75
2ej

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

" INTERVALOS DE CONFIANZA EN REGRESION LINEAL CON VARIABLES INDEPENDIENTES ALEATORIA "

T E S I S

QUE PARA OBTENER EL TITULO DE

A C T U A R I A

P R E S E N T A :

CRISTINA ORTUÑO MOJICA

1997

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

M. en C. Virginia Abrín Batule
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
P r e s e n t e

Comunicamos a usted que hemos revisado el trabajo de Tesis:

"Intervalos de confianza en regresión lineal con variables independientes aleatorias"

realizado por Ortuño Mojica Cristina

con número de cuenta 8736454-7 , pasante de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	Dr. Federico O'Reilly Togno
Propietario	Dr. Eduardo Arturo Gutiérrez Peña
Propietario	Dr. José Rodolfo Mendoza Blanco
Suplente	M. en C. José Salvador Zamora Muñoz
Suplente	M. en C. José Antonio Flores Díaz

Consejo Departamental de Matemáticas

! Por fin !

Pero antes de poner fin a esta novela, es necesario agradecer a todos los que aparecieron en la trama.

Por orden de aparición

A mis padres :

Por todo lo que han hecho para que yo pueda estar aquí Gracias mamá, por tu gran apoyo en todos los momentos (buenos y malos). Gracias papá.

A todos mis hermanos, Mimi, Bety, Homar, Mary, Caro y Jorge, simple y sencillamente, por ser como son.

A J.^{2} M. S. por TODO.

A los cuates.

A Gabriel Huerta por inducirme al vicio de la Estadística.

A Federico O'Reilly por haberme tenido paciencia.

A toda la gente del departamento de Estadística del IIMAS, que mucho tiene que ver con mi poca o mucha *de-formación* estadística. En especial a Salvador Zamora, Raúl Rueda y Eduardo Gutiérrez por sus sugerencias, y a los que tanto insistieron, para que terminara de ocupar la Silicon y los dejara chambear (en especial a ellas).

Al Instituto de Investigación en Matemáticas Aplicadas y Sistemas (IIMAS) y a la Unidad Académica de los Ciclos Profesional y de Posgrado (UACPyP) del C.C.H., por permitirme hacer uso de sus instalaciones.

A los sinodales, por comprender la urgencia de estos momentos y por su colaboración.

A todos los que he olvidado mencionar y algo tengan que ver con esto.

Contenido

Capítulo 1

Introducción

Uno de los modelos estadísticos ampliamente utilizado en aplicaciones es el modelo de regresión, en el cual es usual considerar que la medición de la variable dependiente (o respuesta) se hace con error y que las variables independientes (o regresores) se miden sin error.

Existe una gran cantidad de literatura para modelos de regresión, dentro de la cual destaca la de los modelos lineales, en los que, como su nombre lo indica, la relación entre el valor esperado de la variable dependiente y las variables independientes se modela por medio de una relación lineal, tomando en cuenta que la linealidad se da en los parámetros y no en las variables.

Los modelos de regresión lineal son importantes debido a que la relación que tratan de describir se puede aproximar localmente y con base en ésto predecir valores fuera del rango de las variables analizadas (aunque las predicciones serán confiables sólo para valores en un rango limitado), además de ésto, la interpretación de los parámetros es más sencilla por incorporarse al modelo en forma lineal.

El análisis usual supone que la distribución de los errores en la medición de la variable dependiente es normal (gaussiana). Dicha suposición se ampara en que en cierto modo, los errores son el resultado de agregar muchas imprecisiones y por ello el Teorema Central del Límite sugiere que ese agregado debiera ser normal.

En algunos casos el suponer que las variables independientes se miden sin error no es lo más conveniente ya que pueden presentar un comportamiento aleatorio y es importante que se vea reflejada en el análisis, la información que pueda producir este hecho.

Se abordará el caso lineal simple, centrándose en el parámetro asociado a la pendiente del modelo (un análisis análogo se puede desarrollar para el parámetro relacionado con la ordenada al origen). Para este parámetro es de interés comparar el intervalo de confianza obtenido en la forma acostumbrada contra el que se obtiene al considerar que la variable independiente es aleatoria. Es decir, al no condicionar el análisis al valor de la variable independiente, que en este trabajo se definen como *análisis condicional* e *incondicional*, respectivamente.

Esta comparación servirá para tratar de determinar si realmente es conveniente utilizar un análisis incondicional cuando la variable independiente es aleatoria o si se debe realizar un análisis condicional.

El desarrollo de este trabajo inició con la búsqueda bibliográfica. Los artículos más relacionados con el problema fueron, Sampson (1974), Kinal and Lahiri (1983), Binkley and Abbott (1987), Judge et al (1988) y Benjamini and Fuchs (1990). Sin embargo, el enfoque que se aborda en estos trabajos está orientado a la reducción de varianza principalmente. Por esta razón se decidió iniciar el estudio desde la parte condicional y con base en ella desarrollar la parte incondicional por medio del Teorema de Bayes, usando además un enfoque fiducial como el que se utiliza en Sprott and Farewell (1993).

La distribución de este trabajo es la siguiente :

- En el **capítulo 2** se explica brevemente el modelo de regresión lineal simple. Debido a que este modelo es tratado ampliamente en la literatura, no se presenta una descripción detallada del mismo, empero, se describe cómo se obtienen los estimadores de los parámetros de la regresión al no suponer una distribución en los errores, como al suponer una distribución normal. Una vez asociada una distribución a los errores, se desarrolla la obtención de intervalos de confianza. Por último, se plantea la posibilidad de que la variable independiente sea aleatoria y se realice un análisis incondicional.

• En el **capítulo 3** se obtiene la distribución incondicional inducida del parámetro de interés. Debido a que no se pudo expresar esta distribución en forma explícita, es decir, no se obtuvo una *expresión cerrada*, se recurrió a evaluarla en forma numérica, para obtener el intervalo de confianza de longitud mínima.

• En el **capítulo 4** se describe el programa que se utilizó para determinar el intervalo de longitud mínima del parámetro estudiado.

• En el **capítulo 5** se encuentran las conclusiones de este trabajo.

• En el **apéndice A** presenta el programa escrito en lenguaje Fortran que utiliza algunas subrutinas de IMSL[1].

• En el **apéndice B** se presentan los resultados de los casos considerados en las simulaciones.

• En el **apéndice C** se presenta el desarrollo algebraico para evaluar la función de densidad numéricamente.

[1]International Mathematical System Library.

Capítulo 2

Modelo de Regresión Lineal Simple

2.1 Descripción del modelo

En muchos problemas prácticos o de investigación es de interés analizar cómo cambios en una variable afectan cuantitativamente a otra variable. Estos cambios en algunas ocasiones se presentan como una relación funcional simple entre las variables o bien, como una relación funcional más complicada. En el caso de que la relación sea complicada, se puede aproximar por una relación matemática simple, como puede ser un polinomio, que involucre a las variables de interés y que se aproxime a la verdadera relación funcional en un rango limitado de las variables analizadas.

Si la relación matemática que se desea aproximar es una linea recta, se podrá considerar un análisis de regresión lineal. En el cual se define como variable independiente a la variable en la que se dan los cambios y como variable dependiente a aquella en la que ocurre algún efecto al presentarse los cambios en la variable independiente, es decir, se tiene la siguiente relación

$$Y = \beta_0 + \beta_1 X \tag{2.1}$$

donde X es la variable independiente, Y la variable dependiente, β_0 la ordenada al origen y β_1 la pendiente, los parámetros β_0 y β_1 son desconocidos.

Aún cuando la verdadera relación funcional entre las dos variables sea una linea recta, al tomar observaciones de estas variables y graficarlas, probablemente no se obtendrá exactamente una linea recta, pero las observaciones se encontrarán en forma aleatoria alrededor de la misma.

Este comportamiento se presenta en los datos por defectos en la medición ya sea por errores en los aparatos de medición o bien por el efecto de otras variables que no se controlaron al realizar el análisis, éste es un error para el que no se tiene un control y por lo mismo presenta variaciones aleatorias. Por lo cual, el modelo que se ajusta a la relación funcional es

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{2.2}$$

donde ε es el error aleatorio.

La variación aleatoria está presente en la variable dependiente y no en la variable independiente, es por ésto que la variable independiente se considera fija. Con este modelo y una vez observada una muestra de tamaño n

$$(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n),$$

se puede expresar el modelo en forma muestral como:

$$y_i = \beta_0 + \beta_1 x_i \, ; \quad i = 1, \ldots, n.$$

Por considerar sólo una variable independiente, a este modelo se le conoce como **Modelo de Regresión Lineal Simple**, donde la linealidad se relaciona con los parámetros y no con las variables, es decir, no se requiere que la relación sea una función lineal de X y Y. Se consideran algunos supuestos en los errores, los cuales son:

$$\begin{aligned} \mathbb{E}[\varepsilon_i] &= 0 \quad \text{para toda } i = 1, \ldots, n \\ Var[\varepsilon_i] &= \sigma^2 \quad \text{para toda } i = 1, \ldots, n \\ \mathbb{E}[\varepsilon_i \varepsilon_j] &= 0 \quad \text{para toda } i \neq j. \end{aligned} \tag{2.3}$$

De esta forma, al condicionar el análisis en los valores observados de X, es decir $(\underline{X} = \underline{x})$, se tiene lo siguiente

$$
\begin{aligned}
\mathbb{E}[Y \mid (\underline{X} = \underline{x})] &= \beta_0 + \beta_1 x \\
Var[Y \mid (\underline{X} = \underline{x})] &= \sigma^2,
\end{aligned}
\tag{2.4}
$$

por notación, se define $\sigma_y^2 = \sigma^2$.

Para desarrollar un análisis estadístico del modelo (2.2) es necesario explorar el comportamiento de los parámetros β_0 y β_1, a los cuales se les conoce como *coeficientes de la regresión*. Una forma de realizarlo es hacer inferencias sobre el valor de los mismos, por ejemplo, a través de la estimación, la cual puede ser en forma puntual o por intervalo. Para la estimación por intervalo se requiere de una distribución asociada a los errores. Este estudio se centrará en la obtención de intervalos de confianza para los coeficientes de la regresión, en particular para β_1.

Para la estimación de los parámetros, se obtiene "la mejor" recta que ajuste los datos, es decir, los estimadores de β_0 y β_1 deben ser tales que la diferencia entre y_i y $\beta_0 + \beta_1 x_i$ sea mínima, para toda $i = 1, \ldots, n$.

