

**EL EFECTO DE LA HETEROCEDASTICIDAD EN LA  
CONSTRUCCIÓN DE LOS ESTIMADORES DEL MODELO  
DE REGRESIÓN LINEAL**

**MARTHA EUGENIA SOTELO MERA**

Seminario de grado presentado como requisito parcial para optar al título de  
Licenciada en Educación con Especialidad en Matemáticas

**AURA LUCÍA STERLING LÓPEZ**

Seminario de grado presentado como requisito parcial para optar al título de  
Matemática

**Director**

**Mg. YILTON OVIRNE RIASCOS FORERO**

**UNIVERSIDAD DEL CAUCA**

**FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA  
EDUCACIÓN**

**DEPARTAMENTO DE MATEMÁTICAS**

**POPAYÁN**

**2003**

# CONTENIDO

<b>INTRODUCCIÓN</b>	<b>1</b>
<b>JUSTIFICACIÓN Y PROPÓSITOS</b>	<b>2</b>
<b>1. PRELIMINARES</b>	<b>3</b>
1.1. DISTRIBUCIONES DE PROBABILIDAD MULTIVARIANTE . . . . .	4
Función de Probabilidad Conjunta . . . . .	4
Función de Probabilidad Marginal . . . . .	6
Función de Probabilidad Condicional . . . . .	7
Función de Distribución Condicional Conjunta . . . . .	8
Variables Aleatorias Independientes . . . . .	9
El Valor Esperado de una Función de Variables Aleatorias . . . . .	10
Valor Esperado Condicional . . . . .	11
Covarianza de dos Variables Aleatorias . . . . .	12
Distribución Normal $n$ - Dimensional . . . . .	13
Distribuciones muestrales y el Teorema del Límite central . . . . .	13
Distribuciones Muestrales Relacionadas Con La Distribución Normal . . . .	14
1.2. ESTIMACIÓN . . . . .	18
Propiedades de los Estimadores Puntuales . . . . .	19
Evaluación de la Bondad de un Estimador Puntual . . . . .	19
Suficiencia Mínima y Estimadores insesgados de Mínima Varianza . . . . .	22
Intervalos de Confianza . . . . .	23
Intervalos de Confianza con Muestras Grandes . . . . .	24

Intervalos de Confianza para $\mu, \mu_1 - \mu_2$ . . . . .	25
Intervalos de Confianza para $\sigma^2$ . . . . .	26
Método de Máxima Verosimilitud . . . . .	27
<b>1.3. PRUEBAS DE HIPÓTESIS</b> . . . . .	<b>28</b>
Pruebas Comunes Con Muestras Grandes . . . . .	30
Error de Tipo II y Determinación del Tamaño de la Muestra . . . . .	31
Niveles De Significación . . . . .	32
Pruebas De Hipótesis Referentes a varianzas . . . . .	33
La potencia de las pruebas . . . . .	34
<b>2. EL MODELO DE REGRESIÓN LINEAL</b>	<b>35</b>
2.1. MODELO DE REGRESIÓN LINEAL SIMPLE . . . . .	39
2.2. EL MODELO DE REGRESIÓN LINEAL GENERAL . . . . .	45
Método de los Mínimos cuadrados . . . . .	45
Propiedades de los estimadores de mínimos cuadrados . . . . .	46
Inferencias para los parámetros $\beta_i$ . . . . .	48
Técnica del análisis de varianza . . . . .	49
Prueba de hipótesis para $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$ . . . . .	50
Análisis de Residuos . . . . .	51
<b>3. HETEROCEDASTICIDAD</b>	<b>53</b>
3.1. MÍNIMOS CUADRADOS ORDINARIOS . . . . .	55
3.2. MÍNIMOS CUADRADOS GENERALIZADOS . . . . .	57
Propiedades de los estimadores de $MCG$ . . . . .	58
Estimación de $\sigma^2$ . . . . .	60
3.3. ALGUNAS PRUEBAS DE DETECCIÓN . . . . .	61
Prueba de Bartlett . . . . .	62
Prueba de Cochran . . . . .	63
Prueba de Hartley . . . . .	63
Prueba gráficas . . . . .	64

3.4. CAUSAS MÁS FRECUENTES . . . . .	65
3.5. EFECTOS SOBRE EL MODELO . . . . .	66
3.6. CARACTERIZACIÓN DE LA FORMA DE LA HETEROCEDASTICIDAD	68
3.7. CONTRASTES . . . . .	70
Contráste de Goldfeldt y Quandt . . . . .	70
Contraste de Breusch y Pagan . . . . .	72
Contraste de Glesjer . . . . .	73
Contraste de White . . . . .	74
3.8. CÓMO SE CORRIGE LA HETEROCEDASTICIDAD . . . . .	76
<b>CONCLUSIONES</b>	<b>77</b>
<b>BIBLIOGRAFÍA</b>	<b>80</b>

# INTRODUCCIÓN

En muchas situaciones es necesario representar la realidad utilizando modelos matemáticos con el fin de descubrir la relación existente entre diversos factores y de esta manera hacer predicciones o tomar decisiones convenientes para un determinado propósito. Este es el caso de los modelos estadísticos, que son modelos matemáticos que involucran en su estructura una componente aleatoria (error) que procura representar la presencia del azar en cierto nivel de predicción o en la estimación de la magnitud de algún parámetro o efecto. Estos modelos se construyen experimentalmente a partir de la observación de una parcela de la realidad de la cual se extrae un conjunto finito de información en forma de datos. Las discrepancias entre los resultados obtenidos por el modelo y los que se espera observar en la realidad están sujetas a un conjunto de supuestos que se establecen al momento de la construcción del modelo. El incumplimiento de estos supuestos ocasiona diversos trastornos en el mismo, trastornos que se evidencian a través de sus parámetros mediante análisis de las componentes aleatorias (errores) y de la correlación de las variables de predicción.

Una de las hipótesis establecidas en la construcción de los modelos estadísticos, específicamente en el Modelo de Regresión Lineal (MRL), es que las varianzas de los errores sean constantes. El incumplimiento de este supuesto, denominado heterocedasticidad, y su efecto en la construcción de los estimadores de los parámetros del modelo, es el objetivo central de este seminario. Algunos autores escriben: "...el investigador debe reconocer que las situaciones de autocorrelación y/o heterocedasticidad son la regla y no la excepción, y que debe hacerse siempre un análisis tan exhaustivo como sea posible acerca del gra-

do en que estos problemas están presentes en una aplicación empírica<sup>1</sup>, observación que complementa nuestro interés.

Se presentarán entonces los procesos de estimación de parámetros, las propiedades de los estimadores y los efectos que la heterocedasticidad les ocasiona, en la construcción del denominado *MRL*. También se describirán sus posibles causas y algunos tratamientos para corregirla. Para abordar estos temas se presentarán algunos resultados de probabilidad e inferencia estadística sin llegar a profundizar en tales aspectos.

---

<sup>1</sup>Econometría, A. Novales, 1993,pp169

# JUSTIFICACIÓN Y PROPÓSITOS

El éxito en la aplicación del modelo de regresión lineal en una determinada investigación, la validez de los hallazgos y conclusiones obtenidas, dependen, en gran parte del cumplimiento de cada uno de los supuestos determinados en su construcción, ya que al no cumplirse alguno de estos se pueden perder importantes características de los estimadores de los parámetros; lo cual redundará en errores al momento de tomar decisiones o hacer predicciones.

Por esta razón es conveniente conocer algunos instrumentos y/o métodos que permitan contrastar la plausibilidad de estas suposiciones y en algunos casos valorar el impacto de alejarse de su cumplimiento, especialmente cuando en el MRL no se cumple el supuesto de Homocedasticidad.

Para aproximarnos al estudio del problema nos hemos propuesto cumplir los siguientes objetivos:

1. Estudiar y detallar el proceso de modelación estadística para el *MRL*.
2. Describir matemáticamente la heterocedasticidad en la modelación estadística, sus consecuencias y posibles causas.
3. Presentar los efectos que en la modelación estadística ocasiona la heterocedasticidad al momento construir los estimadores para el *MRL* y sus consecuencias en las predicciones.

# 1. PRELIMINARES

## 1.1. DISTRIBUCIONES DE PROBABILIDAD MULTIVARIANTE

Para poder cumplir con los propósitos establecidos, es aconsejable recordar aspectos estadísticos que tienen gran relación con la problemática a tratar, y en muchos casos son requisito sine qua non para poder abordar algunos desarrollos en la Heterocedasticidad. Se ha decidido obviar las demostraciones de la mayoría de los teoremas por no considerarlo fundamental para esta parte y porque además se encuentran en los textos escolares que abordan estos temas. Estos presupuestos, considerados mas relevantes; sin pretender ser exhaustivos, se presentan a continuación:

### Función de Probabilidad Conjunta

En muchas situaciones resulta de interés medir mas de una característica de algún fenómeno aleatorio, es entonces cuando surge la necesidad de estudiar modelos de probabilidad que contengan mas de una variable aleatoria. Estos modelos se conocen como modelos multivariados.

**Definición 1.1.** Sean  $Y_1, Y_2, \dots, Y_n$ ,  $n$  variables aleatorias discretas, la función de probabilidad conjunta para  $Y_1, Y_2, \dots, Y_n$  es una función  $p$  definida de  $\mathbb{R}^n$  en  $\mathbb{R}$  está dada por

$$p(y_1, y_2, \dots, y_n) = P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n)$$



Esta función de probabilidad conjunta satisface las siguientes condiciones:

1.  $p(y_1, y_2, \dots, y_n) \geq 0$  para todo  $y_1, y_2, \dots, y_n$ .
2.  $\sum_{y_1, y_2, \dots, y_n} p(y_1, y_2, \dots, y_n) = 1$

**Definición 1.2.** Sean  $Y_1, Y_2, \dots, Y_n$ ,  $n$  variables aleatorias continuas, si existe una función definida de  $\mathbb{R}^n$  en  $\mathbb{R}$  que cumple con las siguientes propiedades:

1.  $f(y_1, y_2, \dots, y_n) \geq 0$  para todo  $y_1, y_2, \dots, y_n$
2.  $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(y_1, y_2, \dots, y_n) dy_n dy_{n-1} \dots dy_1 = 1$

entonces a esta función  $f(y_1, y_2, \dots, y_n)$  se le llama función de densidad de probabilidad conjunta.

**Definición 1.3.** Sean  $Y_1, Y_2, \dots, Y_n$ ,  $n$  variables aleatorias discretas, con función de probabilidad  $p(y_1, y_2, \dots, y_n)$ , se define la función de distribución de probabilidad conjunta acumulada  $F(a_1, a_2, \dots, a_n)$  como

$$\begin{aligned} F(a_1, a_2, \dots, a_n) &= P(Y_1 \leq a_1, Y_2 \leq a_2, \dots, Y_n \leq a_n) \\ &= \sum_{y_1 \leq a_1} \dots \sum_{y_n \leq a_n} p(y_1, y_2, \dots, y_n) \end{aligned}$$

Siempre y cuando las sumas sean convergentes.

Si las variables aleatorias  $Y_1, Y_2, \dots, Y_n$  son continuas, con función de densidad de probabilidad conjunta  $f(y_1, y_2, \dots, y_n)$ , su función de distribución conjunta acumulada está dada por la expresión:

$$F(a_1, a_2, \dots, a_n) = \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_n} f(y_1, y_2, \dots, y_n) dy_1 dy_2 \dots dy_n$$

siempre y cuando la integral múltiple exista.

La función de distribución de probabilidad conjunta acumulada tiene las siguientes propiedades:

1.  $F(\infty, \infty, \dots, \infty) = 1$ , esto es; que la función de distribución de probabilidad acumulada en todos los valores de las variables aleatorias.
2.  $F(-\infty, \infty, \dots, \infty) = F(-\infty, y_2, \dots, \infty) = \dots = F(y_1, \infty, \dots, -\infty) = 0$ , que es equivalente a decir que la función de distribución acumulada de  $Y_1, Y_2, \dots, Y_n$ , evaluada en algún valor que no está en el rango donde están definidas las variables, es cero.
3. si  $a_1 \geq b_1, a_2 \geq b_2, \dots, a_n \geq b_n$  se cumple que
 
$$F(a_1, a_2, \dots, a_n) - F(b_1, b_2, \dots, b_n) \geq 0$$

## Función de Probabilidad Marginal

**Definición 1.4.** Sean  $Y_1, Y_2, \dots, Y_n$ , variables aleatorias discretas con función de probabilidad conjunta  $p(y_1, y_2, \dots, y_n)$ . Entonces la función de probabilidad marginal de  $Y_i$ , se define como:

1.  $p_i(y_i) = \sum_{y_1} \dots \sum_{y_{i-1}} \sum_{y_{i+1}} \dots \sum_{y_n} p(y_1, y_2, \dots, y_n)$ .
2. Si  $Y_1, Y_2, \dots, Y_n$  son variables aleatorias continuas con función de densidad de probabilidad conjunta  $f(y_1, y_2, \dots, y_n)$ , se define la función de densidad de probabilidad marginal de  $Y_i$  por

$$f_i(y_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n) dy_1 \dots dy_{i-1} dy_{i+1} \dots dy_n$$

Estos conceptos se pueden generalizar a la función de probabilidad marginal conjunta como sigue:

**Definición 1.5.** Sean  $Y_1, Y_2, \dots, Y_n$  variables aleatorias discretas con función de probabilidad conjunta  $p(y_1, y_2, \dots, y_n)$ , entonces la función de probabilidad marginal conjunta para un número  $k$  de variables  $Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}$ , con  $k < n$  está dada por:

$$p_{Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}}(y_{i_1}, y_{i_2}, \dots, y_{i_k}) = \sum_{B_k} p(y_1, y_2, \dots, y_n)$$

*Siempre y cuando las sumas converjan y donde  $B_k$  es el conjunto de puntos en el rango de  $y_1, y_2, \dots, y_n$  para los que  $Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}$  permanecen fijas.*

*Es decir, que la función de probabilidad marginal conjunta de un subconjunto de  $k$  variables aleatorias de  $Y_1, Y_2, \dots, Y_n$ , es la suma de su función de probabilidad conjunta evaluada en todos los valores del rango de  $Y_1, Y_2, \dots, Y_n$  para los cuales las variables  $Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}$  permanecen fijas, siempre que las sumas sean convergentes.*

*Si las variables aleatorias  $Y_1, Y_2, \dots, Y_n$  son continuas con función de densidad de probabilidad conjunta*

*$f(y_1, y_2, \dots, y_n)$ , entonces la función de densidad de probabilidad marginal conjunta para un subconjunto de variables  $Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}$ , con  $k < n$  está dada por:*

$$f_{Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}}(y_{i_1}, y_{i_2}, \dots, y_{i_k}) = \int_{B_k} f(y_1, y_2, \dots, y_n) d_{y^{(n-k)}}$$

*Siempre que la integral múltiple exista, donde  $B_k$  es el conjunto de puntos en el rango de  $y_1, y_2, \dots, y_n$  para los que  $Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}$  permanecen fijas y  $y^{n-k}$  es un subconjunto de  $n - k$  variables de  $Y_1, Y_2, \dots, Y_n$  en las que no se incluyen las variables  $Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}$ .*

*Esto es, que la función de densidad de probabilidad marginal conjunta de un subconjunto de  $k$  variables aleatorias de  $Y_1, Y_2, \dots, Y_n$ , es la integral múltiple de su función de densidad de probabilidad conjunta evaluada en todos los valores del rango de  $Y_1, Y_2, \dots, Y_n$  para los cuales las variables  $Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}$  permanecen fijas, siempre que las integrales existan.*

## **Función de Distribución Condicional**

**Definición 1.6.** *Sean  $Y_1, Y_2, \dots, Y_n$  variables aleatorias discretas que tienen función de probabilidad conjunta  $p(y_1, y_2, \dots, y_n)$ , se define la función de probabilidad condicional de*

$Y_i$  dado que

$Y_1 = y_1, \dots, Y_{i-1} = y_{i-1}, Y_{i+1} = y_{i+1}, \dots, Y_n = y_n$ , como

$$p(y_i/y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n) = \frac{p(y_1, y_2, \dots, y_n)}{p(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)}$$

donde  $p(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n) > 0$ , es la función de probabilidad marginal conjunta de las variables

$Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n$

Para el caso en que las variables aleatorias sean continuas con función de densidad de probabilidad conjunta  $f(y_1, y_2, \dots, y_n)$  se define la función de densidad condicional de  $Y_i$  dado que

$Y_1 = y_1, \dots, Y_{i-1} = y_{i-1}, Y_{i+1} = y_{i+1}, \dots, Y_n = y_n$ , para  $i = 1, 2, \dots, n$ , como

$$f(y_i/y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n) = \begin{cases} \frac{f(y_1, y_2, \dots, y_n)}{f(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)} & \text{para } f(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n) > 0 \\ 0 & \text{en otro punto} \end{cases}$$

Donde  $f(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$  es la función de probabilidad marginal conjunta de las variables  $Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n$

## **Función de Distribución Condicional Conjunta**

Sean  $Y_1, Y_2, \dots, Y_n$  variables aleatorias discretas con función de probabilidad conjunta  $p(y_1, y_2, \dots, y_n)$ , la función de probabilidad condicional de un conjunto de variables  $Y_{i1}, Y_{i2}, \dots, Y_{ik}$  con  $k < n$  dado que  $Y_{j1} = y_{j1}, \dots, Y_{jn} = y_{jn}$  con  $i \neq j$  se define como:

$$p(y_{i1}, y_{i2}, \dots, y_{ik}/Y_{j1} = y_{j1}, \dots, Y_{jn} = y_{jn}) = \frac{p(y_1, y_2, \dots, y_n)}{p(y_{j1}, \dots, y_{jn})}$$

Donde  $p(y_{j1}, \dots, y_{jn}) > 0$  y es la función de probabilidad marginal conjunta de las variables  $Y_{j1}, \dots, Y_{jn}$ .

