

52

2EJ



UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE CIENCIAS

INTRODUCCION A LOS MODELOS
LINEALES BAYESIANOS

T E S I S

QUE PARA OBTENER EL TITULO DE

A C T U A R I O

P R E S E N T A N :

ADRIANA LOPEZ GARCIA

GABRIEL RUIZ ANTONIO



FACULTAD DE CIENCIAS
SECCION ESCOLAR

MEXICO, D. F.

JUNIO DE 1985

FALLA DE ORIGEN

TESIS CON
FALLA DE ORIGEN





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

M. en C. Virginia Abrín Baule
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis:

" Introducción a los Modelos Lineales Bayesianos "

realizado por Adriana López García y Gabriel Nuñez Antonio

con número de cuenta 9052086-0 y , pasante de la carrera de Actuaría
8715317-2

Dicho trabajo cuenta con nuestro voto aprobatorio.

Acreditados

Director de Tesis	M.en C. José Gabriel Huerta Gómez
Propietario	
Propietario	Act. José Salvador Zamora Muñoz
Propietario	M.en C. José Antonio Flores Díaz
Suplente	
Suplente	Act. María del Pilar Alonso Reyes
Suplente	Mat. Hugo Villaseñor Hernández

Gabriel Huerta Gómez
Zamora Muñoz J.S.
[Firma]
[Firma]

Consejo Departamental de Matemáticas
MAT. CESAR GUEVARA BRAVO

VIVIR

**Se un hombre útil más que un
hombre hábil,
honesto aunque no te vean.**

Se alguien que viva como piensa.

**Vivir no es sólo existir,
sino existir y crear,
saber sufrir y gozar,
y en vez de dormir, soñar.**

Descansar es un poquito morir.

A Dios.
Por todo cuanto tengo.

A mis padres. Pilar y Virgilio, quienes son lo máspreciado que tengo en esta vida, por su ejemplo, apoyo y confianza que siempre me han dado y sin cuya ayuda no podría haber alcanzado esta meta.

A mis hermanos. Ale, Chío, Lety y Luis quienes junto con mis padres son una de las cosas mas importantes en mi existencia y cuya amistad será siempre lo más valioso que pueda tener.

A mi Abuelo. A quien siempre recordare.

**A mis amigos y compañeros.
Gaby, Carlos, Jorge, Dante,
José, Chucho, Tofio y a todas
aquellas personas quienes de
una u otra forma han
contribuido para lograr esta
gran meta a través de su
amistad, conocimiento o
simplemente de un aliento en
los momentos cruciales de mi
vida.**

**A mi compañero de Tesis.
Gabriel, por su comprensión y
amistad.**

Adriana.

SI

*Si puedes estar firme cuando en tu derredor
todo el mundo se ofusca y tacha tu entereza;
si cuando dudan todos, fías en tu valor
y al mismo tiempo sabes excusar su flaqueza;
si puedes esperar y a tu afán poner brida,
o blanco de mentiras, esgrimir la verdad,
o siendo odiado, al odio no dejarle cabida
y ni ensalzas tu juicio ni ostentas tu bondad;*

*Si sueñas pero el sueño no se vuelve tu rey;
si piensas y el pensar no mengua tus ardores;
si el Triunfo y el Desastre no te imponen su ley
y los tratas lo mismo, como a dos impostores;
si puedes soportar que tu frase sincera
sea trampa de necios en boca de malvados,
o mirar hecha trizas tu adorada quimera
y tornar a forjarla con útiles mellados;*

*Si todas tus ganancias poniéndolas en un montón
las arriesgas osado en un golpe de azar,
y las pierdes, y luego con bravo corazón
sin hablar de tus pérdidas vuelves a comenzar;
si puedes mantener en la ruda pelea
alerta el pensamiento y el músculo tirante
para emplearlos cuando en ti todo flaquea,
menos la voluntad que te dice "Adelante";*

*Si entre la turba das a la virtud abrigo;
si marchando junto a reyes del orgullo has
triunfado;
si no pueden herirte amigo ni enemigo;
si eres bueno con todos, pero no demasiado;
y si puedes llenar los preciosos minutos
con sesenta segundos de combate bravío,
tuya es la Tierra y todos sus codiciados frutos,
y lo que mas importa, serás Hombre hijo mío.*

A mis padres.

Con la mayor gratitud por los esfuerzos realizados para que yo logrará terminar mi carrera profesional. Siendo para mi la mejor herencia.

A mi madre que es el ser más maravilloso del mundo, gracias por el apoyo moral, su cariño y comprensión que desde pequeño me a brindado, por guiar mi camino y estar siempre junto a mi en los momentos más difíciles.

A mi padre porque desde pequeño ha sido para mi un hombre grande y maravilloso al que siempre he admirado. Gracias por guiar mi vida con energía, esto es lo que ha hecho que sea lo que soy.

Gracias por todo lo que me han dado.

A mis hermanos.

Tania, Alejandro, Mauricio, Leonel e Ivan. Porque juntos formamos un gran equipo. Gracias por su esfuerzo y sacrificios compartidos.

A los amigos de toda una vida.

A Lulú y Emeterio, a Olivia P., a Olivia E., a Adriana, a Moisés, a Oscar, a Juan Luis, a Leonel, ... y a todos aquéllos que me han dado la oportunidad de conocer lo que es la verdadera amistad. Gracias, simplemente, por ser mis amigos.

A Angélica.

Por su cariño y comprensión.

A Dios.

Por darme la oportunidad de vivir y de ser mejor cada día.

¡Juntos lo hemos logrado!

Gabriel Nuñez Antonio

Quisiéramos externar nuestros agradecimientos a:

- **La Universidad Nacional Autónoma de México, en particular a la Facultad de Ciencias, por darnos la oportunidad de cursar nuestros estudios universitarios.**
- **A nuestros sinodales. Quienes participaron en la revisión de este trabajo, por sus valiosos comentarios, especialmente a nuestro director de tesis Gabriel H. cuya conducción hizo posible la realización de la misma.**
- **A Salvador Zamora por su gran interés en la revisión de este proyecto.**
- **Al Instituto de Investigaciones en Matemáticas Aplicadas (IIMAS) y a la Unidad Académica de los Ciclos Profesionales y de Posgrado del CCH, por permitirnos el uso de sus instalaciones y equipos para la elaboración de este material.**
- **A nuestros profesores y amigos de la Maestría en Estadística por sus valiosos comentarios y el apoyo mostrado durante el desarrollo del trabajo.**
- **Y finalmente a todos aquéllos que hacen posible la educación pública, en México.**

Indice

Introducción	3
1 Teoría de la Decisión y Estadística	7
1.1 Introducción	7
1.2 Introducción a la Teoría de la Decisión.	8
1.3 Problemas de Decisión Estadísticos	15
1.4 Predicción	24
2 El Modelo de Regresión de Rango Completo	31
2.1 Introducción	31
2.2 Planteamiento y Definición del Modelo	32
2.3 Análisis Bayesiano Conjugado	35
2.3.1 Estimación puntual de los parámetros del modelo	37
2.3.2 Intervalos de Probabilidad	42
2.3.3 Prueba de Hipótesis	47
2.4 Análisis Bayesiano no Informativo	52
2.4.1 Estimación puntual de los parámetros del modelo	55
2.4.2 Intervalos de Probabilidad	56
2.4.3 Prueba de Hipótesis	57
3 Predicción y Diagnóstico del Modelo de Regresión de Rango Completo	59
3.1 Introducción	59
3.2 Análisis Bayesiano Conjugado	60
3.2.1 Análisis Predictivo	60
3.2.2 Diagnóstico	62

3.3	Análisis Bayesiano no Informativo	71
3.3.1	Análisis Predictivo	71
3.3.2	Diagnóstico	72
4	Una Aplicación de Regresión Lineal Múltiple	77
4.1	Introducción	77
4.2	Planteamiento del problema	78
4.3	Análisis Bayesiano	79
4.3.1	Estimación puntual de los parámetros del modelo	80
4.3.2	Intervalos de Probabilidad	81
4.3.3	Prueba de Hipótesis	84
4.3.4	Diagnóstico	87
4.3.5	Predicción	94
5	Análisis Bayesiano de un Modelo de Diseño de Experi-	
	mentos	97
5.1	Introducción	97
5.2	Modelos de Diseño de Experimentos.	98
5.3	Planteamiento y Análisis Bayesiano de un modelo de diseño de experimentos.	102
5.3.1	Aplicación	111
	Conclusiones	121
	Bibliografía	123

Introducción

Dentro de los modelos estadísticos una clase importante, de gran diversidad, la constituyen los modelos lineales, dentro de los cuales se encuentran los modelos de series de tiempo, los modelos de diseño de experimentos y los modelos de regresión lineal, entre otros. La importancia de estos modelos se debe a su gran aplicación en los diferentes campos del conocimiento como la medicina, la economía, el control de calidad, etc. Lo anterior ya que una gran variedad de situaciones o fenómenos que se dan en estos campos se pueden modelar y analizar a través de los modelos lineales.

Dentro de la estadística existen dos grandes enfoques para el análisis de modelos estadísticos, el enfoque clásico y el enfoque bayesiano. Como los modelos lineales son una clase particular de modelos estadísticos pueden analizarse con ambos enfoques. Desde el punto de vista clásico, la teoría de los modelos lineales ha sido ampliamente discutida. Por el contrario, los métodos bayesianos para el análisis de los modelos lineales tienen un desarrollo más reciente.

El enfoque bayesiano, a diferencia del enfoque clásico, se caracteriza por la actualización de la información inicial que se tenga sobre los parámetros desconocidos. Dicha actualización se logra al incorporar en el análisis la información que proporciona una muestra aleatoria. Esta incorporación de información muestral como se verá se logra a través de la aplicación del teorema de Bayes.

Dentro de la teoría bayesiana existen dos formas relevantes de llevar a cabo el análisis de problemas estadísticos. Una es sin considerar y la otra es considerando los elementos de la teoría de la decisión, esta última permite combinar la información inicial y muestral con algunos otros aspectos importantes en el problema; por ejemplo, incorporar

información sobre las posibles consecuencias que ocurran al tomar cierta decisión, las cuales pueden ser cuantificadas mediante la especificación de las llamadas funciones de utilidad.

En este trabajo se presenta el material correspondiente al análisis bayesiano, considerando teoría de la decisión, del modelo de regresión lineal múltiple de rango completo, y se introducen las bases para el análisis de los modelos de diseño de experimentos, modelos de rango incompleto. El material de esta tesis está organizado de la siguiente manera.

En el capítulo 1 se presentan los conceptos básicos de la estadística bayesiana y de la teoría de la decisión, que permitan ubicar un problema de inferencia como un problema de decisión en ambiente de incertidumbre, y que además sirva como apoyo a los desarrollos de los siguientes capítulos.

En el capítulo 2 se define el modelo de regresión lineal múltiple de rango completo y se muestra la forma de obtener las estimaciones puntuales y por intervalos correspondientes a los parámetros desconocidos del modelo. En este capítulo también se presenta el problema de prueba de hipótesis sobre los parámetros de regresión. Para el desarrollo de los temas involucrados se proponen dos formas de análisis, una considerando que se cuenta con información inicial sobre los parámetros involucrados y otra en la que no se considera este tipo de información.

Dentro del análisis del modelo de regresión lineal múltiple es importante determinar si éste cumple con los supuestos requeridos y qué tan bien se ajusta el modelo a los datos. Así como, el poder predecir valores futuros del fenómeno en estudio, para ello en el capítulo 3 se desarrollan los tópicos correspondientes a diagnóstico y predicción del modelo de regresión lineal múltiple.

Como una forma de ejemplificar los conceptos presentados en los tres primeros capítulos, en el capítulo 4 se muestra una aplicación práctica sobre el modelo de regresión lineal múltiple de rango completo.

En el capítulo 5 se muestran algunos de los modelos de diseño de experimentos como variantes de los modelos de regresión de rango completo, y se presentan las bases para el análisis bayesiano del modelo de diseño de experimentos con dos factores aditivos. Finalmente, para ilustrar los desarrollos de este capítulo, se muestra un ejemplo práctico.

Este trabajo pretende ser una referencia a toda persona, con co-

nocimientos de inferencia estadística, que esté interesado en aspectos teóricos de la estadística bayesiana y en particular de su aplicación al análisis de regresión.

Capítulo 1

Teoría de la Decisión y Estadística

1.1 Introducción

En este capítulo se trata de dar una visión general de lo que es la estadística bayesiana desde un punto de vista de la teoría de decisiones. Para ello el capítulo se divide en tres secciones básicas, en la sección 1.2 se presenta la forma de resolver un problema de decisión considerando el criterio de Bayes. En la sección 1.3 se plantean algunos problemas estadísticos como problemas de decisión bajo incertidumbre y se presenta el teorema de Bayes como la forma de incorporar la información proporcionada por una muestra aleatoria al análisis estadístico.

Finalmente, en la sección 1.4 se trata el problema estadístico de predicción desde dos puntos de vista, predecir el resultado de un experimento antes de tener cualquier observación y hacer dicha predicción después de haber registrado n datos.

1.2 Introducción a la Teoría de la Decisión.

La mayoría de los problemas de carácter estadístico para los cuales los métodos tradicionales han dado solución, pueden reducirse a problemas de decisión en ambiente de incertidumbre y así dichas soluciones pueden verse desde otra perspectiva.

Ramsey (1926) manifiesta que unos cuantos principios básicos sobre comportamiento coherente, tomados éstos como axiomas, son suficientes para deducir de ellos una teoría general de la decisión y de la inferencia estadística.

Así, es posible establecer una metodología general y unificada para resolver problemas estadísticos a partir del análisis cuidadoso del proceso lógico que debe seguirse para tomar una decisión.

Por lo anterior, se presenta a continuación la estructura general de un problema de decisión y las formas en que podría resolverse, así como, la metodología que proporcione la manera en que debería resolverse un problema de decisión bajo incertidumbre.

Lo primero que hay que hacer cuando uno se enfrenta a un *problema de decisión* bajo incertidumbre, es eliminar cuanta incertidumbre sea posible tratando de obtener la mayor información sobre el problema. Posteriormente, se debe construir el conjunto de *todas* las posibles alternativas, a dicho conjunto se le conoce como espacio de decisiones y se denotará por D . La forma de determinar el espacio de decisiones debe ser exhaustiva para abarcar todas las opciones de interés y de tal forma que la elección de un elemento en D excluya la elección de cualquier otro.

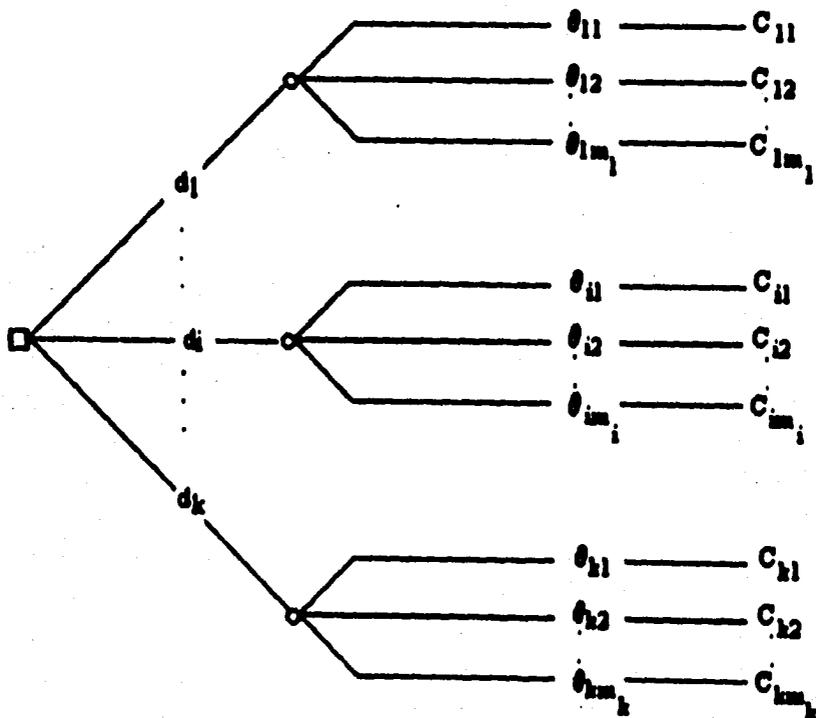
Determinar la mejor de las alternativas sería, en principio, inmediato, si se tuviera la información completa sobre los posibles sucesos inciertos, además de un orden de preferencia entre las posibles consecuencias. Sin embargo, es poco frecuente que ésto suceda al plantearse un problema de decisión, es por ésto que el problema se plantea en un ambiente de *incertidumbre*.

En presencia de incertidumbre una vez determinado el espacio de decisiones, se debe determinar para cada decisión d_i , el conjunto de sucesos inciertos que se denotará por Θ_i , el cual deberá ser exhaus-

tivo y sus elementos mutuamente excluyentes, así como sus eventuales consecuencias y las preferencias entre las mismas.

Esquemáticamente, un problema de decisión en ambiente de incertidumbre, en el caso de un número finito de alternativas y un número finito de sucesos inciertos, puede representarse mediante un "árbol de decisiones" (ver figura 1.2.1), donde $D = \{d_1, \dots, d_k\}$ es el espacio de decisiones, $\Theta_i = \{\theta_{i1}, \dots, \theta_{im_i}\}$ el espacio de sucesos inciertos, C_{ij} denota la correspondiente consecuencia que ocurriría si se elige la decisión d_i y ocurre el suceso incierto θ_{ij} , (\square) representa el nodo de decisión, cada (o) representa un nodo aleatorio, las ramas (\rightarrow) que se desprenden de (\square) están asociadas a decisiones y las ramas (\rightarrow) que se desprenden de (o) se asocian a sucesos inciertos.

Figura 1.2.1



La mayoría de los problemas de decisión tienen, aparentemente, una estructura más compleja que la que se ha descrito. Sin embargo, tales problemas complejos se pueden resolver considerando la idea básica aquí presentada. Se supondrá que el conjunto de sucesos inciertos es el mismo para cualquier elemento d_i del espacio de decisiones, por lo que $\Theta_i = \Theta$ para todo i . Si tal caso no sucediera se tomarían los correspondientes espacios producto de los Θ_i . Si además, sólo hay un número finito de alternativas y sucesos inciertos, entonces el problema de decisión también puede representarse mediante una *tabla de decisiones* (ver Figura 1.2.2).

Figura 1.2.2

	$\theta_1 \dots \theta_j \dots \theta_m$
d_1	$C_{11} \dots C_{1j} \dots C_{1m}$
\vdots	$\vdots \quad \quad \quad \vdots$
d_i	$C_{i1} \dots C_{ij} \dots C_{im}$
\vdots	$\vdots \quad \quad \quad \vdots$
d_k	$C_{k1} \dots C_{kj} \dots C_{km}$

Una vez presentada la estructura general de un problema de decisión se discute a continuación como resolver un problema de decisión bajo incertidumbre. Un criterio que se utiliza frecuentemente para resolver un problema de esta naturaleza, es el *criterio minimax*. Una de las versiones de este criterio consiste en determinar para cada decisión la peor consecuencia y posteriormente elegir aquella decisión que produzca, dentro de las peores consecuencias, la más preferible. Como se puede ver, para utilizar este criterio el decisor tiene que determinar su espacio de decisiones, el espacio de sucesos inciertos y su grado de preferencia entre las posibles consecuencias. Sin embargo, este criterio puede ser incoherente en sus resultados. Por ejemplo, si se considera el problema de decisión definido por la Tabla 1.2.1, donde los valores 0.32, 0.30, 0.38 representan la probabilidad de ocurrencia de los sucesos inciertos θ_1, θ_2 y θ_3 respectivamente, y el orden de preferencia entre las consecuencias está definido por el orden de los números reales.

Tabla 1.2.1

	θ_1 0.32	θ_2 0.30	θ_3 0.38
d_1	.1	.1	.4
d_2	.4	.4	.1

Según el criterio minimax, tanto la decisión d_1 como la decisión d_2 son igualmente preferibles. Sin embargo, considerando las diferentes probabilidades de ocurrencia de los sucesos θ_i ; ningún decisor consideraría ambas decisiones igualmente preferibles. Se puede decir que la causa de la incoherencia de este criterio reside en su excesivo pesimismo, así como, en el hecho de no considerar la información que se dispone sobre la verosimilitud relativa de los sucesos inciertos.

Otro criterio que se usa frecuentemente es el *criterio de la consecuencia más probable*, que consiste en tomar la decisión que resultaría óptima si el suceso más probable ocurriera efectivamente. En este caso el decisor además de determinar el espacio de decisiones, su grado de preferencia entre las posibles consecuencias y el espacio de sucesos inciertos, debe de establecer una medida de verosimilitud entre los elementos de este último que refleje su grado de conocimiento sobre ellos. Este criterio puede ser menos razonable de lo que parece, ya que una decisión que no sea óptima para el suceso más probable puede ser mejor para el resto de los sucesos inciertos. Por ejemplo, si se considera el problema de decisión representado por la Tabla 1.2.1, se puede observar que bajo este criterio, la decisión que se debe elegir es d_1 puesto que si el suceso más probable es θ_3 la decisión d_1 produce una mejor consecuencia. Ahora bien, se puede ver que como las consecuencias correspondientes a ambas decisiones para los sucesos θ_1 y θ_2 son iguales, el problema de decisión puede reformularse mediante la Tabla 1.2.2.

Tabla 1.2.2

	$\theta_1 \cup \theta_2$ 0.62	θ_3 0.38
d_1	0.1	0.4
d_2	0.4	0.1

Ahora el suceso más probable es $\theta_1 \cup \theta_2$ y por tanto, según este criterio la decisión que se debe tomar es d_2 , lo cual representa incoherencia en la toma de decisiones.

Los criterios hasta ahora expuestos pueden dar la pauta para resolver un problema de decisión con incertidumbre. Sin embargo, como se ha visto, estos criterios resultan incoherentes en algunos casos.

A continuación se presenta una metodología a través de la cual siempre es posible encontrar una solución coherente a los problemas de decisión, lo que permitirá obtener un mecanismo general y unificado para resolver problemas estadísticos.

Si el decisor es capaz de aceptar unos cuantos principios de razonamiento coherente o principios de coherencia (Bernardo, 1971) como fundamentos axiomáticos básicos, entonces siempre dispondrá de una forma razonable y única para tomar decisiones de manera coherente. Para lograr ésto el decisor debe ser capaz de determinar el conjunto de todas las posibles decisiones, D , el conjunto Θ de los sucesos inciertos y las consecuencias, C_{ij} . Posteriormente, el decisor debe de cuantificar y evaluar la información que posea de los diferentes sucesos inciertos, así como sus preferencias entre las posibles consecuencias.

La manera de cuantificar la incertidumbre de Θ es mediante la asignación de probabilidades a sus diferentes elementos. Aquí se debe tomar en cuenta que si Θ es finito, aún más, si Θ es numerable, la cuantificación de la incertidumbre sobre Θ se puede realizar asignando números $P(\theta_j|H)$ tales que:

$$0 \leq P(\theta_j|H) \leq 1 \quad \text{para toda } \theta_j \text{ en } \Theta$$

$$y \quad \sum_j P(\theta_j|H) = 1 .$$

$P(\theta_j|H)$ denota la probabilidad de que el suceso θ_j ocurra bajo las condiciones H . Por otro lado, si Θ no es numerable se puede definir una medida de probabilidad sobre el sigma álgebra generado por Θ , $\sigma(\Theta)$. En el caso en que Θ esté representado por un subconjunto en \mathfrak{R}^k , basta definir una función de densidad de probabilidad, (f.d.p.), para cuantificar la incertidumbre sobre Θ .

Las preferencias entre las posibles consecuencias pueden ser evaluadas numéricamente a través de una medida de utilidad $U(\cdot)$. Así, si se tiene que C es el conjunto de todas las posibles consecuencias y (\mathcal{L}) el orden de preferencia entre ellas, una utilidad es una función definida como $U : C \rightarrow \mathfrak{R}$, donde $c_1 \mathcal{L} c_2$ si y sólo si $U(c_1) \mathcal{L} U(c_2)$, para todo par c_1, c_2 de consecuencias en C .

Por otro lado, en un problema de decisión se tiene implícito la existencia de una función

$$F : DX\Theta \rightarrow C.$$

Ya que para cada decisión y suceso incierto se tiene asociada una consecuencia.

Así, para cada decisión d_i en D y cada suceso θ en Θ se tiene una utilidad $U(d_i, \theta)$ definida a través de la composición $U \circ F$.

De esta manera, si se toma la decisión d en las condiciones H se puede obtener la utilidad esperada $U^*(d)$ para cada decisión. En caso de que Θ sea finito o numerable $U^*(d)$ se calcula como

$$U^*(d) = \sum_j U(d, \theta_j) P(\theta_j|H).$$

Cuando Θ es no numerable y $U(d, \theta)$ es p-medible, la utilidad esperada se obtiene a través de una integral de Lebesgue

$$U^*(d) = \int_{\Theta} U(d, \theta) dP.$$

Finalmente, los axiomas de coherencia garantizan que la solución óptima al problema de decisión será aquella decisión que maximice la utilidad media o utilidad esperada. A esto se le conoce como el *criterio de decisión de Bayes*, y la bondad del mismo reside esencialmente en su fundamento axiomático. A continuación se presenta un ejemplo en el que se usarán las diferentes formas, que se han discutido, para resolver un problema de decisión en ambiente de incertidumbre.

Ejemplo 1.2.1

Una compañía de seguros debe decidir si crea una póliza de seguros que cubra completamente uno de los dos tipos de riesgos a los que estarán expuestas las nuevas compañías textiles que se instalarán en México el próximo año, o bien, crear una póliza que cubra los dos tipos de riesgos parcialmente. La compañía de seguros considera que lo más probable es que las nuevas compañías textiles estarán más expuestas al riesgo tipo 2.

El espacio de decisiones en este problema es:

d_1 = crear una póliza que cubra completamente el riesgo tipo 1.

d_2 = crear una póliza que cubra completamente el riesgo tipo 2.

d_3 = crear una póliza que cubra los dos tipos de riesgo pero con una cobertura menor para ambos riesgos.

El espacio de sucesos inciertos podría describirse como:

θ_1 = las nuevas compañías textiles estarán propensas al riesgo tipo 1.

θ_2 = las nuevas compañías textiles estarán propensas al riesgo tipo 2.

θ_3 = las nuevas compañías textiles estarán propensas tanto al riesgo tipo 1 como al riesgo tipo 2.

De acuerdo a la compañía de seguros se presenta la siguiente tabla de utilidades

	θ_1	θ_2	θ_3	mín
d_1	0.9	0.2	0.5	0.2
d_2	0.2	0.9	0.5	0.2
d_3	0.6	0.6	0.7	0.6
			máx	0.6

El criterio *minimax* recomendaría elegir siempre d_3 , independientemente de las verosimilitudes relativas de θ_1 y θ_2 .

El criterio de la consecuencia más probable recomienda elegir siempre d_2 , ya que por hipótesis $P(\theta_2) > P(\theta_3)$ y $P(\theta_2) > P(\theta_1)$

Ahora si se consideran las utilidades de las tres decisiones se tiene:

$$U^*(d_1) = 0.9p + 0.2q + 0.5(1 - p - q) = 0.5 + 0.4p - 0.3q$$

$$U^*(d_2) = 0.2p + 0.9q + 0.5(1 - p - q) = 0.5 - 0.3p + 0.4q$$

$$U^*(d_3) = 0.6p + 0.6q + 0.7(1 - p - q) = 0.7 - 0.1p - 0.1q$$

donde $p = P(\theta_1)$ y $q = P(\theta_2)$ y además, por hipótesis, $q > p$ y $q > 1 - p - q$, esto implica que $p > 1 - 2q$.

Se puede ver que $U^*(d_1) > U^*(d_2)$, si y solamente si $p > q$; como por hipótesis $q > p$, d_2 siempre es mejor que d_1 .

Además

$$U^*(d_2) > U^*(d_3) \text{ si y sólo si } 0.2p - 0.5q + 0.2 < 0$$

es decir, si por ejemplo $p = 1/3$ y $q = 1/2$, entonces $0.2p - 0.5q + 0.2 > 0$ en este caso el criterio de Bayes recomienda tomar la decisión d_3 y si $p = 1/3$ pero $q = 3/5$, entonces se tiene que $0.2p - 0.5q + 0.2 < 0$ por lo que en este caso se recomienda tomar la decisión d_2 .

1.3 Problemas de Decisión Estadísticos

Los problemas de inferencia estadística paramétrica se pueden ver como problemas de decisión bajo incertidumbre.

Para el problema estadístico de estimación puntual del parámetro θ , donde θ es el parámetro de un modelo probabilístico $f(x|\theta)$, el espacio de decisiones es igual a el espacio parametral, es decir, $D = \Theta$ y el conjunto de sucesos inciertos estará representado por los posibles elementos de Θ . Para el problema estadístico de prueba de hipótesis el espacio de decisión estará dado por las hipótesis a contrastar, y el conjunto de sucesos inciertos, igual que en el de estimación puntual, está representado por los posibles valores del espacio parametral. En ambos casos la función de utilidad es una cierta medida de información.

Se dejó ver anteriormente que para resolver un problema de decisión es importante tratar de incorporar toda la información que se disponga sobre el problema. Como lo desconocido en un problema estadístico es el valor de θ , en principio, es necesario expresar el poco o mucho

conocimiento inicial que se tenga sobre θ a través de una f.d.p., $\pi(\theta)$, a esta función se le conoce como *distribución inicial*.

Una manera como se puede elegir una distribución inicial, es considerar una familia paramétrica de densidades de probabilidad suficientemente amplia que pueda contener una densidad que aproxime adecuadamente a $\pi(\theta)$. Posteriormente establecer condiciones (éstas pueden ser sobre medias, cuantiles, etc.), que permitan determinar los parámetros de la distribución. Logrado esto es recomendable proceder a calcular algunas probabilidades para verificar que los resultados sean consistentes con la información inicial con que se contaba.

Para mejorar la información inicial es necesario determinar la relación existente entre los datos y el parámetro de interés θ . Esto se puede lograr especificando el modelo probabilístico $f(x|\theta)$.

Como función de x para un θ fijo, $f(x|\theta)$ es una densidad que describe la probabilidad de obtener los distintos valores de x . Para un $x = x_0$ fijo, $f(x_0|\theta)$ describe la verosimilitud de los distintos valores de θ a la luz del resultado experimental observado x_0 . A la función de θ , $f(x_0|\theta)$, se le conoce como función de *verosimilitud* de θ dado x_0 .

Así, la función de verosimilitud correspondiente a una muestra de variables aleatorias independientes e idénticamente distribuídas (v.a. i.i.d) $\{x_1, \dots, x_n\}$ extraída de una población cuyos elementos tienen una densidad $f(x|\theta)$ que depende de θ , es

$$f(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta).$$

Una vez que se determina la forma de cuantificar o describir la información inicial que se posee sobre el parámetro θ , a través de su distribución inicial, y suponiendo que existe la función de probabilidad conjunta $f(x_1, \dots, x_n|\theta)$, se está en condiciones de obtener la distribución final de θ . La cual describe el conocimiento que se tiene sobre θ tras incorporar a la información inicial la información que proporcionan los resultados experimentales. El camino que permite obtener dicha distribución final $\pi(\theta|x_1, \dots, x_n)$ es el *teorema de Bayes*.

