

00384
2
1983

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO
FACULTAD DE CIENCIAS

REGRESION BAYESIANA CON
OBSERVACIONES INCOMPLETAS.

00384
1983.

T E S I S
QUE PARA OBTENER EL TÍTULO DE
DOCTOR EN CIENCIAS (MATEMÁTICAS)
P R E S E N T A
GUSTAVO JAVIER VALENCIA RAMÍREZ.

MÉXICO, D.F.

1983.

TESIS CON
FALLA DE ORIGEN



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

INDICE

	PAG.
PRÓLOGO	
CAPITULO I. METODOLOGIA CLASICA.	
INTRODUCCIÓN.	1
1.1. LOS ARGUMENTOS CLÁSICOS.	2
1.2. EL PROCESO GENERADOR DE DATOS FALTANTES.	5
1.3. PROCEDIMIENTOS CLÁSICOS PARA EL TRATAMIENTO DE DATOS INCOMPLETOS.	8
CAPITULO II. METODOLOGIA BAYESIANA.	
INTRODUCCIÓN.	20
2.1. LA ARGUMENTACIÓN BAYESIANA.	20
2.2. EL PROCESO GENERADOR DE DATOS FALTANTES.	24
2.3. SOLUCIONES BAYESIANAS A PROBLEMAS DE DATOS INCOMPLETOS.	29
CAPITULO III. UNA SOLUCION BAYESIANA AL PROBLEMA DE REGRESION MULTIPLE CON OBSERVACIONES INCOMPLETAS.	
INTRODUCCION.	35
3.1. PROCEDIMIENTO SECUENCIAL DE INCLUSIÓN DE OBSERVACIONES.	36
3.2. ANÁLISIS DE UNA OBSERVACIÓN COMPLETA.	37

	PAG.
3.3 ANÁLISIS DE UNA OBSERVACIÓN INCOMPLETA.	50
3.3.1 PÉRDIDA DEL VALOR DE LA VARIABLE DEPENDIENTE.	50
3.3.2 OBSERVACIÓN DEL VALOR DE LA VARIABLE DEPENDIENTE.	51
3.4 TRATAMIENTO DE UN CONJUNTO CON MÁS DE UNA OBSERVACIÓN.	61
3.5 PARTICULARIDADES OPERATIVAS EN EL PROCESAMIENTO DE LOS DATOS.	63
3.5.1 ANÁLISIS DE LOS SUPUESTOS HECHOS.	67
CAPITULO IV. DISCUSION Y EJEMPLOS.	72
4.1 COMPORTAMIENTO DEL COEFICIENTE DE DETERMINACIÓN.	77
4.2 EJEMPLOS.	93
EJEMPLO 1.	93
EJEMPLO 2.	105
EJEMPLO 3.	130
EJEMPLO 4.	138
APENDICE 1.	146
APENDICE 2.	148
APENDICE 3.	150

APENDICE 4.	PAG.
APENDICE 5.	151
INDICE POR TEMAS.	152
INDICE POR AUTORES.	166
BIBLIOGRAFIA.	177
	182

PROLOGO

Es claro que "la observación" forma parte del proceso cognositivo humano, que es acto ineludible del llamado Método Científico, e imprescindible para la toma de decisiones.

La Estadística, como todas las demás ciencias es, ante todo, un sistema de conocimientos y conceptos que expresan y generalizan los resultados del conocimiento humano en áreas específicas. En su contenido, existe un método determinado de estudio del mundo material y de una ligazón indestructible con la práctica; "la observación" forma parte de la acción de transposición de lo fenomenológico (realidad objetiva) a la cabeza humana y de su transformación en material empírico y pensamiento teórico.

En la Estadística, sin embargo, "la observación" va más allá de lo externo y además de ser elemento de la metodología científica, se interna en ella, como elemento del sistema de conceptos propiamente estadísticos. En efecto, el concepto de "observación" es un elemento más, del sistema teórico estadístico, y como tal, está interrelacionado con otros conceptos (población, variable, muestra, etc.) en los que encuentra su definición.

En el sistema de conceptos estadísticos (marco teórico), "po-

blación", se define como un conjunto de elementos individuales, de cualquier índole, del que se precisa información. El concepto de "observación" es entonces, el resultado de observar cada elemento seleccionado de la población. Ahora, se puede considerar una población como "el universo de todas las observaciones pertinentes que pueden hacerse en un problema de decisión dado" (Chou, 1969, p.3). Puede darse el caso de que durante la investigación de una población determinada, existan observaciones, consideradas como pertinentes, que por una u otra razón no pueden hacerse, entonces se tiene que no se cuenta con el universo de observaciones y que se producen "observaciones faltantes".

Este trabajo sobre observaciones faltantes se inscribe en el marco teórico del análisis de regresión con todas las definiciones, conceptos y variables usualmente aceptados, añadiendo a éstos la definición de observaciones incompletas, así como de los sinónimos empleados para éstos.

En análisis de regresión es usual considerar que las observaciones obtenidas para determinar la regresión de una variable dependiente en una o más variables independientes o explicativas consisten en una muestra aleatoria de n (el número de observaciones) "observaciones vectoriales", donde las componentes de estos vectores son los valores de las variables dependiente y explicativas.

Se define una observación incompleta como aquella observación vectorial en la que hay datos faltantes.

Las causas de que existan datos faltantes son principalmente: (i) que haya existido la imposibilidad de efectuar la observación correspondiente ya sea por defecto técnico o por características específicas del diseño empleado o (ii) que habiendo efectuado la observación no se tenga registro de ésta, ya sea porque no se hizo registro alguno o porque el registro se extravió.

Una matriz con algunas observaciones incompletas da origen a lo que se ha denominado un problema con datos incompletos. Por lo tanto cuando se habla de datos u observaciones incompletas se está hablando del mismo problema. Además, se han utilizado los términos, "observaciones faltantes", "observaciones perdidas" o "datos perdidos" como sinónimos de "datos faltantes", utilizando el adjetivo "perdido" no para indicar que se tenía y se extravió, sino para indicar la falta de la observación o el dato.

En todos los casos en que se trata con observaciones incompletas, se parte del hecho de que los datos faltantes no pueden ser recuperados, ya que de lo contrario lo natural sería recuperarlos desapareciendo así el problema y por otra parte se ha supuesto que los datos no faltantes son de fiar.

El enfoque del presente trabajo es bayesiano. Entre su contenido se incluye la revisión y el análisis de algunos trabajos que ilustran el tratamiento que ha sido dado por algunos autores al problema de análisis de regresión con observaciones faltantes. La clasificación y discusión del contenido de los trabajos, se hace en términos de su pertenencia a la corriente "clásica" o a la bayesiana (capítulos I y II respectivamente).

En el capítulo III, se desarrolla y explica ampliamente la solución bayesiana que se propone en este trabajo. Aunque, en un principio, se resuelve el problema de regresión con observaciones incompletas, en forma secuencial, se demuestra en este mismo capítulo, que la forma secuencial no es inherente a la solución. Es decir, en el tratamiento del problema, se puede ir incluyendo, una a una, las observaciones, sean éstas incompletas o completas (procedimiento secuencial), o ir las incorporando en bloque. Esta es una particularidad de la solución presentada que la hace muy operativa y ventajosa sobre otros muchos métodos.

El capítulo IV contiene la discusión y ejemplificación del método presentado con el fin de concretar la solución que ofrece. Comienza con una lista de las ventajas y principales características de la solución bayesiana y continúa con la presentación de varios ejemplos en los que se muestran las bondades de la solución y se pone a "prueba" su operatividad. En

algunos de estos ejemplos se comparan los resultados obtenidos mediante soluciones alternativas con los de esta solución bayesiana.

Por último, y como conclusión, puede decirse que la solución que aquí se desarrolla tiene como principales atributos: (i) ser resultado de una conceptualización bayesiana, (ii) ser de muy fácil manejo y (iii) produce resultados intuitivamente aceptables. Cuando la estadística se vincula con las diferentes disciplinas del conocimiento humano y se va más allá del desarrollo puramente estadístico, estas atribuciones resultan de carácter vital.

CAPITULO I METODOLOGIA CLASICA.

INTRODUCCIÓN.

El método estadístico se puede definir como un proceso de producción de conocimiento y por lo tanto con las siguientes etapas:

- a) Identificación del problema a tratar. Se incluye la determinación de la precisión deseada, la ubicación temporal y espacial, etc.;
- b) Elaboración del marco teórico en el que se apoya la búsqueda del conocimiento;
- c) Captación de la información;
- d) Tratamiento y procesamiento de los datos;
- e) Diseminación de la información.

El análisis de la historia del método estadístico y en específico del tratamiento de observaciones incompletas descubre dos tendencias en el desarrollo del problema: la tradicionalmente llamada clásica y la tendencia bayesiana. Las divergencias entre uno y otro tratamiento apuntan sobre todo al contenido fi

lógico en torno al problema pero éste se traduce en importantes diferencias en su tratamiento lógico.

Este capítulo se concreta a la presentación crítica de la conceptualización clásica del problema de tratamiento de datos incompletos, a través de las secciones siguientes:

- 1.1.- Los argumentos clásicos;
- 1.2.- El proceso generador de datos incompletos;
- 1.3.- Procedimientos clásicos para el tratamiento de datos incompletos.

En lo que concierne al último punto "procedimientos clásicos para el tratamiento de datos incompletos", no se pretende realizar una revisión exhaustiva sobre el tema. Se remite al lector interesado, al trabajo de Hoyle (1971) ya que contiene tanto una bibliografía extensa como la revisión de los principales métodos.

1.1. LOS ARGUMENTOS CLASICOS.

Es necesario primero, comprender claramente lo que se entiende por "estadística clásica" con el fin de encontrar las fronteras y conexiones existentes entre esta corriente y la bayesiana. Los tres volúmenes escritos por Kendall & Stuart (1967, 1968, 1969) contienen un panorama amplio y claro de la

estadística clásica por lo que se recomienda su lectura.

La escuela estadística a la que ahora se hace referencia ha recibido también los nombres de ortodoxa, estándar, frecuentista, de muestras o de Berkeley. Se origina con R.A. Fisher, J. Neyman, E.G. Pearson etc.; y considera que la información cuantitativa relevante, puede ser obtenida únicamente a partir de una muestra de comportamiento probabilístico conocida, y modelada en la verosimilitud (Barnett [1973]). En lo que se refiere a la construcción y establecimiento de procedimientos estadísticos se adopta como base el "comportamiento a la larga" (Long Term Behavior) que supone "circunstancias iguales". Bajo esta óptica, resulta que el único marco probabilístico válido es aquel que interpreta a la probabilidad como un límite de frecuencias.

Entonces, pueden definirse como "argumentos clásicos" a aquellos argumentos que se basan en la interpretación frecuentista de la probabilidad, y como consecuencia en distribuciones muestrales y en procedimientos "a la larga".

Como ejemplos del comportamiento "a la larga", pueden citarse, respecto a estimación puntual, las propiedades de insesgamiento, consistencia o eficiencia; respecto a estimación por intervalos, los conceptos de confianza y distancia esperada; y, respecto a contraste de hipótesis el nivel de significancia

y la potencia de la prueba.

El modelo general que se considera bajo el enfoque clásico, puede establecerse en los siguientes términos: considere un espacio muestral X de elementos x , dotado con un σ -campo (apropiado) de conjuntos sobre los cuales se especifica una familia de medidas de probabilidad. Estas medidas están identificadas mediante el índice θ (escalar o vector) que se conoce con el nombre de parámetro y que pertenece al "espacio paramétrico" H . Los valores de x se conocen bajo los nombres de muestra, observaciones o datos.

Es usual suponer que estas medidas de probabilidad están dominadas por una medida σ -finita de tal forma que pueden ser descritas mediante sus funciones de densidad $\{p(x|\theta)\}$ respecto a la medida empleada $\{\mu(x)\}$, de forma que:

$$P(A|\theta) = \int_A p(x|\theta) d\mu(x)$$

donde A es cualquier miembro del σ -campo y $P(A|\theta)$ es la probabilidad de A respecto a la medida identificada mediante θ .

Las medidas dominantes comúnmente empleadas son la de Lebesgue y la de contar, y en cuanto a notación, se emplea dx en vez de $d\mu(x)$. Es costumbre designar a la función $p(x|\theta)$ con el

nombre de "función de verosimilitud" cuando se considera como función del parámetro θ para x fijo.

Es importante mencionar que θ resulta simplemente un índice para las probabilidades, resultando como consecuencia que en esta discusión se ha incluido también a la estadística no-paramétrica.

1.2. EL PROCESO GENERADOR DE DATOS FALTANTES.

El origen de los datos faltantes no siempre es el mismo. Las causas por las que se producen pueden ser diferentes para cada caso que se estudia y afectan el análisis estadístico. En esta sección se presentan las ideas y resultados de Rubin (1976 c) respecto a las circunstancias en las que el proceso generador de observaciones incompletas debe ser tomado en cuenta en el análisis.

Sea $U = (U_1, \dots, U_n)$ un vector de variables aleatorias con función de densidad de probabilidades $p(\cdot | \theta)$, donde θ es el parámetro. Sea $M = (M_1, \dots, M_n)$ el vector de variables aleatorias asociado a U que indica qué variables faltan. Esto es, M_i toma el valor cero o uno; cero si la correspondiente componente de U falta, y uno, si ésta ha sido observada. La probabilidad de que M tome el valor $m = (m_1, \dots, m_n)$ dado que U toma el valor $u = (u_1, \dots, u_n)$ es $p(m|u, \phi)$, donde ϕ es el parámetro correspondiente al "proceso generador o causante de datos faltantes".

Sea X el vector de variables aleatorias (X_1, \dots, X_n) con rango extendido de tal manera que comprenda al símbolo "*" para las observaciones faltantes.

Se define a X por medio de las siguientes reglas: $X_i = U_i$ si $M_i = 1$ y $X_i = *$ si $M_i = 0$. Es importante no perder de vista que X es observable mientras que U no lo es, y que el interés se centra en inferir sobre la distribución de U .

El valor observado de M , es decir m , produce una partición de los vectores de variables aleatorias y de observaciones, de tal forma que corresponden a los valores faltantes ($m=0$) y los no faltantes o registrados ($m=1$):

$$u = (u_{(0)}, u_{(1)}), \quad x = (x_{(0)}, x_{(1)})$$

$$u = (u_{(0)}, u_{(1)}) \text{ y } x = (x_{(0)}, x_{(1)})$$

donde por definición $x_{(0)} = (*, *, \dots, *)$ y $x_{(1)} = u_{(1)}$.

El procedimiento clásico de inferencia en un problema con observaciones faltantes, debe considerar a la densidad de probabilidad obtenida mediante la distribución marginal condicional de $U_{(1)}$ dados $M=m$ θ y ϕ . Esto es:

$$\int p(u|m, \theta, \phi) du_{(1)} = \int \{p(u|\theta) p(m|u, \phi) / K(m|\theta, \phi)\} du_{(0)}$$

donde $K(m|\theta, \phi) = \int p(u|\theta) p(m|u, \phi) du$.

A partir de esta última expresión se observa que las conclusiones o inferencias que se realicen en presencia de datos faltantes dependen no solo de $p(u|\theta)$, sino de $p(m|u, \phi)$.

Rubin (1976 c) proporciona las siguientes definiciones:

1. Los datos faltantes se pierden aleatoriamente si, para cada valor ϕ , $p(m|u, \phi)$ toma el mismo valor para toda $u_{(0)}$.
2. Los datos observados se observan aleatoriamente si, para cada valor de ϕ y $u_{(0)}$, $p(m|u, \phi)$ toma el mismo valor para toda $u_{(1)}$.

Demostrando que en el caso en que los datos se pierdan y observen aleatoriamente se tiene la igualdad entre la densidad marginal de la variable aleatoria $U_{(1)}$ dada por $\int p(u|\theta) du_{(0)}$ y la densidad marginal condicional de $U_{(1)}$ dado el valor de M, θ y ϕ .

Resulta entonces, como consecuencia de los resultados anteriormente establecidos, que el procedimiento clásico de inferencia en un problema con observaciones faltantes debe considerar a la densidad de probabilidad obtenida mediante la distribución marginal condicional de $U_{(1)}$ dados M, θ y ϕ a menos que se suponga que los datos se han perdido y observado aleatoria

mente, en cuyo caso la densidad de probabilidad puede obtenerse directamente mediante la distribución marginal de $U_{(1)}$.

1.3. PROCEDIMIENTOS CLASICOS PARA EL TRATAMIENTO DE DATOS INCOMPLETOS.

Una vez tomado en cuenta el proceso generador de observaciones faltantes se prosigue al análisis estadístico del conjunto de datos incompletos.

En todos aquellos casos en que el proceso generador de observaciones faltantes sea tal, que este deba ser explícitamente modelado se produce una situación particular que debe ser tratada en forma específica y que por tanto, sólo se discutirá en forma muy breve al final de esta sección. En los casos en que el proceso generador de observaciones faltantes sea tal, que permita suponer que las observaciones faltantes se pierden aleatoriamente y los datos observados se observan aleatoriamente, entonces el procedimiento clásico a utilizarse es el discutido en la sección anterior y consiste en obtener la verosimilitud correspondiente a los datos observados. Desgraciadamente, en la gran mayoría de las situaciones prácticas no es posible determinar la densidad de probabilidad apropiada debido a la complejidad del problema, debiéndose continuar el análisis mediante alguno de los siguientes procedimientos que intentan, con mayor o menor éxito, reproducir o

aproximar al descrito en la sección anterior.

Los principales procedimientos estadísticos para el tratamiento de datos incompletos desde un enfoque clásico son:

1.3.1. Omitir, descartar o ignorar todas las observaciones incompletas:

Este procedimiento usualmente no es satisfactorio, sobre todo si en cada observación incompleta se conocen los valores de muchas variables.

En las técnicas conocidas como "diseño de experimentos", donde la única variable a observar es la llamada respuesta o variable dependiente, ocurre que en el caso de que se produzcan pérdidas, se aplique este procedimiento y a continuación se analicen los datos mediante técnicas para diseños desbalanceados. Este es un caso en el que este procedimiento es satisfactorio y correcto. Algunas referencias a este respecto pueden ser: Searle (1971), Kendall & Stuart (1968), Milliken & McDonald (1976), Milliken & Johnson (1981), Freund (1980), Overall & Spiegel (1969) y Speed et al. (1978).

Respecto a técnicas de agrupamiento (Clustering) jerárquico, está el trabajo de Baker (1978) y respecto a regresión lineal el de Haitovsky (1968).

1.3.2 Estimar por separado cada elemento de $X'X$, donde X es la matriz de datos incompletos.

Sea X la matriz de datos incompletos, de dimensión $(n \times p)$ con elemento (i, j) dado por X_{ij} . En algunas ocasiones es conveniente considerar una variable tal que siempre tome el valor uno, esto es una columna de X de unos.

Cada elemento de $X'X$ es de la forma $\sum_{i=1}^n X_{ij} X_{ik}$ y se estima por medio de $\sum_{i \in P} X_{ij} X_{ik}$ donde P es el conjunto de índices donde tanto X_{ij} como X_{ik} fueron observados.¹

Algunas referencias interesantes sobre este procedimiento son las siguientes: en el análisis de regresión considerando pérdidas entre las variables llamadas independientes, están los trabajos de Glasser (1964) y Haitovsky (1968). En el trabajo de Haitovsky se muestra, por medio de simulación en computadora, que este procedimiento no brinda buenas soluciones; cabe mencionar que en el trabajo de Haitovsky, sólo se compara este procedimiento y el discutido en el inciso anterior (1.3.1), en base al insesgamiento y la eficacia de las estimaciones. Además se contradicen algunas de las afirmaciones de Glasser.

Kendall (1975) defiende este procedimiento en el caso en que el número de observaciones con pérdidas sea "pequeño" comparado con el número total de observaciones realizadas.

[1] Una posible mejora a este procedimiento podría ser considerar, al igual que en el muestreo, factores de expansión.

1.3.3. Sustituir observaciones perdidas por valores apropiadamente estimados.

La substitución de observaciones perdidas por valores estimados es un procedimiento utilizado de muy diversas formas. Por ejemplo, Yates (1933), Healy & Westmacott (1956), Cochran & Cox (1957), Wines (1962), Agigi & Elashoff (1966), (1967) y (1967) y Brown (1975), lo utilizan substituyendo cada valor perdido de la variable j , por la media de la variable j calculada con los valores conocidos.

Rubin (1972) y (1976 b) substituye en el "Análisis de Varianza" los valores perdidos, mediante diversas combinaciones lineales de los valores observados utilizando para ello residuos. Aplica este procedimiento a problemas con sólo una fuente de error, con lo que se excluyen modelos como "Parcelas divididas". El trabajo de Rubin resulta sólo una solución parcial al problema, ya que su método no genera correctamente todos los renglones de la tabla de análisis de varianza, tiene problemas con las matrices singulares y además, es un método complicado y laborioso en su utilización. En el trabajo de John & Prescott (1975) se defiende y rediscute, en base al trabajo de Rubin (1972) la utilización de variables ficticias (Dummy) en un análisis de covarianza para estimar los valores perdidos, método que originalmente fue propuesto por Bartlett (1937) y que por ser considerado altamente laborioso (Shearen (1973)) no ha

tenido aceptación práctica.

En el trabajo de Jarret (1978) se discuten ampliamente los trabajos de Rubin (1972) y (1976 b) así como los de Wilkinson (1958 a) y (1958 b), modificando estos últimos en un intento por subsanar los inconvenientes de los resultados de Rubin.

Respecto al análisis de regresión, se tienen los trabajos de Buck (1960), Kruskal (1961), Dagenais (1975) y (1976) y Beale & Little (1975) que a grandes rasgos se presentan a continuación:

Buck (1960), estima los valores faltantes de la siguiente forma: sea X la matriz del diseño, de dimensión $(n \times p)$ y elementos x_{ij} , donde cada renglón representa los valores de las variables independientes en una observación. Sean X_1, \dots, X_p las variables cuyas observaciones son las columnas de X . Suponga que los primeros m renglones de X están completos. Sean las regresiones

$$E(x_n | X_k, I) = \sum_{k \in I} \beta_k X_k$$

donde I es un conjunto de índices.

Si x_{rj} está perdida ($n > m$) entonces, se estimará por medio de

$$\hat{E}(x_j | X_k, I_r) = \sum_{k \in I_r} \hat{\beta}_k X_k$$

donde I_x es el conjunto de índices que identifica a las variables no perdidas en la observación r -ésima, $\hat{\beta}_k$ es el estimador de β_k obtenido por mínimos cuadrados a partir de las primeras m observaciones con la regresión $E(X_j | X_k, I_x)$ y los valores de X_k empleados son aquellos observados en el renglón r -ésimo.

Kruskal (1961) utiliza las ideas expuestas por Vates (1933) donde se estiman los valores perdidos de tal forma que la suma de cuadros del residuo sea mínima.

Dagenais (1973) y (1976) trabaja suponiendo que las pérdidas han sido exclusivamente entre las variables independientes. En estos artículos se estima el valor de las variables faltantes en la observación i por medio de las variables observadas completamente a lo largo de todo el estudio, utilizando para ello una serie de regresiones "auxiliares". Por esta razón los resultados de Dagenais con una aplicación del método de Buck. La mayor aportación aparece en el tratamiento analítico del problema mediante la aplicación del método de mínimos cuadrados generalizados y la determinación de propiedades de los estimadores de los parámetros.

Beale y Little (1975) presentan un método iterativo basado en la investigación de Buck y en el que la única diferencia con este estriba en que al calcular la estimación de la matriz

de varianzas y covarianzas se divide por $(N-1)$ en vez de N .

Ligny et al. (1981), presentan una aplicación iterativa del método de Buck al análisis de factores con observaciones incompletas.

La principal ventaja al utilizar el procedimiento descrito en este inciso estriba en permitir el uso de métodos estadísticos y programas de computadora clasificados como "estandar", colocándose así al alcance de un mayor número de usuarios. Las desventajas son: riesgo de realizar un pésimo análisis si al existir una fuerte correlación entre las variables esta no es tomada en cuenta en forma apropiada y si al llevar a cabo el análisis estadístico no se toma en cuenta el que se han efectuado estimaciones de los valores perdidos y que por tanto deben realizarse "modificaciones apropiadas" por ejemplo en las varianzas de los parámetros estimados. Sobre este último punto Kendall (1975) indica que la determinación de estos ajustes o modificaciones apropiadas no es en forma alguna un trabajo sencillo o un problema resuelto. Además menciona que no debe olvidarse el hecho de que no existe sustituto para el conocimiento y que reemplazar valores perdidos por cantidades estimadas, no añade información, simplemente facilita el análisis estadístico al permitir que este resulte un análisis "estandar" con datos incompletos.

1.3.4. Suponer que los datos proceden de una distribución Normal multivariada y que además la inferencia de interés es la estimación de los parámetros involucrados, realizada mediante el método de máxima verosimilitud.

Como ejemplos de trabajos en los que se utiliza este procedimiento pueden citarse los de Wilks (1932), Rao (1952), Anderson (1957), Afifi & Elashoff (1966), Orchard & Woodbury (1972), Beale & Little (1975), Little (1976), Rubin (1976a) Dempster et al. (1977), Sanathan & Blumenthal (1978), Little (1979) y Hasselblad et al. (1980).

Sobre estos artículos puede decirse brevemente que Orchard y Woodbury establecen el llamado "Principio de la información perdida", que Beale y Little presentan en forma clara los resultados de Orchard y Woodbury, que Little aplica este método a la estimación de modelos autoregresivos en series de tiempo y a la inferencia sobre combinaciones lineales de medias en una población Normal. Dempster et al., generalizan los resultados mencionados anteriormente. Este método es iterativo y produce colateralmente la modificación "apropiada" a la varianza de los parámetros estimados. Los estimadores son aproximadamente máximo verosimiles. Por otra parte es importante señalar que Dempster et al., discuten aplicaciones de este método a problemas de datos incompletos en series de tiempo, and-

lisis de factores y análisis discriminante, que Sanathan y Blumenthal lo aplican a un modelo logístico y que Hasselblad et al., lo aplican durante un análisis Probit múltiple.

Little (1979), para el caso de regresión múltiple compara entre sí por medio de simulación en computadora los procedimientos en el inciso 1.3.3) y el procedimiento de mínimos cuadrados ordinarios aplicado en dos casos: 1) después del procedimiento (descrito en el inciso 1.3.1) consistente en eliminar las observaciones incompletas; y, 2) con los datos completos, esto es, el conjunto de datos sin pérdida alguna.

Rubin (1976 a) generaliza el coeficiente de correlación múltiple (Draper & Smith (1968)) a situaciones en las que haya observaciones perdidas sólo entre las variables independientes. El cálculo de este coeficiente se lleva a cabo utilizando las técnicas de Dempster et al. (1977). En los casos en que no haya pérdidas coincide con el coeficiente definido en forma usual. Con este coeficiente plantea Rubin la posibilidad de comparar variables independientes, y por tanto, diferentes modelos de regresión como posibles predictores de la variable dependiente, ayudando en esta forma a la selección de variables.

1.3.5. Factorizar la verosimilitud de forma tal que sea posible factorizar o separar el problema de estimación original en problemas más pequeños y sencillos de estimación.

En este procedimiento se enmarcan los resultados de Anderson (1957), Orchard & Woodbury (1973) y Rubin (1974).

El principio de la información perdida de Orchard y Woodbury mencionado en el inciso anterior, se aplica en este caso a modelos lineales y a la Normal Multivariada. La idea es, considerar a los valores perdidos como variables aleatorias y factorizar la verosimilitud en:

$$p(x|\theta) = p(u_{(1)}|\theta) p(u_{(0)}|\theta, u_{(1)})$$

según la notación empleada en la sección 1.2.

El trabajo de Rubin es una generalización del de Anderson. En ambos se descompone el problema original de estimación de parámetros en problemas más "pequeños" de estimación, mediante factorizaciones de la verosimilitud de los datos observados, en un producto de verosimilitudes cuyos parámetros son "distintos".

1.3.6. Otros procedimientos.

En problemas con datos binarios multivariados con observaciones incompletas, se tiene el trabajo de Titterton (1977), que utilizando el método de Aitchison & Aitken (1976) propone un tratamiento no paramétrico. Desde un enfoque paramétrico se tiene los resultados de Hocking & Oxpring (1974) y Koch et al. (1972).

En situaciones en las que se quiere contrastar hipótesis sobre una población Normal bivariada con coeficiente de correlación no cero, cuando la muestra tiene datos perdidos correspondientes a una o las dos variables, se puede mencionar los trabajos de Mehta & Gurland (1969), Naik (1975), Lin & Selvers (1975), Ekbohm (1976) y Dahlya & Korwar (1980) respecto a inferencias sobre las diferencias de medias y a Bhoj (1979) en cuanto a inferencias sobre la diferencia de varianzas.

En estadística no paramétrica, se tiene el trabajo de Klotz (1980) en el que se modifica la prueba de Friedman a situaciones con datos categóricos ordinales e incompletos.

Por último, Drygas (1976) trabajando modelos lineales "sin coordenadas" (Eaton (1970)) encuentra, al considerar observaciones incompletas, que los estimadores del vector de medias (X_{θ} en la notación usual) son estimadores de Gauss Markov sólo en condiciones restrictivas y difíciles de satisfacer.

En el caso en que el proceso generador de observaciones perdidas no permita suponer que tanto las pérdidas como las observaciones suceden aleatoriamente, entonces el problema estadístico resultante suele no ser catalogado como un problema de observaciones perdidas, sino que dependiendo del proceso gene-

rador de observaciones incompletas el problema resultará por ejemplo, un caso con muestras censuradas o con un diseño restringido, etc. Se recomienda por lo tanto consultar los trabajos de Rubin (1976 c) y (1978 a) así como la bibliografía apropiada. Cabe mencionar que en el trabajo de Dempster et al. (1977) se discuten aplicaciones del "principio de la información perdida" a problemas con observaciones truncadas, de los censurados o agrupados, proporcionando además, algunas citas bibliográficas.

Resumiendo, desde un enfoque clásico, el procedimiento apropiado ante un problema con datos faltantes consiste en obtener la verosimilitud correspondiente a los datos observados. Debido a que la gran mayoría de las situaciones prácticas no es posible determinar esta verosimilitud, se lleva a cabo el análisis estadístico mediante alguno de los procedimientos descritos someramente en este capítulo.

El panorama que se ha planteado en este capítulo resulta confuso como consecuencia de la no existencia en el enfoque clásico de un criterio de optimalidad que permita juzgar tanto las diferentes opciones como sus diferentes variantes.

CAPITULO II

METODOLOGIA BAYESIANA

INTRODUCCIÓN.

En este capítulo se trata de presentar un panorama (no exhaustivo) sobre el tratamiento de observaciones perdidas utilizando argumentos bayesianos. Para esto, se ha considerado apropiado definir, primero, lo que debe entenderse por argumentos bayesianos, después discutir la necesidad de modelar el proceso generador de observaciones faltantes y por último, presentar y ocasionalmente discutir algunos de los trabajos que se han publicado y que presentan soluciones bayesianas. Lo anterior se logra mediante la presentación del material de la siguiente manera:

- 2.1. La argumentación bayesiana.
- 2.2 El proceso generador de datos faltantes.
- 2.3 Soluciones bayesianas a problemas de datos incompletos.

2.1 LA ARGUMENTACION BAYESIANA.

Es necesario definir qué se entiende por "estadística bayesiana" con el fin de comprender las conexiones y similitudes, así como las diferencias, entre esta corriente y la clásica. Como referencias clave para conocer y estudiar la inferencia con

base en argumentos bayesianos se recomienda a los siguientes trabajos: Raiffa & Schlaifer (1961), Lindley (1965), De Groot (1970) y Box & Tiao (1973).

Lindley (1971, pg. 2) define como "argumento bayesiano" a cualquier argumento basado en la existencia de un σ campo sobre el espacio parametral y una medida de probabilidad sobre este. Este mismo autor define "una solución bayesiana" como una solución basada en argumentos bayesianos y que por lo tanto utiliza distribuciones de probabilidad sobre el espacio del parámetro.

A partir de lo anterior es posible definir como "estadística bayesiana" al método estadístico basado en argumentos bayesianos. El modelo general que se considera bajo el enfoque bayesiano puede establecerse de la manera siguiente:

Considere un espacio muestral X de elementos x , dotado con un σ -campo (apropiado) de conjuntos sobre las cuales se especifica una familia de medidas de probabilidad. Estas medidas están identificadas mediante el índice θ (escalar o vector) que se conoce como parámetro y que pertenece al espacio parametral H . Los valores x se conocen bajo los nombres de muestra, observaciones o datos. Al resultar θ simplemente un índice para las probabilidades resulta que este modelo incluye a la llamada estadística no-paramétrica.

Es usual suponer que estas medidas de probabilidad estan dominadas por una medida q -finita, de forma tal que pueden ser descritas mediante sus funciones de densidad $(p(x|\theta))$ respecto a la medida $(\mu(x))$, de tal forma que

$$P(A|\theta) = \int_A p(x|\theta) d\mu(x)$$

donde A es cualquier miembro del σ -campo y $P(A|\theta)$ es la probabilidad de A respecto a la medida identificada mediante θ . Las medidas dominantes comunmente utilizadas son la medida de contar y la medida de Lebesgue, y en cuanto a notación, en lugar de $d\mu(x)$, se escribe simplemente dx .