Si las variables aleatorias son continuas con función de densidad de probabilidad  $f(y_1, y_2, \dots, y_n)$  la función de densidad de probabilidad condicional de un conjunto de variables  $Y_{i1}, Y_{i2}, \dots, Y_{ik}$  con  $k < n$  dado que  $Y_{j1} = y_{j1}, \dots, Y_{jn} = y_{jn}$  se define como:

$$f(y_{i1}, y_{i2}, \dots, y_{ik} / Y_{j1} = y_{j1}, \dots, Y_{jn} = y_{jn}) = \begin{cases} \frac{f(y_1, y_2, \dots, y_n)}{f(y_{j1}, \dots, y_{jn})} & \text{para } f(y_{j1}, \dots, y_{jn}) > 0 \\ 0 & \text{en cualquier otro punto} \end{cases}$$

Si existen las funciones de densidad de probabilidad conjunta de  $Y_1, Y_2, \dots, Y_n$  y  $Y_{j1}, \dots, Y_{jn}$ .

## VARIABLES ALEATORIAS INDEPENDIENTES

Las variables aleatorias  $Y_1, Y_2, \dots, Y_n$  son mutuamente independientes si los eventos  $Y_1 \leq y_1, Y_2 \leq y_2, \dots, Y_n \leq y_n$  son mutuamente independientes para cada  $n$ -úpla de valores  $(y_1, y_2, \dots, y_n)$ . Esto significa que la distribución de probabilidad de los valores de  $Y_i$  no es afectada por la distribución de probabilidad de los valores de las otras variables. Este concepto se puede expresar como:

**Definición 1.7.** Sean  $Y_1, Y_2, \dots, Y_n$ ,  $n$  variables aleatorias, que tienen funciones de distribución marginal  $F_1(y_1), F_2(y_2), \dots, F_n(y_n)$  respectivamente, y tienen función de distribución de probabilidad conjunta  $F(y_1, y_2, \dots, y_n)$ . Se dice que  $Y_1, Y_2, \dots, Y_n$  son independientes si

$$F(y_1, y_2, \dots, y_n) = F_1(y_1) \cdot F_2(y_2) \cdot \dots \cdot F_n(y_n)$$

Si  $Y_1, Y_2, \dots, Y_n$  son variables aleatorias discretas que tienen función de distribución de probabilidad conjunta  $p(y_1, y_2, \dots, y_n)$  y con respectivas funciones de probabilidad marginal  $p_1(y_1), p_2(y_2), \dots, p_n(y_n)$ , entonces las variables aleatorias son independientes si y sólo si:

$$p(y_1, y_2, \dots, y_n) = p_1(y_1) \cdot p_2(y_2) \cdot \dots \cdot p_n(y_n)$$

para todos los puntos en el rango donde están definidas las variables aleatorias.

Si  $Y_1, Y_2, \dots, Y_n$  son variables aleatorias continuas que tienen función de densidad de probabilidad conjunta  $f(y_1, y_2, \dots, y_n)$  y funciones de densidad de probabilidad marginales  $f_1(y_1), f_2(y_2), \dots, f_n(y_n)$  respectivamente, entonces las variables aleatorias son independientes si, y sólo si,

$$f(y_1, y_2, \dots, y_n) = f_1(y_1) \cdot f_2(y_2) \cdot \dots \cdot f_n(y_n)$$

para todo  $(y_1, y_2, \dots, y_n) \in \mathbb{R}^n$

Otra forma de determinar si las variables aleatorias  $Y_1, Y_2, \dots, Y_n$  continuas son independientes es de la siguiente forma: sea la función de densidad de probabilidad conjunta  $f(y_1, y_2, \dots, y_n)$  que es positiva si  $a_1 \leq y_1 \leq b_1, a_2 \leq y_2 \leq b_2, \dots, a_n \leq y_n \leq b_n$ , y es cero en otro punto. Entonces las variables aleatorias  $Y_1, Y_2, \dots, Y_n$  son independientes si existen funciones  $g_1(y_1), g_2(y_2), \dots, g_n(y_n)$ , donde  $g_i(y_i)$  es una función que sólo depende de  $y_i$  (para  $i = 1, 2, \dots, n$ ) tales que la función de densidad de probabilidad conjunta de las variables aleatorias se pueda expresar como el producto de las  $n$  funciones  $g_1(y_1), g_2(y_2), \dots, g_n(y_n)$ , es decir:

$$f(y_1, y_2, \dots, y_n) = g_1(y_1) \cdot g_2(y_2) \cdot \dots \cdot g_n(y_n)$$

## El Valor Esperado de una Función de Variables Aleatorias

Se le llama también esperanza matemática, momento de primer orden o un valor promedio o ponderado.

**Definición 1.8.** *Considérese a  $g(Y_1, Y_2, \dots, Y_n)$  como una función de las  $n$  variables aleatorias  $Y_1, Y_2, \dots, Y_n$ :*

1. *Si las variables aleatorias  $Y_1, Y_2, \dots, Y_n$  son discretas con función de probabilidad conjunta  $p(y_1, y_2, \dots, y_n)$ , entonces el valor esperado o Esperanza matemática de  $g(Y_1, Y_2, \dots, Y_n)$  denotado por  $E[g(Y_1, Y_2, \dots, Y_n)]$  y está dado por:*

$$E[g(Y_1, Y_2, \dots, Y_n)] = \sum_{y_1} \cdots \sum_{y_n} g(y_1, y_2, \dots, y_n) p(y_1, y_2, \dots, y_n)$$

*siempre que haya convergencia.*

2. Si las variables aleatorias  $Y_1, Y_2, \dots, Y_n$  son continuas con función de densidad de probabilidad conjunta  $f(y_1, y_2, \dots, y_n)$ , el valor esperado de  $g(Y_1, Y_2, \dots, Y_n)$  se define por:

$$E[g(Y_1, Y_2, \dots, Y_n)] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(y_1, y_2, \dots, y_n) f(y_1, y_2, \dots, y_n) dy_n \dots dy_2 dy_1$$

si existe la integral múltiple.

**Teorema 1.1.** Sea  $g(Y_1, Y_2, \dots, Y_n)$  una función de las  $n$  variables aleatorias  $Y_1, Y_2, \dots, Y_n$ , la esperanza matemática cumple con las siguientes propiedades:

1.  $E[c] = c$ , donde  $c$  es un número real.
2.  $E[cg(Y_1, Y_2, \dots, Y_n)] = cE[g(Y_1, Y_2, \dots, Y_n)]$
3. Sean  $g_1(Y_1, Y_2, \dots, Y_n), g_2(Y_1, Y_2, \dots, Y_n), \dots, g_k(Y_1, Y_2, \dots, Y_n)$  funciones de las variables aleatorias  $Y_1, Y_2, \dots, Y_n$ , tales que  $E[g_i(Y_1, Y_2, \dots, Y_n)]$  existe para cada  $i = 1, 2, \dots, k$ , entonces:

$$E\left[\sum_{i=1}^k g_i(Y_1, Y_2, \dots, Y_n)\right] = \sum_{i=1}^k E[g_i(Y_1, Y_2, \dots, Y_n)]$$

## Valor Esperado Condicional

**Definición 1.9.** Sean  $Y_1, Y_2, \dots, Y_n$ ,  $n$  variables aleatorias, se define el valor esperado de  $Y_1, Y_2, \dots, Y_k$  con  $k < n$  dado que  $Y_{k+1} = y_{k+1}, \dots, Y_n = y_n$  por:

1.  $E[Y_1, Y_2, \dots, Y_k / Y_{k+1} = y_{k+1}, \dots, Y_n = y_n]$   
 $= \sum_{y_1} \dots \sum_{y_k} y_1 \cdot y_2 \cdot \dots \cdot y_k p(y_1 \cdot y_2 \cdot \dots \cdot y_k / y_{k+1}, \dots, y_n)$

si  $Y_1, Y_2, \dots, Y_n$  son variables aleatorias discretas con función de probabilidad conjunta  $p(y_1, y_2, \dots, y_n)$  y donde  $p(y_1 \cdot y_2 \cdot \dots \cdot y_k / y_{k+1}, \dots, y_n)$  es la función de probabilidad condicional conjunta de  $Y_1, Y_2, \dots, Y_k$  dado que  $Y_{k+1} = y_{k+1}, \dots, Y_n = y_n$ .

$$\begin{aligned}
2. \quad & E [Y_1, Y_2, \dots, Y_k / Y_{k+1} = y_{k+1}, \dots, Y_n = y_n] \\
&= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} y_1 \cdot y_2 \cdot \dots \cdot y_k f(y_1 \cdot y_2 \cdot \dots \cdot y_k / y_{k+1}, \dots, y_n) dy_1 \dots dy_k
\end{aligned}$$

Si  $Y_1, Y_2, \dots, Y_n$  son variables aleatorias continuas con función de densidad de probabilidad conjunta  $f(y_1, y_2, \dots, y_n)$ , y donde  $f(y_1 \cdot y_2 \cdot \dots \cdot y_k / y_{k+1}, \dots, y_n)$  es la función de densidad de probabilidad condicional conjunta de  $Y_1, Y_2, \dots, Y_k$  dado que  $Y_{k+1} = y_{k+1}, \dots, Y_n = y_n$ .

Esto quiere decir que el valor esperado condicional de  $Y_1, Y_2, \dots, Y_k$  dado que  $Y_{k+1} = y_{k+1}, \dots, Y_n = y_n$  es la suma (en el caso discreto) o la integral (en el caso continuo) del producto de las variables  $Y_1, Y_2, \dots, Y_k$  por su respectivo función de probabilidad condicional o función de densidad de probabilidad condicional, según sea el caso, sumando o integrando para los posibles valores del rango de  $Y_1, Y_2, \dots, Y_k$

## Covarianza de dos Variables Aleatorias

**Definición 1.10.** La covarianza de dos variables aleatorias  $Y_1, Y_2$  se define como

$$cov(Y_1, Y_2) = E [(Y_1 - \mu_1)(Y_2 - \mu_2)] = \sigma_{12}$$

con  $\mu_1 = E[Y_1]$  y  $\mu_2 = E[Y_2]$  finitos

**Definición 1.11.** El coeficiente de correlación lineal entre dos variables  $Y_1, Y_2$  está dado por

$$\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

donde  $\sigma_1$  y  $\sigma_2$  son las desviaciones estándar de  $Y_1$  y  $Y_2$  respectivamente.

**Teorema 1.2.** Si  $Y_1, Y_2, \dots, Y_n$  y  $X_1, X_2, \dots, X_n$  son variables aleatorias para las cuales  $E[Y_i] = \mu_i$ , y  $E[X_i] = \xi_i$ , y sean las funciones  $U_1 = \sum_{i=1}^n a_i Y_i$   $U_2 = \sum_{i=1}^n b_i X_i$ , donde  $a_i$  y  $b_i$  son constantes, entonces:



1.  $E(U_1) = \sum_{i=1}^n a_i \mu_i$
2.  $V(U_i) = \sum_{i=1}^n a_i^2 V(Y_i) + 2 \sum_{i < j} \sum a_i a_j \text{cov}(Y_i, Y_j)$
3.  $\text{cov}(U_1, U_2) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{cov}(Y_i, X_j)$

## Distribución Normal $n$ - Dimensional

Las  $n$  variables aleatorias  $Y_1, Y_2, \dots, Y_n$  tienen distribución normal  $n$ - dimensional, si su función de densidad conjunta es:

$$= f(y_1, y_2, \dots, y_n) = \frac{e^{\frac{1}{2}(Y-\mu)^T \Sigma^{-1}(Y-\mu)}}{(2\pi)^{n/2} (|\Sigma|)^{1/2}}$$

donde:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \quad y \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix}$$

Donde  $|\Sigma|$  es el determinante de  $\Sigma$ ,  $Y$  es un vector formado por los valores observables de las variables aleatorias  $Y_1, Y_2, \dots, Y_n$ ,  $\mu$  es su correspondiente vector de medias y  $\Sigma$  es la correspondiente matriz de varianzas y covarianzas de  $Y$

## Distribuciones Muestrales y el Teorema del

### Límite Central

Se verán algunas funciones de las variables  $Y_1, Y_2, \dots, Y_n$  observadas en una muestra aleatoria seleccionada de una población, con variables aleatorias independientes e idénticamente distribuidas.

**Definición 1.12.** *‘La totalidad de los elementos en discusión y acerca de los cuales se desea información, se denomina población objetivo.*

En este momento es importante estructurar el concepto de *Muestra Aleatoria (MA)*, utilizando para ello los conceptos anteriormente vistos. Se diseña un experimento y se lleva a cabo, obteniendo una observación  $Y_1$  de la característica medible  $Y$ , se repite el experimento bajo las mismas condiciones obteniendo el valor  $Y_2$ , y se continua el experimento hasta obtener  $n$  observaciones  $Y_1, Y_2, \dots, Y_n$  de la característica  $Y$ . Este procedimiento es a lo que comúnmente se conoce como muestra aleatoria, donde las observaciones muestrales se colectan a través de  $n$  ensayos independientes que ocurren cada vez que se repite el experimento, bajo condiciones idénticas para todos los factores que son controlables además cada una de las observaciones  $Y_1, Y_2, \dots, Y_n$ , es una variable aleatoria, cuya distribución de probabilidad es idéntica a la de la población.

