UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Introducción a la Regresión No Paramétrica

T E S I S

QUE PARA OBTENER EL TÍTULO DE

A C T U A R I O

P R E S E N T A

JOSUÉ DAMIÁN ROJAS TREJO

Tutor:
M.EN A.P. MARÍA DEL PILAR ALONSO REYES.

FACULTAD DE CIENCIAS
UNAM

2009

Hoja de Datos del Jurado

1. Datos del alumno
Rojas
Trejo
Josué Damián
56 74 48 36
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
098000519

2. Datos del tutor
M. en A.P.
María del Pilar
Alonso
Reyes

3. Datos del sinodal 1
Mat.
Margarita Elvira
Chávez
Cano

4. Datos del sinodal 2
M. en C.
José Antonio
Flores
Díaz

5. Datos del sinodal 3
Act.
Francisco
Sánchez
Villarreal

6. Datos del sinodal 4
M. en I.D.O.
María Isabel
Escalante
Membrillo

7. Datos del trabajo escrito.
Introducción a la Regresión No Paramétrica
116 P
2009

ÍNDICE GENERAL

INTRODUCCIÓN

Cuando se presentan bases de datos y se quiere pronosticar el comportamiento de una variable, como puede ser el índice de cotización de la bolsa, uno hace referencia a la técnica de Análisis de Regresión de inmediato, sin considerar si se tienen los elementos estadísticos que requiere ésta y sin saber si el método puede ser considerado bajo otro marco, como es la Estadística No Paramétrica.

El objetivo de este trabajo es presentar de manera general el análisis de la regresión no paramétrica y hacer una comparación con los métodos de regresión paramétrica, para finalmente poder concluir cuál es el que más conviene utilizar, cuáles son sus ventajas y desventajas, así como, qué oportunidad tendrá el investigador en usar métodos no paramétricos.

El presente trabajo se divide en tres capítulos:

En el primer capítulo llamado Regresión Lineal Paramétrica, se hará una revisión general de los conceptos básicos de Regresión Lineal como introducción al lector de las cualidades que este método debe llevar y qué debe saber la persona que lo aplique.

El segundo capítulo llamado Regresión no Paramétrica, abarcará algunos métodos para calcular la recta de regresión, para hacer pruebas de hipótesis, así como también calcular intervalos de confianza, utilizando técnicas relativamente simples y que llevan los nombres de los que la generaron.

Finalmente, en el último capítulo se aplicarán todas las pruebas presentadas en la sección anterior, así como una comparación con el modelo de regresión conocido.

Por último se presentan las conclusiones del trabajo, la bibliografía empleada, un apartado de anexos donde se incluyen las tablas de la base de análisis.

CAPITULO I. Regresión Lineal Paramétrica.

1.1 Introducción

El Análisis de Regresión, es una de las técnicas estadísticas más utilizadas para investigar y modelar la relación entre dos o más variables cualitativas o cuantitativas de tal forma que una variable cuantitativa puede ser descrita por otra u otras variables.

Ésta tiene un sin fin de aplicaciones y es utilizada en casi todos los campos como por ejemplo: la ingeniería, las ciencias físicas, la economía, las ciencias biológicas, sociales, entre otros.

En otras palabras la regresión sirve para poner en evidencia las relaciones que existen entre diversas variables.

Su objetivo es investigar la relación estadística que existe entre una variable dependiente y una o más variables independientes. Para poder realizar esta investigación, se debe postular una relación funcional entre las variables. Debido a su simplicidad analítica, la forma funcional que más se utiliza en la práctica es la relación lineal.

1.2 Tipos de variables y escalas

Una variable es una característica que puede tomar más de un valor entre los miembros de una muestra o población.

Las variables se pueden clasificar según:

Criterio	***Clasificación***
Su naturaleza	-Cuantitativas -Cualitativas
La escala en que se mida	-Ordinal -Nominal -De razón -Intervalo

Según su naturaleza:

Variables Cualitativas: Son aquellas que no aparecen en forma numérica, sino como categorías o atributos (profesión, color de ojos, entre otras).

Variables Cuantitativas: Son aquellas en donde las variables se pueden expresar numéricamente (temperatura, salario, número de goles en un partido).

Otra manera en la que las variables cuantitativas se pueden clasificar tiene que ver con el posible número de valores que teóricamente puedan asumir.

Las variables cuantitativas según el tipo de valores que puedan tomar pueden ser discretas o continuas.

Una variable es *discreta* si puede tomar un valor en un conjunto que sea numerable, es decir es el resultado de contar y sólo tomar valores enteros (ej. número de hijos).

Una variable es *continua* si puede tomar su valor de un conjunto de posibles valores continuos, en otras palabras es el resultado de medir, y éstas pueden contener decimales (temperatura, peso, altura, etc).Además se pueden subdividir a voluntad. En conclusión pueden tomar cualquier valor de un determinado intervalo.

Según la escala en que se mida:

Las escalas de medidas de la variable son: Nominal, Ordinal, de Intervalo, de Razón o Proporción.

Escala Nominal: Asigna valores a las observaciones que realizamos. Así clasifica a los individuos estudiados en grupos distintos, atendiendo a las diferentes clases o categorías de la variable en estudio. (ej. sexo: 1.-masculino; 2.-femenino).

Escala Ordinal: Clasifica y ordena a los sujetos de un estudio en grupos que tienen diferente cantidad o grado de la categoría de la variable en estudio. (ej. Primero, Segundo, Tercero…).

Escala de Intervalo: Clasifica y ordena las categorías, asignando números. La asignación de estos números se realiza en base al valor numérico de unidades de medida que posee cada individuo. Esta unidad de media es constante, es decir se aplica la misma a todos los individuos y el investigador decide cuál es. (ej. La medición del IQ).

Escala de Razón: Esta escala es la más precisa. Se diferencia de la de intervalo en que el punto 0 corresponde exactamente a la ausencia de atributo. Las cifras obtenidas de aplicar estas escalas pueden operarse en cualquier sentido y gozan de proporcionalidad (ej. un árbol que mide cuatro metros mide el doble que el que mide dos metros).

1.3 Regresión Lineal

El objetivo principal de ésta es ajustar la mejor recta.

1.- La "mejor" en términos de que represente bien los datos (es decir, a la gran mayoría).

2.- "La mejor recta" será aquella cuyos parámetros tengan las mejores cualidades estadísticas.

Tipos de Regresión Lineal:

Existen dos tipos de regresión lineal, las cuales se clasifican de acuerdo a sus parámetros:

- Regresión Lineal Simple
- Regresión Lineal Múltiple

1.3.1 Regresión Lineal Simple

En la regresión lineal simple se maneja solamente una variable independiente o explicativa, por lo que se puede decir que solamente se cuenta con dos parámetros β_0 y β_1.

De tal forma que el modelo lineal simple es el siguiente:

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ (modelo teórico) en este caso ésta sería una regresión de y en x.

donde:

y_i = Variable dependiente (variable aleatoria)

x_i = Variable independiente (variable no aleatoria)

ε_i = Error de medida (variable aleatoria) y debe cumplir las siguientes condiciones:

1. $E(\varepsilon_i) = 0$

2. $Var(\varepsilon_i) = \sigma^2$

3. $\varepsilon_i \perp \varepsilon_j \ \forall \ i \neq j$ (es decir ε_i y ε_j tienen covarianza cero)

Para obtener el modelo ajustado lo único que se debe hacer es obtener la esperanza del modelo teórico teniendo como resultado lo siguiente:

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

(Observando las condiciones del error de medida queda lo siguiente)

$$= \beta_0 + \beta_1 x_i + 0$$
$$= \beta_0 + \beta_1 x_i$$
$$\therefore \quad E(y_i) = \beta_0 + \beta_1 x_i$$

De esta forma se puede decir que el modelo ajustado queda como sigue:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

La diferencia del modelo teórico y el modelo ajustado, da como resultado los residuales, es decir $y_i - \hat{y}_i = e_i$.

1.3.1.1 Estimación del Modelo

¿Cómo estimar β_0 y β_1?

Para estimar β_0 y β_1 se usan valores muestrales, estos valores son obtenidos de estudios experimentales o estudios observacionales.

De cualquiera de estas dos formas se tiene un conjunto de n observaciones

{ (x_1,y_1) , (x_2,y_2) , ..., (x_i,y_i) , ..., (x_n,y_n) }

1.3.1.1.1 Método de Mínimos Cuadrados

Es un método que sirve para poder estimar estos parámetros en donde lo que se busca es minimizar la suma de cuadrados de los errores.

En este caso la función que se utilizará para aplicar este método será:

$$G = \sum (y_i - \hat{y}_i)^2 = \sum e_i^2 = \sum \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

Obteniendo las derivadas parciales con respecto a $\hat{\beta}_0, \hat{\beta}_1$

$$\frac{\partial G}{\partial \hat{\beta} o} = 2\sum \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) \ (-1)$$

$$\frac{\partial G}{\partial \hat{\beta} 1} = 2\sum \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) \ (-x_i)$$

Desarrollando e igualando a cero, se llega al siguiente sistema de ecuaciones:

$n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i = \sum y_i$ ....(ec. 1)

$\sum x_i \hat{\beta}_0 + \hat{\beta}_1 \sum x_i^2 = \sum y_i x_i$ .... (ec. 2)

Haciendo las operaciones correspondientes se obtiene lo siguiente:

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{S_{xy}}{S_{xx}} \ *$$

---

* $S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$

1.3.1.1.2 Estimación por Máxima Verosimilitud

Cuando se supone $\varepsilon_i \sim N\left(0, \sigma^2\right)$ e independientes:

La función de verosimilitud está definida de la siguiente manera:

$$L = L\left(\beta_0, \beta_1, \underline{Y}\right)^{*} = \frac{e^{-\sum_{i=1}^{n}\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2}}}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}$$

Aplicando *ln* a la función queda lo siguiente:

$$\ln L = -\sum_{i=1}^{n}\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2} - \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2)$$

Entonces obteniendo las derivadas parciales con respecto a β_0, β_1 y σ^2 e igualando a "0" se llega a los siguientes resultados

$$\frac{\partial}{\partial \beta_0}\ln L = \frac{2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)}{2\sigma^2} = 0$$

$$\frac{\partial}{\partial \beta_1}\ln L = \frac{2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)(x_i)}{2\sigma^2} = 0$$

$$\frac{\partial}{\partial \sigma^2}\ln L = \frac{\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^4} - \frac{n}{2\sigma^2} = \frac{\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2 - \sigma^2 n}{2\sigma^4} = 0$$

---

* $\underline{Y} = y_1, \ldots\ldots, y_n$

Desarrollando se llega a las siguientes ecuaciones:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)(x_i) = 0$$

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 - \sigma^2 n = 0$$

Resolviendo el sistema de ecuaciones se tienen los siguientes resultados:

$$\hat{\beta}_0 = \bar{y} - \bar{x}\,\hat{\beta}_1, \quad \hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{S_{xy}}{S_{xx}}, \quad \hat{\sigma}^2 = \frac{\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n}$$

1.3.1.2 Propiedades de $\hat{\beta}_0$ y $\hat{\beta}_1$

Para $\hat{\beta}_1$ se tiene lo siguiente:

Se toma $\hat{\beta}_1 = \dfrac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$

$$= \frac{\sum x_i y_i - \sum \bar{x} y_i}{\sum x_i^2 - n\bar{x}^2}$$

$$= \frac{\sum (x_i y_i - \bar{x} y_i)}{\sum x_i^2 - n\bar{x}^2}$$

$$= \frac{\sum y_i (x_i - \bar{x})}{\sum x_i^2 - n\bar{x}^2}$$

$$= \frac{\sum y_i (x_i - \bar{x})}{S_{xx}}$$

Ahora se define $c_i = \dfrac{x_i - \bar{x}}{S_{xx}}$, entonces se tiene el siguiente resultado

$\hat{\beta}_1 = \sum c_i y_i$, con las siguientes características: $\sum_{i=1}^{n} c_i = 0$ y $\sum_{i=1}^{n} c_i\ x_i = 1$

$$
\begin{aligned}
\Rightarrow\ E\left(\hat{\beta}_1\right) &= E\left(\sum c_i y_i\right) \\
&= \sum_{i=1}^{n} E\left(c_i y_i\right) \\
&= \sum_{i=1}^{n} c_i\ E\left(y_i\right) \\
&= \sum_{i=1}^{n} c_i\ \left(\beta_0 + \beta_1\ x_i\right) \\
&= \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i\ x_i \\
&= \beta_0\ (0) + \beta_1\ (1) \\
&= \beta_1
\end{aligned}
$$

Teniendo esto ahora se procederá a calcular

$$
\begin{aligned}
Var\left(\hat{\beta}_1\right) &= Var\left(\sum c_i y_i\right) \\
&= \sum_{i=1}^{n} c_i^2\ Var\left(y_i\right) \\
&= \sum_{i=1}^{n} c_i^2\ \left(\sigma^2\right) \\
&= \sigma^2 \sum_{i=1}^{n} c_i^2 \\
&= \sigma^2 \sum_{i=1}^{n} \frac{\left(x_i - \bar{x}\right)^2}{S_{xx}^2} \\
&= \frac{\sigma^2}{S_{xx}}
\end{aligned}
$$

Revisando las propiedades de $\hat{\beta}_0$

$$
\begin{aligned}
E\left(\hat{\beta}_0\right) &= E\left(\overline{\mathrm{Y}} - \overline{X}\,\hat{\beta}_1\right) \\
&= E\left(\overline{\mathrm{Y}}\right) - \overline{X}\; E\left(\hat{\beta}_1\right) \\
&= E\left(\hat{\beta}_0 + \overline{X}\,\hat{\beta}_1\right) - \overline{X}\; E\left(\hat{\beta}_1\right) \\
&= \hat{\beta}_0 + \overline{X}\,\hat{\beta}_1 - \overline{X}\,\hat{\beta}_1 \\
&= \hat{\beta}_0
\end{aligned}
$$

$$
\begin{aligned}
Var\left(\hat{\beta}_0\right) &= Var\left(\overline{\mathrm{Y}} - \overline{X}\,\hat{\beta}_1\right) \\
&= Var\left(\overline{\mathrm{Y}}\right) + Var\left(\overline{X}\,\hat{\beta}_1\right) - 2\,\overline{X}\,\operatorname{cov}\left(\overline{\mathrm{Y}}, \hat{\beta}_1\right) \\
&= Var\left(\sum_{i=1}^{n} \frac{y_i}{n}\right) + \overline{X}^2\; Var\left(\hat{\beta}_1\right) \\
&= \frac{\sigma^2}{n} + \overline{X}^2\left(\frac{\sigma^2}{S_{xx}}\right) \\
&= \sigma^2\left(\frac{1}{n} + \frac{\overline{X}^2}{S_{xx}}\right)
\end{aligned}
$$

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

Intervalo de confianza para β_1

La siguiente cantidad pivotal con una distribución t

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{V\hat{a}r(\hat{\beta}_1)}} \sim t_{(n-2)}{}^{*}$$

De ahí se deriva lo siguiente:

$$\Rightarrow \ \text{P}\left(-t_0 \leq \frac{\hat{\beta}_1 - \beta_1}{\sqrt{V\hat{a}r(\hat{\beta}_1)}} \leq t_0\right) = 1-\alpha$$

$$\Rightarrow \ \text{P}\left(-t_0 \sqrt{V\hat{a}r\left(\hat{\beta}_1\right)} \leq \hat{\beta}_1 - \beta_1 \leq t_0 \sqrt{V\hat{a}r\left(\hat{\beta}_1\right)}\right) = 1-\alpha$$

$$\Rightarrow \ \text{P}\left(\hat{\beta}_1 - t_0 \sqrt{V\hat{a}r\left(\hat{\beta}_1\right)} \leq \beta_1 \leq \hat{\beta}_1 + t_0 \sqrt{V\hat{a}r\left(\hat{\beta}_1\right)}\right) = 1-\alpha$$

$\therefore$ El intervalo de confianza al (1-α) x 100% para β_1 es:

$$\left(\hat{\beta}_1 - t_0 \sqrt{V\hat{a}r\left(\hat{\beta}_1\right)} \ , \ \hat{\beta}_1 + t_0 \sqrt{V\hat{a}r\left(\hat{\beta}_1\right)} \right)$$

donde t_0 = cuantil de orden (1- $\frac{\alpha}{2}$) de una $t_{(n-2)}$.

[*] Véase Anexo 1

Intervalo de confianza para β_0

Para obtener el intervalo se realiza el mismo procedimiento anterior, obteniéndose el siguiente resultado[1]

$$\left(\hat{\beta}_0 - t_0 \sqrt{V\hat{a}r\left(\hat{\beta}_0\right)}\ ,\ \hat{\beta}_0 + t_0 \sqrt{V\hat{a}r\left(\hat{\beta}_0\right)} \right)$$

donde t_0 = cuantil de orden (1- $\frac{\alpha}{2}$) en $t_{(n-2)}$.

Intervalo de confianza para σ^2

Si los errores se distribuyen normal e independientemente entonces:

$$\frac{\sum (y_i - \hat{y}_i)^2}{\sigma^2} \sim \chi^2_{n-2}$$

Entonces:

$$P\left(\chi^2_{1-\frac{\alpha}{2},n-2} \le \frac{\sum (y_i - \hat{y}_i)^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2},n-2} \right) = 1-\alpha$$

Haciendo operaciones se tiene que intervalo de confianza al (1-α) x 100% para σ^2 es:

$$\left(\frac{\sum (y_i - \hat{y}_i)^2}{\chi^2_{\frac{\alpha}{2},n-2}} \le \sigma^2 \le \frac{\sum (y_i - \hat{y}_i)^2}{\chi^2_{1-\frac{\alpha}{2},n-2}} \right)$$

---

[1] $Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$, pero como σ^2 es desconocida se propone s^2 su estimador entonces se tiene lo siguiente $V\hat{a}r(\hat{\beta}_0) = s^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$; con $s^2 = \frac{\sum (y_i - \bar{y})^2}{n-2}$

1.3.1.4 Análisis de Varianza

El análisis de varianza es una prueba que permite medir la variación de las respuestas numéricas como valores de evaluación de diferentes variables nominales.

1.3.1.4.1 Precisión de los errores

Residuales $y_i - \hat{y}_i = e_i$

a) Suma de cuadrados total: SCT= $\sum_{i=1}^{n} (y_i - \bar{y})^2$

$$\sum (y_i - \bar{y})^2 = (y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + (y_3 - \bar{y})^2 + ... + (y_n - \bar{y})^2$$

Se sabe que $\sum_{i=1}^{n} \frac{y_i}{n} = \bar{y}$

$$\Rightarrow y_n = n\bar{y} - \sum_{i=1}^{n-1} y_i$$

En esta suma n-1 términos están libres y uno está restringido.

Por lo tanto esta suma tiene n-1 grados de libertad (g.l).

b) Suma de cuadrados del error o de los residuales: SCE $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \ldots\ldots\ldots\ldots + (y_n - \hat{y}_n)^2$$ donde $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

En esta suma n-2 términos están libres y 2 de ellos están restringidos.

La suma satisface dos restricciones, por lo tanto tiene n-2 g.l

c) Suma de cuadrados de la regresión: SCR= $\sum(\hat{y}_i - \bar{y})^2$

A continuación se demostrará que SCT= SCE + SCR

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

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

$$= \sum(y_i - \hat{y}_i)^2 + 2\left[\sum(y_i - \hat{y}_i)\hat{y}_i - \bar{y}\sum(y_i - \hat{y}_i)\right] + \sum(\hat{y}_i - \bar{y})^2$$

$$= \sum(y_i - \hat{y}_i)^2 + 2\left[\sum e_i y_i - \bar{y}\sum e_i\right] + \sum(\hat{y}_i - \bar{y})^2$$

Ahora se demostrará que $\sum e_i = 0$

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

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

$$= \sum y_i - \sum\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$$

$$= n\bar{y} - \sum\left(\bar{y} - \hat{\beta}_1\bar{x} + \hat{\beta}_1 x_i\right)$$

$$= n\bar{y} - n\bar{y} - \hat{\beta}_1\sum(x_i - \bar{x})$$

$$= -\hat{\beta}_1\left(\sum x_i - n\bar{x}\right)$$

$$= -\beta_1\left(n\bar{x} - n\bar{x}\right)$$

$$= 0$$

Por otro lado se demostrará que $\sum e_i \hat{y}_i = 0$

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

$$= \sum \left(y_i \hat{y}_i - \hat{y}_i^2\right)$$

$$= \sum \left(y_i\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right) - \sum \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)^2$$

$$= \sum \left(y_i\left(\bar{y} + \hat{\beta}_1(x_i - \bar{x})\right)\right) - \sum \left(\bar{y} + \hat{\beta}_1(x_i - \bar{x})\right)^2$$

$$= \bar{y}\sum y_i + \hat{\beta}_1 \sum y_i (x_i - \bar{x}) - \sum \bar{y}^2 - 2\sum \bar{y}\hat{\beta}_1 (x_i - \bar{x}) - \sum \hat{\beta}_1^2 (x_i - \bar{x})^2$$

$$= n\bar{y}^2 + \hat{\beta}_1 \sum y_i (x_i - \bar{x}) - n\bar{y}^2 - 2\hat{\beta}_1 \bar{y} \sum (x_i - \bar{x}) - \sum \hat{\beta}_1^2 (x_i - \bar{x})^2$$

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

$$= \hat{\beta}_1 S_{xx} \hat{\beta}_1 - \hat{\beta}_1^2 S_{xx}\ ^*$$

$$= \hat{\beta}_1^2 S_{xx} - \hat{\beta}_1^2 S_{xx}$$

$$= 0$$

Teniendo estos resultados se regresa a la siguiente ecuación

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + 2\left[\sum e_i y_i - \bar{y}\sum e_i\right] + \sum (\hat{y}_i - \bar{y})^2$$

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

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

---

* $\hat{\beta}_1 = \dfrac{\sum x_i y_i - n\bar{x}\bar{y}}{S_{xx}} = \dfrac{S_{xy}}{S_{xx}}$ entonces despejando se tiene lo siguiente $S_{xy} = \hat{\beta}_1 S_{xx}$

De esta forma se puede decir que

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

En cuanto a grados de libertad se refiere, se tiene:

n-1 = 1 + n-2

1.3.1.4.2 Tabla de Análisis de Varianza (ANOVA)

Hipótesis: $H_0:\beta_1=0$

A continuación se presenta la tabla de análisis de varianza.

Fuente de Variación	**Grados de libertad (g.l)**	**Suma de Cuadrados (SCE)**	**Suma de cuadrados medios (MSCE)**	**F**	**Valor p**[2]
Regresión	1	$\sum(\hat{y}_i-\bar{y})^2$	$\frac{\sum(\hat{y}_i-\bar{y})^2}{1}$	$\frac{MSCE_{regresión}}{MSCE_{residuales}}$	
Residuales	n-2	$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$	$\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}$		
Total	n-1	$\sum_{i=1}^{n}(y_i-\bar{y})^2$			

[2] Veáse Anexo II

1.3.1.5 Pruebas de Hipótesis

a) Pruebas de Hipótesis sobre la significancia de la regresión

Hipótesis: $H_0 : \beta_1 = 0$ *vs* $H_1 : \beta_1 \neq 0$

Estadístico de Prueba: $T = \dfrac{\sum(\hat{y}_i - \bar{y})^2}{\dfrac{\sum(y_i - \hat{y}_i)^2}{n-2}}$

Regla de decisión: Rechace H_0 al nivel α si $T > F_{(1,n-2)}^{1-\alpha}$

b) Prueba de Hipótesis para β_1

Hipótesis: $H_0 : \beta_1 = \beta_1^*$ *vs* $H_1 : \beta_1 \neq \beta_1^*$

Estadístico de Prueba: $T = \dfrac{\hat{\beta}_1 - \beta_1^*}{\sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}}}$

Regla de decisión: Rechace H_0 al nivel α si $|T| > t_{n-2}^{1-\frac{\alpha}{2}}$

c) Prueba de Hipótesis para β_0

Hipótesis: $H_0 : \beta_0 = \beta_0^*$ *vs* $H_1 : \beta_0 \neq \beta_0^*$

Estadístico de Prueba: $T = \dfrac{\hat{\beta}_0 - \beta_0^*}{\hat{\sigma}\sqrt{\dfrac{\sum x_i^2}{nS_{xx}}}}$

Regla de decisión: Rechace H_0 al nivel α si $|T| > t_{n-2}^{1-\frac{\alpha}{2}}$

1.3.1.6 Medidas de Asociación entre X y Y

a) Coeficiente de determinación: Indica el porcentaje de la variabilidad total que se explica a través del modelo lineal estimado.

$$
\begin{aligned}
R^2 &= \frac{SCT - SCE}{SCT} \\
&= \frac{SCR}{SCT} \\
&= 1 - \frac{SCE}{SCT} \\
&= \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}
\end{aligned}
$$

b) Coeficiente de correlación lineal: Mide el grado de asociación lineal entre las variables x y y.