Hasta aquí las coincidencias entre los modelos clásicos y bayesiano es absoluta. Pasemos ahora a sus diferencias.

La conceptualización de la probabilidad constituye ciertamente el punto básico de disociación de las dos corrientes estadísticas. La metodología bayesiana la interpreta como una medida del conocimiento, es decir, la probabilidad es ahora una medida subjetiva y como tal se le acepta. La probabilidad subjetiva permite, en forma natural, la obtención del σ -campo y de la medida de probabilidad sobre el espacio parametral. La probabilidad subjetiva considera en forma natural el concepto de probabilidad condicional debido a que todas las probabilidades son condicionales (De Finetti, 1970, pg. 74-134). El marco filosófico del presente trabajo es bayesiano. Por esta razón, aunque en algu-

nos casos la notación empleada no lo manifieste en forma explícita, y aún en tratamientos clásicos, se considera que todas las probabilidades son condicionales.

La descripción de la medida de probabilidad puede hacerse a través de su función de densidad $p(\theta)$ con respecto a alguna medida σ -finita dominante, hecho que resulta matemáticamente conveniente. La integración respecto a esta medida dominante se denotará como $\int \dots d\theta$.

A partir de los hechos expuestos con anterioridad, es posible generar, utilizando para ello el teorema de Bayes, una familia de medidas sobre el espacio parametral $p(\theta|x)$ identificadas respecto a x .

La forma de obtener $p(\theta|x)$ mediante el teorema de Bayes (utilizando funciones de densidad) es:

$$p(\theta|x) = p(x|\theta) p(\theta) / p(x)$$

donde:

$$p(x) = \int_{\mathcal{H}} p(x|\theta) p(\theta) d(\theta)$$

y todas las distribuciones $p(\theta|x)$ están dominadas por la misma medida.

En términos bayesianos se utiliza la siguiente nomenclatura: verosimilitud de θ dado x , para la función de θ expresada median

te $p(x|\theta)$, densidad inicial o a priori (antes de observar x) de θ para $p(\theta)$ y densidad posterior (después de observar x) de θ para $p(\theta|x)$.

En algunos casos, el argumento bayesiano considera explícitamente un espacio de decisiones y una función de utilidad sobre el espacio producto, formado por el producto directo de dos espacios: el paramétrico y el de decisiones. Conociéndose a la teoría resultante como "Teoría de la Decisión".

El uso de la teoría de la decisión y como consecuencia de la estadística bayesiana y la probabilidad subjetiva, cuenta en este siglo con un gran estímulo para la divulgación de trabajos conducentes a fundamentar axiomáticamente la teoría de decisiones. Los axiomas se establecen sobre las preferencias del interesado (en la toma de decisiones), resultando que los axiomas pueden ser vistos como criterios de consistencia y coherencia sobre las preferencias del interesado.

Algunos ejemplos claves sobre trabajos encausados a proporcionar una base axiomática a la teoría de decisiones son los de Ramsey (1931) y Savage (1954); mientras que trabajos que presentan un panorama de los resultados obtenidos en este campo son los de Fishburn (1970) y (1981).

2.2 EL PROCESO GENERADOR DE DATOS FALTANTES.

En la introducción a este trabajo, se discutieron los posi-

bles orígenes de las observaciones perdidas y la manera en que estas afectan al análisis estadístico. Ahora, por lo que se refiere a las circunstancias en que el proceso generador de observaciones faltantes debe ser tomado en cuenta en el análisis estadístico, se presentan las ideas de Rubin (Rubin (1976 c)).

De igual forma que en el proceso clásico, se tiene:

Sea $U = (U_1, \dots, U_n)$ un vector de variables aleatorias con función de densidad de probabilidad $p(\cdot | \theta)$, donde θ es el parámetro. Sea $M = (M_1, \dots, M_n)$ el vector de variables aleatorias asociadas a U y que indica qué variables faltan, esto es, M_i toma el valor cero o uno, cero, si la correspondiente componente de U falta, y uno, si ésta ha sido observada. La probabilidad de que M tome el valor $m = (m_1, \dots, m_n)$ dado que U toma el valor $u = (u_1, \dots, u_n)$ es $p(m | u, \theta)$, donde θ es el parámetro correspondiente al "proceso generador o causante de datos faltantes". Sea X el vector de variables aleatorias (X_1, \dots, X_n) con rango extendido de tal manera que comprenda el símbolo "*" para las observaciones faltantes.

Se define a X mediante las siguientes reglas:

$X_i = U_i$ si $M_i = 1$ y $X_i = *$ si $M_i = 0$. Es importante no perder de vista que X es observable mientras que U no lo es, y que el interés se centra en inferir sobre la distribución de U o de

una función suya.

El valor observado de M , es decir m , produce una partición de los vectores de variables aleatorias y de observaciones, de tal forma que corresponden a los valores faltantes ($m=0$) y los no faltantes o registrados ($m=1$):

$$u = (u_{(0)}, u_{(1)}) \quad , \quad x = (x_{(0)}, x_{(1)}) \quad ,$$

$$u = (u_{(0)}, u_{(1)}) \quad , \quad x = (x_{(0)}, x_{(1)}) \quad .$$

donde por definición, $x_{(0)} = (*, *, \dots, *)$ y $x_{(1)} = u_{(1)}$.

La solución bayesiana a un problema con observaciones faltantes debe considerar a la verosimilitud obtenida de la siguiente manera:

$$\int p(u | \theta) p(m | u, \phi) d u_{(0)} \quad .$$

Si $p(\theta, \phi)$ es la distribución conjunta inicial de θ y ϕ , entonces la distribución posterior será (lea = como "es proporcional a"):

$$p(\theta, \phi | u_{(1)}, m) = p(\theta, \phi) \int p(u | \theta) p(m | u, \phi) d u_{(0)}$$

de donde se observa que la solución bayesiana depende tanto de $p(u|\theta)$, $p(m|u,\phi)$ como de $p(\theta,\phi)$

Rubin (1976 c), proporciona las siguientes definiciones:

1. los datos faltantes se pierden aleatoriamente, si para cada valor de ϕ , $p(m|u,\phi)$ toma el mismo valor para toda $u_{(0)}$;
2. el parámetro ϕ es distinto de θ , si el espacio parametral conjunto se puede factorizar en el espacio de θ y el espacio de ϕ , y cuando se especifican distribuciones a priori para θ y para ϕ , estas son independientes.

Entonces, al considerar a la distribución marginal de θ dado $u_{(1)}$ y m se tiene que:

$$p(\theta|u_{(1)}, m) = \int p(\theta, \phi) \int p(u|\theta) p(m|u, \phi) du_{(0)} d\phi$$

Si los datos faltantes se pierden aleatoriamente, entonces $p(m|u, \phi)$ es una constante respecto a $u_{(0)}$ por lo que

$$p(\theta|u_{(1)}, m) = \int p(\theta, \phi) p(m|u, \phi) \int p(u|\theta) du_{(0)} d\phi$$

y si θ es distinto de ϕ , se tiene

$$p(\theta, \phi) = p(\theta) p(\phi)$$

de donde

$$p(\theta | u_{(1)}, m) = p(\theta) \int p(u | \theta) du_{(0)}$$

En ocasiones, cuando no quiere introducirse información subjetiva, se consideran las llamadas distribuciones iniciales de referencia o "no informativas"⁽¹⁾ (Bernardo, 1979) y en relación a ellas, se hacen las siguientes observaciones:

- a) Es posible suponer que la pérdida de las observaciones sucede aleatoriamente, y si θ diferente de ϕ , entonces la distribución inicial de referencia a usarse es la que se obtendría con el procedimiento de Bernardo (1979) al considerar a la verosimilitud $\int p(u | \theta) du_{(0)}$;
- b) Si no es posible suponer que las observaciones se pierden aleatoriamente o que θ es diferente de ϕ , entonces el conocimiento que se tiene y que no permite realizar estas suposiciones, tiene que ser modelado e incluido apropiadamente en la verosimilitud. En este caso el procedimiento para obtener la inicial de referencia de θ y ϕ , deberá ser aplicado a la verosimilitud de θ y ϕ , donde se contemple apropiadamente el procedimiento de pérdida de observaciones.

(1) Es costumbre dar el nombre de distribución posterior de referencia a la distribución posterior obtenida en base a una inicial de referencia.

2.3 SOLUCIONES BAYESIANAS A PROBLEMAS CON DATOS INCOMPLETOS.

Una vez considerado el proceso generador de observaciones faltantes, debe determinarse si debe o no modelarse este proceso. Si se considera que los datos faltantes se pierden aleatoriamente y que el parámetro θ es diferente del parámetro ϕ (ver sección anterior), entonces una solución bayesiana será aquella que obtenga la distribución de θ condicional a los valores observados, que según la notación de la sección anterior es:

$$p(\theta | u_{(1)}, m) = p(\theta) \int p(u|\theta) du_{(0)}$$

A continuación se comentan los trabajos de Mehta & Swamy (1974), Dagenais (1974), Press & Scott (1974 y 1976), Stewart & Sorensen (1981) y Laird & Louis (1982).

Press & Scott, consideran un modelo de regresión lineal múltiple y discuten el problema de estimación de los parámetros involucrados por medio de máxima verosimilitud generalizada (De Groot, 1970, p. 236). Esto es, calcular la función de distribución posterior de referencia de los parámetros involucrados y las observaciones faltantes condicionado sobre las observaciones no perdidas, y calculan la moda de esta distribución de forma tal que simultáneamente estiman las observaciones faltantes y los

parámetros involucrados. Consideran que pueden haber pérdidas tanto en la variable dependiente como en las independientes o explicativas siempre y cuando al haber pérdidas en la variable dependiente se haya observado al menos, el valor de una de las variables independientes. Suponen también, que las variables independientes perdidas se distribuyen como una Normal. En el trabajo publicado en 1974, suponen que todas las variables independientes tienen la misma media y la misma precisión. En cambio, en el artículo de 1976, suponen que la matriz del diseño se puede expresar (permutando sus columnas) de la siguiente forma $X=(U, V)$ donde en U han quedado todas aquellas variables independientes que han sufrido alguna pérdida, mientras que en V han quedado todas las variables "completas". Además, considero que cada renglón de U se distribuye Normalmente con media una combinación lineal del correspondiente renglón de V y matriz de precisión constante. En ambos trabajos, Press & Scott, obtienen las expresiones analíticas de $p(u|\theta)$, pero hasta el momento de escribir estas líneas ha sido imposible llevar a cabo la integración analítica de estas.

Dagenais (1974) por su parte, considera el problema de obtener $p(\theta|u_{(1)}, m)$ en el análisis de regresión múltiple. Para ello, considera el modelo, que en notación matricial usual puede establecerse como:

$$Y = X \beta + \varepsilon$$

con todas las suposiciones distribucionales usuales. Añade los supuestos siguientes: 1) no existe ordenada al origen; 2) existen otras variables que no han sido incluidas en el modelo, que también han sido observadas y que están relacionadas con las variables independientes y 3) existe una liga entre las variables no incluidas en el modelo de regresión y algunas de las variables explicativas. Por ejemplo, (señala el autor) las variables explicativas X_1 y X_2 , se relacionan con las variables no incluidas en la matriz del diseño Z_1 y Z_2 , de la siguiente forma:

$$\begin{pmatrix} X_{1T} \\ X_{2T} \end{pmatrix} = \begin{pmatrix} Z_{1T} & 0 \\ 0 & Z_{2T} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \begin{pmatrix} v_{1T} \\ v_{2T} \end{pmatrix}$$

donde v_{iT} ($i=1, 2$), se distribuye conjuntamente en forma Normal, con media cero y matriz de covarianza

$$S = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

Supone también que las v 's son independientes de las ε 's, que pueden haber pérdidas en Y , en X y en las variables Z_1 y Z_2 ,

y que la distribución inicial es de referencia.

Dagenais, expresa la verosimilitud como un producto de verosimilitudes, una para cada situación específica de pérdida. Por ejemplo, asocia una verosimilitud para los casos en que no haya habido pérdidas, otra verosimilitud para aquellas observaciones en que se hayan perdido Y y X_2 pero no X_1, Z_1 ó Z_2 , etc. Por medio de este procedimiento, llega a plantear las funciones de densidad posterior de β , σ^2 (la varianza común del error) y S , así como las de β_1, β_2 y $\Lambda^{-1}(\sigma^{-2}) S$, e indica que a partir de estas expresiones se pueden obtener las distribuciones marginales de interés mediante la integración numérica apropiada.

Es importante comentar que el procedimiento de Dagenais presenta las restricciones siguientes: la existencia de las variables Z_1 y Z_2 , la importancia desproporcionada de las variables X_1 y X_2 respecto al resto de las variables independientes, las dificultades técnicas y el gasto en tiempo de computadora para realizar la integración numérica sobre un número de variables que puede resultar excesivo.

Por su parte, Lairds Louis (1982) discuten algunas aproximaciones a problemas con observaciones perdidas, aplicables en diferentes situaciones que los autores ejemplifican con mues-

tras censuradas por la derecha en una población exponencial y en muestras extraídas de una población multinomial. El método de aproximación que utilizan consiste básicamente en igualar la moda y la información de Fisher en la distribución usada (Normal o Gaussiana, o un miembro de la familia conjugada de distribuciones asociadas a la distribución de la que se muestra) con la moda y la información de Fisher observadas en la muestra.

El trabajo de Mehta & Sanny [1974] sobre el problema de realizar inferencias sobre la diferencia de medias en una población Normal bivariada con coeficiente de correlación posiblemente mayor que cero, cuando la muestra tiene algunos valores perdidos tanto en una como en otra variable, es resuelto al obtener la distribución posterior del parámetro de interés mediante integración numérica.

Por último, en problemas de respuesta múltiples, Steward & Sorensen [1981], realizan una aproximación a la función de densidad posterior para casos en que las pérdidas sólo ocurren entre las variables respuesta. La verosimilitud empleada es $\int p(u|\theta) du_{(a)}$ reexpresada adecuadamente.

Para el caso en que las pérdidas no sean aleatorias (según la definición de Rubin (1976 c) presentada en la sección anterior), se tiene, por ejemplo, el trabajo de Susarla & Van Rysin

[1976] donde tratan el problema de estimación de curvas de supervivencia basadas en muestras censuradas por la derecha. En este trabajo, se presenta una solución bayesiana no paramétrica y la distribución a priori usada es un proceso Dirichlet de los presentados por Ferguson (1973).

Resumiendo, el problema fundamental encontrado hasta ahora en el tratamiento del análisis de regresión con datos faltantes desde un enfoque bayesiano ha sido la dificultad de obtener la distribución posterior de los parámetros de interés en condiciones no restrictivas, siendo esta distribución la base de cualquier análisis bayesiano o toma de decisiones, resulta entonces muy importante el obtener una solución bayesiana que brinde resultados satisfactorios ante problemas con datos incompletos.

CAPITULO III

UNA SOLUCION BAYESIANA AL PROBLEMA DE REGRESION MULTIPLE CON OBSERVACIONES INCOMPLETAS.

INTRODUCCIÓN.

Este capítulo presenta una solución bayesiana al problema de regresión múltiple con observaciones perdidas mediante un procedimiento secuencial para el tratamiento de las observaciones, utilizando el concepto de familias conjugadas de distribuciones, que en el caso de un problema de regresión con "errores" Normales es la distribución Normal Gamma. En este mismo capítulo se demuestra que la forma secuencial no es inherente a la solución y que el conjunto de observaciones (sean completas o incompletas) puede ser tratado en bloque. Para facilitar la exposición, se ha partido este capítulo en las siguientes secciones:

- 3.1.- Procedimiento secuencial de inclusión de observaciones;
- 3.2.- Análisis de una observación completa;
- 3.3.- Análisis de una observación incompleta;
 - 3.3.1.- Pérdida del valor de la variable dependiente;
 - 3.3.2.- Observación del valor de la variable dependiente;
- 3.4.- Tratamiento de un conjunto con más de una observación;
- 3.5.- Particularidades operativas en el procesamiento de los datos.

3.1. PROCEDIMIENTO SECUENCIAL DE INCLUSION DE OBSERVACIONES.

La conclusión principal que se extrae del capítulo anterior, en relación a los trabajos realizados respecto al problema de regresión con datos faltantes, es la necesidad de reestructurar el problema de tal manera que sea posible integrar las variables faltantes en la densidad posterior conjunta de parámetros y valores faltantes pudiendo entonces obtener la función de densidad posterior de los parámetros de interés. En Press & Scott (1976) la complejidad del problema se manifiesta en la imposibilidad de integrar analíticamente una expresión en la que aparece una matriz en la que algunas componentes están perdidas y otras no, cuando la integración se plantea respecto a los componentes perdidas. En Dagenais (1974) se factoriza la función de verosimilitud pero no se logra simplificar la integración analítica de las variables perdidas.

Por lo anteriormente expuesto se decidió simplificar la situación a enfrentar lo más posible, manteniendo el compromiso de no perder generalidad en el problema tratado. Se encontró que la forma más simple consiste en ir incluyendo las observaciones una a una, en el análisis, utilizando para ello la idea de un modelo secuencial presentada en Harrison & Stevens (1976). De esta manera resulta posible obtener analíticamente la distribución posterior de los parámetros de interés, resolviendo los problemas antes mencionados.

Por otra parte, se ha considerado importante el que la distribución posterior resultante sea un miembro de la familia de distribuciones Normal Gamma que es la familia conjugada utilizada en problemas de regresión. Para una discusión amplia sobre las familias

conjugadas de distribución en general y sobre la distribución Normal Gamma en particular se recomienda consultar el libro de Raiffa & Schlaifer (1961, pp. 43-78).

En lo que resta de este trabajo se subrayan los vectores para mayor claridad y se considera que los datos faltantes se pierden aleatoriamente y que el parámetro ϕ correspondiente al proceso generador de observaciones faltantes y $\theta = (\beta', w)$ son distintos, donde β y w son parámetros que definen la media y la precisión en un problema de regresión. Estas suposiciones permiten no considerar explícitamente al proceso generador de observaciones faltantes.

3.2. ANALISIS DE UNA OBSERVACION COMPLETA.

En el caso en que la matriz del diseño X y el vector de observaciones de la variable respuesta Y estén completas, la obtención de la distribución posterior de los parámetros involucrados condicionada a los datos, ha sido publicada en muchos lugares y es una solución bayesiana ampliamente conocida. Esta solución que puede encontrarse, por ejemplo, en Raiffa & Schlaifer (1961, pp. 334-353), se presenta a continuación básicamente por dos razones: i) con ella se obtendrá la forma de incluir una sola observación como un caso particular de ésta; y ii) servirá para tener la presente al probar que en el caso de contar con varias observaciones completas, es equivalente incluirlas en bloque o de una en una.

Sea $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$, donde A' es el tranpuesto de A , un vector de parámetros cuyos valores están en \mathbb{R}^k [$k \geq 1$] y sea w otro parámetro real positivo.

Sean X la matriz de dimensión $[n \times p]$ y elementos x_{ij} , conocida con el nombre de matriz del diseño y \underline{y} el vector de dimensión $[n \times 1]$ y elementos y_i que representa las observaciones de la variable dependiente.

Suponga que la distribución de \underline{y} dado $X, \underline{\beta}$ y w es Normal (n variada) con media $X\underline{\beta}$ y precisión wI_n (I_n matriz identidad de dimensión n), que $\underline{\beta}$ y w son desconocidos y que X está fija (Lindley 1965, pg. 263).

El problema de inferir sobre los valores de $\underline{\beta}$ y w basándose en los valores de X y \underline{y} se conoce como el problema de regresión múltiple.

Un estimador de mínimos cuadrados de $\underline{\beta}$ se define como el vector $\hat{\underline{\beta}}$ que minimiza la forma cuadrática $(\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta})$ que aparece en el exponente de la función de verosimilitud obtenida a partir de la distribución de \underline{y} dado $X, \underline{\beta}$ y w mencionada anteriormente. Por diferenciación, se demuestra (Graybill, 1961, p. 11-12) que $\hat{\underline{\beta}}$ minimizará esta forma cuadrática si y sólo si satisface la ecuación conocida como "ecuación normal":

$$X'X\hat{\underline{\beta}} = X'\underline{y}$$

En el libro de Rao [1965], pg. 24-25) se establece la existencia de, al menos una solución a este sistema de ecuaciones. En general la solución de este sistema de ecuaciones no es única. Una solución es $\hat{\beta} = (X'X)^{-} X'Y$ donde $(X'X)^{-}$ es la inversa generalizada de la matriz $X'X$; hay que recordar que en el caso en que $X'X$ sea invertible entonces $(X'X)^{-} = (X'X)^{-1}$...

Es importante comentar que aunque se utilizan inversas generalizadas a lo largo de los desarrollos presentados en este capítulo, al final se demuestra que la solución obtenida es invariante ante elecciones de inversas generalizadas.

La distribución inicial conjunta de β y ω que se considera es una distribución Normal Gamma con parámetros μ de dimensión $(p \times 1)$, t de dimensión $(p \times p)$, simétrica y definida positiva, y escalares α y γ positivos. En este trabajo se denota a esta distribución como $NG(p, \mu, t, \gamma, \alpha)$. La función de densidad de esta distribución es:

$$p(\beta, \omega) = w^{p/2} \exp \left\{ -\frac{1}{2} (\beta - \mu)' t (\beta - \mu) / 2 \right\}$$

$$\times w^{\alpha-1} \exp \{-w \gamma\} \text{ si } \beta \in \mathbb{R}^k \text{ y } w > 0$$

$p(\underline{\beta}, w) = 0$ en otro caso.

La distribución posterior conjunta de $\underline{\beta}$ y w cuando se han observado a \underline{y} y X es:

$$\begin{aligned} p(\underline{\beta}, w | \underline{y}, X) &= p(\underline{\beta}, w) p(\underline{y} | \underline{\beta}, w, X) = \\ &= w^{p/2} \exp \{-w(\underline{\beta} - \underline{\mu})' \underline{x} (\underline{\beta} - \underline{\mu}) / 2\} \\ &\quad \cdot w^{n-1} \exp \{-w\gamma\} \cdot w^{n/2} \\ &\quad \cdot \exp \{-w(\underline{y} - X\underline{\beta})' (\underline{y} - X\underline{\beta}) / 2\} \end{aligned}$$

Como $X(X'X)^{-1}(X'X) = X$ (apéndice 1)⁽¹⁾ entonces:

$$\begin{aligned} (\underline{y} - X\underline{\beta})' (\underline{y} - X\underline{\beta}) &= (\underline{y} - X\hat{\underline{\beta}})' (\underline{y} - X\hat{\underline{\beta}}) + (\underline{\beta} - \hat{\underline{\beta}})' (X'X) (\underline{\beta} - \hat{\underline{\beta}}) \\ &= \underline{y}'\underline{y} - \hat{\underline{\beta}}' (X'X) \hat{\underline{\beta}} + (\underline{\beta} - \hat{\underline{\beta}})' (X'X) (\underline{\beta} - \hat{\underline{\beta}}) \end{aligned}$$

de donde

$$\begin{aligned} p(\underline{\beta}, w | \underline{y}, X) &= w^{p/2} \exp \{-w(\underline{\beta} - \underline{\mu})' \underline{x} (\underline{\beta} - \underline{\mu}) / 2 + \\ &\quad + (\underline{\beta} - \hat{\underline{\beta}})' (X'X) (\underline{\beta} - \hat{\underline{\beta}}) / 2\} \cdot w^{(n+n/2)-1} \\ &\quad \cdot \exp \{-w[\gamma + (\underline{y}'\underline{y} - \hat{\underline{\beta}}' (X'X) \hat{\underline{\beta}}) / 2]\} . \end{aligned}$$

(1) Este resultado puede obtenerse en forma equivalente si se considera simplemente que $\hat{\underline{\beta}}$ es solución de la ecuación normal.

Es fácil demostrar que:

$$\begin{aligned} & (\underline{\beta} - \underline{\mu}_1)' \underline{x} (\underline{\beta} - \underline{\mu}_1) + (\underline{\beta} - \underline{\beta})' (X'X) (\underline{\beta} + \underline{\beta}) \\ &= (\underline{\beta} - \underline{\mu}_1)' \underline{x}_1 (\underline{\beta} - \underline{\mu}_1) + \underline{\mu}' \underline{x} \underline{\mu} + \underline{\beta}' (X'X) \underline{\beta} - \underline{\mu}' \underline{x}_1 \underline{\mu}_1, \end{aligned}$$

donde $\underline{x}_1 = \underline{x} + X'X$ y $\underline{\mu}_1 = \underline{x}_1^{-1} (\underline{x} \underline{\mu} + X' \underline{y})$.

Es importante observar que \underline{x}_1 resulta definida positiva y por tanto invertible debido a que \underline{x} es definida positiva.

Utilizando este resultado, se tiene que:

$$\begin{aligned} p(\underline{\beta}, w | \underline{y}, X) &= w^{p/2} \exp \left\{ -w \left[(\underline{\beta} - \underline{\mu}_1)' \underline{x}_1 (\underline{\beta} - \underline{\mu}_1) / 2 \right] \right\} \cdot \\ &\quad \cdot w^{\alpha_1 - 1} \exp \{ -w \gamma_1 \}, \end{aligned}$$

donde $\alpha_1 = \alpha + n/2$ y

$$\begin{aligned} \gamma_1 &= \gamma + (\underline{y}' \underline{y} + \underline{\mu}' \underline{x} \underline{\mu} - \underline{\mu}' \underline{x}_1 \underline{\mu}_1) / 2 \\ &= \gamma + ((\underline{y} - X \underline{\mu}_1)' \underline{y} + (\underline{\mu} - \underline{\mu}_1)' \underline{x} \underline{\mu}) / 2; \end{aligned}$$

entonces, puede concluirse que la distribución posterior conjunta de $\underline{\beta}$ y w dados \underline{y} y X es $NG(p; \underline{\mu}_1, \underline{x}_1, \gamma_1, \alpha_1)$.

Es importante saber que esta distribución posterior es inva-

riante ante elecciones de Inversa generalizada de $X'X$.

Si la inicial de referencia sobre $\underline{\beta}$ y w es la comunmente aceptada por diversos autores, esto es:

$$p(\underline{\beta}, w) \propto w^{-1},$$

entonces la distribución posterior

$$p(\underline{\beta}, w | Y, X) \text{ es } NG(p, \hat{\underline{\beta}}, X'X, \gamma_1, (n-p)/2)$$

donde

$$\begin{aligned} \gamma_1 &= (\underline{Y} - X\hat{\underline{\beta}})' \underline{Y} / 2 = (\underline{Y} - X\hat{\underline{\beta}})' (\underline{Y} - X\hat{\underline{\beta}}) / 2 - \\ &= (\underline{Y}' \underline{Y} - \underline{Y}' X (X'X)^{-1} X' \underline{Y}) / 2 \end{aligned}$$

Este resultado puede verificarse fácilmente ya que la inicial de referencia puede expresarse como una $NG(p, \underline{\mu}, t, \gamma, \alpha)$ si $t \rightarrow 0$, $\gamma \rightarrow 0$ y $\alpha \rightarrow -p/2$.

Con los resultados anteriores quedar demostrados los siguientes teoremas:

Teorema 1.- Si X es una matriz de dimensión $(n \times p)$, y si para X fija el vector \underline{Y} de dimensión $(n \times 1)$ es tal que $\underline{Y} \sim N(X\underline{\beta}, wI)$ donde w es parámetro de precisión y el vector de $\underline{\beta}$ de dimensión $(p \times 1)$ es parámetro de localización, entonces si la distribución inicial conjunta de $\underline{\beta}$ y w es $NG(p, \underline{\mu}, t, \gamma, \alpha)$ se tiene que la distribución posterior conjunta de $\underline{\beta}$ y w dados \underline{Y} y X es $NG(p, \underline{\mu}_1, t_1, \gamma_1, \alpha_1)$ donde:

$$\underline{x}_1 = \underline{x} + \underline{X}'\underline{X},$$

$$\underline{\mu}_1 = \underline{x}_1^{-1} (\underline{x} \underline{\mu} + \underline{X}'\underline{y}),$$

$$\alpha_1 = \alpha + n/2$$

$$\begin{aligned} \gamma & \quad \underline{\gamma}_1 = \underline{\gamma} + (\underline{y}'\underline{y} + \underline{\mu}'\underline{x}\underline{\mu} - \underline{\mu}_1' \underline{x}_1 \underline{\mu}_1) / 2 \\ & \quad = \underline{\gamma} + ((\underline{y} - \underline{X} \underline{\mu}_1)' \underline{y} + (\underline{\mu} - \underline{\mu}_1)' \underline{x} \underline{\mu}) / 2. \end{aligned}$$

Teorema 2.- Bajo las condiciones del teorema anterior se tiene que si se toma una distribución inicial de referencia para $\underline{\beta}$ y w tal que

$$p(\underline{\beta}, w) = w^{-1}$$

entonces la distribución posterior de $\underline{\beta}$ y w dados \underline{y} y \underline{X} es $NG(p, \hat{\underline{\beta}}, \underline{X}'\underline{X}, \underline{\gamma}_1, (n-p)/2)$

donde

$$\begin{aligned} \hat{\underline{\beta}} &= (\underline{X}'\underline{X})^{-1} \underline{X}'\underline{y} \\ \gamma & \quad \underline{\gamma}_1 = (\underline{y} - \underline{X} \hat{\underline{\beta}})' \underline{y} / 2 = (\underline{y} - \underline{X} \hat{\underline{\beta}})' (\underline{y} - \underline{X} \hat{\underline{\beta}}) / 2 \\ & \quad = (\underline{y}'\underline{y} - \underline{y}'\underline{X} (\underline{X}'\underline{X})^{-1} \underline{X}'\underline{y}) / 2 \end{aligned}$$

En el caso en que $\underline{X}'\underline{X}$ no sea de rango máximo (esto es, invertible), se tiene que la distribución posterior obtenida con una a priori de referencia es invariante ante la elección de la inversa generalizada de $\underline{X}'\underline{X}$. Esta invarianza se consigue con base en que $\underline{X}(\underline{X}'\underline{X})^{-}\underline{X}'$ es invariante (ver lema 2.26 en Rao & Mitra (1971)).

Considerando el problema de regresión de tal forma que se

calcule la distribución posterior de $\underline{\beta}$ y w observación a observación, se tiene que:

$$\underline{y} = (y_1, y_2, \dots, y_n)'; \quad X = (\underline{X}_1, \dots, \underline{X}_n)';$$

y

$$\underline{\beta} = (\beta_1, \dots, \beta_p)';$$

donde $\underline{X}_i = (X_{i1}, \dots, X_{ip})$ con $i=1, \dots, n$.

Con esta notación la i -ésima observación es (y_i, \underline{X}_i) ; $i=1, \dots, n$.

Si la distribución conjunta de $\underline{\beta}$ y w es Normal Gamma $(p, G_0, t_0, \gamma_0, \alpha_0)$, entonces se tiene que al considerar la observación uno:

$$p(\underline{\beta}, w | y_1, \underline{X}_1) \text{ es } NG(p, \underline{\mu}_1, t_1, \gamma_1, \alpha_1)$$

donde

$$t_1 = t_0 + \underline{X}_1' \underline{X}_1,$$

$$\underline{\mu}_1 = t_1^{-1} (t_0 \underline{\mu}_0 + \underline{X}_1' y),$$

$$\gamma_1 = \gamma_0 + [(y_1 - \underline{X}_1' \underline{\mu}_1)' y_1 + (\underline{\mu}_0 - \underline{\mu}_1)' t_0 \underline{\mu}_0] / 2$$

y

$$\alpha_1 = \alpha_0 + 0.5.$$

Este resultado se obtiene como un caso particular del desarrollo en anteriores párrafos si se considera a y_1 como un vec

tor respuesta de orden $[1 \times 1]$ y a X_1 como una matriz de diseño de orden $[1 \times p]$.

En este punto debe tomarse en cuenta que siempre que se considere solamente una observación $X'X$ es de dimensión $p \times p$ y de rango 1, por lo tanto no es invertible y resulta suficiente que t_0 sea definida positiva para que t_1 resulte definida positiva e invertible. Por lo tanto no puede usarse una distribución de referencia como inicial. Sin embargo, en el caso en que haya que usar una distribución a priori de referencia, siempre es posible esperar hasta disponer de h observaciones ($h > p$) tales que X , la matriz del diseño formada por estas h observaciones sea de rango completo y entonces obtener la posterior de β y w de las estas h observaciones, prosiguiendo el análisis estadístico con esta distribución como la distribución inicial para la inclusión de la siguiente observación.

