



**UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO**

**FACULTAD DE ESTUDIOS SUPERIORES
ACATLÁN**

ECONOMETRÍA BAYESIANA

TESIS

QUE PARA OBTENER EL TÍTULO DE

**LICENCIADO EN MATEMÁTICAS APLICADAS Y
COMPUTACIÓN**

PRESENTA

ANA LILIA ÁLVAREZ HERRERA

ASESOR: LIC. SERGIO ALEJANDRO MATÍAS HERNÁNDEZ

MARZO, 2010



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Agradecimientos

Este trabajo representa la culminación de un sueño y el comienzo de nuevos retos que pretenden enriquecer mi vida en todos los sentidos. Encapsula una maravillosa experiencia universitaria por lo cual deseo agradecer a todas las personas que me apoyaron e impulsaron en la culminación de esta etapa.

Agradezco a mis padres, Anita y Amando, por su valioso apoyo en todo momento y por haber confiado en la realización de este maravilloso sueño. Gracias a la vida y a Dios por permitirme compartir esta grandiosa aventura con ellos.

A mis hermanos, Liz, Mary y Quique, gracias por su cariño y por inspirarme a salir adelante en los momentos más difíciles.

De igual forma agradezco a mi asesor Alejandro Matías todo el apoyo, la paciencia y los consejos que me brindó durante toda esta etapa pero más aún por la valiosa amistad que me ha brindado y las enseñanzas durante todo este tiempo. Mil gracias por tu paciencia y por impulsarme a salir adelante en circunstancias difíciles.

A todos mis amigos por su amistad desinteresada, por todos los consejos y el apoyo que me brindaron en todo momento por impulsarme a ser una mejor persona día a día y por confiar en mí, mil gracias a todos los que fueron y han sido parte de esta maravillosa aventura que emprendimos juntos el día que ingresamos por primera vez a la universidad.

Agradezco infinitamente a todos mis profesores, por compartir sus conocimientos y experiencias y por ayudarme a crecer en todos los sentidos.

Mi más infinito agradecimiento a la UNAM por permitirme ser parte de ella.

Muchas gracias a todos por permitirme ser parte de su vida y por ayudarme a crecer día a día.

AGRADECIMIENTOS	2
INTRODUCCIÓN	8
1. PREELIMINARES DE PROBABILIDAD Y ESTADÍSTICA	12
1.1 Introducción	12
1.2 Concepto de Probabilidad	12
1.3 Definición de variable aleatoria	13
1.4 Distribuciones de Probabilidad	14
1.5 Distribución de probabilidad multivariada	15
1.6 Distribución de probabilidad condicional	18
1.7 Variables aleatorias independientes	19
1.8 Momentos de una distribución de probabilidad	20
1.8.1 Valor esperado	21
1.8.2 Varianza	22
1.8.3 Covarianza	23
1.8.4 Esperanza condicional y varianza condicional	24
1.9 Inferencia Estadística	25
1.9.1 Estimación Puntual	25
1.9.2 Estimación de Intervalos	29
1.9.3 Prueba de Hipótesis	29
2. MODELOS ECONOMETRÍCOS CLÁSICOS	32
2.1 Introducción	32
2.2 ¿Qué es Economía?	33

2.2.1 Ramas de la Economía	33
2.3 ¿Qué es Econometría?	34
2.4 Modelo de Regresión Lineal Simple	35
2.4.1 Método de mínimos cuadrados	37
2.4.2 Tabla de Análisis de Varianza	40
2.4.3 Pruebas de Hipótesis	44
2.4.4 Intervalos de confianza para β_0 , β_1 y σ^2	49
2.4.5 Intervalo de confianza para $\mu_{y x}$	50
2.4.6 Predicción de nuevas observaciones	51
2.4.7 Aplicación: El modelo CAPM (Capital Asset Pricing Model)	52
2.5 Modelo de Regresión Lineal Múltiple	57
2.5.1 Estimación del vector de parámetros B: Método de Mínimos Cuadrados Ordinarios	58
2.5.2 Tabla de Análisis de Varianza (ANOVA)	61
2.5.3 Intervalo de confianza para los parámetros del MRLM	63
2.5.4 Prueba de hipótesis acerca de β	65
2.5.5 Predicción de nuevas observaciones	69
2.5.6 Aplicación: Estimación del modelo CAPM en dos etapas	69
2.6 Modelo de Regresión Logística	73
2.6.1 Estimación de los parámetros β_0 y β_1	75
2.6.2 Prueba de Hipótesis	77
2.6.3 Estimación de intervalos de confianza	79
2.7 Aplicación: Quiebra de las compañías noruegas	80
3. PARADIGMA BAYESIANO	83
3.1 Introducción	83
3.2 Crítica a los Modelos Clásicos	84

3.3 Filosofía de la Ciencia	85
3.4 Paradigma Bayesiano	87
3.4.1 Revisión del Teorema de Bayes	89
3.4.2 Ventajas y Desventajas del Paradigma Bayesiano	92
3.5 Distribución a priori	93
3.5.1 Distribución a priori basada en datos	93
3.5.2 Aproximación del histograma	94
3.5.3 Aproximación de la Verosimilitud relativa	94
3.5.4 Elección de una forma funcional dada	95
3.5.5 Distribución a priori conjugada	95
3.6 Inferencia Estadística	96
3.6.1 Estimación puntual	97
3.6.2 Estimación por intervalos	98
3.6.3 Prueba de hipótesis	100
3.6.4 Distribución predictiva	102
4. MODELOS ECONOMETRICOS BAYESIANOS	104
4.1 Modelo de Regresión Lineal Simple	105
4.1.1 Análisis con una distribución a priori no informativa	107
4.1.2 Análisis con una distribución a priori informativa	113
4.2 Modelo de Regresión Lineal Múltiple	118
4.2.1 Análisis con una distribución a priori no informativa	120
4.2.2 Análisis con una distribución a priori informativa	125
4.3 Modelo de Regresión Logística	127
4.3.1 Análisis con una distribución a priori no informativa	128
4.3.2 Análisis con una distribución a priori informativa	129
4.4 Radios de probabilidad posterior	130

4.5 Observaciones finales	131
4.6 Aplicaciones	132
4.6.1 Índice de precios para servicios de radio	132
5. ESTUDIO COMPARATIVO: MODELOS ECONOMETRICOS CLÁSICOS VS MODELOS ECONOMETRICOS BAYESIANOS	135
5.1 Introducción	135
5.2 Estudio comparativo: valuación de bienes raíces	136
5.2.1 El modelo	137
5.2.2 Distribución a priori	138
5.2.3 Resultados	141
5.3 Importancia del tamaño muestral	143
5.4 Modelos Econométricos Clásicos vs Modelos Econométricos Bayesianos	144
CONCLUSIONES	147
APÉNDICE 1: INVERSOS GENERALIZADOS	149
GLOSARIO	151
BIBLIOGRAFÍA	154

INTRODUCCIÓN

Dentro del ámbito económico, la Estadística ha sido de gran utilidad para analizar, interpretar y predecir variables económicas por medio del empleo de modelos matemáticos. Por esta razón resulta importante analizar las ventajas que nos ofrece cada uno de los enfoques de la Estadística para determinar bajo qué circunstancias resulta conveniente el empleo de cada una de ellas.

Existen diferentes enfoques de la Estadística, dentro de los cuales encontramos algunos como son: el clásico y el bayesiano. La diferencia principal entre ellos es la definición de probabilidad, que es la base de su metodología. Mientras que para el clásico, la probabilidad se define como el límite de su frecuencia relativa, para el Bayesiano, es un grado de certeza o creencia, es decir utiliza la poca o mucha información que se tenga. Como se puede apreciar, este concepto le da un sentido distinto a la metodología que cada enfoque maneja.

El enfoque más comúnmente utilizado por los econométricos es el clásico, que ofrece una amplia gama de métodos para pronosticar, por ejemplo: el modelo de regresión lineal simple, el modelo de regresión lineal múltiple, series de tiempo, entre otros. Estos

métodos están sustentados en supuestos, tales como: normalidad, independencia, muestras grandes y con ello el uso de información histórica; desafortunadamente en algunas ocasiones estos supuestos no se cumplen y más aún en el área de las Finanzas, donde no es conveniente utilizar datos históricos puesto que esto implica que las condiciones en las que se dieron esos datos son similares a las que se tienen en el momento en que se hace el análisis, y en la mayoría de los casos esto no sucede. Esto nos conlleva a formular las siguientes cuestiones:

- ¿Qué pasa cuando no tengo información histórica, o cuando el tamaño muestral es pequeño?
- ¿Cómo afecta esto a los parámetros?
- ¿Qué tan eficientes son los pronósticos que se realizan con base en los modelos econométricos?

Hoy en día estos modelos econométricos han sido aplicados a diferentes problemas financieros y económicos, en algunos de ellos se han obtenido resultados aceptables, pero en otros se ha abusado de los supuestos, sin tomar en cuenta las consecuencias que esto conlleva.

El segundo enfoque es el bayesiano, que surge como una alternativa a la Econometría Clásica, esto no quiere decir que es mejor que el enfoque anterior, pero nos ofrece una perspectiva diferente de los modelos que resulta ser más consistente en el análisis del problema. Al igual que el enfoque clásico, tiene algunos problemas, tales como: la elección de una distribución a priori y los cálculos que se tienen que realizar, puesto que se requiere evaluar integrales en varias dimensiones que no se pueden resolver analíticamente. Esto provocó que durante varios años, los econométricos adoptaran lentamente esta metodología, pero afortunadamente, en la actualidad, existen métodos que nos ayudan a resolver numéricamente integrales multidimensionales con un alto grado de exactitud, por ejemplo el método de integración numérica Monte Carlo, y además existen algunos paquetes estadísticos diseñados para realizar análisis bayesiano.

Cabe señalar que este enfoque nos permite incorporar el conocimiento que se tiene acerca de los parámetros, información que se obtiene a veces de manera subjetiva u

objetiva; a ésta se le llama información a priori de los parámetros disponible, por medio de la distribución a priori. En algunos modelos econométricos a menudo, se tiene información a priori acerca de los parámetros, por lo que la adopción de la Econometría Bayesiana, resultaría ser el camino más viable que nos permita incorporar dicha información directamente al modelo. Aunque, existirán casos en los que no se disponga de información a priori, pero para ello, existen distribuciones a priori no informativas que pueden ser empleadas, algunas de las cuales, conducen a resultados similares o iguales a los de la Econometría Clásica, pero su interpretación es diferente. Esto le da al econometrista la libertad de elegir incorporar o no la información a priori de los parámetros, en caso de que la tenga.

La actualización de la información a priori acerca de los parámetros, se realiza por medio del Teorema de Bayes, que es la pieza fundamental del análisis bayesiano, y que combina la información contenida en los datos y la información a priori para dar origen a la distribución a posteriori, que nos permite realizar inferencias acerca de los parámetros.

Al igual que en la Econometría Clásica, el enfoque bayesiano hace uso de algunos de los supuestos, pero estrictamente hablando del tamaño muestral, aquí se puede trabajar con pocos datos, en cuyo caso la distribución a priori tendrá mayor peso en el análisis, mientras que cuando se dispone de muestras grandes, los datos son los que tienen mayor peso.

Por lo que se ha mencionado acerca de la Econometría Bayesiana, resulta interesante analizar la metodología que maneja y realizar una comparación con el enfoque clásico, manejando diferentes escenarios para poder determinar cuando resulta conveniente utilizar cada una de ellas.

En el presente trabajo se efectuará una presentación de la metodología bayesiana aplicada al análisis de regresión. En el capítulo uno se presentan algunos conceptos de Probabilidad y Estadística Clásica que resultarán útiles en los capítulos posteriores. En el capítulo dos se analizan desde el punto de vista Clásico los modelos de regresión lineal simple, múltiple y logístico, cada uno con un respectivo ejemplo. En el capítulo tres se introducirá al paradigma bayesiano presentando algunos conceptos básicos y la forma en la que se efectúa la estimación puntual, de intervalos y las pruebas de hipótesis bajo este

enfoque para visualizar las principales diferencias que existen con respecto al enfoque clásico. En el capítulo cuatro, se analizarán los modelos de regresión lineal simple, múltiple y logístico utilizando la metodología bayesiana. Finalmente en el capítulo cinco se efectuará un análisis comparativo de ambas metodologías para determinar las ventajas y desventajas que ofrece cada enfoque así como determinar cuándo es conveniente utilizar cada una de ellas.

1. Preeliminares de Probabilidad y Estadística

1.1 Introducción

En este capítulo, se presentarán los conceptos más importantes de Probabilidad y Estadística que se utilizarán en capítulos sucesivos. Se comenzará por definir el término probabilidad y variable aleatoria, para posteriormente analizar las distribuciones de probabilidad, momentos de una distribución, entre otros y se concluirá con inferencia estadística en donde se abordarán el problema de estimación y el procedimiento de prueba de hipótesis.

1.2 Concepto de Probabilidad

La probabilidad es una rama de las Matemáticas que nos permite cuantificar numéricamente la incertidumbre asociada con algún suceso, es decir, a un evento A , se le

asocia $P(A)$ que es la probabilidad de que ocurra A . Entendiendo por evento, un conjunto de algunos de los posibles resultados del experimento. Mientras que al conjunto que contempla todos los posibles resultados del experimento, se le denomina Espacio Muestral (S).

Por ejemplo, en el lanzamiento de una moneda, el espacio muestral S es: $S = \{\text{Águila}, \text{Sol}\}$, y los eventos son: $E_1 = \{\text{Águila}\}$ y $E_2 = \{\text{Sol}\}$.

Cabe mencionar que, para que $P(A)$ sea una medida de probabilidad consistente, debe cumplir las siguientes propiedades:

1. Es un número no negativo, es decir: $P(A) \geq 0$
2. $P(S) = 1$, es decir, la probabilidad máxima es 1 y sólo la tiene el espacio muestral, puesto que todos los eventos están contenidos en él.
3. La probabilidad de la unión de una secuencia de eventos mutuamente excluyentes es la suma de sus probabilidades, es decir, si $A_i \in S$ (donde, $i = 1, \dots$) y $A_i \cap A_j = 0$, para $i \neq j$, entonces:

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = \sum_{i=0}^{\infty} P(A_i)$$

1.3 Definición de variable aleatoria

Retomando el ejemplo de la moneda, supongamos que realizamos diez lanzamientos y estamos interesados en conocer, el número de soles y de águilas que se observaron, para ello definimos:

X = Número de águilas

Y = Número de soles

Si se obtuvieron seis águilas y cuatro soles, entonces: $x = 6$ y $y = 4$.

Se observa que, “ X ” y “ Y ”, se definieron en función de los elementos del espacio muestral S y de acuerdo con esto, es posible asignarles valores numéricos dependiendo de los resultados totales que arroje el experimento. Por lo que se puede decir

que, una variable aleatoria, es una regla que asocia un número a cada elemento del espacio muestral del experimento.

En este ejemplo, “ X ” y “ Y ” son variables aleatorias, y la regla de asociación para X , es el número de águilas, mientras que para Y es el número de soles.

Existen dos tipos de variables aleatorias, los cuales son:

- **Discretas:** Son aquellas variables cuyo conjunto de valores que pueden tomar, es finito o infinito numerable, es decir, los elementos del conjunto se pueden numerar. En este caso, la variable aleatoria sólo toma valores enteros. Por ejemplo, el número de soles observados en diez lanzamientos de una moneda, el número de hijos en una familia determinada.
- **Continuas:** Son aquellas variables cuyo conjunto de valores que pueden tomar, es un intervalo de números reales, es decir, para $a < b$, X puede tomar cualquier valor entre a y b . El conjunto de valores, es infinito no numerable, por ejemplo, la esperanza de vida de una población (72.5 años, 75.45 años, 72, 73.423 años, o en forma resumida $X \in (72,76)$), la temperatura ambiente, la estatura media de un grupo de personas, la tasa de interés, entre otras.

1.4 Distribuciones de Probabilidad

La distribución de probabilidad de una variable aleatoria, $f(x)$, es una función que nos indica cómo se distribuye la probabilidad total de uno entre todos los valores que puede tomar la variable aleatoria, es decir, es una función que especifica la probabilidad de ocurrencia de cada uno de los posibles valores que puede tomar la variable aleatoria.

De acuerdo con el tipo de variable aleatoria, las distribuciones de probabilidad se pueden clasificar en:

- **Distribución de masa de probabilidad:** Es la distribución de probabilidad de una variable aleatoria discreta, que está definida para toda X como: $p(x) = P(X = x)$. Por ejemplo, en el caso del lanzamiento de una moneda,

si definimos a $X = 0$, cuando se obtiene águila y $X = 1$ cuando se obtiene sol, su función de distribución de probabilidad de masa es:

$$p(x) = \begin{cases} \frac{1}{2} & x = 0 \\ \frac{1}{2} & x = 1 \end{cases}$$

- **Función de densidad de probabilidad:** Es la distribución de probabilidad de una variable aleatoria continua, y se define de la siguiente forma:

$$p(a < x < b) = \int_a^b f(x)dx$$

Ambas cumplen con las siguientes propiedades:

1. $f(x) \geq 0$; $p(x) \geq 0$
2. $\sum_x p(x) = 1$ ó $\int_{-\infty}^{\infty} f(x)dx = 1$

Cabe señalar que, en el caso de la función de densidad de probabilidad, la probabilidad de que la variable aleatoria tome un valor específico ($X = x$) es cero, debido a que el conjunto de los posibles resultados de la variable aleatoria es incontable e infinito.

1.5 Distribución de probabilidad multivariada

En algunas ocasiones, resulta útil y necesario conocer simultáneamente el comportamiento de dos o más variables aleatorias, en estos casos, se emplea la función de probabilidad conjunta. Primero se tratará el caso de dos variables aleatorias y posteriormente, se extenderá la definición al caso de n variables aleatorias.

La función de probabilidad conjunta para dos variables aleatoria X y Y se define de la siguiente forma:

- En el caso discreto:

$$P(X = x, Y = y) = p(x, y) = \sum \sum p(x, y)$$

- En el caso continuo:

$$P(x_1 < X < x_2, y_1 < Y < y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x, y) dy dx$$

Cabe señalar que para que las funciones $p(x,y)$ y $f(x,y)$ puedan ser empleadas como funciones de distribución conjunta deben cumplir con las siguientes propiedades:

1. $p(x,y) \geq 0$
 $f(x,y) \geq 0$
2. $\sum_x \sum_y p(x, y) = 1$ Caso Discreto

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx = 1 \text{ Caso continuo}$$

Una vez que se conoce la función de probabilidad conjunta, a partir de ella, se pueden calcular las distribuciones de probabilidad que describen el comportamiento individual de cada una de las variables aleatorias involucradas, las cuales se conocen como distribuciones de probabilidad marginal.

En el caso bivariado, las distribuciones de probabilidad marginal, se obtienen de la siguiente manera:

- Distribución de probabilidad marginal de X .

$$p_X(x) = \sum_y p(x, y) \text{ Caso discreto}$$

$$f_Y(x) = \int_{-\infty}^{\infty} f(x, y) dy \text{ Caso continuo}$$

- Distribución de probabilidad marginal de Y .

$$p_Y(y) = \sum_x p(x, y) \quad \text{Caso discreto}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad \text{Caso continuo}$$

Como se puede observar, la función de distribución de probabilidad para una sola de las variables se obtiene al sumar $p(x,y)$ sobre todos los valores de la otra variable, en el caso discreto, mientras que en el caso continuo, se obtiene al integrar con respecto a la otra variable.

Para analizar el comportamiento de n variables aleatorias, se procede en forma similar al caso bivariado, es decir, si tenemos X_1, X_2, \dots, X_n variables aleatorias, entonces su respectiva función de probabilidad conjunta se define de la siguiente forma:

- En el caso discreto:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = p(x_1, x_2, \dots, x_n)$$

- En el caso continuo:

$$P(a_1 < x_1 < b_1, a_2 < x_2 < b_2, \dots, a_n < x_n < b_n) =$$

$$= \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f(x_1, x_2, \dots, x_n) dx_n \dots dx_1$$

Al igual que en el caso bivariado, la función de probabilidad conjunta, cumple con las siguientes propiedades:

1. $p(x_1, x_2, \dots, x_n) \geq 0$

$$f(x_1, x_2, \dots, x_n) \geq 0$$

2. $\sum_{x_1} \dots \sum_{x_n} p(x_1, x_2, \dots, x_n) = 1 \quad \text{Caso Discreto}$

$$\int_{-\infty_1}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_n \dots dx_1 = 1 \quad \text{Caso continuo}$$

Y sus respectivas distribuciones marginales, se obtienen en forma similar al caso bivariado. En el caso discreto, se suma $p(x_1, x_2, \dots, x_n)$ sobre todos los valores de las otras variables, por ejemplo;

$$p_{X_1}(x_1) = \sum_{x_2} \sum_{x_3} \dots \sum_{x_n} p(x_1, x_2, \dots, x_n)$$

Mientras que en el caso continuo, se integra $f(x_1, x_2, \dots, x_n)$ con respecto a las otras variables, por ejemplo:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_n dx_{n-1} \dots dx_2$$

1.6 Distribución de probabilidad condicional

En el pronóstico del tiempo, es común escuchar enunciados como: “Si el frente frío número x toca costas mexicanas, entonces, la probabilidad de que llueva esta tarde es alta, en caso contrario la probabilidad de lluvia es muy baja”. Al igual que este caso, existen otros similares, en los cuales, la ocurrencia de un suceso depende de la ocurrencia de otro. Para cuantificar la incertidumbre asociada a este tipo de situaciones, se utiliza la distribución de probabilidad condicional.

La distribución de probabilidad condicional, se puede calcular fácilmente utilizando la función de distribución de probabilidad conjunta y la función de distribución de probabilidad marginal de la variable aleatoria condicional. Para el caso bivariado, se define de la siguiente forma:

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} \quad \text{siempre que } f_Y(y) > 0$$

Al extender esta definición al caso multivariado se obtiene:

Sean las siguientes variables aleatorias: X_1, X_2, \dots, X_n con función de probabilidad conjunta $f(x_1, x_2, \dots, x_n)$. Si se toma el subconjunto de variables aleatorias: $X_1, X_2, \dots, X_m (1 \leq m < n)$ con función de probabilidad marginal:

$f(x_1, x_2, \dots, x_m)$. Entonces la distribución de probabilidad condicional de X_1, X_2, \dots, X_n dado X_1, X_2, \dots, X_m es:

$$f(x_1, x_2, \dots, x_n | x_1, x_2, \dots, x_m) = \frac{f(x_1, x_2, \dots, x_n)}{f(x_1, x_2, \dots, x_m)}$$

siempre que $f(x_1, x_2, \dots, x_m) > 0$

1.7 Variables aleatorias independientes

Un concepto estadístico importante que puede simplificar el trabajo en el análisis multivariado es el de independencia estocástica. Dos o más variables aleatorias son independientes estadísticamente o están distribuidas independientemente, si el conocer el valor que toma una de ellas no influye en el valor que pueden tomar las demás. Por ejemplo, en el lanzamiento de dos dados, tenemos que X_1 es el número observado en el primer dado, mientras que X_2 es el número observado en el segundo dado, el hecho de que en el primer dado se observe un “dos”, no nos dice absolutamente nada acerca del valor observado en el segundo dado, por lo que X_1 y X_2 son independientes.

Formalmente para el caso bivariado, la independencia estocástica se define de la siguiente forma:

Sean X_1 y X_2 dos variables aleatorias con función de distribución conjunta $f(x_1, x_2)$ y cuyas funciones de distribución marginal son $f_{X_1}(x_1)$ y $f_{X_2}(x_2)$. Entonces, X_1 y X_2 son independientes si y sólo si $f(x_1, x_2) = f(x_1)f(x_2)$ para todos los valores de X_1 y X_2 .

Como consecuencia de esto, tenemos que dos variables aleatorias son independientes si y sólo si $f(x_1 | x_2) = f(x_1)$ o si y sólo si $f(x_2 | x_1) = f(x_2)$ para todos los valores de X_1 y X_2 .

En caso de que para algún valor de X_1 o de X_2 no se cumpla lo anterior, entonces X_1 y X_2 no son independientes.

Análogamente, para el caso multivariado la independencia estocástica se define de la siguiente manera:

Sean las variables aleatorias X_1, X_2, \dots, X_n con función de distribución conjunta $f(x_1, x_2, \dots, x_n)$ y dos subconjuntos de ellas como $Z = \{X_1, X_2, \dots, X_m\}$ donde $1 \leq m < n$ y $W = \{X_{m+1}, X_{m+2}, \dots, X_n\}$ con funciones de probabilidad marginal $f(x_1, x_2, \dots, x_m)$ y $f(x_{m+1}, x_{m+2}, \dots, x_n)$. Entonces Z y W son independientes si y sólo si $f(x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_m)f(x_{m+1}, x_{m+2}, \dots, x_n)$ para todos los valores de Z y W .

1.8 Momentos de una distribución de probabilidad

Los momentos de una distribución de probabilidad, resumen algunas de las características de dicha distribución. Existen dos tipos de momentos: los momentos centrales (con respecto a la media) y los momentos con respecto al origen.

Sea X una variable aleatoria con función de densidad o de masa, $f(x)$, y r un entero no negativo, entonces para el caso univariado, se tiene:

- El r -ésimo momento central para el caso discreto está definido como

$$M_r = \sum_x (x - \mu)^r f(x) \quad \text{y} \quad \text{para} \quad \text{el} \quad \text{caso} \quad \text{continuo}$$

$$M_r = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx, \text{ donde } \mu \text{ es la media de la distribución.}$$

- El r -ésimo momento con respecto al origen para el caso discreto está definido como $M'_r = \sum_x x^r f(x)$ y para el caso continuo

$$M'_r = \int_{-\infty}^{\infty} x^r f(x) dx.$$

Los momentos más utilizados son la media que es el primer momento con respecto al origen y la varianza que es el segundo momento central.

1.8.1 Valor esperado

El valor esperado para una variable aleatoria X con función de densidad o de masa $f(x)$ se define como:

$$E(x) = \sum_x xf(x) \quad \text{para el caso discreto}$$

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx \quad \text{para el caso continuo}$$

Las propiedades del valor esperado son las siguientes:

1. El valor esperado de una constante es la constante misma. Sea a una constante, entonces $E(a) = a$
2. Sea a una constante y X una variable aleatoria, entonces:
 $E(aX) = aE(X)$
3. Sean X y Y variables aleatorias entonces $E(X + Y) = E(X) + E(Y)$
4. Si X y Y son variables aleatorias independientes, entonces:
 $E(XY) = E(X)E(Y)$
5. Si X es una variable aleatoria con función de densidad o de masa $f(x)$ y si $g(x)$ es una función de X , entonces:

$$E[g(x)] = \begin{cases} \sum_x g(x)f(x) & \text{para el caso discreto} \\ \int_{-\infty}^{\infty} g(x)f(x)dx & \text{para el caso continuo} \end{cases}$$

6. Sean las variables aleatorias X_1, X_2, \dots, X_n y las constantes a_1, a_2, \dots, a_n , entonces:

$$E(a_1X_1 + a_2X_2 + \dots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \dots + a_nE(X_n)$$

es decir, la esperanza de una función lineal de variables aleatorias es igual a la función lineal de sus esperanzas.

1.8.2 Varianza

La varianza es una medida de dispersión de los valores de X alrededor de su media, se define de la siguiente manera:

Sea X una variable aleatoria con $E(X) = \mu$, entonces:
 $\text{var}(X) = \sigma^2 = E(X - \mu)^2$ y se calcula:

$$\sigma^2 = \begin{cases} \sum_x (x - \mu)^2 f(x) & \text{para el caso discreto} \\ \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx & \text{para el caso continuo} \end{cases}$$

Para facilitar su cálculo, se puede expresar, en términos de esperanzas:
 $\sigma^2 = E(X^2) - [E(X)]^2$.

La raíz de la varianza se conoce como desviación estándar.

Las propiedades de la varianza son:

1. La varianza de una constante es cero. Sea a una constante con $E(a) = a$, entonces: $\text{var}(a) = E[a - E(a)]^2 = 0$

2. Si a es una constante y X es una variable aleatoria, entonces: $\text{var}(aX) = a^2 \text{var}(X)$.

3. Si X y Y son variables aleatorias, entonces:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y)$$

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2 \text{cov}(X, Y)$$

4. Si X y Y son variables aleatorias independientes, entonces:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$$

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y)$$

1.8.3 Covarianza

Es una medida que cuantifica numéricamente la relación que existe entre dos variables. Se define de la siguiente forma:

Sean X y Y dos variables aleatorias, con sus respectivas medias μ_X y μ_Y , entonces la covarianza entre estas dos variables es:

$$\text{Cov}(X, Y) = E\{(X - \mu_X)(Y - \mu_Y)\} = E(XY) - \mu_X \mu_Y$$

Cabe señalar que, si X y Y son variables aleatorias independientes, entonces su covarianza es cero:

$$\text{Cov}(X, Y) = E\{(X - \mu_X)(Y - \mu_Y)\} = E(XY) - \mu_X \mu_Y = E(X)E(Y) - \mu_X \mu_Y = 0$$

Pero no se cumple lo contrario, es decir, si la covarianza de dos variables aleatorias es cero, esto no implica que sean independientes.

1.8.4 Esperanza condicional y varianza condicional

En el caso de que se tengan dos variables aleatorias, se puede estar interesado en conocer la esperanza de una de ellas dado algún valor específico de la otra. Para resolver este tipo de problema es necesario utilizar la esperanza condicional. También puede suceder una situación similar a la anterior para la varianza, por lo que a continuación se presenta la definición de la esperanza y de varianza condicional.

- **Esperanza condicional:** Sean X y Y dos variables aleatorias con función de distribución conjunta $f(x, y)$, entonces la esperanza condicional de X dado $Y = y$ se define como:

$$E(X | Y = y) = \sum_x xf(x | Y = y) \quad \text{Caso discreto}$$

$$= \int_{-\infty}^{\infty} xf(x | Y = y)dx \quad \text{Caso continuo}$$

La esperanza condicional de Y , se define en forma similar.

- **Varianza condicional:** Sean X y Y dos variables aleatorias con función de distribución conjunta $f(x, y)$, entonces la varianza condicional de X dado $Y = y$ se define como:

$$\text{var}(X | Y = y) = E\{[X - E(X | Y = y)]^2 | Y = y\}$$

$$= \sum_x [X - E(X | Y = y)]^2 f(x | Y = y) \quad \text{Caso discreto}$$

$$= \int_{-\infty}^{\infty} [X - E(X | Y = y)]^2 f(x | Y = y)dx \quad \text{Caso continuo}$$

La varianza condicional de Y , se define en forma similar.

1.9 Inferencia Estadística

Esta parte de la Estadística se ocupa de obtener conclusiones acerca de una o más características de una población utilizando una muestra (un subconjunto de la población). Es de gran utilidad en diferentes áreas puesto que en la mayoría de los casos no se cuenta con toda la información poblacional para realizar un estudio, es entonces cuando se emplea la información muestral disponible para hacer inferencias acerca de la población.

La inferencia estadística se concentra en dos grandes áreas: la estimación y los procedimientos de prueba de hipótesis, los cuales serán tratados en las siguientes secciones. Cabe señalar que, el problema de estimación a su vez se divide en dos, que son la estimación puntual y la de intervalos.

Antes de pasar a la siguiente sección, es necesario presentar las siguientes definiciones que serán utilizadas más adelante:

1. **Función de Verosimilitud:** La función de verosimilitud de una muestra aleatoria proveniente de la densidad $f(x; \theta)$ es la densidad conjunta de las variables muestrales, la cual también depende del parámetro θ , es decir:

$$g(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdots f(x_n; \theta)$$

2. **Estadístico:** Es una función de las variables aleatorias observables y que a su vez es una variable aleatoria observable que no contiene parámetros desconocidos.
3. **Espacio paramétrico:** Contiene todos los posibles valores que puede adoptar el parámetro θ y se denota por la letra Ω .