**Definición 1.13.** *Si las variables aleatorias  $Y_1, Y_2, \dots, Y_n$  tienen la misma función (densidad) de probabilidad, que la distribución de la población y su función (distribución) conjunta de probabilidad es igual al producto de las marginales, entonces  $Y_1, Y_2, \dots, Y_n$  forman un conjunto de  $n$  variables aleatorias independientes e idénticamente distribuidas que constituyen una Muestra Aleatoria (MA) de la población.*

**Definición 1.14.** *Sea  $Y_1, Y_2, \dots, Y_n$  una (MA), de una población con densidad  $f(\bullet)$ , está población recibe el nombre de población muestreada.*

**Definición 1.15.** *Un estadístico es cualquier función de las variables aleatorias que se observan en una muestra, de tal manera que está no contiene cantidades desconocidas.*

## Distribuciones Muestrales

### Relacionadas Con La Distribución Normal

Se presentarán las distribuciones muestrales de algunos estadísticos que se utilizan para hacer inferencias con respecto a los parámetros de una distribución Normal.

**Teorema 1.3.** Sea  $Y_1, Y_2, \dots, Y_n$  una muestra aleatoria de tamaño  $n$  de una distribución Normal, con media  $\mu$  y varianza  $\sigma^2$  entonces el estadístico  $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$  tiene distribución Normal con media  $\mu$  y varianza  $\sigma^2/n$ .

**Teorema 1.4.** Sea  $Y_1, Y_2, \dots, Y_n$  una muestra aleatoria de tamaño  $n$  de una distribución normal, con media  $\mu$  y varianza  $\sigma^2$  entonces  $Z_i = (Y_i - \mu)/\sigma$ , con  $i = 1, 2, \dots, n$ , son variables aleatorias normales estándar e independientes, y

$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n \left( \frac{Y_i - \mu}{\sigma} \right)^2$  tiene una distribución  $\chi^2$  con  $n$  grados de libertad.

**Teorema 1.5.** Sea  $Y_1, Y_2, \dots, Y_n$  una muestra aleatoria de tamaño  $n$  de una distribución normal con media  $\mu$  y varianza  $\sigma^2$ , entonces

$\frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2 = \frac{(n-1)S^2}{\sigma^2}$  tiene una distribución  $\chi^2$  con  $(n-1)$  grados de libertad, y  $\bar{y}$  y  $s^2$  son variables aleatorias independientes.

**Definición 1.16.** Sea  $Z$  una variable aleatoria normal estándar y sea  $\chi^2$  una variable aleatoria ji-cuadrada con  $v$  grados de libertad, entonces si  $Z$  y  $\chi^2$  son independientes

$$T = \frac{Z}{\sqrt{\chi^2/v}}$$

tiene una distribución  $t$  con  $v$  grados de libertad.

**Definición 1.17.** Sean  $\chi_1^2$  y  $\chi_2^2$  variables aleatorias ji-cuadrada con  $v_1$  y  $v_2$  grados de libertad respectivamente, entonces si  $\chi_1^2$  y  $\chi_2^2$  son independientes  $F = \frac{\chi_1^2/v_1}{\chi_2^2/v_2}$  tiene una distribución  $F$  con  $v_1$  grados de libertad en numerador y  $v_2$  grados de libertad de en denominador.

El término *Teorema del Límite Central* significa en la teoría de probabilidades cualquier afirmación a cerca de que al cumplirse ciertas condiciones, la función de distribución de una suma de magnitudes aleatorias individualmente pequeñas converge con el crecimiento del número de sumandos hacia una función de distribución normal. "La importancia exclusiva del teorema del Límite Central se debe al hecho de que explica teóricamente la siguiente observación confirmada reiteradamente en la práctica: si el resultado de un experimento aleatorio se determina con un gran número de factores aleatorios y las influencia de cada uno es tan pequeña que puede despreciarse, entonces tal experimento se

aproxima con éxito mediante una distribución normal, siendo escogidas de manera adecuada la esperanza matemática y la varianza” [?].

**Teorema 1.6. (Del Límite Central)** Sean  $Y_1, Y_2, \dots, Y_n$  variables aleatorias independientes distribuidas idénticamente con  $E(Y_i) = \mu$ ,  $V(Y_i) = \sigma^2$  finita entonces la distribución del estadístico:

$$U_n = \sqrt{n} \left( \frac{\bar{Y} - \mu}{\sigma} \right)$$

converge a la distribución normal estándar cuando  $n \rightarrow \infty$ .

Su demostración se realizará solo para el caso en el cual existe la función generadora de momentos de las variables aleatorias y se basa en el Teorema que presentaremos a continuación.

**Teorema 1.7.** Sean  $Y, Y_2$  variables aleatorias con funciones generadoras de momentos  $m(t)$  y  $m_n(t)$  respectivamente, si  $\lim_{n \rightarrow \infty} m_n(t) = m(t)$  para todo  $t$  real, entonces la función de distribución de  $Y_2$  converge a la función de distribución de  $Y$ , cuando  $n$  tiende a infinito.

*Demostración. (del teorema del límite central)* Defina una variable aleatoria  $Z_i$  como:  $Z_i = \frac{(y_i - \mu)}{\sigma}$  con  $E(Z_i) = 0$ ,  $V(Z_i) = 1$  y  $m(t) = e^{t^2/2}$ , luego podemos escribir la función generadora de momentos de  $Z_i$  como

$$\begin{aligned} m_z(t) &= E[e^{tZ_i}] = E \left[ \frac{(tZ_i)^0}{0!} + \frac{(tZ_i)}{1!} + \frac{(tZ_i)^2}{2!} + \dots \right] \\ &= E \left[ 1 + tZ_i + \frac{t^2 Z_i^2}{2!} + \dots \right] = 1 + tE[Z_i] + \frac{t^2}{2!} E[Z_i^2] + \dots \\ &= 1 + \frac{t^2}{2!} + \frac{t^3 E(Z_i^3)}{3!} + \dots \end{aligned}$$

además si definimos el estadístico

$$U_n = \sqrt{n} \left( \frac{\bar{Y} - \mu}{\sigma} \right) = \frac{1}{\sqrt{n}} \left( \frac{\sum_{i=1}^n Y_i - n\mu}{\sigma} \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i$$

Y como las variables  $Y_i$  son independientes, las variables aleatorias  $Z_i$  también lo son, para  $i = 1, 2, \dots, n$ . Se obtiene que la función generadora de momentos de  $U_n$  es igual a la función generadora de momentos de  $\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i$  y además sabiendo que la función generadora de momentos de la suma de variables aleatorias independientes, es el producto de las funciones generadoras de momentos individuales, entonces

$$m_n(t) = \left[ m_z \left( \frac{t}{\sqrt{n}} \right) \right]^n = \left( 1 + \frac{t^2}{2n} + \frac{t^3 k_j}{3! n^{3/2}} + \dots \right)^n$$

donde  $k_j = E[Z_i^j]$ ,  $j = 3, \dots$

Luego

$$\begin{aligned} m_n(t) &= \left( 1 + \frac{1}{n} \left( \frac{t^2}{2} + \frac{t^3 k_j}{3! n^{1/2}} + \dots \right) \right)^n \\ &= \left( 1 + \frac{r}{n} \right)^n \end{aligned}$$

Donde  $r = \left( \frac{t^2}{2} + \frac{t^3 k_j}{3! n^{1/2}} + \dots \right)$

Ahora tenemos que

$$\lim_{n \rightarrow \infty} [m_n(t)] = \lim_{n \rightarrow \infty} \left( 1 + \frac{r}{n} \right)^n$$

Ya que en  $r$  el único término que no depende de  $n$ , es  $\frac{t^2}{2}$  y todos los demás quedan con un exponente positivo de  $n$  en el denominador se tiene que

$$\lim_{n \rightarrow \infty} r = \frac{t}{2}$$

por lo que

$$\lim_{n \rightarrow \infty} \left( 1 + \frac{r}{n} \right)^n = e^{\frac{t}{2}}$$

La cual es la función generadora de momentos de una variable normal estándar, y por lo tanto se demuestra que la distribución de  $U_n$  converge a la distribución normal estándar cuando  $n$  es grande. ■

## 1.2. ESTIMACIÓN

Sea  $X$  una variable aleatoria que toma valores en un espacio medible  $(A, \Omega)$ , en donde la distribución de la variable  $X$  depende de un parámetro  $\theta$ . Se supone que el valor real del parámetro es desconocido y el problema consiste en estimar a base del experimento y de la variable  $X$  el valor real del parámetro. En la práctica el experimento de esta índole consiste en obtener una muestra de tamaño  $n$ , es decir,  $n$  observaciones independientes sobre la variable aleatoria  $X$ . El resultado de la  $i$ -ésima observación se denotará como  $X_i$ , con lo cual después de  $n$  observaciones se obtendrá un punto en el espacio  $A^n$  llamado espacio muestral.

A partir de las observaciones  $X_1, X_2, \dots, X_n$  se construye la estimación del valor real del parámetro. De este modo la estimación  $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ , es una función dada en el espacio muestral  $A^n$  y Sustituyendo los resultados de las observaciones en vez de los argumentos, obtendremos el valor del parámetro  $\hat{\theta}$ .

Se pueden construir un número infinito de tales funciones y surge la pregunta de cuáles de ellas son preferibles. La respuesta no es unívoca, pues se puede introducir diversos criterios de la calidad de las estimaciones.

El objetivo de la estadística es hacer inferencias con respecto a una población objetivo basándose en la información contenida en una muestra. La generalidad de los procedimientos de la inferencia estadística utilizan en gran parte la estimación. La estimación de un parámetro como hemos dicho anteriormente involucra el uso de los datos muestrales utilizando algún estadístico, Para llevar a cabo lo anterior se utiliza la estimación puntual

y la estimación por intervalo. En la primera se busca un estimador que con base en las datos muestrales de origen a una estimación unívoca del valor del parámetro y que recibe el nombre de *estimado puntual*. Para la segunda se determina un intervalo en el que, en forma probable se encuentre el valor del parámetro, este intervalo recibe el nombre de *intervalo de confianza estimado*.

Se examinarán algunas propiedades de los estimadores, donde se supone la existencia de un solo parámetro desconocido, sin embargo debe aclararse que bajo condiciones más generales estos conceptos pueden extenderse para incluir un número mayor de parámetros.

## Propiedades de los Estimadores Puntuales

**Definición 1.18.** *Sea el estadístico  $\hat{\theta}$  un estimador puntual de un parámetro  $\theta$  :*

1. *Se dice que  $\hat{\theta}$  es un estimador insesgado si  $E(\hat{\theta}) = \theta$ . De lo contrario se dice que  $\hat{\theta}$  es sesgado.*
2. *El sesgo de un estimador puntual  $\hat{\theta}$  está dado por  $B = E(\hat{\theta}) - \theta$ .*
3. *La media del cuadrado del error (MCE) (o error cuadrático medio ECM) de un estimador puntual se define como*

$$MCE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

*donde  $MCE(\hat{\theta})$  también se puede definir como  $MCE(\hat{\theta}) = V(\hat{\theta}) - B^2$  y es una función de la varianza y el sesgo de  $\hat{\theta}$*

## Evaluación de la Bondad de un Estimador Puntual

**Definición 1.19.** *se define el error de estimación  $\epsilon$  de  $\hat{\theta}$  como  $\epsilon = |\hat{\theta} - \theta|$ .*

Es deseable que  $\epsilon$  sea lo mas pequeño posible, pues lo que se quiere es que en un muestreo repetitivo el estimador esté muy cerca del parámetro.

Sea  $\hat{\theta}$  un estimador insesgado de  $\theta$  el cual se distribuye normalmente. Se eligen dos puntos  $\theta - b$  y  $\theta + b$  localizados cerca de las colas de la distribución de probabilidad, tal que  $-b < \theta < b$  entonces se tiene que la probabilidad de que  $\epsilon < b$  está dado por:

$$P(|\hat{\theta} - \theta| < b) = P[-b < (\hat{\theta} - \theta) < b] = P(\theta - b < \hat{\theta} < \theta + b)$$

.

El cual determina un intervalo para el parámetro  $\theta$ .

Si  $b = k\sigma_{\hat{\theta}}$ ,  $k \geq 1$ , se sabe por el teorema de Tchebysheff que:

$$P(\epsilon < k\sigma_{\hat{\theta}}) \leq 1 - \frac{1}{k^2} \text{ donde } \sigma_{\hat{\theta}} \text{ es la desviación estándar muestral de } \hat{\theta}.$$

**Definición 1.20.** *Dados  $\hat{\theta}_1$  y  $\hat{\theta}_2$  dos estimadores insesgados de un parámetro  $\theta$ , con varianzas  $V(\hat{\theta}_1)$  y  $V(\hat{\theta}_2)$  respectivamente, entonces se define la eficiencia relativa de  $\hat{\theta}_1$  con respecto a  $\hat{\theta}_2$  como:*

$$\mathbf{Eficiencia} = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$$

*Si la eficiencia es mayor que 1, se dice que  $\hat{\theta}_1$  es mejor estimador insesgado que  $\hat{\theta}_2$  para  $\theta$ .*

Antes de presentar las demás propiedades de los estimadores, se dará la siguiente definición:

**Definición 1.21.** *Una sucesión de variables aleatorias  $\{Y_n\}_{n \geq 1}$  independientes o no, convergen en probabilidad<sup>2</sup> a la variable aleatoria  $Y$  cuando para todo  $\epsilon > 0$ ,*

$$\lim_{n \rightarrow \infty} P(|Y_n - Y| \geq \epsilon) = 0$$

*y se denota por  $Y_n \xrightarrow{P} Y$*

Esta definición también se puede dar por el complemento

$$\lim_{n \rightarrow \infty} P(|Y_n - Y| \leq \epsilon) = 1$$

---

<sup>2</sup>La convergencia en probabilidad no es equivalente a la convergencia de las variables  $Y_n$  en el sentido matemático



**Definición 1.22.** Se dice que el estimador  $\hat{\theta}_n$  (para un tamaño de muestra  $n$ , dado) es consistente para  $\theta$ , si:

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| \leq \epsilon) = 1$$

**Teorema 1.8.** Un estimador insesgado  $\hat{\theta}_n$  es consistente para  $\theta_n$  si

$$\lim_{n \rightarrow \infty} V(\hat{\theta}_n) = 0$$

**Teorema 1.9.** Si  $\hat{\theta}_n$  y  $\hat{\theta}'_n$  son estimadores consistentes para  $\theta_n$  y  $\theta'_n$  respectivamente, entonces:

- a)  $\hat{\theta}_n + \hat{\theta}'_n$  es un estimador consistente de  $\theta + \theta'$ .
- b)  $\hat{\theta}_n \cdot \hat{\theta}'_n$  es un estimador consistente de  $\theta \cdot \theta'$ .
- c)  $\hat{\theta}_n / \hat{\theta}'_n$  es un estimador consistente de  $\theta / \theta'$ , si  $\theta' \neq 0$ .
- d)  $\sqrt{\hat{\theta}_n}$  es un estimador consistente de  $\sqrt{\theta}$  si  $P(\hat{\theta}_n \geq 0) = 1$ .

**Teorema 1.10.** Si  $U_n$  tiene una función de distribución convergente a la distribución Normal Estándar cuando  $n$  tiende a infinito y  $W_n$  converge en probabilidad a 1, entonces la función de distribución de  $U_n/W_n$  converge a la distribución Normal Estándar.

**Definición 1.23.** Sea  $Y_1, Y_2, \dots, Y_n$  una muestra aleatoria de una distribución de probabilidad con parámetro desconocido  $\theta$ , se dice que el estadístico  $U = g(Y_1, Y_2, \dots, Y_n)$ , es suficiente para  $\theta$  si la distribución condicional de  $Y_1, Y_2, \dots, Y_n$  dado  $U$  no depende de  $\theta$ .