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

$$= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \cdot \frac{\sum (x_i - \bar{x})^2}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

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

$$= \hat{\beta}_1 \left(\frac{S_{xx}}{S_{yy}} \right)^{\frac{1}{2}}$$

De otra forma se puede ver que $r_{xy}^2 = R^2$

$$r_{xy}^2 = \hat{\beta}_1^2 \cdot \frac{S_{xx}}{S_{yy}}$$

$$= \frac{S_{xy}^2}{S_{xx}^2} \cdot \frac{S_{ss}}{S_{yy}}$$

$$= \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

Por otro lado se sabe que

$$\sum (\hat{y}_i - \bar{y})^2 = \sum \left(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y} \right)^2$$

$$= \sum \left(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \bar{y} \right)^2$$

$$= \hat{\beta}_1^2 \sum (x_i - \bar{x})^2$$

$$= \hat{\beta}_1^2 S_{xx}$$

$$= \frac{S_{xy}^2}{S_{xx}^2} . S_{xx}$$

$$= \frac{S_{xy}^2}{S_{xx}}$$

Retomando el desarrollo de r_{xy}^2 se tiene lo siguiente

$$\frac{S_{xy}^2}{S_{xx} S_{yy}} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

$$= \frac{SCR}{SCT}$$

$$= R^2$$

$$\therefore r_{xy}^2 = R^2$$

1.3.1.7 Análisis de Residuales

Sea e_i el estimador de ε_i, donde ε_i es v.a

Se tienen las siguientes propiedades

1) $E(\varepsilon_i) = 0$
2) $Var(\varepsilon_i) = \sigma^2$ cte
3) $\varepsilon_i \perp \varepsilon_j \quad \forall \quad i \neq j$
4) $\varepsilon_i \sim N(0,1)$

Tipos de Residuales

1) $e_i = y_i - \hat{y}_i$ = residual

2) $d_i = \dfrac{e_i}{\sqrt{\dfrac{\sum\left(y_i - \hat{y}_i\right)^2}{n-2}}}$ = residual estandarizado

3) $r_i = \dfrac{e_i}{\tilde{\sigma}\sqrt{1 - \dfrac{1}{n} - \dfrac{\left(x_i - \bar{x}\right)^2}{s_{xx}}}}$ = residuales estudentizados

donde $\tilde{\sigma} = \sqrt{\dfrac{\sum\left(y_i - \hat{y}_i\right)^2}{n-2}}$

Con los diferentes tipos de residuales debe procederse a revisar si los supuestos impuestos al modelo de regresión lineal simple se cumplen. No se revisan en este caso, ya que el enfoque está en la regresión no paramétrica.

1.3.2 Regresión Lineal Múltiple

1.3.2.1 Modelo Lineal Múltiple

$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + ... + \beta_{p-1} x_{p-1i} + \varepsilon_i$

Donde x_{ji} son las regresoras $j = 1,2,......, p-1$ y

β_k son los coeficientes $k = 0,1,......., p-1$

$$n \text{ ecuaciones} \quad \begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + ... + \beta_{p-1} x_{p-11} + \varepsilon_1 \\ . \\ . \\ . \\ . \\ y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + ... + \beta_{p-1} x_{p-1n} + \varepsilon_n \end{cases}$$

Escribiendo en forma matricial.

$$\underline{Y}_{nx1} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{pmatrix} \quad , \quad \underline{\beta}_{px1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_{p-1} \end{pmatrix} \quad , \quad \underline{\varepsilon}_{nx1} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & . & . & . & . & x_{p-11} \\ 1 & x_{12} & x_{22} & . & . & . & . & x_{p-12} \\ . & . & . & & & & & . \\ . & . & . & & & & & . \\ . & . & . & & & & & . \\ . & . & . & & & & & . \\ 1 & x_{1n} & x_{2n} & . & . & . & . & x_{p-1n} \end{pmatrix}$$

Por otro lado sea $\underline{\varepsilon}$ vector aleatorio donde

$E(\varepsilon_i) = 0 \ \forall i$

$$E(\underline{\varepsilon}) = E\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ . \\ . \\ . \\ E(\varepsilon_n) \end{pmatrix} = \underline{0}$$

$E(\varepsilon_i, \varepsilon_j) = 0 \ \ \forall \ \ i \neq j$

$\varepsilon_i \perp \varepsilon_j \ \ \forall \ \ i \neq j$

$$\text{var}-\text{cov}(\underline{\varepsilon}) = \begin{pmatrix} \text{var}(\varepsilon_i) & \text{cov}(\varepsilon_1, \varepsilon_2) & \cdot & \cdot & \cdot & \cdot & \text{cov}(\varepsilon_1, \varepsilon_n) \\ \text{cov}(\varepsilon_2, \varepsilon_1) & \text{var}(\varepsilon_2) & & & & & \text{cov}(\varepsilon_2, \varepsilon_n) \\ . & . & & & & & . \\ . & . & & & & & . \\ . & . & & & & & . \\ \text{cov}(\varepsilon_n, \varepsilon_1) & \text{cov}(\varepsilon_n, \varepsilon_2) & \cdot & \cdot & \cdot & \cdot & \text{var}(\varepsilon_n) \end{pmatrix}$$

$$\text{var}-\text{cov}(\underline{\varepsilon}) = \sigma^2 \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & \cdot & 0 \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ 0 & 0 & \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

$$= \sigma^2 I_n$$

$\therefore \ \text{var}-\text{cov}(\underline{\varepsilon}) = \sigma^2 I_n$ ....(ec. 3)

Teniendo de esta forma

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

$$E(\underline{Y}) = E\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{pmatrix} = \begin{pmatrix} E(y_1) \\ E(y_2) \\ . \\ . \\ . \\ E(y_n) \end{pmatrix}$$

$$= E(X\underline{\beta} + \varepsilon)$$

$$= E(X\underline{\beta}) + E(\varepsilon)$$

$$= X\underline{\beta} + \underline{0}$$

$E(Y) = X\underline{\beta}$ ....(ec. 4)

Por lo cual se puede decir que

$\hat{\underline{Y}} = X\hat{\underline{\beta}}$ es el modelo ajustado................................. (ec. 5)

$e_i = y_i - \hat{y}_i$ es el residual

$\underline{e} = \underline{Y} - \hat{\underline{Y}}$ es el vector de residuales

$$\text{var}-\text{cov}(\underline{Y}) = \begin{pmatrix} \text{var}(y_i) & \text{cov}(y_1, y_2) & \cdot \;\cdot \;\cdot \;\cdot & \text{cov}(y_1, y_n) \\ \text{cov}(y_2, y_1) & \text{var}(y_2) & & \text{cov}(y_2, y_n) \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ \text{cov}(y_n, y_1) & \text{cov}(y_n, y_2) & \cdot \;\cdot \;\cdot \;\cdot & \text{var}(y_n) \end{pmatrix}$$

Sea $G = \underline{e}'_{1xn}\,\underline{e}_{\,nx1}$

$$G = (\underline{Y} - \hat{\underline{Y}})'(\underline{Y} - \hat{\underline{Y}})$$

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

$$= \left(\underline{Y}' - \hat{\underline{\beta}}'X'\right)(\underline{Y} - X\hat{\underline{\beta}})$$

$$= \underline{Y}'\underline{Y} - \underline{Y}'X\hat{\underline{\beta}} - \hat{\underline{\beta}}X'\underline{Y} + \hat{\underline{\beta}}'X'X\hat{\underline{\beta}}$$

$$= \underline{Y}'\underline{Y} - 2\underline{Y}'X\hat{\underline{\beta}} + \hat{\underline{\beta}}'X'X\hat{\underline{\beta}}$$

$$= \underline{Y}'\underline{Y} - 2\hat{\underline{\beta}}'X'\underline{Y} + \hat{\underline{\beta}}'X'X\hat{\underline{\beta}}$$

$\underline{Y}'\underline{Y} =$ Forma cuadrática sobre las y_i

$\underline{\hat{\beta}}' X' X \underline{\hat{\beta}} =$ Forma cuadrática de las betas

$\underline{\hat{\beta}}' X' \underline{Y} =$ Forma Lineal

1.3.2.2 Estimador por mínimos cuadrados

Se hace el supuesto p=2 entonces se tiene lo siguiente

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

De la misma forma con la que se obtuvieron los estimadores para la regresión lineal simple, se obtiene el siguiente sistema de ecuaciones:

$$n\beta_0 + \beta_1 \sum x_i = \sum y_i$$
$$\sum x_i \beta_0 + \beta_1 \sum x_i^2 = \sum y_i x_i$$

Si este sistema se representa de forma matricial se tiene lo siguiente

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_i \end{pmatrix}$$

Despejando β_0 y β_1

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_i \\ \sum y_i x_i \end{pmatrix}$$

Por otro lado se hacen los siguientes cálculos

$$X'X = \begin{pmatrix} 1 & \cdot & \cdot & \cdot & 1 \\ x_1 & \cdot & \cdot & \cdot & x_n \end{pmatrix} \begin{pmatrix} 1 & \cdot & \cdot & x_1 \\ . & & & . \\ . & & & . \\ 1 & \cdot & \cdot & x_n \end{pmatrix}$$

$$= \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

$$X'\underline{Y} = \begin{pmatrix} 1 & \cdot & \cdot & \cdot & 1 \\ x_1 & \cdot & \cdot & \cdot & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum y_i \\ \sum y_i x_i \end{pmatrix}$$

Teniendo estos resultados se puede decir que $\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = (X'X)^{-1} X'\underline{Y}$

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

Por la ec. 5 se tiene:

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

$$= (X'X)^{-1}(X'X)\underline{\beta} + (X'X)^{-1} X'\underline{\varepsilon}$$

$$= \underline{\beta} + (X'X)^{-1} X'\underline{\varepsilon}$$

$\therefore$ De otra forma $\underline{\hat{\beta}} = \underline{\beta} + \left(X^{\prime} X\right)^{-1} X^{\prime} \underline{\varepsilon}$ .... (ec. 6)

Revisando las propiedades de $\underline{\hat{\beta}}$

$$E\left(\underline{\hat{\beta}}\right) = E\left(\left(X^{\prime} X\right)^{-1} X^{\prime} \underline{Y}\right)$$

$$= \left(X^{\prime} X\right)^{-1} X^{\prime} E\left(\underline{Y}\right)$$

Por ec. (4)

$$E\left(\underline{\hat{\beta}}\right) = \left(X^{\prime} X\right)^{-1} X^{\prime} X \underline{\beta}$$

$$= \left(X^{\prime} X\right)^{-1} \left(X^{\prime} X\right) \underline{\beta}$$

$$= I \underline{\beta}$$

$$\therefore \; E\left(\underline{\hat{\beta}}\right) = \underline{\beta}$$

$$Var\left(\underline{\hat{\beta}}\right) = Var - \operatorname{cov}\left(\underline{\hat{\beta}}\right)$$

$$= E\left[\left(\underline{\hat{\beta}} - \underline{\beta}\right)\left(\underline{\hat{\beta}} - \underline{\beta}\right)^{\prime}\right]$$

$$= E\left[\left(\underline{\hat{\beta}} - \underline{\beta}\right)\left(\underline{\hat{\beta}}^{\prime} - \underline{\beta}^{\prime}\right)\right]$$

Utilizando ec. 6

$$Var\left(\underline{\hat{\beta}}\right) = E\left[\left(\underline{\beta} + \left(X^{\prime} X\right)^{-1} X^{\prime} \underline{\varepsilon} - \underline{\beta}\right)\left(\underline{\beta}^{\prime} + \underline{\varepsilon}^{\prime} X \left(X^{\prime} X\right)^{-1} - \underline{\beta}^{\prime}\right)\right]$$

$$= E\left[\left(\left(X^{\prime} X\right)^{-1} X^{\prime} \underline{\varepsilon}\underline{\varepsilon}^{\prime} X \left(X^{\prime} X\right)^{-1}\right)\right]$$

$$= \left(X^{\prime} X\right)^{-1} X^{\prime} E\left(\underline{\varepsilon}\underline{\varepsilon}^{\prime}\right) X \left(X^{\prime} X\right)^{-1\,*}$$

* $E\left(\underline{\varepsilon}\underline{\varepsilon}^{\prime}\right) = \mathrm{var} - \mathrm{cov}\left(\underline{\varepsilon}\right)$

Por la ec. 3

$$
\begin{aligned}
Var(\underline{\hat{\beta}}) &= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} \\
&= \sigma^2 (X'X)^{-1}(X'X)(X'X)^{-1} \\
&= \sigma^2 (X'X)^{-1}
\end{aligned}
$$

$$\therefore Var(\underline{\hat{\beta}}) = \sigma^2 (X'X)^{-1}$$

1.3.2.3 Pruebas de Hipótesis

Para una sola β.

a) *Hipótesis:* $H_0 : \beta_i = 0$ *vs* $H_1 : \beta_i \neq 0 \quad \forall \quad i = 1,........,n$

Estadístico de Prueba: $T = \dfrac{\hat{\beta}_i}{\tilde{\sigma}\sqrt{V_{ii}}} \sim t_{(n-1)}$

Regla de decisión: Rechace H_0 al nivel α si $T > t_{(n-1)}^{1-\frac{\alpha}{2}}$

Para $(\beta_1,......,\beta_p) = \underline{\beta}^*$

b) *Hipótesis:* $H_0 : \underline{\beta}^* = 0$ vs $H_1 : \underline{\beta}^* \neq 0$

El estadístico de prueba será la Tabla de ANOVA

1.3.2.4 Tabla de Análisis de Varianza (ANOVA)

A continuación se presenta la tabla de análisis de varianza.

Fuente de Variación	Grados de libertad (g.l)	Suma de Cuadrados (SCE)	Suma de cuadrados medios (MSCE)	F	Valor p
Regresión	$p = \#\beta's - 1$	$\sum(\hat{y}_i - \bar{y})^2 = \underline{\hat{\beta}}'X'\underline{Y} - \frac{1}{n}\underline{Y}'11'\underline{Y}$	$\frac{\sum(\hat{y}_i - \bar{y})^2}{p}$	$\frac{MSCE_{regresión}}{MSCE_{residuales}}$	
Residuales	$n - p - 1 = \#obs - \#\beta's$	$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \underline{Y}\underline{Y}' - \underline{\hat{\beta}}'X'\underline{Y}$	$\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n - p - 1}$		
Total	$n - 1$	$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \underline{Y}'\underline{Y} - \frac{1}{n}\underline{Y}'11'\underline{Y}$			

Regla de decisión: Rechace H_0 al nivel α si $F > F_{(1,n-p-1)}^{1-\frac{\alpha}{2}}$ o si el valor p es muy pequeño

1.3.2.5 R² ajustada

Mide el número de regresoras y si éstas son suficientes o no.

$$R_{adj}^2 = 1 - \frac{\dfrac{SCE_{res}}{g.l_{res}}}{\dfrac{SCE_{total}}{g.l_{total}}} = 1 - \frac{MSCE_{res}}{MSCE_{total}} = 1 - R^2\left(\frac{n-1}{n-p-1}\right)$$

Si R_{adj}^2 es grande $\Rightarrow$ las regresoras en el modelo explican en su conjunto.

Si R_{adj}^2 es pequeña $\Rightarrow$ las regresoras deben revisarse.

1.3.2.6 Análisis de Residuales

Residuales

1) $e_i = y_i - \hat{y}_i$ = residual

2) $d_i = \dfrac{e_i}{\sqrt{\dfrac{\sum (y_i - \hat{y}_i)^2}{n-p-1}}}$ = residual estandarizado

3) $r_i = \dfrac{e_i}{\sqrt{\tilde{\sigma}^2 (1-h_{ii})}}$ = residuales estudentizados

donde $1-h_{ii} =$ entrada ii de la matriz I-H*

---

* $H = X\left(X`X\right)^{-1} X`$

CAPÍTULO II. Regresión no Paramétrica

2.1 Introducción

Como se vio en el capítulo anterior la teoría clásica de la regresión se basa, en gran parte, en el supuesto de que las observaciones son independientes e idéntica y normalmente distribuidas. Si bien existen muchos fenómenos del mundo real que pueden modelarse de esta manera, para el tratamiento de ciertos problemas, la normalidad de los datos es insostenible. En el intento de eliminar esa restricción se diseñaron métodos que hacen un número mínimo de supuestos sobre los modelos que describen las observaciones.

La teoría de los métodos no paramétricos trata, esencialmente, del desarrollo de procedimientos de inferencia estadística, que no realizan una suposición explícita con respecto a la forma funcional de la distribución de probabilidad de las observaciones de la muestra. Si bien en la estadística no paramétrica también aparecen modelos y parámetros, estos están definidos de una manera más general que en su contrapartida paramétrica.

La regresión no paramétrica es una colección de técnicas para el ajuste de funciones de regresión cuando existe poco conocimiento a priori acerca de su forma. Proporciona funciones suavizadas de la relación y el procedimiento se denomina suavizado.

Los fundamentos de los métodos de suavizamiento son antiguos, pero sólo lograron el estado actual de desarrollo gracias a los avances de la computación y los estudios por simulación han permitido evaluar sus comportamientos.

La técnica más simple de suavizamiento es llamada, promedios móviles, fue la primera en usarse, sin embargo han surgido nuevas técnicas como la estimación mediante núcleos (“kernel”) o la regresión local ponderada.

Estos estimadores de regresión no paramétrica son herramientas poderosas para el análisis de datos, tanto como una técnica de estimación para resumir una relación compleja que no puede ser expresada por un modelo paramétrico, como para suplementar un análisis de regresión paramétrico.

2.2 Necesidad de usar la Regresión no Paramétrica

En los análisis paramétricos se comienza haciendo supuestos rígidos sobre la estructura básica de los datos, luego se estiman de la forma más eficiente posible los parámetros que definen la estructura y por último se comprueba si los supuestos iniciales se cumplen.

La regresión no paramétrica, en cambio, desarrolla un modelo libre para predecir la respuesta sobre el rango de valores de los datos. Básicamente está constituida por métodos que proporcionan una estimación suavizada de la relación para un conjunto de valores (denominado ventana) de la variable explicativa. Estos valores son ponderados de modo que, por ejemplo, los vecinos más cercanos tengan mayor peso que los más alejados dentro de una ventana de datos. Se pueden utilizar diversas funciones de ponderación, que son los pesos en que se basan los estimadores. La combinación de la función de ponderación y el ancho de la ventana inciden sobre la bondad de la estimación resultante.

En la mayoría de las aplicaciones sobre regresión no paramétrica se considera el caso de una sola regresora a pesar de que, a simple vista no pareciera de gran utilidad, ya que las aplicaciones más interesantes involucran varias variables explicativas.

La regresión no paramétrica es importante por dos motivos:

- En etapas preliminares del análisis de datos o en pruebas de diagnóstico se utilizan gráficos de dispersión en los cuales puede ser muy útil ajustar una curva suavizada. Por ejemplo, para explorar la forma de la función respuesta, para confirmar una función respuesta en particular que haya sido ajustada a los datos, para obtener estimaciones de la respuesta media sin especificar la forma de la función respuesta, para estudiar el cumplimiento de supuestos, etc.

- Forma la base a partir de la cual se extienden los conceptos para regresión no paramétrica múltiple.

2.3 Planteamiento Matemático

A continuación se mostrarán algunos procedimientos no paramétricos que son utilizados en el análisis de regresión lineal simple ($y_i = \beta_0 + \beta_1 x_i$).

2.3.1 Método Brown-Mood

Brown y Mood describen un método para determinar β_0 y β_1.
Por este método, primero se dividen los valores de y en dos grupos:

- Aquellos valores de x que son menores o iguales que la mediana de x.
- Aquellos valores de x que son mayores que la mediana de x.

Después los valores deseados de β_0 y β_1 son aquéllos que forman una línea en la cual la mediana de las desviaciones sobre la línea es cero en cada uno de los dos grupos.

Para encontrar β_0 y β_1 se siguen los siguientes pasos:

1 Hacer un diagrama de dispersión de la muestra de datos.

2 Dibujar una línea vertical a través de la mediana de los valores de x. Si uno o más puntos caen en esta línea, pasarlos a la izquierda o derecha como sea necesario de tal forma que el número de puntos en cualquiera de los dos lados de la mediana sean tan cercanos o iguales como sea posible.

3 Determinar la mediana de x y y en cada uno de los dos grupos formados en el paso 2. Esto significa calcular un total de cuatro medianas.

4 En el primer grupo de observaciones, trazar un punto que represente la intersección de la mediana de x con la mediana de y. Trazar un punto similar para el segundo grupo de observaciones.

5 Dibujar una línea conectando los dos puntos trazados en el paso 4. Esta línea es una primera aproximación de la línea deseada.

6 Si la mediana de las desviaciones verticales de los puntos de esta línea es distinta de cero en ambos grupos, cambiar la línea a una nueva posición hasta que las desviaciones en cada grupo tengan una mediana igual a cero. Esto puede ser completado de una mejor forma utilizando una regla transparente. Si se necesita una mayor precisión, el procedimiento de iteración descrito por Mood puede ser utilizado.

7 El valor de β_0 está dado por la intersección de y al final de la línea y $\beta_1 = \frac{y_1 - y_2}{x_1 - x_2}$ donde (x_1, y_1) y (x_2, y_2) son las coordenadas de dos puntos cualesquiera que se encuentren sobre la línea.

A continuación se presenta un ejemplo (E1) que ilustra este procedimiento.

Ejemplo E1

Clark examinó las características de filtración de la grasa de un envasado de poliéster y lana filtrada en las líneas arteriales durante una hemodilución clínica. Se obtuvieron como datos de los filtros de recuperación los sólidos de 10 pacientes quienes se sometieron a cirugía. La siguiente tabla muestra los porcentajes eliminados de lípidos y colesterol. Lo que se desea es ajustar una regresión lineal de los datos, donde el colesterol es la variable dependiente (y) y los lípidos se considera la variable independiente (x).

Datos recuperados de las líneas arteriales filtradas de 10 pacientes.*
Porcentajes eliminados, mg/kg/L*10^{-2}

Paciente	Lípidos (x)	Colesterol (y)
1	3.81	1.9
2	2.1	1.03
3	0.79	0.44
4	1.99	1.18
5	1.03	0.62
6	2.07	1.29
7	0.74	0.39
8	3.88	2.3
9	1.43	0.93
10	0.41	0.29

La mediana de los valores de x es 1.71, cuando se dividen las observaciones en dos grupos de cinco cada grupo. Por debajo de la mediana se tienen los puntos (0.79, 0.44), (1.03, 0.62), (0.74, 0.39), (1.43, 0.93), (0.41, 0.29). Por arriba de la mediana se tienen (3.81, 1.90), (2.10, 1.03), (1.99, 1.18), (2.07, 1.29), (3.88, 2.30).

