UNIVERSIDAD DE SEVILLA

FACULTAD DE MATEMATICAS

ANALISIS DE INFLUENCIA EN EL MODELO LINEAL GENERAL: SESGO CONDICIONADO

Memoria dirigida por:
Prof. Dr. D. Joaquín Muñoz García
Prof. Dr. D. Rafael Infante Macías

Memoria presentada por:
Juan M. Muñoz Pichardo

UNIVERSIDAD DE SEVILLA

FACULTAD DE MATEMATICAS

ANALISIS DE INFLUENCIA EN EL MODELO LINEAL GENERAL: SESGO CONDICIONADO

Visado en Sevilla a 15 de Julio de 1992

Memoria dirigida por:
Prof. Dr. D. Joaquín Muñoz García
Prof. Dr. D. Rafael Infante Macías

Memoria presentada
para optar al Grado de
Doctor en Matemáticas

Fdo. Juan M. Muñoz Pichardo

A mis padres,
por su esfuerzo.
A María Dolores y Marta,
por su apoyo.

Quiero expresar mi más sincero agradecimiento al Departamento de Estadística e Investigación Operativa, por despertar en mí el interés por la investigación, y en especial al Prof. Dr. D. Rafael Infante Macías por su inestimable labor de ilusión y ayuda, al Prof. Dr. D. Joaquín Muñoz García, a quién debo gran parte de mi formación, y al Prof. Dr. D. Juan Luis Moreno Rebollo, por su constante apoyo moral y científico para la realización de este trabajo.

Sevilla, Julio de 1992

I N D I C E

INTRODUCCION

El objetivo de todo análisis estadístico es obtener conclusiones fiables a partir de los datos resultantes de una experimentación. La fiabilidad de las conclusiones del proceso es de especial interés, ya que el análisis se realiza sobre codificaciones del fenómeno natural en estudio y las técnicas estadísticas que se apliquen pueden verse fuertemente afectadas por algunas de las observaciones realizadas.

Este problema ha provocado que diversos autores hayan desarrollado métodos enfocados bien al desarrollo de nuevas técnicas que no se vean influenciadas excesivamente por la modelización del fenómeno natural, bien al análisis de la calidad de los datos, o bien al estudio de aquellas observaciones que afectan considerablemente a los resultados del análisis.

En este último aspecto es donde se sitúa el Análisis de Observaciones Influencia, el cual ha sido fundamentalmente desarrollado en las dos últimas décadas.

El objetivo de este trabajo se centra en la obtención de diversas medidas de influencia en el Modelo Lineal General como base para abordar de una manera unificada los distintos modelos lineales que pueden ser considerados como casos particulares de éste.

En el Capítulo I se plantea la problemática de las observaciones influencia, analizándose las diversas definiciones que aparecen en la literatura, destacándose los rasgos comunes presentes en las mismas.

Asímismo, se presenta distintas metodologías para abordar el problema de la influencia y se recogen diversos resultados en el Modelo de Regresión Lineal Múltiple, por el hecho de ser el más tratado desde el punto de vista de la influencia y servir los métodos aplicados en el mismo como pautas para el estudio en otros modelos.

En el capítulo II se obtiene una descomposición de los estimadores del Modelo Lineal General y se analizan los estimadores resultantes al provocar algunas perturbaciones en dicho modelo, como base para el análisis de la influencia que se realiza en el capítulo siguiente.

En el capítulo III se presenta un nuevo enfoque para el estudio de la influencia, basado en el sesgo que sobre un estadístico provocan las observaciones consideradas individual o conjuntamente. Este planteamiento se justifica relacionándolo con las descomposiciones obtenidas en el capítulo anterior.

Con este enfoque se analiza la influencia en el Modelo Lineal General, y se trasladan los resultados obtenidos a los modelos de ANOVA de un factor y de Regresión Lineal Múltiple, observando que algunas de las medidas que se obtienen para este último son semejantes a medidas de influencia ya propuestas por otros autores.

C A P I T U L O I

OBSERVACIONES INFLUENCIA:
CONCEPTO Y TECNICAS DE ANALISIS

CAPITULO I : OBSERVACIONES INFLUENCIA: CONCEPTOS Y TECNICAS

Con uno de los problemas que generalmente se enfrenta el estadístico al realizar un estudio es, como se ha indicado en la introducción, el analizar el comportamiento de las observaciones frente al modelo estadístico, así como la fiabilidad y precisión de las mismas.

Este análisis puede realizarse, entre otros, bajo los dos enfoques siguientes:

1.- La identificación de aquellas observaciones que pueden considerarse en algún sentido erróneas al desviarse marcadamente del comportamiento del resto. ANALISIS DE OUTLIERS.

2.- La identificación de aquellas observaciones que tengan un efecto considerable sobre los resultados de las técnicas estadísticas que se aplican en el estudio. ANALISIS DE OBSERVACIONES INFLUENCIA.

En este capítulo se recogen los rasgos principales que caracterizan las distintas definiciones que de observación influencia han dado diversos autores, indicándose los inconvenientes que presentan. Del análisis de estos rasgos y de los factores que intervienen en el concepto se concluye la imposibilidad de dar una definición del mismo sin ambigüedad, por lo que el análisis de influencia debe tener cómo objetivo el proporcionar medidas de la influencia que permitan la ordenación de las observaciones en función de las misma, y no considerarlo como técnicas de identificación.

Por último, se realiza un estudio general de los distintos métodos existentes para abordar el problema de la influencia, recogiendo los resultados más significativos obtenidos para el modelo de Regresión Lineal Múltiple por ser en éste donde más se han desarrollado.

1.1.- EL PROBLEMA DE LAS OBSERVACIONES INFLUENCIA

Un gran número de autores han presentado situaciones prácticas en las que existen observaciones experimentales que inciden considerablemente en los resultados del análisis, motivando la necesidad de identificar tales observaciones, denominadas en la literatura observaciones influencia, y evaluar sus efectos en el estudio estadístico que se pretende realizar.

Además, siguiendo a Cook y Weisberg (1982), puede decirse, que el análisis de los datos experimentales, con objeto de encontrar estos casos relevantes, es de gran interés para las conclusiones que se obtengan de la experiencia por dos motivos:

1.- Da información referida a la fiabilidad de las conclusiones y resultados obtenidos.

2.- Puede indicar áreas del espacio muestral con efecto informativo inadecuado para una inferencia fiable y estable.

No obstante, a pesar del gran número de autores que han estudiado este problema en distintos modelos estadísticos, y del número de técnicas propuestas para la identificación de estas observaciones, no existe una definición objetiva de las mismas, quizás motivado por los numerosos factores que intervienen en el problema, y por la dificultad de clasificar los efectos de las observaciones sin utilizar un criterio en el que intervenga la subjetividad del experimentador.

En este apartado se intenta una aproximación al concepto de *observación influencia*, analizando las distintas definiciones que aparecen en la literatura.

Johnson y Kotz (1982) recogen:

> « *Las observaciones son consideradas como influencia si su omisión de los datos da lugar a cambios sustanciales en rasgos importantes de un análisis* »

En este mismo sentido, Gray (1988), en su trabajo sobre la clasificación de medidas de influencia, redunda en esta cuestión al referirse a las observaciones influencia en regresión:

> « *Una observación se dice influencia si uno o más estadísticos de la regresión son alterados sustancialmente por la omisión del caso en cuestión* en el *análisis* »

Ambas definiciones utilizan el caracter de la omisión de la observación, cuando éste debe ser considerado cómo un método para cuantificar el efecto que dicha observación tiene en los estadísticos.

En la definición dada por Belsley y otros (1980) se observan matices distintos, pues ya no se habla de una observación individual, sino que apunta la posibilidad de una influencia conjunta entre varias observaciones, y considera el efecto sobre los resultados en relación con el provocado por el resto de observaciones:

> « *Una observación influencia es aquélla que, bien individualmente o bien conjuntamente con otras, tiene un mayor impacto que las otras observaciones sobre los valores aislados de varias estimaciones.* »

Esta definición, aún siendo probablemente de las que más se acercan al concepto, presenta ambigüedad por la no concreción de la influencia sobre un rasgo determinado del análisis y por la falta de un criterio claro sobre qué se debe considerar por "*un mayor impacto*". Es, por tanto, una definición que no evita la subjetividad del experimentador, como recogen Hossaim y Naik (1989).

El resto de referencias que sobre el concepto de observación influencia recogen los distintos autores que han estudiado el problema, presentan alguna de las ambigüedades señaladas.

En base a las consideraciones realizadas anteriormente, en el concepto de observación influencia se deben tener en cuenta los siguientes aspectos:

1.- <u>¿Influencia sobre qué?</u>, puesto que una observación puede ejercer un fuerte impacto sobre algunos rasgos del análisis estadístico y no ejercerlo sobre otros. Por ello se deberá fijar previamente aquellos estadísticos, estimadores..., sobre los cuales se pretenda estudiar la influencia.

2.- <u>El impacto o efecto sobre los resultados considerado en relación con el resto de observaciones</u>, dado que, como indican Beckman y Cook (1983)

> « ... *la fiabilidad probable de una observación se refleja mediante su relación con otras observaciones que se obtienen bajo condiciones similares.* »

3.- <u>El caracter de influencia individual o conjunta</u>, dado que una observación puede tener un fuerte impacto considerada conjuntamente con otras. Johnson y Kotz recogen (1982):

> «*Las observaciones pueden ser influyentes individual o conjuntamente con una o más observaciones. Sin embargo, no se da siempre el caso de que observaciones conjuntamemte influyentes sean también individualemte influyentes.*»

Y se podría también añadir el caso contrario, es decir, una observación individualmente influyente puede no serlo al considerarla conjuntamente con otras.

No obstante, aunque se especificase en la definición el rasgo estadístico sobre el que se estudia la influencia y el caracter individual o conjunto de la misma, y se pueda medir el impacto o incidencia de cada observación sobre el rasgo en estudio, siempre quedaría por fijar un criterio que, en función del efecto, determinase la consideración o no de observación influencia.

Por tanto, se puede concluir la imposibilidad de dar una definición concreta y explícita del concepto de observación influencia, aunque sí se pueda hablar de observaciones más influyentes que otras sobre un rasgo determinado del análisis. Y en consecuencia, no se deberá utilizar el término de métodos de identificación de observaciones influencia, sino de medidas de influencia.

Sin embargo, aún queda otra cuestión por resolver: ¿cómo se mide el impacto?. En este sentido Cook y Weisberg (1982) recogen:

> « *La idea básica en el análisis de influencia es muy simple. Introducimos pequeñas perturbaciones en la formulación del problema y entonces calculamos cuánto cambian los resultados del análisis por la perturbación.* »

Por lo tanto, para medir el efecto que sobre un aspecto de interés del análisis tiene una observación o un conjunto de ellas, se introducen pequeñas perturbaciones y se cuantifica el cambio producido. Así aparece un nuevo factor de ambigüedad, la <u>perturbación</u>, que se suma a los anteriormente comentados.

En la literatura, el esquema de perturbación más extendido es el de la omisión de las observaciones a las que se le pretende estudiar su influencia, cuantificando la diferencia entre los resultados obtenidos para el modelo postulado inicialmente y el modelo perturbado.

Cook (1987), teniendo en cuenta las consideraciones sobre el análisis de influencia anteriormente expuestas, trata de unificar el problema bajo una formulación general válida para los distintos planteamientos y enfoques realizados sobre él.

Sea un conjunto de datos D, un modelo M postulado a priori, y R(D,M) un resultado seleccionado de una síntesis de los datos y el modelo M, y sea $\underline{\omega}$ un vector de perturbaciones perteneciente a un conjunto Ω de perturbaciones relevantes y $M(\underline{\omega})$ el modelo perturbado, de forma que

$$\exists\ \underline{\omega}_0 \in \Omega \quad / \quad M \equiv M(\underline{\omega}_0)$$

Entonces el análisis de influencia consiste en comparar $R[D,M(\underline{\omega})]$ y R(D,M), variando $\underline{\omega}$ sobre el conjunto de perturbaciones Ω, para lo que es necesario elegir el esquema de perturbación y el método de comparación de resultados.

Se apuntan como cuestiones claves le elección del esquema de perturbación y el método de comparación. Fijados ambos, el conjunto de observaciones puede ser ordenado en base a su influencia, es decir, en función de la desviación entre los dos resultados.

Por lo tanto, se tendrá que hablar de observaciones (o conjuntos de observaciones) más o menos influyentes que otras sobre un resultado de un análisis estadístico, bajo un esquema de perturbación y un método de comparación determinados. Quedando, pues, en manos del experimentador, después de valorar los resultados del análisis, la decisión de considerar o no una observación como influencia.

Sin embargo, siempre se pueden interpretar dichos resultados y extraer conclusiones tanto sobre la fiabilidad del modelo, como sobre el comportamiento de algunas zonas del espacio muestral para otros posibles análisis.

1.2.- TECNICAS DE ANALISIS DE INFLUENCIA

El desarrollo de técnicas de análisis de influencia ha tenido un gran auge en la última década. La mayoría de ellas bajo el esquema de perturbación de la omisión de aquellas observaciones a las que se le pretende estudiar su impacto sobre algún resultado del análisis, y eligiendo un método de comparación en base a dicho resultado.

No obstante, la formulación del problema de la influencia realizado por Cook (1987), ha permitido nuevos enfoques y nuevos métodos de cuantificación de la relevancia de una observación en la determinación del estadístico objeto de estudio.

1.2.1.- MEDIDAS DE INFLUENCIA BASADAS EN LA OMISION

El esquema de perturbación de la omisión de las observaciones a las que se desea analizar la influencia, además de ser el más utilizado y aceptado por los distintos autores, posiblemente sea el más intuitivo en relación al concepto de influencia. Cabe resaltar que, siempre ha estado ligado a distintas definiciones que de observación influencia se han dado [Johnson y Kotz(1982),Gray (1988)].

Su aplicación para el caso de estadísticos o estimadores de parámetros del modelo, si lleva asociada la diferencia de resultados como esquema de comparación, está muy relacionada con el concepto de curva de influencia muestral [Cook y Weisberg (1982)], dado que ésta se define, salvo constantes, como la diferencia entre el valor del estimador obtenido con todos los datos y el valor obtenido con la omisión de uno o de un conjunto de ellos.

Así, dado un estadístico T obtenido con todas las observaciones muestrales y $T_{(i)}$ obtenido con la omisión de la i-ésima observación, la curva de influencia muestral viene definida por :

$$CIM_i(T) = -(n-1)\ [\ T_{(i)} - T\ ]$$

Y puede considerarse como una versión muestral de la curva de influencia [Hampel (1974)], y está, por tanto, estrechamente ligada a este esquema de perturbación, considerándose como una herramienta útil para el análisis de la influencia que sobre T ejerce la i-ésima observación.

No obstante, plantea el problema de su valoración cuando el estadístico T toma valores en $\mathbb{R}^p$, con p>1.

Cook y Weisberg (1982) proponen, de una forma general, una solución al problema planteado, considerando la norma:

$$\frac{CIM_i(T)'\ \mathbf{Q}\ \ CIM_i(T)}{c}$$

donde Q es una matriz simétrica definida positiva y c un escalar positivo. Para cada elección del escalar y de la matriz se obtiene una medida de influencia distinta.

Este planteamiento se generaliza para el análisis de la influencia conjunta, con la omisión de dos o más observaciones, considerando la curva de influencia muestral asociada a un conjunto de observaciones B

$$\mathrm{CIM}_{B}(T) = -\,(n-m)\ [\ T_{(B)} - T\ ]$$

siendo $T_{(B)}$ el valor del estadístico obtenido con la omisión de las m observaciones de B.

Cook, Peña y Weisberg (1988) tratan de dar un criterio que unifique el planteamiento del análisis de la influencia bajo el esquema de omisión englobando todo el modelo. Para ello utilizan el concepto de verosimilitud, tomando como criterio de comparación la diferencia entre los valores del logaritmo de la función de verosimilitd para los estimadores del modelo postulado y el modelo perturbado, lo que se denomina DESPLAZAMIENTO DE VEROSIMILITUD [Cook y Weisberg (1982), Cook (1987)].

Así para un estimador $\hat{\underline{\vartheta}}$ de máxima verosimilitud de un vector paramétrico $\underline{\vartheta}$, se define por

$$\mathrm{LD}_{i}[\ \underline{\vartheta}\ ] = 2\ \Big[\ L\ [\ \hat{\underline{\vartheta}}\ ]\ -\ L\ [\ \hat{\underline{\vartheta}}_{(i)}]\ \Big]$$

siendo L[.] el logaritmo de la función de verosimilitud del modelo.

1.2.2.- EL ESQUEMA DE LA OMISION EN REGRESION LINEAL MULTIPLE

El modelo más tratado desde el punto de vista del análisis de influencia es el de Regresión Lineal Múltiple, y generalmente bajo el esquema de la omisión. Incluso, el estudio de este modelo parece servir como punto de partida para el desarrollo de técnicas en otros modelos y desde otros enfoques.

Así, como extensión del análisis de influencia en éste modelo, y siguiendo el mismo esquema de perturbación y de comparación, se analiza la influencia del modelo de Regresión Multivariante por autores como Caroni (1987), Barrett y Ling (1992), y en los modelos lineales generalizados por autores como Seeber (1986), de Gruttola y otros (1987), Williams (1987), Lee (1987,1988), Schall y Dunne (1988) y Thomas (1990). Putterman (1988) traslada los resultados obtenidos en el modelo de regresión para el caso de errores autocorrelados.

El hecho de que sea el modelo de regresión en el que más se ha avanzado en el estudio de la influencia puede venir provocado por circunstancias como la descrita por Cook (1982)

> « *Es bien conocido, por ejemplo, que inferencias basadas en regresión por mínimos cuadrados ordinarios pueden ser fuertemente influenciadas por unos pocos casos de los datos* »

o como indica Atkinson (1984), por la inadecuación del modelo lineal para describir el comportamiento de los datos obtenidos de un fenómeno natural que puede ser mucho más complejo y amplio. Estas circunstancias, posiblemente, han provocado el desarrollo de un gran número de técnicas, modificando el método de comparación de resultados. Chatterjee y Hadi (1986, 1988) recogen la gran mayoría de estas técnicas, aplicándolas a diversas situaciones prácticas.

Considérese el modelo de Regresión Lineal Múltiple definido en forma matricial por:

$$\underset{n\times 1}{Y} = \underset{n\times p}{X}\ \underset{p\times 1}{\underline{\beta}} + \underset{n\times 1}{\underline{\varepsilon}} \qquad \underline{\varepsilon} \simeq N_n\,[\,0\,,\,\sigma^2\,\mathbf{I}_n\,]$$

$$\text{rango }[X] = p$$

Algunas medidas de influencia sobre el estimador mínimo-cuadrático del vector de coeficientes de regresión propuestas bajo este esquema son las siguientes:

* DISTANCIA DE COOK (D_i) :

[Cook y Weisberg (1982)]

Considerando $\mathbf{Q} = \mathbf{X'X}$ y $c = (n-1)^2\,\hat{\sigma}^2$

siendo $\hat{\sigma}^2$ la estimación mínimo cuadrático de σ^2

* DISTANCIA DE WELSCH-KUH ($DFFITS_i$):

[Belsley, Kuh y Welsch (1980)]

Considerando $\mathbf{Q} = \mathbf{X'X}$ y $c = (n-1)^2\,\hat{\sigma}^2_{(i)}$

siendo $\hat{\sigma}^2_{(i)}$ la estimación de mínimos cuadrados de σ^2 obtenido con la omisión de la i-ésima observación.

* DISTANCIA MODIFICADA DE COOK

[Atkinson (1982)]

Considerando $\mathbf{Q} = \mathbf{X'X}$ y $c = \left[p(n-1)^2/(n-p)\right] \hat{\sigma}^2_{(i)}$

* DISTANCIA DE WELSCH

[Chatterjee y Hadi (1986)]

Considerando $\mathbf{Q} = \mathbf{X}'_{(i)}\,\mathbf{X}_{(i)}$ y $c = (n-1)\,\hat{\sigma}^2_{(i)}$

donde $\mathbf{X}_{(i)}$ es la matriz $\mathbf{X}$ en la que se ha omitido la i-ésima fila.

Siguiendo el mismo esquema se han obtenido medidas de la influencia sobre una componente del vector de coeficientes de regresión y sobre un grupo de dichos coeficientes [Belsley y otros (1980)], denominándose *Influencia Parcial.*

Los autores proporcionan criterios para tomar la decisión de etiquetar o no a una observación como influencia a través de las mencionadas medidas. Dichos criterios, aunque con cierto fundamento estadístico, no dejan de ser subjetivos.

Para otros resultados del modelo de regresión y, por tanto, con la elección de otros métodos de comparación de los mismos, aparecen en la literatura otras medidas de influencia bajo el mismo esquema de perturbación provocada por la omisión de observaciones.

En particular, para la matriz de varianzas y covarianzas de $\hat{\underline{\beta}}$ que viene determinada por

$$\sigma^2 \, [\, X'X \,]^{-1}$$

y para el elipsoide de confianza de $\underline{\beta}$, determinado por

$$\left\{ \underline{\beta} \;/\; \frac{\left[\hat{\underline{\beta}} - \underline{\beta} \right]' X'X \left[\hat{\underline{\beta}} - \underline{\beta} \right]}{p\,\hat{\sigma}^2} \leq F_{1-\alpha,p,n-p} \right\}$$

se han propuesto las siguientes medidas:

- Influencia sobre el elipsoide de confianza de $\underline{\beta}$, basados en el cambio producido en el volumen del mismo:

* ESTADISTICO DE ANDREWS-PREGIBON (AP_i)

[Chatterjee y Hadi (1986)]

$$AP_i = \frac{\det\left[X^{*\prime}_{(i)}\, X^*_{(i)} \right]}{\det\left[X^{*\prime}\, X^* \right]}$$

donde $X^* = [\; X \quad Y \;]$ y $X^*_{(i)}$ es la correspondiente en la que se ha omitido la i-ésima fila.

* ESTADISTICO DE COOK-WEISBERG

[Cook y Weisberg (1982)]

definido como el logaritmo de la razón de los volúmenes del elipsoide de confianza obtenidos con y sin la i-ésima observación.

- Influencia sobre la matriz de varianzas y covarianzas del estimador de mínimos cuadrados de $\underline{\beta}$.

$$\text{COVRATIO}_i = \frac{\det\left[\hat{\sigma}^2_{(i)}[X_{(i)}' X_{(i)}]^{-1}\right]}{\det\left[\hat{\sigma}^2 \; [X' X]^{-1}\right]}$$

[Belsley, Kuh y Welsch (1980)]

Las medidas anteriores también son aplicadas al análisis de la influencia conjunta de un grupo de observaciones, realizando las modificaciones oportunas.

Se observa la gran diversidad de medidas existentes, provocada por la elección del método de comparación y del resultado del análisis estadístico que se pretende estudiar. Esta diversidad ha motivado que algunos autores realicen estudios comparativos [Chatterjee y Hadi (1986), Balasooriya y otros (1987)] y clasificaciones de las mismas [Gray (1988)].

1.2.3.- OTROS ESQUEMAS DE PERTURBACION . INFLUENCIA LOCAL

El esquema de perturbación de la omisión de observaciones valora la influencia a partir del salto cualitativo que supone el considerar o no considerar la observación en cuestión. No obstante, a través de él, no se pueden estudiar comportamientos intermedios que permitan analizar la influencia con magnitudes de perturbación más pequeñas, lo cual es interesante considerar en algunos casos.

Así, Cook (1987), a través de la formulación del problema de la influencia comentado en el punto 1.1, plantea otros posibles modelos de perturbación basados en ligeras modificaciones en los datos, en la estructura de la varianza del modelo o, incluso, en la estructura de correlación de las variables.

Por ejemplo, para el modelo de Regresión apunta dos posibles esquemas de perturbación:

I.- Considerar el modelo perturbado

$$\mathbf{Y} = \mathbf{X}\,\underline{\beta} + \underline{\varepsilon} \quad , \ \underline{\varepsilon} \simeq N_n[0,\sigma^2\mathbf{W}]$$

donde $\mathbf{W} = \text{diag}\left[1 \ , 1 \ , \dots , \underset{i)}{1/\omega} \ , \dots , 1 \right] \ , \ \omega > 0$

Es decir, propone una perturbación que provoca la falta de homocedasticidad del modelo. Se observa que para ω=1 el modelo perturbado coincide con el clásico de regresión.

II.- Considerar el modelo perturbado

$$\mathbf{Y} = \Big[\mathbf{X} + [\ 0 \quad \underline{\omega}\ ] \Big]\, \underline{\beta} + \underline{\varepsilon} \quad , \ \underline{\varepsilon} \simeq N_n[0,\sigma^2 I_n]$$

donde $\underline{\omega}$ es el vector de perturbaciones de $\mathbb{R}^n$.

En estos dos modelos perturbados, o en cualquier otro que siga el esquema planteado por Cook en la formulación del problema de la influencia, haciendo variar $\underline{\omega}$ convenientemente, se puede obtener un análisis del comportamiento de los resultados del modelo frente a dichas perturbaciones $\underline{\omega}$.

En el caso del esquema I, estudiando ω en el intervalo (0,1), permite analizar comportamientos intermedios entre el considerar o no la i–ésima observación en el modelo.

Este tipo de análisis, que el autor denomina *Análisis de Influencia Local*, utiliza como herramientas para su desarrollo, fundamentalmente, la realización de representaciones gráficas de los cambios producidos por los resultados frente a ω, a partir de las cuales se pueden extraer conclusiones sobre fiabilidad de las observaciones y de los resultados frente a éstas.

