ESTADISTICA ESPAÑOLA
núm. 116, 1988, págs. 55 a 74

# Un modelo para el estudio de efectos sobre la dispersión en ausencia de replicaciones

por
L. R. ZUNICA RAMAJO
y
R. ROMERO VILLAFRANCA
Departamento de Estadística e Investigación Operativa
de la Universidad Politécnica de Valencia.
Apartado 22012
46071 - Valencia

**RESUMEN**

Las aportaciones japonesas al campo de la aplicación de la Estadística a la mejora de la calidad y de la productividad, han originado un interés creciente respecto al estudio de efectos sobre la dispersión. Las técnicas existentes exigen, en general, disponer de replicaciones para las distintas condiciones estudiadas, lo que dificulta el análisis a partir de datos no experimentales, especialmente cuando algunas de las variables fluctúan de forma continua en la muestra. En el trabajo se estudia un modelo que permite el análisis de efectos sobre la dispersión en condiciones muy generales, desarrollándose un algoritmo sencillo de estimación. El modelo y la técnica propuestos se aplican a dos casos reales, uno extraído de la experiencia personal de los autores y otro estudiado por Taguchi y Yu y reanalizado posteriormente por Box y Meyer.

*Palabras clave:* Efectos sobre la dispersión. Modelos lineales generalizados. Heterocedasticidad. Mejora de la calidad y de la productividad. Control "off-line".

Clasificación AMS: 62N10, 62K15.

## 1. INTRODUCCION

El problema de establecer la significación y cuantificar la magnitud de la influencia de un conjunto de variables explicativas sobre el valor medio de cierta variable que depende de las mismas, constituye posiblemente el tema más importante y mejor estudiado dentro del campo de la estadística aplicada. Bajo el nombre general de Modelo Lineal, un conjunto de modelos y técnicas permiten estimar los efectos de diferentes variables sobre dicho valor medio y contrastar hipótesis al respecto, bajo distintos supuestos respecto a la distribución de la variable dependiente y a la naturaleza de dichos efectos.

Estos últimos años, sin embargo, han visto nacer un interés creciente respecto al problema del estudio de efectos sobre la dispersión.

Indudablemente la raíz de este nuevo interés, hay que buscarla en el fuerte impulso originado por las recientes aportaciones japonesas en el campo de la aplicación de técnicas estadísticas para la mejora de la calidad y de la productividad. En particular los trabajos de Taguchi (1979) sobre Control de Calidad "off-line" han resaltado la importancia, con el fin de optimizar la capacidad de un determinado proceso, de obtener las condiciones operativas que, manteniendo el valor medio del output en el objetivo fijado, minimicen la varianza del mismo alrededor de dicho objetivo.

En sus trabajos Taguchi propone un conjunto de diseños experimentales, estrechamente relacionados en muchos casos con las fracciones factoriales, cuya finalidad es la identificación y estimación de los efectos de los factores estudiados sobre la media y sobre la dispersión de la respuesta obtenida. Con el fin de poder estudiar los efectos sobre la dispersión, los diseños de Taguchi incluyen replicaciones de las diferentes condiciones experimentales, sea mediante replicas genuinas o mediante la introducción en el diseño de factores ruido.

El análisis de los efectos sobre la dispersión lo realiza Taguchi aplicando técnicas convencionales de análisis de la varianza a estadísticos derivados a partir de la media y la varianza correspondientes a cada replicación. Taguchi utiliza la denominación genérica de "ratios señal/ruido" para referirse a estos estadísticos, cuyo empleo indiscriminado ha sido objeto de críticas por parte de diversos autores como Box (1986), Box y Ramírez (1986) o Tort-Martorell (1985). Este último autor propone un método alternativo al utilizado por Taguchi para estudiar efectos sobre la dispersión.

Las técnicas desarrolladas en los trabajos mencionados sólo pueden aplicarse si se dispone de replicas para los diferentes valores de los factores o variables explicativas.

Box y Meyer (1986) proponen un método para estudiar efectos sobre la dispersión a partir de los resultados de fracciones factoriales no replicadas. En síntesis el método consiste en eliminar previamente los efectos de posición estimados como significativos, y en evaluar el efecto de cada factor sobre la dispersión a partir del ratio entre las varianzas de las observaciones correspondientes a los dos niveles del mismo.

El método de Box y Meyer sólo resulta aplicable a planes o fracciones factoriales con factores a dos niveles. Frecuentemente, sin embargo, resulta necesario estudiar la existencia de efectos sobre la dispersión en condiciones mucho más generales. Este es el caso, por ejemplo, cuando los datos disponibles no son el resultado de una experiencia diseñada, y muy especialmente cuando algunas de las variables explicativas varían de forma continua en la muestra.

En el presente trabajo se estudia un modelo alternativo que permite la identificación y estimación simultánea de efectos de posición y de dispersión a partir de datos no necesariamente replicados. Las posibilidades de aplicación del modelo expuesto son muy amplias, incluyendo tanto el análisis de datos procedentes de diseños experimentales como el de observaciones en las que las variables explicativas varían de forma continua.

El modelo, que asume efectos multiplicativos de las variables explicativas sobre la varianza de la variable dependiente, ha sido propuesto por Harvey (1976), que desarrolla en la referencia citada la estimación máximo verosímil de sus parámetros. En el presente trabajo se propone un método de estimación alternativo, que presenta la ventaja de su mayor sencillez al sólo exigir la aplicación iterada de un algoritmo ordinario de mínimos cuadrados.

El modelo propuesto ha sido utilizado por los autores para el estudio de diversos problemas relacionados con la mejora de la calidad y la productividad en la industria, especialmente en el sector automovilístico. Uno de estos problemas, relativo al estudio de desajustes entre diferentes unidades que trabajan en paralelo en un proceso de montaje, es analizado en el presente trabajo con el fin de poner de manifiesto la aplicabilidad práctica de la técnica desarrollada.

También se analizan mediante esta técnica unos datos procedentes de un estudio inicial de Taguchi y Wu (1979) posteriormente revisado por Box y Meyer (1986). El análisis pone de manifiesto la sensibilidad de la metodología propuesta al identificar efectos sobre la variabilidad que no habían sido detectados en anteriores estudios.