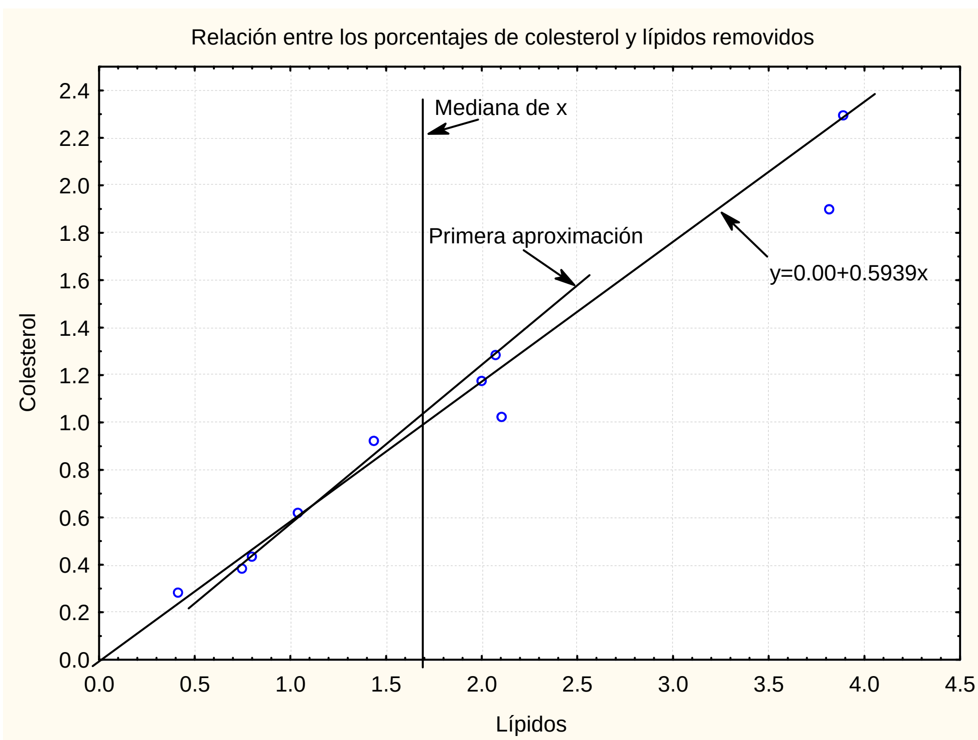
La mediana de los valores de x y y que se encuentran por debajo de la mediana es 0.79 y 0.44 respectivamente. Por arriba de la mediana de x,

* Fuente: Richard E. Clark, Harry W. Margrat y Richard A. Beauchamp. “Fat and Solid Filtration in Clinical Perfusions”. Surgery, 77(1975), 216-224.

las medianas son 2.10 para x y 1.29 para y. Estos datos producen una primera aproximación para calcular el parámetro β_1, donde $\beta_1^{`} = \dfrac{1.29 - 0.44}{2.10 - 0.79} = 0.6489$. Entonces como se puede ver las medianas de los subgrupos son distintas de cero, por lo que se ajusta la línea visualmente. Este procedimiento conduce al valor final de $\beta_1 = 0.5939$ y $\beta_0 = 0.00$.

De esta manera la ecuación de la línea es $y = 0.00 + 0.5939x$.

En esta figura se muestra el diagrama de dispersión de los datos originales para este ejemplo, la mediana de x, la línea de la primera aproximación y la línea de la regresión final.



La línea de regresión calculada de la muestra de datos en este ejemplo describe la relación entre el porcentaje de colesterol y el total de lípidos medidos en la muestra.

2.3.2 Método Theil

El método sugerido por Brown y Mood es rápido y puede generar suspicacia, por lo que quizá el más utilizado por muchos investigadores es el método propuesto por Theil para obtener un punto estimado del coeficiente β_1. Se supone que los datos se ajustan para el modelo clásico de regresión.

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

donde x_i son constantes conocidas, β_0 y β_1 son parámetros desconocidos y y_i es un valor observado de la variable aleatoria continua y sobre el valor x_i. Para cada valor de x_i, se asume una subpoblación de los valores de y, y las ε_i son mutuamente independientes.

Las $x_i^{\prime s}$ son distintas, y se toma $x_1 < x_2 < \ldots < x_n$.

Los datos consisten de n pares de observaciones $(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_n, y_n)$, donde el i-ésimo par representa las medidas tomadas en la i-ésima unidad de asociación.

Para obtener un estimador de β_1, primero se forman todas las posibles muestras $S_{ij} = \dfrac{(y_j - y_i)}{(x_j - x_i)}$, donde $i < j$. Habrá $N = \binom{n}{2}$ valores de S_{ij}. El estimador de β_1, el cual se designa por $\hat{\beta}_1$, es la mediana de los valores de S_{ij}; esto es: $\hat{\beta}_1 = mediana(S_{ij})$

A continuación se muestra un ejemplo (E2) para ilustrar este método.

Ejemplo E2

Yunginger y Gleich estudiaron los cambios secuenciales en las concentraciones del suero y la inmuglobulina nasal (IgE) en adultos sin alergias y en adultos con rinitis alérgica estacional. Los datos obtenidos se muestran en la siguiente tabla. Ésta contiene medidas de (IgE) y de la proteína total nasal (TP) tomadas en 10 pacientes con historial de fiebre típica del heno ambrosía. Debido a la naturaleza del tratamiento recibido en esos 10 pacientes, éste hizo que la dosis del tratamiento disminuyera en el grupo de estudio.

Determinaciones IgE y Nasal TP en 10 pacientes con fiebre típica del heno ambrosia*

(y) Nasal (TP)	(x) Nasal IgE/TP(*10^{-7})
206	9.7
453	408
141	106
131	7.5
203	49.3
172	5.8
153	6.5
356	2.8
297	16.8
425	7

Lo que se desea calcular es $\hat{\beta}_1$, siguiendo las fórmulas se calcula S_{ij} y ordenándolos se obtienen los siguientes resultados:

-588.0000000	-13.0612245	-0.67497404	0.69696125	11.3636364
-81.1111111	-5.24822695	-0.30938124	0.69865738	12.8169014
-61.3333333	-4.21428571	-0.12060302	0.71264368	13.9805825
-54.8648649	-3.29032258	-0.07575758	0.74719801	16.4285714
-47.8723404	-2.89230769	0.06982544	0.80399501	16.5625000
-27.1428571	-2.86868687	0.10152284	1.03311258	17.8494624
-24.1176471	-2.08333333	0.23938796	1.16822430	34.0909091
-22.0000000	-1.74887892	0.39877301	1.72248804	210.833333
-21.7391304	-1.09347443	0.62013558	8.71794872	544.000000

* Fuente: John W. Yunginger and Gerald J. Gleich, “Seasonal Changes in Serum and Nasal IgE Concentrations”, J. Allergy Clin. Immunol., 51 (1973), 174-186.

Se puede ver que existen $N=\binom{10}{2}=45$ valores de S_{ij}, y la mediana de estos valores es 0.070.

Entonces el estimador de β_1 es $\hat{\beta}_1=0.070$.

2.3.3 Pruebas de Hipótesis sobre β_0 y β_1

Algunos investigadores se han interesado en las pruebas de hipótesis sobre uno o ambos parámetros (β_0 y β_1). En esta sección se presentará un método para hacer pruebas simultáneas sobre la hipótesis nula de $\beta_0=\beta_0^*$ y $\beta_1=\beta_1^*$, y dos métodos de pruebas de hipótesis nula de $\beta_1=\beta_1^*$.

2.3.3.1 Método de Brown-Mood

Supuesto:

Los datos se conforman por "n" pares de observaciones (x_1,y_1), (x_2,y_2), ..., (x_i,y_i), ..., (x_n,y_n), donde cada par de observaciones (x_i,y_i) es una medida sobre la misma unidad de asociación.

Hipótesis: $H_0:\beta_0=\beta_0^*,\beta_1=\beta_1^* \quad vs \quad H_1:\beta_0\neq\beta_0^* \;\; y/o \;\; \beta_1\neq\beta_1^*$

Estadístico de Prueba

Para calcular el estadístico de prueba, se procede de la siguiente manera:

1. Hacer un diagrama de dispersión con los datos.
2. Dibujar la línea $y = \beta_0^* + \beta_1^* x$ sobre el diagrama de dispersión.
3. Dibujar sobre el diagrama de dispersión una línea vertical que atraviese la mediana de los valores de x.
4. Sea n_1 el número de puntos que se encuentran por arriba de la línea de regresión y a la izquierda de la línea vertical que atraviesa la mediana de los valores de x. Además sea n_2 el número de puntos por arriba de la línea de regresión y a la derecha de la línea vertical que atraviesa la mediana de los valores de x.

Mood estableció que los puntos que no pertenezcan a n_1 y n_2 tienen una distribución binomial con parámetro 0.5.

Por lo tanto el *estadístico de prueba* =

$$X^2 = \frac{8}{n}\left[\left(n_1 - \frac{n}{4}\right)^2 + \left(n_2 - \frac{n}{4}\right)^2\right]$$

el cual se distribuye aproximadamente como una ji cuadrada con dos grados de libertad cuando H_0 es verdadera y n no es muy pequeña. Tate y Clelland establecieron que la aproximación tiende a ser buena para la práctica cuando n es mayor o igual que 10.

Regla de decisión: Rechace H_0 al nivel α si $X^2 > \chi^2_{(\alpha,2)}$ *

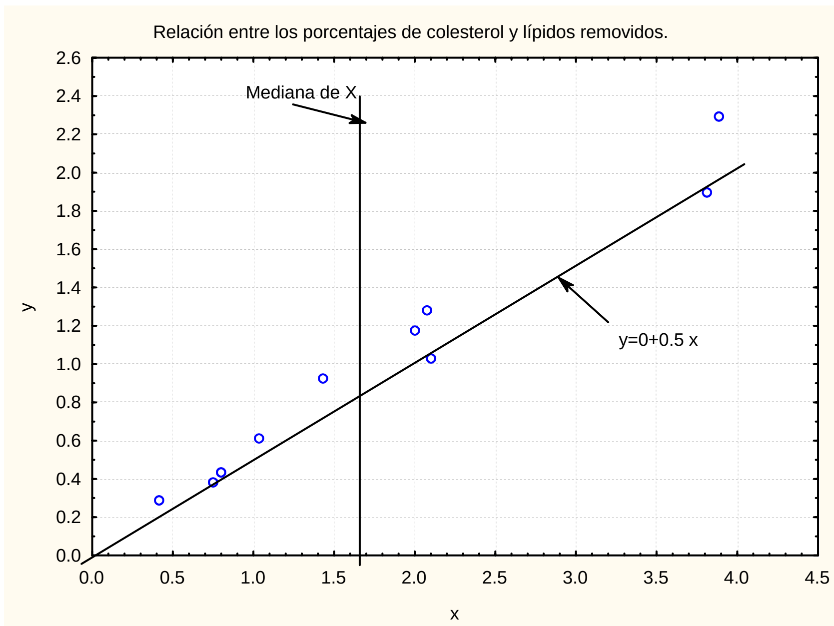
A continuación, se ilustrará este procedimiento, para ello se usará el ejemplo (E1).

Hipótesis: $H_0 : \beta_0 = 0, \beta_1 = 0.5 \quad vs \quad H_1 : \beta_0 \neq 0 \,, \beta_1 \neq 0.5$

* $\chi^2_{(\alpha,2)}$ = ji cuadrada con dos grados de libertad y un nivel de significancia α

Estadístico de Prueba

En la siguiente figura se muestra el diagrama de dispersión, la recta de regresión y la línea que atraviesa la mediana de los valores de x.



Como se puede observar $n_1 = 5$ y $n_2 = 3$, con esto se calcula

$$X^2 = \frac{8}{10}\left[\left(5 - \frac{10}{4}\right)^2 + \left(3 - \frac{10}{4}\right)^2\right] = 5.2$$

Observando en tablas se puede decir que cuando H_0 es verdadero, la probabilidad de obtener un valor de X^2 similar o mayor que 5.2 es mayor que 0.05. Por lo que no se rechaza H_0 y se concluye que la muestra puede provenir de una población en la cual la regresión lineal tiene una pendiente de 0.5 y una intersección de cero.

En el análisis de regresión normalmente hay más interés por β_1, la pendiente la recta de regresión. Cuando se desea hacer una prueba de hipótesis sobre β_1 (es decir $H_0 : \beta_1 = \beta_1^*$ vs $H_1 : \beta_1 \neq \beta_1^*$) se utiliza el siguiente procedimiento descrito por Brown y Mood.

1. Hacer un diagrama de dispersión con los datos.
2. Dibujar sobre el diagrama de dispersión una línea vertical que atraviese la mediana de los valores de x.
3. Ajustar la línea $y = \beta_0 + \beta_1 x$ a los datos, donde β_0 es la mediana de las desviaciones $y_i - \beta_1^* x$ para todos los valores observados de y y β_1^* es el valor hipotético de β_1. Usualmente se puede determinar esta línea de una manera más fácil trazando la línea $y = \beta_1^* x$ y dibujando una línea paralela a ésta misma, la cual divide los puntos en dos grupos iguales.
4. Contar el número de puntos n_1 que se encuentran por arriba de la línea $y = \beta_0 + \beta_1^* x$ y la izquierda de la mediana de los valores de x.

En este caso el estadístico de prueba es:

$$X_b^2 = \frac{16}{n}\left(n_1 - \frac{n}{4}\right)^2$$

Si H_0 es verdadera, el estadístico de prueba se distribuye aproximadamente como una ji-cuadrada con un grado de libertad, para n muy grande. Tate y Clelland recomendaron usar la ji-cuadrada aproximadamente para $n \geq 20$.

En otras palabras la regla de decisión es:

Rechace H_0 al nivel α si: $X_b^2 > \chi^2_{(\alpha,1)}$

A continuación se presenta un ejemplo (E3) que ilustra el procedimiento para una prueba de hipótesis sobre β_1.

Ejemplo E3

Pilkey y Hower examinaron los cambios en la concentración de magnesio y estronio en pruebas de una reciente colección de especies *echinoid* provenientes de un ambiente variable sobre muchos de esos rangos geográficos. Ellos usaron técnicas de rayos x para analizar el magnesio y estroncio en especies de "*Dendraster excéntricas*", la comunidad de la costa del Pacífico, coleccionó de 24 localidades entre la isla de Vancouver, Columbia Británica y la bahía de Santa Rosalía, Baja California, y calcularon el porcentaje de calcio. En la siguiente tabla se muestra la temperatura media en el verano (x) sobre las 24 localidades y el porcentaje medio de $MgCO_3$ contenido (y) en la colección de especímenes. Lo que se desea es probar la hipótesis nula sobre $\beta_1 = 0$ en la regresión lineal para la población representada en la siguiente muestra*.

Localidad número	1	2	3	4	5	6	7	8
Temperatura media en el verano ºC (x)	23.0	18.7	17.5	21.0	20.0	19.0	15.3	14.0
Porcentaje Medio de $MgCO_3$ (y)	9.5	9.0	9.2	9.2	9.4	9.3	9.0	8.5
Localidad número	9	10	11	12	13	14	15	16
Temperatura media en el verano ºC (x)	14.0	13.7	13.3	13.6	13.1	13.0	13.6	14.2
Porcentaje Medio de $MgCO_3$ (y)	9.0	8.4	8.8	8.9	8.5	8.7	8.6	8.7
Localidad número	17	18	19	20	21	22	23	24
Temperatura media en el verano ºC (x)	13.9	14.8	14.2	13.0	16.1	15.9	13.0	11.7
Porcentaje Medio de $MgCO_3$ (y)	8.5	9.1	9.1	8.0	8.1	8.5	8.4	8.7

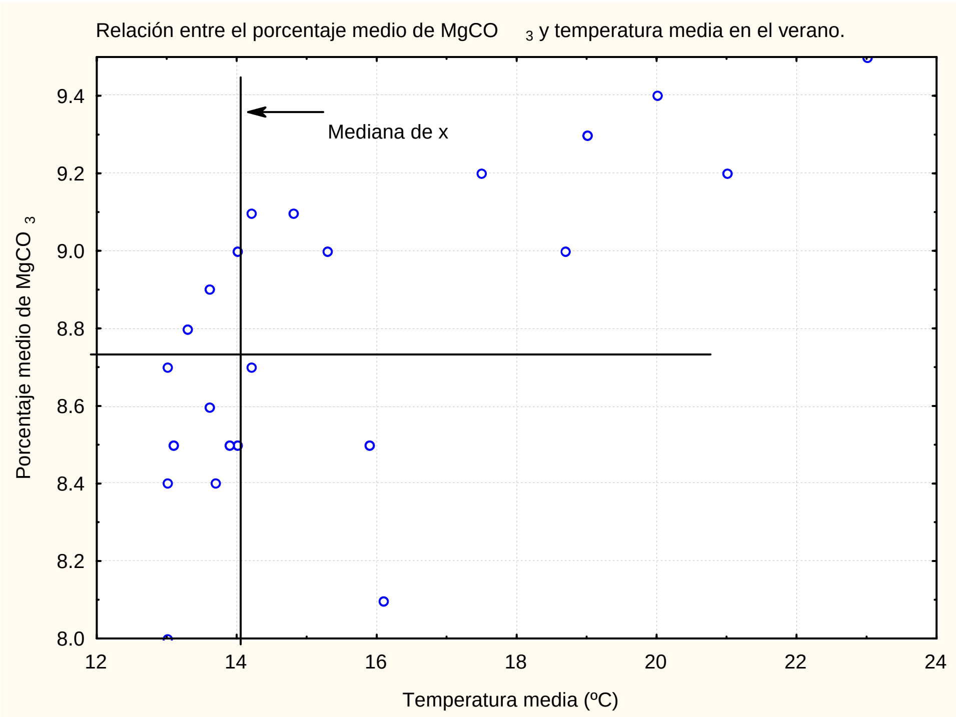
* Fuente: Orrin H. Pikey and John Hower, "The Effect of Enviroment on the Concentrations ok Skeletal Magnesium and Strontium in Dendraster", J. Geol, 68(1960), 203-214. Copyright 1960, the University of Chicago.

Al hacer cálculos se obtiene la mediana de los valores de x = 14.1

Por otro lado si $\beta_1^* = 0$, la línea $y = \beta_1^* x$ es el eje x.

En la siguiente figura se muestra el diagrama de dispersión, la mediana de los valores de x, y se puede observar la línea paralela al eje x en la cual se dividen los puntos en dos grupos iguales.

Observando la figura se puede ver que hay n_1 = 3 puntos por arriba la línea y a la izquierda de la mediana de los valores de x.



Entonces con estos datos se procede a calcular el estadístico de prueba:

$$X_b^2 = \frac{16}{24}\left(3 - \frac{24}{4}\right)^2 = 6$$

Observando en tablas se aprecia que la probabilidad de obtener un valor de X_b^2 tan grande como 6, y de tal forma que se pueda aceptar H_0 es menor que 0.025.

2.3.3.2 Método de Theil

Otro método para probar $H_0: \beta_1 = \beta_1^*$ ha sido propuesto por Theil. El procedimiento está basado en el estadístico Tau Kendall.

Supuestos:

A. El modelo apropiado es $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1,...,n$

donde las x_i son constantes conocidas y β_0 y β_1 son parámetros desconocidos.

B. Para cada valor de x_i existe una subpoblación de valores de y.

C. y_i es un valor observado de una variable aleatoria continua y sobre el valor x_i.

D. Las x_i son todas distintas, y se toma $x_1 < x_2 < ... < x_n$.

E. Las e_i son mutuamente independientes y provienen de la misma población continua.

Los datos obtenidos para el análisis consisten de n pares de observaciones, $(x_1, y_1),(x_2, y_2),....(x_n, y_n)$, donde el i-ésimo par representa las medidas tomadas sobre la i-ésima unidad de asociación.

Hipótesis

A. (Dos colas) $H_0: \beta_1 = \beta_1^* \quad vs \quad H_1: \beta_1 \neq \beta_1^*$

B. (Una cola) $H_0: \beta_1 \leq \beta_1^* \quad vs \quad H_1: \beta_1 > \beta_1^*$

C. (Una cola) $H_0: \beta_1 \geq \beta_1^* \quad vs \quad H_1: \beta_1 < \beta_1^*$

Estadístico de Prueba

Como se mencionó, el procedimiento descrito aquí está basado en el estadístico Tau Kendall. Específicamente para calcular éste, se comparan todos los posibles pares de observaciones de la forma $(x_i, y_i - \beta_1^* x_i)$. Se puede resumir el procedimiento como sigue:

1. Ordenar los pares de observaciones $(x_i, y_i - \beta_1^* x_i)$ en una columna en orden natural respecto a los valores de x.
2. Comparar cada $y_i - \beta_1^* x_i$ con cada $y_j - \beta_1^* x_j$ apareciendo por debajo de ese.
3. Sea P el número de comparaciones semejantes que resultan en un par $(y_i - \beta_1^* x_i, y_j - \beta_1^* x_j)$ que está en orden natural, y sea Q el número de comparaciones semejantes que resultan en un par que está en sentido contrario del orden natural.
4. Sea $S = P - Q$. El estadístico de prueba, entonces se tiene que:

$$\hat{\tau} = \frac{S}{n(n-1)/2}$$

Regla de Decisión

Las reglas de decisión para las tres situaciones de hipótesis son como sigue:

A. (Dos colas): Haciendo referencia a tablas (A.1). Rechazar H_0 al nivel de significancia α si el valor calculado de $\hat{\tau}$ es positivo y mayor que $\hat{\tau}^*$ con n y $\frac{\alpha}{2}$, o negativo y menor que el negativo de $\hat{\tau}^*$ con n y $\frac{\alpha}{2}$.

B. (Una cola): Haciendo referencia a tablas (A.1). Rechazar H_0 al nivel de significancia α si el valor calculado de $\hat{\tau}$ es positivo y mayor que $\hat{\tau}^*$ con n y α.

C. (Una cola): Haciendo referencia a tablas (A.1). Rechazar H_0 al nivel de significancia α si el valor calculado de $\hat{\tau}$ negativo y menor que el negativo de $\hat{\tau}^*$ con n y α.

A continuación se presenta un ejemplo (E4).

Ejemplo E4

Gruñid y Metzger describieron la metodología involucrada en un procedimiento psicológico para estimar la secreción hepática de los líquidos biliares de un hombre. Los autores reportaron los datos en el horario de producción de ácidos biliares y fosfolípidos biliares que se muestran en la siguiente tabla. Lo que se desea es probar la hipótesis nula sobre $\beta_1 = 0$ en la ecuación de regresión de la población describiendo la relación entre dos variables.

	Producción de acidos biliares y fosfolípidos biliares en 10 personas*				
Ácidos biliares (mg/hr) x	940	594	1200	1440	1112
Fosfolípidos biliares (mg/hr) y	311	391	414	542	387
Ácidos biliares (mg/hr) x	625	1385	1035	931	742
Fosfolípidos biliares (mg/hr) y	485	502	458	345	346

* Fuente: Scout M. Gruñid and Allan L. Metzger. “A Physiological Method for Estimation of Hepatic Secretion of Biliary Lipids in Man”, Gastroenterology, 62 (1972), 1200-1217, William & Wilkins, Baltimore.

Hipótesis: $H_0 : \beta_1 = \beta_1^* \;\; vs \quad H_1 : \beta_1 \neq \beta_1^*$

Estadístico de Prueba

El primer paso para calcular $\hat{\tau}$ es calcular $y_i - \beta_1^* x_i$ para cada par de observaciones. Con $\beta_1^* = 0$, $y_i - \beta_1^* x_i$ es igual a y_i para cada par. El resto de los pasos descritos anteriormente para calcular $\hat{\tau}$ son resumidos en la siguiente tabla.

x_i	$y_i - \beta_1^* x_i$	Número de pares mayores que $y_i - \beta_1^* x_i$ en orden natural	número de pares menores que $y_i - \beta_1^* x_i$ en orden natural
594	391	5	4
625	485	2	6
742	346	5	2
931	345	5	1
940	311	5	0
1035	458	2	2
1112	387	3	0
1200	414	2	0
1385	502	1	0
1440	542	0	0
		P=30	Q=15

Con los datos de esta tabla, se calcula $S = 30 - 15 = 15$. Finalmente se tiene que:

$$\hat{\tau} = \frac{15}{10(9)/2} = 0.33$$

Regla de decisión

Observando en tablas se puede ver que cuando H_0 es verdadero y $n = 10$, la probabilidad de observar un valor de $\hat{t}$ tan grande como 0.33 es aproximadamente 0.20. Esto no es una evidencia por la cual se tenga que rechazar H_0. Por lo que se concluye que la pendiente de la regresión de la población debería de ser cero.