Analógicamente a la forma en que se pasa de la información a priori a la asimilación de la observación número uno, se pasa de haber incorporado $T-1$ observaciones a incorporar la observación T , obteniendo que $p(\beta, w | y_1, \dots, y_T, X_1, \dots, X_T)$ es una distribución $NG(p, \underline{\mu}_T, t_T, \gamma_T, a_T)$ donde

$$t_T = t_{T-1} + \frac{X_T' X_T}{\sigma^2}$$

$$\underline{\mu}_T = t_T^{-1} (t_{T-1} \underline{\mu}_{T-1} + \frac{X_T' y_T}{\sigma^2})$$

$$\gamma_T = \gamma_{T-1} + \left[(\underline{y}_T - \underline{X}_T \underline{\mu}_T)' \underline{y}_T + (\underline{\mu}_{T-1} - \underline{\mu}_T)' \underline{t}_{T-1} \underline{\mu}_{T-1} \right] / 2$$

$$y \quad \alpha_T = \alpha_{T-1} + 0.5$$

para $T=1, 2, \dots$

Si se tienen n observaciones y un conocimiento inicial $NG(p, \underline{\mu}_0, \underline{t}_0, \gamma_0, \alpha_0)$ puede optarse por incluir en conjunto a las n observaciones utilizando la matriz del diseño X y el vector \underline{y} , o incluir observación a observación analizando, si se desea, el impacto que causa cada observación en la distribución posterior. Desde luego que estas dos maneras de proceder coinciden, y como las distribuciones resultantes en cada caso son Normales Gamma, basta con comparar los parámetros para demostrarlo, siendo obvio en el caso de observaciones.

Como ya se demostró en el caso de trabajar con la matriz X y el vector \underline{y} se tiene que la distribución posterior es una $NG(p, \underline{\mu}_I, \underline{t}_I, \gamma_I, \alpha_I)$ donde

$$\underline{t}_I = \underline{t}_0 + X'X,$$

$$\underline{\mu}_I = \underline{t}_I^{-1} (\underline{t}_0 \underline{\mu}_0 + X' \underline{y}),$$

$$\gamma_I = \gamma_0 + \left[(\underline{y} - X \underline{\mu}_I)' \underline{y} + (\underline{\mu}_0 - \underline{\mu}_I)' \underline{t}_0 \underline{\mu}_0 \right] / 2$$

$$y \quad \alpha_I = \alpha_0 + n/2.$$

En el caso de incluir observación a observación se tiene que

una vez incluidas las n observaciones, la distribución posterior resultante es:

$NG(\mu_n, \tau_n, \gamma_n, \alpha_n)$ donde

$$\tau_n = \tau_{n-1} + \frac{X_n' X_n}{n}$$

$$\mu_n = \tau_n^{-1} (\tau_{n-1} \mu_{n-1} + X_n' y_n)$$

$$\gamma_n = \gamma_{n-1} + \left[y_n - X_n \mu_{n-1} \right]' y_n + (\mu_{n-1} - \mu_n)' \tau_{n-1} \mu_{n-1} / 2$$

$$\alpha_n = \alpha_{n-1} + 0.5$$

entonces

$$a) \alpha_n = \alpha_{n-1} + 1/2 = \alpha_{n-2} + 2/2 = \dots = \alpha_{n-K} + K/2 = \dots =$$

$$= \alpha_0 + n/2 = \alpha_T$$

$$b) \tau_n = \tau_{n-1} + \frac{X_n' X_n}{n} = \tau_{n-2} + \frac{X_{n-1}' X_{n-1}}{n-1} + \frac{X_n' X_n}{n} =$$

$$\dots = \tau_0 + \frac{1}{n} \sum_{i=1}^n X_i' X_i$$

y como $X = (X_1', \dots, X_n')$ entonces $X'X = \sum_{i=1}^n X_i' X_i$

y por lo tanto $\tau_n = \tau_T$.

$$c) \quad t_n \underline{\mu}_n = t_{n-1} \underline{\mu}_{n-1} + \underline{X}'_n \underline{y}_n = t_{n-2} \underline{\mu}_{n-2} + \underline{X}'_{n-1} \underline{y}_{n-1} + \underline{X}'_n \underline{y}_n = \dots = t_0 \underline{\mu}_0 + \sum_{i=1}^n \underline{X}'_i \underline{y}_i$$

y como $\underline{y} = (y_1, \dots, y_n)'$ entonces $\underline{X}' \underline{y} = \sum_{i=1}^n \underline{X}'_i y_i$

y por lo tanto $t_n \underline{\mu}_n = t_I \underline{\mu}_I$ y como $t_n = t_I$ y existe t_I^{-1} entonces $\underline{\mu}_n = \underline{\mu}_I$.

$$d) \quad \gamma_n = \gamma_{n-1} + \{y_n^2 + \underline{\mu}'_{n-1} t_{n-1} \underline{\mu}_{n-1} - \underline{\mu}'_n t_n \underline{\mu}_n\} / 2$$

$$= \gamma_{n-2} + \{y_{n-1}^2 + y_n^2 + \underline{\mu}'_{n-2} t_{n-2} \underline{\mu}_{n-2} - \underline{\mu}'_n t_n \underline{\mu}_n\} / 2$$

$$= \dots = \gamma_0 + \left\{ \sum_{i=1}^n y_i^2 + \underline{\mu}'_0 t_0 \underline{\mu}_0 - \underline{\mu}'_n t_n \underline{\mu}_n \right\} / 2$$

y como $\underline{\mu}_n = \underline{\mu}_I$ y $t_I = t_n$ entonces

$$\gamma_n = \gamma_0 + \{ \underline{y}' \underline{y} + \underline{\mu}'_0 t_0 \underline{\mu}_0 - \underline{\mu}'_n t_n \underline{\mu}_n \} / 2$$

$$= \gamma_I$$

Por lo tanto es equivalente incluir las observaciones una a una o todas a la vez. Esta es una propiedad de los procedimientos bayesianos.

El resultado anterior demuestra el siguiente teorema:

Teorema 3.- Bajo las condiciones del teorema 1 y considerando a \underline{y} y X como n observaciones de la forma (y_i, X_i) para $i=1, \dots, n$, donde X_i es el i -ésimo renglón de X , se tiene que la distribución posterior conjunta de $\underline{\beta}$ y w puede obtenerse mediante cualquiera de los siguientes procedimientos:

- a) Incluir en conjunto las n observaciones, utilizando la matriz X y el vector \underline{y} β
- b) Incluir observación a observación calculando cada vez la distribución posterior de $\underline{\beta}$ y w y utilizandola para la siguiente observación como distribución inicial, realizando este proceso hasta haber incluido las n observaciones.

Respecto a las distribuciones marginales asociadas a una distribución Normal Gamma, el lector interesado puede consultar el libro de De Groot (1970), pg. 183.

3.3. ANALISIS DE UNA OBSERVACION INCOMPLETA.

En el caso en que una observación (y, \underline{X}) sea incompleta por haber sufrido pérdidas en sus componentes, es necesario conocer la manera de incorporar al análisis estadístico la información contenida en la observación. Para esto, se presentan a continuación las diferentes maneras en que una observación puede ser incompleta, discutiendo en cada caso la distribución posterior de $\underline{\beta}$ y w dada la observación incompleta.

3.3.1. Pérdida del valor de la variable dependiente.

Suponga que se ha perdido el valor de la variable dependiente Y y que además se han perdido q ($0 \leq q \leq p$) componentes del vector de observaciones \underline{X} correspondiente a los valores de las variables independientes. Puede suponerse sin pérdida de generalidad que han sido las q primeras componentes y expresar a \underline{X} como $(\underline{v}, \underline{X})$, donde \underline{v} y \underline{X} son vectores de orden $(1 \times q)$ y $(1 \times (p-q))$ respectivamente, además \underline{v} es aleatorio y representa a las observaciones perdidas de \underline{X} y \underline{X} representa a las observaciones no perdidas de \underline{X} .

Por medio de una aplicación directa del teorema de Bayes, se tiene que si $q=0$

$$p(\underline{\beta}, w, y | \text{datos}) = p(\underline{\beta}, w, y | \underline{X}) = \\ = p(\underline{\beta}, w) p(y | \underline{\beta}, w) p(\underline{X} | \underline{\beta}, w, y),$$

y en esta situación es aceptable suponer⁽²⁾ que $p(\underline{X}|\underline{\beta}, w, y)$ es constante. Integrando sobre y se tiene que:

$$p(\underline{\beta}, w | \underline{X}) = p(\underline{\beta}, w)$$

concluyendo que si se pierde el valor de la variable dependiente y se observa todo el vector \underline{X} , la distribución posterior de $\underline{\beta}$ y w coincide con la distribución inicial y por lo tanto \underline{X} no añade conocimiento sobre $\underline{\beta}$ o w .

En el caso en que $q > 0$ se tiene, al igual que cuando $q=0$, que \underline{X} no añade información alguna sobre $\underline{\beta}$ o w cuando el valor de y se pierde. Esto puede entenderse si se piensa que si \underline{X} no aporta nueva información sobre $\underline{\beta}$ y w , tampoco lo hará si se han perdido algunos de sus componentes.

3.3.2 Observación del valor de la variable dependiente.

En esta sección se supone que el valor de la variable dependiente y ha sido observado y que, por tanto, si hay pérdidas en una observación (y, \underline{X}_i) , estas deben estar en el vector \underline{X}_i , de orden $(1 \times p)$.

Supóngase que q ($0 < q < p$) componentes de \underline{X} se han perdido y (sin pérdida de generalidad) que estas son las q primeras componentes de \underline{X} que serán denotadas por \underline{v} , un vector de dimensión $(1 \times q)$. Sea \underline{X} el vector de dimensión $(1 \times (p-q))$ correspondiente

(2) Recordar que se trata de un problema de regresión y que el modelo es $E(Y) = \underline{X}\underline{\beta}$.

a los componentes de \underline{X} no extraviadas.

Supóngase que las variables explicativas tienen media cero.

Considere, como es usual, que la distribución condicional de Y dados $\underline{\beta}, w$ y \underline{X} es Normal, con media $\underline{X} \underline{\beta}$ y precisión w .

Sean \underline{a} y $\underline{\beta}$ vectores de parámetros de orden $(q \times 1)$ y $((p-q) \times 1)$ respectivamente y tales que $\underline{\beta} = (\underline{a}', \underline{\beta}')'$, con \underline{a} representando a los parámetros correspondientes a las observaciones extraviadas (\underline{y}) y $\underline{\beta}$ representando a las correspondientes a \underline{X} .

Entonces re-expresando a la media de Y dado \underline{X} , $\underline{\beta}$ y w se tiene

$$E\{Y \mid \underline{X}, \underline{\beta}, w\} = \underline{X} \underline{\beta} = \underline{y} \underline{a} + \underline{X} \underline{\beta}.$$

Sea $\underline{0}$ un vector cuyas componentes son cero y que tiene orden $q \times 1$. Si $\underline{a} = \underline{0}$ entonces no intervienen realmente en el modelo las observaciones perdidas, mientras que si $\underline{a} \neq \underline{0}$ interviene al menos una. Para facilitar la discusión siguiente, se consideran estos dos casos por separado.

Si $\underline{a} \neq \underline{0}$, se tiene que:

$$p(\underline{\beta}, w, \underline{y} \mid \text{datos}) = p(\underline{\beta}, w, \underline{y}) p(\underline{y} \mid \underline{\beta}, w, \underline{X})$$

$$= p(\underline{\beta}, w, \underline{v} | w^{1/2} \exp\{-w(\underline{y}-\underline{X}\underline{\beta})'(\underline{y}-\underline{X}\underline{\beta})/2\}) .$$

Si $y^0 = \underline{y}'\underline{\beta}$, entonces se tiene que

$$(\underline{y}-\underline{X}\underline{\beta}) = (\underline{y}-\underline{v}\underline{a} - \underline{X}\underline{\beta}) = y^0 - \underline{v}\underline{a} ,$$

y como $\underline{v}\underline{a} = (\underline{v}\underline{a}' = \underline{a}'\underline{v}')$, entonces $(\underline{y}-\underline{X}\underline{\beta}) = y^0 - \underline{a}'\underline{v}'$,

de donde:

$$p(\underline{\beta}, w, \underline{v} | \text{datos}) = p(\underline{\beta}, w, \underline{v} | w^{1/2} \exp\{-w(y^0 - \underline{a}'\underline{v}')' (y^0 - \underline{a}'\underline{v}')/2\}) .$$

Utilizando inversas generalizadas y los resultados del apéndice J, que afirman que $\underline{a}'(\underline{a}\underline{a}' + \underline{I})^{-1}(\underline{a}\underline{a}' + \underline{I})^{-1}\underline{a} = \underline{a}$ y $(\underline{a}\underline{a}' + \underline{I})(\underline{a}\underline{a}' + \underline{I})^{-1}\underline{a} = \underline{a}$, se tiene que

$$(y^0 - \underline{a}'\underline{v}')' (y^0 - \underline{a}'\underline{v}') = (\underline{v} - \underline{a})' D (\underline{v} - \underline{a})$$

donde $D = \underline{a}\underline{a}'$, $\underline{a} = y^0 \underline{a}'(\underline{a}\underline{a}' + \underline{I})^{-1}$ y $\underline{a}' D \underline{a} = y^0$.

Sustituyendo en $p(\underline{\beta}, w, \underline{v} | \text{datos})$ la igualdad

$$(y^0 - \underline{a}'\underline{v}')' (y^0 - \underline{a}'\underline{v}') = (\underline{v} - \underline{a})' D (\underline{v} - \underline{a})$$

se tiene:

$$p(\underline{\beta}, w, y | \text{datos}) \propto p(\underline{\beta}, w, y) \cdot w^{1/2} \exp \left\{ -w(y - \underline{\alpha})' D (y - \underline{\alpha}) / 2 \right\}$$

Como $p(\underline{\beta}, w, y) = p(\underline{\beta}, w) p(y | \underline{\beta}, w)$ entonces puede procederse a especificar $p(y | \underline{\beta}, w)$.

La distribución a priori sobre las observaciones perdidas deberá tener media cero debido a que se ha supuesto que las variables explicativas tienen media cero, puede suponerse que su matriz de precisión es diagonal si se considera que se está trabajando bajo la suposición de pérdidas aleatorias resultando que una forma general de los elementos diagonales es d/w , con d constante. Puede suponerse una distribución Normal q -variada al considerar las "bondades explicativas" del modelo Normal y la conveniencia de trabajar con una familia de distribuciones, en este caso la Normal Gamma, cerrada ante el muestreo esto es, una familia conjugada de distribuciones.

Para permitir la obtención analítica de la distribución posterior de $\underline{\beta}$ y w (es necesario en este punto suponer que $d = (\underline{\alpha}' \underline{\alpha}) / 2$ donde (como se recorda) $\underline{\alpha}$ son las β 's correspondientes a las observaciones faltantes. Como se verá en el capítulo 4, la distribución posterior obtenida con base en esta forma de especificar el conocimiento sobre las observaciones perdidas resulta intuitivamente aceptable. Por otra parte, considerando por simplicidad $q=1$, si se mide la contribución de una variable explicativa a la formación de la variable dependiente mediante el valor absoluto del coeficiente β respectivo se tiene, que entre más contribuye la variable más concentrada en el cero estará la distribución inicial, disminuyendo con esto la contribución de la variable explicativa en los casos en que

haya pérdidas. En el caso en que el valor absoluto del coeficiente sea "pequeño", se tendrá que la variable explicativa contribuye "poco" y que entre menos contribuye menos concentrada en el cero estará la distribución a priori. Por lo tanto la elección de d resulta razonable.

Considerando a $p(y|\beta, w)$ de la forma antes descrita se tiene:

$$p(\beta, w, v | \text{datos}) = p(\beta, w) \cdot d^{n/2} w^{n/2} \exp\{-w \underline{v} (d I) \underline{v}' / 2\} \cdot \\ \cdot w^{1/2} \exp\{-w(\underline{y}-\underline{a})' D (\underline{y}-\underline{a})' / 2\}.$$

Por otro lado

$$(\underline{y}-\underline{a})' D (\underline{y}-\underline{a})' + \underline{v} (d I) \underline{v}' = \\ = (\underline{y}-\underline{a}_1)' D_1 (\underline{y}-\underline{a}_1)' + \underline{a} D \underline{a}' - \underline{a}_1 D \underline{a}'$$

$$\text{donde } D_1 = d I + D, D_1^{-1} = d^{-1} I - (2 d^2)^{-1} (\underline{\alpha} \underline{\alpha}'), \text{ y } \underline{a}_1 = \underline{a} D D_1^{-1}.$$

Además se demuestra en el apéndice 2 que si A es una matriz cuadrada de orden K , X es un vector de orden $(K \times 1)$ y existe A^{-1} , entonces $|A \cdot X X'| = |A| \cdot |I + X' A^{-1} X|$, de aquí que

$$|D_1| = |d I + \underline{\alpha} \underline{\alpha}'| = |d I| \cdot |I + \underline{\alpha}' (\frac{1}{d} I) \underline{\alpha}| \\ = d^q (3).$$

Hay que notar que $d \neq 0$ ya que $\underline{\alpha} \neq 0$.

Entonces:

$$p(\underline{\beta}, w, \underline{y} \mid \text{datos}) = p(\underline{\beta}, w) \cdot$$

$$\cdot |\mathcal{D}_1|^{1/2} w^{q/2} \exp \{-w(\underline{y} - \underline{a}_1)' \mathcal{D}_1 (\underline{y} - \underline{a}_1) / 2\}$$

$$\cdot w^{1/2} \exp \{-w(\underline{a}_1' \mathcal{D}_1 \underline{a}_1 - \underline{a}_1' \mathcal{D}_1 \underline{a}_1) / 2\}$$

Si se integra $p(\underline{\beta}, w, \underline{y} \mid \text{datos})$ respecto a \underline{y} , que al observar la expresión anterior se concluye que es un vector Normal q -variado con vector de medias \underline{a}_1 y matriz de precisión $w \mathcal{D}_1$, se tiene que

$$p(\underline{\beta}, w \mid \text{datos}) = p(\underline{\beta}, w) \cdot$$

$$\cdot w^{1/2} \exp \{-w(\underline{a}_1' \mathcal{D}_1 \underline{a}_1 - \underline{a}_1' \mathcal{D}_1 \underline{a}_1) / 2\}.$$

Usando el resultado $\underline{a}_1' \mathcal{D}_1 \underline{a}_1 = 0$ (demostrado en el apéndice 3) y recordando que $\underline{a}_1' \mathcal{D}_1 \underline{a}_1 = y^{*2}$, se obtiene

$$p(\underline{\beta}, w \mid \text{datos}) = p(\underline{\beta}, w) \cdot w^{1/2} \exp \{-w y^{*2} / 2\}. \quad \dots (1)$$

Ahora bien definiendo a $\underline{\hat{\beta}} = (\underline{X}' \underline{X})^{-1} \underline{X}' y$, se tiene que:

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

con lo que se obtiene al considerar que $y^* = y - \underline{X} \underline{\hat{\beta}}$

$$p(\underline{\beta}, w \mid \text{datos}) = p(\underline{\beta}, w) \cdot w^{1/2} \exp \{-w(y - \underline{X} \underline{\hat{\beta}})'(y - \underline{X} \underline{\hat{\beta}}) / 2\}$$

$$\begin{aligned}
 &= p(\underline{\beta}, \omega) \cdot \omega^{1/2} \exp\{-\omega(\underline{\beta} - \hat{\underline{\beta}})'(\underline{X}'\underline{X})(\underline{\beta} - \hat{\underline{\beta}})/2\} \\
 &= p(\underline{\beta}, \omega) \omega^{1/2} \exp\{-\omega(\underline{\beta} - \underline{\mu}^0)'T(\underline{\beta} - \underline{\mu}^0)/2\}
 \end{aligned}$$

donde $\underline{\mu}^0$ y T son matrices de orden $(p \times 1)$ y $(p \times p)$ respectivamente y están definidas de la siguiente manera:

$\underline{\mu}^0 = (\underline{0}', \hat{\underline{\beta}}')'$, donde $\underline{0}$ es el vector de ceros de orden $(q \times 1)$ y

$$T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}$$

donde T_{11} , T_{12} , T_{21} y T_{22} son matrices de orden $(q \times q)$, $q \times (p-q)$, $(p-q) \times q$ y $(p-q) \times (p-q)$ respectivamente, tales que T_{11} , T_{12} y T_{21} son idénticamente cero y $T_{22} = \underline{X}'\underline{X}$.

Si $p(\underline{\beta}, \omega)$ es una distribución Normal Gamma denotada mediante

$$NG(p, \underline{\mu}_0, t_0, \gamma_0, \alpha_0)$$

entonces

$$p(\underline{\beta}, \omega | \text{datos}) \propto$$

$$\begin{aligned}
 &\omega^{p/2} \exp\{-\omega(\underline{\beta} - \underline{\mu}_0)'t_0(\underline{\beta} - \underline{\mu}_0)/2\} \\
 &\cdot \omega^{\alpha_0 - 1} \exp\{-\omega \gamma_0\}
 \end{aligned}$$

$$\cdot w^{1/2} \exp \left\{ -w(\underline{\beta} - \underline{\mu}^0)' T (\underline{\beta} - \underline{\mu}^0) / 2 \right\}$$

Como $(\underline{\beta} - \underline{\mu}_1)' \underline{x}_0 (\underline{\beta} - \underline{\mu}_1) + (\underline{\beta} - \underline{\mu}^0)' T (\underline{\beta} - \underline{\mu}^0) =$

$$= (\underline{\beta} - \underline{\mu}_1)' \underline{x}_1 (\underline{\beta} - \underline{\mu}_1) + \underline{\mu}_1' \underline{x}_0 \underline{\mu}_1 + \underline{\mu}^0' T \underline{\mu}^0 - \underline{\mu}_1' \underline{x}_1 \underline{\mu}_1,$$

donde

$$\underline{x}_1 = \underline{x}_0 + T$$

(que resulta definida positiva por serlo \underline{x}_0) y

$$\underline{\mu}_1 = \underline{x}_1^{-1} (\underline{x}_0 \underline{\mu}_1 + T \underline{\mu}^0) \quad (2)$$

Sustituyendo este resultado en $p(\underline{\beta}, w | \text{datos})$ se tiene que

$$p(\underline{\beta}, w | \text{datos}) = w^{p/2} \exp \left\{ -w(\underline{\beta} - \underline{\mu}_1)' \underline{x}_1 (\underline{\beta} - \underline{\mu}_1) / 2 \right\}$$

$$\cdot w^{\alpha_1 - 1} \exp \left\{ -w \gamma_1 \right\}$$

donde

$$\gamma_1 = \gamma_0 + (\underline{\mu}_1' \underline{x}_0 \underline{\mu}_1 + y^2 - \underline{\mu}_1' \underline{x}_1 \underline{\mu}_1) / 2 \quad (4)$$

(ya que $\underline{\mu}^0' T \underline{\mu}^0 = y^2$) y

$$\alpha_1 = \alpha_0 + 0.5 \quad (5)$$

Estos resultados demuestran el siguiente:

Teorema 4.- Sea \underline{X} un vector de dimensión $(1 \times p)$ y y un escalar. Suponga que las primeras q ($1 < q < p$) componentes de \underline{X} se han perdido aleatoriamente (sección 2.2) y que estas componentes se denotan mediante \underline{V} un vector de dimensión $(1 \times q)$ mientras que \underline{X} un

vector de dimensión $(1 \times (p-q))$ denota las componentes no perdidas. Suponga que la distribución de V dado \underline{X} es $N(\underline{X} \underline{\beta}, w)$ donde $\underline{\beta}$ es de dimensión $(p \times 1)$ y w es la precisión y que $\underline{\beta}$ y w son diferentes (sección 2.2) del parámetro del proceso generador de observaciones faltantes. Sean $\underline{\alpha}$ y $\underline{\beta}$ vectores de orden $(q \times 1)$ y $((p-q) \times 1)$ respectivamente tales que $\underline{\beta} = (\underline{\alpha}', \underline{\beta}')$. Entonces si la distribución inicial conjunta de $\underline{\beta}$, w , y V es tal que su densidad

$$p(\underline{\beta}, w, V) = P(\underline{\beta}, w, V) p(V | \underline{\beta}, w)$$

donde

$p(V | \underline{\beta}, w)$ es la densidad de una distribución Normal con media $\underline{\beta}(1 \times q)$ y matriz de precisión dwI ; con $d = (\underline{\alpha}' \underline{\alpha})/2$, y $p(\underline{\beta}, w)$ es la densidad de una distribución NG $(p, \underline{\mu}_0, t_0, \gamma_0, \nu_0)$ entonces la distribución posterior conjunta de $\underline{\beta}$ y w dados V, \underline{X} es una distribución NG $(p, \underline{\mu}_1, t_1, \gamma_1, \alpha_1)$ donde

$$t_1 = t_0 + T$$

$$T(p \times p) = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix},$$

$T_{11} (q \times q)$, $T_{12} (q \times (p-q))$, $T_{21} ((p-q) \times q)$ son matrices de ceros,

$$T_{22} ((p-q) \times (p-q)) = \underline{X}' \underline{X},$$

$$\underline{\mu}_1 = t_1^{-1} (t_0 \underline{\mu}_0 + T \underline{\mu}^*),$$

$$\underline{\mu}^* (p \times 1) = (\underline{0}', \underline{\beta}'),$$

$$\underline{0}' (q \times 1) \text{ vector de ceros,}$$

$$\underline{\beta}' ((p-q) \times 1) = (\underline{X}' \underline{X})^{-1} \underline{X}' y,$$

$$\gamma_1 = \gamma_0 + (\underline{\mu}_0' t_0 \underline{\mu}_0 + y^2 - \underline{\mu}_1' t_1 \underline{\mu}_1) / 2$$

$$y \quad \alpha_1 = \alpha_0 + 0.5.$$

Hay que notar que esta distribución posterior no depende de la elección de inversa generalizada de $(\underline{X}'\underline{X})$, ya que

$$T_{\underline{\beta}}^{\circ} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} \underline{\hat{\beta}} \\ \underline{\hat{\beta}} \end{pmatrix} = \begin{pmatrix} \underline{\hat{\beta}} \\ T_{22}\underline{\hat{\beta}} \end{pmatrix} = \begin{pmatrix} \underline{\hat{\beta}} \\ \underline{X}'\underline{y} \end{pmatrix}$$

Si $\underline{a} = \underline{0}$, entonces \underline{y} no tiene ingerencia en el modelo y $p(\underline{\beta}, w | \text{datos}, \underline{a} = \underline{0}) = p(\underline{\beta}, w | \underline{a} = \underline{0}) w^{1/2} \exp\{-\frac{w}{2} (y - \underline{X}\underline{\beta})^2\}$, esta expresión es similar a la expresión (1) obtenida en el caso $\underline{a} \neq \underline{0}$, por lo tanto la distribución obtenida en este caso tiene la misma forma la obtenida cuando $\underline{a} \neq \underline{0}$.

Esta solución ha supuesto que son las primeras q componentes de \underline{X} las que se han perdido. Esta suposición no es restrictiva en forma alguna, ya que sólo habría que proceder a llevar a cabo las permutaciones apropiadas en \underline{X} , $\underline{\beta}$, $\underline{\mu}_1$ y t_1 para que fuera adecuada, procediendo una vez incorporada la observación al análisis, a permutar a \underline{X} , $\underline{\beta}$, $\underline{\mu}_1$ y t_1 de tal forma que se recuperara el orden que originalmente se habla especificado.

En el caso en que $q=p$, se tendría que los resultados anteriores seguirían siendo válidos hasta la expresión (1), la cual en este caso sería

$$p(\underline{\beta}, w | \text{datos}) = p(\underline{\beta}, w) w^{1/2} \exp\{-\frac{w}{2} y^2\}$$

ya que $y^0 = y$ y por no existir \underline{x} .

Si $p(\underline{\beta}, w)$, la distribución inicial de $\underline{\beta}$ y w es Normal Gamma tal que se denote mediante $NG(p, \underline{\mu}_0, t_0, \gamma_0, \alpha_0]$, entonces la distribución posterior de $\underline{\beta}$ y w será también Normal Gamma y se de notarla mediante $NG(p, \underline{\mu}_1, t_1, \gamma_1, \alpha_1]$ donde

$$\gamma_1 = \gamma_0 + 0.5 [y^2] \quad (6)$$

y

$$\alpha_1 = \alpha_0 + 0.5 \quad (7)$$

En el siguiente capítulo, se discute ampliamente este caso en el ejemplo 1.

3.4 TRATAMIENTO DE UN CONJUNTO CON MAS DE UNA OBSERVACION.

En el caso en que se cuente con una distribución inicial, no hay ningún problema al tratar cualquier número de observaciones, sin importar si son completas o no. Pero la situación es diferente si no se tiene una distribución inicial (la del interesado), o no quiere o puede utilizarse esta. En esta situación, pueden distinguirse dos casos:

- a) Existen al menos $p+1$ observaciones completas. Como puede observarse a partir de las expresiones obtenidas para los parámetros de la distribución posterior (Normal Gamma)

en el caso de incluir en el análisis una sola observación sea esta completa o incompleta, la distribución posterior que se obtendría con un conjunto de n observaciones no depende en absoluto del orden en que estas n observaciones se haya realizado.

Debido a que en el sentido mencionado en el párrafo anterior las observaciones son intercambiables, se puede proceder en este caso a reordenar el conjunto de observaciones, de tal forma que primero entren al análisis las observaciones completas y a continuación las incompletas. Las observaciones completas se analizarían en bloque mediante el uso de una distribución de referencia y acto seguido se analizarían las incompletas una a una.

- b) Existen menos de $p+1$ observaciones completas. En este caso no pueden utilizarse las distribuciones iniciales de referencia usuales y es forzoso e imprescindible contar con información inicial subjetiva. Pero ya que se está suponiendo que esta información no existe, o no se quiere o puede usarse, se procederá a presentar una salida matemática al problema mediante la construcción de una distribución "inicial" lo menos informativa posible. Se sugiere por facilidad que sea miembro de la familia conjugada de distribucio-

nea, que en este caso es la Normal Gamma, $NG(p, \mu_0, \sigma_0^2, \gamma_0, \alpha_0)$. Se sugiere que los parámetros sean tales que: $\alpha_0 = \alpha I$, $\mu_0 = \underline{0}$ y se escoja a σ_0^2 , α y γ_0 como reales positivas "pequeñas". Esto permite iniciar matemáticamente el proceso de asimilación de la información y proporcionar una distribución posterior Normal-Gamma basada "poco" en los valores iniciales de los parámetros y por tanto "fuertemente" en la información muestral.

En el caso en que se tuvieran al menos $p+1$ observaciones de $|y|$ la variable dependiente, se podría elegir α_0 igual a $[-p/2]$, en un intento por tener una "distribución inicial" (impropia) parecida lo más posible a la inicial de referencia usual en problemas de regresión y propuesta entre otros por Bernardo (1975).

3.5 PARTICULARIDADES OPERATIVAS EN EL PROCESAMIENTO DE LOS DATOS.

Para llevar a cabo el procedimiento descrito en este capítulo, es necesario, en el caso de tener observaciones incompletas (sección 3.3.2), realizar permutaciones en X , β , μ_0 y t_0 antes de incorporar al análisis la observación, y en β , t_1 y μ_1 después de haberla incorporado. Operacionalmente, es posible simplificar el procedimiento descrito. A continuación se demuestra esto.

Considere la notación empleada en la sección 3.3.4 y defina

a \underline{z} de orden $[n \times q]$ de tal forma que $\underline{z} = [\underline{a}, \underline{x}]$, donde \underline{a} es el vector de ceros de orden $[n \times q]$. Resulta entonces que:

$$T = \underline{z}' \underline{z} \quad \text{y} \quad \underline{\mu}^0 = \begin{cases} \underline{a} \text{ si } \underline{z} = \underline{a} \\ (\underline{z}' \underline{z})^{-1} \underline{z}' \underline{y} \text{ si } \underline{z} \neq \underline{a} \end{cases}$$

y si $\underline{z} \neq \underline{a}$, entonces

$$(\underline{z}' \underline{z})^{-1} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & (\underline{x}' \underline{x})^{-1} \end{pmatrix} \quad \text{y} \quad \underline{z}' \underline{y} = \begin{pmatrix} \underline{a}' \underline{y} \\ \underline{x}' \underline{y} \end{pmatrix}$$

Las expresiones (2) y (3) de la sección 3.3.2, pueden expresarse de la siguiente manera:

$$\underline{t}_1 = \underline{t}_0 + \underline{z}' \underline{z} \quad (1)$$

$$\underline{\mu}_1 = \underline{t}_1^{-1} (\underline{t}_0 \underline{\mu}_0 + \underline{z}' \underline{y}) \quad (2)$$

ya que $T \underline{\mu}^0 = \underline{z}' \underline{y}$.

A partir de estas expresiones y de las expresiones (4) a (7) de la sección 3.3.2 puede observarse la equivalencia de las siguientes tres formas de proceder a la incorporación de una observación incompleta (en el sentido de la sección 3.3.2):

- a) Permutar apropiadamente a \underline{X} , y en consecuencia a $\underline{\beta}$, $\underline{\mu}_0$ y t_0 de forma que las primeras q componentes de \underline{X} sean las observaciones faltantes, incorporar la observación incompleta y proceder a re-permutar a $\underline{\beta}$, $\underline{\mu}_1$ y t_1 de forma tal que se consiga de nuevo el orden original.
- b) Definir a \underline{Z} en la forma especificada anteriormente, considerando simplemente a \underline{X} como el vector cuyas componentes son las observaciones de las variables explicativas no extraviadas. Permutar a \underline{Z} de tal forma que los elementos de \underline{X} queden en la posición que originalmente ocupan en \underline{X} . Calcular a $\underline{\mu}_1$ y t_1 mediante las expresiones (1) y (2) de esta sección y con estos dos valores a γ_1 y α_1 mediante las expresiones (4) y (5) de la sección 3.3.2.
- c) Sustituir en \underline{X} las observaciones faltantes por ceros y definir al vector resultante como \underline{Z} . Calcular a $\underline{\mu}_1$ y t_1 con las expresiones (1) y (2) de esta sección y a γ_1 y α_1 con las expresiones (4) y (5) de la sección 3.3.2.