1.9.1 Estimación Puntual

Se concentra en estimar a partir de la información muestral el (o los) valor(es) de un(os) parámetro(s) poblacional(es) desconocido(s). Formalmente, sea X_1, X_2, \dots, X_n una muestra aleatoria de la función de densidad $f(x_1, x_2, \dots, x_n; \theta)$ cuya forma es conocida pero el parámetro θ se desconoce, entonces, se trata de encontrar un estadístico $\hat{\theta}$ que

pueda ser utilizado como un estimador de θ . Entonces, se tienen las siguientes definiciones:

1. **Estimador:** Algún estadístico $\hat{\theta}$ cuyos valores pueden ser utilizados para estimar a θ . Como es una función de la información muestral, entonces, un estimador es una variable aleatoria. En lo sucesivo se denotará a un estimador por: $\hat{\theta}$ en lugar de $\hat{\theta} = u(x_1, x_2, \dots, x_n)$
2. **Estimación:** Es una realización del estimador, es decir, es el valor numérico que toma el estimador cuando se sustituye la información muestral disponible en la función.

Para tener un criterio de selección de un estimador, es necesario conocer las propiedades más importantes de los estimadores, las cuales se resumen a continuación.

A. Error Cuadrático Medio

Sea $\hat{\theta}$ un estimador de θ , entonces el error cuadrático medio del estimador $\hat{\theta}$ se define como $E[(\theta - \hat{\theta})^2]$ que se puede describir como:

$$E[(\hat{\theta} - \theta)^2] = \text{Var}(\theta) + [\theta - E(\hat{\theta})]^2$$

donde $[\theta - E(\hat{\theta})]$ se conoce como sesgo del estimador.

De manera que, un criterio para seleccionar al mejor estimador de θ , consistiría en escoger a aquel que posea el error cuadrático medio más pequeño, pero, generalmente, estos no existen, por lo que es necesario considerar otros criterios.

B. Estimadores Insesgados

Considere a $\hat{\theta}$ un estimador de θ , entonces, $\hat{\theta}$ es un estimador insesgado de θ si y sólo si la esperanza del estimador $\hat{\theta}$ es igual al parámetro θ , es decir, $E(\hat{\theta}) = \theta$ para todos los valores de $\theta \in \Omega$ (Espacio paramétrico).

C. Eficiencia

Sean dos estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ del parámetro θ , entonces si:

$$\frac{E[(\hat{\theta}_1 - \theta)^2]}{E[(\hat{\theta}_2 - \theta)^2]} < 1$$

$\hat{\theta}_1$ es relativamente más eficiente que $\hat{\theta}_2$. Si la expresión anterior en lugar de ser menor fuera mayor que 1 entonces $\hat{\theta}_2$ es relativamente más eficiente que $\hat{\theta}_1$.

D. Suficiencia

Un estadístico $\hat{\theta}$ es suficiente si contiene toda la información acerca de θ presente en la muestra X_1, X_2, \dots, X_n de la densidad $f(x_1, x_2, \dots, x_n; \theta)$. De manera formal, se puede definir a un estadístico suficiente de la siguiente manera:

Sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n de la densidad $f(x_1, x_2, \dots, x_n; \theta)$ y un estadístico $\hat{\theta}$ del parámetro θ . El estadístico $\hat{\theta}$ es un estadístico suficiente de θ si y sólo si $f(x_1, x_2, \dots, x_n | \hat{\theta})$ no depende de θ , es decir, la muestra aleatoria no facilita más información acerca del parámetro que no haya sido proporcionada por $\hat{\theta}$.

Un criterio para determinar si un estadístico es suficiente, es proporcionado por el siguiente teorema:

Teorema de Factorización: Sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n de la densidad $f(x_1, x_2, \dots, x_n; \theta)$ y cuya función de distribución conjunta está dada por:

$$g(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdots f(x_n; \theta)$$

El estadístico $\hat{\theta}$ es un estadístico suficiente si la función de distribución conjunta se puede factorizar de la siguiente forma:

$$g(x_1, x_2, \dots, x_n; \theta) = h(\hat{\theta}; \theta) \cdot k(x_1, x_2, \dots, x_n)$$

donde $k(x_1, x_2, \dots, x_n)$ es una función no negativa que no depende del parámetro θ .

E. Consistencia

Un estimador $\hat{\theta}$ del parámetro θ es consistente, si a medida que aumente el tamaño muestral, el estimador $\hat{\theta}$ se acerca más al verdadero valor del parámetro θ . A continuación se definen la consistencia en error cuadrático medio y la consistencia simple.

- **Estimador consistente en ECM:** Sea $\hat{\theta}_n$ un estimador del parámetro θ basado en una muestra de tamaño n , entonces, $\hat{\theta}_n$ es un estimador consistente en error cuadrático medio si se cumple:

$$\lim_{n \rightarrow \infty} E[(\hat{\theta}_n - \theta)^2] = 0 \quad \forall \theta \in \Omega$$

Es decir, el ECM tiende a cero conforme el tamaño muestral n aumenta indefinidamente y considerando que el ECM es igual a la suma de la varianza de θ más el cuadrado del sesgo, entonces ambos tienden a cero conforme el tamaño muestral n aumente indefinidamente.

- **Consistencia simple:** Sea $\hat{\theta}_n$ un estimador del parámetro θ basado en una muestra de tamaño n , entonces, $\hat{\theta}_n$ es un estimador consistente simple, si:

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1 \quad \forall \varepsilon > 0$$

Es decir, la probabilidad de que $\hat{\theta}_n$ se encuentre alejado del verdadero valor del parámetro θ , por una cantidad inferior a ε , se aproxima a uno conforme el tamaño muestral aumenta indefinidamente.

F. Teorema de Rao – Blackwell

Este teorema permite encontrar al mejor estimador insesgado de mínima varianza utilizando un estadístico suficiente y un estimador insesgado.

Teorema: Sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n de la densidad $f(x_1, x_2, \dots, x_n; \theta)$, $\hat{\theta}$ un estadístico suficiente de θ y T un estimador insesgado de $g(\theta)$ que

no es función de $\hat{\theta}$. Además sea $U = E(T|\theta)$, entonces U cumple con las siguientes propiedades:

1. U es un estadístico y es función del estadístico suficiente $\hat{\theta}$
2. U es un estimador insesgado de $g(\theta)$, es decir, $E[U]=g(\theta)$
3. $Var(U) \leq Var(T)$

1.9.2 Estimación de Intervalos

Se refiere al procedimiento de construcción de un intervalo de confianza ($U \leq \theta \leq V$), que consiste en un rango de valores que contenga al parámetro θ con una probabilidad de $1 - \alpha$.

A partir de la muestra aleatoria X_1, X_2, \dots, X_n de tamaño n , se calculan los límites de confianza superior V e inferior U , ambos son estadísticos que cumplen $U \leq V$. Se está interesado en obtener un intervalo que satisfaga:

$$P(U \leq \theta \leq V) = 1 - \alpha \quad 0 < \alpha < 1 \quad \dots \mathbf{1}$$

Donde $1 - \alpha$ es el coeficiente de confianza y al intervalo $U \leq \theta \leq V$ se le conoce como intervalo de confianza al $(1 - \alpha)100\%$.

La expresión $\mathbf{1}$ indica que el intervalo (U, V) contendrá al parámetro θ con una probabilidad de $1 - \alpha$, su interpretación es la siguiente: si se obtienen muestras repetidas del mismo tamaño y de la misma población y si se construye un intervalo con un coeficiente de confianza $1 - \alpha$ para cada muestra, entonces, el $(1 - \alpha)100\%$ de estos intervalos contendrá al parámetro θ .

1.9.3 Prueba de Hipótesis

Una hipótesis estadística es un enunciado o conjetura acerca de la función de distribución de una o más variables aleatorias o acerca de la población de la que se obtuvo la muestra en cuestión. La hipótesis es compuesta si el enunciado no especifica completamente los parámetros de la distribución ($H_0: \sigma \geq 3.5$), en caso contrario, la hipótesis es simple ($H_0: \sigma = 3.5$).

El procedimiento de prueba de hipótesis consiste en determinar la validez de la hipótesis nula H_0 frente a la hipótesis alternativa H_1 utilizando la información muestral mediante el empleo del estadístico de prueba, para ello es necesario, determinar la región de rechazo que es el conjunto de valores que puede tomar el estadístico de prueba para los cuales H_0 debe ser rechazada, las fronteras de esta región se conocen como valores críticos del estadístico de prueba.

En este procedimiento se puede incurrir en los siguientes errores:

1. **Error de tipo I:** Consiste en rechazar la hipótesis nula H_0 cuando es verdadera. La probabilidad de cometer un error de tipo I se denota por α y se denomina nivel de significancia de la prueba.
2. **Error de tipo II:** Consiste en no rechazar la hipótesis nula H_0 cuando es falsa. La probabilidad de cometer un error de tipo I se denota por β . La probabilidad de no cometer un error de tipo II, es decir $1 - \beta$, se denomina potencia de la prueba.

En el enfoque clásico, se fija la probabilidad de cometer un error de tipo I, α , a un nivel bajo, a menudo al 0.05, 0.01 ó 0.1, y posteriormente se trata de maximizar la potencia de la prueba, de esta manera se intenta minimizar ambos errores.

Existen diferentes tipos de prueba de hipótesis, cada una de las cuales tiene un objetivo específico y cuya manera de proceder se puede resumir en los siguientes pasos:

1. La prueba seleccionada debe concordar con el objetivo de nuestra investigación. Se deben revisar los requerimientos o supuestos de la prueba referentes a la muestra de datos o a la variable aleatoria de interés como: la independencia, la escala de medición, entre otras.
2. Postular la hipótesis nula H_0 y la alternativa H_1 .
3. Especificar el nivel de significancia α .
4. Especificar el estadístico de prueba y su distribución bajo la hipótesis nula H_0 .
5. Determinar la regla de decisión, es decir, se debe especificar la región de rechazo.

6. Calcular el valor del estadístico de prueba utilizando la muestra de datos disponible.
7. Determinar si existe evidencia estadística con base en la muestra para rechazar o no la hipótesis nula H_0 .

Esta fue sólo una breve presentación de los conceptos básicos relacionados con los procedimientos de prueba de hipótesis, si desea profundizar en el tema y conocer los diferentes tipos de pruebas que existen, le sugiero que consulte Rodríguez [2005].

2. Modelos Econométricos Clásicos

2.1 Introducción

La Econometría nos ofrece una metodología para analizar los fenómenos económicos a partir de unos supuestos, ya sea para pronosticar, para probar una teoría económica o ambas. Uno de los enfoques más ampliamente utilizados en este ámbito ha sido la metodología clásica, la cual se abordará en el presente capítulo.

Los modelos econométricos clásicos que se analizarán serán los siguientes: el modelo de regresión simple, el modelo de regresión múltiple y el modelo de regresión logístico, para cada uno de ellos se presentarán los supuestos correspondientes, así como el método de estimación de los parámetros y pruebas de hipótesis e intervalos de confianza con sus respectivas interpretaciones.

Un aspecto importante que no se debe perder de vista es la importancia de que se cumplan los supuestos que hay detrás de cada modelo, ya que esto nos garantiza de alguna

forma la validez de los resultados obtenidos, en caso de que no se cumpla alguno de los supuestos, se tiene que emplear alguna herramienta alternativa.

Como introducción se presentará la definición de Economía, Microeconomía, Macroeconomía y Econometría.

2.2 ¿Qué es Economía?

La Economía se encarga de estudiar la asignación de los recursos escasos entre diferentes actividades con el objetivo de obtener bienes y servicios que son distribuidos entre los individuos para la satisfacción de sus necesidades.

Entre los diferentes problemas que aborda se encuentran: los procesos de producción, distribución, comercialización y consumo de bienes y servicios. De manera que brinda un panorama general para la toma de decisiones a nivel personal, por ejemplo, invertir o no en acciones de una empresa, o a nivel nacional como puede ser, las medidas adoptadas para enfrentar una crisis económica. Además nos permite conocer la situación económica de un país así como el funcionamiento de las políticas económicas adoptadas.

En la carrera de MAC, Economía se imparte como una asignatura más y no se le da la importancia que se debe, esto ocasiona que el alumno pierda el interés en el área y no vea en ella la oportunidad de desarrollar la habilidad para modelar problemas económicos y resolverlos utilizando conocimientos adquiridos en otras áreas, e incluso puede rescatar algunos de ellos como ejemplos en otras asignaturas y de esta forma darle un giro distinto a otras clases.

2.2.1 Ramas de la Economía

Para lograr su objetivo, se divide en dos grandes áreas que son: la microeconomía y la macroeconomía, las cuales serán definidas a continuación:

Microeconomía: Es la parte de la economía que se encarga de analizar el comportamiento de las unidades productivas y del consumidor individual, es decir, se encarga de estudiar comportamiento dentro del mercado de los pequeños grupos tales como

asociaciones, la familia, las empresas y la industria. Estudia los factores que determinan qué y cuánto producir así como el precio del producto, en el caso de las empresas, mientras en el caso de las familias se enfoca a las interrogantes de qué y cuánto comprar. Los aspectos que trata son:

- La teoría del consumo,
- Teoría de producción y costos
- Teoría de los precios
- El mercado y los mecanismos de formación de precios
- Teoría de la demanda

Macroeconomía: Es la parte de la economía que se encarga de analizar el funcionamiento de la economía en su conjunto. Nos permite conocer el desarrollo económico de un país para poder tomar decisiones adecuadas, es decir, nos indica la situación socio – económica actual. Entre los aspectos que abarca se encuentran: el producto nacional total, la inflación, devaluación, crisis económicas, inversión, el empleo, entre otros a nivel nacional.

En lugar de enfocarse al estudio del comportamiento de las unidades económicas dirige su atención a la suma de las decisiones de dichas unidades, es decir, estudia el comportamiento económico de la totalidad de las empresas y las familias en conjunto.

La econometría analiza tanto modelos microeconómicos como macroeconómicos, utilizando la información histórica disponible, este análisis puede servir para conocer cómo se comportan las variables, para simular políticas de control, es decir, para manipular las variables y ver cómo se comportan con base en los cambios que se realizan o tal vez para realizar pronósticos y a partir de esto tomar decisiones.

2.3 ¿Qué es Econometría?

La Econometría se encarga de analizar, interpretar y pronosticar cuantitativamente los fenómenos económicos aplicando conocimientos del área de Cálculo, Probabilidad e

Inferencia Estadística. A menudo se utiliza para probar o refutar la validez de las teorías económicas. Para hacerlo, se requiere especificar el modelo econométrico de la teoría así como obtener la información histórica disponible acerca de las variables económicas involucradas en el modelo; posteriormente, se estiman los parámetros del modelo, se realizan las pruebas de hipótesis pertinentes y finalmente se obtienen los pronósticos que pueden ser útiles en la toma de decisiones, aunque otro aspecto importante es la validación de los supuestos, si alguno de los supuestos del modelo que se esté utilizando, no se cumple, entonces es necesario, buscar otro modelo que se ajuste a las características del problema.

Como podrá notar uno de los principales factores que influye en la calidad de los resultados obtenidos son los datos, las variables económicas no son controlables, por lo que la información histórica empleada proviene de alguna agencia o institución pública o privada. Desafortunadamente en ocasiones no se tiene demasiada información, lo que dificulta la tarea de estimación del modelo econométrico.

Entre las herramientas que emplea para el análisis de la información se encuentran los modelos de series de tiempo, de regresión lineal simple y múltiple así como el logístico, estos tres últimos serán analizados en este trabajo.

2.4 Modelo de Regresión Lineal Simple

Una de las técnicas más ampliamente utilizadas en diferentes áreas, para investigar y modelar la relación existente entre las variables de estudio, es el análisis de regresión. Existen diferentes tipos de análisis de regresión, entre los cuales encontramos: el lineal simple, lineal múltiple, polinomial, no lineal entre otros. En el caso de dos variables, para poder determinar cuál es el modelo de regresión más adecuado en principio, es necesario graficar los datos que se tienen, a esta gráfica se le conoce como gráfica de dispersión. Esto nos mostrará la relación que existe entre las variables en cuestión.

En el modelo de regresión lineal simple, la variable dependiente, es explicada por sólo una variable independiente, y se expresa de la siguiente forma:

$$y = \beta_0 + \beta_1 x + \varepsilon \dots 2$$

donde:

y: variable dependiente o de respuesta.

x: variable independiente o regresora.

ε : Error aleatorio

β_0, β_1 : Coeficientes de regresión (constantes desconocidas).

El error aleatorio, ε , se considera en el modelo 2, debido a que no todas las observaciones caen en la línea recta, y es la diferencia entre el valor observado “y”, y la línea recta. Además, representa la influencia de las variables omitidas en el modelo y los errores de medición.

El modelo 2, se conoce como el modelo de regresión poblacional, y cuando se ajusta a la muestra aleatoria disponible, entonces, se tiene el modelo de regresión muestral.

Antes de presentar la forma de estimar los valores de β_0 y β_1 , es importante conocer los supuestos que sustentan al modelo, los cuales son:

1. **Linealidad¹**: El modelo de regresión es lineal en los parámetros β_0 y β_1 , es decir: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
2. La esperanza de los errores es cero: $E(\varepsilon_i) = 0$
3. **Homocedasticidad²**: La varianza de los errores es constante: $E(\varepsilon_i^2) = \sigma^2$, $i = 1, 2, \dots, n$
4. Los errores no están correlacionados, es decir, el valor de un error no afecta el valor de otro: $E(\varepsilon_i \varepsilon_j) = 0$ para $i \neq j$

¹ Un ejemplo de un modelo no lineal es: $y = \beta_0 + \exp(\beta_1 x_1) + \varepsilon$

² Cuando la varianza para cada error no es la misma, es decir, cambia para cada uno de los errores: $E(\varepsilon_i^2) = \sigma_i^2$, se le conoce como heterocedasticidad.

5. La variable x es no estocástica. Las únicas variables aleatorias involucradas en el modelo son “ y ” y “ ε ”. Las constantes desconocidas β_0 y β_1 son parámetros con valores desconocidos pero específicos para una población particular. Para la variable “ y ” existe una distribución de probabilidad y su respectiva media es:

$$E(y | x) = \beta_0 + \beta_1 x$$

y su varianza es: $Var(y | x) = Var(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2$

6. Los errores están distribuidos independientemente de forma Normal con media cero y varianza σ^2 , es decir: $\varepsilon \sim N(0, \sigma^2)$.

Este último supuesto, es necesario para poder hacer inferencias acerca de los parámetros del modelo, como pruebas de hipótesis y la construcción intervalos de confianza.

La primera fase del análisis de regresión consiste en estimar los parámetros β_0 y β_1 , los métodos empleados para hacerlo son: máxima verosimilitud y mínimos cuadrados, en este caso, ambos proporcionan los mismos resultados.

2.4.1 Método de mínimos cuadrados

Para poder estimar los parámetros desconocidos β_0 y β_1 , se requiere de una muestra de observaciones de tamaño $n > 2$. Suponga que se tienen n pares de datos: $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, entonces el modelo de regresión muestral se escribe de la siguiente forma:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, 2, \dots, n \quad \dots \mathbf{3}$$

El método de mínimos cuadrados nos permite encontrar la recta estimada: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ que mejor se ajuste a los datos, minimizando la suma de los cuadrados de los errores, es decir:

$$u(\beta_0, \beta_1) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

Para minimizar esta suma, se procede a obtener las derivadas parciales y se igualan a cero:

$$\frac{\partial u}{\partial \hat{\beta}_0} = \sum_{i=1}^n (-2)(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = 0 \quad \dots 4$$

$$\frac{\partial u}{\partial \hat{\beta}_1} = \sum_{i=1}^n (-2)x_i(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = -2 \sum_{i=1}^n x_i \varepsilon_i = 0 \quad \dots 5$$

Realizando algunos despejes, en ambas, se obtiene:

$$\sum_{i=1}^n y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n x_i y_i = \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2$$

A este sistema de ecuaciones se le conoce como las ecuaciones normales. Resolviendo el sistema obtenemos los estimadores de β_0 y β_1 , los cuales son:

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad \dots 6$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \bar{y} - \hat{\beta}_1 \bar{x} \quad \dots 7$$

donde:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i ; \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

entonces, el modelo de regresión lineal simple ajustado es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Cabe señalar que, al numerador de la expresión 6 se le conoce como la suma corregida de los productos cruzados de x_i con y_i , mientras que al denominador se le conoce como la suma corregida de los cuadrados de las x_i , y se pueden reescribir de la siguiente forma:

$$S_{xy} = \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} = \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^n y_i (x_i - \bar{x}) \quad \dots 8$$

y

$$S_{xx} = \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad \dots 9$$

De igual forma podemos reescribir la expresión 6 como:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad \dots 10$$

Como se puede observar los estimadores de mínimos cuadrados son funciones de la muestra, consecuentemente pueden variar de muestra en muestra, por lo que resulta útil calcular sus respectivas varianzas, las cuales están dadas como sigue:

$$Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \quad \dots 11$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}} \quad \dots 12$$

Por otro lado, resulta importante conocer las propiedades estadísticas de los estimadores de mínimos cuadrados $\hat{\beta}_0$ y $\hat{\beta}_1$, las cuales se resumen en un Teorema muy importante conocido como de Gauss – Markov, el cual se enuncia a continuación:

Teorema de Gauss – Markov: Dado el modelo de regresión (2) con sus respectivos supuestos, los estimadores de mínimos cuadrados son insesgados y tienen mínima varianza dentro de la clase de los estimadores lineales insesgados, por lo que son los Mejores Estimadores Lineales Insesgados (MELI) o Best Linear Unbiased Estimator (BLUE) en inglés.

A su vez, el modelo de regresión muestral tiene las siguientes propiedades:

1. Pasa a través de las medias muestrales de “ x ” y “ y ”.
2. La suma de los residuales es cero, siempre que $\hat{\beta}_0 \neq 0$
3. La media de los valores estimados “ \hat{y} ” es igual a la media de “ y ”.

$$\frac{\sum_{i=1}^n \hat{y}_i}{n} = \frac{\sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i)}{n} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y}$$

4. La suma del producto de los residuales por la variable independiente o regresora “ x ” es igual a cero.

$$\sum_{i=1}^n e_i x_i = 0$$

5. La suma del producto de los residuales por los valores estimados “ \hat{y} ” es igual a cero.

$$\sum_{i=1}^n e_i \hat{y}_i = 0$$

2.4.2 Tabla de Análisis de Varianza

Una vez que se ha obtenido el modelo de regresión lineal muestral, la siguiente fase consiste en analizar el ajuste de la línea de regresión estimada a los datos, para ello primero se analizará la variación de los datos estimados así como las fuentes que la provocan. Para ello se utilizará la siguiente figura:

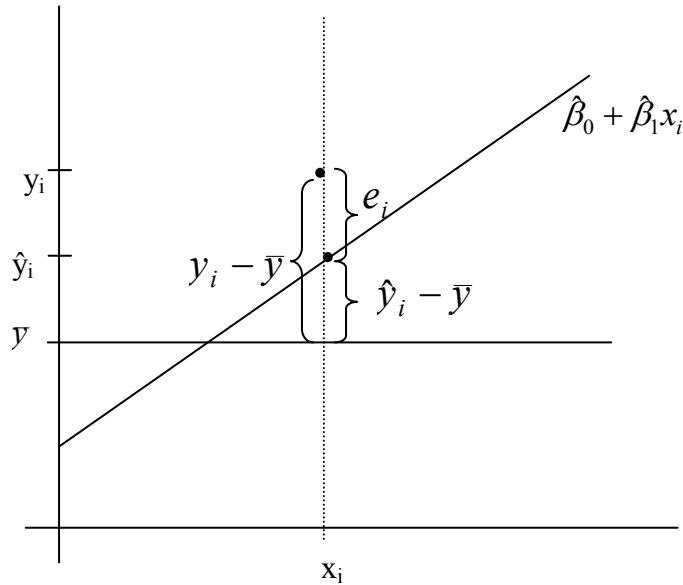


Figura 2.4-1 Modelo de Regresión Muestral

Se observa que los residuales e_i , se componen de dos partes: la primera es la desviación de los valores observados “ y_i ” con respecto a su media muestral “ \bar{y} ” y la segunda es la desviación de los valores estimados \hat{y}_i con respecto a su media muestral “ \bar{y} ”. Esto se resume en la siguiente identidad:

$$y_i - \hat{y}_i = (y_i - \bar{y}) - (\hat{y}_i - \bar{y})$$

que se puede reescribir como sigue:

$$y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

elevando ambas partes de la expresión al cuadrado y calculando la suma total se obtiene:

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + 2 \sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum (\hat{y}_i - \bar{y})^2 \dots 13$$

ahora se analizará el producto cruzado, se tiene:

$$\begin{aligned} 2 \sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= 2 \sum [y_i - \bar{y} - \hat{\beta}_1(x_i - \bar{x})][\hat{\beta}_1(x_i - \bar{x})] \\ &= 2 \sum [\hat{\beta}_1(x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}_1^2(x_i - \bar{x})^2] \end{aligned}$$

considerando las identidades 8, 9 y 10, se obtiene:

$$= 2 \sum [\hat{\beta}_1 S_{xy} - \hat{\beta}_1^2 S_{xx}]$$

$$2 \sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 2 \sum [\hat{\beta}_1^2 S_{xx} - \hat{\beta}_1^2 S_{xx}] = 0$$

por lo tanto, la ecuación 13 se puede reescribir de la siguiente forma:

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 \quad \dots \mathbf{14}$$

Donde:

Es la suma del cuadrado de las desviaciones de “y” respecto a su media muestral “ \bar{y} ” y representa la variación total de “y” antes de tomar en cuenta el efecto lineal de la variable “x”. Conocida como la Suma Total de Cuadrados (SST).

$$\sum_{i=1}^n (y_i - \bar{y})^2$$

Es la suma del cuadrado de las desviaciones de los valores estimados “ \hat{y}_i ” con respecto a su media muestral “ \bar{y} ” y representa la variación de “y” después de tomar en cuenta el efecto lineal de la variable “x”, es decir, es la variación explicada por la regresión. Se le denomina Suma de Cuadrados debida a la Regresión (SSR).

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

Es la suma del cuadrado de las desviaciones de los valores estimados “ \hat{y}_i ” con respecto a los valores observados “ y_i ”. Representa la variación residual inexplicada por el modelo de regresión y es conocida como la Suma de cuadrados de los residuales (SSE).

$$\sum_{i=1}^n e_i^2$$

Entonces: **SST = SSR + SSE**

Es decir, la variación total inexplicada SST se descompone en dos partes, la primera es la variación debida a la recta de regresión (SSR) y la segunda es la variación residual inexplicada SSE. A esto se le conoce como descomposición de la varianza.

Con esta información se construye la tabla de Análisis de Varianza (ANOVA), como se muestra a continuación:

Fuente de variación	Grados de Libertad	Suma de cuadrados	Cuadrados Medios	F ₀ (Radio de Varianza)
Debida a la regresión	1	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$MSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\frac{MSR}{MSE}$
Debida a los residuales	n - 2	$\sum (y_i - \hat{y}_i)^2$	$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n - 2}$	
Total	n - 1	$\sum_{i=1}^n (y_i - \bar{y})^2$		

Donde la columna de cuadrados medios, se obtiene al dividir la suma de cuadrados entre sus respectivos grados de libertad.

Cabe señalar que,

$$\hat{\sigma}^2 = s^2 = \frac{SSE}{n - 2}$$

es un estimador de σ^2 si los datos se ajustan a una línea recta.

El radio de varianza o estadístico F_0 será utilizado en secciones posteriores para realizar pruebas de hipótesis respecto al parámetro β_1 .

Por otra parte, a partir de la ecuación 14 se obtiene el coeficiente de determinación r^2 , una medida de bondad de ajuste que cuantifica la variación total de “y” explicada por el modelo de regresión, es decir, mide la fuerza de la relación entre “x” y “y”. Este coeficiente se obtiene dividiendo ambos lados de la ecuación 14, entre SST, como se indica a continuación:

$$1 = \frac{SSR}{SST} + \frac{SSE}{SST}$$

$$1 - \frac{SSR}{SST} = \frac{SSE}{SST}$$

Entonces:
$$r^2 = 1 - \frac{SSR}{SST} = \frac{SSE}{SST}$$

Que a su vez es equivalente también a:

$$r^2 = \hat{\beta}_1^2 \left(\frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2} \right) = \hat{\beta}_1^2 \left(\frac{S_{xx}}{S_{yy}} \right)$$

Las propiedades más importantes de r^2 son:

1. Es una cantidad positiva.
2. $0 \leq r^2 \leq 1$.
3. $r^2 = 1$, ocurre cuando $\beta_1 \neq 0$ y $SSE = 0$, es decir, la recta estimada pasa por todos los puntos, existe un ajuste perfecto: $\hat{y}_i = y_i$.
4. $r^2 = 0$, ocurre cuando la variable independiente no ayuda a explicar el modelo, es decir, hay ausencia de alguna relación lineal entre “ x ” y “ y ”. En este caso, $\beta_1 = 0$ y la mejor predicción para “ y ” será la media muestral: $\hat{y}_i = \bar{y}$. Puede existir una relación no lineal.

De acuerdo con lo anterior, valores muy cercanos a la unidad, implican que la mayor parte de la variabilidad en “ y ” es explicada por el modelo.

2.4.3 Pruebas de Hipótesis

La siguiente fase del Análisis de Regresión Lineal contempla la construcción de intervalos de confianza y los procedimientos de prueba de hipótesis acerca de los parámetros desconocidos, pero antes es necesario considerar lo siguiente:

1. Como $\varepsilon_i \sim N(0, \sigma^2)$, entonces “ y_i ” se distribuye de forma Normal con la siguiente media y varianza:

$$E(y_i) = \beta_0 + \beta_1 x_i$$

$$Var(y_i) = \sigma^2$$

en forma más compacta: $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$

2. Como ya se mencionó, los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ son funciones lineales de la variable Y , entonces siguen una distribución normal y sus respectivas varianzas están dadas por 11 y 12:

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$$

Cabe señalar que en muy pocas ocasiones se conoce el valor de la varianza σ^2 , por lo que se tiene que utilizar el estimador s^2 , entonces para obtener los respectivos estimadores de las varianzas de $\hat{\beta}_0$ y $\hat{\beta}_1$ es necesario sustituir σ^2 por s^2 . Estos estimadores de las varianzas se utilizarán en los estadísticos de prueba.

3. $(n-2) \frac{s^2}{\sigma^2}$ sigue una distribución χ^2 con $n-2$ grados de libertad.

Los puntos 1 y 2 se derivan de una propiedad de la distribución Normal que dice que una función lineal de variables normalmente distribuidas sigue también una distribución Normal.

Por otro lado, no hay que perder de vista que el supuesto de normalidad en los errores es de vital importancia en los procedimientos de prueba de hipótesis y construcción de intervalos de confianza acerca de los parámetros desconocidos β_0 y β_1 . En esta sección se abordarán los procedimientos de prueba de hipótesis y se divide en tres subsecciones en la “A” se abordará las pruebas de hipótesis para β_0 y β_1 , la “B” se dedicará a la prueba pendiente cero y la “C” a las pruebas relacionadas con σ^2 .

A. Prueba de hipótesis acerca de la pendiente β_1 y el intercepto β_0

El procedimiento de prueba más común relacionado con la pendiente β_1 , consiste en suponer que ésta es igual a una constante β_1^* las respectivas hipótesis nula y alternativa son:

$$H_0: \beta_1 = \beta_1^*$$

$$H_1: \beta_1 \neq \beta_1^*$$

Se trata de una hipótesis de dos colas. El estadístico de prueba utilizado en este caso es:

$$T = \frac{\hat{\beta}_1 - \beta_1^*}{\sqrt{MSE/S_{xx}}}$$

que sigue una distribución t – Student con $n - 2$ grados de libertad si la hipótesis nula es verdadera. Note que el denominador del estadístico de prueba es el estimador de la desviación estándar de $\hat{\beta}_1$.