## 2. UN MODELO PARA EL ESTUDIO DE EFECTOS SOBRE LA DISPERSION

Sea $y_j$ ($j$=1...J) el valor obtenido para una determinada variable en la *j-ava* observación, y sean $x_{ji}$ ($i$=1...I) los valores en la misma de I variables explicativas, donde $x_{j1}$=1 para todo $j$ en el caso de incluirse una constante en los modelos. Asumiremos que las variables explicativas son no aleatorias, pudiendo generalizarse los modelos al caso de regresores estocásticos de la forma habitual.

Los modelos de regresión lineal asumen, en su forma clásica, que $y_j$ es el valor observado de una variable $Y_j$ que se distribuye normalmente con

$$E(Y_j) = \Sigma\, \beta_i\, x_{ji}$$
$$\sigma^2(Y_j) = \sigma^2 \text{ constante}$$

siendo las $Y_j$ independientes entre sí. Los parámetros $\beta_i$ recogen el efecto de las diferentes variables explicativas sobre el valor medio de la variable dependiente.

Matricialmente, denominando **y** al vector cuyas componentes son las $y_j$, éste resulta ser el valor observado de una variable **Y** normal J-dimensional cuyo vector medio es $\mathbf{X\beta}$ y cuya matriz de covarianzas es $\sigma^2\mathbf{I}$, siendo **X** la matriz formada por las $x_{ji}$, $\boldsymbol{\beta}$ el vector de las $\beta_i$ e **I** la matriz unitaria J-dimensional.

Dentro del ámbito del estudio de modelos econométricos heterocedásticos, se han propuesto diversas generalizaciones del modelo clásico que permiten considerar la posible existencia de efectos sobre la dispersión. En este trabajo nos limitamos a considerar un modelo que asume efectos multiplicativos sobre la dispersión, modelo que presenta, entre otras, la ventaja de no permitir estimas negativas de las varianzas.

Sean $Z_1$, $Z_2$,...,$Z_k$ las variables cuyo efecto sobre la dispersión de la Y se desea analizar. (Obviamente las $Z_k$ pueden coincidir total o parcialmente con las $X_i$) y sea $z_{jk}$ el valor en la *j-ava* observación de la variable $Z_k$.

Harvey (1976), propone modelizar el efecto de las $z_{jk}$ sobre las $\sigma_j^2$ asumiendo una relación del siguiente tipo:

$$\sigma_j^2 = e^{\Sigma \alpha_k z_{jk}}$$

o, de forma equivalente

$$L\, n\, (\sigma_j^2) = \Sigma \alpha_k\, z_{jk}$$

formulación que implica un efecto aditivo sobre el logaritmo de $\sigma_j^2$ y, por tanto, multiplicativo sobre $\sigma_j^2$.

Bajo este modelo, **y** resulta ser el valor observado de un vector aleatorio

$$\mathbf{Y} \sim \text{Normal}\ (\mathbf{X\beta}\ ,\ \mathbf{D}(e^{\mathbf{Z}\alpha})\ )$$

donde **Z** es la matriz de las $z_{jz}$, $\alpha$ es el vector de las $\alpha_k$ y $\mathbf{D}(e^{\mathbf{Z}\alpha})$ simboliza una matriz diagonal en la que el elemento *j-avo* de la diagonal principal viene dado por $e$ elevado a la *j-ava* componente de $\mathbf{Z}\alpha$.

El modelo contendrá en general una primera columna de unos en **Z**, es decir una variable $Z_1$ tal que $z_{j1}=1$ para todo $j$, e incluirá como caso particular el modelo ordinario de regresión lineal sin más que hacer $\alpha_k=0$ para $k=2...K$.

## 3. ESTIMACION DEL MODELO

En el presente apartado consideramos el problema de la estimación de los parámetros **ß** y $\alpha$ del modelo. Exponemos en primer lugar dos procedimientos sugeridos por Harvey, uno bastante sencillo y otro basado en la maximización mediante un algoritmo numérico de la verosimilitud. Seguidamente proponemos un método alternativo, que siendo más eficiente que el primero de los mencionados resulta sensiblemente más sencillo de aplicar que el de máxima verosimilitud.

### 3.1. *Estimación bietápica*

Harvey sugiere estimar el vector $\alpha$ mediante una regresión ordinaria sobre la $z_{jk}$ utilizando como variable dependiente el logaritmo neperiano del cuadrado de los residuos estimados en la regresión ordinaria del vector **y** sobre las $x_{ji}$. El autor bautiza este método como "Two-Step Procedure" y nosotros utilizaremos las siglas TSP para referirnos al mismo.

El estimador propuesto es, en consecuencia

$$\mathbf{a} = (\mathbf{Z'Z})^{-1}\,\mathbf{Z'}L_n\mathbf{e}^2 \qquad (3.1.)$$

donde

$$\mathbf{e} = \mathbf{y}-\mathbf{Xb} \quad \text{y} \quad \mathbf{b} = (\mathbf{X'X})^{-1}\,\mathbf{X'y}$$

y donde $L_n\mathbf{e}^2$ es el vector constituído por los $L_n(e_j^2)$.

Por los motivos que se expondrán en el apartado 3.3. el valor de $a_o$ dado por (3.1.) debe corregirse añadiéndole la cantidad 1.27 para obtener un estimador consistente. La consistencia del estimador se deduce del hecho de que, bajo ciertas condiciones generales (ver Harvey 1976), los residuos estimados $e_j$ convergen en probabilidad a las $u_j$.

La matriz de covarianzas asintótica del vector **a** es $4.93(\mathbf{Z'Z})^{-1}$, aunque en nuestra opinión las varianzas de las $a_k$ en muestras pequeñas deben ser sensiblemente superiores a las dadas por la anterior expresión.

### 3.2. *Estimación máximo-verosímil*

Denominado $\mathbf{x}_j$ y $\mathbf{z}_j$ a los vectores que contienen los valores de las variables explicativas en la *j-ava* observación, se deduce inmediatamente de las hipótesis del modelo que el logaritmo de la función de verosimilitud tiene la siguiente expresión:

$$L_n(\mathrm{L}) = cte - 0.5(\Sigma \mathbf{z}_j'\alpha + \Sigma e^{-\mathbf{z}_j'\alpha}\,(y_j-\mathbf{x}_j'\mathbf{ß})^2) \qquad (3.2.)$$

donde los sumatorios se extienden para todas las observaciones ($j$=1...J)

La aplicación del método de Newton para la maximización de la anterior expresión conduce a un procedimiento iterativo en el que, como es sabido, si denominamos $\Gamma$ al

vector I+K dimensional cuyos elementos son los estimadores de lo $\beta_i$ y de las $\alpha_k$, su valor en la iteración $t+1$ viene dado por

$$\Gamma(t+1) = \Gamma(t) - \mathbf{Hes}^{-1}\mathbf{g}$$

donde **Hes** y **g** son respectivamente la matriz hessiana y el vector gradiente de $Ln(L)$ evaluados en la iteración $t$.