**Definición 1.24.** Sean  $y_1, y_2, \dots, y_n$  las  $n$  observaciones muestrales de las variables aleatorias  $Y_1, Y_2, \dots, Y_n$  respectivamente, entonces la verosimilitud  $L(\theta)$  está dada por

$$L(\theta) = L(y_1, y_2, \dots, y_n; \theta) = f(y_1, y_2, \dots, y_n; \theta)$$

Donde  $L(\theta) = L(y_1, y_2, \dots, y_n; \theta)$  se define como la probabilidad conjunta evaluada en  $y_1, y_2, \dots, y_n$  para variables aleatorias discretas y como la densidad conjunta evaluada en  $y_1, y_2, \dots, y_n$  para variables aleatorias continuas.

**Teorema 1.11.** Sea  $U$  un estadístico basado en una muestra aleatoria  $Y_1, Y_2, \dots, Y_n$ , entonces  $U$  es un estadístico suficiente para la estimación de un parámetro  $\theta$  si y sólo si  $L(\theta) = L(y_1, y_2, \dots, y_n; \theta) = g(u, \theta) \cdot h(y_1, y_2, \dots, y_n)$ , donde  $g$  depende sólo de  $u$  y de  $\theta$ , y  $h$  no depende de  $\theta$ .

## Suficiencia Mínima y Estimadores insesgados de Mínima Varianza

Sea  $Y_1, Y_2, \dots, Y_n$  una muestra aleatoria de una función de probabilidad  $p(y)$ , o una función de densidad  $f(y)$  con un parámetro desconocido  $\theta$  al comparar las verosimilitudes de dos observaciones;  $x_1, x_2, \dots, x_n$  y  $y_1, y_2, \dots, y_n$ , el método de Lehmann y Scheffé enuncia que si se puede encontrar una función  $g$  tal que la razón  $L(x_1, x_2, \dots, x_n)/L(y_1, y_2, \dots, y_n)$  no presente el parámetro desconocido  $\theta$  si y sólo si  $g(x_1, x_2, \dots, x_n) = g(y_1, y_2, \dots, y_n)$ , entonces  $g(y_1, y_2, \dots, y_n)$  es un estadístico de *suficiencia mínima* para  $\theta$ .

Si tenemos dos estimadores insesgados  $\hat{\theta}_1$  y  $\hat{\theta}_2$  de un parámetro  $\theta$  y aunque las distribuciones de cada uno están centradas en el verdadero valor de  $\theta$ , las varianzas de estas distribuciones pueden ser diferentes. Si se tiene que  $\hat{\theta}_1$  tiene una varianza más pequeña que  $\hat{\theta}_2$ , lo más probable es que el estimador  $\hat{\theta}_1$  produzca un estimado más cercano al verdadero valor de  $\theta$ . Y si se quiere elegir entre dos o más de estos estimadores, lo más

lógico sería seleccionar aquel que tenga la varianza más pequeña.

**Definición 1.25.** Sea  $Y_1, Y_2, \dots, Y_n$  una muestra aleatoria de una función de probabilidad  $p(y)$ , o una función de densidad  $f(y)$ . Sea el estadístico  $\hat{\theta}$  un estimador insesgado de  $\theta$  y si la varianza de este estimador es menor que la varianza de cualquier otro estimador insesgado de  $\theta$ , para todos los posibles valores de  $\theta$ , se dice que  $\hat{\theta}$  es un estimador insesgado de varianza mínima.

Además, para funciones de probabilidad y funciones de densidad normales, cualquier estadístico insesgado que sea función de un estadístico de mínima suficiencia, será un estimador insesgado de mínima varianza

## Intervalos de Confianza

En algunos casos es importante no solo dar la estimación para un parámetro, si no también indicar un intervalo donde supuestamente debe estar el valor real del parámetro. Lo ideal sería que el intervalo contenga al parámetro y que éste sea lo más estrecho posible. El intervalo se construye a partir de las observaciones y por ello puede variar de una muestra a otra, por lo que los extremos, la longitud y la localización del intervalo son cantidades aleatorias.

A los estimadores por intervalo se les denomina intervalos de confianza, a los extremos superior e inferior de estos intervalos, se les llama *límite de confianza superior* y *límite de confianza inferior* respectivamente, la probabilidad que un intervalo de confianza contenga al parámetro objetivo se llama *coeficiente de confianza*, que se denota por  $(1 - \alpha)$ , donde  $\alpha$  es determinado por el investigador.

**Definición 1.26.** Supóngase que  $\hat{\theta}_s$  y  $\hat{\theta}_i$  son los límites superior e inferior respectivamente, de un intervalo que contiene al parámetro  $\theta$  con un nivel de confianza de  $(1 - \alpha)\%$ , entonces el intervalo de confianza para  $\theta$  está dado por

$$P(\hat{\theta}_i < \theta < \hat{\theta}_s) = 1 - \alpha$$

Donde  $0 \leq \alpha \leq 1$

De lo anterior podemos decir que: se tiene una probabilidad de  $(1 - \alpha)\%$  de seleccionar una muestra que produzca un intervalo que contiene el verdadero valor de  $\theta$ , es decir, que si se recopila un número infinito de muestras aleatorias y se calcula un intervalo de confianza del  $100(1 - \alpha)$  por ciento para  $\theta$ , para cada una de las muestras, entonces el  $100(1 - \alpha)$  por ciento de esos intervalos contienen el valor verdadero de  $\theta$ . Además a estos intervalos se les conoce como *intervalos de confianza bilaterales*

Se pueden construir *intervalos de confianza unilaterales*, como:

$$P(\hat{\theta} < \hat{\theta}_s) = 1 - \alpha$$

que se denominan *intervalos de confianza unilaterales superiores* y

$$P(\hat{\theta}_i < \theta) = 1 - \alpha$$

que se denominan *intervalos de confianza unilaterales inferiores*

## Intervalos de Confianza con Muestras Grandes

Si  $\hat{\theta}$  es un estimador puntual para  $\theta$ , el cual ha sido calculado a partir de una muestra grande con valor esperado  $\theta$  y varianza  $\sigma_{\hat{\theta}}^2$ , se puede demostrar que  $z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$  posee aproximadamente una distribución normal estándar, entonces se desea construir un intervalo de confianza para un parámetro  $\theta$ , con un coeficiente de confianza de  $(1 - \alpha)$ . Veamos: Se toman dos valores  $z_{\alpha/2}$  y  $-z_{\alpha/2}$  de los extremos de la distribución normal estándar, de tal manera que

$$p(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

Sustituyendo  $Z$  y realizando los desarrollos correspondientes tenemos:

$$\begin{aligned} &= p\left(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} < z_{\alpha/2}\right) = 1 - \alpha \\ &= P(-z_{\alpha/2}\sigma_{\hat{\theta}} < \hat{\theta} - \theta < z_{\alpha/2}\sigma_{\hat{\theta}}) = 1 - \alpha \\ &= P(-\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} < -\theta < -\hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}) = 1 - \alpha \\ &P(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} < \theta < \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}) = 1 - \alpha \end{aligned}$$

donde los extremos  $-z_{\alpha/2}$  y  $z_{\alpha/2}$  son los valores de la abscisa de la curva normal entre los cuales está comprendido un área de  $1 - \alpha$ .

Y así obtenemos el intervalo de confianza para  $\theta$ , determinado por:

$$\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$$

### **Intervalos de Confianza para $\mu$ , $\mu_1 - \mu_2$**

Supóngase que  $Y_1, Y_2, \dots, Y_n$  es una muestra aleatoria obtenida de una población con distribución normal, y sean  $\bar{Y}$  y  $S^2$  la media y la varianza muestral respectivamente. Se desea construir un intervalo de confianza para la media poblacional  $\mu$  conociendo  $V(Y_i) = \sigma^2$  con  $i = 1, 2, \dots, n$  y cuando el tamaño de la muestra es demasiado pequeño. De los teoremas (1.3), (1.5) y de la definición (1.16) se tiene que

$$T = \frac{\bar{Y} - \mu}{S/\sqrt{n}}$$

tiene una distribución  $t$  con  $(n - 1)$  grados de libertad. Se pueden encontrar los valores  $-t_{\alpha/2}$  y  $t_{\alpha/2}$  de tal manera que:

$$P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1 - \alpha$$

Y sustituyendo  $T$  y realizando los desarrollos correspondientes como en el caso anterior, se obtiene el intervalo de confianza para  $\mu$  por:

$$\bar{Y} \pm t_{\alpha/2} \left( \frac{S}{\sqrt{n}} \right)$$

Si se quiere comparar las medias de dos poblaciones normales con medias  $\mu_1$ ,  $\mu_2$ , y con varianzas  $\sigma_1^2$  y  $\sigma_2^2$  respectivamente tal que  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ , y si  $\bar{Y}_1$ , y  $\bar{Y}_2$  son las medias muestrales de tamaños  $n_1$  y  $n_2$ , respectivamente entonces el intervalo de confianza con muestras pequeñas para  $\mu_1 - \mu_2$  está dado por

$$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

que tiene una distribución  $t$  con  $(n_1 + n_2 - 2)$  grados de libertad y el intervalo de confianza para  $\mu_1 - \mu_2$  tendrá la forma

$$(\bar{Y}_1 - \bar{Y}_2) \pm t_{\alpha/2} S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

donde  $-t_{\alpha/2}$  y  $t_{\alpha/2}$  se obtienen de la distribución  $t$  con  $(n - 1)$  grados de libertad, entre los cuales está comprendido un área de  $1 - \alpha$ . Siendo

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Cuando el tamaño de la muestra crece los intervalos de confianza de estas estas distribuciones son equivalentes a los intervalos de confianza con muestra grandes, vistos anteriormente.

## Intervalos de Confianza para $\sigma^2$

Sea  $Y_1, Y_2, \dots, Y_n$  una muestra aleatoria, de una población con distribución normal con media  $\mu$  y varianza  $\sigma^2$ , ambas desconocidas, por el teorema 1.5 se sabe que

$$\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2}$$

tiene una distribución  $\chi^2$  con  $(n-1)$  grados de libertad, de donde se obtiene el intervalo de confianza para  $\sigma^2$  que está dado por

$$P\left(\frac{(n-1)S^2}{\chi^2_s} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_i}\right) = (1 - \alpha)$$

con  $\chi^2_i$  y  $\chi^2_s$  límites inferior y superior de confianza, que son los valores de la abscisa de la curva  $ji^2$  entre los cuales está comprendido un área de  $1 - \alpha$ , y se pueden hallar a través de las tablas de la distribución  $\chi^2$  con  $(n-1)$  grados de libertad, para una probabilidad de  $(1 - \alpha)$ .

## Método de Máxima Verosimilitud

En dos artículos publicados a principios del siglo pasado, R.A. Fisher, el prominente estadístico, propuso un método general de estimación llamado *El método de máxima verosimilitud*. También mostró las ventajas de este método al demostrar que producían estimadores suficientes siempre que estos existieran y que los estimadores de máxima verosimilitud suelen ser estimadores insesgados de mínima varianza, además estos estimadores tienen buenas propiedades asintóticas.

**Definición 1.27.** *El estimador de máxima verosimilitud del parámetro  $\theta$ , es el valor de  $\theta$  que maximiza la función de verosimilitud  $L(\theta) = f(y_1, y_2, \dots, y_n; \theta)$ .*

En el caso de una variable aleatoria discreta, la función de verosimilitud es  $P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n)$  y el estimador de máxima verosimilitud es un estimador que maximiza la probabilidad de ocurrencia de los valores muestrales.

Los estimadores de máxima verosimilitud, no son necesariamente insesgados, por ejemplo puede demostrarse que el estimador de máxima verosimilitud  $\hat{\theta}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$  no es insesgado para  $\sigma^2$ .

El estimador de máxima verosimilitud, en general, de cualquier parámetro  $\theta$  es insesgado para  $n$  grande y tiene una varianza tan pequeña como la que puede obtenerse con otro estimador, además tienen una propiedad de invarianza, es decir, si  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s$ , son los estimadores de máxima verosimilitud de los parámetros  $\theta_1, \theta_2, \dots, \theta_s$ , entonces el estimador de máxima verosimilitud de cualquier función  $h(\theta_1, \theta_2, \dots, \theta_s)$  de los parámetros, es la misma función  $h(\bar{\theta}_1, \bar{\theta}_2, \dots, \bar{\theta}_s)$  de los estimadores.

El método de máxima verosimilitud puede emplearse en situaciones donde existen varios parámetros desconocidos, en estos casos la función de verosimilitud, es una función de estos parámetros. Los estimadores de máxima verosimilitud se obtienen entonces igualando a cero las derivadas parciales de la función de máxima verosimilitud respecto a cada parámetro y resolviendo el sistema de ecuaciones resultante.

### 1.3. PRUEBAS DE HIPÓTESIS

La esencia de una prueba de hipótesis es decidir o probar si una afirmación con respecto a alguna característica desconocida de una población, se encuentra apoyada por la evidencia experimental obtenida a través de una muestra aleatoria. Esta afirmación involucra algún parámetro o alguna forma no conocida de la distribución de la población de la cual se obtiene la muestra. La decisión de si los datos apoyan tal afirmación se toma con base en la probabilidad y si esta es mínima entonces se rechaza la afirmación.

Una prueba de hipótesis involucra los siguientes elementos:

**Hipótesis Nula ( $H_0$ ):** Es una suposición de que un parámetro de una población tome



un valor o un conjunto de valores determinados y es considerada verdadera a menos que los datos arrojados por la muestra demuestren lo contrario.

**Hipótesis alternativa ( $H_a$ ):** (o hipótesis de estudio) Es una hipótesis alterna a la hipótesis nula, lanzada por el investigador, con la cual pretende probar la hipótesis nula, es decir si puede aceptar la hipótesis alternativa cuando los datos de arrojados por la muestra determinen que se debe rechazar  $H_0$ . Esta debe reflejar el valor posible o intervalo de valores del parámetro de interés si la hipótesis nula es falsa.

**Estadístico de prueba:** Es (un estimador) una función de las mediciones muestrales, en el cual se fundamenta la decisión estadística. Para ciertos valores del estadístico de prueba, la decisión será rechazar la hipótesis nula.

**Región de rechazo ( $RR$ ):** Especifica los valores del estadístico de la prueba para los cuales rechaza la hipótesis nula, si en una determinada prueba el valor arrojado por el estadístico de se localiza en la región de rechazo, se rechazará  $H_0$ , de lo contrario se dice que no hay suficiente evidencia para rechazarla.

Ya habiendo definido los elementos de una prueba de hipótesis se presentará la siguiente definición:

**Definición 1.28.** *Una prueba de hipótesis estadística con respecto a alguna característica desconocida de una población de interés es cualquier regla para decidir si se rechaza la hipótesis nula con base en una muestra aleatoria de la población.*

Observe que en una prueba de hipótesis se evalúa la hipótesis nula en donde: puede rechazarse  $H_0$  y aceptarse  $H_a$ , o en forma alternativa; puede no rechazarse la hipótesis nula. Nunca se acepta una hipótesis nula, lo que debe decir es que no se encontró evidencia

para rechazarla. posteriormente puede ser probada como incorrecta al recolectar nuevos datos de la población.

Al realizar una prueba de hipótesis se puede tomar dos decisiones con respecto a  $H_0$  ; rechazarla o no rechazarla, lo cual implica que se puedan cometer uno de los dos tipos de errores siguientes:

**Error de Tipo I:** se comete al rechazar  $H_0$  cuando es cierta y la probabilidad de cometer este error se denota por  $\alpha$ , llamado *nivel de significación*.

**Error de Tipo II:** se comete cuando se acepta  $H_0$  siendo cierta  $H_a$  y la probabilidad de cometer el error de tipo *II* se denota por  $\beta$ .

Además  $\alpha$  y  $\beta$  están relacionados inversamente.