Posteriormente se obtiene el valor crítico correspondiente $t_{\alpha/2, n-2}$ y se compara con el estadístico calculado utilizando la siguiente regla de decisión:

“Si $|T| > t_{\alpha/2, n-2}$ entonces concluiremos que existe evidencia estadística con base en la muestra para Rechazar H_0 ”

De igual forma se pueden formular pruebas de hipótesis de cola superior e inferior para β_1 , el siguiente cuadro resume para cada caso la formulación de hipótesis con su respectiva regla de decisión:

Tipo de Prueba	Hipótesis Nula	Hipótesis Alternativa	Regla de Decisión Rechazar H_0 si:
Cola Superior	$H_0: \beta_1 \geq \beta_1^*$	$H_1: \beta_1 < \beta_1^*$	$T < -t_{\alpha, n-2}$

Tipo de Prueba	Hipótesis Nula	Hipótesis Alternativa	Regla de Decisión Rechazar H_0 si:
Cola Inferior	$H_0: \beta_1 \leq \beta_1^*$	$H_1: \beta_1 > \beta_1^*$	$T > t_{\alpha, n-2}$

Por otro lado, el procedimiento de prueba de hipótesis para β_0 , es semejante al descrito anteriormente excepto por el estadístico de prueba. A continuación se describe la prueba de dos colas correspondiente a β_0 :

$$H_0: \beta_0 = \beta_0^*$$

$$H_1: \beta_0 \neq \beta_0^*$$

donde β_0^* es una constante dada. En este caso el estadístico de prueba que se utilizará es el siguiente:

$$T = \frac{\hat{\beta}_0 - \beta_0^*}{\sqrt{MSE \left(\frac{1}{n} + \frac{x^2}{S_{xx}} \right)}}$$

que sigue una distribución t – Student con $n - 2$ grados de libertad si la hipótesis nula es verdadera.

La regla de decisión asociada es: “Si $|T| > t_{\alpha/2, n-2}$ entonces, concluiremos que existe evidencia estadística con base en la muestra para Rechazar H_0 ”

La siguiente tabla resume el procedimiento de prueba para el caso de cola superior y el inferior:

Tipo de Prueba	Hipótesis Nula	Hipótesis Alternativa	Regla de Decisión Rechazar H_0 si:
Cola Superior	$H_0: \beta_0 \geq \beta_0^*$	$H_1: \beta_0 < \beta_0^*$	$T < -t_{\alpha, n-2}$

Tipo de Prueba	Hipótesis Nula	Hipótesis Alternativa	Regla de Decisión Rechazar H_0 si:
Cola Inferior	$H_0: \beta_0 \leq \beta_0^*$	$H_1: \beta_0 > \beta_0^*$	$T > t_{\alpha, n-2}$

B. Prueba pendiente cero ($H_0: \beta_1 = 0$)

Este es un caso especial de la prueba de hipótesis para β_1 descrita anteriormente y ayuda a determinar la significancia de la regresión, es decir, permite establecer si “y” está relacionada linealmente con “x”. La formulación de la hipótesis, es la siguiente:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

Existen dos formas de efectuar esta prueba de hipótesis: la primera es el procedimiento que se describió en la sección anterior, en el cual se utiliza la distribución t – Student, la segunda utiliza la distribución F y se describe a continuación.

El estadístico de prueba que se utilizará se obtiene de la última columna de la tabla ANOVA:

$$F_0 = \frac{MSR}{MSE} = \frac{SSR/1}{SSE/n-2}$$

donde MSR y MSE son variables aleatorias independientes y sus correspondientes valores esperados son:

$$E(MSR) = \sigma^2 + \beta_1^2 S_{xx} \quad \dots \mathbf{15}$$

$$E(MSE) = \sigma^2 \quad \dots \mathbf{16}$$

Como puede notar, MSE, es un estimador insesgado de σ^2 .

El estadístico F_0 sigue una distribución F con 1 y $n - 2$ grados de libertad para el numerador y el denominador respectivamente.

Una vez obtenido el valor crítico de tablas $F_{\alpha,1,n-2}$, la regla de decisión correspondiente consiste en :

“Si $F_0 > F_{\alpha,1,n-2}$ entonces concluiremos que existe evidencia estadística con base en la muestra para Rechazar H_0 ”

Por lo que si F_0 es muy grande, entonces es probable que $\beta_1 \neq 0$. En caso de que $\beta_1=0$, entonces las ecuaciones 15 y 16 serán estimadores idénticos de σ^2 , es decir, “ x ” no tiene influencia lineal en “ y ”.

Las interpretaciones de los resultados obtenidos para esta prueba, se resumen a continuación:

1. No rechazar H_0 puede significar:
 - i. La contribución de la variable “ x ” para explicar la variación en “ y ” es poca o nula. En este caso el modelo $\hat{y}_i = \bar{y}$ es tan bueno como $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, entonces la mejor predicción por el principio de parsimonia es el primero.
 - ii. Las variables “ x ” y “ y ” están relacionadas de forma no lineal.
2. Rechazar H_0 puede significar:
 - i. El modelo de regresión lineal simple es adecuado, es decir, la mayor parte de las observaciones caen en la línea recta.
 - ii. Existe un efecto lineal en “ x ” pero tal vez la gran parte de las observaciones no caen en la línea recta, por lo que para obtener un mejor ajuste en los datos, resulta necesario añadir términos polinomiales de orden mayor.

2.4.4 Intervalos de confianza para β_0 , β_1 y σ^2

Se pueden construir intervalos de confianza para β_0 y β_1 utilizando la distribución t – Student. En el caso específico de β_1 , se tiene:

$$P\left(-t_{\alpha/2} \leq \frac{\hat{\beta}_1 - \beta_1}{\sqrt{MSE/S_{XX}}} \leq t_{\alpha/2}\right) = 1 - \alpha$$

donde $t_{\alpha/2}$ es el valor crítico t al nivel de significancia $\alpha/2$. Entonces el intervalo de confianza al $(1 - \alpha)100\%$ para β_1 está dado por:

$$\hat{\beta}_1 - t_{\alpha/2} \sqrt{MSE/S_{XX}} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2} \sqrt{MSE/S_{XX}} \dots 17$$

Se procede de manera similar para β_0 y el correspondiente intervalo de confianza al $(1 - \alpha)100\%$ para β_0 está dado por:

$$\hat{\beta}_0 - t_{\alpha/2} \sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)} \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2} \sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)} \dots 18$$

Mientras que, el intervalo de confianza para σ^2 al $(1 - \alpha)100\%$ está dado por:

$$\frac{(n-2)MSE}{\chi_{\alpha/2, n-2}^2} \leq \sigma^2 \leq \frac{(n-2)MSE}{\chi_{1-\alpha/2, n-2}^2} \dots 19$$

No hay que perder de vista la interpretación frecuentista de estos intervalos, puesto que no se puede afirmar que la probabilidad de que el intervalo 17 contenga al parámetro β_1 es del 95%. Si se toman otras muestras del mismo tamaño que la actual y se construye un intervalo de confianza correspondiente a β_1 para cada muestra, entonces el 95% de estos intervalos contendrá al parámetro β_1 , por lo que la interpretación es a largo plazo. Lo mismo sucede para los intervalos 18 y 19.

2.4.5 Intervalo de confianza para $\mu_{y|x}$

Como podrá notar, dado un valor específico " x_0 ", el estimador puntual de la media condicional de "y" dado un valor específico " x_0 ", $\mu_{y|x_0}$, es el siguiente:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

que se distribuye de manera normal con media $\mu_{y|x_0} = \beta_0 + \beta_1 x_0$ y con varianza:

$$\text{Var}(\hat{y}_0) = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)$$

Estos datos son importantes para construir un intervalo de confianza para la media condicional de “y” dado un valor específico “x₀”, $\mu_{y|x_0}$, para ello considere la siguiente variable:

$$\frac{\hat{y}_0 - \mu_{y|x_0}}{\sqrt{\text{MSE} \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}}$$

que se distribuye como una t – Student con n – 2 grados de libertad. A partir de esta expresión, se construye el siguiente intervalo de confianza al (1 – α)100% para $\mu_{y|x_0}$:

$$\hat{y}_0 - t_{\alpha/2, n-2} \sqrt{\text{MSE} \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \leq \mu_{y|x_0} \leq \hat{y}_0 + t_{\alpha/2, n-2} \sqrt{\text{MSE} \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}$$

Cabe señalar que además de obtener un intervalo de confianza para la media condicional de “y” dado el valor deseado “x₀”, se tiene la opción de calcular este intervalo para cada valor de “x” de la muestra, con lo que se obtendrían las bandas de confianza para el modelo de regresión muestral.

2.4.6 Predicción de nuevas observaciones

El modelo de regresión lineal simple también puede ser empleado para predecir el valor individual de “y₀” correspondiente a un valor específico “x₀”, en este caso, el estimador de “y₀” es el siguiente:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

que también se utiliza como estimador de la media condicional $\mu_{y|x_0}$. Cabe señalar que es necesario considerar la variabilidad que tiene con respecto al verdadero valor “y₀”, denominado error de predicción o de pronóstico y que se expresa de la siguiente forma:

$$\varepsilon_p = y_0 - \hat{y}_0$$

y se distribuye de forma Normal con media cero y varianza:

$$\text{Var}(y_0 - \hat{y}_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

de manera que la variable:

$$\frac{y_0 - \hat{y}_0}{\sqrt{\text{MSE} \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}}$$

tiene una distribución t – Student con $n - 2$ grados de libertad y es empleada para construir el intervalo de predicción al $(1 - \alpha)100\%$ para una observación futura “ y_0 ” en “ x_0 ” que es el siguiente:

$$\hat{y}_0 - t_{\alpha/2, n-2} \sqrt{\text{MSE} \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \leq y_0 \leq \hat{y}_0 + t_{\alpha/2, n-2} \sqrt{\text{MSE} \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}$$

Si en lugar de predecir una nueva observación futura “ y_0 ”, se ocupa de predecir cada una de las “ y ’s” actuales observadas y se construye el respectivo intervalo de predicción, notará que éste es más ancho que el intervalo de confianza correspondiente a la media condicional $\mu_{y|x_0}$, esto se debe a que el intervalo de predicción depende de dos fuentes de error, la primera es asociada al modelo ajustado y la segunda involucra el error de la observación futura. También observará que a medida que “ x_0 ” se aleja de la media muestral \bar{x} , el ancho del intervalo de predicción aumenta y como consecuencia la capacidad predictiva del modelo decrece.

2.4.7 Aplicación: El modelo CAPM (Capital Asset Pricing Model)

El CAPM es un modelo financiero, una de sus aplicaciones es la valuación de activos o acciones que cotizan en la bolsa, y se expresa de la siguiente forma:

$$(ER_i - r_f) = \beta_i (ER_m - r_f) \dots \mathbf{20}$$

Donde:

ER_i : es el rendimiento esperado de la acción i .

r_f : es el rendimiento requerido libre de riesgo, un ejemplo son los bonos del tesoro a 90 días.

ER_m : es el rendimiento de la cartera del mercado, por ejemplo, el índice de acciones Standard And Poors S&P 500.

β_i : Coeficiente beta de la acción i . Se emplea como una medida del riesgo sistemático. El riesgo se compone de dos partes: la que puede ser eliminada por medio de la diversificación y que se denomina diversificable o no sistemático, y la parte que se atribuye a factores políticos, inflación guerras, entre otros, que afectan a todas las acciones y que no puede ser eliminado por medio de la diversificación, recibe el nombre de sistemático o no diversificable.

Si el CAPM se mantiene, entonces el premio al riesgo de la acción i , $(ER_i - r_f)$, es igual al coeficiente beta de dicha acción por el premio al riesgo del mercado, $(ER_m - r_f)$.

El modelo es conocido como línea del mercado y para fines empíricos se suele escribir de la siguiente forma:

$$(ER_i - r_f) = \beta_i(ER_m - r_f) + \varepsilon_i$$

Como puede notar, se trata de un modelo de regresión lineal sin el intercepto. En este caso, la variable dependiente es $(ER_i - r_f)$, mientras que la variable independiente o explicativa es β_i y no $(ER_m - r_f)$. Para poder estimar este modelo, es necesario conocer las β_i correspondientes a cada acción, las cuales se obtienen mediante la línea característica.

La línea característica es el siguiente modelo de regresión:

$$R_{jt} = r_{ft} + \beta_j(R_{mt} - r_{ft}) + \varepsilon_{jt}$$

donde ε_{jt} es el error estocástico y se consideran los supuestos del MRLS. Sea $R_{jt}^* = R_{jt} - r_{ft}$ (el retorno en exceso de la acción i) y $R_{Mt}^* = R_{Mt} - r_{ft}$ (el retorno en exceso del mercado), entonces, se tiene:

$$R_{jt}^* = \beta_j R_{Mt}^* + \varepsilon_{jt}$$

El cual es un modelo de regresión sin el intercepto, y donde la variable dependiente es R_{jt}^* , la variable independiente o explicativa es R_{Mt}^* y β_j es la pendiente del modelo. Este modelo puede estimarse en presencia del intercepto, es decir, se considera el siguiente modelo:

$$R_{jt}^* = \alpha_j + \beta_j R_{Mt}^* + \varepsilon_{jt} \dots 21$$

Si el CAPM es válido, entonces, el término α_j es igual a cero, en caso de que $\alpha_j \neq 0$, existe la posibilidad de una mala cotización.

Para estimar el modelo 21, se requiere conocer los valores R_{jt} , r_{ft} y R_{Mt} para $t = 1, 2, \dots, n$, posteriormente, se calculan R_{jt}^* y R_{Mt}^* y se efectúa la regresión.

Ruppert [2004], ejemplifica este modelo utilizando los siguientes datos diarios durante el periodo del 1 de Noviembre de 1993 al 3 de Abril del 2003:

- El logaritmo del precio de cierre de las acciones de Microsoft, R_{jt} .
- El logaritmo del precio de cierre del índice S&P 500, R_{Mt} .

Y como retorno libre de riesgo utilizó el rendimiento de los bonos del tesoro a 90 días, éste fue dividido entre 100 y después entre 253 para convertirlo a un porcentaje diario (se tomaron en cuenta 253 días hábiles en un año), este dato es r_{ft} .

A partir de estos datos, se calcularon R_{jt}^* y R_{Mt}^* y se efectuó la regresión 21, los resultados obtenidos fueron los siguientes:

Estimador	Desviación estándar estimada
$\hat{\alpha}_j = 0.00069605$	0.00040936
$\hat{\beta}_j = 1.24449$	0.03543

La interpretación de $\hat{\beta}_j$ es la siguiente: un incremento (disminución) del 1% en la tasa de retorno del mercado conduce en promedio a un incremento (disminución) de cerca del 1.24449% en la tasa de retorno de las acciones de Microsoft.

Entonces, el modelo estimado queda de la siguiente forma:

$$\hat{R}_{jt}^* = 0.00069605 + 1.24449 R_{mt}^*$$

Cabe señalar que, si los datos son consistentes con el CAPM, entonces, α_j debe ser igual a cero, por lo que para checar la validez del CAPM, se efectúa la siguiente prueba de hipótesis para el intercepto, la cual fue analizada en la sección 2.4.3:

$$H_0 : \alpha_j = 0 \quad H_1 : \alpha_j \neq 0$$

El estadístico de prueba calculado es: $T = \frac{0.00069605}{0.00040936} = 1.7$

Si fijamos el nivel de significancia en 0.05, entonces, el valor de tablas es: $t_{0.025,2360} = 1.96$, y como $|T| < t_{0.05,2360}$, entonces concluimos que existe evidencia estadística con base en la muestra para no rechazar la hipótesis nula, es decir, los datos son consistentes con el CAPM.

Pero si fijamos el nivel de significancia en 0.1, el valor de tablas es $t_{0.05,2360} = 1.645$, y como $|T| > t_{0.05,2360}$, en este caso, se concluye que existe evidencia estadística con base en la muestra para rechazar la hipótesis nula, lo cual implica que los datos no son consistentes con el CAPM.

Como puede notar, en este caso, el CAPM es válido para un nivel de significancia de 0.05 pero no para 0.1.

Por otra parte, la tabla ANOVA que obtuvo es la siguiente:

Fuente de variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	F_0
Debida a la regresión	1	0.48843	0.48843	1234.07
Debida a los residuales	2360	0.93407	0.00039579	
Total	2361	1.42250		

De aquí se obtiene que el estimador de la varianza es $s^2 = 0.00039579$.

Y se puede efectuar la prueba de significancia: $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$, el estadístico de prueba es: $F_0 = 1234.07$. Si fijamos el nivel de significancia en 0.05, entonces el valor crítico es: $F_{0.05,1,2360} = 3.84$. Como $F_0 > F_{0.05,1,2360}$ entonces se concluye que existe evidencia estadística con base en la muestra para rechazar la hipótesis nula, es decir, la variable R_{mt}^* sí explica a la variable dependiente.

El coeficiente de determinación es $R^2 = 0.3434$. En el modelo CAPM R^2 indica la proporción del riesgo que se atribuye al riesgo sistemático o de mercado, en este caso, corresponde al 34.4%. Mientras que, $1 - R^2$ indica la proporción del riesgo que no es de mercado y que se puede eliminar por medio de la diversificación, en este caso es 65.6%, como puede notar, es muy alto, por lo que tal vez convendría realizar la diversificación.

Un aspecto importante que se debe resaltar es que la β_j se toma como una constante desconocida en el modelo, pero ¿realmente se mantendrá constante si se toma un periodo largo como muestra para estimar el modelo?

2.5 Modelo de Regresión Lineal Múltiple

El modelo de regresión lineal múltiple es una extensión del modelo de regresión lineal simple, en este caso, existe una variable dependiente sujeta a “k” variables independientes o regresoras, es decir:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad \dots \quad 22$$

para simplificar el tratamiento del modelo se utilizará la notación matricial:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}$$

Donde:

Y: Vector de $n \times 1$ de observaciones de la variable dependiente

X: Matriz de $n \times (k+1)$ que contiene las n observaciones correspondientes a las k variables.

B: Vector de $(k+1) \times 1$ de los parámetros desconocidos $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ o coeficientes de regresión, estos son constantes.

E: Vector de $n \times 1$ de los errores o perturbaciones aleatorias.

Para ilustrar la ecuación 22 en forma matricial, se tiene:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\begin{array}{ccccccc} \mathbf{Y} & = & & \mathbf{X} & & \mathbf{B} & + & \mathbf{E} \\ n \times 1 & & & n \times (k+1) & & (k+1) \times 1 & & n \times 1 \end{array}$$

Los supuestos asociados al modelo son los siguientes:

1. **Linealidad**: El término lineal implica que el modelo es lineal en los parámetros, es decir: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$

2. La esperanza del vector de los errores es cero: $E(\mathbf{E}) = \mathbf{0}$, donde \mathbf{E} y $\mathbf{0}$ son vectores de $n \times 1$.
3. **Homocedasticidad y no correlación serial:** La varianza es la misma para todos los errores y además no existe correlación entre los errores, es decir, $E(\varepsilon_i \varepsilon_j) = 0$ para $i \neq j$. Estos dos supuestos se resumen en la siguiente notación matricial: $\text{var}(\mathbf{E}) = \sigma^2 \mathbf{I}$ donde \mathbf{I} es la matriz identidad de $n \times n$.
4. La matriz \mathbf{X} es no estocástica, las únicas variables aleatorias incluidas en el modelo son \mathbf{Y} y \mathbf{E} .
5. El rango de la matriz \mathbf{X} es igual a $k+1$, es decir, no existe multicolinealidad, dicho de otra manera, las columnas de la matriz son linealmente independientes, no existe una relación lineal exacta entre las variables x .
6. El vector de los errores se distribuye de forma normal con un vector de medias cero y una matriz de varianza – covarianza $\sigma^2 \mathbf{I}$, es decir: $\mathbf{E} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$.

2.5.1 Estimación del vector de parámetros B: Método de Mínimos Cuadrados Ordinarios

Para aplicar el método de mínimos cuadrados ordinarios, es necesario que el número de observaciones “n” sea mayor que el número de parámetros a estimar, es decir, $n > k+1$.

Se procede igual que en el caso de dos variables, buscamos un modelo $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}}$ que mejor se ajuste a los datos, esto se logra, minimizando la suma de cuadrados de los errores, es decir:

$$S(\beta_0, \beta_1, \dots, \beta_k) = \sum \varepsilon_i^2 = \mathbf{E}' \mathbf{E} \text{ donde } \mathbf{E} = \mathbf{Y} - \hat{\mathbf{Y}}, \text{ entonces:}$$

$$\begin{aligned} S(\beta_0, \beta_1, \dots, \beta_k) &= \mathbf{E}' \mathbf{E} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})' (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) \\ &= \mathbf{Y}' \mathbf{Y} - \mathbf{Y}' \mathbf{X}\hat{\mathbf{B}} - \hat{\mathbf{B}}' \mathbf{X}' \mathbf{Y} + \hat{\mathbf{B}}' \mathbf{X}' \mathbf{X}\hat{\mathbf{B}} \end{aligned}$$

como $\mathbf{Y}' \mathbf{X}\hat{\mathbf{B}}$ y $\hat{\mathbf{B}}' \mathbf{X}' \mathbf{Y}$ son escalares, entonces

$$S(\beta_0, \beta_1, \dots, \beta_k) = \mathbf{Y}' \mathbf{Y} - 2\hat{\mathbf{B}}' \mathbf{X}' \mathbf{Y} + \hat{\mathbf{B}}' \mathbf{X}' \mathbf{X} \hat{\mathbf{B}}$$

Ahora, derivando con respecto a \mathbf{B} e igualando a cero, se tiene:

$$\frac{\partial S}{\partial \mathbf{B}} = -2\mathbf{X}' \mathbf{Y} + \hat{\mathbf{B}}' \mathbf{X}' \mathbf{X} + \mathbf{X}' \mathbf{X} \hat{\mathbf{B}}$$

$$= -2\mathbf{X}' \mathbf{Y} + 2\mathbf{X}' \mathbf{X} \hat{\mathbf{B}}$$

$$\mathbf{X}' \mathbf{X} \hat{\mathbf{B}} = \mathbf{X}' \mathbf{Y} \dots 23$$

La ecuación resultante representa las ecuaciones normales, cuya forma matricial es la siguiente:

$$\begin{bmatrix} n & \sum x_{i2} & \cdots & \sum x_{ik} \\ \sum x_{i2} & \sum x_{i2}^2 & \cdots & \sum x_{i2}x_{ik} \\ \vdots & \vdots & \vdots & \vdots \\ \sum x_{ik} & \sum x_{ik}x_{i2} & \cdots & \sum x_{ik}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i2}y_i \\ \vdots \\ \sum x_{ik}y_i \end{bmatrix}$$

$$\mathbf{X}' \mathbf{X} \hat{\mathbf{B}} = \mathbf{X}' \mathbf{Y}$$

Para resolver este sistema de ecuaciones se multiplica ambos lados de la expresión 23 por $(\mathbf{X}' \mathbf{X})^{-1}$:

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

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

Una vez que se ha obtenido el vector de estimadores $\hat{\mathbf{B}}$ del vector de parámetros \mathbf{B} , se puede escribir: $\mathbf{Y} = \mathbf{X}\hat{\mathbf{B}} + \mathbf{E}$. Como puede notar este resultado, está condicionado en la existencia de la matriz inversa $(\mathbf{X}' \mathbf{X})^{-1}$, si el supuesto 5 se cumple, entonces efectivamente dicha inversa existe, pero ¿qué pasa cuando la matriz \mathbf{X} no es de rango completo y por consecuencia la matriz inversa $(\mathbf{X}' \mathbf{X})^{-1}$ no puede calcularse? Una forma de resolver este problema, consiste en calcular el inverso generalizado de la matriz $(\mathbf{X}' \mathbf{X})$, para mayores detalles, le sugiero consultar el apéndice 1.

Al igual que en el caso del modelo de regresión lineal simple, nos interesa conocer la varianza de los estimadores $\hat{\mathbf{B}}$, utilizando notación matricial, se obtendrá una matriz en cuya diagonal se encontrarán las varianzas de los estimadores $\hat{\beta}$ y los elementos que no pertenecen a la diagonal son las covarianzas entre los estimadores. A esta matriz se le denomina matriz de varianzas - covarianzas, es una matriz simétrica de tamaño $(k+1) \times (k+1)$ y es la siguiente:

$$\text{var-cov}(\hat{\mathbf{B}}) = \sigma^2 (\mathbf{X}' \mathbf{X})^{-1}$$

En el MRLS se obtuvo un estimador de la varianza σ^2 utilizando a MSE, en este caso, se procederá de manera similar, sin perder de vista que este estimador será denotado por s^2 :

$$\begin{aligned} SSE &= \sum e_i^2 = \mathbf{E}' \mathbf{E} \\ &= (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})' (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) \end{aligned}$$

$$SSE = \mathbf{Y}' \mathbf{Y} - \hat{\mathbf{B}}' \mathbf{X}' \mathbf{Y}$$

El estimador s^2 de σ^2 es MSE entre el número de grados de libertad, que es el número de datos menos el número de parámetros estimados, en este caso es $n - (k+1)$, es decir:

$$s^2 = MSE = \frac{SSE}{n - (k + 1)}$$

Las propiedades de los estimadores de mínimos cuadrados son las siguientes:

1. Insesgamiento: $E(\hat{\mathbf{B}}) = \mathbf{B}$
2. El teorema de Gauss – Markov se extiende para el caso del Modelo de Regresión Lineal Múltiple, es decir, el vector de estimadores $\hat{\mathbf{B}}$ es el mejor estimador lineal insesgado de \mathbf{B} dentro de la clase de estimadores lineales insesgados de mínima varianza.

2.5.2 Tabla de Análisis de Varianza (ANOVA)

Este procedimiento es una generalización del que fue utilizado en el caso del MRLS, por lo que se conservará la notación. Sabemos que la suma total de cuadrados se descompone en dos partes, la primera es la variación debida a la regresión SSR, y la otra es la variación total inexplicada SSE, es decir:

$$SST = SSR + SSE$$

En notación matricial se tiene:

$$SSE = \mathbf{Y}' \mathbf{Y} - \hat{\mathbf{B}}' \mathbf{X}' \mathbf{Y}$$

$$SST = \mathbf{Y}' \mathbf{Y} - n\bar{y}^2 = \mathbf{Y}' \mathbf{Y} - \frac{\left(\sum y_i\right)^2}{n}$$

$$SSR = \hat{\mathbf{B}}' \mathbf{X}' \mathbf{Y} - n\bar{y}^2 = \hat{\mathbf{B}}' \mathbf{X}' \mathbf{Y} - \frac{\left(\sum y_i\right)^2}{n}$$

Con esta información se construye la tabla ANOVA:

Fuente de variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	F ₀ (Radio de Varianza)
Debida a la regresión	K	$SSR = \hat{\mathbf{B}}' \mathbf{X}' \mathbf{Y} - n\bar{y}^2$	$MSR = \frac{SSR}{k}$	$\frac{MSR}{MSE}$
Debida a los residuales	n - k - 1	$SSE = \mathbf{Y}' \mathbf{Y} - \hat{\mathbf{B}}' \mathbf{X}' \mathbf{Y}$	$MSE = \frac{SSE}{n - k - 1}$	
Total	n - 1	$SST = \mathbf{Y}' \mathbf{Y} - n\bar{y}^2$		

La Tabla ANOVA está conformada de formas cuadráticas. Respecto a los grados de libertad, la SSR tiene k, porque el espacio en el que están definidos los parámetros es (k+1) – dimensional y se pierde un grado de libertad en el cálculo de la media de la variable dependiente que como sabemos es igual a la media de la variable dependiente estimada, la

SSE tiene asociados $n - k - 1$ grados de libertad debido a que se estimaron $k + 1$ parámetros, entonces al total de observaciones se le resta $k + 1$ grados de libertad, y finalmente la SST tiene asociados $n - 1$ grados de libertad que es la suma de los grados de libertad de SSR y SSE. La columna de cuadrados medios se calcula al dividir la suma de cuadrados entre sus respectivos grados de libertad. El radio de varianza será utilizado en las pruebas de significancia.

A. Coeficiente de determinación múltiple

El coeficiente de determinación múltiple se construye en forma similar al que se obtuvo en el MRLS y mide la proporción de la variación total de la variable dependiente “y” explicada por las variables regresoras x_1, x_2, \dots, x_k y se define de la siguiente forma:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

y conserva las mismas propiedades que en caso del MRLS.

Se puede utilizar para comparar la capacidad explicativa de dos modelos con la misma variable dependiente, el mismo número de observaciones n y que solo difieren por el número de regresores involucrados en el modelo. Aunque se debe tener cuidado ya que la inclusión de una nueva variable regresora en el modelo aumenta a R^2 , es decir, a medida que aumenta el número de variables regresoras en el modelo aumenta R^2 o en el peor de los casos no se modifica, por lo que R^2 se considera una función no decreciente del número de variables regresoras. Debido a este inconveniente, muchos investigadores prefieren utilizar el coeficiente de determinación múltiple ajustado, que se define de la siguiente forma:

$$\bar{R}^2 = 1 - \frac{SSE / (n - v)}{SST / (n - 1)} = 1 - \frac{(n - 1)}{(n - v)} (1 - R^2)$$

Donde v es el número de parámetros en el modelo incluyendo al intercepto. Su interpretación se similar a la de R^2 ya que tiene las mismas propiedades del R^2 original y cumple con la función de medida de bondad de ajusta del modelo, solo que a diferencia del R^2 original contempla los grados de libertad que se pierden al incluir a una nueva variable

regresora, de manera que $(n - 1)/(n - v)$ aumenta cuando el número de variables regresoras aumenta mientras que $(1 - R^2)$ disminuye y como consecuencia el \bar{R}^2 aumenta en menor cantidad que R^2 . Se debe tener en cuenta lo siguiente:

1. Cuando $R^2=1$, entonces $\bar{R}^2=1$.
2. Cuando $R^2=0$, entonces $\bar{R}^2=(1-v)/(n-v)$, y si $v > 1$ \bar{R}^2 puede ser negativo. En este caso \bar{R}^2 se toma como cero.

2.5.3 Intervalo de confianza para los parámetros del MRLM

Antes de analizar los procedimientos de construcción de intervalos de confianza para \mathbf{B} y de prueba de hipótesis acerca de \mathbf{B} , es necesario considerar los siguientes puntos que son importantes:

1. Como $\mathbf{E} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, entonces \mathbf{Y} se distribuye de forma Normal con la siguiente media:

$$\mu_{y|x_1, x_2, \dots, x_k} = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

y varianza: $\sigma^2 \mathbf{I}$.

2. El vector de estimadores $\hat{\mathbf{B}}$, son variables aleatorias distribuidas normalmente con media $E(\hat{\mathbf{B}}) = \mathbf{B}$ y matriz de varianzas - covarianzas $\sigma^2 (\mathbf{X}' \mathbf{X})^{-1}$, es decir: $\hat{\mathbf{B}} \sim N(\mathbf{B}, \sigma^2 (\mathbf{X}' \mathbf{X})^{-1})$

Para obtener un estimador de la matriz de varianzas - covarianzas, se reemplaza σ^2 por su respectivo estimador s^2 .

Ahora, se procederá a presentar el procedimiento para la construcción de intervalos de confianza.

A. Intervalo de confianza para β_i

Para obtener un intervalo de confianza para un coeficiente de regresión individual, β_i , al igual que en el caso del MRLS, se considera la siguiente variable:

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{s^2 M_{ii}}} \quad i = 0, 1, 2, \dots, k$$

que sigue una distribución t - Student con $n - k - 1$ grados de libertad, donde M_{ii} es el elemento de la diagonal principal de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ y el denominador es un estimador de la desviación estándar de $\hat{\beta}_i$.

Considerando lo anterior, el intervalo de confianza al $(1 - \alpha)100\%$ para el coeficiente de regresión β_i está dado por la siguiente expresión:

$$\hat{\beta}_i - t_{\alpha/2, n-k-1} \sqrt{s^2 M_{ii}} \leq \beta_i \leq \hat{\beta}_i + t_{\alpha/2, n-k-1} \sqrt{s^2 M_{ii}}$$

B. Intervalo de confianza para la media condicional

Al igual que en el MRLS, se puede construir un intervalo de confianza para la media condicional en un punto particular: $x_{01}, x_{02}, \dots, x_{0k}$, que en forma matricial se representa de la siguiente forma:

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ x_{01} \\ \vdots \\ x_{0k} \end{bmatrix}$$

por lo que la media condicional en este punto está dada por:

$$y_0 = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} + \dots + \beta_k x_{0k} = \mathbf{x}_0' \mathbf{B}$$

y un estimador puntual para y_0 es:

$$\hat{y}_0 = \mathbf{x}_0' \hat{\mathbf{B}}$$

Su respectiva varianza está dada por la siguiente expresión:

$$\text{var}(\hat{y}_0) = E[(\mathbf{x}_0' \hat{\mathbf{B}} - \mathbf{x}_0' \mathbf{B})(\mathbf{x}_0' \hat{\mathbf{B}} - \mathbf{x}_0' \mathbf{B})] = \sigma^2 \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0$$

El intervalo de confianza al $(1 - \alpha)100\%$ para la media condicional, es el siguiente:

$$\hat{y}_0 - t_{\alpha/2, n-k-1} \sqrt{s^2 \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0} \leq y_0 \leq \hat{y}_0 + t_{\alpha/2, n-k-1} \sqrt{s^2 \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0}$$

2.5.4 Prueba de hipótesis acerca de β

En esta sección se analizarán las pruebas de significancia de la regresión, las cuales permiten determinar si las variables regresoras ayudan a explicar la variable dependiente “y”. Se presentarán tres pruebas de significancia: la Total, la Parcial o individual y la Parcial múltiple, en la primera y la tercera se utilizará el estadístico F, el cual es un ratio de varianzas, donde el denominador es un estimador de la varianza σ^2 sea o no sea verdadera la hipótesis nula, mientras que el numerador es un estimador de σ^2 sólo si la hipótesis nula es verdadera.