2.3.4 Intervalo de confianza para β_1

Mood sugirió una técnica de prueba y error para obtener un intervalo de confianza para β_1 basados en la prueba de hipótesis de Brown- Mood. Esta técnica está basada en el hecho de que el $(1-\alpha)\,\mathrm{x}\,100\%$ del intervalo de confianza consiste de los valores de β_1^* que no fueron rechazados al nivel de significancia α. Mood también sugirió un método alternativo aproximado basado en la prueba de hipótesis de Brown- Mood para β_1.

El método para construir un intervalo de confianza para β_1, está basado en el procedimiento de la prueba de hipótesis de β_1. Los supuestos fundamentales para el procedimiento de prueba de hipótesis también aplican para la construcción de los intervalos. Este método se construye con los siguientes pasos:

1. De los n pares (x_i, y_i), calcular todas las posibles muestras de las pendientes $S_{ij} = \frac{(y_j - y_i)}{(x_j - x_i)}$, donde $i < j$. En total habrá $N = \binom{n}{2}$ valores de S_{ij}.
2. Ordenar los valores de S_{ij} en orden de magnitud, del más pequeño al más grande. Los límites inferior y superior del intervalo de confianza para β consiste de dos valores, $\hat{\beta}_I$ y $\hat{\beta}_U$
3. Ingresar los valores dados en la tabla A.1[1] con n y $\frac{\alpha}{2}$ para encontrar $S_{\frac{\alpha}{2}}$.[2]
4. Restar 2 de $S_{\frac{\alpha}{2}}$, para obtener $C_{\frac{\alpha}{2}}$, esto significa $C_{\frac{\alpha}{2}} = S_{\frac{\alpha}{2}} - 2$.
5. Sea $k = \frac{N - C_{\frac{\alpha}{2}}}{2}$
6. El límite inferior $\hat{\beta}_L$ del intervalo de confianza para β_1 es el k-ésimo valor de S_{ij}, contando a partir del valor más pequeño en el orden presentado en el paso 2. El límite superior $\hat{\beta}_U$ es el k-ésimo valor de S_{ij}, contando hacia atrás a partir del valor más grande. En otras palabras se puede escribir el intervalo de confianza obtenido para β_1 con una confiabilidad de $1-\alpha$ como:

$$\beta \;\; \varepsilon \;\; (\hat{\beta}_L, \hat{\beta}_U) \text{ con una confiabilidad de } 1-\alpha$$

---

[1] Critical values for use with the Kendall tau statistic

[2] Para los tamaños de muestras que no se encuentren en tablas, se puede calcular $C_{\frac{\alpha}{2}}$ de la siguiente forma:

$$C_{\frac{\alpha}{2}} \approx z_{\frac{\alpha}{2}} \sqrt{\frac{n(n-1)(2n+5)}{18}}$$

Donde $z_{\frac{\alpha}{2}}$ es el valor de z proveniente de la tabla de la normal que tiene $\frac{\alpha}{2}$ del área debajo de la curva.

Ejemplo: Para ilustrar la construcción de un intervalo de confianza para β_1, se hará referencia al ejemplo 2 (E2). En este ejemplo lo que se desea es construir un intervalo de confianza para β_1 con una confiabilidad de 0.95. Entonces $\frac{\alpha}{2} = 0.025$

1. Los datos originales se encuentran en la tabla del ejemplo (E2).
2. De igual forma las muestras de las pendientes ordenadas se encuentran en la tabla del ejemplo (E2).
3. Observando los valores en tablas (Tabla A.1) con $n = 10$ y $\frac{\alpha}{2} = 0.025$, se obtiene $S_{\frac{\alpha}{2}} = 23$.
4. Restando 2 de $S_{\frac{\alpha}{2}}$ se obtiene 23-2=21
5. Calculando $k = \frac{(45 - 21)}{2} = 12$
6. Entonces $\hat{\beta}_L$ doceavo valor de S_{ij}, contando a partir del valor más pequeño en el orden presentado en el paso 2. Entonces $\hat{\beta}_L = -4.214$ y el límite superior $\hat{\beta}_U = 1.168$. Por lo que el intervalo de confianza queda como sigue:

β ε (-4.214,1.168) con una confiabilidad del 95%

2.3.5 Prueba de paralelismo de dos regresiones lineales

Los investigadores frecuentemente hacen análisis de regresión sobre dos escenarios de datos que son similares. Una pregunta muy común en estas situaciones es si las dos regresiones tienen la misma pendiente. Por ejemplo un investigador médico desearía saber si la pendiente de la regresión lineal que describe la relación lineal entre la presión de la sangre y edad es igual en hombres como en mujeres. Un psicólogo quisiera saber si la pendiente de la regresión lineal que describe la relación entre la intensidad de estímulos y la velocidad de respuesta es la misma en dos diferentes grupos de edad.

Como este tipo de pruebas indican si existe una igualdad entre las pendientes de las regresiones lineales de las poblaciones, éstas son referidas como pruebas de homogeneidad de pendientes o pruebas de paralelismo. En la estadística paramétrica las pruebas son basadas en las distribuciones t y F. Cuando los supuestos de las distribuciones no son conocidos, se necesita algún procedimiento alternativo. Una alternativa es la prueba no paramétrica de paralelismo sugerida por Hollander.

Supuestos

a) Los datos provienen de dos grupos de observaciones. Para cada uno, el modelo es:

$$y_j = \beta_0 + \beta_1 x_j + \varepsilon_j$$

En esta ecuación las x_j´s son constantes conocidas, β_0 y β_1 son parámetros desconocidos y y_j es el valor de la variable aleatoria y observada sobre los puntos x_j.

b) Los ε_j son mutuamente independientes. Por cada línea, los ε_j provienen de una misma población continua, pero las dos poblaciones no tienen que ser la misma.

Hipótesis

Se tienen dos grupos de datos, el grupo 1 y el grupo 2. El modelo que se tiene para el grupo 1 es:

$$y_{1j} = \beta_0^* + \beta_1^* x_{1j} + \varepsilon_{1j} \qquad j = 1,......,n_1, \quad x_{11} \leq x_{12} \leq ... \leq x_{1n}$$

Y el modelo para el grupo 2 es:

$$y_{2j} = \beta_0^{**} + \beta_1^{**} x_{2j} + \varepsilon_{2j} \qquad j = 1,......,n_2, \quad x_{21} \leq x_{22} \leq ... \leq x_{2n}$$

Se pueden probar las siguientes hipótesis nulas (H_0) contra las alternativas H_1.

A. (Dos colas) $H_0 : \beta_1^* = \beta_1^{**} \quad vs \quad H_1 : \beta_1^* \neq \beta_1^{**}$

B. (Una cola) $H_0 : \beta_1^* \geq \beta_1^{**} \quad vs \quad H_1 : \beta_1^* < \beta_1^{**}$

C. (Una cola) $H_0 : \beta_1^* \leq \beta_1^{**} \quad vs \quad H_1 : \beta_1^* > \beta_1^{**}$

Estadístico de Prueba

Para calcular el estadístico de prueba, se procede de la siguiente manera:

1. La prueba requiere que el número de pares de observaciones en los dos grupos sean iguales. Consecuentemente el primer paso es descartar las observaciones aleatorias de uno o ambos grupos como sea necesario para satisfacer los requerimientos del tamaño de la muestra.
2. Colocar el número resultante de observaciones en cada grupo igual a $2n$.
3. De los datos del grupo 1, formar n pares $\left(x_{1j}, x_{1,j+n}\right)$, con $j = 1,2,...,n$.
4. Calcular las n pendientes estimadas del grupo 1. Teniendo cada pendiente la siguiente forma:

$$u_{1j} = \frac{y_{1,t+n} - y_{1,t}}{x_{1,t+n} - x_{1t}} \quad t = 1,2,....,n$$

5. De los datos del grupo 2, formar n pares como se describe en el paso 3.
6. Calcular las n pendientes estimadas del grupo 2. Teniendo cada pendiente la siguiente forma:

$$u_{2t} = \frac{y_{2,t+n} - y_{2,t}}{x_{2,t+n} - x_{2t}} \quad t = 1,2,....,n$$

7. Organizar aleatoriamente las u_{1j}\`$s$ con las $u_{2t}$\`s de tal forma que cada u aparezca en uno y sólo un par.
8. Calcular las n diferencias de la siguiente forma:

$$Z_t = u_{1j} - u_{2t}$$

Nombrando a estas n diferencias como $Z_1, Z_2,...., Z_n$.

El resto de los pasos para calcular el estadístico de prueba son los siguientes:

9. Clasificar los valores absolutos de las Z de la más pequeña a la más grande, numerándolas del 1 a n.

10. Asignar a cada clasificación de los resultados el signo de la diferencia que se haya producido al calcular cada Z.

11. Calcular

$T_+ =$ La suma de los rangos que tienen signo positivo.

$T_- =$ La suma de los rangos que tienen signo negativo.

El estadístico de prueba es T_+ o T_-, dependiendo de la hipótesis alternativa.

Regla de Decisión

Las reglas de decisión para los tres escenarios de las posibles hipótesis son como sigue:

a) (Dos colas): Sea $T = T_+$ ó T_-, la que sea más pequeña. Rechazar H_0 al nivel de significancia α si la T calculada es más pequeña o igual que la T calculada en la tabla A.2[1] con n y $\frac{\alpha}{2}$.

b) (Una cola): Cuando $\beta_1^* < \beta_1^{**}$, los estimadores calculados de los datos de la muestra 1 tienden a ser más pequeños que los datos calculados de la muestra 2. Esta situación arroja más $Z\acute{}s$ negativas, lo cual conduce a un valor grande de T_-. Cuando T_- es grande, T_+ es más pequeña. Los valores pequeños de T_+ tienden a soportar la hipótesis alternativa. Los valores suficientes de T_+ causan que se rechace H_0. Por lo tanto rechazar H_0 al nivel de significancia α si

[1] Probability levels for the Wilcoxon signed-rank test

los valores calculados de T_+ son menores o iguales que la T calculada en la tabla A.2 con n y α.

c) (Una cola): Cuando $\beta_1^* > \beta_1^{**}$, los estimadores calculados de los datos de la muestra 1 tienden a ser más grandes que los datos calculados de la muestra 2. Esta situación arroja más Z positivas, lo cual conduce a un valor grande de T_+. Cuando T_+ es grande, T_- es más pequeña. Los valores pequeños de T_- tienden a soportar la hipótesis alternativa. Los valores suficientes de T_- causan que se rechace H_0. Por lo tanto rechazar H_0 al nivel de significancia α si los valores calculados de T_- son menores o iguales que la T calculada en la tabla A.2 con n y α.

A continuación se muestra un ejemplo (E5):

Ejemplo E5

Cuando los primates eran usados para hacer investigaciones, en algunas ocasiones era necesario conocer la edad del animal. Reed coleccionó medidas del cráneo, hocico, hueso largo, y dientes, desarrollados en babuinos *Papio cynocephalus*. El utilizó estos datos, coleccionados a lo largo de cinco años, para desarrollar una regresión entre esas medidas y la edad. Reed concluyó con base a los resultados que el desarrollo y el crecimiento del húmero produjo una herramienta precisa para determinar la edad del babuino. Parte de los datos recabados por Reed se muestran en la siguiente tabla.

Edad, en meses, y la suma de medidas de cráneo, hocico y hueso largo, en milímetros, para babuinos Papio cynocephalus hembras y machos*

Macho	Suma(x)	Edad (y)	Hembra	Suma (x)	Edad (y)
1	175.0	1.36	1	175.0	1.58
2	183.0	2.20	2	183.0	2.48
3	190.0	3.05	3	190.0	3.40
4	200.0	4.45	4	200.0	4.92
5	211.0	6.19	5	211.0	6.87
6	220.0	7.78	6	220.0	8.66
7	230.0	9.70	7	230.0	10.86
8	239.5	11.66	8	239.5	13.14
9	245.5	12.96	9	245.5	14.67
10	260.0	16.33	10	260.0	18.68
11	271.5	19.21	11	271.5	22.14
12	284.0	22.52	12	284.0	26.18
13	291.0	24.46	13	291.0	28.57
14	302.5	27.78	14	302.5	32.68
15	314.0	31.25	15	314.0	37.03
16	318.5	32.65	16	318.5	38.80
17	327.0	35.36	17	327.0	42.22
18	337.0	38.65			
19	345.5	41.52			
20	360.0	46.61			
21	375.0	52.10			
22	384.5	55.69			
23	397.0	60.55			
24	411.0	66.18			
25	419.5	69.68			
26	428.5	73.47			
27	440.0	78.41			
28	454.5	84.81			

Los datos provienen de colonias africanas existentes en la Fundación Sureste para la Investigación y Educación, San Antonio, Texas. Lo que se desea es concluir que las pendientes resultantes tanto para machos como para hembras son diferentes.

* Fuente: O.M. Reed, “Papio Cynocephalus Age Determinations”, Amer. J. Phys. Anthropol.,38(1973), 309-314.

Hipótesis: $H_0 : \beta_M = \beta_F$ vs $H_1 : \beta_M \neq \beta_F$

Estadístico de Prueba:

Para obtener un número equitativo de pares de observaciones en cada grupo, aleatoriamente se descartan 1 par de observaciones del grupo de las hembras y 12 pares de observaciones del grupo de los machos. Después se ordenan los 16 grupos de observaciones restantes para cada grupo en orden de magnitud de los valores de x. Quedando la siguiente tabla:

Edad, en meses, y la suma de medidas de cráneo, hocico y hueso largo, en milímetros, para babuinos Papio cynocephalus hembras y machos

Macho	Suma(x)	Edad (y)	Hembra	Suma (x)	Edad (y)
1	200	4.45	1	175	1.58
2	211	6.19	2	183	2.48
3	220	7.78	3	190	3.4
4	230	9.7	4	200	4.92
5	239.5	11.66	5	211	6.87
6	260	16.33	6	220	8.66
7	291	24.46	7	239.5	13.14
8	314	31.25	8	245.5	14.67
9	327	35.36	9	260	18.68
10	345.5	41.52	10	271.5	22.14
11	360	46.61	11	284	26.18
12	384.5	55.69	12	291	28.57
13	397	60.55	13	302.5	32.68
14	411	66.18	14	314	37.03
15	428.5	73.47	15	318.5	38.8
16	440	78.41	16	327	42.22

Ahora se tiene que el número de observaciones es 16 entonces $2n = 16$.
Con $n = 8$, se hacen pares de la siguiente manera: $x_{1,1}$ con $x_{1,9}$; $x_{1,2}$ con $x_{1,10}$; hasta llegar a $x_{1,8}$ con $x_{1,16}$.

Ahora se forman ocho pendientes estimadas que pertenecen al grupo 1.

$$u_{11} = \frac{y_{1,9} - y_{1,1}}{x_{1,9} - x_{1,1}} = \frac{35.36 - 4.45}{327 - 200} = 0.243$$

$$u_{12} = \frac{y_{1,10} - y_{1,2}}{x_{1,10} - x_{1,2}} = \frac{41.52 - 6.19}{345.5 - 211.0} = 0.263$$

$$u_{13} = \frac{y_{1,11} - y_{1,3}}{x_{1,11} - x_{1,3}} = \frac{46.61 - 7.78}{360.0 - 220.0} = 0.277$$

$$u_{14} = \frac{y_{1,12} - y_{1,4}}{x_{1,12} - x_{1,4}} = \frac{55.69 - 9.70}{384.5 - 230.0} = 0.298$$

$$u_{15} = \frac{y_{1,13} - y_{1,5}}{x_{1,13} - x_{1,5}} = \frac{60.55 - 11.66}{397.0 - 239.5} = 0.310$$

$$u_{16} = \frac{y_{1,14} - y_{1,6}}{x_{1,14} - x_{1,6}} = \frac{66.18 - 16.33}{411.0 - 260.0} = 0.330$$

$$u_{17} = \frac{y_{1,15} - y_{1,7}}{x_{1,15} - x_{1,7}} = \frac{73.47 - 24.46}{428.5 - 291.0} = 0.356$$

$$u_{18} = \frac{y_{1,16} - y_{1,8}}{x_{1,16} - x_{1,8}} = \frac{78.41 - 31.25}{440.0 - 314.0} = 0.374$$

Para los datos del grupo 2 (hembras), se hacen pares de la siguiente manera: $x_{2,1}$ con $x_{2,9}$, $x_{2,2}$ con $x_{2,10}$ hasta llegar a $x_{2,8}$ con $x_{2,16}$.

Ahora para el grupo 2 se forman ocho pendientes estimadas.

$$u_{21} = \frac{y_{2,9} - y_{2,1}}{x_{2,9} - x_{2,1}} = \frac{18.68 - 1.58}{260.0 - 175.0} = 0.201$$

$$u_{22} = \frac{y_{2,10} - y_{2,2}}{x_{2,10} - x_{2,2}} = \frac{22.14 - 2.48}{271.5 - 183.0} = 0.222$$

$$u_{23}=\frac{y_{2,11}-y_{2,3}}{x_{2,11}-x_{2,3}}=\frac{26.18-3.40}{284.0-190.0}=0.242$$

$$u_{24}=\frac{y_{2,12}-y_{2,4}}{x_{2,12}-x_{2,4}}=\frac{28.57-4.92}{291.0-200.0}=0.260$$

$$u_{25}=\frac{y_{2,13}-y_{2,5}}{x_{2,13}-x_{2,5}}=\frac{32.68-6.87}{302.5-211.0}=0.282$$

$$u_{26}=\frac{y_{2,14}-y_{2,6}}{x_{2,14}-x_{2,6}}=\frac{37.03-8.66}{314.0-220.0}=0.302$$

$$u_{27}=\frac{y_{2,15}-y_{2,7}}{x_{2,15}-x_{2,7}}=\frac{38.80-13.14}{318.5-239.5}=0.325$$

$$u_{28}=\frac{y_{2,16}-y_{2,8}}{x_{2,16}-x_{2,8}}=\frac{42.22-14.67}{327.0-245.5}=0.338$$

Aleatoriamente se forman pares con $u_{1j}´s$ y $u_{2t}´s$, quedando los pares de la siguiente forma:

$$\begin{array}{cccc} u_{14},u_{21} & u_{11},u_{24} & u_{16},u_{27} & u_{15},u_{23} \\ u_{18},u_{25} & u_{17},u_{22} & u_{12},u_{28} & u_{13},u_{26} \end{array}$$

Las diferencias entre los miembros de cada par son las siguientes:

$$\begin{array}{ll} Z_1=0.298-0.201=0.097 & Z_5=0.374-0.282=0.092 \\ Z_2=0.243-0.260=-0.017 & Z_6=0.356-0.222=0.134 \\ Z_3=0.330-0.325=0.005 & Z_7=0.263-0.338=-0.075 \\ Z_4=0.310-0.242=0.068 & Z_8=0.277-0.302=-0.025 \end{array}$$

Los pasos 9 y 10 son resumidos en la siguiente tabla:

Z_i	$\lvert Z_i \rvert$	Rango de $\lvert Z_i \rvert$	Rangos con signo
0.097	0.097	7	7
-0.017	0.017	2	-2
0.005	0.005	1	1
0.068	0.068	4	4
0.092	0.092	6	6
0.134	0.134	8	8
-0.075	0.075	5	-5
-0.025	0.025	3	-3

Con las entradas de la última columna de la tabla, se calcula T_+ y T_- teniendo como resultado $T_+ = 26$ y $T_- = 10$.

Regla de Decisión

Como $T_- = 10$ es la más pequeña, entonces se tomará $T = 10$, cuando se consulta en tablas (A.2) con $n = 8$, se encuentra que, cuando H_0 es verdadero, la probabilidad de que el valor de T sea tan grande como 10 es 2(0.1563)=0.3162 (probabilidad dos colas). Por lo tanto se concluye que los datos no muestran una evidencia a favor de la hipótesis alternativa en donde las dos regresiones de las poblaciones tienen diferentes pendientes.

2.3.6 Estimador para la diferencia entre los parámetros de la pendiente

Hollander y Wolfe han propuesto un procedimiento para estimar la diferencia entre las pendientes de dos regresiones lineales. Estos parámetros se pueden designar como $\Theta = \beta_1^* - \beta_1^{**}$ donde β_1^* y β_1^{**} son las pendientes de las dos regresiones lineales. El procedimiento está basado en la prueba de Hollander para el paralelismo, los supuestos y el modelo para la prueba son válidos. Para obtener el estimador se hace lo siguiente:

1 Obtener lo valores de Z como se describió en el paso 8 de la prueba anterior.

2 Calcular los $\frac{n(n+1)}{2}$ valores de $\frac{(Z_i + Z_j)}{2}$, donde $i \leq j$, y j toma los valores entre 1 y n.

3 Calcular el estimador de $\hat{\Theta} = mediana\left(\frac{Z_i + Z_j}{2}\right)$, $i \leq j$.

Para ilustrar el cálculo de $\hat{\Theta}$, se considerara el ejemplo (E5). Lo que se desea es estimar la diferencia entre β_M y β_F.

1 Los valores de Z obtenidos son los siguientes: 0.097, -0.017, 0.005, 0.068, 0.092, 0.134, -0.075 y -0.025.

2 Los $\frac{n(n+1)}{2} = 36$ valores de $\frac{(Z_i + Z_j)}{2}$, donde $i \leq j$, obtenidos son los siguientes:

-0.0750	-0.0170	0.0110	0.0365	0.0585	0.0945
-0.0500	-0.0100	0.0215	0.0375	0.0680	0.0970
-0.0460	-0.0060	0.0255	0.0400	0.0695	0.1010
-0.0350	-0.0035	0.0295	0.0485	0.0800	0.1130
-0.0250	0.0050	0.0335	0.0510	0.0825	0.1155
-0.0210	0.0085	0.0360	0.0545	0.0920	0.1340

3 El estimador de la diferencia entre las pendientes es la mediana de esos 36 valores. Esto es:

$$\hat{\Theta} = \frac{0.0360 + 0.0365}{2} = 0.03625$$

2.3.7 Intervalo de confianza para la diferencia entre los parámetros de la pendiente

Hollander y Wolfe describieron un procedimiento para construir un intervalo de confianza al $100(1-\alpha)\%$ para $\Theta = \beta_1^{*} - \beta_1^{**}$.

El modelo propuesto por Hollander consiste de los siguientes pasos:

1 Obtener los valores de Z como se describió en el paso 8 de la prueba anterior.

2 Calcular los $\frac{n(n+1)}{2}$ valores de $\frac{Z_i + Z_j}{2}$, donde $i \leq j$, y j toma los valores entre 1 y n. Ordenar los valores del más pequeño al más grande.

3 Localizar en la tabla (A.2), el tamaño de la muestra y el valor apropiado de P, así como determinar el nivel de confianza. Cuando se decide $(1-\alpha)$, entonces se calcula $P = \frac{\alpha}{2}$. Cuando el valor exacto de $\frac{\alpha}{2}$ no se encuentra en la tabla, se elige el valor que más se aproxime.

4 Los puntos finales del intervalo de confianza son los k-ésimos valores más grandes y más pequeños de $\frac{Z_i + Z_j}{2}$. $K = T + 1$, donde T es el valor en la columna marcada correspondiente al valor de P seleccionado en el paso 3.

Para ilustrar la construcción de un intervalo de confianza para $\beta_1^* - \beta_1^{**}$, en este caso se hará referencia de nueva cuenta la ejemplo (E5). Se construirá un intervalo de confianza del 95 % para $\beta_M - \beta_F$.

1 Los valores de Z obtenidos son los siguientes: 0.097, -0.017, 0.005, 0.068, 0.092, 0.134, -0.075 y -0.025.

2 Los valores ordenados se encuentran en el ejemplo anterior en el paso 2.

3 El nivel de confianza es $0.95 = (1 - 0.05)$, por lo que $\alpha = 0.05$ y entonces $\frac{\alpha}{2} = 0.025$. Revisando en la tabla (A.2), para n=8, el valor de P más cercano a 0.025 es 0.0273. Entonces se obtienes T=4 y K=4+1=5.

4 Entonces el límite inferior de el intervalo de confianza es el 5° número más pequeño de $\frac{(Z_i + Z_j)}{2}$, y el límite superior es el 5° número más grande. Entonces los valores obtenidos son: -0.0250 y 0.0970. Entonces $100[1 - 2(0.0273)] = 94.5\%$ es el nivel de confianza que se tiene con el siguiente intervalo $P(-0.0250 \leq \beta_M - \beta_F \leq 0.0970) = 0.945$.