Teorema 1.3.1 (Teorema de Bayes). Sean x_1, \dots, x_n los resultados del experimento ξ con f.d.p. conjunta $f(x_1, \dots, x_n|\theta)$ y sea $\pi(\theta)$ la

distribución inicial de θ . La distribución final de θ está dada por:

$$\pi(\theta|x_1, \dots, x_n) = f(x_1, \dots, x_n|\theta)\pi(\theta)/\pi(x_1, \dots, x_n)$$

en donde

$$\pi(x_1, \dots, x_n) = \int_{\Theta} f(x_1, \dots, x_n|\theta)\pi(\theta)d\theta.$$

$\pi(x_1, \dots, x_n)$ es la distribución conjunta de x_1, \dots, x_n sin condicionar a θ y también se le conoce como *distribución predictiva final* de X , con $X = (x_1, \dots, x_n)$. Además como $\pi(X)$ no depende de θ , sin pérdida de generalidad, el teorema de Bayes puede expresarse en la forma

$$\pi(\theta|x_1, \dots, x_n) \propto f(x_1, \dots, x_n|\theta)\pi(\theta),$$

donde \propto denota "proporcional a". Por lo que si dos modelos probabilísticos son proporcionales para todo θ , dan necesariamente lugar a la misma distribución final, y se puede escribir

Distribución final \propto verosimilitud \times Distribución inicial.

Ejemplo 1.3.1

Con objeto de mejorar la información sobre una magnitud θ , se realizan n medidas independientes x_1, \dots, x_n cada una de las cuales tiene una distribución Bernoulli con parámetro θ . Si la información inicial sobre θ se describe a través de la distribución inicial $\pi(\theta)$, aplicando el teorema de Bayes, la distribución final de θ será de la forma:

$$\pi(\theta|x_1 \dots x_n) \propto [\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}] \pi(\theta).$$

Si además la distribución inicial de θ es de la familia Beta, es decir, $\pi(\theta) = \text{Be}(\alpha, \beta)$ $\alpha, \beta > 0$, resulta

$$\begin{aligned} \pi(\theta|x_1, \dots, x_n) &\propto \{\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}\} \theta^{\alpha-1} (1-\theta)^{\beta-1} \\ &\propto \theta^{(\alpha+\sum_{i=1}^n x_i)-1} (1-\theta)^{(\beta+n-\sum_{i=1}^n x_i)-1} \end{aligned}$$

en donde la constante de proporcionalidad, está dada por

$$\frac{\Gamma(\alpha + \sum_{i=1}^n x_i + \beta + n - \sum_{i=1}^n x_i)}{\Gamma(\alpha + \sum_{i=1}^n x_i) \Gamma(\beta + n - \sum_{i=1}^n x_i)}.$$

Por tanto

$$\pi(\theta|x_1, \dots, x_n) = Be(\alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i)$$

$$\text{con } \alpha + \sum_{i=1}^n x_i > 0 \quad \text{y} \quad \beta + n - \sum_{i=1}^n x_i > 0.$$

Ejemplo 1.3.2

Ahora suponga que el parámetro de interés es la precisión r de una distribución $N(\mu, r)$ con $\mu \in (-\infty, \infty)$ conocida, y además suponga que la apreciación inicial de r puede modelarse por medio de una distribución Gamma, es decir $\pi(r) = \text{Gamma}(\alpha, \beta)$ con $\alpha, \beta > 0$ conocidos. Con el objeto de mejorar la apreciación inicial de r , se obtiene una m.a. x_1, \dots, x_n de la distribución $N(\mu, r)$. Observe como se mejora la apreciación inicial $\pi(r)$ obteniendo la correspondiente distribución final de r .

Se sabe que

$$\pi(r) = \frac{\beta^\alpha}{\Gamma(\alpha)} r^{\alpha-1} \exp\{-r\beta\} I_{(0,\infty)}(r)$$

$$f(x|\mu, r) = \frac{r^{1/2}}{\sqrt{2\pi}} \exp\left\{-\frac{r}{2}(x - \mu)^2\right\} I_{(-\infty,\infty)}(x)$$

entonces

$$\begin{aligned} \pi(r|x_1, \dots, x_n) &\propto f(x_1, \dots, x_n|r)\pi(r) \\ &\propto r^{\frac{n}{2}} \exp\left\{-\frac{r}{2} \sum_{i=1}^n (x_i - \mu)^2\right\} r^{\alpha-1} \exp\{-r\beta\} \\ &\propto r^{\alpha+\frac{n}{2}-1} \exp\left\{-\frac{r}{2} \left\{2\beta + \sum_{i=1}^n (x_i - \mu)^2\right\}\right\}. \end{aligned}$$

Así, la apreciación final sobre r , después de haber observado la m.a. x_1, \dots, x_n está dada por una

$$\text{Gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2}\right)$$

Ejemplo 1.3.3

Sean x_1, \dots, x_n n observaciones independientes cada una de las cuales tiene una distribución $N(\mu, \sigma_0^2)$ con σ_0^2 conocida. Si la información inicial sobre μ se describe a través de una distribución $N(\mu_0, \sigma_1^2)$, con μ_0 y σ_1^2 conocidos, la distribución final será de la forma

$$\begin{aligned}\pi(\mu|x_1, \dots, x_n) &\propto \frac{1}{\sigma_0^{n/2}} \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)^2\right\} \frac{1}{\sigma_1} \exp\left\{-\frac{1}{2\sigma_1^2} (\mu - \mu_0)^2\right\} \\ &\propto \exp\left\{\frac{-1}{2\sigma_0^2} [n(\mu - \bar{x})^2] - \frac{1}{2\sigma_1^2} (\mu - \mu_0)^2\right\} \\ &\propto \exp\left\{\frac{-1}{2} \left[\frac{n}{\sigma_0^2} (\mu - \bar{x})^2 + \frac{1}{\sigma_1^2} (\mu - \mu_0)^2\right]\right\}\end{aligned}$$

Completando el cuadrado para μ , la última expresión puede verse de la siguiente manera

$$\pi(\mu|x_1, \dots, x_n) \propto \exp\left\{-\frac{1}{2} \left[\mu - \left(\frac{n\bar{x}}{\sigma_0^2} + \frac{\mu_0}{\sigma_1^2}\right) / \left(\frac{n}{\sigma_0^2} + \frac{1}{\sigma_1^2}\right)\right]^2 \left(\frac{n}{\sigma_0^2} + \frac{1}{\sigma_1^2}\right)\right\}.$$

De aquí se puede concluir que la distribución final de μ dado x_1, \dots, x_n observaciones independientes, es una Normal con media

$$\mu^* = \left(\frac{n\bar{x}}{\sigma_0^2} + \frac{\mu_0}{\sigma_1^2}\right) / \left(\frac{n}{\sigma_0^2} + \frac{1}{\sigma_1^2}\right)$$

y varianza

$$\frac{1}{\frac{n}{\sigma_0^2} + \frac{1}{\sigma_1^2}}$$

Como se puede ver de los párrafos anteriores, la elección de una determinada familia para representar la información inicial sobre el parámetro θ puede conducir a planteamientos y cálculos más sencillos. Así, si se tiene x_1, \dots, x_n una m.a. de una distribución Bernoulli con parámetro θ desconocido y la distribución inicial de θ es una distribución Beta con parámetros $\alpha, \beta > 0$, entonces la distribución final

de θ pertenece a la misma familia y toda la información proporcionada por los datos, para hacer inferencia sobre θ , puede resumirse en el par $(\sum_{i=1}^n x_i, n)$. Análogamente, si la información inicial sobre la media μ de una distribución Normal con varianza conocida, se puede describir mediante otra distribución Normal con varianza y media también conocida, entonces la correspondiente distribución final de μ resulta así mismo Normal y toda la información proporcionada por los datos pueden resumirse en el par (\bar{x}, n) .

Los ejemplos anteriores son casos particulares de una amplia gama de situaciones en las que la aproximación de la distribución inicial, mediante distribuciones pertenecientes a una determinada familia (familia conjugada de distribuciones) permite obtener resultados sencillos de tipo analítico en problemas que, de otra manera, requerirían técnicas de integración numérica. Es de interés notar que en este caso, el paso de la distribución inicial del parámetro a la distribución final del mismo se hace a través de simples reglas de actualización de los parámetros.

Definición 1.3.1 *Se dice que una familia de distribuciones de θ es conjugada con respecto a un modelo probabilístico $f(x|\theta)$ si para cualquier distribución inicial perteneciente a cierta familia de distribuciones F , se obtiene una distribución final que también pertenece a la misma familia de distribuciones F .*

En el siguiente teorema se presenta un resultado que muestra la existencia de familias conjugadas, pero antes se define que es una estadística suficiente.

Definición 1.3.2 *Una estadística $T = T(x_1, \dots, x_n)$ (esto es, una función de la muestra), es suficiente para hacer inferencia sobre el parámetro $\theta = (\theta_1, \dots, \theta_k)$ de un modelo probabilístico $f(x|\theta)$, si cualquiera que sea la distribución inicial de θ , $\pi(\theta)$, su distribución final sólo depende de la muestra x_1, \dots, x_n a través de T , esto es*

$$\pi(\theta|x_1, \dots, x_n) = \pi(\theta|T)$$

Teorema 1.3.1 *Sea $\pi(\theta)$ la distribución inicial del parámetro θ la cual pertenece a una cierta familia paramétrica. Si $f(x|\theta)$ es una función de*

densidad correspondiente a un modelo probabilístico, para el cual existe una estadística T suficiente de dimensión fija y además $\int_{\Theta} \pi(\theta) d\theta < \infty$ entonces existe una familia conjugada paramétrica para θ . (De Groot, 1970)

Finalmente, se mostrará un ejemplo de familias conjugadas en el caso multivariado; pero antes de realizar lo anterior, se introducirá de manera rápida la distribución Normal multivariada.

Se dice que un vector aleatorio K -dimensional $X = (X_1, \dots, X_k)^t$ tiene una distribución Normal multivariada no singular con media el vector $\underline{\mu}$ y matriz de covarianza Σ si X tiene una distribución absolutamente continua cuya función de densidad de probabilidad $f(\cdot | \underline{\mu}, \Sigma)$, para cualquier punto $x \in \mathbb{R}^k$, está dada por

$$f(x | \underline{\mu}, \Sigma) = (2\pi^{-k/2}) |\Sigma|^{-1/2} \exp\{-\frac{1}{2}(x - \underline{\mu})^t \Sigma^{-1} (x - \underline{\mu})\}.$$

Aquí la media $\underline{\mu} = (\mu_1, \dots, \mu_k)^t$ es un vector k -dimensional de números reales y la matriz de covarianzas Σ es de $k \times k$, simétrica y definida positiva.

Ejemplo 1.3.4

Supóngase que x_1, \dots, x_n es una m.a. de una distribución Normal multivariada con vector de medias M desconocido y una matriz de precisión conocida r ($r > 0$). Supóngase también que la distribución inicial de M es una normal multivariada con media $\underline{\mu}$ y matriz de precisión τ ambas conocidas, tal que $\underline{\mu} \in \mathbb{R}^k$ y τ es una matriz definida positiva simétrica.

Calcule entonces la distribución final de M , para $M = m$. La función de verosimilitud $f_n(x_1, \dots, x_n | m)$ satisface la siguiente relación:

$$f_n(x_1, \dots, x_n | m) \propto \exp\{-\frac{1}{2} \sum_{i=1}^n (x_i - m)^t r (x_i - m)\}. \quad (1.1)$$

Además

$$\sum_{i=1}^n (x_i - m)^t r (x_i - m) =$$

$$= \sum_{i=1}^n (x_i - \bar{x})' r (x_i - \bar{x}) + n(m - \bar{x})' r (m - \bar{x}). \quad (1.2)$$

Por lo que la relación 1.1 puede escribirse de la siguiente forma:

$$f_n(x_1, \dots, x_n | m) \propto \exp\left\{-\frac{1}{2}(m - \bar{x})'(nr)(m - \bar{x})\right\}. \quad (1.3)$$

Ahora la función de densidad de probabilidad inicial $\pi(\cdot)$ de M satisface que

$$\pi(m) \propto \exp\left\{-\frac{1}{2}(m - \underline{\mu})' \tau (m - \underline{\mu})\right\}. \quad (1.4)$$

Como la f.d.p. final $\pi(\cdot | x_1, \dots, x_n)$ de M es proporcional al producto de las funciones especificadas en las relaciones 1.3 y 1.4, haciendo una extensión del caso univariado se tiene:

$$(m - \underline{\mu})' \tau (m - \underline{\mu}) + (m - \bar{x})'(nr)(m - \bar{x}) =$$

$$= (m - \underline{\mu}^*)' (\tau + nr) (m - \underline{\mu}^*) + c$$

donde

$$\underline{\mu}^* = (\tau + nr)^{-1} (\tau \underline{\mu} + nr \bar{x})$$

y c representa a términos que no involucran a m , por lo que se pueden considerar en el factor de proporcionalidad, obteniéndose que

$$\pi(m | x_1, \dots, x_n) \propto \exp\left\{-\frac{1}{2}(m - \underline{\mu}^*)' (\tau + nr) (m - \underline{\mu}^*)\right\}.$$

Así, se tiene que la distribución final de m es una Normal multivariada con media el vector $\underline{\mu}^*$ y matriz de precisión Σ^* donde

$$\Sigma^* = \tau + nr.$$

En resumen, con lo que se ha discutido en este capítulo, un problema de inferencia estadística formulado como uno de decisión y sin

incorporar información muestral se resuelve al determinar la opción que maximiza

$$E_{\pi(\theta)}(U(d, \theta)) = \int_{\Theta} U(d, \theta) \pi(\theta) d\theta.$$

Obtenidos los datos y especificada la función de verosimilitud $f(x_1, \dots, x_n | \theta)$ se puede usar el teorema de Bayes para calcular $\pi(\theta | x_1, \dots, x_n)$ y así poder determinar la opción que maximiza

$$E_{\pi(\theta | x_1, \dots, x_n)}(U(d, \theta)) = \int_{\Theta} U(d, \theta) \pi(\theta | x_1, \dots, x_n) d\theta.$$

Ejemplo 1.3.5

Sea x_1, \dots, x_n una m.a. donde $x_i \sim \text{Bernoulli}(\theta)$, se quiere estimar θ y se desea usar una función de utilidad cuadrática. Además se sabe que la información inicial, $\pi(\theta)$, se puede describir por una distribución Beta con parámetros $\alpha, \beta > 0$ conocidos.

Sea $U(d, \theta) = -(d - \theta)^2$ la función de utilidad cuadrática, si $P(\theta)$ es cualquier función de probabilidad, definida en el intervalo $(0, 1)$, el estimador óptimo es aquel valor que maximice

$$E_{P(\theta)}(U(d, \theta)) = E_{P(\theta)}(-(d - \theta)^2)$$

Así

$$E_{P(\theta)}(-(d - \theta)^2) = -(d^2 - 2dE_{P(\theta)}(\theta) + E_{P(\theta)}(\theta^2))$$

$$\frac{\partial E_{P(\theta)}(U(d, \theta))}{\partial d} = -(2d - 2E_{P(\theta)}(\theta)).$$

Por tanto la decisión

$$d_o = E_{P(\theta)}(\theta)$$

maximiza la utilidad esperada respecto de $P(\theta)$. Como la distribución inicial $\pi(\theta)$ es una distribución Beta(α, β) entonces

$$d_o = \frac{\alpha}{\alpha + \beta}.$$

Si se tiene $\{x_1, \dots, x_n\}$ una m.a. de n observaciones Bernoulli(θ), la distribución final de θ con una inicial Beta(α, β), es una

Beta($\alpha + y, \beta + n - y$) con $y = \sum_{i=1}^n x_i$ (ver ejemplo 1.3.1). De aquí se obtiene que la decisión que maximiza

$$E_{\pi(\theta|x_1, \dots, x_n)}(U(d, \theta)) = E_{\pi(\theta|x_1, \dots, x_n)}(-(d - \theta)^2)$$

es

$$d_o = \frac{\alpha + y}{\alpha + \beta + n}.$$

Parámetros de Ruido.

Tradicionalmente el problema de inferencia se formula como el problema de extraer conclusiones probabilísticas sobre los valores de los parámetros desconocidos de una cierta f.d.p.

Sin embargo, hasta el momento se ha analizado el caso en que la distribución de los resultados experimentales x sólo depende del parámetro de interés θ , lo cual no siempre ocurre, sino que frecuentemente dicha distribución también depende de otro parámetro ν que se denomina parámetro marginal o de ruido. De tal forma que los resultados experimentales x presentan una f.d.p $f(x|\theta, \nu)$.

En este caso el problema de hacer inferencia sobre θ , se resuelve obteniendo la distribución final $\pi(\theta, \nu|x_1, \dots, x_n)$ y marginalizando con respecto al parámetro de ruido, ν .

1.4 Predicción

Uno de los problemas que también se presentan en inferencia es el querer estimar una observación futura. Aquí se presentan dos situaciones, la primera cuando inicialmente se tiene un experimento ξ y se quiere predecir en cierta forma el resultado del experimento. Ahora, si se considera que x es el resultado del experimento ξ con una f.d.p. $f(x|\theta)$ y θ desconocido, se observa que no es posible utilizar $f(x|\theta)$ para determinar los valores de x que resulten más probables. Sin embargo, aunque el valor exacto de θ es desconocido, se dispone de cierta información a través de su distribución inicial $\pi(\theta)$ la que junto con la distribución $f(x|\theta)$ permite describir la información que se posee sobre

los posibles valores de x , $\pi(x)$, de la siguiente manera

$$\pi(x) = \int_{\theta \in \Theta} f(x|\theta) \pi(\theta) d\theta \quad (1.5)$$

si θ es una cantidad aleatoria continua.

En el caso de que θ sea una cantidad aleatoria discreta, se tiene

$$\pi(x) = \sum_i f(x|\theta_i) \pi(\theta_i) \quad (1.6)$$

Así, las expresiones 1.5 y 1.6 proporcionan una distribución de x totalmente conocida, la cual es posible usar, entre otras cosas, para hacer predicciones sobre los valores de x a que daría lugar el experimento ξ , por lo que dicha distribución $\pi(x)$ recibe el nombre de *distribución predictiva inicial o a priori*.

La segunda situación consiste en predecir un valor futuro x , contando ya con n observaciones $\{x_1, \dots, x_n\}$ del experimento ξ . La información acumulada sobre el valor θ puede utilizarse para predecir el resultado futuro mediante la distribución predictiva final, la cual se obtiene de la siguiente manera

$$\pi(x|x_1, \dots, x_n) = \int_{\Theta} f(x|\theta) \pi(\theta|x_1, \dots, x_n) d\theta, \quad (1.7)$$

si θ es una cantidad aleatoria continua. En el caso de que θ sea una cantidad aleatoria discreta se tiene

$$\pi(x|x_1, \dots, x_n) = \sum_j f(x|\theta_j) \pi(\theta_j|x_1, \dots, x_n). \quad (1.8)$$

También un problema de predicción puede plantearse como un problema de decisión. A continuación se presenta una breve discusión de este enfoque a través de los conceptos de la sección anterior.

En un problema de predicción visto como uno de decisión, el espacio de decisiones está representado por el conjunto de todos los posibles resultados x del experimento ξ para los cuales $f(x|\theta) > 0$, para alguna $\theta \in \Theta$. El espacio de sucesos inciertos está representado por el producto cruz entre el espacio de resultados x y el espacio de los valores del parámetro θ . En general, la función de utilidad está definida por

$U(\hat{x}, x, \theta)$ que depende de \hat{x} el valor de la predicción, x el valor del resultado que ocurra y θ el parámetro del modelo $f(x|\theta)$.

De esta manera si no se han registrado observaciones del experimento ξ , la mejor predicción \hat{x} es aquella que maximice

$$E_{\pi(x,\theta)}(U(\hat{x}, x, \theta)). \quad (1.9)$$

Si ya se han registrado n observaciones $\{x_1, \dots, x_n\}$ la mejor predicción \hat{x} es aquella que maximiza

$$E_{\pi(x,\theta|x_1, \dots, x_n)}(U(\hat{x}, x, \theta)). \quad (1.10)$$

donde $\pi(x, \theta|\cdot)$ es la función de distribución conjunta de x y θ . Considerando que comúnmente $U(\hat{x}, x, \theta)$ se expresa como $U(\hat{x}, x)$ y que para expresar el conocimiento sobre (x, θ) basta con hacerlo sobre θ ya que

$$\pi(x, \theta) = f(x|\theta)\pi(\theta).$$

Entonces maximizar 1.9 o 1.10, considerando el caso continuo, es equivalente a maximizar

$$\int \int U(\hat{x}, x) f(x|\theta) \pi(\theta) d\theta dx \quad (1.11)$$

o bien maximizar

$$\int \int U(\hat{x}, x) f(x|\theta) \pi(\theta|x_1, \dots, x_n) d\theta dx \quad (1.12)$$

respectivamente.

Ahora, si se cumplen las condiciones del teorema de Fubini, maximizar 1.11 y 1.12 es equivalente a maximizar

$$E_{\pi(x)}\{U(\hat{x}, x)\}$$

y maximizar

$$E_{\pi(x|x_1, \dots, x_n)}\{U(\hat{x}, x)\}$$

respectivamente.

Ejemplo 1.4.1

Considerando que se ha obtenido una muestra aleatoria de tamaño n de una población y que se ha encontrado que k elementos de la muestra poseían una determinada característica, se desea saber la probabilidad de que un nuevo elemento de la población escogido al azar tenga ($x = 1$) o no tenga ($x = 0$) la característica mencionada. Si θ es la proporción (desconocida) de elementos de la población que tienen la característica, se tiene

$$\begin{aligned} f(x = 1|\theta) &= \theta. \\ f(x = 0|\theta) &= 1 - \theta. \end{aligned}$$

En consecuencia, si $\pi(\theta|k, n)$ es la distribución final que describe la información de que se dispone sobre θ tras observar los resultados muestrales, se tiene por la ecuación 1.7

$$\begin{aligned} \pi(x = 1|k, n) &= \int f(x = 1|\theta)\pi(\theta|k, n)d\theta \\ &= \int_0^1 \theta\pi(\theta|k, n)d\theta = E(\theta|k, n) \end{aligned}$$

y

$$\pi(x = 0|k, n) = 1 - \pi(x = 1|k, n).$$

Por tanto se tiene que la probabilidad de que un nuevo elemento de la población tenga la característica estudiada es simplemente la *media* de la distribución final de θ .

Ejemplo 1.4.2

Considérese finalmente que se ha obtenido una muestra aleatoria x_1, \dots, x_n de una población Normal con desviación estándar σ conocida. Se desea determinar la distribución predictiva final de una nueva observación x .

Por hipótesis

$$f(x|\mu) = N(x|\mu, \sigma).$$

Si $\pi(\mu|x_1, \dots, x_n)$ es la distribución final que describe la información disponible sobre μ tras observar los resultados experimentales, se tiene que

$$\pi(x|x_1, \dots, x_n) = \int N(x|\mu, \sigma)\pi(\mu|x_1, \dots, x_n)d\mu.$$

En particular, si la distribución inicial de μ es $N(\mu|\mu_0, \sigma_0)$ con μ_0 y σ_0 conocidas, por el ejemplo 1.3.3., la distribución final de μ es $N(\mu|\mu_n, \sigma_n)$ donde

$$\mu_n = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \quad y$$

$$\sigma_n = \sqrt{\frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}}$$

por lo que

$$\begin{aligned} \pi(x|x_1, \dots, x_n) &= \int N(x|\mu, \sigma)N(\mu|\mu_n, \sigma_n)d\mu \\ &= N(x|\mu_n, \sqrt{\sigma^2 + \sigma_n^2}). \end{aligned}$$

Ejemplo 1.4.3

Si se considera el ejemplo 1.4.1 con una función de utilidad cuadrática $U(\hat{x}, x) = -(\hat{x} - x)^2$ se tendría que la mejor predicción \hat{x} dada la muestra $\{x_1, \dots, x_n\}$, será aquella que maximice.

$$\begin{aligned} E_{\pi(x|x_1, \dots, x_n)}[U(\hat{x}, x)] &= E_{\pi(x|x_1, \dots, x_n)}[-(\hat{x} - x)^2] \\ &= E_{\pi(x|x_1, \dots, x_n)}[-(\hat{x}^2 - 2\hat{x}x - x^2)] \\ &= -\hat{x}^2 + 2\hat{x}E_{\pi(x|x_1, \dots, x_n)}(x) - E_{\pi(x|x_1, \dots, x_n)}(x^2) \end{aligned} \tag{1.13}$$

Derivando la ecuación 1.13 con respecto a \hat{x} , e igualando a cero se obtiene que

$$\hat{x} = E_{\pi(x|x_1, \dots, x_n)}(x)$$

maximiza

$$E_{\pi(x|x_1, \dots, x_n)}[-(\hat{x} - x)^2]$$

y sin datos

$$\hat{x} = E_{\pi(x)}$$

maximiza

$$E_{\pi(x)}[-(\hat{x} - x)^2].$$

Se puede decir que la distribución predictiva inicial $\pi(x)$ permite hacer inferencias sobre el resultado experimental x antes de realizar alguna observación y puede utilizarse para analizar las implicaciones de la distribución inicial $\pi(\theta)$.

Mientras que la distribución predictiva final $\pi(x|x_1, \dots, x_n)$ recoge la información proporcionada por las observaciones x_1, \dots, x_n y permite inferir sobre el resultado de la próxima observación x .

Por lo anterior es posible darse cuenta que el problema de inferir sobre un posible valor x de un experimento ξ , habiéndose registrado ya n observaciones $\{x_1, \dots, x_n\}$ es un problema de inferencia que puede ser resuelto una vez que se ha obtenido la distribución final de θ .

Capítulo 2

El Modelo de Regresión de Rango Completo

2.1 Introducción

En este capítulo se presenta el modelo de regresión lineal múltiple de rango completo, su planteamiento y el correspondiente análisis bayesiano.

El capítulo consta de tres secciones básicas. En la sección 2.2 se introduce y se define el modelo de regresión lineal múltiple. En la sección 2.3 se considera una distribución inicial Normal-Gamma para los parámetros $\underline{\beta}$ y τ del modelo, se identifican distribuciones marginales y se obtienen las correspondientes distribuciones finales conjuntas y marginales. Se presenta la forma de estimar puntualmente los parámetros $\underline{\beta}$ y τ considerando dos casos, cuando se tienen funciones de utilidad y cuando no se cuenta con ellas. En esta sección también se plantea el problema de estimación por intervalos como un problema de decisión, así como la forma de resolverlo, lo que lleva a definir las llamadas regiones de alta densidad final, regiones HPD. Por lo que se presenta la manera de obtener este tipo de regiones para los parámetros del modelo.

Finalmente, para terminar la sección 2.3 se analiza el problema de contraste de hipótesis para el parámetro $\underline{\beta}$ y se resuelve éste considerando como un problema de decisión. Además, se presentan procedimientos bayesianos alternativos para resolver este problema.

Por último, en la sección 2.4 se muestran los resultados análogos de la sección 2.3, considerando una distribución inicial no informativa para los parámetros β y τ .

2.2 Planteamiento y Definición del Modelo

Es común que cuando se miden varias características para cada individuo o unidad de un conjunto en estudio, se esté interesado en observar el efecto que producen algunas de estas características sobre una de ellas en particular. Por ejemplo, en un estudio médico se miden, a cada individuo de un grupo, las variables edad, peso, estatura y presión arterial; y se está interesado en determinar si la presión arterial de un individuo depende de su edad, peso y estatura. En un segundo ejemplo se registran, en un experimento industrial, el tiempo y la temperatura de calentamiento, el tiempo y temperatura de enfriamiento y la dureza de 15 artículos de cierto material, y se está interesado en ver de qué manera afecta a la dureza del artículo la temperatura y tiempo de calentamiento y enfriamiento.

Una técnica estadística que se usa comúnmente para el estudio de estas relaciones entre variables es el análisis de regresión. Para introducir la notación y terminología necesaria se considera el ejemplo del estudio médico en donde

- Y_i = presión arterial en el i -ésimo individuo
- X_{i1} = edad del i -ésimo individuo
- X_{i2} = peso del i -ésimo individuo
- X_{i3} = estatura del i -ésimo individuo

Cuando se refiera a un individuo particular para el cual los datos han sido obtenidos, se usará la notación $Y_i, X_{i1}, X_{i2}, X_{i3}$. Cuando se hable en general de un individuo cualquiera que forma parte del estudio, se usará la notación Y, X_1, X_2, X_3 .

La variable Y , la cual se piensa que depende de las otras variables, es llamada la *variable dependiente o variable respuesta*, X_1, X_2 y X_3 son llamadas *variables independientes o variables explicativas*.

La relación de las variables de interés puede establecerse a través de un modelo. Así, se puede escribir $F(X_1, X_2, X_3)$ para expresar la parte de Y explicada por las variables independientes, y considerar que Y difiere de $F(X_1, X_2, X_3)$ en una cantidad aleatoria ϵ . No se espera que este modelo describa completamente al fenómeno de interés pero sí que lo haga al menos aproximadamente.

Por lo anterior y con la idea de que la presión arterial de un individuo (Y) es aproximadamente una función de su edad (X_1), de su peso (X_2) y de su estatura (X_3) un modelo posible para este ejemplo es

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon.$$

En general, si hay n variables independiente X_1, \dots, X_n , a la ecuación

$$Y = F(X_1, \dots, X_n) + \epsilon \quad (2.1)$$

se le conoce como modelo de regresión, en el cual F es la función de regresión y ϵ es un error aleatorio. En lo sucesivo, se trabajará con el caso particular en que $F(X_1, \dots, X_n)$ es una función lineal en los parámetros. Esto da lugar a la siguiente definición

Definición 2.2.1 *El modelo de regresión lineal múltiple (MRLM) que relaciona una respuesta aleatoria Y con un conjunto de variables independientes X_1, X_2, \dots, X_p es*

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon.$$

Los coeficientes $\beta_0, \beta_1, \dots, \beta_p$ son parámetros desconocidos y son llamados *coeficientes de regresión o parámetros de regresión*. X_1, \dots, X_p son variables que se miden sin error, ϵ es un vector aleatorio con distribución Normal multivariada de media $\mathbf{0}$ y matriz de varianza-covarianza $\sigma^2 I_n$ con σ^2 desconocida.

En términos de los datos observados, el modelo es

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + \epsilon_i \quad i = 1, \dots, n$$

y además

$$E[Y] = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p. \quad (2.2)$$

Si $p = 1$ al modelo anterior se le conoce como modelo de regresión lineal simple.

Una manera conveniente de expresar al MRLM es a través de matrices. Si $\underline{\beta} = (\beta_0, \dots, \beta_p)^t$ es un vector de $p \times 1$, $Y = (Y_1, \dots, Y_n)^t$ un vector de $n \times 1$, X una matriz de $n \times p$ con i -ésimo renglón (X_{i1}, \dots, X_{ip}) y $\underline{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^t$ un vector aleatorio de $n \times 1$, resulta que el MRLM se puede escribir como

$$Y = X\underline{\beta} + \underline{\epsilon}. \quad (2.3)$$

Por supuesto $\underline{\epsilon} \sim N(\underline{0}, \tau^{-1} I_n)$, con τI_n la matriz de precisión de $\underline{\epsilon}$, X matriz de rango completo, $\sigma^2 = \tau^{-1}$ es la varianza de cada error ϵ_i y la matriz de varianzas y covarianzas de $\underline{\epsilon}$ es $\sigma^2 I_n$.