A. Prueba de Significancia Total

Esta prueba permite determinar si existe una relación lineal entre la variable dependiente “y” y alguna de las k variables regresoras. Formalmente, la hipótesis nula y la alternativa se enuncian de la siguiente forma:

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

$$H_1 : \beta_j \neq 0$$

Es decir, se confronta la hipótesis nula de que las k variables regresoras consideradas en conjunto no explican a la variable dependiente “y”, contra la alternativa, que al menos una de las variables regresoras explica a “y”.

El estadístico de prueba a utilizar es el siguiente:

$$F_0 = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE}$$

que sigue una distribución F con k y n – k – 1 grados de libertad bajo la hipótesis nula.

La regla de decisión es:” Si $F_0 > F_{\alpha, n-k-1}$ entonces concluiremos que existe evidencia estadística con base en la muestra para Rechazar H_0 “.

Las interpretaciones de los resultados obtenidos por la prueba se resumen a continuación:

1. Rechazar H_0 significa que alguno de los regresores contribuye significativamente a explicar la variable dependiente “ y ”, pero esto no quiere decir que las k variables sean necesarias para obtener un buen modelo, para poder seleccionar las variables más significantes, es necesario aplicar las pruebas de significancia que se detallan en las siguientes secciones.
2. No rechazar H_0 implica que las k variables regresoras ayudan poco o nada a explicar la variable dependiente “ y ”, razón por la cual el modelo reducido $y = \beta_0 + \varepsilon$ y el modelo completo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$ son buenos para explicar a “ y ”, pero de acuerdo al principio de parsimonia, en este caso, el modelo más adecuado sería el reducido.

B. Prueba de Significancia Parcial o individual

Esta prueba ayuda a determinar si es significativa la contribución de la variable “ x_i ” dado que las otras variables regresoras están presentes en el modelo, de manera que los modelos que se comparan son el completo que incluye a todas las variables regresoras y el reducido que incluye a todos los regresores excepto a “ x_i ”, de esta forma se pueden ir eliminando del modelo aquellas variables regresoras que ayudan poco o nada en la predicción de “ y ”.

Formalmente, la hipótesis nula y la alternativa se enuncian de la siguiente manera:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

El estadístico de prueba a utilizar es el siguiente:

$$T_0 = \frac{\hat{\beta}_i}{\sqrt{s^2 M_{ii}}}$$

donde M_{ii} es el elemento de la diagonal principal de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. Este estadístico sigue una distribución t – Student con $n - k - 1$ grados de libertad bajo la hipótesis nula.

La regla de decisión es: “Si $|T_0| > t_{\alpha/2, n-k-1}$, entonces concluiremos que existe evidencia estadística con base en la muestra para Rechazar H_0 “. La correspondiente interpretación a cada resultado es la siguiente:

1. No rechazar H_0 implica que la variable regresora “ x_i ” puede ser eliminada del modelo ya que su contribución a la predicción de “ y ” es poca o nula. De esta forma se obtendría un modelo más apropiado.
2. Rechazar H_0 implica que la variable regresora “ x_i ” contribuye significativamente a la predicción de “ y ”, por lo que el modelo completo es apropiado.

C. Prueba de Significancia Parcial – Múltiple

Esta prueba permite determinar si es significativa la contribución de un conjunto de dos o más variables regresoras a la predicción de “ y ” dado que las otras variables regresoras están incluidas en el modelo. Para efectuar esta prueba se utiliza el método de la suma extra de cuadrados, el cual se describe a continuación.

Como se desea determinar la contribución de un grupo de p variables regresoras al modelo, entonces es necesario particionar en dos al vector de parámetros y la matriz \mathbf{X} , entonces el modelo completo se puede describir de la siguiente forma:

$$\mathbf{y} = \mathbf{X}_1\mathbf{B}_1 + \mathbf{X}_2\mathbf{B}_2 + \mathbf{E} \quad \text{y} \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix}$$

donde \mathbf{B}_1 es de $(k+1-p) \times 1$, \mathbf{B}_2 es de $p \times 1$, \mathbf{X}_1 es de $n \times (k+1-p)$ y \mathbf{X}_2 es de $n \times p$.

Entonces, la hipótesis nula y la alternativa quedan de la siguiente forma:

$$H_0 : \mathbf{B}_2 = \mathbf{0}$$

$$H_1 : \mathbf{B}_2 \neq \mathbf{0}$$

dado que como se mencionó anteriormente se desea determinar si la contribución de las p variables regresoras es significativa en el modelo completo.

Para efectuar la prueba de hipótesis se utiliza el método de la suma extra de cuadrados que se describe a continuación:

1. Se asume que la hipótesis nula es verdadera para poder determinar si la contribución de las p variables regresoras es significativa al modelo que incluye las otras $k+1-p$ variables regresoras. Entonces el modelo reducido se puede escribir de la siguiente forma: $\mathbf{y} = \mathbf{X}_1\mathbf{B}_1 + \mathbf{E}$
2. Se calcula la suma extra de cuadrados resultante de incluir a \mathbf{B}_2 en el modelo reducido, mediante la siguiente fórmula:

$$SSR(\mathbf{B}_2 | \mathbf{B}_1) = SSR(\mathbf{B}) - SSR(\mathbf{B}_1)$$

Donde:

$SSR(\mathbf{B}_2 | \mathbf{B}_1)$: Suma Extra de Cuadrados de \mathbf{B}_2 dado \mathbf{B}_1

$SSR(\mathbf{B})$: Suma de Cuadrados de la regresión del modelo completo (incluye a \mathbf{B}_1 y \mathbf{B}_2).

$SSR(\mathbf{B}_1)$: Suma de Cuadrados de la regresión del modelo reducido (cuando \mathbf{B}_2 no está incluido).

Esta suma extra de cuadrados mide el incremento que sufre la $SSR(\mathbf{B}_1)$ al añadir a \mathbf{B}_2 al modelo

3. Se obtiene el estadístico de prueba, utilizando la suma extra de cuadrados que se calculó en el paso anterior:

$$F_0 = \frac{SSR(\mathbf{B}_2 | \mathbf{B}_1) / p}{MSE}$$

sigue una distribución F con p y $n - k - 1$ grados de libertad. Como puede observar, el número de grados de libertad asociados a $SSR(\mathbf{B}_2 | \mathbf{B}_1)$ es p , el número de variables regresoras que se suponen son cero bajo la hipótesis nula.

4. La regla de decisión es: “Si $F_0 > F_{\alpha, p, n-k-1}$ entonces concluiremos que existe evidencia estadística con base en la muestra para Rechazar H_0 “. Si se rechaza H_0 , significa que al menos una de las p variables regresoras

contribuye significativamente al modelo, claro que para poder obtener un modelo más sencillo, resulta útil aplicar la prueba de significancia individual.

2.5.5 Predicción de nuevas observaciones

El modelo también puede ser utilizado para predecir una nueva observación y_0 correspondiente a un punto particular \mathbf{x}_0 :

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ x_{01} \\ \vdots \\ x_{0k} \end{bmatrix}$$

el estimador puntual para y_0 es el siguiente:

$$\hat{y}_0 = \mathbf{x}_0' \hat{\mathbf{B}}$$

y su respectivo intervalo de confianza al $(1 - \alpha)100\%$ está dado por la siguiente expresión:

$$\hat{y}_0 - t_{\alpha/2, n-k-1} \sqrt{s^2 (1 + \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0)} \leq y_0 \leq \hat{y}_0 + t_{\alpha/2, n-k-1} \sqrt{s^2 (1 + \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0)}$$

En este intervalo de confianza están involucradas dos fuentes de error, la primera es debida al modelo ajustado y la segunda es debida al error de la observación futura.

2.5.6 Aplicación: Estimación del modelo CAPM en dos etapas

Como recordara, en la sección 2.4.7, se describió a grandes rasgos, en qué consiste el modelo CAPM, que se expresa de la siguiente forma:

$$(ER_i - r_f) = \beta_i (ER_m - r_f) \dots 24$$

Este modelo puede ser estimado en dos etapas las cuales son:

1. Etapa I: Considere que se tiene una muestra de N activos y que a su vez, para cada una de ellas se tiene una muestra de datos correspondiente a sus

respectivos rendimientos durante un periodo de tiempo. La primera etapa, consiste en efectuar la regresión a través del tiempo para cada activo:

$$R_{it} = \hat{\alpha}_i + \hat{\beta}_i R_{mt} + \varepsilon_{it} \quad t = 1, 2, \dots, k$$

en esta se estima el riesgo sistemático, $\hat{\beta}_i$, para cada activo i. En total, se realizarán N regresiones, una para cada acción, por lo que se tendrán, N valores estimados para β_i .

2. Etapa II: Se efectúa una regresión de corte transversal, es decir, se corre la siguiente regresión para los N activos:

$$\bar{R}_i - r_f = \hat{\gamma}_0 + \hat{\gamma}_1 \hat{\beta}_i + \varepsilon_i \dots \mathbf{25}$$

Donde:

\bar{R}_i : es el promedio o tasa media de retorno del i – ésimo activo riesgoso.

$\hat{\beta}_i$: es el estimador del riesgo sistemático del i – ésimo activo riesgoso que se obtuvo en la primera etapa.

$\hat{\gamma}_0$ y $\hat{\gamma}_1$: son los coeficientes estimados de la regresión.

Comparando las ecuaciones 24 y 25, se deduce, que si el CAPM es válido, entonces se espera que $\hat{\gamma}_0 = 0$ y $\hat{\gamma}_1 = \bar{R}_m - r_f$, donde \bar{R}_m es el porcentaje promedio observado de retorno en la cartera de mercado.

Aunque en estudios realizados, se ha encontrado que cuando los inversionistas incluyen pocos activos en sus carteras de inversión, otra variable explicativa que puede ser incluida en el modelo 25 es la varianza residual obtenida en la primera etapa para cada activo, en este caso si el CAPM es válido entonces el coeficiente de regresión para dicha variable debe ser cero.

Levy [1978] analizó los siguientes modelos:

$$\bar{R}_i - r_f = \gamma_0 + \gamma_1 \hat{\beta}_i + \varepsilon_i \dots \mathbf{26}$$

$$\bar{R}_i - r_f = \gamma_0 + \gamma_3 \hat{\sigma}_i^2 + \varepsilon_i \dots \mathbf{27}$$

$$\bar{R}_i - r_f = \gamma_0 + \gamma_2 \hat{S}_{e_i}^2 + \varepsilon_i \quad \dots \quad \mathbf{28}$$

$$\bar{R}_i - r_f = \gamma_0 + \gamma_1 \hat{\beta}_i + \gamma_3 \hat{\sigma}_i^2 + \varepsilon_i \quad \dots \quad \mathbf{29}$$

$$\bar{R}_i - r_f = \gamma_0 + \gamma_1 \hat{\beta}_i + \gamma_2 \hat{S}_{e_i}^2 + \varepsilon_i \quad \dots \quad \mathbf{30}$$

Donde:

$\hat{\beta}_i$: Riesgo sistemático estimado del activo i, en la primera etapa.

$\hat{\sigma}_i^2$: Varianza estimada para el activo i.

$\hat{S}_{e_i}^2$: Varianza residual estimada para el activo i, en la primera etapa.

Como puede notar, estos modelos corresponden a la etapa II, es decir, son regresiones de corte transversal, cada uno de ellos se corre para los N activos. Otro aspecto importante de resaltar, es que los primeros tres modelos son de regresión lineal simple, mientras que los últimos dos corresponden a modelos de regresión lineal múltiple.

Levy examinó estos modelos utilizando datos mensuales, semestrales y anuales de una muestra de 101 acciones cotizadas en NYSE (New York Stock Exchange) durante el periodo 1948 – 1968. Los resultados de su estudio son los siguientes:

1. Datos mensuales: La mayoría de los coeficientes de regresión para los modelos estimados son insignificantes y el coeficiente de determinación para cada modelo es menor al 5%.
2. Datos semestrales: Existe una mejoría en el coeficiente de determinación de cada modelo de regresión estimado, aunque sigue siendo bajo. Los resultados obtenidos indican que el modelo 27 explica mejor la variable dependiente que el modelo 26, es decir, la varianza estimada de cada activo i es mejor predictor que el riesgo sistemático estimado para cada activo i. Mientras que en el modelo 30, el coeficiente de regresión de $\hat{\beta}_i$ y el de $\hat{S}_{e_i}^2$, son estadísticamente significantes, lo que indica que el CAPM tradicional no se cumple. Finalmente, en el modelo 29, el coeficiente de regresión de $\hat{\sigma}_i^2$ es positivo y estadísticamente significativo, mientras que el coeficiente de

regresión de $\hat{\beta}_i$ es insignificante, es decir, esta variable puede ser omitida en el modelo.

3. Datos anuales: El coeficiente de determinación para cada uno de los modelos estimados es mayor que el que se obtuvo utilizando datos semestrales, además de que todos los coeficientes de regresión son positivos y significantes. De aquí se obtuvieron las siguientes conclusiones:
 - i. El coeficiente de determinación para el modelo 26, fue 21%, mientras que para el modelo 27 fue de 38%, por lo que el desempeño de este modelo, cuya única variable explicativa es la varianza estimada de cada activo i , es mejor que el primero, cuya única variable explicativa es el riesgo sistemático estimado de cada activo i .
 - ii. El coeficiente de determinación del modelo 30, fue de 39% tan sólo 1% mayor que el correspondiente al modelo 27.
 - iii. Al efectuar la regresión 29, el coeficiente del riesgo sistemático estimado, $\hat{\beta}_i$, es muy pequeño y estadísticamente insignificante, es decir, esta variable no contribuye en la explicación del comportamiento del precio y nuevamente encontramos que el CAPM tradicional no es válido.

Como puede notar, los resultados de las regresiones efectuadas son muy sensitivos al horizonte de inversión adoptado, en este caso, se obtuvieron mejores resultados al utilizar datos anuales.

Otro aspecto interesante, es que cuando los inversionistas incluyen pocos activos en sus carteras de inversión, como consecuencia de factores tales como la presencia de costos de transacción, entre otros, entonces el CAPM tradicional no es válido, por esta razón, en los modelos de regresión 27, 28, 29 y 30, los coeficientes de regresión de \hat{S}_e^2 y $\hat{\sigma}_i^2$ son estadísticamente significantes, e incluso explican mejor el comportamiento de los precios que $\hat{\beta}_i$, tal es de caso del modelo 29.

2.6 Modelo de Regresión Logística

Existen problemas en los cuales la variable dependiente o de respuesta, y , es dicótoma o binaria, es decir, sólo puede tomar de 1 ó 0, indicando la presencia o ausencia de la característica bajo estudio, respectivamente. Por ejemplo, si deseamos estudiar la compra de una casa considerando el nivel de ingresos de la persona, en este caso, la variable dependiente es cualitativa y se codifica como, $y = 1$, denota comprar la casa, mientras que, $y = 0$, denota no comprar la casa, la variable independiente es el nivel de ingresos de la persona.

Algunas de las herramientas más comúnmente utilizadas para resolver este tipo de problemas son: el modelo lineal de probabilidad, el modelo probit y el modelo de regresión logística (o modelo logit), este último será analizado en esta sección y se caracteriza por utilizar la función de distribución logística para relacionar la variable dependiente con las respectivas variables independientes.

El modelo de regresión logística se expresa de la siguiente forma:

$$\pi(x) = \frac{e^{p(x)}}{1 + e^{p(x)}} \dots \mathbf{31}$$

cuando existe sólo una variable independiente, entonces $p(x) = \beta_0 + \beta_1 x$, en este caso, 31, es conocido como el modelo de regresión logística simple. Cuando existen k variables independientes ($k > 1$), entonces $p(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_i x_i$, y la expresión 31, se denomina modelo de regresión logística múltiple.

En esta sección, se analizará el modelo de regresión logística simple, aunque se debe señalar que todos los procedimientos que se presentarán a continuación pueden generalizarse para el caso múltiple.

Una transformación que será útil en lo sucesivo es la siguiente:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x$$

es conocida como transformación logit. Las propiedades más importantes del logit son las siguientes: es lineal en los parámetros β_0 y β_1 , puede ser continua y su rango depende de los valores que asuma la variable independiente.

Un aspecto importante que no hay que perder de vista es que, la variable dependiente es codificada como 0 ó 1, indicando la presencia o ausencia de la característica en cuestión, tomando en cuenta esto, la expresión 31 representa la probabilidad condicional de que la variable y sea igual a 1 dado x , por consiguiente, $1 - \pi(x)$ es la probabilidad condicional de que y sea igual a 0 dado x . De manera que, el valor de la variable de respuesta, y , dada la variable independiente x está determinado por:

$$y = \pi(x) + \varepsilon$$

Donde ε es el error o perturbación estocástica y al igual que la variable dependiente sólo asume dos valores, como se muestra a continuación:

$$\varepsilon = \begin{cases} 1 - \pi(x) & \text{con probabilidad } \pi(x), \text{ cuando } y = 1 \\ -\pi(x) & \text{con probabilidad } 1 - \pi(x), \text{ cuando } y = 0 \end{cases}$$

Como consecuencia, ε sigue una distribución con media cero y varianza $\pi(x)[1 - \pi(x)]$. Mientras que la distribución condicional de la variable dependiente es una binomial con media: $E[y | x] = \pi(x)$ y varianza: $Var[y | x] = \pi(x)[1 - \pi(x)]$.

Entre las características más importantes del modelo de regresión logística destacan las siguientes:

1. Es un modelo no lineal
2. La variable dependiente al igual que ε no siguen una distribución Normal.
3. Existe heterocedasticidad, es decir, la varianza para cada error, ε_i , no es la misma.

2.6.1 Estimación de los parámetros β_0 y β_1

Como recordará en los modelos de regresión lineal simple y múltiple los parámetros se estimaron utilizando el método de mínimos cuadrados ordinarios, y se señaló que también podía utilizarse el método de máxima verosimilitud, en el caso del modelo de regresión logística, no se puede aplicar el método de mínimos cuadrados ordinarios debido al problema de heterocedasticidad que se presenta. Entre los métodos de estimación más utilizados para este modelo, se encuentran el de mínimos cuadrados ponderados y el de máxima verosimilitud, el cual será analizado en esta sección.

Suponga que se tienen una muestra de n observaciones independientes: (x_1, y_1) , $(x_2, y_2), \dots, (x_n, y_n)$, considerando las características de la variable dependiente, su respectiva función de densidad de probabilidad puede expresarse de la siguiente forma:

$$f(y_i) = \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

esta función representa la contribución del par (x_i, y_i) a la función de verosimilitud. Como las n observaciones de la muestra son independientes, entonces, la función de densidad de probabilidad conjunta de la muestra es el producto de las n funciones de densidad de probabilidad $f(y_i)$. La expresión obtenida, es una función de los parámetros desconocidos considerando que los elementos muestrales (x_i, y_i) son conocidos y es denominada la función de verosimilitud, la cual se expresa de la siguiente forma:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \dots \quad 32$$

El método de máxima verosimilitud, consiste en escoger como estimadores aquellos valores de β_0 y β_1 que maximicen la probabilidad de obtener la muestra actual. Es decir, procede a maximizar la función de verosimilitud dada por 32 (para facilitar los cálculos, a menudo se trabaja con el logaritmo de la función de verosimilitud), para ello primero se obtienen las primeras derivadas del logaritmo de $l(\beta)$ con respecto a β_0 y β_1 , y se igualan a cero, las ecuaciones resultantes son denominadas ecuaciones de verosimilitud y son las siguientes:

$$\begin{aligned}\sum [y_i - \pi(x_i)] &= 0 \\ \sum x_i [y_i - \pi(x_i)] &= 0\end{aligned}$$

Finalmente, los estimadores de β_0 y β_1 se obtienen al resolver este sistema de ecuaciones, debido a que este sistema es no lineal en los parámetros β_0 y β_1 , para resolverlo se debe emplear métodos numéricos. Los estimadores de β_0 y β_1 obtenidos son denotados por $\hat{\beta}_0$ y $\hat{\beta}_1$.

El estimador de máxima verosimilitud del logit es: $\hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$, mientras que el estimador de $\pi(x)$ es el siguiente:

$$\hat{\pi}(x) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x}}$$

Se debe tener cuidado con la interpretación del valor ajustado, $\hat{\pi}(x)$, ya que no representa una probabilidad a nivel individual, por ejemplo, en el caso de la compra de la casa, $\hat{\pi}(x_i)$ no representa la probabilidad de que el sujeto i con un nivel de ingreso x_i compre la casa. La interpretación correcta es la siguiente: $\hat{\pi}(x)$ es un estimador de la media, es decir, representa la proporción de sujetos con la característica x en la población muestreada, para los cuales $y = 1$, en el ejemplo de la compra de la casa, $\hat{\pi}(x_i)$ representa la proporción de personas con un nivel de ingresos x_i que comprarían la casa.

Los estimadores de las varianzas y covarianzas de los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ son aproximados utilizando la matriz de información estimada, que en este caso es la siguiente:

$$\hat{I}(\hat{\beta}_0, \hat{\beta}_1) = \begin{bmatrix} \sum_{i=1}^n \hat{\pi}_i (1 - \hat{\pi}_i) & \sum_{i=1}^n x_i \hat{\pi}_i (1 - \hat{\pi}_i) \\ \sum_{i=1}^n x_i \hat{\pi}_i (1 - \hat{\pi}_i) & \sum_{i=1}^n x_i^2 \hat{\pi}_i (1 - \hat{\pi}_i) \end{bmatrix}$$

El estimador de la matriz de varianzas y covarianzas de $\hat{\beta}_0$ y $\hat{\beta}_1$ está dado por la inversa de la matriz de información estimada, es decir: $\hat{V}(\hat{\beta}_0, \hat{\beta}_1) = [\hat{I}(\hat{\beta}_0, \hat{\beta}_1)]^{-1}$. El elemento

(1,1) es el estimador de la varianza de $\hat{\beta}_0$, denotado por $\hat{V}(\hat{\beta}_0)$, el elemento (2,2) es el estimador de la varianza de $\hat{\beta}_1$, denotado por $\hat{V}(\hat{\beta}_1)$ y el estimador de la covarianza de $\hat{\beta}_0$ y $\hat{\beta}_1$ denotado por $\hat{Cov}(\hat{\beta}_0, \hat{\beta}_1)$ está dada por el elemento (1,2).

A su vez, el estimador de la varianza del logit estimado, $\hat{g}(x)$ está dado por:

$$\hat{V}[\hat{g}(x)] = \hat{V}(\hat{\beta}_0) + x^2 \hat{V}(\hat{\beta}_1) + 2x \hat{Cov}(\hat{\beta}_0, \hat{\beta}_1)$$

2.6.2 Prueba de Hipótesis

Entre las pruebas más utilizadas para determinar la significancia de la variable independiente en el modelo, se encuentran, la prueba del radio de verosimilitud y la prueba de Wald, las cuales serán descritas a continuación.

Las hipótesis nula y alternativa relacionadas con la prueba de significancia de la variable independiente son las siguientes:

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

A. Prueba del Radio de Verosimilitud

Esta prueba utiliza el siguiente estadístico de prueba:

$$G = D(\text{modelo sin la variable}) - D(\text{modelo con la variable})$$

Donde:

$$D(\text{modelo sin la variable}) = -2 \ln \left[\frac{\text{verosimilitud}_{\text{del_modelo_sin_la_var_x}}}{\text{verosimilitud}_{\text{del_modelo_saturado}}} \right]$$

$$D(\text{modelo con la variable}) = -2 \ln \left[\frac{\text{verosimilitud}_{\text{del_modelo_con_la_var_x}}}{\text{verosimilitud}_{\text{del_modelo_saturado}}} \right]$$

Como puede notar, el estadístico G, determina el cambio en D debido a la inclusión de la variable independiente, x, y se puede describir de la siguiente forma:

$$G = -2 \ln \left[\frac{\text{verosimilitud_del_modelo_sin_la_var_x}}{\text{verosimilitud_del_modelo_con_la_var_x}} \right]$$

Cuando el modelo no incluye a la variable independiente, x , es decir,

$$\pi(x) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$$

el estimador de máxima verosimilitud de β_0 es: $\hat{\beta}_0 = \ln\left(\frac{n_1}{n_0}\right)$; donde $n_1 = \sum y_i$ y

$n_0 = \sum (1 - y_i)$, mientras que el estimador de máxima verosimilitud de $\pi(x)$ es:

$\hat{\pi}(x) = \frac{n_1}{n}$, el cual es un valor constante. Tomando en cuenta esto, G , puede simplificarse

y queda de la siguiente forma:

$$G = 2 \left[\sum_{i=1}^n [y_i \ln(\hat{\pi}_i) + (1 - y_i) \ln(1 - \hat{\pi}_i)] - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)] \right]$$

Este estadístico de prueba bajo la hipótesis nula y con un tamaño muestral grande, sigue una distribución Chi – Cuadrada con 1 grado de libertad, χ_1^2 .

Entonces si $G \geq \chi_{1,\alpha}^2$ concluimos que existe evidencia estadística para rechazar la hipótesis nula con un nivel de significancia α .

B. Prueba de Wald

El estadístico de prueba Wald es el siguiente:

$$W = \frac{\hat{\beta}_1}{\hat{se}(\hat{\beta}_1)}$$

donde el denominador es el estimador de la desviación estándar de $\hat{\beta}_1$, que es la raíz cuadrada del estimador de la varianza de $\hat{\beta}_1$, dado por $\hat{V}(\hat{\beta}_1)$. Este estadístico de prueba sigue una distribución normal estándar si H_0 es verdadera y el tamaño muestral es grande.

Entonces, si $|W| \geq z_{\alpha/2}$, concluimos que existe evidencia estadística para rechazar la hipótesis nula, con un nivel de significancia α .

Cabe señalar que, no rechazar H_0 implica que la variable independiente puede ser eliminada del modelo, es decir, no ayuda a explicar a la variable dependiente.

2.6.3 Estimación de intervalos de confianza

La estimación de intervalos de confianza para los parámetros β_0 y β_1 , el logit $g(x)$ y $\pi(x)$ utiliza el respectivo estadístico de Wald, definido en la sección anterior.

Los límites superior e inferior del intervalo de confianza al $(1 - \alpha)100\%$ para β_0 están dados por: $\hat{\beta}_0 \pm z_{\alpha/2} \hat{se}(\hat{\beta}_0)$.

De igual forma, los límites superior e inferior del intervalo de confianza al $(1 - \alpha)100\%$ para β_1 están dados por: $\hat{\beta}_1 \pm z_{\alpha/2} \hat{se}(\hat{\beta}_1)$.

Mientras que para el logit, $g(x)$, los límites superior e inferior del respectivo intervalo de confianza al $(1 - \alpha)100\%$ son: $\hat{g}(x) \pm z_{\alpha/2} \hat{se}(\hat{g}(x))$.

Un aspecto que no se debe perder de vista es que, se está utilizando el cuartil $z_{\alpha/2}$ de la distribución normal estándar y que estos intervalos de confianza son válidos cuando el tamaño muestral es grande.

Para obtener un intervalo de confianza al $(1 - \alpha)100\%$ para $\pi(x)$ evaluado en un punto específico, se utiliza el intervalo de confianza calculado para el logit $g(x)$ evaluado en el mismo punto, los límites superior e inferior del intervalo, se obtienen mediante la siguiente expresión:

$$\frac{e^{\hat{g}(x) \pm z_{\alpha/2} \hat{se}(\hat{g}(x))}}{1 + e^{\hat{g}(x) \pm z_{\alpha/2} \hat{se}(\hat{g}(x))}}$$

2.7 Aplicación: Quiebra de las compañías noruegas

En este estudio, se aplica el modelo de regresión logística para examinar el impacto de las variables financieras y las características de la firma, tales como la edad, tamaño y clasificación de la industria, en la determinación de la firma.

La quiebra de una compañía sucede cada vez que el flujo de efectivo realizado (en un sólo periodo) o el flujo de efectivo realizado más el valor del flujo de efectivo futuro esperado (multiperiodo) es inferior al tamaño de las obligaciones de deudas.

La información muestral consta de datos de incumplimientos y de la contabilidad de compañías noruegas en el periodo 1995 – 1999, que fueron recolectados por Dun & Bradstreet., de los cuales, se removieron todas las firmas del sector público y financiero porque no tuvieron quiebras.

El modelo logit se utilizó para estimar las probabilidades de quiebra. En este modelo, los dos estados de la variable dependiente, son: Estado 1 = Quiebra, Estado 2 = No quiebra. Las variables dependientes o explicativas son ratios de variables financieras, relacionadas con las propiedades del flujo de efectivo en combinación con las obligaciones de deuda y el valor de la firma después del periodo actual, y son las siguientes:

- CASHDEBT:= Flujo de efectivo dividido entre la deuda total. Da una medida directa de la afluencia de efectivo en relación de la deuda total de la compañía.
- FINANCOV:= Protección financiera, que es medida como los resultados operacionales dividido entre los costos financieros.
- LIQUIDEZ:= Es medida como el activo actual entre la deuda actual.
- SOLIDEZ:= Es medida como el valor líquido entre el capital total.

Además de estas variables, se agregaron la Edad, el Tamaño de la compañía y las siguientes variables dummy relacionadas con la clasificación de la industria y la ubicación geográfica:

- REALSERV:= Variable dummy que indica si la compañía pertenece o no a la industria de bienes raíces y servicios.
- HOTRESTA:= Variable dummy que indica si la compañía pertenece o no a la industria de hoteles y restaurantes.
- MNORWAY:= Variable dummy que indica si la compañía se ubica en Noruega media.
- NNORWAY:= Variable dummy que indica si la compañía se ubica en el norte de Noruega.

La muestra disponible fue dividida en dos, una parte para el proceso de estimación y la otra para pronosticar. La siguiente tabla, muestra los resultados obtenidos al efectuar la regresión logística:

Parámetro	Estimador	Desv. Estándar
CASHDEBT	-1.7698	0.1679
LOG(TAMAÑO)	-0.1204	0.0289
LIQUIDEZ	-0.2775	0.0480
FINANCOV	-0.00891	0.00238
SOLIDEZ	-0.5807	0.0564
EDAD	-0.3232	0.0114
REALSERV	-0.9530	0.1094
HOTRESTA	0.2455	0.1337
MNORWAY	0.3631	0.1208
NNORWAY	0.2470	0.1075

Como puede notar, los ratios financieros y las variables EDAD y LOG(TAMAÑO) son negativas y significativamente diferentes de cero, al menos para un nivel de significancia del 5%. Las variables REALSERV y las relacionadas con la ubicación geográfica son estadísticamente significantes al nivel del 5%, mientras que la variable HOTRESTA no lo es.

Este tipo de modelo es útil para poder tomar una decisión, en cuanto a invertir o no invertir en determinada compañía en función de la probabilidad de quiebra estimada.

3. Paradigma Bayesiano

3.1 Introducción

Existen situaciones en las cuales los problemas a analizar cuentan con una muestra de datos muy pequeña y se puede explotar la experiencia del experto en el área y/o existe información de algún estudio previo, en este caso resultaría útil la incorporación de ambas fuentes de información al modelo, desafortunadamente la metodología clásica sólo nos permite considerar la información muestral, por lo que es necesario contemplar una metodología alternativa que nos permita obtener un buen análisis del modelo utilizando tanto la información muestral como la del experto, este es el caso de la Metodología Bayesiana. Una de las ventajas que ofrece es trabajar con muestras pequeñas, lo que resulta conveniente en el tratamiento de la información económica ya que al tomar una muestra de datos históricos grande, se está asumiendo que las condiciones en las que se dieron esos datos son las mismas que las actuales y eso no sucede.