CAPITULO III. Aplicación del Modelo de Regresión no Paramétrica

En este capítulo se analizarán y se llevarán a cabo los métodos descritos en el capítulo 2, cabe destacar que la base de datos fue tomada de la tesis "Regresión Dinámica y la Relación entre precios de crudos ligeros y pesados" presentada por la Act. Leticia Eugenia Corona Callejas en el año 2003.

La base de datos que se tomó fue la de Precios de los crudos Maya y WTI, desde enero de 1980 hasta marzo del 2002.

Se muestra un ejemplo de la forma en que se presentan los datos

Precios de los crudos Maya y WTI

Mes	Precios Corrientes		Precios Constantes (dll.2001)		CPI
	WTI (x)	Maya (y)	WTI (x)	Maya (y)	
Ene-80	32.50	26.59	73.97	60.52	0.43938
Feb-80	37.00	28.00	83.04	62.84	0.44559
Mar-80	38.00	28.00	84.00	61.90	0.45237
Abr-80	39.50	27.36	86.35	59.82	0.45745
May-80	39.50	28.00	85.50	60.61	0.46197
Jun-80	39.50	28.00	84.57	59.95	0.46706
Jul-80	39.50	29.00	84.57	62.09	0.46706
Ago-80	38.00	29.00	80.77	61.64	0.47044
Sep-80	36.00	29.00	75.89	61.13	0.47440
Oct-80	36.00	28.99	75.17	60.54	0.47892
Nov-80	36.00	29.00	74.55	60.06	0.48287
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
Dic-01	19.33	13.67	19.37	13.70	0.99793
Ene-02	19.67	14.06	19.67	14.08	1.00019
Feb-02	20.74	15.06	20.65	15.02	1.00414
Mar-02	24.24	18.50	24.01	18.32	1.00979

Donde:

WTI(x): West Texas Intermediate (WTI), también conocido como Texas Sweet Light, es un petróleo que contiene el promedio de características del petróleo extraído en campos occidentales de Texas (USA). Debido a su poco contenido de azufre, es catalogado como petróleo dulce y en relación a su densidad, el petróleo WTI es catalogado como ligero. Esto hace del WTI un petróleo de alta calidad. El precio del petróleo WTI es utilizado como referencia principalmente en el mercado norteamericano (Nueva York).

Maya: Se caracteriza por su alta viscosidad y contenido de azufre, metales y asfaltenos, además de un bajo rendimiento de fracciones ligeras en la destilación.

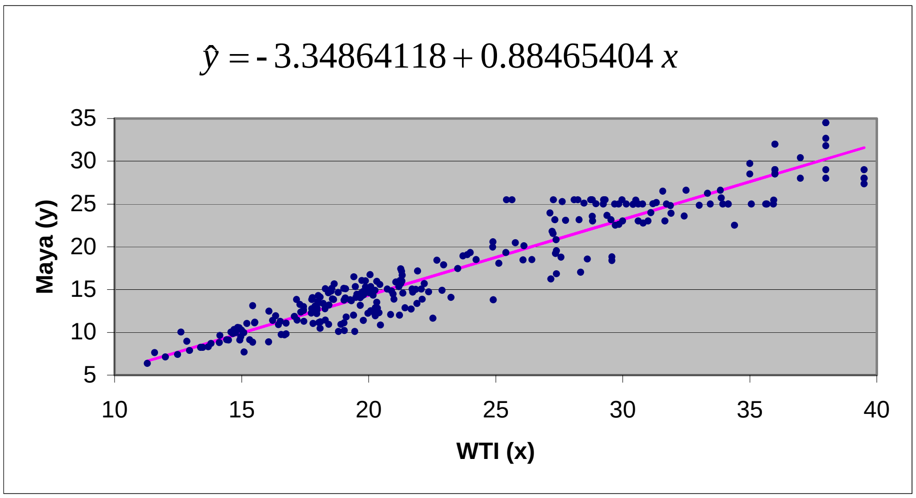
En México se producen crudos de cuatro diferentes calidades: Crudo Istmo, crudo Olmeca, Pesado de Altamira y crudo Maya, donde este último es considerado como pesado.

Antes de comenzar a evaluar los métodos descritos, se realizará el método de regresión paramétrica con los datos antes mencionados. Con el objetivo de comparar y analizar que método es el que da una mejor aproximación.

Los datos con lo que se evaluarán los métodos serán los correspondientes a los precios corrientes, se utilizan los precios constantes en la prueba de paralelismo para dos regresiones lineales.

3.1 Aplicación del método de Análisis de Regresión Paramétrica

Se evaluaron los 267 datos, con los métodos presentados en el capítulo 1, aunque cabe destacar que hay muchos programas que pueden arrojar resultados como los siguientes:



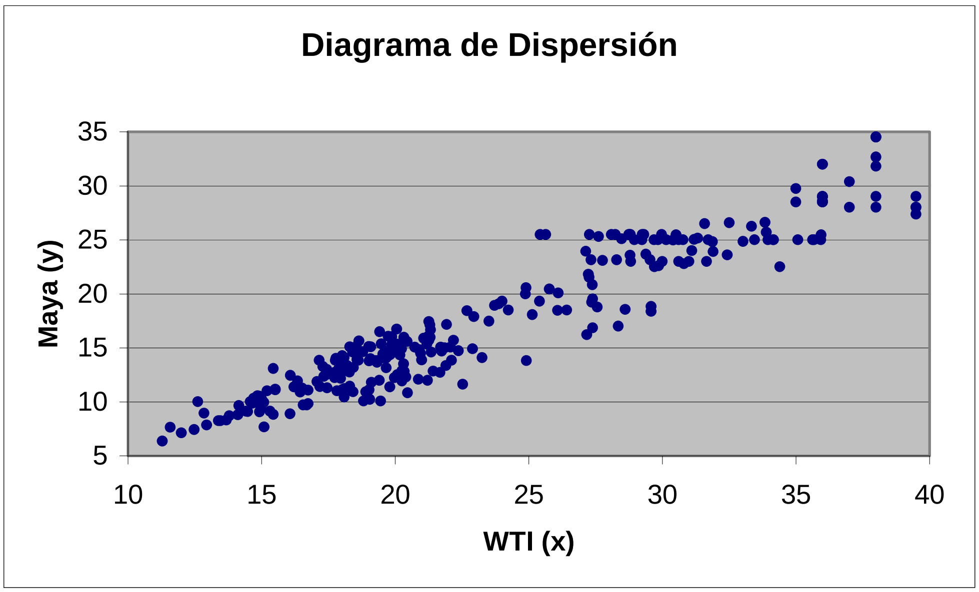
La ecuación de regresión resultante es: $\hat{y} = -3.34864118 + 0.88465404\,x$ con $r_{xy} = 0.95364705$, que indica que hay una buena asociación entre los datos; sin embargo es más importante tomar en cuenta $R^2 = 0.9094427$, que indica que aproximadamente el 90.94 % de la variabilidad total está representado por esta ecuación de regresión y $R^2{}_{adj} = 0.90910098$, que en este caso se toma como un número grande y por ende las regresoras están bien.

Y finalmente el intervalo de confianza al 95 % para β_1 es (0.85088951,0.91841856).

3.2 Aplicación del Método Brown-Mood

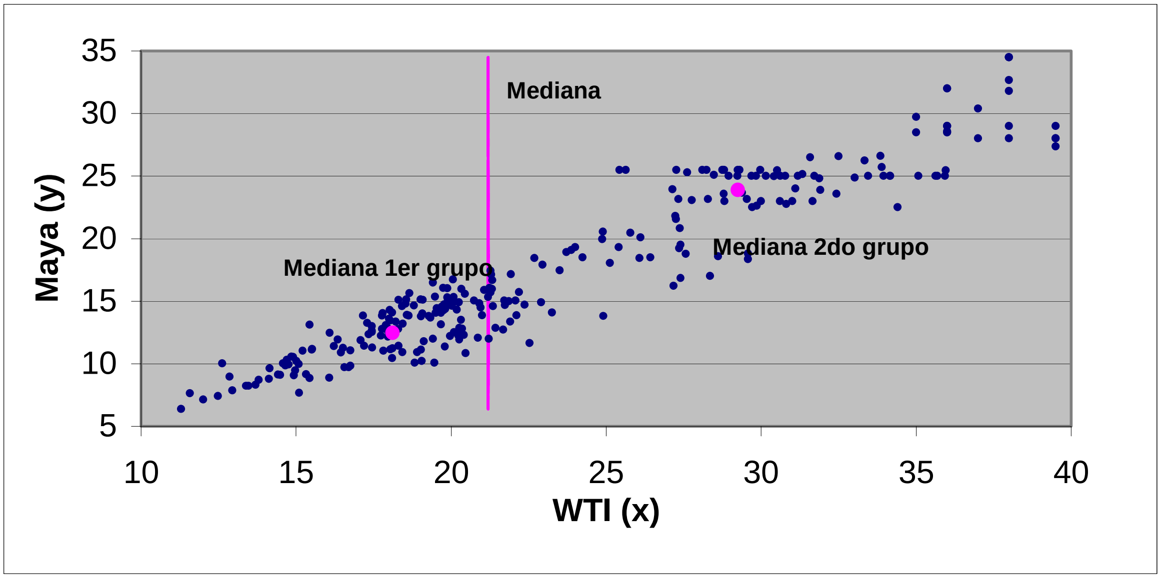
Para encontrar β_0 y β_1 se llevan a cabo los siguientes pasos:

1 Hacer un diagrama de dispersión de la muestra de datos.

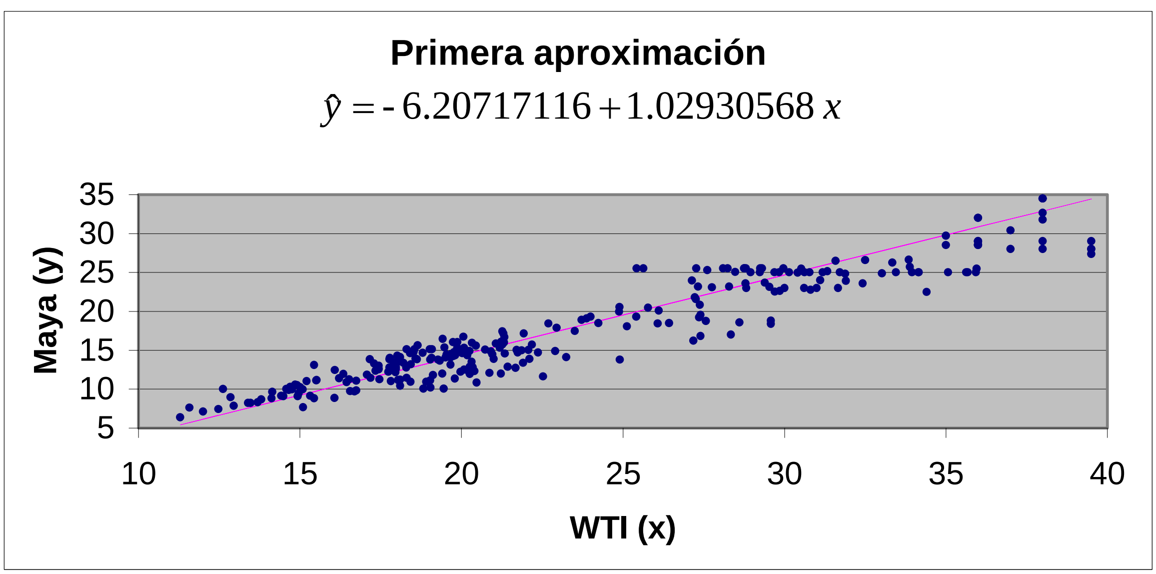


Hay que dibujar una línea vertical a través de la mediana de los valores de x igual a 21.19.

Enseguida hay que determinar la mediana de x y y en cada uno de los dos grupos, esto es: para el primer grupo las medianas son: 18.16 y 12.49, para el segundo grupo las medianas son 29.25 y 23.29 respectivamente.

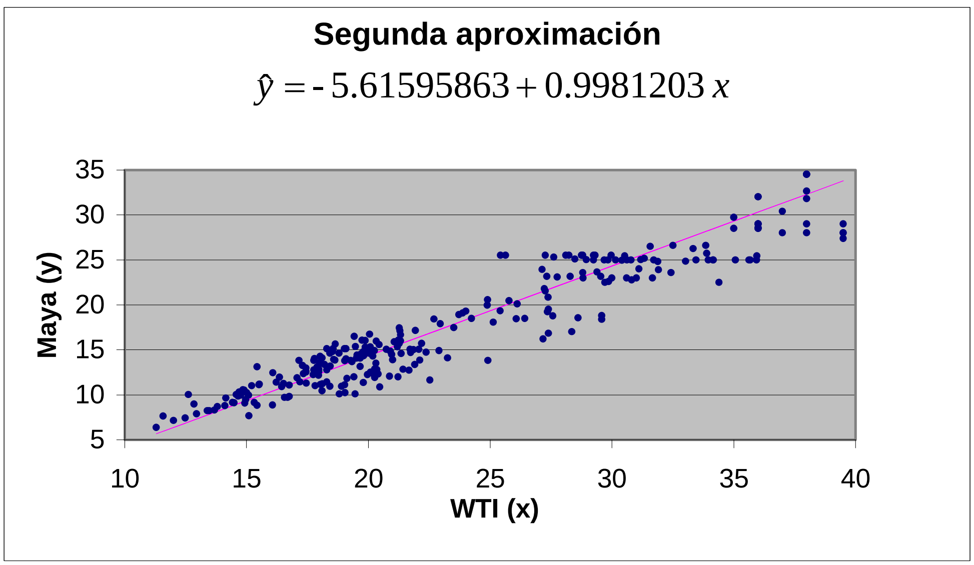


2 Dibujar una línea conectando los dos puntos trazados en el paso anterior. Esta línea es una primera aproximación de la línea deseada.



3 Si la mediana de las desviaciones verticales de los puntos de esta línea es distinta de cero en ambos grupos, cambiar la línea a una nueva posición hasta que las desviaciones en cada grupo tengan una mediana igual a cero.

4 El valor de β_0 está dado por la intersección de y al final de la línea y $\beta_1 = \frac{y_1 - y_2}{x_1 - x_2}$ donde (x_1, y_1) y (x_2, y_2) son las coordenadas de dos puntos cualesquiera que se encuentren sobre la línea.



Por lo tanto utilizando este método la mejor recta que se pudo obtener es $\hat{y} = -5.61595863 + 0.9981203x$.

3.3 Aplicación del Método Theil

Los datos consisten de 267 pares de observaciones $(x_1, y_1),(x_2, y_2),....,(x_i, y_i),.....,(x_n, y_n)$.

Cabe destacar que este procedimiento es bastante cansado y requiere de mucha paciencia.

Para obtener un estimador de β_1, primero se forman todas las posibles muestras $S_{ij} = \frac{(y_j - y_i)}{(x_j - x_i)}$, donde $i < j$. En total habrá $N = 35511$ valores de S_{ij}.

El estimador de β_1, el cual se designa por $\hat{\beta}_1$, es la mediana de los valores de S_{ij}; esto es,

$$\hat{\beta}_1 = mediana(S_{ij}) = 0.88562092$$

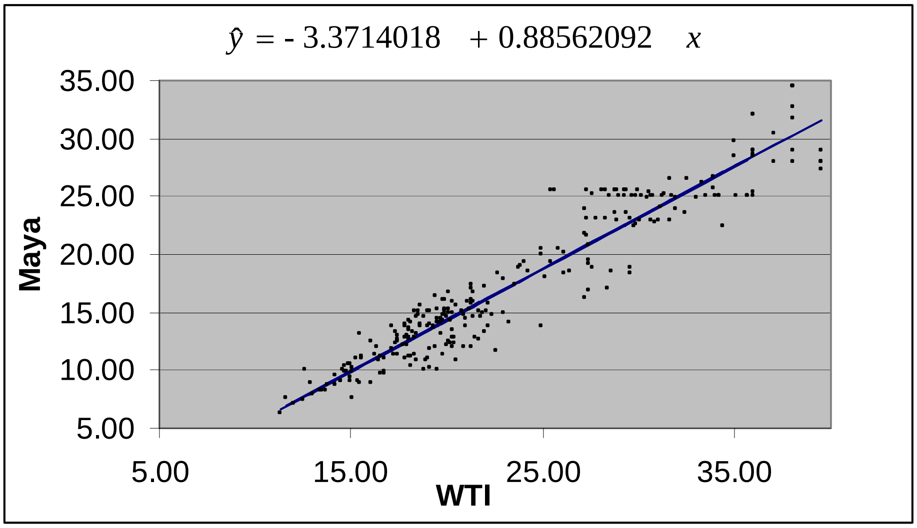
El valor de β_0 se obtuvo evaluando todos los datos que se tenían en un principio, es decir, se formaron un total de 267 posibles β_0, para obtener solamente una, la cuál se encontró promediando éstas, dando como resultado:

$$\beta_0 = -3.371401802$$

Para finalmente obtener la siguiente ecuación:

$$\hat{y}_i = -3.371401802 + 0.88562092 x_i$$

A continuación, se muestra la gráfica con la recta de regresión generada.



3.4 Aplicación de Pruebas de Hipótesis sobre β_0 y β_1

En esta sección se harán las pruebas de hipótesis por el método Brown-Mood y por el método Theil.

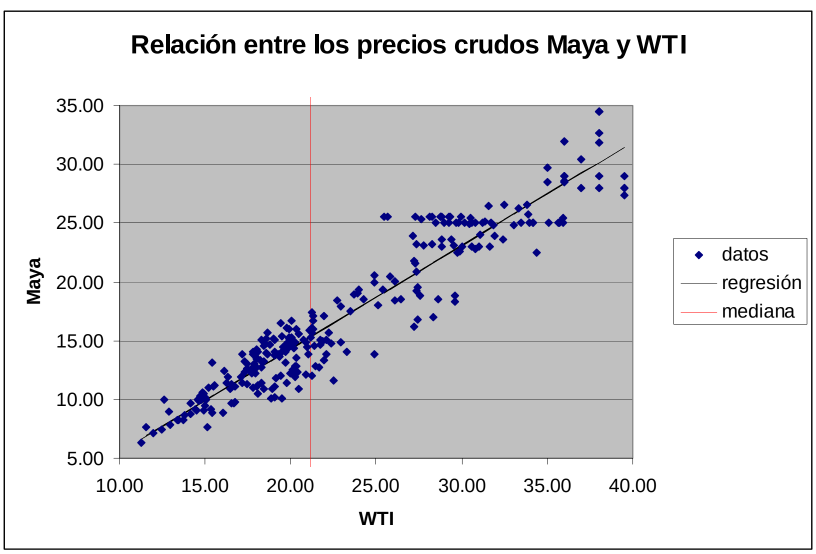
3.4.1 Método Brown-Mood

Hipótesis: (Se toman estos valores porque son similares a los obtenidos con el método paramétrico)

$$H_0 : \beta_0 = -3.3, \beta_1 = 0.88 \quad vs \quad H_1 : \beta_0 \neq -3.3 \, , \beta_1 \neq 0.88$$

Estadístico de Prueba

En la siguiente figura se muestra el diagrama de dispersión, la recta de regresión y la línea que atraviesa la mediana de los valores de x.



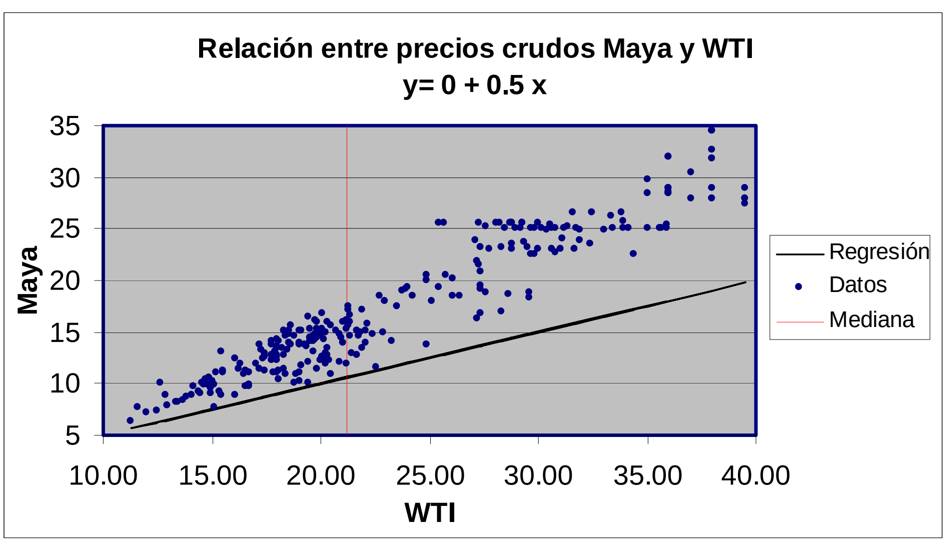
Como se puede observar en la gráfica anterior, $n_1 = 77$ y $n_2 = 70$, con estos datos se calcula

$$X^2 = \frac{8}{267}\left[\left(77 - \frac{267}{4}\right)^2 + \left(70 - \frac{267}{4}\right)^2\right] = 2.831460674$$

Observando en tablas se puede decir que cuando H_0 es verdadero, la probabilidad de obtener un valor de X^2 similar a 2.8314 es mayor que 0.10. Por lo que no se rechaza H_0, concluyéndose que la muestra puede provenir de una población en la cual la regresión lineal tiene una pendiente de 0.88 y una intersección de -3.3.

Enseguida, para determinar la confiabilidad de esta prueba, se establecerá una nueva hipótesis para los parámetros. Así:

Hipótesis: $H_0 : \beta_0 = 0, \beta_1 = 0.5 \quad vs \quad H_1 : \beta_0 \neq 0 \, , \beta_1 \neq 0.5$



Donde $n_1 = 133$ y $n_2 = 133$, entonces

$$X^2 = \frac{8}{267}\left[\left(133 - \frac{267}{4}\right)^2 + \left(133 - \frac{267}{4}\right)^2\right] = 263.0149813$$

Observando en tablas se puede decir que cuando H_0 es verdadero, la probabilidad de obtener un valor de X^2 similar a 263.015 es 0. Por lo anterior se rechaza H_0 y llegando a la conclusión de que la muestra no puede provenir de una población en la cual la regresión lineal tiene una pendiente de 0.5 y una intercepción de 0.

Por otro lado ahora lo que se desea hacer es una prueba de hipótesis sobre β_1 por lo que se utiliza el siguiente procedimiento:

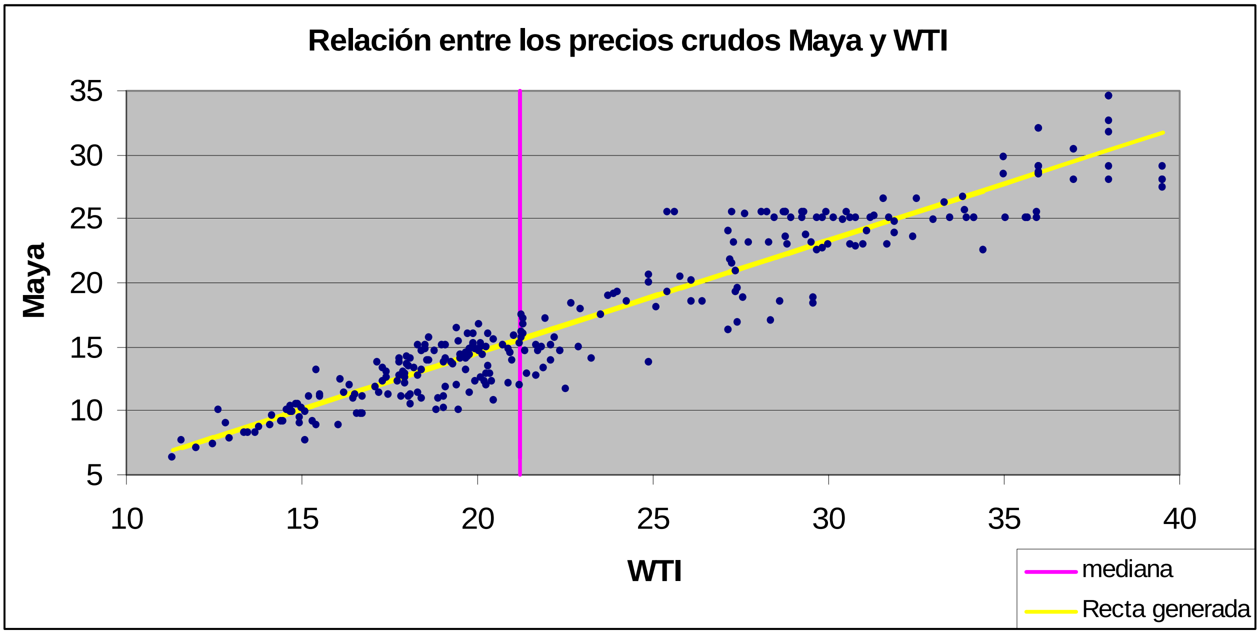
Prueba de Hipótesis

$H_0 : \beta_1 = 0.88 \;\; vs \;\; H_1 : \beta_1 \neq 0.88$

Se calculo la mediana de los valores de x obteniendo como resultado 21.19.