Cook (1987) desarrolla el esquema de perturbación II, considerando como modelo de comparación el desplazamiento de verosimilitud. La elección del modelo de comparación y el esquema de perturbación permiten una gran variedad de análisis de influencia local. No obstante, es el esquema I el más adecuado para extraer conclusiones referentes a la influencia, dado que el esquema II y el también apuntado por Cook referente a perturbaciones en la estructura de correlación entre la variables explicativas están dirigidos, fundamentalmente, al estudio de la adecuación del modelo a la estructura de datos.

C A P I T U L O I I

DESCOMPOSICION DE LOS ESTIMADORES DEL MODELO LINEAL GENERAL. MODELOS PERTURBADOS

CAPITULO II.- DESCOMPOSICION DE LOS ESTIMADORES DEL MODELO LINEAL GENERAL. MODELOS PERTURBADOS

En este capítulo se obtienen diversos resultados que se aplican en el análisis de influencia del Modelo Lineal General. Así, el objetivo fundamental es obtener la descomposición de los estimadores de los parámetros del Modelo Lineal General en base a las esperanzas condicionadas a las observaciones muestrales.

Inicialmente se presentan generalidades sobre el Modelo Lineal General, a partir de las cuales se obtiene la descomposición de los estimadores de las funciones linealmente estimables y del estimador mínimo-cuadrático de la varianza, trasladándose los resultados al modelo de Regresión Lineal Múltiple y Análisis de la Varianza de un Factor, como modelos que pueden ser formulados a través del Modelo Lineal General.

Se estudia un modelo perturbado, relacionándose los estimadores de máxima verosimilitud del mismo con los correspondientes del Modelo Lineal General. Asímismo, se repite este esquema para el modelo de omisión de una o varias observaciones.

Finalmente se estudian dos transformaciones de los resíduos, denominados *resíduos estandarizados* y *resíduos studentizados*, los cuales van a desempeñar un papel relevante en el análisis de influencia.

2.1.- GENERALIDADES SOBRE EL MODELO LINEAL GENERAL

El estudio del Modelo Lineal General es de gran importancia dentro de la estadística fundamentalmente por la gran variedad de sus aplicaciones, al servir como modelo base para el desarrollo de cualquier modelo lineal (Modelo de Regresión Múltiple, Modelos de Diseño de Experimentos, Modelo de Análisis de la Covarianza,...).

El Modelo Lineal General se considerará definido por

$$Y = X \underline{\beta} + \underline{\varepsilon} \;, \quad E[\underline{\varepsilon}] = 0 \;, \; Var(\underline{\varepsilon}) = \sigma^2 I_n \qquad \textbf{(Mod.1)}$$

donde **Y** es un vector aleatorio de dimension n; **X** una matriz de constantes conocidas de dimensión nxp y de rango r ($r \leq p$); $\underline{\beta}$ un vector paramétrico desconocido de dimensión p y $\underline{\varepsilon}$ un vector de dimensión n que representa la perturbación aleatoria no observable.

Los resultados fundamentales obtenidos sobre este modelo, recogidos por Kshirsagar (1983), y que serán utilizados en desarrollos posteriores son los que a continuación se especifican.

TEOREMA 2.1.1.- Los estimadores mínimo cuadráticos del vector paramétrico $\underline{\beta}$ son:

$$\hat{\underline{\beta}} = S^{-} X' Y + (I_p - H) z$$

donde I_p es la matriz identidad de dimensión p, $S=X'X$ y S^{-} es una inversa generalizada de S, con $H = S^{-}S$ y $z \in \mathbb{R}^p$ arbitrario.

Por otro lado, $\hat{Y} = V\ Y$ representa el vector de valores ajustados, siendo

$$V = X\ S^{-}X' = [\ \underline{V}_1,\ \underline{V}_2, \ldots,\ \underline{V}_n\ ] = ((v_{ij}))_{i,j=1,\ldots,n}$$

la matriz de predicción, la cual verifica:

TEOREMA 2.1.2.- La matriz de predicción **V** es simétrica, idempotente, de rango r y única para cualquier elección de S^{-}.

Al vector $e = Y - \hat{Y} = M\ \underline{\varepsilon}$ se denomina vector de resíduos, donde $M = I_n - V$, que también es simétrica, idempotente y de rango n-r.

DEFINICION 2.1.3.- Una función lineal del vector paramétrico, $\underline{\lambda}'\ \underline{\beta}$, se denomina linealmente estimable si

$$\exists\ u \in \mathbb{R}^n\ /\ E\ [\ u'Y\ ] = \underline{\lambda}'\ \underline{\beta}$$

Se obtienen las siguientes condiciones necesarias y suficientes de estimabilidad:

TEOREMA 2.1.4.- Una función lineal paramétrica, $\underline{\lambda}'\ \underline{\beta}$, es linealmente estimable sii $\exists\ \mathbf{a} \in \mathbb{R}^n\ /\ \underline{\lambda}' = \mathbf{a}'\mathbf{X}$.

TEOREMA 2.1.5.- Una función lineal paramétrica, $\underline{\lambda}'\ \underline{\beta}$, es linealmente estimable sii $\underline{\lambda}' = \underline{\lambda}\ \mathbf{H}$.

COROLARIO 2.1.6.- $\underline{x}_i\underline{\beta}$ y $X\ \underline{\beta}$ son linealmente estimables, siendo $\underline{x}_i$ la i-ésima fila de X,

TEOREMA 2.1.7.- $\underline{\lambda}'\hat{\underline{\beta}}$ es único para cualquier estimador mínimo-cuadrático $\hat{\underline{\beta}}$ sii $\underline{\lambda}'= \underline{\lambda}'\mathbf{H}$.

Obteniéndose, finalmente, el Teorema de Gauss-Markov, que proporciona el estimador lineal insesgado uniformemente de mínima varianza (BLUE) de una función linealmente estimable de $\underline{\beta}$.

TEOREMA 2.1.8.- (*TEOREMA DE GAUSS-MARKOV*)

El BLUE de una función $\underline{\lambda}'\underline{\beta}$ linealmente estimable viene dado por $\underline{\lambda}'\hat{\underline{\beta}}'$, para cualquier $\hat{\underline{\beta}}$ solución del método de mínimos cuadrados.

Por extensión, dada una matriz Λ de dimension qxp y de rango q, la función lineal del vector paramétrico, $\Lambda\ \underline{\beta}$, es linealmente estimable sí y sólo sí $\Lambda\ \mathbf{H} = \Lambda$, verificándose que $\Lambda\hat{\underline{\beta}}$ es única para cualquier elección de $\hat{\underline{\beta}}$, y además es un estimador insesgado tal que cada una de sus componentes es el BLUE de la componente correspondiente de $\Lambda\underline{\beta}$.

Como estimador del parámetro desconocido σ^2 se obtiene

TEOREMA 2.1.9.- El estadístico

$$\hat{\sigma}^2 = \frac{1}{n-r}\, e'e = \frac{1}{n-r}\, [Y - X\,\hat{\underline{\beta}}\,]'[Y - X\,\hat{\underline{\beta}}\,]$$

es único para cualquier elección de $\hat{\underline{\beta}}$ y es un estimador insesgado del parámetro σ^2

El estimador $\hat{\sigma}^2$ también se representa $SC_\varepsilon/(n-r)$, cuadrado medio debido al error.

Si se supone que el vector aleatorio $\underline{\varepsilon}$ sigue una distribución $N_n[0,\sigma^2 I_n]$, el modelo Mod.1 se representará por **(Mod.1-N)** obteniéndose que los estimadores de máxima verosimilitud del vector paramétrico, coinciden con los obtenidos por el método de mínimos cuadrados. Y el correspondiente estimador máximo-verosímil de σ^2 es

$$\tilde{\sigma}^2 = \frac{1}{n} e'e = \frac{1}{n} [Y - X \hat{\underline{\beta}}]'[Y - X \hat{\underline{\beta}}]$$

Tal supuesto de normalidad, conduce a obtener los siguientes resultados:

TEOREMA 2.1.10.-

1.- $e \simeq N_n[0,\sigma^2 V]$

2.- Si $\Lambda \underline{\beta}$ es linealmente estimable, siendo Λ una matriz qxp de rango q, entonces

$$\Lambda \hat{\underline{\beta}} \simeq N_q[\Lambda\underline{\beta}, \sigma^2 \Lambda S^- \Lambda']$$

3- $SC_\varepsilon \simeq \sigma^2 \chi^2_{n-r}$ y es independiente del BLUE de cualquier función linealmente estimable.

4.- El elipsoide de confianza , a un nivel del 100(1-α)% , de una función linealmente estimble $\Lambda\underline{\beta}$ viene dado por:

$$\left\{ \Lambda\underline{\beta} \ / \ \frac{[\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta}]' \ [\Lambda \ S^- \ \Lambda']^{-1} \ [\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta}]}{q \ \hat{\sigma}^2} \le F_{q,n-r,1-\alpha} \right\}$$

2.2.- DESCOMPOSICION DE LOS ESTIMADORES DEL MODELO LINEAL GENERAL

Con el objetivo de obtener un criterio para la identificación de observaciones influencia en la estimación de un parámetro, se considera el siguiente lema de descomposición de Efron y Stein (1981):

LEMA 2.2.1.- Toda variable aleatoria $S(X_1,X_2,\ldots,X_n)$, función de n variables aleatorias independientes $X_1,X_2,\ldots,X_n$, puede ser expresada por

$$S(X_1,X_2,\ldots,X_n) = E[S] + \sum_i A_i(X_i;S) + \sum_{i<j}\sum B_{ij}(X_i,X_j;S) + \\ + \sum_{i<j<k}\sum\sum C_{ijk}(X_i,X_j,X_k;S) + \ldots + H(X_1,X_2,\ldots, X_n;S)$$

donde las 2^n-1 variables aleatorias del miembro derecho de la expresión tiene esperanza nula y están mutuamente incorreladas, siendo

- $A_i(x_i;S) = E[S|X_i= x_i] - E[S]$, lo que representa el i-ésimo efecto medio.

- $B_{ij}(x_i,x_j;S)= E[S|\ X_i= x_i,X_j=x_j] - E[S|\ X_i= x_i] -$
  $- E[S|\ X_j= x_j] - E[S]$, lo que representa la (i,j)-interacción de segundo orden

etc...

A través de este resultado, se puede obtener una descomposición para los estimadores de $\Lambda\underline{\beta}$ y σ^2 del Modelo Lineal General, repectivamente $\Lambda\hat{\underline{\beta}}$ y $\hat{\sigma}^2$.

2.2.1.-DESCOMPOSICION DEL ESTIMADOR DE UNA FUNCION LINEAL ESTIMABLE DEL VECTOR PARAMETRICO.

En este apartado se obtiene la descomposición de los estimadores obtenidos para las funciones linealmente estimables $\Lambda\underline{\beta}$.

LEMA 2.2.2.- Sea $\Lambda\underline{\beta}$ una función linealmente estimable, donde Λ es una matriz de dimensión qxp de rango q, entonces

1) $A_i[y_i;\Lambda\,\hat{\underline{\beta}}\,] = \Lambda\, S^- \,\underline{x}_i'\,(\,y_i - \underline{x}_i\,\underline{\beta})$, i=1,...,n

2) $B_{ij}[y_i,y_j;\Lambda\,\hat{\underline{\beta}}\,] = 0$ i,j=1,...,n , i≠j.

Demostración.-

1) $$E[\Lambda\hat{\underline{\beta}} \mid Y_i=y_i] = E[\Lambda S^- X'Y \mid Y_i=y_i] = \Lambda S^- E[X'Y \mid Y_i=y_i] =$$

$$= \Lambda\, S^-\, E\Big[\sum_k \underline{x}_k'Y_k \mid Y_i=y_i\Big] = \Lambda\, S^-\Big[\underline{x}_i'\,y_i + E\Big[\sum_{k\neq i}\underline{x}_k'\,Y_k\Big]\Big] =$$

$$= \Lambda\, S^-\Big[\underline{x}_i'\,y_i + \sum_{k\neq i}\underline{x}_k'\,\underline{x}_k\,\underline{\beta}\Big] = \Lambda\, S^-\Big[\underline{x}_i'\,y_i + (S - \underline{x}_i'\underline{x}_i)\,\underline{\beta}\Big] =$$

$$= \Lambda\, S^-\Big[S\,\underline{\beta} + \underline{x}_i'\,(y_i - \underline{x}_i\underline{\beta})\Big] = \Lambda\underline{\beta} + \Lambda\, S^-\,\underline{x}_i'\,(\,y_i - \underline{x}_i\,\underline{\beta}\,)$$

Luego

$$A_i[y_i;\Lambda\,\hat{\underline{\beta}}\,] = \Lambda\, S^-\,\underline{x}_i'\,(\,y_i - \underline{x}_i\,\underline{\beta}) \text{ , i=1,...,n}$$

2) $E[\Lambda\hat{\underline{\beta}} \mid Y_i=y_i, Y_j=y_j] = \Lambda\, S^- \, E\left[\underline{x}_i'\, y_i + \underline{x}_j'\, y_j + \sum_{k\neq i,j} \underline{x}_k'\, Y_k \right] =$

$\Lambda S^-\left[\underline{x}_i' y_i + \underline{x}_j' y_j + \sum_{k\neq i,j} \underline{x}_k' \underline{x}_k \underline{\beta} \right] = \Lambda S^- \left[\underline{x}_i' y_i + \underline{x}_j' y_j + (S - \underline{x}_i'\underline{x}_i - \underline{x}_j'\underline{x}_j)\underline{\beta} \right]$

$= \Lambda\underline{\beta} + \Lambda S^- \underline{x}_i'(y_i - \underline{x}_i\underline{\beta}) + \Lambda S^- \underline{x}_j'(y_j - \underline{x}_j\underline{\beta})$

Luego, $B_{ij}[y_i, y_j; \Lambda\hat{\underline{\beta}}\ ] = 0.$

■

TEOREMA 2.2.3.- Sea $\Lambda\underline{\beta}$ una función linealmente estimable, donde Λ es una matriz de dimensión qxp de rango q, entonces

$$\Lambda\hat{\underline{\beta}} = \Lambda\underline{\beta} + \sum_i \Lambda\, S^-\, \underline{x}_i' \,(\, Y_i - \underline{x}_i\, \underline{\beta}\,)$$

Demostración.-

Evidente, dado que

$$\sum_i \Lambda\, S^-\, \underline{x}_i'\ \underline{x}_i\, \underline{\beta} = \Lambda\underline{\beta}$$

■

Por lo tanto, la descomposición del estimador $\Lambda\hat{\underline{\beta}}$ únicamente contiene los efectos medios de primer orden.

COROLARIO 2.2.4.- Sea $\hat{Y}$ el vector de valores ajustados, entonces:

$$\hat{Y} = X\, \underline{\beta} + \sum_i \underline{V}_i (\, Y_i - \underline{x}_i\, \underline{\beta}\,)$$

donde $\underline{V}_i$ es la i-ésima columna de la matriz de predicción **V**

2.2.2.-DESCOMPOSICION DEL ESTIMADOR DE LA VARIANZA

El estimador insesgado mínimo-cuadrático de σ^2 obtenido en el teorema 2.1.9, también puede ser expresado por

$$\hat{\sigma}^2 = \frac{1}{n-r}\, \mathbf{Y'(I-V)Y}$$

y utilizando esta igualdad se obtienen los siguientes resultados:

LEMA 2.2.5..- Sea $\hat{\sigma}^2$ el estimador mínimo-cuadrático de σ^2, entonces

$$A_i[y_i;\hat{\sigma}^2] = \frac{1-v_{ii}}{n-r} \left[[y_i - \underline{x}_i\, \underline{\beta}\,]^2 - \sigma^2 \right] \quad , \; i=1,\dots,n$$

<u>Demostración</u>.-

$$E\,[(n-r)\hat{\sigma}^2 \mid Y_i=y_i\,] = E\,[\,\mathbf{Y'(I-V)Y} \mid Y_i=y_i] =$$

$$= E\left[\sum_k (1-v_{kk})Y_k^2 - \sum\sum_{k\neq s} v_{ks}\, Y_k Y_s \mid Y_i=y_i\right] =$$

$$= (1-v_{ii})y_i^2 + \sum_{k\neq i}(1-v_{kk})\,[\sigma^2 + (\underline{x}_k\underline{\beta})^2] - 2y_i \sum_{s\neq i} v_{is}(\underline{x}_s\underline{\beta}) -$$

$$- \sum\sum_{\substack{k\neq s\\ k,\,s\neq i}} v_{ks}(\underline{x}_s\underline{\beta})(\underline{x}_k\underline{\beta})$$

Desarrollando cada uno de los sumandos de la expresión anterior, teniendo en cuenta, además, que $\mathbf{V\ X = X}$, y el caracter idempotente de $\mathbf{V}$, se obtiene

* $\sum_k v_{ik}\, \underline{x}_k = \underline{x}_i$, luego $\sum_{k\neq i} v_{ik}\, \underline{x}_k = (1-v_{ii})\, \underline{x}_i$;
* traza$(\mathbf{V}) = \sum_k v_{kk} = r$

Así,

$$\sum_{k\neq i}(1-v_{kk})\,[\sigma^2 + (\underline{x}_k\underline{\beta})^2] = (n-1-r+v_{ii})\,\sigma^2 + \sum_{k\neq i}(1-v_{kk})\,(\underline{x}_k\underline{\beta})^2$$

Por otra parte

$$2\, y_i \sum_{s\neq i} v_{is}(\underline{x}_s \underline{\beta}) = 2\,(1-v_{ii})\, y_i\, (\underline{x}_i\, \underline{\beta})$$

y, ademas,

$$\sum_{\substack{k\neq s\\ k,\, s\neq i}}\sum v_{ks}(\underline{x}_s\underline{\beta})(\underline{x}_k\underline{\beta}) = \sum_{k\neq i}\left[\sum_{s\neq k,\, i} v_{ks}(\underline{x}_s\underline{\beta})\right](\underline{x}_k\underline{\beta}) =$$

$$= \sum_{k\neq i}\left[(1-v_{kk})\,\underline{x}_k\,\underline{\beta} - v_{ki}\,\underline{x}_i\,\underline{\beta}\right](\underline{x}_k\underline{\beta}) =$$

$$= \sum_{k\neq i}(1-v_{kk})\quad (\underline{x}_k\underline{\beta})^2 - (1-v_{ii})\,(\underline{x}_i\,\underline{\beta}\,)^2$$

Por lo tanto,

$$E\,[(n-r)\hat{\sigma}^2 \mid Y_i=y_i] = (n-1-r+v_{ii})\,\sigma^2 \;+ (1-v_{ii})\;[\,y_i - \underline{x}_i\,\underline{\beta}\,]^2$$

De donde se deduce lo enunciado .

■

LEMA 2.2.6.- Sea $\hat{\sigma}^2$ el estimador mínimo-cuadrático de σ^2, entonces

$$B_{ij}[y_i,y_j;\hat{\sigma}^2] = -\frac{v_{ij}}{n-r}\left[y_i - \underline{x}_i\underline{\beta}\right]\left[y_j - \underline{x}_j\underline{\beta}\right]$$

$$i,j=1,\ldots,n\ ,\ i\neq j$$

<u>Demostración</u>.-

Siguiendo un razonamiento análogo al teorema anterior se tiene que

$$E\,[(n-r)\hat{\sigma}^2 \mid Y_i=y_i,\ Y_j=y_j\,] = E\,[\ \mathbf{Y'(I-V)Y} \mid Y_i=y_i,\ Y_j=y_j] =$$

$$= E\left[\sum_{k} (1-v_{kk})Y_k^2 - \sum_{k\neq s}\sum v_{ks}\, Y_k Y_s \;\middle|\; Y_i=y_i,\ Y_j=y_j \right] =$$

$$= (1-v_{ii})y_i^2 + (1-v_{jj})y_j^2 + \sum_{k\neq i,j}(1-v_{kk})\ [\sigma^2 + (\underline{x}_k\underline{\beta})^2] -$$

$$- \sum_{k\neq s}\sum v_{ks}\ E\,[\ Y_k\ Y_s |\ Y_i=y_i,\ Y_j=y_j\ ]$$

Desarrollando individualmente los sumatorios que aparecen en la última expresión, se tiene en primer lugar que

$$\sum_{k\neq i,j}(1-v_{kk})\ [\sigma^2 + (\underline{x}_k\underline{\beta})^2] = (n-2-r+v_{ii}+v_{jj})\sigma^2 + \sum_{k\neq i,j}(1-v_{kk})\ \ (\underline{x}_k\underline{\beta})^2$$

Y por otra parte, descomponiendo el segundo sumando:

$$\sum_{k\neq s}\sum v_{ks} E\,[\ Y_k\ Y_s |\ Y_i=y_i,\ Y_j=y_j\ ]\ =$$

$$\sum_{s\neq i,j} v_{is} E\,[\ Y_i\ Y_s | Y_i=y_i, Y_j=y_j] + \sum_{s\neq i,j} v_{js} E\,[\ Y_j\ Y_s | Y_i=y_i, Y_j=y_j] +$$

$$\sum_{k\neq i,j} v_{ki} E\,[\ Y_i\ Y_k | Y_i=y_i, Y_j=y_j] + \sum_{k\neq i,j} v_{kj} E\,[\ Y_j\ Y_k | Y_i=y_i, Y_j=y_j] +$$

$$+ v_{ij}\ E\,[\ Y_i\ Y_j | Y_i=y_i, Y_j=y_j] + v_{ji}\ E[\ Y_i\ Y_j | Y_i=y_i, Y_j=y_j] +$$

$$+ \sum_{\substack{k\neq s\\ k,s\neq i,j}}\sum v_{ks} E\,[\ Y_k\ Y_s |\ Y_i=y_i,\ Y_j=y_j\ ]$$

Por la simetría de **V** y la simetría de las expresiones, los sumandos primero y tercero son iguales, así como el segundo y el cuarto, y el quinto y el sexto. Se calculan, pues, los sumandos primero, segundo, quinto y séptimo.

$$\sum_{s\neq i,j} v_{is} E[\, Y_i\, Y_s | Y_i=y_i, Y_j=y_j] = \sum_{s\neq i,j} v_{is}\, y_i\, E[Y_s] =$$

$$= y_i \sum_{s\neq i,j} v_{is}\, (\underline{x}_s\, \underline{\beta}) = (1-v_{ii})\, y_i\, (\underline{x}_i\, \underline{\beta}) - v_{ij}\, y_i\, (\underline{x}_j\, \underline{\beta})$$

Análogamente, se tiene

$$\sum_{s\neq i,j} v_{js} E[\, Y_j\, Y_s | Y_i=y_i, Y_j=y_j] = (1-v_{jj})\, y_j\, (\underline{x}_j\, \underline{\beta}) - v_{ji}\, y_j\, (\underline{x}_i\, \underline{\beta})$$

Por otra parte,

$$v_{ij}\, E[\, Y_i\, Y_j | Y_i=y_i, Y_j=y_j] = v_{ij}\, y_i\, y_j$$

Además,

$$\sum_{k,s\neq i,j}\sum_{k\neq s} v_{ks} E[\, Y_k\, Y_s | \, Y_i=y_i,\, Y_j=y_j\,] = \sum_{k,s\neq i,j}\sum_{k\neq s} v_{ks} E[\, Y_k\, Y_s\,] =$$

$$\sum_{k,s\neq i,j}\sum_{k\neq s} v_{ks}\, (\underline{x}_k\, \underline{\beta})\, (\underline{x}_s\, \underline{\beta}) = \sum_{k\neq i,j} \left[\sum_{s\neq i,j,k} v_{ks}\, (\underline{x}_s\, \underline{\beta}) \right] (\underline{x}_k\, \underline{\beta}) =$$

$$= \sum_{k\neq i,j} \left[(1-v_{kk})\, (\underline{x}_k\, \underline{\beta}) - v_{ki}\, (\underline{x}_i\, \underline{\beta}) - v_{kj}\, (\underline{x}_j\, \underline{\beta}\,) \right] (\underline{x}_k \underline{\beta}) =$$

$$= \sum_{k\neq i,j} (1-v_{kk})\, (\underline{x}_k\, \underline{\beta})^2 - \left[\sum_{k\neq i,j} v_{ki}\, (\underline{x}_k\, \underline{\beta}) \right] (\underline{x}_i\, \underline{\beta}) -$$

$$- \left[\sum_{k\neq i,j} v_{kj}\, (\underline{x}_k\, \underline{\beta}) \right] (\underline{x}_j\, \underline{\beta}) =$$

$$= \sum_{k\neq i,j} (1-v_{kk})\, (\underline{x}_k\, \underline{\beta})^2 - \left[(1-v_{ii})\, (\underline{x}_i\, \underline{\beta})^2 - v_{ij}\, (\underline{x}_i\, \underline{\beta})\, (\underline{x}_j\, \underline{\beta}) \right] -$$

$$- \left[(1-v_{jj})\, (\underline{x}_j\, \underline{\beta})^2 - v_{ji}\, (\underline{x}_i\, \underline{\beta})\, (\underline{x}_j\, \underline{\beta}) \right] =$$

$$= \sum_{k\neq i,j} (1-v_{kk})\, (\underline{x}_k\, \underline{\beta})^2 - (1-v_{ii})\, (\underline{x}_i\, \underline{\beta})^2 - (1-v_{jj})\, (\underline{x}_j\, \underline{\beta})^2 +$$

$$+ 2\, v_{ij}\, (\underline{x}_i\, \underline{\beta})\, (\underline{x}_j\, \underline{\beta})$$

Por lo tanto, se tiene

$$\sum_{k\neq s}\sum v_{ks} E\, [\, Y_k\, Y_s |\, Y_i = y_i,\, Y_j = y_j\,] \, =$$

$$2\,(1-v_{ii})\, y_i\, (\underline{x}_i\, \underline{\beta}) - 2\, v_{ij}\, y_i\, (\underline{x}_j\, \underline{\beta}) + 2\,(1-v_{jj})\, y_j\, (\underline{x}_j\, \underline{\beta}) -$$

$$- 2\, v_{ji}\, y_j\, (\underline{x}_i\, \underline{\beta}) + 2\, v_{ij}\, y_i\, y_j + \sum_{k\neq i,j} (1-v_{kk})\, (\underline{x}_k\, \underline{\beta})^2 -$$

$$- (1-v_{ii})\, (\underline{x}_i\, \underline{\beta})^2 - (1-v_{jj})\, (\underline{x}_j\, \underline{\beta})^2 + 2\, v_{ij}\, (\underline{x}_i\, \underline{\beta})\, (\underline{x}_j\, \underline{\beta}) =$$

$$= (1-v_{ii}) \left[2\, y_i(\underline{x}_i\, \underline{\beta}) - (\underline{x}_i\, \underline{\beta})^2\right] + (1-v_{jj}) \left[2\, y_j(\underline{x}_j\, \underline{\beta}) - (\underline{x}_j\, \underline{\beta})^2\right] +$$

$$+ 2\, v_{ij} \left[y_i - \underline{x}_i\underline{\beta}\right] \left[y_j - \underline{x}_j\underline{\beta}\right] + \sum_{k\neq i,j} (1-v_{kk})\, (\underline{x}_k\, \underline{\beta})^2$$

Obteniéndose así

$$E\, [(n-r)\hat{\sigma}^2 |\, Y_i = y_i,\, Y_j = y_j\,] = (n-2-r+v_{ii}+v_{jj})\, \sigma^2 +$$

$$+ (1-v_{ii})\left[y_i - \underline{x}_i\underline{\beta}\right]^2 + (1-v_{jj})\left[y_j - \underline{x}_j\underline{\beta}\right]^2 - 2\, v_{ij} \left[y_i - \underline{x}_i\underline{\beta}\right]\left[y_j - \underline{x}_j\underline{\beta}\right]$$

y, por tanto:

$$B_{ij}[y_i, y_j; \hat{\sigma}^2] = - \frac{v_{ij}}{n-r} \left[y_i - \underline{x}_i\underline{\beta}\right] \left[y_j - \underline{x}_j\underline{\beta}\right]$$

$$i,j=1,\ldots,n\ ,\ i\neq j$$

■

Una vez obtenidos los términos de segundo orden en la descomposición, en el siguiente lema se comprueba que los términos de tercer orden son nulos.