De esta manera, se define la **Suma de Cuadrados de los Errores** (SCE) como:

$$
SCE(\beta_0, \beta_1) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 \tag{2.5}
$$

Una forma de obtener los estimadores de β_0 y β_1 es el **Método de Mínimos Cuadrados**, en el cual se busca minimizar la SCE.

Ya que la SCE (ver (2.5)) es una función convexa y diferenciable con respecto a β_0 y β_1, los estimadores se obtienen al igualar a cero las correspondientes derivadas

$$
\begin{aligned}
\frac{\partial}{\partial \beta_0} SCE(\beta_0, \beta_1) &= 0 \\
\frac{\partial}{\partial \beta_1} SCE(\beta_0, \beta_1) &= 0,
\end{aligned}
\tag{2.6}
$$

de estas igualdades se desprenden las conocidas ecuaciones normales,

$$
\begin{aligned}
\sum_{i=1}^{n} Y_i - n\widehat{\beta}_0 - \widehat{\beta}_1 \sum_{i=1}^{n} X_i &= 0 \\
\sum_{i=1}^{n} X_i Y_i - \widehat{\beta}_0 \sum_{i=1}^{n} X_i - \widehat{\beta}_1 \sum_{i=1}^{n} X_i^2 &= 0
\end{aligned} \tag{2.7}
$$

cuya solución es

$$
\begin{aligned}
\widehat{\beta}_0 &= \overline{Y} - \widehat{\beta}_1 \overline{X} \\
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \quad .
\end{aligned} \tag{2.8}
$$

Una vez evaluados $\widehat{\beta}_0$ y $\widehat{\beta}_1$ en la SCE, se obtiene su esperanza, es decir,

$$
\mathbb{E}\Big[SCE(\widehat{\beta}_0, \widehat{\beta}_1)\Big] = (n-2)\sigma_y^2, \tag{2.9}
$$

de donde se obsserva que un estimador insesgado de la varianza es:

$$
\widetilde{\sigma_y^2} = \frac{SCE(\widehat{\beta}_0, \widehat{\beta}_1)}{n-2}.
$$

Por otra parte, si el interés radica en la construcción de estimadores por intervalo o en poder contrastar hipótesis sobre los parámetros, entonces se requiere necesariamente de una distribución para los errores. En la literatura, el caso más estudiado y que obedece al gran número de aplicaciones es en el que los errores tienen distribución normal.

Si existe un supuesto distribucional en los errores, otro método para realizar las estimaciones de los parámetros es el **Método de Máxima Verosimilitud**. En este caso, debido a que las observaciones son condicionalmente independientes dado $\underline{X} = \underline{x}$, la función de verosimilitud tiene la forma general

$$L(\beta_0, \beta_1, \sigma_y^2; \underline{y}) \quad = \quad \prod_{i=1}^{n} f(y_i; \beta_0, \beta_1, \sigma_y^2). \tag{2.10}$$

Bajo el modelo (2.2), la variable aleatoria Y_i, para X_i habiendo tomado el valor fijo y conocido x_i, tiene distribución normal con media $\beta_0 + \beta_1 x_i$ y varianza σ_y^2, es decir,

$$Y_i | (X_i = x_i) \quad \sim \quad N(\beta_0 + \beta_1 x_i, \sigma_y^2); \quad \forall\, i = 1, \ldots, n \tag{2.11}$$

De esta forma (2.10) es

$$L(\beta_0, \beta_1, \sigma_y^2; \underline{y}) \quad = \quad (2\pi\sigma_y^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_y^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2\right\}, \tag{2.12}$$

la maximización de L produce exactamente los mismos estimadores $\widehat{\beta}_0$ y $\widehat{\beta}_1$ que el **Método de Mínimos Cuadrados**. Adicionalmente, el estimador máximo verosímil de σ_y^2 es

$$\widehat{\sigma_y^2} = \frac{SCE(\widehat{\beta}_0\, ,\, \widehat{\beta}_1)}{n}\,.$$

En el caso de mínimos cuadrados se sabe que $\widehat{\beta}_0$ y $\widehat{\beta}_1$ cumplen con la propiedad de ser de varianza mínima dentro de la clase de los estimadores lineales e insesgados.

Para el caso de máxima verosimilitud, las distribuciones asociadas a los estimadores $\widehat{\beta}_0$, $\widehat{\beta}_1$ y $\widehat{\sigma_y^2}$ son:

$$\begin{aligned}
\widehat{\beta}_0 \quad &\sim \quad N\left(\beta_0\, ,\, \sigma_y^2 \left(\frac{\overline{X}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} + \frac{1}{n}\right)\right) \\
\widehat{\beta}_1 \quad &\sim \quad N\left(\beta_1\, ,\, \frac{\sigma_y^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right) \\
\frac{(n-2)\widehat{\sigma_y^2}}{\sigma_y^2} \quad &\sim \quad \chi^2_{(n-2)}
\end{aligned} \tag{2.13}$$

donde $\widehat{\beta}_0$ y $\widehat{\beta}_1$ son independientes de $\widehat{\sigma_y^2}$ y $\chi^2_{(n-2)}$ representa la distribución *Chi-cuadrada* con $(n-2)$ grados de libertad.

Además de la estimación puntual de β_0, β_1 y σ_y^2, se puede obtener una estimación por medio de intervalos de confianza. Para ésto, se requiere utilizar cantidades pivotales, donde una **cantidad pivotal** se define como:

> *Una función de la muestra y el parámetro de interés cuya distribución no depende del parámetro de interés.* [1]

De esta forma

$$\frac{\beta_0-\widehat{\beta}_0}{\sqrt{\widehat{\sigma_y^2}\left(\frac{\overline{X}^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}+\frac{1}{n}\right)}} \sim t_{(n-2)}$$

$$\frac{\beta_1-\widehat{\beta}_1}{\sqrt{\frac{\widehat{\sigma_y^2}}{\sum_{i=1}^{n}(X_i-\overline{X})^2}}} \sim t_{(n-2)} \qquad \textbf{(2.14)}$$

$$\frac{(n-2)\widehat{\sigma_y^2}}{\sigma_y^2} \sim \chi^2_{(n-2)}$$

son cada una, una cantidad pivotal (donde $t_{(n-2)}$ representa la distribución t de Student con $(n-2)$ grados de libertad).

Con estas cantidades se pueden obtener intervalos de confianza de longitud mínima para cada uno de los parámetros al "pivotear". Donde un **intervalo de confianza de longitud mínima** se define como:

> *Dado el parámetro de interés* (θ), *una cantidad pivotal* $(q(x_1, x_2, \ldots, x_n; \theta)$ *y un nivel de confianza* α *fijo* $(0 \leq \alpha \leq 1)$, *existen* Q_1 *y* Q_2 *que dependen de* α *tales que*,

[1] Véase Mood pag. 379

$$P[\,Q_1 \leq q(x_1, x_2, \ldots, x_n\,;\,\theta) \leq Q_2\,] = 1 - \alpha.$$

Es decir, para cada muestra $(x_1, x_2, \ldots, x_n)$

$$Q_1 \leq q(x_1, x_2, \ldots, x_n\,;\,\theta) \leq Q_2$$

sí y sólo si

$$t_1\,(x_1, x_2, \ldots, x_n) \leq \tau(\theta) \leq t_2\,(x_1, x_2, \ldots, x_n)$$

para ciertas funciones t_1 *y* t_2 *que no dependen de* θ. *De esta forma* $(T_1\,,\,T_2)$ *es el intervalo al* $(1-\alpha)\times 100\%$ *de confianza para* $\tau(\theta)$, *con* $T_i \;=\; t_i(x_1, x_2, \ldots, x_n) : i \;=\; \{1, 2\}$.

El intervalo de longitud mínima es aquel en el que Q_1 *y* Q_2 *se eligen de tal forma que* $t_2(x_1, x_2, \ldots, x_n) - t_1(x_1, x_2, \ldots, x_n)$ *sea de longitud mínima.* [2]

De esta forma, los intervalos de confianza de longitud mínima para β_0 , β_0 y σ_y^2 a un nivel de confianza α son, respectivamente

$$\widehat{\beta}_0 - t_{(n-2)}^{1-\frac{\alpha}{2}}\left(\frac{\widehat{\sigma_y^2}\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i-\overline{x})^2}\right)^{\frac{1}{2}} \quad\leq\quad \beta_0 \quad\leq\quad \widehat{\beta}_1 + t_{(n-2)}^{1-\frac{\alpha}{2}}\left(\frac{\widehat{\sigma_y^2}\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i-\overline{x})^2}\right)^{\frac{1}{2}}$$

$$\widehat{\beta}_1 - t_{(n-2)}^{1-\frac{\alpha}{2}}\left(\frac{\widehat{\sigma_y^2}}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right)^{\frac{1}{2}} \quad\leq\quad \beta_1 \quad\leq\quad \widehat{\beta}_1 + t_{(n-2)}^{1-\frac{\alpha}{2}}\left(\frac{\widehat{\sigma_y^2}}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right)^{\frac{1}{2}}$$

$$\frac{(n-2)\widehat{\sigma_y^2}}{\chi^{2\,(1-\frac{\alpha}{2})}_{(n-2)}} \quad\leq\quad \sigma_y^2 \quad\leq\quad \frac{(n-2)\widehat{\sigma_y^2}}{\chi^{2\,\frac{\alpha}{2}}_{(n-2)}} \qquad (2.15)$$

Este análisis para el caso de la regresión lineal simple con una variable independiente fija es elemental y aparece en varios libros de regresión, incluyendo Weisberg (1980), Draper and Smith (1981) y Montgomery and Peck (1992).

---

[2] Véase Mood pag. 379

2.2 Planteamiento del Problema

En algunas aplicaciones de la técnica de regresión lineal, la variable independiente que se pretenden relacionar con la variable dependiente mediante un modelo simple, es aleatoria. Sin embargo, en la práctica para ajustar este tipo de modelos se considera fija y así se lleva a cabo el análisis, argumentando que es condicional.

Resulta de interés preguntarse, ¿qué ocurriría si se considera la variable independiente aleatoria al hacer el correspondiente análisis del modelo de regresión lineal?. En particular ¿qué ocurriría si la distribución conjunta de X y Y es normal?

El desarrollo de este trabajo será bajo el supuesto de que la variable independiente, en el modelo de regresión lineal simple, es aleatoria. En particular, se estudiará el intervalo de confianza para β_1 suponiendo que el análisis es incondicional.