Se ajusto la línea $y = \beta_0 + \beta_1^* x$ a los datos, donde β_0 es la mediana de las desviaciones $y_i - \beta_1^* x$; teniendo la siguiente línea: $y_i = -3.0728 + 0.88 x_i$.

A continuación se presenta un resumen en la siguiente gráfica.



El total de puntos que se encuentran por arriba de la línea generada y a la izquierda de mediana es de 65 contra un total de 267 puntos.

En este caso el estadístico de prueba es:

$$X_b^2 = \frac{16}{267}\left(65 - \frac{267}{4}\right)^2 = 0.183521$$

Rechace H_0 al nivel α si: $X_b^2 > \chi^2_{(\alpha,1)}$

Tomando $\alpha = 0.05$ y teniendo que $\chi^2_{(0.05,1)} = 0.00393$

Por lo tanto se rechaza la hipótesis $H_0 : \beta_1 = 0.88$.

3.4.2 Método Theil

Se hará la evaluación para la hipótesis nula es decir, $\beta_1 = 0$, con un nivel de significancia de $\alpha = 0.05$

Estadístico de Prueba

Haciendo los cálculos correspondientes se obtuvo la siguiente tabla, cabe destacar que no se presentan todos los resultados, dado que son alrededor de 267 valores.

x_i	$y_i - \beta_1 x_i$	Número de pares mayores que $y_i - \beta_1 x_i$ en orden natural	Número de pares menores que $y_i - \beta_1 x_i$ en orden natural
11.29	6.37	266	0
11.58	7.63	263	2
12	7.13	264	0
12.48	7.42	263	0
12.62	10.02	240	22
12.85	8.96	252	9
12.95	7.87	259	1
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
38	31.8	0	0
39.5	27.36	3	0
39.5	28	1	0
39.5	28	1	0
39.5	29	0	0
		P=31366	Q=3940

Con los datos de esta tabla, se calcula $S = 31366 - 3940 = 27426$. Para finalmente obtener:

$$\hat{\tau} = \frac{27426}{267(266)/2} = 0.77232407$$

A continuación se probará las siguientes hipótesis.

Hipótesis

A. (Dos colas) $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$ con $\alpha = 0.05$

B. (Una cola) $H_0 : \beta_1 \leq \beta_1^* \quad vs \quad H_1 : \beta_1 > \beta_1^*$ con $\alpha = 0.05$

Regla de Decisión

Para la hipótesis (A) la regla de decisión es:

Rechazar H_0 al nivel de significancia α = 0.05 si el valor calculado de $\hat{\tau}$ es positivo y mayor que $\hat{\tau}^*$ con n =267 y $\frac{\alpha}{2}$ = 0.025, o negativo y menor que el negativo de $\hat{\tau}^*$ con n y $\frac{\alpha}{2}$.

Como el número de datos es muy grande no se alcanza a obtener un valor específico de $\hat{\tau}^*$ por lo que la opción fue extrapolar obteniendo $\hat{\tau}^*$ = -0.5084. Si esto fuera válido la prueba se rechazaría pero como se genera un valor negativo y el último valor que arroja la tabla (A.1) es para n=40 con $\hat{\tau}^*$=0.185 por lo que no es tan confiable afirmar que la hipótesis se acepta o se rechaza.

Lo mismo acontece para la prueba de una cola.

3.5 Aplicación del Intervalo de Confianza para β_1

Se tienen los siguientes datos, calculados anteriormente:

$n = 267$; $N = 35511$; $\frac{\alpha}{2} = 0.025$

1. Los datos originales se encuentran en el anexo.
2. De igual forma las muestras de las pendientes ordenadas se elaboraron anteriormente.
3. Observando los valores en tablas (Tabla A.1) con $n = 267$ y $\frac{\alpha}{2} = 0.025$, se obtiene $S_{\frac{\alpha}{2}} = 1577$
4. Restando 2 de $S_{\frac{\alpha}{2}}$ se obtiene 1577-2=1575
5. Calculando $k = \frac{(35511 - 1575)}{2} = 16968$

Entonces $\hat{\beta}_L$ es el valor que está en la posición 16968 de S_{ij}, contando a partir del valor más pequeño en el orden presentado en el paso 2.

Entonces $\hat{\beta}_L = 0.86302251$ y el límite superior $\hat{\beta}_U = 0.91037344$. Por lo que el intervalo de confianza queda como sigue:

$(0.86302, 0.91037)$ al 95% de confianza

3.6 Aplicación de la Prueba de paralelismo de dos regresiones lineales

En esta ocasión la prueba se hará entre los precios corrientes y precios constantes de Maya y WTI.

Hipótesis

Se tienen dos grupos de datos, el grupo de precios corrientes y el grupo de precios constantes.

Aplicando regresión paramétrica al grupo de precios corrientes, el modelo es el siguiente:

$$\hat{y}_{1j} = -3.3486411 + 0.884654 x_{1j}$$

Y el modelo para el grupo de precios constantes es:

$$\hat{y}_{2j} = -3.294643 + 0.84792\, x_{2j}$$

Se probarán las siguientes hipótesis nulas (H_0) contra las alternativas H_1.

A. (Dos colas) $H_0 : \beta_1^{*} = \beta_1^{**} \quad vs \quad H_1 : \beta_1^{*} \neq \beta_1^{**}$

B. (Una cola) $H_0 : \beta_1^{*} \geq \beta_1^{**} \quad vs \quad H_1 : \beta_1^{*} < \beta_1^{**}$

Estadístico de Prueba:

Se obtuvieron 266 pares de observaciones, es decir se descartó aleatoriamente un par de cada grupo.
Por lo tanto $2n = 266$.

Con $n = 133$, se hacen pares de la manera en que se describió en el capítulo anterior. Por ejemplo:

$$u_{11} = \frac{y_{1,134} - y_{1,1}}{x_{1,134} - x_{1,1}} = \frac{15.32 - 6.37}{21.19 - 11.29} = 0.9040404$$

Aleatoriamente se forman pares con u_{1j}´s y u_{2t}´s.

Las diferencias entre los miembros de cada par son las siguientes:

$$\begin{array}{lcl} Z_1 = 0.35796667 & & Z_{130} = 0.060169788 \\ Z_2 = \text{-}0.354824761 & & Z_{131} = 0.077002199 \\ Z_3 = 0.134182922 & \ldots\ldots & Z_{132} = 0.792142938 \\ Z_4 = 0.29414306 & & Z_{133} = \text{-}0.034536176 \end{array}$$

A continuación se presentan los resultados finales

Z_i	$\vert Z_i \vert$	Rango de $\vert Z_i \vert$	Rangos con signo
0.35796667	0.35796667	103	103
-0.35482476	0.35482476	102	-102
0.13418292	0.13418292	42	42
0.29414306	0.29414306	86	86
.	.	.	.
.	.	.	.
0.06016979	0.06016979	18	18
0.0770022	0.0770022	23	23
0.79214294	0.79214294	130	130
-0.03453618	0.03453618	8	-8

Para tener como resultado final $T_+ = 5278$ y $T_- = 3633$.

Regla de Decisión

(Dos colas): Como $T_- = 3633$ es la más pequeña. Rechazar H_0 al nivel de significancia $\alpha = 0.05$ si T_- es más pequeña o igual que la T calculada en la tabla A.2 con n y $\frac{\alpha}{2}$.

Pero como $n > 100$, no es posible realizar esta prueba por el método anterior por lo que se aplicará un método alterno[1]

Finalmente se tiene que Z = 1.84605055 < $N(0,1)^{\left(1-\frac{\alpha}{2}\right)}$ =1.96

$\therefore$ Se rechaza H_0

(Una cola) $H_0 : \beta_1^* \geq \beta_1^{**} \quad vs \quad H_1 : \beta_1^* < \beta_1^{**}$

En este caso $T_+ = 5278$, por lo tanto Z= 1.84605055 > $N(0,1)^{(1-\alpha)}$ =1.645

$\therefore$No se rechaza H_0

[1] Veáse anexo III

3.7 Aplicación del Estimador para la diferencia entre los parámetros de la pendiente

Lo que se desea es estimar la diferencia entre β_1^{MAYA} y β_1^{WTI}.

1 Los valores de Z obtenidos se presentaron en la sección anterior.

2 Los $\frac{n(n+1)}{2} = 8911$ valores de $\frac{(Z_i + Z_j)}{2}$, donde $i \leq j$, obtenidos son los siguientes:

-0.64809025	-0.64286183	-0.6376334	-0.63733011	-0.63391175
-0.63210169	-0.62868332	-0.62656997	-0.6231516	-0.61973324
.	.	.	.	.
.	.	.	.	.
0.82892668	0.8363268	0.86281756	0.86582539	0.88120943
0.89231615	0.91070802	0.93349218	0.96299077	0.99248936

3 El estimador de la diferencia entre las pendientes es la mediana de esos 36 valores. Esto es:

$$\hat{\Theta} = \frac{0.052509824 + 0.052559449}{2} = 0.05255945$$

3.8 Aplicación del Intervalo de confianza para la diferencia entre los parámetros de la pendiente

Sea $\Theta = \beta_1^{MAYA} - \beta_1^{WTI}$:

Se construirá un intervalo de confianza del 95 %

El nivel de confianza es $0.95 = (1 - 0.05)$, por lo que $\alpha = 0.05$ y entonces $\frac{\alpha}{2} = 0.025$. Revisando en la tabla (A.2), para n=8911 el valor de P no se puede obtener ya que no es posible calcular un valor que sea más cercano a 0.025 por la forma en como se encuentra la tabla. Por lo tanto no se pueden obtener los valores de T y K.

Y por ende no es posible conocer el intervalo de confianza para Θ.

CONCLUSIONES

Comparando los dos métodos de regresión no paramétrica Brown-Mood y Theil, el método de Theil se aproxima más a los resultados arrojados por la regresión paramétrica, la diferencia que existe entre ambos es mínima.

En cuanto al intervalo de confianza el método que se utilizó fue muy efectivo dado que es muy similar al intervalo que se calcula en la regresión paramétrica.

Regresión Paramétrica	**Regresión No Paramétrica (Método de Theil)**
$\hat{y} = -3.34864118 + 0.88465404\,x$	$\hat{y} = -3.371401802 + 0.88562092x$
Intervalo de confianza (0.85089,0.91842) *al 95% de confianza*	*Intervalo de confianza* (0.86302,0.91037) *al 95% de confianza*

Con la experiencia que se tuvo al realizar este trabajo, se llegaron a las siguientes conclusiones donde se mencionan las ventajas y desventajas de la regresión no paramétrica.

Ventajas:

Como se mencionó anteriormente es muy conveniente utilizar estos métodos cuando no se tiene una idea acerca de la forma funcional de la distribución de probabilidad de las observaciones de la muestra.

Si no se cuenta con un paquete estadístico son muy útiles estos métodos, finalmente se vio que los resultados son muy similares.

Desventajas:

Cuando hay demasiados datos, hay ciertas pruebas que no se pueden calcular, por ejemplo calcular la prueba de paralelismo entre dos regresiones lineales y el estimador e intervalo de confianza para la diferencia entre los parámetros de la pendiente, esto debido a que son demasiados datos y no es posible encontrar los valores en tablas.

Hacer los cálculos a través del método de Theil es muy pesado, debido a que son muchos cálculos, la forma más eficiente y rápida para este método es a través de una macro en excel.

De todo esto queda decir que los métodos presentados son muy útiles cuando los datos son pocos y requieren de elementos estadísticos más simples que los métodos tradicionales.

ANEXOS

ANEXO I

$Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$, pero como σ^2 es desconocida se propone s^2, entonces se tiene lo siguiente:

$$V\hat{a}r(\hat{\beta}_1) = \frac{s^2}{S_{xx}}.$$

Por lo que se llega a la siguiente aseveración

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{V\hat{a}r(\hat{\beta}_1)}} \sim t_{(n-2)}$$

Ésta se basa en el siguiente teorema:

Se sabe que si:

$$x_i \sim N(\mu, \sigma^2)$$

$$\Rightarrow \frac{x_i - \mu}{\sigma} \sim N(0,1)$$

$$\Rightarrow \frac{(x_i - \mu)^2}{\sigma^2} \sim \chi^2_{(1)}$$

$$\Rightarrow \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} \sim \chi^2_{(n)}$$

Entonces retomando $V\hat{a}r(\hat{\beta}_1) = \frac{s^2}{S_{xx}}$ y $s^2 = \sum \frac{(y_i - \bar{y})^2}{n-2}$, por lo anterior se puede decir que:

$$\frac{\sum(y_i-\bar{y})^2}{(n-2)\sigma^2} \sim \frac{\chi^2_{(n-2)}}{n-2}$$

Usando este resultado se tiene que:

$$\frac{\hat{\beta}_1-\beta_1}{\sqrt{V\hat{a}r(\hat{\beta}_1)}} = \frac{\hat{\beta}_1-\beta_1}{\sqrt{Var(\hat{\beta}_1)}}\left(\frac{\sqrt{Var(\hat{\beta}_1)}}{\sqrt{V\hat{a}r(\hat{\beta}_1)}}\right)$$

$$= \frac{\hat{\beta}_1-\beta_1}{\sqrt{Var(\hat{\beta}_1)}} \div \left(\frac{\sqrt{V\hat{a}r(\hat{\beta}_1)}}{\sqrt{Var(\hat{\beta}_1)}}\right)$$

Por un lado

$$\frac{\sqrt{V\hat{a}r(\hat{\beta}_1)}}{\sqrt{Var(\hat{\beta}_1)}} = \frac{\dfrac{\sum(y_i-\bar{y})^2}{(n-2)} \Big/ \sum(x_i-\bar{x})^2}{\sigma^2 \Big/ \sum(x_i-\bar{x})^2}$$

$$= \frac{\sum(y_i-\bar{y})^2}{(n-2)\sigma^2}$$

$$\Rightarrow \frac{\sum(y_i-\bar{y})^2}{(n-2)\sigma^2} \sim \frac{\chi^2_{(n-2)}}{n-2}$$

Por otro lado

$$\frac{\hat{\beta}_1-\beta_1}{\sqrt{Var(\hat{\beta}_1)}} \sim N(0,1)$$

Finalmente

$$\therefore \frac{\hat{\beta}_1-\beta_1}{\sqrt{V\hat{a}r(\hat{\beta}_1)}} \sim \frac{N(0,1)}{\sqrt{\dfrac{\chi^2_{(n-2)}}{n-2}}} \sim t_{(n-2)}$$

ANEXO II

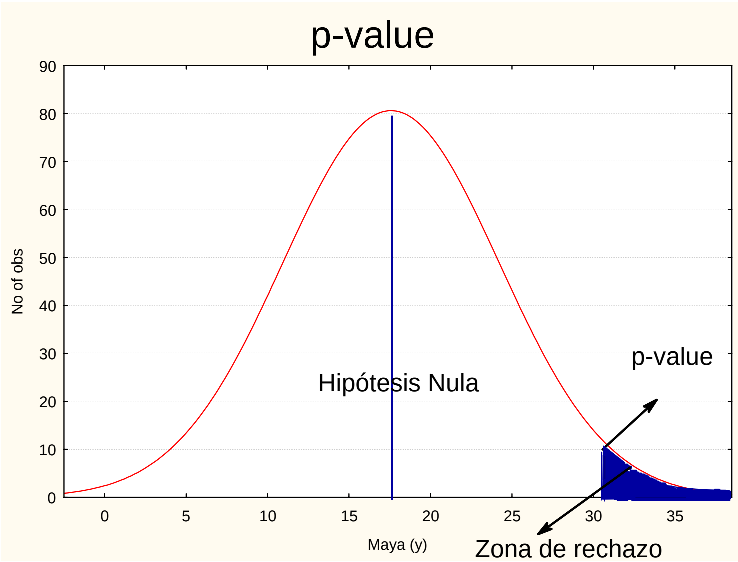
P – Value

El p-value es el nivel de significancia más pequeño posible al que se podría rechazar la hipótesis nula para observaciones dadas.

Al p-value también se le conoce como nivel de significancia observado, que es el mínimo nivel al cual Ho puede ser rechazada para un conjunto dado de datos.

Si el p-value es $\leq \alpha^{*} \Rightarrow$ Se Rechaza la Hipótesis Nula (H_0).

Si el p-value es $> \alpha \Rightarrow$ No se Rechaza la Hipótesis Nula (H_0).



[*] Nivel de significancia; usualmente es de 0.05

ANEXO III

Para n>100, se aplica lo siguiente

Sea T ~ Función de distribución

Sea $Z = \dfrac{|T-\mu_T| - 0.5}{\sigma_T} \sim N(0,1)$

donde $\mu_T = \dfrac{n(n+1)}{4}$ y $\sigma_T = \sqrt{\dfrac{n(n+1)(2n+1)}{24}}$.

Con esto se siguen las siguientes pruebas de hipótesis:

Hipótesis (Dos colas): $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$

Regla de decisión: Rechace H_0 si $|Z| > N(0,1)$ al nivel $\left(1-\dfrac{\alpha}{2}\right)$.

Hipótesis (Una cola): $H_0 : \mu \leq \mu_0 \quad vs \quad H_1 : \mu > \mu_0$

Regla de decisión: Rechace H_0 si $\mathrm{Z} > N(0,1)$ al nivel $(1-\alpha)$.

TABLAS

Precios de los crudos Maya y WTI

Mes	Precios Corrientes		Precios Constantes (dll.2001)		CPI
	WTI (x)	Maya (y)	WTI (x)	Maya (y)	
Ene-80	32.50	26.59	73.97	60.52	0.43938
Feb-80	37.00	28.00	83.04	62.84	0.44559
Mar-80	38.00	28.00	84.00	61.90	0.45237
Abr-80	39.50	27.36	86.35	59.82	0.45745
May-80	39.50	28.00	85.50	60.61	0.46197
Jun-80	39.50	28.00	84.57	59.95	0.46706
Jul-80	39.50	29.00	84.57	62.09	0.46706
Ago-80	38.00	29.00	80.77	61.64	0.47044
Sep-80	36.00	29.00	75.89	61.13	0.47440
Oct-80	36.00	28.99	75.17	60.54	0.47892
Nov-80	36.00	29.00	74.55	60.06	0.48287
Dic-80	37.00	30.39	75.92	62.35	0.48739
Ene-81	38.00	34.50	77.34	70.22	0.49134
Feb-81	38.00	34.50	76.55	69.50	0.49642
Mar-81	38.00	34.50	76.03	69.03	0.49981
Abr-81	38.00	32.66	75.52	64.91	0.50320
May-81	38.00	31.80	74.93	62.69	0.50715
Jun-81	36.00	32.00	70.36	62.54	0.51167
Jul-81	36.00	32.00	69.59	61.86	0.51732
Ago-81	36.00	28.58	69.06	54.82	0.52127
Sep-81	36.00	28.50	68.39	54.15	0.52636
Oct-81	35.00	29.73	66.35	56.37	0.52748
Nov-81	36.00	28.50	68.03	53.86	0.52918
Dic-81	35.00	28.50	65.93	53.69	0.53087
Ene-82	33.85	26.61	63.56	49.96	0.53257
Feb-82	31.58	26.50	59.07	49.60	0.53426
Mar-82	28.48	25.08	53.36	47.00	0.53370
Abr-82	33.45	25.00	62.41	46.65	0.53596
May-82	35.93	25.00	66.41	46.21	0.54104
Jun-82	35.07	25.00	64.02	45.64	0.54782
Jul-82	34.16	25.00	62.04	45.40	0.55064
Ago-82	33.95	25.00	61.53	45.31	0.55177
Sep-82	35.63	25.00	64.44	45.22	0.55290
Oct-82	35.68	25.00	64.34	45.08	0.55459
Nov-82	34.15	25.00	61.70	45.17	0.55346
Dic-82	31.72	25.00	57.55	45.36	0.55120
Ene-83	31.19	25.02	56.47	45.30	0.55233
Feb-83	28.95	25.01	52.36	45.24	0.55290
Mar-83	28.82	23.00	52.13	41.60	0.55290
Abr-83	30.61	23.00	54.97	41.30	0.55685
May-83	30.00	23.00	53.55	41.05	0.56024
Jun-83	31.00	23.00	55.17	40.93	0.56194

Mes	Precios Corrientes		Precios Constantes (dll.2001)		CPI
	WTI (x)	Maya (y)	WTI (x)	Maya (y)	
Jul-83	31.66	23.00	56.12	40.77	0.56419
Ago-83	31.91	23.90	56.39	42.23	0.56589
Sep-83	31.11	24.00	54.70	42.20	0.56871
Oct-83	30.41	24.97	53.31	43.77	0.57041
Nov-83	29.84	25.00	52.21	43.74	0.57154
Dic-83	29.24	25.00	51.11	43.70	0.57210
Ene-84	29.69	25.00	51.59	43.44	0.57549
Feb-84	30.15	25.00	52.13	43.23	0.57831
Mar-84	30.78	25.00	53.09	43.14	0.57944
Abr-84	30.62	25.00	52.58	42.94	0.58227
May-84	30.52	25.46	52.26	43.61	0.58396
Jun-84	29.97	25.50	51.17	43.54	0.58566
Jul-84	28.75	25.50	48.91	43.37	0.58791
Ago-84	29.25	25.50	49.56	43.21	0.59017
Sep-84	29.31	25.50	49.43	43.00	0.59300
Oct-84	28.77	25.50	48.37	42.88	0.59469
Nov-84	28.10	25.50	47.25	42.88	0.59469
Dic-84	25.43	25.50	42.76	42.88	0.59469
Ene-85	25.64	25.50	43.03	42.80	0.59582
Feb-85	27.27	25.50	45.55	42.60	0.59864
Mar-85	28.24	25.50	46.99	42.44	0.60090
Abr-85	28.81	25.50	47.71	42.24	0.60373
May-85	27.62	25.29	45.58	41.73	0.60599
Jun-85	27.14	23.95	44.67	39.42	0.60768
Jul-85	27.33	23.16	44.89	38.04	0.60881
Ago-85	27.76	23.08	45.50	37.85	0.60994
Sep-85	28.29	23.16	46.25	37.86	0.61163
Oct-85	29.54	23.15	48.12	37.71	0.61389
Nov-85	30.81	22.77	50.05	36.99	0.61559
Dic-85	27.23	21.80	44.11	35.32	0.61728
Ene-86	22.95	17.90	37.07	28.92	0.61898
Feb-86	15.44	13.10	25.02	21.23	0.61728
Mar-86	12.62	10.02	20.54	16.30	0.61446
Abr-86	12.85	8.96	20.95	14.65	0.61333
May-86	15.44	8.84	25.11	14.37	0.61502
Jun-86	13.47	8.24	21.78	13.32	0.61841
Jul-86	11.58	7.63	18.72	12.34	0.61841
Ago-86	15.09	9.96	24.36	16.08	0.61954
Sep-86	14.91	10.53	23.96	16.93	0.62236
Oct-86	14.85	10.55	23.84	16.93	0.62293
Nov-86	15.21	11.02	24.39	17.68	0.62349
Dic-86	16.08	12.46	25.76	19.96	0.62406
Ene-87	18.63	13.84	29.67	22.04	0.62801
Feb-87	17.79	14.03	28.23	22.28	0.63027
Mar-87	18.30	15.10	28.91	23.85	0.63309