Al modelo 2.3, se le conoce como MRLM de rango completo. Dada la muestra $Y = \{y_1, \dots, y_n\}$, la función de verosimilitud es

$$f(Y|\underline{\beta}, \tau) = \left(\frac{\tau}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\tau}{2}(Y - X\underline{\beta})^t(Y - X\underline{\beta})\right\}. \quad (2.4)$$

El tratamiento bayesiano para el análisis estadístico es diferente del tratamiento clásico usual. En el enfoque bayesiano, una estimación o prueba de hipótesis es el resultado de la combinación de los datos actuales y de la información que cada decisor tenga para realizar sus inferencias.

En el capítulo anterior se presentó la fórmula básica que se usa para incorporar el conocimiento inicial en un análisis estadístico, la cual fue propuesta por *Thomas Bayes* alrededor de 1760. Sin embargo, la primera aplicación a problemas de regresión fue hecha por *Harold Jeffreys* en 1939. En las siguientes secciones se presenta un análisis del MRLM desde el punto de vista bayesiano.

De acuerdo a lo presentado en el capítulo anterior, para poder realizar un análisis bayesiano del MRLM, es necesario especificar

una distribución inicial $\pi(\underline{\beta}, \tau)$. En el resto del capítulo se presenta este análisis considerando dos distribuciones iniciales, una distribución Normal-Gamma y una no informativa, para los parámetros $\underline{\beta}$ y τ .

2.3 Análisis Bayesiano Conjugado

Es común considerar como distribución inicial, en el caso de un análisis conjugado, un miembro de la familia Normal-Gamma de la forma

$$\pi(\underline{\beta}, \tau) \propto \tau^{\frac{\alpha}{2}} \exp\left\{-\frac{\tau}{2}(\underline{\beta} - \underline{\mu})' A(\underline{\beta} - \underline{\mu})\right\} \tau^{\alpha-1} \exp\{-\tau\lambda\},$$

es decir, $\pi(\underline{\beta}, \tau)$ una Normal-Gamma($\underline{\mu}, A, \alpha, \lambda$), ya que esta familia ofrece varias posibilidades para expresar el conocimiento inicial sobre $\underline{\beta}$ y τ y como se verá más adelante, esta familia es conjugada para el MRLM. Además, cualquiera distribución inicial se puede aproximar a través de combinaciones convexas (mezclas) de distribuciones conjugadas (Dalal, 1983).

Considerando que el enfoque bayesiano basa todas sus inferencias en la distribución final, se obtendrán las correspondientes distribuciones finales $\pi(\underline{\beta}, \tau|Y)$, $\pi(\underline{\beta}|Y)$ y $\pi(\tau|Y)$, usando como distribución inicial conjunta una Normal-Gamma.

Para obtener $\pi(\underline{\beta})$ y $\pi(\tau)$ bastaría con marginalizar $\pi(\underline{\beta}, \tau)$. Sin embargo

$$\pi(\underline{\beta}, \tau) = \pi(\underline{\beta}|\tau)\pi(\tau) \quad (2.5)$$

y de acuerdo a la forma de la distribución Normal-Gamma

$$\begin{aligned} \pi(\underline{\beta}|\tau) &= \text{Normal}(\underline{\mu}, \tau A) \text{ y} \\ \pi(\tau) &= \text{Gamma}(\alpha, \lambda) \end{aligned}$$

donde A es una matriz conocida de $p \times p$ definida positiva, α y λ reales positivos también conocidos.

Por otro lado, se sabe que

$$\pi(\underline{\beta}) \propto \int_0^{\infty} \pi(\underline{\beta}, \tau) d\tau$$

es decir

$$\pi(\underline{\beta}) \propto \int_0^{\infty} \tau^{\frac{n+2\alpha}{2}-1} \exp\left\{-\frac{\tau}{2}[2\lambda + (\underline{\beta} - \underline{\mu})' A(\underline{\beta} - \underline{\mu})]\right\} d\tau.$$

Completando dentro de la integral una distribución Gamma con parámetros

$$\alpha_1 = \frac{p+2\alpha}{2},$$

$$\lambda_1 = \frac{2\lambda + (\underline{\beta} - \underline{\mu})' A(\underline{\beta} - \underline{\mu})}{2},$$

se obtiene

$$\pi(\underline{\beta}) \propto [2\lambda + (\underline{\beta} - \underline{\mu})' A(\underline{\beta} - \underline{\mu})]^{-\frac{n+2\alpha}{2}}.$$

Por lo tanto, $\pi(\underline{\beta})$ es una distribución t-multivariada con 2α grados de libertad, vector de localización $\underline{\mu}$ y matriz de precisión $(2\alpha)(2\lambda)^{-1}A$.

Una vez conocidas las distribuciones iniciales $\pi(\underline{\beta}, \tau)$, $\pi(\underline{\beta})$ y $\pi(\tau)$, se procederá a calcular las correspondientes distribuciones finales.

Para $\underline{\beta}$ y τ la distribución final se puede determinar como

$$\begin{aligned} \pi(\underline{\beta}, \tau | Y) &\propto f(Y | \underline{\beta}, \tau) \pi(\underline{\beta}, \tau) \\ &\propto \tau^{\frac{n+2\alpha}{2}-1} \exp\left\{-\frac{\tau}{2}[2\lambda + (\underline{\beta} - \underline{\mu})' A(\underline{\beta} - \underline{\mu}) \right. \\ &\quad \left. + (Y - X\underline{\beta})'(Y - X\underline{\beta})]\right\}. \end{aligned} \quad (2.6)$$

Dado que

$$\begin{aligned} &-\frac{\tau}{2}[2\lambda + (\underline{\beta} - \underline{\mu})' A(\underline{\beta} - \underline{\mu}) + (Y - X\underline{\beta})'(Y - X\underline{\beta})] = \\ &-\tau\lambda - \frac{\tau}{2}[\underline{\beta}' A \underline{\beta} - 2\underline{\mu}' A \underline{\beta} + \underline{\mu}' A \underline{\mu} + Y'Y - 2Y'X\underline{\beta} + \underline{\beta}' X'X \underline{\beta}] \end{aligned}$$

y completando en el lado derecho de la igualdad anterior una forma cuadrática en $(\underline{\beta} - (A + (X'X)^{-1}(A\underline{\mu} + X'Y)))$, se obtiene

$$\begin{aligned} \pi(\underline{\beta}, \tau | Y) &\propto \\ &\tau^{\frac{n+2\alpha}{2}-1} \exp\left\{-\tau \left\{ \lambda + \frac{Y'Y - (A\underline{\mu} + X'Y)'(A + X'X)^{-1}(A\underline{\mu} + X'Y)}{2} \right\} \right\} \\ &\tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2}[\underline{\beta} - (A + X'X)^{-1}(A\underline{\mu} + X'Y)]'(A + X'X) \right. \\ &\quad \left. [\underline{\beta} - (A + X'X)^{-1}(A\underline{\mu} + X'Y)]\right\}. \end{aligned}$$

Por lo que $\pi(\underline{\beta}, \tau|Y)$ es una distribución Normal-Gamma con parámetros

$$\begin{aligned}\alpha_1^* &= \frac{n + 2\alpha}{2} \\ \lambda_1^* &= \lambda + \frac{Y'Y - (A\underline{\mu} + X'Y)'(A + X'X)^{-1}(A\underline{\mu} + X'Y)}{2} \\ \underline{\mu}_1^* &= (X'X + A)^{-1}(X'Y + A\underline{\mu}) \\ A_1^* &= (X'X + A).\end{aligned}\tag{2.7}$$

Para obtener la distribución final de $\underline{\beta}$ basta integrar la distribución anterior con respecto a τ , es decir

$$\pi(\underline{\beta}|Y) \propto \int \pi(\underline{\beta}, \tau|Y) d\tau.$$

Completando dentro de la integral una Gamma con parámetros

$$\begin{aligned}\lambda_2 &= \frac{2\lambda_1^* + (\underline{\beta} - \underline{\mu}_1^*)'A_1^*(\underline{\beta} - \underline{\mu}_1^*)}{2} \\ \alpha_2 &= \alpha_1^*\end{aligned}$$

se obtiene $\pi(\underline{\beta}|Y) \propto [2\lambda_1^* + (\underline{\beta} - \underline{\mu}_1^*)'A_1^*(\underline{\beta} - \underline{\mu}_1^*)]^{-\frac{n+2\alpha+2}{2}}$ por lo que $\pi(\underline{\beta}|Y)$ es una densidad t p -dimensional con $n + 2\alpha$ grados de libertad, vector de localización

$$\underline{\mu}^* = \underline{\mu}_1^*$$

y matriz de precisión

$$\begin{aligned}A^* &= A_1^*(n + 2\alpha)(2\lambda_1^*)^{-1} \\ &= A_1^*\alpha_1^*(\lambda_1^*)^{-1}\end{aligned}$$

Por otro lado, considerando la forma de $\pi(\underline{\beta}, \tau|Y)$, se tiene que $\pi(\tau|Y)$ es una Gamma con parámetros α_1^*, λ_1^* .

2.3.1 Estimación puntual de los parámetros del modelo

Estimación puntual sin función de utilidad

Una vez determinadas las distribuciones finales de los parámetros $\underline{\beta}$ y τ se está en condiciones de obtener sus correspondientes estimaciones.

La técnica clásica más común para estimar puntualmente un parámetro θ , es considerar estimaciones máximo verosimiles, las cuales seleccionan como estimador de θ el valor $\hat{\theta}$ tal que maximice la función de verosimilitud $f(Y|\theta)$. El estimador bayesiano análogo se define como sigue:

Definición 2.3.1 *El estimador de máxima verosimilitud generalizado de θ es la moda, $\hat{\theta}$, de $\pi(\theta|Y)$ (es decir el valor $\hat{\theta}$ el cual maximiza $\pi(\theta|Y)$ considerada como una función de θ) (Berger, 1985).*

De esta manera $\hat{\theta}$ se puede interpretar como el valor más parecido de θ dada la distribución inicial y la muestra Y .

Otros estimadores bayesianos de θ pueden ser la media o la mediana de $\pi(\theta|Y)$. La media, moda y mediana son, en general, relativamente fáciles de encontrar cuando la distribución inicial y por lo tanto la distribución final pertenece a una familia conjugada de distribuciones. Si es posible se deben calcular y comparar los tres estimadores en un análisis bayesiano, especialmente considerando su robustez ante cambios de la distribución inicial, es decir, que ante pequeños cambios en la especificación de los parámetros de la distribución inicial se obtengan pequeños cambios en las inferencias de interés.

Así, si se quiere estimar $\underline{\beta}$ y se considera a τ un parámetro de ruido, entonces $\pi(\underline{\beta}|Y)$ resulta una Normal-Gamma la cual es simétrica y por lo tanto se tiene que la media, la moda y la mediana coinciden, por lo que

$$\begin{aligned} E_{\pi(\underline{\beta}|Y)}(\underline{\beta}) &= (X'X + A)^{-1}(X'Y + A\underline{\mu}) \\ &= \underline{\mu}^* \end{aligned}$$

es un estimador de $\underline{\beta}$.

Una vez obtenida la distribución final de τ , se pueden obtener las estimaciones puntuales correspondientes. Sin embargo, como τ es el inverso de σ^2 resulta de interés realizar inferencias sobre τ^{-1} . Por ello en este trabajo, sólo se presentan desarrollos inferenciales sobre el parámetro τ^{-1} , así también es conveniente mencionar que en el caso de que se desee hacer inferencias sobre τ estas se pueden realizar de manera análoga a las de τ^{-1} .

Tomando en cuenta que la distribución final de τ , $\pi(\tau|Y)$, es una f.d.p Gamma con parámetros α_1^* y λ_1^* , definidas en las ecuaciones 2.7, se tiene que

$$\pi(\tau^{-1}|Y) = \frac{\lambda_1^{*\alpha_1^*}}{\Gamma(\alpha_1^*)} (\tau^{-1})^{-(\alpha_1^*+1)} \exp\left\{-\frac{\lambda_1^*}{\tau^{-1}}\right\} \quad (2.8)$$

la cual es una f.d.p Gamma inversa con parámetros $(\alpha_1^*, \lambda_1^*)$. Dado que esta distribución no necesariamente es simétrica, el estimador de τ^{-1} se podrá elegir entre la media, moda o mediana de $\pi(\tau^{-1}|Y)$. En este caso, la media esta dada por

$$\begin{aligned} E_{\pi(\tau^{-1}|Y)}(\tau^{-1}) &= \int_0^\infty \tau^{-1} \pi(\tau^{-1}|Y) d\tau \\ &= \lambda_1^* (\alpha_1^* - 1)^{-1} \\ &= \frac{2\lambda_1^*}{n + 2\alpha_1^* - 2} \end{aligned} \quad (2.9)$$

Para determinar la moda de τ^{-1} basta maximizar 2.8 lo cual da como resultado que la moda sea

$$\begin{aligned} \hat{\tau}^{-1} &= (\alpha_1^* + 1)^{-1} \lambda_1^* \\ &= \frac{2\lambda_1^*}{n + 2\alpha_1^* + 2} \end{aligned} \quad (2.10)$$

Como se puede apreciar, para n suficientemente grande, la esperanza y la moda de τ^{-1} prácticamente son iguales.

Por otro lado la mediana q se puede obtener calculando el cuantil de orden .5 de la Gamma inversa (2.8), o bien, considerando que $\pi(\tau|Y)$ es una Gamma(α_1^*, λ_1^*) entonces $q = \frac{1}{\hat{\tau}}$ donde q^1 es la mediana de una Gamma(α_1^*, λ_1^*).

Estimación puntual con funciones de utilidad

Otra forma de estimar puntualmente los parámetros del MRLM es considerar este problema como un problema de decisión y utilizar el principio de la utilidad esperada máxima visto en el capítulo anterior.

Como una generalización de la utilidad cuadrática se tiene

$$U(\underline{\beta}, \hat{\underline{\beta}}) = -(\underline{\beta} - \hat{\underline{\beta}})' Q (\underline{\beta} - \hat{\underline{\beta}}) \quad (2.11)$$

con Q una matriz conocida definida positiva y simétrica de $p \times p$ y $\hat{\underline{\beta}}$ un estimador de $\underline{\beta}$. Considerando esta función de utilidad se sabe que la mejor opción para estimar $\underline{\beta}$, es encontrar $\hat{\underline{\beta}}$ que maximice

$$E_P[U(\underline{\beta}, \hat{\underline{\beta}})] = E_P[-(\underline{\beta} - \hat{\underline{\beta}})' Q (\underline{\beta} - \hat{\underline{\beta}})] \quad (2.12)$$

con P una distribución de probabilidad sobre $\underline{\beta}$. Desarrollando la forma cuadrática y tomando los correspondientes valores esperados

$$\begin{aligned} E_P[U(\underline{\beta}, \hat{\underline{\beta}})] &= -E_P[\underline{\beta}' Q \underline{\beta}] + E_P[\hat{\underline{\beta}}' Q \underline{\beta} + \underline{\beta}' Q \hat{\underline{\beta}}] - E_P[\hat{\underline{\beta}}' Q \hat{\underline{\beta}}] \\ &= -E_P[\underline{\beta}' Q \underline{\beta}] + 2E_P[\underline{\beta}' Q \hat{\underline{\beta}}] - E_P[\hat{\underline{\beta}}' Q \hat{\underline{\beta}}], \end{aligned}$$

de donde

$$\frac{\partial E_P(U(\underline{\beta}, \hat{\underline{\beta}}))}{\partial \hat{\underline{\beta}}} = 2(Q E_P(\underline{\beta})) - 2Q \hat{\underline{\beta}}$$

igualando a cero esta última ecuación finalmente se tiene que

$$\hat{\underline{\beta}} = E_P(\underline{\beta}).$$

De lo anterior, se puede concluir que el estimador óptimo de $\underline{\beta}$ cuando no se tienen datos es

$$\hat{\underline{\beta}} = E_{\pi(\underline{\beta})}(\underline{\beta}) = \underline{\mu}$$

y con datos

$$\hat{\underline{\beta}} = E_{\pi(\underline{\beta}|Y)}(\underline{\beta}) = \underline{\mu}^*.$$

Otra función de utilidad que se puede usar para estimar $\underline{\beta}$ es

$$U(\underline{\beta}, \hat{\underline{\beta}}) = -\delta(\underline{\beta}, \tau; \hat{\underline{\beta}}, \tau) \quad (2.13)$$

donde

$$\delta(\underline{\beta}, \tau; \hat{\underline{\beta}}, \tau) = \int \pi(y|\underline{\beta}, \tau) \log \frac{\pi(y|\underline{\beta}, \tau)}{\pi(y|\hat{\underline{\beta}}, \tau)} dY,$$

es la llamada divergencia logarítmica de Kullback-Leibler entre $\pi(y|\underline{\beta}, \tau)$ y $\pi(y|\hat{\underline{\beta}}, \tau)$. Es razonable usar 2.13 como función de utilidad para estimar $\hat{\underline{\beta}}$, ya que ésta representa la discrepancia entre $\pi(y|\underline{\beta}, \tau)$ y $\pi(y|\hat{\underline{\beta}}, \tau)$. De esta manera la utilidad de elegir $\hat{\underline{\beta}}$ aumentará conforme disminuya la discrepancia entre los modelos involucrados.

En este caso el estimador de $\underline{\beta}$ resulta también ser la media de su distribución final, es decir $\hat{\underline{\beta}} = \underline{\mu}^*$. (Gutiérrez, 1991)

Ahora para estimar τ^{-1} considerando una función de utilidad cuadrática

$$U(\tau^{-1}, \hat{\tau}^{-1}) = -(\hat{\tau}^{-1} - \tau^{-1})^2,$$

la mejor decisión para estimar τ^{-1} , sin incorporar información muestral, es encontrar $\hat{\tau}^{-1}$ que maximice

$$E_{\pi(\tau)}(U(\tau^{-1}, \hat{\tau}^{-1}))$$

e incorporando información muestral es $\hat{\tau}^{-1}$ que maximice

$$E_{\pi(\tau|Y)}(U(\tau^{-1}, \hat{\tau}^{-1})).$$

Siguiendo un procedimiento similar al que se efectuó con $\underline{\beta}$, se tiene que cuando no hay datos, el estimador óptimo para τ^{-1} es

$$\hat{\tau}^{-1} = E_{\pi(\tau)}(\tau^{-1}) = \lambda(\alpha - 1)^{-1}$$

y con datos

$$\hat{\tau}^{-1} = E_{\pi(\tau|Y)}(\tau^{-1}) = \lambda_1^*(\alpha_1^* - 1)^{-1}$$

Usando la divergencia de Kullback-Leibler como función de utilidad, para estimar a τ^{-1} , la cual es de la forma

$$U(\tau^{-1}, \hat{\tau}^{-1}) = -\delta(\tau^{-1}; \hat{\tau}^{-1}).$$

Se tiene que el estimador óptimo de τ^{-1} es la moda de su distribución final, es decir

$$\hat{\tau}^{-1} = \lambda_1^*(\alpha_1^* + 1)^{-1}$$

(Gutiérrez, 1991).

Como se puede observar los estimadores obtenidos en esta sección, considerando ciertas funciones de utilidad, como la cuadrática y la de la divergencia logarítmica resultaron ser la media y la moda, de la respectiva distribución final, los cuales coinciden con los estimadores propuestos inicialmente. Sin embargo, hay que notar que usando otras funciones de utilidad es posible que los estimadores resultantes sean diferente a los obtenidos en esta sección.

2.3.2 Intervalos de Probabilidad

En la sección anterior se obtuvieron estimaciones puntuales de los parámetros desconocidos del MRLM. Sin embargo, una estimación puntual sobre algún parámetro desconocido, generalmente, no es de mucha utilidad ya que es improbable que la estimación puntual sea igual al verdadero valor del parámetro. Es más razonable considerar que el verdadero valor del parámetro esté contenido en un intervalo.

En el capítulo anterior se mostró cómo los problemas estadísticos de estimación puntual y contraste de hipótesis se pueden plantear como problemas de decisión. El problema de intervalos de probabilidad no es la excepción, en este caso, el espacio de decisiones D es el conjunto de todos los posibles intervalos de probabilidad para el parámetro en cuestión que tengan ciertas características de interés para el decisor. El conjunto de sucesos inciertos está representado por el espacio parametral Θ , y la función de utilidad será una medida que cuantifique las consecuencias de acuerdo a las preferencias de cada decisor.

Si se considera como primera opción un espacio de decisiones formado por cualquier intervalo o región, A , para θ y una función de utilidad que de uno si θ pertenece al intervalo y cero si no pertenece al intervalo, es decir $U(A, \theta) = I_A(\theta)$. Se puede ver que la solución a este problema de decisión sería la región formada por todo el espacio parametral, la cual no es muy informativa.

Otra forma de plantear este problema es tomando el espacio de decisiones D y la función de utilidad U como

$$D = \left\{ \begin{array}{l} A: A \text{ cualquier intervalo o región para } \theta \text{ con} \\ P_{\pi(\theta|Y)}(A, \theta) = 1 - \gamma \end{array} \right\}$$

$$U(A, \theta) = I_A(\theta)$$

ya que no sólo es de interés que el intervalo o región contenga a θ si no, que lo contenga con cierta probabilidad $(1 - \alpha)$. En este caso se tiene que la decisión óptima es aquella que maximiza

$$\begin{aligned} E_{\pi(\theta|Y)}[U(A, \theta)] &= \int_{\Theta} I_A(\theta) \pi(\theta|Y) d\theta \\ &= P_{\pi(\theta|Y)}(\theta \in A) = 1 - \gamma \end{aligned}$$

como todo elemento de D cumple con la propiedad $P_{\pi(\theta|Y)}(\theta \in A) = 1 - \gamma$ se tiene que cualquier elemento de D es solución óptima.

Como se puede ver, en este caso no hay una manera única de elegir algún intervalo o región, cuando esto sucede debe decidirse que propiedades necesita cumplir una región para ser seleccionada. Como se ha mencionado, desde el punto de vista bayesiano, todos los problemas de inferencia concernientes a los parámetros del MRLM pueden resolverse a través del análisis de las correspondientes distribuciones finales. Así, es razonable pedir que la región seleccionada tenga la propiedad que la densidad final para cada punto dentro de la región, sea al menos tan grande como para cualquier punto fuera de ella. Una segunda propiedad deseable es que para un contenido de probabilidad dado, la región fuera la más pequeña posible. Se puede ver que ambas propiedades son equivalentes. Por lo que finalmente si se considera que la utilidad será mayor a medida que el volumen de la región disminuya, se tiene que el problema de regiones de probabilidad para θ , visto como un problema de decisión, se puede plantear como

$$D = \{A : A \text{ es cualquier región para } \theta \text{ con } P_{\pi(\theta|Y)}(\theta \in A) = 1 - \gamma\}$$

$$U(A, \theta) = I_A(\theta) - V(A) \quad (2.14)$$

donde $V(A)$ es la longitud (para $A \in \mathfrak{R}$), el área (para $A \in \mathfrak{R}^2$), el volumen (para $A \in \mathfrak{R}^3$), etc., de la región o intervalo. La solución óptima en este caso está determinada a través de

$$\begin{aligned} \max_D E_{\pi(\theta|Y)}(U(A, \theta)) &= \max_D [P(\theta \in A) - V(A)] \\ &= (1 - \gamma) - \min_D V(A) \end{aligned}$$

por lo que la solución del problema se reduce a encontrar aquel intervalo o región de volumen mínimo que contenga a θ con probabilidad $1 - \gamma$. En

este tipo de regiones todo punto tendrá una densidad de probabilidad mayor que cualquier punto excluido, comúnmente a estas regiones se les llama *regiones de alta densidad final o regiones HPD*, por lo que la decisión óptima se reduce a encontrar regiones HPD

Box y Tiao (1973) definen las regiones HPD de la siguiente manera:

Definición 2.3.2 Sea $\pi(\theta|Y)$ una función de densidad de probabilidad final de θ . Una región R contenida en el espacio parametral de θ , es llamada una región HPD de contenido $(1-\gamma)$ para θ si:

- a) $P\{\theta \in R|Y\} = 1 - \gamma$
- b) Si $\theta_1 \in R$ y $\theta_2 \in R^c$,

entonces

$$\pi(\theta_1|Y) \geq \pi(\theta_2|Y)$$

Algunas propiedades (Box y Tiao, 1973) de las regiones HPD que se siguen de la definición son:

1. Para un contenido de probabilidad $(1 - \gamma)$, la región HPD tiene el volumen más pequeño posible en el espacio parametral de θ .

2. R es una región HPD única si la densidad final de θ no es uniforme sobre toda la región del espacio parametral.

3. Sean θ_1, θ_2 dos puntos en Θ tales que $\pi(\theta_1|Y) = \pi(\theta_2|Y)$, entonces estos dos puntos están simultáneamente incluidos (o no incluidos) en una región HPD de contenido $(1-\gamma)$. Ahora si $\pi(\theta_1|Y) \neq \pi(\theta_2|Y)$, entonces existe una región HPD $(1-\gamma)$ la cual incluye un punto pero no al otro.

Regresando al modelo general de regresión lineal múltiple, el interés será construir regiones HPD $(1-\gamma)$ para los parámetros desconocidos $\underline{\beta}$ y τ del modelo.

Para el vector de parámetros $\underline{\beta}$ se encontró su distribución final $\pi(\underline{\beta}|Y)$; como esta distribución es una t p -dimensional con $(2\alpha + n)$ grados de libertad, vector de localización $\underline{\mu}^*$ y matriz de precisión A^* , los contornos de nivel de $\pi(\underline{\beta}|Y)$ en el espacio parametral de $\underline{\beta}$ son elipses ($p=2$), elipsoides ($p=3$) o hiperelipsoides ($p > 3$) definidos por

$$F(\underline{\beta}) = p^{-1}(\underline{\beta} - \underline{\mu}^*)' A^* (\underline{\beta} - \underline{\mu}^*). \quad (2.15)$$

Además, estos contornos delimitan una región de cierto contenido $(1 - \gamma)$ de probabilidad para $\underline{\beta}$

$$R_{1-\gamma}(\underline{\beta}) = \{\underline{\beta} : F(\underline{\beta}) \leq K(1 - \gamma)\}.$$

Ahora, como la distribución t p -dimensional es simétrica, unimodal, con colas monótonamente decrecientes, la región $R_{1-\gamma}(\underline{\beta})$ es una región HPD de contenido $(1 - \gamma)$. Lo único que falta por determinar es $K(1 - \gamma)$ que delimite completamente la región HPD.

Como $\underline{\beta}$ tiene una distribución t p -dimensional, se puede escribir a $\underline{\beta}$ de la siguiente manera

$$\underline{\beta} = \underline{\mu}^* + X^{-1/2}((2\alpha + n)A^*)^{1/2}Z \quad (2.16)$$

con X una variable aleatoria que se distribuye $\chi^2_{(2\alpha+n)}$ y Z un vector aleatorio que se distribuye $N_p(0, I_p)$.

Considerando lo anterior se tiene que

$$p^{-1}(\underline{\beta} - \underline{\mu}^*)' A^* (\underline{\beta} - \underline{\mu}^*) = \frac{2\alpha + n}{p} \frac{Z'Z}{X}$$

es una v.a. que tiene una distribución F con p y $2\alpha + n$ grados de libertad (Muirhead, 1982) por que $Z'Z \sim \chi^2_{(p)}$. Así

$$R_{1-\gamma}(\underline{\beta}) = \{\underline{\beta} : F(\underline{\beta}) \leq F_{p,2\alpha+n,1-\gamma}\}$$

determina una región HPD de contenido $(1 - \gamma)$ para $\underline{\beta}$, donde $F_{(p,2\alpha+n,1-\gamma)}$ es el percentil superior de orden $1 - \gamma$ de una distribución F con p y $2\alpha + n$ grados de libertad. Una vez encontrada la región HPD para $\underline{\beta}$, resta encontrar una región HPD para τ^{-1} .

Ya que $\pi(\tau|Y)$ es una densidad $\text{Gamma}(\alpha_1^*, \lambda_1^*)$ se tiene que

$$2\lambda_1^*(\tau^{-1})^{-1} \sim \chi^2_{(2\alpha_1^*)}$$

por lo que

$$\left(\frac{2\lambda_1^*}{\chi^2_{(2\alpha_1^*, 1-\frac{\gamma}{2})}}, \frac{2\lambda_1^*}{\chi^2_{(2\alpha_1^*, \frac{\gamma}{2})}} \right) \quad (2.17)$$

es un intervalo de contenido $(1 - \gamma)$ de probabilidad final para τ^{-1} , con $(\chi^2_{2\alpha_1^*, \frac{\gamma}{2}})$, $(\chi^2_{2\alpha_1^*, 1 - \frac{\gamma}{2}})$ los porcentiles de orden $(\frac{\gamma}{2})$ y $(1 - \frac{\gamma}{2})$ de una v.a. chi - cuadrada con $2\alpha_1^*$ grados de libertad. Como la f.d.p. Gamma inversa no es necesariamente simétrica entonces, si se requiere de una región HPD de contenido $(1 - \gamma)$ para τ^{-1} , se tendrán que emplear métodos numéricos, ya que es necesario encontrar a y b tal que

$$\pi_{(\tau^{-1}|Y)}(a) \doteq \pi_{(\tau^{-1}|Y)}(b)$$

Donde $\pi_{(\tau^{-1}|Y)}(\cdot)$ es la f.d.p final de τ^{-1} , ecuación (2.8), de esta forma quedará caracterizada la región H.P.D para τ^{-1} .

Otra forma de obtener un intervalo HPD de contenido $1 - \gamma$ de probabilidad para τ^{-1} , es considerando la f.d.p $\pi(\tau|Y)$. En este caso se tendrían que encontrar a' y b' tal que

$$P(a' < \tau < b') = P\left(\frac{1}{b'} < \tau^{-1} < \frac{1}{a'}\right) = 1 - \gamma$$

y cuya longitud del intervalo $L = \frac{1}{a'} - \frac{1}{b'}$ fuera mínima. Lo cual implica minimizar

$$L = \frac{1}{a'} - \frac{1}{b'}$$

sujeto a

$$\int_{a'}^{b'} \pi(\tau|Y) = 1 - \gamma$$

Resolviendo este problema de minimización con restricción, se tiene que

a' y b' son solución si

$$\begin{aligned} (a')^2 \pi_{(\tau|Y)}(a') &\doteq (b')^2 \pi_{(\tau|Y)}(b') \\ \text{y} \quad \int_{a'}^{b'} \pi(\tau|Y) &= 1 - \gamma. \end{aligned} \quad (2.18)$$

Así, la región HPD, de contenido de probabilidad $1 - \gamma$, para τ^{-1} estará dada por

$$\left(\frac{1}{b'}, \frac{1}{a'}\right)$$

donde a' y b' cumplen las condiciones 2.18.

2.3.3 Prueba de Hipótesis

Sea Y una variable aleatoria con función de densidad $\pi(Y|\theta)$ y supóngase que existe una familia de densidades de probabilidad

$$F = \{\pi(Y|\theta) : \theta \in \Theta\} \quad \Theta \subseteq \mathfrak{R}^p$$

que contiene a la densidad de probabilidad de Y , es decir, $\pi(Y|\theta)$ es conocida excepto por un número finito de parámetros.