Ahora bien en la sección 3.2, se ha demostrado que si la a priori de $\underline{\beta}$ y w es NG $(\underline{\mu}_0, t_0, \gamma_0, \alpha_0)$ entonces la distribución posterior de $\underline{\beta}$ y w obtenida de la incorporación al análi-

sis de una observación completa (y, X) es una distribución Normal Gamma, denotada por $NG(p, \underline{\mu}_1, t_1, \gamma, \alpha)$ donde:

$$t_1 = t_0 + \underline{X}'\underline{X} ,$$

$$\underline{\mu}_1 = t_1^{-1} (t_0 \underline{\mu}_0 + \underline{X}'y) ,$$

$$\gamma = \gamma_0 + 0.5 (\underline{\mu}'_0 t_0 \underline{\mu}_0 + y^2 - \underline{\mu}'_1 t_1 \underline{\mu}_1) ,$$

$$\alpha = \alpha_0 + 0.5 .$$

Además se ha demostrado que debe eliminarse la observación si el valor de y se ha perdido (sección 3.3.1).

En esta sección se ha visto que si el valor de y se observa, entonces la incorporación de la observación incompleta proporciona una distribución posterior $NG(p, \underline{\mu}, t, \gamma, \alpha)$ donde

$$t = t_0 + \underline{Z}'\underline{Z} ,$$

$$\underline{\mu} = t^{-1} (t_0 \underline{\mu}_0 + \underline{Z}'y) ,$$

$$\gamma = \gamma_0 + 0.5 (\underline{\mu}'_0 t_0 \underline{\mu}_0 + y^2 - \underline{\mu}' t \underline{\mu}) ,$$

y

$$\alpha = \alpha_0 + 0.5$$

en la que \underline{Z} es el vector de dimensión $(1 \times p)$ que se obtiene al sustituir en \underline{X} los valores perdidos por ceros.

A partir de estas expresiones para los parámetros de las distribuciones posteriores de $\underline{\beta}$ y w , en los casos en que (y, X)

o sea completa o las pérdidas no afecten a y , resulta fácil concluir que las expresiones para calcular los parámetros de la distribución posterior son similares y además, como se ha demostrado en la sección 3.2, con estas expresiones es equivalente proceder a la incorporación de las observaciones mediante una matriz de diseño, que proceder incorporándolas una a una. Entonces, a lo largo de esta sección se ha encontrado una forma de proceder operativamente en el tratamiento de observaciones incompletas, de tal manera que este tratamiento se puede llevar a cabo fácilmente mediante un programa de computadora que realice un análisis "estandar" de regresión múltiple.

3.5.1. ANALISIS DE LOS SUPUESTOS HECHOS.

Se ha supuesto en este trabajo que las variables explicativas son tales que no hay multicolinealidad. En el caso de observaciones incompletas con multicolinealidad el problema de regresión puede complicarse mucho, dependiendo del número de datos faltantes y la estructura de pérdidas en la matriz de datos. Este problema no se discute en este trabajo.

Respecto a la suposición de que las variables explicativas

tienen media cero, cabe decir que no es restrictiva ya que puede conseguirse la validez de esta suposición mediante una transformación de las variables.

Considere el modelo expresado de la siguiente forma:

$$Y = \beta_1 X_1 + \dots + \beta_p X_p + \epsilon$$

entonces un resultado ampliamente conocido en el caso en que no hay observaciones incompletas, es que al obtener la media aritmética de las Y 's, se tiene que:

$$\bar{Y} = \beta_1 \bar{X}_1 + \dots + \beta_p \bar{X}_p + \bar{\epsilon}$$

de donde

$$Y - \bar{Y} = \beta_1 (X_1 - \bar{X}_1) + \dots + \beta_p (X_p - \bar{X}_p) + \epsilon - \bar{\epsilon}$$

y en este caso las "nuevas" variables explicativas definidas por $Z_{ji} = (X_{ji} - \bar{X}_j)$; $j=1, \dots, p$ y $i=1, \dots, n$ tienen media cero.*

En el caso de observaciones incompletas, esta transformación no puede aplicarse directamente debido a que al menos algunas de las medias aritméticas involucradas no pueden calcularse.

Ante esta dificultad se plantea el siguiente procedimiento: con la variable explicativa j ($j=1, \dots, p$) calcular la media arit

(*) Los grados de libertad de la distribución marginal de β son en el caso en que se contemple en el modelo un término constante $n-p$ y en caso contrario $n-p+1$.

métrica con las observaciones disponibles de esa variable y denotarla por a_j . Hacer lo mismo con la variable dependiente y , denotando a la media por a . Si I_p es el símbolo utilizado para las observaciones perdidas y se define la suma de I_p con un real c de la siguiente manera $I_p + c = c + I_p = I_p$, entonces puede construirse nuevas variables explicativas (Z_i) y nueva variable dependiente (B_i) de la siguiente forma:

$$B_i = y_j - a_i = \begin{cases} I_p - a_i & \text{si } y_j = I_p \\ y_j - a_i & \text{si } y_j \neq I_p \end{cases}$$

$$Z_{ij} = x_{ij} - a_j = \begin{cases} I_p - a_j & \text{si } x_{ij} = I_p \\ x_{ij} - a_j & \text{si } x_{ij} \neq I_p \end{cases}$$

con $j=1, 2, \dots, n$ y $i=1, \dots, p$.

De esta manera el modelo se escribe como:

$$y - a_i = (c_i + \sum_{j=1}^n \beta_j a_j) + \sum_{j=1}^p \beta_j (x_j - a_j) + \epsilon$$

o equivalentemente como:

$$B_i = \beta_{0i} + \beta_1 Z_{i1} + \dots + \beta_p Z_{ip} + \epsilon_i$$

con $i=1, \dots, n$ y donde $\beta_{0i} = -a_i + \sum_{j=1}^p \beta_j a_{ij}$, β_j , a_{ij} , y la media de las variables explicativas es cero.⁵

El procedimiento discutido a lo largo de este capítulo es consecuencia matemática de la aplicación de la metodología bayesiana a un problema de regresión con observaciones faltantes. Una manera sencilla de proceder a la determinación de la distribución posterior de $\underline{\beta}$ y w puede resumirse de la siguiente forma:

- restar a cada variable (explicativas y dependiente) la media aritmética calculada con los valores disponibles en esa variable, recordando que en caso de haber observaciones perdidas se mantienen como tales;
- eliminar del análisis a aquellas observaciones en las que se perdió el valor de la variable dependiente;
- sustituir en las observaciones restantes, los datos perdidos por ceros;
- llevar a cabo un análisis de regresión bayesiano estándar con la matriz del diseño y el vector de valores de la variable respuesta, que resultan de las eliminaciones y sustituciones anteriores, considerando siempre un modelo de regresión con ordenada al origen.

(5) Hay que notar que β_{0i} asume el papel de ordenada al origen y el número de parámetros es entonces igual al número de variables explicativas (no constantes) más uno (por β_{0i}) y los grados de libertad son en el caso en que se contemple originalmente el modelo con ordenada al origen $n-p$ y en caso contrario $n-p-1$.

Es conveniente resaltar el hecho de que el tratamiento bayesiano aplicado a las observaciones faltantes de la variable explicativa K , ($K=1, 2, \dots, p$) produce un resultado equivalente a la sustitución de las variables perdidas por la media de la variable explicativa K , siguiendo a continuación con un análisis de regresión bayesiano estándar. Este hecho es de gran importancia ya que entonces no solo se dispone de una distribución posterior (obtenida analíticamente) de β y w sino de un resultado intuitivamente aceptable de fácil aplicación y respaldado teóricamente con la metodología bayesiana.

CAPITULO IV

DISCUSION Y EJEMPLOS

En el capítulo anterior se ha presentado una solución bayesiana al problema de regresión con observaciones incompletas. Aunque esta solución inicialmente es secuencial, se demuestra al final del mismo capítulo, que esta forma de proceder (incluyendo una a una las observaciones sean éstas completas o incompletas) es equivalente al siguiente procedimiento (que por énfasis se presenta una vez más):

- restar a cada observación de la variable (explicativas y dependiente) la media aritmética calculada con los valores disponibles de esa variable, recordando que en caso de haber observaciones perdidas se mantienen como tales;
- eliminar del análisis a aquellas observaciones en las que se perdió el valor de la variable dependiente;
- sustituir en las observaciones restantes, los datos faltantes por ceros;
- llevar a cabo un análisis de regresión bayesiano estándar con la distribución inicial conjunta de β y w en la familia conjugada de distribuciones Normal Gamma. Utilizar, para ello, la matriz del diseño y el vector de valores de la va-

stable respuesta resultantes de las eliminaciones y sustituciones anteriores, considerando siempre un modelo de regresión con ordenada al origen (representada por β_0 en la notación del capítulo tres).

La solución bayesiana dada, al quedar expresada en esta forma, reúne las siguientes ventajas:

- 1) Es posible tratar observación a observación o tratar todo el conjunto de observaciones en bloque.
- 2) No supone distribución alguna sobre las variables explicativas, sólo supone que la matriz del diseño está fija y que las variables explicativas tienen media cero (aunque se supone una a priori específica para $p(\gamma|\beta, w)$).
- 3) La solución es inmediata. No es necesario recurrir a iteraciones del procedimiento para obtener la solución.
- 4) No requiere de programas especiales de computadora. El tratamiento se hace utilizando programas para realizar análisis de regresión y en ocasiones basta con usar simples calculadoras.
- 5) Si se utiliza esta solución con un conjunto de datos completos el resultado es el mismo que se hubiera obtenido empleando el procedimiento para observaciones completas (sección 3.2).

6] La distribución posterior resultante es miembro de la familia conjugada de distribuciones.

7] Permite la formulación inmediata de una solución "clásica".

... La solución clásica que se obtiene es una mezcla de los procedimientos discutidos en el capítulo uno, secciones 1.3.1, 1.3.2 y 1.3.3. Del procedimiento descrito en 1.3.2., porque calcula las medias de las variables involucradas a partir de las observaciones disponibles; del procedimiento descrito en 1.3.1., porque se eliminan las observaciones en que falta el valor de la variable dependiente; del procedimiento descrito en 1.3.3., porque se sustituye los valores faltantes por cero después de haber restado, a cada variable, la media correspondiente.

8] El tratamiento resultante en el caso en que el valor de la variable dependiente se pierde muestra el camino correcto a seguir, en el caso de observaciones perdidas en diseño de experimentos. Esto es, ignorar todas las observaciones incompletas; esto puede desbalancear o complicar el diseño original, quedando como principal problema la determinación y manejo [numérico en computadora] de la matriz del diseño [Kendall & Stuart (1965)].

Es importante comentar en este punto que no todos los casos de modelos desbalanceados en "diseño de experimentos" son producidos por observaciones faltantes, hay algunos experimentos que por restricciones en el diseño son desbalanceados. Healy & Westmacott (1956) proponen que en estos últimos se "inventen" observaciones perdidas (y por lo tanto renglones en la matriz del diseño) de tal manera, que una vez solucionado el problema de las observaciones "faltantes ficticias" se contará con un diseño balanceado. Con esta solución bayesiana, se demuestra que como era lógico esperar, "inventar" observaciones faltantes no cambia en forma alguna el problema de desbalanceo existente.

Por otro lado en el caso de observaciones aberrantes (outliers) desde un enfoque "clásico", cuando han sido identificadas y se ha realizado un reporte adecuado de la presencia de estas observaciones, se plantea el problema de proseguir el análisis estadístico. Es claro que por ser una observación aberrante sólo pueden pasar dos cosas: o la población es en realidad una mezcla de poblaciones y esas observaciones aberrantes son miembros de algunas de esas poblaciones y por tanto no pueden ni debenser substituidas para hacerlas pasar como observaciones sólo de una población, o ha habido alguna causa fortuita que ha distorsionado al verdadero valor ocasionando su pérdida y por lo

tanto en este caso debe tratarse como observación faltante y entonces eliminarla del análisis.

- 2) Con esta solución se puede explicar adecuadamente la relación entre las variables explicativas y la variable dependiente utilizando para ello la función de distribución de probabilidad marginal conjunta posterior de β y la de w . Además, pueden predecirse los valores de la variable dependiente dado un determinado vector de valores de las variables explicativas mediante la utilización de la función de distribución predictiva (apéndice 4). Es importante recordar que el vector de valores de las variables explicativas debe estar transformado (los valores faltantes sustituidos por ceros y a las no faltantes se les debe restar la media correspondiente) y de esta manera la predicción se hace sobre B ($y - a_0$, utilizando la notación del capítulo anterior). Hay que observar también que la distribución predictiva queda centrada en un valor calculado exclusivamente con los valores no faltantes del vector de variables explicativas.

Con esta solución al estar en posibilidad de explicar la relación entre las variables explicativas y la variable dependiente, se está en posibilidad de satisfacer los objetivos de cualquier problema de regresión.

10] Permite la utilización, en problemas con datos incompletos, de técnicas y procedimientos ligados al análisis de regresión, como por ejemplo en la sección de variables (ver sección 4.2 ejemplo 4] o la utilización del coeficiente de determinación.

Contar con un coeficiente de determinación en el caso de observaciones incompletas es de gran utilidad, ya que permite valorar fácil y rápidamente el éxito que ha tenido el modelo en la explicación de la variación observada en los datos. Es por esto que se decidió estudiar (en la siguiente sección] el comportamiento del coeficiente de correlación en problemas de regresión con datos faltantes.

A continuación se presentan en la sección 4.1 los resultados de un estudio somero sobre la forma de la distribución de probabilidad del coeficiente de determinación para diferentes números de pérdidas y para diferentes conjuntos de variables con pérdidas. En la sección 4.2 se presentan ejemplos de aplicación de la solución al problema de regresión con datos faltantes, con conjuntos de diferentes estructuras y números de pérdidas, se proporcionan las distribuciones marginales posteriores de β (el vector de parámetros] y de w (la precisión de la variable dependiente), intervalos de mayor densidad sobre las componentes de β y el coeficiente de determinación. Esto se hace con algunos conjuntos de datos simulados y con los conjuntos de datos publicados en Dagenais y Rubin (1976a).

En los ejemplos basados en los datos de Dagenais y Rubin se comparan y discuten los resultados obtenidos por estos autores con los obtenidos aquí con este procedimiento.

Se utilizó para las simulaciones una computadora Hewlett Packard modelo 9845 B.

Es conveniente en este punto recordar que la solución presentada es consecuencia de la metodología bayesiana. Este hecho es de gran importancia ya que entonces se dispone de un resultado respaldado teóricamente y que resulta además intuitivamente aceptable y de fácil aplicación.

4.1 Comportamiento del Coeficiente de Determinación. Considere el conjunto de datos de la tabla 1. La primera columna es la variable respuesta o dependiente y las siguientes tres son las variables explicativas.

El modelo empleado fue:

$$Y = 2 X_1 + 14 X_2 + 0.5 X_3 + c$$

donde c se simuló Normal, mediante el método de Box & Muller [1958], con media cero y precisión uno. Para las variables explicativas se escogieron dos valores para cada una $X_1: 1.5$ y 2.5 , $X_2: 10.5$ y 11.5 y $X_3: 1.5$ y 2.5 . Estos valores se asignaron

TABLA 1. DATOS UTILIZADOS EN PARTE 1.

Y	X ₁	X ₂	X ₃
152.642	2.5	10.5	2.5
151.502	2.5	10.5	1.5
166.033	2.5	11.5	1.5
162.336	1.5	11.5	1.5
164.082	1.5	11.5	2.5
151.534	2.5	10.5	2.5
164.749	1.5	11.5	1.5
165.285	2.5	11.5	1.5
149.907	1.5	10.5	2.5
150.163	1.5	10.5	1.5
150.090	1.5	10.5	2.5
164.320	1.5	11.5	2.5
153.785	2.5	10.5	2.5
163.405	1.5	11.5	2.5
143.933	1.5	10.5	1.5
150.013	2.5	10.5	2.5
164.732	2.5	11.5	1.5
167.264	2.5	11.5	2.5
164.083	1.5	11.5	1.5
150.142	1.5	10.5	2.5
149.770	1.5	10.5	1.5
164.530	1.5	11.5	2.5
153.252	1.5	10.5	1.5
165.657	2.5	11.5	2.5
163.822	1.5	11.5	1.5
150.542	1.5	10.5	2.5
162.066	1.5	11.5	1.5
170.707	2.5	11.5	2.5
163.543	1.5	11.5	2.5
152.513	2.5	10.5	1.5

aleatoriamente en cada observación con probabilidad 0.5.

El coeficiente de determinación se define (Draper & Smith (1968) pg. 57-62) como:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} .$$

La manera de calcularlo con un conjunto de observaciones in completas es la siguiente: aplicar el procedimiento de eliminaciones y sustituciones indicado en el capítulo III, y con la matriz del diseño y el vector de valores de la variable respuesta resultantes obtener directamente $\sum (y_i - \bar{y})^2$ y $\sum (y_i - \hat{y}_i)^2$.

En el caso de utilizar una distribución inicial de referencia se simplifica la obtención de la suma de cuadrados del residuo ($\sum (y_i - \hat{y}_i)^2$) ya que (ver sección 3.2)

$$\hat{y} = X(X'X)^{-1}X' y ,$$

y

$$\hat{y}'(y - \hat{y}) = 0 ,$$

entonces

$$\gamma_1 = 0.5 (y - \hat{y})'(y - \hat{y}) = 0.5 \sum (y_i - \hat{y}_i)^2$$

de donde

$$R^2 = 1 - 2\gamma_1 / \sum (y_i - \bar{y})^2 .$$

Una propiedad conocida del coeficiente de determinación es que $0 < R^2 < 1$.

Estadísticas básicas para este conjunto de datos se presentan en la tabla 2. Dentro de estas es importante destacar las correlaciones de las variables explicativas con Y:

$$r_{x_1, y} = 0.121$$

$$r_{x_2, y} = 0.957$$

$$r_{x_3, y} = 0.007$$

El coeficiente de determinación que se obtiene al ajustar un modelo de regresión con este conjunto de datos es $R^2 = 0.95$ y se utiliza a lo largo de esta parte uno, como punto de comparación para los valores obtenidos en las simulaciones.

A continuación se procedió a simular pérdidas dentro de este conjunto de datos. Realizándose en dos formas: variable a variable y en grupos de variables.

La simulación de las pérdidas se hizo aleatoriamente. Se determinaron cuantas (denotadas por c ; $1 < c < 3$) y cuales eran las columnas en que se simularían las pérdidas, a continuación se procedía a escoger aleatoriamente una columna y un renglón. Los renglones se escogieron con probabilidad $1/30$ y las columnas con probabilidad $1/c$. En el caso en que la posición obtenida mediante la

TABLA 2. CORRELACIONES, MEDIAS Y VARIANZAS DE LOS DATOS DE LA TABLA 1.

MATRIZ DE CORRELACIÓN			
Y	X ₁	X ₂	X ₃
1	0.121	0.957	0.007
	1	-0.546	0.082
		1	-0.071
			1
MEDIAS			
Y	X ₁	X ₂	X ₃
158.216	1.900	11.033	2.033
VARIANZAS			
Y	X ₁	X ₂	X ₃
55.812	0.248	0.258	0.258

elección de columnas y fila estuviera ya identificada como perdida o faltante, se procedió a escoger de nuevo fila y columna. Este procedimiento se repitió hasta conseguir el número deseado de observaciones faltantes en el conjunto. Ato seguido se calculó el coeficiente de determinación.

Se simularon 5, 10 y 15 pérdidas en cada variable por separado, empezando por Y [la dependiente] y siguiendo con X_1, X_2 y X_3 . Además se simularon 5, 10, 15, 30 y 60 pérdidas entre los conjuntos de variables $\{X_1, X_2\}$ y $\{X_1, X_2, X_3\}$, haciendo 100 repeticiones cada vez (a excepción de cuando hubo 60 pérdidas entre X_1 y X_2).

Los resultados de estas simulaciones se expresan mediante 100 veces el coeficiente de determinación y se comentan a continuación con la ayuda de las figuras 1 a 21. Estas figuras muestran los histogramas obtenidos con la información resultante de estas simulaciones (los intervalos de clase, sus límites y las frecuencias correspondientes se muestran en el apéndice 5).

En las figuras 1, 2 y 3 puede apreciarse el comportamiento del coeficiente de determinación (CD) al haber respectivamente 5, 10 y 15 pérdidas en la variable dependiente Y .

Como era de esperar, dependiendo de los valores que se pierden pueden cambiar el valor del CD [en este caso aumentando hasta 0.99 o disminuyendo hasta 0.91]. Esto implica

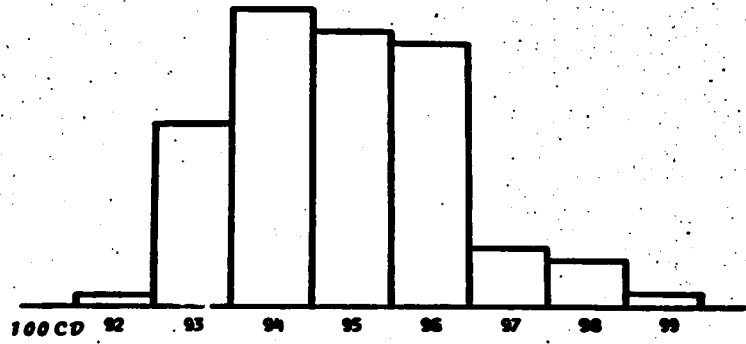
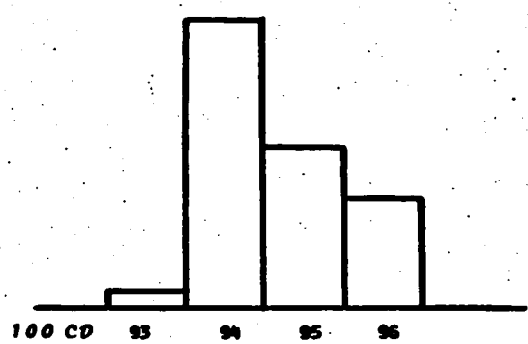
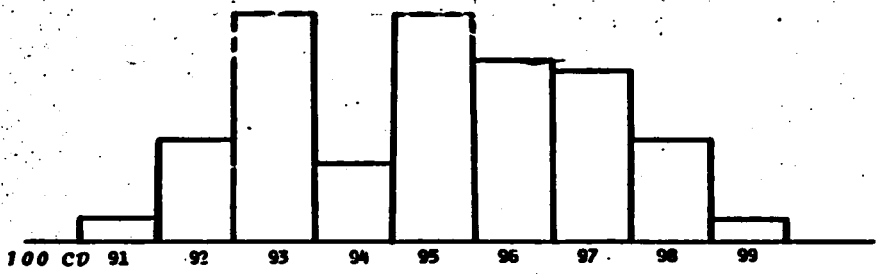


FIGURA 2. 15 PERDIDAS EN Y



que siempre es posible mejorar un ajuste si se extravían las observaciones "adecuadas" de la variable dependiente. Como este efecto puede resultar engañoso y de difícil interpretación al presentarse confundido con los efectos causados por la existencia de pérdidas entre las variables explicativas, se decidió no simular en este trabajo pérdidas sobre la variable dependiente. Sin embargo, en el ejemplo 3 de la segunda parte, se presenta y analiza un conjunto de datos con pérdidas entre los valores de la variable dependiente.

En las figuras 4 a 12 puede apreciarse el comportamiento del coeficiente de determinación al haber faltantes entre las diferentes variables explicativas. Puede decirse después de observar estas figuras, que la reducción en la cantidad de variación explicada por el modelo es directamente proporcional a la correlación de la variable explicativa (que sufre las pérdidas) con la variable dependiente y desde luego al número de pérdidas sufridas.

En las figuras 13 a 16 se observa el comportamiento del coeficiente de determinación al haber pérdidas en dos variables, en este caso X_1 y X_2 . Puede concluirse que el comportamiento es similar al observado al ocurrir pérdidas sólo en X_1 ó en X_2 . Se trató también un caso en que hubiera 60 pérdidas en X_1 y X_2 , esto es, que las dos variables perdieran todas sus obser

FIGURA 4. 5 PERDIDAS EN X₁

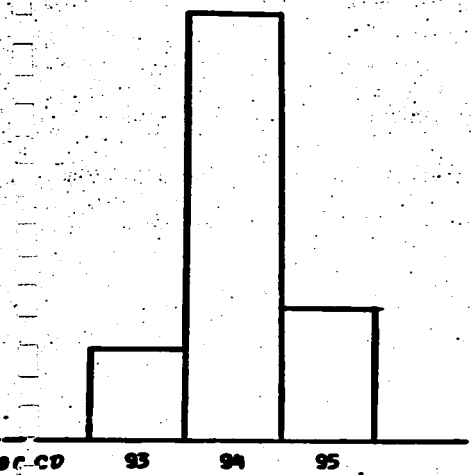


FIGURA 5. 10 PERDIDAS EN X₁

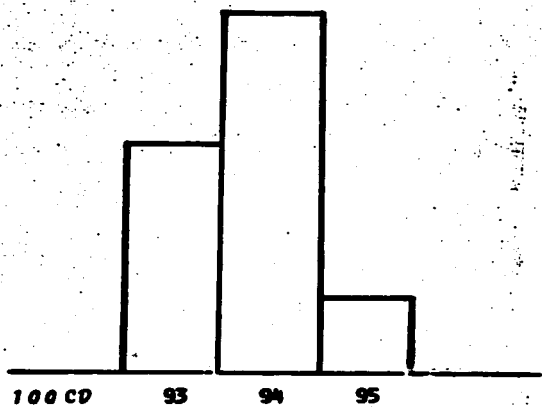


FIGURA 6. 15 PERDIDAS EN X₁

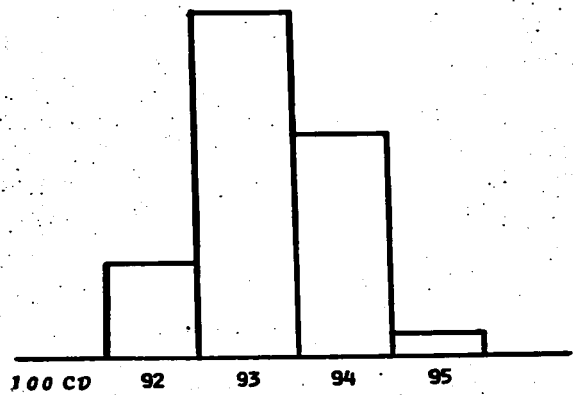


FIGURA 1. 5 PERDIDAS X.

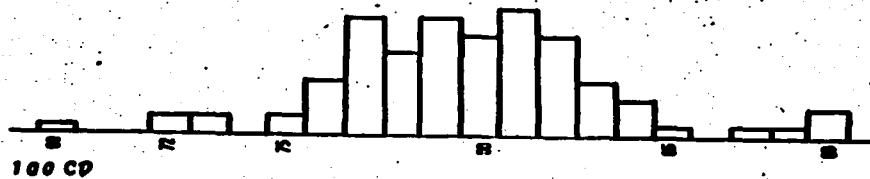


FIGURA 2. 20 PERDIDAS X.

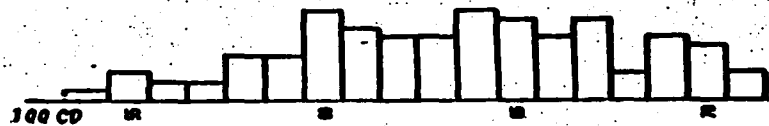


FIGURA 3. 25 PERDIDAS X.

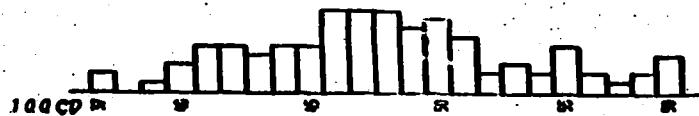


FIGURA 10. 5 PERDIDAS EN X.

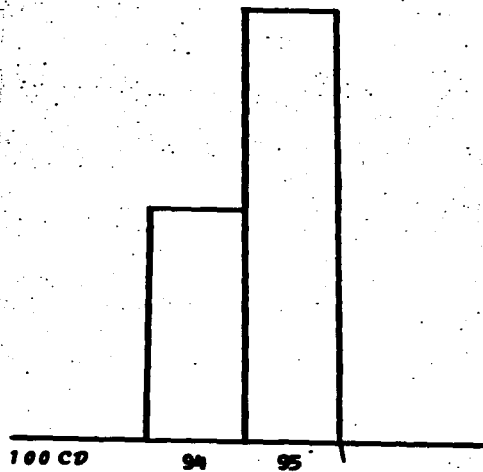


FIGURA 11. 10 PERDIDAS EN X.

288.

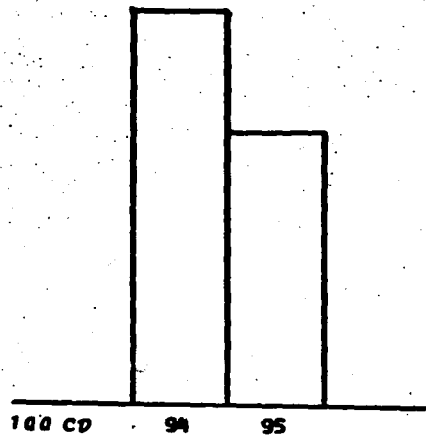


FIGURA 12. 15 PERDIDAS EN X.

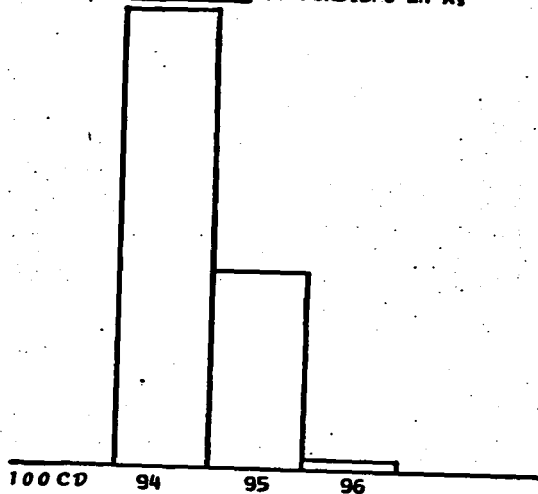


FIGURA 17. 5 PERDIDAS ENTRE X.v X₀.

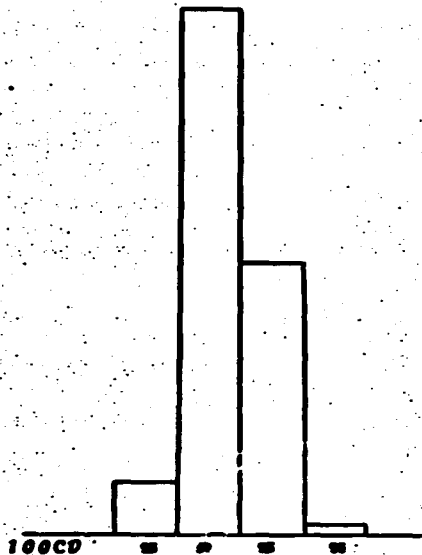
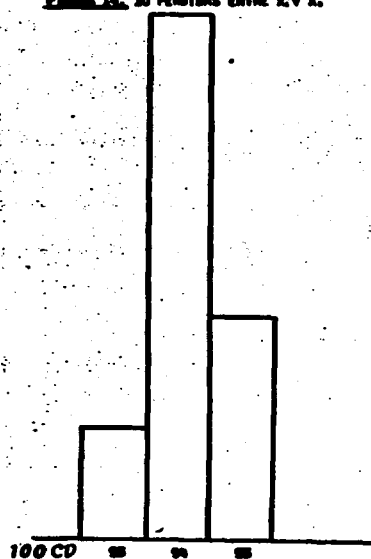


FIGURA 18. 30 PERDIDAS ENTRE X.v X₀.



89.

FIGURA 19. 1% PERDIDAS ENTRE X.v X₀.

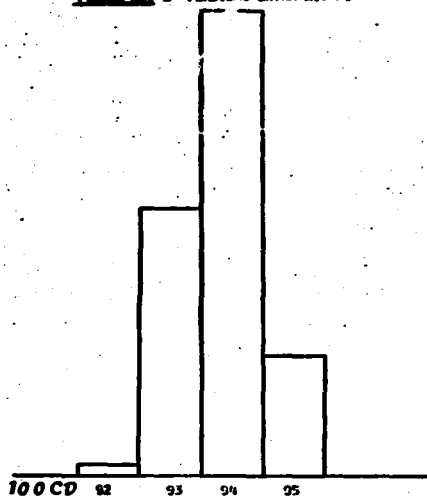
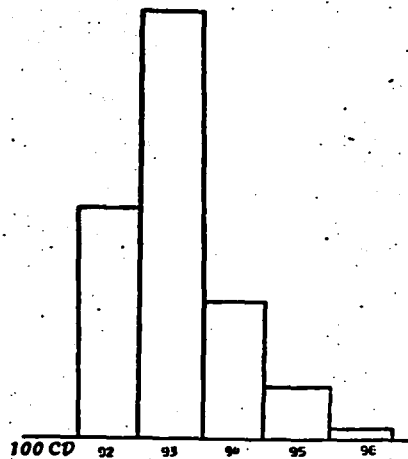


FIGURA 20. 30 PERDIDAS ENTRE X.v X₀.



vaciones y resulto como es de esperarse que el valor del coeficiente de determinación obtenido fue 0.916 que es el cuadrado del coeficiente de correlación de X_2 con Y . Esto es intuitivamente lógico ya que debe explicarse la misma cantidad de variación observada tanto si sólo se tiene a la variable X_2 como si se consideran las tres variables explicativas pero se pierden todos los valores de X_1 y X_3 .