En vez de considerar la probabilidad de cometer un error de tipo I, se puede considerar también la probabilidad de no rechazar  $H_0$  cuando sea verdadera, a esta probabilidad la llamamos nivel de confianza de la prueba y se denota por  $1 - \alpha$ , en las pruebas de hipótesis el investigador fija este nivel de acuerdo al conocimiento que tenga a cerca del problema que tenga a tratar.

## Pruebas Comunes Con Muestras Grandes

Si se quiere probar una hipótesis estadística referente a un parámetro  $\theta$ , basada en una muestra aleatoria  $Y_1, Y_2, \dots, Y_n$ . Se basará la prueba en el estimador  $\hat{\theta}$  que tiene una distribución aproximadamente normal con media  $\theta$  y desviación estándar  $\sigma_{\hat{\theta}}$ . Sea  $\theta_0$  un valor específico de  $\theta$ , entonces bajo la hipótesis nula  $H_0 : \theta = \theta_0$ , se puede tener una de las siguientes hipótesis alternativas:

$$H_a = \begin{cases} \theta > \theta_0, & \text{Alternativa de cola superior,} \\ \theta < \theta_0, & \text{Alternativa de cola inferior} \\ \theta \neq \theta_0, & \text{Alternativa de dos colas.} \end{cases}$$

bajo es supuesto de que la muestra tiene distribución normal; el estadístico de prueba es:

$$Z = \left( \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \right)$$

y para una probabilidad  $\alpha$  de cometer un error de tipo I, la región de rechazo correspondiente a cada una de las hipótesis alternativas estará determinada por:

$$RR = \begin{cases} z > z_{\alpha}, & \text{region de rechazo de cola superior,} \\ z < -z_{\alpha}, & \text{region de rechazo de cola inferior,} \\ |z| > z_{\alpha/2}, & \text{region de rechazo de dos colas.} \end{cases}$$

Por ejemplo, para la hipótesis alternativa de cola superior, si  $z$  cae muy alejado en la cola superior de la distribución normal estándar, es decir  $z > z_{\alpha/2}$ , entonces se rechazará la hipótesis que se tenía inicialmente de que  $\theta = \theta_0$ .

## Error de Tipo II y Determinación del Tamaño de la Muestra

Si la hipótesis nula es falsa y esta se rechaza, entonces se está tomando una decisión acertada, a esta probabilidad se la llama poder de la prueba y está dado por  $1 - \beta$ .

Si se tiene una prueba  $H_0 : \theta = \theta_0$  frente a  $H_a : \theta > \theta_0$  y si se dispone de una alternativa específica  $\theta = \theta_a$ , ( $\theta_a > \theta_0$ ), con estadístico de prueba es  $\hat{\theta}$  y cuya respectivo región de rechazo  $RR = \hat{\theta} > k$  para un valor específico de  $\alpha$ , entonces la probabilidad de cometer un error de tipo II es

$$\beta = P(\hat{\theta} \leq k \text{ cuando } \theta = \theta_a) = P\left(\frac{\hat{\theta} - \theta_a}{\sigma_{\hat{\theta}}} \leq \frac{k - \theta_a}{\sigma_{\hat{\theta}}} \text{ cuando } \theta = \theta_a\right)$$

si  $\theta = \theta_a$ , entonces  $\left[ \frac{\hat{\theta} - \theta_a}{\sigma_{\hat{\theta}}} \right]$  tiene una distribución normal estándar.

Si se fijan los valores de  $\alpha$  y  $\beta$  la prueba depende del tamaño de la muestra y del punto  $k$  dónde empieza la región de rechazo, y debido a que la desviación estándar del estimador,

$\sigma_{\hat{\theta}}$  depende del tamaño  $n$  de la muestra, entonces las probabilidades para  $\alpha$  y  $\beta$  pueden escribirse en dos ecuaciones que involucran a  $k$  y  $n$

$$\alpha = P\left(\frac{Y - \mu_0}{\sigma/\sqrt{n}} > \frac{k - \mu_0}{\sigma/\sqrt{n}} \text{ cuando } \mu = \mu_0\right) = P(z > z_\alpha)$$

$$\beta = P\left(\frac{Y - \mu_a}{\sigma/\sqrt{n}} > \frac{k - \mu_a}{\sigma/\sqrt{n}} \text{ cuando } \mu = \mu_a\right) = P(z \leq -z_\beta)$$

donde  $\beta$  se evalúa para un valor particular de  $\mu$ ,  $\mu_a > \mu_0$

De estas ecuaciones se obtiene que

$$\left[\frac{k - \mu_0}{\sigma/\sqrt{n}} = z_\alpha\right] \text{ y } \left[\frac{k - \mu_a}{\sigma/\sqrt{n}} = -z_\beta\right]$$

al resolver este sistema se obtiene el tamaño de muestra

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_a - \mu_0)^2}$$

## Niveles De Significación

**Definición 1.29.** Si  $W$  es un estadístico de prueba, el valor  $p$  o nivel de significación alcanzado es el mínimo valor de  $\alpha$  para el cual los datos observados indican que se debe rechazar la hipótesis nula.

Una forma de calcular los valores  $p$  es la siguiente: si se tiene que rechazar  $H_0$  en favor de  $H_a$ , cuando el estadístico  $W$  arroja valores pequeños, y sea  $w_0$ , el valor observado por el estadístico  $W$ , entonces:

$$\text{valor } p = P(W \leq w_0 \text{ cuando } H_0 \text{ es verdadera})$$

Si se rechaza  $H_0$  para valores grandes de  $W$ , el valor  $p$  es:

$$\text{valor } p = P(W \geq w_0 \text{ cuando } H_0 \text{ es verdadera})$$

## Pruebas De Hipótesis Referentes a varianzas

**Definición 1.30.** Sea  $Y_1, Y_2, \dots, Y_n$  una muestra aleatoria distribuidas normalmente con  $E(Y_i) = \mu$  y  $V(Y_i) = \sigma^2$  desconocida, si se desea probar  $H_0 : \sigma^2 = \sigma_0^2$  frente a una de las alternativas:

$$H_a = \begin{cases} \sigma^2 > \sigma_0^2 & \text{alternativa de cola superior} \\ \sigma^2 < \sigma_0^2 & \text{alternativa de cola inferior} \\ \sigma^2 \neq \sigma_0^2 & \text{alternativa de dos colas} \end{cases}$$

El estadístico de prueba será:  $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$

Y las correspondientes regiones de rechazo

$$RR = \begin{cases} \chi^2 > \chi_\alpha^2 & \text{región de rechazo de cola superior} \\ \chi^2 < \chi_{1-\alpha}^2 & \text{región de rechazo de cola inferior} \\ \chi^2 > \chi_{\alpha/2}^2 \text{ o } \chi^2 < \chi_{1-\alpha/2}^2 & \text{región de rechazo de dos colas} \end{cases}$$

donde  $\chi_\alpha^2$  se halla de  $P(\chi^2 > \chi_\alpha^2) = \alpha$  para un valor de  $\alpha$  dado.

**Definición 1.31.** Si se tienen dos muestras aleatorias independientes de tamaño  $n_1$  y  $n_2$  distribuidas normalmente con medias desconocidas y varianzas  $\sigma_1^2$  y  $\sigma_2^2$  desconocidas y se quiere probar  $H_0 : \sigma_1^2 = \sigma_2^2$  frente a la alternativa  $H_a : \sigma_1^2 > \sigma_2^2$ , entonces el estadístico de prueba será

$F = \frac{S_1^2}{S_2^2}$  donde  $S_1^2$  y  $S_2^2$  son las varianzas muestrales.

La región de rechazo es  $F > F_\alpha$ , donde  $F_\alpha$  se obtiene de la distribución  $F$  con  $(n_1 - 1)$  grados de libertad en el numerador y  $(n_2 - 1)$  grados de libertad en el denominador, de tal forma que  $P(F > F_\alpha) = \alpha$ . Si la hipótesis alternativa es  $H_a : \sigma_1^2 \neq \sigma_2^2$  el estadístico seguirá siendo

$F = \frac{S_1^2}{S_2^2}$  y la región de rechazo apropiada es

$$RR = \left\{ F > F^{n_1-1, n_2-1, \alpha/2} \quad \text{o} \quad F < \left( F^{n_2-1, n_1-1, \alpha/2} \right)^{-1} \right\}$$

## La potencia de las pruebas

**Definición 1.32.** Supóngase que  $W$  es el estadístico de una prueba y sea  $RR$  la región de rechazo para una prueba referente al parámetro  $\theta$ . Entonces la potencia de la prueba, denotado por  $\text{potencia}(\theta)$  está dada por:

$$\text{potencia}(\theta) = P(\text{de que } W \text{ esté en } RR \text{ cuando el valor del parámetro es } \theta)$$

Para cualquier valor de  $\theta$  para  $H_a$  la potencia de una prueba mide la capacidad de la prueba para detectar que  $H_0$  es falsa, esto es  $\theta = \theta_a$

$$\text{Potencia}(\theta_a) = P(\text{rechazar } H_0 \text{ cuando } \theta = \theta_a)$$

donde la potencia de la prueba para  $\theta_a$  y la probabilidad del error de tipo II se relacionan como:

$$\text{Potencia}(\theta_a) = 1 - \beta.$$

**Teorema 1.12.** (Lema de Neyman-Pearson) Si se desea probar una hipótesis nula simple  $H_0 : \theta = \theta_0$  frente a una hipótesis alternativa  $H_a : \theta = \theta_a$  basados en una muestra aleatoria

$Y_1, Y_2, \dots, Y_n$  de una distribución con parámetro  $\theta$ . Entonces para un valor dado de  $\alpha$ , la prueba que maximiza la potencia en  $\theta_a$  tiene una región de rechazo determinada por

$$\frac{L(\theta_0)}{L(\theta_a)} < k$$

tal prueba será la más potente para  $H_0$  frente a  $H_a$ .

La ampliación de estos conceptos, así como otros, se pueden encontrar en la mayoría de los textos citados en la bibliografía de este documento.

# 2. EL MODELO DE REGRESIÓN LINEAL

El simposio de Estadística de la Universidad Nacional de Colombia en el año de 1990 estuvo dedicado al análisis de Regresión y abrió su acto de conferencias con el trabajo “Perspectiva Histórica” (MAYORGA, J.H., SOTO, O. F. 1990) en el cual se presenta un esbozo histórico del desarrollo de la teoría de la Regresión desde Legendre.

En su presentación Mayorga y Soto, hacen una división de la historia en tres eras, señalando a través de estas, los principales aspectos de los trabajos de Legendre y Gauss, así como los alcances relevantes que a través de otras disciplinas alcanzó la Estadística.

Hemos querido presentar un aporte de la introducción de este trabajo para ambientar los desarrollos Teóricos que en este trabajo se han puntualizado.

“... El triunvirato de Gauss en Alemania, Legendre en Francia y el menos nombrado, Adrian en América, inicia la primera era de la historia del desarrollo de los conceptos, métodos y usos del Análisis de Regresión, la cual cubre el periodo que se inicia en 1805 y que termina alrededor de 1935, año que en el cual; Fisher publica su texto de Diseño Experimental y simultáneamente Aiken formula la teoría de los Mínimos Cuadrados en forma matricial.

La segunda época se caracteriza por el desarrollo conceptual, impulsado por el hecho de producirse en este periodo, en la teoría Estadística, un fulgurante avance que contribuye con la consolidación de la estructura de la Teoría de la Regresión.



Alrededor de 1960 como consecuencia del desarrollo tecnológico en el cómputo electrónico, comienza una tercera época en la historia de la Regresión en la cual el interés se centra en el aspecto metodológico y los resultados de estos procesos, son rápidamente implementados en el “software” de los sistemas de cómputo”.

Este capítulo presenta los desarrollos teóricos relacionados con el Modelo de Regresión Lineal.

Queras, E.L. (1990) presenta una definición acerca de lo que se puede entender por modelo general (univariado) propiamente dicho así:

**Definición 2.33.** *Un modelo general (univariado) propiamente dicho es aquella ecuación (lineal o no lineal) mediante la cual se expresan los resultados de una variable o respuesta de interés (variable aleatoria dependiente) en términos (función) de las variables explicativas o independientes [constantes, variables cuantitativas(variables) y/o variables cualitativas (factores)], para los elementos o unidades en observación, teniendo en cuenta los objetivos de la investigación y orientados por un marco conceptual (tórico o de referencia) de acuerdo con el área de conocimiento en donde se hace la evaluación y se realizará la decisión.*

Sean  $x_1, x_2, \dots, x_k$  un conjunto de observaciones de las variables de predicción con las cuales se pretende obtener información acerca del valor promedio de una variable aleatoria  $Y$ , de la cual también se realiza observación y supóngase que estas variable están relacionadas por el modelo teórico desconocido

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k$$

El cual se puede aproximar a partir del modelo estadístico:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \tag{2.1}$$

Donde  $Y$  es la variable aleatoria de respuesta para los valores fijos  $x_1, x_2, \dots, x_k$ ,  $\varepsilon$  es una variable aleatoria no observable asociada a  $Y$  y  $\beta_0, \beta_1, \dots, \beta_k$  son los parámetros lineales desconocidos del modelo. El modelo planteado en 2.1 recibe el nombre de **Modelo de Regresión Lineal (MRL)**.

Al tomar una muestra aleatoria de tamaño  $n$ , a partir del modelo 2.1 se generan  $n$  ecuaciones las cuales se pueden representar por el modelo matricial propuesto por Aiken en 1935 como sigue:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.2)$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

para el cual se tienen los siguientes supuestos:

- i. El modelo está correctamente especificado, lo cual implica que una variación en la variable de respuesta que no sea atribuible al modelo es debido a un error aleatorio, para lo cual se espera que  $E(\varepsilon_i) = 0$ , para todo  $i = 1, 2, \dots, n$  lo cual equivale a

$$E(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

que en forma matricial se expresa así

$$E(\boldsymbol{\varepsilon}) = \mathbf{0} \quad E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$$

donde  $\mathbf{0}$  es un vector de ceros.

De acuerdo con esto se puede decir que el promedio de  $Y$  para un número determinado de observaciones se puede expresar como una combinación lineal de los  $x_i$ .

- ii. La homogeneidad de las varianzas de los errores; esto es  $V(\varepsilon_i) = \sigma^2$  para todo  $i = 1, 2, \dots, n$  y como las variables de predicción son fijas, entonces la varianza de  $Y_i$  es también  $\sigma^2$ , para todos los valores involucrados de  $x$ .
- iii. La incorrelación de los errores, esto es que  $Cov(\varepsilon_i, \varepsilon_j) = 0$  para todo  $i \neq j$ , o escrito en forma matricial  $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \sigma^2\mathbf{I}_n$  donde  $\sigma^2$  es la varianza de  $\varepsilon_i$  e  $\mathbf{I}_n$  es la matriz identidad. Cuando también se cumple el supuesto (ii.) se tiene que  $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \sigma^2\mathbf{I}_n$ .
- iv. Los errores están normalmente distribuidos  $\varepsilon_i \sim \mathbf{N}(0, \sigma^2)$  y su equivalente en forma matricial es

$$\boldsymbol{\varepsilon} \sim \mathbf{N}_n(\mathbf{0}, \sigma^2\mathbf{I}_n) \quad \text{lo cual implica que} \quad \mathbf{Y} \sim \mathbf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$$

Además se supone que los valores de la variable de predicción son fijos o se han seleccionados con anticipación, y por tanto se miden sin error<sup>3</sup>

## 2.1. MODELO DE REGRESIÓN LINEAL SIMPLE

Una herramienta para comprender los conceptos del *MRL* (2.2), es el llamado Modelo de regresión lineal simple, que es un caso particular del *MRL*, el cual expresa a  $E[Y]$  como una función lineal de los parámetros  $\beta_0$  y  $\beta_1$ , o de igual manera:

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{2.3}$$

Para encontrar los estimadores de los parámetros  $\beta_0$  y  $\beta_1$ , utilizaremos el Método de los Mínimos cuadrados, el cual pretende ajustar un conjunto de datos a una línea recta. Dado

---

<sup>3</sup> Se debe tener en cuenta que para el *MRL*, las variables  $x_{i1}, x_{i2}, \dots, x_{ik}$  son variables aleatorias y para el Modelo Lineal General *MLG*, estas variables son no aleatorias; Sin embargo la construcción de cada uno de ellos conlleva a obtener un Modelo matricial, que en forma Matemática es idéntico. Luego la mayoría de los autores realiza Inferencias para el *MRL* sin distinguirlo del *MLG*. Conforme a esta especificación, se ha seguido la metodología de los autores y en este escrito se tomarán las variables de predicción como no aleatorias.

el Modelo (2.3), donde  $\varepsilon$  tiene una distribución con  $E[\varepsilon] = 0$  y si  $\hat{\beta}_0$  y  $\hat{\beta}_1$ , son los estimadores de  $\beta_0$  y  $\beta_1$ , entonces  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$  es un estimador de  $E[Y]$ .

