]=bq[2]; pseudo[m,3]=aq[1];
pseudo[m,4]=aq[2]; pseudo[m,5]=CME; pseudo[m,6]=iq;
M=M+1;
ENDO;
print pseudo;

```

Quasiverosimilitud

```
/* PROGRAMA DE SIMULACION PARA ESTIMAR LOS PARAMETROS DE
UN MODELO
```

```
  CON VARIANZA NO CONSTANTE */
  library maxlik;
  #include maxlik.ext;
  maxset;
  let TV=1000;           /*tamaño del vector*/
  m=1;
  n=8;                 /* tamaño de muestra*/
  ni=7;               /* número de tratamientos */
  TM=ni*n;           /*número total de observaciones : n repeticiones
por 7 tratamientos*/
  DIFER=ZEROS(TV,1);
  let beta=1.5 0.6;
  let alfa=0.5 3.5;
  QUASI=ZEROS(TV,4);  /*MATRIZ DE RESULTADOS*/
  id=eye(tm);
  /* MATRIZ DISEÑO */
  proc matdis(tt);
    local x1,j;
    i=1;
    x1=zeros(tt,2);
    x1[:,1]=ones(tt,1);
    do while i<ni+1;
      x1[i,2]=2.3+0.23*(i-1);
      i=i+1;
    endo;
  retp(x1);
  endp;
  {x1}=matdis(ni);
  x=x1; j=1;
  do while j<n;
    x=x|x1;
    j=j+1;
  endo;
  DO WHILE M<TV+1;
```

```

/* VECTOR DE OBSERVACIONES */
proc observ(tt);
  local y1,e,e1;
  e=rndn(tt,1);
  e1=e.*exp((alfa[1]+alfa[2]*x[.,2])/2);
  y1=beta[1]+beta[2]*x[.,2]+e1;
retp(y1);
endp;
{y}=observ(tm);
output file=mod13q8.dat on;
var=exp(x*alfa);
w1=diagrv(id,(1/var));
/* ESTIMACION POR MINIMOS CUADRADOS */
proc (2)= mincuad(x1, y1);
  local mcb1,r1,yr,mct1,s1,r2;
  mcb1=inv(x1'x1)*x1'y1; /*estimacion de beta*/
  r1=y1-x1*mcb1; /*residuos del modelo de medias*/
retp(mcb1,r1);
endp;
{mcb,res}=mincuad(x,y);
/* PONDERACIONES */
proc pondera(x1,tt);
  local bpon1;
  bpon1=inv(x1'w1*x1)*(x1'w1*res);
retp(bpon1);
endp;
/*ESTIMACION POR MINIMOS CUADRADOS PONDERADOS*/
difb=1; bmcp=mcb; i=1;
do until abs(difb)<1e-05 or i>1000;
  {bpon}=pondera(x,tm);
  bquasi=mcb+bpon;
  res=y-x*bquasi;
  difb=mcb-bquasi;
  mcb=bquasi;
  i=i+1;
endo;
cme=((res./var)'*res)/(tm-2);
quasi[m,1]=bquasi[1]; quasi[m,2]=bquasi[2]; quasi[m,3]=cme;
quasi[m,4]=i;
M=M+1;
ENDO;
print quasi;

```