Por último, en las figuras 17 a 21 se muestra el comportamiento del coeficiente de determinación al producirse pérdidas en X_1 , X_2 y X_3 . Se observa en estas figuras que X_2 , por ser la variable más correlacionada con Y , determina en este caso el comportamiento del coeficiente. Si X_1 es la variable que más pérdidas sufre entonces el coeficiente de determinación se afecta más, mientras que si X_3 sufre pocas pérdidas el coeficiente se afecta poco, resultando en consecuencia que el rango de variación es mayor en este caso que en los anteriores.

Resumiendo, en base a la discusión anterior puede conjeturarse que al aplicar la técnica desarrollada en este trabajo a un conjunto de datos incompletos el coeficiente de determinación se comporta de la siguiente manera (al compararlo con el obtenido con los datos completos): en el caso de pérdidas en

FIGURA 17. 5 PERDIDAS X_1, X_2, \dots, X_5 .

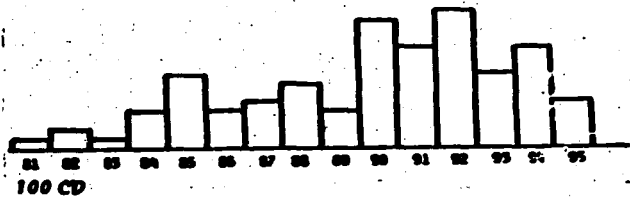


FIGURA 18. 20 PERDIDAS X_1, X_2, \dots, X_{20} .

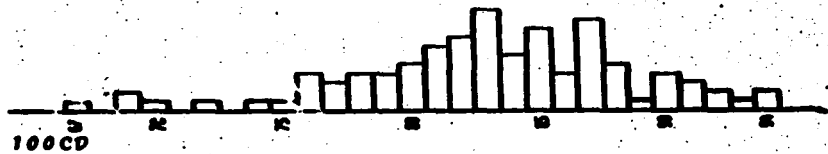


FIGURA 19. 25 PERDIDAS X_1, X_2, \dots, X_{25} .

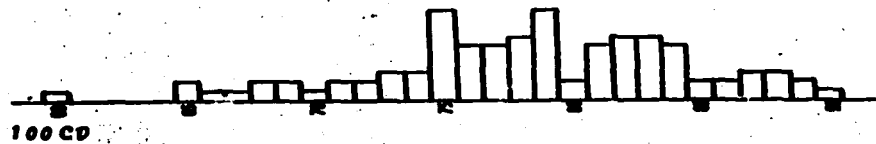


FIGURA 20. 30 PERDIDAS X_1, X_2, \dots, X_{30} .

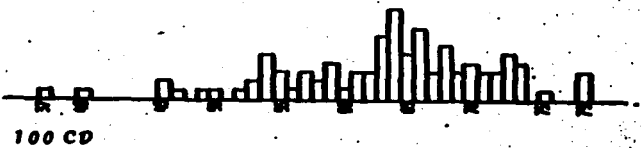
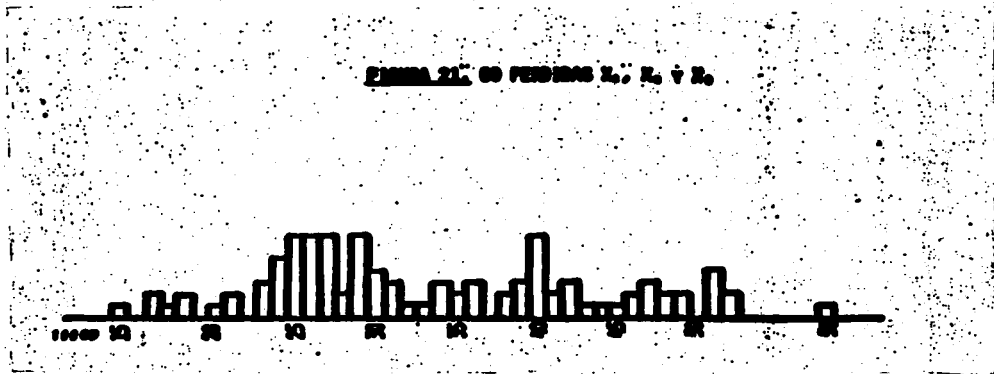


FIGURA 21. DO FENÔMENO X, Y, Z.



tre las variables explicativas se tiene que lo más probable es que baje la cantidad de variación explicada por el modelo. La magnitud del decremento depende tanto del número de pérdidas como de la importancia (medida por la correlación con la variable dependiente) de la variable explicativa. En muy pocos casos se observó que subiera el valor del coeficiente de determinación al haber pérdidas y nunca subió en más de una centésima, lo que sugiere que el valor del coeficiente de determinación en el caso de datos sin pérdidas se encuentra en el extremo derecho de las distribuciones bajo estudio.

En el caso de pérdidas sobre la variable dependiente, el modelo puede tanto mejorar como empeorar el ajuste que se obtendría si no hubiera pérdidas, dependiendo de qué valores de esta variable son los que se pierden.

4.2 EJEMPLOS

En esta parte, y a menos que se especifique lo contrario, se han utilizado las distribuciones iniciales de referencia mencionadas en el capítulo anterior. En los ejemplos 1 y 2, se presentan conjuntos de datos generados utilizando modelos de regresión en los que el error se simula Normal con media cero y precisión 1 mediante el método de Box & Muller [1958].

Ejemplo 1.- A lo largo de este ejemplo se ilustra el com-

portamiento de la solución bayesiana en el caso en que haya observaciones con todos los valores de las variables independientes perdidas, se consideran diferentes situaciones en las que el número de observaciones con todos los valores de las variables explicativas perdidas se incrementa a lo largo del ejemplo.

Por facilidad en la discusión, el modelo de regresión empleado para la simulación de los valores de la variable dependiente, mismos que se presentan en la tabla 3 junto con los valores de la variable explicativa, fue:

$$Y = 10 + 2X + \epsilon.$$

Los valores de X se escogieron aleatoriamente con repetición entre los enteros positivos menores que 100.

En la tabla 4 aparecen las medias y las varianzas de Y y X , así como la correlación entre la Y y la X . Las varianzas se calcularon dividiendo entre $(N-1)$, donde N es el número de observaciones de Y .

Con este conjunto de datos la distribución marginal posterior de β es una T de Student con 28 grados de libertad, localizada en 1.985 y precisión 10336.242. La marginal posterior de w es una distribución Gamma con parámetros $\alpha=14.500$ y $\gamma=41.542$.

Los intervalos de mayor densidad, de probabilidades 0.5, 0.8, 0.9 y 0.95 para β son, respectivamente:

TABLA 3. DATOS EJEMPLO 1. SIN PERDIDAS

Y	X	Y	X
124.524	22	137.888	65
93.427	42	32.659	12
155.958	73	125.139	92
174.261	83	84.008	37
77.964	24	126.017	93
23.409	7	23.695	7
174.674	32	122.385	56
135.210	63	175.184	81
59.782	25	71.532	31
12.088	1	87.747	39
127.965	52	30.417	10
66.264	28	189.991	91
31.660	10	60.582	23
121.280	21	21.416	41
201.858	29	56.438	23

TABLA 4. EJEMPLO 1. CORRELACIONES, MEDIAS Y VARIANZAS DE LOS DATOS.

MATRIZ DE CORRELACIONES

Y	X
J	0,9996
	J

MEDIAS

Y	X
109,183	49,667

VARIANZAS

Y	X
4027,783	1021,126

[1.978, 1.992], [1.972, 1.998], [1.968, 2.002] y [1.965, 2.005].

El valor del coeficiente de determinación es de 0.999 .

En la tabla 5, se presentan los datos de la tabla 3 pero, en este caso, diez valores de X se han perdido (por facilidad, los diez últimos).

Se utiliza a 9999.9999 como identificador de valores perdidos.

Las medias y varianzas muestrales de Y y X, calculadas con los valores disponibles de estas variables, son:

$$\bar{Y} = 109.181, \bar{X} = 54.400, S_Y^2 = 4027.783 \text{ y } S_X^2 = 1109.516 .$$

El modelo considerado para el ajuste es en notación del capítulo anterior,

$$B_i = Y_i - a_0 = \beta_{01} + \beta_{11}(x_i - a_1) + \epsilon_i$$

donde $i=1, 2, \dots, 30$; $a_0 = \bar{Y}$ y $a_1 = \bar{X}$.

Con este conjunto de datos la distribución marginal posterior de $\underline{\theta} = (\beta_{11}, \beta_{01})'$ es una T de Student con 28 grados de libertad, vector de localización $(1.986, 0.000)'$ y matriz de precisión

$$\begin{pmatrix} 17.558 & 0 \\ 0 & 0.020 \end{pmatrix}$$

TABLA 5. DATOS EJEMPLO 1. HAY 10 PERDIDAS ENTRE LOS VALORES DE LA VARIABLE EXPLICATIVA.

Y	X	Y	X
124.524	22	137.888	43
93.427	42	32.659	12
155.958	73	195.139	92
174.261	83	84.008	37
77.964	34	126.017	93
23.409	7	23.695	9999.9999
174.674	82	122.385	9999.9999
135.210	63	175.184	9999.9999
59.782	25	71.532	9999.9599
12.088	1	87.747	9999.9999
127.965	59	30.417	9999.9999
66.264	28	189.991	9999.9999
31.660	10	60.582	9999.9999
191.280	91	91.416	9999.9999
201.858	99	56.438	9999.9999

NOTA. - 9999.9999 representa un valor faltante.

La distribución marginal posterior de w es Gamma con parámetros $\alpha=14$ y $\gamma=16808.724$.

Los intervalos de mayor densidad para β_1 de probabilidades 0.5, 0.8, 0.9 y 0.95 son respectivamente: [1.824, 2.150], [1.673, 2.300], [1.581, 2.392] y [1.498, 2.475].

Estos intervalos de probabilidad permiten obtener fácilmente una idea de la distribución marginal posterior de β_1 la cual es de utilidad, por ejemplo en el caso de estar interesados en hacer estimación o contrastes de hipótesis (Lindley 1965, pg. 6 y 23-24 y De Groot 1970, pg. 226-237), por ejemplo: si $A=[-5, 5]$ entonces podría interesar contrastar $H_1: \beta_1 \in A$ contra $H_2: \beta_1 \notin A$, en este caso se tendría una muy fuerte evidencia contra H_1 y en favor de H_2 (la probabilidad de H_2 sea cierta es prácticamente uno), en el caso en que $B = [1.5, 2.5]$ e interesara probar $H_1: \beta_1 \in B$ contra $H_2: \beta_1 \notin B$ se tendría una fuerte evidencia en favor de H_1 .

En este ejemplo el valor del coeficiente de determinación es 0.712, lo que se interpreta como que el 71.2% de la variación observada de Y , ha sido explicada mediante el modelo. El resulta completamente lógico ya que se perdió una tercera parte de las observaciones de X .

En la tabla 6 se presentan los mismos datos de la tabla 3, pero

TABLA 6. DATOS EJEMPLO 1. HAY 27 PERDIDAS ENTRE LOS VALORES DE LA VARIABLE EXPLICATIVA.

y	x	y	x
194.524	92	137.888	9999.9999
93.427	42	32.659	9999.9999
155.958	73	195.732	9999.9999
174.261	9999.9999	84.008	9999.9999
77.964	9999.9999	196.017	9999.9999
23.409	9999.9999	23.695	9999.9999
174.674	9999.9999	122.385	9999.9999
135.210	9999.9999	175.184	9999.9999
59.782	9999.9999	71.532	9999.9999
12.088	9999.9999	87.747	9999.9999
127.965	9999.9999	30.417	9999.9999
66.264	9999.9999	189.991	9999.9999
31.660	9999.9999	60.582	9999.9999
191.280	9999.9999	91.414	9999.9999
201.858	9999.9999	56.438	9999.9999

Nota. - 9999.9999 representa un valor faltante.

en este caso se han perdido todos los valores de X excepto las correspondientes a las tres primeras observaciones.

La media y la varianza muestrales de Y y X son: $\bar{Y}=109.181$, $S_y^2=4027.783$, $\bar{X}=69.000$ y $S_x^2=637.000$.

El modelo que se uso para el ajuste es:

$$B_{2i} - Y_{2i} - a_0 = \beta_{01} + \beta_{11}(X_{2i} - a_1) + \epsilon_{1i}$$

donde $i=1, 2, \dots, 30$; $a_0 = \bar{Y}$ y $a_1 = \bar{X}$.

En este caso, la distribución posterior marginal de $\underline{\beta} = (\beta_{01}, \beta_{11})'$ es T de Student con 28 grados de libertad, vector de localización $(2.022, 0.000)'$ y matriz de precisión

$$\begin{pmatrix} 0.320 & 0 \\ 0 & 0.008 \end{pmatrix}$$

La distribución marginal posterior de w es una Gamma con $\alpha=14$ y $\gamma=55792.843$. El valor del coeficiente de determinación es 0.0445.

Este ejemplo es particularmente interesante si se considera el rechazo de las observaciones incompletas, ya que entonces só lo quedarían tres observaciones, la distribución de β_1 quedaría centrada en 2.022, la variación observada de la variable dependiente, $\sum(Y_{1i} - \bar{Y})^2$, se reduciría en forma arbitraria y artificial

de 116805.700 a 5206.022 y el coeficiente de determinación tendría un valor de 0.99997.

Este valor del coeficiente de determinación no debe interpretarse directamente puesto que podría pensarse que se ha explicado prácticamente toda la variación observada, lo cual es un error. Debe interpretarse que con el modelo se ha explicado el 99.9997 % de 5206.022 (que es sólo el 4.5% de la variación total observada 116805.700), de donde se concluye que se ha explicado sólo el 4.46% de la variación total observada, valor que corresponde exactamente a la cantidad explicada mediante la solución bayesiana.

Es importante señalar en este ejemplo que el ignorar las observaciones con pérdidas no sólo afecta el coeficiente de determinación si no también a las precisiones involucradas. En este caso se tiene, al considerar sólo a las tres observaciones sin pérdidas, que la precisión posterior marginal del parámetro β_1 es 175 442.341, siendo en realidad de sólo 0.320.

En la tabla 7, se presentan los datos de la tabla 3, pero ahora se han perdido todos los valores de X.

La media y la varianza muestral de Y son: $\bar{Y} = 109.181$ y $S_y^2 = 4027.783$.

TABLA 7. DATOS EJEMPLO 1. TODOS LOS VALORES DE LA VARIABLE EXPLICATIVA SE HAN PERDIDO.

Y	X	Y	X
124.524	9999.9999	137.887	9999.9999
93.427	9999.9999	32.659	9999.9999
155.958	9999.9999	195.139	9999.9999
174.261	9999.9999	84.008	9999.9999
77.964	9999.9999	196.017	9999.9999
23.409	9999.9999	23.695	9999.9999
174.674	9999.9999	122.385	9999.9999
135.210	9999.9999	175.184	9999.9999
59.782	9999.9999	71.532	9999.9999
12.088	9999.9999	87.747	9999.9999
127.965	9999.9999	30.417	9999.9999
66.264	9999.9999	189.991	9999.9999
31.660	9999.9999	60.582	9999.9999
191.280	9999.9999	91.416	9999.9999
201.858	9999.9999	50.438	9999.9999

Nota.- 9999.9999 representa un valor faltante.

El modelo que se considera para el ajuste es en notación del capítulo anterior:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

donde $i=1, 2, \dots, 30$; $y \bar{Y}$ y se considera $a_1 = 0$.

Suponiendo que no se tiene una distribución inicial informativa se podrá proceder como se sugiere en el inciso b) de la sección 3.4 y considerar que la distribución inicial de $\beta = (\beta_0, \beta_1)'$ y w es una distribución Normal Gamma $(2, \underline{\mu}_0, t_0, \gamma_0, \alpha_0)$ donde

$$\underline{\mu}_0 = \underline{0}, t_0 = [1/300] I, \gamma_0 = 1/300 \text{ y } \alpha_0 = -1.$$

En este caso la distribución posterior marginal de β es una T de Student con 28 grados de libertad, vector de localización $\underline{0}$ y matriz de precisión

$$\begin{pmatrix} 7.99 \times 10^{-7} & 0 \\ 0 & 0.007 \end{pmatrix}$$

La distribución marginal posterior de w es una Gamma con parámetros $\alpha=14$ y $\gamma = 1/300 + 0.5 \sum (Y_i - \bar{Y})^2 = 58402.858$.

El valor del coeficiente de determinación es cero.

El coeficiente de determinación cero indica claramente que el modelo no explica nada de la variación observada de las Y 's,

lo cual es lógico puesto que al no observar ningún valor de X , no se aprendió nada sobre β por lo que no se puede hablar sobre la relación existente entre las X 's y la Y (si es que esta existe) y por otra parte las predicciones sobre Y se centran en V , lo mismo que si no se usara el modelo.

Este ejemplo es interesante ya que permite en forma sencilla la comprobación, de que la solución bayesiana presentada en el capítulo tres, es intuitivamente correcta en esta situación extrema.

EJEMPLO 2.— A partir de un conjunto específico de datos, se muestra en este ejemplo el comportamiento de la solución bayesiana obtenida, al considerar problemas de regresión tanto con grandes porcentajes de observaciones faltantes como con estructuras complicadas de pérdidas. Encontrando en las situaciones estudiadas un comportamiento intuitivamente aceptable.

Se trataja con los datos que se muestran en la tabla 8 y que fueron generados de la siguiente forma. Los valores de la variable dependiente Y se generaron según el modelo:

$$Y = 2 X_1 + 5 X_2 + 3 X_3 + \epsilon$$

donde ϵ se simulo Normal con media cero y precisión 1.

Los valores de X_1 , X_2 y X_3 se escogieron aleatoriamente y por separado, con repetición entre los enteros positivos menores a 100. En la ta

TABLA 8. DATOS EJEMPLO 2. SIN PERDIDAS

Y	X ₁	X ₂	X ₃
615.524	92	42	73
356.427	83	34	7
553.958	82	63	25
379.261	1	59	28
771.964	10	91	99
465.409	65	12	92
560.674	37	93	7
609.210	56	81	31
400.782	39	10	91
320.088	23	41	23
409.965	43	9	93
491.264	15	89	72
572.660	56	45	78
632.280	5	82	71
110.858	6	18	5
686.888	90	43	98
617.659	98	78	11
613.139	66	57	65
572.008	17	83	41
325.017	6	11	86
360.695	3	44	45
765.385	48	90	73
268.184	43	19	28
772.532	57	73	98
530.747	43	83	10
363.417	7	20	83
446.991	24	73	12
718.582	88	50	96
703.416	40	87	63
249.438	55	11	28

bla 9 aparecen la matriz de correlaciones entre las variables Y , X_1 , X_2 y X_3 , así como las medias y la varianzas muestrales respectivas.

El modelo que se ajustó es:

$$B_i = Y_i - a_i = \beta_1(X_{1i} - \bar{X}_1) + \beta_2(X_{2i} - \bar{X}_2) + \beta_3(X_{3i} - \bar{X}_3) + e_i$$

donde $i=1, 2, \dots, 30$.

La distribución marginal posterior de $\underline{\beta} = (\beta_1, \beta_2, \beta_3)'$ es una T de Student con 26 grados de libertad, con parámetro de centralidad $(2.015, 5.004, 3.016)'$ y con matriz de precisión

$$\begin{pmatrix} 9670.657 & -375.584 & 461.955 \\ -375.584 & 9128.644 & -1776.922 \\ 461.955 & -1776.922 & 11927.134 \end{pmatrix} .$$

La distribución marginal posterior de w es una Gamma con $\alpha=13.5$ y $\gamma=37.284$. El valor del coeficiente de determinación es 0.9999.

Intervalos de mayor densidad para β_1, β_2 y β_3 , de probabilidades 0.5, 0.8, 0.9 y 0.95 son:

TABLA 9. DATOS EJEMPLO 2.- CORRELACIONES, MEDIAS Y VARIANZAS DE LOS DATOS DE LA TABLA 8.

MATRIZ DE CORRELACIONES

Y	X_1	X_2	X_3
1	0.341	0.751	0.453
	1	0.004	0.004
		1	-0.170
			1

MEDIAS

Y	X_1	X_2	X_3
514.747	43.267	53.033	54.400

VARIANZAS

Y	X_1	X_2	X_3
30548.795	920.961	869.344	1136.041

P_x	β_1	β_2	β_3
0.5	[2.008, 2.022]	[4.997, 5.011]	[3.010, 3.022]
0.8	[2.002, 2.028]	[4.990, 5.018]	[3.004, 3.028]
0.9	[1.998, 2.032]	[4.986, 5.022]	[3.000, 3.032]
0.95	[1.994, 2.036]	[4.982, 5.026]	[2.997, 3.035]

2.1.- Si se pierden diez valores de X_1 , que por facilidad se han escogido como los últimos, se tiene la tabla 10 de datos.

Con estos datos se tienen las siguientes medias y varianzas muestrales:

$$\begin{aligned}
 \bar{Y} &= 514.747 & , & & S_y^2 &= 30548.795 & , \\
 \bar{X}_1 &= 44.500 & , & & S_{x_1}^2 &= 1105.737 & , \\
 \bar{X}_2 &= 53.033 & , & & S_{x_2}^2 &= 869.344 & , \\
 \bar{X}_3 &= 54.400 & \text{ u } & & S_{x_3}^2 &= 1136.041 & ,
 \end{aligned}$$

El modelo que se ajusta es:

$$Y_i = \beta_0 + \beta_1(X_{1i} - a_1) + \beta_2(X_{2i} - a_2) + \beta_3(X_{3i} - a_3) + \epsilon_i$$

Donde $i=1, 2, \dots, 30$; $a_0 = \bar{Y}$, $a_1 = \bar{X}_1$, $a_2 = \bar{X}_2$ y $a_3 = \bar{X}_3$.

La distribución marginal posterior de $\underline{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)$ es una T de Student multivariada con 26 grados de libertad, vector de localización (2.028, 5.105, 3.178, 0.000)' y matriz de precisión

TABLA 10. EJEMPLO 2. HAY 10 PERDIDAS ENTRE LOS VALORES DE LA VARIABLE X_1 .

V	X_1	X_2	X_3
613.524	92	42	73
556.427	83	34	7
553.958	82	63	25
379.261	1	59	28
771.964	10	91	99
465.409	65	12	92
560.674	37	93	7
609.210	56	81	31
400.782	39	10	91
320.088	23	41	23
409.965	43	9	93
691.264	15	89	72
572.660	56	45	78
632.280	5	82	71
110.858	6	18	5
686.888	90	43	98
617.659	98	78	11
613.139	66	57	65
572.008	17	83	41
325.017	6	11	86
360.695	9999.9999	44	45
765.385	9999.9999	90	73
268.184	9999.9999	19	28
772.532	9999.9999	73	98
530.747	9999.9999	83	10
363.417	9999.9999	20	83
446.991	9999.9999	73	12
718.582	9999.9999	50	96
703.416	9999.9999	87	63
249.438	9999.9999	11	28

Nota. - 9999.9999 representa un valor faltante.

$$\begin{pmatrix} 24.097 & -2.180 & -1.286 & 0.000 \\ -2.180 & 28.917 & -5.629 & 0.000 \\ -1.286 & -5.629 & 37.788 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.034 \end{pmatrix}$$

La distribución marginal posterior de w es Gamma con $\alpha=13$ y $\gamma=11333.919$. El valor del coeficiente de determinación es 0.974.

Intervalos de mayor densidad para β_1 , β_2 y β_3 , de probabilidades 0.5, 0.8, 0.9 y 0.95 son:

P_{Σ}	β_1	β_2	β_3
0.5	[1.888, 2.167]	[4.978, 5.232]	[3.067, 3.290]
0.8	[1.760, 2.296]	[4.861, 5.350]	[2.964, 3.392]
0.9	[1.680, 2.375]	[4.788, 5.423]	[2.901, 3.456]
0.95	[1.609, 2.447]	[4.723, 5.488]	[2.844, 3.513]

2.2.- Ahora se considera que hay diez valores perdidos de X_2 (por facilidad, se escogen en este caso los últimos) se tiene el conjunto de datos de la tabla 11.

Se tiene con estos datos las siguientes medias y varianzas muestrales:

TABLA 11. DATOS EJEMPLO 2. HAY 10 PERDIDAS ENTRE LOS VALORES DE LA VARIABLE X_2 .

Y	X_1	X_2	X_3
613.524	92	42	73
356.427	83	34	7
553.958	82	63	25
379.261	1	59	28
771.964	10	91	99
465.409	65	12	92
560.674	37	93	7
609.210	56	81	31
300.782	39	10	91
320.088	23	41	23
409.965	43	9	93
691.264	15	89	72
572.660	56	45	78
632.280	5	82	71
110.858	6	18	5
686.832	90	43	98
617.659	98	78	11
613.139	66	57	65
572.008	17	83	41
325.017	6	11	86
360.695	3	9999.9999	45
765.385	48	9999.9999	73
268.184	43	9999.9999	28
772.532	57	9999.9999	98
530.747	43	9999.9999	10
363.417	7	9999.9999	83
446.991	24	9999.9999	12
718.582	83	9999.9999	96
703.416	40	9999.9999	63
249.438	55	9999.9999	28

Nota. - 9999.9999 representa un valor faltante.

$$\begin{array}{rcl}
 \bar{Y} = 514.747 & , & S_Y^2 = 30548.795 \\
 \bar{X}_1 = 43.267 & , & S_{X_1}^2 = 920.961 \\
 \bar{X}_2 = 52.050 & , & S_{X_2}^2 = 887.840 \\
 \bar{X}_3 = 54.400 & , & S_{X_3}^2 = 1136.041
 \end{array}$$

El modelo que se ajusta es:

$$B_i = Y_i - \bar{Y} = \beta_{01} + \beta_1(X_{i1} - a_1) + \beta_2(X_{i2} - a_2) + \beta_3(X_{i3} - a_3) + \epsilon_i,$$

donde $i=1, 2, \dots, 30$; $a_1 = \bar{X}_1$, $a_2 = \bar{X}_2$ y $a_3 = \bar{X}_3$.

La distribución marginal posterior de $\underline{\beta} = (\beta_1, \beta_2, \beta_3, \beta_{01})'$ es una T de Student multivariada con 26 grados de libertad, vector de localización $(2.176, 5.074, 3.126, 0.000)'$ y matriz de precisión:

$$\begin{pmatrix}
 3.383 & -0.241 & 0.162 & 0 \\
 -0.241 & 2.137 & -0.709 & 0 \\
 0.162 & -0.709 & 4.173 & 0 \\
 0 & 0 & 0 & 0.004
 \end{pmatrix}$$

La distribución marginal posterior de w es Gamma con parámetros $\alpha=13$ y $\gamma=102621.768$. El valor del coeficiente de determinación es 0.768.

Los intervalos de mayor densidad para β_1, β_2 y β_3 de probabilidades 0.5, 0.8, 0.9 y 0.95 son:

P_x	β_1	β_2	β_3
0.5	[1.804, 2.548]	[4.606, 5.542]	[2.792, 3.461]
0.8	[1.461, 2.893]	[4.174, 5.973]	[2.483, 3.770]
0.9	[1.249, 3.104]	[3.907, 6.241]	[2.291, 3.961]
0.95	[1.858, 3.294]	[3.667, 6.480]	[2.120, 4.133]

Si ahora se considera que hay diez valores perdidos de X_i (por facilidad de escogen los diez últimos) se tiene el conjunto de datos de la tabla 12. Con estos datos se tienen las siguientes medias y varianzas muestrales:

$$\begin{aligned} \bar{Y} &= 514.747, & S_y^2 &= 30548.795 \\ \bar{X}_1 &= 43.267, & S_{x_1}^2 &= 920.961 \\ \bar{X}_2 &= 53.035, & S_{x_2}^2 &= 869.344 \\ \bar{X}_3 &= 54.800, & S_{x_3}^2 &= 1201.116. \end{aligned}$$

El modelo que se ajusta es:

$$B_i = Y_i - V = \beta_{01} + \beta_1(X_{1i} - a_1) + \beta_2(X_{2i} - a_2) + \beta_3(X_{3i} - a_3) + e_i$$

donde $i=1, 2, \dots, 30$; $a_1 = \bar{X}_1$, $a_2 = \bar{X}_2$ y $a_3 = \bar{X}_3$.

La distribución marginal posterior de $\underline{\beta} = (\beta_1, \beta_2, \beta_3, \beta_{01})'$ es T de Student multivariada con 26 grados de libertad, vector de localización $(2.291, 5.106, 3.056, 0.000)'$ y matriz de precisión

$$\begin{pmatrix} 7.728 & -0.300 & -0.324 & 0 \\ -0.300 & 7.295 & -1.619 & 0 \\ -0.324 & -1.619 & 6.603 & 0 \\ 0 & 0 & 0 & 0.009 \end{pmatrix}$$

TABLA 12. DATOS EJEMPLO 2. HAY 10 PERDIDAS ENTRE LOS VALORES DE LA VARIABLE X_3 .

Y	X_1	X_2	X_3
613.524	92	42	73
356.427	83	34	7
553.958	82	63	25
379.261	1	59	28
771.964	10	91	99
465.409	65	12	92
560.674	37	93	7
609.210	56	81	31
400.782	39	10	91
320.088	23	41	23
409.965	43	9	93
691.264	15	89	72
572.660	56	45	78
632.280	5	82	71
110.858	6	18	5
686.888	90	43	98
617.659	98	78	11
613.139	66	57	65
572.008	17	83	41
325.017	6	11	86
360.695	3	44	9999.9999
765.385	48	90	9999.9999
268.184	43	19	9999.9999
772.532	57	73	9999.9999
530.747	43	83	9999.9999
363.417	7	20	9999.9999
446.991	24	73	9999.9999
718.581	88	50	9999.9999
703.416	40	87	9999.9999
249.438	55	11	9999.9999

Nota. - 9999.9999 representa un valor faltante.

La distribución marginal posterior de μ es Gamma con parámetros $\alpha=13$ y $\gamma=44928.302$.

El coeficiente de determinación es 0.899.

Los intervalos de mayor densidad para β_1 , β_2 y β_3 de probabilidad 0.5, 0.8, 0.9 y 0.95 son:

P_x	β_1	β_2	β_3
0.5	[2.045 , 2.537]	[4.853 , 5.359]	[2.790 , 3.322]
0.8	[1.818 , 2.764]	[4.611 , 5.523]	[2.544 , 3.568]
0.9	[1.677 , 2.905]	[4.474 , 5.738]	[2.392 , 3.720]
0.95	[1.552 , 3.031]	[4.345 , 5.867]	[2.256 , 3.856]

Antes de proseguir con este ejemplo, es conveniente hacer algunas observaciones sobre el material presentado. Hay que notar, observando las matrices de precisión de las distribuciones marginales posteriores del vector β , que las precisiones marginales (elementos de la diagonal) se ven muy disminuidas por la presencia de observaciones faltantes y que la disminución es directamente proporcional a la correlación entre la variable dependiente y la variable que sufre las pérdidas. Además, y en concordancia a lo discutido en la primera parte de este capítulo, los coeficientes de determinación han disminuido al haber pérdidas, siendo esta disminución directamente proporcional a

La correlación entre la variable dependiente y la variable con pérdidas. Estos cambios indican, por la dependencia existente entre el coeficiente de determinación y γ , el parámetro de la distribución marginal posterior de w , la manera en que ésta se ve afectada por las pérdidas.

A continuación y en base en el conjunto de datos de la tabla 7, se ejemplifica con conjuntos de datos con estructuras de pérdidas cada vez más complejas.

En la tabla 13, se presenta un conjunto de datos con diez pérdidas en X_1 y ocho pérdidas en X_2 . Se tienen para este conjunto las siguientes medias y varianzas muestrales:

$$\begin{array}{rcl} \bar{Y} & = & 514.747 \quad , \quad S_Y^2 = 30548.795 \quad , \\ \bar{X}_1 & = & 44.500 \quad , \quad S_{X_1}^2 = 1105.737 \quad , \\ \bar{X}_2 & = & 53.364 \quad , \quad S_{X_2}^2 = 936.814 \quad , \\ \bar{X}_3 & = & 54.400 \quad , \quad S_{X_3}^2 = 1136.041 \quad , \end{array}$$

El modelo que se ajusta es:

$$B_i = Y_i - \bar{Y} - \beta_{01} + \beta_1 (X_{1i} - a_1) + \beta_2 (X_{2i} - a_2) + \beta_3 (X_{3i} - a_3) + e_i$$

donde $i=1, 2, \dots, 30$; $a_1 = \bar{X}_1$, $a_2 = \bar{X}_2$ y $a_3 = \bar{X}_3$.