De acuerdo con la definición (1.19), el error de estimación puede escribirse así:

$\varepsilon_i = y_i - \hat{y}_i$ , donde  $\hat{y}_i$  es el valor que se predice a partir del  $i$ -ésimo valor observado de  $y$ . Lo que se quiere es minimizar la suma de los cuadrados de los errores (*SCE*) determinada como:

$$SCE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \left[ y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right]^2$$

Derivando *SCE* respecto a  $\hat{\beta}_0$  y  $\hat{\beta}_1$  e igualando a cero se obtienen las ecuaciones, llamadas **ecuaciones normales**.

$$\sum_{i=1}^n y_i - n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i \quad (2.4)$$

$$\sum_{i=1}^n x_i y_i - \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n (x_i)^2 \quad (2.5)$$

De las ecuaciones 2.4 y 2.5, obtenemos

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x} \quad (2.6)$$

y

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n (x_i)^2 - \left( \sum_{i=1}^n x_i \right)^2} \quad (2.7)$$

Se puede verificar fácilmente que  $\hat{\beta}_0$  y  $\hat{\beta}_1$  minimizan *SCE*, con el criterio de la segunda derivada.

Ahora veamos que efectivamente el modelo (2.3), al obtener las respectivas observaciones de la muestra aleatoria  $Y_1, Y_2, \dots, Y_n$ , puede escribirse en forma matricial dada en la

ecuación (2.2).

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

Veamos

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n (x_i)^2 \end{bmatrix}$$

Es así que

$$\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n (x_i)^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i \\ \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n (x_i)^2 \end{bmatrix}$$

Además

$$\mathbf{X}^T \mathbf{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix}$$

Igualando estos dos últimos resultados, tenemos las **ecuaciones normales** (2.4) y (2.5), por lo que matricialmente se pueden representar por:

$$(\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y}$$

A partir de esta ecuación y si  $\mathbf{X}^T \mathbf{X}$  tiene inversa, entonces:

$$(\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

y así obtenemos el vector de estimación de los parámetros

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

Podemos expresar el vector de predicción, de la siguiente manera:

$$\hat{\mathbf{Y}} = \mathbf{X} \hat{\beta} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

Los Estimadores de Mínimos Cuadrados (*EMC*) tienen algunas propiedades, entre ellas, la de obtener estimadores insesgados. De acuerdo a la ecuación (2.7)

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

Como  $\sum_{i=1}^n (x_i - \bar{x}) = 0$  entonces

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Recordando la definición (1.18).

$$\begin{aligned} E[\hat{\beta}_1] &= \frac{\sum_{i=1}^n (x_i - \bar{x}) E[y_i]}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) (\beta_0 + \beta_1 x_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \beta_0 \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} + \beta_1 \frac{\sum_{i=1}^n (x_i - \bar{x}) x_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \beta_1 \frac{\sum_{i=1}^n (x_i - \bar{x}) x_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta_1 \end{aligned}$$

De acuerdo a la ecuación (2.6)

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

$$\begin{aligned} E[\hat{\beta}_0] &= E[\bar{y}] - \bar{x}E[\hat{\beta}_1] \\ &= \beta_0 + \beta_1 \bar{x} - \bar{x}\beta_1 = \beta_0 \end{aligned}$$

Por lo tanto  $\hat{\beta}_0$  y  $\hat{\beta}_1$  son estimadores insesgados de  $\beta_0$  y  $\beta_1$ .

Encontremos ahora las varianzas de  $\hat{\beta}_0$  y  $\hat{\beta}_1$

$$V[\hat{\beta}_1] = V\left[\frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Encontremos la Varianza de  $\hat{\beta}_0$ , recordando la expresión 2.6

$$\begin{aligned} V[\hat{\beta}_0] &= V[\bar{y} - \hat{\beta}_1 \bar{x}] = V[\bar{y}] + \bar{x}^2 V[\hat{\beta}_1] \\ &= \frac{\sigma^2}{n} + \bar{x}^2 \left[ \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] = \sigma^2 \left[ \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n\bar{x}^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \right] \\ &= \sigma^2 \left[ \frac{\sum_{i=1}^n x_i}{n \sum_{i=1}^n (x_i - \bar{x})^2} \right] = \frac{\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

Obtenidas ya las varianzas, podemos encontrar  $Cov(\hat{\beta}_0, \hat{\beta}_1)$ . Utilizando el Teorema (??) tenemos:

$$\begin{aligned}
Cov(\hat{\beta}_0 - \hat{\beta}_1) &= Cov(\bar{y} - \beta_1 \bar{x}, \beta_1) \\
&= E[(\bar{y} - \beta_1 \bar{x})\beta_1] - E[\bar{y} - \beta_1 \bar{x}]E[\beta_1] \\
&= E[\bar{y}\beta_1 - \beta_1^2 \bar{x}] - E(\bar{y})E(\beta_1) - \bar{x}E(\beta)E(\beta_1) \\
&= E[\bar{y}\beta_1] - \bar{x}E[\beta_1^2] - E(\bar{y})E(\beta_1) + \bar{x}E(\beta)E(\beta_1) \\
&= Cov(\bar{y}, \beta_1) - \bar{x}v(\beta_1) \\
&= \frac{-\bar{x}\sigma^2}{\sum (x_i - \bar{x})^2}
\end{aligned}$$

Se pueden expresar en forma matricial, las anteriores expresiones para  $V[\hat{\beta}_0]$ ,  $V[\hat{\beta}_1]$  y  $Cov(\beta_0, \beta_1)$ , mediante:

$$\mathbf{X}^T \mathbf{X}^{-1} = \begin{bmatrix} \frac{\sum_{i=2}^n (x_i)^2}{n \sum_{i=2}^n (x_i - \bar{x})^2} & \frac{\sum_{i=1}^n x_i}{n \sum_{i=2}^n (x_i - \bar{x})^2} \\ -\frac{\sum_{i=1}^n x_i}{n \sum_{i=2}^n (x_i - \bar{x})^2} & \frac{1}{n \sum_{i=2}^n (x_i - \bar{x})^2} \end{bmatrix} = \begin{bmatrix} C_{00} & C_{01} \\ C_{10} & C_{11} \end{bmatrix}$$

Donde puede verse que para  $i = 0, 1$   $V[\hat{\beta}_i] = C_{i,i} \sigma^2$  y  $Cov[\hat{\beta}_0, \hat{\beta}_1] = C_{01} \sigma^2 = C_{10} \sigma^2$ .

Demostremos que  $s^2$  es un estimador insesgado de  $\sigma^2$ , donde

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} = \frac{(SCE)}{n-2}$$

$$\begin{aligned}
E[s^2] &= \frac{E[SCE]}{n-2} = \frac{E\left[\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2\right]}{n-2} \\
&= \frac{E\left[\sum_{i=1}^n (y_i - \bar{y})^2 - 2\hat{\beta}_1 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) + \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2\right]}{n-2} \\
&= \frac{\sum_{i=1}^n E[y_i^2] - nE[\bar{y}^2] - \sum_{i=1}^n (x_i - \bar{x})^2 E[\hat{\beta}_1^2]}{n-2}
\end{aligned}$$

Lo cual se simplifica como:

$$E[s^2] = \frac{(n-2)\sigma^2}{n-2} = \sigma^2$$



Además  $SCE$ , en forma matricial, se puede representar por medio de:

$$\mathbf{Y}^T \mathbf{Y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y}$$

Veamos por qué

$$\begin{aligned} SCE &= \sum_{i=1}^n (y_i - \hat{y})^2 = \sum_{i=1}^n e_i^2 = \mathbf{e}^T \mathbf{e} \\ &= (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{Y}^T \mathbf{Y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{Y}^T \mathbf{Y} - 2\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y} + \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Y}^T \mathbf{Y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y} \end{aligned}$$

## 2.2. EL MODELO DE REGRESIÓN LINEAL GENERAL

Ahora se extenderán los conceptos del modelo de regresión lineal simple al modelo de regresión lineal general. Considérese el modelo descrito en forma matricial planteado en la ecuación (2.2), para el cual se tienen los supuestos dados en el  $MRL$

### Método de los mínimos cuadrados para estimar los parámetros del modelo de regresión Lineal

Dada la ecuación de regresión, el propósito ahora es estimar los parámetros del modelo, para tal fin se empleará el método de los mínimos cuadrados, el cual se detalló en el modelo de regresión lineal simple. Siguiendo el mismo procedimiento pero con  $k$  variables de predicción y  $k + 1$  parámetros se obtienen las ecuaciones normales para el  $MRL$ , que en forma matricial se expresan de la siguiente manera

$$\mathbf{X}^T \mathbf{Y} = \mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}}$$

Si la matriz  $\mathbf{X}^T \mathbf{X}$  tiene inversa entonces los estimadores de mínimos cuadrados para los parámetros del modelo están dados por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

multiplicando por  $\mathbf{X}$  a la izquierda de cada término de la igualdad se obtiene

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad (2.8)$$

que es el llamado modelo de regresión ajustado, el cual contiene los valores esperados para  $\mathbf{Y}$  para el conjunto de observaciones de las variables de predicción, este modelo se expresa de la siguiente forma

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

De acuerdo con esto se puede definir el vector de residuales que es un estimador del error no observable y está dado por

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} \quad (2.9)$$

Se define la *suma de los cuadrados de los errores (SCE)* por

$$SCE = \mathbf{e}^T\mathbf{e} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Equivalentemente

$$SCE = \mathbf{Y}^T\mathbf{Y} - \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{Y}$$

## Propiedades de los estimadores de mínimos cuadrados

Si se multiplica  $(\mathbf{X}^T\mathbf{X})^{-1}$  por  $\sigma^2$  (la varianza de  $Y$ ) la matriz resultante contiene las varianzas y Covarianzas de los estimadores de mínimos cuadrados del modelo. Es decir la matriz de varianzas y covarianzas del vector de estimaciones  $\boldsymbol{\beta}$  está determinada por

$$V(\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{X}^T\mathbf{X}^{-1}$$

*Prueba.*

$$V(\hat{\boldsymbol{\beta}}) = E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T]$$

$$V(\hat{\boldsymbol{\beta}}) = E[(\mathbf{X}^T\mathbf{X}^{-1}\mathbf{X}^T\boldsymbol{\varepsilon})(\mathbf{X}^T\mathbf{X}^{-1}\mathbf{X}^T\boldsymbol{\varepsilon})^T]$$

$$V(\hat{\boldsymbol{\beta}}) = E[\mathbf{X}^T\mathbf{X}^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{X}\mathbf{X}^T\mathbf{X}^{-1}]$$

$$V(\hat{\boldsymbol{\beta}}) = \mathbf{X}^T\mathbf{X}^{-1}\mathbf{X}^TE[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T]\mathbf{X}\mathbf{X}^T\mathbf{X}^{-1}$$

Como

$$E[\varepsilon\varepsilon^T] = \sigma^2\mathbf{I}_n$$

entonces

$$V(\hat{\boldsymbol{\beta}}) = \mathbf{X}^T\mathbf{X}^{-1}\mathbf{X}^T\sigma^2\mathbf{X}\mathbf{X}^T\mathbf{X}^{-1}$$

Por lo tanto

$$V(\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{X}^T\mathbf{X}^{-1}\mathbf{X}^T$$

■

A continuación se presentarán las propiedades de los estimadores de mínimos cuadrados para el *MRL*.

1. Los estimadores de mínimos cuadrados son insesgados.

*Prueba.* Sea  $\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$  hallando la esperanza de  $\hat{\boldsymbol{\beta}}$  se tiene

$$E(\hat{\boldsymbol{\beta}}) = E[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}]$$

reemplazando  $\mathbf{Y}$  por la expresión 2.2

$$E(\hat{\boldsymbol{\beta}}) = E[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] \quad \text{distribuyendo}$$

$$E(\hat{\boldsymbol{\beta}}) = E[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}]$$

aplicando propiedades de la esperanza, se tiene

$$E(\hat{\boldsymbol{\beta}}) = E[\mathbf{I}\boldsymbol{\beta}] + E[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}]$$

como

$$E[\boldsymbol{\varepsilon}] = \mathbf{0}$$

entonces

$$E(\hat{\beta}) = \mathbf{I}E(\beta) + \mathbf{0} \quad \text{como } \beta \text{ es constante se tiene que}$$

$$E(\hat{\beta}) = \mathbf{I}\beta$$

Por tanto,  $E(\hat{\beta}) = \beta$  ■

Además, como los  $\varepsilon_i$  tienen distribución normal (*supuesto iv*) entonces

2. Cada  $\hat{\beta}_i$  tiene distribución normal, y esto es debido a que cada  $\hat{\beta}_i$  es combinación lineal de los  $Y_i$  que están distribuidos normalmente.
3. un estimador, insesgado de  $\sigma^2$  es

$$S^2 = \frac{SCE}{n - (k + 1)} = \frac{\mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}}{n - (k + 1)}$$

donde  $n$  es el número de observaciones y  $k + 1$  es el número de parámetros del modelo.

4. La variable aleatoria

$$\frac{[n - (k + 1)]S^2}{\sigma^2}$$

tiene una distribución  $\chi^2$  con  $n - (k + 1)$  grados de libertad. Esto se puede verificar por el teorema 1.5.

Además los estimadores de mínimos cuadrados son los mejores estimadores lineales insesgados de los  $\hat{\beta}_i$ , esto se garantiza por el Teorema de *Gauss - Markov* citado por [Nov](Pp. 71).

## Inferencias para los parámetros $\beta_i$

Si se quiere verificar la relación existente entre la variable de respuesta y alguna variable de predicción, se realiza una prueba de hipótesis con  $H_0 : \beta_i = 0$  frente a  $H_a : \beta_i \neq 0$  ya que si  $Y$  no está relacionado con algunas variables de predicción el valor del parámetro de regresión será cero.

Recuérdese que bajo el supuesto de que el error aleatorio tiene distribución normal y partiendo de que los  $\beta_i$  se distribuyen normalmente con  $V(\hat{\beta}_i) = c_{ii}\sigma^2$ , además que es un estimador insesgado de  $\beta_i$ , entonces el estadístico de prueba es

$$Z = \frac{\hat{\beta}_i}{\sigma\sqrt{c_{ii}}}$$

y la correspondiente región de rechazo para una prueba de dos colas es

$$|Z| \geq z_{\alpha/2}$$

Pero como no siempre es posible conocer  $\sigma$  se puede utilizar el estimador insesgado  $S$  y por los resultados obtenidos en el capítulo anterior la variable

$$T = \frac{\hat{\beta}_i - \beta_{i0}}{s\sqrt{c_{ii}}}$$

tiene una distribución t con  $[n - (k + 1)]$  grados de libertad, siguiendo los procedimientos dados en la sección de estimación, se puede verificar que un intervalo de confianza para  $\hat{\beta}_i$  esta dado por

$$\hat{\beta}_i \pm t_{\alpha/2} S \sqrt{c_{ii}}$$

el cual contendrá al parámetro  $\beta$  con un nivel de confianza de  $(1 - \alpha)100$

## Técnica del análisis de varianza

Para el modelo de regresión, esta técnica consiste en descomponer la variación total de las observaciones en dos componentes: la variación debida a los términos no aleatorios  $\beta_i x_i$  y la causada por el error aleatorio  $\varepsilon$ .