El análisis utilizado en la gran mayoría de las aplicaciones para obtener un estimador por intervalo (o un intervalo de confianza) es condicional. En el presente trabajo se hace una comparación entre el intervalo de confianza incondicional para β_1 con el intervalo condicional.

Para este análisis se considerarán los siguientes casos:

- la variabilidad de la variable independiente es mayor que la de la variable dependiente.
- la variabilidad de la variable independiente es menor que la de la variable dependiente.
- la variabilidad de la variable independiente y de la variable dependiente son similares.

Capítulo 3

Obtención de intervalos de confianza incondicionales

Partiendo del modelo de regresión lineal simple en donde la distribución de los parámetros es condicional, se obtendrá la distribución incondicional del parámetro asociado a la pendiente, por medio del Teorema de Bayes. Es decir, el modelo de regresión lineal simple con $(X\,,\,Y)$ conjuntamente normales es:

$$\begin{aligned} \varepsilon &\sim N(\,0\,,\,\sigma_y^2\,) \\ Y\,|\,X &\sim N(\,\beta_0+\beta_1 X\,,\,\sigma_y^2\,) \\ X &\sim N(\,\mu_x\,,\,\sigma_x^2\,). \end{aligned} \tag{3.1}$$

Dada una muestra, $\underline{X}=\underline{x}$, la distribución de $\widehat{\beta}_1 \,\big|\, (\underline{X}=\underline{x})$ es

$$\widehat{\beta}_1 \,\Big|\, (\underline{X}=\underline{x}) \;\sim\; N\left(\beta_1\,,\,\frac{\sigma_y^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}\right). \tag{3.2}$$

Obsérvese que esta distribución depende de $\underline{X}$ sólo a través de $V \;=\; \sum_{i=1}^{n}(X_i-\,\overline{X}\,)^2$ de esta forma

$$\widehat{\beta}_1 \,\Big|\, \underline{X} \overset{D}{=} \widehat{\beta}_1 \,\Big|\, V,$$

donde el símbolo $\stackrel{D}{\equiv}$ significa que son equivalentes en distribución.

Sea $g_{\widehat{\beta}_1|V}$ la función de densidad de $\widehat{\beta}_1 \Big| V$ y h_V la función de densidad de V. Al marginalizar la distribución conjunta de $(\widehat{\beta}_1, V)$ se obtendrá la distribución incondicional de $\widehat{\beta}_1$ (por notación $g_{\widehat{\beta}_1}$), es decir,

$$g_{\widehat{\beta}_1}(\widehat{\beta}_1) \quad = \quad \int g_{\widehat{\beta}_1|V}(\widehat{\beta}_1 \Big| v) \cdot h_V(v) dv. \tag{3.3}$$

para hacer inferencias incondicionales sobre β_1.

Dada la distribución de X (ver (3.1)), se tiene que

$$\frac{(n-1)S_x^2}{\sigma_x^2} \quad \sim \quad \chi^2_{(n-1)} \tag{3.4}$$

con

$$S_x^2 \quad = \quad \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}. \tag{3.5}$$

De esta forma

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad \sim \quad \sigma_x^2 \cdot \chi^2_{(n-1)} \tag{3.6}$$

y [1]

$$\begin{aligned} \sigma_x^2 \cdot \chi^2_{(n-1)} \quad &\equiv \quad \sigma_x^2 \cdot \mathrm{Gama}(\tfrac{n-1}{2}, \tfrac{1}{2}) \\ &= \quad \mathrm{Gama}(\tfrac{n-1}{2}, \tfrac{1}{2\sigma_x^2}). \end{aligned} \tag{3.7}$$

[1] $\chi_k^2 \equiv \mathrm{Gama}(\frac{k}{2}, \frac{1}{2})$.

Sea z una v.a. con función de densidad χ_k^2 ent. $cz \sim \mathrm{Gama}(\frac{k}{2}, \frac{1}{2c})$. donde

$$\mathrm{Gama}(a, b) = \frac{b^a}{\Gamma(a)} z^{a-1} \exp\{-bz\} \ \ ; \ \ z \in (0, \infty) \ , \ a, b > 0 \text{ y } c > 0$$

por lo que $h_V(v)$ se puede expresar como:

$$h_V(v) \;=\; \frac{v^{\frac{n-1}{2}-1}\exp\left\{-\frac{v}{2\sigma_x^2}\right\}}{(2\sigma_x^2)^{\frac{n-1}{2}}\,\Gamma(\frac{n-1}{2})}. \tag{3.8}$$

Aunado a esto, dado que $\widehat{\beta}_1\big| V \sim N\left(\beta_1, \frac{\sigma_y^2}{V}\right)$, es decir,

$$g_{\widehat{\beta}_1|V}(\widehat{\beta}_1\big| v) \;=\; \sqrt{\frac{v}{2\pi\sigma_y^2}}\exp\left\{-\frac{v}{2\sigma_y^2}(\widehat{\beta}_1-\beta_1)^2\right\}\;, \tag{3.9}$$

al realizar el producto entre (3.8) y (3.9) e integrar con respecto a v, se obtiene la densidad incondicional de $\widehat{\beta}_1$

$$\begin{aligned} g_{\widehat{\beta}_1}(\widehat{\beta}_1) \;&=\; \frac{1}{\sqrt{2\pi\sigma_y^2}(2\sigma_x^2)^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} \\ &\quad\times\int_0^\infty v^{\frac{n}{2}-1}\exp\left\{-\frac{1}{2}\left\{\frac{1}{\sigma_x^2}+\frac{1}{\sigma_y^2}(\widehat{\beta}_1-\beta_1)^2\right\}v\right\}dv\,, \end{aligned} \tag{3.10}$$

donde el integrando es el núcleo de una distribución Gama(a,b) con

$$\begin{aligned} a \;&=\; \frac{n}{2} \\ b \;&=\; \frac{1}{2}\left\{\frac{1}{\sigma_x^2}+\frac{1}{\sigma_y^2}(\widehat{\beta}_1-\beta_1)^2\right\}. \end{aligned} \tag{3.11}$$

Al completar esta distribución Gama, la densidad de $\widehat{\beta}_1$ resulta ser

$$g_{\widehat{\beta}_1}(\widehat{\beta}_1) \;=\; \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi}\Gamma(\frac{n-1}{2})}\left(\frac{\sigma_x^2}{\sigma_y^2}\right)^{\frac{1}{2}}\left\{1+\frac{\sigma_x^2}{\sigma_y^2}(\widehat{\beta}_1-\beta_1)^2\right\}^{-\frac{n}{2}}. \tag{3.12}$$

Con el objeto de estudiar esta densidad, se define

$$\rho \;=\; \frac{\sigma_y^2}{\sigma_x^2}\,. \tag{3.13}$$

ya que la función (3.12) depende de las varianzas sólo a través de ρ, se puede reescribir como:

$$g_{\widehat{\beta}_1}(\widehat{\beta}_1\,;\beta_1\,,\rho\,) \;=\; \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi}\Gamma(\frac{n-1}{2})}\rho^{-\frac{1}{2}}\left\{1+\tfrac{1}{\rho}(\widehat{\beta}_1-\beta_1)^2\right\}^{-\frac{n}{2}}\,. \tag{3.14}$$

Para analizar la función de densidad de $\widehat{\beta}_1$ se consideraron dos casos:

- ρ conocido
- ρ desconocido.

3.1 Primer caso, ρ conocido .

El primer caso corresponde a la situación práctica en la que se podría suponer un valor para el cociente de varianzas, es decir cuando se puede saber qué proporción guarda la variabilidad de Y con respecto a la de X.

Sea $\rho = \rho_o$ conocido, de manera que la distribución de $\widehat{\beta}_1$ sólo depende de β_1, es decir

$$g_{\widehat{\beta}_1}(\widehat{\beta}_1\,;\,\beta_1) \;=\; \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi\rho_o}\,\Gamma(\frac{n-1}{2})}\left\{1+\tfrac{1}{\rho_o}(\widehat{\beta}_1-\beta_1)^2\right\}^{-\frac{n}{2}}, \tag{3.15}$$

en esta función el núcleo corresponde a una densidad t de Student con $(n-1)$ grados de libertad, parámetro de localización β_1 y parámetro de escala $\frac{1}{(n-1)}\rho_o$. [2] Por lo cual

$$\widehat{\beta}_1 \;\sim\; t_{(n-1)} \tag{3.16}$$

con

[2]Sea z una v.a. con función de densidad t de Student con k grados de libertad, parámetro de localización μ y parámetro de escala σ^2 entonces

$$f_z(z) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\left(k\pi\sigma^2\right)^{\frac{1}{2}}}\left\{1+\tfrac{1}{k\sigma^2}\left(z-\mu\right)^2\right\}^{-\frac{k+1}{2}}$$

$$
\begin{aligned}
\mathbb{E}\left[\widehat{\beta}_1\right] &= \beta_1 \\
Var\left[\widehat{\beta}_1\right] &= \tfrac{\rho_o}{(n-3)}\,, \quad \text{si } n > 3\ .
\end{aligned}
\tag{3.17}
$$

De esta distribución se obtiene el intervalo de confianza, de longitud mínima, para β_1 por el camino usual.

Sea

$$\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\frac{\rho_o}{n-1}}}$$

una cantidad pivotal, con distribución t de Student con parámetro de escala 0, parámetro de localización 1 y $(n-1)$ grados de libertad. Así, se puede definir el intervalo al $(1-\alpha) \times 100\%$ de confianza de longitud mínima para β_1 como

$$Q_1 \;\leq\; \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\frac{\rho_o}{n-1}}} \;\leq\; Q_2, \tag{3.18}$$

donde $0 \leq \alpha \leq 1$ y Q_1, Q_2 $(Q_1 = -Q_2)$ son percentiles de una $t_{(n-1)}$. Es decir, si $t_{(n-1)}^{1-\frac{\alpha}{2}}$ es el percentil de una distribución t de Student, entonces el intervalo de longitud mínima para β_1 es:

$$-t_{(n-1)}^{1-\frac{\alpha}{2}} \;\leq\; \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\frac{\rho_o}{n-1}}} \;\leq\; t_{(n-1)}^{1-\frac{\alpha}{2}}\ . \tag{3.19}$$

Al "pivotear" se obtiene el siguiente intervalo de confianza

$$\widehat{\beta}_1 - t_{(n-1)}^{1-\frac{\alpha}{2}}\sqrt{\frac{\rho_o}{n-1}} \;\leq\; \beta_1 \;\leq\; \widehat{\beta}_1 + t_{(n-1)}^{1-\frac{\alpha}{2}}\sqrt{\frac{\rho_o}{n-1}}\ . \tag{3.20}$$

Para poder hacer una comparación de este intervalo con el correspondiente intervalo condicional, es necesario aproximar σ_x^2 por medio de $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ esta aproximación es adecuada sólo si n es grande. Por la relación que guarda σ_x^2 con σ_y^2, se puede aproximar σ_y^2 por medio de $\frac{\rho_o}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, por lo cual, se considerará σ_y^2 conocida.