En la práctica, por su menor complejidad computacional, se utiliza una ligera modificación del procedimiento anterior, consistente en sustituir la matriz hessiana por su esperanza matemática que no es más que la matriz de información relativa a los parámetros del modelo.

Derivando (3.2.) se obtienen como componentes del vector gradiente:

$$\delta Ln(L) / \delta\beta_i = \Sigma e^{-z_j'\alpha} (y_j - \mathbf{x}_j'\beta) x_{ji} \tag{3.3.}$$

$$\delta Ln(L) / \delta\alpha_k = 0.5\Sigma z_{jk} (e^{-z_j'\alpha}(y_j - \mathbf{x}_j'\beta)^2 - 1) \tag{3.4.}$$

Derivando nuevamente y calculando la esperanza matemática de las expresiones resultantes se obtienen las siguientes expresiones para los elementos de la matriz de información:

$$E(\delta^2 Ln(L) / \delta\beta_i\delta\beta_l) = \Sigma e^{-z_j'\alpha} x_{ji} x_{jl} \tag{3.5.}$$

$$E(\delta^2 Ln(L) / \delta\alpha_k\delta\alpha_m) = 0{,}5 \Sigma z_{jk} z_{jm} \tag{3.6.}$$

$$E(\delta^2 Ln(L) / \delta\beta_i\delta\alpha_k) = 0$$

La estructura diagonal en bloques de la matriz de información (y en consecuencia de su inversa) implica que en este caso cada iteración del método de Newton se descomponga en dos fases, la primera de las cuales afecta a la reestimación de las $\beta_i$ y la segunda a la de las $\alpha_k$. Las expresiones correspondientes se obtienen con cierta sencillez si se tiene en cuenta que (3.6.) no es más que el término general de la matriz $0.5\mathbf{Z'Z}$ y que (3.5.) lo es de la matriz $\mathbf{X'D^{-1}X}$, donde **D** es la matriz diagonal J×J cuyo elemento $d_{jj}$ es $e^{-\mathbf{j}'\alpha}$. En consecuencia las ecuaciones cuya aplicación iterativa conducen a la obtención de los estimadores máximo verosímiles son:

$$\mathbf{b}(t+1) = \mathbf{b}(t) + (\mathbf{X'D^{-1}X})^{-1}\mathbf{g1} \tag{3.7.}$$

$$\mathbf{a}(t+1) = \mathbf{a}(t) + 2(\mathbf{Z'Z})^{-1}\mathbf{g2} \tag{3.8.}$$

donde **g1** y **g2** son los vectores I y K dimensionales cuyos elementos genéricos vienen dados por (3.3.) y (3.4.).

Como estimadores iniciales **b**(0) y **a**(0) Harvey propone utilizar el estimador TSP definido en el apartado anterior.

De acuerdo con las propiedades generales de la estimación máximo verosímil, la matriz de covarianzas asintótica de los estimadores no es más que la inversa de la matriz de información. En particular la matriz de covarianzas asintótica de **a** resulta ser $2(\mathbf{Z'Z})^{-1}$.

3.3. *Estimación por mínimos cuadrados ponderados iterados*

Un procedimiento alternativo de estimación que sólo exige la aplicación iterativa de una rutina ordinaria de mínimos cuadrados puede desarrollarse de acuerdo con el siguiente razonamiento.

Las hipótesis del modelo multiplicativo implican, para $j=1,...,J$

$$y_j = \Sigma\, \beta_i\, x_{ji} + u_j$$

donde las $u_j$ tienen media nula y varianzas $\sigma_j^2 = e^{\Sigma\, z\, k^2\, jk}$

En consecuencia $u_j^2$ se distribuirá como $\sigma_j^2 G$, siendo G una G$i$-2 con un grado de libertad, y su logaritmo neperiano.

$$Ln\,(u_j^2) = (\Sigma\, \alpha_k\, z_{jk}) + Ln\,(G) \qquad (3.8.)$$

tendrá por media y varianza

$$E\,(Ln\,(u_j^2)) = (\Sigma\, \alpha_k\, z_{jk}) + E\,(Ln\,(G)) = (\Sigma\, \alpha_k\, z_{jk}) - 1.27$$

$$\sigma^2\,(Ln\,(u_j^2) = \sigma^2\,(Ln\,(G)) = 4.93$$

al ser -1.27 y 4.93 el valor medio y la varianza del logaritmo de una G$i$-2 con un grado de libertad, como puede deducirse fácilmente por integración numérica.

Si las $u_j$ fueran conocidas, las $\alpha_k$ podrían estimarse a partir de una regresión de los $Ln(u_j^2)$ sobre las $z_{jk}$, rectificando la ordenada en el origen obtenida en el ajuste restándole E (L$n$ (G) ), o sea adicionándole 1.27. Los estimadores obtenidos de esta forma serían óptimos en el sentido de Markov dado que la ecuación (3.8) define un modelo lineal homocedástico.

El método TSP, desarrollado en el apartado 3.1., se basa en esta idea, pero sustituyendo el vector **u** desconocido por su estimador

$$\mathbf{e} = \mathbf{y} - \mathbf{Xb}$$

siendo **b** el estimador mínimo cuadrático ordinario de **ß**.

Sin embargo, dado que el modelo $\mathbf{y} = \mathbf{X\beta} + \mathbf{u}$ a partir del que se estima **b** es heterocedástico, siendo$\mathbf{V}=\mathbf{D}(e^{z\,z})$ la matriz de covarianzas de su vector de residuos **u**,

parece preferible estimarlo mediante mínimos cuadrados ponderados, utilizando alguna estimación **V*** de dicha matriz de covarianzas.

La idea más natural es utilizar $\mathbf{V}^*=\mathbf{D}(e^{\mathbf{Za}})$ siendo **a** una estimación previa de $\alpha$, lo que conduce a un proceso iterativo que puede esquematizarse en los siguientes pasos:

0 – Hacer inicialmente $t=0$ $\mathbf{b}(0)=(\mathbf{X'X})^{-1}\mathbf{X'y}$
1 – $t=t+1$ $\mathbf{e}=\mathbf{y}-\mathbf{Xb}(t-1)$
2 – $\mathbf{a}=(\mathbf{Z'Z})^{-1}\mathbf{Z'}Ln(\mathbf{e}^2)$ $a(0)=a(0)+1.27$
3 – $\mathbf{V}=\mathbf{D}(e^{\mathbf{Za}})$ $\mathbf{b}(t)=(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'V}^{-1}\mathbf{y}$
4 – Si $\| \mathbf{b}(t)-\mathbf{b}(t-1) \| < \delta$ ir a 5. En caso contrario ir a 1.
5 – Fin de la estimación.