La suma de los cuadrados de los errores se puede descomponer de la siguiente forma

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

donde  $\sum_{i=1}^n (y_i - \bar{y})^2 = STC$  (suma total de cuadrados),  $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = SCR$  (Suma de los cuadrados de la regresión) y  $\sum_{i=1}^n (y_i - \hat{y}_i)^2 = SCE$  (Suma de los cuadrados de los errores),

que en forma matricial se pueden escribir de la siguiente manera

$$\begin{aligned} SCE &= \mathbf{Y}^T \mathbf{Y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y} \\ SCR &= \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y} - \frac{(\sum y_i)^2}{n} = \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y} - \mathbf{n} \bar{Y}^2 \\ STC &= \mathbf{Y}^T \mathbf{Y} - \frac{(\sum y_i)^2}{n} = \mathbf{Y}^T \mathbf{Y} - \mathbf{n} \bar{Y}^2 \end{aligned}$$

Donde el número total de grados de libertad es  $n - 1$  pero los grados de libertad de  $SCE$  es  $n - m$  y de  $SCR$  es  $m - 1$  (donde  $m = k + 1$ , número de parámetros del modelo).

De acuerdo con esto se define la varianza residual o error cuadrático medio  $CME = \frac{SCE}{n-m}$  y el cuadrado medio de regresión  $CMR = \frac{SCR}{m-1}$ .

### Prueba de hipótesis para $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$

Si se desea probar que no todos los parámetros de regresión son cero, o sea, que el modelo planteado explica la relación entre la variables de respuesta y las variables de predicción, se debe realizar una prueba de hipótesis  $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$  contra la alternativa  $H_a : \beta_i \neq 0$  para algún  $i = 1, 2, \dots, k$ .

Bajo la hipótesis nula, el estadístico de prueba es

$$F = \frac{CMR}{CME}$$

Que tiene una distribución  $F$  con  $m - 1$  y  $n - m$  grados de libertad, por teorema 1.5 y definición 1.16.

Si  $F > F_\alpha$  para una distribución  $F$  con una probabilidad de  $\alpha$  y  $m - 1$  y  $n - m$  grados de libertad, entonces se dice que la variación de la variable  $Y$  se puede atribuir a las variables de predicción, en tal caso se rechazará la hipótesis nula. Esto es debido a que si las variables de predicción explican bien el comportamiento de la variable de respuesta entonces  $SCE$  se hace muy pequeña y entonces la variación de  $Y$  dependerá de la regresión.

**Nota 1.** *El hecho de que se rechace  $H_0$ , quiere decir que la ecuación de regresión es útil para realizar las predicciones.*

**Definición 2.34.** *El coeficiente de correlación múltiple es una medida de que tanto explican las variables de predicción la variación de respuesta, y está dado por*

$$R^2 = \frac{SCR}{STC} = 1 - \frac{SCE}{STC}$$

$0 \leq R^2 \leq 1$  y entre mas cercano esté  $R^2$  de uno mayor es la variación que puede explicar las variables de predicción, pero el hecho de que  $R^2$  sea muy cercano a uno no implica que la ecuación de regresión sea la mas apropiada.

## Análisis de Residuos

La importancia de realizar un análisis de residuos para el modelo de regresión radica en que, es posible detectar las deficiencias del modelo o la violación de las suposiciones del mismo. Entre las fallas mas comunes que se pueden encontrar por medio del análisis de residuos están:

Que el modelo no sea lineal en las variables de predicción.

Que una o mas variables importantes no sean incluidas en el modelo y

Que las varianzas del error  $\sigma^2$  sea constante; es decir, que el modelo sea *homocedástico*.

Recordemos que los residuales  $e_i$  son la diferencia entre los valores observados de la variable de respuesta y sus correspondientes valores estimados por medio de la ecuación de regresión, el análisis de residuos consiste en analizar las gráficas de los residuos. Si al realizar la gráfica de los residuos contra una variable  $x_i$  incluida en el modelo, no se observa ninguna tendencia entonces se tiene que el modelo está correctamente especificado, (Figura 2.1) pero si esta gráfica presenta algún patrón de comportamiento entonces se debe esperar que el modelo de regresión tiene algunas fallas y para detectarlas se emplean los *residuos estandarizados*

$$e_{i_s} = \frac{e_i}{s}, \text{ donde } s = \sqrt{ECM}$$

Si en la ecuación de regresión se debe incluir un término cuadrático, entonces la gráfica de los residuos estandarizados contra la variable  $x_i$ ; para algún  $i = 1, 2, \dots, k$ , tendrá la forma de una parábola, en este caso el remedio es incluir el factor cuadrático de  $x_i$  en el

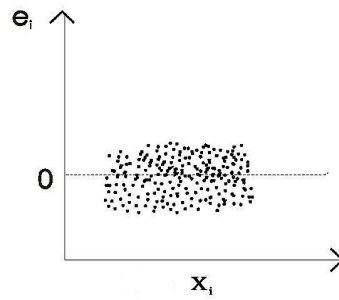


Figura 2.1

modelo, también se puede detectar si hace falta algún término de orden superior; (figura 2.2).

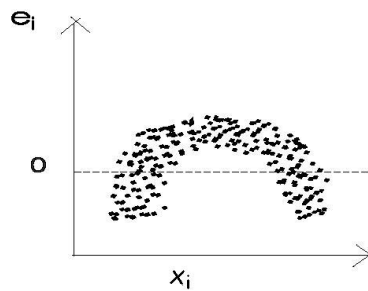


Figura 2.2

Si al realizar la gráfica de  $e_{i_s}$  contra alguna variable que no haya sido incluida en el modelo, y esta presenta una tendencia lineal, entonces dicha variable debe ser incluida en el modelo.

Cuando al realizar la gráfica de  $e_{i_s}$  contra su correspondiente valor estimado de la respuesta y esta presenta forma de cuña (tendencia a aumentar o a disminuir  $e_{i_s}$  al aumentar  $\hat{y}_i$ ), (Figura 2.3); es un indicador de que se ha violado la suposición de que las varianzas de los errores sean constantes, es decir hay presencia de heterocedastidad.

Para remediar este problema es necesario emplear mínimos cuadrados con factores de peso, donde los pesos son inversamente proporcionales a las varianzas. Este caso se desarrollará con mayor profundidad en el siguiente capítulo.



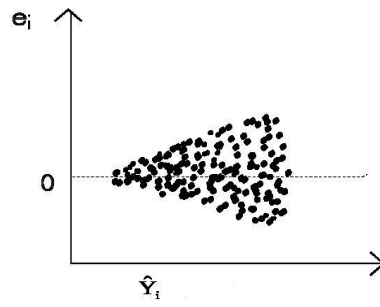


Figura 2.3

Por medio del análisis de residuos también se puede detectar la presencia de puntos que se encuentran alejados de las demás observaciones; llamado outliers, además es posible detectar el incumplimiento del supuesto **(iii)**. Es por esto que realizar un análisis detallado de residuos es indispensable para cualquier buen intento por construir un modelo estadístico que permita realizar predicciones con una confianza aceptable.

### 3. HETEROCEDASTICIDAD

En el Modelo de Regresión Lineal hay presencia de Heterocedasticidad si la varianza de los errores aleatorios no son constantes, es decir; si se incumple el supuesto (ii). Recordemos que en el análisis de residuos, si los residuales estandarizados tienden a aumentar o a disminuir conforme aumentan los valores estimados de la respuesta; la varianza del error no puede considerarse constante y por tanto no es conveniente estimar los parámetros del modelo por el método de los mínimos cuadrados como se ha venido haciendo, pues no se tendría la misma precisión, ya que estos estimadores dejarán de ser los mejores estimadores lineales insesgados de los parámetros del modelo.

Para el modelo dado en (2.2)

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

En presencia de heterocedasticidad, la varianza del vector aleatorio  $\boldsymbol{\varepsilon}$  es  $V(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Sigma}$  donde  $\boldsymbol{\Sigma}$  es la matriz de variaciones de la varianza y la debe especificar el investigador de acuerdo con alguna información que tenga acerca de la relación entre la varianza y las variables de predicción; y  $\sigma^2$  es el factor constante de la varianza.

Donde el vector aleatorio  $\boldsymbol{\varepsilon}$  tiene matriz de varianzas y covarianzas

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \begin{bmatrix} E(\varepsilon_1)^2 & 0 & \cdots & 0 \\ 0 & E(\varepsilon_2)^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E(\varepsilon_n)^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} = \sigma^2\boldsymbol{\Sigma}$$

Asumiendo el cumplimiento de los demás supuestos. La matriz  $\Sigma$  recibe el nombre de matriz de ponderaciones. Entonces el modelo se estandariza de la siguiente manera:

$$y_i/\sigma_i^2 = \beta_0/\sigma_i^2 + \beta_1 x_{i1}/\sigma_i^2 + \beta_2 x_{i2}/\sigma_i^2 + \cdots + \beta_k x_{ik}/\sigma_i^2 + \varepsilon_i/\sigma_i^2 \quad (3.10)$$

donde  $\sigma_i^2$  es la varianza correspondiente al  $i$ -ésimo error aleatorio.

Como  $\Sigma$  es una matriz simétrica, definida positiva, entonces existe una matriz cuadrada, no singular  $\mathbf{V}$  tal que  $\Sigma = \mathbf{V}\mathbf{V}^T$  cuya inversa es de la forma  $\Sigma^{-1} = (\mathbf{V}^{-1})^T \mathbf{V}^{-1}$ .

Se puede utilizar la matriz  $\mathbf{V}^{-1}$  para transformar el modelo heterocedástico (2.2) en el modelo

$$\mathbf{Y}^* = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \quad (3.11)$$

Donde  $\mathbf{Y}^* = \mathbf{V}^{-1} \mathbf{Y}$ ,  $\mathbf{X}^* = \mathbf{V}^{-1} \mathbf{X}$  y  $\boldsymbol{\varepsilon}^* = \mathbf{V}^{-1} \boldsymbol{\varepsilon}$ . pudiéndose demostrar que para este modelo se cumple que

$$E(\boldsymbol{\varepsilon}^*) = E(\mathbf{V}^{-1} \boldsymbol{\varepsilon}) = \mathbf{0}_n.$$

### 3.1. MÍNIMOS CUADRADOS ORDINARIOS (MCO) EN PRESENCIA DE HETEROCEDASTICIDAD

Cuando la matriz de varianzas y covarianzas del vector aleatorio  $\boldsymbol{\varepsilon}$  es  $V(\boldsymbol{\varepsilon}) = \sigma^2 \Sigma$ , el estimador de mínimos cuadrados ordinarios para el vector de parámetros  $\boldsymbol{\beta}$  sigue siendo solución a las ecuaciones normales:

$$(\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y}$$

Cuyo vector de estimaciones es

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

Si reemplazamos  $\mathbf{Y}$  por  $\mathbf{X}\beta + \varepsilon$ , el estimador para  $\beta$  se puede escribir de la siguiente manera

$$\hat{\beta} = \beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon \quad (3.12)$$

Se puede observar, a partir de esta expresión que:

- El estimador de mínimos cuadrados sigue siendo insesgado.  $E[\hat{\beta}] = \beta$
- La matriz de varianzas y covarianzas ya no tiene la forma  $\sigma^2(\mathbf{X}^T \mathbf{X}^{-1})$  sino que es

$$V(\hat{\beta}) = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Sigma \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \quad (3.13)$$

Afirmación que se puede demostrar de la siguiente forma:

$$V(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T]$$

por la expresión (3.12) se tiene que

$$V(\hat{\beta}) = E[(\beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon - \beta)(\beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon - \beta)^T]$$

$$V(\hat{\beta}) = E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon)^T]$$

$$V(\hat{\beta}) = E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon \varepsilon^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}]$$

por propiedades de la esperanza

$$V(\hat{\beta}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E[\varepsilon \varepsilon^T] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$$

$$V(\hat{\beta}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 \Sigma \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$$

Por tanto

$$V(\hat{\beta}) = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Sigma \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$$

- El estimador de mínimos cuadrados es una combinación lineal del error  $\varepsilon$ , y como el vector aleatorio  $\varepsilon$  tiene distribución normal, o sea  $\varepsilon \sim N_n(\beta; \sigma^2 \Sigma)$ , entonces el estimador de *MCO* también se distribuye normalmente

$$\hat{\beta}_{MCO} \sim N(\beta, \sigma^2(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Sigma \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1})$$

Cabe anotar que realizar la estimación de los parámetros  $\beta_i$  con *MCO* en presencia de heterocedasticidad es incorrecto ya que puede producir errores de inferencia arbitrarios, además no podría asegurarse que los estadísticos presentados en el capítulo 2 tengan distribuciones *t* y *F*.

Observe que la expresión (3.13) es la misma matriz de varianzas y covarianzas  $\sigma^2(\mathbf{X}^T \mathbf{X})^{-1}$  del estimador *MCO* cuando  $\Sigma = \mathbf{I}_n$

## 3.2. EL ESTIMADOR DE MÍNIMOS CUADRADOS GENERALIZADOS (MCG)

Vimos que bajo la presencia de heterocedasticidad se puede transformar el modelo original en la ecuación (3.11); el cual es un modelo con los mismos parámetros que el modelo inicial pero cuyo error aleatorio tiene varianza constante.

Entonces, por los resultados presentados en el capítulo 2, para obtener las estimaciones de los parámetros del modelo; se debe minimizar la suma de los cuadrados de los errores en el modelo transformado

$$\varepsilon^{*T} \varepsilon^* = (\mathbf{Y}^* - \mathbf{X}^* \beta)^T (\mathbf{Y}^* - \mathbf{X}^* \beta)$$

De donde se obtienen las ecuaciones normales

$$(\mathbf{X}^{*T} \mathbf{X}^*) \hat{\boldsymbol{\beta}}_{MCG} = \mathbf{X}^{*T} \mathbf{Y}^*$$

reemplazando cada transformación por su respectiva ecuación se tiene

$$\begin{aligned} (\mathbf{V}^{-1} \mathbf{X})^T (\mathbf{V}^{-1} \mathbf{X}) \hat{\boldsymbol{\beta}}_{MCG} &= (\mathbf{V}^{-1} \mathbf{X})^T (\mathbf{V}^{-1} \mathbf{Y}) \\ \mathbf{X}^T (\mathbf{V}^{-1})^T (\mathbf{V}^{-1} \mathbf{X}) \hat{\boldsymbol{\beta}}_{MCG} &= \mathbf{X}^T (\mathbf{V}^{-1})^T \mathbf{V}^{-1} \mathbf{Y} \\ (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}) \hat{\boldsymbol{\beta}}_{MCG} &= \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y} \end{aligned}$$

Cuyo vector de estimaciones siempre que exista la inversa de la matriz  $(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})$ , es

$$\hat{\boldsymbol{\beta}}_{MCG} = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y} \quad (3.14)$$

### Propiedades de los estimadores de *MCG*

Los estimadores *MCG* tienen las mismas propiedades que los estimadores de mínimos cuadrados ordinarios, ya que los estimadores de mínimos cuadrados generalizados son los de mínimos cuadrados en el modelo transformado (3.11).