Sean $\Theta_0 \subset \Theta$ y $\Theta_1 \subset \Theta$ tales que $\Theta_0 \cap \Theta_1 = \emptyset$ y sea $Y = \{y_1, \dots, y_n\}$ un conjunto de observaciones de la variable Y . El problema general de contraste de hipótesis

$$H_0 : \pi \in F_0 \quad \text{vs.} \quad H_1 : \pi \in F_1$$

donde

$$F_i = \{\pi(Y|\theta) : \theta \in \Theta_i\} \quad (i = 0, 1)$$

es equivalente a

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1. \quad (2.19)$$

En problemas de regresión lineal múltiple ciertas pruebas de hipótesis acerca de los parámetros del modelo son útiles para medir la adecuación del modelo. Una de las pruebas que es importante realizar es aquella para determinar si alguna variable independiente influya para explicar la variable respuesta. En este caso las hipótesis apropiadas son

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \beta_j \neq 0 \quad \text{para al menos una } j$$

Rechazar H_0 implica que al menos una variable independiente contribuye a explicar a la variable respuesta. Comunmente esta prueba se rechaza, ya que de lo contrario se pensaría que la elección de las variables independientes fué realizada arbitrariamente, sin considerar su posible influencia sobre la variable respuesta.

Por otro lado, son de interés las pruebas de hipótesis sobre los coeficientes de regresión individuales. Las hipótesis para probar la significancia de cualquier coeficiente de regresión, β_j , son de la forma

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0.$$

Si H_0 se rechaza, esto indica que la variable independiente X_j puede ser eliminada del modelo, ya que no contribuye a explicar a la variable respuesta.

Muchas hipótesis, incluyendo las anteriores, acerca de los coeficientes de regresión pueden probarse considerando la prueba de hipótesis general

$$H_0 : M\beta = b \quad \text{vs.} \quad H_1 : M\beta \neq b \quad (2.20)$$

donde M es una matriz de $k \times p$ de rango k ($1 \leq k \leq p$) y b vector de constantes conocidas de $k \times 1$.

Como se mencionó en el capítulo uno el problema de contraste de hipótesis puede plantearse como un problema de decisión donde

$D = \{d_0, d_1\}$, con d_i la decisión a elegir H_i , $i = 0, 1$
 Θ , es igual al espacio parametral, que define el espacio de sucesos inciertos.

$U(d, \theta)$, es una función de utilidad que cuantifica las consecuencias de tomar la decisión d , de acuerdo a las preferencias del decisor.

Una vez planteado el problema de contraste de hipótesis como un problema de toma de decisiones, se tiene que la decisión óptima consiste en elegir H_i si y sólo si

$$E_{\pi(\theta|Y)}(U(d_j, \theta)) < E_{\pi(\theta|Y)}(U(d_i, \theta)) \\ \text{con } i \neq j \quad i, j = 0, 1$$

Por lo que finalmente para resolver el problema de prueba de hipótesis, considerándolo como un problema de decisión, resta determinar una función de utilidad que estará dada por el decisor, y la correspondiente distribución final de θ .

Así, para realizar el contraste de hipótesis general para el parámetro de regresión β , ecuación 2.20, considérese la siguiente familia de funciones de utilidad

$$U(d_0, M\beta) = \begin{cases} a_0 & \text{si } M\beta = b \\ 0 & \text{si } M\beta \neq b \end{cases} \quad (2.21)$$

$$U(d_1, M\beta) = \begin{cases} a_1(M\beta) & \text{si } M\beta \neq b \\ 0 & \text{si } M\beta = b \end{cases} \quad (2.22)$$

con a_0 y a_1 funciones no negativas de $M\beta$.

Como la distribución final de β es una t p -dimensional. También $\underline{\nu} = M\beta$ tiene una f.d.p $\pi(\underline{\nu}|Y)$, t k -dimensional con $n + 2\alpha$ grados de libertad, media $M\underline{\mu}^*$ y matriz de precisión $(M(A^*)^{-1}M^t)^{-1}$ de $k \times k$. M es tal que $M(A^*)^{-1}M^t$ es no singular ya que de lo contrario la distribución final de $\underline{\nu}$ sería impropia.

Como $\underline{\nu} = M\beta$ es una variable aleatoria continua se tiene que

$$\begin{aligned} E_{\pi(\underline{\nu}|Y)}(U(d_0, \underline{\nu})) &= a_0 \pi(\underline{\nu} = b|Y) \\ E_{\pi(\underline{\nu}|Y)}(U(d_1, \underline{\nu})) &= \int_{\{\underline{\nu}: \underline{\nu} \neq b\}} a_1(\underline{\nu}) \pi(\underline{\nu}|Y) d\underline{\nu} \end{aligned}$$

Sin embargo, ya que la distribución de $\underline{\nu}$ es continua $P(\underline{\nu} = b|Y) = 0$ para todo Y , y nunca se tendría suficiente evidencia para no rechazar H_0 .

En general cuando θ es un vector aleatorio continuo y se quiere contrastar las hipótesis

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0$$

se consideran familias de distribuciones iniciales de la forma

$$\pi_0(\theta) = \begin{cases} q & \text{si } \theta = \theta_0 \\ (1 - q)f(\theta) & \text{si } \theta \neq \theta_0 \end{cases}$$

con $f(\theta)$ una función de densidad definida sobre $\Theta_1 = \Theta - \{\theta_0\}$, ya que de esta forma se evita que $P(\theta = \theta_0)$ sea cero, por ser θ v. a. continua. Es decir, lo único que se hace es asignarle una probabilidad al evento $\theta = \theta_0$ y el complemento distribuirlo en el resto del espacio parametral. En este caso tomando $f(\underline{\nu}) = \pi(\underline{\nu})$ se puede considerar a

$$\pi_0(\underline{\nu}) = \begin{cases} q & \text{si } \underline{\nu} = b \\ (1 - q)\pi(\underline{\nu}) & \text{si } \underline{\nu} \neq b \end{cases} \quad (2.23)$$

como una distribución inicial de ν , entonces se rechazará H_0 si y sólo si

$$E_{\pi_0(\nu|Y)}[U(d_0, \nu)] < E_{\pi_0(\nu|Y)}[U(d_1, \nu)]$$

con $\pi_0(\nu|Y)$ la distribución final de ν , es decir, si

$$a_0 \pi_0(b|Y) < \int_{\Theta_1} a_1(\nu) \pi_0(\nu|Y) d\nu$$

$$\frac{a_0 \pi_0(b) f(Y|\nu = b)}{\int_{\Theta_1} a_1(\nu) \pi_0(\nu) f(Y|\nu) d\nu} < 1$$

usando la ecuación 2.23 se tiene que

$$\frac{a_0 q f(Y|\nu = b)}{\int_{\Theta_1} a_1(\nu) (1 - q) \pi(\nu) f(Y|\nu) d\nu} < 1$$

por lo que finalmente se rechazará H_0 si y sólo si

$$\frac{f(Y|\nu = b)}{\int_{\Theta_1} a_1(\nu) \pi(\nu) f(Y|\nu) d\nu} < \frac{1 - q}{a_0 q}$$

Como $a_1(\nu)$ es cualquier función de ν , en particular se pueden considerar funciones de utilidad de la forma 4.3, con $a_1(\nu)$ una función cuadrática. Esto porque puede considerarse que la utilidad de elegir d_1 (rechazar H_0) será menor cuando el valor de ν sea más próximo a b .

Propuestas y alternativas

Considerando que el problema de contraste de hipótesis (2.20) se puede plantear como un problema de toma de decisiones, en donde la decisión d_1 (rechazar H_0) será preferible a la decisión d_0 (aceptar H_0) si y sólo si

$$E_{\pi(\theta|Y)}(U(d_1, \theta)) > E_{\pi(\theta|Y)}(U(d_0, \theta))$$

o, equivalentemente si

$$\int \{U(d_1, \theta) - U(d_0, \theta)\} \pi(\theta|Y) d\theta > 0,$$

se puede pensar que sólo basta especificar la diferencia de las utilidades respectivas para considerar la decisión óptima.

En Bernardo (1984) se sugiere que la ventaja de elegir d_1 debe depender solamente de la discrepancia entre $\pi(x|\theta)$ y $\pi(x|\theta_0)$. Por lo que si se denota por $\delta(\theta, \theta_0)$ a una medida de la discrepancia entre $\pi(x|\theta)$ y $\pi(x|\theta_0)$, puede considerarse a la diferencia $\{U(d_1, \theta) - U(d_0, \theta)\}$ como una función no decreciente de $\delta(\theta; \theta_0)$.

Por otro lado, en Gutiérrez (1991) se propone una función de utilidad de la forma

$$U(d_i, \theta) = B_i - C\delta(\theta; \theta_i^*) \quad C \in \mathbb{R}^+, B_i \in \mathbb{R}$$

donde $\theta_i^* \in \Theta_i$ es un valor representativo de Θ_i ; ($i = 0, 1$)

En estas condiciones, la acción d_1 será preferible a la decisión d_0 si y sólo si

$$E(\delta(\theta; \theta_0^*)|Y) - E(\delta(\theta; \theta_1^*)|Y) > \frac{B_0 - B_1}{C}.$$

Considerando como medida de discrepancia $\delta(\theta, \theta_0)$, la divergencia logarítmica de Kullback-Leibler, y una inicial de la forma 2.23 con $\pi(\underline{\mu})$ una t-multivariada($n+2\alpha$, $M\mu^*$, $(MA^{*-1}M^t)^{-1}$), para el problema de contraste de hipótesis general del MRLM, definido anteriormente, se tiene que la hipótesis H_0 debe ser rechazada si y sólo si

$$\alpha_1 \frac{[M\mu^* - b]^t [M(X^t X)^{-1} M^t]^{-1} [M\mu^* - b]}{2\lambda + Y^t [I_n - X(X^t X + A)^{-1} X^t] Y + R(Y, \underline{\mu}, A)} > \delta^* \quad (2.24)$$

donde

$$R(Y, \underline{\mu}, A) = \underline{\mu}^t [A + A(X^t X + A)^{-1} A] \underline{\mu} - 2\underline{\mu}^t A \mu^* \text{ y } \delta^* = \frac{(B_0 - B_1)}{C}$$

(Gutiérrez, 1991).

Otros procedimientos bayesianos

Hasta ahora para probar $H_0 : M\underline{\beta} = \underline{b}$ se ha utilizado el principio de la utilidad esperada máxima, que requiere de la especificación de una distribución inicial y de una función de utilidad. Si sólo se cuenta con $\pi(\underline{\beta})$ y se acepta la utilidad 2.14, es posible probar la hipótesis $M\underline{\beta} = \underline{b}$ mediante el uso de regiones HPD.

Así, si S es una región HPD de contenido $(1 - \gamma)$ para $M\underline{\beta}$, se rechazará la hipótesis nula cuando \underline{b} no pertenezca a S.

Como $\pi(\underline{\nu}|Y)$ es una distribución t k -dimensional con $n + 2\alpha$ grados de libertad, media $M\underline{\mu}^*$ y matriz de precisión $(M(A^*)^{-1}M^t)^{-1}$, la variable aleatoria

$$H(\underline{\nu}) = k^{-1}[G(\underline{\nu}) - M\underline{\mu}^*]'(M(A^*)^{-1}M^t)^{-1}[G(\underline{\nu}) - M\underline{\mu}^*],$$

donde $G(\underline{\nu}) = M\underline{\beta}$ es una función lineal de $\underline{\beta}$, tiene una distribución F con k y $n + 2\alpha$ grados de libertad (ver sección 2.3.2).

De esta manera

$$S_{1-\gamma}(\underline{\nu}) = \{\underline{\nu} : H(\underline{\nu}) \leq F_{k,n+2\alpha,1-\gamma}\}$$

determina una región HPD de contenido $1 - \gamma$ para $\underline{\beta}$, con $F_{(k,2\alpha+n,1-\gamma)}$ el percentil superior de orden $1 - \gamma$ de una distribución F con k y $n + 2\alpha$ grados de libertad. Por lo que la hipótesis nula se puede rechazar si \underline{b} no pertenece a $S_{1-\gamma}(g)$ o equivalentemente cuando $H(\underline{b}) > F_{k,n+2\alpha,1-\gamma}$.

2.4 Análisis Bayesiano no Informativo

En la sección 2.3 se presentó el análisis del modelo de regresión lineal múltiple (MRLM) de rango completo, considerando una distribución inicial Normal-Gamma(μ, A, α, λ) para los parámetros $\underline{\beta}$ y τ . En esta sección se presenta el análisis bayesiano correspondiente al MRLM de rango completo considerando lo que en la literatura se conoce como una distribución inicial no informativa.

Es posible hacer un análisis bayesiano, aún cuando no se cuente con información inicial o se esté limitado para especificarla, en este caso lo que se debe emplear es una distribución inicial *no informativa*, esto significa considerar una distribución inicial que contenga la mínima información sobre los parámetros de interés (o bien no tener preferencia sobre ningún valor posible de los parámetros). Por ejemplo, si se lanza una moneda, la distribución inicial que asigna una probabilidad de $(1/2)$ al evento de caer cualquiera de las caras, es claramente no informativa. Es común que la distribución inicial natural sea impropia. La situación más simple de considerar es cuando el espacio parametral Θ , es un conjunto finito que consta de n elementos. Aquí, la distribución inicial no informativa natural es aquella que da a cada elemento de Θ

una probabilidad de $(1/n)$. Cuando Θ es infinito la especificación de la distribución inicial no es tan natural, y se deben considerar otros argumentos. Sin embargo, la mayoría de los métodos útiles para determinar una inicial no informativa toman en consideración la estructura del problema que se esté trabajando.

La forma de considerar distribuciones iniciales no informativas surge con Jeffreys (ver por ejemplo Jeffreys, 1961).

La regla de Jeffreys para problemas multiparamétricos establece que la distribución inicial no informativa para un conjunto de parámetros se toma proporcional a la raíz cuadrada del determinante de la matriz de información de Fisher, es decir

$$\pi(\theta) \propto |I_n(\theta)|^{\frac{1}{2}}$$

donde $I_n(\theta)$ es la información de Fisher que está dada por

$$I_n(\theta) = -E\left(\frac{\partial^2 \ln f(Y|\theta)}{\partial \theta_i \partial \theta_j}\right)$$

Debido a que la regla de Jeffreys satisface ciertas propiedades de invarianza ante transformaciones uno a uno de los parámetros de interés y contiene la mínima información acerca de los parámetros (Bromeling, 1985; Box y Tiao, 1973), se realizará el análisis del MRLM de rango completo considerando la distribución inicial no informativa desarrollada por Jeffreys

$$\pi(\underline{\beta}, \tau) \propto \frac{1}{\tau} \quad \tau > 0, \underline{\beta} \in \mathbb{R}^p \quad (2.25)$$

la que implica que inicialmente $\underline{\beta}$ y τ son independientes, que $\underline{\beta}$ tiene una distribución inicial constante sobre todo \mathbb{R}^p y que la densidad marginal de τ es

$$\pi(\tau) \propto \frac{1}{\tau} \quad \tau > 0.$$

Dada la muestra $Y = \{y_1, \dots, y_n\}$ la función de verosimilitud, considerando el MRLM, modelo 2.3, es

$$f(Y|\underline{\beta}, \tau) = \left(\frac{\tau}{2\pi}\right)^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2}(Y - X\underline{\beta})'(Y - X\underline{\beta})\right\} \quad (2.26)$$

la cual se puede escribir como

$$f(Y|\underline{\beta}, \tau) \propto (s^2\tau)^{n/2} \exp\left\{-\frac{(n-p)s^2\tau}{2} - \frac{s^2\tau(\underline{\beta} - \hat{\underline{\beta}})'(X'X)(\underline{\beta} - \hat{\underline{\beta}})}{2s^2}\right\}$$

donde

$$s^2 = \frac{1}{(n-p)}(Y - \hat{Y})'(Y - \hat{Y}),$$

$$\hat{Y} = X\hat{\underline{\beta}} \quad \text{y}$$

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

por lo que considerando la distribución inicial 2.25 y el teorema de Bayes, se tiene que la distribución conjunta final para $\underline{\beta}$ y τ es

$$\pi(\underline{\beta}, \tau|Y) \propto \tau^{\frac{n-2}{2}-1+\frac{p}{2}} \exp\left\{-\frac{(n-p)s^2\tau}{2} - \frac{s^2\tau(\underline{\beta} - \hat{\underline{\beta}})'(X'X)(\underline{\beta} - \hat{\underline{\beta}})}{2s^2}\right\}$$

la cual es una distribución Normal-Gamma con parámetros

$$\begin{aligned} \hat{\underline{\beta}} &= (X'X)^{-1}X'Y \\ A_1^* &= (X'X) \\ \alpha_1^* &= \frac{n-p}{2} \\ \lambda_1^* &= \frac{(n-p)s^2}{2}. \end{aligned} \quad (2.27)$$

Dado que

$$\pi(\underline{\beta}, \tau|Y) = \pi(\tau|Y)\pi(\underline{\beta}|\tau, Y)$$

se tiene que $\pi(\underline{\beta}|\tau, Y)$ es una Normal($\hat{\underline{\beta}}, \tau A_1^*$) y $\pi(\tau|Y)$ es una Gamma con parámetros α_1^* y λ_1^* . Así, integrando $\pi(\underline{\beta}, \tau|Y)$ con respecto a τ se tiene

$$\pi(\underline{\beta}|Y) \propto \left(\frac{(n-p)s^2}{2} + \frac{(\underline{\beta} - \hat{\underline{\beta}})'(X'X)(\underline{\beta} - \hat{\underline{\beta}})}{2} \right)^{-\frac{n-p+2}{2}} \quad (2.28)$$

Es decir, la final de $\underline{\beta}$ es una t p -dimensional con $n - p$ grados de libertad, vector de localización $\hat{\underline{\beta}}$ y matriz de precisión

$$A^* = \frac{X'X}{s^2}$$

2.4.1 Estimación puntual de los parámetros del modelo

Estimación puntual sin función de utilidad

Si se quiere estimar $\underline{\beta}$ y se considera a τ un parámetro de ruido, entonces como $\pi(\underline{\beta}|Y)$ es una distribución simétrica, se tiene que la media, la moda y la mediana coinciden, por lo que

$$\begin{aligned} E_{\pi(\underline{\beta}|Y)}(\underline{\beta}) &= (X'X)^{-1}(X'Y) \\ &= \hat{\underline{\beta}} \end{aligned}$$

es un estimador de $\underline{\beta}$, el cual coincide con el estimador clásico de mínimos cuadrados.

Ahora, si se desea estimar τ^{-1} considerando a $\underline{\beta}$ como un parámetro de ruido, las inferencias estarán basadas en $\pi(\tau|Y)$, dado que esta distribución no necesariamente es simétrica, un estimador de τ^{-1} se podrá elegir entre la media, moda o mediana.

Así

$$\begin{aligned} E_{\pi(\tau|Y)}(\tau^{-1}) &= \lambda_1^*(\alpha_1^* - 1)^{-1} \\ &= \frac{(n-p)s^2}{n-p-2} \end{aligned} \quad (2.29)$$

y de acuerdo a la subsección 2.3.1 la moda de $\pi(\tau^{-1}|Y)$ es

$$(\alpha_1^* + 1)^{-1} \lambda_1^* = \frac{(n-p)s^2}{n-p+2}. \quad (2.30)$$

En este caso también, para n suficientemente grande, la esperanza y la moda de τ^{-1} prácticamente son iguales. La mediana q se puede obtener calculando el cuantil de orden .5 de la Gamma inversa $(\alpha_1^*, \lambda_1^*)$.

Estimación puntual con funciones de utilidad

Considerando el problema de estimación puntual como un problema de decisión se pueden realizar estimaciones de $\underline{\beta}$ utilizando funciones de utilidad.

Así, una estimación para $\underline{\beta}$ considerando una función de utilidad cuadrática de la forma

$$U(\underline{\beta}, \underline{\beta}^*) = -(\underline{\beta} - \underline{\beta}^*)' Q (\underline{\beta} - \underline{\beta}^*) \quad (2.31)$$

con Q una matriz simétrica definida positiva de $p \times p$ conocida, y $\underline{\beta}^*$ un estimador de $\underline{\beta}$ se tiene que el mejor estimador de $\underline{\beta}$ es

$$\underline{\beta}^* = E_{\pi(\underline{\beta}|Y)}(\underline{\beta}) = \hat{\underline{\beta}}. \quad (2.32)$$

Otra función de utilidad que se puede usar para estimar $\underline{\beta}$ es

$$U(\underline{\beta}, \underline{\beta}^*) = B - A \delta(\underline{\beta}, \tau; \underline{\beta}^*, \tau), \quad B \in \mathfrak{R}, A \in \mathfrak{R}^+ \quad (2.33)$$

donde $\delta(\underline{\beta}, \tau; \underline{\beta}^*, \tau)$ es la divergencia logarítmica de Kullback-Leibler entre $\pi(y|\underline{\beta}, \tau)$ y $\pi(y|\underline{\beta}^*, \tau)$. En este caso el estimador de $\underline{\beta}$ resulta también ser la media de la distribución final de $\underline{\beta}$, es decir $\underline{\beta}^* = \hat{\underline{\beta}}$ (Gutiérrez, 1991)

Ahora, para estimar τ^{-1} considerando la función de utilidad cuadrática

$$U(\tau^{-1}, \hat{\tau}^{-1}) = -(\hat{\tau}^{-1} - \tau^{-1})^2$$

se tiene que el estimador óptimo es

$$\hat{\tau}^{-1} = E_{\pi(\tau|Y)}(\tau^{-1}) = \lambda_1^*(\alpha_1^* - 1)^{-1}$$

y usando como función de utilidad a la divergencia de Kullback-Leibler para estimar τ^{-1} , se tiene que el estimador óptimo es la moda de la distribución final, es decir

$$\begin{aligned} \hat{\tau}^{-1} &= \lambda_1^*(\alpha_1^* + 1)^{-1} \\ &= \frac{(n-p)s^2}{n-p+2}. \end{aligned} \quad (2.34)$$

2.4.2 Intervalos de Probabilidad

Considerando que la distribución final de $\underline{\beta}$, $\pi(\underline{\beta}|Y)$, es una t p -dimensional con $(n-p)$ grados de libertad, vector de localización $\hat{\underline{\beta}}$ y matriz

de precisión A^* se tiene que una región HPD de contenido $1 - \gamma$ para $\underline{\beta}$ es

$$R_{1-\gamma}(\underline{\beta}) = \{\underline{\beta} : F(\underline{\beta}) \leq F_{p, n-p, 1-\gamma}\}$$

donde $F(\underline{\beta}) = p^{-1}(\underline{\beta} - \hat{\underline{\beta}})' A^* (\underline{\beta} - \hat{\underline{\beta}}) = (ps^2)^{-1}(\underline{\beta} - \hat{\underline{\beta}})' (X'X) (\underline{\beta} - \hat{\underline{\beta}})$ y $F_{(p, n-p, 1-\gamma)}$ es el percentil superior de orden $1 - \gamma$ de una distribución F con p y $n - p$ grados de libertad.

Una vez encontrada la región HPD para $\underline{\beta}$, resta encontrar una región HPD para τ^{-1} . Dado que $\pi(\tau|Y)$ es una densidad Gamma(α_1^* , λ_1^*) es posible sustituir los valores correspondientes en los desarrollos de la subsección 2.3.2 de este capítulo, y de esta forma obtener los intervalos HPD para τ^{-1} .

2.4.3 Prueba de Hipótesis

Para realizar el contraste de hipótesis lineal general

$$H_0 : M\underline{\beta} = \underline{b} \quad \text{vs} \quad H_1 : M\underline{\beta} \neq \underline{b} \quad (2.35)$$

con M una matriz de $k \times p$ de rango k ($1 \leq k \leq p$) y \underline{b} vector de constantes conocidas de $k \times 1$. Considerando la familia de funciones de utilidad de la forma 2.21 y 2.22, de la subsección 2.3.3 y una inicial de la forma 2.23, con $q = 1/2$, donde $\underline{\nu} = M\underline{\beta}$ y $\pi(\underline{\nu})$ es una distribución t k -dimensional con $n - p$ grados de libertad, media $M\hat{\underline{\beta}}$ y matriz de precisión $(MA^*M')^{-1}$ de $k \times k$. La regla de decisión consiste en rechazar H_0 si y sólo si

$$\frac{f(Y|\underline{b})}{\int_{\Theta_1} a_1(\underline{\nu}) \pi(\underline{\nu}) f(Y|\underline{\nu}) d\underline{\nu}} < \frac{1-q}{a_0 q}$$

Por otro lado, considerando la función de utilidad

$$U(d_i, \underline{\nu}) = B_i - C\delta(\underline{\nu}; \underline{\nu}_i^*) \quad C \in \mathfrak{R}^+, B_i \in \mathfrak{R}$$

donde $\underline{\nu}_i^* \in \Theta_i$ es un valor representativo de Θ_i ; ($i = 0, 1$) y $\delta(\underline{\nu}; \underline{\nu}_i^*)$ es la divergencia logarítmica de Kullback-Leibler. Para el problema de contraste 2.35 se tiene que la hipótesis H_0 se rechaza sí y sólo si

$$\frac{n-p [M\hat{\underline{\beta}} - \underline{b}]' [M(X'X)^{-1}M']^{-1} [M\hat{\underline{\beta}} - \underline{b}]}{2 Y' [I_n - X(X'X)^{-1}X'] Y} > \delta^* \quad (2.36)$$

donde $\delta^* = \frac{(B_0 - B_1)}{C}$ (Gutiérrez, 1991) o equivalentemente siempre y cuando

$$\frac{k}{2} F(Y) > \delta^*,$$

donde

$$F(Y) = \frac{n-p}{k} \frac{[M\hat{\beta} - b]' \{M(X'X)^{-1}M'\}^{-1} [M\hat{\beta} - b]}{Y' [I_n - X(X'X)^{-1}X'] Y}.$$

Obsérvese que $F(Y)$ es la estadística utilizada en el procedimiento clásico, cuya región de rechazo tiene la forma

$$R = \{Y \in \mathfrak{R}^n : F(Y) > C^*\}$$

con C^* una constante (ver por ejemplo Montgomery, 1992).

Otros procedimientos bayesianos

Si se quiere probar $H_0 : M\underline{\beta} = \underline{b}$ considerando una función de utilidad de la forma 2.14 y además se cuenta con $\pi(\underline{\beta})$, es posible probar la hipótesis $M\underline{\beta} = \underline{b}$ mediante el uso de regiones HPD, como se vio en la sección 2.3.3.

Así, si S es una región HPD de contenido $(1 - \gamma)$ para $\underline{\nu} = M\underline{\beta}$ se rechazará la hipótesis nula cuando \underline{b} no pertenezca a S la cual, por propiedades de la distribución t , esta dada por

$$S_{1-\gamma}(\underline{\nu}) = \{\underline{\nu} : H(\underline{\nu}) \leq F_{k,n-p,1-\gamma}\}$$

donde $H(\underline{\nu}) = k^{-1}(G(\underline{\nu}) - M\hat{\beta})'(M(A^*)^{-1}M')^{-1}(G(\underline{\nu}) - M\hat{\beta})$ y la hipótesis nula se puede rechazar si \underline{b} no pertenece a $S_{1-\gamma}(g)$ o equivalentemente cuando $H(M\underline{\beta} = \underline{b}) > F_{k,n-p,1-\gamma}$.

Capítulo 3

Predicción y Diagnóstico del Modelo de Regresión de Rango Completo

3.1 Introducción

Dentro del análisis del MRLM un tópico de gran importancia, es el poder predecir k valores futuros de la variable dependiente dados ciertos valores de las variables independientes. Desde el punto de vista bayesiano este problema se puede resolver obteniendo la correspondiente distribución final de los k valores que se quieren predecir, es decir, obtener la distribución predictiva final. En la sección 3.2.1 y 3.3.1 se determina esta distribución tomando en cuenta una distribución inicial Normal-Gamma, para los parámetros desconocidos del MRLM, y una distribución inicial no informativa, respectivamente. Así, también se trata el problema de estimación de la correspondiente predicción usando una función de utilidad cuadrática. Otro tópico de interés sobre el MRLM, es poder evaluar lo adecuado del modelo, dentro de lo que se encuentra la validación de los supuestos del modelo y la detección de observaciones atípicas. En este capítulo se realiza este estudio desde un punto de vista bayesiano usando para los parámetros desconocidos del MRLM una distribución inicial Normal-Gamma, sección 3.2.2, y una distribución inicial no informativa, sección 3.3.2.

3.2 Análisis Bayesiano Conjugado

3.2.1 Análisis Predictivo

Como se mencionó en el capítulo uno el problema de estimación de una observación futura después de haber observado el resultado de un experimento, se resuelve determinando su distribución predictiva final, por ello si se desea predecir k valores futuros del MRLM es necesario determinar la distribución predictiva final de estos k valores futuros. A continuación se presenta la forma de obtener la distribución predictiva final, considerando para $\underline{\beta}$ y τ una distribución inicial Normal-Gamma. Si $Y = (y_1, \dots, y_n)^t$ es un vector de observaciones que conforman una muestra del MRLM y $W = (W_1, \dots, W_k)^t$ es el vector de las k observaciones que se quieren predecir, con

$$\underline{W} = Z\underline{\beta} + \underline{\varepsilon}$$

donde Z es una matriz conocida de $k \times p$, $\underline{\varepsilon}$ un vector de errores que se distribuyen $N(0, \tau^{-1}I)$ y $\underline{\beta}$ el vector de parámetros del modelo

$$Y = X\underline{\beta} + \underline{\varepsilon},$$

suponiendo que $\underline{\varepsilon}$ y $\underline{\varepsilon}$ son independientes, la distribución predictiva final de W está determinada por

$$\begin{aligned} \pi(W|Y) &= \int \pi(W, \underline{\beta}, \tau|Y) d\underline{\beta} d\tau \\ &= \int f(W|\underline{\beta}, \tau) \pi(\underline{\beta}, \tau|Y) d\underline{\beta} d\tau \end{aligned}$$

donde

$$f(W|\underline{\beta}, \tau) \propto \tau^{\frac{k}{2}} \exp\left\{-\frac{\tau}{2}(W - Z\underline{\beta})'(W - Z\underline{\beta})\right\}.$$

Ahora, si se considera que la distribución final de $\underline{\beta}$ y τ es una Normal-Gamma, ecuación 2.7, se tiene que

$$\begin{aligned} \pi(W, \underline{\beta}, \tau|Y) &\propto \tau^{(n+k+p+2\alpha)/2-1} \exp\left\{-\frac{\tau}{2}[2\lambda + (\underline{\beta} - \underline{\mu})'A(\underline{\beta} - \underline{\mu}) \right. \\ &\quad \left. + (Y - X\underline{\beta})'(Y - X\underline{\beta}) + (W - Z\underline{\beta})'(W - Z\underline{\beta})\right\}. \end{aligned}$$

Desarrollando la expresión anterior y completando las formas cuadráticas correspondientes, se obtiene

$$\begin{aligned} \pi(W, \underline{\beta}, \tau|Y) &\propto \tau^{(k+n+2\alpha)/2-1} \exp\left\{-\frac{\tau}{2}[2\lambda + Y'Y + W'W - \varphi'v\varphi]\right\} \\ &\times \tau^{\frac{k}{2}} \exp\left\{-\frac{\tau}{2}[\underline{\beta} - v\varphi]'v[\underline{\beta} - v\varphi]\right\} \end{aligned} \quad (3.1)$$

en donde

$$\begin{aligned} \varphi &= (Y'X + \underline{\mu}'A + W'Z)' \\ v &= (A + X'X + Z'Z)^{-1}. \end{aligned}$$

Finalmente integrando 3.1 con respecto a $\underline{\beta}$ y τ

$$\pi(W|Y) \propto \{[W - \Lambda^{-1}\Upsilon]' \Lambda [W - \Lambda^{-1}\Upsilon] + \Psi - \Upsilon' \Lambda^{-1} \Upsilon\}^{(n+k+2\alpha)/2}$$

con

$$\begin{aligned} \Lambda &= (I - ZvZ') \\ \Upsilon &= Zv(A\underline{\mu} + X'Y) \\ \Psi &= Y'Y + \underline{\mu}'A\underline{\mu} - (X'Y + A\underline{\mu})'v(X'Y + A\underline{\mu}) + 2\lambda. \end{aligned}$$

Por lo que la distribución predictiva final de W es una distribución t k -dimensional con $n + 2\alpha$ grados de libertad, vector de localización $\Lambda^{-1}\Upsilon$, y matriz de precisión $(n + 2\alpha)\Lambda(\Psi - \Upsilon'\Lambda^{-1}\Upsilon)^{-1}$. Una vez que se ha obtenido la distribución predictiva final, por lo visto en la sección 2.4.1, si se tienen n observaciones y se considera una función de utilidad cuadrática la mejor predicción puntual del vector W de las k observaciones que se quieren predecir, será

$$\hat{W} = E_{\pi(W|Y)}(W).$$

Sin embargo, como se mencionó en el capítulo dos (sección 2.3.2) una mejor estimación, en este caso para predicción, es a través de intervalos HPD. Estos se pueden construir fácilmente tomando en cuenta los desarrollos del capítulo dos y el hecho que la distribución predictiva final es una t k -dimensional.