LEMA 2.2.7.- Sea $\hat{\sigma}^2$ el estimador mínimo-cuadrático de σ^2, entonces

$$C_{ijk}\left[y_i,y_j,y_k;\ \hat{\sigma}^2\right] = 0 \quad \text{para todo} \quad i,j,k = 1,..,n\ , \quad \text{distintos.}$$

Demostración.-

Expresando $(n-r)\hat{\sigma}^2$ en su representación matricial **Y'(I-V)Y**, se tiene:

$$E\,[Y'(I-V)Y \mid Y_i = y_i, Y_j = y_j,\ Y_k = y_k] =$$

$$= \sum_s (1-v_{ss})\ E\,[\ Y_s^2 \mid Y_i = y_i, Y_j = y_j,\ Y_k = y_k\ ] -$$

$$- \sum_{s\neq t}\sum v_{st}\ E[Y_sY_t \mid Y_i = y_i, Y_j = y_j,\ Y_k = y_k] \qquad (*)$$

Se denota los dos sumatorios por SUMA1 y SUMA2 respectivamente, y calculándolos individualmente, se obtiene:

* *Cálculo de* SUMA1

$$\text{SUMA1} = (1-v_{ii})y_i^2 + (1-v_{jj})y_j^2 + (1-v_{kk})y_k^2 + \sum_{s\neq i,j,k} (1-v_{ss})[\sigma^2+(\underline{x}_s\ \underline{\beta})^2] =$$

$$(1-v_{ii})y_i^2 + (1-v_{jj})y_j^2 + (1-v_{kk})y_k^2 + [n-3-r+v_{ii}+v_{jj}+v_{kk}]\ \sigma^2 +$$

$$+ \sum_{s\neq i,j,k} (1-v_{ss})[\sigma^2+(\underline{x}_s\ \underline{\beta})^2]$$

* *Cálculo de* SUMA2.

Para obtener el segundo sumatorio se descompone en función de los subíndices, s y t, como sigue:

$\{ (s,t) \mid s,t=1,..,n;\ s\neq t \} = \{ (s,t) \mid s,t=i,j,k;\ s\neq t \} \cup$

$\cup \{ (s,t) \mid [s=i;\ t\neq i,j,k\ ;t=1,..,n] \text{ ó } [t=i;\ s\neq i,j,k\ ;s=1,..,n] \} \cup$

$\cup \{ (s,t) \mid [s=j;\ t\neq i,j,k\ ;t=1,..,n] \text{ ó } [t=j;\ s\neq i,j,k\ ;s=1,..,n] \} \cup$

$\cup \{ (s,t) \mid [s=k;\ t\neq i,j,k\ ;t=1,..,n] \text{ ó } [t=k;\ s\neq i,j,k\ ;s=1,..,n] \} \cup$

$\cup \{ (s,t) \mid s,t\neq i,j,k\ ;t=1,..,n;\ s\neq t \}$

Denotando estos sumandos por SUMA2(1), SUMA2(2), SUMA2(3), SUMA2(4) Y SUMA2(5), respectivamente, el cálculo de cada uno de ellos es como sigue:

- *Cálculo de* SUMA2(1)

Como $E[Y_i Y_j \mid Y_i = y_i, Y_j = y_j,\ Y_k = y_k] = y_i\ y_j$, entonces, por simetría, se obtiene

$$\text{SUMA2(1)} = 2\ v_{ij}\ y_i\ y_j + 2\ v_{ik}\ y_i\ y_k + 2\ v_{jk}\ y_j\ y_k$$

- *Cálculo de* SUMA2(2) , SUMA2(3), SUMA2(4).

Por simetría entre los subíndices i,j,k los tres sumandos indicados se calculan de forma semejante, obteniéndose:

$$\text{SUMA2(2)} = \sum_{t\neq i,j,k} v_{it} E[Y_i Y_t \mid Y_i=y_i, Y_j=y_j, Y_k=y_k] +$$

$$+ \sum_{s\neq i,j,k} v_{si} E[Y_s Y_i \mid Y_i=y_i, Y_j=y_j, Y_k=y_k] = 2\ y_i \sum_{t\neq i,j,k} v_{it} (\underline{x}_t\ \underline{\beta}) =$$

$$= 2\ y_i \Big[(1-v_{ii})\ (\underline{x}_i\ \underline{\beta}) - v_{ij}\ (\underline{x}_j\ \underline{\beta}) - v_{ik}\ (\underline{x}_k\ \underline{\beta}) \Big]$$

Por lo tanto, por la simetría entre los subíndices se obtiene:

$$\text{SUMA2(3)} = 2\ y_j \Big[(1-v_{jj})\ (\underline{x}_j\ \underline{\beta}) - v_{ji}\ (\underline{x}_i\ \underline{\beta}) - v_{jk}\ (\underline{x}_k\ \underline{\beta}) \Big]$$

$$SUMA2(4) = 2\, y_k \left[(1-v_{kk})\, (\underline{x}_k\, \underline{\beta}) - v_{ki}\, (\underline{x}_i\, \underline{\beta}) - v_{kj}\, (\underline{x}_j\, \underline{\beta}) \right]$$

- *Cálculo de* SUMA2(5)

$$SUMA2(5) = \sum_{\substack{s \\ s,t\neq i,j,k \\ s\neq t}} \sum_{t} v_{st}\; E[Y_s Y_t \,|\, Y_i=y_i, Y_j=y_j, Y_k=y_k] =$$

$$= \sum_{\substack{s \\ s,t\neq i,j,k \\ s\neq t}} \sum_{t} v_{st}\, (\underline{x}_s\, \underline{\beta})\, (\underline{x}_t\, \underline{\beta}) = \sum_{s\neq i,j,k} \left[\sum_{t\neq i,j,k,s} v_{st}\, (\underline{x}_t\, \underline{\beta}) \right] (\underline{x}_s\, \underline{\beta})$$

$$= \sum_{s\neq i,j,k} \left[(1-v_{ss})\, (\underline{x}_s \underline{\beta}) - v_{si} (\underline{x}_i \underline{\beta}) - v_{sj} (\underline{x}_j \underline{\beta}) - v_{sk} (\underline{x}_k \underline{\beta}) \right] (\underline{x}_s \underline{\beta}) =$$

$$= \sum_{s\neq i,j,k} (1-v_{ss})\, (\underline{x}_s \underline{\beta})^2 - (\underline{x}_i \underline{\beta}) \left[(1-v_{ii})(\underline{x}_i \underline{\beta}) - v_{ij} (\underline{x}_j \underline{\beta}) - v_{ik} (\underline{x}_k \underline{\beta}) \right] -$$

$$- (\underline{x}_j \underline{\beta}) \left[(1-v_{jj})(\underline{x}_j \underline{\beta}) - v_{ji} (\underline{x}_i \underline{\beta}) - v_{jk} (\underline{x}_k \underline{\beta}) \right] -$$

$$- (\underline{x}_k \underline{\beta}) \left[(1-v_{kk})(\underline{x}_k \underline{\beta}) - v_{ki} (\underline{x}_i \underline{\beta}) - v_{kj} (\underline{x}_j \underline{\beta}) \right]$$

Simplificando se obtiene:

$$SUMA2(5) = = \sum_{s\neq i,j,k} (1-v_{ss})(\underline{x}_s \underline{\beta})^2 -$$

$$- \left[(1-v_{ii})(\underline{x}_i \underline{\beta})^2 + (1\; v_{jj})(\underline{x}_j \underline{\beta})^2 - (1-v_{kk})(\underline{x}_k \underline{\beta})^2 \right] +$$

$$+ 2 \left[v_{ij}\, (\underline{x}_i\, \underline{\beta})(\underline{x}_j\, \underline{\beta}) + v_{ik}\, (\underline{x}_i\, \underline{\beta})(\underline{x}_k\, \underline{\beta}) + v_{jk}\, (\underline{x}_j\, \underline{\beta})\, (\underline{x}_k\, \underline{\beta}) \right]$$

Por lo tanto, sustituyendo las distintas expresiones obtenidas que componen el sumatorio que se ha denotado por SUMA2, se obtiene:

$$\text{SUMA2} = 2\left[v_{ij}\, y_i\, y_j + v_{ik}\, y_i\, y_k + v_{jk}\, y_j\, y_k\right] +$$

$$+ \; 2\, y_i\left[(1-v_{ii})\,(\underline{x}_i\,\underline{\beta}) - v_{ij}\,(\underline{x}_j\,\underline{\beta}) - v_{ik}\,(\underline{x}_k\,\underline{\beta})\right] +$$

$$+ \; 2\, y_j\left[(1-v_{jj})\,(\underline{x}_j\,\underline{\beta}) - v_{ji}\,(\underline{x}_i\,\underline{\beta}) - v_{jk}\,(\underline{x}_k\,\underline{\beta})\right] +$$

$$+ \; 2\, y_k\left[(1-v_{kk})\,(\underline{x}_k\,\underline{\beta}) - v_{ki}\,(\underline{x}_i\,\underline{\beta}) - v_{kj}\,(\underline{x}_j\,\underline{\beta})\right] +$$

$$+ \sum_{s\neq i\,,j,k}(1-v_{ss})\,(\underline{x}_s\underline{\beta})^2 - (\underline{x}_i\underline{\beta})\left[(1-v_{ii})(\underline{x}_i\underline{\beta}) - v_{ij}(\underline{x}_j\underline{\beta}) - v_{ik}(\underline{x}_k\underline{\beta})\right] -$$

$$- (\underline{x}_j\underline{\beta})\left[(1-v_{jj})(\underline{x}_j\underline{\beta}) - v_{ji}(\underline{x}_i\underline{\beta}) - v_{jk}(\underline{x}_k\underline{\beta})\right] -$$

$$- (\underline{x}_k\underline{\beta})\left[(1-v_{kk})(\underline{x}_k\underline{\beta}) - v_{ki}(\underline{x}_i\underline{\beta}) - v_{kj}(\underline{x}_j\underline{\beta})\right]$$

En consecuencia,

$$\text{SUMA2} = \sum_{s\neq i\,,j,k}(1-v_{ss})\,(\underline{x}_s\,\underline{\beta})^2 + (1-v_{ii})\left[2\,y_i\,(\underline{x}_i\,\underline{\beta}) - (\underline{x}_i\,\underline{\beta})^2\right] +$$

$$+ (1-v_{jj})\left[2\,y_j\,(\underline{x}_j\,\underline{\beta}) - (\underline{x}_j\,\underline{\beta})^2\right] + (1-v_{kk})\left[2\,y_k\,(\underline{x}_k\,\underline{\beta}) - (\underline{x}_k\,\underline{\beta})^2\right] +$$

$$+ 2\,v_{ij}\left[y_i - \underline{x}_i\underline{\beta}\right]\left[y_j - \underline{x}_j\underline{\beta}\right] + 2\,v_{ik}\left[y_i - \underline{x}_i\underline{\beta}\right]\left[y_k - \underline{x}_k\underline{\beta}\right] +$$

$$+ 2\,v_{jk}\left[y_j - \underline{x}_j\underline{\beta}\right]\left[y_k - \underline{x}_k\underline{\beta}\right]$$

Obteniéndose finalmente, por la igualdad (*), que

$$E[\;(n-r)\hat{\sigma}^2 \;|\; Y_i = y_i, Y_j = y_j,\; Y_k = y_k] = [n-3-r+v_{ii}+v_{jj}+v_{kk}]\;\sigma^2$$

$$+ \;(1-v_{ii})\;[y_i - \underline{x}_i\;\underline{\beta}]^2 + (1-v_{jj})\;[y_j - \underline{x}_j\;\underline{\beta}]^2 + (1-v_{kk})\;[y_k - \underline{x}_k\;\underline{\beta}]^2 \;-$$

$$- \;\; 2\left[v_{ij}\;[y_i - \underline{x}_i\;\underline{\beta}]\;[y_j - \underline{x}_j\;\underline{\beta}] + v_{ik}\;[y_i - (\underline{x}_i\;\underline{\beta})]\;[y_k - \underline{x}_k\;\underline{\beta}] +\right.$$

$$+ v_{jk} [y_j - \underline{x}_j \underline{\beta}] [y_k - (\underline{x}_k \underline{\beta})] \Big]$$

Pudiéndose comprobar, por tanto, el enunciado del lema.

■

TEOREMA 2.2.8.- Sea $\hat{\sigma}^2$ el estimador insesgado mínimo-cuadrático de σ^2, entonces

$$\hat{\sigma}^2 = \sigma^2 + \sum_i \frac{1-v_{ii}}{n-r} \left[[Y_i - \underline{x}_i \underline{\beta}]^2 - \sigma^2 \right] +$$

$$+ \sum\sum_{i<j} \frac{-2v_{ij}}{n-r} [Y_i - \underline{x}_i \underline{\beta}] [Y_j - \underline{x}_j \underline{\beta}]$$

<u>Demostración</u>.- Evidente, dado que

$$\sum_i (1-v_{ii}) = n-r$$

■

En consecuencia, la descomposición de Efron y Stein del estimador de σ^2 queda dependiendo de los términos de primer y segundo orden, en contraposición a la descomposición de los estimadores insesgados obtenidos de las funciones lineales estimables $\Lambda \underline{\beta}$, en la que tan sólo aparecían los términos de primer orden.

2.3.- DESCOMPOSICION DE LOS ESTIMADORES EN MODELOS PARTICULARES DEL MODELO LINEAL GENERAL

Los resultados obtenidos en los dos apartados anteriores para el Modelo Lineal General (Mod.1), pueden ser trasladados a modelos que se derivan de él. Se estudiarán el Modelo de Regresión Lineal Múltiple y el Modelo de ANOVA de un factor.

2.3.1.- DESCOMPOSICION DE LOS ESTIMADORES EN EL MODELO DE REGRESION LINEAL MULTIPLE.

El Modelo de Regresión Lineal Múltiple (**Mod.1-R**) puede ser formulado como el Modelo Lineal General bajo hipótesis de normalidad en el que la matriz **X** es de rango total [rango(X)=p]. Lo que permite que exista la inversa de la matriz $S = X'X$, y el estimador mínimo cuadrático de $\underline{\beta}$ será único, $\hat{\underline{\beta}} = S^{-1}X'Y$, insesgado, y cualquier función lineal del vector paramétrico es linealmente estimable.

Así, tomando $\Lambda = I$ en la descomposición obtenida para el modelo Mod.1 , se puede enunciar el siguiente lema:

LEMA 2.3.1 .- En el modelo Mod.1-R, se tiene:

- $A_i[\, y_i \,;\, \hat{\underline{\beta}} \,] = S^{-1}\, \underline{x}_i' \,[\, y_i - \underline{x}_i\, \underline{\beta} \,]$ $\quad i=1,\dots,n$
- $B_{ij}[\, y_i, y_j \,;\, \hat{\underline{\beta}} \,] = 0$ $\quad i,j=1,\dots,n \;,\; i\neq j$

<u>Demostración</u>.- En base al resultado 2.2.2

Y por lo tanto, el teorema:

TEOREMA 2.3.2.- En el Mod.1-R, el estimador de máxima verosimilitud del vector paramétrico $\underline{\beta}$ se puede descomponer de la siguiente forma:

$$\hat{\underline{\beta}} = \underline{\beta} + \sum_{i} S^{-1} \underline{x}'_i [Y_i - \underline{x}_i \underline{\beta}]$$

Demostración.- En base al resultado 2.2.3

De este último resultado se pueden extraer las siguientes consecuencias:

COROLARIO 2.3.3.- En el Mod.1-R se verifican:

1.- Si $\hat{Y} = X \hat{\underline{\beta}}$, entonces

$$\hat{Y} = X \underline{\beta} + \sum_{i} \underline{V}_i [Y_i - \underline{x}_i \underline{\beta}]$$

donde $\underline{V}_i$ es la i-ésima columna de la matriz $V = X S^{-1} X'$.

2.- Dado un vector $\underline{x}'_0$ de observaciones de las variables explicativas, $\hat{Y}_0 = \underline{x}_0 \hat{\underline{\beta}}$ sería la predicción de la variable aleatoria Y para tales valores de las variables predictoras. Así:

$$\hat{Y}_0 = \underline{x}_0 \underline{\beta} + \sum_{i} \underline{x}_0 S^{-1} \underline{x}'_i [Y_i - \underline{x}_i \underline{\beta}]$$

3.- Si $\hat{\beta}_k$ es la k-ésima componente de $\hat{\underline{\beta}}$, k=1,..,p, entonces

$$\hat{\beta}_k = \beta_k + \sum_{i} \underline{\delta}'_k S^{-1} \underline{x}'_i [Y_i - \underline{x}_i \underline{\beta}]$$

Para el estimador de σ^2, la descomposición que se obtiene es idéntica al modelo lineal general, es decir,

$$\hat{\sigma}^2 = \sigma^2 + \sum_i \frac{1-v_{ii}}{n-p} \left[[Y_i - \underline{x}_i \underline{\beta}]^2 - \sigma^2 \right] -$$

$$- \sum_{i<j}\sum \frac{2\, v_{ij}}{n-p} [Y_i - \underline{x}_i \underline{\beta}] [Y_j - \underline{x}_j \underline{\beta}]$$

con la salvedad de que en este caso $v_{ii} = \underline{x}_i\ S^{-1}\ \underline{x}_i'$.

2.3.2.- DESCOMPOSICION DE LOS ESTIMADORES DEL ANOVA DE UN FACTOR

El modelo ANOVA de un factor con k tratamientos **(Mod.1-A)**, en el que se han tomado n_j observaciones para el j-ésimo tratamiento, viene representado por

$$y_{js} = \mu + \alpha_j + \varepsilon_{js}$$

donde $E\ [\varepsilon_{js}] = 0$, $Var[\varepsilon_{js}] = \sigma^2$, independientes

Esta representación se puede reformular como el modelo lineal general y ser también representado por:

$$\mathbf{Y = X\ \underline{\beta} + \underline{\varepsilon}\quad ,\quad E\ [\ \underline{\varepsilon}\ ] = 0\ ,\ Var\ [\underline{\varepsilon}\ ] = \sigma^2 I}$$

donde $Y = [y_{11}..y_{1n_1}\ y_{21}..y_{2n_2}\ \dots\ y_{k1}..y_{kn_k}]'$

$\underline{\beta} = [\ \mu\ \ \alpha_1 \dots \alpha_k\ ]'$

$\underline{\varepsilon} = [\varepsilon_{11}..\varepsilon_{1n_1}\ \varepsilon_{21}..\varepsilon_{2n_2}\ \dots\ \varepsilon_{k1}..\varepsilon_{kn_k}]'$

y X , a la que se denomina matriz de diseño, es de dimensiones

$n\times(k+1)$, $n = \Sigma\, n_j$, de rango k

$$
X = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{array}{l} \left.\right]n_1 \\ \\ \\ \left.\right]n_2 \\ \\ \\ \\ \left.\right]n_k \end{array} = \left[\underline{x}'_{11} \dots \underline{x}'_{1n_1}\ \underline{x}'_{21} \dots \underline{x}'_{2n_2} \dots\dots \underline{x}'_{k1} \dots \underline{x}'_{kn_k}\right]'
$$

LEMA 2.3.4.- En el Mod.1-A se tiene:

1.-
$$
S = X'X = \begin{bmatrix} n & n_1 & n_2 & \dots & n_k \\ n_1 & n_1 & 0 & \dots & 0 \\ n_2 & 0 & n_2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ n_k & 0 & 0 & \dots & n_k \end{bmatrix}
$$

2.- La matriz

$$
S^- = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & D \end{array}\right] \quad \text{donde } \mathbf{D}=\text{diag}\left[\frac{1}{n_1}, \frac{1}{n_2}, \dots, \frac{1}{n_k}\right]
$$

es una inversa generalizada de S.

<u>Demostración</u>.- [Kshirsagar (1983)]

■

La función lineal definida por la matriz

$$
\Lambda_o = \begin{bmatrix} 1 & 1 & 0 & .. & 0 \\ 1 & 0 & 1 & .. & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & & 1 \end{bmatrix}
$$

es linealmente estimable, siendo el estimador cuyas componentes son los BLUEs de la componentes de la función lineal:

$$\hat{\underline{\eta}} = \Lambda_0 \hat{\underline{\beta}} = \left[(\widehat{\mu+\alpha_1}) \quad (\widehat{\mu+\alpha_2}) \; .... \; (\widehat{\mu+\alpha_k}) \right]' = \left[\bar{y}_{1.} \quad \bar{y}_{2.} \; .... \; \bar{y}_{k.} \right]'$$

donde

$$\bar{y}_{j.} = \frac{1}{n_j} \sum_{s=1}^{n_j} y_{js}$$

Entonces se puede comprobar, sustituyendo en la descomposición de $\Lambda\hat{\underline{\beta}}$, el siguiente teorema

TEOREMA 2.3.5.- En el Mod.1-A se tiene

$$\hat{\underline{\eta}} = \begin{bmatrix} \bar{y}_{1.} \\ \bar{y}_{2.} \\ \vdots \\ \bar{y}_{k.} \end{bmatrix} = \begin{bmatrix} \mu+\alpha_1 \\ \mu+\alpha_2 \\ \vdots \\ \mu+\alpha_k \end{bmatrix} + \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{1}{n_i} \underline{\delta}_i \left[Y_{ij} - (\mu+\alpha_i) \right]$$

<u>Demostración</u>.- En base al resultado 2.2.3

Por otra parte, otras funciones linealmente estimables de interés en este modelo son las denominadas combinaciones contrastes de los efectos de los tratamientos:

$$\sum_{s=1}^{k} c_s \alpha_s \qquad \text{tal que} \qquad \sum_{s=1}^{k} c_s = 0$$

Estas pueden ser representadas en foma vectorial como sigue:

$$[0 \; c_1 \; c_2 \; ... \; c_k \;] \hat{\underline{\beta}} = [\; 0 \quad c' \;] \hat{\underline{\beta}} = \sum_{s=1}^{k} c_s \hat{\alpha}_s = \sum_{s=1}^{k} c_s \bar{y}_{s.}$$

Por tanto, puede enunciarse el siguiente resultado referente a la descomposición de su BLUE.