En el presente capítulo se analizarán las ventajas y desventajas que ofrece la metodología bayesiana y posteriormente se presentarán los conceptos básicos que se utilizarán en el siguiente capítulo.

3.2 Crítica a los Modelos Clásicos

Los modelos econométricos clásicos han sido ampliamente utilizados, pese a la problemática que presentan, esto no quiere decir, que sean malos sino que simplemente se debe tener conocimiento de las desventajas o problemas que presenta la metodología que se está utilizando.

En algunos problemas, resulta difícil obtener muchos datos y los que se disponen resultan insuficientes para que los estimadores de los parámetros tengan las propiedades muestrales deseables, tal es el caso de los estimadores de máxima verosimilitud, los cuales tienen buenas propiedades muestrales siempre que el tamaño de la muestra sea grande, en caso contrario, resulta conveniente utilizar otro método de estimación. Otro aspecto interesante es que, las propiedades muestrales de los estimadores, son justificadas en términos de su comportamiento en muestras repetidas no observadas, tal interpretación resulta poco adecuada si consideramos que la Economía es una ciencia no experimental y de la escasez de los datos para algunos problemas, esto no quiere decir, que tales propiedades muestrales no sean importantes, sino que se necesita de un procedimiento que se justifique utilizando la información actual que es la que nos interesa.

De igual forma, existen problemas de interpretación en los intervalos de confianza al $(1 - \alpha)100\%$, muchos afirman que el intervalo calculado contiene al parámetro con una probabilidad de $(1 - \alpha)$, lo cual es falso, en realidad lo que indica es que si se tuvieran otras muestras similares a la actual bajo las mismas condiciones y para cada una de ellas se calculara el respectivo intervalo de confianza al $(1 - \alpha)100\%$, entonces el $(1 - \alpha)100\%$ de estos intervalos contendrá al parámetro en cuestión. Una vez que el intervalo de confianza es calculado, contendrá al parámetro con una probabilidad de 1 ó 0.

Actualmente, existen un gran número de paquetes computacionales estadísticos enfocados a facilitar los cálculos en el análisis, desafortunadamente muchas personas han abusado de este hecho, un claro ejemplo es cuando se efectúa una prueba de hipótesis, existen personas que se limitan a obtener el p - value de la tabla final que muestran estos paquetes para determinar si se rechaza o no la hipótesis nula, sin ni siquiera conocer la definición de p - value. Esta acción le resta importancia a la metodología que existe detrás

del procedimiento de prueba de hipótesis y la limita a rechazar o no la hipótesis nula en función del p – value. Lo que deseo resaltar es la importancia de los conceptos estadísticos y el conocimiento que se debe tener de los principales procedimientos de inferencia, para poder hacer un uso correcto de los resultados que nos brindan los paquetes estadísticos, esto nos ayudará a tener un panorama más claro y considerar otras opciones, ya que saber de Estadística no se limita a saber utilizar un paquete estadístico.

Por otra parte, ¿qué pasa cuando aparte de la información muestral, existe información acerca de los valores que pueden tomar los parámetros, la cual proviene de consideraciones teóricas, estudios previos o de la experiencia del investigador?, ¿se debe ignorar este tipo de información o cómo debe incorporarse al análisis? Existen algunos métodos que permiten incorporar de manera informal parte de esta información, desafortunadamente, no son muy conocidos por la mayoría, lo cual limita su uso, como consecuencia de esto, algunas veces tal información no es incorporada en el análisis.

Como puede notar, sólo se han mencionado algunas problemáticas que presenta la metodología clásica así como algunos malos usos que se hacen de ella, el objetivo de esto es resaltar la importancia de conocer cómo funciona la metodología, los supuestos que la sustentan así como los problemas y limitaciones que presenta y en función de esto determinar la calidad de los resultados obtenidos.

3.3 Filosofía de la Ciencia

¿La Ciencia es objetiva o subjetiva? Algunos consideran que la Ciencia debe ser objetiva, es decir, el investigador debe observar al mundo sin dejarse influenciar por sentimientos, creencias y opiniones, sin embargo, esto no concuerda con la realidad en algunos campos de la ciencia, tales como la Estadística donde el subjetivismo está incluido de alguna forma en el análisis, como lo muestran los siguientes ejemplos:

1. En el procedimiento clásico de prueba de hipótesis, se elige subjetivamente el nivel de significancia para rechazar la hipótesis nula, entonces, si el investigador de acuerdo a su experiencia cree que H_0 es verdadera, elegirá un nivel de significancia que soporte su creencia.

2. Cuando en el conjunto de datos disponibles existen algunas observaciones ya sea muy grandes o muy pequeñas comparadas con las demás, en ocasiones son excluidas del análisis por razones subjetivas (ya sea porque son consideradas errores o que las condiciones experimentales para esos datos fueron diferentes).

De acuerdo con lo anterior, no podemos afirmar que la Ciencia sea completamente objetiva, ya que el subjetivismo ya está instituido de alguna manera en ella. Entonces, lo importante es encontrar una metodología que nos permita incorporar de manera formal al subjetivismo en el análisis, la metodología bayesiana cumple con esta característica ya que nos permite combinar la experiencia del investigador con la información muestral para actualizar nuestro conocimiento.

Zellner [1988], considera los siguientes postulados de vital importancia en la Econometría y otras áreas de la Ciencia, los cuales son cumplidos por la metodología bayesiana:

1. “El principio de unidad de la Ciencia es válido”, esto quiere decir, que la unidad de la Ciencia radica en su método no en su material u objeto de estudio, lo que implica que, los principios de inferencia: estimación, prueba de hipótesis, predicción entre otros, se deben aplicar de la misma forma en ciencias experimentales y no experimentales. Y en efecto, la inferencia que se realiza en la investigación económica no difiere de la que se efectúa en alguna otra área de la ciencia.
2. “El postulado de Simplicidad de Jeffreys – Wrinch es un principio de trabajo muy útil en Econometría aplicada y otras ciencias”. Este postulado nos dice, que las leyes más simples son las que tienen las probabilidades a priori más altas. Una vez que se han asignado las probabilidades a priori a los modelos, éstos deben ser ordenados de forma decreciente con respecto a la probabilidad asignada y que además, se empezaría probando con el modelo más sencillo y posteriormente se va incrementando el nivel de complejidad.
3. “El principio de predicción cuyo desempeño predictivo es central en la evaluación de hipótesis y modelos es de vital importancia en Econometría y

otras ciencias”. El principio de predicción consiste en predecir experiencias futuras a partir de las experiencias pasadas, y resulta muy útil ya que tales predicciones pueden influir en la toma de decisiones o en la evaluación del desempeño del modelo.

4. “Un concepto subjetivo de probabilidad es más útil en la investigación y desarrollo de aplicaciones en Econometría y otras ciencias, que otros conceptos de probabilidad”. Los Bayesianos interpretan a la probabilidad como una cuantificación del grado razonable de confianza que se tiene en una proposición, esta definición resulta más útil y familiar para el investigador. En la práctica, la mayor parte de los estadísticos utilizan el concepto subjetivo de probabilidad, aunque a veces no se percaten de ello, por ejemplo al interpretar un intervalo de confianza al $(1 - \alpha)100\%$, muchos dicen que el intervalo contiene al parámetro con una probabilidad de $(1 - \alpha)100\%$ lo cual difiere de su verdadera interpretación como se observó en el capítulo anterior.
5. “El modelo de aprendizaje involucrado en el Teorema de Bayes es muy útil en Econometría y otras ciencias”. El teorema de Bayes nos provee de un mecanismo formal para incorporar la información proveniente de la experiencia del investigador, de estudios previos o de consideraciones teóricas en el análisis, y combinarlo con la información muestral disponible, de esta forma actualizamos el conocimiento que se tiene acerca de los parámetros.

3.4 Paradigma Bayesiano

Cuando se analiza un modelo econométrico, a menudo se cuenta con dos fuentes de información: una muestra de datos históricos y la experiencia de un experto o datos de algún experimento previo, a esta segunda fuente, se le denomina información a priori. El paradigma Bayesiano nos ofrece una metodología que da la opción de incorporar o no la

información a priori disponible acerca de los parámetros del modelo al análisis mediante el empleo de la distribución a priori.

El principal sustento de esta metodología consiste en definir a la probabilidad como un grado razonable de creencia en una proposición, además de que el término aleatoriedad es asociado a la incertidumbre, es decir, algo es aleatorio no porque sea resultado de un experimento aleatorio sino porque es desconocido, por lo que la palabra aleatoria se utiliza para representar nuestra ignorancia acerca del fenómeno. Esta interpretación subjetiva de la probabilidad se adapta muy bien a una gran diversidad de problemas y facilita la interpretación de los resultados.

Por lo que se acaba de mencionar, puede deducir que existen diferencias entre este enfoque y el clásico, algunas de ellas se resumen a continuación:

1. **Inferencia:** La metodología clásica se considera totalmente objetiva puesto que el análisis sólo depende de los datos, mientras que la bayesiana es tanto objetiva como subjetiva ya que además de depender de los datos, depende de la información a priori disponible acerca de los parámetros del modelo.
2. **Parámetros:** La metodología los considera cantidades desconocidas fijas o constantes. Debido al concepto de probabilidad del enfoque Bayesiano, los parámetros se consideran variables aleatorias y se les puede asociar una distribución de probabilidad.
3. **Datos:** En la metodología clásica son variables aleatorias debido a que se consideran diferentes de ensayo a ensayo. En la metodología bayesiana los datos son aleatorios antes de observarlos y una vez observados son considerados fijos o constantes.
4. En la metodología clásica siempre se condiciona en los parámetros desconocidos, en este caso la pregunta a responder es ¿Dado el valor del parámetro qué conocemos acerca de los datos? En la metodología bayesiana como no se puede condicionar en algo que se desconoce entonces se condiciona en los datos, con esto se pretende responder la pregunta ¿Dados los datos qué conocemos acerca de los parámetros?

5. El motor de la metodología bayesiana es el Teorema de Bayes mediante el cual se combina la información muestral y la información a priori para dar origen a la distribución a posteriori a partir de la cual se realiza la inferencia estadística (estimación puntual, por intervalos pruebas de hipótesis, entre otras).

Cabe señalar que en las siguientes secciones se mencionarán otras diferencias entre ambos enfoques. A continuación se definirán los elementos del Análisis Bayesiano:

1. **Distribución a priori:** Captura la información a priori disponible acerca de los parámetros del modelo antes de recoger o ver los datos y se denota por $p(\theta | I_0)$ donde I_0 es la información a priori o inicial y θ el parámetro, aunque para efectos de simplificar la notación se suele escribir como $p(\theta)$, ésta será la notación que se utilizará en lo sucesivo. La información a priori puede provenir de la experiencia de un experto en el área, de consideraciones teóricas o de datos de algún experimento previo.
2. **Función de Verosimilitud:** Conocida también como el proceso generador de los datos y resume la información contenida en los datos observados, se denota como $p(y | \theta)$ donde y representa a los datos.
3. **Distribución a posteriori:** Resume el conocimiento que se tiene acerca de los parámetros después de haber observado los datos. Es el resultado de combinar la información a priori con la muestral, de manera que puede considerarse como una regla de actualización, donde los datos nos permiten actualizar nuestro conocimiento a priori acerca de los parámetros. Se denota por: $p(\theta | y)$.

3.4.1 Revisión del Teorema de Bayes

El Teorema de Bayes representa el corazón del Paradigma Bayesiano y nos provee de un mecanismo para actualizar el conocimiento que se tiene acerca del parámetro θ .

Sea $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ un vector de n observaciones con distribución de probabilidad $p(\mathbf{y} | \boldsymbol{\theta})$ que depende de $\boldsymbol{\theta}$, además considere un vector de k parámetros $\boldsymbol{\theta}' = (\theta_1, \theta_2, \dots, \theta_k)$ con distribución de probabilidad $p(\boldsymbol{\theta})$, puesto que como ya se mencionó los parámetros se consideran variables aleatorias. Entonces la distribución de probabilidad conjunta de \mathbf{y} y $\boldsymbol{\theta}$ está dada por la siguiente expresión:

$$p(\mathbf{y}, \boldsymbol{\theta}) = p(\boldsymbol{\theta})p(\mathbf{y} | \boldsymbol{\theta}) = p(\boldsymbol{\theta})p(\boldsymbol{\theta} | \mathbf{y})$$

El Teorema de Bayes afirma que la distribución condicional de $\boldsymbol{\theta}$ dadas las observaciones está dada por:

$$p(\boldsymbol{\theta} | \mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y} | \boldsymbol{\theta})}{p(\mathbf{y})}; \quad p(\mathbf{y}) > 0$$

donde $p(\mathbf{y})$ es la distribución marginal de \mathbf{y} , es una cantidad que no depende de $\boldsymbol{\theta}$ y es conocida como constante de normalización. Se puede calcular de la siguiente forma:

$$p(\mathbf{y}) = \begin{cases} \int p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} & \boldsymbol{\theta} \text{ continuo} \\ \sum p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta}) & \boldsymbol{\theta} \text{ discreto} \end{cases}$$

esta constante de normalización puede ser omitida y entonces el Teorema de Bayes puede reescribirse de la siguiente manera:

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{y} | \boldsymbol{\theta}) \dots \text{ 33}$$

donde:

$p(\boldsymbol{\theta})$: Es la distribución a priori de $\boldsymbol{\theta}$, representa lo que se sabe acerca de $\boldsymbol{\theta}$ previo al conocimiento de los datos.

$p(\mathbf{y} | \boldsymbol{\theta})$: Es la función de verosimilitud y representa la información acerca de $\boldsymbol{\theta}$ contenida en los datos. Es una función a través de la cual los datos modifican el conocimiento que se tiene acerca de $\boldsymbol{\theta}$.

$p(\boldsymbol{\theta} | \mathbf{y})$: Es la distribución a posteriori y representa lo que se sabe acerca de $\boldsymbol{\theta}$ después del conocimiento de los datos.

Además el símbolo \propto significa proporcionalidad, por lo que la expresión 33 significa que la distribución a posteriori es proporcional a la distribución a priori veces la función de verosimilitud, es decir:

$$\text{distribución a posteriori} \propto \text{distribución a priori} \times \text{función de verosimilitud}$$

La distribución a posteriori juega un papel central en el paradigma Bayesiano puesto que todos los procedimientos de inferencia, por ejemplo, estimación puntual, por intervalos, pruebas de hipótesis, entre otros, se basan en ella.

El Teorema de Bayes puede considerarse como un proceso de aprendizaje de la experiencia, puesto que nos permite actualizar continuamente el conocimiento que se tiene acerca de los parámetros a medida que se tienen nuevas observaciones. Cuando no se tienen observaciones disponibles y sólo se cuenta con información previa, entonces para realizar inferencias acerca de los parámetros, se utiliza la distribución a priori. En el caso de que se disponga tanto de la información muestral y de la experiencia previa, entonces se utiliza el Teorema de Bayes para combinarlas y obtener la distribución a posteriori.

La actualización del conocimiento acerca del parámetro θ conforme se dispone de nueva información muestral, se realiza de la siguiente forma:

Suponga que tiene una muestra inicial \mathbf{y}_1 , entonces la distribución a posteriori correspondiente es: $p(\boldsymbol{\theta} | \mathbf{y}_1) \propto p(\boldsymbol{\theta})p(\mathbf{y}_1 | \boldsymbol{\theta})$. Posteriormente, se toma una segunda muestra \mathbf{y}_2 , que se distribuye de manera independiente de la primera muestra, entonces, ahora la distribución a priori será $p(\boldsymbol{\theta} | \mathbf{y}_1)$, la distribución a posteriori de la primera muestra, es decir: $p(\boldsymbol{\theta} | \mathbf{y}_1, \mathbf{y}_2) \propto p(\boldsymbol{\theta} | \mathbf{y}_1)p(\mathbf{y}_2 | \boldsymbol{\theta})$. Entonces, para la m – ésima muestra obtenida que se distribuye independientemente de las anteriores, la correspondiente distribución a posteriori será la siguiente:

$$p(\boldsymbol{\theta} | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m) \propto p(\boldsymbol{\theta} | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{m-1})p(\mathbf{y}_m | \boldsymbol{\theta})$$

donde: $p(\boldsymbol{\theta} | \mathbf{y}_1) \propto p(\boldsymbol{\theta})p(\mathbf{y}_1 | \boldsymbol{\theta})$

3.4.2 Ventajas y Desventajas del Paradigma Bayesiano

Una vez que se conoce de manera general cómo opera la metodología bayesiana, resulta útil conocer las ventajas y desventajas que ofrece, algunas de las cuales se mencionarán a continuación.

Entre algunas de las ventajas que ofrece la metodología bayesiana, se encuentran las siguientes:

1. Provee un mecanismo para formalizar el proceso de aprendizaje de los datos, mediante el cual se actualiza el conocimiento que se tiene acerca de los parámetros del modelo, incorporando la nueva información.
2. Permite incorporar la información a priori que se tiene acerca de los parámetros del modelo, proveniente ya sea de estudios previos o de la experiencia de un experto en el área, de manera flexible en los procedimientos de prueba de hipótesis, estimación (puntual y por intervalos), predicción, entre otros.
3. Puede trabajar con muestras pequeñas. Si en un principio sólo se dispone de la información a priori, entonces las inferencias acerca de los parámetros del modelo, se realizan con respecto a la distribución a priori, y conforme se dispone de información muestral, ésta se incluye en el modelo y de esta forma el modelo va mejorando considerando la nueva información.
4. Ofrece procedimientos alternativos al de la metodología clásica para prueba de hipótesis y estimación por intervalos basados en supuestos de normalidad.
5. Permite asociar probabilidades a las hipótesis y provee herramientas que permiten actualizar tales probabilidades conforme se disponga de nueva información.
6. Facilita la interpretación de los resultados obtenidos en los procedimientos de prueba de hipótesis y estimación por intervalos, ya que corresponde a la interpretación bajo el sentido común del investigador.

Entre algunas de las desventajas de la metodología bayesiana se encuentran:

1. Se pueden presentar problemas operacionales en el cálculo de la distribución a posteriori cuando se trabaja en más de tres dimensiones, aunque afortunadamente se han desarrollado métodos numéricos estadísticos que combinados con la computadora logran resolver esta dificultad.
2. La elección o construcción de una distribución a priori en algunos casos resulta ser una tarea no muy fácil de resolver.

3.5 Distribución a priori

Como ya se mencionó, la distribución a priori representa la información a priori que se tiene acerca de los parámetros del modelo, esta distribución no depende de la muestra de datos actuales con los que se está trabajando.

Una de las tareas antes de efectuar un análisis Bayesiano, consiste en determinar la distribución a priori de los parámetros de acuerdo a la información existente referente a ellos. En esta sección se mencionarán algunos métodos para construir una distribución a priori, para mayor información le sugiero consulte Robert [2001].

3.5.1 Distribución a priori basada en datos

Esta distribución a priori recibe este nombre debido a que para obtenerla, se utiliza la información contenida en muestras de datos históricos provenientes de algún estudio previo.

Suponga que tiene una serie de experimentos independientes, para el primer experimento cuenta con una muestra de datos \mathbf{y}_1 , entonces, la distribución a posteriori es la siguiente:

$$p(\boldsymbol{\theta} | \mathbf{y}_1) \propto p(\boldsymbol{\theta})p(\mathbf{y}_1 | \boldsymbol{\theta})$$

para el segundo experimento, tiene una muestra de datos \mathbf{y}_2 que se distribuye de forma independiente de la primera muestra, entonces, la distribución a posteriori es:

$$p(\boldsymbol{\theta} | \mathbf{y}_1, \mathbf{y}_2) \propto p(\boldsymbol{\theta})p(\mathbf{y}_1 | \boldsymbol{\theta})p(\mathbf{y}_2 | \boldsymbol{\theta})$$

$$\propto p(\boldsymbol{\theta} | \mathbf{y}_1) p(\mathbf{y}_2 | \boldsymbol{\theta})$$

Para el n -ésimo experimento, se tiene una muestra de datos \mathbf{y}_n que se distribuye de forma independiente de las anteriores, entonces la distribución a posteriori es la siguiente:

$$p(\boldsymbol{\theta} | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m) \propto p(\boldsymbol{\theta} | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{m-1}) p(\mathbf{y}_m | \boldsymbol{\theta})$$

Como podrá notar, se cumple que: la distribución posterior de experimentos previos = distribución a priori del siguiente experimento.

De aquí su denominación de distribución a priori basada en datos. En caso de que no se disponga de datos de algún experimento previo, entonces se utiliza la información proporcionada por expertos en el área, a este tipo de distribución a priori se le denomina no basada en datos.

3.5.2 Aproximación del histograma

Este método resulta útil cuando el espacio paramétrico Θ consiste en un intervalo de la línea real, y procede de la siguiente forma:

1. Divide a Θ en intervalos.
2. Asigna las probabilidades de manera subjetiva a cada intervalo.
3. Construye el correspondiente histograma de probabilidad y obtiene una aproximación de la distribución a priori

Entre los principales inconvenientes que presenta este método resalta la no existencia de una regla que determine cuántos intervalos y de qué tamaño deben utilizarse para obtener una buena aproximación de la distribución a priori.

3.5.3 Aproximación de la Verosimilitud relativa

Al igual que la aproximación anterior, es de mayor utilidad cuando el espacio paramétrico Θ consiste en un intervalo de la línea real, y opera de la siguiente forma:

1. Selecciona elementos significativos de Θ .
2. Compara y asigna las verosimilitudes de los elementos seleccionados.

3. Grafica los elementos con sus respectivas verosimilitudes y traza directamente una densidad a priori.

Cabe señalar que en ocasiones la distribución a priori resultante es impropia, es decir, su integral no es uno. En este caso, se puede encontrar una constante tal que la distribución a priori sea igual a $cp(\theta)$ la cual constituye una densidad propia.

3.5.4 Elección de una forma funcional dada

Este método consiste en asumir que la distribución a priori tiene una forma funcional dada o conocida, entonces, se debe escoger una densidad que se aproxime lo más que se pueda a la información o creencias que se tiene a priori, posteriormente, se deben especificar el valor de los parámetros de la densidad elegida.

Uno de los métodos más utilizados para determinar el valor de los parámetros consiste en estimar subjetivamente algunos fractiles de la distribución a priori y posteriormente, escoger los parámetros de la forma funcional especificada, de manera que se obtenga la distribución a priori que mejor se aproxime a los fractiles estimados.

3.5.5 Distribución a priori conjugada

Una distribución a priori conjugada, es aquella que combinada con la función de verosimilitud da como resultado una distribución a posteriori que es de la misma clase que la distribución a priori, una definición más formal es la siguiente:

Sea $q(\theta)$ una distribución a priori que pertenece a la familia de distribuciones Q, la muestra tiene una distribución que pertenece a F, y si la distribución posterior resultante pertenece también a la familia Q, entonces, se dice que Q es una familia de distribuciones a priori conjugadas a F.

Optar por esta aproximación, restringe la elección de la distribución a priori a una clase limitada de distribuciones, en este caso, la información a priori disponible se utiliza para determinar el valor de los respectivos hiperparámetros (nombre que reciben los parámetros de la distribución a priori). Como puede notar, este tipo de distribución

preserva la estructura de la distribución a priori y además reduce la complejidad matemática de los procedimientos.

A continuación se presentan algunas de las distribuciones a priori conjugadas.

Función de Verosimilitud	Distribución a priori	Distribución a posteriori
$X \sim \text{Bin}(n, \theta)$	Beta(p, q)	Beta(p + x, q + n - x)
$x_1, \dots, x_n \sim \text{Geométrica}(\theta)$	Beta(p, q)	Beta(p + n, q + $\sum x_i - n$)
$x \sim \text{BinNeg}(n, \theta)$	Beta(p, q)	Beta(p + n, q + x - n)
$X_1, \dots, x_n \sim \text{Poisson}(\theta)$	Gamma(p, q)	Gamma(p + $\sum x_i$, q + n)
$x_1, \dots, x_n \sim \text{Gamma}(k, \theta)$ k conocida	Gamma(p, q)	Gamma(p + nk, q + $\sum x_i$)
$x_1, \dots, x_n \sim \text{Normal}(\theta, \lambda^{-1})$ λ conocida	Normal(b, c^{-1})	Normal $\left(\frac{cb + n\lambda\bar{x}}{c + n\lambda}, \frac{1}{c + n\lambda}\right)$

3.6 Inferencia Estadística

Como ya se mencionó, la distribución a posteriori resume toda la información disponible acerca de los parámetros involucrados en el modelo, por lo que los procedimientos de inferencia se efectúan con respecto a ella.

Cabe señalar que cuando el modelo consta de varios parámetros y se desea efectuar algún procedimiento de inferencia con respecto a un parámetro en específico, entonces, éste se efectuará utilizando su respectiva distribución marginal posterior, la cual se definirá a continuación.

Sea Θ un vector de k parámetros, la distribución marginal posterior del elemento θ_i , denotada por $p(\theta_i | y)$, se obtiene de la siguiente forma:

$$p(\theta_1 | y) = \int_{R\theta_k} \dots \int_{R\theta_2} p(\theta_1, \theta_2, \dots, \theta_k | y) d\theta_2 \dots d\theta_k$$

donde: $p(\theta_1, \theta_2, \dots, \theta_k | y)$ es la distribución a posteriori de los k parámetros y $R\theta_i$ es la región del parámetro θ_i .

Como puede notar, la distribución marginal posterior se obtiene integrando la distribución posterior de los k parámetros con respecto a los parámetros que no son de interés.

En las siguientes secciones se presentan los procedimientos de estimación puntual, por intervalos, prueba de hipótesis y la distribución predictiva.

3.6.1 Estimación puntual

Una opción para obtener un estimador puntual, $\hat{\theta}$, del parámetro θ , consiste en seleccionar un rasgo característico de la distribución posterior, $p(\theta|y)$, como la media, la moda o la mediana.

Cuando la distribución posterior es simétrica, entonces, la media y la mediana serán las mismas, y si además la distribución posterior, es unimodal, entonces, las tres medidas coincidirán. Cuando la distribución posterior, es asimétrica, entonces, se utiliza la mediana como estimador del parámetro θ .

Para resolver el problema de obtener un estimador puntual del parámetro de interés, desde el punto de vista de Teoría de decisiones, es necesario especificar:

1. Una distribución a priori, $p(\theta)$
2. Función de verosimilitud, $p(y|\theta)$
3. Función de pérdida, $L(\theta, \hat{\theta})$, que cuantifica las consecuencias de actuar como si el estimador $\hat{\theta}$ fuera el verdadero valor del parámetro θ .

Para poder obtener un estimador puntual, primero es necesario obtener la pérdida posterior esperada, que se expresa de la siguiente forma:

$$E[L(\theta, \hat{\theta}) | y] = \int_{\Theta} L(\theta, \hat{\theta}) p(\theta | y) d\theta$$

para el caso continuo, en el caso discreto, en lugar de utilizar la integral se utiliza la suma.

De manera que, un estimador puntual del parámetro θ , es aquel valor $\hat{\theta}$ que minimice la pérdida posterior esperada.

Como podrá notar el estimador seleccionado dependerá de la función de pérdida que se utilice, a continuación se presenta un cuadro en el que se presentan los estimadores que se obtienen con algunas de las funciones de pérdida más utilizadas.

	Función de Pérdida	Estimador
Cuadrática	$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$	Media posterior: $E[\theta y]$
Error absoluto	$L(\theta, \hat{\theta}) = \theta - \hat{\theta} $	Mediana posterior
Error absoluto Generalizado	$L(\theta, \hat{\theta}) = \begin{cases} c_1(\hat{\theta} - \theta) & \text{si } \hat{\theta} \geq \theta \\ c_2(\theta - \hat{\theta}) & \text{si } \hat{\theta} < \theta \end{cases}$	Fractil $\frac{c_2}{(c_2 + c_1)}$ de la distribución posterior

3.6.2 Estimación por intervalos

Una vez que se ha obtenido la distribución posterior del parámetro θ , puede resultar útil obtener un intervalo que contenga a θ con determinada probabilidad, en la metodología bayesiana, éste recibe el nombre de intervalo de credibilidad o posterior, a continuación se presenta una definición más formal de dicho intervalo.

Un intervalo $(1 - \alpha)100\%$ creíble para θ , es un subconjunto R de Θ , $R \subset \Theta$, tal que cumple la siguiente condición:

$$p(\theta \in R | y) = \int_R p(\theta | y) d\theta = 1 - \alpha \quad \dots \quad 34$$

para el caso continuo, en el caso del caso discreto, se utiliza la suma en lugar de la integral. Donde α es una constante cuyo valor se encuentra en el intervalo $[0,1]$ y que se determina subjetivamente.

Una aproximación utilizada para obtener un intervalo de credibilidad, consiste en tomar los cuantiles $\alpha/2$ y $(1 - \alpha/2)$ de la distribución posterior, como los límites del intervalo creíble al $(1 - \alpha)100\%$ para θ .

La interpretación de la expresión 34, es la siguiente: dados los datos observados y la información a priori actual, el parámetro θ pertenece a R con probabilidad $1 - \alpha$. Cabe señalar que el intervalo que se obtiene mediante la definición anterior no necesariamente es único.

Cuando la distribución posterior es unimodal, es decir, la moda es única, se puede obtener un intervalo único de tamaño mínimo que incluye a aquellos puntos con la densidad posterior más alta, esto significa que sólo incluye a los valores más probables de θ , este intervalo recibe el nombre de intervalo con la más alta densidad posterior o HPD (Highest Posterior Density). A continuación se presenta una definición más formal de un intervalo HPD.

Un intervalo con la más alta densidad posterior o HPD, es un subconjunto $S=[a,b]$ de Θ , $S \subset \Theta$, que cumple las siguientes condiciones:

1.
$$p(\theta \in S | y) = \int_S p(\theta | y) d\theta = 1 - \alpha$$
2. Para $a \leq \theta \leq b$, la densidad posterior, $p(\theta | y)$, para cada punto del intervalo es más grande que para cualquier punto que pertenezca a otro intervalo con el mismo contenido de probabilidad, $1 - \alpha$.

De todos los intervalos de credibilidad para θ con una probabilidad posterior de $1 - \alpha$, el intervalo HPD es el que tiene el tamaño más pequeño.

Cabe señalar que, cuando la distribución posterior es unimodal y simétrica, entonces el intervalo HPD y el obtenido utilizando los cuantiles $\alpha/2$ y $(1 - \alpha/2)$, con iguales.

Como podrá notar, a diferencia de la interpretación de los intervalos de confianza clásicos, la metodología bayesiana provee un intervalo con una interpretación más natural para el investigador.

3.6.3 Prueba de hipótesis

En la metodología bayesiana, el procedimiento de prueba de hipótesis se puede efectuar calculando ratios de probabilidad posterior, mediante éstos, se decide cuál hipótesis es la más probable considerando los datos y la información a priori disponible. A continuación se definirá el ratio de probabilidad posterior.

Suponga que desea confrontar la hipótesis nula $H_0 : \theta \in \Theta_0$ contra la alternativa $H_1 : \theta \in \Theta_1$, donde $\Theta_0 \cap \Theta_1 = \emptyset$. A cada una se le asigna una probabilidad a priori $p(H_i | I_0)$, $i = 0,1$ tal que $p(H_0 | I_0) + p(H_1 | I_0) = 1$.

El ratio de probabilidad posterior en favor de H_0 se define de la siguiente forma:

$$\frac{p(H_0 | x)}{p(H_1 | x)} = \left[\frac{p(H_0)}{p(H_1)} \right] \left[\frac{p(x | H_0, \theta)}{p(x | H_1, \theta)} \right]$$

De acuerdo a lo anterior, el ratio de probabilidad posterior es igual al ratio de probabilidad a priori veces el ratio de verosimilitudes promediado bajo H_0 y H_1 . De manera que:

Si el ratio de probabilidad posterior > 1 , entonces se acepta H_0 .

Si el ratio de probabilidad posterior < 1 , entonces se rechaza H_0 en favor de H_1 .

Si el ratio de probabilidad posterior $= 1$ entonces existe indecisión.

Como puede notar, esta decisión se centra en el cálculo de las probabilidades actuales de cada hipótesis (las respectivas probabilidades posteriores) considerando las observaciones y la información a priori disponible.