3.2.2 Diagnóstico

Dentro del análisis del MRLM otro aspecto que es de interés analizar son las discrepancias entre el modelo y los datos para los cuales se ajusta. Por lo que en esta sección se presentan algunas técnicas de diagnóstico desde un punto de vista bayesiano para el análisis de supuestos del modelo y detección de observaciones atípicas (aberrantes e influyentes), tomando en cuenta una inicial Normal-Gamma para los parámetros $\underline{\beta}$ y τ .

Diagnóstico para Supuestos del Modelo

Como se mencionó en el capítulo dos los supuestos del MRLM son los siguientes:

- a) $E(\epsilon_i) = 0$
- b) $\text{Var}(\epsilon_i) = \sigma^2$
- c) $\text{Cov}(\epsilon_i, \epsilon_j) = 0 \quad \forall i \neq j$
- d) Los errores ϵ_i tienen una distribución Normal.

Las violaciones a estas suposiciones y a la bondad del modelo pueden ser detectadas realizando un análisis de residuos.

El residuo de la observación i -ésima se define mediante

$$r_i = y_i - \hat{y}_i$$

en donde \hat{y}_i es una estimación de la observación y_i correspondiente, calculada por

$$\hat{y}_i = X_i \underline{\hat{\beta}} \quad \text{con} \quad i = 1, \dots, n$$

donde el estimador de $\underline{\beta}$ se obtiene aplicando el método de máxima utilidad esperada (ver subsección 2.3.1). Un análisis de residuos puede realizarse a través de una exploración gráfica de éstos y con base a ciertos patrones determinar si los supuestos del modelo se cumplen o si existe evidencia de que se están violando. A continuación se presentan varias gráficas de residuos que son útiles para detectar desviaciones a los supuestos.

* **Gráficas de r_i sobre papel normal.** Estas gráficas, como su nombre lo indica, muestran si se cumple el supuesto de normalidad en los errores. El papel normal está diseñado de tal manera que

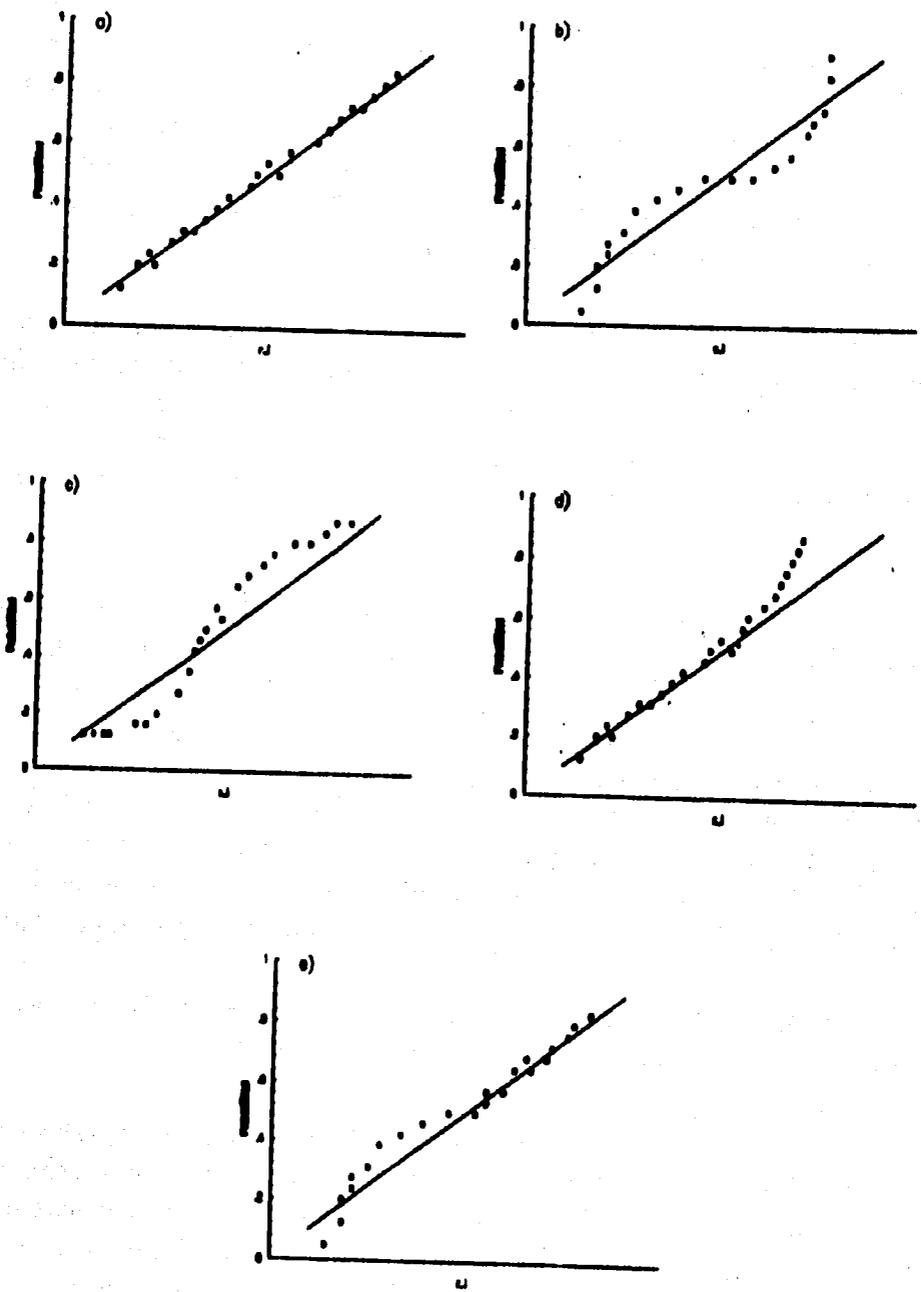


Figura 3.1: Gráficas de r ; sobre papel normal

desviaciones substanciales de una línea recta, determinada visualmente, indican que la distribución de los errores no es normal. En la figura 3.1.a se muestra una gráfica de probabilidad normal "ideal". Se puede notar que los puntos se ajustan aproximadamente a lo largo de una línea recta. Las figuras 3.1.b, 3.1.c, 3.1.d y 3.1.e muestran problemas comunes sobre desviaciones de la normalidad.

* **Gráficas de r_i contra \hat{y}_i .** Estas gráficas son útiles para detectar algunos tipos comunes de desviaciones a los supuestos del modelo. Si esta gráfica tiene un comportamiento parecido a la presentada en la figura 3.2.a es decir, que los residuos pueden estar contenidos en una banda horizontal, esto indicaría que no hay desviaciones obvios a los supuestos del modelo.

Gráficas con patrones como en las figuras 3.2.b, 3.2.c y 3.2.d indican deficiencias del modelo. Los patrones que se muestran en las figuras 3.3.b y 3.3.c indican que la varianza de los errores no es constante.

La forma de embudo de la figura 3.2.b implica que la varianza es una función creciente de y . La gráfica 3.2.d indica no linealidad, esto podría significar que otras variables independientes son necesarias en el modelo. Por ejemplo, un término cuadrático. Las gráficas de residuos contra \hat{y}_i también pueden revelar aquellos residuos de gran magnitud.

* **Gráficas de residuos contra X_i .** Estas gráficas tienen la misma interpretación que las gráficas de residuos contra \hat{y}_i .

* **Gráficas de r_i contra el tiempo.** Para obtener estas gráficas es necesario conocer el orden en el que fueron tomadas las observaciones. Estas gráficas ayudan a identificar cuando la varianza no es constante, es decir, pueden indicar si la varianza cambia con el tiempo o bien si términos lineales o cuadráticos en el tiempo deberían ser adicionados al modelo.

Un análisis más detallado de estas gráficas se puede ver en Draper y Smith, 1966.

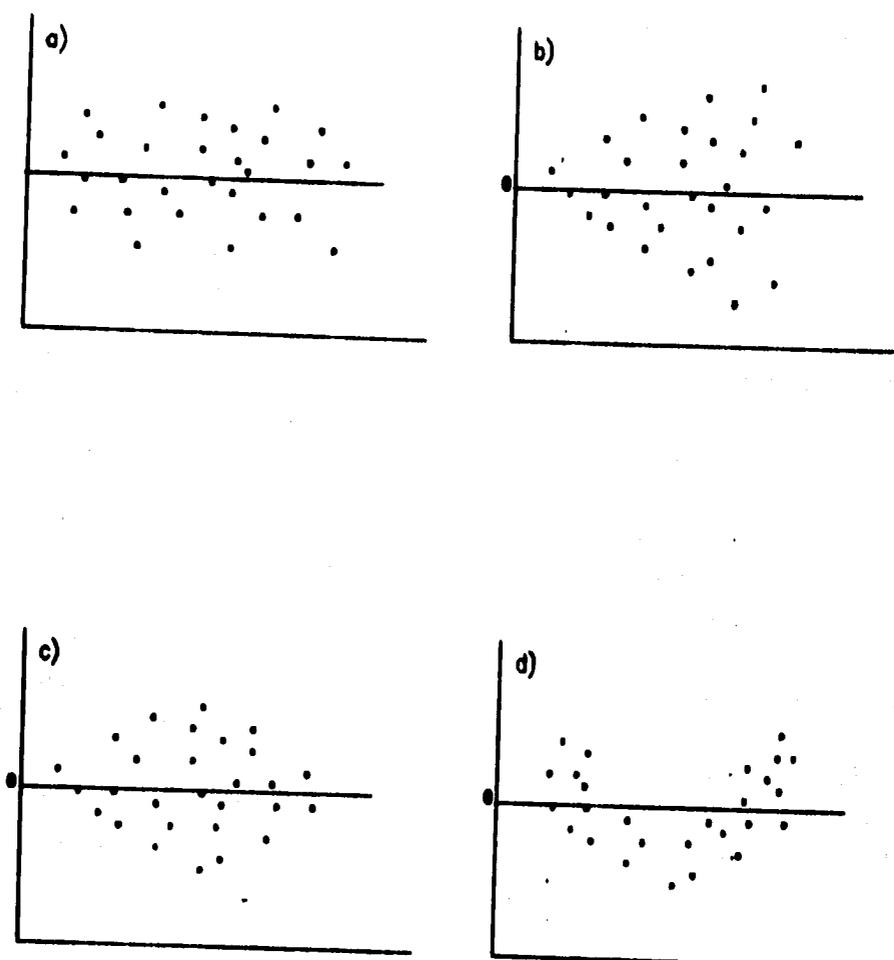


Figura 3.2: Gráficas de Residuos r_i contra \hat{y}_i

Diagnóstico para observaciones Atípicas

En algunos conjuntos de datos a veces se observa que el modelo ajustado depende de alguna o algunas de las observaciones en tal forma que si dichas observaciones fueran eliminadas, las distribuciones finales de los parámetros, predicciones e inferencias serían muy diferentes. A estas observaciones se les considera influyentes por el efecto que tiene su ausencia o presencia en el modelo ajustado. En otras ocasiones, lo que ocurre es que ciertas observaciones parecen que no siguen los patrones de comportamiento establecidos por el resto de las observaciones, de tal manera que la predicción que el modelo hace de ellas es muy diferente al valor actual de la variable respuesta. A estas observaciones se les conoce como observaciones aberrantes. La detección de observaciones atípicas (aberrantes e influyentes) es de interés en el análisis del modelo MRLM ya que éstas pueden indicar condiciones bajo las cuales el modelo no se ajuste adecuadamente.

A continuación se presenta un análisis bayesiano para detectar observaciones aberrantes e influyentes. La atención del análisis se centrará en dos problemas básicamente, el primero consiste en decidir cuando una o más observaciones y_i , del conjunto de datos Y , es aberrante; el segundo problema comprende, determinar cuando una observación o subconjunto de observaciones influye sobre la estimación de los parámetros.

Como antes, sea $Y = \{y_1, \dots, y_n\}$ una muestra aleatoria del MRLM, una idea intuitiva para detectar si una observación y_i es aberrante, es considerar si ésta puede ser predicha por el resto de las observaciones, es decir, si su densidad final

$$\pi(y_i|Y_{(i)}) = \int_{\theta \in \Theta} f(y|\theta)\pi(\theta|Y_{(i)})d\theta$$

es muy pequeña se puede considerar y_i como aberrante. Donde $Y_{(i)}$ es la muestra de observaciones iniciales sin la observación i -ésima. Sin embargo, aún cuando se tiene una idea intuitiva para la detección de observaciones aberrantes es necesario la presentación de conceptos más formales. En Girón, Martínez y Morcillo (1992) se consideran las siguientes definiciones.

Definición 3.2.1 Una observación y_i es aberrante, si y_i no pertenece a una región HPD de $\pi(y|Y_{(i)})$ de contenido de probabilidad $1 - \gamma$.

La extensión de esta definición para cualquier subconjunto de observaciones es inmediata y se presenta a continuación.

Definición 3.2.2 *Un subconjunto de observaciones indexadas por $I = \{i_1, \dots, i_m\} \subset \{1, \dots, n\}$ es aberrante si el vector Y_I no pertenece a una región HPD de contenido de probabilidad $1 - \gamma$ de la densidad predictiva final $\pi(y_1, \dots, y_m | Y_{(I)})$.*

Una vez que se ha determinado si una observación o un conjunto de observaciones es aberrante, se debe realizar una investigación cuidadosa para ver si se puede encontrar la razón de este comportamiento inusual, o si la observación es producto de un mal registro de los datos. Finalmente, si se considera θ^* un estimador bayesiano, la siguiente definición hace precisa la influencia de una observación sobre un estimador particular y además la generalización para un subconjunto de observaciones es inmediata.

Definición 3.2.3 *La observación y_i es influyente en la estimación de θ^* si θ^* no pertenece a una región HPD de contenido de probabilidad $1 - \gamma$ de la distribución final $\pi(\theta | Y_{(i)})$.*

A continuación se establece el análisis correspondiente para la detección de observaciones atípicas en el MRLM, considerando una distribución inicial Normal-Gamma para $\underline{\beta}$ y τ . Como antes, sea $Y = \{y_1, \dots, y_n\}$ una muestra aleatoria de n observaciones, I cualquier subconjunto de Y de m observaciones. Se identificará con el subíndice I cualquier submatriz o subvector involucrado en el modelo que tome en cuenta las m observaciones indexadas por I , y por el subíndice (I) el subvector o submatriz que involucre el resto de las $n-m$ observaciones. En vista de que la distribución final de $\underline{\beta}$ y τ es una Normal-Gamma, la distribución final de $\underline{\beta}$ y τ dado $Y_{(I)}$, $\pi(\underline{\beta}, \tau | Y_{(I)})$, es Normal-Gamma con parámetros $(\alpha_{1(I)}^*, \lambda_{1(I)}^*, \mu_{1(I)}^*, A_{1(I)}^*)$. Donde

$$\begin{aligned} \alpha_{1(I)}^* &= \frac{n + 2\alpha - m}{2} \\ \lambda_{1(I)}^* &= \lambda + \frac{Y_{(I)}^t Y_{(I)} - (A_{\underline{\mu}} + X_{(I)}^t Y_{(I)})^t A_{1(I)}^* (A_{\underline{\mu}} + X_{(I)}^t Y_{(I)})}{2} \\ \mu_{1(I)}^* &= (X_{(I)}^t X_{(I)} + A)^{-1} (X_{(I)}^t Y_{(I)} + A_{\underline{\mu}}) \\ A_{1(I)}^* &= (X_{(I)}^t X_{(I)} + A), \end{aligned}$$

la distribución predictiva, $\pi(W|Y_{(I)})$, del vector W que consta de las m observaciones dado $Y_{(I)}$ es una t k -dimensional con $n - m + 2\alpha$ grados de libertad (donde α es el mismo de la distribución inicial de τ), vector de localización $(\Lambda_{(I)}^{-1}\Upsilon_{(I)})$ y matriz de precisión $B = (n - m + 2\alpha)\Lambda_{(I)}(\Psi_{(I)} - \Upsilon_{(I)}^t\Lambda_{(I)}^{-1}\Upsilon_{(I)})^{-1}$. Por las propiedades de la t multivariada y por lo visto en la sección 2.3.2 se sigue que la región HPD de contenido de probabilidad $(1 - \gamma)$ para W esta determinada por

$$\{W \in \mathfrak{R}^m : m^{-1}(W - \Lambda_{(I)}^{-1}\Upsilon_{(I)})^t(B)(W - \Lambda_{(I)}^{-1}\Upsilon_{(I)}) < F_{(m, n-m+2\alpha, 1-\gamma)}\}$$

donde

$$\begin{aligned}\Lambda_{(I)} &= (I_{(I)} - X_I v_{(I)} X_I^t) \\ \Upsilon_{(I)} &= X_I v_{(I)} (A_{\underline{\mu}} + X_{(I)}^t Y_{(I)}) \\ \Psi_{(I)} &= Y_{(I)}^t Y_{(I)} + \underline{\mu}^t A_{\underline{\mu}} - (X_{(I)}^t Y_{(I)} + A_{\underline{\mu}})^t v_{(I)} (X_{(I)}^t Y_{(I)} + A_{\underline{\mu}}) + 2\lambda \\ v_{(I)} &= (A + X_{(I)}^t X_{(I)} + X_I^t X_I)^{-1}\end{aligned}$$

y como antes, $F_{(m, n-m+2\alpha, 1-\gamma)}$ es el cuantil de orden $1 - \gamma$ de una distribución F con m y $n - m + 2\alpha$ grados de libertad. Así, de acuerdo a la definición 3.2.2 el subconjunto de observaciones indexado por I , Y_I , es aberrante si

$$m^{-1}(Y_I - \Lambda_{(I)}^{-1}\Upsilon_{(I)})^t(B)(Y_I - \Lambda_{(I)}^{-1}\Upsilon_{(I)}) > F_{(m, n-m+2\alpha, 1-\gamma)}.$$

Una vez que se ha determinado si un subconjunto de observaciones es aberrante, se analizará cuándo un conjunto de observaciones puede influir sobre la estimación de los parámetros del MRLM. Primero se analizará la influencia sobre la estimación del parámetro de regresión $\underline{\beta}$, posteriormente se analizará la influencia sobre la estimación de τ^{-1} .

Como la distribución de $\pi(\underline{\beta}, \tau|Y_{(I)})$ es una Normal-Gamma con parámetros $\alpha_{1(I)}^*$, $\lambda_{1(I)}^*$, $\mu_{1(I)}^*$ y $A_{1(I)}^*$, la distribución final de $\underline{\beta}$ dado $Y_{(I)}$ es una t p -dimensional con $n - m + 2\alpha$ grados de libertad, vector de localización $\underline{\mu}_{(I)}^*$ y matriz de precisión $A_{(I)}^* = A_{1(I)}^* \alpha_{1(I)}^* (\lambda_{1(I)}^*)^{-1}$. Por las propiedades de una t p -dimensional, las regiones HPD para $\underline{\beta}$ de contenido de probabilidad $1 - \gamma$ son de la forma:

$$\{\underline{\beta} \in \mathfrak{R}^p, p^{-1}(\underline{\beta} - \underline{\mu}_{(I)}^*)^t A_{(I)}^* (\underline{\beta} - \underline{\mu}_{(I)}^*) \leq F_{(p, n+2\alpha-m, 1-\gamma)}\}$$

y de acuerdo a la definición 3.2.3, el subconjunto de observaciones indexadas por I influirá sobre $\underline{\beta}$, con respecto al estimador $\hat{\underline{\beta}}$, si

$$BD(I) = p^{-1}(\hat{\underline{\beta}} - \underline{\mu}_{(I)}^*)' A_{(I)}^* (\hat{\underline{\beta}} - \underline{\mu}_{(I)}^*) > F_{(p, n-m+2\alpha, 1-\gamma)}.$$

Las medidas de influencia de subconjuntos de observaciones sobre los elementos β_i , con $i=1, \dots, p$, del parámetro de regresión $\underline{\beta}$ o combinaciones lineales de éstos, se pueden derivar encontrando su correspondiente distribución final y aplicando el procedimiento presentado arriba. A continuación se presenta un análisis similar para τ^{-1} .

De acuerdo a lo visto en la sección 2.3 se tiene que la distribución final de τ^{-1} dado $Y_{(I)}$ es una Gamma inversa con parámetros $\alpha_{1(I)}^*$ y $\lambda_{1(I)}^*$, por lo que

$$\hat{\tau}_{(I)}^{-1} (\tau^{-1})^{-1} 2(\alpha_{1(I)}^* - 1) \sim \chi_{(2\alpha_{1(I)}^*)}^2$$

donde

$$\hat{\tau}_{(I)}^{-1} = \lambda_{1(I)}^* (\alpha_{1(I)}^* - 1)^{-1}$$

así, un intervalo HPD para τ^{-1} (Girón, Martínez y Morcillo, 1992) es

$$\left(\frac{2\hat{\tau}_{(I)}^{-1}(\alpha_{1(I)}^* - 1)}{\chi_{(2\alpha_{1(I)}^*, 1-\frac{\gamma}{2})}^2}, \frac{2\hat{\tau}_{(I)}^{-1}(\alpha_{1(I)}^* - 1)}{\chi_{(2\alpha_{1(I)}^*, \frac{\gamma}{2})}^2} \right).$$

Por lo tanto el subconjunto de observaciones indexadas por I influirá en la estimación de τ^{-1} , $\hat{\tau}^{-1}$, si

$$\frac{2\hat{\tau}_{(I)}^{-1}(\alpha_{1(I)}^* - 1)}{\hat{\tau}^{-1}} \geq \chi_{(2\alpha_{1(I)}^*, 1-\frac{\gamma}{2})}^2$$

ó

$$\frac{2\hat{\tau}_{(I)}^{-1}(\alpha_{1(I)}^* - 1)}{\hat{\tau}^{-1}} \leq \chi_{(2\alpha_{1(I)}^*, \frac{\gamma}{2})}^2.$$

donde $\hat{\tau}^{-1} = \lambda_1^* (\alpha_1^* - 1)^{-1}$. Existen diferentes formas o técnicas para considerar posibles candidatos de observaciones atípicas. A continuación se presenta otra forma bayesiana para detectar observaciones de este tipo, la cual considera el hecho de que los errores realizados son

parámetros con valores desconocidos. Considerando que \underline{Y} es un vector aleatorio de $n \times 1$, generado por el modelo

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon},$$

una muestra dada de observaciones $Y^t = y_1, \dots, y_n$ del modelo anterior (MRLM) es una realización de éste y satisface

$$Y = X\underline{\beta} + \epsilon \quad (3.2)$$

donde ϵ es un vector de $n \times 1$ cuyos elementos son realizaciones del proceso generado por $\underline{\epsilon}$ y por lo tanto no son aleatorios para un X y $\underline{\beta}$ dados. Como $\epsilon = Y - X\underline{\beta}$ si los elementos de $\underline{\beta}$, β_i , fueran conocidos se podría establecer un análisis de los valores realizados ϵ_i para detectar posibles candidatos de observaciones atípicas. Ya que $\underline{\beta}$ es desconocido el valor de ϵ también lo es. Así, si se considera una distribución final para $\underline{\beta}$ t p -dimensional con $n + 2\alpha$ grados de libertad, vector de localización $\underline{\mu}^*$ y matriz de precisión A^* se puede derivar la distribución final para $\epsilon = Y - X\underline{\beta}$ fijando Y y X .

Por lo que, considerando que las y_i y las x_i son fijas, se tiene que la f.d.p. para cada elemento de $\underline{\epsilon}$, ϵ_i

$$\epsilon_i = y_i - x_i \underline{\beta} \quad i = 1, 2, \dots, n$$

con x_i el i -ésimo renglón de la matriz X , tiene la forma de una f.d.p. t de Student univariada con media $\hat{\epsilon}_i = y_i - x_i \underline{\mu}^*$ y precisión $(x_i(A^*)^{-1}x_i^t)^{-1}$. Por lo que los residuos

$$u_i = \frac{(\epsilon_i - \hat{\epsilon}_i)}{(x_i(A^*)^{-1}x_i^t)^{1/2}}$$

del MRLM tienen una distribución t -univariada estándar con $n + 2\alpha$ grados de libertad. Es deseable que al menos la $\Sigma \hat{\epsilon}_i \doteq 0$ y aún más que cada $\hat{\epsilon}_i \doteq 0$ dado que ésto representaría un buen ajuste. Una vez determinada la distribución de u_i se pueden obtener intervalos HPD para cada ϵ_i y analizar si el intervalo correspondiente contiene a $\hat{\epsilon}_i$, de esta forma si todos los intervalos HPD de cierto contenido $1 - \gamma$ contienen a $\hat{\epsilon}_i$, se tendría un buen ajuste. Por otro lado para analizar $\Sigma \hat{\epsilon}_i$ es necesario encontrar la distribución de Σu_i y realizar un análisis

similar al de u_i . Dado que $\epsilon = Y - X\beta$ tiene vector de localización $\hat{\epsilon} = Y - X\hat{\mu}^*$ y matriz de precisión $(X(A^*)^{-1}X')$. Si l' es un vector de $k \times 1$ de rango k , $Z = l'\epsilon$ se distribuye t de Student con vector de localización $\hat{Z} = l'\hat{\epsilon}$ y matriz de precisión

$$C^2 = [l'X(A^*)^{-1}X'l],$$

así

$$\frac{Z - \hat{Z}}{C} \sim t_{n-p} \quad (\text{Zellner, 1975})$$

de esta forma, como ya se mencionó, bastará con obtener una región HPD de contenido $1 - \gamma$ para Z y verificar que \hat{Z} esté contenido en la región, lo cual implicaría tener un buen ajuste.

3.3 Análisis Bayesiano no Informativo

3.3.1 Análisis Predictivo

En esta sección se presenta la distribución predictiva final para el MRLM de rango completo considerando la distribución inicial

$$\pi(\underline{\beta}, \tau) = \frac{1}{\tau} \quad \tau > 0. \quad (3.3)$$

Se sabe que

$$\begin{aligned} \pi(W|Y) &= \int \pi(W, \underline{\beta}, \tau|Y) d\underline{\beta} d\tau \\ &= \int f(W|\underline{\beta}, \tau) \pi(\underline{\beta}, \tau|Y) d\underline{\beta} d\tau. \end{aligned}$$

Asumiendo la distribución inicial (3.3) y de acuerdo a los resultados de las secciones 3.2.1 y 3.2.4, se tiene

$$\begin{aligned} \pi(W, \underline{\beta}, \tau|Y) &\propto \tau^{(k+n-p)/2-1} \exp\left\{-\frac{\tau}{2}[(W - Z\underline{\beta})'(W - Z\underline{\beta})]\right\} \\ &\tau^{\frac{n-p}{2}} \exp\left\{-\frac{(n-p)s^2\tau}{2} - \frac{s^2\tau(\underline{\beta} - \hat{\underline{\beta}})'(X'X)(\underline{\beta} - \hat{\underline{\beta}})}{2s^2}\right\} \end{aligned}$$

con

$$\begin{aligned}\hat{\beta} &= (X^t X)^{-1} X^t Y, \\ s^2 &= \frac{(Y - \hat{Y})^t (Y - \hat{Y})}{(n - p)}\end{aligned}$$

por lo que

$$\pi(W|Y) \propto \{[W - Z\hat{\beta}]^t \tilde{A} [W - Z\hat{\beta}] + 1\}^{(n+k-p)/2}$$

donde

$$\tilde{A} = (n - p)(Y^t Y - Y^t X (X^t X)^{-1} X^t Y)^{-1} (I + Z^t (X^t X)^{-1} Z^t)^{-1}$$

Así, la distribución predictiva final de W es una distribución t k -dimensional con $n - p$ grados de libertad, vector de localización $Z\hat{\beta}$ y matriz de precisión \tilde{A} .

Así, si se tienen n observaciones la mejor predicción \hat{W} , considerando una función de utilidad cuadrática, es

$$\hat{W} = E_{\pi(W|Y)}(W) = Z\hat{\beta}.$$

Se puede observar que, en este caso, la estimación puntual del vector de k observaciones que se quieren predecir coincide con la que se obtiene en el enfoque clásico.

3.3.2 Diagnóstico

En esta sección se presenta el análisis de diagnóstico para el MRLM de rango completo considerando la distribución inicial no informativa 3.3.

La parte correspondiente al diagnóstico para supuestos del modelo es similar a la presentada en la sección 3.2.2, sólo se debe considerar que en este caso $\hat{\beta} = (X^t X)^{-1} X^t Y$. Sin embargo, en el caso de detección de observaciones atípicas los resultados no son tan inmediatos por lo que a continuación se presenta su correspondiente análisis.

Detección de observaciones atípicas

Considerando la notación y las definiciones de la sección 3.2.2 se tiene que la distribución predictiva final para un subconjunto de m observaciones indexado por I , dado $Y_{(I)}$, es una t k -dimensional de $n - p - m$ grados de libertad, media $X_I \hat{\beta}_{(I)}$ y matriz de precisión

$$\tilde{A} = (n - p - m)(Y_{(I)}^t Y_{(I)} - Y_{(I)}^t X_{(I)}(X_{(I)}^t X_{(I)})^{-1} X_{(I)}^t Y_{(I)})^{-1} \times \\ (I + X_I(X_{(I)}^t X_{(I)})X_I^t)^{-1}$$

De acuerdo a la definición 3.2.2 el subconjunto de observaciones indexado por I , Y_I , es aberrante si

$$PD(I) = \frac{(Y_I - X_I \hat{\beta}_{(I)})^t \tilde{A}_{(I)} (Y_I - X_I \hat{\beta}_{(I)})}{m} \geq F_{(m, n-p-m, 1-\gamma)} \quad (3.4)$$

la expresión anterior se puede reescribir como:

$$PD(I) = \frac{r_I^t \tilde{A}_{(I)} r_I}{m} \geq F_{(m, n-p-m, 1-\gamma)}$$

donde $r_I = Y_I - X_I \hat{\beta}_{(I)}$ es conocido como el vector de residuos de jackknife. Para el caso de una sola observación (y_i, x_i) , la distribución predictiva del i -ésimo residuo de jackknife es

$$\pi(r_i | Y_{(i)}) \sim t(0, \tilde{A}_{(i)}, n - p - 1)$$

por lo tanto la i -ésima observación se considera como aberrante si

$$t_i^* = |r_i \sqrt{\tilde{A}_{(i)}}| \geq t_{(n-p-1, 1-\gamma)}$$

donde $t_{(n-p-1, 1-\gamma)}$ es el cuantil de orden $1 - \gamma$ de una distribución t estándar con $n - p - 1$ grados de libertad. La cantidad t_i^* es comúnmente conocida como el i -ésimo residuo de jackknife estudentizado. Este resultado es equivalente al obtenido en la ecuación 3.4 cuando $I = i$, tomando en cuenta que el cuantil de orden γ de una distribución F con 1 y $n - p - 1$ grados de libertad, es el cuadrado del cuantil correspondiente de una distribución t con $n - p - 1$ grados de libertad.