TEOREMA 2.3.6.- En el Mod.1-A, se tiene

$$\sum_{s=1}^{k} c_s \hat{\alpha}_s = \sum_{s=1}^{k} c_s \alpha_s + \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{1}{n_i} c_i [Y_{ij} - (\mu + \alpha_i)]$$

Demostración.- En base al resultado 2.2.3

Para la descomposición del estimador de la varianza poblacional, es necesario obtener la expresión de los elementos de la matriz de predicción **V** para este caso particular del Modelo Lineal General. Se puede comprobar que

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \Theta & \dots & \Theta \\ \Theta & \mathbf{V}_2 & \dots & \Theta \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \Theta & \Theta & \dots & \mathbf{V}_k \end{bmatrix}$$

donde $\mathbf{V}_i = \frac{1}{n_i} \mathbf{E}_{(n_i x n_i)}$, siendo $\mathbf{E}_{(n_i x n_i)}$ la matriz de unos de dimensiones subindicadas. Por lo tanto, se obtiene

LEMA 2.3.7 .- En el Mod.1-A se verifica:

1.-

$$A_{ij}[y_{ij};\hat{\sigma}^2] = E\left[\hat{\sigma}^2 \mid Y_{ij}=y_{ij}\right] - E\left[\hat{\sigma}^2\right] = \frac{n_i - 1}{n_i (n-k)} \left[[y_{ij} - (\mu+\alpha_i)]^2 - \sigma^2 \right]$$

para todo $i=1,\dots,k$; $j=1,\dots,n_i$

2.-Los elementos de segundo orden de la descomposición, $B_{ij,i'j'}[y_{ij},y_{i'j'};\hat{\sigma}^2]$, para $i \neq i'$ son nulos . Y para el caso $i=i'$, se tiene:

$$B_{ij,ij'}\ [y_{ij},y_{ij'}\ ;\hat{\sigma}^2] = \frac{-2}{n_i(n-k)}\left[[y_{ij}-(\mu+\alpha_i)]\ [y_{ij'}-(\mu+\alpha_i)]\right]$$

para todo $i=1,\dots,k$; $j,j'=1,\dots,n_i$, $j<j'$

Demostración.- En base a los resultados 2.2.5 y 2.2.6

Obteniéndose finalmente:

TEOREMA 2.3.8.- En el Mod.1-A, el estimador insesgado de σ^2 se puede descomponer de la siguiente forma:

$$\hat{\sigma}^2 = \sigma^2 + \sum_{i=1}^{k}\ \sum_{j=1}^{n_i} \frac{n_i-1}{n_i(n-k)}\left[[Y_{ij}-(\mu+\alpha_i)]^2 - \sigma^2\right] -$$

$$-\sum_{i=1}^{k}\ \sum_{1\le j<j'\le n_i} \frac{2}{n_i(n-k)}\left[[Y_{ij}-(\mu+\alpha_i)]\ [Y_{ij'}-(\mu+\alpha_i)]\right]$$

Demostración.- En base al resultado 2.2.8

2.4.- MODELO LINEAL GENERAL PERTURBADO

Según la descripción del problema del análisis de influencia realizada por D. Cook [1987] comentada en el capítulo anterior, es necesario estudiar el esquema de perturbación en Modelo Lineal General, para posteriormente comparar los resultados. La perturbación considerada está basada en dar un peso variable a la i-ésima observación.

Así, se denominará Modelo Perturbado **(Mod.2)** al Modelo Lineal General bajo hipótesis de normalidad con la particularidad de que

$$\underline{\varepsilon} \simeq N_n[\, 0, \sigma^2 \mathbf{W}]$$

donde

$$\mathbf{W} = \mathrm{diag}[1,1,..,\underset{i)}{1/\omega},...,1] \ , \ \omega > 0.$$

LEMA 2.4.1.- Los estimadores de máxima verosimilitud para el Mod.2 son

$$\hat{\underline{\beta}}_\omega = S_\omega^- X' W^{-1} Y + (I_p - H_\omega)\underline{z} \quad , \ \underline{z} \in \mathbb{R}^p \ \text{arbitrario}$$

donde $S_\omega = X' W^{-1} X$ y $H_\omega = S_\omega^- S_\omega$.

$$\tilde{\sigma}_\omega^2 = \frac{1}{n}\left[Y - X\hat{\underline{\beta}}_\omega \right]' W^{-1} \left[Y - X\hat{\underline{\beta}}_\omega \right]$$

Se observa que al igual que ocurría en el **Mod.1-N**, el estimador de $\underline{\beta}$ no será único, y sí el de la varianza σ^2.

En consecuencia, y dado que el objetivo es poder comparar resultados con el modelo no perturbado, Mod.1-N, , se deben estudiar en el modelo perturbado Mod.2 las funciones lineales estimables del modelo Mod.1-N, del tipo $\Lambda\ \underline{\beta}$, donde Λ es una matriz de dimensiones qxp de rango q.

LEMA 2.4.2 .- Sean **A** y **L** matrices de dimensiones pxp y mxp ($m\leq p$), respectivamente, y tales que **L** pertenece al espacio fila generado por la matriz **A**. Entonces, si $\gamma\in\mathbb{R}$, $\gamma\neq 0$, y existe la inversa de la matriz $[\mathbf{I} + \gamma\ \mathbf{L}\ \mathbf{A}^{-}\ \mathbf{L}']$, entonces las matrices

$$\left[\mathbf{A}^{-} - \gamma\ \mathbf{A}^{-}\ \mathbf{L}'\ [\mathbf{I} + \gamma\ \mathbf{L}\ \mathbf{A}^{-}\ \mathbf{L}']^{-1}\mathbf{L}\ \mathbf{A}^{-}\right]$$

son inversas generalizadas de $[\mathbf{A} + \gamma\ \mathbf{L}'\mathbf{L}]$

Demostración.-

La demostración es evidente en base a laa igualdades

$$\mathbf{L}\ \mathbf{A}^{-}\ \mathbf{A} = \mathbf{L}\ ,\quad \mathbf{A}\ \mathbf{A}^{-}\ \mathbf{L}' = \mathbf{L}'$$

que la verifican por pertenecer **L** al espacio fila de **A**.

■

LEMA 2.4.3.- 1.- Si $v_{ii}\neq 1$, elemento i-ésimo diagonal de **V**, entonces

$$1+(\omega-1)v_{ii}\neq 0 \qquad \forall\ \omega > 0.$$

2.- Si $\underline{\lambda}'\underline{\beta}$ es linealmente estimable en el modelo postulado Mod.1-N entonces $\underline{\lambda}'\mathbf{H}_{\omega}=\underline{\lambda}'$.

3.- $\mathbf{X}\ \mathbf{S}_{\omega}^{-}\ \mathbf{X}'$ para cualquier elección de la inversa generalizada de $\mathbf{S}_{\omega}$

Demostración.-

1.- Evidente, dado que por ser **V** idempotente, entonces $0 \le v_{ii} \le 1$.

2.- Inicialmente se calcula S_ω^- utilizando el lema anterior. Como

$$S_\omega = X' \, W^{-1} \, X = X'X + (\omega-1) \, \underline{x}_i' \, \underline{x}_i$$

entonces considerando $L = \underline{x}_i$, $\gamma = \omega-1$, $A = S = X'X$, y dado que $[I + \gamma \, L \, A^- \, L'] = 1+(\omega-1)v_{ii}$ que es no nulo por el primer apartado, se puede aplicar el lema anterior, obteniéndose:

$$S_\omega^- = S^- - \frac{\omega - 1}{1+(\omega-1)v_{ii}} \, S^- \, \underline{x}_i' \, \underline{x}_i \, S^-$$

Así,

$$\underline{\lambda}'H_\omega = \underline{\lambda}' \left[S^- - \frac{\omega - 1}{1+(\omega-1)v_{ii}} \, S^- \, \underline{x}_i' \, \underline{x}_i \, S^- \right] \left[S + (\omega-1) \, \underline{x}_i' \, \underline{x}_i \right] =$$

$$= \underline{\lambda}' \left[H + \left[(\omega-1) \, \frac{\omega - 1}{1+(\omega-1)v_{ii}} - \frac{(\omega-1)^2 \, v_{ii}}{1+(\omega-1)v_{ii}} \right] S^- \, \underline{x}_i' \, \underline{x}_i \right] = \underline{\lambda}'H = \underline{\lambda}'$$

3.- Evidente por ser $X \, S^- \, X'$ única para cualquier elección de S^-

■

TEOREMA 2.4.4.- Sea $\underline{\lambda}'\underline{\beta}$ linealmente estimable en el modelo postulado, y sean $\hat{\underline{\beta}}_\omega$ y $\hat{\underline{\beta}}_\omega^*$ dos estimadores cualesquiera de máxima verosimilitud de $\underline{\beta}$ en el modelo perturbado. Entonces

$$\underline{\lambda}'\hat{\underline{\beta}}_\omega = \underline{\lambda}'\hat{\underline{\beta}}_\omega^*$$

y pueden ser expresados por:

$$\underline{\lambda}'\hat{\underline{\beta}}_\omega = \underline{\lambda}'\hat{\underline{\beta}} - \frac{\omega - 1}{1+(\omega-1)v_{ii}} \, \underline{\lambda}' \, S^- \, \underline{x}_i' \left[Y_i - \underline{x}_i \, \hat{\underline{\beta}} \right]$$

Demostración.-

Por ser $\underline{\lambda}'\underline{\beta}$ estimable en Mod.1-N se tiene por el apartado 2 del lema anterior que

$$\underline{\lambda}'\hat{\underline{\beta}}_\omega = \underline{\lambda}'\ S_\omega^-\ X'\ W^{-1}\ Y$$
$$\underline{\lambda}'\hat{\underline{\beta}}_\omega^* = \underline{\lambda}'\ T_\omega^-\ X'\ W^{-1}\ Y$$

donde S_ω^- y T_ω^- son dos inversas generalizadas de S_ω . Y además, por el resultado 2.1.4 se tiene $\underline{\lambda}' = u'\ X$, luego

$$\underline{\lambda}'\hat{\underline{\beta}}_\omega = u'X\ S_\omega^-\ X'\ W^{-1}\ Y$$
$$\underline{\lambda}'\hat{\underline{\beta}}_\omega^* = u'X\ T_\omega^-\ X'\ W^{-1}\ Y$$

que son iguales por el apartado 3 del lema 2.4.4.

Por otra parte,

$$\underline{\lambda}'\hat{\underline{\beta}}_\omega = \underline{\lambda}'\ S_\omega^-\ X'\ W^{-1}\ Y =$$

$$= \left[\underline{\lambda}'\ S^- - \frac{\omega-1}{1+(\omega-1)v_{ii}}\ \underline{\lambda}'\ S^-\ \underline{x}_i'\ \underline{x}_i\ S^-\right]\left[\ X'\ Y - (\omega-1)\ \underline{x}_i'\ Y_i\right] =$$

$$= \underline{\lambda}'\hat{\underline{\beta}} - \frac{\omega-1}{1+(\omega-1)v_{ii}}\ \underline{\lambda}'\ S^-\ \underline{x}_i'\ \underline{x}_i\ \hat{\underline{\beta}} + (\omega-1)\left[1- \frac{(\omega-1)\ v_{ii}}{1+(\omega-1)v_{ii}}\right]\ \underline{\lambda}'S^-\underline{x}_i'Y_i =$$

$$= \underline{\lambda}'\hat{\underline{\beta}} - \frac{\omega-1}{1+(\omega-1)v_{ii}}\ \underline{\lambda}'\ S^-\ \underline{x}_i'\left[\ Y_i - \underline{x}_i\ \hat{\underline{\beta}}\ \right]$$

■

TEOREMA 2.4.5.- Si $\underline{\lambda}'\underline{\beta}$ es linealmente estimable en el modelo postulado, entonces el BLUE en el modelo perturbado Mod.2 de dicha función lineal viene dado por

$$\underline{\lambda}'\hat{\underline{\beta}}_\omega$$

para cualquier $\hat{\underline{\beta}}_\omega$ estimador de máxima verosimilitud de $\underline{\beta}$ en dicho modelo perturbado.

demostración.-

Es un estimador insesgado dado que:

$$E\,[\,\underline{\lambda}'\hat{\underline{\beta}}_{\omega}\,] = \underline{\lambda}'\,S_{\omega}^{-}\,X'\,W^{-1}\,E\,[\,Y\,] = \underline{\lambda}'\,S_{\omega}^{-}\,S_{\omega}\,\underline{\beta} = \underline{\lambda}'\underline{\beta}$$

Por otra parte, sea $l \in \mathbb{R}^{p}$, tal que

$$E\,[\,l'Y\,] = \underline{\lambda}'\underline{\beta}$$

Como $E\,[\,l'Y\,] = l'\,X\,\underline{\beta}$, entonces $l'\,X = \underline{\lambda}'$.

Por otra parte,

$$Var[l'Y] = Var[l'Y - \underline{\lambda}'\hat{\underline{\beta}}_{\omega}] + Var\,[\underline{\lambda}'\hat{\underline{\beta}}_{\omega}] - 2\,Cov[(l'Y - \underline{\lambda}'\hat{\underline{\beta}}_{\omega})\,,\,\underline{\lambda}'\hat{\underline{\beta}}_{\omega}]$$

Desarrollando la covarianza, se tiene

$$Cov[(l'Y - \underline{\lambda}'\hat{\underline{\beta}}_{\omega})\,,\,\underline{\lambda}'\hat{\underline{\beta}}_{\omega}] = Cov\left[(l' - \underline{\lambda}'S_{\omega}^{-}\,X'W^{-1})\,Y\,,\,\underline{\lambda}'S_{\omega}^{-}\,X'W^{-1}\text{-}\,Y\right] =$$

$$[l' - \underline{\lambda}'S_{\omega}^{-}\,X'W^{-1}]\,Var(Y)\,[\underline{\lambda}'S_{\omega}^{-}\,X'W^{-1}]' = \sigma^{2}\,[l' - \underline{\lambda}'S_{\omega}^{-}\,X'W^{-1}]\,X\,S_{\omega}^{-}\,\underline{\lambda} =$$

$$= \sigma^{2}\,[\,l'\,X\,S_{\omega}^{-}\,\underline{\lambda} - \underline{\lambda}'S_{\omega}^{-}\,\underline{\lambda}\,] = 0$$

Por lo tanto,

$$Var[l'Y] = Var[l'Y - \underline{\lambda}'\hat{\underline{\beta}}_{\omega}] + Var\,[\underline{\lambda}'\hat{\underline{\beta}}_{\omega}]$$

lo cual implica que

$$Var[l'Y] \geq Var\,[\underline{\lambda}'\hat{\underline{\beta}}_{\omega}] \quad \text{para todo } l \in \mathbb{R}^{p}$$

Además como se tiene que $E[l'Y - \underline{\lambda}'\hat{\underline{\beta}}_{\omega}]=0$, y la igualdad se alcanza cuando $Var[l'Y - \underline{\lambda}'\hat{\underline{\beta}}_{\omega}]=0$ entonces han de ser idénticos con probabilidad uno.

■

En resumen, dada una función linealmente estimable en Mod.1-N postulado, del tipo $\Lambda\underline{\beta}$, entonces el vector aleatorio

$$\Lambda\hat{\underline{\beta}}_{\omega} = \Lambda\hat{\underline{\beta}} + \frac{\omega - 1}{1+(\omega-1)v_{ii}} \Lambda S^{-} \underline{x}_i' \left[Y_i - \underline{x}_i \hat{\underline{\beta}} \right]$$

es único para cualquier solución de la ecuación de verosimilitud del modelo perturbado, es un estimador insesgado de $\Lambda\underline{\beta}$ en dicho modelo, y tal que cada componente es el BLUE de cada componente de $\Lambda\underline{\beta}$.

LEMA 2.4.6.- Sea $\underline{\xi} \simeq N_n[0 ,\sigma^2 \Sigma]$ y sea **A** una matriz simétrica, entonces

$$(\sigma^2/2)\ \underline{\xi}'A\ \underline{\xi} \simeq \chi^2_f \quad \text{sii} \quad A\Sigma \text{ es idempotente de rango f}$$

Demostración.- Graybill [1961]

■

LEMA 2.4.7.- En el Mod.2 se tiene

1.- $[I - X S_{\omega}^{-} X'W^{-1}]'\ W^{-1}\ [I - X S_{\omega}^{-} X'W^{-1}] = [I - W^{-1}X S_{\omega}^{-} X']\ W^{-1}$
$= W^{-1}\ [I - X S_{\omega}^{-} X'\ W^{-1}]$

y es única para cualquier elección de S_{ω}^{-}

2.- $[I - W^{-1}X S_{\omega}^{-} X']$ es idempotente de rango n-r

Demostración.-

1.- Sea $A = [I - X S_{\omega}^{-} X'W^{-1}]'\ W^{-1}\ [I - X S_{\omega}^{-} X'W^{-1}]$

Simplificando la expresión de **A**,

$$A = [I - X S_{\omega}^{-} X'W^{-1}]'\ W^{-1}\ [I - X S_{\omega}^{-} X'W^{-1}] =$$
$$= [I - X S_{\omega}^{-} X'W^{-1}]'\ W^{-1} - [I - W^{-1}X\ (S_{\omega}^{-})'X']\ W^{-1}X\ S_{\omega}^{-}\ X'W^{-1} =$$

$$= [I - W^{-1}X\, S_{\omega}^{-}\, X'] \, W^{-1} - W^{-1}X\, S_{\omega}^{-}\, X'W^{-1} + W^{-1}X\, S_{\omega}^{-}\, X'\, W^{-1}X\, S_{\omega}^{-}\, X'W^{-1} =$$

$$= [I - W^{-1}X\, S_{\omega}^{-}\, X'] \, W^{-1} - W^{-1}X\, S_{\omega}^{-}\, X'W^{-1} + W^{-1}X\, S_{\omega}^{-}\, S_{\omega}\, S_{\omega}^{-}\, X'W^{-1} =$$

$$= [I - W^{-1}X\, S_{\omega}^{-}\, X'] \, W^{-1}$$

De forma análoga se tiene:

$$A = [I - X\, S_{\omega}^{-}\, X'W^{-1}]'\, W^{-1}\, [I - X\, S_{\omega}^{-}\, X'W^{-1}] =$$

$$= W^{-1}[I - X\, S_{\omega}^{-}\, X'W^{-1}] - W^{-1}X\, S_{\omega}^{-}\, X'W^{-1}\, [I - X\, S_{\omega}^{-}\, X'W^{-1}] =$$

$$= W^{-1}[I - X\, S_{\omega}^{-}\, X'W^{-1}].$$

Luego

$$A = [I - W^{-1}X\, S_{\omega}^{-}\, X'] \, W^{-1} = W^{-1}\, [I - X\, S_{\omega}^{-}\, X'\, W^{-1}]$$

y es única por el apartado 3 de 2.4.3.

2.- El caracter idempotente es evidente. Por otra parte

$$\text{traza}[I - W^{-1}X\, S_{\omega}^{-}\, X'] = n - \text{traza}\,(W^{-1}X\, S_{\omega}^{-}\, X')$$

Como

$$S_{\omega}^{-} = S^{-} - \frac{\omega - 1}{1+(\omega-1)v_{ii}}\, S^{-}\, \underline{x}_i'\, \underline{x}_i\, S^{-}$$

entonces

$$W^{-1}X\, S_{\omega}^{-}\, X' = W^{-1}V - \frac{\omega - 1}{1+(\omega-1)v_{ii}}\, W^{-1}\, \underline{V}_i\, \underline{V}_i'$$

donde $\underline{V}_i$ es la i-ésima columna de la matriz de predicción **V**.

Calculando la traza de cada uno de los dos sumandos, tenemos:

* $\text{traza}(W^{-1}V) = \text{traza}(V) + (\omega-1)v_{ii} = r + (\omega-1)v_{ii}$

* $\text{traza}\left[\frac{\omega - 1}{1+(\omega-1)v_{ii}}\, W^{-1}\, \underline{V}_i\, \underline{V}_i' \right] = \frac{\omega - 1}{1+(\omega-1)v_{ii}}\, \text{traza}\left[W^{-1}\, \underline{V}_i\, \underline{V}_i' \right] =$

$$= \frac{\omega - 1}{1+(\omega-1)v_{ii}} \left[\sum_j v_{ij}^2 + (\omega-1)v_{ii}^2 \right] = \frac{\omega - 1}{1+(\omega-1)v_{ii}} \left[v_{ii}^2 + (\omega-1)v_{ii}^2 \right] =$$

$$= (\omega-1)\ v_{ii}$$

Por lo tanto, traza ($W^{-1}X\ S_\omega^- X'$) = r , y así

$$\text{traza}[I - W^{-1}X\ S_\omega^-\ X'] = n- r$$

■

TEOREMA 2.4.8.- En el modelo perturbado (Mod.2),

$$\hat{\sigma}_\omega^2 = \frac{1}{n-r} \left[Y - X\ \hat{\underline{\beta}}_\omega \right]' W^{-1} \left[Y - X\ \hat{\underline{\beta}}_\omega \right]$$

es un estimador insesgasdo de σ^2 y es independiente de la elección de $\hat{\underline{\beta}}_\omega$

Demostración.- $\left[Y - X\ \hat{\underline{\beta}}_\omega \right]' W^{-1} \left[Y - X\ \hat{\underline{\beta}}_\omega \right] =$

$$= Y'\ [I - X\ S_\omega^-\ X'W^{-1}]'\ W^{-1}\ [I - X\ S_\omega^-\ X'W^{-1}]\ Y = Y'\ A\ Y$$

siendo $A = [I - X\ S_\omega^-\ X'W^{-1}]'\ W^{-1}\ [I - X\ S_\omega^-\ X'W^{-1}]$

Por el lema anterior

$$A= [I - W^{-1}X\ S_\omega^-\ X']\ W^{-1}= W^{-1}\ [I - X\ S_\omega^-\ X'\ W^{-1}]$$

Además, $AX=\Theta$, siendo Θ la matriz nula, luego

$$Y'A\ Y = [X\ \underline{\beta} + \underline{\varepsilon}]'\ A\ [X\ \underline{\beta} + \underline{\varepsilon}] = \underline{\varepsilon}'A\ \underline{\varepsilon}\ .$$

Por lo que aplicando los lemas 2.4.6 y 2.4.7 se obtiene que

$$Y'A\ Y \simeq \sigma^2\ \chi_{n-r}^2$$

verificándose el teorema.

■

Anteriormente se ha obtenido la relación existente entre el estimador de una función lineal estimable del modelo perturbado con el correspondiente al modelo postulado. De forma análoga, el siguiente teorema proporciona la relación entre los estimadores insesgados de σ^2 de los dos modelos.

TREOREMA 2.4.9.- $\hat{\sigma}^2_\omega = \hat{\sigma}^2 + \frac{1}{n-r} \frac{\omega - 1}{1+(\omega-1)v_{ii}} e_i^2$ para todo $\omega > 0$

donde e_i es la i-ésima componente del vector de resíduos **e** .

<u>Demostración</u>.-

$$Y - X\,\hat{\underline{\beta}}_\omega = Y - X\,\hat{\underline{\beta}} + X\,\hat{\underline{\beta}} - X\,\hat{\underline{\beta}}_\omega = e - X\,[\,\hat{\underline{\beta}} - \hat{\underline{\beta}}_\omega\,] =$$

$$= e - \frac{\omega - 1}{1+(\omega-1)v_{ii}}\, X\, S^-\, \underline{x}_i'\, e_i$$

Luego,

$$\left[Y - X\,\hat{\underline{\beta}}_\omega\right]' W^{-1}\left[Y - X\,\hat{\underline{\beta}}_\omega\right] = e'\, W^{-1} e \;-\; \frac{2(\omega - 1)}{1+(\omega-1)v_{ii}}\, e'\, W^{-1} X\, S^-\, \underline{x}_i'\, e_i \;+$$

$$+ \left[\frac{\omega - 1}{1+(\omega-1)v_{ii}}\right]^2 e_i^2\, \underline{x}_i\, S^-\, X'\, W^{-1}\, S^-\, \underline{x}_i'$$

Desarrollando los tres sumandos de la expresión anterior se tiene:

* $e'\, W^{-1} e = e'\, e + (\omega-1)\, e_i^2$

* $e'\, W^{-1} X\, S^-\, \underline{x}_i'\, e_i = e_i \sum_j e_j\, v_{ij} + (\omega-1)\, e_i^2\, v_{ii} = (\omega-1)\, e_i^2\, v_{ii}$

Por lo tanto:

$$\frac{2(\omega - 1)}{1+(\omega-1)v_{ii}} \; e' \; W^{-1}X \; S^{-} \; \underline{x}_i' \; e_i = \frac{2(\omega-1)^2 v_{ii}}{1+(\omega-1)v_{ii}} \; e_i^2$$

$$* \left[\frac{\omega - 1}{1+(\omega-1)v_{ii}} \right]^2 e_i^2 \; \underline{x}_i \; S^{-} \; X' \; W^{-1} \; S^{-} \; \underline{x}_i' =$$

$$= \left[\frac{\omega - 1}{1+(\omega-1)v_{ii}} \right]^2 e_i^2 \; \underline{x}_i \; S^{-} \; [S +(\omega-1) \; \underline{x}_i' \; \underline{x}_i] \; S^{-} \; \underline{x}_i' =$$

$$= \left[\frac{\omega - 1}{1+(\omega-1)v_{ii}} \right]^2 e_i^2 \; [v_{ii} + (\omega-1) \; v_{ii}^2] = \frac{(\omega-1)^2 v_{ii}}{1+(\omega-1)v_{ii}} \; e_i^2$$

De donde se deduce el resultado propuesto.

■

Estos resultados obtenidos en el modelo perturbado van a servir posteriormente para el análisis de influencia en el modelo Mod.1-N. No obstante, el esquema de perturbación más utilizado en la literatura es el basado en la omisión de una observación o de un conjunto de observaciones. Por ello, es conveniente desarrollar el modelo resultante de la omisión de un conjunto de observaciones, y estudiar su relación con el modelo perturbado, así como la relación entre los estimadores tanto de la varianza σ^2, como de las funciones paramétricas linealmente estimables, en ambos modelos.

2.5.- MODELO LINEAL GENERAL BAJO LA OMISION DE OBSERVACIONES

Sea el Modelo Lineal General, Mod.1, en el cual se omiten del estudio la observaciones subindicadas por la colección de subíndices:

$$\mathfrak{I} = \left\{ i_1, i_2, \ldots, i_m \right\} \subset \left\{ 1,2,\ldots,n \right\}$$

Se notará por $\mathbf{X}_{\mathfrak{I}}$ la matriz formada por las filas de $\mathbf{X}$ correspondientes a $\mathfrak{I}$ y por $\mathbf{V}_{\mathfrak{I}}$ la submatriz correspondiente a $\mathbf{V}$, siendo por tanto $\mathbf{V}_{\mathfrak{I}} = \mathbf{X}_{\mathfrak{I}}\, \mathbf{S}^{-}\, \mathbf{X}'_{\mathfrak{I}}$; asímismo, con el subíndice $(\mathfrak{I})$ se indica la omisión de las i_j-ésimas observaciones.