Al sustituir σ_y^2 por su aproximación en (3.2), se conserva la media pero la varianza cambia, siendo ahora $\frac{\rho_o}{n-1}$.

Por considerar fijos y conocidos a σ_y^2 y $\sum_{i=1}^{n}(X_i - \overline{X})^2$ se tiene que

$$\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\frac{\rho_o}{n-1}}} \sim N(0, 1), \tag{3.21}$$

de esta manera, el intervalo de confianza condicional para β_1 es:

$$\widehat{\beta}_1 - z^{1-\frac{\alpha}{2}}\sqrt{\frac{\rho_o}{n-1}} \leq \beta_1 \leq \widehat{\beta}_1 + z^{1-\frac{\alpha}{2}}\sqrt{\frac{\rho_o}{n-1}} \tag{3.22}$$

donde $z^{1-\frac{\alpha}{2}}$ es el percentil de una distribución normal estándar, con $0 \leq \alpha \leq 1$.

La motivación de tener estos dos intervalos de confianza (3.20 y 3.22) es poder compararlos. Dado que ambos están centrados en $\widehat{\beta}_1$, es suficiente con comparar sus longitudes.

La longitud del intervalo incondicional es

$$2t_{(n-1)}^{1-\frac{\alpha}{2}}\left(\frac{\rho_o}{n-1}\right)^{\frac{1}{2}}$$

y la longitud del intervalo condicional es

$$2z^{1-\frac{\alpha}{2}}\left(\frac{\rho_o}{n-1}\right)^{\frac{1}{2}}.$$

Obsérvese que $z^{1-\frac{\alpha}{2}} \leq t_{(n-1)}^{1-\frac{\alpha}{2}}$, $\forall\, n \in I\!N$ y el límite cuando $n \to \infty$ de $t_{(n-1)}^{1-\frac{\alpha}{2}}$ es $z^{1-\frac{\alpha}{2}}$.

Como se mencionó en párrafos anteriores, para aproximar σ_x^2 se requiere que n sea grande, por lo cual, la longitud de ambos intervalos es esencialmente la misma. Para tamaños de muestra pequeños, esta comparación no es adecuada, ya que se fundamenta en la aproximación de σ_x^2, que en este caso no es una buena aproximación, por la poca información que se obtiene al tener pocas observaciones.

3.2 Segundo caso, ρ desconocido

Como $g_{\widehat{\beta}_1}$ depende de un parámetro desconocido (ρ), su distribución se considerará *condicionada* a este valor. Para obtener la distribución incondicional se requiere un procedimiento análogo al que se desarrolló cuando $\widehat{\beta}_1$ dependía de V, con la diferencia de que V es una variable aleatoria y ρ es un parámetro fijo. Este tipo de consideración es similar al utilizado en Sprott and Farewell (1993) en el problema de *Behrens-Fisher*, en el cual se obtiene una distribución que depende solamente de un parámetro y a éste se le *induce* una distribución. Este es un enfoque *fiducial*, en el que el papel de la cantidad observable (estadística) y el parámetro son, en cierto modo, intercambiables ya que los une una cantidad pivotal.

Al considerar que la distribución de $\widehat{\beta}_1$ (ver (3.14)) depende de ρ, utilizando como notación $g_{\widehat{\beta}_1|\rho}(\widehat{\beta}_1 ; \beta_1)$ en lugar de $g_{\widehat{\beta}_1}(\widehat{\beta}_1 ; \beta_1 , \rho)$, se *promediará* $g_{\widehat{\beta}_1|\rho}$ con respecto a la densidad inducida sobre ρ (denotada como f_ρ), es decir,

$$g_{\widehat{\beta}_1}(\widehat{\beta}_1) = \int_0^\infty g_{\widehat{\beta}_1|\rho}(\widehat{\beta}_1;\beta_1,\rho)\, f_\rho(\rho)\, d\rho \, .$$

Dado que

$$\begin{aligned} \frac{(n-1)\widehat{\sigma^2_x}}{\sigma^2_x} &\sim \chi^2_{(n-1)} \\ \frac{(n-2)\widehat{\sigma^2_y}}{\sigma^2_y} &\sim \chi^2_{(n-2)} \end{aligned} \tag{3.23}$$

con

$$\begin{aligned} \widehat{\sigma^2_x} &= \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \overline{x})^2 \\ \widehat{\sigma^2_y} &= \frac{1}{(n-2)} \sum_{i=1}^n \left(y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right)\right)^2 . \end{aligned} \tag{3.24}$$

en el análisis condicional la distribución de $\frac{(n-2)\widehat{\sigma^2_y}}{\sigma^2_y}$ no depende de $\underline{X}$. Entonces $\widehat{\sigma^2_y}$ es independiente de $\underline{X}$, por lo tanto es independiente de

$\sum_{i=1}^{n}(X_i - \overline{X})^2$, es decir,

$$\widehat{\sigma_y^2} \perp \sum_{i=1}^{n}(X_i - \overline{X})^2 \Rightarrow \frac{(n-2)\widehat{\sigma_y^2}}{\sigma_y^2} \perp \frac{(n-1)\widehat{\sigma_x^2}}{\sigma_x^2}, \qquad (3.25)$$

donde el símbolo $\perp$ significa que son independientes.

Sea $s_x = \frac{(n-1)\widehat{\sigma_x^2}}{\sigma_x^2}$ y $s_y = \frac{(n-2)\widehat{\sigma_y^2}}{\sigma_y^2}$, por el enfoque fiducial se tiene que

$$\sigma_x^2 \sim \frac{(n-1)\widehat{\sigma_x^2}}{\chi^2_{(n-1)}} \qquad \sigma_y^2 \sim \frac{(n-2)\widehat{\sigma_y^2}}{\chi^2_{(n-2)}}$$

$$(3.26)$$

$$\sigma_x^2 = \frac{(n-1)\widehat{\sigma_x^2}}{s_x} \qquad \sigma_y^2 = \frac{(n-2)\widehat{\sigma_y^2}}{s_y}.$$

Al sustituir σ_x^2 y σ_y^2 en (3.13), se obtiene

$$\rho = \frac{\widehat{\sigma_y^2}\left(\frac{s_x}{(n-1)}\right)}{\widehat{\sigma_x^2}\left(\frac{s_y}{(n-2)}\right)}.$$

Si se define $\widehat{\rho} = \frac{\widehat{\sigma_y^2}}{\widehat{\sigma_x^2}}$ entonces

$$\rho = \widehat{\rho}\,\frac{\left(\frac{s_x}{(n-1)}\right)}{\left(\frac{s_y}{(n-2)}\right)}. \qquad (3.27)$$

Por la relación dada en (3.23), se tiene que $s_x \sim \chi^2_{(n-1)}$ y $s_y \sim \chi^2_{(n-2)}$, además, $s_x \perp s_y$ (ver (3.25)), por lo cual [3]

[3]Sean U y V variables aleatorias independientes con distribución *Chi-cuadrada* con u y v grados de libertad respectivamente , entonces $W = \frac{U/u}{V/v}$ se distribuye como una F con (u, v) grados de libertad.

$$\frac{\left(\frac{s_x}{(n-1)}\right)}{\left(\frac{s_y}{(n-2)}\right)} \sim F_{(n-1,n-2)}\,. \tag{3.28}$$

Sea $q = \frac{\left(\frac{s_x}{(n-1)}\right)}{\left(\frac{s_y}{(n-2)}\right)}$ entonces $\rho = \widehat{\rho} q \Rightarrow q = \frac{\rho}{\widehat{\rho}}$, es decir,

$$\frac{\rho}{\widehat{\rho}} \sim F_{(n-1,n-2)},$$

de esta manera la cantidad $\frac{\rho}{\widehat{\rho}}$ es una cantidad pivotal.

Por notación, sea $Q_q(q)$ la función de densidad de q,[4]

$$Q_q(q) = \frac{\Gamma(\frac{2n-3}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{n-2}{2})}\left(\frac{n-1}{n-2}\right)^{\frac{n-1}{2}} q^{\frac{n-3}{2}}\left(1+\frac{n-1}{n-2}q\right)^{-\frac{2n-3}{2}}. \tag{3.29}$$

Dado que se obtuvo la densidad de $\frac{\rho}{\widehat{\rho}}$, es conveniente hacer una transformación para poder *promediar* $g_{\widehat{\beta}_1|\rho}$. Al hacer el cambio de variable $q = \frac{\rho}{\widehat{\rho}}$ se tiene que

$$g_{\widehat{\beta}_1}(\widehat{\beta}_1) = \int_0^\infty g_{\widehat{\beta}_1|q}(\widehat{\beta}_1;\beta_1,q)\,Q_q(q)\,dq$$

con

$$g_{\widehat{\beta}_1|q}(\widehat{\beta}_1;\beta_1,q) = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})(\pi q\widehat{\rho})^{\frac{1}{2}}}\left\{1+\frac{1}{q\widehat{\rho}}(\widehat{\beta}_1-\beta_1)^2\right\}^{-\frac{n}{2}} \tag{3.30}$$

[4] Si $z \sim F_{(u,v)}$ entonces

$$f_z(z) = \frac{\Gamma(\frac{u+v}{2})}{\Gamma(\frac{u}{2})\Gamma(\frac{v}{2})}\left(\frac{u}{v}\right)^{\frac{u}{2}} z^{\frac{u-2}{2}}\left(1+\frac{u}{v}z\right)^{-\frac{u+v}{2}} \text{ con } z \in (0,\infty)$$