A continuación se analizará la influencia de un subconjunto de observaciones sobre la estimación del parámetro $\underline{\beta}$, posteriormente se analizará la influencia sobre la estimación de τ^{-1} . Como la distribución final de $\underline{\beta}$ dado $Y_{(I)}$ es una t p -dimensional con $n-p-m$ grados de libertad, vector de localización $\hat{\underline{\beta}}_{(I)}$ y matriz de precisión $A_{(I)}^*$, de acuerdo a la definición 3.2.3 y por propiedades de la distribución t , el subconjunto de observaciones indexadas por I influirá sobre el estimador $\hat{\underline{\beta}}^*$ si

$$BD(I) = p^{-1}(\hat{\underline{\beta}}^* - \hat{\underline{\beta}}_{(I)})^t A_{(I)}^* (\hat{\underline{\beta}}^* - \hat{\underline{\beta}}_{(I)}) \geq F_{(p, n-p-m, 1-\gamma)}$$

A continuación se presenta un análisis similar para τ^{-1} . Por lo visto en la sección 3.2.2. se tiene que la distribución final de τ^{-1} dado $Y_{(I)}$ es una Gamma inversa $(\alpha_{1(I)}^*, \lambda_{1(I)}^*)$. Por lo que el subconjunto de observaciones indexadas por I , Y_I influirá en la estimación de τ^{-1} , $\hat{\tau}^{-1}$, si

$$\frac{2\hat{\tau}_{(I)}^{-1}(\alpha_{1(I)}^* - 1)}{\hat{\tau}^{-1}} \leq \chi_{(2\alpha_{1(I)}^*, \frac{\gamma}{2})}^2$$

ó

$$\frac{2\hat{\tau}_{(I)}^{-1}(\alpha_{1(I)}^* - 1)}{\hat{\tau}^{-1}} \geq \chi_{(2\alpha_{1(I)}^*, 1-\frac{\gamma}{2})}^2$$

donde

$$\begin{aligned} \alpha_{1(I)}^* &= \frac{n-p-m}{2} \\ \hat{\tau}^{-1} &= \lambda_1^*(\alpha_1^*)^{-1} \\ \lambda_{(I)}^* &= \alpha_{1(I)}^* s_{(I)}^2 \end{aligned}$$

Una técnica bayesiana para considerar la bondad del modelo ajustado es la planteada en la sección 3.2.2, en donde se considera que \underline{Y} es un vector aleatorio de $n \times 1$, generado por una realización del modelo

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon},$$

Por lo que si se considera que las y_i y las x_i son fijas se tiene que los residuos

$$r_i = \frac{(\epsilon_i - m_i)}{(x_i^t (A^*)^{-1} x_i)^{1/2}}$$

tienen una distribución t-univariada estándar con $n - p$ grados de libertad. Donde $m_i = y_i - x_i \hat{\beta}$ y $A^* = \frac{X'X}{p}$. Como antes, es deseable que al menos la $\Sigma \hat{\epsilon}_i \doteq 0$ y aún más que cada $\hat{\epsilon}_i \doteq 0$ dado que esto representa un buen ajuste. Estas pruebas pueden llevarse a cabo a través de la construcción de regiones HPD y hacer el respectivo análisis por lo que para ello es necesario determinar las distribuciones finales de r_i y de Σr_i que se pueden determinar de la misma manera como se hizo en la sección anterior, la única diferencia son los grados de libertad, los cuales en este caso son $n - p$ y el vector de localización $\underline{\mu}^* = \hat{\beta}$. Otra medida que en la práctica se emplea es $\epsilon' \epsilon = \Sigma \epsilon_i^2$, si $\epsilon' \epsilon \doteq 0$ se tendría un buen ajuste del modelo. Por lo que, si se considera la distribución inicial no informativa 3.3 para $\underline{\beta}$ y τ , se obtiene que

$$\begin{aligned} \epsilon' \epsilon &= (Y - X \hat{\beta})'(Y - X \hat{\beta}) + (\underline{\beta} - \hat{\beta})' X' X (\underline{\beta} - \hat{\beta}) \\ \frac{\epsilon' \epsilon}{ps^2} &= \frac{(Y - X \hat{\beta})'(Y - X \hat{\beta})}{ps^2} + \frac{(\underline{\beta} - \hat{\beta})' X' X (\underline{\beta} - \hat{\beta})}{ps^2} \\ \frac{\epsilon' \epsilon}{ps^2} &= \frac{n - p}{p} + \frac{(\underline{\beta} - \hat{\beta})' X' X (\underline{\beta} - \hat{\beta})}{ps^2} \end{aligned}$$

como $\underline{\beta}$ tiene una distribución final t p-dimensional con media $\hat{\beta}$ y matriz de precisión $\frac{X'X}{p}$ se tiene que

$$\frac{(\underline{\beta} - \hat{\beta})' X' X (\underline{\beta} - \hat{\beta})}{ps^2} \sim F_{(p, n-p, 1-\gamma)}$$

entonces

$$\frac{\epsilon' \epsilon}{ps^2} - \frac{n - p}{p} = \frac{\epsilon' \epsilon - \hat{\epsilon}' \hat{\epsilon}}{ps^2} \sim F_{(p, n-p, 1-\gamma)}$$

tiene una distribución F con p y $n - p$ grados de libertad, donde $\hat{\epsilon}' = Y - X \hat{\beta}$ y $\hat{\beta}$ es el estimador de mínimos cuadrados. Así, se pueden obtener intervalos HPD para $\epsilon' \epsilon$ y analizar si este intervalo contiene a $\hat{\epsilon}' \hat{\epsilon}$ en cuyo caso se tendría un buen ajuste.

Capítulo 4

Una Aplicación de Regresión Lineal Múltiple

4.1 Introducción

En los capítulos anteriores se presentó el análisis bayesiano del modelo de regresión lineal múltiple de rango completo, considerando para los parámetros de éste dos tipos de distribuciones iniciales, una distribución Normal-Gamma y una no informativa. Se mencionó que cuando no se cuenta con la información inicial o se está limitado para especificarla, se puede llevar a cabo un análisis bayesiano empleando una distribución inicial no informativa. En este capítulo se presenta una aplicación del MRLM de rango completo, con la finalidad de mostrar, en un contexto práctico, los desarrollos presentados en los capítulos anteriores. Debido a que no se dispone de información inicial sobre los parámetros involucrados, se realiza un análisis bayesiano considerando la distribución inicial no informativa presentada en la sección 2.4 del capítulo dos.

4.2 Planteamiento del problema

Supóngase que una embotelladora de refrescos desea estudiar el tiempo en que se realiza el abasto de productos (refrescos) y el mantenimiento general de un sistema automatizado de expendio de estas bebidas. La embotelladora está interesada en predecir la cantidad de tiempo que requiere un conductor en cubrir una ruta para dar este servicio.

Tabla 4.1.1

observación	Y	X ₁	X ₂
1	16.68	-7	560
2	11.5	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8	2	110
8	17.83	7	210
9	40.33	16	688
10	21	10	215
11	13.5	4	255
12	19.75	6	462
13	24	9	448
14	29	10	776
15	15.35	6	200
16	19	7	132
17	9.5	3	36
18	35.1	17	770
19	17.9	10	140
20	18.75	9	450
21	19.83	8	635
22	10.75	4	150

El responsable de este estudio sugiere que las dos variables más importantes que afectan al tiempo de entrega son, el número de productos

surtidos y la distancia recorrida por el conductor de la ruta, además se cree que la relación se puede modelar de la siguiente manera

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \quad (4.1)$$

donde

- Y = tiempo de entrega en minutos
- X_1 = número de productos que se surten
- X_2 = distancia en pies recorrida por el conductor de ruta,

β_0, β_1 y β_2 son parámetros desconocidos y ϵ un error que se supone se distribuye $N(0, \tau I)$.

Para el estudio se cuenta con 22 observaciones las cuales se muestran en la tabla 4.1.1. Los datos para esta aplicación fueron tomados básicamente de Montgomery, 1992.

A continuación se ajustará y analizará un modelo (4:1) considerando la siguiente distribución inicial no informativa

$$\pi(\underline{\beta}, \tau) \propto \frac{1}{\tau} \quad \tau \in \mathbb{R}, \underline{\beta} \in \mathbb{R}^3.$$

4.3 Análisis Bayesiano

Usando los resultados del capítulo dos, sección 2.3, se tiene que la distribución final conjunta de $\underline{\beta}$ y τ , $\pi(\underline{\beta}, \tau | Y)$, es una Normal-Gamma con parámetros

$$\alpha_1^* = \frac{n-p}{2} = 9.5$$

$$\lambda_1^* = \frac{Y'Y - Y'X(X'X)^{-1}X'Y}{2} = 55.045$$

$$\hat{\underline{\beta}} = (X'X)^{-1}X'Y = \begin{pmatrix} 4.417 \\ 1.5613 \\ 0.008519 \end{pmatrix}$$

$$A_1^* = (X'X) = \begin{pmatrix} 22 & 158 & 7357 \\ 158 & 1454 & 66014 \\ 7357 & 66014 & 3571963 \end{pmatrix}$$

Por lo que la distribución final de $\underline{\beta}$ es una distribución t de tres dimensiones con $n - p = 19$ grados de libertad, vector de localización $\hat{\underline{\beta}}$ y matriz de precisión

$$A^* = \frac{A_1^*}{s^2} = \frac{A_1^*}{\frac{Y'Y - Y'X(X'X)^{-1}X'Y}{n-p}}$$

$$= \frac{A_1^*}{5.7942} = \begin{pmatrix} 3.7969 & 27.268 & 1269.7 \\ 27.268 & 250.94 & 11393 \\ 1269.7 & 11393 & 616470 \end{pmatrix}$$

También se tiene que la distribución final de τ es una Gamma con parámetros

$$\alpha_1^* = 9.5$$

$$\lambda_1^* = 55.045.$$

Una vez determinadas las distribuciones finales de $\underline{\beta}$ y τ se está en condiciones de realizar las inferencias necesarias.

4.3.1 Estimación puntual de los parámetros del modelo

Estimación puntual sin función de utilidad

Considerando τ un parámetro de ruido, de acuerdo a la sección 2.4.1, se tiene que un estimador puntual de $\underline{\beta}$ es:

$$E_{\pi(\underline{\beta}|Y)}(\underline{\beta}) = \hat{\underline{\beta}} = \begin{pmatrix} 4.4176 \\ 1.5613 \\ 0.008519 \end{pmatrix}$$

que coincide con la moda y la mediana de $\pi(\underline{\beta}|Y)$ ya que esta distribución es simétrica.

Para τ^{-1} se tienen los siguientes estimadores

$$\hat{\tau}_1^{-1} = E_{\pi(\tau|Y)}(\tau^{-1}) = \lambda_1^*(\alpha_1^* - 1)^{-1}$$

$$= 6.4759$$

$$\hat{\tau}_2^{-1} = \text{Moda}[\pi(\tau^{-1}|Y)] = \lambda_1^*(\alpha_1^* + 1)^{-1}$$

$$= 5.2424.$$

$$\hat{\tau}_3^{-1} = \text{Mediana}[\pi(\tau^{-1}|Y)] = 6.003495$$

Es importante observar que en este caso la media y la moda de τ^{-1} son diferentes, esto era de esperarse, ya que el tamaño de muestra no es lo suficientemente grande como para que las dos estimaciones puedan considerarse prácticamente iguales.

Estimación puntual con funciones de utilidad

Considerando una función de utilidad cuadrática se tiene que los estimadores óptimos para $\underline{\beta}$ y τ_1^{-1} son $\hat{\underline{\beta}}$ y $\hat{\tau}^{-1}$ respectivamente. Y usando como función de utilidad la divergencia logarítmica de Kullback-Leibler, los estimadores óptimos son, para $\underline{\beta}$ nuevamente $\hat{\underline{\beta}}$ y para τ^{-1} , $\hat{\tau}_2^{-1}$.

Una vez obtenido el estimador puntual de $\underline{\beta}$ se tiene que el modelo estimado es

$$Y = 4.4176 + 1.5613X_1 + .008519X_2 \quad (4.2)$$

Una primera forma de determinar si cada variable X_i contribuye a explicar la respuesta y es a través de intervalos HPD para los parámetros del modelo, los cuales se construyen a continuación.

4.3.2 Intervalos de Probabilidad

De acuerdo a lo presentado en la sección 2.4.2 se tiene que

$$R_{.95}(\underline{\beta}) = \{\underline{\beta} : F(\underline{\beta}) \leq F_{3,19,.95}\}$$

con $F(\underline{\beta}) = \frac{(\underline{\beta} - \hat{\underline{\beta}})' A^*(\underline{\beta} - \hat{\underline{\beta}})}{3}$ determina una región HPD de contenido .95 de probabilidad para $\underline{\beta}$. Las regiones HPD de contenido .95 de probabilidad para parejas de elementos de $\underline{\beta}$, $(\beta_i, \beta_j) = \underline{\beta}_{ij}$, con $i \neq j$, $i, j = 0, 1, 2$ están determinadas por $R_{.95}(\underline{\beta}_{ij}) = \{\underline{\beta}_{ij} : F(\underline{\beta}_{ij}) \leq F_{2,19,.95}\}$. Estas regiones están delimitadas por elipses cuyas ecuaciones son

para β_0 y β_1 ,

$$\frac{(\beta_0 - 4.4176)^2}{.26454^2} + \frac{(\beta_1 - 1.561300)^2}{2.4043^2} = 1$$

para β_0 y β_2 ,

$$\frac{(\beta_0 - 4.4176)^2}{.0053452^2} + \frac{(\beta_2 - 0.008519)^2}{2.3672^2} = 1$$

y para β_1 y β_2

$$\frac{(\beta_1 - 1.5613)^2}{.005345^2} + \frac{(\beta_2 - 0.008519)^2}{.29804^2} = 1.$$

Por otro lado, se pueden obtener intervalos HPD, al 95 %, para cada β_i , considerando las distribuciones finales marginales correspondientes. Puesto que $\underline{\beta}^t = (\beta_0, \beta_1, \beta_2)$ tiene una distribución final t de tres dimensiones con 19 grados de libertad vector de localización $\hat{\underline{\beta}}$ y matriz de precisión A^* , la distribución final de cada β_i es una distribución t con 19 grados de libertad (gl), vector de localización

$$m_i = V_i^t(\hat{\underline{\beta}})$$

matriz de precisión

$$P_i = [V_i^t(A)^{-1}V_i]^{-1},$$

y varianza

$$var(\beta_i) = \frac{gl}{P_i(gl - 2)},$$

donde V_i es un vector en \mathfrak{R}^3 , con todas sus entradas iguales a cero, excepto en el lugar i -ésimo donde tiene un 1. Usando esto se tiene que un intervalo para β_i esta dado por (a, b) donde

$$a = (T_{(19, .025)})var(\beta_i) + m_i$$

$$b = (T_{(19, .975)})var(\beta_i) + m_i$$

con $T_{(19, \gamma)}$ el percentil de orden γ de una distribución t con 19 gl . Por lo anterior se tiene que β_0 tiene una distribución t con 19 grados de libertad, media

$$\hat{\beta}_0 = (1, 0, 0)\hat{\underline{\beta}} = 4.417$$

y precisión

$$P_1 = ((1, 0, 0)(A^*)^{-1}(1, 0, 0)^t)^{-1} = 3.7969.$$

Por tanto el intervalo HPD al 95 % para β_0 es

$$(3.801353, 5.033846).$$

De igual manera, β_1 tiene una distribución t con 19 grados de libertad media= 1.5613 y precisión=250.94 por lo que el intervalo HPD al 95% para β_2 es

$$(1.06288, 1.654542).$$

Finalmente, se tiene que β_2 también se distribuye como t con 19 grados de libertad, media=8.5190 y precisión= 616470 y así el intervalo HPD al 95% está dado por

$$(8.518996, 8.51900379).$$

Como puede observarse los intervalos correspondientes a los β_i no contienen al cero y desde un punto de vista práctico los parámetros se pueden considerar diferentes de cero.

A continuación se obtendrán intervalos y regiones HPD para τ^{-1} . Ya que $\pi(\tau|Y)$ es una Gamma(9.5,55.045) se tiene entonces que

$$2(55.045)(\tau^{-1})^{-1} \sim \chi^2_{(19)}$$

y por lo tanto

$$\left(\frac{2(55.045)}{\chi^2_{(19, .975)}}, \frac{2(55.045)}{\chi^2_{(19, .025)}} \right) = (3.351293, 12.3599)$$

será un intervalo de contenido .95 de probabilidad para τ^{-1} Para determinar la región HPD de contenido .95 de probabilidad para τ^{-1} se procede numéricamente y se tiene el siguiente intervalo HPD al 95% para τ^{-1}

$$(2.869003, 11.14168)$$

el cual cumple que $\pi(\tau^{-1} = 2.86900) \doteq \pi(\tau^{-1} = 11.1416863) = .0209755$ y además

$$P(2.869003 < \tau^{-1} < 11.14168) = .95.$$

Con esto se tiene que τ^{-1} es diferente de cero, ya que al igual que para $\hat{\beta}$ su respectivo HPD no contiene al cero. Otra forma de verificar si los parámetros de regresión del modelo son diferentes de cero es a través de pruebas de hipótesis sobre los elementos de $\underline{\beta}$.

A continuación se presentan pruebas de hipótesis, considerando en primer lugar funciones de utilidad y posteriormente utilizando regiones HPD.

4.3.3 Prueba de Hipótesis

Como se mencionó en la sección 2.3.3. una prueba de hipótesis de interés es

$$H_0 : \beta_0 = \beta_1 = \beta_2 = 0 \quad \text{vs} \quad H_1 : \text{al menos una } \beta_i \neq 0$$

Esta hipótesis se probará reescribiéndola en la forma •

$$H_0 : M\underline{\beta} = \underline{b} \quad \text{vs} \quad H_1 : M\underline{\beta} \neq \underline{b}$$

con

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ y } \underline{b} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Para esta prueba se propone la siguiente función de utilidad

$$U(d_0, M\underline{\beta}) = \begin{cases} .2 & \text{si } M\underline{\beta} = \underline{b} \\ 0 & \text{si } M\underline{\beta} \neq \underline{b} \end{cases}$$

$$U(d_1, M\underline{\beta}) = \begin{cases} 1 & \text{si } M\underline{\beta} \neq \underline{b} \\ 0 & \text{si } M\underline{\beta} = \underline{b} \end{cases} \quad (4.3)$$

que le asigna a H_1 cinco veces más utilidad que a H_0 . Esto por que si se rechaza la hipótesis nula se tendría que llevar a cabo un nuevo proceso para poder apreciar otras variables que pudieran estar influyendo en el tiempo de entrega de los productos.

Además, si se considera una distribución inicial de la forma

$$\pi_0(M\underline{\beta}) = \begin{cases} .5 & \text{si } M\underline{\beta} = \underline{b} \\ (.5)\pi(M\underline{\beta}) & \text{si } M\underline{\beta} \neq \underline{b} \end{cases} \quad (4.4)$$

donde $\pi(M\underline{\beta})$ es una distribución t de tres dimensiones con media

$$M\underline{\hat{\beta}} = \begin{pmatrix} 4.417 \\ 1.5613 \\ 0.008519 \end{pmatrix}$$

y matriz de precisión

$$(M(A^*)^{-1}M^t)^{-1} = \begin{pmatrix} 3.7969 & 27.268 & 1269.7 \\ 27.268 & 250.94 & 11393 \\ 1269.7 & 11393 & 616470 \end{pmatrix}$$

por lo visto en la sección 2.4.3 se rechazará H_0 sí y sólo si

$$T = \frac{f(Y|M\underline{\beta} = \underline{b})}{\int_{\Theta} a_1(M\underline{\beta})\pi(M\underline{\beta})f(Y|M\underline{\beta})dM\underline{\beta}} < \frac{1-q}{a_0q}.$$

En este caso

$$\begin{aligned} a_0 &= .2 \\ a_1(M\underline{\beta}) &= 1 \\ q &= p = .5, \end{aligned}$$

entonces

$$T = (3.7239)(10)^{-68} \leq 5,$$

y por tanto se debe rechazar H_0 .

Ahora considerando como función de utilidad la divergencia logarítmica de Kullback-Leibler.

$$\begin{aligned} U(d_0, \theta) &= .2 - \delta(M\underline{\beta}; M\underline{\beta} = \underline{b}) \\ U(d_1, \theta) &= 1 - \delta(M\underline{\beta}, M\underline{\beta} \neq \underline{b}) \end{aligned}$$

y con la misma distribución inicial, $\pi_0(M\underline{\beta})$, se tiene que el estadístico de prueba

$$T = (3)(19) \frac{[M\underline{\beta} - \underline{b}]^t [M(X^tX)^{-1}M^t]^{-1} [M\underline{\beta} - \underline{b}]}{Y^t [I_n - X(X^tX)^{-1}X^t] Y} = 1.9498(10)^5$$

y como $T > \delta^* = \frac{2(2-1)}{(3)(1)} = -.053333$ se rechaza H_0 .

De acuerdo a los resultados de las pruebas anteriores, se tiene que para el modelo propuesto al menos uno de los coeficientes de las variables independientes es diferente de cero.

A continuación se realizan pruebas de hipótesis del tipo

$$H_0 : \beta_i = 0 \quad \text{vs} \quad H_1 : \beta_i \neq 0$$

usando regiones HPD.

Desde que β_0 tiene una distribución t con 19 grados de libertad, media = 4.417 y precisión $P_1 = 3.7969$ se sigue que, la región HPD de .95 de probabilidad para β_0 es

$$R_{.95}(\beta_0) = \{\beta_0 : (\beta_0 - 4.417)^2 (3.7969)(\beta_0 - 4.417) \leq F_{(1,19,.95)}\}$$

por lo que H_0 se rechazará si y sólo si

$$T = (4.417)(3.7969)(4.417) > F_{(1,19,.95)}$$

como $T = 74.077$ y $F_{(1,19,.95)} = 4.38288$, se rechaza la hipótesis de que $\beta_0 = 0$.

Para probar la hipótesis

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0$$

se sigue el mismo procedimiento que para β_0 y como en este caso $T = 611.7058 > F_{(1,19,.95)}$, entonces se rechaza H_0 .

Finalmente, también en este caso, la hipótesis $H_0 : \beta_2 = 0$ se rechaza ya que

$$T = 44.7392 > F_{(1,19,.95)}.$$

Aún cuando ya se verificó que los parámetros del modelo son estadísticamente diferentes de cero, resta verificar si el modelo es adecuado, en el sentido de si realmente existe una relación lineal entre la variable dependiente y las variables independientes. Además es necesario corroborar si se cumplen los supuestos presentados en el capítulo tres, así como identificar si existen observaciones aberrantes y/o influyentes. Por lo que a continuación se presenta un análisis de diagnóstico para el modelo propuesto.

4.3.4 Diagnóstico

Como ya se mencionó, el diagnóstico del modelo se puede llevar a cabo a través de una exploración gráfica de los residuos los cuales están determinados por

$$r_i = y_i - \hat{y}_i.$$

Tabla 4.3.4.1

y_i	\hat{y}_i	$r_i = y_i - \hat{y}_i$
16.68	20.117	-3.437
11.5	10.976	0.52444
12.03	11.998	0.032156
14.88	11.344	3.5358
13.75	15.063	-1.313
18.11	18.158	-0.04767
8	8.4772	-0.47722
17.83	17.135	0.69462
40.33	35.259	5.0713
21	21.862	-0.86174
13.5	12.835	0.66502
19.75	17.721	2.0291
24	22.285	1.7146
29	26.641	2.3591
15.35	15.489	-0.13894
19	16.471	2.5291
9.5	9.4081	0.091937
35.1	37.519	-2.4186
17.9	21.223	-3.3228
18.75	22.302	-3.5525
19.83	22.317	-2.4872
10.75	11.94	-1.1905

Los residuos obtenidos del modelo ajustado se muestran en la tabla 4.3.4.1 junto con las observaciones y_i y sus correspondientes valores

ajustados \hat{y}_i . En esta tabla se puede apreciar que los residuos de mayor magnitud están asociados a las observaciones 1, 4, 9 y 20. Las cuales se pueden considerar observaciones potencialmente aberrantes.

A continuación se presenta una exploración gráfica, realizada con la ayuda del paquete Statistica (CSS), para determinar si los supuestos sobre el modelo son correctos.

Gráfica de papel normal. En esta gráfica, figura 4.1, se puede observar que la distribución del comportamiento de los residuos no presenta fuertes desviaciones a lo largo de la línea recta, por lo que se puede considerar que no hay indicios de violación al supuesto de normalidad. Por otro lado se puede apreciar la presencia de un residuo distanciado, el que corresponde a la observación Y_9 la cual es un candidato a ser considerado observación atípica.

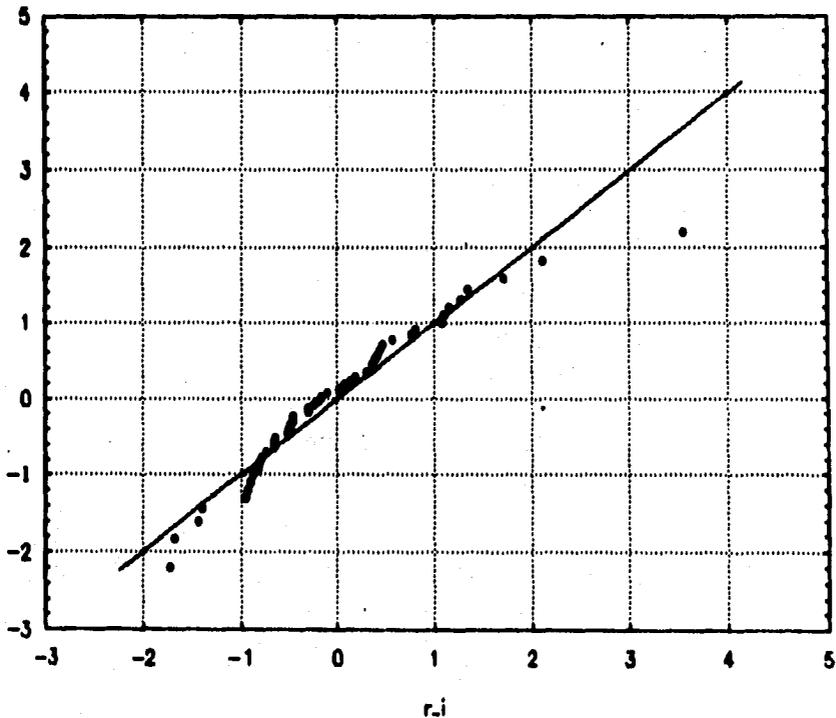


Figura 4.1: Gráfica de papel normal

Gráfica de r_i vs t_i . En este problema no es posible presentar la gráfica de los residuos contra el tiempo, porque no se cuenta con el orden (tiempo) en el que fueron tomadas las observaciones.

Gráfica de r_i vs \hat{y}_i . La gráfica de residuos contra los valores ajustados, figura 4.2, no muestra ningún patrón de comportamiento obvio, por lo que se puede suponer que la varianza de los errores es constante.

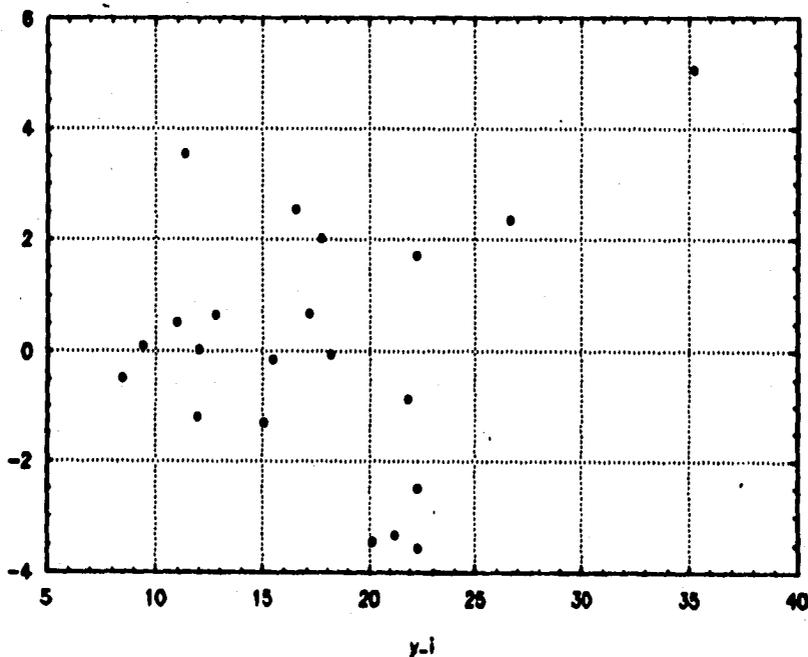


Figura 4.2 : Gráfica de r_i vs \hat{y}_i

Gráfica de r_i vs X_i . A continuación se presenta las gráficas de residuos contra cada variable independiente.

En las figura 4.3, no se observa ningún patrón significativo por lo que se puede decir que la suposición de linealidad del modelo es correcta.