El modelo lineal resultante al omitir tales observaciones es:

$$\mathbf{Y}_{(\mathfrak{I})} = \mathbf{X}_{(\mathfrak{I})}\, \underline{\beta} + \underline{\varepsilon}_{(\mathfrak{I})} \qquad \textbf{(Mod.3)}$$

siendo $\mathbf{Y}_{(\mathfrak{I})}$, $\mathbf{X}_{(\mathfrak{I})}$, $\underline{\varepsilon}_{(\mathfrak{I})}$ los correspondientes subvectores o submatrices asociadas al conjunto de subíndices $\mathfrak{I}$. Para este modelo, los estimadores mínimo cuadrático vienen expresados por

$$\hat{\underline{\beta}}_{(\mathfrak{I})} = \mathbf{S}^{-}_{(\mathfrak{I})}\, \mathbf{X}'_{(\mathfrak{I})}\, \mathbf{Y}_{(\mathfrak{I})} + [\mathbf{I}_k - \mathbf{H}_{(\mathfrak{I})}]\underline{\nu} \quad , \quad \underline{\nu} \in \mathbf{R}^p$$

siendo $\mathbf{S}_{(\mathfrak{I})} = \mathbf{X}'_{(\mathfrak{I})}\mathbf{X}_{(\mathfrak{I})}$ y $\mathbf{H}_{(\mathfrak{I})} = \mathbf{S}^{-}_{(\mathfrak{I})}\mathbf{S}_{(\mathfrak{I})}$.

En este modelo se supondrá:

(i) $rg(\mathbf{X}_{\mathfrak{I}}) = m$, por lo que $rg(\mathbf{V}_{\mathfrak{I}}) = m$

(ii) $\mathbf{M}_{\mathfrak{I}} = \mathbf{I}_m - \mathbf{V}_{\mathfrak{I}}$ es no singular

(iii) $n > r+m$

Y, sin pérdida de generalidad el subconjunto $\mathfrak{I}$ de índices tomará la forma $\mathfrak{I} = \{1,2,\ldots,m\}$.

TEOREMA 2.5.1- Si $\Lambda\underline{\beta}$ es linealmente estimable en el modelo postulado, entonces también lo es en el modelo Mod.3

Demostración.- Se considera la descomposición

$$X = \begin{bmatrix} X_{\mathfrak{I}} \\ X_{(\mathfrak{I})} \end{bmatrix} \qquad V = \begin{bmatrix} V_{\mathfrak{I}} & V_a \\ V'_a & V_{(\mathfrak{I})} \end{bmatrix}$$

Al ser VX=X, se tiene $V_{\mathfrak{I}}\, X_{\mathfrak{I}} + V_a\, X_{(\mathfrak{I})} = X_{\mathfrak{I}}$. Así, si $\Lambda\underline{\beta}$ es linealmente estimable en el modelo postulado, existe una matriz **A** de dimensión qxn tal que

$$\Lambda = \mathbf{A} \begin{bmatrix} M_{\mathfrak{I}}^{-1} V_a \\ I_m \end{bmatrix} X_{(\mathfrak{I})} \Rightarrow \exists\ \mathbf{B} \,/\, \Lambda = \mathbf{B}\, X_{(\mathfrak{I})}$$

luego $\Lambda\underline{\beta}$ es linealmente estimable en el modelo Mod.3.

■

TEOREMA 2.5.2.-Los estimadores de mínimos cuadrados de $\underline{\beta}$ en Mod.3 se pueden expresar:

$$\hat{\underline{\beta}}_{(\mathfrak{I})} = S^{-}X'Y - S^{-}X'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1} e_{\mathfrak{I}} + (I_k - H)\underline{\nu}\ ,\ \underline{\nu} \in R^p \text{ arbitrario.}$$

Demostración.-

Por el lema 2.4.2, se tiene

$$S^{-}_{(\mathfrak{I})} = (\, S - X'_{\mathfrak{I}} X_{\mathfrak{I}}\,)^{-} = S^{-} + S^{-}\, X'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, X_{\mathfrak{I}}\, S^{-}$$

Por otra parte,

$$H_{(\mathfrak{I})} = H - S^{-}\, X'_{\mathfrak{I}}\, X_{\mathfrak{I}} + S^{-}\, X'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, X_{\mathfrak{I}} - S^{-}\, X'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, V_{\mathfrak{I}}\, X_{\mathfrak{I}} = H$$

Por tanto:

$$\hat{\beta}_{(\mathfrak{J})} = (S^- + S^- X'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} X_{\mathfrak{J}} S^-)(X'Y - X'_{\mathfrak{J}} Y_{\mathfrak{J}}) + (I_k - H) \underline{\nu} =$$
$$= S^- X'Y - S^- X'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} (Y_{\mathfrak{J}} - X_{\mathfrak{J}} S^- X'Y) + (I_k - H) \underline{\nu} =$$
$$= S^- X'Y - S^- X'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} e_{\mathfrak{J}} + (I_k - H) \underline{\nu}$$

■

COROLARIO 2.5.3.- En el Mod.3 se tiene:

1) $\hat{\underline{\beta}}_{(\mathfrak{J})} = \hat{\underline{\beta}} - S^- X'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} e_{\mathfrak{J}}$

2) Para $\mathfrak{J} = \{ i \}$,

$$\hat{\underline{\beta}}_{(i)} = \hat{\underline{\beta}} - \frac{1}{1-v_{ii}} S^- \underline{x}'_i [Y_i - \underline{x}_i \hat{\underline{\beta}}]$$

3) $\hat{\underline{\beta}}_{\omega} \xrightarrow[\omega \longrightarrow 0]{} \hat{\underline{\beta}}_{(i)}$

En cuanto al estimador del parámetro σ^2, se tiene

$$\hat{\sigma}^2_{(\mathfrak{J})} = \frac{1}{n-r-m} \left[Y_{(\mathfrak{J})} - X_{(\mathfrak{J})} \hat{\underline{\beta}}_{(\mathfrak{J})} \right]' \left[Y_{(\mathfrak{J})} - X_{(\mathfrak{J})} \hat{\underline{\beta}}_{(\mathfrak{J})} \right]$$

En el siguiente teorema se obtiene la relación con el estimador obtenido en el modelo postulado, Mod.1, es decir, en el modelo que contempla todas las observaciones.

TEOREMA 2.5.4.- En el modelo Mod.3

$$\hat{\sigma}^2_{(\mathfrak{J})} = \frac{1}{n-r-m} \left[(n-r)\, \hat{\sigma}^2 - e'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} e_{\mathfrak{J}} \right]$$

<u>Demostración</u>.-

$$\left[Y_{(\mathfrak{J})} - X_{(\mathfrak{J})} \hat{\underline{\beta}}_{(\mathfrak{J})} \right]' \left[Y_{(\mathfrak{J})} - X_{(\mathfrak{J})} \hat{\underline{\beta}}_{(\mathfrak{J})} \right] = Y'_{(\mathfrak{J})} Y_{(\mathfrak{J})} - Y'_{(\mathfrak{J})} X_{(\mathfrak{J})} \hat{\underline{\beta}}_{(\mathfrak{J})} =$$

$$= \left[Y'Y - Y'_{\mathfrak{J}} Y_{\mathfrak{J}}\right] - \left[Y'X - Y'_{\mathfrak{J}} X_{\mathfrak{J}}\right] \left[S^{-} X'Y - S^{-} X'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} e_{\mathfrak{J}} + (I-H) \underline{\nu}\right] =$$

$$= Y'(I-V)Y - Y'_{\mathfrak{J}} Y_{\mathfrak{J}} + \hat{\underline{\beta}} X'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} e_{\mathfrak{J}} + Y'_{\mathfrak{J}} X_{\mathfrak{J}} \hat{\underline{\beta}} - Y'_{\mathfrak{J}} V_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} e_{\mathfrak{J}} =$$

$$= Y'(I-V)Y - Y'_{\mathfrak{J}} e_{\mathfrak{J}} + \hat{\underline{\beta}} X'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} e_{\mathfrak{J}} - Y'_{\mathfrak{J}} V_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} e_{\mathfrak{J}} =$$

$$= Y'(I-V)Y - \left[Y'_{\mathfrak{J}} [I + V_{\mathfrak{J}} M_{\mathfrak{J}}^{-1}] - \hat{\underline{\beta}} X'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1}\right] e_{\mathfrak{J}} =$$

$$= Y'(I-V)Y - \left[Y'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} - \hat{\underline{\beta}} X'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1}\right] e_{\mathfrak{J}} = Y'(I-V)Y - e'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} e_{\mathfrak{J}}$$

$$= (n-r)\, \hat{\sigma}^2 - e'_{\mathfrak{J}} M_{\mathfrak{J}}^{-1} e_{\mathfrak{J}}$$

De donde se deduce el enunciado.

■

COROLARIO 2.5.5.- En el modelo Mod.3 se verifica:

1) Para $\mathfrak{J} = \{ i \}$, se tiene

$$\hat{\sigma}^2_{(i)} = \frac{1}{n-r-1} \left[(n-r)\, \hat{\sigma}^2 - e_i^2/(1-v_{ii})\right]$$

2) $$(n-r)\, \hat{\sigma}^2_{\omega} \xrightarrow[\omega \to 0]{} (n-r-1)\, \hat{\sigma}^2_{(i)}$$

2.5 .- LOS RESIDUOS EN EL MODELO LINEAL GENERAL

En la análisis de influencia en el Modelo Lineal General, el vector e de resíduos juega un papel relevante, dado que la gran mayoría de las medidas de influencia, que se obtienen a través de los distintos esquemas de perturbación y comparación, dependen de algunas de sus componentes.

Bajo el supuesto de normalidad en el modelo, la distribución del vector de resíduos $\underline{e} = (I_n - V)Y = (I_n - V)\,\underline{\varepsilon}$ se distribuye según una $N_n[0, \sigma^2(I_n - V)]$. Al depender su distribución del parámetro σ^2, el cual es generalmente desconocido, se hace necesario trabajar con alguna función de los mismos que no dependa del mencionado parámetro.

DEFINICION 2.6.1.- Se denomina vector de resíduos estandarizados al vector n-dimensional $\underline{r}$ cuyas componentes son:

$$r_i = \frac{e_i}{\hat{\sigma}\ [1 - v_{ii}]^{1/2}}$$

DEFINICION 2.6.2.- Se denomina vector de resíduos studentizados al vector n-dimensional $\underline{r}$ cuyas componentes son:

$$t_i = \frac{e_i}{\hat{\sigma}_{(i)}\ [1 - v_{ii}]^{1/2}}$$

La relación entre los resíduos estandarizados y studentizados viene dada por:

$$t_i = r_i \left[\frac{n - r - 1}{n - r - r_i^2} \right]^{1/2}$$

Luego, el cuadrado del resíduo estandarizado es una transformación monótona del cuadrado del studentizado.

LEMA 2.6.3.- Sea $\mathfrak{J} = \left\{ i_1 , i_2 , \dots , i_m \right\}$ una colección de subíndices. En el Modelo Lineal General bajo hipotesis de normalidad $\left[\underline{\varepsilon} \simeq N_n[0,\sigma^2 I] \right]$, si la submatriz $\mathbf{M}_{\mathfrak{J}}$ de $\mathbf{M}$ formada por las filas y columnas subindicadas por $\mathfrak{J}$ es de rango m entonces

$$\mathbf{e'e} - \mathbf{e}'_{\mathfrak{J}}\, \mathbf{M}_{\mathfrak{J}}^{-1}\, \mathbf{e}_{\mathfrak{J}} \quad \text{y} \quad \mathbf{e}_{\mathfrak{J}} \quad \text{son independientes.}$$

<u>Demostración</u>.- Sin perdida de generalidad se puede considerar

$$\mathbf{e} = \begin{bmatrix} \mathbf{e}_{\mathfrak{J}} \\ \mathbf{e}_{(\mathfrak{J})} \end{bmatrix} = (\mathbf{I}_n - \mathbf{V})\, \underline{\varepsilon} = \begin{bmatrix} \mathbf{M}_{\mathfrak{J}} & \mathbf{M}_0 \\ \mathbf{M}'_0 & \mathbf{M}_{(\mathfrak{J})} \end{bmatrix} \begin{bmatrix} \underline{\varepsilon}_{\mathfrak{J}} \\ \underline{\varepsilon}_{(\mathfrak{J})} \end{bmatrix}$$

Por lo tanto,

$$\mathbf{e}_{\mathfrak{J}} = \begin{bmatrix} \mathbf{M}_{\mathfrak{J}} & \mathbf{M}_0 \end{bmatrix} \underline{\varepsilon}$$

y

$$\mathbf{e'e} - \mathbf{e}'_{\mathfrak{J}}\, \mathbf{M}_{\mathfrak{J}}^{-1}\, \mathbf{e}_{\mathfrak{J}} = \underline{\varepsilon}' \begin{bmatrix} \mathbf{M} - \mathbf{M}\,\mathbf{M}^{*}\,\mathbf{M} \end{bmatrix} \underline{\varepsilon}$$

donde

$$\mathbf{M}^{*} = \begin{bmatrix} \mathbf{M}_{\mathfrak{J}}^{-1} & \Theta \\ \Theta & \Theta \end{bmatrix}$$

Siendo, por tanto, independientes, ya que

$$\begin{bmatrix} \mathbf{M}_{\mathfrak{J}} & \mathbf{M}_0 \end{bmatrix} \begin{bmatrix} \mathbf{M} - \mathbf{M}\,\mathbf{M}^{*}\,\mathbf{M} \end{bmatrix} = \Theta$$

■

TEOREMA 2.6.4.-

1.- Si $v_{ii} \neq 1$, entonces $\hat{\sigma}^2_{(i)}$ y e_i son independientes

2.- Si $M_{\mathfrak{I}}$ es no singular entonces $\hat{\sigma}^2_{(\mathfrak{I})}$ y $e_{\mathfrak{I}}$ son independientes.

Demostración.-

Resulta evidente en base al lema 2.5.3

■

COROLARIO 2.6.5.- Si $v_{ii} \neq 1$, el i-ésimo resíduo studentizado t_i se distribuye según una *t* de Student con n-r-1 grados de libertad.

C A P I T U L O I I I

ANALISIS DE INFLUENCIA EN EL MODELO LINEAL GENERAL

CAPITULO III,- ANALISIS DE INFLUENCIA EN EL MODELO LINEAL GENERAL

El Análisis de Influencia cada día está tomando mayor importancia debido, fundamentalmente, a la necesidad de buscar la fiabilidad de los resultados en un análisis estadístico. Dicha necesidad viene provocada por el hecho de que las conclusiones de cualquier análisis se extraen a través de métodos basados en observaciones muestrales y en supuestos sobre los modelos que subyacen en la experimentación.

En este capítulo, se propone un nuevo enfoque en el estudio de la influencia basado en el sesgo que sobre el estadístico provoca cada observación considerada individual o conjuntamente. Este sesgo viene descrito por la esperanza condicionada a tal observación, por lo que se obtendrán expresiones relacionadas a las ya obtenidas en la descomposición de los estimadores.

Posteriormente, se relaciona este enfoque con el planteamiento clásico de la omisión de observaciones, a través de la curva de influencia muestral.

El enfoque propuesto se aplica en el Modelo Lineal General, trasladando los resultados obtenidos a los modelos de Regresión Lineal Múltiple y de Análisis de la Varianza de un Factor.

Finalmente se estudia la Influencia Local en el Modelo Lineal General, como método útil para complementar las técnicas de análisis de influencia obtenidas a través del sesgo condicionado.

3.1.- DESCOMPOSICION DE LOS ESTIMADORES Y ANALISIS DE INFLUENCIA

DEFINICION 3.1.1.- Sea $\underline{X}_1,\underline{X}_2,\ldots,\underline{X}_n$ una muestra aleatoria simple extraida de una población descrita por una vector aleatorio $\underline{X}$, sea $T=T(\underline{X}_1,\underline{X}_2,..,\underline{X}_n)$ un estadístico definido sobre la muestra, y sea $\underline{x}_1$, $\underline{x}_2,\ldots,$ $\underline{x}_n$ una realización de la muestra. Se denomina <u>sesgo condicionado a la i-ésima observación del estadístico T</u>, y se denota por $\mathcal{S}[\underline{x}_i;T]$ a:

$$\mathcal{S}[\underline{x}_i;T]=E\left[\ T \mid \underline{X}_i=\underline{x}_i\right] - E\left[\ T\ \right]$$

Sobre esta definición se pueden realizar las siguientes consideraciones:

- El sesgo condicionado a una observación muestral coincide con el término de primer orden en la descomposición del estadístico propuesta por Efron y Stein (1981).

$$A_i[\underline{x}_i;T] = \mathcal{S}[\underline{x}_i;T]$$

- El sesgo condicionado a una observación muestral del estadístico T puede ser considerado como una medida de la de la influencia de dicha observación sobre T, puesto que si se considera la descomposición:

$$T(\underline{x}_1,\underline{x}_2,..,\underline{x}_n) = E[T] + \sum_{i=1}^{n} A_i[\underline{x}_i;T] + \sum_{1\leq i<j\leq n}\sum B_{ij}[\underline{x}_i,\underline{x}_j;T] + \ldots.$$

la presencia de la i-ésima observación, considerada individualmente, provoca una desviación del valor del estadístico a su esperanza, obtenida bajo el modelo teórico considerado, que puede ser medido por $A_i[\underline{x}_i;T]$. Cuanto mayor sea esta desviación, mayor será el impacto que sobre T ejerce dicha observación.

Por extensión, se puede definir el sesgo condicionado a un conjunto de observaciones, el cual podrá ser considerado como una medida de la influencia conjunta que ejerce las mismas sobre el estadístico.

DEFINICION 3.1.2.- En las condiciones de la Definicion 3.1.1, se denomina <u>sesgo condicionado al conjunto de observaciones</u> $\{ x_{i_1}, x_{i_2}, \dots , x_{i_m} \}$ <u>del estadístico T</u> a:

$$E\left[T \mid \underline{X}_{i_1} = \underline{x}_{i_1}, \underline{X}_{i_2} = \underline{x}_{i_2}, \dots, \underline{X}_{i_m} = \underline{x}_{i_m} \right] - E\left[T \right]$$

Y se denotará por $\mathscr{S}\left[\underline{x}_{i_1}, \underline{x}_{i_2}, \dots, \underline{x}_{i_m} ; T \right]$

En base a estas definiciones, para el Modelo Lineal General, como consecuencia de los resultados 2.2.2 y 2.2.5, se tiene

<u>Sesgo condicinado a la i-ésima observacion del BLUE de una función linealmente estimable $\Lambda\underline{\beta}$</u>

$$\mathscr{S}\left[y_i ; \Lambda\hat{\underline{\beta}} \right] = \Lambda S^- \underline{x}_i' \left[y_i - \underline{x}_i \underline{\beta} \right]$$

<u>Sesgo condicionado a la i-ésima observación del estimador insesgado de σ^2</u>

$$\mathcal{S}\left[\, y_i \,;\, \hat{\sigma}^2 \,\right] = \frac{1-v_{ii}}{n-r} \left[[y_i - \underline{x}_i \underline{\beta}\,]^2 - \sigma^2 \right]$$

El sesgo condicionado a un conjunto de observaciones para los estimadores del Modelo Lineal General se obtienen en los dos teoremas siguientes. Se considera la colección de subíndices

$$\mathfrak{J} = \left\{ i_1 \,,\, i_2 \,,\, \dots \,,\, i_m \right\} \subset \left\{ 1,\, 2,\, \dots,\, n \right\}$$

Con la notación utilizada en el apartado 2.5 para la partición de vectores y matrices, se puede expresar el sesgo condicionado al conjunto de observaciones subíndicadas por los elementos de $\mathfrak{J}$ de la forma:

$$\mathcal{S}\left[\, y_{\mathfrak{J}};\, T \,\right] = E\left[\, T \mid Y_{\mathfrak{J}} = y_{\mathfrak{J}} \right] - E\left[\, T \,\right]$$

Reordenando convenientemente las matrices y vectores, se puede utilizar la descomposición siguiente:

$$X = \begin{bmatrix} X_{(\mathfrak{J})} \\ X_{\mathfrak{J}} \end{bmatrix} \;;\; Y = \begin{bmatrix} Y_{(\mathfrak{J})} \\ Y_{\mathfrak{J}} \end{bmatrix} \;;\; V = \begin{bmatrix} V_{(\mathfrak{J})} & V_0 \\ V'_0 & V_{\mathfrak{J}} \end{bmatrix} \;;\; \mathbf{M} = \mathbf{I}_n - \mathbf{V} = \begin{bmatrix} \mathbf{M}_{(\mathfrak{J})} & \mathbf{M}_0 \\ \mathbf{M}'_0 & \mathbf{M}_{\mathfrak{J}} \end{bmatrix}$$

En este apartado, y en los sucesivos, siempre se va a suponer que $\mathbf{V}_{\mathfrak{J}}$ y $\mathbf{M}_{\mathfrak{J}}$ son no singulares, y que $n > r + m$.

TEOREMA 3.1.3.- Sea $\Lambda\underline{\beta}$ una función linealmente estimable. El sesgo condicionado al conjunto de observaciones muestrales subindicadas por $\mathfrak{J}$ de $\Lambda\hat{\underline{\beta}}$, es:

$$\mathscr{S}\left[\, y_{\mathfrak{J}} \,;\, \Lambda\, \hat{\underline{\beta}} \,\right] = \Lambda\, S^{-}\, X'_{\mathfrak{J}}\left[y_{\mathfrak{J}} - X_{\mathfrak{J}}\, \underline{\beta} \,\right]$$

Demostración.-

$$E\left[\Lambda\, \hat{\underline{\beta}} \mid Y_{\mathfrak{J}} = y_{\mathfrak{J}}\right] - E\left[\Lambda\, \hat{\underline{\beta}}\right] =$$

$$= \Lambda\, S^{-} \left[E\left[X'_{\mathfrak{J}}\, Y_{\mathfrak{J}} \mid Y_{\mathfrak{J}} = y_{\mathfrak{J}}\right] + E\left[X'_{(\mathfrak{J})}\, Y_{(\mathfrak{J})} \mid Y_{\mathfrak{J}} = y_{\mathfrak{J}}\right]\right] - \Lambda\, \underline{\beta} =$$

$$= \Lambda\, S^{-} \left[\, X'_{\mathfrak{J}}\, y_{\mathfrak{J}} + X'_{(\mathfrak{J})} X_{(\mathfrak{J})}\, \underline{\beta} \,\right] - \Lambda\, \underline{\beta} =$$

$$= \Lambda\, S^{-} \left[\, X'X\, \underline{\beta} + X'_{\mathfrak{J}}\, [y_{\mathfrak{J}} - X_{\mathfrak{J}}\, \underline{\beta}\,] \,\right] - \Lambda\, \underline{\beta} =$$

$$= \Lambda\, S^{-}\, X'_{\mathfrak{J}} \left[\, y_{\mathfrak{J}} - X_{\mathfrak{J}}\, \underline{\beta} \,\right]$$

■

LEMA 3.1.4.- Dada la descomposición en cajas de las matrices **M, V** y **X** en base a la colección de subíndices $\mathfrak{J}$, se tiene:

$$M'_0\, X_{(\mathfrak{J})} = -\, [\, I_m - V_{\mathfrak{J}}\,]\, X_{\mathfrak{J}}$$

Demostración.- Como $V\,X = X$, entonces $M\,X = \Theta$, siendo Θ la matriz nula, por lo tanto

$$M'_0\, X_{(\mathfrak{J})} = -\, M_{\mathfrak{J}}\, X_{\mathfrak{J}} = -\, [\, I_m - V_{\mathfrak{J}}\,]\, X_{\mathfrak{J}}$$

■

TEOREMA 3.1.5.- El sesgo condicionado al conjunto de observaciones muestrales subindicadas por $\mathfrak{J}$ del estimador $\hat{\sigma}^2$ es

$$\mathcal{S}\left[y_{\mathfrak{J}};\hat{\sigma}^2\right] = \frac{1}{n-r}\left[\left[y_{\mathfrak{J}} - X_{\mathfrak{J}}\,\underline{\beta}\right]'\left[I_m - V_{\mathfrak{J}}\right]\left[y_{\mathfrak{J}} - X_{\mathfrak{J}}\,\underline{\beta}\right] - \sigma^2 \text{traza}\left[I_m - V_{\mathfrak{J}}\right]\right]$$

Demostración.-

El estimador insesgado de σ^2 puede ser expresado como sigue:

$$\hat{\sigma}^2 = \frac{1}{n-r}\left[\,Y'_{(\mathfrak{J})}\,M_{(\mathfrak{J})}\,Y_{(\mathfrak{J})} + 2\,Y'_{\mathfrak{J}}\,M'_0\,Y_{(\mathfrak{J})} + Y'_{\mathfrak{J}}\,M_{\mathfrak{J}}\,Y_{\mathfrak{J}}\,\right]$$

Además,

$$* \; E\left[Y'_{\mathfrak{J}}\,M_{\mathfrak{J}}\,Y_{\mathfrak{J}} \mid Y_{\mathfrak{J}}=y_{\mathfrak{J}}\right] = y'_{\mathfrak{J}}\left[I_m - V_{\mathfrak{J}}\right] y_{\mathfrak{J}}$$

$$* \; E\left[Y'_{\mathfrak{J}}\,M'_0\,Y_{(\mathfrak{J})} \mid Y_{\mathfrak{J}}=y_{\mathfrak{J}}\right] = y'_{\mathfrak{J}}\,M'_0\,X_{(\mathfrak{J})}\underline{\beta} = -\,y'_{\mathfrak{J}}\,[I_m - V_{\mathfrak{J}}\,]\,X_{\mathfrak{J}}\,\underline{\beta}$$

$$* \; E\left[\,Y'_{(\mathfrak{J})}\,M_{(\mathfrak{J})}\,Y_{(\mathfrak{J})} \mid Y_{\mathfrak{J}}=y_{\mathfrak{J}}\right] = E\left[\,Y'_{(\mathfrak{J})}\,M_{(\mathfrak{J})}\,Y_{(\mathfrak{J})}\,\right] =$$

$$= \sum_{k\notin\mathfrak{J}} (1-v_{kk})\,E[Y_k^2] - \sum_{k\notin\mathfrak{J}}\sum_{\substack{s\notin\mathfrak{J}\\ k\neq s}} v_{ks}\,E[Y_k Y_s] = \sigma^2\left[\sum_{k\notin\mathfrak{J}}(1-v_{kk})\right] +$$

$$+ \sum_{k\notin\mathfrak{J}}(1-v_{kk})(\underline{x}_k\underline{\beta})^2 - \sum_{k\notin\mathfrak{J}}\left[\sum_{\substack{s\notin\mathfrak{J}\\ s\neq k}} v_{ks}\,(\underline{x}_s\underline{\beta})\right](\underline{x}_k\,\underline{\beta}) =$$

$$= \sigma^2\left[n-r-\text{traza}\left[I_m - V_{\mathfrak{J}}\right]\right] + \sum_{k\notin\mathfrak{J}}(1-v_{kk})(\underline{x}_k\underline{\beta})^2 -$$

$$- \sum_{k\notin\mathfrak{J}}\left[(1-v_{kk})\,(\underline{x}_k\,\underline{\beta}) - \sum_{j\in\mathfrak{J}} v_{kj}\,(\underline{x}_j\underline{\beta})\right](\underline{x}_k\,\underline{\beta}) =$$

$$= \sigma^2\left[n-r-\text{traza}\left[I_m - V_{\mathfrak{J}}\right]\right] + \sum_{j\in\mathfrak{J}}\left[\sum_{k\notin\mathfrak{J}} v_{jk}\,(\underline{x}_k\underline{\beta})\right](\underline{x}_j\,\underline{\beta}) =$$

$$= \sigma^2\left[n-r-\text{traza}\left[I_m - V_{\mathfrak{J}}\right]\right] + \sum_{j\in\mathfrak{J}}\left[(1-v_{jj})(\underline{x}_j\,\underline{\beta}) + \sum_{\substack{t\in\mathfrak{J}\\ t\neq j}} v_{jt}\,(\underline{x}_t\underline{\beta})\right](\underline{x}_j\,\underline{\beta}) =$$

$$= \sigma^2\left[n-r-\text{traza}\left[I_m - V_{\mathfrak{J}}\right]\right] + \sum_{j\in\mathfrak{J}}(1-v_{jj})(\underline{x}_j\,\underline{\beta})^2 + \sum_{\substack{t\in\mathfrak{J}\\ t\neq j}} v_{jt}\,(\underline{x}_t\underline{\beta})\,(\underline{x}_j\,\underline{\beta}) =$$

$$= \sigma^2\left[n-r-\text{traza}\left[I_m - V_{\mathfrak{J}}\right]\right] + [X_{\mathfrak{J}}\,\underline{\beta}\,]'\left[I_m - V_{\mathfrak{J}}\right]\,[X_{\mathfrak{J}}\,\underline{\beta}\,]$$

Por lo tanto, se tiene que:

$$E\left[\hat{\sigma}^2 \mid Y_{\mathfrak{J}} = y_{\mathfrak{J}}\right] = \sigma^2\left[1 - \frac{1}{n-r}\,\text{traza}\left[I_m - V_{\mathfrak{J}}\right]\right] +$$

$$+ \frac{1}{n-r}\left[y_{\mathfrak{J}} - X_{\mathfrak{J}}\,\underline{\beta}\right]'\left[I_m - V_{\mathfrak{J}}\right]\left[y_{\mathfrak{J}} - X_{\mathfrak{J}}\,\underline{\beta}\right]$$

De donde se deduce el resultado propuesto.