La distribución marginal posterior de $\underline{\beta} = (\beta_1, \beta_2, \beta_3, \beta_{01})'$ es una distribución T de Student multivariada con 26 grados de li-

TABLA 13. DATOS EJEMPLO 2. ENTRE LOS VALORES DE LAS VARIABLES X Y X HAY 10 Y 8 PERDIDAS RESPECTIVAMENTE.

Y	X ₁	X ₂	X ₃
613.524	22	42	73
356.427	83	34	7
553.958	82	63	25
379.261	1	59	28
771.964	10	91	99
465.409	65	12	92
560.674	37	23	7
609.210	56	81	31
400.782	39	10	91
320.088	23	41	23
409.965	43	9	23
691.264	15	89	72
572.660	56	9999.9999	78
632.280	5	9999.9999	71
110.858	6	9999.9999	5
686.888	90	9999.9999	98
617.659	98	9999.9999	11
613.139	66	9999.9999	65
572.008	17	9999.9999	41
325.017	6	9999.9999	86
360.695	9999.9999	44	45
765.385	9999.9999	20	73
268.184	9999.9999	12	28
772.532	9999.9999	73	98
530.747	9999.9999	83	10
363.417	9999.9999	20	83
446.991	9999.9999	73	12
718.582	9999.9999	50	96
703.416	9999.9999	87	63
249.438	9999.9999	11	28

Nota. - 9999.9999 representa un valor faltante.

bertad, vector de localización $(2.488, 5.177, 3.017, 0.000)'$
y matriz de precisión:

$$\begin{pmatrix} 3.439 & -0.618 & -0.184 & 0 \\ -0.618 & 3.221 & -0.608 & 0 \\ -0.184 & -0.608 & 5.394 & 0 \\ 0 & 0 & 0 & 0.005 \end{pmatrix}$$

La distribución marginal posterior de w es Gamma con $\alpha=13$ y $\gamma=79402.495$. El valor del coeficiente de determinación es 0.821.

Intervalos de mayor densidad para β_1 , β_2 y β_3 de probabilidad en 0.5, 0.8, 0.9 y 0.95

P_x	β_1	β_2	β_3
0.5	[2.119, 2.856]	[4.796, 5.558]	[2.723, 3.312]
0.8	[3.779, 3.197]	[4.444, 5.910]	[2.457, 3.583]
0.9	[3.568, 3.406]	[4.226, 6.128]	[2.283, 3.752]
0.95	[3.379, 3.596]	[4.031, 6.323]	[2.132, 3.903]

En la tabla 14, se presenta el conjunto de datos de la tabla 13 ahora con ocho pérdidas más en la variable X_3 . Con este conjunto de datos se tiene las siguientes medias y varianzas:

TABLA 14. DATOS EJEMPLO 2. ENTRE LOS VALORES DE LA VARIABLE X_1 , X_2 Y X_3 HAY 10, 8 Y 8 PERDIDAS RESPECTIVAMENTE

y	X_1	X_2	X_3
613.524	22	42	73
356.427	83	34	7
553.958	82	63	25
379.261	1	59	28
771.964	10	91	9999.9999
465.409	65	12	9999.9999
560.674	37	93	9999.9999
609.210	56	81	9999.9999
400.782	39	10	9999.9999
320.088	23	41	9999.9999
409.965	43	9	9999.9999
691.264	15	89	9999.9999
572.660	56	9999.9999	78
632.280	5	9999.9999	71
110.858	6	9999.9999	5
686.888	90	9999.9999	98
617.659	28	9999.9999	11
613.139	66	9999.9999	65
572.008	17	9999.9999	41
325.017	6	9999.9999	86
360.695	9999.9999	44	45
765.384	9999.9999	90	73
268.183	9999.9999	19	28
772.532	9999.9999	73	98
530.747	9999.9999	83	10
363.417	9999.9999	20	83
446.991	9999.9999	73	12
718.582	9999.9999	50	96
703.416	9999.9999	87	63
249.438	9999.9999	11	28

Nota. - 9999.9999 representa un valor faltante.

$$\begin{array}{rcl}
 \bar{Y} & = & 514.747, \quad S_Y^2 = 30548.795, \\
 \bar{X}_1 & = & 44.500, \quad S_{X_1}^2 = 1105.737, \\
 \bar{X}_2 & = & 53.364, \quad S_{X_2}^2 = 936.814, \\
 \bar{X}_3 & = & 51.091, \quad S_{X_3}^2 = 1065.801.
 \end{array}$$

El modelo que se ajusta es:

$$B_i = Y_i - \bar{Y} = \beta_{0i} + \beta_1 (X_{1i} - a_1) + \beta_2 (X_{2i} - a_2) + \beta_3 (X_{3i} - a_3) + \epsilon_i$$

donde $i=1, 2, \dots, 30$; $a_1 = \bar{X}_1$, $a_2 = \bar{X}_2$ y $a_3 = \bar{X}_3$.

La distribución marginal posterior de $\underline{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)$ es una T de Student multivariada con 26 grados de libertad, vector de localización $(2.204, 4.405, 2.982, 0.000)'$ y matriz de precisión:

$$\begin{pmatrix}
 2.165 & -0.389 & -1.146 & 0 \\
 -0.389 & 2.027 & 0.101 & 0 \\
 -1.146 & 0.101 & 2.307 & 0 \\
 0 & 0 & 0 & 0.003
 \end{pmatrix}$$

La distribución marginal posterior de w es Gamma con $\alpha=13$ y $\gamma=126149.436$. El valor del coeficiente de determinación es 0.715.

Los intervalos de mayor densidad para β_1 , β_2 y β_3 de probabilidades 0.5, 0.8, 0.9 y 0.95 son:

P_x	β_1	β_2	β_3
0.5	[1.739 , 2.669]	[3.925 , 4.886]	[2.531 , 3.432]
0.8	[1.310 , 3.097]	[3.482 , 5.329]	[2.116 , 3.848]
0.9	[1.044 , 3.363]	[3.207 , 5.603]	[1.858 , 4.105]
0.95	[0.806 , 3.601]	[2.961 , 5.849]	[1.628 , 4.335]

2.6.-En la tabla 15 se presenta el conjunto de datos de la tabla 14, ahora con una pérdida mas en cada renglón (de los que presentan pérdidas en la tabla 13). Con estos datos se calculan las medias y varianzas muestrales, resultando:

$$\begin{aligned} \bar{Y} &= 514.747 & , & & S_y^2 &= 30548.795 & , \\ \bar{X}_1 &= 47.000 & , & & S_{x_1}^2 &= 1224.909 & , \\ \bar{X}_2 &= 55.923 & , & & S_{x_2}^2 &= 675.410 & , \\ \bar{X}_3 &= 63.923 & , & & S_{x_3}^2 &= 979.577 & . \end{aligned}$$

El modelo que se ajustó es:

$$B_i = Y_i - \bar{Y} = \beta_{0i} + \beta_1(X_{1i} - a_1) + \beta_2(X_{2i} - a_2) + \beta_3(X_{3i} - a_3) + \epsilon_i ,$$

donde $i=1, 2, \dots, 30$; $a_1 = \bar{X}_1$, $a_2 = \bar{X}_2$ y $a_3 = \bar{X}_3$.

La distribución marginal posterior de $\underline{\beta} = (\beta_1, \beta_2, \beta_3, \beta_{01})'$ es una T de Student multivariada con 26 grados de libertad, vector de localización $(2.079, 3.890, 3.407, 0.000)'$ y matriz de precisión:

**TABLA 15. EJEMPLO 2. ENTRE LOS VALORES DE LAS VARIABLES
X, X, Y X HAY 18,17 Y 17 PERDIDAS RESPECTIVAMENTE.**

Y	X₁	X₂	X₃
613.524	22	42	73
356.427	83	34	7
553.958	82	63	25
579.261	1	59	28
771.964	10	9999.9999	9999.9999
465.409	9999.9999	12	9999.9999
560.674	37	9999.9999	9999.9999
609.210	9999.9999	81	9999.9999
400.782	39	9999.9999	9999.9999
320.088	9999.9999	41	9999.9999
409.965	43	9999.9999	9999.9999
691.264	9999.9999	19	9999.9999
572.660	56	9999.9999	9999.9999
632.280	9999.9999	9999.9999	71
110.858	6	9999.9999	9999.9999
686.888	9999.9999	9999.9999	98
617.659	98	9999.9999	9999.9999
613.139	9999.9999	9999.9999	65
572.008	17	9999.9999	9999.9999
325.017	9999.9999	9999.9999	86
360.695	9999.9999	44	9999.9999
765.385	9999.9999	9999.9999	73
268.184	9999.9999	19	9999.9999
772.532	9999.9999	9999.9999	98
530.747	9999.9999	83	9999.9999
363.417	9999.9999	9999.9999	83
446.991	9999.9999	73	9999.9999
718.582	9999.9999	9999.9999	96
703.416	9999.9999	87	9999.9999
249.438	9999.9999	9999.9999	28

Nota.- 9999.9999 representa un valor faltante.

$$\begin{pmatrix} 0.514 & -0.058 & -0.060 & 0 \\ -0.058 & 0.358 & 0.032 & 0 \\ -0.060 & 0.032 & 0.519 & 0 \\ 0 & 0 & 0 & 0.001 \end{pmatrix}$$

La distribución marginal posterior de w es Gamma con parámetros $\alpha=13$ y $\gamma=294673.252$. El valor del coeficiente de determinación es 0.335.

Los intervalos de mayor densidad para β_1, β_2 y β_3 son:

P_x	β_1	β_2	β_3
0.5	[1.192, 2.966]	[2.747, 5.034]	[2.458, 4.357]
0.8	[0.373, 3.785]	[1.691, 6.090]	[1.581, 5.233]
0.9	[0.134, 4.292]	[1.038, 6.744]	[1.038, 5.776]
0.95	[0.588, 4.746]	[0.452, 7.329]	[0.552, 6.262]

2.7.- En la tabla 16, se presenta el conjunto de datos de la tabla 15 ahora con los cuatros primeros renglones con pérdidas.

Calculando las medias y varianzas muestrales de estos datos se tiene:

$$\begin{aligned} \bar{Y} &= 514.747 & , & & S_y^2 &= 30548.795 & , \\ \bar{X}_1 &= 47.100 & , & & S_{x_1}^2 &= 1036.989 & , \\ \bar{X}_2 &= 58.800 & , & & S_{x_2}^2 &= 810.844 & , \\ \bar{X}_3 &= 63.167 & , & & S_{x_3}^2 &= 1060.515 & . \end{aligned}$$

TABLA 16. EJEMPLO 2. ENTRE LOS VALORES DE LAS VARIABLES X_1 , X_2 Y X_3 , HAY 20, 20 Y 13 PERDIDAS RESPECTIVAMENTE

Y	X_1	X_2	X_3
613.524	9999.9999	9999.9999	9999.9999
356.427	83	9999.9999	7
553.958	82	9999.9999	25
379.261	9999.9999	59	28
771.964	10	9999.9999	9999.9999
465.409	9999.9999	12	9999.9999
560.674	37	9999.9999	9999.9999
609.210	9999.9999	81	9999.9999
400.782	39	9999.9999	9999.9999
320.088	9999.9999	41	9999.9999
409.965	43	9999.9999	9999.9999
691.264	9999.9999	89	9999.9999
572.660	56	9999.9999	9999.9999
632.280	9999.9999	9999.9999	71
110.858	6	9999.9999	9999.9999
686.888	9999.9999	9999.9999	98
617.659	98	9999.9999	9999.9999
613.139	9999.9999	9999.9999	65
572.008	17	9999.9999	9999.9999
325.017	9999.9999	9999.9999	86
360.695	9999.9999	44	9999.9999
765.385	9999.9999	9999.9999	73
268.184	9999.9999	19	9999.9999
772.532	9999.9999	9999.9999	98
530.747	9999.9999	83	9999.9999
363.417	9999.9999	9999.9999	83
446.991	9999.9999	73	9999.9999
718.582	9999.9999	9999.9999	96
703.416	9999.9999	87	9999.9999
249.438	9999.9999	9999.9999	28

Nota. 9999.9999 representa un valor faltante.

El modelo que se ajustó es:

$$B_{1i} = y_{1i} - \bar{y} = \beta_{01} + \beta_1(X_{1i} - a_1) + \beta_2(X_{2i} - a_2) + \beta_3(X_{3i} - a_3) + e_{1i}$$

donde $i=1, 2, \dots, 30$; $a_1 = \bar{X}_1$, $a_2 = \bar{X}_2$ y $a_3 = \bar{X}_3$.

La distribución marginal posterior de $\underline{\beta} = (\beta_1, \beta_2, \beta_3, \beta_{01})'$ es una T de Student con 26 grados de libertad, vector de localización $(2.267, 4.117, 4.030, 0.000)'$ y matriz de precisión:

$$\begin{pmatrix} 0.414 & 0 & -0.148 & 0 \\ 0 & 0.324 & 0 & 0 \\ -0.148 & 0 & 0.517 & 0 \\ 0 & 0 & 0 & 0.401 \end{pmatrix}$$

La distribución marginal posterior de w es Gamma con $\alpha = 13$ y $\gamma = 293131.055$. El coeficiente de determinación tomó el valor 0.338.

Los intervalos de mayor densidad para β_1 , β_2 y β_3 son:

	β_1	β_2	β_3
0.5	[1.203, 3.330]	[2.914, 5.319]	[3.079, 4.983]
0.8	[0.223, 4.311]	[1.805, 6.428]	[2.202, 5.858]
0.9	[-0.385, 4.918]	[1.118, 7.115]	[1.658, 6.402]
0.95	[0.929, 5.462]	[0.503, 7.731]	[1.171, 6.888]

Resumiendo los resultados antes presentados (Tablas 17 a 19)

**TABLA 17. EJEMPLO 2. DISTRIBUCIONES MARGINALES POSTERIORES DE θ_1 .
 LAS DISTRIBUCIONES SON T DE STUDENT CON 26 GRADOS DE LIBERTAD. EL VALOR VERDADERO DE θ_1 ES 2.**

NÚMERO DE PERDIDAS EN LAS VARIABLES			PORCENTAJE DE OB- SERVACIONES CON PÉRDIDAS.	LOCALIZACIÓN	PRECISIÓN (ESCALA)
X_1	X_2	X_3			
0	0	0	0	2.015	9670.657
10	0	0	33.33	2.028	24.097
0	10	0	33.33	2.176	3.383
0	0	10	33.33	2.291	7.728
10	8	0	60.00	2.488	3.439
10	8	8	86.66	2.204	2.165
18	18	17	86.66	2.079	0.594
20	20	18	100.00	2.267	0.414

**TABLA 18. EJEMPLO 2. DISTRIBUCIONES MARGINALES POSTERIORES DE θ_1 .
 LAS DISTRIBUCIONES SON T DE STUDENT CON 26 GRADOS DE LIBERTAD. EL VALOR VERDADERO DE θ_1 ES 5.**

NÚMERO DE PERDIDAS EN LAS VARIABLES			PORCENTAJE DE OB- SERVACIONES CON PÉRDIDAS.	LOCALIZACIÓN	PRECISIÓN (ESCALA)
X_1	X_2	X_3			
0	0	0	0	5.004	9128,644
10	0	0	33,33	5,105	28,917
0	10	0	33,33	5,074	2,137
0	0	10	33,33	5,106	7,295
10	8	0	60,00	5,177	3,221
10	8	8	86,66	4,405	2,027
18	18	17	86,66	3,890	0,358
20	20	18	100,00	4,117	0,324

TABLA 19. EJEMPLO 2. DISTRIBUCIONES MARGINALES POSTERIORES DE μ . LAS DISTRIBUCIONES SON T DE STUDENT CON 26 GRADOS DE LIBERTAD. EL VALOR VERDADERO DE μ ES 3.

NÚMERO DE PERDIDAS EN LAS VARIABLES			PORCENTAJE DE OBSERVACIONES CON PÉRDIDAS.	LOCALIZACIÓN	PRECISIÓN (ESCALA)
X_1	X_2	X_3			
0	0	0	0	3.016	11929.134
10	0	0	33.33	3.178	37.788
0	10	0	33.33	3.126	4.173
0	0	10	33.33	3.056	6.603
10	8	0	60.00	3.017	5.394
10	8	8	86.66	2.982	2.307
18	18	17	86.66	3.407	0.519
20	20	18	100.00	4.030	0.517

puede concluirse al utilizar la solución bayesiana que al haber pérdidas entre las observaciones de las variables explicativas las precisiones de las distribuciones marginales posteriores se disminuyen en forma drástica y se puede conjeturar que la magnitud de la disminución depende tanto del porcentaje de observaciones con pérdidas como del número de observaciones faltantes en cada una de las variables explicativas.

Sin embargo la disminución en la precisión moldea de tal forma a las distribuciones marginales posteriores de β_1 , β_2 y β_3 que resultan contenidos los verdaderos valores de estos parámetros en regiones de mayor densidad para probabilidades que en la mayoría de los casos son tan bajas como 0.5 y en el resto de los casos ya es tan contenidos para probabilidades 0.8 .

Esto indica que en todos los casos estudiados los verdaderos valores de los parámetros son valores "cercanos" a la localización de la distribución.

EJEMPLO 3.- En este ejemplo se presenta el conjunto de datos simulados y publicados por Dagenais [1974]. La importancia

de trabajar con este conjunto de datos, está en el hecho de ser uno de los pocos conjuntos de datos publicados en la literatura en relación al problema de observaciones faltantes. Al final de este ejemplo se discuten las conclusiones obtenidas por Dagenais (1974).

En la tabla 17, se presenta el conjunto de datos en los que la variable dependiente se simuló con el modelo:

$$Y_i = 2X_{1i} + 3X_{2i} + \epsilon_i; \quad i = 1, 2, \dots, 30$$

La información referente a la manera en que se obtuvieron ϵ , X_1 y X_2 no es precisa o completa, lo que se sabe es que ϵ y X_1 se simulan Normales, ϵ con media cero y X_1 con media uno. Los valores de X_2 se escogieron de tal forma que X_2 resultara "medianamente" correlacionada con X_1 . Las varianzas de ϵ y X_1 se ajustaron para producir los coeficientes de correlación que reproducimos en la tabla 17.

Los valores perdidos se escogieron al azar con probabilidades $1/15$ para Y y X_2 , y $1/4$ para X_1 , produciendo entonces el conjunto de datos que reproducimos en la tabla 18.

Las medias y varianzas muestrales calculadas con los datos completos (tabla 17) son:

TABLA 17. DATOS EJEMPLO 3. SIN PERDIDAS

Y	X ₁	X ₂	Y	X ₁	X ₂
83.09	-0.66	17.81	-23.86	-0.41	-14.01
19.56	-5.55	13.94	57.47	18.09	10.38
36.26	-3.68	16.96	-79.15	7.04	-13.06
-35.68	-8.62	-11.16	-99.99	0.60	-9.94
115.07	6.77	6.03	-102.12	-11.70	-29.00
-23.42	-2.22	-5.84	-49.27	-9.14	0.21
-5.64	-4.24	-13.04	47.71	7.53	2.38
-10.41	22.91	-8.16	147.30	29.09	17.96
-129.72	-5.04	-25.01	4.38	6.25	10.13
27.53	-2.46	9.65	-49.39	-8.12	-17.99
-132.60	-14.27	-36.08	121.00	21.96	-24.59
-17.80	10.88	0.96	57.91	11.90	0.19
-21.69	-4.70	-6.67	-79.05	1.38	-10.86
13.11	-15.55	0.90	4.34	-4.26	3.92
-46.23	-11.46	3.42	-75.22	-15.79	-15.90

(CONTINUACIÓN) CORRELACIONES, MEDIAS Y VARIANZAS.

MATRIZ DE CORRELACIONES		
Y	X ₁	X ₂
1	0.624	0.805
	1	0.497
		1
MEDIAS		
Y	X ₁	X ₂
-16.284	-0.670	-5.380
VARIANZAS		
Y	X ₁	X ₂
4946.795	136.118	192.825

Fuente: Dagenais (1974).

$$\begin{array}{rcl} \bar{Y} = 16.284 & , & S_Y^2 = 4246.725 \\ \bar{X}_1 = 0.670 & , & S_{X_1}^2 = 136.118 \\ \bar{X}_2 = 5.380 & y & S_{X_2}^2 = 192.825 \end{array}$$

Se ajustó el siguiente modelo a los datos completos:

$$B_i = Y_i - \bar{Y} = \beta_1(X_{1i} - \bar{X}_1) + \beta_2(X_{2i} - \bar{X}_2) + \varepsilon_i$$

con $i=1, 2, \dots, 30$.

La distribución inicial que se utilizó fue una de referencia. La distribución marginal posterior de $\underline{\beta} = (\beta_1, \beta_2)$ es una T Student multivariada con 27 grados de libertad, vector de localización $[1.720, 3.328]'$ y matriz de precisión

$$\begin{pmatrix} 2.692 & 1.593 \\ 1.593 & 3.813 \end{pmatrix}$$

La distribución marginal posterior de w es Gamma con $\alpha=14$ y $\gamma=20530.940$. El coeficiente de determinación tomó el valor 0.714.

Es importante notar que debido a la forma en que los datos fueron simulados, el modelo de regresión que se obtiene al ajustar los datos completos sólo explica el 71.4% de la variación observada en la variable dependiente.

Los intervalos de mayor densidad para β_1 y β_2 son:

	β_1	β_2
0.5	[1.373 , 2.207]	[2.278 , 3.678]
0.8	[0.282 , 2.591]	[2.655 , 4.001]
0.9	[0.752 , 2.828]	[2.456 , 4.200]
0.95	[0.539 , 3.041]	[2.277 , 4.379]

Considerando ahora los datos incompletos (tabla 18), se tienen las siguientes medias y varianzas muestrales:

$$\begin{aligned} \bar{Y} &= 6.553 & , & & S_y^2 &= 4262.113 \\ \bar{X}_1 &= 0.603 & , & & S_{X_1}^2 &= 122.272 \\ \bar{X}_2 &= 6.606 & , & & S_{X_2}^2 &= 183.357 \end{aligned}$$

Se ajustó el modelo:

$$B_i = Y_i - a_0 = \beta_{01} + \beta_1 (X_{1i} - a_1) + \beta_2 (X_{2i} - a_2) + \epsilon_i$$

donde $i=1, 2, \dots, 30$ y $a_0 = \bar{Y}$, $a_1 = \bar{X}_1$ y $a_2 = \bar{X}_2$.

La distribución inicial es de referencia y la distribución marginal posterior de $\underline{\beta} = (\beta_1, \beta_2, \beta_{01})'$ es una T de Student multivariada con 24 grados de libertad, vector de localización (1.476, 3.692, -5.922)' y matriz de precisión

$$\begin{pmatrix} 1.454 & , & 0.872 & & 0.003 \\ 0.872 & & 2.119 & & 0.024 \\ 0.003 & & 0.024 & & 0.015 \end{pmatrix}$$

**TABLA 18. DATOS EJEMPLO 3. ENTRE LOS VALORES DE LAS VARIABLES
Y, X₁ Y X₂, HAY 3, 8 Y 2 PERDIDAS RESPECTIVAMENTE**

Y	X ₁	X ₂
83.09	-.66	17.81
19.56	-5.55	9999.9999
36.26	3.68	-16.96
-35.68	-8.62	-11.16
115.07	6.77	6.03
-23.42	-2.22	-5.04
-5.64	-4.24	-13.04
-10.41	9999.9999	-8.16
9999.9999	-5.04	-25.01
21.53	-2.46	9999.9999
9999.9999	9999.9999	-36.88
17.8	10.88	.96
-21.69	9999.9999	-6.67
13.11	-15.55	.9
-46.23	-11.46	3.42
-23.86	-.41	-14.01
57.47	18.09	10.38
-79.15	7.04	-13.06
-99.99	9999.9999	-9.94
-102.12	11.7	-29
9999.9999	9999.9999	.21
47.71	7.53	2.38
147.3	29.89	17.76
4.38	6.25	10.13
-49.32	9999.9999	17.99
-121	9999.9999	-24.59
57.91	11.2	.19
-72.05	9999.9999	-10.86
.34	-4.26	3.92
-75.22	-15.72	-15.9

Fuente: DAGENATS [1974].

La distribución marginal posterior de w es Gamma con parámetros $\alpha=12$ y $\gamma=20935.867$. El valor del coeficiente de determinación es 0.63).

Los intervalos de mayor densidad sobre β_1 , β_2 y β_{01} son :

P_T	β_1	β_2	β_{01}
0.50	[0.908 , 2.044]	[3.221 , 4.162]	[-11.429, -0.416]
0.80	[0.383 , 2.570]	[2.786 , 4.597]	[-16.517, 4.672]
0.90	[0.057 , 2.896]	[2.516 , 4.867]	[-19.676, 7.831]
0.95	[-0.236 , 3.188]	[2.274 , 5.109]	[-22.514, 10.669]

Conviene notar que en este ejemplo (y a diferencia de los ejemplos anteriores) β_{01} ha quedado con una distribución marginal posterior con las siguientes características: centrada en un valor diferente de cero (en este caso -5.922) y con una precisión de 0.015. Esto es debido a la existencia de valores faltantes entre los valores de Y , y se refleja en la predicción de valores de la variable dependiente.

Los verdaderos valores de los parámetros β_1 y β_2 (2 y 3 respectivamente) están incluidos en regiones (marginales) de mayor densidad de probabilidad desde probabilidades .47 y .68 respectivamente. Esto puede utilizarse como un indicador de la proximidad de los parámetros de localización respecto a los verdaderos valo-

rea de los parámetros.

En Dagenais [1974] pag. 287, se concluye en relación a este conjunto de datos mediante dos afirmaciones, mismas que se reproducen y discuten a continuación:

a) Las modas de las distribuciones [marginales posteriores de β_1 y β_2] están más cerca de los verdaderos valores de los parámetros, para las densidades obtenidas mediante este método (el de Dagenais), que las obtenidas con las densidades correspondientes a la muestra completa y a la muestra con solo las 19 observaciones completas.

b) Como se esperaba, la dispersión de las densidades posteriores obtenidas es menor que la de las densidades posteriores basadas en las 19 observaciones completas.

En ambas afirmaciones se comparan los resultados de Dagenais contra los obtenidos con el conjunto de 19 observaciones completas producido al ignorar a las observaciones incompletas.

Es importante recordar que en el ejemplo 1, se muestra que utilizar al conjunto de observaciones completas ignorando las observaciones incompletas puede producir una distribución posterior no confiable. En el ejemplo 1 la distribución posterior correspondiente presentó (entre otras fallas) una precisión "inflada" [75 442, 341], que resultó aún más grande que la obtenida con el

conjunto completo de datos (10 336. 242). El que la distribución posterior, obtenida al ignorar a las observaciones incompletas, resulte no confiable se debe a que es incorrecto ignorar a las observaciones incompletas.

En base a la discusión anterior resulta que en un problema (de regresión) con datos faltantes es inapropiado utilizar como punto de referencia a la distribución posterior obtenida al ignorar las observaciones incompletas y no es conveniente en base a ella "esperar" resultados que permitan juzgar la conveniencia de un procedimiento estadístico.

La afirmación de que las modas de las distribuciones posteriores obtenidas mediante el método de Dagénais están (aparentemente, ya que sólo se juzga mediante una gráfica) más cerca de los verdaderos valores de los parámetros, que las modas obtenidas con las densidades posteriores correspondientes a la muestra completa. Es intuitivamente inaceptable como una propiedad general del método de Dagénais (o de cualquier procedimiento para problemas con datos incompletos), ya que entonces resultaría preferible tener observaciones perdidas que no tenerlas.

EJEMPLO 4.- En este caso se ilustra la posibilidad de utilizar técnicas asociadas al análisis de regresión en problemas con datos faltantes. Esto se hace específicamente mediante la utilización del

coeficiente de determinación como un indicador para la selección de variables dependientes. Se comparan los resultados obtenidos por Rubin [1976 a] con los obtenidos con la solución propuesta en este trabajo.

Se presenta en la tabla 20 el conjunto de datos que originalmente aparecen en el capítulo 6 del libro de Draper y Smith (1968) y que Rubin [1976 a] utiliza para ilustrar el comportamiento de la estadística que propone para estimar al coeficiente de determinación en problemas de regresión con datos incompletos.

TABLA 20. DATOS EJEMPLO 4. SIN PERDIDAS

y	x_1	x_2	x_3	x_4
78.5	7	26	6	60
74.3	1	29	15	52
104.3	11	56	8	20
87.6	11	31	8	47
95.9	7	52	6	33
109.2	11	55	9	22
102.7	3	71	17	6
72.5	1	31	22	44
93.1	2	54	18	22
115.9	21	47	4	26
83.8	1	40	23	34
113.3	11	66	9	12
109.4	10	68	8	12

FUENTE: RUBIN (1976 a)

El método propuesto por Rubín para estimar el coeficiente de determinación (ver sección 1.3.4) se basa en los resultados de Hartley & Hocking (1973), Rubín (1974) o Beale & Little (1975) que son procedimientos iterativos, que suponen Normalidad y que producen estimadores aproximadamente máximo verosímiles. Estos métodos requieren para su utilización de programas de computadora especiales y complicados. Este estimador se denota en este trabajo mediante Q^2 . Rubín afirma que $Q^2 = R_c^2$ sólo cuando las variables explicativas no presentan pérdidas y que en general $Q < Q^2 < R_c^2$, donde R_c^2 es el coeficiente de determinación obtenido con el conjunto de datos completo. Como ya se mencionó en la sección 1.3.4, Rubín sólo trata pérdidas entre las variables explicativas.

En la parte 1, se ha estudiado el comportamiento del coeficiente de determinación definido en forma usual y calculado a partir de las distribuciones posteriores obtenidas en el capítulo 3. Se ha encontrado que este coeficiente de determinación, que se ha denotado por R^2 , es tal que $0 < R^2 < 1$ y que $R^2 = R_c^2$ cuando no hay pérdidas entre los valores de la variable dependiente, R^2 puede ser mayor o igual que R_c^2 y que cuando hay pérdidas entre las variables explicativas se tiene con una probabilidad "alta" que $R^2 < R_c^2$ y con una probabilidad "baja" $R^2 > R_c^2$. Entonces resulta al menos posible que $R^2 > Q^2$ en algunos casos.

Rubín utilizando los datos de la tabla 20, descarta algunas observaciones y presenta los datos de la tabla 21.

En la tabla 22 se presentan, en la primer columna los indicadores de qué variables explicativas se usaron en la regresión, por ejemplo 234, indica que en ese caso la regresión tuvo como variables explicativas a X_2, X_3 y X_4 ; en la columna dos se presentan los coeficientes de determinación obtenidos con los datos completos [tabla 20.], en la columna tres los resultados obtenidos con el estimador de Rubín y en la columna cuatro los coeficientes de correlación (R^2) calculados a partir de las distribuciones posteriores obtenidas en el capítulo tres de este trabajo.

Se puede observar en esta tabla que todas las Q^2 y las R^2 son menores que las correspondientes R^2_c .

En cuanto a las indicaciones que se obtienen con la Q^2 y la R^2 sobre los mejores modelos con una, dos y tres variables, se tiene que coinciden en señalar que entre las regresiones con sólo una variable la mejor es la que se forma con la variable explicativa 2. y que entre las regresiones con dos variables la mejor es la que se forma con las variables 2 y 3. Entre las regresiones con tres variables explicativas, con las Q^2 se apunta como la mejor a la formada con las variables 1, 2, 3 y después de ésta a la formada con las variables 2, 3 y 4, mientras que con la R^2 se apunta como la mejor a la

TABLA 21. DATOS EJEMPLO 4. ENTRE LOS VALORES DE LAS VARIABLES X_1 , X_2 , Y X_3 HAY 4,4 Y 7 PERDIDAS RESPECTIVAMENTE

Y	X_1	X_2	X_3	X_4
78.5	7	26	6	60
74.3	1	29	15	52
104.3	11	56	8	20
87.6	11	31	8	47
95.9	7	52	6	33
109.2	11	55	9	22
102.7	3	71	17	9999.9999
72.5	1	31	22	9999.9999
93.1	2	54	18	9999.9999
115.9	9999.9999	9999.9999	4	9999.9999
83.8	9999.9999	9999.9999	23	9999.9999
113.3	9999.9999	9999.9999	9	9999.9999
109.4	9999.9999	9999.9999	8	9999.9999

FUENTE: RUBIN (1976 a)

Nota.- 9999.9999 representa un valor faltante.

TABLA 22. COEFICIENTES DE DETERMINACION.

Variables Explicativas Incluidas en Regresión	COEFICIENTE DETERMINACIÓN		
	Datos Completos R^2	Estimador de Rubin Q^2	Datos Incomple- tos R^2
0	0.000	0.000	0.000
1	0.534	0.335	0.171
2	0.666	0.532	0.405
3	0.286	0.286	0.286
4	0.675	0.272	0.327
12	0.979	0.679	0.526
13	0.548	0.435	0.296
14	0.973	0.546	0.263
23	0.847	0.728	0.727
24	0.680	0.554	0.440
34	0.935	0.597	0.581
123	0.982	0.767	0.727
124	0.982	0.679	0.527
134	0.981	0.652	0.618
234	0.973	0.745	0.741
1234	0.982	0.775	0.749

formada con las variables 2., 3., 4. y en segundo lugar a la formada con las variables 1., 2., 3.. Es importante notar que prácticamente los valores de R^2 correspondientes a las variables 1., 2., 3. y 2., 3., 4. son iguales indicando esto que realmente en base a la R^2 son prácticamente iguales ambos modelos de regresión.