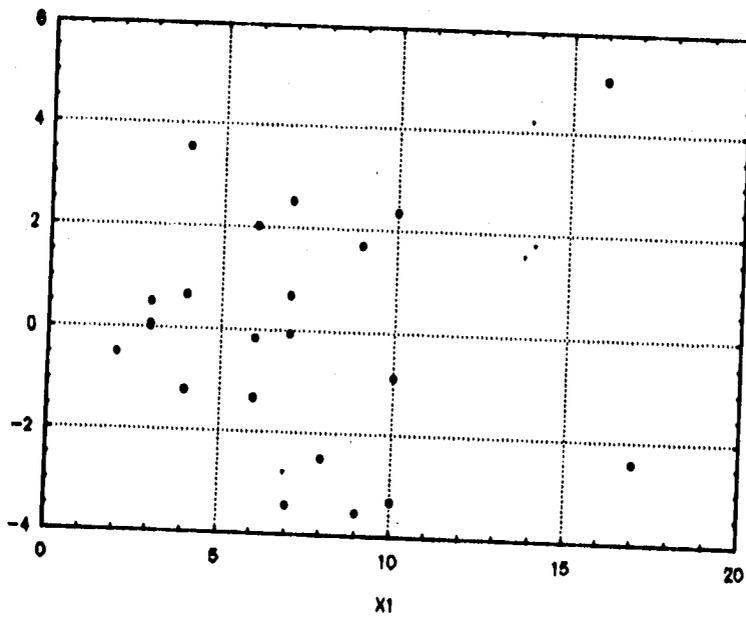


Figura 4.3: Gráfica de r_i vs X_1

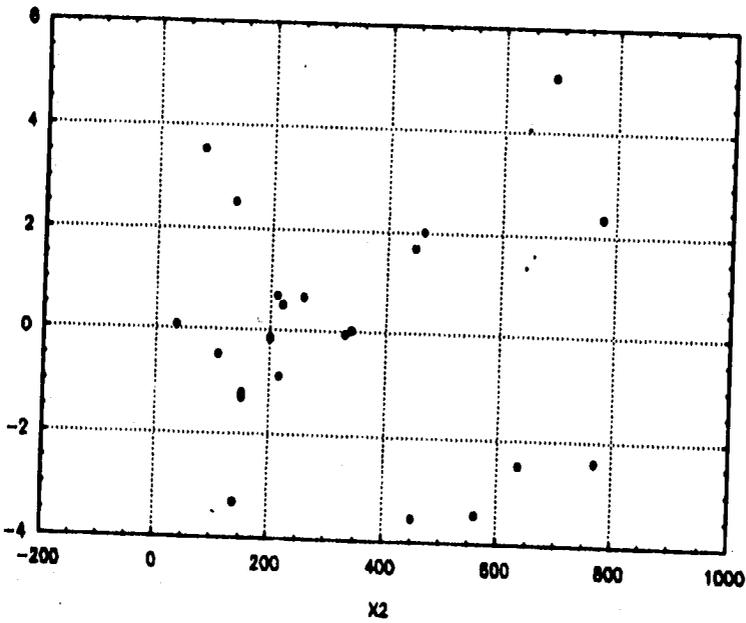


Figura 4.3: Gráfica de r_i vs X_2

De la exploración gráfica anterior, se puede concluir que no hay evidencias significativas que indiquen violaciones a los supuestos presentados en la sección 3.2.2.

Diagnóstico para observaciones atípicas

En esta sección se usarán las definiciones presentadas en la sección 3.2.2. para la identificación de observaciones aberrantes y/o influyentes.

Se aplicará la definición 3.2.1. para determinar si alguna observación y_i es aberrante o no. Para ésto es necesario construir regiones HPD de la distribución predictiva final correspondiente, $\pi(y|Y_{(i)})$, de contenido .95 de probabilidad.

El procedimiento para encontrar la distribución predictiva de la observación i -ésima es el mismo para $i = 1, \dots, 22$ por ello sólo se muestra el caso para la primera observación.

Sea

$$X_1 = (1, 7, 560), \quad y_1 = (16.68),$$

se tiene que $\pi(y_1|Y_{(1)})$ es una distribución t con 18 grados de libertad, media

$$X_1 \hat{\beta}_{(1)} = X_1 ((X_{(1)}^t X_{(1)})^{-1} X_{(1)}^t Y_{(1)}) = 20.682$$

y precisión

$$\begin{aligned} \tilde{A}_1 &= 18(Y_{(1)}^t Y_{(1)} - Y_{(1)}^t X_{(1)}(X_{(1)}^t X_{(1)})^{-1} X_{(1)}^t Y_{(1)})^{-1} \times \\ &\quad (1 + X_1(X_{(1)}^t X_{(1)})X_1^t)^{-1} \\ &= 1.6046. \end{aligned}$$

Donde $X_{(1)}$ y $Y_{(1)}$ son la matriz X sin el primer renglón y el vector Y sin la primera entrada, respectivamente. Por lo que la región predictiva final HPD con .95 de probabilidad para y_1 es

$$R_{.95}(y_1) = \left\{ y_1 : PD(1) = (y_1 - 20.682)^t (1.6046)(y_1 - 20.682) \leq F_{(1,18,.95)} \right\}$$

donde $F_{(1,18,.95)} = 4.416$.

Dado que $PD(1) \leq 4.416$ se tiene que y_1 pertenece a $R_{.95}(y_1)$ y por lo tanto no se considera una observación aberrante.

Procediendo de forma análoga para las 21 observaciones restantes se tiene que la única observación que resulta ser aberrante es la observación y_9 ya que $PD(9) = 8.8116$ y entonces $PD(9) \geq 4.416$.

Como se desconoce el proceso de obtención de los datos y la manera en que fueron registrados, no se puede aplicar ninguna técnica de corrección o bien eliminar el dato y_9 del conjunto de observaciones.

Una vez detectadas las observaciones aberrantes se realizará un análisis para determinar que observaciones son influyentes en las estimaciones de los parámetros del MRLM. Para esto se aplicará la definición 3.2.3. En principio será necesario construir regiones HPD de la distribución final $\pi(\underline{\beta}|Y_{(i)})$.

Se tiene que la distribución final de $\underline{\beta}$ sin considerar la observación y_1 , $\pi(\underline{\beta}|Y_{(1)})$, es una t de tres dimensiones con 18 grados de libertad, media

$$\underline{\hat{\beta}}_{(1)} = ((X_{(1)}^t X_{(1)})^{-1} X_{(1)}^t Y_{(1)}) = \begin{pmatrix} 4.5535 \\ 1.4919 \\ 0.010162 \end{pmatrix}$$

y precisión

$$A_{(1)}^* = \frac{X_{(1)}^t X_{(1)}}{s_{(1)}^2} = \begin{pmatrix} 3.9238 & 28.214 & 1270 \\ 28.214 & 262.52 & 11602 \\ 1270 & 11602 & 608820 \end{pmatrix}$$

De aquí se sigue que es

$$R_{.95}(\underline{\hat{\beta}}) = \{\underline{\hat{\beta}} : BD(1) \leq F_{(1,18,.95)}\}$$

donde $BD(1) = (\underline{\hat{\beta}} - \underline{\hat{\beta}}_{(1)})^t A_{(1)}^* (\underline{\hat{\beta}} - \underline{\hat{\beta}}_{(1)})$. Evaluando, se tiene que

$$BD(1) =$$

$$\left[\begin{pmatrix} 4.417 - 4.5535 \\ 1.5613 - 1.4912 \\ 0.008519 - 0.010162 \end{pmatrix}^t \begin{pmatrix} 3.9238 & 28.214 & 1270 \\ 28.214 & 262.52 & 11602 \\ 1270.0 & 11602 & 608820 \end{pmatrix} \right] \times$$

$$\begin{pmatrix} 4.417 - 4.5535 \\ 1.5613 - 1.4912 \\ 0.008519 - 0.010162 \end{pmatrix} = .012101.$$

Como $BD(1) = .012101 \leq 4.416 = F_{(1,18,.95)}$, y_1 no influye sobre $\hat{\beta}$.

De manera similar se procede para las 21 observaciones restantes y resulta que ninguna observación es influyente sobre $\hat{\beta}$.

Una vez realizado el análisis sobre la influencia en la estimación de $\hat{\beta}$ resta reproducirlo para la estimación de τ^{-1} .

Para ello es necesario determinar regiones HPD para $\hat{\tau}^{-1}$ usando la distribución final de τ^{-1} dado $Y_{(i)}$, $\pi(\tau^{-1}|Y_{(i)})$.

De acuerdo a la sección 2.4 se tiene que la distribución final $\pi(\tau^{-1}|Y_{(1)})$ es una distribución Gamma Inversa con parámetros

$$\begin{aligned}\alpha_{1(1)}^* &= 9 \\ \lambda_{1(1)}^* &= 48.167.\end{aligned}$$

Por lo visto en la sección 3.3.2, el intervalo HPD para τ^{-1} , sin considerar la observación y_1 , está dado por

$$\left(\frac{\hat{\tau}_{1(1)}^{-1} 2(\alpha_{1(1)}^* - 1)}{\chi_{(2\alpha_{1(1)}^*, .975)}^2}, \frac{\hat{\tau}_{1(1)}^{-1} 2(\alpha_{1(1)}^* - 1)}{\chi_{(2\alpha_{1(1)}^*, .025)}^2} \right)$$

donde

$$\begin{aligned}\hat{\tau}_{1(1)}^{-1} &= (\alpha_{1(1)}^* - 1)^{-1} \lambda_{1(1)}^* = 6.020875, \\ \chi_{(18, .975)}^2 &= 38.08 \\ \text{y } \chi_{(18, .025)}^2 &= 8.231.\end{aligned}$$

Es decir, el intervalo HPD para τ^{-1} es

$$(2.529779 \leq \tau^{-1} \leq 11.703802)$$

Como el estimador por pérdida cuadrática $\hat{\tau}_1^{-1} = (\alpha_1^* - 1)^{-1} \lambda_1^* = 6.4759$ queda dentro de el intervalo HPD se tiene que y_1 no influye en el estimador $\hat{\tau}_1^{-1}$. Realizando un análisis similar para las 21 observaciones restantes se obtuvo que ninguna de las 22 observaciones y_i , $i = 1, \dots, 22$, es influyente sobre el estimador $\hat{\tau}_1^{-1}$. Por lo anterior se tiene que sólo existió una observación aberrante, Y_9 , y ninguna observación influyente, en la estimación de los parámetros del modelo.

Resumiendo esta parte se tiene que de acuerdo a las pruebas de hipótesis realizadas sobre los parámetros de regresión, los intervalos de

probabilidad obtenidos y el análisis de diagnóstico realizado, se puede considerar que el modelo

$$\hat{Y} = 4.4176 + 1.5613X_1 + .008519X_2 \quad (4.5)$$

es apropiado para representar la relación entre el tiempo de entrega (Y), el número de productos surtidos (X_1) y la distancia recorrida para la entrega (X_2). Por lo que se está en condiciones de obtener las predicciones del tiempo de entrega como una función del número de productos surtidos y la distancia requerida.

4.3.5 Predicción

Supóngase que la embotelladora desea obtener predicciones sobre el tiempo de entrega de tres recorridos, donde en el primer recorrido se tienen que entregar 8 productos a una distancia de 275 pies. En el segundo recorrido se tienen que entregar 15 productos a una distancia de 700 pies y en el tercer recorrido se tienen que entregar 28 productos a una distancia de 1000 pies.

De los datos se tiene que

$$Z = \begin{pmatrix} 1 & 8 & 275 \\ 1 & 15 & 700 \\ 1 & 28 & 1000 \end{pmatrix}$$

Por lo tanto, de la sección 3.3.1, se tiene que la distribución predictiva final para los tres tiempos de entrega $W^t = (w_1, w_2, w_3)$ es una t de tres dimensiones con 19 grados de libertad, vector de localización

$$Z\hat{\beta} = (19.250, 33.8, 56.652)^t$$

y matriz de precisión

$$\hat{A} = \begin{pmatrix} 0.16357 & -0.0043534 & -0.0076975 \\ -0.0043534 & 0.15397 & -0.033483 \\ -0.0076975 & -0.033483 & 0.077641 \end{pmatrix}$$

De esto, se tiene que la mejor predicción \hat{W} , usando una función de utilidad cuadrática, es

$$\hat{W} = E_{\pi(W|Y)}(W) = \begin{pmatrix} 19.250 \\ 33.8 \\ 56.652 \end{pmatrix}$$

Es decir, para el primer recorrido se estima, considerando una función de utilidad cuadrática, que el tiempo de entrega será de 19.25 minutos. Para el segundo recorrido se estima un tiempo de entrega de 33.8 minutos y para el tercer recorrido, un tiempo de 56.652 minutos.

Por otro lado, se pueden obtener intervalos HPD, al 95%, para las estimaciones individuales de los tiempos de entrega considerando la distribución final de cada w_i . De las propiedades de una distribución t se tiene los siguientes intervalos de predicción individuales al 95% de probabilidad,

para w_1

(18.55051, 19.94949),

para w_2

(33.12135, 34.47865)

y para w_3

(56.17008, 57.1339).

Capítulo 5

Análisis Bayesiano de un Modelo de Diseño de Experimentos

5.1 Introducción

En los capítulos dos y tres se presentaron los desarrollos teóricos del análisis bayesiano para el MRLM, en el cual se consideró que la matriz X era de rango completo. Sin embargo, existen problemas de regresión asociados con variables independientes categóricas y cuya matriz X no es de rango completo. Estos problemas dan origen a los llamados modelos de diseños de experimentos. En este capítulo se establecen las bases para el análisis bayesiano de este tipo de modelos, para ésto en la sección 5.2 se presenta una introducción a los modelos de diseños de experimentos y en la sección 5.3 se muestran las bases para el análisis bayesiano de un modelo de diseño de experimentos con dos factores aditivos. Finalmente, como una forma de ilustrar los desarrollos presentados en el capítulo, en la sección 5.4 se presenta un ejemplo de un diseño de experimentos con dos factores aditivos, así como el correspondiente análisis bayesiano.

5.2 Modelos de Diseño de Experimentos.

Una clase especial de problemas estadísticos surge cuando se tiene un modelo lineal de rango incompleto asociado con variables independientes categóricas. En estos modelos es posible asignar códigos a las categorías de las variables independientes y llevar a cabo el correspondiente análisis de regresión. Sin embargo, la asignación de códigos a las categorías de las variables independientes es en general poco realista y está cargada de problemas conceptuales (Searle, 1971). Estos problemas de asignación se pueden evitar usando la técnica de análisis de regresión sobre variables *dummy*, es decir variables que sólo toman 2 valores, por ejemplo 0 y 1. La discusión del análisis de regresión sobre variables *dummy* involucra la noción de factores y niveles. La palabra *factor* denota lo que hasta ahora se ha llamado una variable. Las categorías en las cuales cada factor(variable) se divide son llamadas *niveles del factor*.

El uso de "factor" en lugar de "variable" enfatiza que las variables no pueden ser medidas precisamente por valores numéricos. El término "niveles" enfatiza que las categorías de un factor son justamente divisiones arbitrarias sin una imposición de valores asignados. A continuación se presenta un ejemplo para introducir la notación y terminología con variables *dummy*.

Ejemplo 5.2.1

Varios experimentos se realizan cada año en agricultura para investigar los efectos sobre el crecimiento de diferentes variedades de plantas a las cuales se les aplican varios tipos de fertilizantes (tratamientos). Supóngase que se tienen datos de 6 plantas, correspondientes a 3 variedades de plantas en combinación con 2 tipos de fertilizante. La tabla 5.1 muestra los datos.

La observación y_{ijk} representa el crecimiento de la k -ésima planta de variedad i que recibió el fertilizante j . Aquí la variedad de planta, es el factor a , y el fertilizante es el factor b , el tipo de variedad y de fertilizante corresponden a los niveles de cada factor, respectivamente. En este caso el factor a tiene 3 niveles, así $i = 1, 2, 3$ y el factor b tiene dos niveles, por lo que $j = 1, 2$ a continuación se escribe el modelo de

Tabla 5.1

Variedad	Fertilizante	
	1	2
1	y_{111}, y_{112}	y_{121}
2	y_{211}	y_{221}
3	y_{311}	

regresión usando 5 variables dummy $x_{ijk,l}$, con $l = 1, \dots, 5$ y 5 coeficientes de regresión a_1, a_2, a_3, b_1 y b_2 .

Por otro lado el coeficiente de regresión β_0 será escrito como μ . Así, el vector de parámetros $\underline{\beta}$ será

$$\underline{\beta} = (\mu, a_1, a_2, a_3, b_1, b_2)^t.$$

La ecuación de regresión para y_{ijk} usando esta notación es

$$y_{ijk} = \mu + a_1 x_{ijk,1} + a_2 x_{ijk,2} + a_3 x_{ijk,3} + b_1 x_{ijk,4} + b_2 x_{ijk,5} + \epsilon_{ijk}.$$

Así, para la planta 1 de variedad 1 que recibe el tratamiento 2 se tiene que $x_{121,1} = 1$, $x_{121,2} = 0$, $x_{121,3} = 0$, $x_{121,4} = 0$ y $x_{121,5} = 1$. De esta manera las ecuaciones de regresión para los datos de los crecimientos de la tabla 5.1 son

$$\begin{aligned} y_{111} &= \mu + a_1(1) + a_2(0) + a_3(0) + b_1(1) + b_2(0) + \epsilon_{111} \\ y_{112} &= \mu + a_1(1) + a_2(0) + a_3(0) + b_1(1) + b_2(0) + \epsilon_{112} \\ y_{121} &= \mu + a_1(1) + a_2(0) + a_3(0) + b_1(0) + b_2(1) + \epsilon_{121} \\ y_{211} &= \mu + a_1(0) + a_2(1) + a_3(0) + b_1(1) + b_2(0) + \epsilon_{211} \\ y_{221} &= \mu + a_1(0) + a_2(1) + a_3(0) + b_1(0) + b_2(1) + \epsilon_{221} \\ y_{311} &= \mu + a_1(0) + a_2(0) + a_3(1) + b_1(1) + b_2(0) + \epsilon_{311}. \end{aligned} \quad (5.1)$$

Si Y y ϵ denotan los vectores de observaciones y términos de error respectivamente, las ecuaciones anteriores se pueden representar de la

siguiente manera

$$Y = X\underline{\beta} + \epsilon \quad (5.2)$$

donde

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \quad y$$

$$\underline{\beta} = \begin{pmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ b_1 \\ b_2 \end{pmatrix}$$

Se puede apreciar que X no es de rango completo ya que, en particular, la suma de sus columnas 2, 3 y 4 es igual a la columna 1.

Es importante resaltar que la matriz X se conoce como *matriz de incidencia o matriz de diseño*, debido a que la localización de los ceros y unos en sus entradas representa la incidencia de los términos (μ, a_i, b_i) en la respuesta y_{ijk} . Las ecuaciones 5.1 se obtuvieron desde el punto de vista de regresión sobre variables dummy. Ahora supóngase que éstas se reescriben sin mostrar explícitamente las $x_{ijk,l}$. Así, las ecuaciones 5.1 se pueden ver como

$$y_{111} = \mu + a_1 + b_1 + \epsilon_{111}$$

$$y_{112} = \mu + a_1 + b_1 + \epsilon_{112}$$

$$y_{121} = \mu + a_1 + b_2 + \epsilon_{121}$$

$$\begin{aligned}
 y_{211} &= \mu + a_2 + b_1 + \epsilon_{211} \\
 y_{221} &= \mu + a_2 + b_2 + \epsilon_{221} \\
 y_{311} &= \mu + a_3 + b_1 + \epsilon_{311}.
 \end{aligned}
 \tag{5.3}$$

Por lo que el modelo estadístico lineal para este ejemplo se puede escribir como

$$y_{ijk} = \mu + a_i + b_j + \epsilon_{ijk} \tag{5.4}$$

con $i = 1, 2, 3$, $j = 1, 2$ y k el número de observaciones (plantas), de la i -ésima variedad que recibe el j -ésimo tratamiento es decir, $k = 1, 2$.

Además los elementos de $\underline{\beta}$ (en este caso μ, a_i, b_j) no necesariamente deben verse como coeficientes de regresión ya que su significado se puede dar en sus propios términos, es decir, a μ se le puede considerar como la media general básica de crecimiento de la población, a a_i se le puede considerar como el efecto de la variedad i sobre el crecimiento, esto ya que a_1 aparece sólo en aquellas ecuaciones que pertenecen al crecimiento de la planta de la variedad 1. Similarmente, a_2 sólo aparece en las ecuaciones de crecimiento de planta de la variedad 2, y lo mismo para a_3 . Así, a_i se puede considerar como el efecto de la variedad i , sobre el crecimiento de la planta. Análogamente b_j se puede considerar el efecto del fertilizante j sobre el crecimiento. En general, al modelo

$$y_{ijk} = \mu + a_i + b_j + \epsilon_{ijk} \tag{5.5}$$

con $i = 1, \dots, t$, $j = 1, \dots, b$, $k = 1, \dots, r$, se le conoce como modelo de diseño de experimentos con dos factores aditivos.

Si el efecto de los niveles de un factor sobre la variable respuesta, no se puede medir independientemente del efecto de los niveles del otro factor, se dice que los factores no son aditivos y que existe interacción entre ellos. Así, si en el modelo 5.4 no se considera que los dos factores son aditivos, sino que existe interacción entre ellos, se tienen los

llamados *modelos de dos factores con interacción*

$$y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \epsilon_{ijk} \quad (5.6)$$

con $i = 1, \dots, t$, $j = 1, \dots, b$, $k = 1, \dots, r$ y $(ab)_{ij}$ es llamada la interacción entre el i -ésimo y el j -ésimo nivel del primer y segundo factor, respectivamente.

El modelo 5.4 puede usarse para analizar resultados experimentales de lo que se conoce como un diseño de bloques aleatorizados. En un diseño de bloques aleatorizados, las n unidades experimentales se organizan en b bloques de tal forma que cada bloque contenga t unidades, donde a cada una de las t unidades de cada bloque se le asigna aleatoriamente uno de los t tratamientos. Si no hay interacción entre los tratamientos y los bloques, un diseño de bloques aleatorizado es modelado por 5.5 tomando $r = 1$, los niveles del primer factor a_1, \dots, a_t son llamados efectos de tratamientos mientras que los efectos de bloque b_1, \dots, b_b son los niveles del segundo factor. Así, el *modelo de diseño de bloques aleatorizados* es

$$y_{ij} = \mu + a_i + b_j + \epsilon_{ij} \quad (5.7)$$

donde los ϵ_{ij} son independientes e idénticamente distribuidos $N(0, \tau^{-1})$, $\mu \in \mathfrak{R}$, $a_i, b_j \in \mathfrak{R}$.

Los modelos 5.4, 5.6 y 5.7 son ejemplos de la gran variedad de modelos de diseños de experimentos. En la siguiente sección se presenta el análisis bayesiano para el modelo 5.4.

5.3 Planteamiento y Análisis Bayesiano de un modelo de diseño de experimentos.

Sea i el i -ésimo nivel del primer factor y j el j -ésimo nivel del segundo factor, con $i = 1, \dots, t$ y $j = 1, \dots, b$.

Supóngase que para cada una de las bt combinaciones de los niveles de los factores se tienen n_{ij} medidas de la variable respuesta y supóngase que las $n = \sum_i^t \sum_j^b n_{ij}$ observaciones son independientes. Entonces si y_{ijk} es la k -ésima medida, cuando el nivel i del primer factor y el nivel j del segundo están presentes, el modelo de diseño de experimentos para dos factores es

$$y_{ijk} = m_{ijk} + \epsilon_{ijk}$$

donde los ϵ_{ijk} son variables aleatorias independientes con media cero y precisión τ , es decir, la respuesta promedio es $E(y_{ijk}) = m_{ijk}$ para toda i, j y k .

Si se considera que los dos factores son aditivos, esto es

$$m_{ijk} = \mu + a_i + b_j$$

entonces se referirá a a_i como el efecto del i -ésimo nivel del primer factor y a b_j como el efecto del j -ésimo nivel del segundo factor. Así, el modelo de diseño de experimentos con dos factores aditivos es

$$y_{ijk} = \mu + a_i + b_j + \epsilon_{ijk} \quad (5.8)$$

con $i = 1, \dots, t, j = 1, \dots, b, k = 1, \dots, r, \epsilon_{ijk} \sim N(0, \tau^{-1}), a_i, b_j$ y $\mu \in \mathfrak{R}$.

Ahora, si $\underline{\beta} = (\beta_0, \beta_1, \beta_2)'$ es un vector de $p \times 1$, con $p = 1 + t + b$, donde

$$\begin{aligned} \beta_0 &= \mu, \\ \beta_1 &= (a_1, \dots, a_t) \quad y \\ \beta_2 &= (b_1, \dots, b_b), \end{aligned}$$

el modelo 5.8 se puede escribir como

$$Y = X\underline{\beta} + \epsilon \quad (5.9)$$

donde X es la matriz de $n \times p$ ($p = 1 + b + t$)

$$X = \left(\begin{array}{cccc}
 1; 1, 0, 0, \dots, 0; 1, 0, 0, \dots, 0 & \left. \vphantom{\begin{array}{c} 1; 1, 0, 0, \dots, 0; 1, 0, 0, \dots, 0 \\ 1; 1, 0, 0, \dots, 0; 1, 0, 0, \dots, 0 \\ 1; 1, 0, 0, \dots, 0; 0, 1, 0, \dots, 0 \end{array}} \right\} n_{11} \\
 \vdots \\
 1; 1, 0, 0, \dots, 0; 1, 0, 0, \dots, 0 \\
 1; 1, 0, 0, \dots, 0; 0, 1, 0, \dots, 0 \\
 \vdots \\
 1; 1, 0, 0, \dots, 0; 0, 1, 0, \dots, 0 & \left. \vphantom{\begin{array}{c} 1; 1, 0, 0, \dots, 0; 0, 1, 0, \dots, 0 \\ 1; 1, 0, 0, \dots, 0; 0, 0, 0, \dots, 1 \\ 1; 1, 0, 0, \dots, 0; 0, 0, 0, \dots, 1 \end{array}} \right\} n_{12} \\
 \vdots \\
 1; 1, 0, 0, \dots, 0; 0, 0, 0, \dots, 1 \\
 1; 1, 0, 0, \dots, 0; 0, 0, 0, \dots, 1 & \left. \vphantom{\begin{array}{c} 1; 1, 0, 0, \dots, 0; 0, 0, 0, \dots, 1 \\ 1; 0, 0, 0, \dots, 1; 1, 0, 0, \dots, 0 \\ 1; 0, 0, 0, \dots, 1; 1, 0, 0, \dots, 0 \end{array}} \right\} n_{1b} \\
 \vdots \\
 1; 0, 0, 0, \dots, 1; 1, 0, 0, \dots, 0 \\
 1; 0, 0, 0, \dots, 1; 1, 0, 0, \dots, 0 & \left. \vphantom{\begin{array}{c} 1; 0, 0, 0, \dots, 1; 1, 0, 0, \dots, 0 \\ 1; 0, 0, 0, \dots, 1; 0, 0, 0, \dots, 1 \\ 1; 0, 0, 0, \dots, 1; 0, 0, 0, \dots, 1 \end{array}} \right\} n_{t1} \\
 \vdots \\
 1; 0, 0, 0, \dots, 1; 0, 0, 0, \dots, 1 \\
 1; 0, 0, 0, \dots, 1; 0, 0, 0, \dots, 1 & \left. \vphantom{\begin{array}{c} 1; 0, 0, 0, \dots, 1; 0, 0, 0, \dots, 1 \\ 1; 0, 0, 0, \dots, 1; 0, 0, 0, \dots, 1 \end{array}} \right\} n_{tb}
 \end{array} \right)$$

Y el vector de $n \times 1$

$$Y = \begin{pmatrix} Y_{111} \\ Y_{112} \\ \cdot \\ \cdot \\ Y_{11n_{11}} \\ Y_{121} \\ \cdot \\ \cdot \\ Y_{12n_{12}} \\ \cdot \\ \cdot \\ Y_{i01} \\ Y_{i02} \\ \cdot \\ \cdot \\ Y_{in_{1i}} \end{pmatrix}$$

y ϵ es el correspondiente vector de errores de dimensión n .

Se puede notar que la matriz X del modelo 5.9 es de rango menor que p , en realidad es de rango $p-2$. Ya que X no es de rango completo, no se pueden hacer inferencias directamente sobre $\underline{\beta}$ pero lo que si es posible es hacer inferencias sobre ciertas funciones lineales de $\underline{\beta}$ que sean de interés, esto se puede lograr reparametrizando el modelo 5.9. Las reparametrizaciones se pueden hacer de tal forma que a partir del modelo 5.9, de rango incompleto, se obtengan modelos de interés de rango completo.

A continuación se muestra la forma de obtener, a partir del modelo 5.9, un modelo de la forma

$$Y = D\underline{\zeta} + \epsilon \quad (5.10)$$

donde D es una matriz de $n \times k$ ($k < p$) de rango completo, $\epsilon \sim N(0, \tau^{-1}I_n)$, además Y y ϵ son igual que antes y

$$\underline{\zeta} = U\underline{\beta} \quad (5.11)$$

con U una matriz conocida de $k \times p$. Pero antes se presentan un par de definiciones.

Definición 5.3.1 Por una reparametrización del modelo $Y = X\beta + \epsilon$ se entiende una transformación del vector β al vector \underline{r} , dada por $\underline{r} = U\beta$, donde cada elemento de $\underline{r} = U\beta$ es una función estimable.

Definición 5.3.2 Un parámetro (o función de parámetros) se dice que es estimable si existe un estimador insesgado del parámetro (o de la función de parámetros).

Como $X'X$ es una matriz de $p \times p$, simétrica semidefinida positiva, existe una matriz no singular V^* de $p \times p$ tal que

$$(V^*)'X'XV^* = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}$$

donde B es una matriz de $k \times k$ de rango completo (Graybill, 1961). Si $V^* = (V, V_1)$, con V una matriz de $p \times k$, entonces

$$\begin{pmatrix} V' \\ V_1' \end{pmatrix} X'X(V, V_1) = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}$$

de lo que se tiene que $V'X'XV = B$ y $V_1'X'XV_1 = 0$. Esto implica que $V'X'$ sea de rango k y que $V_1'X' = 0$.

Por lo anterior el modelo $Y = X\beta + \epsilon$, de rango incompleto, se puede escribir como $Y = XV^*(V^*)^{-1}\beta + \epsilon$, y si $(V^*)^{-1} = U^* = (U, U_1)'$, se tiene que

$$Y = X(V, V_1) \begin{pmatrix} U \\ U_1 \end{pmatrix} \beta + \epsilon \quad \text{ó}$$

$$Y = (XV)(U\beta) + (XV_1)(U_1\beta) + \epsilon,$$

y en virtud de que $XV_1 = 0$ el modelo anterior se reduce a

$$Y = (XV)(U\beta) + \epsilon.$$

Haciendo $D = XV$ y $\underline{r} = U\beta$, el modelo $Y = X\beta + \epsilon$ se puede escribir finalmente como

$$Y = D\underline{r} + \epsilon$$

donde D es la matriz de $n \times k$ de rango completo k y por consiguiente se tiene, a través de la reparametrización 5.11, un modelo de rango completo a partir de un modelo de rango incompleto.

Como en un diseño de experimentos con dos factores aditivos es de interés investigar el efecto de los niveles de los dos factores sobre la respuesta promedio $\mu + a_i + b_j$, el punto de interés en este momento es la selección de \underline{r} , una forma de seleccionar \underline{r} es tomando.

$$\underline{r} = \begin{pmatrix} \mu + a_1 + b_1 \\ a_1 - a_2 \\ a_2 - a_3 \\ \vdots \\ a_{t-1} - a_t \\ b_1 - b_2 \\ b_2 - b_3 \\ \vdots \\ b_{s-1} - b_s \end{pmatrix} = \begin{pmatrix} \underline{r}_1 \\ \underline{r}_2 \\ \underline{r}_3 \end{pmatrix} \quad (5.12)$$

es decir, \underline{r} un vector de $k \times 1$, con $k = p - 2$, donde $\underline{r}_1 = \mu + a_1 + b_1$, $\underline{r}_2 = (a_1 - a_2, \dots, a_{t-1} - a_t)^t$ y $\underline{r}_3 = (b_1 - b_2, \dots, b_{s-1} - b_s)^t$. Esto significa que U debe ser una matriz de $k \times p$ de la forma

$$U = \begin{pmatrix} 1; 1, 0, & \dots, 0; 1, 0 & 0, \dots, 0 \\ 0; 1, -1, 0, & \dots, 0; 0, 0 & 0, \dots, 0 \\ 0; 0, 1, -1, & \dots, 0; 0, 0 & 0, \dots, 0 \\ \vdots \\ \vdots \\ 0; 0, \dots, & 0, 0; 1, -1, 0 & \dots, 0 \\ 0; 0, \dots, & 0, 0; 0, 1, -1 & \dots, 0 \\ \vdots \\ \vdots \\ 0; 0, \dots, & 0, 0; 0 & 0, \dots, 1, -1 \end{pmatrix} \quad (5.13)$$

Una vez especificada la reparametrización $\underline{r} = U\underline{\beta}$, la matriz V^* queda determinada de acuerdo a la selección de \underline{r} y con ello queda especificado

el modelo reparametrizado de rango completo. A continuación se presenta el análisis bayesiano correspondiente al modelo 5.11, con \underline{z} como en 5.12, considerando la distribución inicial no informativa

$$\pi(\underline{z}, \tau) \propto \frac{1}{\tau} \quad \underline{z} \in \mathfrak{R}, \quad \tau > 0. \quad (5.14)$$

Se mencionó anteriormente que el análisis también se puede llevar a cabo considerando una distribución inicial Normal-Gamma.