Mes	Precios Corrientes		Precios Constantes (dll.2001)		CPI
	WTI (x)	Maya (y)	WTI (x)	Maya (y)	
Abr-87	18.65	15.64	29.30	24.58	0.63648
May-87	19.42	16.48	30.40	25.80	0.63874
Jun-87	20.06	16.73	31.33	26.10	0.64100
Jul-87	21.30	17.13	33.14	26.65	0.64270
Ago-87	20.33	15.96	31.47	24.70	0.64608
Sep-87	19.48	15.35	29.99	23.63	0.64947
Oct-87	19.88	15.30	30.53	23.50	0.65117
Nov-87	19.03	13.77	29.20	21.12	0.65173
Dic-87	17.19	11.42	26.38	17.53	0.65173
Ene-88	17.08	11.87	26.14	18.17	0.65343
Feb-88	16.75	11.07	25.57	16.89	0.65512
Mar-88	16.22	11.39	24.65	17.31	0.65794
Abr-88	17.77	12.76	26.87	19.30	0.66133
May-88	17.44	12.63	26.28	19.03	0.66359
Jun-88	16.52	11.27	24.79	16.92	0.66642
Jul-88	15.52	11.17	23.18	16.70	0.66924
Ago-88	15.51	11.11	23.07	16.53	0.67206
Sep-88	14.58	10.03	21.54	14.83	0.67658
Oct-88	13.80	8.70	20.32	12.82	0.67884
Nov-88	14.15	9.64	20.83	14.18	0.67941
Dic-88	16.45	10.90	24.18	16.02	0.68053
Ene-89	17.96	12.17	26.26	17.79	0.68392
Feb-89	17.97	12.60	26.17	18.34	0.68675
Mar-89	19.51	14.07	28.25	20.38	0.69070
Abr-89	21.19	15.32	30.47	22.03	0.69522
May-89	20.25	14.91	28.97	21.33	0.69917
Jun-89	20.03	14.63	28.58	20.88	0.70087
Jul-89	19.81	14.36	28.20	20.44	0.70256
Ago-89	18.58	13.91	26.41	19.77	0.70369
Sep-89	19.54	14.42	27.68	20.42	0.70595
Oct-89	20.09	14.93	28.33	21.04	0.70934
Nov-89	19.90	15.14	27.99	21.29	0.71103
Dic-89	21.07	15.88	29.59	22.30	0.71216
Ene-90	22.90	14.90	31.83	20.71	0.71950
Feb-90	22.11	13.86	30.59	19.17	0.72289
Mar-90	20.41	12.30	28.08	16.92	0.72684
Abr-90	18.43	10.93	25.32	15.01	0.72797
May-90	18.10	10.45	24.80	14.32	0.72967
Jun-90	16.75	9.82	22.83	13.38	0.73362
Jul-90	18.44	13.19	25.05	17.91	0.73645
Ago-90	27.38	20.83	36.84	28.02	0.74322
Sep-90	33.34	26.24	44.49	35.01	0.74944
Oct-90	35.95	25.46	47.68	33.77	0.75395
Nov-90	32.43	23.59	42.92	31.22	0.75565
Dic-90	27.36	19.22	36.20	25.44	0.75565

Mes	Precios Corrientes		Precios Constantes (dll.2001)		CPI
	WTI (x)	Maya (y)	WTI (x)	Maya (y)	
Ene-91	24.91	13.80	32.77	18.15	0.76017
Feb-91	20.47	10.85	26.89	14.25	0.76130
Mar-91	19.80	11.37	25.97	14.91	0.76242
Abr-91	20.87	12.07	27.33	15.81	0.76355
May-91	21.22	11.99	27.71	15.66	0.76581
Jun-91	20.26	11.93	26.37	15.53	0.76807
Jul-91	21.43	12.85	27.86	16.71	0.76920
Ago-91	21.68	12.72	28.11	16.49	0.77146
Sep-91	21.91	13.36	28.27	17.24	0.77485
Oct-91	23.25	14.09	29.96	18.16	0.77598
Nov-91	22.53	11.64	28.95	14.96	0.77824
Dic-91	19.46	10.08	24.99	12.95	0.77880
Ene-92	18.82	10.08	24.13	12.92	0.77993
Feb-92	19.05	10.22	24.33	13.06	0.78276
Mar-92	18.91	10.93	24.03	13.89	0.78671
Abr-92	20.23	12.26	25.68	15.57	0.78784
May-92	21.00	13.88	26.61	17.60	0.78897
Jun-92	22.37	14.72	28.26	18.59	0.79179
Jul-92	21.74	14.69	27.40	18.52	0.79349
Ago-92	21.35	14.59	26.83	18.34	0.79575
Sep-92	21.86	15.00	27.39	18.80	0.79800
Oct-92	21.71	15.05	27.12	18.79	0.80083
Nov-92	20.32	13.50	25.33	16.83	0.80196
Dic-92	19.41	12.00	24.23	14.97	0.80139
Ene-93	19.12	11.80	23.74	14.66	0.80535
Feb-93	20.09	12.51	24.86	15.47	0.80817
Mar-93	20.35	12.81	25.09	15.80	0.81099
Abr-93	20.26	12.87	24.91	15.83	0.81325
May-93	19.97	12.22	24.52	15.01	0.81438
Jun-93	19.03	11.11	23.34	13.62	0.81551
Jul-93	17.82	11.02	21.85	13.52	0.81551
Ago-93	18.04	11.15	22.05	13.64	0.81777
Sep-93	17.46	11.28	21.31	13.76	0.81947
Oct-93	18.11	11.22	22.01	13.63	0.82285
Nov-93	16.56	9.72	20.11	11.80	0.82342
Dic-93	14.49	9.10	17.60	11.05	0.82342
Ene-94	15.01	10.23	18.18	12.39	0.82568
Feb-94	14.76	9.95	17.82	12.02	0.82850
Mar-94	14.70	10.31	17.68	12.40	0.83133
Abr-94	16.35	11.94	19.65	14.35	0.83245
May-94	17.97	12.82	21.57	15.39	0.83302
Jun-94	19.08	14.01	22.83	16.76	0.83584
Jul-94	19.70	14.53	23.50	17.33	0.83810
Ago-94	18.29	12.76	21.74	15.17	0.84149
Sep-94	17.44	12.53	20.66	14.85	0.84375

Mes	Precios Corrientes		Precios Constantes (dll.2001)		CPI
	WTI (x)	Maya (y)	WTI (x)	Maya (y)	
Oct-94	17.77	13.83	21.04	16.38	0.84431
Nov-94	18.10	14.10	21.41	16.68	0.84544
Dic-94	17.16	13.83	20.30	16.36	0.84544
Ene-95	18.02	14.28	21.23	16.83	0.84883
Feb-95	18.53	14.80	21.74	17.37	0.85222
Mar-95	18.55	15.12	21.69	17.68	0.85505
Abr-95	19.88	16.03	23.18	18.69	0.85787
May-95	19.74	16.05	22.96	18.67	0.85956
Jun-95	18.42	14.61	21.39	16.97	0.86126
Jul-95	17.30	13.26	20.09	15.39	0.86126
Ago-95	18.03	13.45	20.88	15.57	0.86352
Sep-95	18.21	13.36	21.05	15.44	0.86521
Oct-95	17.44	12.99	20.09	14.97	0.86803
Nov-95	18.00	13.58	20.74	15.65	0.86747
Dic-95	19.02	15.12	21.94	17.44	0.86691
Ene-96	18.80	14.64	21.56	16.79	0.87199
Feb-96	19.09	15.10	21.83	17.26	0.87481
Mar-96	21.33	16.68	24.25	18.97	0.87933
Abr-96	23.51	17.46	26.63	19.78	0.88272
May-96	21.24	16.07	24.02	18.17	0.88441
Jun-96	20.45	15.58	23.11	17.58	0.88498
Jul-96	21.32	15.98	24.04	18.02	0.88667
Ago-96	21.93	17.15	24.68	19.31	0.88837
Sep-96	24.00	19.31	26.93	21.67	0.89119
Oct-96	24.90	20.56	27.85	23.00	0.89401
Nov-96	23.72	18.91	26.48	21.12	0.89571
Dic-96	25.41	19.32	28.37	21.57	0.89571
Ene-97	25.13	18.06	27.97	20.10	0.89853
Feb-97	22.19	15.71	24.61	17.43	0.90136
Mar-97	20.96	14.48	23.19	16.03	0.90361
Abr-97	19.75	14.26	21.83	15.76	0.90474
May-97	20.91	14.83	23.13	16.41	0.90418
Jun-97	19.28	13.80	21.29	15.24	0.90531
Jul-97	19.63	14.16	21.66	15.62	0.90644
Ago-97	19.93	14.82	21.95	16.32	0.90813
Sep-97	19.78	14.75	21.72	16.20	0.91039
Oct-97	21.27	15.70	23.31	17.20	0.91265
Nov-97	20.18	14.32	22.12	15.70	0.91209
Dic-97	18.30	11.44	20.09	12.56	0.91096
Ene-98	16.69	9.71	18.29	10.64	0.91265
Feb-98	16.07	8.88	17.58	9.71	0.91434
Mar-98	15.10	7.68	16.48	8.38	0.91604
Abr-98	15.32	9.14	16.69	9.96	0.91773
May-98	14.93	9.08	16.24	9.88	0.91943
Jun-98	13.69	8.30	14.87	9.01	0.92056

Mes	Precios Corrientes		Precios Constantes (dll.2001)		CPI
	WTI (x)	Maya (y)	WTI (x)	Maya (y)	
Jul-98	14.12	8.80	15.32	9.55	0.92169
Ago-98	13.39	8.24	14.51	8.93	0.92282
Sep-98	14.97	9.47	16.21	10.24	0.92395
Oct-98	14.42	9.13	15.57	9.86	0.92620
Nov-98	12.95	7.87	13.98	8.49	0.92620
Dic-98	11.29	6.37	12.20	6.89	0.92564
Ene-99	12.48	7.42	13.45	7.99	0.92790
Feb-99	12.00	7.13	12.91	7.67	0.92903
Mar-99	14.66	9.86	15.74	10.58	0.93185
Abr-99	17.34	12.35	18.47	13.16	0.93863
May-99	17.74	12.24	18.90	13.04	0.93863
Jun-99	17.90	13.07	19.07	13.93	0.93863
Jul-99	20.08	15.32	21.33	16.27	0.94145
Ago-99	21.27	17.43	22.53	18.47	0.94371
Sep-99	23.88	19.08	25.19	20.12	0.94823
Oct-99	22.69	18.43	23.89	19.41	0.94992
Nov-99	24.88	19.97	26.17	21.01	0.95049
Dic-99	26.11	20.09	27.47	21.14	0.95049
Ene-00	27.26	21.55	28.59	22.60	0.95331
Feb-00	29.39	23.66	30.64	24.67	0.95896
Mar-00	29.86	22.61	30.89	23.38	0.96687
Abr-00	25.78	20.45	26.65	21.14	0.96743
May-00	28.80	23.57	29.73	24.33	0.96856
Jun-00	31.88	24.81	32.74	25.48	0.97364
Jul-00	29.71	22.51	30.44	23.06	0.97590
Ago-00	31.33	25.16	32.10	25.78	0.97590
Sep-00	33.89	25.71	34.54	26.21	0.98099
Oct-00	33.02	24.86	33.60	25.29	0.98268
Nov-00	34.40	22.50	34.98	22.89	0.98325
Dic-00	28.35	17.01	28.85	17.31	0.98268
Ene-01	29.58	18.82	29.89	19.03	0.98889
Feb-01	29.58	18.37	29.78	18.50	0.99285
Mar-01	27.18	16.22	27.32	16.30	0.99511
Abr-01	27.40	16.84	27.42	16.86	0.99906
May-01	28.61	18.57	28.51	18.51	1.00558
Jun-01	27.57	18.78	27.42	18.68	1.00527
Jul-01	26.43	18.50	26.37	18.45	1.00245
Ago-01	27.40	19.52	27.33	19.47	1.00245
Sep-01	26.08	18.45	25.90	18.32	1.00697
Oct-01	22.08	15.05	22.00	15.00	1.00358
Nov-01	19.67	13.15	19.63	13.12	1.00188
Dic-01	19.33	13.67	19.37	13.70	0.99793
Ene-02	19.67	14.06	19.67	14.08	1.00019
Feb-02	20.74	15.06	20.65	15.02	1.00414
Mar-02	24.24	18.50	24.01	18.32	1.00979

TABLA A.1

Critical values for use with the Kendall tau statistic

n \ α	0.005		0.010		0.025		0.050		0.100	
	S	τ^*	S	τ^*	S	τ^*	S	τ^*	S	τ^*
4	8	1.000	8	1.000	8	1.000	6	1.000	6	1.000
5	12	1.000	10	1.000	10	1.000	8	.800	8	.800
6	15	1.000	13	.867	13	.867	11	.733	9	.600
7	19	.905	17	.810	15	.714	13	.619	11	.524
8	22	.786	20	.714	18	.643	16	.571	12	.429
9	26	.722	24	.667	20	.556	18	.500	14	.389
10	29	.644	27	.600	23	.511	21	.467	17	.378
11	33	.600	31	.564	27	.491	23	.418	19	.345
12	38	.576	36	.545	30	.455	26	.394	20	.303
13	44	.564	40	.513	34	.436	28	.359	24	.308
14	47	.516	43	.473	37	.407	33	.363	25	.275
15	53	.505	49	.467	41	.390	35	.333	29	.276
16	58	.483	52	.433	46	.383	38	.317	30	.250
17	64	.471	58	.426	50	.368	42	.309	34	.250
18	69	.451	63	.412	53	.346	45	.294	37	.242
19	75	.439	67	.392	57	.333	49	.287	39	.228
20	80	.421	72	.379	62	.326	52	.274	42	.221
21	86	.410	78	.371	66	.314	56	.267	44	.210
22	91	.394	83	.359	71	.307	61	.264	47	.203
23	99	.391	89	.352	75	.296	65	.257	51	.202
24	104	.377	94	.341	80	.290	68	.246	54	.196
25	110	.367	100	.333	86	.287	72	.240	58	.193
26	117	.360	107	.329	91	.280	77	.237	61	.188
27	125	.356	113	.322	95	.271	81	.231	63	.179
28	130	.344	118	.312	100	.265	86	.228	68	.180
29	138	.340	126	.310	106	.261	90	.222	70	.172
30	145	.333	131	.301	111	.255	95	.218	75	.172
31	151	.325	137	.295	117	.252	99	.213	77	.166
32	160	.323	144	.290	122	.246	104	.210	82	.165
33	166	.314	152	.288	128	.242	108	.205	86	.163
34	175	.312	157	.280	133	.237	113	.201	89	.159
35	181	.304	165	.277	139	.234	117	.197	93	.156
36	190	.302	172	.273	146	.232	122	.194	96	.152
37	198	.297	178	.267	152	.228	128	.192	100	.150
38	205	.292	185	.263	157	.223	133	.189	105	.149
39	213	.287	193	.260	163	.220	139	.188	109	.147
40	222	.285	200	.256	170	.218	144	.185	112	.144

Source: L. Kaarsemaker and A. van Wijngaarden, "Tables for Use in Rank Correlation," *Statistica Neerlandica*, 7 (1953), 41–54.

* The column labeled S contains, for each n, the smallest value of S for which $P(S \geq S) \leq \alpha$. The column labeled τ^* contains, for each n, the smallest value of τ^* for which $P(\tau \geq \tau^*) \leq \alpha$.

TABLA A.2

Probability levels for the Wilcoxon signed-rank test

n = 5

T	P
*0	.0313
1	.0625
2	.0938
3	.1563
4	.2188
5	.3125
6	.4063
7	.5000

n = 6

T	P
0	.0156
1	.0313
*2	.0469
3	.0781
4	.1094
5	.1563
6	.2188
7	.2813
8	.3438
9	.4219
10	.5000

n = 7

T	P
0	.0078
1	.0156
2	.0234
*3	.0391
4	.0547
5	.0781
6	.1094
7	.1484
8	.1875
9	.2344
10	.2891
11	.3438
12	.4063
13	.4688
14	.5313

n = 8

T	P
0	.0039
1	.0078
2	.0117
3	.0195
4	.0273
*5	.0391
6	.0547
7	.0742
8	.0977
9	.1250
10	.1563
11	.1914
12	.2305
13	.2734
14	.3203
15	.3711
16	.4219
17	.4727
18	.5273

n = 9

T	P
0	.0020
1	.0039
2	.0059
3	.0098
4	.0137
5	.0195
6	.0273
7	.0371
*8	.0488
9	.0645
10	.0820
11	.1016
12	.1250
13	.1504
14	.1797
15	.2129
16	.2480
17	.2852
18	.3262
19	.3672
20	.4102
21	.4551
22	.5000

n = 10

T	P
0	.0010
1	.0020
2	.0029
3	.0049
4	.0068
5	.0098
6	.0137
7	.0186
8	.0244
9	.0322
*10	.0420
11	.0527
12	.0654
13	.0801
14	.0967
15	.1162
16	.1377
17	.1611
18	.1875
19	.2158
20	.2461
21	.2783
22	.3125
23	.3477
24	.3848
25	.4229
26	.4609
27	.5000

n = 11

T	P
0	.0005
1	.0010
2	.0015
3	.0024
4	.0034
5	.0049
6	.0068
7	.0093
8	.0122
9	.0161
10	.0210
11	.0269
12	.0337
*13	.0415
14	.0508
15	.0615
16	.0737
17	.0874
18	.1030
19	.1201
20	.1392
21	.1602
22	.1826
23	.2065
24	.2324
25	.2598
26	.2886
27	.3188
28	.3501
29	.3823
30	.4155
31	.4492
32	.4829
33	.5171

n = 12

T	P
0	.0002
1	.0005
2	.0007
3	.0012
4	.0017
5	.0024
6	.0034
7	.0046
8	.0061
9	.0081
10	.0105
11	.0134
12	.0171
13	.0212
14	.0261
15	.0320
16	.0386
*17	.0461
18	.0549
19	.0647
20	.0757
21	.0881
22	.1018
23	.1167
24	.1331
25	.1506
26	.1697
27	.1902
28	.2119
29	.2349
30	.2593
31	.2847
32	.3110
33	.3386
34	.3667
35	.3955
36	.4250
37	.4548
38	.4849
39	.5151

n = 13

T	P
0	.0001
1	.0002
2	.0004
3	.0006
4	.0009
5	.0012
6	.0017
7	.0023
8	.0031
9	.0040
10	.0052
11	.0067
12	.0085
13	.0107
14	.0133
15	.0164
16	.0199
17	.0239
18	.0287
19	.0341
20	.0402
*21	.0471
22	.0549
23	.0636
24	.0732
25	.0839
26	.0955
27	.1082
28	.1219
29	.1367
30	.1527
31	.1698
32	.1879
33	.2072
34	.2274
35	.2487
36	.2709
37	.2939
38	.3177
39	.3424
40	.3677
41	.3934
42	.4197
43	.4463
44	.4730
45	.5000

* For given n, the smallest rank total for which the probability level is equal to or less than 0.0500.

TABLA A.2 (continuación)

T	P	T	P	T	P	T	P	T	P	T	P
$n = 14$		$n = 14$		$n = 15$		$n = 16$		$n = 17$		$n = 17$	
0	.0001	50	.4516	47	.2444	39	.0719	25	.0064	74	.4633
2	.0002	51	.4758	48	.2622	40	.0795	26	.0075	75	.4816
3	.0003	52	.5000	49	.2807	41	.0877	27	.0087	76	.5000
4	.0004			50	.2997	42	.0964	28	.0101		
5	.0006	$n = 15$		51	.3193	43	.1057	29	.0116	$n = 18$	
6	.0009	1	.0001	52	.3394	44	.1156	30	.0133	6	.0001
7	.0012	3	.0002	53	.3599	45	.1261	31	.0153	10	.0002
8	.0015	5	.0003	54	.3808	46	.1372	32	.0174	12	.0003
9	.0020	6	.0004	55	.4020	47	.1489	33	.0198	14	.0004
10	.0026	7	.0006	56	.4235	48	.1613	34	.0224	15	.0005
11	.0034	8	.0008	57	.4452	49	.1742	35	.0253	16	.0006
12	.0043	9	.0010	58	.4670	50	.1877	36	.0284	17	.0008
13	.0054	10	.0013	59	.4890	51	.2019	37	.0319	18	.0010
14	.0067	11	.0017	60	.5110	52	.2166	38	.0357	19	.0012
15	.0083	12	.0021	$n = 16$		53	.2319	39	.0398	20	.0014
16	.0101	13	.0027	3	.0001	54	.2477	40	.0443	21	.0017
17	.0123	14	.0034	5	.0002	55	.2641	*41	.0492	22	.0020
18	.0148	15	.0042	7	.0003	56	.2809	42	.0544	23	.0024
19	.0176	16	.0051	8	.0004	57	.2983	43	.0601	24	.0028
20	.0209	17	.0062	9	.0005	58	.3161	44	.0662	25	.0033
21	.0247	18	.0075	10	.0007	59	.3343	45	.0727	26	.0038
22	.0290	19	.0090	11	.0008	60	.3529	46	.0797	27	.0045
23	.0338	20	.0108	12	.0011	61	.3718	47	.0871	28	.0052
24	.0392	21	.0128	13	.0013	62	.3910	48	.0950	29	.0060
*25	.0453	22	.0151	14	.0017	63	.4104	49	.1034	30	.0069
26	.0520	23	.0177	15	.0021	64	.4301	50	.1123	31	.0080
27	.0594	24	.0206	16	.0026	65	.4500	51	.1218	32	.0091
28	.0676	25	.0240	17	.0031	66	.4699	52	.1317	33	.0104
29	.0765	26	.0277	18	.0038	67	.4900	53	.1421	34	.0118
30	.0863	27	.0319	19	.0046	68	.5100	54	.1530	35	.0134
31	.0969	28	.0365	20	.0055			55	.1645	36	.0152
32	.1083	29	.0416	21	.0065	$n = 17$		56	.1764	37	.0171
33	.1206	*30	.0473	22	.0078	4	.0001	57	.1889	38	.0192
34	.1338	31	.0535	23	.0091	8	.0002	58	.2019	39	.0216
35	.1479	32	.0603	24	.0107	9	.0003	59	.2153	40	.0241
36	.1629	33	.0677	25	.0125	11	.0004	60	.2293	41	.0269
37	.1788	34	.0757	26	.0145	12	.0005	61	.2437	42	.0300
38	.1955	35	.0844	27	.0168	13	.0007	62	.2585	43	.0333
39	.2131	36	.0938	28	.0193	14	.0008	63	.2738	44	.0368
40	.2316	37	.1039	29	.0222	15	.0010	64	.2895	45	.0407
41	.2508	38	.1147	30	.0253	16	.0013	65	.3056	46	.0449
42	.2708	39	.1262	31	.0288	17	.0016	66	.3221	*47	.0494
43	.2915	40	.1384	32	.0327	18	.0019	67	.3389	48	.0542
44	.3129	41	.1514	33	.0370	19	.0023	68	.3559	49	.0594
45	.3349	42	.1651	34	.0416	20	.0028	69	.3733	50	.0649
46	.3574	43	.1796	*35	.0467	21	.0033	70	.3910	51	.0708
47	.3804	44	.1947	36	.0523	22	.0040	71	.4088	52	.0770
48	.4039	45	.2106	37	.0583	23	.0047	72	.4268	53	.0837
49	.4276	46	.2271	38	.0649	24	.0055	73	.4450	54	.0907

TABLA A.2 (continuación)