Puede entonces concluirse que, al menos en este ejemplo han coincidido las indicaciones obtenidas mediante la Q^2 y las obtenidas con el coeficiente de correlación calculado a partir de la solución bayesiana. Sin embargo, cabe resaltar la superioridad de las R^2 sobre Q^2 , tanto por su sencillez en el cálculo como por la lógica de sus resultados mostrada a lo largo de los ejemplos presentados en este capítulo.

Resumiendo de los resultados obtenidos en este capítulo puede concluirse que al menos en los casos estudiados la solución bayesiana ha proporcionado resultados intuitivamente aceptables que al estar respaldados metodológicamente resultan doblemente convincentes.

APENDICE 1

Una inversa generalizada de la matriz A de orden $(m \times k)$ es una matriz A^- , de orden $(k \times m)$ tal que

$$A A^- A = A .$$

Es fácil demostrar que si A' es la matriz transpuesta de A , entonces una elección de $(A')^-$ es $(A^-)'$.

A continuación se demuestra el siguiente resultado:

$$A(A'A)^-(A'A) = A \quad \text{y} \quad (A'A)(A'A)^-A' = A' .$$

Considere el producto

$$\begin{aligned} & (A(A'A)^-(A'A)-A)'(A(A'A)^-(A'A)-A) = \\ & = ((A'A)(A'A)^-A'-A')(A(A'A)^-(A'A)-A) = \\ & = ((A'A)(A'A)^-I)(A(A'A)^-(A'A)-A) = 0 \end{aligned}$$

y como $E'E=0$ implica que $E=0$, para cualquier matriz E , entonces se tiene que:

$$A(A'A)^-(A'A) = A .$$

Análogamente se procede para demostrar que $(A'A)(A'A)^-A' = A'$.

A continuación se demuestra que si A^- existe, entonces $H = A A^-$

es una matriz idempotente y su rango es igual a la traza de H , que a su vez es igual al rango de A , esto es:

$$R(H) = R(A A^{-1}) = T_x(A A^{-1}) = R(A).$$

Si A^{-1} existe, se tiene por definición que $A A^{-1} A = A$ de donde $A A^{-1} A A^{-1} = A A^{-1}$, lo que demuestra que H es idempotente.

Como el rango del producto de dos matrices es menor o igual al rango de cualquiera de las dos matrices, entonces:

$$R(A) \geq R(A A^{-1}) = R(H) \geq R(HA) = R(A)$$

por lo tanto $R(H) = R(A)$ y como H es idempotente se tiene que $R(H)$ es igual a traza de H .

Si $\underline{\alpha}$ es un vector de dimensión (qx) , no idénticamente cero, entonces $(\underline{\alpha} \underline{\alpha}')$ es una matriz de orden (qxq) , de rango uno y tal que no todos sus elementos son cero, por lo tanto existe $(\underline{\alpha} \underline{\alpha}')$ y utilizando los resultados que arriba se demuestran se tiene que:

$$a) \underline{\alpha}' (\underline{\alpha} \underline{\alpha}')$$

$$b) (\underline{\alpha} \underline{\alpha}')$$

$$c) \underline{\alpha}' (\underline{\alpha} \underline{\alpha}')$$

APENDICE 2

Por $|A|$ se denota el determinante de la matriz A .

Como:

$$\text{CCI} \quad |AB| = |A| |B|$$

$$\text{CCII} \quad \begin{vmatrix} E & F \\ G & H \end{vmatrix} = |E| |G|$$

y

$$\text{CCCI} \quad \begin{vmatrix} I & Q \\ -BA^{-1} & T \end{vmatrix} = J$$

entonces:

$$\begin{aligned} \begin{vmatrix} A & C \\ B & D \end{vmatrix} &= \begin{vmatrix} I & Q \\ -BA^{-1} & T \end{vmatrix} \begin{vmatrix} A & C \\ B & D \end{vmatrix} = \begin{vmatrix} A & C \\ 0 & D-BA^{-1}C \end{vmatrix} \\ &= |A| |D-BA^{-1}C| \end{aligned}$$

de este resultado se desprende que:

$$\begin{vmatrix} A & X \\ X' & -J \end{vmatrix} = |A| |-J-X'A^{-1}X| = -|A|(J+X'A^{-1}X)$$

ya que $J+X'A^{-1}X$ es un escalar.

Por otra parte se tiene que:

$$\begin{vmatrix} I & 0 \\ X' & J \end{vmatrix} = J$$

entonces:

$$\begin{vmatrix} A & X \\ X' & -I \end{vmatrix} = \begin{vmatrix} A & X \\ X' & -I \end{vmatrix} \begin{vmatrix} I & 0 \\ X' & I \end{vmatrix} = \begin{vmatrix} A+X X' & X \\ 0 & -I \end{vmatrix} = |A+X X'| (-1)$$

Por lo tanto:

$$|A+X X'| = |A| |I+X' A^{-1} X|$$

si A ($n \times n$), X ($n \times j$) y $|A| \neq 0$.

APENDICE 3.

En este apéndice se demuestra que $\underline{a}_1 D_1 \underline{a}_1' = 0$. Como \underline{a}_1 y D_1 se definen de la siguiente manera:

$$\underline{a}_1 = \underline{a} D D_1^{-1} \quad \text{y} \quad D_1 = d I + D$$

donde

$$\underline{a} = y^* \underline{\alpha}' (\underline{\alpha} \underline{\alpha}')^{-1}$$

$$D = \underline{\alpha} \underline{\alpha}'$$

$$d = (\underline{\alpha}' \underline{\alpha}) / 2$$

y en consecuencia a la definición de D_1 se tiene que $D_1^{-1} = d^{-1} I - (2d^2)^{-1} (\underline{\alpha} \underline{\alpha}')$, entonces utilizando los resultados del apéndice 2:

$$\underline{a} D = y^* \underline{\alpha}' (\underline{\alpha} \underline{\alpha}')^{-1} (\underline{\alpha} \underline{\alpha}') = y^* \underline{\alpha}'$$

de donde:

$$\underline{a}_1 D_1 \underline{a}_1' = \underline{a} D D_1^{-1} D \underline{a}' = y^{*2} \underline{\alpha}' D_1^{-1} \underline{\alpha}$$

y como

$$\begin{aligned} \underline{\alpha}' D_1^{-1} \underline{\alpha} &= d^{-1} (\underline{\alpha}' \underline{\alpha}) - (2d^2)^{-1} \underline{\alpha}' (\underline{\alpha} \underline{\alpha}') \underline{\alpha} \\ &= 0 \end{aligned}$$

se concluye que $\underline{a}_1 D_1 \underline{a}_1' = 0$.

APENDICE 4

DISTRIBUCION PREDICTIVA DE Y.

Se tiene que la distribución de Y dados los valores de β, w y X es Normal $(X\beta, w)$ de donde la densidad es:

$$f(y|X, \beta, w) = w^{1/2} \exp\{-w(y - X\beta)'(y - X\beta)/2\}$$

y si la posterior de β y w es una Normal-Gamma con parámetros $\underline{\mu}, t, \gamma$ y α entonces la función de densidad de probabilidad conjunta de Y, β y w es tal que

$$p(y, \beta, w | X) = w^{1/2} w^{\alpha/2} \exp\{-w(y - X\beta)'(y - X\beta)/2\} \cdot \exp\{-w(\beta - \underline{\mu})'t(\beta - \underline{\mu})/2\} \cdot w^{\alpha-1} \exp\{-\gamma w\}$$

que a su vez es proporcional a

$$[w^{1/2} w^{\alpha/2} w^{\alpha-1} \exp\{-w\gamma\}] \cdot \exp\{-w[(y - X\beta)'(y - X\beta) + (\beta - \underline{\mu})'t(\beta - \underline{\mu})]/2\}$$

y si $H = t + X'X$ entonces

$$H^{-1} = t^{-1} - \frac{t^{-1}(X'X)t^{-1}}{1 + X't^{-1}X}$$

y

$$J = \underline{X} H^{-1} \underline{X}' = \frac{1}{J + \underline{X} \underline{X}'^{-1} \underline{X}'}$$

de donde, es fácil demostrar que:

$$(\underline{y} - \underline{X} \underline{\beta})' (\underline{y} - \underline{X} \underline{\beta}) + (\underline{\beta} - \underline{\mu})' \underline{I} (\underline{\beta} - \underline{\mu}) =$$

$$= (\underline{\beta} - \underline{b})' H (\underline{\beta} - \underline{b}) + (\underline{y} - \underline{X} \underline{\mu})' (J - \underline{X} H^{-1} \underline{X}') (\underline{y} - \underline{X} \underline{\mu})$$

con $\underline{b} = H^{-1} (\underline{X} \underline{\mu} + \underline{X}' \underline{y})$

entonces:

$$p(\underline{y}, \underline{\beta}, w | \underline{X}) = [w^{1/2} w^{p/2} w^{\alpha-1} \exp\{-w\gamma\}]$$

$$\cdot \exp\{-w(\underline{\beta} - \underline{b})' H (\underline{\beta} - \underline{b})/2\}$$

$$\cdot \exp\{-w(\underline{y} - \underline{X} \underline{\mu})' (J - \underline{X} H^{-1} \underline{X}') (\underline{y} - \underline{X} \underline{\mu})/2\}$$

e integrando esta densidad respecto a $\underline{\beta}$ (una Normal p variada), se tiene que la densidad de \underline{y} y w dado \underline{X} es tal que

$$p(\underline{y}, w | \underline{X}) = [w^{1/2} w^{\alpha-1} \exp\{-w\gamma\}]$$

$$\cdot \exp\{-w(\underline{y} - \underline{X} \underline{\mu})' (J - \underline{X} H^{-1} \underline{X}') (\underline{y} - \underline{X} \underline{\mu})/2\}$$

e integrando esta densidad respecto a w (una Gamma) se tiene que

La función de densidad de probabilidad de y dado X es tal que

$$p(y|X) = \frac{\Gamma\left(\frac{2n+1}{2}\right)}{\Gamma\left(\frac{2n}{2}\right)} \left(\frac{\alpha}{\pi}\right)^{1/2} \left\{1 - \frac{(y-X)^2}{\alpha}\right\}^{-\frac{2n+1}{2}}$$

Por lo tanto, la distribución predictiva de y dado X es una T de Student con $2n$ grados de libertad localizada en X y con precisión $\frac{\alpha}{\pi}$.

APENDICE 5

FIGURA 1

clase	Límites	Frecuencia	Frecuencia relativa	Frecuencia acumulada
23	[22.95, 23.95 >	3	0.03	0.03
24	[23.95, 24.95 >	50	0.50	0.53
25	[24.95, 25.95 >	28	0.28	0.81
26	[25.95, 26.95 >	19	0.19	1.00

FIGURA 2

clase	Límites	Frecuencia	Frecuencia relativa	Frecuencia acumulada
22	[21.95, 22.95 >	1	0.01	0.01
23	[22.95, 23.95 >	16	0.16	0.17
24	[23.95, 24.95 >	26	0.26	0.43
25	[24.95, 25.95 >	24	0.24	0.67
26	[25.95, 26.95 >	23	0.23	0.90
27	[26.95, 27.95 >	5	0.05	0.95
28	[27.95, 28.95 >	4	0.04	0.99
29	[28.95, 29.95 >	1	0.01	1.00

FIGURA 3

clase	Límites	Frecuencia	Frecuencia relativa	Frecuencia acumulada
21	[20.95, 21.95 >	2	0.02	0.02
22	[21.95, 22.95 >	9	0.09	0.11
23	[22.95, 23.95 >	20	0.20	0.31
24	[23.95, 24.95 >	7	0.07	0.38
25	[24.95, 25.95 >	20	0.20	0.58
26	[25.95, 26.95 >	16	0.16	0.74
27	[26.95, 27.95 >	15	0.15	0.89
28	[27.95, 28.95 >	9	0.09	0.98
29	[28.95, 29.95 >	2	0.02	1.00

FIGURA 4

clase	Llmites	frecuencia	frecuencia relativa	frecuencia acumulada
93	[92.95, 93.95>	13	0.13	0.13
94	[93.95, 94.95>	65	0.65	0.78
95	[94.95, 95.95>	22	0.22	1.00

FIGURA 5

clase	Llmites	frecuencia	frecuencia relativa	frecuencia acumulada
93	[92.95, 93.95>	35	0.35	0.35
94	[93.95, 94.95>	54	0.54	0.89
95	[94.95, 95.95>	11	0.11	1.00

FIGURA 6

clase	Llmites	frecuencia	frecuencia relativa	frecuencia acumulada
92	[91.95, 92.95>	14	0.14	0.14
93	[92.95, 93.95>	52	0.52	0.66
94	[93.95, 94.95>	31	0.31	0.97
95	[94.95, 95.95>	3	0.03	1.00

FIGURA 7.

<i>clase</i>	<i>Límites</i>	<i>frecuencia</i>	<i>Frecuencia relativa</i>	<i>Frecuencia acumulada</i>
69	[68.95, 69.95>	1	0.01	0.01
70	[69.95, 70.95>	0	0.00	0.01
71	[70.95, 71.95>	0	0.00	0.01
72	[71.95, 72.95>	2	0.02	0.03
73	[72.95, 73.95>	2	0.02	0.05
74	[73.95, 74.95>	0	0.00	0.05
75	[74.95, 75.95>	2	0.02	0.07
76	[75.95, 76.95>	6	0.06	0.13
77	[76.95, 77.95>	13	0.13	0.26
78	[77.95, 78.95>	9	0.09	0.35
79	[78.95, 79.95>	13	0.13	0.48
80	[79.95, 80.95>	11	0.11	0.59
81	[80.95, 81.95>	14	0.14	0.73
82	[81.95, 82.95>	11	0.11	0.84
83	[82.95, 83.95>	6	0.06	0.90
84	[83.95, 84.95>	4	0.04	0.94
85	[84.95, 85.95>	1	0.01	0.95
86	[85.95, 86.95>	0	0.00	0.95
87	[86.95, 87.95>	1	0.01	0.96
88	[87.95, 88.95>	1	0.01	0.97
89	[88.95, 89.95>	3	0.03	1.00

FIGURA 8

<i>clase</i>	<i>límites</i>	<i>frecuencia</i>	<i>Frecuencia relativa</i>	<i>Frecuencia acumulada</i>
54	[53, 25, 54, 25 >	1	0, 01	0, 01
55	[54, 25, 55, 25 >	3	0, 03	0, 04
56	[55, 25, 56, 25 >	2	0, 02	0, 06
57	[56, 25, 57, 25 >	2	0, 02	0, 08
58	[57, 25, 58, 25 >	5	0, 05	0, 13
59	[58, 25, 59, 25 >	5	0, 05	0, 18
60	[59, 25, 60, 25 >	10	0, 10	0, 28
61	[60, 25, 61, 25 >	8	0, 08	0, 36
62	[61, 25, 62, 25 >	3	0, 03	0, 39
63	[62, 25, 63, 25 >	7	0, 07	0, 46
64	[63, 25, 64, 25 >	10	0, 10	0, 56
65	[64, 25, 65, 25 >	9	0, 09	0, 65
66	[65, 25, 66, 25 >	7	0, 07	0, 72
67	[66, 25, 67, 25 >	9	0, 09	0, 81
68	[67, 25, 68, 25 >	3	0, 03	0, 84
69	[68, 25, 69, 25 >	7	0, 07	0, 91
70	[69, 25, 70, 25 >	6	0, 06	0, 97
71	[70, 25, 71, 25 >	3	0, 03	1, 00

FIGURA 9

<i>clase</i>	<i>Límites</i>	<i>frecuencia</i>	<i>Frecuencia relativa</i>	<i>Frecuencia acumulada</i>
37	[36.95, 37.95>	2	0.02	0.02
38	[37.95, 38.95>	1	0.01	0.03
39	[38.95, 39.95>	1	0.01	0.04
40	[39.95, 40.95>	3	0.03	0.07
41	[40.95, 41.95>	5	0.05	0.12
42	[41.95, 42.95>	5	0.05	0.17
43	[42.95, 43.95>	4	0.04	0.21
44	[43.95, 44.95>	5	0.05	0.26
45	[44.95, 45.95>	5	0.05	0.31
46	[45.95, 46.95>	9	0.09	0.40
47	[46.95, 47.95>	9	0.09	0.49
48	[47.95, 48.95>	9	0.09	0.58
49	[48.95, 49.95>	7	0.07	0.65
50	[49.95, 50.95>	8	0.08	0.73
51	[50.95, 51.95>	6	0.06	0.79
52	[51.95, 52.95>	2	0.02	0.81
53	[52.95, 53.95>	3	0.03	0.84
54	[53.95, 54.95>	2	0.02	0.86
55	[54.95, 55.95>	5	0.05	0.91
56	[55.95, 56.95>	2	0.02	0.93
57	[56.95, 57.95>	1	0.01	0.94
58	[57.95, 58.95>	2	0.02	0.96
59	[58.95, 59.95>	4	0.04	1.00

FIGURA 10

<i>clase</i>	<i>Límites</i>	<i>frecuencia</i>	<i>Frecuencia relativa</i>	<i>Frecuencia acumulada</i>
94	[93.95, 94.95>	35	0.35	0.35
95	[94.95, 95.95>	65	0.65	1.00

FIGURA 11

<i>clase</i>	<i>Límites</i>	<i>frecuencia</i>	<i>Frecuencia relativa</i>	<i>Frecuencia acumulada</i>
94	[93.95, 94.95>	59	0.59	0.59
95	[94.95, 95.95>	41	0.41	1.00

FIGURA 12

<i>clase</i>	<i>Límites</i>	<i>frecuencia</i>	<i>Frecuencia relativa</i>	<i>Frecuencia acumulada</i>
94	[93.95, 94.95>	69	0.69	0.69
95	[94.95, 95.95>	30	0.30	0.99
96	[95.96, 96.95>	1	0.01	1.00

FIGURA 13

<i>clase</i>	<i>Límites</i>	<i>frecuencia</i>	<i>Frecuencia relativa</i>	<i>Frecuencia acumulada</i>
93	[92.95, 93.95>	6	0.06	0.06
94	[93.95, 94.95>	61	0.61	0.67
95	[94.95, 95.95>	32	0.32	0.99
96	[95.95, 96.95>	1	0.01	1.00

FIGURA 14

<i>clase</i>	<i>Límites</i>	<i>frecuencia</i>	<i>Frecuencia relativa</i>	<i>Frecuencia acumulada</i>
93	[92.95, 93.95>	13	0.13	0.13
94	[93.95, 94.95>	61	0.61	0.74
95	[94.95, 95.95>	26	0.26	1.00

FIGURA 15

<i>clase</i>	<i>Límites</i>	<i>frecuencia</i>	<i>Frecuencia relativa</i>	<i>Frecuencia acumulada</i>
92	[91.95, 92.95 >	1	0.01	0.01
93	[92.95, 93.95 >	31	0.31	0.32
94	[93.95, 94.95 >	54	0.54	0.86
95	[94.95, 95.95 >	14	0.14	1.00

FIGURA 16

<i>clase</i>	<i>Límites</i>	<i>frecuencia</i>	<i>Frecuencia relativa</i>	<i>Frecuencia acumulada</i>
92	[91.95, 92.95 >	27	0.27	0.27
93	[92.95, 93.95 >	50	0.50	0.77
94	[93.95, 94.95 >	16	0.16	0.93
95	[94.95, 95.95 >	6	0.06	0.99
96	[95.95, 96.95 >	1	0.01	1.00

FIGURA 17

<i>clase</i>	<i>Límites</i>	<i>frecuencia</i>	<i>Frecuencia relativa</i>	<i>Frecuencia acumulada</i>
81	[80.95, 81.95 >	1	0.01	0.01
82	[81.95, 82.95 >	2	0.02	0.03
83	[82.95, 83.95 >	1	0.01	0.04
84	[83.95, 84.95 >	4	0.04	0.08
85	[84.95, 85.95 >	4	0.08	0.16
86	[85.95, 86.95 >	4	0.04	0.20
87	[86.95, 87.95 >	5	0.05	0.25
88	[87.95, 88.95 >	7	0.07	0.32
89	[88.95, 89.95 >	4	0.04	0.36
90	[89.95, 90.95 >	14	0.14	0.50
91	[90.95, 91.95 >	11	0.11	0.61
92	[91.95, 92.95 >	15	0.15	0.76
93	[92.95, 93.95 >	8	0.08	0.84
94	[93.95, 94.95 >	11	0.11	0.95
95	[94.95, 95.95 >	5	0.05	1.00

FIGURA 18

clase	Límites	Frecuencia	Frecuencia relativa	Frecuencia acumulada
67	[66.95, 67.95>	1	0.01	0.01
68	[67.95, 68.95>	0	0.00	0.01
69	[68.95, 69.95>	2	0.02	0.03
70	[69.95, 70.95>	1	0.01	0.04
71	[70.95, 71.95>	0	0.00	0.04
72	[71.95, 72.95>	1	0.01	0.05
73	[72.95, 73.95>	0	0.00	0.05
74	[73.95, 74.95>	1	0.01	0.06
75	[74.95, 75.95>	1	0.01	0.07
76	[75.95, 76.95>	4	0.04	0.11
77	[76.95, 77.95>	3	0.03	0.14
78	[77.95, 78.95>	4	0.04	0.18
79	[78.95, 79.95>	4	0.04	0.22
80	[79.95, 80.95>	5	0.05	0.27
81	[80.95, 81.95>	7	0.07	0.34
82	[81.95, 82.95>	8	0.08	0.42
83	[82.95, 83.95>	11	0.11	0.53
84	[83.95, 84.95>	6	0.06	0.59
85	[84.95, 85.95>	9	0.09	0.68
86	[85.95, 86.95>	4	0.04	0.72
87	[86.95, 87.95>	10	0.10	0.82
88	[87.95, 88.95>	5	0.05	0.87
89	[88.95, 89.95>	1	0.01	0.88
90	[89.95, 90.95>	4	0.04	0.92
91	[90.95, 91.95>	3	0.03	0.95
92	[91.95, 92.95>	2	0.02	0.97
93	[92.95, 93.95>	1	0.01	0.98
94	[93.95, 94.95>	2	0.02	1.00

FIGURA 19

clase	Límites	Frecuencia	Frecuencia relativa	Frecuencia acumulada
60	[59.95, 60.95>	1	0.01	0.01
61	[60.95, 61.95>	0	0.00	0.01
62	[61.95, 62.95>	0	0.00	0.01
63	[62.95, 63.95>	0	0.00	0.01
64	[63.95, 64.95>	0	0.00	0.01
65	[64.95, 65.95>	2	0.02	0.03
66	[65.95, 66.95>	1	0.01	0.04
67	[66.95, 67.95>	1	0.01	0.05
68	[67.95, 68.95>	2	0.02	0.07
69	[68.95, 69.95>	2	0.02	0.09
70	[69.95, 70.95>	1	0.01	0.10
71	[70.95, 71.95>	2	0.02	0.12
72	[71.95, 72.95>	2	0.02	0.14
73	[72.95, 73.95>	3	0.03	0.17
74	[73.95, 74.95>	3	0.03	0.20
75	[74.95, 75.95>	10	0.10	0.30
76	[75.95, 76.95>	6	0.06	0.36
77	[76.95, 77.95>	6	0.06	0.42
78	[77.95, 78.95>	7	0.07	0.49
79	[78.95, 79.95>	10	0.10	0.59
80	[79.95, 80.95>	2	0.02	0.61
81	[80.95, 81.95>	6	0.06	0.67
82	[81.95, 82.95>	7	0.07	0.74
83	[82.95, 83.95>	7	0.07	0.81
84	[83.95, 84.95>	6	0.06	0.87
85	[84.95, 85.95>	2	0.02	0.89
86	[85.95, 86.95>	2	0.02	0.91
87	[86.95, 87.95>	3	0.03	0.94
88	[87.95, 88.95>	3	0.03	0.97
89	[88.95, 89.95>	2	0.02	0.99
90	[89.95, 90.95>	1	0.01	1.00

FIGURA 20

Clase	Límites	Frecuencia	Frecuencia relativa	Frecuencia acumulada
37	[36.95, 37.95 >	1	0.01	0.01
38	[37.95, 38.95 >	0	0.00	0.01
39	[38.95, 39.95 >	0	0.00	0.01
40	[39.95, 40.95 >	1	0.01	0.02
41	[40.95, 41.95 >	0	0.00	0.02
42	[41.95, 42.95 >	0	0.00	0.02
43	[42.95, 43.95 >	0	0.00	0.02
44	[43.95, 44.95 >	0	0.00	0.02
45	[44.95, 45.95 >	0	0.00	0.02
46	[45.95, 46.95 >	2	0.02	0.04
47	[46.95, 47.95 >	1	0.01	0.05
48	[47.95, 48.95 >	0	0.00	0.05
49	[48.95, 49.95 >	1	0.01	0.06
50	[49.95, 50.95 >	1	0.01	0.07
51	[50.95, 51.95 >	0	0.00	0.07
52	[51.95, 52.95 >	1	0.01	0.08
53	[52.95, 53.95 >	2	0.02	0.10
54	[53.95, 54.95 >	5	0.05	0.15
55	[54.95, 55.95 >	3	0.03	0.18
56	[55.95, 56.95 >	1	0.01	0.19
57	[56.95, 57.95 >	3	0.03	0.22
58	[57.95, 58.95 >	2	0.02	0.24
59	[58.95, 59.95 >	4	0.04	0.28
60	[59.95, 60.95 >	1	0.01	0.29
61	[60.95, 61.95 >	3	0.03	0.32
62	[61.95, 62.95 >	3	0.03	0.35
63	[62.95, 63.95 >	7	0.07	0.42
64	[63.95, 64.95 >	10	0.10	0.52
65	[64.95, 65.95 >	5	0.05	0.57
66	[65.95, 66.95 >	8	0.08	0.65
67	[66.95, 67.95 >	3	0.03	0.68
68	[67.95, 68.95 >	6	0.06	0.74

FIGURA 20 (CONTINUACIÓN).

69	[68.95, 69.95 >	3	0.03	0.77
70	[69.95, 70.95 >	4	0.04	0.81
71	[70.95, 71.95 >	3	0.03	0.84
72	[71.95, 72.95 >	3	0.03	0.87
73	[72.95, 73.95 >	5	0.05	0.92
74	[73.95, 74.95 >	4	0.04	0.96
75	[74.95, 75.95 >	0	0.00	0.96
76	[75.95, 76.95 >	1	0.01	0.97
77	[76.95, 77.95 >	0	0.00	0.97
78	[77.95, 78.95 >	0	0.00	0.97
79	[78.95, 79.95 >	3	0.03	1.00

FIGURA 21

Clase	Límites	Frecuencia	Frecuencia relativa	Frecuencia acumulada
14	[13.95, 14.95 >	1	0.01	0.01
15	[14.95, 15.95 >	0	0.00	0.01
16	[15.95, 16.95 >	2	0.02	0.03
17	[16.95, 17.95 >	1	0.01	0.04
18	[17.95, 18.95 >	2	0.02	0.06
19	[18.95, 19.95 >	0	0.00	0.06
20	[19.95, 20.95 >	1	0.01	0.07
21	[20.95, 21.95 >	2	0.02	0.09
22	[21.95, 22.95 >	0	0.00	0.09
23	[22.95, 23.95 >	3	0.03	0.12
24	[23.95, 24.95 >	5	0.05	0.17
25	[24.95, 25.95 >	7	0.07	0.24
26	[25.95, 26.95 >	7	0.07	0.31
27	[26.95, 27.95 >	7	0.07	0.38
28	[27.95, 28.95 >	2	0.02	0.40
29	[28.95, 29.95 >	7	0.07	0.47
30	[29.95, 30.95 >	4	0.04	0.51
31	[30.95, 31.95 >	3	0.03	0.54
32	[31.95, 32.95 >	1	0.01	0.55
33	[32.95, 33.95 >	1	0.01	0.56
34	[33.95, 34.95 >	3	0.03	0.59

FIGURA 21 (CONTINUACIÓN).

35	[34.95, 35.95 >	2	0.02	0.61
36	[35.95, 36.95 >	3	0.03	0.64
37	[36.95, 37.95 >	0	0.00	0.64
38	[37.95, 38.95 >	2	0.02	0.66
39	[38.95, 39.95 >	3	0.03	0.69
40	[39.95, 40.95 >	7	0.07	0.76
41	[40.95, 41.95 >	2	0.02	0.78
42	[41.95, 42.95 >	3	0.03	0.81
43	[42.95, 43.95 >	1	0.01	0.82
44	[43.95, 44.95 >	1	0.01	0.83
45	[44.95, 45.95 >	1	0.01	0.84
46	[45.95, 46.95 >	2	0.02	0.86
47	[46.95, 47.95 >	3	0.03	0.89
48	[47.95, 48.95 >	2	0.02	0.91
49	[48.95, 49.95 >	2	0.02	0.93
50	[49.95, 50.95 >	0	0.00	0.93
51	[50.95, 51.95 >	4	0.04	0.97
52	[51.95, 52.95 >	2	0.02	0.99
53	[52.95, 53.95 >	0	0.00	0.99
54	[53.95, 54.95 >	0	0.00	0.99
55	[54.95, 55.95 >	0	0.00	0.99
56	[55.95, 56.95 >	0	0.00	0.99
57	[56.95, 57.95 >	0	0.00	0.99
58	[57.95, 58.95 >	1	0.01	1.00

INDICE POR TEMAS

A

AGRUPADOS, DATOS . . . , 19

AGRUPAMIENTO (CLUSTER) JERARQUICO, 9

ALEATORIA, MUESTRA . . . , II

ANALISIS

DE COVARIANZA, 11

DISCRIMINANTE, 16

DE FACTORES, 14, 16

**DE REGRESION, II, IV, 9, 10, 12, 16, 29, 30, 31, 35, 36, 37, 67
70, 138**

DE VARIANZA, 11

ARGUMENTOS BAYESIANOS, 20, 21

**AUTOREGRESIVOS, MODELOS . . . , 15
(VER SERIES DE TIEMPO)**

B

BAYES, TEOREMA DE . . . , 23, 50

BAYESIANA

ESCUELA ESTADISTICA . . . , IV

ESTADISTICA . . . , 20, 24

INFERENCIA . . . , 20

METODOLOGIA . . . , 71

REGRESION . . . , 35- 72

SOLUCION . . . , 21, 26, 73

BAYESIANO, PROCEDIMIENTO PARA TRATAR OBSERVACIONES INCOMPLETAS, 70
BAYESIANOS, ARGUMENTOS . . . , 20, 21
BERKELEY, ESCUELA ESTADISTICA DE . . . , 3
BINARIOS, DATOS . . . , 17

C

CATEGORICOS ORDINALES, DATOS . . . , 18
CAUSAS DE EXISTENCIA DE DATOS FALTANTES, III
CENSURADAS, MUESTRAS . . . , 19, 33, 34
CLASICA, ESCUELA ESTADISTICA . . . , IV, 3
ESTADISTICA . . . , 2
CLASICOS, PROCEDIMIENTOS . . . , 2, 8-19
COEFICIENTE DE DETERMINACION, 77-93, 134-145
COMBINACIONES LINEALES DE MEDIAS, 15
COMPORTAMIENTO A LA LARGA, 3
CONDICIONAL, PROBABILIDAD . . . , 22
CONCEPTOS DE LA ESTADISTICA, I
SISTEMA DE . . . , I
CONFIANZA, 3
CONJUGADA DE DISTRIBUCIONES, FAMILIA . . . , 33, 35, 36, 62, 72, 74
CONOCIMIENTO, MEDIDA DE . . . , 22
CONSISTENCIA, 3
CONTRASTE DE HIPOTESIS, 3, 18, 99
COVARIANZA, ANALISIS DE . . . , 11
CUADRADOS, MINIMOS . . . , 13, 16, 38
CURVAS DE SUPERVIVENCIA, ESTIMACION DE . . . , 34

D

DATOS 4.21**AGRUPADOS. 19****BINARIOS MULTIVARIADOS. 17****CENSURADOS. 19****FALTANTES. III****CAUSAS DE QUE EXISTAN. . . . III****PERDIDA ALEATORIA DE 7. 27. 37****PROCESO GENERADOR DE 5-8. 24-28****RECUPERACION DE III****INCOMPLETOS. III. 2****PERDIDOS. III****DECISION. I****ESPACIO DE 24****PROBLEMA DE II****TEORIA DE 24****DENSIDAD. FUNCION DE 4. 22. 23****INICIAL. 24. 26. 28. 32. 39. 42****APROXIMACIONES. 33****DESCARTAR OBSERVACIONES INCOMPLETAS. 9****DETERMINACION. COEFICIENTE DE 77-93. 138-145****DIFERENCIAS DE****MEDIAS. 18. 33****UARIANZAS. 18****DIRICHLET. DISTRIBUCION INICIAL. . . . 34****DISCRIMINANTE. ANALISIS DE 16****DISTANCIA ESPERADA. 3**

EXPONENCIAL.33

INICIAL DIRICHLET.34

MULTINOMIAL.33

DISTRIBUCIONES.FAMILIA CONJUGADA DE...33,35,36,62,72,74

E

EFICIENCIA.3.10

ESCUELA ESTADISTICA

BAYESIANA.IV.1

CLASICA.IV.1.3

DE BERKELEY.3

DE MUESTRAS.3

ESTANDAR.3

FRECIENTISTA.3

ESPACIO DE DECISIONES.24

ESTADISTICA.I

BAYESIANA.20.24

CLASICA.2

CONCEPTOS DE LA...I

MARCO TEORICO DE LA...I

METODO DE LA...I

SISTEMA DE CONCEPTOS DE LA...I

ESTIMACION.15,16,29,99

DE CURVAS DE SUPERVIVENCIA.34

DE VALORES FALTANTES.14

GAUSS-MARKOV.18

INTERVALO.3

PUNTUAL.3

ESTIMAR ELEMENTOS DE $X'X$, 10
ETAPAS DEL METODO ESTADISTICO, 1
EXPERIMENTOS, DISEÑO DE . . . 9, 74, 75
 DESBALANCEADOS, 9, 75
EXPONENCIAL, DISTRIBUCION . . . 33

170.