La estimación de la matriz de covarianzas de **b** es $(\mathbf{X'V}^{-1}\mathbf{X})^{-1}$ siendo $\mathbf{V}=\mathbf{D}(e^{\mathbf{Xa}})$. Una estimación aproximada de la matriz de covarianzas de **a** puede obtenerse de la forma habitual, al ser este modelo asintóticamente homocedástico, o utilizar el valor teórico asintótico $4.93(\mathbf{Z'Z})^{-1}$, dado que por converger en probabilidad los $Ln(e_j^2)$ a $Ln(u_j^2)$ sus varianzas convergen a 4.93.

La relación entre el método de estimación propuesto y el de máxima verosimilitud es más estrecha de lo que pudiera parecerlo en un análisis superficial. En efecto puede comprobarse, tras unas transformaciones aritméticas sencillas, que la iteración sobre **b** indicada en el punto 3 del anterior esquema coincide con la primera de las dos que se realizan en la que se realizan en la estimación máximo verosimil y que se reflejó en (3.7.). La iteración sobre **a** es, sin embargo, diferente en un caso y en otro.

Una ventaja importante del método de estimación propuesto es que sólo exige realmente el recurso a una rutina standard de mínimos cuadrados ordinarios. La estimación por mínimos cuadrados ponderados que aparece en el paso 3 puede llevarse a cabo, en efecto, aplicando mínimos cuadrados ordinarios a las variables transformadas que resultan de dividir en cada observación las variables dependiente e independientes por la correspondiente desviación típica.

Es aconsejable, muy especialmente en el caso de que el número de observaciones sea escaso en relación al de parámetros a estimar, ir eliminando a lo largo del proceso de estimación de los dos submodelos (el de efectos de posición y el de dispersión) aquellas variables que aparezcan como claramente no significativas.

En la estimación del modelo mediante el algoritmo propuesto, puede presentarse el problema de que algunos de los $e_j$ estimados sean nulos, lo que no permite calcular el $Ln(e_j^2)$. Como regla heurística para abordar este problema proponemos sustituir los residuos nulos, en el caso de que existan, por la mitad del valor del menor valor absoluto de los residuos no nulos.

Con el fin de poner de manifiesto la sencillez final del algoritmo propuesto, se incluye en el Apéndice un programa APL para estimar mediante esta técnica los parámetros de

los submodelos de posición y dispersión. Como puede apreciarse el programa consta sólo de 6 instrucciones.

Señalemos por último que ya en la primera iteración, que coincide simplemente con el resultado del método TSP, es posible obtener una información valiosa sobre el orden de magnitud de los posibles efectos sobre dispersión, sin más que ajustar un modelo lineal ordinario utilizando como variable dependiente $Ln(\mathbf{e}^2)$, donde **e** es el vector de los residuos del modelo de regresión utilizado para estimar efectos sobre las medias. Dada la sencillez de este análisis, consideramos que es una práctica recomendable con carácter general.

## 4. UNA APLICACION: ESTUDIO DE DESAJUSTES EN UN PROCESO DE UNIDADES MULTIPLES

En el control de procesos industriales, es frecuente la necesidad de analizar si las diferentes unidades que están realizando en paralelo una determinada operación funcionan con la misma media y varianza. Este análisis puede ser complicado en el caso de que sean varias las operaciones que se verifican secuencialmente por distintos grupos de máquinas que trabajan en paralelo en cada una de ellas. El caso que se expone a continuación trata un problema de este tipo, que fue estudiado por los autores en una importante factoría de automóviles.

Entre las distintas operaciones que se realizan en la Planta de Carrocerías para el montaje de la caja de toma de aire, dos sucesivas, el montaje del panel exterior con un panel lateral (realizado en la estación A) y el montaje del panel exterior con el panel interior (realizado en la estación B) son críticas en la obtención de ciertas dimensiones relevantes para la funcionalidad final de la caja. Tanto la estación A como la B constan a su vez de 4 subestaciones, que numeraremos de 1 a 4, que realizan la operación correspondiente sobre 4 unidades, trabajando en paralelo. No existe correspondencia entre la subestación por la que pasa una unidad en la estación A y aquella por la que pasa en la estación B.

Los datos de la Tabla 1 recogen los valores de una determinada cota, a la que nos referiremos como X4, en 180 unidades fabricadas consecutivamente, indicando en cada caso la subestación por la que pasó dicha unidad en la estación A y en la estación B. Las cotas están medidas en desviaciones sobre el valor nominal.

La media de las cotas es -0.168 y la desviación típica 0.362.

TABLA I

VALORES DE X4 EN 180 OBSERVACIONES (MONTAJE DE LA TOMA DE AIRE)