■

Las medidas de influencia expuestas presentan dos problemas que deben ser considerados. En primer lugar las expresiones obtenidas dependen de los parámetros $\underline{\beta}$ y σ^2 que son desconocidos . En segundo lugar, dichas expresiones tienen la misma dimensión que el parámetro, por lo que en el caso de la influencia sobre $\Lambda\hat{\underline{\beta}}$ es de dimensión mxl. Esto provoca la necesidad de determinar métodos que caractericen la influencia y que permitan ordenar las observaciones en función de la influencia.

Una solución al primer problema planteado viene dada por el siguiente teorema.

TEOREMA 3.1.6.- Sea $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ una muestra aleatoria simple extraida de una población descrita por una vector aleatorio $\underline{X}$, cuya distribución depende de un vector de parámetros $\underline{\vartheta} \in \Omega \subseteq \mathbb{R}^p$. Sea $\hat{\underline{\vartheta}} = \hat{\underline{\vartheta}}(\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n)$ un estimador insesgado, cualquiera que sea el tamaño muestral, del vector $\underline{\vartheta}$ y sea $\hat{\underline{\vartheta}}_{(\mathfrak{J})}$ el estimador insesgado obtenido de la muestra con la omisión de las observaciones $\underline{X}_k$, donde $k \in \mathfrak{J}$. Entonces

$$E\left[\hat{\underline{\vartheta}} - \hat{\underline{\vartheta}}_{(\mathfrak{J})} \mid \underline{X}_{\mathfrak{J}} = \underline{x}_{\mathfrak{J}} \right] = \mathscr{S}\left[\underline{x}_{\mathfrak{J}} \; ; \; \hat{\underline{\vartheta}} \right]$$

<u>Demostración</u>.- Evidente dado que $\hat{\underline{\vartheta}}_{(\mathfrak{J})}$ es insesgado y no depende de $\underline{X}_{\mathfrak{J}}$

■

En este punto se ha de hacer notar que el estimador del sesgo condicionado obtenido es proporcional a la curva de influencia muestral definida por D. Cook y S. Weisberg [1982].

$$CIM_{\mathfrak{J}} (\hat{\underline{\vartheta}}) = - (n-m) \left[\hat{\underline{\vartheta}}_{(\mathfrak{J})} - \hat{\underline{\vartheta}} \right]$$

No obstante, ésta siempre ha sido considerada definida sobre una realización de la muestra aleatoria, y por lo tanto como un valor o vector numérico, y no como variable o vector aleatorio, dado que su aplicación se reduce a las observaciones muestrales.

Aplicando el último teorema al Modelo Lineal General, se puede enunciar el siguiente corolario.

COROLARIO 3.1.7 .-

$$1.-\ E\left[\Lambda\, S^{-}\, X'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, e_{\mathfrak{I}} \,\middle|\, Y_{\mathfrak{I}} = y_{\mathfrak{I}}\right] = \mathcal{S}\,[\, y_{\mathfrak{I}}\, ;\, \Lambda\, \underline{\hat{\beta}}\,]$$

$$2.-\ E\left[\frac{1}{n-r-m}\, e'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, e_{\mathfrak{I}} - \frac{m}{n-r-m}\, \hat{\sigma}^{2} \,\middle|\, Y_{\mathfrak{I}} = y_{\mathfrak{I}}\right] = \mathcal{S}\,[\, y_{\mathfrak{I}}\, ;\, \hat{\sigma}^{2}\,]$$

<u>Demostración</u>.-

1.- Por el Teorema 2.5.2 , se tiene que

$$\Lambda\, \underline{\hat{\beta}}_{(\mathfrak{I})} = \Lambda\, S^{-}\, X'\, Y - \Lambda\, S^{-}\, X'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, e_{\mathfrak{I}}$$

por lo que

$$\Lambda\, \underline{\hat{\beta}} - \Lambda\, \underline{\hat{\beta}}_{(\mathfrak{I})} = \Lambda\, S^{-}\, X'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, e_{\mathfrak{I}}$$

2.- Por el teorema 2.5.4 , se tiene que

$$\hat{\sigma}^{2}_{(\mathfrak{I})} = \frac{1}{n-r-m}\left[(n-r)\, \hat{\sigma}^{2} - e'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, e_{\mathfrak{I}}\right]$$

y así

$$\hat{\sigma}^{2} - \hat{\sigma}^{2}_{(\mathfrak{I})} = \frac{1}{n-r-m}\, e'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, e_{\mathfrak{I}} - \frac{m}{n-r-m}\, \hat{\sigma}^{2}$$

■

Por lo tanto, desde un punto de vista práctico pueden considerarse las siguientes medidas de influencia obtenidas en función del sesgo condicionado:

INFLUENCIA SOBRE $\Lambda\hat{\underline{\beta}}$

1.- Medida de la influencia que sobre $\Lambda\hat{\underline{\beta}}$ ejerce el conjunto de observaciones subindicadas por $\mathfrak{I}$

$$\Lambda\, S^{-}\, X'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, e_{\mathfrak{I}} \in \mathbb{R}^{q}$$

2.- Medida de la influencia que sobre $\Lambda\hat{\underline{\beta}}$ ejerce la i–ésima observación

$$\frac{e_i}{1-v_{ii}}\, \Lambda\, S^{-}\, \underline{x}'_i \in \mathbb{R}^{q}$$

INFLUENCIA SOBRE $\hat{\sigma}^2$

3.- Medida de la influencia que sobre $\hat{\sigma}^2$ ejerce el conjunto de observaciones subindicadas por $\mathfrak{I}$

$$\frac{1}{n-r-m}\, e'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, e_{\mathfrak{I}} - \frac{m}{n-r-m}\, \hat{\sigma}^2 \in \mathbb{R}$$

4.- Medida de la influencia que sobre $\hat{\sigma}^2$ ejerce la i–ésima observación

$$\frac{1}{n-r-1}\, \frac{e_i^2}{1-v_{ii}} - \frac{1}{n-r-1}\, \hat{\sigma}^2 = \frac{1}{n-r-1}\, \hat{\sigma}^2 \left[r_i^2 - 1 \right] \in \mathbb{R}$$

3.1.1.-EL SESGO CONDICIONADO EN EL MODELO DE REGRESION LINEAL MULTIPLE

Dado que el Modelo de Regresión Lineal Múltiple (Mod.1–R) es un caso particular del Modelo Lineal General, se obtienen sobre él los siguientes resultados para el sesgo condicionado de los distintos estimadores del modelo, con el objeto de obtener medidas de influencia sobre ellos.

TEOREMA 3.1.8.- En el modelo Mod.1-R se tiene:

1.- El sesgo condicionado al conjunto de observaciones subindicadas por $\mathfrak{J}$ del estimador $\underline{\hat{\beta}}$ es

$$\mathcal{S}\,[y_{\mathfrak{J}}\ ;\ \underline{\hat{\beta}}\,] = S^{-1}\, X'_{\mathfrak{J}}\ \ [\, y_{\mathfrak{J}} - X_{\mathfrak{J}}\, \underline{\beta}\,]$$

2.- $E\left[\ S^{-1}X'_{\mathfrak{J}}\, M_{\mathfrak{J}}^{-1}\, e_{\mathfrak{J}}\ \mid Y_{\mathfrak{J}} = y_{\mathfrak{J}}\right] = \mathcal{S}\,[\, y_{\mathfrak{J}}\ ;\ \underline{\hat{\beta}}\,]$

Demostración.- En base a los resultados 3.1.3 y 3.1.7

■

COROLARIO 3.1.9.- En el modelo Mod.1-R se tiene:

1.- El sesgo condicionado a la i-ésima observación del estimador $\underline{\hat{\beta}}$ es

$$\mathcal{S}\,[y_i;\ \underline{\hat{\beta}}\,] = S^{-1}\, \underline{x}'_i\ [\, y_i - \underline{x}_i\, \underline{\beta}\,]$$

2.- $E\left[\ \frac{1}{1-v_{ii}}\ S^{-1}\underline{x}'_i\ e_i \,|\, Y_i = y_i\right] = \mathcal{S}\,[\, y_i\ ;\ \underline{\hat{\beta}}\,]$

TEOREMA 3.1.10.- En el modelo Mod.1-R se tiene:

1.- Dado un vector $\underline{x}'_0 \in \mathbb{R}^p$ de observaciones de las variables explicativas, el sesgo condicionado al conjunto de observaciones subindicadas por $\mathfrak{J}$ de la predicción $\hat{Y}_0 = \underline{x}_0\, \underline{\hat{\beta}}$ es

$$\mathcal{S}\,[y_{\mathfrak{J}}\ ;\ \hat{Y}_0] = \underline{x}_0\ S^{-1}\, X'_{\mathfrak{J}}\ \ [\, y_{\mathfrak{J}} - X_{\mathfrak{J}}\, \underline{\beta}\,]$$

2.- $E\left[\ \underline{x}_0\ S^{-1}X'_{\mathfrak{J}}\, M_{\mathfrak{J}}^{-1}\, e_{\mathfrak{J}}\ |Y_{\mathfrak{J}} = y_{\mathfrak{J}}\right] = \mathcal{S}\,[\, y_{\mathfrak{J}}\ ;\ \hat{Y}_0\,]$

Demostración.- En base a los resultados 3.1.3 y 3.1.7

■

COROLARIO 3.1.11.- En el modelo Mod.1-R se tiene:

1.- Dado un vector $\underline{x}'_0 \in \mathbb{R}^p$ de observaciones de las variables explicativas, el sesgo condicionado a la i-ésima observación de la predicción $\hat{Y}_0 = \underline{x}_0 \hat{\underline{\beta}}$ es

$$\mathcal{S}\,[y_i \,;\, \hat{Y}_0\,] = \underline{x}_0\, S^{-1}\, \underline{x}'_i \;\; [\, y_i - \underline{x}_i\, \underline{\beta}\,]$$

2.- $E\left[\frac{1}{1-v_{ii}}\, \underline{x}'_0\, S^{-1} \underline{x}'_i\, e_i \;\middle|\; Y_i = y_i\right] = \mathcal{S}\,[\, y_i \,;\, \hat{Y}_0\,]$

TEOREMA 3.1.12.- En el modelo Mod.1-R se tiene:

1.- El sesgo condicionado al conjunto de observaciones subindicadas por $\mathfrak{I}$ de $\hat{\beta}_k$ (k-ésima componente de $\hat{\underline{\beta}}$) es

$$\mathcal{S}\,[y_{\mathfrak{I}} \,;\, \hat{\beta}_k] = \underline{\delta}'_k\, S^{-1}\, X'_{\mathfrak{I}} \;\; [\, y_{\mathfrak{I}} - X_{\mathfrak{I}}\, \underline{\beta}\,]$$

siendo $\underline{\delta}_k$ el vector p-dimensional con todos sus componentes nulas salvo la k-ésima.

2.- $E\left[\, \underline{\delta}'_k\, S^{-1} X'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, e_{\mathfrak{I}} \;\middle|\; Y_{\mathfrak{I}} = y_{\mathfrak{I}}\right] = \mathcal{S}\,[\, y_{\mathfrak{I}} \,;\hat{\beta}_k\,]$

<u>Demostración</u>.- En base a los resultados 3.1.3 y 3.1.7.

■

COROLARIO 3.1.13.- En el modelo Mod.1-R se tiene

1.- El sesgo condicionado a la i-ésima observación de $\hat{\beta}_k$ es

$$\mathcal{S}\,[y_i \,;\, \hat{\beta}_k] = \underline{\delta}'_k\, S^{-1}\, \underline{x}'_i \;\; [\, y_i - \underline{x}_i\, \underline{\beta}\,]$$

siendo $\underline{\delta}_k$ el vector p-dimensional con todos sus componentes nulas salvo la k-ésima.

2.- $E\left[\frac{1}{1-v_{ii}}\, \underline{\delta}'_k\, S^{-1} \underline{x}'_i\, e_i \;\middle|\; Y_i = y_i\right] = \mathcal{S}\,[\, y_i \,;\hat{\beta}_k\,]$

TEOREMA 3.1.14.- En el Mod.1-R se tiene:

1.- El sesgo condicionado al conjunto de observaciones subindicadas por $\mathfrak{J}$ del estimador $\hat{\sigma}^2$ es

$$\mathcal{S}[y_{\mathfrak{J}};\hat{\sigma}^2]= \frac{1}{n-p}\left[[y_{\mathfrak{J}}-X_{\mathfrak{J}}\ \underline{\beta}]'[I_m-V_{\mathfrak{J}}]^{-1}[y_{\mathfrak{J}}- X_{\mathfrak{J}}\ \underline{\beta}] - \sigma^2\ \text{traza}[I_m-V_{\mathfrak{J}}]\right]$$

2.- $$E\left[\frac{1}{n-p-m}\ e'_{\mathfrak{J}}\ M_{\mathfrak{J}}^{-1}\ e_{\mathfrak{J}} - \frac{m}{n-p-m}\ \hat{\sigma}^2\ |Y_{\mathfrak{J}}= y_{\mathfrak{J}}\right] = \mathcal{S}[\ y_{\mathfrak{J}}\ ;\ \hat{\sigma}^2\ ]$$

<u>Demostración</u>.- En base a los resultados 3.1.5 y 3.1.7

■

COROLARIO 3.1.15.- En el modelo Mod.1-R se tiene:

1.- El sesgo condicionado a la i-ésima observación del estimador $\hat{\sigma}^2$ es

$$\mathcal{S}\ [y_i;\ \hat{\sigma}^2] = \frac{1-v_{ii}}{n-p}\left[\ [\ y_i - \underline{x}_i\ \underline{\beta}\ ]^2 - \sigma^2\right]$$

2.- $$E\left[\frac{1}{n-p-1}\left[\frac{1}{1-v_{ii}}\ e_i^2 - \hat{\sigma}^2\right]\middle|\ Y_i= y_i\right]= \mathcal{S}\ [y_i;\ \hat{\underline{\beta}}\ ]$$

Por lo tanto, desde un punto de vista práctico, se pueden considerar las siguientes medidas de influencia basadas en el sesgo condicionado:

<u>INFLUENCIA SOBRE</u> $\hat{\underline{\beta}}$

1.- Medida de la influencia que sobre $\hat{\underline{\beta}}$ ejerce el conjunto observaciones subindicadas por $\mathfrak{J}$

$$S^{-1}X'_{\mathfrak{J}}\ M_{\mathfrak{J}}^{-1}\ e_{\mathfrak{J}}\ \in \mathbb{R}^p$$

2.- Medida de la influencia de la i-ésima observación sobre $\hat{\underline{\beta}}$

$$\frac{1}{1-v_{ii}} \; S^{-1}\underline{x}_i' \; e_i \in \mathbb{R}^p$$

INFLUENCIA SOBRE LA PREDICCION

3.- Medida de la influencia que sobre una predicción $\underline{x}_0\hat{\underline{\beta}}$ ejerce el conjunto observaciones subindicadas por $\mathfrak{J}$

$$\underline{x}_0 \; S^{-1}X_{\mathfrak{J}}' \; M_{\mathfrak{J}}^{-1} \; e_{\mathfrak{J}} \in \mathbb{R}$$

4.- Medida de la influencia de la i-ésima observación sobre una predicción $\underline{x}_0\hat{\underline{\beta}}$

$$\frac{e_i}{1-v_{ii}} \; \underline{x}_0' \; S^{-1}\underline{x}_i' \in \mathbb{R}$$

INFLUENCIA PARCIAL

5.- Medida de la influencia que sobre la k-ésima componente de $\hat{\underline{\beta}}$ ejerce el conjunto observaciones subindicadas por $\mathfrak{J}$

$$\underline{\delta}_k' \; S^{-1}X_{\mathfrak{J}}' \; M_{\mathfrak{J}}^{-1} \; e_{\mathfrak{J}} \in \mathbb{R}$$

6.- Medida de la influencia de la i-ésima observación sobre la k-ésima componente de $\hat{\underline{\beta}}$

$$\frac{e_i}{1-v_{ii}} \; \underline{\delta}_k' \; S^{-1}\underline{x}_i' \; e_i \in \mathbb{R}$$

INFLUENCIA SOBRE $\hat{\sigma}^2$

7.- Medida de la influencia que sobre $\hat{\sigma}^2$ ejerce el conjunto observaciones subindicadas por $\mathfrak{J}$

$$\frac{1}{n-p-m}\left[e_{\mathfrak{J}}' \; M_{\mathfrak{J}}^{-1} \; e_{\mathfrak{J}} - m\,\hat{\sigma}^2 \right] \in \mathbb{R}$$

8.- Medida de la influencia de la i-ésima observación sobre $\hat{\sigma}^2$

$$\frac{1}{n-p-1}\left[\frac{1}{1-v_{ii}} e_i^2 - \hat{\sigma}^2\right] = \frac{1}{n-p-1}\hat{\sigma}^2 \left[r_i^2 - 1\right] \in \mathbb{R}$$

La primera y segunda medida, ambas correspondientes al estimador del vector de coeficientes de regresión, pertenecen al espacio real p-dimensional, por lo que plantean el problema comentado anteriormente de la imposibilidad de su ordenación. Por ello, han de ser transformadas al espacio $\mathbb{R}$.

El resto de medidas no plantean dicho problema, por lo tanto son válidas para la ordenación de las observaciones en función de la influencia, pudiéndose, además, considerar un caso como observación influencia, o un grupo de observaciones como conjuntamente influyentes, si su medida asociada es elevada en relación con el resto de observaciones.

3.1.2.- EL SESGO CONDICIONADO EN EL MODELO ANOVA DE UN FACTOR

El estudio de la influencia sobre el modelo de Análisis de la Varianza apenas ha sido tratado, en contraposición de otros modelos lineales como el de regresión. Sin embargo, con el esquema general que se ha planteado para el análisis unificado de los modelos lineales, este estudio será posible, para lo cual será suficiente trasladar los resultados obtenido para el Mod.1.

Así, si se desea estudiar en el Mod.1-A el sesgo condicionado a un conjunto de observaciones, de las cuales m_1 corresponden al primer tratamiento, m_2 al segundo tratamiento etc..., se puede reordenar el modelo de forma que las m_i observaciones correspondientes al i-ésimo tratamiento sean las m_i primeras. Así, se puede considerar la colección de subíndices:

$$\mathfrak{I} = \left\{ (1,1),\ldots,(1,m_1),(2,1),\ldots,(2,m_2), \ldots\ldots,(k,1),\ldots,(k,m_k) \right\}$$

$$m_i < n_i \ ; \ \sum m_i = m$$

La colección $\mathfrak{I}$ se puede descomponer como unión de colecciones disjuntas:

$$\mathfrak{I} = \bigsqcup_{i=1}^{k} \mathfrak{I}_i \quad , \text{ donde } \mathfrak{I}_i = \left\{ (i,1),\ldots,(i,m_i) \right\}$$

Por otra parte, $\mathbf{V}$ resulta en este caso ser una matriz diagonal por cajas, en la que la caja i-ésima es $\mathbf{V}_i = \frac{1}{n_i}\mathbf{E}_{(n_i x n_i)}$, siendo $\mathbf{E}_{(n_i x n_i)}$ la matriz de unos con las dimensiones indicadas. Si se denota por $\mathbf{V}_{i\mathfrak{I}_i}$ la submatriz de $\mathbf{V}_i$ formada por los filas y columnas subindicadas en $\mathfrak{I}_i$, entonces se tiene que $\mathbf{V}_{i\mathfrak{I}_i} = \frac{1}{n_i}\mathbf{E}_{(m_i x m_i)}$.

En base a esto, se pueden enunciar los siguientes resultados:

TEOREMA 3.1.16 .- En el Mod.1-A

1.- El sesgo condicionado al conjunto de observaciones subindicadas por $\mathfrak{J}$ del estimador $\underline{\hat{\eta}}=\left[\widehat{\mu+\alpha_1}\ \widehat{\mu+\alpha_2}\ \dots\ \widehat{\mu+\alpha_k}\right]'$ es

$$\mathscr{S}\,[y_{\mathfrak{J}}\ ;\ \underline{\hat{\eta}}] = \sum_{i=1}^{k} \frac{1}{n_i-1}\underline{\delta}_i \left[\sum_{j=1}^{m_i} \left[y_{ij} - (\mu+\alpha_i) \right] \right]$$

donde $\underline{\delta}_i$ es el vector k-dimensional con el i-ésimo elemento la unidad y el resto ceros.

2.-

$$E\left[\sum_{i=1}^{k} \frac{1}{n_i} E^{*}_{(k\times m_i)} \left[I_{m_i} - \frac{1}{n_i} E_{(m_i\times m_i)} \right]^{-1} \left[Y_{\mathfrak{J}_i} - \bar{Y}_{i.} E_{(m_i\times 1)} \right] \middle|\ Y_{\mathfrak{J}} = y_{\mathfrak{J}} \right] =$$

$$= \mathscr{S}\,[y_{\mathfrak{J}}\ ;\ \underline{\hat{\eta}}]$$

donde $E^{*}_{(k\times m_i)}$ es la matriz con la i-ésima fila de unos y el resto nulas, con las dimensiones indicadas, e $\bar{Y}_{i.}$ es el estadístico media muestral del grupo i-ésimo.