T	P	T	P	T	P	T	P	T	P	T	P
n = 18		n = 19		n = 19		n = 20		n = 20		n = 21	
55	.0982	30	.0036	79	.2706	48	.0164	97	.3921	61	.0298
56	.1061	31	.0041	80	.2839	49	.0181	98	.4062	62	.0323
57	.1144	32	.0047	81	.2974	50	.0200	99	.4204	63	.0351
58	.1231	33	.0054	82	.3113	51	.0220	100	.4347	64	.0380
59	.1323	34	.0062	83	.3254	52	.0242	101	.4492	65	.0411
60	.1419	35	.0070	84	.3397	53	.0266	102	.4636	66	.0444
61	.1519	36	.0080	85	.3543	54	.0291	103	.4782	*67	.0479
62	.1624	37	.0090	86	.3690	55	.0319	104	.4927	68	.0516
63	.1733	38	.0102	87	.3840	56	.0348	105	.5073	69	.0555
64	.1846	39	.0115	88	.3991	57	.0379	n = 21		70	.0597
65	.1964	40	.0129	89	.4144	58	.0413	14	.0001	71	.0640
66	.2086	41	.0145	90	.4298	59	.0448	20	.0002	72	.0686
67	.2211	42	.0162	91	.4453	*60	.0487	22	.0003	73	.0735
68	.2341	43	.0180	92	.4609	61	.0527	24	.0004	74	.0786
69	.2475	44	.0201	93	.4765	62	.0570	26	.0005	75	.0839
70	.2613	45	.0223	94	.4922	63	.0615	27	.0006	76	.0895
71	.2754	46	.0247	95	.5078	64	.0664	28	.0007	77	.0953
72	.2899	47	.0273			65	.0715	29	.0008	78	.1015
73	.3047	48	.0301	n = 20		66	.0768	30	.0009	79	.1078
74	.3198	49	.0331	11	.0001	67	.0825	31	.0011	80	.1145
75	.3353	50	.0364	16	.0002	68	.0884	32	.0012	81	.1214
76	.3509	51	.0399	19	.0003	69	.0947	33	.0014	82	.1286
77	.3669	52	.0437	20	.0004	70	.1012	34	.0016	83	.1361
78	.3830	*53	.0478	22	.0005	71	.1081	35	.0019	84	.1439
79	.3994	54	.0521	23	.0006	72	.1153	36	.0021	85	.1519
80	.4159	55	.0567	24	.0007	73	.1227	37	.0024	86	.1602
81	.4325	56	.0616	25	.0008	74	.1305	38	.0028	87	.1688
82	.4493	57	.0668	26	.0010	75	.1387	39	.0031	88	.1777
83	.4661	58	.0723	27	.0012	76	.1471	40	.0036	89	.1869
84	.4831	59	.0782	28	.0014	77	.1559	41	.0040	90	.1963
85	.5000	60	.0844	29	.0016	78	.1650	42	.0045	91	.2060
		61	.0909	30	.0018	79	.1744	43	.0051	92	.2160
n = 19		62	.0978	31	.0021	80	.1841	44	.0057	93	.2262
9	.0001	63	.1051	32	.0024	81	.1942	45	.0063	94	.2367
13	.0002	64	.1127	33	.0028	82	.2045	46	.0071	95	.2474
15	.0003	65	.1206	34	.0032	83	.2152	47	.0079	96	.2584
17	.0004	66	.1290	35	.0036	84	.2262	48	.0088	97	.2696
18	.0005	67	.1377	36	.0042	85	.2375	49	.0097	98	.2810
19	.0006	68	.1467	37	.0047	86	.2490	50	.0108	99	.2927
20	.0007	69	.1562	38	.0053	87	.2608	51	.0119	100	.3046
21	.0008	70	.1660	39	.0060	88	.2729	52	.0132	101	.3166
22	.0010	71	.1762	40	.0068	89	.2853	53	.0145	102	.3289
23	.0012	72	.1868	41	.0077	90	.2979	54	.0160	103	.3414
24	.0014	73	.1977	42	.0086	91	.3108	55	.0175	104	.3540
25	.0017	74	.2090	43	.0096	92	.3238	56	.0192	105	.3667
26	.0020	75	.2207	44	.0107	93	.3371	57	.0210	106	.3796
27	.0023	76	.2327	45	.0120	94	.3506	58	.0230	107	.3927
28	.0027	77	.2450	46	.0133	95	.3643	59	.0251	108	.4058
29	.0031	78	.2576	47	.0148	96	.3781	60	.0273	109	.4191

TABLA A.2 (continuación)

T	*P*	*T*	*P*	*T*	*P*	*T*	*P*	*T*	*P*	*T*	*P*
***n* = 21**		***n* = 22**		***n* = 22**		***n* = 23**		***n* = 23**		***n* = 24**	
110	.4324	67	.0271	116	.3751	68	.0163	117	.2700	62	.0053
111	.4459	68	.0293	117	.3873	69	.0177	118	.2800	63	.0058
112	.4593	69	.0317	118	.3995	70	.0192	119	.2902	64	.0063
113	.4729	70	.0342	119	.4119	71	.0208	120	.3005	65	.0069
114	.4864	71	.0369	120	.4243	72	.0224	121	.3110	66	.0075
115	.5000	72	.0397	121	.4368	73	.0242	122	.3217	67	.0082
		73	.0427	122	.4494	74	.0261	123	.3325	68	.0089
		74	.0459	123	.4620	75	.0281	124	.3434	69	.0097
***n* = 22**		*75	.0492	124	.4746	76	.0303	125	.3545	70	.0106
18	.0001	76	.0527	125	.4873	77	.0325	126	.3657	71	.0115
23	.0002	77	.0564	126	.5000	78	.0349	127	.3770	72	.0124
26	.0003	78	.0603			79	.0274	128	.3884	73	.0135
29	.0004	79	.0644	***n* = 23**		80	.0401	129	.3999	74	.0146
30	.0005	80	.0687	21	.0001	81	.0429	130	.4115	75	.0157
32	.0006	81	.0733	28	.0002	82	.0459	131	.4231	76	.0170
33	.0007	82	.0780	31	.0003	*83	.0490	132	.4348	77	.0183
34	.0008	83	.0829	33	.0004	84	.0523	133	.4466	78	.0197
35	.0010	84	.0881	35	.0005	85	.0557	134	.4584	79	.0212
36	.0011	85	.0935	36	.0006	86	.0593	135	.4703	80	.0228
37	.0013	86	.0991	38	.0007	87	.0631	136	.4822	81	.0245
38	.0014	87	.1050	39	.0008	88	.0671	137	.4941	82	.0263
39	.0016	88	.1111	40	.0009	89	.0712	138	.5060	83	.0282
40	.0018	89	.1174	41	.0011	90	.0755			84	.0302
41	.0021	90	.1240	42	.0012	91	.0801	***n* = 24**		85	.0323
42	.0023	91	.1308	43	.0014	92	.0848	25	.0001	86	.0346
43	.0026	92	.1378	44	.0015	93	.0897	32	.0002	87	.0369
44	.0030	93	.1451	45	.0017	94	.0948	36	.0003	88	.0394
45	.0033	94	.1527	46	.0019	95	.1001	38	.0004	89	.0420
46	.0037	95	.1604	47	.0022	96	.1056	40	.0005	90	.0447
47	.0042	96	.1685	48	.0024	97	.1113	42	.0006	*91	.0475
48	.0046	97	.1767	49	.0027	98	.1172	43	.0007	92	.0505
49	.0052	98	.1853	50	.0030	99	.1234	44	.0008	93	.0537
50	.0057	99	.1940	51	.0034	100	.1297	45	.0009	94	.0570
51	.0064	100	.2030	52	.0037	101	.1363	46	.0010	95	.0604
52	.0070	101	.2122	53	.0041	102	.1431	47	.0011	96	.0640
53	.0078	102	.2217	54	.0046	103	.1501	48	.0013	97	.0678
54	.0086	103	.2314	55	.0051	104	.1573	49	.0014	98	.0717
55	.0095	104	.2413	56	.0056	105	.1647	50	.0016	99	.0758
56	.0104	105	.2514	57	.0061	106	.1723	51	.0018	100	.0800
57	.0115	106	.2618	58	.0068	107	.1802	52	.0020	101	.0844
58	.0126	107	.2723	59	.0074	108	.1883	53	.0022	102	.0890
59	.0138	108	.2830	60	.0082	109	.1965	54	.0024	103	.0938
60	.0151	109	.2940	61	.0089	110	.2050	55	.0027	104	.0987
61	.0164	110	.3051	62	.0098	111	.2137	56	.0029	105	.1038
62	.0179	111	.3164	63	.0107	112	.2226	57	.0033	106	.1091
63	.0195	112	.3278	64	.0117	113	.2317	58	.0036	107	.1146
64	.0212	113	.3394	65	.0127	114	.2410	59	.0040	108	.1203
65	.0231	114	.3512	66	.0138	115	.2505	60	.0044	109	.1261
66	.0250	115	.3631	67	.0150	116	.2601	61	.0048	110	.1322

TABLA A.2 (continuación)

T	P	T	P	T	P	T	P	T	P	T	P
$n = 24$		$n = 25$		$n = 25$		$n = 25$		$n = 26$		$n = 26$	
111	.1384	50	.0008	99	.0452	148	.3556	81	.0076	130	.1289
112	.1448	51	.0009	*100	.0479	149	.3655	82	.0082	131	.1344
113	.1515	52	.0010	101	.0507	150	.3755	83	.0088	132	.1399
114	.1583	53	.0011	102	.0537	151	.3856	84	.0095	133	.1457
115	.1653	54	.0013	103	.0567	152	.3957	85	.0102	134	.1516
116	.1724	55	.0014	104	.0600	153	.4060	86	.0110	135	.1576
117	.1798	56	.0015	105	.0633	154	.4163	87	.0118	136	.1638
118	.1874	57	.0017	106	.0668	155	.4266	88	.0127	137	.1702
119	.1951	58	.0019	107	.0705	156	.4370	89	.0136	138	.1767
120	.2031	59	.0021	108	.0742	157	.4474	90	.0146	139	.1833
121	.2112	60	.0023	109	.0782	158	.4579	91	.0156	140	.1901
122	.2195	61	.0025	110	.0822	159	.4684	92	.0167	141	.1970
123	.2279	62	.0028	111	.0865	160	.4789	93	.0179	142	.2041
124	.2366	63	.0031	112	.0909	161	.4895	94	.0191	143	.2114
125	.2454	64	.0034	113	.0954	162	.5000	95	.0204	144	.2187
126	.2544	65	.0037	114	.1001			96	.0217	145	.2262
127	.2635	66	.0040	115	.1050	$n = 26$		97	.0232	146	.2339
128	.2728	67	.0044	116	.1100	34	.0001	98	.0247	147	.2417
129	.2823	68	.0048	117	.1152	42	.0002	99	.0263	148	.2496
130	.2919	69	.0053	118	.1205	46	.0003	100	.0279	149	.2577
131	.3017	70	.0057	119	.1261	49	.0004	101	.0297	150	.2658
132	.3115	71	.0062	120	.1317	51	.0005	102	.0315	151	.2741
133	.3216	72	.0068	121	.1376	53	.0006	103	.0334	152	.2826
134	.3317	73	.0074	122	.1436	55	.0007	104	.0355	153	.2911
135	.3420	74	.0080	123	.1498	56	.0008	105	.0376	154	.2998
136	.3524	75	.0087	124	.1562	57	.0009	106	.0398	155	.3085
137	.3629	76	.0094	125	.1627	58	.0010	107	.0421	156	.3174
138	.3735	77	.0101	126	.1694	59	.0011	108	.0445	157	.3264
139	.3841	78	.0110	127	.1763	60	.0012	109	.0470	158	.3355
140	.3949	79	.0118	128	.1833	61	.0013	*110	.0497	159	.3447
141	.4058	80	.0128	129	.1905	62	.0015	111	.0524	160	.3539
142	.4167	81	.0137	130	.1979	63	.0016	112	.0553	161	.3633
143	.4277	82	.0148	131	.2054	64	.0018	113	.0582	162	.3727
144	.4387	83	.0159	132	.2131	65	.0020	114	.0613	163	.3822
145	.4498	84	.0171	133	.2209	66	.0021	115	.0646	164	.3918
146	.4609	85	.0183	134	.2289	67	.0023	116	.0679	165	.4014
147	.4721	86	.0197	135	.2371	68	.0026	117	.0714	166	.4111
148	.4832	87	.0211	136	.2454	69	.0028	118	.0750	167	.4208
149	.4944	88	.0226	137	.2539	70	.0031	119	.0787	168	.4306
150	.5056	89	.0241	138	.2625	71	.0033	120	.0825	169	.4405
		90	.0258	139	.2712	72	.0036	121	.0865	170	.4503
$n = 25$		91	.0275	140	.2801	73	.0040	122	.0907	171	.4602
29	.0001	92	.0294	141	.2891	74	.0043	123	.0950	172	.4702
37	.0002	93	.0313	142	.2983	75	.0047	124	.0994	173	.4801
41	.0003	94	.0334	143	.3075	76	.0051	125	.1039	174	.4900
43	.0004	95	.0355	144	.3169	77	.0055	126	.1086	175	.5000
45	.0005	96	.0377	145	.3264	78	.0060	127	.1135		
47	.0006	97	.0401	146	.3360	79	.0065	128	.1185		
48	.0007	98	.0426	147	.3458	80	.0070	129	.1236		

TABLA A.2 (continuación)

T	*P*	*T*	*P*	*T*	*P*	*T*	*P*	*T*	*P*	*T*	*P*
$n = 27$		$n = 27$		$n = 27$		$n = 28$		$n = 28$		$n = 28$	
39	.0001	105	.0218	154	.2066	74	.0012	123	.0349	172	.2466
47	.0002	106	.0231	155	.2135	75	.0013	124	.0368	173	.2538
52	.0003	107	.0246	156	.2205	76	.0015	125	.0387	174	.2611
55	.0004	108	.0260	157	.2277	77	.0016	126	.0407	175	.2685
57	.0005	109	.0276	158	.2349	78	.0017	127	.0428	176	.2759
59	.0006	110	.0292	159	.2423	79	.0019	128	.0450	177	.2835
61	.0007	111	.0309	160	.2498	80	.0020	129	.0473	178	.2912
62	.0008	112	.0327	161	.2574	81	.0022	*130	.0496	179	.2990
64	.0009	113	.0346	162	.2652	82	.0024	131	.0521	180	.3068
65	.0010	114	.0366	163	.2730	83	.0026	132	.0546	181	.3148
66	.0011	115	.0386	164	.2810	84	.0028	133	.0573	182	.3228
67	.0012	116	.0407	165	.2890	85	.0030	134	.0600	183	.3309
68	.0014	117	.0430	166	.2972	86	.0033	135	.0628	184	.3391
69	.0015	118	.0453	167	.3055	87	.0035	136	.0657	185	.3474
70	.0016	*119	.0477	168	.3138	88	.0038	137	.0688	186	.3557
71	.0018	120	.0502	169	.3223	89	.0041	138	.0719	187	.3641
72	.0019	121	.0528	170	.3308	90	.0044	139	.0751	188	.3725
73	.0021	122	.0555	171	.3395	91	.0048	140	.0785	189	.3811
74	.0023	123	.0583	172	.3482	92	.0051	141	.0819	190	.3896
75	.0025	124	.0613	173	.3570	93	.0055	142	.0855	191	.3983
76	.0027	125	.0643	174	.3659	94	.0059	143	.0891	192	.4070
77	.0030	126	.0674	175	.3748	95	.0064	144	.0929	193	.4157
78	.0032	127	.0707	176	.3838	96	.0068	145	.0968	194	.4245
79	.0035	128	.0741	177	.3929	97	.0073	146	.1008	195	.4333
80	.0038	129	.0776	178	.4020	98	.0078	147	.1049	196	.4421
81	.0041	130	.0812	179	.4112	99	.0084	148	.1091	197	.4510
82	.0044	131	.0849	180	.4204	100	.0089	149	.1135	198	.4598
83	.0048	132	.0888	181	.4297	101	.0096	150	.1180	199	.4687
84	.0052	133	.0927	182	.4390	102	.0102	151	.1225	200	.4777
85	.0056	134	.0968	183	.4483	103	.0109	152	.1273	201	.4866
86	.0060	135	.1010	184	.4577	104	.0116	153	.1321	202	.4955
87	.0065	136	.1054	185	.4670	105	.0124	154	.1370	203	.5045
88	.0070	137	.1099	186	.4764	106	.0132	155	.1421		
89	.0075	138	.1145	187	.4859	107	.0140	156	.1473	$n = 29$	
90	.0081	139	.1193	188	.4953	108	.0149	157	.1526	50	.0001
91	.0087	140	.1242	189	.5047	109	.0159	158	.1580	59	.0002
92	.0093	141	.1292			110	.0168	159	.1636	65	.0003
93	.0100	142	.1343	$n = 28$		111	.0179	160	.1693	68	.0004
94	.0107	143	.1396	44	.0001	112	.0190	161	.1751	71	.0005
95	.0115	144	.1450	53	.0002	113	.0201	162	.1810	73	.0006
96	.0123	145	.1506	58	.0003	114	.0213	163	.1870	75	.0007
97	.0131	146	.1563	61	.0004	115	.0226	164	.1932	76	.0008
98	.0140	147	.1621	64	.0005	116	.0239	165	.1995	78	.0009
99	.0150	148	.1681	66	.0006	117	.0252	166	.2059	79	.0010
100	.0159	149	.1742	68	.0007	118	.0267	167	.2124	80	.0011
101	.0170	150	.1804	69	.0008	119	.0282	168	.2190	81	.0012
102	.0181	151	.1868	70	.0009	120	.0298	169	.2257	82	.0013
103	.0193	152	.1932	72	.0010	121	.0314	170	.2326	83	.0014
104	.0205	153	.1999	73	.0011	122	.0331	171	.2395	84	.0015

TABLA A.2 (continuación)

T	P	T	P	T	P	T	P	T	P	T	P
$n = 29$		$n = 29$		$n = 29$		$n = 30$		$n = 30$		$n = 30$	
85	.0016	134	.0362	183	.2340	90	.0013	139	.0275	188	.1854
86	.0018	135	.0380	184	.2406	91	.0014	140	.0288	189	.1909
87	.0019	136	.0399	185	.2473	92	.0015	141	.0303	190	.1965
88	.0021	137	.0418	186	.2541	93	.0016	142	.0318	191	.2022
89	.0022	138	.0439	187	.2611	94	.0017	143	.0333	192	.2081
90	.0024	139	.0460	188	.2681	95	.0019	144	.0349	193	.2140
91	.0026	*140	.0482	189	.2752	96	.0020	145	.0366	194	.2200
92	.0028	141	.0504	190	.2824	97	.0022	146	.0384	195	.2261
93	.0030	142	.0528	191	.2896	98	.0023	147	.0402	196	.2323
94	.0032	143	.0552	192	.2970	99	.0025	148	.0420	197	.2386
95	.0035	144	.0577	193	.3044	100	.0027	149	.0440	198	.2449
96	.0037	145	.0603	194	.3120	101	.0029	150	.0460	199	.2514
97	.0040	146	.0630	195	.3196	102	.0031	•151	.0481	200	.2579
98	.0043	147	.0658	196	.3272	103	.0033	152	.0502	201	.2646
99	.0046	148	.0687	197	.3350	104	.0036	153	.0524	202	.2713
100	.0049	149	.0716	198	.3428	105	.0038	154	.0547	203	.2781
101	.0053	150	.0747	199	.3507	106	.0041	155	.0571	204	.2849
102	.0057	151	.0778	200	.3586	107	.0044	156	.0595	205	.2919
103	.0061	152	.0811	201	.3666	108	.0047	157	.0621	206	.2989
104	.0065	153	.0844	202	.3747	109	.0050	158	.0647	207	.3060
105	.0069	154	.0879	203	.3828	110	.0053	159	.0674	208	.3132
106	.0074	155	.0914	204	.3909	111	.0057	160	.0701	209	.3204
107	.0079	156	.0951	205	.3991	112	.0060	161	.0730	210	.3277
108	.0084	157	.0988	206	.4074	113	.0064	162	.0759	211	.3351
109	.0089	158	.1027	207	.4157	114	.0068	163	.0790	212	.3425
110	.0095	159	.1066	208	.4240	115	.0073	164	.0821	213	.3500
111	.0101	160	.1107	209	.4324	116	.0077	165	.0853	214	.3576
112	.0108	161	.1149	210	.4408	117	.0082	166	.0886	215	.3652
113	.0115	162	.1191	211	.4492	118	.0087	167	.0920	216	.3728
114	.0122	163	.1235	212	.4576	119	.0093	168	.0955	217	.3805
115	.0129	164	.1280	213	.4661	120	.0098	169	.0990	218	.3883
116	.0137	165	.1326	214	.4745	121	.0104	170	.1027	219	.3961
117	.0145	166	.1373	215	.4830	122	.0110	171	.1065	220	.4039
118	.0154	167	.1421	216	.4915	123	.0117	172	.1103	221	.4118
119	.0163	168	.1471	217	.5000	124	.0124	173	.1143	222	.4197
120	.0173	169	.1521			125	.0131	174	.1183	223	.4276
121	.0183	170	.1572	$n = 30$		126	.0139	175	.1225	224	.4356
122	.0193	171	.1625	55	.0001	127	.0147	176	.1267	225	.4436
123	.0204	172	.1679	66	.0002	128	.0155	177	.1311	226	.4516
124	.0216	173	.1733	71	.0003	129	.0164	178	.1355	227	.4596
125	.0228	174	.1789	75	.0004	130	.0173	179	.1400	228	.4677
126	.0240	175	.1846	78	.0005	131	.0182	180	.1447	229	.4758
127	.0253	176	.1904	80	.0006	132	.0192	181	.1494	230	.4838
128	.0267	177	.1963	82	.0007	133	.0202	182	.1543	231	.4919
129	.0281	178	.2023	84	.0008	134	.0213	183	.1592	232	.5000
130	.0296	179	.2085	85	.0009	135	.0225	184	.1642		
131	.0311	180	.2147	87	.0010	136	.0236	185	.1694		
132	.0328	181	.2210	88	.0011	137	.0249	186	.1746		
133	.0344	182	.2274	89	.0012	138	.0261	187	.1799		

Source: Frank Wilcoxon, S. K. Katti, and Roberta A. Wilcox, "Critical Values and Probability Levels for the Wilcoxon Rank Sum Test and the Wilcoxon Signed Rank Test." Originally prepared and distributed by Lederle Laboratories Division, American Cyanamid Company, Pearl River, New York, in cooperation with the Department of Statistics, The Florida State University, Tallahassee, Florida. Revised October 1968. Copyright 1963 by the American Cyanamid Company and The Florida State University. Reproduced by permission of S. K. Katti.

BIBLIOGRAFÍA

Agresti, Alan; Finlay, Barbara.

Statistical Methods for the Social Sciences.

Prentice Hall, 1997. 706 p.

Conover, W.J.

Practical Nonparametric Statistics

Universidad de California, 1971. 462 p.

Corona Callejas, Leticia E.

Tesis de licenciatura

Regresión Dinámica y la Relación entre precios de crudos ligeros y pesados.

Universidad Nacional Autónoma de México, 2003.

Draper, Norman R; Smith, Harry.

Applied Regression Analysis.

2da edición. Universidad de Michigan, 1981. 709 p.

Eck, David; Ryan, Jim.

Regression Lines

National Science Foundation DUE-9950473

http://math.hws.edu/javamath/ryan/Regression.html

Estadística: Nociones Generales

http://www.liccom.edu.uy/bedelia/cursos/metodos/material/estadistica/estadistica.html

García Salinero, Julia

El estudio de las variables

http://www.nureinvestigacion.es/FICHEROS_ADMINISTRADOR/F_METODOLOGICA/formacion%2013%5B1%5D.para%20pdf.pdf

Servy. Elsa; Garcia, María del Carmen; Paccapelo, Valeria

REGRESIÓN NO PARAMÉTRICA: UNA APLICACIÓN

Instituto de Investigaciones Teóricas y Aplicadas, de la Escuela de Estadística, Universidad Nacional de Rosario, Argentina, 2007.

http://www.fcecon.unr.edu.ar/investigacion/jornadas/archivos/servyregresion.PDF

Walters, Elizabeth J; Morrell, Christopher H. and Auer, Richard E.

An Investigation of the Median-Median Method of Linear Regression

Loyola College of Maryland; Journal of Statistics Education *Volume 14, Number 2 (2006).*

http://www.amstat.org/publications/jse/v14n2/morrell.html

Wayne W, Daniel.

Applied Nonparametric Statistics.

Cengage Learning, 2000. 635 p.