De esta forma $g_{\widehat{\beta}_1}(\widehat{\beta}_1)$ queda expresada como

$$g_{\widehat{\beta}_1}(\widehat{\beta}_1) = \frac{\Gamma(\frac{n}{2})\Gamma(\frac{2n-3}{2})}{\Gamma(\frac{n-1}{2})^2\Gamma(\frac{n-2}{2})(\pi\widehat{\rho})^{\frac{1}{2}}}\left(\frac{n-1}{n-2}\right)^{\frac{n-1}{2}}$$

$$\times \int_0^\infty \left\{\left(1+\frac{1}{q\rho}(\widehat{\beta}_1-\beta_1)^2\right)^{-\frac{n}{2}} q^{\frac{n-3}{2}-\frac{1}{2}}\left(1+\frac{n-1}{n-2}q\right)^{-\frac{2n-3}{2}}\right\}dq\ . \tag{3.31}$$

Obsérvese que el integrando es el núcleo de una mezcla infinita de distribuciones t de Student, con el mismo parámetro de localización (β_1) y factor de escala (q) que varía de 0 a ∞, ponderadas por la densidad del factor de escala. Dado que esta mezcla es simétrica con respecto a $\widehat{\beta}_1$, se puede pensar en ésta como la distribución usual para $\widehat{\beta}_1$ con parámetro β_1, así, como en la *inducida* para β_1, en donde ahora $\widehat{\beta}_1$ es el "parámetro" (en adelante se utilizará esta nomenclatura), es decir,

$$g_{\beta_1}(\beta_1) = C \times \int_0^\infty \left\{\frac{1}{(\pi q\widehat{\rho})^{\frac{1}{2}}} q^{\frac{n-3}{2}}\left(1+\frac{1}{q\rho}(\widehat{\beta}_1-\beta_1)^2\right)^{-\frac{n}{2}}\left(1+\frac{n-1}{n-2}q\right)^{-\frac{2n-3}{2}}\right\}dq \tag{3.32}$$

donde

$$C = \frac{\Gamma(\frac{n}{2})\Gamma(\frac{2n-3}{2})}{\Gamma(\frac{n-1}{2})^2\Gamma(\frac{n-2}{2})}\left(\frac{n-1}{n-2}\right)^{\frac{n-1}{2}}\ .$$

La integral (3.32) no se pudo resolver en forma analítica, por lo que la distribución de β_1 no se obtuvo en forma explícita. Para dar una expresión del intervalo de confianza de β_1 alrededor de $\widehat{\beta}_1$, se recurrió a resolver la integral en forma numérica.

Dicho de otra forma, se desea obtener $\Delta \in I\!R^+$ (fijo), tal que

$$\int_{\widehat{\beta}_1-\Delta}^{\widehat{\beta}_1+\Delta} g_{\beta_1}(\beta_1)\, d\beta_1 = 1-\alpha, \tag{3.33}$$

con $0 \le \alpha \le 1$, es decir,

$$1 - \alpha = C \int_{\widehat{\beta}_1 - \Delta}^{\widehat{\beta}_1 + \Delta} \int_0^{\infty} \left\{ \frac{1}{(\pi q \widehat{\rho})^{\frac{1}{2}}} q^{\frac{n-3}{2}} \left(1 + \frac{1}{q\widehat{\rho}}(\beta_1 - \widehat{\beta}_1)^2\right)^{-\frac{n}{2}} \left(1 + \frac{n-1}{n-2} q\right)^{-\frac{2n-3}{2}} \right\} dq\, d \tag{3.34}$$

El resultado obtenido en (3.34) podrá compararse con el correspondiente intervalo condicional para β_1 obtenido en el capítulo anterior. Esta comparación indicará qué tan distintas pueden ser las estimaciones por intervalo al utilizar un análisis condicional o uno incondicional.

Capítulo 4

Programa

La función de densidad *inducida* para β_1, obtenida en el capítulo anterior, es

$$
\begin{aligned}
g_{\beta_1}(\beta_1) \;=\;& \frac{\Gamma(\frac{n}{2})\Gamma(\frac{2n-3}{2})}{\Gamma(\frac{n-1}{2})^2\Gamma(\frac{n-2}{2})}\left(\frac{n-1}{n-2}\right)^{\frac{n-1}{2}} \\
&\times \int_0^\infty \left\{ \frac{1}{(\pi q\widehat{\rho})^{\frac{1}{2}}} q^{\frac{n-3}{2}} \left(1+\frac{1}{q\rho}(\widehat{\beta}_1-\beta_1)^2\right)^{-\frac{n}{2}} \left(1+\frac{n-1}{n-2}q\right)^{-\frac{2n-3}{2}} \right\} dq.
\end{aligned}
\tag{4.1}
$$

Dicha densidad depende de las siguientes cantidades o parámetros relevantes para la simulación

$$
\begin{array}{rl}
n & \text{tamaño de la muestra} \\
\beta_1 & \text{valor del parámetro de interés} \\
 & \quad \text{en la simulación} \\
\rho = q\widehat{\rho} & \text{parámetro de ruido.}
\end{array}
\tag{4.2}
$$

Para determinar el intervalo al $(1-\alpha)\times 100\%$ de confianza para β_1, obsérvese que α se incorpora como un parámetro más en el problema de búsqueda numérica.

Además, habrá que recordar que tanto la variable independiente como la variable dependiente involucran a otros parámetros, ya que $X \sim N(\mu_x, \sigma_x^2)$ y $Y \mid (X = x) \sim N(\beta_0 + \beta_1 x, \sigma_y^2)$, por lo que se anexan a la lista de cantidades a controlar ya sea como parámetros fijos o de interés.

Los parámetros que se mantuvieron fijos son los relacionados con la localización, es decir $\mu_x = 0$ y $\beta_0 = 9.0 \times 10^{-25}$ (este valor se tomó relativamente igual a cero, pero para poder utilizar las expresiones de los estimadores tal como se definieron anteriormente era necesario que fuera diferente de cero). El valor de ρ se restringió a que tomara los valores $\{0.1, 1, 2\}$ que tienen un sentido razonable en esta exploración, ya que representan la relación que existe entre la variabilidad de la variable dependiente y la independiente. La desviación estándar de X (σ_x) se generó en forma pseudoaleatoria y el valor de σ_y se obtuvo por medio de la relación $\sigma_y = \sqrt{\rho}\,\sigma_x$. El parámetro que falta por analizar es β_1 (parámetro de interés), para este análisis se decidió asignarle valores de $\{0.1, 1, 10\}$.

Debe quedar claro que el fijar las cantidades en pocas combinaciones, fue con la intención de poder tener un mejor control del análisis.

Para poder realizar las simulaciones se hizo un programa (ver apéndice A) con el que se generaron muestras de normales de tamaño $n = \{10, 50, 100, 500\}$, para cada n y parámetros σ_x, β_1 y ρ, se generó una muestra con media μ_x y varianza σ_x^2 y una secuencia de observaciones generadas por medio de una normal con media $\beta_0 + \beta_1 x_i$ y varianza $\rho\sigma_x^2$ para cada $i = 1, \ldots, n$. Por último, los dos valores de α utilizados fueron 0.05 y 0.10. De esta manera, los casos a analizar fueron:

$$(n, \alpha, \rho, \beta_1) \in \{10, 50, 100, 500\} \times \{0.5, 0.10\} \times \{0.1, 1.0, 2.0\} \times \{0.1, 1.0, 10\}.$$

Con cada una de estas cuartetas se generaron 1000 muestras de normales bivariadas (x, y) de tamaño n.

Para cada una de las muestras generadas de la misma normal bivariada se obtuvieron los estimadores $\widehat{\sigma_x^2}$, $\widehat{\sigma_y^2}$, $\widehat{\beta}_0$, $\widehat{\beta}_1$ y $\widehat{\rho}$, es decir,

$$
\begin{aligned}
\widehat{\sigma_x^2} &= \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1} \\
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2} \\
\widehat{\beta}_0 &= \overline{y} - \widehat{\beta}_1\overline{x} \\
\widehat{\sigma_y^2} &= \frac{\sum_{i=1}^{n}(y_i-\widehat{\beta}_0-\widehat{\beta}_1 x_i)^2}{n-2} \\
\widehat{\rho} &= \frac{\widehat{\sigma_y^2}}{\widehat{\sigma_x^2}}.
\end{aligned}
\tag{4.3}
$$

Con los estimadores $\widehat{\beta}_1$, $\widehat{\sigma_y^2}$ y valores en tablas de una t de Student con $(n-2)$ grados de libertad y un nivel de confianza α, se calculó el intervalo de confianza *condicional* de β_1, de éste se obtuvo su longitud, es decir,

$$
2\, t_{(n-2)}^{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{\sigma_y^2}}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}\,. \tag{4.4}
$$

El intervalo incondicional se calculó de la siguiente forma, dada la función de densidad incondicional de β_1 (véase (4.1)), el estimador $\widehat{\beta}_1$ y un valor inicial de Δ (Δ_o), se inició una búsqueda recursiva para encontrar el intervalo de longitud mínima que cubriera el $(1-\alpha)\times 100\%$ del área, por medio del siguiente algoritmo.

ALGORITMO DE BÚSQUEDA

1.— Asignar $\Delta = \Delta_o$

2.— Evaluar

$$\int_{\widehat{\beta}_1-\Delta}^{\widehat{\beta}_1+\Delta} g_{\beta_1}(\beta_1)\, d\beta_1$$

3.— Comparar con $1-\alpha$

si es mayor

Asignar $\Delta = \Delta - \left|\frac{\Delta}{2}\right|$
Regresar a 2.—

si es menor

Asignar $\Delta = \Delta + \left|\frac{\Delta}{2}\right|$
Regresar a 2.—

si es igual

Ir a 4.—

4.— FIN

(4.5)

La aproximación al valor $(1-\alpha)$ se realizó con 3 dígitos significativos, [1] es decir,

$$\frac{\left|(1-\alpha) - \int_{\widehat{\beta}_1-\Delta}^{\widehat{\beta}_1+\Delta} g_{\beta_1}(\beta_1)\, d\beta_1\right|}{|(1-\alpha)|} < 5 \times 10^{-3} \tag{4.6}$$

por lo que, para llegar al punto (4), la integral evaluada en Δ deberá estar entre

$$(1-\alpha) \times [1 - (5 \times 10^{-3})] \qquad \text{y} \qquad (1-\alpha) \times [1 + (5 \times 10^{-3})],$$

es decir, una aproximación (Δ^*) al verdadero valor de Δ, debe ser tal que

[1] Véase Burden (1985), p.24

$$(1-\alpha)\times(0.995) \;\leq\; \int_{\widehat{\beta}_1-\Delta^*}^{\widehat{\beta}_1+\Delta^*} g_{\beta_1}(\beta_1)\,d\beta_1 \;\leq\; (1-\alpha)\times(1.005) \qquad (4.7)$$

Dado el valor de Δ^*, se evaluó la longitud del intervalo incondicional $(2\Delta^*)$, para después poder compararla con (4.4).