Demostración.- En base a los resultados 3.1.3 y 3.1.7

■

COROLARIO 3.1.17.- En el modelo Mod.1-A se tiene:

1.- El sesgo condicionado a la j-ésima observación del i-ésimo tratamiento del estimador $\underline{\hat{\eta}} = \left[\widehat{\mu+\alpha_1}\ \ \widehat{\mu+\alpha_2}\ \ \dots\ \ \widehat{\mu+\alpha_k}\right]'$ es

$$\mathscr{S}\,[y_{ij}\ ;\underline{\hat{\eta}}\ ] = \frac{1}{n_i}\ \underline{\delta}_i\ [y_{ij} - (\mu + \alpha_i)\ ]$$

$$2.-\ E\left[\frac{1}{n_i-1}\ \underline{\delta}_i\ [Y_{ij} - \bar{Y}_{i.}\ ]\ \middle|\ Y_{ij} = y_{ij}\right] = \mathscr{S}\,[y_{ij}\ ;\underline{\hat{\eta}}\ ]$$

TEOREMA 3.1.18.- En el modelo Mod.1-A se tiene:

1.- El sesgo condicionado al conjunto de observaciones subindicadas por $\mathfrak{J}$ del BLUE de una combinación lineal contraste de los efectos de los tratamientos, α_i , es

$$\mathscr{S}\left[y_{\mathfrak{J}};\sum_{s=1}^{k} c_s\hat{\alpha}_s\right] = \sum_{i=1}^{k} \frac{c_i}{n_i}\left[\sum_{j=1}^{m_i}[y_{ij} - (\mu + \alpha_i)]\right]$$

2.-

$$E\left[\sum_{i=1}^{k}\frac{c_i}{n_i}E_{(1xm_i)}\left[I_{m_i} - \frac{1}{n_i}E_{(m_i x m_i)}\right]^{-1}\left[Y_{\mathfrak{J}_i} - \bar{Y}_{i.}E_{(m_i x1)}\right] \,\middle|\, Y_{\mathfrak{J}}=y_{\mathfrak{J}}\right] =$$

$$= \mathscr{S}\left[y_{\mathfrak{J}};\sum_{s=1}^{k} c_s\hat{\alpha}_s\right]$$

<u>Demostración</u>.- En base a los resultados 3.1.3 y 3.1.7

∎

COROLARIO 3.1.19.- En el modelo Mod.1-A se tiene:

1.- El sesgo condicionado a la j-ésima observación del i-ésimo tratamiento del BLUE de una combinación lineal contraste de los efectos de los tratamientos, α_i , es

$$\mathscr{S}\left[y_{ij};\sum_{s=1}^{k} c_s\hat{\alpha}_s\right] = \frac{c_i}{n_i}\left[y_{ij} - (\mu + \alpha_i)\right]$$

2.- $$E\left[\frac{c_i}{n_i - 1}\left[Y_{ij} - \bar{Y}_{i.}\right] \,\middle|\, Y_{ij}=y_{ij}\right] = \mathscr{S}\left[y_{ij};\sum_{s=1}^{k} c_s\hat{\alpha}_s\right]$$

TEOREMA 3.1.20.- En el modelo Mod.1-A se tiene:

1.- El sesgo condicionado al conjunto de observaciones subindicadas por $\mathfrak{J}$ del estimador $\hat{\sigma}^2$ es

$$\mathscr{S}[y_{\mathfrak{J}};\hat{\sigma}^2]= \frac{1}{n-k}\left\{ \sum_{i=1}^{k}\left[\frac{2n_i-1}{n_i}\left[\sum_{j=1}^{m_i}\left[y_{ij}-(\mu+\alpha_i)\right]^2\right] - \left[\sum_{j=1}^{m_i}[y_{ij}-(\mu+\alpha_i)]\right]^2\right]-\right.$$

$$\left. - \sigma^2 \sum_{i=1}^{k} \frac{m_i\ (n_i-1)}{n_i}\right\}$$

2.-

$$E\left[\sum_{i=1}^{k}\left[Y_{\mathfrak{J}_i}-\bar{Y}_{i.}E_{(m_i x1)}\right]'\left[I_{m_i}-\frac{1}{n_i}E_{(m_i x m_i)}\right]^{-1}\left[Y_{\mathfrak{J}_i}-\bar{Y}_{i.}E_{(m_i x1)}\right]-m\hat{\sigma}^2 \,\middle|\, Y_{\mathfrak{J}}=y_{\mathfrak{J}}\right]=$$

$$= \mathscr{S}[y_{\mathfrak{J}};\hat{\sigma}^2]$$

<u>Demostración</u>.- En base a los resultados 3.1.5 y 3.1.7

■

COROLARIO 3.1.21.- En el modelo Mod.1-A se tiene

1.- El sesgo condicionado a la j-ésima observación del i-ésimo tratamiento del estimador $\hat{\sigma}^2$ es

$$\mathscr{S}\ [y_{ij}\ ;\ \hat{\sigma}^2] = \frac{n_i-1}{n_i(n-k)}\left[\ [y_{ij} - (\mu + \alpha_i)\ ]^2-\sigma^2\ \right]$$

2.- $E\left[\frac{n_i}{n_i-1}\ [Y_{ij} - \bar{Y}_{i.}]^2 - \hat{\sigma}^2 \mid Y_{ij}= y_{ij}\right] = \mathscr{S}\ [y_{ij}\ ;\hat{\sigma}^2\ ]$

Teniendo en cuenta que los resíduos en este modelo son

- Resíduos: $e_{ij} = Y_{ij} - (\hat{\mu} + \hat{\alpha}_i) = Y_{ij} - \bar{Y}_{i.}$

- Resíduos estandarizados : $r_{ij} = \left[\dfrac{n_i}{n_i - 1}\right]^{1/2} \dfrac{[Y_{ij} - \bar{Y}_{i.}]}{\hat{\sigma}}$

- Resíduos studentizados : $t_{ij} = \left[\dfrac{n_i}{n_i - 1}\right]^{1/2} \dfrac{[Y_{ij} - \bar{Y}_{i.}]}{\hat{\sigma}_{(i)}}$

entonces, desde un punto de vista práctico se pueden considerar las siguientes medidas de influencia basadas en el sesgo condicionado.

INFLUENCIA SOBRE $\hat{\underline{\eta}}$

1.- Medida de influencia que ejerce el conjunto de observaciones subindicadas por $\mathfrak{I}$ sobre el estimador de $\underline{\eta}$

$$\sum_{i=1}^{k} \frac{1}{n_i} E^{*}_{(k\times m_i)} \left[I_{m_i} - \frac{1}{n_i} E_{(m_i \times m_i)} \right]^{-1} \left[y_{\mathfrak{I}_i} - \bar{y}_{i.} E_{(m_i \times 1)} \right] \in \mathbb{R}^k$$

2.- Medida de influencia que ejerce la j-ésima observación del i-ésimo tratamiento sobre el estimador de $\underline{\eta}$

$$\frac{1}{n_i - 1} \underline{\delta}_i \, [y_{ij} - \bar{y}_{i.}] \in \mathbb{R}^k$$

INFLUENCIA SOBRE UNA COMBINACION LINEAL CONTRASTE

3.- Medida de influencia que ejerce el conjunto de observaciones subindicadas por $\mathfrak{I}$ sobre una combinación lineal contraste

$$\sum_{i=1}^{k} \frac{c_i}{n_i} E_{(1\times m_i)} \left[I_{m_i} - \frac{1}{n_i} E_{(m_i \times m_i)} \right]^{-1} \left[y_{\mathfrak{I}_i} - \bar{y}_{i.} E_{(m_i \times 1)} \right] \in \mathbb{R}$$

4.- Medida de influencia que ejerce la j-ésima observación del i-ésimo tratamiento sobre una combinación lineal contraste.

$$\frac{c_i}{n_i - 1} \left[y_{ij} - \bar{y}_{i.} \right] \in \mathbb{R}$$

INFLUENCIA SOBRE $\hat{\sigma}^2$

5- Medida de influencia que ejerce el conjunto de observaciones subindicadas por $\mathfrak{I}$ sobre $\hat{\sigma}^2$

$$\sum_{i=1}^{k} \left[y_{\mathfrak{I}_i} - \bar{y}_{i.} E_{(m_i x 1)} \right]' \left[I_{m_i} - \frac{1}{n_i} E_{(m_i x m_i)} \right]^{-1} \left[y_{\mathfrak{I}_i} - \bar{y}_{i.} E_{(m_i x 1)} \right] - m\hat{\sigma}^2 \in \mathbb{R}$$

6- Medida de influencia que ejerce la j-ésima observación del i-ésimo tratamiento ejerce sobre $\hat{\sigma}^2$

$$\frac{n_i}{n_i - 1} [Y_{ij} - \bar{Y}_{i.}]^2 - \hat{\sigma}^2 = \hat{\sigma}^2 [r_i^2 - 1] \in \mathbb{R}$$

Las medidas correspondientes a la estimación de $\underline{\eta}$ son vectores de $\mathbb{R}^k$, por lo que se le aplicarán técnicas que las transformen al espacio unidimensional.

3.2.- MEDIDAS DE INFLUENCIA SOBRE LA ESTIMACION DE UNA FUNCION LINEALMENTE ESTIMABLE.

En el apartado 3.1 se ha presentado el sesgo condicionado a las observaciones como medida de influencia, el cual se puede interpretar como el cambio en la estimación provocado por la observación o conjunto de observaciones en estudio, observándose dos problemas para su tratamiento práctico. Uno de ellos consistía en su dependencia de los parámetros desconocidos, lo cual quedó resuelto por el teorema 3.1.6, en el que se obtiene, bajo la condición de los valores muestrales observados correspondientes a la muestra aleatoria, que se verifica la insesgadez del cambio provocado en la estimación por la omisión de dichas observaciones.

El segundo problema se presentaba en el sesgo condicionado a la/as observación/es de $\Lambda\hat{\underline{\beta}}$, al pertenecer a $\mathbb{R}^q$, con lo que no es posible una caracterización de la influencia, así como una ordenación de las observaciones.

Para dar solución a este problema, se pueden usar estrategias semejantes a las comentadas en el capítulo I para el modelo de regresión. Es decir, definir normas o distancias sobre el sesgo o sobre su estimación, que transformen éstos en puntos de $\mathbb{R}$. La elección de una distancia para caracterizarlos es, por lo tanto, un apartado fundamental en el análisis de influencia.

Aunque para la definición y determinación de las distancias no es necesario asumir hipótesis sobre la distribución del modelo, para la obtención de puntos límites a partir de los cuales considerar la existencia del caracter de influencia sí lo es. Por lo tanto, en este apartado se va a suponer el modelo lineal general bajo hipótesis de normalidad, es decir, $\underline{\varepsilon} \simeq N_n[0,\sigma^2 I_n]$.

DEFINICION 3.2.1.- En el espacio vectorial $\mathbb{R}^q$ se define la (Q,c)-norma por :

$$\| \underline{x} \|_{(M,c)} = \frac{1}{c}\, \underline{x}'\, Q\, \underline{x}$$

donde **Q** es una matriz qxq, simétrica, definida positiva y c un escalar positivo.

Por lo tanto, en las condiciones del Teorema 3.1.6, la norma del sesgo condicionado al conjunto de observaciones subindicadas por $\mathfrak{I}$ de un estimador $\hat{\underline{\vartheta}}$ de $\underline{\vartheta}$ será:

$$\mathcal{D}_{\mathfrak{I}}\ [Q,c;\hat{\underline{\vartheta}}] = \frac{1}{c}\ \mathcal{S}\left[\underline{x}_{\mathfrak{I}}\ ;\ \hat{\underline{\vartheta}}\right]'\ Q\ \mathcal{S}\left[\underline{x}_{\mathfrak{I}}\ ;\ \hat{\underline{\vartheta}}\right]$$

Para el caso concreto del estimador $\Lambda\hat{\underline{\beta}}$ en el modelo lineal general será:

$$\mathcal{D}_{\mathfrak{I}}\ [Q,c;\Lambda\hat{\underline{\beta}}] = \frac{1}{c}\left[y_{\mathfrak{I}}-X_{\mathfrak{I}}\underline{\beta}\right]' X_{\mathfrak{I}}\ S^-\ \Lambda'\ Q\ \Lambda\ S^-\ X'_{\mathfrak{I}}\left[y_{\mathfrak{I}}-X_{\mathfrak{I}}\underline{\beta}\right]$$

Si se denota por $P_{\mathfrak{I}}[Q]$ a la matriz $X_{\mathfrak{I}}\ S^-\Lambda'\ Q\ \Lambda\ S^-\ X'_{\mathfrak{I}}$, se obtiene que

$$\mathcal{D}_{\mathfrak{I}}\,[\mathbf{Q},c;\Lambda\hat{\underline{\beta}}] = \frac{1}{c}\left[y_{\mathfrak{I}}-X_{\mathfrak{I}}\underline{\beta}\right]' P_{\mathfrak{I}}[\mathbf{Q}]\left[y_{\mathfrak{I}}-X_{\mathfrak{I}}\underline{\beta}\right]$$

lo que lleva a algunos autores [Cook y Weisberg (1982)] a considerarlo como el potencial relativo a **Q**, para que dichas observaciones sean conjuntamente influyentes.

Y el caso de que $\mathfrak{I} = \{ i \}$, entonces:

$$\mathcal{D}_{i}\,[\mathbf{Q},c;\Lambda\hat{\underline{\beta}}] = \frac{1}{c}\left[y_{i}-\underline{x}_{i}\underline{\beta}\right]' P_{i}[\mathbf{Q}]\left[y_{i}-\underline{x}_{i}\underline{\beta}\right]$$

donde $P_{\mathfrak{I}}[\mathbf{Q}] = \underline{x}_i\ S^{-}\ \Lambda'\ \mathbf{Q}\ \Lambda\ S^{-}\ \underline{x}_i'$, es el potencial relativo a **Q** para que la i-ésima observación sea influencia.

Se trata, por tanto, de elegir aquellas matrices **Q** y aquellos escalares c que permitan obtener medidas de la influencia que sobre la estimación de una función linealmente estimable ejerza una o un conjunto de observaciones.

3.2.1.- DISTANCIAS PARTICULARES

En este apartado se obtienen distintas medidas de influencia sobre $\Lambda\hat{\underline{\beta}}$ aplicando las (**Q**,c)-normas a las estimaciones del sesgo condicionado.

LEMA 3.2.2-

1.- Si Λ es una matriz de dimensiones qxp (q<p), de rango q, de forma que $\Lambda\underline{\beta}$ es linealmente estimable, entonces la matriz $(\Lambda\ S^{-}\Lambda')$ es definida positiva.

2.- Si Λ es una matriz de dimensiones qxp (q<p), de rango q, de forma que $\Lambda\underline{\beta}$ es linealmente estimable, entonces la matriz $(\Lambda\ S^{-}\Lambda')^{-1}$ es simétrica y definida positiva.

<u>Demostración</u>.-

1.- Por ser $\Lambda\underline{\beta}$ linealmente estimable entonces, por el resultado 2.1.4, Λ=**BX** , siendo **B** una matriz de dimensiones qxn, de rango q. Luego

$$(\Lambda\ S^{-}\ \Lambda') = \mathbf{B'\ V\ B} = [\mathbf{V\ B'}]'[\mathbf{V\ B'}]$$

por lo que para todo $\underline{u} \in \mathbb{R}^q$, se tiene

$$\underline{u}'(\Lambda\ S^{-}\ \Lambda')\ \underline{u} = [\mathbf{V\ B'}\underline{u}]'\ [\mathbf{V\ B'}\underline{u}] \geq 0$$

Y además, la igualdad a cero se alcanza para aquellos $\underline{u} \in \mathbb{R}^q$ soluciones del sistema de ecuaciones

$$\mathbf{V\ B'}\underline{u} = 0$$

cuya matriz es **V B'** de dimensiones nxq y rango q (n>q), luego es un sistema de ecuaciones consistente y determinado, por lo que su única solución es el vector nulo.

2.- En base al punto 1 dado que si una matriz es simétrica definida positiva, entonces también lo es su inversa.

■

Por lo tanto, se puede definir la (**Q**,c)-norma con

$$\mathbf{Q}_1 = (\Lambda\ S^{-}\ \Lambda')^{-1} \qquad \text{y} \qquad c_1 = q\ \hat{\sigma}^2$$

DEFINICION 3.2.3.- Sea $\Lambda\underline{\beta}$ linealmente estimable. La (Q_1,c_1)-norma del sesgo condicionado al conjunto de observaciones subindicadas por $\mathfrak{I}$ de $\Lambda\hat{\underline{\beta}}$ se denomina $\mathcal{D}_{\mathfrak{I}}$-DISTANCIA asociada al conjunto de observaciones $y_{\mathfrak{I}}$ y se denota por $\mathcal{D}_{\mathfrak{I}}[\Lambda\hat{\underline{\beta}}]$.

$$\mathcal{D}_{\mathfrak{I}}[\Lambda\hat{\underline{\beta}}] = \frac{1}{q\,\hat{\sigma}^2}\left[y_{\mathfrak{I}}-X_{\mathfrak{I}}\underline{\beta}\right]' X_{\mathfrak{I}}\, S^-\, \Lambda'\left[\Lambda\, S^-\, \Lambda'\right]^{-1}\Lambda\, S^-\, X'_{\mathfrak{I}}\left[y_{\mathfrak{I}}-X_{\mathfrak{I}}\underline{\beta}\right]$$

Así, se tiene que la $\mathcal{D}_i$-distancia asociada a y_i viene dada por:

$$\mathcal{D}_i[\Lambda\hat{\underline{\beta}}] = \frac{1}{q\,\hat{\sigma}^2}\, \underline{x}_i\, S^-\, \Lambda'\left[\Lambda\, S^-\, \Lambda'\right]^{-1}\Lambda\, S^-\, \underline{x}'_i\left[y_i-\underline{x}_i\underline{\beta}\right]^2$$

En estas distancias se puede observar uno de los problemas planteados en la definición del sesgo condicionado: su dependencia del vector paramétrico $\underline{\beta}$ desconocido. Esta misma distancia se puede aplicar al estimador del sesgo, definiéndose::

DEFINICION 3.2.4.- Para el estimador $\Lambda\hat{\underline{\beta}}$ de una función linealmente estimable $\Lambda\underline{\beta}$ se denomina $D_{\mathfrak{I}}$-DISTANCIA asociada al conjunto de observaciones $y_{\mathfrak{I}}$ y se denota por $D_{\mathfrak{I}}[\Lambda\underline{\beta}]$ a

$$D_{\mathfrak{I}}[\Lambda\hat{\underline{\beta}}] = \frac{1}{q\,\hat{\sigma}^2}\, e'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, X_{\mathfrak{I}}\, S^-\, \Lambda'\left[\Lambda\, S^-\, \Lambda'\right]^{-1}\Lambda\, S^-\, X'_{\mathfrak{I}}\, M_{\mathfrak{I}}^{-1}\, e_{\mathfrak{I}}$$

Por lo tanto, para el caso del estudio de la influencia de una observación se tiene que la D_i-distancia asociada a la i-ésima observación viene dada por

$$D_i[\Lambda\hat{\underline{\beta}}] = \frac{1}{q(1-v_{ii})}\ \underline{x}_i\ S^-\ \Lambda'\left[\Lambda\ S^-\ \Lambda'\right]^{-1}\Lambda\ S^-\ \underline{x}_i'\ r_i^2$$

El elipsoide de confianza para $\Lambda\underline{\beta}$, con un nivel de significación α, viene dado, según el resultado 2.1.10, por el conjunto de vectores $\mathbf{d}^*$ tales que:

$$\frac{[\Lambda\ \hat{\underline{\beta}} - \mathbf{d}^*\ ]'\ [\Lambda\ S^-\ \Lambda']^{-1}\ [\Lambda\ \hat{\underline{\beta}} - \mathbf{d}^*\ ]}{q\ \hat{\sigma}^2} \leq F_{1-\alpha,q,n-r}$$

Dada la semejanza con la distancia propuesta

$$D_{\mathfrak{J}}[\Lambda\underline{\beta}] = \frac{[\Lambda\ \hat{\underline{\beta}} - \Lambda\ \hat{\underline{\beta}}_{(\mathfrak{J})}\ ]'\ [\Lambda\ S^-\ \Lambda']^{-1}\ [\Lambda\ \hat{\underline{\beta}} - \Lambda\ \hat{\underline{\beta}}_{(\mathfrak{J})}\ ]}{q\ \hat{\sigma}^2}$$

se puede utilizar como criterio para determinar si el conjunto de observaciones subindicadas por $\mathfrak{J}$ son conjuntamente influeyentes o no el siguiente:

- Si el valor de la $D_{\mathfrak{J}}$-distancia es mayor que el punto crítico de la distribución, $F_{1-\alpha,q,n-r}$, entonces el cambio producido en la estimación de $\Lambda\underline{\beta}$ se da fuera del elipsoide de confianza a un nivel de significación $(1-\alpha)100\%$, siendo, por tanto, un cambio suficiente como para considerar influencia conjunta de las observaciones.

Planteamiento análogo se puede hacer para estudiar la influencia individual de una observación.

Trasladando esta medida de influencia para los modelos de Regresión Lineal Múltiple y Análisis de la Varianza de un factor, se obtienen los siguientes resultados:

TEOREMA 3.2.5 .- En el modelo Mod.1-R la $D_{\mathfrak{J}}$-distancia sobre $\hat{\underline{\beta}}$ coincide con la distancia de Cook, verificándose:

1.- $D_{\mathfrak{J}}[\hat{\underline{\beta}}] = \dfrac{1}{p\,\hat{\sigma}^2}\, \mathbf{e}'_{\mathfrak{J}}\, \mathbf{M}_{\mathfrak{J}}^{-1}\, \mathbf{V}_{\mathfrak{J}}\, \mathbf{M}_{\mathfrak{J}}^{-1}\, \mathbf{e}_{\mathfrak{J}}$

2.- $D_i[\hat{\underline{\beta}}] = \dfrac{1}{p}\, \dfrac{v_{ii}}{(1-v_{ii})}\, r_i^2$

TEOREMA 3.2.6.- En el modelo Mod.1-A se verifica:

1.- $D_{\mathfrak{J}}(\hat{\underline{\eta}}) =$

$$\frac{\displaystyle\sum_{i=1}^{k} \frac{1}{n_i}\left[y_{\mathfrak{J}_i} - \bar{y}_{i.}\mathbf{E}_{m_i x1}\right]'\left[\mathbf{I}_{m_i} - \frac{1}{n_i}\mathbf{E}_{m_i x m_i}\right]^{-1}\mathbf{E}_{m_i x m_i}\left[\mathbf{I}_{m_i} - \frac{1}{n_i}\mathbf{E}_{m_i x m_i}\right]^{-1}\left[y_{\mathfrak{J}_i} - \bar{y}_{i.}\mathbf{E}_{m_i x1}\right]}{k\,\hat{\sigma}^2}$$

2.- $D_i(\hat{\underline{\eta}}) = \dfrac{1}{k(n_i-1)}\, r_{ij}^2$

donde r_{ij} es el resídual estandarizado correspondiente a la j-ésima observación del i-ésimo tratamiento.