Un enfoque alternativo al ya descrito, consiste en utilizar el Factor de Bayes en lugar del ratio de probabilidad posterior, el cual se define como sigue:

El Factor de Bayes a favor de H_0 , es el radio de probabilidad posterior sobre el radio de probabilidad a priori, es decir:

$$B_{01} = \frac{p(H_0 | x)}{p(H_1 | x)} \bigg/ \frac{p(H_0)}{p(H_1)}$$

Este factor es considerado como una respuesta bayesiana más objetiva debido a que elimina parte de la influencia de la distribución a priori considerada en la distribución posterior y enfatiza el papel de las observaciones en el procedimiento.

Resulta interesante confrontar dos hipótesis simples debido a que se llega a que el Factor de Bayes respectivo es igual al radio de verosimilitud empleado en la metodología clásica, como se verá a continuación:

Se desea confrontar las siguientes hipótesis simples:

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta = \theta_1$$

la probabilidad posterior de H_0 es la siguiente:

$$p(H_0 | x) = \frac{p(x | H_0)p(H_0)}{p(x | H_0)p(H_0) + p(x | H_1)p(H_1)}$$

y la probabilidad posterior para H_1 es:

$$p(H_1 | x) = \frac{p(x | H_1)p(H_1)}{p(x | H_0)p(H_0) + p(x | H_1)p(H_1)}$$

Entonces el Factor de Bayes para este caso es el siguiente:

$$B_{01} = \frac{p(H_0 | x)}{p(H_1 | x)} \bigg/ \frac{p(H_0)}{p(H_1)} = \frac{p(x | H_0)}{p(x | H_1)} = \frac{p(x | \theta_0)}{p(x | \theta_1)}$$

Como puede observar, efectivamente en el caso de que ambas hipótesis sean simples, entonces el Factor de Bayes coincidirá con el radio de verosimilitud utilizado por la metodología clásica.

Cabe señalar que en el caso de que la distribución posterior contenga más de un parámetro y se esté interesado en efectuar una prueba de hipótesis respecto a un solo

parámetro, entonces, primero se debe obtener la distribución marginal del parámetro de interés y posteriormente efectuar la prueba de hipótesis. Ejemplo:

Sea $p(\theta, \lambda | x)$ la distribución posterior de los parámetros θ y λ , y se desea efectuar la siguiente prueba de hipótesis: $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \notin \Theta_0$, entonces el numerador del Factor de Bayes o ratio de probabilidad posterior es el siguiente:

$$\frac{\int \int_{H_0 G} p(x | \theta, \lambda) p(\theta, \lambda | H_0) d\lambda d\theta}{\int \int_{H_1 G} p(x | \theta, \lambda) p(\theta, \lambda | H_1) d\lambda d\theta}$$

donde G es el espacio paramétrico de λ .

3.6.4 Distribución predictiva

En muchas ocasiones resulta útil obtener pronósticos acerca de datos aún no observados, para hacerlo mediante la metodología bayesiana, se debe obtener la distribución predictiva de los datos aun no observados, que es la distribución condicional de los datos aun no observados dadas las observaciones previas, a continuación se presenta una definición más formal de esta distribución.

Sea $\{y_1, \dots, y_n\}$ una serie de observaciones cuya distribución conjunta $f(y_1, \dots, y_n | \theta)$ y la distribución a priori $f(\theta)$ se conocen, como consecuencia la distribución posterior $f(\theta | y_1, \dots, y_n)$ también se conoce o puede ser calculada. Se desea pronosticar la siguiente observación y_{n+1} , para hacerlo, se debe encontrar la distribución predictiva de y_{n+1} dadas las observaciones previas, de la siguiente forma:

$$f(y_{n+1} | y_1, \dots, y_n) = \int_{\Theta} f(y_{n+1} | y_1, \dots, y_n, \theta) f(\theta | y_1, \dots, y_n) d\theta$$

Donde:

$f(y_{n+1} | y_1, \dots, y_n)$: Distribución predictiva de y_{n+1} dadas las observaciones previas.

$f(y_{n+1} | y_1, \dots, y_n, \theta)$: Distribución condicional de y_{n+1} dado el parámetro θ y las observaciones previas.

$f(\theta | y_1, \dots, y_n)$; Distribución posterior de θ

Una vez que se obtiene la distribución predictiva, se pueden hacer inferencias acerca de y_{n+1} en forma similar a cuando se efectuaba para la distribución posterior. En caso de que se desee obtener un estimador puntual para y_{n+1} , entonces se puede emplear la media, la moda o la mediana de la distribución predictiva, o en su defecto si se dispone de una función de pérdida se busca el valor que minimice la esperanza matemática de la función de pérdida, en caso de que exista. También puede obtener un intervalo de credibilidad para y_{n+1} en forma similar a los que se obtienen para los parámetros del modelo.

4. Modelos Econométricos Bayesianos

Un análisis de regresión bayesiano, nos permite incorporar al modelo la información a priori disponible acerca de los parámetros. A diferencia de la metodología clásica, los Bayesianos tratan a los parámetros β como variables aleatorias, no porque sean el resultado de un experimento, sino en el sentido de que se les puede asignar una distribución de probabilidad que condense el estado de conocimiento que se tiene acerca de ellos antes de tomar en cuenta la información muestral. Esta distribución a priori es actualizada mediante los datos vía el Teorema de Bayes y se obtiene la distribución posterior, la cual es empleada para obtener estimadores puntuales, por intervalo y para realizar pruebas de hipótesis.

Como ya se señaló, la información a priori puede ser obtenida de dos fuentes:

1. De un estudio previo: Puede darse el caso en el que se disponga de un análisis previo con una muestra diferente de datos, entonces, la distribución posterior obtenida pasa a ser la distribución a priori del nuevo análisis, y los estimadores obtenidos son ahora los hiperparámetros de la distribución a

priori. Posteriormente, se procede a utilizar la muestra actual de datos para obtener la nueva distribución posterior y efectuar los procedimientos de inferencia.

2. De la experiencia de un experto: Se le puede hacer una serie de preguntas al investigador acerca del comportamiento del parámetro, por ejemplo, el tipo de valores que adopta el parámetro, su respectivo signo, etc., y posteriormente construir la distribución a priori utilizando alguna de las técnicas mencionadas en el capítulo anterior.

Aunque puede darse el caso en el que no se disponga de información a priori, entonces se utiliza una distribución a priori uniforme para el análisis. Dentro de este contexto, puede suceder que solo se conozca el signo o el conjunto de valores que puede adoptar el parámetro, entonces, se restringe la región correspondiente al parámetro.

En este capítulo se analizarán los modelos de regresión tratados en el capítulo dos, pero ahora bajo la metodología bayesiana.

4.1 Modelo de Regresión Lineal Simple

El modelo de regresión Lineal Simple involucra a una variable dependiente de una variable regresora, en esta sección se analizará este modelo utilizando una distribución a priori informativa y una no informativa. Antes de iniciar el análisis bayesiano es necesario recordar el modelo y a los supuestos que lo respaldan.

Del capítulo 2, sabemos que el modelo de regresión lineal simple es:

$$y = \beta_0 + \beta_1 x + \varepsilon \dots 35$$

con sus respectivos supuestos:

1. Linealidad: $y = \beta_0 + \beta_1 x + \varepsilon$
2. Homocedasticidad: $Var(y_i) = \sigma^2 \quad i = 1 \dots n$
3. La variable x es no estocástica.

4. La esperanza de los errores es cero, $E(\varepsilon_i) = 0$.
5. Los errores no están correlacionados, es decir: $E(\varepsilon_i \varepsilon_j) = 0$ para $i \neq j$
6. La variable $\varepsilon_i \quad i = 1 \dots n$, se distribuye independientemente, de forma normal, es decir, $\varepsilon_i \sim N(0, \sigma^2)$. Como consecuencia de este supuesto, “y” es una variable aleatoria tal que, “y” dado x , β_0 , β_1 y σ^2 se distribuye de forma normal con la siguiente media y varianza:

$$E[y_i | x_i, \beta_0, \beta_1, \sigma^2] = \beta_0 + \beta_1 x$$

$$Var[y_i | x_i, \beta_0, \beta_1, \sigma^2] = \sigma^2$$

y además $y_i | x_i, \beta_0, \beta_1, \sigma^2$ se distribuye de forma independiente.

El primer paso para realizar un análisis de regresión bayesiano, consiste en especificar la función de verosimilitud, para ello, considere a $\mathbf{y}' = (y_1, \dots, y_n)'$ y a $\mathbf{x}' = (x_1, \dots, x_n)'$, considerando que las y_i se distribuyen de forma independiente y el supuesto cuatro, la función de verosimilitud es la siguiente:

$$p(\mathbf{y} | \mathbf{x}, \beta_0, \beta_1, \sigma) = \frac{1}{(\sqrt{2\pi})^n \sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_i)^2 \right\}$$

donde la constante $(\sqrt{2\pi})^n$ puede ser omitida, rescribiendo la función de verosimilitud se tiene:

$$p(\mathbf{y} | \mathbf{x}, \beta_0, \beta_1, \sigma) = \frac{1}{\sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_i)^2 \right\}$$

Una vez que se tiene la función de verosimilitud, se procede a especificar la distribución a priori de los parámetros del modelo, en este caso, β_0 , β_1 y σ^2 . Cabe señalar que en caso de que no se disponga de información a priori o de que el investigador no desee incluirla en el modelo, entonces, se utiliza una distribución a priori no informativa, este caso será tratado en la siguiente sección. En caso de que se disponga de información a priori, ésta se condensa en una distribución a priori denominada informativa.

4.1.1 Análisis con una distribución a priori no informativa

A. Distribución a priori

La única información que se tiene en este caso es que σ debe ser positiva, considerando esto, se requiere especificar una distribución a priori conjunta para β_0 , β_1 y σ^2 que refleje nuestra ignorancia acerca de los posibles valores de los parámetros, para ello, se asume lo siguiente:

$$p(\beta_0, \beta_1, \sigma) = p(\beta_0)p(\beta_1)p(\sigma)$$

es decir, β_0 , β_1 y $\log\sigma$ se distribuyen de forma independiente. De acuerdo con lo anterior, resulta lógico asumir que antes de tomar en cuenta la información muestral, todos los valores de β_0 son equiprobables, lo mismo sucede con los valores de β_1 y $\log\sigma$, entonces, se toma como distribución a priori para cada parámetro una distribución uniforme, esto es:

$$\begin{aligned} p(\beta_0) &\propto k_1 & -\infty < \beta_0, \beta_1 < \infty \\ p(\beta_1) &\propto k_2 & 0 < \sigma < \infty \\ p(\sigma) &\propto \frac{1}{\sigma} \end{aligned}$$

Entonces la distribución a priori conjunta resulta ser la siguiente:

$$p(\beta_0, \beta_1, \sigma) \propto \frac{1}{\sigma} \quad -\infty < \beta_0, \beta_1 < \infty \quad 0 < \sigma < \infty$$

Como podrá notar, esta distribución es impropia, ya que su integral evaluada en la región correspondiente a cada parámetro no es uno sino ∞ ; sin embargo, el producto de esta distribución con la función de verosimilitud proporciona una distribución válida para el análisis.

B. Distribución posterior³

Como ya se mencionó, la distribución posterior es proporcional al producto de la distribución a priori con la función de verosimilitud, en este caso, se tiene lo siguiente:

$$\begin{aligned}
 p(\beta_0, \beta_1, \sigma | \mathbf{y}, \mathbf{x}) &\propto \frac{1}{\sigma} \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_i)^2\right\} \\
 &\propto \frac{1}{\sigma^{n+1}} \exp\left\{-\frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_i)^2\right\} \dots \mathbf{36}
 \end{aligned}$$

Esta distribución representa nuestro conocimiento actual acerca de los parámetros después de que la distribución a priori es actualizada utilizando la información muestral.

Para obtener los estimadores de los parámetros, se calculan las respectivas medias posteriores, $E[\beta_0 | y]$ y $E[\beta_1 | y]$, en este caso, los respectivos estimadores son los siguientes:

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}; \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

En este caso, los estimadores de mínimos cuadrados ordinarios, coinciden con las medias posteriores, sin embargo, en la metodología clásica, los parámetros son tratados como constantes desconocidas.

Un estimador de σ^2 está dado por:

$$s^2 = \frac{\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2}$$

Para hacer inferencias con respecto a los parámetros se utiliza la distribución posterior, de ésta se obtienen las distribuciones marginales posteriores de cada parámetro.

Para obtener la distribución posterior conjunta de β_0 y β_1 , se debe integrar la distribución posterior dada por la expresión 36, con respecto a σ , la distribución resultante

³ Véase Zellner [1971]

es una t – Student bivariada con $n - 2$ grados de libertad. De manera que, la distribución marginal posterior para β_0 y β_1 es una distribución t – Student univariada con $\nu = n - 2$ grados de libertad, es decir:

$$p(\beta_0 | \mathbf{y}, \mathbf{x}) \propto \left[\nu + \frac{\sum (x_i - \bar{x})^2}{s^2 \sum x_i^2 / n} (\beta_0 - \hat{\beta}_0) \right]^{-(\nu+1)/2} \quad -\infty < \beta_0 < \infty$$

$$p(\beta_1 | \mathbf{y}, \mathbf{x}) \propto \left[\nu + \frac{\sum (x_i - \bar{x})^2}{s^2} (\beta_1 - \hat{\beta}_1) \right]^{-(\nu+1)/2} \quad -\infty < \beta_1 < \infty$$

En forma similar, la distribución marginal posterior de σ se obtiene al integrar la distribución posterior con respecto a β_0 y β_1 , tal que:

$$p(\sigma | \mathbf{y}, \mathbf{x}) \propto \frac{1}{\sigma^{\nu+1}} \exp \left[-\frac{\nu s^2}{2\sigma^2} \right]$$

es una distribución gamma invertida.

Las distribuciones marginales posteriores representan el estado de conocimiento actual con respecto a cada parámetro individual, una vez que la información muestral es incorporada al modelo.

C. Intervalos de credibilidad

Las siguientes transformaciones son útiles para construir intervalos de credibilidad para β_0 y β_1 , las cuales siguen una distribución t – Student univariada con $n - 2$ grados de libertad⁴:

$$\frac{(\beta_0 - \hat{\beta}_0)}{\left[\frac{s^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2} \right]^{1/2}} | \mathbf{y}, \mathbf{x} \sim t_{n-2} \dots \mathbf{37}$$

Donde $\hat{\beta}_0$ es la media posterior de β_0 y el denominador es la desviación estándar de β_0 .

⁴Veáse Press [1989]

$$\left[\frac{(\beta_1 - \hat{\beta}_1)}{\frac{s}{\left(\sum (x_i - \bar{x})^2\right)^{1/2}}} \right] | \mathbf{y}, \mathbf{x} \sim t_{n-2} \dots \mathbf{38}$$

Donde $\hat{\beta}_1$ es la media posterior de β_1 y el denominador es la desviación estándar de β_1 .

Se puede construir un intervalo de credibilidad al $(1 - \alpha)100\%$ para β_1 utilizando un α - fractil de la distribución t - Student con $n - 2$ grados de libertad, $t_{\alpha/2, n-2}$ y la transformación 38:

$$-t_{\alpha/2, n-2} \leq \frac{(\beta_1 - \hat{\beta}_1)}{A} \leq t_{\alpha/2, n-2}$$

Donde: $A = \frac{s}{\left(\sum (x_i - \bar{x})^2\right)^{1/2}}$

Entonces el intervalo de credibilidad al $(1 - \alpha)100\%$ para β_1 es el siguiente:

$$\hat{\beta}_1 - t_{\alpha/2, n-2} A \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2, n-2} A$$

que cumple: $P(\hat{\beta}_1 - t_{\alpha/2, n-2} A \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2, n-2} A | \mathbf{y}, \mathbf{x}) = 1 - \alpha$.

Para obtener un intervalo de credibilidad al $(1 - \alpha)100\%$ para β_0 , se procede en forma similar utilizando la transformación 37:

El intervalo de credibilidad al $(1 - \alpha)100\%$ para β_0 es el siguiente:

$$\hat{\beta}_0 - t_{\alpha/2, n-2} C \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2, n-2} C$$

Donde, C es el denominador de la transformación 37.

que cumple: $P(\hat{\beta}_0 - t_{\alpha/2, n-2} C \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2, n-2} C | \mathbf{y}, \mathbf{x}) = 1 - \alpha$

Como podrá notar, estos intervalos de credibilidad son similares a los que se obtuvieron en la metodología clásica, sin embargo, existen diferencias entre las cuales podemos destacar las siguientes:

1. En la Metodología Clásica, los parámetros son considerados constantes desconocidas, por lo que al calcular los respectivos intervalos de confianza, no se puede hablar de que existe una probabilidad $(1 - \alpha)100\%$ de que el parámetro involucrado pertenezca al intervalo de confianza, en este caso, el intervalo es aleatorio. En cambio, en el enfoque bayesiano, al calcular los respectivos intervalos de credibilidad, dada la información muestral disponible, se puede decir, que la probabilidad de que el parámetro involucrado pertenezca al intervalo de credibilidad es $(1 - \alpha)100\%$.
2. En la Metodología Clásica, la obtención de los intervalos de confianza se hace en función de la distribución de probabilidad de los estimadores de los parámetros, en el enfoque bayesiano, esto no es así.

D. Distribución predictiva

Como ya se mencionó anteriormente, uno de los objetivos del MRLS, es el pronóstico de la variable dependiente y_{n+1} dada la nueva observación x_{n+1} , para ello, la metodología bayesiana procede a encontrar la distribución predictiva, que es la distribución condicional de la variable respuesta, y_{n+1} dado el pasado. Posteriormente, se utiliza la esperanza posterior predictiva de esta distribución como un estimador de y_{n+1} , además se puede construir el respectivo intervalo predictivo. En la sección 3.5.4, se vio que para obtener la distribución predictiva, se integra con respecto a los parámetros del modelo, la distribución conjunta de y_{n+1} y los parámetros.

Se asume que y_{n+1} es generada por el modelo 35 con una distribución a priori no informativa, de acuerdo con esto, se tiene lo siguiente:

$$y_{n+1} = \beta_0 + \beta_1 x_{n+1} + \varepsilon; \quad \varepsilon \sim N(0, \sigma^2)$$

Para obtener la distribución conjunta de y_{n+1} y los parámetros β_0, β_1, σ , se utiliza la siguiente expresión:

$$p(y_{n+1}, \beta_0, \beta_1, \sigma | \mathbf{y}, \mathbf{x}, x_{n+1}) = p(y_{n+1} | \beta_0, \beta_1, \sigma, x_{n+1})p(\beta_0, \beta_1, \sigma | \mathbf{y}, \mathbf{x})$$

Donde: $p(\beta_0, \beta_1, \sigma | \mathbf{y}, \mathbf{x})$ es la distribución posterior de β_0, β_1, σ , dada por la expresión 36, mientras que, $p(y_{n+1} | \beta_0, \beta_1, \sigma, x_{n+1})$ es la función de verosimilitud de y_{n+1} , que es la siguiente:

$$p(y_{n+1} | \beta_0, \beta_1, \sigma, x_{n+1}) \propto \frac{1}{\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_{n+1} - \beta_0 - \beta_1 x_{n+1})^2\right\}$$

Entonces la distribución conjunta de y_{n+1} y los parámetros β_0, β_1, σ está dada por:

$$p(y_{n+1}, \beta_0, \beta_1, \sigma | \mathbf{y}, \mathbf{x}) \propto \frac{1}{\sigma^{n+2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum (y_i - \beta_0 - \beta_1 x_i)^2 + (y_{n+1} - \beta_0 - \beta_1 x_{n+1})^2\right)\right\}$$

La distribución predictiva de y_{n+1} dado el pasado se obtiene al integrar esta última expresión con respecto a β_0, β_1, σ :

$$p(y_{n+1} | \mathbf{y}, \mathbf{x}, x_{n+1}) = \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty p(y_{n+1}, \beta_0, \beta_1, \sigma | \mathbf{y}, \mathbf{x}, x_{n+1}) d\beta_0 d\beta_1 d\sigma$$

La expresión resultante sigue una distribución t – Student con $n - 2$ grados de libertad. Una transformación útil para construir el respectivo intervalo predictivo de y_{n+1} es la siguiente:

$$\frac{(y_{n+1} - \hat{\beta}_0 - \hat{\beta}_1 x_{n+1})}{s \left[1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]^{1/2}} | \mathbf{y}, \mathbf{x}, x_{n+1} \sim t_{n-2}$$

Como puede notar en el denominador, existen dos fuentes de variación, la primera debida a la variación muestral mientras que la segunda se atribuye a la incertidumbre debida a los parámetros β_0 y β_1 .

Para construir el respectivo intervalo predictivo se utiliza un α – fractil de la distribución t – Student con $n - 2$ grados de libertad, $t_{\alpha/2, n-2}$. El intervalo predictivo al $(1-\alpha)100\%$ para y_{n+1} es el siguiente:

$$\hat{\beta}_0 + \hat{\beta}_1 x_{n+1} - t_{\alpha/2, n-2} D \leq y_{n+1} \leq \hat{\beta}_0 + \hat{\beta}_1 x_{n+1} + t_{\alpha/2, n-2} D$$

$$\text{Donde: } D = s \left[1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]^{1/2}$$

Este intervalo cumple:

$$p(\hat{\beta}_0 + \hat{\beta}_1 x_{n+1} - t_{\alpha/2, n-2} D \leq y_{n+1} \leq \hat{\beta}_0 + \hat{\beta}_1 x_{n+1} + t_{\alpha/2, n-2} D \mid \mathbf{y}, \mathbf{x}, x_{n+1}) = 1 - \alpha$$

Nuevamente, estos resultados son similares a los obtenidos en el capítulo 2, aunque debe enfatizarse que esto sólo sucede cuando se utiliza una distribución a priori no informativa. Bajo la metodología clásica, se estima el modelo utilizando datos históricos y después éste se utiliza para pronosticar valores futuros suponiendo que el modelo estimado es correcto. Mientras que bajo el enfoque bayesiano, se obtiene la distribución predictiva la cual incorpora tanto la incertidumbre relacionada con el valor futuro como la asociada con los parámetros del modelo, y a partir de ésta se encuentra un estimador para y_{n+1} .

4.1.2 Análisis con una distribución a priori informativa

Como ya se mencionó al inicio de este capítulo, la información a priori puede ser obtenida de la experiencia de un experto mediante una serie de preguntas, que nos permitan determinar la forma de la distribución a priori y sus respectivos parámetros, los cuales son denominados hiperparámetros. Aunque también, puede existir información proveniente de un estudio previo, ya sean algunas características tales como los estimadores de los parámetros del respectivo estudio previo, los cuales deben ser incorporados adecuadamente en el estudio actual, o tal vez, el estudio previo fue realizado bajo el enfoque bayesiano, entonces la distribución posterior obtenida, ahora en el estudio actual, será la distribución a priori.

Por conveniencia matemática, un enfoque alternativo a los ya mencionados, consiste en utilizar una distribución a priori conjugada, en este caso, la información a priori existente es utilizada para especificar los valores de los hiperparámetros, es decir, se utiliza la experiencia de un experto y/o las características aportadas por un estudio previo. En esta sección se analizará este enfoque, para ello, expresaremos el MRLS, dado por 35, así como la función de verosimilitud de los datos, en términos matriciales:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}$$

Donde:

Y: vector de nx1 de observaciones de la variable dependiente

X: matriz de nx2, cuya primera columna está conformada por unos y la segunda contiene las n observaciones correspondientes a la variable regresora x.

B: vector de 2x1 de los parámetros desconocidos β_0, β_1 o coeficientes de regresión.

E: vector de nx1 de los errores o perturbaciones aleatorias.

La función de verosimilitud de los datos, expresión 36, se reexpresa como:

$$p(\mathbf{Y} | \mathbf{X}, \mathbf{B}, \sigma) = \frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{XB})' (\mathbf{Y} - \mathbf{XB})\right] \dots 39$$

A. Distribución a priori

Utilizaremos una distribución a priori conjugada definida en dos pasos de la siguiente manera:

$$p(\mathbf{B}, \sigma) = p(\mathbf{B} | \sigma) p(\sigma)$$

Donde $p(\mathbf{B} | \sigma)$ es una distribución Normal con media a priori \mathbf{b} y matriz de varianzas – covarianzas a priori $\sigma^2 \mathbf{V}^{-1}$, es decir:

$$p(\mathbf{B} | \sigma) \propto \frac{|\mathbf{V}|^{1/2}}{\sigma^2} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{B} - \mathbf{b})' \mathbf{V} (\mathbf{B} - \mathbf{b})\right]$$

En este caso, los hiperparámetros \mathbf{b} y \mathbf{V}^{-1} se deben especificar utilizando la información a priori disponible.

El hiperparámetro \mathbf{b} representa la información a priori que se tiene acerca del modelo de regresión, puede determinarse de dos formas:

1. En caso de que exista un estudio previo y se disponga del correspondiente vector de estimadores de \mathbf{B} , entonces puede ser utilizado como el vector de medias a priori \mathbf{b} .
2. Cuando no se dispone de un estudio previo, el vector \mathbf{b} se determina de acuerdo a la experiencia de un experto, es decir, él determina cuáles valores de \mathbf{B} son más probables a priori.

El hiperparámetro \mathbf{V}^{-1} es una medida de precisión de la evaluación que se hace al determinar el vector \mathbf{b} , es decir, refleja la confianza que tenemos en el vector de medias a priori. Este puede ser determinado de la siguiente forma:

1. Del estudio previo del cual se obtuvo \mathbf{b} se puede utilizar la respectiva matriz de varianzas – covarianzas estimada de \mathbf{B} , para determinar a \mathbf{V}^{-1} . Birkes[1993], utiliza la siguiente expresión para determinar a \mathbf{V}^{-1} :

$$\mathbf{V}^{-1} = \frac{a}{s^2} \mathbf{W}$$

Donde:

\mathbf{W} es la matriz de varianzas – covarianzas estimada del estudio previo;

s^2 : es la varianza estimada del estudio previo

a : es un factor mayor a uno, que refleja nuestra incertidumbre acerca de qué tan relevante es el estudio previo en el análisis actual.

2. Cuando no se dispone de un estudio previo, se utiliza la experiencia del experto para determinar a \mathbf{V}^{-1} , pero debido a que no es fácil formular opiniones respecto a \mathbf{V}^{-1} , frecuentemente es especificada como una matriz diagonal, lo que implica que a priori, los elementos de \mathbf{B} no están correlacionados.

La distribución a priori de σ es una Gamma Invertida con hiperparámetros ν_0 y c_0^2 , es decir:

$$p(\sigma) \propto \frac{1}{\sigma^{\nu_0+1}} \exp\left(-\frac{\nu_0 c_0^2}{2\sigma^2}\right) \quad \text{para } \nu_0 > 0$$

En este caso, la información del experto referente a la variabilidad de los datos, puede ser utilizada para determinar el valor de c_0^2 , mientras que ν_0 que representa los grados de libertad a priori, es determinado por el experto en términos del tamaño de una muestra imaginaria.

Si se dispone de información de un estudio previo, entonces la respectiva varianza estimada puede ser utilizada como el valor de c_0^2 y ν_0 es determinado a partir del tamaño muestral de dicho estudio.

De manera que la distribución a priori conjunta queda de la siguiente forma:

$$p(\mathbf{B}, \sigma) \propto \frac{|\mathbf{V}|^{1/2}}{\sigma^2} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{B} - \mathbf{b})' \mathbf{V} (\mathbf{B} - \mathbf{b})\right] \times \frac{1}{\sigma^{\nu_0+1}} \exp\left(-\frac{\nu_0 c_0^2}{2\sigma^2}\right)$$

efectuando el producto se tiene:

$$p(\mathbf{B}, \sigma) \propto \frac{1}{\sigma^{\nu_0+3}} \exp\left[-\frac{1}{2\sigma^2} (\nu_0 c_0^2 + (\mathbf{B} - \mathbf{b})' \mathbf{V} (\mathbf{B} - \mathbf{b}))\right] \dots 40$$

Como puede notar, la distribución a priori es determinada independientemente de la información muestral actual.

B. Distribución posterior

Al combinar la función de verosimilitud, 39, con la distribución a priori, 40, se obtiene la distribución posterior de \mathbf{B} y σ dada la información muestral actual:

$$p(\mathbf{B}, \sigma | \mathbf{Y}, \mathbf{X}) \propto \left(\frac{1}{\sigma^{\nu_0+3}} \exp\left[-\frac{1}{2\sigma^2} (\nu_0 c_0^2 + (\mathbf{B} - \mathbf{b})' \mathbf{V} (\mathbf{B} - \mathbf{b}))\right] \right) \\ \times \left(\frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{XB})' (\mathbf{Y} - \mathbf{XB})\right] \right)$$

$$p(\mathbf{B}, \sigma | \mathbf{Y}, \mathbf{X}) \propto \left(\frac{1}{\sigma^{\nu_0+n+3}} \exp \left[-\frac{1}{2\sigma^2} (\nu_0 c_0^2 + (\mathbf{B} - \mathbf{b})' \mathbf{V} (\mathbf{B} - \mathbf{b}) + (\mathbf{Y} - \mathbf{X}\mathbf{B})' (\mathbf{Y} - \mathbf{X}\mathbf{B})) \right] \right)$$

Esta distribución representa nuestro conocimiento actual referente a los parámetros \mathbf{B} y σ una vez que la información a priori es actualizada por la información muestral mediante el Teorema de Bayes. Esta distribución puede reescribirse de la siguiente forma:

$$p(\mathbf{B}, \sigma | \mathbf{Y}, \mathbf{X}) \propto \left(\frac{1}{\sigma^{\nu_p+n+3}} \exp \left[-\frac{1}{2\sigma^2} (\nu_p s_p^2 + (\mathbf{B} - \bar{\mathbf{B}})' (\mathbf{V} + \mathbf{X}'\mathbf{X}) (\mathbf{B} - \bar{\mathbf{B}})) \right] \right)$$

Donde:

$$\nu_p = \nu_0 + n$$

$$\nu_p s_p^2 = \nu_0 c_0^2 + (\mathbf{Y} - \mathbf{X}\mathbf{B})' (\mathbf{Y} - \mathbf{X}\mathbf{B}) + (\mathbf{b} - \mathbf{B})' \mathbf{V} (\mathbf{b} - \mathbf{B})$$

$$\bar{\mathbf{B}} = (\mathbf{V} + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{V}\mathbf{b} + \mathbf{X}'\mathbf{X}\hat{\mathbf{B}}_{LS})$$

$\hat{\mathbf{B}}_{LS}$: Estimador de mínimos cuadrados

Esta distribución es una Normal – Gamma Invertida. La distribución condicional de \mathbf{B} dado σ es una Normal con media posterior $\bar{\mathbf{B}}$ y matriz de varianzas – covarianzas $\sigma^2 (\mathbf{V} + \mathbf{X}'\mathbf{X})^{-1}$, mientras que la distribución de σ es una Gamma Invertida con parámetros posteriores ν_p y s_p^2 (se obtiene al dividir $\nu_p s_p^2$ entre ν_p).

La distribución marginal posterior de \mathbf{B} es una t – Student bivariada con ν_p grados de libertad y las siguientes características:

- Media posterior: $E[\mathbf{B} | \mathbf{Y}, \mathbf{X}] = \bar{\mathbf{B}}$, la cual se utiliza como estimador del vector de parámetros \mathbf{B} .
- Matriz de precisión posterior: $s_p^{-2} (\mathbf{V} + \mathbf{X}'\mathbf{X})$
- Matriz de covarianza posterior: $\left[\frac{\nu_p}{\nu_p - 2} \right] s_p^2 (\mathbf{V} + \mathbf{X}'\mathbf{X})^{-1}$

Como puede notar, el estimador del vector de parámetros \mathbf{B} , $\bar{\mathbf{B}}$, es un promedio ponderado del vector de medias a priori y el estimador de mínimos cuadrados, esto se debe a que cada uno de ellos, está multiplicado por las matrices \mathbf{V} y $\mathbf{X}'\mathbf{X}$ respectivamente.

C. Intervalos de credibilidad

La distribución marginal de cada elemento del vector \mathbf{B} es una t – Student univariada con ν_p grados de libertad, este resultado es útil para construir intervalos de credibilidad para cada β_i utilizando α – fractiles de la distribución t – Student, $t_{\alpha/2, \nu_p}$.