Una vez obtenida la simulación se evaluó el cociente de las longitudes

$$P = \frac{\Delta^*}{t_{(n-2)}^{1-\frac{\alpha}{2}} \sqrt{\dfrac{\widehat{\sigma_y^2}}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}},$$

para luego determinar con todas las simulaciones en qué porcentaje el intervalo incondicional resultó ser mayor que el intervalo condicional.

Estos resultados se presentan sintetizados en el apéndice B.

Capítulo 5

Conclusiones

En este trabajo se contrastaron, el intervalo de confianza condicional y el intervalo de confianza incondicional (ambos de longitud mínima), asociados a la pendiente de la recta del modelo de Regresión Lineal Simple.

Del análisis se derivaron los siguientes casos.

- ρ conocido.

  En el capítulo 3 se obtuvo el intervalo de confianza incondicional en forma analítica, el cual se comparó con el intervalo de confianza condicional.

  Se observó, que para tamaños de muestra grandes, la longitud de los intervalos resultó ser esencialmente la misma, ya que la diferencia entre las longitudes esta asociada al cuantil de la distribución que involucra cada uno de los intervalos. Es decir, para el intervalo incondicional se asocian cuantiles de una distribución t de Student y para el intervalo condicional corresponden a una normal, por lo cual la diferencia se minimiza conforme n aumenta.

  Pero para tamaños de muestra pequeõs, esta comparación no es del todo adecuada por los supuestos que respaldan a el criterio de comparación, en este caso es necesario determinar otra forma de compararlos.

- ρ desconocido.

En el capítulo 4 se obtuvo el intervalo de confianza incondicional en forma numérica, que se comparó con el intervalo de confianza condicional.

Se pudo observar, que a medida que n crece la longitud de los intervalos es, cada vez, más pequeña llegando a oscilar entre dos por ciento por abajo de la longitud condicional y dos por ciento por arriba de la misma, esta variación por abajo de la longitud del intervalo codicional se debe a redondeos numéricos.

Para tamaños de muestra pequeños ($n = 10$), la diferencia entre la longitud del intervalo incondicional con respecto a la del intervalo condicional va disminuyendo, a medida que el nivel de confianza aumenta, sin embargo no llega a ser tan pequeña esta diferncia, ya que el cociente es al menos 1.08, lo cual puede conducir a conclusiones diferentes de las que se puedan obtener al utilizar un análisis condicional.

Por lo tanto, se puede concluir que el incorporar el supuesto de que la variable independiente es aleatoria produce un incremento sustancial en la longitud del intervalo de confianza del parámetro β_1 si el tamaño de muestra es pequeño. Para tamaños de muestra grandes al considerar fija la variable independiente o aleatoria, los intervalos de confianza presentan esencialmente la misma longitud.

Por último, es necesario mencionar que en el artículo de Gull (1989) se presenta una solución a este problema desde un enfoque Bayesiano.

Este problema se puede extender al caso múltiple, que, por supuesto, esta abierto para desarrollarse.

Apéndice A

Listado del Programa

```
cccccc
cccccc Este es el progama principal que evalua el intervalo de confianza
cccccc para beta1.
cccccc
cccccc Consta de tres partes
cccccc
cccccc 1a. Subrutina REG_COND
cccccc     Que genera dos muestras de normales y calcula los estimadores
cccccc     de la regresion en forma usual. A su vez hace la llamada de otra
cccccc     subrutina (Subrutina T_ST)
cccccc               Subrutina T_ST
cccccc               Obtiene el cuantil de una t de Student
cccccc 2a. Subrutina CTTE
cccccc     Que evalua la constante ''de normalizacion'' para la densidad
cccccc     incondicional de beta1
cccccc 3a. Subrutina INTERV_BETA1
cccccc     Que evalua el intervalo de beta1 en limites fijos
cccccc

      integer NUM,ISEED,KCONT,KSIM,KTOT

      double precision ALPHA,PARAM(6),BETA1EST,RHOEST,DELTA,CTE,AREA,
     *       DELTA1,DELTA2,AREA2,COTAINF,COTASUP,COTA,APROX,A1,B1,D0,
     *       D1,PROMCON,DSVCON,PROMINC,DSVINC,TOT,PORCNT,XVAR,YVAR

      ISEED = 124007
```

```
      KSIM = 1
      KTOT = 1000
      APROX = 5.0d-03
      NUM = 10
      ALPHA = 0.10d0
      COTA = (1.0d0 - ALPHA)/2.0d0
      COTAINF = -COTA*APROX
      COTASUP = COTA*APROX
      PROMCON = 0.0d0
      PROMINC = 0.0d0
      DSVCON = 0.0d0
      DSVINC = 0.0d0

      PARAM(1) = 0.0d0
      PARAM(2) = 9.0d-25
      PARAM(3) = 10.0d0
      PARAM(4) = 2.0d0
      call rnset(ISEED)
    4 if (KSIM .le. KTOT) then
      KCONT = 0
      do 1 j = 1,100
        a = rnunf()
    1 continue
      PARAM(5) = 10.0d0*rnunf()
      PARAM(6) = PARAM(5)*sqrt(PARAM(4))

      call REG_COND(NUM,ALPHA,PARAM,BETA1EST,RHOEST,DELTA1,
     *          XVAR,YVAR)

cccccc genera las dos normales y calcula los estimadores de la regresion,
cccccc ademas la longitud del intervalo de confianza para beta1 y el
cccccc estimador de el cociente de varianzas.

      PROMCON = PROMCON + DELTA1
      DSVCON = DSVCON + (abs(DELTA1))**2.0d0

      call CTTE(NUM,RHOEST,CTE)

cccccc evalua la cte de la densidad de beta1.
```

```
    A1 = BETA1EST + 9.0d-25
    B1 = BETA1EST + abs(BETA1EST/2.0d0)

  2 if (KCONT .le. 200) then
      call INTERV_BETA1(NUM,BETA1EST,RHOEST,CTE,A1,B1,AREA,
   *       DELTA2)
      DELTA2 = 2.0d0*DELTA2
      AREA2 = AREA - COTA

      if ((AREA2 .ge. COTAINF).and.(AREA2 .le. COTASUP)) then
        go to 3
      else
        D1 = B1
        if (KCONT .eq. 0 ) then
          D0 = A1
          KCONT = 1
        else
          KCONT = KCONT + 1
        end if
        if (AREA2 .gt. COTASUP) then
          B1 = D1 - abs((D1-D0)/2.0d0)
          go to 2
        else
          B1 = D1 + abs((D1-D0)/2.0d0)
          go to 2
        end if
      end if
 3    PROMINC = PROMINC + DELTA2
      DSVINC = DSVINC + (abs(DELTA2))**2.0d0
      write(*,5)(PARAM(i)**2.0d0,i=5,6),XVAR,YVAR,BETA1EST,RHOEST,
   *           DELTA1,DELTA2
 5    format(8f15.5)
      KSIM = KSIM + 1
      go to 4
    else
      go to 4
    end if
    end if
```

```
      TOT = dble(KTOT)
      PROMCON = PROMCON/TOT
      PROMINC = PROMINC/TOT
      DSVCON = (DSVCON - TOT*((abs(PROMCON))**2.0d0))/(TOT-2.0d0)
      DSVINC = (DSVINC - TOT*((abs(PROMINC))**2.0d0))/(TOT-2.0d0)
      write(*,6)ISEED,(PARAM(i),i=3,4)
    6 format(/,i6,2x,2f15.5)
      write(*,7)PROMCON,PROMINC,DSVCON,DSVINC
    7 format(4f15.5)
      end


      subroutine REG_COND(num,alpha,param,beta1est,rhoest,delta1,
     *           xxvar,yyvar)

cccccc
cccccc Se genera una muestra de tamanio num de una Normal con media
cccccc param(1) y desviacion param(5) que se guarda en el vector xx.
cccccc
cccccc Se genera una muestra de tamanio num de una Normal con media
cccccc param(2)+param(3)*xx y desviacion param(6) que se guarda en el
cccccc vector yy.
cccccc
cccccc El cociente de varianzas es param(4).
cccccc
cccccc Los valores que se obtienen son:
cccccc     estimador de param(3) es decir beta1est
cccccc     estimador de param(4) es decir rhoest
cccccc     longitud del intervalo condicional al (1-alpha)x100% de
cccccc     confianza de beta1est es decir delta
cccccc

      integer num
      parameter (max = 500)
      double precision alpha,param(6),delta1,xx(max),yy(max),xxmed,
     *       yymed,sxx,sxy,syy,beta0est,beta1est,rhoest,xxvar,yyvar,
     *       yxinterv,dn,tst
```

```
tst = 0.0d0
xxmed = 0.0d0
yymed = 0.0d0

do j = 1,max
  xx(j) = 0.0d0
  yy(j) = 0.0d0
end do

call drnnoa(num,xx)
call dscal(num,param(5),xx,1)
call dadd(num,param(1),xx,1)
call drnnoa(num,yy)

do j = 1,num
  yy(j) = param(2) + param(3)*xx(j) + param(6)*yy(j)
end do
do j = 1,num
  xxmed = xxmed + xx(j)
  yymed = yymed + yy(j)
end do

dn = dble(num)
xxmed = xxmed/dn
yymed = yymed/dn

sxx = 0.0d0
sxy = 0.0d0
do j = 1,num
  sxx = sxx + (abs(xx(j)-xxmed))**2.0d0
  sxy = sxy + (xx(j)-xxmed)*yy(j)
end do

beta1est = sxy/sxx
beta0est = yymed - xxmed*beta1est

syy = 0.0d0
do j = 1,num
  syy = syy + (abs(yy(j)-beta0est-beta1est*xx(j)))**2.0d0
```

```
      end do

      xxvar = sxx/(dn-1.0d0)
      yyvar = syy/(dn-2.0d0)
      rhoest = yyvar/xxvar
      yxinterv = sqrt(yyvar/sxx)

      call T_ST(num,alpha,tst)

cccccc tst es el cuantil de una t de Student con num-2 g.l y (1-alpha/2)

      delta1 = 2.0d0*tst*yxinterv
      return
      end


      subroutine T_ST(nu,alp,ts)

CCCCCC
CCCCCC Evalua el cuantil de una t de Student con num-2 g.l. y 1-alpha/2
CCCCCC

      integer nu
      double precision alp,ts,p,df,dtin,dn
      external dtin

      dn = dble(nu)
      df = dn - 2.0d0
      p = 1.0d0 - alp/2.0d0
      ts = dtin(p,df)
      return
      end


     subroutine CTTE(num,rhoest,cte)

cccccc
cccccc El valor de constante que involucra gamas se calcula por
cccccc descomposicion de las mismas.
```

```
cccccc

      integer num
      double precision dn,rhoest,cte,cte1,aux1,aux3,aux4,aux5,
     *        aux6,aux8,aux9,aux10,pi2

      pi2 = 8.0d0*atan(1.0d0)
      dn = dble(num)
      aux1 = dn - 2d0
      aux2 = 2*num - 5
      aux3 = dn - 1d0
      aux7=num - 3
      aux4 = 0.5d0*aux3*(log(aux3)-log(aux1))
      aux1 = log(aux1)
      aux5 = 0.0d0

      do i = 3,aux2,2
        aux5 = aux5 + log(dble(i))
      end do
      aux6 = aux1 + aux4 + aux5
      aux8 = 0.0d0

      if ( mod(num,2.0) .eq. 0.0) then
        do i = 3,aux7,2
         aux8 = aux8 + log(dble(i))
        end do
        aux8 = 2.*d0*aux8
        aux9 = log(pi2) + 0.5*log(rhoest)
        go to 20
      else
        do i = 2,(aux7/2.0)
          aux8 = aux8 + log(dble(i))
        end do
        aux8 = 2.0d0*aux8
        aux9 = aux3*log(2.0d0) + 0.5*log(rhoest)
        go to 20
      end if
   20 aux10 = aux8 + aux9
      cte = exp(aux6-aux10)
```

```
      return
      end


      subroutine INTERV_BETA1(num,beta1est,rhoest,cte,a1,b1,
     *           area,delta2)

cccccc
cccccc utiliza la subrutina dqdag (basada en la regla de Guass-Kronrod)
cccccc Este programa evalua la funcion de densidad de beta1
cccccc

      integer irule,maxsub,num
      double precision g,a1,b1,errabs,errest1,errrel,result1,cte,area,
     *       beta1est,rhoest,delta2,beta,rho,aux0,exp1,exp2,exp3,exp4,
     *       alist1(800),blist1(800),rlist1(800),elist1(800),iord1(800)

      external g,dq2ag
      common /est/ beta,rho
      common /param/ aux0,exp1,exp2,exp3,exp4
      common /err/ errabs,errrel
      common /int/ irule,maxsub

      dn = dble(num)

      irule = 1
      maxsub = 800
      errabs = 1.0d-03
      errrel = 1.0d-05
      aux0 = (dn-1.0d0)/(dn-2.0d0)
      exp1 = dn-2.0d0
      exp2 = dn/2.0d0
      exp3 = (2.0d0*dn-3.0d0)/2.0d0
      exp4 = (dn-3.0d0)/2.0d0
      beta = beta1est
      rho = rhoest

      call dq2ag(g,a1,b1,errabs,errrel,irule,result1,errest1,maxsub,
     *     neval1,nsubin1,alist1,blist1,rlist1,elist1,iord1)
```

```
  area =  (exp(log(cte) + log(result1)))
  delta2 = abs(b1 - a1)
  return
  end


  double precision function g(y)

  integer irule,maxsub
  double precision y,f,a2,b2,result2,errest2,beta12,
 *       errabs,errrel,beta,rho,
 *       alist2(800),blist2(800),rlist2(800),elist2(800),iord2(800)
  external f,dq2ag
  common /est/ beta,rho
  common /err/ errabs,errrel
  common /int/ irule,maxsub
  common /var/ beta12

  a2 = 9.0d-25
  b2 = 1.0d0
  beta12=((abs(y-beta))**2.0d0)/rho

  call dq2ag(f,a2,b2,errabs,errrel,irule,result2,errest2,maxsub,
 *     neval2,nsubin2,alist2,blist2,rlist2,elist2,iord2)
  g = result2
  return
  end


  double precision function f(x)

  double precision x,aux1,aux2,aux3,aux4,aux5,aux6,f1,f2,
 *       beta12,aux0,exp1,exp2,exp3,exp4

  common /var/ beta12
  common /param/ aux0,exp1,exp2,exp3,exp4

  aux1 = exp1*log(x)
```

```
aux2 = exp2*log(abs(x+beta12))
aux3 = exp3*log(abs(1+x*aux0))
f1 = exp(aux1 - aux2 - aux3)
aux4 = exp4*log(x)
aux5 = exp2*log(abs(1+x*beta12))
aux6 = exp3*log(abs(x+aux0))
f2 = exp(aux4 - aux5 - aux6)
f = f1 + f2
return
end
```