| EST.B SUBEST. | EST.A SUBEST. | X4 | EST.B SUBEST. | EST.A SUBEST. | X4 | EST.B SUBEST. | EST.A SUBEST. | X4 |
|---|---|---|---|---|---|---|---|---|
| 1 | 3 | -.4 | 1 | 3 | -.8 | 1 | 4 | -.3 |
| 1 | 3 | -.4 | 1 | 3 | -.5 | 1 | 3 | .1 |
| 1 | 3 | -1.0 | 1 | 4 | -.6 | 1 | 4 | .3 |
| 1 | 1 | .1 | 1 | 4 | -.7 | 1 | 1 | .4 |
| 1 | 4 | -.4 | 1 | 2 | -.2 | 1 | 1 | .1 |
| 2 | 1 | -.2 | 2 | 3 | -.2 | 2 | 3 | -.3 |
| 2 | 4 | -.2 | 2 | 2 | -.2 | 2 | 1 | -.1 |
| 2 | 3 | -.2 | 2 | 1 | .1 | 2 | 1 | -.5 |
| 2 | 4 | -.2 | 2 | 3 | -.6 | 2 | 4 | -.3 |
| 2 | 4 | .0 | 2 | 2 | -.3 | 2 | 3 | -.4 |
| 3 | 2 | -.2 | 3 | 4 | -.2 | 3 | 3 | -.2 |
| 3 | 2 | -.4 | 3 | 3 | -.5 | 3 | 1 | .3 |
| 3 | 3 | -.3 | 3 | 2 | .3 | 3 | 4 | -.3 |
| 3 | 2 | -.3 | 3 | 1 | .7 | 3 | 1 | -.6 |
| 3 | 2 | .0 | 3 | 4 | -.2 | 3 | 2 | -.1 |
| 4 | 2 | -.1 | 4 | 1 | .6 | 4 | 1 | -.1 |
| 4 | 2 | .0 | 4 | 3 | -.3 | 4 | 4 | -.3 |
| 4 | 1 | -.2 | 4 | 2 | .0 | 4 | 4 | -.4 |
| 4 | 4 | -.1 | 4 | 4 | -.4 | 4 | 2 | -.2 |
| 4 | 2 | -.1 | 4 | 1 | .3 | 4 | 3 | -.2 |
| 1 | 1 | .0 | 1 | 1 | .1 | 1 | 4 | .3 |
| 1 | 2 | -.1 | 1 | 1 | .2 | 1 | 2 | .1 |
| 1 | 4 | -.4 | 1 | 1 | .1 | 1 | 4 | .7 |
| 1 | 1 | -.3 | 1 | 3 | -.2 | 1 | 1 | .3 |
| 1 | 3 | -1.0 | 1 | 4 | -.2 | 1 | 3 | .2 |
| 2 | 2 | .2 | 2 | 4 | .2 | 2 | 3 | -.4 |
| 2 | 2 | -.2 | 2 | 4 | .3 | 2 | 4 | -.5 |
| 2 | 4 | -.1 | 2 | 3 | .0 | 2 | 1 | .0 |
| 2 | 1 | .2 | 2 | 1 | .5 | 2 | 2 | -.1 |
| 2 | 3 | -.3 | 2 | 2 | -.2 | 2 | 1 | -.7 |
| 3 | 1 | -.2 | 3 | 3 | -.1 | 3 | 4 | -.4 |
| 3 | 3 | -.7 | 3 | 1 | .6 | 3 | 3 | -.6 |
| 3 | 4 | -.5 | 3 | 4 | -.7 | 3 | 3 | .0 |
| 3 | 2 | -.3 | 3 | 3 | -.5 | 3 | 3 | -.5 |
| 3 | 1 | -.1 | 3 | 2 | .3 | 3 | 2 | -.3 |
| 4 | 3 | .0 | 4 | 3 | .1 | 4 | 1 | -.1 |
| 4 | 3 | -.7 | 4 | 1 | .7 | 4 | 1 | -.8 |
| 4 | 1 | -.1 | 4 | 4 | .0 | 4 | 1 | -.4 |
| 4 | 2 | .2 | 4 | 3 | -.3 | 4 | 3 | -.2 |
| 4 | 1 | -.2 | 4 | 4 | -.4 | 4 | 4 | -.5 |
| 1 | 4 | -1.0 | 1 | 1 | -.5 | 1 | 2 | .7 |
| 1 | 4 | -.5 | 1 | 4 | -.1 | 1 | 3 | .4 |
| 1 | 4 | -.2 | 1 | 4 | -.2 | 1 | 4 | .5 |
| 1 | 4 | -1.0 | 1 | 2 | .5 | 1 | 1 | .6 |
| 1 | 1 | -.4 | 1 | 2 | .5 | 1 | 1 | .4 |
| 2 | 2 | .1 | 2 | 2 | -.2 | 2 | 4 | -.2 |
| 2 | 4 | .0 | 2 | 3 | -.6 | 2 | 4 | -.2 |
| 2 | 2 | -.6 | 2 | 3 | -.5 | 2 | 3 | .1 |
| 2 | 1 | .0 | 2 | 3 | -.6 | 2 | 3 | .1 |
| 2 | 3 | -.8 | 2 | 4 | .1 | 2 | 2 | .1 |
| 3 | 3 | .4 | 3 | 2 | -.6 | 3 | 3 | -.5 |
| 3 | 2 | .2 | 3 | 1 | -.5 | 3 | 4 | .3 |
| 3 | 4 | -.6 | 3 | 1 | -.2 | 3 | 1 | .0 |
| 3 | 1 | .0 | 3 | 1 | -.2 | 3 | 2 | -.4 |
| 3 | 3 | -.5 | 3 | 3 | -.5 | 3 | 1 | .2 |
| 4 | 4 | .3 | 4 | 4 | -.1 | 4 | 4 | .0 |
| 4 | 1 | .0 | 4 | 3 | -.5 | 4 | 1 | .3 |
| 4 | 3 | -.7 | 4 | 1 | -.5 | 4 | 4 | -.5 |
| 4 | 2 | .1 | 4 | 1 | -.2 | 4 | 2 | -.1 |
| 4 | 4 | -.5 | 4 | 2 | .0 | 4 | 3 | -.4 |

Con el fin de centrar el proceso y reducir en lo posible su variabilidad se analizó si existía algún efecto de las diferentes subestaciones, en una y otra estación, sobre la media y varianza de la cota estudiada.

Se definieron en principio 3 variables dummy (XA1, XA2 y XA3) asociadas a las tres primeras subestaciones de la estación A, y otras 3 variables dummy (XB1, XB2 y XB3) para las correspondientes de la estación B. Como es habitual en este tipo de parametrizaciones, los parámetros asociados a dichas variables miden la diferencia entre el efecto de la subestación correspondiente y el de la subestación 4 de la misma estación, que es aquella cuyo efecto no se ha incluído explícitamente en el modelo.

La Tabla 2 recoge los resultados obtenidos en los ajustes iniciales realizados para identificar las variables con efectos relevantes sobre la media de X4.

En el modelo completo ninguno de los parámetros asociados a la estación B resulta significativo, mientras que en la estación A las subestaciones 1 y 2 difieren significativamente de la 4, no llegando a ser significativo el efecto de la 3.

El análisis del signo y orden de magnitud de la ordenada en el origen, que en este caso no es más que el descentrado promedio de las unidades que pasan por la subestación 4 en ambas estaciones, sugiere que un modelo adecuado a los datos y mucho más sencillo puede obtenerse postulando que las únicas variables con efecto sobre X4 son XA3 y XA4, y que la ordenada en el origen (que en este nuevo modelo corresponde al descentrado promedio de las unidades que no pasan por estas subestaciones) es nula.