Un intervalo de credibilidad al $(1 - \alpha)100\%$ para β_i es el siguiente:

$$\bar{\mathbf{B}}_i \pm t_{\alpha/2, \nu_p} \sqrt{s_p^2 \mathbf{M}_{ii}}$$

donde $\bar{\mathbf{B}}_i$ es el elemento i del vector $\bar{\mathbf{B}}$ y \mathbf{M}_{ii} es el elemento (i, i) de la matriz $(\mathbf{V} + \mathbf{X}'\mathbf{X})^{-1}$. Este intervalo cumple: $P(\beta_i \in \bar{\mathbf{B}}_i \pm t_{\alpha/2, \nu_p} \sqrt{s_p^2 \mathbf{M}_{ii}} | \mathbf{Y}, \mathbf{X}) = 1 - \alpha$, es decir, la probabilidad de que β_i esté en el intervalo es de $1 - \alpha$.

4.2 Modelo de Regresión Lineal Múltiple

Como sabemos, el MRLM involucra a más de una variable regresora, en esta sección, se analizará este modelo bajo el enfoque bayesiano utilizando una distribución a priori informativa y una no informativa.

Al utilizar una distribución a priori no informativa notará que los resultados obtenidos son similares a los de la metodología clásica, sin embargo, difieren tanto en interpretación como en procedimientos.

El MRLM en notación matricial es el siguiente:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \dots 41$$

Donde:

\mathbf{Y} : vector de $n \times 1$ de observaciones de la variable dependiente

X: matriz de $n \times k$ que contiene las n observaciones correspondientes a las $k - 1$ variables regresoras.

B: vector de $k \times 1$ de los parámetros desconocidos $\beta_0, \beta_1, \beta_2, \dots, \beta_{k-1}$ o coeficientes de regresión.

E: vector de $n \times 1$ de los errores o perturbaciones aleatorias.

Y los supuestos que respaldan a este modelo son los siguientes:

1. Linealidad
2. La esperanza del vector de los errores es cero: $E(\mathbf{E}) = \mathbf{0}$, donde \mathbf{E} y $\mathbf{0}$ son vectores de $n \times 1$.
3. **Homocedasticidad y no correlación serial**: La varianza es la misma para todos los errores, no existe correlación entre los errores, es decir, $E(\varepsilon_i \varepsilon_j) = 0$ para $i \neq j$, esto se resume en la siguiente expresión: $\text{var}(\mathbf{E}) = \sigma^2 \mathbf{I}$ donde \mathbf{I} es la matriz identidad de $n \times n$.
4. La matriz **X** es no estocástica.
5. El rango de la matriz **X** es igual a k , es decir, no existe multicolinealidad, dicho de otra manera, las columnas de la matriz son linealmente independientes, no existe una relación lineal exacta entre las variables x .
6. El vector de los errores se distribuye de forma normal con un vector de medias cero y una matriz de varianzas – covarianzas $\sigma^2 \mathbf{I}$, es decir: $\mathbf{E} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$.

Ignorando una constante de proporcionalidad, la función de verosimilitud es la siguiente:

$$p(\mathbf{Y} | \mathbf{X}, \mathbf{B}, \sigma) = \frac{1}{\sigma^n} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{XB})' (\mathbf{Y} - \mathbf{XB}) \right]$$

4.2.1 Análisis con una distribución a priori no informativa

A. Distribución a priori

Se requiere especificar una distribución a priori que exprese nuestra ignorancia acerca de los valores que pueden tomar los parámetros del modelo, en este caso, \mathbf{B} y σ^2 , para ello, se asume que \mathbf{B} y $\log\sigma$ se distribuyen de forma independiente, lo que implica que su distribución a priori conjunta es el producto de la distribución a priori individual de cada parámetro:

$$p(\mathbf{B}, \sigma) = p(\mathbf{B})p(\sigma)$$

Cabe señalar que, la única información disponible es que σ debe tomar valores positivos. Considerando esto, se puede asumir una distribución uniforme para cada parámetro, de manera que la distribución a priori conjunta es la siguiente:

$$p(\mathbf{B}, \sigma) \propto \frac{1}{\sigma} \quad -\infty < \beta_i < \infty \quad 0 < \sigma < \infty$$

lo que significa que todos los valores de β_i y $\log\sigma$ son equiprobables antes de tomar en cuenta la información muestral. Esta distribución es impropia, ya que su integral evaluada en la región correspondiente a cada parámetro, no es uno.

B. Distribución posterior

Como sabemos, la distribución posterior es proporcional al producto de la distribución a priori y la función de verosimilitud, por lo que se tiene:

$$p(\mathbf{B}, \sigma | \mathbf{Y}, \mathbf{X}) \propto \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB})\right] \dots 42$$

Debido a que la distribución a priori empleada es impropia, entonces para que la distribución posterior sea propia se debe cumplir lo siguiente:

1. El número de observaciones, n , debe ser mayor que el número de parámetros, k , a estimar: $n > k$.

2. El rango de la matriz \mathbf{X} debe ser igual a k , esto implica que las columnas de la matriz \mathbf{X} deben ser independientes, es decir, ninguna columna puede expresarse como una combinación lineal de las otras columnas.

Los estimadores de los parámetros se obtienen al calcular las respectivas esperanzas posteriores:

$$E[\mathbf{B} | \mathbf{Y}, \mathbf{X}] = \hat{\mathbf{B}}; \quad \hat{\mathbf{B}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}$$

y un estimador de σ^2 está dado por: $s^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})}{n - k}$

La distribución marginal posterior de σ , se obtiene al integrar la distribución posterior 42, con respecto a cada uno de los elementos del vector \mathbf{B} :

$$p(\sigma | \mathbf{Y}, \mathbf{X}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\mathbf{B}, \sigma | \mathbf{Y}, \mathbf{X}) d\beta_1 \cdots d\beta_{k-1}$$

La distribución resultante es una Gamma invertida con $\nu = n - k$ grados de libertad.

Cabe señalar que la distribución condicional posterior de \mathbf{B} dado σ , \mathbf{Y} y \mathbf{X} es una distribución Normal multivariada con media $\hat{\mathbf{B}}$ y matriz de varianzas – covarianzas $\sigma^2 (\mathbf{X}' \mathbf{X})^{-1}$. Considerando que la distribución posterior también se puede factorizar de la siguiente manera:

$$p(\mathbf{B}, \sigma | \mathbf{Y}, \mathbf{X}) = p(\mathbf{B} | \sigma, \mathbf{Y}, \mathbf{X}) p(\sigma | \mathbf{Y}, \mathbf{X})$$

entonces la distribución posterior 42 es una Normal – Gamma Invertida.

Por otra parte, la distribución marginal posterior del vector \mathbf{B} , se obtiene al integrar la distribución posterior, 42, con respecto a σ :

$$p(\mathbf{B} | \mathbf{Y}, \mathbf{X}) = \int_0^{\infty} p(\mathbf{B}, \sigma | \mathbf{Y}, \mathbf{X}) d\sigma$$

la distribución resultante es una t – Student multivariada con $\nu = n - k$ grados de libertad.

Debido a que se desea realizar inferencias con respecto a cada parámetro, es necesario, obtener la distribución marginal de cada elemento del vector \mathbf{B} . Esto se puede realizar de dos formas: la primera consiste en integrar la distribución posterior, primero con respecto a los parámetros β_i que no son de interés y posteriormente con respecto a σ . La segunda forma, consiste en integrar la distribución marginal posterior de \mathbf{B} con respecto a los parámetros β_i que no son de interés. Como resultado, se tiene que cada parámetros β_i sigue una distribución t – Student univariada con $\nu = n - k$ grados de libertad.

Nuevamente es importante señalar que la distribución posterior resume todo el conocimiento actualizado referente a los parámetros \mathbf{B} y σ , una vez que la información muestral ha sido tomada en cuenta. A su vez, las distribuciones marginales posteriores representan el conocimiento actual que se tiene referente al parámetro en cuestión, una vez que se incorporó la información muestral.

C. Intervalos de Credibilidad

Como se mencionó en el apartado anterior, cada β_i sigue una distribución t – Student univariada con $\nu = n - k$ grados de libertad, de manera que la siguiente transformación:

$$\frac{\beta_i - \hat{\beta}_i}{s(m_{ii})^{1/2}} \Big| \mathbf{Y}, \mathbf{X} \sim t_\nu$$

sigue una distribución t – Student con $\nu = n - k$ grados de libertad. Donde: $\hat{\beta}_i$ es el i – ésimo elemento del vector $\hat{\mathbf{B}}$, el denominador es la desviación estándar posterior y m_{ii} es el elemento (i,i) de la matriz $(\mathbf{X}' \mathbf{X})^{-1}$.

El intervalo de credibilidad al $(1 - \alpha)100\%$ para β_i , utilizando un α – fractil de la distribución t – Student, $t_{\alpha/2, \nu}$, es el siguiente:

$$\hat{\beta}_i - t_{\alpha/2, \nu} s(m_{ii})^{1/2} \leq \beta_i \leq \hat{\beta}_i + t_{\alpha/2, \nu} s(m_{ii})^{1/2}$$

tal que: $p(\hat{\beta}_i - t_{\alpha/2, \nu} s(m_{ii})^{1/2} \leq \beta_i \leq \hat{\beta}_i + t_{\alpha/2, \nu} s(m_{ii})^{1/2} \mid \mathbf{Y}, \mathbf{X}) = 1 - \alpha$

Es decir, la probabilidad de que el parámetro β_i caiga dentro del intervalo especificado, condicionada en la información muestral actual es $1 - \alpha$. Esta interpretación es adecuada, debido a que β_i es considerada una variable aleatoria, aunque no hay que perder de vista el sentido subjetivo de la definición de probabilidad.

En los intervalos de confianza obtenidos por la metodología clásica, no aplica esta interpretación, esto se debe a que los β_i son considerados constantes desconocidas, en este caso, los intervalos de confianza son aleatorios.

D. Distribución Predictiva

Si se desea pronosticar un vector de $q \times 1$ variables dependientes, \mathbf{Y}_q , dada la matriz de $q \times k$ nuevas observaciones, \mathbf{X}_q , entonces se procede en forma similar que en el caso del MRLS, es decir, tenemos que encontrar la distribución predictiva del vector \mathbf{Y}_q dado el pasado.

Asumimos que el vector \mathbf{Y}_q es generado por el MRLM, 41, con una distribución a priori no informativa:

$$\mathbf{Y}_q = \mathbf{X}_q \mathbf{B} + \mathbf{E} \quad \mathbf{E} \sim N(0, \sigma^2)$$

La función de verosimilitud de \mathbf{Y}_q es la siguiente:

$$p(\mathbf{Y}_q | \mathbf{B}, \sigma) \propto \frac{1}{\sigma^q} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{Y}_q - \mathbf{X}_q \mathbf{B})' (\mathbf{Y}_q - \mathbf{X}_q \mathbf{B}) \right]$$

que combinada con la distribución posterior, 42, se obtiene la distribución conjunta de \mathbf{Y}_q, \mathbf{B} y σ , que es la siguiente:

$$p(\mathbf{Y}_q, \mathbf{B}, \sigma | \mathbf{Y}, \mathbf{X}, \mathbf{X}_q) \propto \frac{1}{\sigma^{n+q+1}} \exp \left\{ -\frac{1}{2\sigma^2} \left[(\mathbf{Y} - \mathbf{X}\mathbf{B})' (\mathbf{Y} - \mathbf{X}\mathbf{B}) + (\mathbf{Y}_q - \mathbf{X}_q \mathbf{B})' (\mathbf{Y}_q - \mathbf{X}_q \mathbf{B}) \right] \right\}$$

La distribución predictiva de \mathbf{Y}_q dado el pasado, se obtiene al integrar esta distribución con respecto a \mathbf{B} y σ :

$$p(\mathbf{Y}_q | \mathbf{Y}, \mathbf{X}, \mathbf{X}_q) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_0^{\infty} p(\mathbf{Y}_q, \mathbf{B}, \sigma | \mathbf{Y}, \mathbf{X}, \mathbf{X}_q) d\sigma d\beta_0 \dots d\beta_k$$

la distribución resultante es una t – Student multivariada con $\nu = n - k$ grados de libertad.

Un estimador de \mathbf{Y}_q es la media posterior predictiva que es la siguiente:

$$E[\mathbf{Y}_q] = \mathbf{X}_q \hat{\mathbf{B}} \text{ para } \nu > 1$$

La distribución marginal de cada elemento del vector \mathbf{Y}_q es una t – Student univariada, una transformación útil para realizar inferencias es la siguiente:

$$\frac{y_{qi} - \mathbf{X}_q(i)\hat{\mathbf{B}}}{s\sqrt{h_{ii}}} \Big| \mathbf{Y}, \mathbf{X}, \mathbf{X}_q \sim t_{\nu} \dots \mathbf{43}$$

Donde:

y_{qi} es el i – ésimo elemento del vector \mathbf{Y}_q

$\mathbf{X}_q(i)$ es el i – ésimo renglón de la matriz \mathbf{X}_q

h_{ii} es el elemento (i,i) de la matriz $\mathbf{H} = (\mathbf{I} + \mathbf{X}_q(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_q')$

Cuando se desea pronosticar sólo una variable dependiente, es decir $q = 1$, entonces \mathbf{X}_q es de tamaño 1 x k, y la transformación 43, queda de la siguiente forma:

$$\frac{y_q - \mathbf{X}_q \hat{\mathbf{B}}}{s\sqrt{[1 + \mathbf{X}_q(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_q']}} \Big| \mathbf{Y}, \mathbf{X}, \mathbf{X}_q \sim t_{\nu} \dots \mathbf{44}$$

El denominador de ambas transformaciones es la desviación estándar posterior predictiva, en la cual se observa que existen dos fuentes de error, una debida a la variación muestral y la otra a la incertidumbre asociada con los parámetros \mathbf{B} del modelo.

Se puede construir un intervalo posterior predictivo para cada elemento del vector \mathbf{Y}_q , utilizando la transformación 43 y un α – fractil de la distribución t – Student univariada con $\nu = n - k$ grados de libertad, y está dado por la siguiente expresión:

$$\mathbf{X}_q(i)\hat{\mathbf{B}} - t_{\alpha/2, \nu} s \sqrt{h_{ii}} \leq y_{qi} \leq \mathbf{X}_q(i)\hat{\mathbf{B}} + t_{\alpha/2, \nu} s \sqrt{h_{ii}}$$

que cumple: $P(\mathbf{X}_q(i)\hat{\mathbf{B}} - t_{\alpha/2, \nu} s \sqrt{h_{ii}} \leq y_{qi} \leq \mathbf{X}_q(i)\hat{\mathbf{B}} + t_{\alpha/2, \nu} s \sqrt{h_{ii}} \mid \mathbf{Y}, \mathbf{X}, \mathbf{X}_q) = 1 - \alpha$

Es decir, la probabilidad de que la variable pronosticada, y_{qi} , caiga dentro del intervalo es $1 - \alpha$, claro, sin perder de vista el sentido subjetivo de la definición de probabilidad.

4.2.2 Análisis con una distribución a priori informativa

Puede suceder que se disponga de información de un estudio previo, el cual fue realizado bajo el enfoque bayesiano, entonces, la distribución posterior obtenida, pasa a ser la distribución a priori del nuevo estudio. Esto implica que, la nueva muestra disponible es generada por el mismo modelo de regresión, es decir, en el estudio previo, se tenían los parámetros (\mathbf{B}, σ) y los datos $(\mathbf{X}_1, \mathbf{Y}_1, \nu_1)$, mientras que en el estudio actual, se tienen los mismos parámetros pero los datos son $(\mathbf{X}_2, \mathbf{Y}_2, \nu_2)$. En este caso, la distribución a priori será de la misma forma que la expresión 42, pero reexpresada en términos de $(\mathbf{X}_1, \mathbf{Y}_1, \nu_1)$, es decir:

$$p(\mathbf{B}, \sigma \mid \mathbf{Y}_1, \mathbf{X}_1) \propto \frac{1}{\sigma^{n_1+1}} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{Y}_1 - \mathbf{X}_1\mathbf{B})' (\mathbf{Y}_1 - \mathbf{X}_1\mathbf{B})\right] \dots 45$$

Los resultados obtenidos en el estudio previo reexpresados en términos de $(\mathbf{X}_1, \mathbf{Y}_1, \nu_1)$ son los siguientes:

$$\hat{\mathbf{B}}_1 = (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{Y}_1; \quad s_1^2 = \frac{(\mathbf{Y}_1 - \mathbf{X}_1\hat{\mathbf{B}})' (\mathbf{Y}_1 - \mathbf{X}_1\hat{\mathbf{B}})}{n_1 - k}; \quad \nu_1 = n_1 - k$$

La distribución 4.11 es una Normal – Gamma Invertida, ya que la distribución condicional a priori de \mathbf{B} dado σ es una Normal con media a priori $\hat{\mathbf{B}}_1$ y matriz de covarianza a priori $(\mathbf{X}_1' \mathbf{X}_1)^{-1} \sigma^2$, mientras que la distribución marginal a priori de σ es una Gamma Invertida con hiperparámetros s_1^2 y ν_1 .

La función de verosimilitud considerando la muestra actual, es la siguiente:

$$p(\mathbf{Y}_2 | \mathbf{X}_2, \mathbf{B}, \sigma) = \frac{1}{\sigma^{n_2}} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{Y}_2 - \mathbf{X}_2 \mathbf{B})' (\mathbf{Y}_2 - \mathbf{X}_2 \mathbf{B}) \right]$$

Entonces la distribución posterior del estudio actual, que se obtiene al combinar la función de verosimilitud con la distribución a priori, es la siguiente:

$$p(\mathbf{B}, \sigma | \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{X}_1, \mathbf{X}_2) \propto \frac{1}{\sigma^{n_1+n_2+1}} \exp \left[-\frac{1}{2\sigma^2} [(\mathbf{Y}_1 - \mathbf{X}_1 \mathbf{B})' (\mathbf{Y}_1 - \mathbf{X}_1 \mathbf{B}) + (\mathbf{Y}_2 - \mathbf{X}_2 \mathbf{B})' (\mathbf{Y}_2 - \mathbf{X}_2 \mathbf{B})] \right]$$

Esta distribución es una Normal – Gamma Invertida, lo cual es más visible, si la reescribimos de la siguiente manera:

$$p(\mathbf{B}, \sigma | \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{X}_1, \mathbf{X}_2) \propto \frac{1}{\sigma^{n_1+n_2+1}} \exp \left[-\frac{1}{2\sigma^2} [\nu s^2 + (\mathbf{B} - \tilde{\mathbf{B}})' \mathbf{M} (\mathbf{B} - \tilde{\mathbf{B}})] \right]$$

Donde:

$$\mathbf{M} = \mathbf{X}_1' \mathbf{X}_1 + \mathbf{X}_2' \mathbf{X}_2$$

$$\tilde{\mathbf{B}} = \mathbf{M}^{-1} (\mathbf{X}_1' \mathbf{Y}_1 + \mathbf{X}_2' \mathbf{Y}_2)$$

$$\nu s^2 = (\mathbf{Y}_1 - \mathbf{X}_1 \tilde{\mathbf{B}})' (\mathbf{Y}_1 - \mathbf{X}_1 \tilde{\mathbf{B}}) + (\mathbf{Y}_2 - \mathbf{X}_2 \tilde{\mathbf{B}})' (\mathbf{Y}_2 - \mathbf{X}_2 \tilde{\mathbf{B}})$$

$$\nu = n_1 + n_2 - k$$

La distribución condicional posterior de \mathbf{B} dado σ es una Normal con media posterior $\tilde{\mathbf{B}}$ y matriz de covarianza posterior $\mathbf{M}^{-1} \sigma^2$, mientras que la distribución marginal posterior de σ es una Gamma Invertida con parámetros s^2 y ν .

La distribución marginal de \mathbf{B} es una t – Student multivariada con ν grados de libertad y las siguientes características:

- Media posterior: $E[\mathbf{B} | \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{X}_1, \mathbf{X}_2] = \tilde{\mathbf{B}}$, la cual se utiliza como estimador del vector de parámetros \mathbf{B} .
- Matriz de precisión posterior: $s^{-2} \mathbf{M}$
- Matriz de covarianza posterior: $\left[\frac{\nu}{\nu - 2} \right] s^2 \mathbf{M}^{-1}$

Como podrá notar, la distribución posterior y las distribuciones marginales posteriores son similares a las que se obtuvieron en la sección anterior, por lo que el procedimiento para construir intervalos de credibilidad y la distribución predictiva se efectúa en forma análoga.

Antes de concluir esta sección es importante considerar lo siguiente:

1. Como distribución a priori se tomó la distribución posterior de un análisis bayesiano previo que utilizó una distribución a priori no informativa, esto es importante, debido a que sustenta la validez de los resultados obtenidos. Si el estudio previo se hubiera efectuado con una distribución a priori informativa, entonces los resultados del análisis actual serían distintos.
2. La situación planteada en esta sección ejemplifica la naturaleza iterativa de la distribución posterior, es decir, así como la distribución posterior del estudio previo fue utilizada como distribución a priori en el análisis actual, entonces la distribución posterior obtenida, puede ser utilizada como la distribución a priori de un futuro estudio. Aunque no se debe olvidar, que para efectuar esto, es necesario suponer que la muestra actual es generada por el mismo modelo de regresión utilizado en el estudio previo.
3. La distribución posterior, las distribuciones marginales posteriores están condicionadas tanto en la muestra actual como en la muestra previa. Además como puede notar, la media posterior, la matriz de precisión, la matriz de covarianza de \mathbf{B} , los grados de libertad y s^2 combinan la información de ambas muestras.

4.3 Modelo de Regresión Logística

En el modelo de regresión logística, la variable dependiente es dicótoma, es decir, sólo puede tomar el valor de 1 con probabilidad P , ó el valor de 0 con probabilidad $1 - P$. La media y la varianza de la variable dependiente son: $Ey_i = P_i$ y $Var(y_i) = P_i(1 - P_i)$.

El modelo de regresión logística es el siguiente:

$$P = \frac{e^{XB}}{1 + e^{XB}}$$

donde \mathbf{B} es el vector de los coeficientes de regresión y X es la colección de variables independientes.

Sea (x_i, y_i) , $i = 1, 2, \dots, n$ una muestra de observaciones, la función de verosimilitud es la siguiente:

$$l(\mathbf{B} | \mathbf{y}) = \prod P_i^{y_i} (1 - P_i)^{1 - y_i} \dots \quad 46$$

4.3.1 Análisis con una distribución a priori no informativa

Cuando no se dispone de información acerca de los parámetros β_i , entonces se utiliza la siguiente distribución a priori difusa:

$$p(\mathbf{B}) \propto c$$

donde c es una constante. Lo cual implica que antes de tomar en cuenta la información muestral, todos los valores de cada β_i son igualmente probables.

En este caso, la distribución posterior de \mathbf{B} es la siguiente:

$$\begin{aligned} p(\mathbf{B} | \mathbf{y}, \mathbf{X}) &\propto c \exp\{\ln l(\mathbf{B} | \mathbf{y})\} \\ &\propto \exp\{\ln l(\mathbf{B} | \mathbf{y})\} \end{aligned}$$

Donde $\ln l(\mathbf{B} | \mathbf{y})$ denota el logaritmo natural de la función de verosimilitud dada por la expresión 46.

Una de las técnicas utilizadas para analizar esta distribución posterior es la de expansiones asintóticas, la cual consiste en expandir a $\ln l(\mathbf{B} | \mathbf{y})$ en una serie de Taylor con respecto al valor modal $\hat{\mathbf{B}}$, que es el estimador de máxima verosimilitud. Una vez que se obtiene la respectiva serie de Taylor, se toma el primer término para obtener la siguiente aproximación de la distribución posterior de \mathbf{B} :

$$p(\mathbf{B} | \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -(\mathbf{B} - \hat{\mathbf{B}})' \mathbf{H}(\mathbf{B} - \hat{\mathbf{B}}) / 2 \right\} \dots 47$$

Donde $\mathbf{H} = \mathbf{X}' \mathbf{D} \mathbf{X}$, \mathbf{X} es una matriz de $n \times k$ de observaciones de las variables independientes, mientras que \mathbf{D} es una matriz diagonal con elementos:

$$d_i = \left[y_i / F_i^2 + (1 - y_i) / (1 - F_i)^2 \right] f_i^2 - (y_i - F_i) f'_i / F_i (1 - F_i)$$

$$F_i \equiv F(x'_i \hat{\beta})$$

$$f_i \equiv \left[dF(z_i) / dz_i \right]_{z_i = x'_i \hat{\beta}} \equiv f(x'_i \hat{\beta})$$

$$f'_i \equiv \left[df(z_i) / dz_i \right]_{z_i = x'_i \hat{\beta}}$$

De acuerdo con la expresión 47, \mathbf{B} sigue aproximadamente una distribución normal con media $\hat{\mathbf{B}}$ y matriz de covarianzas \mathbf{H}^{-1} . Cabe señalar que, si se desea obtener una mejor aproximación de la distribución posterior de \mathbf{B} , entonces, se deben tomar en cuenta más términos de la serie de Taylor correspondiente.

Como consecuencia del resultado anterior, la distribución posterior aproximada del logit, $g(x)$ es una Normal, con media $\mathbf{x}'_i \hat{\mathbf{B}}$ y varianza $\mathbf{x}'_i \mathbf{H}^{-1} \mathbf{x}_i$, donde \mathbf{x}'_i es el i -ésimo renglón de la matriz \mathbf{X} .

4.3.2 Análisis con una distribución a priori informativa

Una opción para trabajar con una distribución a priori informativa para \mathbf{B} , consiste en utilizar una distribución Normal con media \mathbf{b} y matriz de covarianzas \mathbf{A}^{-1} , $N(\mathbf{b}, \mathbf{A}^{-1})$, entonces, considerando el logaritmo de la función de verosimilitud, la distribución posterior de \mathbf{B} es la siguiente:

$$p(\mathbf{B} | \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -(\mathbf{B} - \mathbf{b})' \mathbf{A}(\mathbf{B} - \mathbf{b}) / 2 + \log l(\mathbf{B} | \mathbf{y}) \right\} \dots 48$$

De igual forma que en la sección anterior, para analizar esta distribución, se obtiene su respectiva expansión asintótica, en la cual el primer término es el siguiente:

$$p(\mathbf{B} | \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -(\mathbf{B} - \bar{\mathbf{b}})' \mathbf{A}(\mathbf{B} - \bar{\mathbf{b}}) / 2 \right\}$$

Donde:

$$\bar{\mathbf{b}} = (\mathbf{A} + \mathbf{H})^{-1}(\mathbf{A}\mathbf{b} + \mathbf{H}\hat{\mathbf{B}})$$

$$\mathbf{G} = \mathbf{A} + \mathbf{H}$$

$\hat{\mathbf{B}}$: Estimador de máxima verosimilitud

Mientras que \mathbf{H} , se definió en la sección anterior, el valor modal de la distribución es $\bar{\mathbf{b}}$ y la matriz de covarianza es $(\mathbf{A} + \mathbf{H})^{-1}$.

4.4 Radios de probabilidad posterior

Sea β un conjunto de k coeficientes de regresión, que se puede particionar en dos subconjuntos, β_r y $\beta_{r'}$, y considere las siguientes hipótesis:

$$\begin{aligned} H_r &: \beta_r = 0 & -\infty < \beta_{r'} < \infty, 0 < \sigma < \infty \\ H_m &: \beta_r \neq 0 \end{aligned}$$

El radio de probabilidad posterior para H_r relativo a H_m , considerando un radio de probabilidad a priori unitario, es el siguiente:

$$K_{rm} = \frac{\int p(y | \beta_{r'}, \sigma, H_r) p(\beta_{r'}, \sigma | H_r) d\beta_{r'} d\sigma}{\int p(y | \beta, \sigma, H_m) p(\beta, \sigma | H_m) d\beta d\sigma}$$

Donde $p(y | \beta_{r'}, \sigma, H_r)$ y $p(y | \beta, \sigma, H_m)$ son las funciones de verosimilitud bajo las respectivas hipótesis H_r y H_m , mientras que $p(\beta_{r'}, \sigma | H_r)$ y $p(\beta, \sigma | H_m)$ son las distribuciones a priori para los parámetros bajo H_r y H_m .

En el caso del modelo de regresión logística, el radio de probabilidad posterior, se escribe de la siguiente forma:

$$K_{rm} = \frac{\int p(y | \beta_{r'}, \sigma, H_r) p(\beta_{r'}, \sigma | H_r) d\beta_{r'}}{\int p(y | \beta, \sigma, H_m) p(\beta, \sigma | H_m) d\beta}$$

Los radios de probabilidad posterior pueden utilizarse en el procedimiento de selección de modelos, considerando que hay k coeficientes de regresión, entonces asumimos que hay m especificaciones diferentes a considerar, es decir, para cada especificación se tendrán diferentes subconjuntos β_r y $\beta_{r'}$. Ahora considere, la especificación m donde ninguno de los k parámetros está restringido a ser igual a cero, éste será denominado el modelo completo. Si los radios de probabilidad posterior son calculados para las m especificaciones o modelos, entonces, la probabilidad posterior para la i – ésima hipótesis es:

$$p(H_r | y) = \frac{K_{rm}}{\sum_{r=1}^m K_{rm}}$$

De esta forma, se puede visualizar cómo varía la probabilidad posterior para cada modelo, así como las variaciones que existen en los estimadores de los parámetros.

4.5 Observaciones finales

Como pudo notar, cuando se adopta una distribución a priori no informativa, para el MRLS y el MRLM, los estimadores así como los intervalos de credibilidad obtenidos son iguales a los que se obtienen al utilizar la metodología clásica, sin embargo, su interpretación y fundamentos son diferentes.

Otro aspecto que se debe resaltar, es que cuando se emplea una distribución informativa, si el tamaño muestral es pequeño, entonces la distribución a priori tiene mayor peso en los resultados obtenidos, mientras que cuando el tamaño muestral es grande, entonces los datos son los que tienen mayor influencia en los resultados, de aquí la importancia de una buena especificación de la distribución a priori.

Con respecto a las distribuciones posteriores, pudo notar que son centrales en la inferencia acerca de los parámetros. Otra forma para realizar inferencias acerca de los parámetros consiste en obtener las distribuciones posteriores de los parámetros y posteriormente simularlas.

4.6 Aplicaciones

Dentro del ámbito financiero, la metodología bayesiana ha sido empleada para analizar los siguientes problemas: el modelo CAPM, el APT, así como se han aplicado procedimientos de selección de modelos para checar cuáles de todas las variables empleadas para pronosticar el retorno de acciones, en diferentes estudios realmente deben ser incluidas. Algunos de estos estudios ilustran la aplicación de métodos numéricos de integración resaltan la importancia de la especificación de la distribución a priori.

4.6.1 Índice de precios para servicios de radio

Moulton [1991] plantea una aplicación de la aproximación bayesiana para examinar cómo influyen las siguientes características en el precio de los servicios de publicidad en la radio: el tamaño de la audiencia y otras características de la estación, del programa y del anuncio. Para ello, efectuó una regresión de corte transversal para el periodo 0, utilizó datos recolectados por el programa PPI cuando la muestra para el índice de precios de la radio fue iniciado, estos datos consisten en precios y características del anuncio, del anunciante o la estación. La muestra utilizada consta de 87 observaciones.

En el modelo de regresión lineal múltiple que estimó, la variable dependiente, fue el precio del anuncio y como regresores potenciales consideró los siguientes:

1. Tamaño de la audiencia: es tomado de los índices de popularidad, como el promedio de los radioescuchas (edades de 12 en adelante) en el área de cobertura de la estación.
2. Parte del día: mañana, mediodía o tarde.
3. Servicios de producción: no disponible, ninguno, producción incluida.
4. Lugar: no disponible, durante la programación regular, posición constante o rasgos específicos.
5. Unidad: por comercial, simultáneamente en AM y FM.
6. Modulación: no disponible o ambos, AM, FM.
7. Tipo de compra: no disponible, local, anuncio nacional.

8. Semanas de estar al aire: no disponible, 1, de 2 a 4, más de 5 semanas.
9. Tipo de comprador: manufacturero, venta al por menor, servicios, otro, mismo precio a todos los compradores.
10. Estado de sustitución: no disponible, sustituible o garantizado.
11. Duración: no disponible, 60 segundos, 30 segundos, otro.
12. Frecuencia de compra: No disponible, otra, sólo una vez, regular, bajo o volumen promedio.
13. Volumen de compras: mayor a 69 por semana, menor a 13 por semana, mayor o igual a 13 pero menor o igual a 69, no disponible.