Apéndice B

Simulaciones

En cada uno de los casos analizados, se define los rangos en los que cae el cociente de longitudes (P). El intervalo incondicional presenta longitud mayor que el intervalo condicional, en la mayoría de las simulaciones, en algunos casos la longitud del intervalo incondicional llega a ser más pequeña, esto se debe a redondeos númericos.

Las simulaciones realizadas para $n = 10$ con los diferentes valores de ϑ_1, ρ y α se sintetizan en las tablas 1 y 2, para $n = 50$, se localizan en las tablas 3 y 4, para $n = 100$, en las tablas 5 y 6, por último, para $n = 500$, se presentan los resultados en las tablas 7 y 8.

Tamaño de muestra 10

Tabla 1. Comparación del cociente de longitudes.

$\alpha = 0.05$					
ρ	β_1	$1.0 \leq P < 1.12$	$1.12 \leq P < 1.15$	$1.15 \leq P < 1.20$	$P \geq 1.20$
0.01	0.01	0	307	693	0
	1.0	0	311	689	0
	10.0	0	317	683	0
1.0	0.01	0	314	686	0
	1.0	0	302	698	0
	10.0	0	323	677	0
2.0	0.01	0	308	692	0
	1.0	0	313	687	0
	10.0	0	302	698	0

Tabla 2. Comparación del cociente de longitudes.

$\alpha = 0.10$					
ρ	β_1	$1.0 \leq P < 1.08$	$1.08 \leq P < 1.10$	$1.10 \leq P < 1.13$	$P \geq 1.13$
0.01	0.01	0	620	380	0
	1.0	0	599	401	0
	10.0	0	607	393	0
1.0	0.01	0	587	413	0
	1.0	0	603	397	0
	10.0	0	570	430	0
2.0	0.01	0	597	403	0
	1.0	0	566	434	0
	10.0	0	612	388	0

Tamaño de muestra 50

Tabla 3. Comparación del cociente de longitudes.

$\alpha = 0.05$					
ρ	β_1	$P = 1.0$	$1.0 < P \leq 1.03$	$1.03 < P \leq 1.06$	$P > 1.06$
0.01	0.01	0	621	379	0
	1.0	0	597	403	0
	10.0	0	588	412	0
1.0	0.01	0	604	396	0
	1.0	0	604	396	0
	10.0	0	593	407	0
2.0	0.01	0	590	410	0
	1.0	0	609	391	0
	10.0	0	631	369	0

Tabla 4. Comparación del cociente de longitudes.

$\alpha = 0.10$					
ρ	β_1	$P = 1.0$	$1.0 < P \leq 1.01$	$1.01 < P \leq 1.04$	$P > 1.04$
0.01	0.01	0	301	699	0
	1.0	0	296	704	0
	10.0	0	291	709	0
1.0	0.01	0	307	693	0
	1.0	0	284	716	0
	10.0	0	312	688	0
2.0	0.01	0	306	694	0
	1.0	0	303	697	0
	10.0	0	321	679	0

Tamaño de muestra 100

Tabla 5. Comparación del cociente de longitudes.

$\alpha = 0.05$					
ρ	β_1	$P < 0.99$	$0.99 \leq P < 1.0$	$1.0 \leq P < 1.04$	$P \geq 1.04$
0.01	0.01	0	202	798	0
	1.0	0	216	784	0
	10.0	0	312	683	0
1.0	0.01	0	179	821	0
	1.0	0	199	801	0
	10.0	0	203	797	0
2.0	0.01	0	204	796	0
	1.0	0	203	797	0
	10.0	0	186	814	0

Tabla 6. Comparación del cociente de longitudes.

$\alpha = 0.10$					
ρ	β_1	$P < 0.99$	$0.99 \leq P < 1.0$	$1.0 \leq P < 1.03$	$P \geq 1.03$
0.01	0.01	0	251	749	0
	1.0	0	249	751	0
	10.0	0	246	754	0
1.0	0.01	0	241	759	0
	1.0	0	266	734	0
	10.0	0	244	756	0
2.0	0.01	0	250	750	0
	1.0	0	250	750	0
	10.0	0	242	758	0

Tamaño de muestra 500

Tabla 7. Comparación del cociente de longitudes.

$\alpha = 0.05$					
ρ	β_1	$P < 0.98$	$0.98 \leq P < 1.0$	$1.0 \leq P < 1.03$	$P \geq 1.03$
0.01	0.01	0	566	434	0
	1.0	0	573	427	0
	10.0	0	536	464	0
1.0	0.01	0	574	426	0
	1.0	0	578	422	0
	10.0	0	583	417	0
2.0	0.01	0	589	411	0
	1.0	0	568	432	0
	10.0	0	612	388	0

Tabla 8. Comparación del cociente de longitudes.

$\alpha = 0.10$					
ρ	β_1	$P < 0.98$	$0.98 \leq P < 1.0$	$1.0 \leq P < 1.02$	$P \geq 1.02$
0.01	0.01	0	647	353	0
	1.0	0	658	342	0
	10.0	0	672	328	0
1.0	0.01	0	657	343	0
	1.0	0	641	359	0
	10.0	0	629	371	0
2.0	0.01	0	655	345	0
	1.0	0	646	354	0
	10.0	0	675	325	0

Apéndice C

Desarrollos Algebraicos

En este apéndice se presenta el álgebra que se requirió para poder realizar el programa hecho en lenguaje Fortran.