La Tabla 2 recoge también el resultado del ajuste de este nuevo modelo. Puede comprobarse que, pese a que el número de parámetros se reduce de 7 a 2, la suma de cuadrados residual apenas aumenta pasando de 19.521 a 19.984, incremento que no resulta significativo.

Como conclusión de esta fase inicial de identificación del modelo de efectos de posición, se decidió retener únicamente las variables XA3 y XA4 y estimar los parámetros correspondientes en un modelo con ordenada en el origen nula.

TABLA 2

TOMA DE AIRE. ESTIMACION INICIAL EFECTOS DE POSICION

AJUSTE: Modelo completo

| VARIABLE | COEFICIENTE | DESV. TIPICA | VALOR DE T |
|---|---|---|---|
| ORD. ORIG. | -.256 | .065 | -3.912 |
| XA1 | .231 | .069 | 3.349 |
| XA2 | .192 | .074 | 2.583 |
| XA3 | -.115 | .069 | -1.656 |
| XB1 | .082 | .071 | 1.159 |
| XB2 | .023 | .071 | .319 |
| XB3 | -.039 | .071 | -.553 |

SUMA CUADRADOS RESIDUAL 19.521 G.L. 173

AJUSTE: Modelo simplificado

| VARIABLE | COEFICIENTE | DESV. TIPICA | VALOR DE T |
|---|---|---|---|
| XA3 | -.357 | .049 | -7.314 |
| XA4 | -.229 | .048 | -4.738 |

SUMA CUADRADOS RESIDUAL 19.984 G.L. 178

TABLA 3

TOMA DE AIRE. ESTIMACION INICIAL EFECTOS DE DISPERSION

AJUSTE: Modelo completo

| VARIABLE | COEFICIENTE | DESV. TIPICA | VALOR DE T |
|---|---|---|---|
| ORD. ORIG. | -2.658 | .404 | -6.591 |
| XA1 | .295 | .425 | .693 |
| XA2 | .027 | .460 | .059 |
| XA3 | .193 | .429 | .449 |
| XB1 | .824 | .439 | 1.876 |
| XB2 | -.006 | .440 | .014 |
| XB3 | .312 | .440 | .709 |

SUMA CUADRADOS RESIDUAL 745.049 G.L. 173

AJUSTE: Modelo simplificado

| VARIABLE | COEFICIENTE | DESV. TIPICA | VALOR DE T |
|---|---|---|---|
| ORD. ORIG. | -2.421 | .177 | -13.693 |
| XB1 | .718 | .354 | -2.030 |

SUMA CUADRADOS RESIDUAL 750.887 G.L. 178

La Tabla 3 recoge los resultados obtenidos en los ajustes iniciales realizados mediante el procedimiento TSP para la identificación del modelo de efectos sobre la dispersión. El primer ajuste pone de manifiesto la existencia de un posible efecto de dispersión asociado a la subestación 1 de la estación B. La suma de cuadrados residual del ajuste del modelo simplificado, que retiene el efecto de esta variable y la ordenada en el origen como únicos parámetros, no difiere significativamente de la del modelo completo (751 frente a 745) por lo que es éste el modelo retenido con el fin de estimar efectos sobre la dispersión.

La Tabla 4 recoge los resultados obtenidos en la estimación de los parámetros del modelo mediante los tres procedimientos expuestos: TSP, máxima verosimilitud (MV) y mínimos cuadrados ponderados iterados (MCPI). En los dos últimos métodos se ha exigido para la convergencia una discrepancia inferior a 0.001, que resulta suficiente a efectos prácticos.

TABLA 4

RESULTADOS DE LA ESTIMACION DE LOS EFECTOS DE POSICION Y DISPERSION (TOMA DE AIRE)

| | Efectos de posición | | Efectos de dispersión | |
|---|---|---|---|---|
| Método | $b_{A3}$ | $b_{A4}$ | $a_0$ | $a_{B1}$ |
| TSP | -0.357 | -0.229 | -2.421 | 0.718 |
| MCPI | -0.357 | -0.227 | -2.433 | 0.721 |
| MV | -0.357 | -0.227 | -2.440 | 0.741 |
| Desv. tip.* | (0.046) | (0.047) | (0.122) | (0.243) |

* Valores asintóticos para los estimadores MV

Como puede constatarse los tres métodos conducen en este caso a estimaciones muy similares, lo que en cierto sentido era previsible dado el tamaño de la muestra y la consistencia de los tres estimadores. Los resultados obtenidos mediante el método MCPI son, en cualquier caso, más próximos a los máximo verosímiles que los del método TSP. Los cuatro parámetros difieren de forma significativa de cero para $\alpha$=0.05, como se deduce utilizando las expresiones asintóticas para las varianzas de los estimadores MV vistas en 3.2.

Las varianzas del proceso, según las piezas pasen o no por la subestación B1, pueden estimarse a partir de los valores de las $a_j$ recogidos en la Tabla 4 obteniéndose los siguientes resultados:

$$\sigma^2(\text{XB1} = 0) = e^{-2.440} = 0.087$$

$$\sigma^2(\text{XB1} = 1) = e^{-2.440 + 0.741} = 0.183$$

Se deduce, por tanto, que la varianza de la característica analizada es más del doble cuando las piezas pasan por la subestación 1 de B que cuando lo hacen por las otras

tres subestaciones. La conclusión práctica de este resultado es la necesidad de revisar dicha subestación con el fin de corregir el problema.

La corrección de los desajustes de posición y dispersión detectados mediante el modelo, permite por una parte centrar el proceso y por otra reducir en un 20 % su desviación típica, mejoras ambas que redundan en un incremento sensible de la capacidad del mismo.

## 5. UNA APLICACION AL ANALISIS DE UNA FRACCION FACTORIAL

Para ilustrar la aplicabilidad del modelo al estudio de muestras pequeñas y en ausencia de replicaciones, hemos analizado mediante el mismo unos datos procedentes de un trabajo original de Taguchi y Wu (1980) y posteriormente reanalizados por Box y Meyer (1986).

El objetivo del estudio, efectuado para la National Railway Corporation de Japón, era el análisis del posible efecto de nueve factores, que identificaremos con las letras A a I, sobre la resistencia de una soldadura. Todos los factores se estudiaron a dos niveles (que etiquetamos con los signos - y +), siendo el diseño utilizado una fracción factorial $2^{9-5}$ que permite la estimación separada de los nueve efectos simples. La Tabla 5 recoge el diseño utilizado, así como los resultados observados en las dieciseis pruebas.