F

FALTANTES

CAUSAS DE QUE EXISTAN DATOS . . . III

DATOS . . . III

ORIGEN DE DATOS . . . 5

PERDIDA ALEATORIA DE DATOS . . . 7, 27, 37

PROCESO GENERADOR DE DATOS . . . 5-8, 24-28

FACTORES, ANALISIS DE . . . 14, 16

FACTORIZACION DE LA VEROSIMILITUD, 16, 17, 32, 36

FAMILIA CONJUGADA DE DISTRIBUCIONES, 33, 35, 36, 62, 72, 74

FICTICIAS

VARIABLES . . . 11

OBSERVACIONES . . . 75

FISHER, INFORMACION DE . . . 33

FRECUENCIA, PROBABILIDAD COMO LIMITE DE . . . 3

FRECUENTISTA, ESCUELA ESTADISTICA . . . 3

FRIEDMAN, PRUEBA DE . . . 18

FUNCION

DE DENSIDAD, 4, 22, 23

 APROXIMACION A LA . . . 33

DE UTILIDAD, 24

DE VEROSIMILITUD, 5

GAUSS-MARKOU, ESTIMADORES. . . . 18
GENERALIZADA, INVERSA. . . . 39, 53, 60, 146-147
GRADOS DE LIBERTAD. 68, 70

H

HIPOTESIS, CONTRASTES DE. . . . 3, 18, 99

I

INCOMPLETOS, DATOS. . . . III, 2
IGNORAR OBSERVACIONES INCOMPLETAS. 9
INICIAL
 DE REFERENCIA, DISTRIBUCION. . . . 28, 32, 42, 44
 DIRICHLET, 34
 DISTRIBUCION. . . . 24, 26, 28, 32, 39, 42, 45, 61, 62, 63, 72
INFERENCIA BAYESIANA, 28
INFORMACION SUBJETIVA, 28, 62
INSEGAMIENTO. 18
INTEGRACION NUMERICA, 30, 32, 33
INTERCAMBIABLES, OBSERVACIONES. . . . 62
INVARIANZA ANTE INVERSA GENERALIZADA, 41, 42, 60

L

LARGA, COMPORTAMIENTO A LA 3
LIBERTAD, GRADOS DE 68, 70
LINEAL, REGRESION. . . . (VER REGRESION)
LINEALES, COMBINACIONES. . . . 15

M

MARCO TEORICO DE LA ESTADISTICA. I**MAXIMA VEROSIMILITUD. 15. 29****MEDIAS****COMBINACIONES LINEALES DE... 15****DIFERENCIA DE... 18****MEIDAS****DE CONTAR. 4. 22****DE LEBESQUE. 4. 22****DOMINANTES. 4****METODO CIENTIFICO. I****ESTADISTICO. 1****MINIMOS CUADRADOS. 13. 16. 38****GENERALIZADOS. 13****MODELOS AUTORREGRESIVOS. 15
(VER SERIES DE TIEMPO)****MODELOS LINEALES. 16****SIN COORDENADAS. 18****MODIFICACIONES A LAS VARIANZAS. 14. 15****MUESTRA . I. 3. 4. 21****ALEATORIA. II****CENSURADA. 19. 33. 34****ESCUELA ESTADISTICA DE... 3****MULTICOLINEALIDAD. 6 7****MULTINOMIAL. DISTRIBUCION... 33****MULTIPLES. RESPUESTAS... 33****MULTIVARIADOS. DATOS BINARIOS... 17**

NIVEL DE SIGNIFICANCIA, 3

NO-PARAMETRICA, ESTADISTICA. . . . 5, 17, 18, 22, 34

NORMAL

COMBINACIONES LINEALES DE MEDIAS, 15

ECUACION. . . . 38

GAMMA, 35, 36, 37, 39, 72

DISTRIBUCIONES MARGINALES, 50

MULTIVARIADA, 15, 17, 18, 30, 31, 33, 38, 52

NUMERICA, INTEGRACION. . . . 30, 32, 33

O

OBSERVACION 1, 11, 4, 21

ABERRANTE (OUTLIERS), 75

FALTANTES, 11, 111

FICTICIAS, 75

INCOMPLETAS, 11, 111

OMITIR. . . . 9

DESCARTAR. . . . 9

IGNORAR. . . . 9

PERDIDAS, 111

TRUNCADAS, 19

VECTORIALES, 11

OBJETIVOS REGRESION MULTIPLE, 76

OMITIR OBSERVACIONES INCOMPLETAS, 9

ORDENADA AL ORIGEN, 31, 70, 73

ORDINALES, DATOS CATEGORICOS. . . . 18

ORIGEN DE DATOS FALTANTES, 5

ORTODOXA, ESCUELA ESTADISTICA. . . . 3

PARAMETRO.4
PARAMETROS DISTINTOS.27.37
PARCELAS DIVIDIDAS.11
PERDIDA ALEATORIA DE DATOS FALTANTES.7.27.37
PERDIDOS .OBSERVACIONES....III
PERDIDOS.DATOS....III
POTENCIA DE LA PRUEBA.4
POBLACION.I.II
PREDICTIVA.DISTRIBUCION....76
PRINCIPIO DE LA INFORMACION PERDIDA.15.17.19
PROBABILIDAD
 CONCEPTUALIZACION DE LA....22
 CONDICIONAL.22
 FRECUENTISTA.3
 MEDIDAS DE....4
 SUPJETIVA.22.24
PROBIT MULTIPLE.16
PROCEDIMIENTOS CLASICOS.2.8-19
 DE INFERENCIA.6
PROCESO GENERADOR DE DATOS FALTANTES.7.27.37
PROBLEMA DE DECISION.II
PRUEBA DE FRIEDMAN.18

R

RECUPERACION DE DATOS FALTANTES.III
REFERENCIA.DISTRIBUCIONES INICIALES DE....28.32.42.44
REGRESION.ANALISIS DE....II.IV.9.10.12.16.29.30.31.35.36.37.61.70
 72.138
 OBJETIVOS DEL....76

REGRESIONES AUXILIARES. 13

RESIDUOS. 11

175.

RESPUESTAS MULTIPLES. 33

RESTRINGIDO. DISEÑO. . . . 19

S

SECUENCIAL. IV. 35. 72

SELECCION DE VARIABLES. 16, 77, 138-145

SERIES DE TIEMPO. 15
(VER AUTORREGRESIVO)

SIGNIFICANCIA. NIVEL DE. . . . 3

SIMULACION. 10. 16

SOLUCION BAYESIANA. 21. 26

VENTAJAS DE LA. . . . 73

SUBJETIVA. PROBABILIDAD. . . . 22. 24

SUPERVIUENCIA. ESTIMACION DE CURVAS DE. . . . 34

T

TEORIA DE DECISIONES. 24

AXIOMAS DE LA. . . . 24

CRITERIOS DE CONSISTENCIA Y COHERENCIA DE LA. . . . 24

TEOREMA DE BAYES. 23. 50

TIEMPO. SERIES DE. . . . 15
(VER AUTORREGRESIVO)

TOMA DE DECISIONES. I

U

UNIVERSO. II

UTILIDAD. FUNCION DE. . . . 24

VARIABLE . I**FICTICIAS. 11****SELECCION DE... 16, 77, 138-145****VARIANZA****ANALISIS DE LA... 11****DIFERENCIAS DE... 18****MODIFICACIONES A LA... 14, 15****VENTAJAS DE LA SOLUCION BAYESIANA. 73****VEROSIMILITUD. 3, 28, 38****FACTORIZACION DE LA... 16, 17, 32, 36****FUNCION DE... 5, 23****MAXIMA... 15**

INDICE POR AUTORES

AFIFI, A. A. , 11, 15

AITCHISON, J. , 17

AITKEN, C. G. G. , 17

ANDERSON, T. H. , 15, 17

BAKER, F. B. , 9

BARNETT, U. , 3

BARTLETT, M. S. , 11

BEALE, E. M. L. , 12, 13, 15, 141

BERNARDO, J. M. , 28, 63

BHOJ, D. S. , 18

BLUMENTHAL, S. , 15, 16

BOX, G. E. P. , 21, 78, 95

BREDERADE, H. K. , 14

BROWN, H. B. , 11

BUCK, S. F. , 12, 13, 14

COX, G. M. , 11

COCHRAN, W. G. , 11

CHOU, Y. L. , II

CREASON, J. P. , 15, 16

DAGENAIS, H. G. , 12, 13, 29, 30, 31, 32, 36, 78, 126, 131, 132, 135, 136, 138

DAHIYA, R. C. , 18

DE FINETTI, B. , 22

DE GROOT, H. H. , 21, 29, 47, 99

DENPSTER, A. P. , 15, 16, 19

DRAPER, H. R. , 16, 60, 138

DRYGAS, H. , 18

EATON, H. L. , 18

EKBOHM, G. , 18

ELASHOFF, R. M. , 11, 15

FERGUSON, T. S. , 34

FISHBURN, P. C. , 24

FISHER, R. A. , 3

FREUND, R. J. , 9

GLASSER, M. , 18

GRAYBILL, F. , 38

GURLAND, J. , 18

HAIKOUSKY, Y. , 9, 18

HANNERS, H. E. , 14

HARRISON, P. J. , 36

HARTLEY, H. O. , 141

HASSELBLAD, U. , 15, 16

HEALY, H. J. R. , 11, 71

HOCKING, R. R. , 9, 17, 141

HOCKNEY, O. P. , 9

HOYLE, M. H. , 2

JARRET, R. G. , 12

179.

JOHN, J. A. , 11

JOHNSON, D. E. , 9

KENDALL, H. G. , 2, 9, 10, 14, 74

KLOTZ, J. , 18

KOCH, G. G. , 17

KORHAR, R. M. , 18

KRUSKAL, H. H. , 12, 13

LAIRD, H. M. , 29, 32, 15, 16, 19

LIGNY, C. L. , 14

LIN, PI-ERH. 18

LINDLEY, D. U. , 21, 94

LITTLE, R. J. A. , 12, 13, 15, 16, 141

LOUIS, T. A. , 29, 32

MENTA, J. S. , 18, 29, 33

MC DONALD, L. L. , 9

MILLIKEN, G. A. , 9

NITRA, S. K. , 39, 42

MULLER, M. E. , 78, 93

NAIK, U. D. , 18

NEYMAN, J. , 3

NIEUWDOORP, G. H. E. , 14

ORCHARD, T. , 15, 17

OVERALL, J. E. , 9

OXPRING, H. H. , 17

PEARSON, E. G. , 3

PRESCOTT, P. , 11

PRESS, S. J. , 29, 30, 36

RAIFFA, H. , 21, 37

RANSEY, F. P. , 24

RAO, C. R. , 15, 39, 42

RUBIN, D. B. , 5-7, 11-19, 23-28, 33, 72, 72, 140, 141, 144

SAMTHANAN, L. , 15, 16

SCOTT, A. , 29, 30, 36

SAVAGE, L. J. , 24

SEARLE, S. R. , 9

SHEARER, P. R. , 11

SCHLAIFER, R. , 21, 37

SMITH, H. , 16, 76, 138

SPEED, F. H. , 9

SPIEGEL, D. K. , 9

SORENSEN, J. P. , 29, 33

STEAD, A. G. , 15, 16

STEVENS, C. F. , 36

STIVERS, L. E. , 18

STEWARD, W. E. , 29, 33

STUART, A. , 2, 9, 14, 74

SUSARLA, U. , 33

SUNNY, P. A. U. B. , 29, 33

TIAO, G. C. , 21

TITTERINGTON, D. H. , 17

VAN HAUMELINGEN, J. C. , 14

VAN RYZIN, J. , 33

WILKINSON, G. N. , 12

WILKS, S. S. , 15

WINER, B. J. , 11

WESTMACOTT, M. , 11, 71

WOODBURY, M. A. , 15, 17

YATES, F. , 11, 13

BIBLIOGRAFIA

- AFIFI, A. A. & ELASHOFF, R. M. (1966): "MISSING OBSERVATIONS IN MULTI-VARIATE STATISTICS I. REVIEW OF THE LITERATURE". JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, U. 61, PP. 595-604.
- AFIFI, A. A. & ELASHOFF, R. M. (1967): "MISSING OBSERVATIONS IN MULTI-VARIATE STATISTICS II. POINT ESTIMATION IN SIMPLE LINEAR REGRESSION". JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, U. 62, PP.
- AFIFI, A. A. & ELASHOFF, R. M. (1969A): "MISSING OBSERVATIONS IN MULTI-VARIATE STATISTICS III. LARGE SAMPLE ANALYSIS OF SIMPLE LINEAR REGRESSION". JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, U. 64, PP. 337-358.
- AFIFI, A. A. & ELASHOFF, R. M. (1969B): "MISSING OBSERVATIONS IN MULTI-VARIATE STATISTICS IV. A NOTE ON SIMPLE LINEAR REGRESSION". JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, U. 64, PP. 359-365.
- AITCHISON, J. & AITKEN, C. G. G. (1976): "MULTIVARIATE BINARY DISCRIMINATION BY THE KERNEL METHOD". BIOMETRIKA, U. 63, PP. 413-420.
- ANDERSON, T. W. (1957): "MAXIMUM LIKELIHOOD ESTIMATES FOR A MULTI-VARIATE NORMAL DISTRIBUTION WHEN SOME OBSERVATIONS ARE MISSING". JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, U. 52, PP. 200-203.
- BAKER, F. B. (1978): "SENSITIVITY OF THE COMPLETE-LINK CLUSTERING TECHNIQUE TO MISSING INDIVIDUALS". JOURNAL OF EDUCATIONAL STATISTICS, U. 3, PP. 233-252.
- BARNET, U. (1973): "COMPARATIVE STATISTICAL INFERENCE". WILEY, LONDON.
- BARTLETT, M. S. (1937): "SOME EXAMPLES OF STATISTICAL METHODS OF RESEARCH IN AGRICULTURE AND APPLIED BOTANY". JOURNAL OF THE ROYAL STATISTICAL SOCIETY, U. 4, PP. 137-170.
- BEALE, E. M. L. & LITT, E. R. J. A. (1975): "MISSING VALUES IN MULTI-VARIATE ANALYSIS". JOURNAL OF THE ROYAL STATISTICAL SOCIETY, B, U. 37, PP. 129-145.
- BERNARDI, J. M. (1979): "REFERENCE POSTERIOR DISTRIBUTIONS FOR BAYESIAN INFERENCE WITH DISCUSSION". JOURNAL OF THE ROYAL STATISTICAL SOCIETY, B, U. 41, PP. 113-147.
- BHOJ, D. S. (1979): "TESTING EQUALITY OF VARIANCES OF CORRELATED VARIATES WITH INCOMPLETE DATA ON BOTH RESPONSES". BIOMETRIKA, U. 66, PP. 681-683.

- FERGUSON, T. S. (1973): "A BAYESIAN ANALYSIS OF SOME NON-PARAMETRIC PROBLEMS". ANNALS OF STATISTICS, U. 1, PP. 209-230.
- FISHBURN, P. C. (1970): "UTILITY THEORY FOR DECISION MAKING". REPRINTED BY KRIEGER PU. CO., HUNTINGTON, NEW YORK, 1979.
- FISHBURN, P. C. (1981): "SUBJECTIVE EXPECTED UTILITY: A REVIEW OF NORMATIVE THEORIES". THEORY AND DECISION, U. 13, PP. 139-199.
- FREUND, R. J. (1980): "THE CASE OF THE MISSING CELL". AMERICAN STATISTICIAN, U. 34, PP. 94-98.
- GLASSER, M. (1964): "LINEAR REGRESSION ANALYSIS WITH MISSING OBSERVATIONS AMONG THE INDEPENDENT VARIABLES". JOURNAL OF THE AMERICAN ASSOCIATION, U. 59, PP. 834-844.
- GLEASON, T. C. & STAELIN, R. (1975): "A PROPOSAL FOR HANDLING MISSING DATA". PSYCHOMETRICA, U. 40, PP. 229-252.
- GRAYBILL, F. (1961): "AN INTRODUCTION TO LINEAR STATISTICAL MODELS". MC GRAW-HILL.
- HAITOUSKY, Y. (1968): "MISSING DATA IN REGRESSION ANALYSIS". JOURNAL OF THE ROYAL STATISTICAL SOCIETY, B, U. 30, PP. 67-82.
- HARRISON, P. J. & STEVENS, C. F. (1976): "BAYESIAN FORECASTING". JOURNAL OF THE ROYAL STATISTICAL SOCIETY, B, U. 38, PP. 205-247.
- HARTLEY, H. O. & HOCKING, R. R. (1971): "THE ANALYSIS OF INCOMPLETE DATA". BIOMETRICS, U. 27, PP. 783-804.
- HASSELBLAD, U., STEAD, A. G. & CREASON, J. P. (1980): "MULTIPLE PROBIT ANALYSIS WITH NON-ZERO BACKGROUND". BIOMETRICS, U. 36, PP. 655-663.
- HEALY, M. J. R. & WESTMACOTT, M. (1956): "MISSING VALUES IN EXPERIMENTS ANALYSED ON AUTOMATIC COMPUTERS". APPLIED STATISTICS, U. 5, PP. 203-206.
- HOCKING, R. R. & OXPRING, H. H. (1971): "MAXIMUM LIKELIHOOD ESTIMATION WITH INCOMPLETE MULTINOMIAL DATA". JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, U. 66, PP. 65-70.
- HOYLE, M. H. (1971): "SPOILT DATA - AN INTRODUCTION AND BIBLIOGRAPHY". JOURNAL OF THE ROYAL STATISTICAL SOCIETY, A, U. 134, PP. 429-439.
- JARRET, R. G. (1978): "THE ANALYSIS OF DESIGNED EXPERIMENTS WITH MISSING OBSERVATIONS". APPLIED STATISTICS, U. 24, PP. 190-192.
- JOHN, J. A. & PRESCOTT, P. (1975): "ESTIMATING MISSING VALUES IN EXPERIMENTS". APPLIED STATISTICS, U. 24, PP. 190-192.

- **BOX, G. E. P. & MULLER, M. E. (1958): "A NOTE ON THE GENERATION OF RANDOM NORMAL DEVIATES". ANNALS OF MATHEMATICAL STATISTICS, U. 29, PP. 610-611.**
- **BOX, G. E. P. & TIAO, G. C. (1973): "BAYESIAN INFERENCE IN STATISTICAL ANALYSIS". ADDISON-WESLEY PUB. CO., READING, MASS.**
- **BROWN, M. B. (1975): "EXPLORING INTERACTION EFFECTS IN THE ANOVA". APPLIED STATISTICS, U. 24, PP. 288-298.**
- **BUCK, S. F. (1960): "A METHOD OF ESTIMATION OF MISSING VALUES IN MULTIVARIATE DATA SUITABLE FOR USE WITH AN ELECTRONIC COMPUTER". JOURNAL OF THE ROYAL STATISTICAL SOCIETY, B, U. 22, PP. 302-306.**
- **COCHRAN, M. G. & COX, G. M. (1957): "EXPERIMENTAL DESIGNS". SECOND EDITION. WILEY, NEW YORK.**
- **CHOU, Y. L. (1969): "STATISTICAL ANALYSIS: WITH BUSINESS AND ECONOMICS APPLICATIONS". HOLT, RINEHART & WINSTON, NEW YORK.**
- **DAGENAIS, M. G. (1973): "THE USE OF INCOMPLETE OBSERVATIONS IN MULTIPLE REGRESSION ANALYSIS. A GENERALIZED LEAST SQUARES APPROACH". JOURNAL OF ECONOMETRICS, U. 1, PP. 317-328.**
- **DAGENAIS, M. G. (1974): "MULTIPLE REGRESSION ANALYSIS WITH INCOMPLETE OBSERVATIONS, FROM A BAYESIAN POINT OF VIEW". IN FIENBERG S. E. & ZELLNER, A. (EDS.): "STUDIES IN BAYESIAN ECONOMETRICS AND STATISTICS". NORTH-HOLLANDER PUB. CO.**
- **DAGENAIS, M. G. (1976): "INCOMPLETE OBSERVATIONS AND SIMULTANEOUS EQUATIONS MODELS". JOURNAL OF ECONOMETRICS, U. 4, PP. 231-242.**
- **DAHIYA, R. C. & CORNAR, R. M. (1980): "MAXIMUM LIKELIHOOD ESTIMATES FOR A BIVARIATE NORMAL DISTRIBUTION WITH MISSING DATA". ANNALS OF STATISTICS, U. 8, PP. 687-692.**
- **DE FINETTI, B. (1970): "THEORY OF PROBABILITY". 2 VOL. WILEY, NEW YORK.**
- **DE GROOT, M. H. (1970): "OPTIMAL STATISTICAL DECISIONS". MC GRAM HILL.**
- **DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977): "MAXIMUM LIKELIHOOD FROM INCOMPLETE DATA VIA THE EM ALGORITHM". JOURNAL OF THE ROYAL STATISTICAL SOCIETY, B, U. 39, PP. 1-37.**
- **DRAPER, N. R. & SMITH, H. (1968): "APPLIED REGRESSION ANALYSIS". WILEY, NEW YORK.**
- **DRYGAS, H. (1978): "GAUSS-MARKOV ESTIMATION FOR MULTIVARIATE LINEAR MODELS". ANNALS OF STATISTICS, U. 4, PP. 779-787.**
- **EATON, M. L. (1970): "GAUSS-MARKOV ESTIMATION FOR MULTIVARIATE LINEAR MODELS: A COORDINATE-FREE APPROACH". ANNALS OF MATHEMATICAL STATISTICS, U. 41, PP. 528-538.**
- **EKBOHM, G. (1976): "ON COMPARING MEANS IN THE PAIRED CASE WITH INCOMPLETE DATA ON BOTH RESPONSES". BIOMETRIKA, U. 63, PP. 299-304.**

- KENDALL, M. G. (1975): "MULTIVARIATE ANALYSIS". GRIFFIN PUB. CO. LONDON. PP. 104-107.
- KENDALL, M. G. & STUART, A. (1967): "THE ADVANCED THEORY OF STATISTICS". VOL. II, SECOND EDITION. GRIFFIN PUB. CO., LONDON.
- KENDALL, M. G. & STUART, A. (1968): "THE ADVANCED THEORY OF STATISTICS". VOL. III, GRIFFIN PUB. CO., LONDON.
- KENDALL, M. G. & STUART, A. (1969): "THE ADVANCED THEORY OF STATISTICS". VOL. I, THIRD EDITION, GRIFFIN PUB. CO., LONDON.
- KLOTZ, J. (1980): "A MODIFIED COCHRAN-FRIEDMAN TEST WITH MISSING OBSERVATIONS AND ORDERED CATEGORICAL DATA". BIOMETRICS, U. 36, PP. 665-670.
- KOCH, G. G., INREY, P. B. & REINFURT, D. W. (1972): "LINEAR MODEL ANALYSIS OF CATEGORICAL DATA WITH INCOMPLETE RESPONSE VECTORS". BIOMETRICS, U. 28, PP. 663-692.
- KRUSKAL, W. H. (1961): "THE COORDINATE FREE APPROACH TO GAUSS MARKOV ESTIMATION AND ITS APPLICATION TO MISSING AND EXTRA OBSERVATIONS". IN BERKELEY SYMPOSIUM, U. 1, PP. 435-451.
- LAIRD, N. M. & LOUIS, T. A. (1982): "APPROXIMATE POSTERIOR DISTRIBUTIONS FOR INCOMPLETE DATA PROBLEMS". JOURNAL OF THE ROYAL STATISTICAL SOCIETY, B, U. 44, PP. 190-200.
- LIGNY, C. L., NIEUWDRP, G. H. E., BREDERODE, W. K., HANNERS, W. E. & VAN HAUMELINGEN, J. C. (1981): "AN APPLICATION OF FACTOR ANALYSIS WITH MISSING DATA". TECHNOMETRICS, U. 23, PP. 91-96.
- LIN, PI-ERH & STIERS, L. E. (1975): "TESTING FOR EQUALITY OF MEANS WITH INCOMPLETE DATA ON ONE VARIABLE: A MONTE CARLO STUDY". JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, U. 70, PP. 190-193.
- LINDLEY, D. V. (1965): "INTRODUCTION TO PROBABILITY AND STATISTICS FROM A BAYESIAN VIEWPOINT". PART 1: PROBABILITY, PART 2: INFERENCE. CAMBRIDGE UNIVERSITY PRESS, CAMBRIDGE.
- LINDLEY, D. V. (1972): "BAYESIAN STATISTICS, A REVIEW". SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS (SIAM), PHILADELPHIA.
- LITTLE, R. J. A. (1976): "INFERENCE ABOUT MEANS FROM INCOMPLETE MULTIVARIATE DATA". BIOMETRIKA, U. 63, PP. 593-604.
- LITTLE, R. J. A. (1979): "MAXIMUM LIKELIHOOD INFERENCE FOR MULTIPLE REGRESSION WITH MISSING VALUES". JOURNAL OF THE ROYAL STATISTICAL SOCIETY, B, U. 41, PP. 76-87.
- MEHTA, J. S. & GURLAND, J. (1969): "SOME PROPERTIES AND APPLICATIONS OF A STATISTICS ARISING IN TESTING CORRELATION". ANNALS OF MATHEMATICAL STATISTICS, U. 40, PP. 1736-1745.
- MEHTA, J. S. & SHAMY, P. A. U. B. (1974): "BAYESIAN ANALYSIS OF A BIVARIATE DISTRIBUTION WHEN SOME OBSERVATIONS ARE MISSING". IN FIENBERG, S. E. & ZELLNER, A. (EDS.): "STUDIES IN BAYESIAN ECONOMETRICS AND STATISTICS". NORTH-HOLLANDER PUB. CO., AMSTERDAM.

- WILLIKEN, G. A. & JOHNSON, D. E. (1981): "ANALYSIS OF MESSY DATA". PREPARED FOR THE INSTITUTE OF PROFESSIONAL EDUCATION. WASHINGTON.
- WILLIKEN, G. A. & MC DONALD, L. L. (1976): "LINEAR MODELS AND THEIR MISSING OR INCOMPLETE DATA: A UNIFYING APPROACH". BIOMETRISCHE ZEITSCHRIFT, U. 18, PP. 381-386.
- NAIK, V. D. (1975): "ON TESTING EQUALITY OF MEANS OF CORRELATED VARIABLES WITH INCOMPLETE DATA". BIOMETRIKA, U. 62, PP. 615-622.
- ORCHARD, T. & WOODBURY, M. A. (1973): "A MISSING INFORMATION PRINCIPLE: THEORY AND APPLICATIONS". UI BERKELEY SYMPOSIUM ON MATHEMATICS AND STATISTICS, U. 1, PP. 697-715.
- OVERALL, J. E. & SPIEGEL, D. K. (1969): "CONCERNING LEAST SQUARES ANALYSIS OF EXPERIMENTAL DATA". PSYCHOLOGICAL BULLETIN, U. 72, PP. 311-322.
- PRESS, S. J. & SCOTT, A. (1974): "MISSING VARIABLES IN BAYESIAN REGRESSION". IN FIENBERG, S. E. & ZELLNER, A. (EDS.): "STUDIES IN BAYESIAN ECONOMETRICS AND STATISTICS". NORTH-HOLLANDER PUB. CO., AMSTERDAM.
- PRESS, S. J. & SCOTT, A. (1976): "MISSING VARIABLES IN BAYESIAN REGRESSION, II". JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, U. 71, PP. 366-369.
- RAIFFA, H. & SCHLAIFER, R. (1961): "APPLIED STATISTICAL DECISION THEORY". M. I. T. PRESS, MASS.
- RAMSEY, F. P. (1931): "TRUTH AND PROBABILITY". IN KYBURG, H. E. & SHOKLER, H. E. (EDS.) (1964): "STUDIES IN SUBJECTIVE PROBABILITY". WILEY, NEW YORK.
- RAO, C. R. (1952): "ADVANCED STATISTICAL METHODS IN BIOMETRIC RESEARCH". WILEY, NEW YORK.
- RAO, C. R. (1965): "LINEAR STATISTICAL INFERENCE AND ITS APPLICATIONS" SECOND EDITION. WILEY, NEW YORK.
- RAO, C. R. & MITRA, S. K. (1971): "GENERALIZED INVERSE OF MATRICES AND ITS APPLICATIONS". WILEY, NEW YORK.
- RUBIN, D. B. (1972): "A NON-ITERATIVE ALGORITHM FOR LEAST SQUARES ESTIMATION OF MISSING VALUES IN ANY ANALYSIS OF VARIANCE DESIGNS". APPLIED STATISTICS, U. 21, PP. 136-141.
- RUBIN, D. B. (1974): "CHARACTERIZING THE ESTIMATION OF PARAMETERS IN INCOMPLETE DATA PROBLEMS". JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, U. 69, PP. 467-474.
- RUBIN, D. B. (1976A): "COMPARING REGRESSIONS WHEN SOME PREDICTOR VARIABLES ARE MISSING". TECHNOMETRICS, U. 18, PP. 281-286.
- RUBIN, D. B. (1976B): "NON ITERATIVE LEAST SQUARES ESTIMATES, STANDARD ERRORS AND F-TESTS FOR ANALYSIS OF VARIANCE WITH MISSING DATA". JOURNAL OF THE ROYAL STATISTICAL SOCIETY, B, U. 38, PP. 278-274.

- RUBIN, D. B. (1976C): "INFERENCE AND MISSING DATA". BIOMETRIKA, U. 63, PP. 581-598.
- RUBIN, D. B. (1978): "A NOTE ON BAYESIAN LIKELIHOOD AND SAMPLING DISTRIBUTION INFERENCES". JOURNAL OF EDUCATIONAL STATISTICS, U. 3, PP. 189-201.
- SANATHANAN, L. & BLUMENTHAL, S. (1978): "THE LOGISTIC MODEL AND ESTIMATION OF LATENT STRUCTURE". JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, U. 73, PP. 794-799.
- SAUSAGE, L. J. (1954): "THE FOUNDATIONS OF STATISTICS". WILEY, NEW YORK.
- SEARLE, S. R. (1971): "LINEAR MODELS". WILEY, NEW YORK.
- SHEARER, P. R. (1973): "MISSING DATA IN QUANTITATIVE DESIGNS". APPLIED STATISTICS, U. 22, PP. 135-140.
- SPEED, F. M., HOCKING, R. R. & HOCKNEY, O. P. (1978): "METHODS OF ANALYSIS OF LINEAR MODELS WITH UNBALANCED DATA". JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, U. 73, PP. 105-112.
- STEWARD, W. E. & SORENSEN, J. P. (1981): "BAYESIAN ESTIMATION OF COMMON PARAMETERS FROM MULTIRESPONSE DATA WITH MISSING OBSERVATIONS". TECHNOMETRICS, U. 23, PP. 131-142.
- SUSARLA, U. & VAN RYZIN, J. (1976): "NON PARAMETRIC BAYESIAN ESTIMATION OF SURVIVAL CURVES FROM INCOMPLETE OBSERVATIONS". JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, U. 71, PP. 897-902.
- TITTERINGTON, D. M. (1977): "ANALYSIS OF INCOMPLETE MULTIVARIATE BINARY DATA BY THE KERNEL METHOD". BIOMETRIKA, U. 64, PP. 455-460.
- WILKINSON, G. N. (1958A): "ESTIMATION OF MISSING VALUES FOR THE ANALYSIS OF INCOMPLETE DATA". BIOMETRICS, U. 14, PP. 257-286.
- WILKINSON, G. N. (1958B): "THE ANALYSIS OF VARIANCE AND DERIVATION OF STANDARD ERRORS FOR INCOMPLETE DATA". BIOMETRICS, U. 14, PP. 360-384.
- MILKS, S. S. (1932): "MOMENTS AND DISTRIBUTIONS OF ESTIMATES OF POPULATION PARAMETERS FROM FRAGMENTARY SAMPLES". ANNALS OF MATHEMATICAL STATISTICS, U. 3, PP. 163-195.
- MINER, B. J. (1962): "STATISTICAL PRINCIPLES IN EXPERIMENTAL DESIGN". MC GRAM-HILL, NEW YORK.
- YATES, F. (1933): "THE ANALYSIS OF REPLICATED EXPERIMENTS WHEN THE FIELD RESULTS ARE INCOMPLETE". EMPIRE JOURNAL OF EXPERIMENTAL AGRICULTURE, U. 1, PP. 129-142.