Recuérdese que la función que se va a evaluar numéricamente es

$$
\begin{aligned}
g_{\beta_1}(\beta_1) &= \frac{\Gamma(\frac{n}{2})\Gamma(\frac{2n-3}{2})}{\Gamma(\frac{n-1}{2})^2\Gamma(\frac{n-2}{2})}\left(\frac{n-1}{n-2}\right)^{\frac{n-1}{2}} \\
&\times \int_0^\infty \left\{ \frac{1}{(\pi q\widehat{\rho})^{\frac{1}{2}}} q^{\frac{n-3}{2}} \left(1 + \frac{1}{q\rho}(\widehat{\beta}_1 - \beta_1)^2\right)^{-\frac{n}{2}} \left(1 + \frac{n-1}{n-2}q\right)^{-\frac{2n-3}{2}} \right\} dq.
\end{aligned} \tag{0.1}
$$

la cual se puede reescribir como

$$
\begin{aligned}
g_{\beta_1}(\beta_1) &= \frac{\Gamma(\frac{n}{2})\Gamma(\frac{2n-3}{2})}{\Gamma(\frac{n-1}{2})^2\Gamma(\frac{n-2}{2})\Gamma(\frac{1}{2})}\left(\frac{n-1}{n-2}\right)^{\frac{n-1}{2}} \widehat{\rho}^{-\frac{1}{2}} \\
&\times \int_0^\infty \left\{ \frac{q^{\frac{n-4}{2}}}{\left(1+\frac{(\widehat{\beta}_1-\beta_1)^2}{q\rho}\right)^{\frac{n}{2}} \left(1+\frac{n-1}{n-2}q\right)^{\frac{2n-3}{2}}} \right\} dq.
\end{aligned} \tag{0.2}
$$

Se expresará la constante, que sólo depende de n en forma más sencilla , es decir si

$$
C = \frac{\Gamma(\frac{n}{2})\Gamma(\frac{2n-3}{2})}{\Gamma(\frac{n-1}{2})^2\Gamma(\frac{n-2}{2})\Gamma(\frac{1}{2})}\left(\frac{n-1}{n-2}\right)^{\frac{n-1}{2}}, \tag{0.3}
$$

se puede notar que

$$\Gamma\left(\frac{i}{2}\right) = \left(\frac{i}{2} - 1\right)\Gamma\left(\frac{i}{2} - 1\right), \quad \forall\, i \in I\!N.$$

Así, de la ecuación (0.3),

$$\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-2}{2})} = \left(\frac{n-2}{2}\right). \tag{0.4}$$

De igual forma en el caso en el que el argumento de la función Gama no sea natural, si se puede expresar de la forma $i + \frac{1}{2}$ con $i \in I\!N$,[1] entonces

$$\Gamma(i + \frac{1}{2}) = \frac{\{1 \cdot 3 \cdot 5 \cdots (2i - 1)\}\sqrt{\pi}}{2^i}.$$

En la ecuación (0.3) se observa que

$$\begin{aligned} \Gamma(\tfrac{2n-3}{2}) &= \Gamma((n-2) + \tfrac{1}{2}) \\ &= \frac{\{1{\cdot}3{\cdot}5\cdots(2(n-2)-1)\}\Gamma(\frac{1}{2})}{2^{n-2}} \end{aligned} \tag{0.5}$$

por lo cual

$$\frac{\Gamma(\frac{2n-3}{2})}{\Gamma(\frac{1}{2})} = \frac{\{1{\cdot}3{\cdot}5\cdots(2n-5)\}}{2^{n-2}}, \tag{0.6}$$

por otro lado, si n es par se tiene

$$\begin{aligned} \Gamma(\tfrac{n-1}{2}) &= \Gamma(\tfrac{n-2}{2} + \tfrac{1}{2}) \\ &= \frac{\{1{\cdot}3{\cdot}5\cdots(n-3)\}\Gamma(\frac{1}{2})}{2^{\frac{n-2}{2}}} \end{aligned} \tag{0.7}$$

mientras, que si es impar se obtiene

$$\begin{aligned} \Gamma(\tfrac{n-1}{2}) &= (\tfrac{n-1}{2} - 1)! \\ &= \{1 \cdot 2 \cdot 3 \cdots \tfrac{(n-3)}{2}\}. \end{aligned} \tag{0.8}$$

[1] Véase Mood pag. 534, ecuación (40)

De esta forma, dependiendo de si n es par o impar, se tiene una expresión particular para la constante (0.3).

Si n es par:

$$C = \frac{(n-2)}{2\pi\widehat{\rho}^{\frac{1}{2}}} \frac{\left\{1\cdot 3\cdot 5\cdots(2n-5)\right\}}{\left\{1\cdot 3\cdot 5\cdots(n-3)\right\}^2}\left(\frac{n-1}{n-2}\right)^{\frac{n-1}{2}}. \qquad (0.9)$$

Si n es impar:

$$C = \frac{(n-2)}{2^{n-1}\widehat{\rho}^{\frac{1}{2}}} \frac{\left\{1\cdot 3\cdot 5\cdots(2n-5)\right\}}{\left\{1\cdot 2\cdot 3\cdots(\frac{n-3}{2})\right\}^2}\left(\frac{n-1}{n-2}\right)^{\frac{n-1}{2}}. \qquad (0.10)$$

Por otro lado, la integral de (0.3) se puede reescribir como

$$\int_0^\infty h_q(q)dq = \int_0^1 h_q(q)dq + \int_1^\infty h_q(q)dq \qquad (0.11)$$

donde

$$h_q(q) = \left\{\frac{q^{\frac{n-4}{2}}}{\left(1+\frac{(\widehat{\beta}_1-\beta_1)^2}{q\widehat{\rho}}\right)^{\frac{n}{2}}\left(1+\frac{n-1}{n-2}q\right)^{\frac{2n-3}{2}}}\right\}.$$

Analizando por separado cada sumando se tiene lo siguiente. el primer término del lado derecho es:

$$\int_0^1\left\{\frac{q^{\frac{n-4}{2}}}{\left(1+\frac{(\beta_1-\widehat{\beta}_1)^2}{q\widehat{\rho}}\right)^{\frac{n}{2}}\left(1+\frac{n-1}{n-2}q\right)^{\frac{2n-3}{2}}}\right\}dq = \int_0^1\left\{\frac{q^{n-2}}{\left(q+\frac{(\beta_1-\widehat{\beta}_1)^2}{\widehat{\rho}}\right)^{\frac{n}{2}}\left(1+\frac{n-1}{n-2}q\right)^{\frac{2n-3}{2}}}\right\}dq$$

y el segundo término del lado derecho bajo un cambio de variable ($v = q^{-1}$) es:

$$\int_1^\infty\left\{\frac{q^{\frac{n-4}{2}}}{\left(1+\frac{(\widehat{\beta}_1-\beta_1)^2}{q\widehat{\rho}}\right)^{\frac{n}{2}}\left(1+\frac{n-1}{n-2}q\right)^{\frac{2n-3}{2}}}\right\}dq = \int_0^1\left\{\frac{v^{\frac{n-3}{2}}}{\left(1+\frac{(\beta_1-\widehat{\beta}_1)^2}{\widehat{\rho}}v\right)^{\frac{n}{2}}\left(v+\frac{n-1}{n-2}\right)^{\frac{2n-3}{2}}}\right\}dv.$$

Renombrando la variable q, la integral de la distribución de β_1 (0.2) se transforma en

$$\int_0^1 \left\{ \frac{v^{n-2}}{\left(v + \frac{(\widehat{\beta}_1 - \beta_1)^2}{\rho}\right)^{\frac{n}{2}} \left(1 + \frac{n-1}{n-2} v\right)^{\frac{2n-3}{2}}} \right\} dv + \int_0^1 \left\{ \frac{v^{\frac{n-3}{2}}}{\left(1 + \frac{(\beta_1 + \widehat{\beta}_1)^2}{\rho} v\right)^{\frac{n}{2}} \left(v + \frac{n-1}{n-2}\right)^{\frac{2n-3}{2}}} \right\} dv.$$

Con estos cambios se programó la función, siendo así más eficiente en la evaluación.

Bibliografía

[1] Benjamini Y. and Fuchs C. (1990). "Conditional Versus Unconditional Analysis in Some Regression Models". *Commun. Statist. Theory Meth.*, Vol. 19, No. 12, 4731-4756.

[2] Berger O. J. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd. ed. Springer-Verlang.

[3] Binkley J. K. and Abbott P. C. (1987). "The Fixed X Assumption in Econometrics: Can the Textbooks be Trusted?". *The American Statistician*, Vol. 41, No. 3, 206-214.

[4] Burden R. L.and Faires J. D. (1985). *Análisis Numérico*, Grupo Editorial Iberoamericana.

[5] Draper N. R. and Smith H. (1981). *Applied Regression Analysis*, 2nd. ed., New York, Wiley.

[6] Feller W. (1989). *Introducción a la Teoría de Probabilidades y sus aplicaciones*, Vol. 1, Editorial Limusa.

[7] Gull, S. F. (1989). "Bayesian data analysis: straigth-line fitting. In Maximum Entropy and Bayesian Methods (J. Skillind, ed.)". *Dordrecht, Netherlands: Kluwer Academic Publishers*, 511-518.

[8] Johnson N. L. and Kotz S. (1970). *Continuous Univariate Distributions 1*, Distributions in Statistics, Houghton Mifflin Series in Statistics, John Wiley and Sons, Inc.

[9] Johnson N. L. and Kotz S. (1970). *Continuous Univariate Distributions 2*, Distributions in Statistics, Houghton Mifflin Series in Statistics, John Wiley and Sons, Inc.

[10] Judge G. G., Hill R. C., Griffiths W. E., Lütkepohl H., and Lee T. C. (1988). *Introduction to the Theory and Practice of Econometrics*, 2nd. ed. New York, Wiley, Chapter 13 *Stochastic regressors*, 571-595.

[11] Kinal T. and Lahiri K. (1983). "Specification Error Analysis With Stochastic Regressors". *Econometrica*, Vol. 51, No. 4, 1209-1219.

[12] Montgomery D. C. and Peck E. A. (1992). *Introduction to Linear Regression Analysis*, 2nd. ed., Wiley, New York.

[13] Mood M. A., Graybill A. F. and Boes C. D. (1974). *Introduction to the Theory of Statistics*, 3th. ed., McGraw Hill International Editions.

[14] Sampson A. R. (1974). "A Tale of Two Regressions". *Journal of the American Statistical Association*, Vol. 69, No. 347, 682-689.

[15] Sprott D. A. and Farewell V. T. (1993). "The Difference Between Two Normal Means". *The American Statistician*, Vol. 47, No. 2, 126-128.

[16] Weisberg S. (1980). *Applied Linear Regression*, Wiley, New York.