Como puede notar, estas variables son categóricas excepto el tamaño de la audiencia, por lo que se utilizaron variables dummy para cada una de las categorías específicas, por ejemplo, para la categoría parte del día, se incluirían dos variable dummy en el modelo de regresión, serían P_1 y P_2 , las cuales asumirán sólo valores de 0 ó 1, cuando la categoría parte del día asuma el valor, mañana, entonces $P_1 = 1$ y $P_2 = 0$, cuando asuma el valor mediodía, entonces $P_1 = 0$ y $P_2 = 1$, y cuando asuma el valor tarde, entonces, $P_1 = 0$ y $P_2 = 0$. El número de variables dummy incluidas por cada categoría, será igual al número de características que presente menos uno, cabe señalar que para la característica no disponible, se asume una variable dummy en cada categoría que la contenga.

En total, se tienen 31 coeficientes de regresión más el intercepto.

Para efectos de estabilizar la varianza y que se cumpla el supuesto de normalidad, Mourton aplicó una transformación a la variable dependiente (precio o cotización del anuncio) y a la variable independiente tamaño de la audiencia, la transformación adecuada fue la raíz cúbica para ambas.

Para aplicar el análisis bayesiano utilizó una distribución a priori no informativa, la función de verosimilitud es una Normal, por lo que la distribuciones marginal para cada parámetro β_i fue una t – Student con 55 grados de libertad, para el modelo completo.

En el análisis del modelo completo, las distribuciones posteriores de muchos coeficientes de regresión, mostraron una gran dispersión, esto como consecuencia de la

sobrep parametrización del modelo, es decir, el número de parámetros es grande relativo al número de observaciones disponibles. La probabilidad posterior de este modelo fue del 0.00002%, la cual es muy baja.

Moulton, calculó los ratios de probabilidad posterior para cada modelo relativo al modelo completo y presentó los nueve con la probabilidad posterior más alta con las respectivas medias y desviaciones posteriores para los coeficientes de regresión de cada modelo. De estos nueve modelos, la probabilidad posterior más alta fue de 1.77% aunque sigue siendo pequeña, se puede notar que es mejor que el modelo completo.

En el modelo con la probabilidad posterior más alta, no se incluyeron las siguientes categorías: servicios de producción, lugar, unidad, frecuencia de compra y volumen de compra. En este caso, los grados de libertad son 69. Otro aspecto importante, es que las desviaciones estándar posteriores de cada parámetro son inferiores a las que se obtuvieron en el modelo completo, del mismo modo, el estimador de la varianza residual es más pequeño.

La variable tamaño de la audiencia fue incluida en todos los modelos.

5. Estudio Comparativo: Modelos Econométricos Clásicos vs Modelos Econométricos Bayesianos

5.1 Introducción

La importancia de conocer herramientas alternativas para resolver un problema, consiste en saber bajo qué circunstancias resulta adecuado aplicar cada una de ellas, esto puede determinarse, una vez que se conocen las ventajas y desventajas que ofrecen así como las características que presenta nuestro problema.

En la primera parte de este capítulo, se presenta una aplicación en la que se compararán los resultados obtenidos bajo la metodología clásica y la bayesiana, esto nos

permitirá visualizar los cambios que hay en los estimadores cuando se utiliza la información a priori existente.

Posteriormente, se resaltarán la importancia del tamaño muestral en cada metodología, este es un punto importante, ya que en ocasiones, no se dispone de suficiente información muestral para estimar los parámetros involucrados en el modelo.

Por último, se presentarán las ventajas y desventajas que ofrecen tanto los modelos econométricos clásicos como los Bayesianos, de acuerdo con lo que se investigó en el presente trabajo, la idea, es que en función de esto, se pueda determinar cuándo resulta conveniente aplicar cada una de las metodologías presentadas en capítulos anteriores.

5.2 Estudio comparativo: valuación de bienes raíces⁵

Se trata de un estudio realizado en California, donde cada año, asesores de bienes raíces, son requeridos para estimar el valor de mercado actual de todas las propiedades gravables en sus respectivos distritos. El estudio se limita a analizar las casas pertenecientes a una sola familia.

La tarea del asesor consiste en estimar el valor de mercado actual de la casa, de acuerdo a las características que presenta. Sea y_e , el valor estimado por el asesor, y_a , el valor actual de la casa y $L(y_e, y_a)$ la pérdida incurrida. Cabe señalar que, sólo se consideran las pérdidas atribuidas al asesor. En el estudio pueden presentarse los siguientes casos:

1. Que el valor de la casa sea subestimado por el asesor, entonces, la pérdida equivale a la cantidad de la subestimación.
2. Que el valor de la casa sea sobreestimado por el asesor, entonces, el propietario, puede quejarse con el asesor y convencerlo de que su casa ha sido sobrevaluada, o en su defecto, puede presentar una queja al Consejo de

⁵ Varian, Hal R.[1974]

Equidad de Evaluación, el cual determina quién tiene la razón. Si el asesor pierde el caso, pierde la cantidad de la sobreestimación y los costos del proceso de apelación. Si gana, el distrito debe pagar a la corte los costos involucrados.

De acuerdo con esto, el asesor desea un estimador del valor actual de la casa que minimice la pérdida total esperada, por lo que, la función de pérdida no puede ser cuadrática, debido a que asigna la misma pérdida a sobreestimar o subestimar el valor de la casa. El autor sugiere la siguiente función de pérdida asimétrica:

$$L(y_e, y_a) = b \exp[a(y_e - y_a)] - c(y_e - y_a) - b$$

a la cual denominó, función Linex. Esta función de pérdida se ajusta mejor al problema. Los valores de a, b y c sugeridos por el autor, son los siguientes: $a=0.0004$, $b=2500$ y $c=1$.

El estimador que minimice la función de pérdida, es aquel valor de y_e que minimice, la pérdida esperada, que está dada por la siguiente expresión:

$$E[L] = \int_{-\infty}^{\infty} L(y_e, y_a) p(y_a) dy_a$$

donde, $p(y_a)$ es la función de distribución de y_a .

5.2.1 El modelo

Se asume que el precio de venta (o el valor de mercado) de una casa, es una función lineal de las características de la casa más un error, por lo que se puede ajustar al modelo de regresión lineal múltiple. Las variables independientes o explicativas, fueron sugeridas por un estudio previo, en donde se encontró que son buenos predictores del precio de venta, y son las siguientes:

X_1 := Ancho del lote

X_2 := Vista (1:= vista promedio o arriba del promedio, 0:= vista debajo del promedio)

X_3 : = Edad

X_4 : = Número de recámaras

X_5 : = Calidad

X_6 : = Área total de la vivienda

X_7 : = Costo de construcción

X_8 : = Costo de la cerca

Entonces, el modelo queda de la siguiente forma:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}$$

donde:

\mathbf{Y} : es un vector de $n \times 1$ de observaciones de los precios de venta de las casas.

\mathbf{X} : es una matriz de $n \times 9$ de observaciones de las características que se acaban de mencionar.

\mathbf{B} : vector de 9×1 de los coeficientes de regresión.

\mathbf{E} : vector de $n \times 1$ de errores.

Los supuestos del modelo de regresión lineal múltiple, pueden consultarse en los capítulos 2 y 4.

Los datos para este estudio, consisten en una muestra de 168 observaciones de ventas de casas en una ciudad de tamaño mediano de California, que se efectuaron en 1965. Esta muestra se dividió en dos grupos, el primero de 125 observaciones, las cuales se utilizaron para estimar los parámetros, y el segundo de 43 observaciones, que se reservaron para proyecciones y comparaciones de propuestas.

5.2.2 Distribución a priori

En este estudio, la información a priori acerca de \mathbf{B} , puede provenir de las siguientes fuentes:

1. **Reglas de evaluación.** Este tipo de información a priori, proviene de la experiencia del evaluador y consiste en juicios o evaluaciones subjetivas acerca de la importancia relativa de los factores que determinan el valor de la casa. Un ejemplo de tales juicios o evaluaciones subjetivas sería: “Se esperaría que el área total de la vivienda sea más importante que el costo de la cerca, y además, que su respectivo coeficiente sea positivo”. Una vez que se dispone de este tipo de evaluaciones, se utilizan para construir la respectiva distribución de probabilidad a priori.
2. **Experiencia previa.** En este caso, si se dispone de algún estudio previo similar, que se haya efectuado exitosamente en otra comunidad, entonces, los estimadores obtenidos en dicho estudio, pueden ser utilizados para construir la distribución a priori, y de esta forma ayudar a determinar con mayor exactitud los estimadores para el estudio actual.
3. **Estimadores históricos.** Puede suceder que se haya efectuado un estudio previo en la misma comunidad y que ahora se desee, incorporar la nueva información disponible, para actualizar los estimadores de los parámetros. Debido a que las condiciones de mercado cambian lentamente entonces, si el estudio previo fue efectuado un año antes, puede considerarse que es probable los estimadores del estudio actual, sean similares a los obtenidos en el estudio previo, entonces, estos estimadores pueden utilizarse como parámetros de la distribución a priori.

En este estudio, se utilizaron los siguientes dos tipos de distribuciones a priori:

1. **Distribución a priori no basada en datos:** Se trata de una distribución Normal – Gamma Inversa, los hiperparámetros fueron determinados a partir de reglas de evaluación.
2. **Distribución a priori basada en datos:** En este caso, se utilizaron las 43 observaciones que se reservaron de la muestra disponible, para construir la distribución a priori, por lo que la aproximación utilizada fue la de estimadores históricos o la de experiencia previa, ya que se asume que las 43

Estudio Comparativo: Modelos Econométricos Clásicos vs Modelos Econométricos
Bayesianos

observaciones constituyen una muestra independiente, posteriormente se efectúa la regresión y los estimadores obtenidos se utilizan como hiperparámetros de la distribución a priori, la cual es una Normal – Gamma Inversa.

La siguiente tabla muestra los hiperparámetros asignados a cada distribución a priori.

Variables	Distribución a priori no basada en datos	Distribución a priori basada en datos
Ancho del lote	50	98.08 (31.16)
Vista	2000	2405.24 (828.87)
Edad	-200	-139.22 (48.21)
Número de recámaras	1000	-12.04 (741.66)
Calidad	2000	2454.36 (626.19)
Área total de la vivienda	10	6.12 (1.56)
Costo de construcción	5	4.52 (3.50)

Variables	Distribución a priori no basada en datos	Distribución a priori basada en datos
Costo de la cerca	2	-0.31 (5.78)
Constante	1000	-4413.44 (4016.91)

Las cantidades que aparecen entre paréntesis es la desviación estándar del parámetro respectivo.

5.2.3 Resultados

A continuación se muestran los resultados obtenidos utilizando la metodología clásica (método de mínimos cuadrados), la distribución a priori no basada en datos y la distribución a priori basada en datos. Nuevamente, la cantidad entre paréntesis se refiere a la desviación estándar del parámetro respectivo.

Variables	Mínimos cuadrados ordinarios	Regresión bayesiana no basada en datos	Regresión bayesiana basada en datos
Ancho del lote	33.11 (20.30)	41.55 (31.18)	53.88 (15.97)
Vista	1879.27 (433.21)	1939.64 (665.25)	2109.04 (379.04)
Edad	-175.34 (27.11)	-187.67 (41.64)	-173.50 (23.02)

Estudio Comparativo: Modelos Econométricos Clásicos vs Modelos Econométricos
Bayesianos

Variables	Mínimos cuadrados ordinarios	Regresión bayesiana no basada en datos	Regresión bayesiana basada en datos
Número de recámaras	-722.48 (32.35)	138.76 (496.74)	-608.90 (289.63)
Calidad	2000.54 (33.61)	2002.73 (516.17)	2119.04 (285.23)
Área total de la vivienda	7.95 (0.78)	8.98 (1.20)	7.33 (0.67)
Costo de construcción	5.05 (1.29)	5.03 (1.98)	4.48 (1.18)
Costo de la cerca	3.89 (1.80)	2.94 (2.80)	3.51 (1.72)
Constante	1356.11 (1840.49)	1178.05 (2826.29)	161.53 (1641.28)

El autor considera que los resultados obtenidos utilizando mínimos cuadrados ordinarios, parecen razonables, excepto el del número de recámaras, ya que se esperaba que el signo de este coeficiente fuera positivo.

Cuando se utiliza una distribución a priori informativa no basada en datos, el signo del coeficiente del número de recámaras es positivo, sin embargo, la desviación estándar obtenida para cada parámetro es más alta que la que se obtuvo utilizando mínimos cuadrados ordinarios, por lo que su capacidad predictiva es más pobre. Una de las causas de estos resultados puede ser una mala especificación de los hiperparámetros de la distribución a priori. De acuerdo con esto, y omitiendo el signo negativo del coeficiente del

número de recámaras, entonces, de estas dos metodologías, la mejor es utilizar mínimos cuadrados ordinarios.

En el caso de la distribución a priori informativa basada en datos si comparamos los resultados que se obtuvieron al efectuar la regresión con los hiperparámetros de la distribución a priori, se observa una mejoría en los estimadores, además de que la desviación estándar de cada parámetro es menor. Ahora, si comparamos los resultados con los obtenidos en mínimos cuadrados ordinarios, se observa que los estimadores mejoraron y las desviaciones estándar disminuyeron, excepto para dos coeficientes, además nuevamente el signo del coeficiente del número de recámaras es negativo, sin embargo, la desviación estándar global, es ligeramente más alta que la de mínimos cuadrados ordinarios. En este caso, el precio de optar por utilizar la metodología bayesiana sería tener una desviación estándar global ligeramente más alta.

En este ejemplo pudo notar, las consecuencias de utilizar una distribución a priori mal especificada, por lo que se debe tener cuidado al interpretar y darle el peso correspondiente a los juicios emitidos por el experto, para construir la distribución a priori, y que los resultados obtenidos al efectuar la regresión sean razonables. Para este caso particular, los resultados obtenidos por mínimos cuadrados y la regresión bayesiana utilizando una distribución a priori informativa basada en datos, fueron similares casi para todos los coeficientes, y la desviación estándar utilizando la metodología bayesiana fue disminuida, aunque a nivel global es ligeramente más alta que la de mínimos cuadrados.

5.3 Importancia del tamaño muestral

Uno de los requerimientos tanto para la metodología clásica como para la bayesiana con respecto al tamaño muestral, es que el número de observaciones, sea mayor que el número de parámetros a estimar, pero ¿cómo afecta el tamaño muestral en los resultados obtenidos en cada metodología? Al trabajar con la metodología clásica, la única fuente de información que se considera el estimar los parámetros es la muestral, por lo que cuando el tamaño muestral apenas supere el número de parámetros a estimar, los resultados obtenidos

no serán óptimos y las respectivas desviaciones estándar suelen ser muy altas, como consecuencia la capacidad predictiva del modelo estimado no es muy buena.

Como ya se vio, en la metodología bayesiana, las dos fuentes de información que se incluyen en el análisis del modelo son la información a priori y la muestral, por lo que cuando el tamaño muestral es pequeño, entonces, la distribución a priori predomina en los resultados obtenidos, mientras que cuando la muestra es grande, el papel de la distribución a priori en el análisis se ve minimizado, los datos predominan en los resultados obtenidos. Entonces, cuando el tamaño muestral sea pequeño, se debe tener mayor cuidado al especificar la distribución a priori ya que como se vio en la sección anterior, una mala especificación influye en la calidad de los resultados. Una de las ventajas que ofrece la metodología bayesiana, es su recursividad, es decir, puede suceder que al efectuarse un análisis bayesiano, se disponga de poca información muestral y los resultados obtenidos no sean muy buenos, pero conforme se disponga de nueva información ésta puede ser incorporada en el análisis y de esta forma, se mejora la estimación del modelo.

Para visualizar cómo afecta el tamaño muestral en los resultados de ambas metodologías, una alternativa, consiste en analizar el problema con ambas y posteriormente comparar los estimadores obtenidos, sus respectivas desviaciones estándar, checar su capacidad predictiva y observar cómo se ven afectados los resultados al incluir la información a priori disponible.

5.4 Modelos Econométricos Clásicos vs Modelos Econométricos Bayesianos

De acuerdo con lo que ya se ha analizado, una de las preguntas que se formularía es la siguiente ¿Cuál de las dos metodologías es mejor? La respuesta adecuada es: lo importante radica no en saber cuál es la mejor sino en saber aplicar e interpretar los resultados que se obtienen en cada metodología, conocer las condiciones bajo las cuales, dichos resultados son razonables así como las ventajas y desventajas que ofrece cada metodología, para que de acuerdo a las características que presenta el problema a analizar, podamos determinar

cuál de las dos es adecuada, o si se puede y se desea, se efectúa un análisis bajo ambas metodologías para visualizar los cambios existentes en los estimadores. La metodología Bayesiana puede ser considerada como una herramienta alternativa que nos permite conceptualizar la solución del problema de una forma diferente.

Pero, ¿Cuándo utilizar cada metodología? Consideremos, primero el caso cuando el problema a analizar dispone de información muestral y de la experiencia de un experto o de un estudio previo:

1. Bajo la metodología clásica, no existe un mecanismo formal para incorporar la información a priori, aunque se disponen de algunas herramientas que nos permiten incorporar parte de la información a priori, pero no son muy conocidas, por lo que si la persona que efectúe el análisis no las conoce optará por no incluir dicha información.
2. La metodología bayesiana provee un mecanismo formal para incluir la información a priori, y siempre se hace de la misma forma, por lo que siempre se sabe cómo incluirla. Aunque se debe tener cuidado al determinar la distribución a priori utilizando la información a priori existente, ya que una mala especificación se verá reflejada en la calidad de los resultados obtenidos, como se vio en la segunda sección de este capítulo.

La importancia de incluir la información a priori en el análisis, radica en que es útil para determinar el valor de los estimadores de manera más precisa, siempre y cuando se elija una distribución a priori adecuada al problema. En este trabajo se presentó el análisis bayesiano utilizando una distribución a priori informativa conjugada, debido a la facilidad de los cálculos, sin embargo, este tipo de distribuciones, no siempre se ajusta a la realidad del problema, por lo que en ocasiones es necesario especificar una distribución a priori que se ajuste mejor a la realidad, afortunadamente ya existe software para construir distribuciones a priori y para efectuar análisis bayesiano, lo que facilita el cálculo de las distribuciones posteriores.

Y ¿qué pasa cuando no se dispone de información a priori? En este caso, la metodología bayesiana emplea una distribución a priori no informativa, los resultados son

Estudio Comparativo: Modelos Econométricos Clásicos vs Modelos Econométricos Bayesianos

iguales a los que se obtienen utilizando la metodología clásica, aunque la interpretación no es la misma.

Si desea visualizar cómo afecta utilizar diferentes tipos de distribuciones a priori y la metodología clásica en los estimadores, puede efectuar un análisis de sensibilidad global y comparar los estimadores, las desviaciones estándar y la capacidad predictiva. El análisis de sensibilidad global, consiste en especificar diferentes distribuciones a priori y visualizar los cambios que existen en los estimadores. Y si desea checar cómo afecta la elección de los hiperparámetros de la distribución a priori, en los estimadores, entonces se efectúa un análisis de sensibilidad local, y se visualiza cómo se ven afectados los estimadores ante pequeños cambios en los hiperparámetros. Esto nos dará una mejor visualización de los resultados y nos permitirá tomar una mejor decisión en la elección del modelo.

Conclusiones

Una de las tareas de un egresado de MAC, específicamente del área de simulación, es proponer modelos probabilísticos y porque no matemáticos para entender cómo se comporta la realidad y de esta manera, tomar decisiones, que es lo que hace el ser humano a diario. La creación de estos modelos es mediante una metodología recursiva, que consiste en tomar datos de la realidad, con ellos crear un modelo que se ajuste a ellos, posteriormente, se efectúa una simulación o pronóstico y finalmente se comparan los resultados obtenidos con los de la realidad, si el modelo aún es pobre, entonces vuelvo al paso del modelado, modifico el modelo y vuelvo a realizar el proceso, hasta obtener un modelo que se aproxime lo mejor posible a la realidad.

Dentro del ámbito económico, existen muchos modelos interesantes, entre los que se encuentran, los financieros, que pueden ser analizados desde el punto de vista estadístico, utilizando la metodología Clásica y la Bayesiana para determinar bajo qué circunstancias trabajan bien los modelos y porque no, proponer modificaciones al modelo para explicar mejor el comportamiento del mercado.

La metodología Bayesiana es un área bastante amplia, la cual considero que podemos explotar, lo que se presentó en este trabajo es una pequeña introducción con la intención de despertar el interés del lector para proponer otras líneas de investigación.

Uno de los objetivos al inicio del trabajo fue tratar de generar nuevas líneas de investigación, por ello, es importante mencionar algunas de las que encontré, las cuales son:

1. Como ya se mencionó, la distribución a priori juega un papel importante en la metodología bayesiana, aunque su especificación no es una tarea fácil, en torno a este punto, podría analizarse la posibilidad de elaborar un programa que mediante una serie de preguntas pueda proponer la forma de la distribución a priori.
2. Otra alternativa, consiste en analizar los métodos de simulación de la distribución posterior.
3. Analizar el diseño de las redes neuronales bayesianas, su funcionamiento y limitaciones en la solución de problemas, así como de presentar las áreas en las que han sido aplicadas y los resultados que se han obtenido.

Considero que es importante conocer a fondo las metodologías que nos enseñan en la escuela e investigar acerca de las alternativas que existen, esto nos dará amplitud de criterio al decidir que metodología utilizar de acuerdo al problema que estemos analizando. De igual forma, es importante verificar la validez de los supuestos que sustentan a los modelos que utilicemos, puesto que esto influirá en la calidad de los resultados obtenidos, en caso de que alguno de los supuestos no se cumpla, entonces, es necesario considerar herramientas alternativas que nos permitan resolver el problema en cuestión.

Finalmente, es importante resaltar, que la intención del presente trabajo es presentar a la metodología bayesiana como una herramienta alternativa, a la Clásica, en el análisis de la información económica, ya que ambas tienen sus ventajas y desventajas, considero que ninguna de las dos es mejor que la otra, simplemente, nos permiten analizar el problema con un enfoque diferente.

Apéndice 1: Inversos Generalizados

Cuando la matriz \mathbf{A} no es de rango completo, entonces su respectiva matriz inversa, bajo la definición usual, no existe, en este caso, se puede calcular su inverso generalizado.

Un inverso generalizado de una matriz \mathbf{A} de tamaño $n \times p$, es una matriz \mathbf{A}^- que cumple con la siguiente condición: $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$.

Como puede notar, esta definición aplica tanto para matrices cuadradas como para rectangulares.

Cabe señalar que, un inverso generalizado no es único, excepto cuando \mathbf{A} es no singular, entonces: $\mathbf{A}^- = \mathbf{A}^{-1}$.

Una forma de obtener un inverso generalizado para la matriz \mathbf{A} , consiste en invertir una submatriz, para ello, se particiona a \mathbf{A} en submatrices, de la siguiente forma:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

Ahora suponga que la submatriz principal \mathbf{A}_{11} , es no singular y cuyo rango es el mismo de la matriz \mathbf{A} , $r_{\mathbf{A}}$. Entonces un inverso generalizado de la matriz \mathbf{A} se puede obtener de la siguiente forma:

$$\mathbf{A}^- = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

donde \mathbf{A}_{11}^{-1} es la matriz inversa de \mathbf{A}_{11} , y las matrices nulas son del tamaño apropiado. Aunque se debe señalar que la matriz \mathbf{A}_{22} se puede utilizar de la misma forma, para obtener un inverso generalizado.

Otra definición importante es la de la inversa de Moore – Penrose que se presenta a continuación:

Inversa de Moore – Penrose: Dada una matriz \mathbf{A} , existe una matriz única \mathbf{M} , que cumple con las siguientes condiciones:

$$\mathbf{AMA} = \mathbf{A}$$

$$\mathbf{MAM} = \mathbf{M}$$

\mathbf{AM} es simétrica

\mathbf{MA} es simétrica

Estas son las condiciones de Penrose, mientras que a \mathbf{M} se le denomina inversa de Moore – Penrose.

GLOSARIO

Distribución a posteriori: Resume el conocimiento que se tiene acerca de los parámetros después de haber observado los datos. Es el resultado de combinar la información a priori con la muestral, de manera que puede considerarse como una regla de actualización, donde los datos nos permiten actualizar nuestro conocimiento a priori acerca de los parámetros.

Distribución a priori: Captura la información a priori disponible acerca de los parámetros del modelo antes de recoger o ver los datos. La información a priori o inicial, puede provenir de la experiencia de un experto en el área, de consideraciones teóricas o de resultados de algún experimento previo.

Distribución a priori conjugada: Es aquella distribución que combinada con la función de verosimilitud, da como resultado una distribución posterior que pertenece a la misma familia o clase de la distribución a priori.

Distribución a priori informativa: Aquella distribución a priori que captura la información a priori existente respecto a parámetro, la cual, no depende de la muestra de datos actual. Esta distribución se construye utilizando resultados provenientes de un estudio previo similar, consideraciones teóricas o la experiencia de un experto en el área.

Distribución a priori no informativa: Es una distribución a priori estándar que se utiliza cuando no se cuenta con información a priori acerca del parámetro (resultados de un estudio previo, consideraciones teóricas, la experiencia de un experto en el área), esta distribución asume que antes de tomar en cuenta la información muestral todos los valores del parámetro son equiprobables.

Econometría: Se encarga de analizar, interpretar y pronosticar cuantitativamente los fenómenos económicos aplicando conocimientos del área de Cálculo, Probabilidad e Inferencia Estadística. A menudo se utiliza para probar o refutar teorías económicas. Entre las principales herramientas que emplea se encuentran los siguientes modelos: de regresión, de ecuaciones simultáneas, series de tiempo, entre otros.

Economía: Es la ciencia que se encarga de estudiar la asignación de los recursos escasos entre diferentes actividades con el objetivo de obtener bienes y servicios que son distribuidos entre los individuos para la satisfacción de sus necesidades.

Espacio paramétrico: Contiene todos los posibles valores que puede adoptar el parámetro θ y se denota por la letra Ω .

Estadístico: Es una función de las variables aleatorias observables y que a su vez es una variable aleatoria observable que no contiene parámetros desconocidos.

Estimación: Es una realización del estimador, es decir, es el valor numérico que toma el estimador cuando se sustituye la información muestral disponible en la función.

Estimador: Algún estadístico $\hat{\theta}$ cuyos valores pueden ser utilizados para estimar a θ . Como es una función de la información muestral, entonces, un estimador es una variable aleatoria.

Función de Verosimilitud: La función de verosimilitud de una muestra aleatoria proveniente de la densidad $f(x; \theta)$ es la densidad conjunta de las variables muestrales, la cual también depende del parámetro θ , se representa de la siguiente manera:
$$g(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdots f(x_n; \theta)$$

Heterocedasticidad: Cuando la varianza para cada error no es la misma, es decir, cambia para cada uno de ellos: $E(\varepsilon_i) = \sigma_i^2$.

Homocedasticidad: Cuando la varianza de los errores es constante.

Macroeconomía: Parte de la Economía que se encarga de analizar el funcionamiento de la economía en su conjunto, es decir, en lugar de enfocarse al comportamiento de las unidades económicas, dirige su atención a la suma de las decisiones de dichas unidades.

Microeconomía: Parte de la Economía que se encarga de analizar el comportamiento de las unidades productivas y del consumidor individual, es decir, se encarga de estudiar el comportamiento dentro del mercado de los grupos pequeños, tales como, asociaciones, la familia y la industria.

Riesgo sistemático: Es la parte del riesgo que puede ser eliminada por medio de la diversificación

Riesgo no sistemático: Es la parte del riesgo que se atribuye a factores políticos, inflación guerras, entre otros, que afectan a todas las acciones y que no puede ser eliminado por medio de la diversificación.

Bibliografía

Bauwens, Luc, Michel Lubrano and Jean – F. Richard.(1999). *Bayesian Inference in Dynamic Econometric Models*, Oxford University.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd edition. Springer – Verlag. New York

Birkes, D. and Dodge Y.(1993). *Alternative Methods of Regression*. John Wiley & Sons.

Box G.E.P., Tiao G. C. (1973) *Bayesian Inference in Statistical Analysis*. Addison-Wesley Publishing Company, Boston, Mass.

Congdon, Peter (2003). *Bayesian Statistical Modelling*. John Wiley & Sons

Draper, Norman Richard. *Applied Regression Analysis*. 3rd edition. John Wiley & Sons.

Gelman, A., John B. Carlin, Hal S. Stern and Donald Rubin (2004). *Bayesian Data Analysis*. 2nd edition Chapman & Hall.

Gibbons, M.R., (1982), *Multivariate Test of Financial Models: A new Approach*. Journal of Financial Economics, 10 3-28

Goldberger, Arthur S. (1964). *Econometric Theory*. John Wiley & Sons.

- Harvey, C.R. and G, Zhou, 1990, *Bayesian Inference in Asset Pricing Test*. Journal of Financial Economics, 26, 221-254
- Jobson J D and B Korkie, 1982, *Potential Performance and Test of Portfolio Efficiency*. Journal of Financial Economics, 10. 433-466
- Judge,G.G., W. E. Griffiths, R.C., Hill, H. Lütkephol and T. C. Lee.(1985). *The theory and practice of Econometrics*, 2nd edition.
- Hosmer, David W. and Stanley Lemeshow (2000). *Applied Logistic Regression*. 2nd edition. John Wiley & Sons.
- Keinbaum, David G. (1987). *Applied Regression Analysis and other Multivariate Methods*.
- Levy, H. (1978). *Equilibrium in an Imperfect Market: a constraint on the number of securities in the portfolio*. American Economic Review, vol. 68, no. 4, 643 – 658.
- Martijn, Cremers K.J. (2002) *Stock Return Predictability: A bayesian Model Selection Perspective*. The review of Financial Studies, vol. 5, núm. 4, 1223 – 1247.
- McCulloh R., and P. E. Rossi, 1991, *A Bayesian Approach, to testing Arbitrage Pricing Theory*, Journal Of Econometrics, 49, 141-168
- Mood, Alexander M. and Franklin A. Graybill (1976) *Introducción a la teoría de la Estadística*. 4^{ta} edición.
- Montgomery, D. C. and Elizabeth A. Peck (1992). *Introduction to Linear Regression analysis*. 2nd edition, Wiley
- Moulton, R. Brent (1991). *A Bayesian Approach to Regression Selection and Estimation with Application to a Price Index for Radio Services*. Journal Of Econometrics, 49, 169-193
- Press, S.J. (1989). *Bayesian Statistics: Principles, Models and Applications*, Wiley, NY
- Rencher, Alvin C. (2000). *Linear Models in Statistics*. John Wiley & Sons.
- Robert, Christian P. (2001). *The Bayesian Choice*. 2nd edition. Springer – Verlag. New York

Rodríguez, López Juan Carlos (2005). *Catálogo de Pruebas Paramétricas y no Paramétricas*, Tesis para obtener el grado de Licenciatura en Matemáticas Aplicadas y Computación, FES Acatlán.

Searle, Shayle R. and Lois Schertz Willett. *Matrix Algebra for Applied Economics*. John Wiley & Sons.

Tiao G. C., Zellner, A. 1964, *Bayes Theorem and the Use of Prior Knowledge in Regression Analysis*, *Biometrika*, 51, 219-230

Varian, H.R.(1975). *A Bayesian Approach to Real Estate of Assessment in*: S. E. Fienberg and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics in honor of Leonard J. Savage* (North – Holland, Amsterdam), 195 –208

Winkler, Robert L. and Christopher B. Barry (1975). *A Bayesian Model for Portfolio Selection and Revision*. *The journal of Finance*, vol 30,Núm. 1, 179 -192

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York.

Zellner, A. and P. E. Rossi. (1984). *Bayesian Analysis of Dichotomous Quantal Response Models*, *Journal of Econometrics* 25, 365-395

Zellner, A. (1988). *Bayesian Analysis in Econometrics*, *Journal of Econometrics* 37, 27-50

Zhou G (1991), *Small Sample Test of Portfolio Efficiency*. *Journal of Financial Economics*, 30, 165-191.