Analizando estos resultados Box y Meyer constatan que sólo los factores B y C tienen un efecto significativo sobre el valor medio de la resistencia. En consecuencia analizan los dieciseis resultados como si constituyeran un plan factorial con sólo dos factores (B y C) y 4 replicaciones. La existencia de estas seudo-replicaciones permite a los autores mencionados poner de manifiesto un efecto importante sobre la dispersión del factor C.

Un inconveniente de esta forma de operar radica en que no permite detectar posibles efectos sobre la dispersión de los factores que no afectan a la media. Este inconveniente es sensible, dado que este tipo de factores, que permiten disminuir la variabilidad sin modificar la media, pueden resultar de gran importancia en la mejora de un proceso.

TABLA 5

RESULTADOS FRACCION FACTORIAL PARA ANALIZAR EL EFECTO DE 9 FACTORES SOBRE LA RESISTENCIA DE UNA SOLDADURA

| | Factor | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Prueba | A | B | C | D | E | F | G | H | I | Resistencia |
| 1 | − | − | + | − | + | + | − | − | − | 43.7 |
| 2 | − | − | − | + | − | − | − | − | + | 40.2 |
| 3 | − | + | − | − | + | + | − | + | − | 42.4 |
| 4 | − | + | + | + | − | − | − | + | + | 44.7 |
| 5 | − | + | − | − | + | − | + | − | + | 42.4 |
| 6 | − | + | + | + | − | + | + | − | − | 45.9 |
| 7 | − | − | + | − | + | − | + | + | + | 42.2 |
| 8 | − | − | − | + | − | + | + | + | − | 40.6 |
| 9 | + | + | − | − | − | + | − | − | + | 42.4 |
| 10 | + | + | + | + | + | − | − | − | − | 45.5 |
| 11 | + | − | + | − | − | + | − | + | + | 43.6 |
| 12 | + | − | − | + | + | − | − | + | − | 40.6 |
| 13 | + | − | + | − | − | − | + | − | − | 44.0 |
| 14 | + | − | − | + | + | + | + | − | + | 40.2 |
| 15 | + | + | − | − | − | − | + | + | − | 42.5 |
| 16 | + | + | + | + | + | + | + | + | + | 46.5 |

Fuente: Box y Meyer (1986)

La Tabla 6 recoge los resultados de nuestro análisis inicial para identificar los efectos de posición. Obviamente se reencuentran los mismos resultados obtenidos por Box y Meyer, apareciendo los factores B y C como los únicos con una influencia sensible al respecto.

Utilizando como variable dependiente los logaritmos de los cuadrados de los residuos del modelo anterior, se identifican en la Tabla 7 los factores con posibles efectos sobre la dispersión. En concordancia con los resultados de Box y Meyer el factor C aparece como el más importante al respecto; adicionalmente otros factores, como A, B, D e I, parecen también tener un cierto efecto por lo que resulta prudente considerarlos como candidatos de cara a la estimación por los métodos de máxima verosimilitud o mínimos cuadrados ponderados iterados.

La Tabla 8 recoge los resultados de la estimación de los parámetros del modelo mediante estos métodos.

TABLA 6

DATOS BOX-MEYER. ESTIMACION INICIAL EFECTOS DE POSICION

AJUSTE: Modelo completo

| VARIABLE | COEFICIENTE | DESV. TIPICA | VALOR DE T |
|---|---|---|---|
| ORD. ORIG. | 42.963 | .136 | 316.178 |
| A | .200 | .136 | 1.472 |
| B | 1.075 | 136 | 7.911 |
| C | 1.550 | .136 | 11.407 |
| D | .063 | .136 | .460 |
| E | -.025 | .136 | -.184 |
| F | .200 | .136 | 1.472 |
| G | .075 | .136 | .552 |
| H | -.075 | .136 | -.552 |
| I | -.187 | .136 | -1.380 |

SUMA CUADRADOS RESIDUAL 1.772 G.L. 6

AJUSTE: Modelo simplificado

| VARIABLE | COEFICIENTE | DESV. TIPICA | VALOR DE T |
|---|---|---|---|
| ORD. ORIG. | 42.963 | .136 | 315.069 |
| B | 1.075 | .136 | 7.884 |
| C | 1.550 | .136 | 11.367 |

SUMA CUADRADOS RESIDUAL 3.867 G.L. 13

TABLA 7

DATOS BOX-MEYER. ESTIMACION INICIAL EFECTOS DE DISPERSION

AJUSTE: Modelo completo

| VARIABLE | COEFICIENTE | DESV. TIPICA | VALOR DE T |
|---|---|---|---|
| ORD. ORIG. | -1.938 | .490 | -3.953 |
| A | -.557 | .490 | -1.137 |
| B | -.668 | .490 | -1.362 |
| C | 1.376 | .490 | 2.807 |
| D | .594 | .490 | 1.212 |
| E | .246 | .490 | .502 |
| F | .057 | .490 | .116 |
| G | .268 | .490 | .548 |
| H | .381 | .490 | .776 |
| I | 544 | .490 | 1.109 |

SUMA CUADRADOS RESIDUAL 23.064 G.L. 6

AJUSTE: Modelo simplificado

| VARIABLE | COEFICIENTE | DESV. TIPICA | VALOR DE T |
|---|---|---|---|
| ORD. ORIG. | -1.938 | .415 | -4.669 |
| A | -.557 | .415 | -1.343 |
| B | -.668 | .415 | -1.609 |
| C | 1.376 | .415 | 3.315 |
| D | .594 | .415 | 1.432 |
| I | .544 | .415 | 1.310 |

SUMA CUADRADOS RESIDUAL 27.555 G.L. 10

TABLA 8

RESULTADOS DE LA ESTIMACION DE LOS EFECTOS DE POSICION Y DISPERSION (DATOS BOX-MEYER)

| Parámetro | TSP | MCPI | MV | Desv. Típica * |
|---|---|---|---|---|
| $b0$ | 42.963 | 43.147 | 43.143 | 0.045 |
| $b$B | 1.075 | 0.937 | 0.935 | 0.021 |
| $b$C | 1.550 | 1.623 | 1.618 | 0.045 |
| $a0$ | -1.938 | -2.528 | -3.167 | 0.354 |
| $a$A | -0.557 | -0.291 | -0.266 | 0.354 |
| $a$B | -0.668 | -0.027 | -0.159 | 0.354 |
| $a$C | 1.376 | 1.744 | 1.427 | 0.354 |
| $a$D | 0.594 | 0.304 | 0.213 | 0.354 |
| $a$I | 0.544 | 1.625 | 1.700 | 0.354 |

* Valores asintóticos para los estimadores MV

Como se aprecia los estimadores MCPI resultan muy próximos a los MV, difiriendo ambos sensiblemente en algunos casos de los estimadores iniciales TSP. En particular el análisis de los valores estimados para las $a_k$ tanto por el método MCPI como por el MV detectan, además del efecto ya anunciado del factor C, un efecto altamente significativo sobre la dispersión del factor I, efecto no identificado en estudios anteriores sobre estos datos y que tampoco había puesto de manifiesto el estimador más sencillo TSP utilizado en la fase inicial.