Dada la muestra $Y = \{y_1, \dots, y_n\}$, se tiene que la función de verosimilitud para \underline{z} y τ es

$$L(Y|\underline{z}, \tau) \propto \tau^{n/2} \exp\left\{-\frac{(n-k)\tau s^2}{2} - \frac{s^2 \tau (\underline{z} - \hat{\underline{z}})' D' D (\underline{z} - \hat{\underline{z}})}{2s^2}\right\}$$

donde

$$\begin{aligned} s^2 &= \frac{(Y - \hat{Y})'(Y - \hat{Y})}{n-k} \\ \hat{\underline{z}} &= (D'D)^{-1} D'Y \\ \hat{Y} &= D\hat{\underline{z}} \\ n &= rbt, \\ k &= b+t-1. \end{aligned}$$

Por lo que la distribución final de \underline{z} y τ es

$$\pi(\underline{z}, \tau|Y) \propto \tau^{\frac{r}{2}t-1} \exp\left\{-\frac{(n-k)\tau s^2}{2} - \frac{s^2 \tau (\underline{z} - \hat{\underline{z}})' D' D (\underline{z} - \hat{\underline{z}})}{2s^2}\right\}.$$

Es decir, $\pi(\underline{z}, \tau)$ es una Normal-Gamma con parámetros

$$\begin{aligned} \alpha_1^* &= \frac{n-k}{2} = \frac{rbt - (t+b-1)}{2} \\ \lambda_1^* &= \frac{(n-k)s^2}{2} \\ A_1^* &= (D'D) \\ \mu_1^* &= \hat{\underline{z}}. \end{aligned}$$

Se puede mostrar (Broemeling, 1985) que $\hat{\underline{z}}$ es tal que su primera componente es $\bar{y}_{..} + \bar{y}_{i..} + \bar{y}_{.j.}$, sus siguientes $t-1$ componentes son

$\bar{y}_{i..} - \bar{y}_{i+1..}$, $i = 1, \dots, t-1$, y sus últimas $b-1$ componentes son $\bar{y}_{.j.} - \bar{y}_{.j+1.}$, con $j = 1, \dots, b-1$ donde

$$\begin{aligned}\bar{y}_{...} &= \sum_i \sum_j \sum_k y_{ijk} / btr \\ \bar{y}_{i..} &= \sum_j \sum_k y_{ijk} / br \\ \bar{y}_{.j.} &= \sum_i \sum_k y_{ijk} / tr.\end{aligned}$$

Una vez obtenida la distribución final $\pi(\underline{r}, \tau)$, se puede llevar a cabo un análisis similar al presentado en los capítulos 2 y 3. A continuación sólo se presentan los desarrollos correspondientes a pruebas de hipótesis con la finalidad de investigar el efecto de los niveles de los dos factores sobre la respuesta promedio.

Si es de interés hacer inferencias sobre el primer factor a , se debe obtener la distribución final marginal de $\underline{r}_2 = (a_1 - a_2, a_2 - a_3, \dots, a_{t-1} - a_t)$, la cual se puede encontrar una vez que se conoce la distribución final de \underline{r} .

Aplicando los resultados del capítulo 2 se tiene que la distribución final de \underline{r} es una t k -dimensional con $rbt - (b+t-1)$ grados de libertad, vector de localización $\hat{\underline{r}}$ y matriz de precisión

$$P = \frac{[(rbt) - (b+t-1)]}{(Y - D\hat{\underline{r}})'(Y - D\hat{\underline{r}})} D' D.$$

Así, como $\underline{r} = (\underline{r}_1, \underline{r}_2, \underline{r}_3)'$, \underline{r}_2 tiene una distribución t $(t-1)$ -dimensional con $rbt - (t+b-1)$ grados de libertad, vector de localización

$$E_2 = (O_1, I_{t-1}, O_3)\hat{\underline{r}},$$

donde O_1 es un vector de $(t-1) \times 1$ cuyas entradas son todas cero, I_{t-1} es la matriz identidad de rango $t-1$, y O_3 es una matriz de $(t-1) \times (b-1)$ de ceros. La matriz de precisión de \underline{r}_2 esta dada por

$$P_2 = [(O_1, I_{t-1}, O_3)P^{-1}(O_1, I_{t-1}, O_3)']^{-1}$$

Todos los resultados anteriores se obtienen por propiedades de la distribución t multivariada (Broemeling, 1985).

Ahora, si se está interesado en hacer inferencias sobre el segundo factor b , entonces se puede obtener la distribución final marginal de $r_3 = (b_1 - b_2, b_2 - b_3, \dots, b_{b-1} - b_b)$. Mediante un procedimiento similar al que se utilizó para r_2 se tiene que la distribución final de r_3 es una t $(b-1)$ -dimensional con $rbt - (t + b - 1)$ grados de libertad, vector de localización

$$E_3 = (O_1, O_2, I_{b-1})\hat{\xi}$$

donde O_2 es una matriz de $(b-1) \times (t-1)$ de ceros, I_{b-1} es la matriz identidad de rango $b-1$ y en este caso O_1 es un vector de $(b-1) \times 1$ de ceros, y matriz de precisión

$$P_3 = [(O_1, O_2, I_{b-1})P^{-1}(O_1, O_2, I_{b-1})']^{-1}.$$

Por otro lado, si se quieren hacer inferencias sobre τ , por lo visto en el capítulo 2, estas se pueden hacer tomando en cuenta que la distribución final de τ es una Gamma con parámetros $(\alpha_1^*, \lambda_1^*)$.

Una vez que se han especificado las distribuciones finales de r_2 y r_3 , se pueden realizar pruebas de hipótesis de acuerdo a las inferencias que sean de interés.

Como se mencionó anteriormente generalmente se está interesado en el efecto de los dos factores sobre la respuesta promedio m_{ijk} y en particular si todos los niveles de un factor tienen el mismo efecto. De esta forma se puede estar interesado en la prueba de hipótesis

$$H_0 : a_1 = a_2 = \dots = a_t$$

la cual es equivalente a

$$H_0 : r_2 = 0 \tag{5.15}$$

es decir, si los t niveles del primer factor tiene el mismo efecto sobre la variable respuesta.

Como r_2 tiene una distribución final t $(t-1)$ -dimensional con media E_2 y matriz de precisión P_2 , por lo visto en el capítulo 2, la variable aleatoria

$$F_2(r_2) = (t-1)^{-1}(r_2 - E_2)'P_2(r_2 - E_2)$$

tiene una distribución F con $(t-1)$ y $rbt - (t+b-1)$ grados de libertad. Así, una región de contenido $1 - \gamma$ de probabilidad para \underline{r}_2 es

$$S_{1-\gamma}(\underline{r}_2) = \{\underline{r}_2 : F_2(\underline{r}_2) \leq F_{(t-1,rbt-(t+b-1),1-\gamma)}\}$$

donde $F_{(t-1,rbt-(t+b-1),1-\gamma)}$ es el cuantil de orden $1-\gamma$ de una distribución F con $(t-1)$ y $rbt - (t+b-1)$, grados de libertad.

Por lo que la hipótesis

$$H_0 : a_1 = a_2 = \dots = a_t$$

se rechaza cuando $\underline{r}_2 = \underline{0}$ no está contenido en $S_{1-\gamma}(\underline{r}_2)$ o equivalentemente cuando $F_2(\underline{0}) > F_{(t-1,rbt-(t+b-1),1-\gamma)}$.

Se puede emplear un desarrollo similar para probar la hipótesis

$$H_0 : b_1 = b_2 = \dots = b_b(\underline{r}_3 = 0)$$

construyendo una región HPD para \underline{r}_3 . Así, si alguna de las pruebas anteriores se rechaza, entonces se puede decir que los niveles del factor correspondiente tienen el mismo efecto sobre la respuesta promedio.

A continuación se presenta un breve ejemplo, para ilustrar los desarrollos presentados en este capítulo.

5.3.1 Aplicación

La impureza presente en un producto químico se ve afectada por dos factores presión y temperatura. Por lo que se tiene interés en saber si los diferentes niveles de presión a los que se somete el producto, tienen el mismo efecto sobre la impureza de éste. Así, también se desea saber si los diferentes grados de temperatura a los que se expone el producto, tienen el mismo efecto sobre su grado de impureza. Para ello se realiza un diseño de experimentos con dos factores aditivos (presión y temperatura) y se observan los resultados que se muestran en la tabla 5.2. Los distintos grados de temperatura son 100,125,150 y los niveles de presión 25,30,35,40 y 45. El modelo de diseño de experimentos asociado a este ejemplo es

$$y_{ij1} = \mu + a_1x_{ij1.1} + a_2x_{ij1.2} + a_3x_{ij1.3} + b_1x_{ij1.4} + b_2x_{ij1.5} + b_3x_{ij1.6} + b_4x_{ij1.7} + b_5x_{ij1.8} + \epsilon_{ij1}, \quad (5.16)$$

Tabla 5.2

Temperatura (°F)	Presión					Y _j
	25	30	35	40	45	
100	5	4	6	3	5	23
125	3	1	4	2	3	13
150	1	1	3	1	2	8
Y _i	9	6	13	6	10	44=Y _{..}

donde $i = 1, 2, 3$, $j = 1, 2, 3, 4, 5$, μ representa el promedio general de impureza en los productos químicos, y las x_{ijl} , con $l = 1, 2, \dots, 6$ son variables dummy que toman valores en $\{0, 1\}$.

En forma matricial el modelo 5.16 se puede ver como

$$Y = X\beta + \epsilon \quad (5.17)$$

con

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad y$$

$$\underline{\beta} = \begin{pmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix}$$

Así, $\underline{\beta}$ es un vector de $p \times q$, con $p=9$. Se puede ver que el rango de X es $k=7$, es decir, la matriz X es de rango incompleto.

Dado que el interés del experimento es investigar si los diferentes niveles de cada factor tienen el mismo efecto sobre la impureza del producto químico, como se mencionó anteriormente, una forma conveniente de seleccionar \underline{r} es tomando

$$\underline{r} = \begin{pmatrix} \mu + a_1 + b_1 \\ a_1 - a_2 \\ a_2 - a_3 \\ b_1 - b_2 \\ b_2 - b_3 \\ b_3 - b_4 \\ b_4 - b_5 \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} \quad (5.18)$$

Así, \underline{r} es un vector de $k \times 1$, con $k=7$, donde $r_1 = \mu + a_1 + b_1$, $r_2 = (a_1 - a_2, a_2 - a_3)^t$ y $r_3 = (b_1 - b_2, b_2 - b_3, b_3 - b_4, b_4 - b_5)^t$. Como $\underline{r} = U\underline{\beta}$, U debe ser la matriz de 7×9 ($k \times p$) de la forma

$$U = \begin{pmatrix} 1; 1, 0, 0; 1, 0, 0, 0, 0 \\ 0; 1, -1, 0; 0, 0, 0, 0, 0 \\ 0; 0, 1, -1; 0, 0, 0, 0, 0 \\ 0; 0, 0, 0; 1, -1, 0, 0, 0 \\ 0; 0, 0, 0; 0, 1, -1, 0, 0 \\ 0; 0, 0, 0; 0, 0, 1, -1, 0 \\ 0; 0, 0, 0; 0, 0, 0, 1, -1 \end{pmatrix} \quad (5.19)$$

Hasta el momento, se tiene el modelo de rango incompleto, 5.17, $Y = X\underline{\beta} + \epsilon$ y se quiere hacer inferencias sobre el vector \underline{r} , con $\underline{r} = U\underline{\beta}$,

asociado a un modelo de la forma $Y = D\underline{r} + \epsilon$ con D una matriz de rango completo. A continuación se muestra la forma de obtener el modelo $Y = D\underline{r} + \epsilon$ a partir del modelo $Y = X\underline{\beta} + \epsilon$.

Para obtener la matriz D es necesario encontrar $V^* = (V, V_1) = (U^*)^{-1}$, donde $U^* = \begin{pmatrix} U \\ U_1 \end{pmatrix}$. Dado que U es de la forma 5.19, es necesario especificar U_1 tal que U^* sea no singular. Una posible elección de U_1 es

$$U_1 = \begin{pmatrix} 0; 1 & 1 & 1; 0, & 0 & 0 & 0 & 0 \\ 0; 0 & 0 & 0; 1, & 1 & 1 & 1 & 1 \end{pmatrix}$$

la cual satisface que U^* sea no singular. Así

$$U^* = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad y$$

$$V^* = (V, V_1) = (U^*)^{-1} =$$

$$\begin{pmatrix} 1.00 & -0.66 & -0.33 & -0.80 & -0.60 & -0.40 & -0.20 & -0.33 & -0.20 \\ 0 & 0.66 & 0.33 & 0 & 0 & 0 & 0 & 0.33 & 0 \\ 0 & -0.33 & 0.33 & 0 & 0 & 0 & 0 & 0.33 & 0 \\ 0 & -0.33 & -0.66 & 0 & 0 & 0 & 0 & 0.33 & 0 \\ 0 & 0 & 0 & 0.80 & 0.60 & 0.40 & 0.20 & 0 & 0.20 \\ 0 & 0 & 0 & -0.20 & 0.60 & 0.40 & 0.20 & 0 & 0.20 \\ 0 & 0 & 0 & -0.20 & -0.40 & 0.40 & 0.20 & 0 & 0.20 \\ 0 & 0 & 0 & -0.20 & -0.40 & -0.60 & 0.20 & 0 & 0.20 \\ 0 & 0 & 0 & -0.20 & -0.40 & -0.60 & -0.80 & 0 & 0.20 \end{pmatrix}$$

Además, V^* satisface

$$(V^*)'X'XV = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}$$

con

$$B = \begin{pmatrix} 15 & -10 & -5 & -12 & -9 & -6 & -3 \\ -10 & 10 & 5 & 8 & 6 & 4 & 2 \\ -5 & 5 & 5 & 4 & 3 & 2 & 1 \\ -12 & 8 & 4 & 12 & 9 & 6 & 3 \\ -9 & 6 & 3 & 9 & 9 & 6 & 3 \\ -6 & 4 & 2 & 6 & 6 & 6 & 3 \\ -3 & 2 & 1 & 3 & 3 & 3 & 3 \end{pmatrix}$$

matriz de rango completo, de 7×7 ($k \times k$).

Como $V^* = (V, V_1)$ y V debe ser una matriz de 9×7 ($p \times k$), entonces

$$V = \begin{pmatrix} 1.0 & -0.66 & -0.33 & -0.80 & -0.60 & -0.40 & -0.20 \\ 0 & 0.66 & 0.33 & 0 & 0 & 0 & 0 \\ 0 & -0.33 & 0.33 & 0 & 0 & 0 & 0 \\ 0 & -0.33 & -0.66 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.80 & 0.60 & 0.40 & 0.20 \\ 0 & 0 & 0 & -0.20 & 0.60 & 0.40 & 0.20 \\ 0 & 0 & 0 & -0.20 & -0.40 & 0.40 & 0.20 \\ 0 & 0 & 0 & -0.20 & -0.40 & -0.60 & 0.20 \\ 0 & 0 & 0 & -0.20 & -0.40 & -0.60 & -0.80 \end{pmatrix}$$

y por lo tanto

$$D = XV = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 & -0 \\ 1 & 0 & 0 & -1 & -1 & -1 & -0 \\ 1 & 0 & 0 & -1 & -1 & -1 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & -1 & -1 & 0 & -0 \\ 1 & -1 & 0 & -1 & -1 & -1 & -0 \\ 1 & -1 & 0 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & -1 & -1 & -1 & -1 & 0 & -0 \\ 1 & -1 & -1 & -1 & -1 & -1 & -0 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix} \quad (5.20)$$

Así, se tiene el siguiente modelo de rango completo

$$Y = D\boldsymbol{\tau} + \epsilon \quad (5.21)$$

con D como en la ecuación 5.20 y $\boldsymbol{\tau}$ como en la ecuación 5.18.

Una vez obtenido el modelo 5.21, se mostrará el análisis bayesiano correspondiente.

Dada la muestra $Y = \{5, 4, 6, 3, 5, 3, 1, 4, 2, 3, 1, 1, 3, 1, 2\}$, considerando el modelo 5.21, se tiene que la función de verosimilitud para $\boldsymbol{\tau}$ y τ es

$$L(Y|\boldsymbol{\tau}, \tau) \propto \tau^{n/2} \exp\left\{-\frac{(n-k)\tau s^2}{2} - \frac{s^2\tau(\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})'D'D(\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})}{2s^2}\right\}$$

donde

$$\hat{\boldsymbol{\tau}} = (D'D)^{-1}D'Y = \begin{pmatrix} 4.66 \\ 2.00 \\ 1.00 \\ 1.00 \\ -2.33 \\ 2.33 \\ -1.33 \end{pmatrix}$$

$$\hat{Y} = D\hat{\boldsymbol{\tau}} = \begin{pmatrix} 4.66 \\ 3.66 \\ 6.00 \\ 3.66 \\ 5.00 \\ 2.66 \\ 1.66 \\ 4.00 \\ 1.66 \\ 3.00 \\ 1.66 \\ 0.66 \\ 3.00 \\ 0.66 \\ 2.00 \end{pmatrix}$$

$$\begin{aligned}
 r &= 1, \quad b = 5, \quad t = 3, \\
 n &= rbt = 15, \\
 k &= b + t - 1 = 7, \\
 s^2 &= \frac{(Y - \hat{Y})'(Y - \hat{Y})}{n - k} = 0.2500.
 \end{aligned}$$

Por lo que la distribución final conjunta de \underline{z} y τ es una Normal-Gamma con parámetros

$$\begin{aligned}
 \mu_1^* &= \hat{\underline{z}}, \\
 A_1^* &= (D^t D) = \begin{pmatrix} 15 & -10 & -5 & -12 & -9 & -6 & -3 \\ -10 & 10 & 5 & 8 & 6 & 4 & 2 \\ -5 & 5 & 5 & 4 & 3 & 2 & 1 \\ -12 & 8 & 4 & 12 & 9 & 6 & 3 \\ -9 & 6 & 3 & 9 & 9 & 6 & 3 \\ -6 & 4 & 2 & 6 & 6 & 6 & 3 \\ -3 & 2 & 1 & 3 & 3 & 3 & 3 \end{pmatrix}
 \end{aligned}$$

$$\alpha_1^* = (n - k)/2 = 4 \quad y$$

$$\lambda_1^* = (n - k)s^2/2 = 1.$$

Por lo presentado en la sección 5.3, la distribución final de \underline{z} es una t de 7 dimensiones con $15 - 7 = 8$ grados de libertad, vector de localización $\hat{\underline{z}}$ y matriz de precisión

$$P = \begin{pmatrix} 60 & -40 & -20 & -48 & -36 & -24 & -12 \\ -40 & 40 & 20 & 32 & 24 & 16 & 8 \\ -20 & 20 & 20 & 16 & 12 & 8 & 4 \\ -48 & 32 & 16 & 48 & 36 & 24 & 12 \\ -36 & 24 & 12 & 36 & 36 & 24 & 12 \\ -24 & 16 & 8 & 24 & 24 & 24 & 12 \\ -12 & 8 & 4 & 12 & 12 & 12 & 12 \end{pmatrix}$$

Así, también $r_2 = (a_1 - a_2, a_2 - a_3)^t$ tiene una distribución t de 2 dimensiones con 8 grados de libertad, vector de localización

$$\begin{aligned}
 E_2 &= (O_1, I_{1-1=2}, O_3)\hat{\underline{z}} \\
 &= \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \hat{\underline{z}}
 \end{aligned}$$

$$= \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

y matriz de precisión

$$\begin{aligned} P_2 &= [(O_1, I_{t-1}, O_3)P^{-1}(O_1, I_{t-1}, O_2)^t]^{-1} \\ &= \begin{pmatrix} 13.33 & 6.66 \\ 6.66 & 13.33 \end{pmatrix}. \end{aligned}$$

Finalmente, $\xi_3 = (b_1 - b_2, b_2 - b_3, b_3 - b_4, b_4 - b_5)^t$ tiene una distribución t de 4 dimensiones con 8 grados de libertad, vector de localización

$$\begin{aligned} E_3 &= (O_1, O_2, I_{b-1})\hat{\xi} \\ &= \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \hat{\xi} \\ &= \begin{pmatrix} 1.00 \\ -2.33 \\ 2.33 \\ -1.33 \end{pmatrix} \end{aligned}$$

y matriz de precisión

$$\begin{aligned} P_3 &= [(O_1, O_2, I_{b-1})P^{-1}(O_1, O_2, I_{b-1})^t]^{-1} \\ &= \begin{pmatrix} 9.6 & 7.2 & 4.8 & 2.4 \\ 7.2 & 14.4 & 9.6 & 4.8 \\ 4.8 & 9.6 & 14.4 & 7.2 \\ 2.4 & 4.8 & 7.2 & 9.6 \end{pmatrix} \end{aligned}$$

Una vez determinadas las distribuciones finales marginales de ξ_2 y ξ_3 , se está en condiciones de inferir si los diferentes niveles de presión tienen el mismo efecto sobre la impureza del producto, y también si los diferentes grados de temperatura tienen el mismo efecto sobre la impureza del producto químico.

Para analizar si los diferentes grados ($^{\circ}F$) de temperatura tienen el mismo efecto sobre la impureza del producto, como se mencionó anteriormente, se puede llevar a cabo la prueba de la hipótesis

$$H_0 : a_1 = a_2 = a_3$$

la cual es equivalente a probar

$$H_0 : \mathcal{L}_2 = 0.$$

Así, si no se rechaza H_0 se puede decir que la exposición del producto químico a los diferentes grados de temperatura tiene el mismo efecto sobre el grado de impureza del producto.

La prueba de hipótesis se realizará considerando regiones HPD para \mathcal{L}_2 . Desde que \mathcal{L}_2 tiene una distribución final t de 2 dimensiones con 8 grados de libertad, vector de medias E_2 y matriz de precisión P_2 , la región HPD de contenido .95 de probabilidad para \mathcal{L}_2 es

$$R_{.95}(\mathcal{L}_2) = \{ \mathcal{L}_2 : F_2(\mathcal{L}_2) \leq F_{(2,8,.95)} \}$$

donde $F_{(2,8,.95)} = 4.45897$ es el cuantil de orden .95 de una distribución F con 2 y 8 grados de libertad, y

$$F_2(\mathcal{L}_2) = (2)^{-1}(\mathcal{L}_2 - E_2)^t P_2(\mathcal{L}_2 - E_2).$$

Como $F(\mathcal{L}_2 = 0) = 46.6667 \geq F_{(2,8,.95)}$ entonces $\mathcal{L}_2 = 0$ no está contenido en $R_{.95}(\mathcal{L}_2)$ y por tanto se puede concluir que las diferentes temperaturas, $100^{\circ}F$, $125^{\circ}F$ y $150^{\circ}F$, a las que se expone el producto químico no tienen el mismo efecto en el grado de impureza que éste presenta.

Ahora, para investigar si los diferentes grados de presión tienen el mismo efecto sobre el grado de impureza del producto químico, se puede probar la hipótesis

$$H_0 : \mathcal{L}_3 = 0$$

o lo que es lo mismo

$$H_0 : b_1 - b_2 = b_2 - b_3 = b_3 - b_4 = b_4 - b_5 = 0.$$

Como \mathbf{L}_3 tiene una distribución t de 4 dimensiones con 8 grados de libertad, vector de localización E_3 y matriz de precisión P_3 se tiene que la región HPD de contenido .95 de probabilidad esta dada por

$$R_{.95}(\mathbf{L}_3) = \{\mathbf{L}_3 : F_3(\mathbf{L}_3) \leq F_{(4,8,.95)}\}$$

donde

$$F_3(\mathbf{L}_3) = (4)^{-1}(\mathbf{L}_3 - E_3)^t P_3 (\mathbf{L}_3 - E_3)$$

y $F_{(4,8,.95)} = 3.837853$ es el cuantil de orden .95 de una distribución F con 4 y 8 grados de libertad. Como $F_3(\mathbf{L}_3 = 0) = 11.60 \geq F_{(4,8,.95)}$, se tiene que $\mathbf{L}_3 = 0$ no pertenece a $R_{.95}(\mathbf{L}_3)$ y por lo tanto se puede inferir que las diferentes presiones a las que es sometido el producto químico, no tienen el mismo efecto sobre su grado de impureza.

Conclusiones

De acuerdo a lo expuesto en este trabajo, se puede decir que el enfoque bayesiano usando teoría de la decisión, ofrece una gran flexibilidad en la solución de problemas estadísticos. Lo anterior porque este enfoque permite la incorporación, al análisis estadístico, de la información que cada decisor tenga sobre el problema y la actualización del conocimiento inicial que se tenga sobre los parámetros desconocidos, al incorporar la información obtenida a partir de una muestra aleatoria.

Se mostró que se pueden tomar dos posturas en el análisis del MRLM. Por un lado, considerar una distribución inicial conjugada que refleje el conocimiento que el decisor tenga sobre los parámetros involucrados y por el otro, considerar una distribución inicial no informativa en caso de no contar con información sobre los parámetros desconocidos o bien, no tener la posibilidad de especificar ésta.

En este trabajo se consideró como distribución inicial conjugada para el MRLM, una distribución Normal-Gamma. Sin embargo, no necesariamente se debe considerar esta distribución cuando se trabaja con el MRLM, ya que esto dependerá de la información inicial con que se cuente. Independientemente de cuál sea la distribución inicial con la que se trabaje se puede llevar a cabo, salvo problemas de tipo numérico, un análisis bayesiano del MRLM considerando la metodología expuesta en este trabajo. Además, cualquier distribución inicial se puede aproximar a través de combinaciones convexas (mezclas) de distribuciones conjugadas, en este caso a través de mezclas de distribuciones Normales Gamma.

Finalmente, es importante notar la similitud de algunos resultados de estadística clásica y los resultados obtenidos desde el punto de vista bayesiano. De acuerdo a los resultados presentados en este trabajo se

puede concluir que considerando una distribución inicial no informativa y en el caso particular de funciones de utilidad cuadrática y usando la divergencia de Kullback-Leibler, se recuperan los resultados clásicos en el caso de estimación puntual del vector de regresión.

Por otro lado, considerando una distribución inicial no informativa, la región HPD para todos los parámetros de regresión, coincide con la región de confianza que se obtiene en el enfoque clásico. En este punto hay que recordar que si bien las regiones obtenidas en los dos enfoques coinciden numéricamente, las interpretaciones de éstos son muy diferentes.

Para probar la hipótesis lineal general del vector de regresión; si se considerará una distribución inicial no informativa y una función de utilidad en términos de la divergencia de Kullback-Leibler, se recupera la estadística de prueba obtenida en el procedimiento clásico usual.

En el caso del problema de predicción, la estimación puntual coincide con los resultados clásicos si se considera una distribución inicial no informativa y una función de utilidad cuadrática.

Bibliografía

1. Berger, J.O. (1985). *Statistical Decision & Bayesian Analysis*. Springer-Verlag, New York.
2. Bernardo, J.M. (1984). *Análisis Bayesiano de los Contrastes de Hipótesis Paramétricas*. Pub. Depto Bioestadística, Universidad de Valencia.
3. Bernardo, J.M. (1981). *Bioestadística. Una Perspectiva Bayesiana*. Vicens-vives, Valencia.
4. Bernardo, J.M. & Smith A. F. M. (1994) *Bayesian Theory*, John Wiley and Sons. Inc. New York.
5. Birks, D. (1992). *Alternative Methods of Regression*, John Wiley and Sons. Inc. New York.
6. Box, G.E.P. & Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Adisson-Wesley, Reading, Mass.
7. Broemeling, D. L. (1985). *Bayesian Analysis of Linear Models* Marcel Dekker, Inc. New, York.
8. Dalal, S. & Hall, W. J. (1983). "Aproximating Priors by Mixtures of Natural Conjugate Priors ". *Journal Royal Statistics Society*, Vol 45. pp 278-286, S. B.
9. DeGroot, H. M. (1986). *Probability and Statistics*, 2nd. Edition., Adisson-Wesley, Massachusetts, E.U.A.
10. DeGroot, M.H. (1970). *Optimal Statistical Decisions*, McGraw-Hill, Nueva York.

11. Draper, N.R. & Smith M. (1966). *Applied Regression Analysis*. John Wiley and Sons, Inc. New York.
12. Girón, F.J. Martínez M.L. & Morcillo C. (1992). "A bayesian Justification for the Analysis of Residuals and Influence Measures". *Bayesian Statistics 4* pp. 651-660, Academic Press University of Málaga Spain.
13. Graybill, F.A. (1961). *An Introduction to Linear Statistics Models*. McGraw Hill, New York.
14. Gutiérrez E. (1991). *Contraste Bayesiano de Hipótesis Paramétricas. Tesis de Maestría U.A.C.P. y P. IIMAS UNAM, México.*
15. Jeffreys, H. (1961) *Theory and Probabilty*, 3rd. ed. Clarendon Press, Oxford.
16. Lindley, D.V. (1980). *Introduction to Probability and Statistics from a Bayesian Viewpoint*. Vol. 1, Cambridge University.
17. Montgomery, D.C. (1991). *Design and Analysis of Experiments*. John Willey and Sons. Inc. New York.
18. Montgomery, D.C. & Peck E. (1992) *Introduction to Linear Regression Analysis* , 2nd Edition, John Willey and Sons. Inc. New York.
19. Mood, A. M., Graybill F.A. & Boes D. (1974). *Introduction to the Theory of Statistics*. Ed. McGraw Hill, New York.
20. Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley and Sons. Inc. New York.
21. Ramsey, F.P. (1964) *Truth and Probability. Studies in Subjetive Probability (Kyburg and Smokler, eds.)*, 61-92. Wiley, New York.
22. Rueda R. (1992). "A Bayesian Alternative to Parametric Hypothesis Testing", *Test* , Vol 1, No.1, pp 61-67.
23. Searle, S.R. (1971). *Linear Models*. Wiley, New York.

24. Winkler, R. L. (1977). *Prior Distributions and Model Building in Regression Analysis*, in *New Developments in the Applications of Bayesian Methods*, eds. Ahmet Aykay and Carlo Baumant, North-MoMand, Amsterdam.
25. Zellner, A. (1975) "Bayesian Analysis of Regresion Error Terms". *Journal of the American Association, Theory and Methods Section*, Vol. 70, No. 349, pp. 138-144.