Siguiendo el esquema anterior para la norma definida por la matriz $Q_1=[\Lambda S^{-}\Lambda']^{-1}$ y por el escalar positivo $c_2=\hat{\sigma}^2_{(\mathfrak{J})}$, es decir, la estimación mínimo-cuadrática de la varianza en la que se omiten las observaciones subindicadas por $\mathfrak{J}$ se tiene

DEFINICION 3.2.7.- Sea $\Lambda\underline{\beta}$ linealmente estimable. La $(\mathbf{Q}_1, c_2)$-norma del sesgo condicionado al conjunto de observaciones subindicadas por $\mathfrak{I}$ de $\Lambda\hat{\underline{\beta}}$ se denomina $W_{\mathfrak{I}}$-DISTANCIA asociada al conjunto de observaciones $\mathbf{y}_{\mathfrak{I}}$ y se denota por $W_{\mathfrak{I}}[\Lambda\hat{\underline{\beta}}]$.

$$W_{\mathfrak{I}}[\Lambda\hat{\underline{\beta}}] = \frac{1}{\hat{\sigma}^2_{(\mathfrak{I})}} \left[y_{\mathfrak{I}}-X_{\mathfrak{I}}\underline{\beta}\right]' X_{\mathfrak{I}}\ S^-\ \Lambda'\left[\Lambda\ S^-\ \Lambda'\right]^{-1}\Lambda\ S^-\ X_{\mathfrak{I}}'\ \left[y_{\mathfrak{I}}-X_{\mathfrak{I}}\underline{\beta}\right]$$

Así, se tiene que la W_i-distancia asociada a y_i viene dada por:

$$W_i[\Lambda\hat{\underline{\beta}}] = \frac{1}{\hat{\sigma}^2_{(i)}}\ \underline{x}_i\ S^-\ \Lambda'\left[\Lambda\ S^-\ \Lambda'\right]^{-1}\Lambda\ S^-\ \underline{x}_i'\ \left[y_i-\underline{x}_i\underline{\beta}\right]^2$$

Análogamente se definen estas distancias sobre los estimadores del sesgo. Así,

DEFINICION 3.2.8 .- Para el estimador $\Lambda\hat{\underline{\beta}}$ de una función linealmente estimable $\Lambda\underline{\beta}$ se denomina $W_{\mathfrak{I}}$-DISTANCIA asociada al conjunto de observaciones $\mathbf{y}_{\mathfrak{I}}$ y se denota por $W_{\mathfrak{I}}[\Lambda\hat{\underline{\beta}}]$ a

$$W_{\mathfrak{I}}[\Lambda\hat{\underline{\beta}}] = \frac{1}{\hat{\sigma}^2_{(\mathfrak{I})}}\ \mathbf{e}_{\mathfrak{I}}'\ \mathbf{M}_{\mathfrak{I}}^{-1}\ X_{\mathfrak{I}}\ S^-\ \Lambda'\left[\Lambda\ S^-\ \Lambda'\right]^{-1}\Lambda\ S^-\ X_{\mathfrak{I}}'\ \mathbf{M}_{\mathfrak{I}}^{-1}\ \mathbf{e}_{\mathfrak{I}}$$

Por lo tanto, para el caso del estudio de la influencia de una observación se tiene que la W_i-distancia asociada a la i-ésima observación viene dada por

$$W_i[\Lambda\hat{\underline{\beta}}] = \frac{1}{(1-v_{ii})}\ \underline{x}_i\ S^-\ \Lambda'\left[\Lambda\ S^-\ \Lambda'\right]^{-1}\Lambda\ S^-\ \underline{x}_i'\ t_i^2$$

Al depender del i-ésimo resíduo studentizado, el cual se distribuye según una t de Student con n-r-1 grados de libertad, puede utilizarse como criterio para la determinación de observaciones influencia aquéllas que

$$W_i[\Lambda\hat{\underline{\beta}}] \geq \left[\frac{1}{(1-v_{ii})} \underline{x}_i\ S^-\ \Lambda' \left[\Lambda\ S^-\ \Lambda'\right]^{-1} \Lambda\ S^-\ \underline{x}'_i \right] F_{1-\alpha,1,n-r-1}$$

Para el caso de los modelos de Regresión Lineal Múltiple y de Análisis de la Varianza de un factor se tiene:

TEOREMA 3.2.9.- En el modelo Mod.1-R la $W_{\mathfrak{I}}$-distancia sobre $\hat{\underline{\beta}}$ coincide con el cuadrado de la distancia de Welsch-Kuh, verificándose:

1.- $W_{\mathfrak{I}}[\hat{\underline{\beta}}] = \dfrac{1}{\hat{\sigma}^2_{(\mathfrak{I})}}\ e'_{\mathfrak{I}}\ M_{\mathfrak{I}}^{-1}\ V_{\mathfrak{I}}\ M_{\mathfrak{I}}^{-1}\ e_{\mathfrak{I}}$

2.- $W_i[\hat{\underline{\beta}}] = \dfrac{v_{ii}}{(1-v_{ii})}\ t_i^2$

TEOREMA 3.2 10.- En el modelo Mod.1-A se verifica:

1.- $W_{\mathfrak{I}}(\hat{\underline{\eta}}) =$

$$\frac{\displaystyle\sum_{i=1}^{k} \frac{1}{n_i}\left[y_{\mathfrak{I}_i} - \bar{y}_{i.} E_{m_i x1}\right]' \left[I_{m_i} - \frac{1}{n_i}E_{m_i x m_i}\right]^{-1} E_{m_i x m_i} \left[I_{m_i} - \frac{1}{n_i}E_{m_i x m_i}\right]^{-1} \left[y_{\mathfrak{I}_i} - \bar{y}_{i.} E_{m_i x1}\right]}{\hat{\sigma}^2_{(\mathfrak{I})}}$$

2.- $W_i(\hat{\underline{\eta}}) = \dfrac{1}{(n_i - 1)}\ t_{ij}^2$

donde t_{ij} es el resíduo studentizado correspondiente a la j-ésima observación del i-ésimo tratamiento

Considerando la norma definida por la matriz $Q_1=[\Lambda S^- \Lambda']^{-1}$ y por el escalar positivo $c_3= [r/(n-r)]\ \hat{\sigma}^2_{(\mathfrak{I})}$, y siguiendo un desarrollo semejante a los anteriores, se tiene:

DEFINICION 3.2.11.- Sea $\Lambda\underline{\beta}$ linealmente estimable. La (Q_1,c_3)-norma del sesgo condicionado al conjunto de observaciones subindicadas por $\mathfrak{I}$ de $\Lambda\hat{\underline{\beta}}$ se denomina $\mathcal{C}_{\mathfrak{I}}$-DISTANCIA asociada al conjunto de observaciones $\mathbf{y}_{\mathfrak{I}}$ y se denota por $\mathcal{C}_{\mathfrak{I}}[\Lambda\hat{\underline{\beta}}]$.

$$\mathcal{C}_{\mathfrak{I}}[\Lambda\hat{\underline{\beta}}] = \frac{1}{\frac{r}{n-r}\hat{\sigma}^2_{(\mathfrak{I})}} \left[y_{\mathfrak{I}}-X_{\mathfrak{I}}\underline{\beta}\right]' X_{\mathfrak{I}}\ S^-\ \Lambda'\left[\Lambda\ S^-\ \Lambda'\right]^{-1}\Lambda\ S^-\ X'_{\mathfrak{I}} \left[y_{\mathfrak{I}}-X_{\mathfrak{I}}\underline{\beta}\right]$$

Así, se tiene que la $\mathcal{C}_1$-distancia asociada a y_1 viene dada por:

$$\mathcal{C}_1[\Lambda\hat{\underline{\beta}}] = \frac{1}{\frac{r}{n-r}\hat{\sigma}^2_{(1)}} \quad \underline{x}_1\ S^-\ \Lambda'\left[\Lambda\ S^-\ \Lambda'\right]^{-1}\Lambda\ S^-\ \underline{x}'_1\ [y_1-\underline{x}_1\underline{\beta}\ ]^2$$

Y de forma análoga , se definen estas distancias sobre los estimadores del sesgo. Así,

DEFINICION 3.2.12.- Para el estimador $\Lambda\hat{\underline{\beta}}$ de una función linealmente estimable $\Lambda\underline{\beta}$ se denomina $C_{\mathfrak{I}}$-DISTANCIA asociada al conjunto de observaciones $\mathbf{y}_{\mathfrak{I}}$ y se denota por $C_{\mathfrak{I}}[\Lambda\hat{\underline{\beta}}]$ a

$$C_{\mathfrak{I}}[\Lambda\hat{\underline{\beta}}] = \frac{1}{\frac{r}{n-r}\hat{\sigma}^2_{(\mathfrak{I})}} \mathbf{e}'_{\mathfrak{I}}\ \mathbf{M}^{-1}_{\mathfrak{I}}\ X_{\mathfrak{I}}\ S^-\ \Lambda'\left[\Lambda\ S^-\ \Lambda'\right]^{-1}\Lambda\ S^-\ X'_{\mathfrak{I}}\ \mathbf{M}^{-1}_{\mathfrak{I}}\ \mathbf{e}_{\mathfrak{I}}$$

Por lo tanto, para el caso del estudio de la influencia de una observación se tiene que la C_i-distancia asociada a la i-ésima observación viene dada por

$$C_i[\Lambda\underline{\hat{\beta}}] = \frac{n-r}{r}\,\frac{1}{(1-v_{ii})}\;\underline{x}_i\; S^{-}\,\Lambda'\left[\Lambda\, S^{-}\,\Lambda'\right]^{-1}\Lambda\, S^{-}\,\underline{x}_i'\;\; t_i^2$$

Al depender del i-ésimo resíduo studentizado, puede utilizarse como criterio para la determinación de observaciones influencia aquéllas que

$$C_i[\Lambda\underline{\hat{\beta}}] \geq \left[\;\frac{n-r}{r}\,\frac{1}{(1-v_{ii})}\;\underline{x}_i\; S^{-}\,\Lambda'\left[\Lambda\, S^{-}\,\Lambda'\right]^{-1}\Lambda\, S^{-}\,\underline{x}_i'\;\right] F_{1-\alpha,1,n-r-1}$$

Para el Modelo de Regresión Lineal Múltiple se tiene

TEOREMA 3.2.13.- En el modelo Mod.1-R la $C_{\mathfrak{I}}$-distancia sobre $\underline{\hat{\beta}}$ coincide con el cuadrado de la distancia Modificada de Cook, verificándose:

1.-$C_{\mathfrak{I}}[\underline{\hat{\beta}}] = \dfrac{1}{\frac{r}{n-r}\,\hat{\sigma}^2_{(\mathfrak{I})}}\;\underline{e}_{\mathfrak{I}}'\; M_{\mathfrak{I}}^{-1}\; V_{\mathfrak{I}}\; M_{\mathfrak{I}}^{-1}\;\underline{e}_{\mathfrak{I}}$

2.-$C_i[\underline{\hat{\beta}}] = \dfrac{n-r}{r}\,\dfrac{v_{ii}}{(1-v_{ii})}\;t_i^2$

Y para el Modelo de Análisis de la Varianza:

TEOREMA 3.2.14 .- En el modelo Mod.1-A se tiene:

1.- $C_{\mathfrak{J}}(\hat{\underline{\eta}}) =$

$$\frac{\displaystyle\sum_{i=1}^{k} \frac{1}{n_i}\left[y_{\mathfrak{J}_i} - \bar{y}_{i.} E_{m_i x1}\right]' \left[I_{m_i} - \frac{1}{n_i} E_{m_i x m_i}\right]^{-1} E_{m_i x m_i} \left[I_{m_i} - \frac{1}{n_i} E_{m_i x m_i}\right]^{-1} \left[y_{\mathfrak{J}_i} - \bar{y}_{i.} E_{m_i x1}\right]}{[\, k/(n-k)\,]\ \hat{\sigma}^2_{(\mathfrak{J})}}$$

2.- $C_i(\hat{\underline{\eta}}) = \dfrac{n-k}{n} \dfrac{1}{(n_i-1)} t^2_{ij}$

donde t_{ij} es el resíduo studentizado correspondiente a la j-ésima observación del i-ésimo tratamiento.

3.3.- INFLUENCIA LOCAL EN EL MODELO LINEAL GENERAL

Los resultados de los apartados 3.1 y 3.2 se han obtenido bajo el fundamento teórico que proporciona el sesgo condicionado por una observación o un conjunto de observaciones. Dicho fundamento conduce en la práctica a utilizar técnicas que coinciden con los obtenidos a través del esquema de perturbación del modelo provocado por la omisión de uno o de un conjunto de datos. Estas técnicas son útiles para detectar casos que deberían ser cuidadosamente inspeccionados por tener asociados errores grandes, producir considerables cambios en los estimadores o grandes diferencias en la verosimilitud.

Otro esquema de perturbación, propuesto por Cook (1986), es el basado en el modelo perturbado (Mod.2):

$$\underset{n\times 1}{Y} = \underset{n\times p}{X}\ \underset{p\times 1}{\underline{\beta}} + \underset{n\times 1}{\underline{\varepsilon}} \qquad \underline{\varepsilon} \simeq N_n[\,0\,,\sigma^2\ W\,]$$

$$rg[\,X\,] = r \le p \qquad W = \text{diag}\left[\,1,\ 1,\ \ldots,1/\omega,\ldots,1\,\right]$$

En los teoremas 2.5.3 y 2.5.5 se demostró que el comportamiento en el límite cuando, ω tiende a cero, de este modelo Mod.2 coincide prácticamente con el Mod.3, modelo en el que se omite una observación o un conjunto de ellas. Y, por otra parte, es evidente que para ω=1 el modelo perturbado Mod.2 es el modelo lineal general bajo condición de normalidad (Mod.1-N).

Cook (1987), argumentando que dos observaciones que pueden ser consideradas igualmente influyentes bajo el esquema de la omisión, pueden tener comportamientos diferentes ante pequeñas modificaciones de ω, concluye que para un estudio completo de la influencia de un caso simple es necesario investigar la conducta de los resultados del análisis para valores de ω no nulos. Este técnica de análisis de influencia se denomina *influencia local.*

Así, en este apartado se estudia, bajo el esquema de perturbación del modelo Mod.2 diversas medidas y distancias semejantes a las obtenidas bajo el fundamento teórico del sesgo. Y como esquema de comparación de resultados se utilizará la diferencia entre los estimadores obtenidos.

3.4.1.- INFLUENCIA LOCAL SOBRE LA ESTIMACION DE UNA FUNCIÓN LINEALMENTE ESTIMABLEBLE

En el apartado 2.4 se obtiene el estimador de $\Lambda\underline{\beta}$, función linealmente estimable, bajo el modelo perturbado y su relación con el el obtenido bajo el modelo postulado (Mod.1) :

$$\Lambda \hat{\underline{\beta}} - \Lambda \hat{\underline{\beta}}_{\omega} = \frac{\omega-1}{1+(\omega-1)v_{ii}} e_i \Lambda S^{-} \underline{x}_i' \in \mathbb{R}^q$$

siendo Λ una matriz de dimensiones qxp , de rango q.

Sobre esta diferencia, por pertenencer al espacio real q-dimensional, se puede aplicar algunas de las (Q,c)-normas indicadas en el apartado 3.2.

TEOREMA 3.4.1.- Sea $Q_1 = [\Lambda\ S^- \ \Lambda']^{-1}$, $c_1 = \frac{1}{q\hat{\sigma}^2}$. La $(\mathbf{M}_1,c_1)$-norma de $\Lambda\ \hat{\underline{\beta}} - \Lambda\ \hat{\underline{\beta}}_\omega$ es:

$$D_i[\ \omega,\ \Lambda\ \hat{\underline{\beta}}\ ] = D_i[\Lambda\ \hat{\underline{\beta}}\ ] \left[\frac{(\omega-1)(1-v_{ii})}{1+\ (\omega-1)v_{ii}} \right]^2$$

Demostración.-

$$D_i[\omega,\Lambda\ \hat{\underline{\beta}}] = \left[\frac{(\omega-1)(1-v_{ii})}{1+\ (\omega-1)v_{ii}} \right]^2 \frac{1}{q\hat{\sigma}^2}\ e_i^2\ \underline{x}_i\ S^-\ \Lambda'\ [\Lambda\ S^-\ \Lambda']^{-1}\ \Lambda\ S^-\ \underline{x}_i' =$$

$$= \left[\frac{(\omega-1)(1-v_{ii})}{1+\ (\omega-1)v_{ii}} \right]^2 \frac{1}{q(1-v_{ii})}\ r_i^2\ \underline{x}_i\ S^-\ \Lambda'\ [\Lambda\ S^-\ \Lambda']^{-1}\ \Lambda\ S^-\ \underline{x}_i' =$$

$$= D_i[\Lambda\ \hat{\underline{\beta}}] \left[\frac{(\omega-1)(1-v_{ii})}{1+\ (\omega-1)v_{ii}} \right]^2$$

■

Propiedades de interés de esta función de ω son:

TEOREMA 3.4.2.- 1.- $D_i[\ 0\ ,\ \Lambda\ \hat{\underline{\beta}}\ ] = D_i[\Lambda\ \hat{\underline{\beta}}\ ]$, $D_i[\ 1\ ,\ \Lambda\ \hat{\underline{\beta}}\ ] = 0$

2.- Si $D_i[\Lambda\ \hat{\underline{\beta}}\ ] = D_j[\Lambda\ \hat{\underline{\beta}}\ ]$ y $v_{ii} > v_{jj}$ entonces

$$D_i[\omega,\Lambda\ \hat{\underline{\beta}}\ ] < D_j[\omega,\Lambda\ \hat{\underline{\beta}}\ ] \qquad \forall\ \omega\ > 0$$

Demostración.- Evidente

■

La Fig.A representa las gráficas correspondientes a las funciones asociadas a dos observaciones con la misma D_i-distancia sobre $\Lambda\hat{\underline{\beta}}$, siendo $v_{ii}>v_{jj}$. Se puede observar como a pesar de tener ambas observaciones la misma influencia bajo el esquema de la omisión, la observación correspondiente a un menor elemento diagonal de la matriz de predicción posee potencialmente mayor influencia.

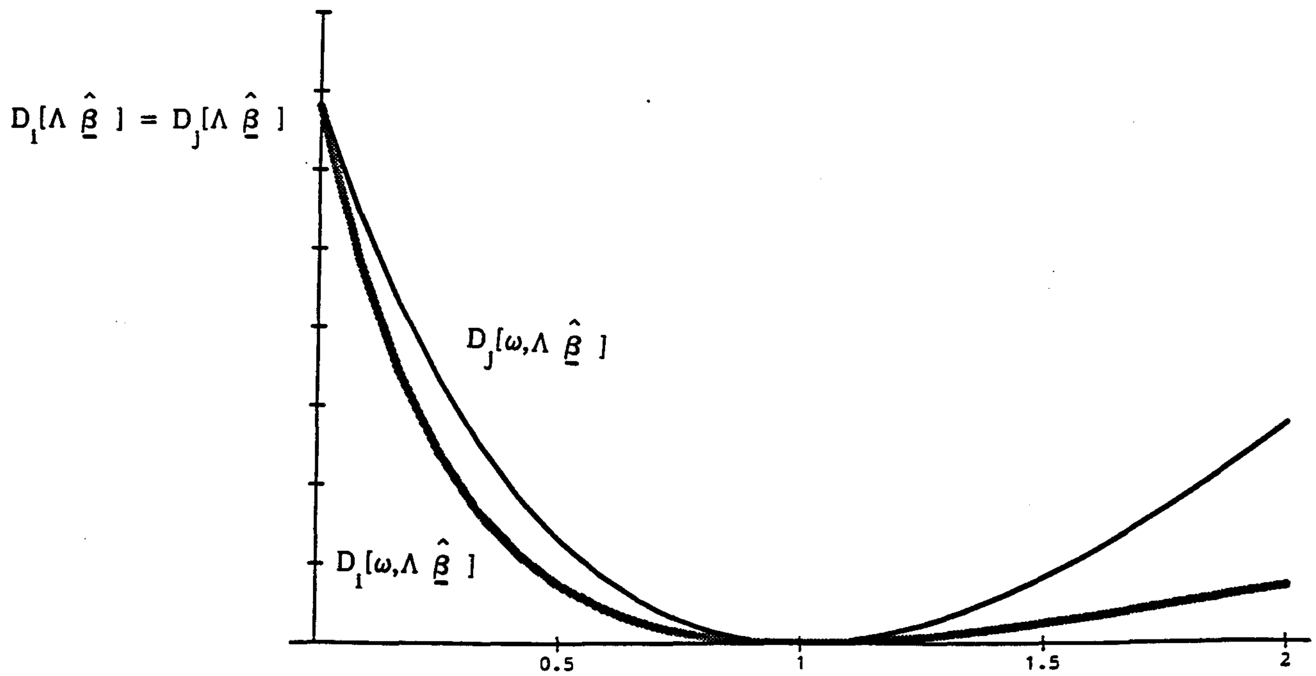

FIG. A

Se pueden realizar otros estudios comparativos conjugando resíduos, elementos diagonales de la matriz de predicción y D_i-distancia.No obstante, desde un punto de vista práctico, se debe realizar la confrontación de las gráficas asociadas a aquellas observaciones con mayor D_i-distancia.

En cuanto a su aplicación en los modelos de regresión lineal múltiple (Mod.1-R) y de análisis de la varianza (Mod.1-A) cabe resaltar que en el primero se tiene

$$D_i[\ \omega,\ \hat{\underline{\beta}}\ ] = D_i[\ \hat{\underline{\beta}}\ ] \left[\frac{(\omega-1)(1-v_{ii})}{1+\ (\omega-1)v_{ii}} \right]^2$$

y en el segundo, al ser los elementos diagonales de **V** los inversos de los tamaños muestrales tomados para el tratamiento al que corresponde la observación en cuestión, la influencia local de dos casos con la D_i-distancia va a depender de tales tamaños muestrales. En consecuencia, en un modelo balanceado bastaría estudiar D_i-distancia sobre la función linealmente estimable que se esté tratando.

3.4.2.- INFLUENCIA LOCAL EN EL ESTIMADOR MINIMO-CUADRATICO DE LA VARIANZA.

En el estudio del modelo perturbado (Mod.2) se obtuvo que

$$\hat{\sigma}^2 - \hat{\sigma}^2_\omega = -\ \frac{e_i^2}{n-r}\ \frac{\omega-1}{1+(\omega-1)v_{ii}} \qquad \omega \in (0,\infty)$$

Esta expresión puede considerarse como una función $\phi_i(.)$ de ω y su representación gráfica es de la forma que indica la figura B.

En la práctica se deberá realizar un estudio comparativo entre aquellas funciones $\phi_i(\omega)$ asociadas a posibles observaciones influencia sobre $\hat{\sigma}^2$ detectadas por otras técnicas.

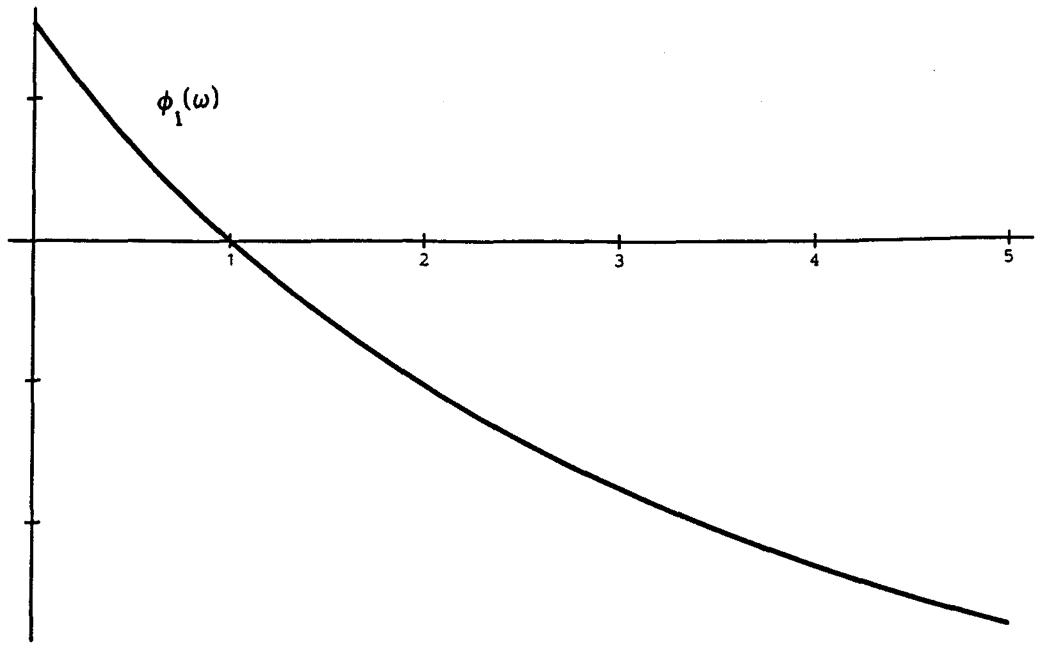

FIG. B

No obstante, un factor importante a tener en cuenta es la pendiente de la curva en el punto $\omega=1$

$$- e_i^2 /(n-r)$$

puesto que cuanto mayor sea en módulo dicha pendiente, que depende del i-ésimo resíduo y del tamaño muestral, mayor será la influencia local que sobre $\hat{\sigma}^2$ ejerce dicha observación.

B I B L I O G R A F I A

ATKINSON, A.C. (1982)

Regression Diagnostics.Transformations and Constructed Variables.
J.R. Statist. Soc., **44**, **No 1**, 1-36.

ATKINSON, A.C. (1984)

Two books on regression diagnostics.
The Annals of Statistics. **Vol.12, No.1**, 392-401.

BALASOORIYA,U.; TSE, Y.K. & LIEW,Y.S. (1987)

An empirical comparison of some statistics for identifying outliers and influential observations in linear regression models
Journal of Applied Statistics, **Vol 14, No 2**, 177-184

BARRETT, B.E. & LING,R.F. (1992)

General Classes of Influence Measures for Multivariate Regression
J.A.S.A. **Vol. 87, No.417**, 184-191

BECKMAN,R.J. y COOK,R.D. (1983)

Outliers....
Technometrics, **Vol. 15**, 119-163

BELSEY, D.A., KUH, E., WELSCH, ,R.E. (1980)

Regression Diagnostics: Identifying Influencial Data an Sources of Collinearity.
John Wiley & Sons

CARONI,C. (1987)

Residuals and influence in the multivariate linear model.
The Statitistician , **Vol.36**, 365-370.

CHATTERJEE, S., HADI, A.S. (1986)

Influencial Observations. High Leverage Points and Outliers in Linear Regression.
Statistical Science, **Vol 1, No 3**, 379-416

GRAY, J.B. (1988)

A classification of influence measures.

J. Statist. Comput. Simul. , **Vol.30,** 159–171.

GRAYBILL, F.A. (1961)

An Introduction to Linear Statistical Models. Vol I.

Mac Graw-Hill.

GRUTTOLA, V. ; WARE,J.H. & LOUIS, T.A. (1987)

Influence analysis of generalized least squares estimators.

J.A.S.A., **Vol.82, No.399** , 911–917.

HAMPEL, F.R. (1974)

The Influence Curve and its Role in Robust Estimation.

J.A.S.A., **Vol 69, No 346,** 383-393

HOSSAIN, A. & NAIK, D.N. (1989)

Detection of influential observations in multivariate regression

Journal of Applied Statistics, **Vol 16, No 1,** 25–38.

JOHNSON , N.L. KOTZ, S. (1982)

Enciclopedy of Stataistical Sciences.

John Wiley & Sons.

KSHIRSAGAR, A.M. (1983)

A Course in Linear Models

Marcel Dekker,Inc.

LEE,H.A. (1987)

Diagnostic displays for assessing leverage and influence in generalized linear models.

Australian Journal Statistics, **Vol.29 ,No.3** , 233–243.

LEE, A.H. (1988)

Assesing Partial Influence in Generalized Linear Models

Biometrics . **Vol.44,** 71-77.

MUÑOZ GARCIA, J., PASCUAL ACOSTA, A. (1986)

Analisis Singular de un Conjunto de Datos. Las Observaciones Outliers.

Revista Thales, **No 6,** 33-40.

PRINGLE, R.M., RAYNER ,A.A. (1971)

Generalized inverse matrices, with applications to statistics.

Griffin.

PUTERMAN, M.L. (1988)

Leverage and influence in autocorrelated regression models.

Applied Statistics, **Vol.37, No.1,** 76-86.

SCHALL, R. & DUNNE,, T.T. (1988)

A unified approach to outliers in the general linear model.

Sankhya, Series B, **Vol.50, pt.2,** 157-167.

SEEBER,G.U.H. (1986)

On the assesment of case influence in generalized linear models

Compstat 1986. 75-80.

THOMAS, W. (1990)

Influence on confidence regions for regressions coefficients in generalizad linear models.

J.A.S.A. , **Vol.85, No.410,** 393-397.

WILLIAMS, D.A. (1987)

Generalized Linear Model Diagnostics using the Deviance and Single Case Deletions.

Appied Statistics, **Vol.36, No.2,** 181-191.

UNIVERSIDAD DE SEVILLA

[illegible] por los abajo firmantes [illegible] Tesis Doctoral de

en el [illegible]

D. Juan M. Muñoz Pichardo

titulada "Análisis de influencia en el modelo lineal general sesgo condicionado"

acordó [illegible] APTO CUM-LAUDE

Sevilla 17 de octubre 19 92

El Vocal, El Vocal, El Vocal

El Presidente El Secretario El Doctorado.