## 6. CONSIDERACIONES FINALES

Diversas cuestiones relacionadas con el modelo y las técnicas desarrolladas en el presente trabajo deben ser objeto de investigaciones adicionales.

En particular consideramos necesario profundizar en el conocimiento de las propiedades de los distintos estimadores cuando los tamaños muestrales son reducidos. En realidad los principales resultados obtenidos por el momento son de carácter asintótico y su validez, incluso como aproximaciones, puede resultar muy discutible en la práctica industrial, contexto en el que consideraciones económicas y organizativas acostumbran a limitar el número de observaciones disponibles.

Los resultados obtenidos en lo dos ejemplos analizados parecen indicar que, si bien la elección entre uno u otro método de estimación no es muy relevante con muestras grandes, las conclusiones obtenidas pueden diferir sensiblemente según el método utilizado cuando el número de datos es reducido. En estas condiciones parece manifestarse una clara superioridad de los estimadores MCPI sobre los TSP, al menos desde el punto de vista de su proximidad a los máximo verosímiles.

También creemos necesario estudiar la robustez de los estimadores propuestos, especialmente frente a la existencia de "outliers". Es bien sabido, en efecto, que los tests respecto a varianzas resultan bastante afectados por el carácter leptocúrtico de los datos analizados. En consecuencia puede resultar interesante desarrollar métodos alternativos que, explotando la misma idea básica, tengan mejor comportamiento en estas situaciones.

La posibilidad de ampliar el modelo con el fin de estudiar efectos sobre los ratios señal/ruido constituye también un tema de gran interés potencial.

A pesar de las consideraciones anteriores queremos precisar que el modelo y las técnicas expuestas en el presente trabajo constituyen en nuestra opinión algo más que una propuesta tentativa en fase de elaboración. Los autores los hemos aplicado a numerosos datos reales, generalmente en el campo de la mejora de la calidad y de la productividad en la industria, habiendo constituído en todos los casos una herramienta sencilla y poderosa para identificar y estimar efectos de interés sobre medias y sobre

varianzas. También hemos reanalizado con estas técnicas los datos de algún trabajo anterior, como es el caso del ejemplo estudiado por Box y Meyer y recogido en el presente artículo, encontrando de forma sencilla y directa resultados similares a los hallados en los mismos por sus autores mediante métodos menos generales.

Consideramos en definitiva que, al margen de que siga siendo necesario profundizar en el estudio de sus limitaciones y potencialidades, el modelo y los métodos analizados en el presente artículo constituyen ya hoy en día una valiosa herramienta para el control y la mejora de la calidad y de la productividad.

## AGRADECIMIENTO

Queremos manifestar nuestro agradecimiento a Ford España S. A., empresa con la que nuestra Universidad mantiene un importante convenio de colaboración en cuyo contexto surgieron algunas de las motivaciones e ideas que inspiraron este estudio.

Los autores también desean hacer patente su reconocimiento al profesor del Departamento D. José Miguel Almenar, por sus valiosas sugerencias a lo largo del presente trabajo.

## APENDICE: PROGRAMACION APL DEL METODO DE ESTIMACION MCPI

El programa adjunto obtiene los estimadores de los efectos de posición y dispersión por el método de Mínimos Cuadrados Ponderados Iterados. Su objetivo es exclusivamente ilustrar el algoritmo de estimación expuesto en 3.3.; en consecuencia apenas se han considerado las cuestiones relativas a la entrada de datos y salida de resultados.

Se asume en el programa que el vector Y, que contiene los valores de la variable dependiente, es una variable externa. Los dos argumentos X y Z son las matrices, posiblemente con una primera columna de unos, que contienen las variables explicativas de los efectos de posición y de dispersión.

## BIBLIOGRAFIA

AMEMIYA T. (1977). *A note on a heterocedastic model* Journal of Econometrics, 6, 365-370.

BOX G. E. P. y MEYER R. D. (1986). *Studies in quality improvement: Dispersion effects from fractional designs.* Center for Quality and Productivity improvement. University of Winsconsin.

BOX G. E. P. (1986). *Studies in quality improvement: signal to noise ratios, performance criteria and statistical analysis: Part I* Center for Quality and Productivity Improvement. University of Winsconsin.

BOX G. E. P. y RAMIREZ J. (1986). *Estudies in quality improvement: signal to noise ratios, performance criteria and statistical analysis: Part II* Center for Quality and Productivity Improvement. University of Winsconsin.

HARVEY A. C. (1976). *Estimating regression models with multiplicative heterocedasticity* Econometrica, Vol 44, n.º 3, 461-465.

TAGUCHI G. y YUIN WU (1979). *Introduction to off-line quality control Central Japan Quality Control Association.*

*TORT-MARTORELL J. (1985). Diseños factoriales fraccionales. Aplicación al control de calidad mediante el diseño de productos y procesos.* Tesis doctoral. Universidad Politécnica de Barcelona

## SUMMARY

### A MODEL FOR DISPERSION EFECTS ANALYSIS FROM UNREPLICATED DATA

Japanese contributions in the field of quality and productivity improvement have originated a great interest on the statistical analysis of dispersión effects. Most techniques proposed require the availability of replicated data for the differents conditions; this requirement makes difficult the analysis from non experimental data, specially if some of the variables investigated are continuous. The aim of this paper is to study a model that make possible the analysis of dispersion effects in a very wide range of situations, and to propose a simple method to estimate their parameters. This model and the technique are applied to the analysis of two real cases, one coming from de writer's personal experience and other previously studied for Taguchi and Yu and reanalysed for Box and Meyer.

*Key words:* Dispersion effects. Generalized linear models. Parametric heterocedasticity. Quality and productivity improvement. Off-line control.