1. **Insesgamiento:** El estimador de mínimos cuadrados generalizados es insesgado, veamos

*Prueba.*

$$\begin{aligned}
E[\hat{\beta}_{MCG}] &= E[(\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \mathbf{Y}] \\
&= E[(\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} (\mathbf{X} \beta + \varepsilon)] \\
&= E[(\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \mathbf{X} \beta + (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \varepsilon] \\
&= E[\beta + (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \varepsilon] \\
&= \beta + (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} E[\varepsilon] \\
&= \beta + (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \mathbf{0}_n \\
&= \beta
\end{aligned}$$

■

## 2. Matriz de varianzas y Covarianzas:

Para el modelo transformado (3.11) se tiene que

$$\begin{aligned}
V(\hat{\beta}_{MCG}) &= \sigma^2 (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \\
&= \sigma^2 (\mathbf{V}^{-1} \mathbf{X})^T (\mathbf{V}^{-1} \mathbf{X})^{-1} \\
&= \sigma^2 (\mathbf{X}^T (\mathbf{V}^{-1})^T \mathbf{V}^{-1} \mathbf{X})^{-1} \\
&= \sigma^2 (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1}
\end{aligned}$$

Por lo tanto

$$V(\hat{\beta}_{MCG}) = \sigma^2 (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \quad (3.15)$$

3. **Eficiencia:** El estimador *MCG* es el estimador lineal insesgado de mínima varianza de los parámetros del modelo.

Esto se puede garantizar por el hecho de que estos estimadores son los mismos *MCO* del modelo (3.11), entonces por el teorema de *Gauss - Markov*, este es el

Mejor estimador lineal insesgado y de mínima varianza de los parámetros del modelo heterocedástico. (*Esta demostración la podemos encontrar en [Nov]*).

4. El estimador de mínimos cuadrados generalizados es mas eficiente que el estimador de mínimos cuadrados ordinarios.

En efecto, si analizamos la diferencia de las respectivas varianzas de cada estimador, expresiones (3.13) y (3.15) encontramos que la matriz:

$$\sigma^2(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} - \sigma^2(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$$

es semidefinida negativa, esto es la varianza del estimador *MCO* es mayor que la varianza del estimador *MCG*.

5. **Obtención:** El estimador de mínimos cuadrados se puede encontrar de dos formas
  - Haciendo la estimación por *MCO* al modelo transformado en (3.11).
  - Utilizando la expresión (3.14).

La forma mas adecuada, o mas corta; es la primera, puesto que después de haber planteado una hipótesis a cerca de la matriz  $\boldsymbol{\Sigma}$  y encontrar  $\mathbf{V}^{-1}$ , sólo se requiere invertir la matriz  $(\mathbf{X}^{*T} \mathbf{X}^*)^{-1}$  que tiene dimensión  $k$  y no aumenta con el tamaño de la muestra. En cambio, en la segunda alternativa se requiere invertir una matriz de dimensión  $n$  y realizar todos los productos, lo cual se hace mucho mas difícil al aumentar el tamaño de la muestra. Además para este caso, los autores previenen en el costo computacional en el que se puede incurrir.

## Estimación de $\sigma^2$

Para hallar las estimaciones de las varianzas de los parámetros del modelo por medio de *MCG*, expresión (3.15); es necesario conocer el valor constante  $\sigma^2$ , pero como esto no es posible entonces se debe hacer una estimación para poder realizar los cálculos pertinentes.



Siguiendo los pasos vistos en el capítulo 2, el vector de estimaciones para el error aleatorio del modelo (3.11) es

$$\hat{\mathbf{e}}^* = \mathbf{Y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}^*$$

y la suma de los cuadrados de los errores es

$$SCE = \hat{\mathbf{e}}^{*T} \hat{\mathbf{e}}^*$$

Entonces por la propiedad (3) de los estimadores *MCO* presentadas en el Capítulo 2; un estimador insesgado de la varianza es

$$\sigma_{MCG}^2 = \frac{\hat{\mathbf{e}}^{*T} \hat{\mathbf{e}}^*}{n - (k + 1)}$$

Si reemplazamos cada variable por las expresiones originales se tiene

$$\begin{aligned} \sigma_{MCG}^2 &= \frac{\hat{\mathbf{e}}_{MCG}^T (\mathbf{V}^{-1})^T \mathbf{V}^{-1} \hat{\mathbf{e}}_{MCG}}{n - (k + 1)} \\ &= \frac{\hat{\mathbf{e}}_{MCG}^T \boldsymbol{\Sigma}^{-1} \hat{\mathbf{e}}_{MCG}}{n - (k + 1)} \end{aligned}$$

Donde  $\hat{\mathbf{e}}_{MCG} = \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{MCG}$ , entonces:

$$\sigma_{MCG}^2 = \frac{(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{MCG})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{MCG})}{n - (k + 1)}$$

Entonces el estimador por *MCG* de la varianza es

$$\sigma_{MCG}^2 = \frac{\mathbf{Y}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y} - \hat{\boldsymbol{\beta}}_{MCG} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y}}{n - (k + 1)} \quad (3.16)$$

### 3.3. ALGUNAS PRUEBAS DE DETECCIÓN

En esta sección se presentan algunas de las pruebas<sup>4</sup> mas usadas para detectar la heterocedasticidad, sin embargo es válido aclarar que aunque es muy importante detectar la

<sup>4</sup>Estas pruebas están basadas en SEBER(1976;pp147), citado en [Be1],(1990;pp29)

presencia de este fenómeno en el modelo de regresión lineal, esto no es suficiente; puesto que se debe conocer la forma como se comporta dicha heterogeneidad para así, poder dar algunos remedios, si esto es posible, pero esto último se tratará más adelante.

Las pruebas que se mencionarán en este apartado parten de la suposición de que los  $n$  errores aleatorios  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  se pueden clasificar en  $p$  grupos, donde además se supone que la varianza de los errores en cada grupo es constante, esto es la varianza de todos los errores pertenecientes al  $j$ -ésimo grupo es  $\sigma_j^2$  con  $j = 1, 2, \dots, p$ . Nótese que estos supuesto resultan obvios si se entiende que se trata de considerar una situación en la cual no hay presencia de la heterocedasticidad

Lo que se pretende es probar la hipótesis  $H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_p^2 = \sigma^2$ , para esta hipótesis se dispone de  $p$  estadísticas independientes

$$S_j^2 = \frac{\sum (y_{ji} - \bar{y}_j)^2}{n_j - 1}$$

con  $j = 1, 2, \dots, p$ , donde  $n_j$  es el tamaño de la muestra dentro de cada grupo y  $\bar{y}_j$  es su promedio, de esta manera, y por los resultados obtenidos en el Capítulo 1; se tiene que la variable aleatoria  $\frac{(n_j-1)S_j^2}{\sigma_j^2}$  tiene distribución  $\chi^2$  con  $n_j - 1$  grados de libertad.

Entonces la hipótesis nula se puede probar por los siguientes métodos:

## Prueba de Bartlett

Bartlett [1.937] demostró que cuando  $H_0$  es cierta, el estadístico

$$T_1 = \frac{\sum (n_j - 1) \log S^2 - \sum (n_j - 1) \log S_j^2}{C}$$

en el cual

$$S^2 = \frac{\sum (n_j - 1) S_j^2}{\sum n_j - 1}$$

$$C = \frac{\sum(n_j - 1)^{-1} - [\sum(n_j - 1)]^{-1}}{3(p - 1)}$$

Entonces bajo la hipótesis nula;  $T_1$  tiene distribución  $\chi^2$  con  $p - 1$  grados de libertad. También afirma que para muestras pequeñas ( $n_j \leq 4$ ) esta prueba sigue siendo efectiva (robusta frente al tamaño de la muestra).

## Prueba de Cochran

En esta prueba, Cochran(1.941); sugiere que en cada uno de los grupos se debe tener el mismo número de observaciones ( $r$ ) y está diseñada para cuando una varianza es más grande que las otras, y propone el estadístico:

$$T_2 = \frac{\max(S_1^2, S_2^2, \dots, S_p^2)}{S_1^2 + S_2^2 + \dots + S_p^2}$$

La distribución de  $T_2$  depende de  $p$  y de  $r$  y está tabulada en DIXON and MASSEY(1.969) y HARTLEY (1.970).

## Prueba de Hartley

Hartley (1.950) también supone que en todos los grupos se tenga el mismo número de observaciones y propone la estadística:

$$F_{MAX} = \frac{\max(S_1^2, S_2^2, \dots, S_p^2)}{\min(S_1^2, S_2^2, \dots, S_p^2)}$$

la cual es una extensión de la prueba F para la igualdad de varianzas.

Algunos comentarios que se presentan al respecto de estas pruebas son:

- El inconveniente de estas pruebas es que son muy sensibles al alejamiento del supuesto de normalidad, arrojando valores que indican que se debe rechazar  $H_0$

aún cuando es cierta, por ello es recomendable realizar un respectivo diagnóstico de normalidad, sin embargo el rechazo de la hipótesis nula sugiere que se incumple cualquiera de los dos supuestos.

- Las pruebas de Hartley y Cochran requieren que el tamaño de las muestras dentro de cada grupo sea la misma o al menos parecidas, y en este caso se usa como tamaño de muestra el valor mayor de las observaciones o el promedio de ellas.
- La prueba  $F$  no se afecta cuando hay varianzas desiguales que no difieren mucho.
- Si el tamaño de la muestra es grande los residuales heredan la homocedasticidad de los errores, y en este caso es mas conveniente realizar las pruebas con éstos, si no, entonces es conveniente trabajar con la distribución de  $Y$  dado  $X$

## Prueba gráficas

Los gráficos de los residuos contra alguna de las variables involucradas en el modelo, como se vio en el capítulo 2, permiten detectar las posibles violaciones de los supuestos establecidos para el modelo. En el problema que abordamos en este capítulo, vemos como los gráficos; además de detectar la presencia de la heterocedasticidad, también pueden dar información a cerca de la forma de la misma y así sugerir una forma para corregirla. Este método consiste en realizar la gráfica de los residuales estandarizados contra la variable de la cual se sospeche que es la causante de producir la heterocedasticidad, si la gráfica la presentada en la figura 2.3 , entonces se dice que la varianza tiene un comportamiento lineal.

Cuando la gráfica es como aparece en la figura ??, entonces se dice que el comportamiento de la varianza no es lineal. Al reconocer la forma de la heterocedasticidad se debe buscar corregirla, y los métodos empleados para tal fin se presentarán mas adelante.

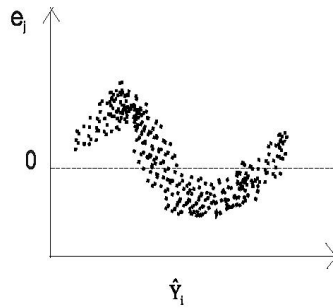


Figura 3.1

### 3.4. CAUSAS MÁS FRECUENTES

Existen muchas causas que conllevan a obtener un modelo heterocedástico; entre las más frecuentes se tienen:

- a. *Cuando las variables explicativas tienen una gran dispersión respecto a su media.*

Los modelos más sensibles a la heterocedasticidad son los modelos de corte transversal, pues las observaciones obtenidas pueden concentrar valores grandes (o valores pequeños) en cierta variable explicativa. Si en particular dicha variable es la que produce la distorsión, entonces para las variables con mayor carga respecto a esta variable, seguramente tendrán mayor peso en la distorsión y de igual manera los valores con menor carga tendrán menor peso. En los modelos de corte temporal se da una situación parecida, pues en uno de los períodos pueden quedar agrupados valores mayores y menores comparándolos respecto a su media.

- b. *Omisión de variables relevantes.*

Cuando en la especificación del modelo se han omitido variables explicativas de importancia, ellas quedarán recogidas en el comportamiento de las perturbaciones aleatorias, pudiendo introducir variaciones no necesariamente fijas de estas.

- c. *Cambio de estructura.*

Cuando se produce un cambio de la estructura del modelo también se produce un ajuste en los parámetros respecto a las observaciones muestrales produciendo

desajustes diferentes en el modelo y así varianzas no constantes por sub-periodos. Además los parámetros estimados en el modelo cuando se produce un cambio de estructura servirán como medida de ponderación de los parámetros estimados del modelo inicial.

Este fenómeno es equivalente al de omisión de variables relevantes, en el cual faltaría precisamente esa variable, ficticia, que realza la diferencia entre las dos estructuras del modelo.

**d.** *Empleo de variables no relativizadas.*

Este caso es similar al presentado en el ítem (a.) en donde las observaciones que contengan valores grandes de la variable explicativa producen probablemente la variación del error.

### **3.5. EFECTOS SOBRE EL MODELO**

Todos los casos presentados antes, tienen algún efecto sobre el modelo, algunos de los más importantes son los siguientes:

**a.** *Incorrecta estimación de los parámetros del modelo*

Como hemos visto anteriormente en la sección (3.1), los estimadores *MCO* aunque siguen siendo insesgados, son ahora ineficientes. Por esto es necesario utilizar los estimadores *MCG* que como se demostró en la sección (3.2); son los mejores estimadores lineales insesgados de mínima varianza y además, este estimador es más eficiente que el estimador *MCO*.

**b.** *Cálculo incorrecto de las varianzas de los parámetros*

Si se realiza la estimación de los parámetros del modelo utilizando *MCO*, haciendo caso omiso a la posible existencia de la heterocedasticidad, se tienen dos opciones:

- Estimar las varianzas como si hubiera homocedasticidad en el modelo (lo que resulta conveniente).

- Estimar los parámetros con *MCO*, calculando la verdadera varianza de los estimadores de los parámetros, o sea calcular la matriz (3.15)(que es más adecuado).

Respecto a esto existe un experimento realizado por Goldfeldt y Quandt(1972)<sup>5</sup> dentro del cual se quería estudiar el comportamiento de esas varianzas y para los casos:

- Estimar los parámetros del modelo inicial mediante *MCG* y calcular las verdaderas varianzas.
- Estimar los parámetros del modelo mediante *MCO* y calcular la varianza bajo el supuesto de homocedasticidad
- Estimar los parámetros del modelo mediante *MCO* y calcular la varianza bajo el supuesto de heterocedasticidad

Obteniéndose la siguiente conclusión:

Cuando se estiman los parámetros con *MCO* en lugar de realizarla por *MCG*, en un modelo heterocedástico, se produce generalmente un incremento de casi diez veces en la estimación de la varianza del parámetro constante y cuatro veces en la varianza de los parámetros que multiplican a las variables explicativas; además cuando se calcula esta varianza se produce una subvaloración en su media real.

- c. *Invalidación de los contrastes de significación* Como ya se ha dicho, se sub-valora la estimación de la varianza de los parámetros al emplear *MCO*, en el caso en que hubiera heterocedasticidad. Los contraste de significación de los parámetros rechazarán la hipótesis nula con mayor frecuencia de la que se espera; la validéz de alguna variable para explicar la variable de respuesta se aceptaría siendo que esto es falso. Ahora si se utilizan *MCO* bajo heterocedasticidad, como ya se ha mencionado, los parámetros estimados pierden eficiencia comparados con los estimadores por

---

<sup>5</sup>citado por [Ar](2001;pp5)

*MCG* por lo que la hipótesis nula en los contrastes de significación se aceptaría más veces de las que se espera.

### 3.6. CARACTERIZACIÓN DE LA FORMA DE LA HETEROCEDASTICIDAD

Al probar que las varianzas de los errores no son constantes, solo se identifica la forma de la Heterocedasticidad, pero esto no es suficiente mientras no se detecte la forma o la naturaleza de la heterocedasticidad, lo que se quiere decir es que es muy importante y conveniente tener información acerca de la matriz de varianzas y covarianzas de los errores para realizar las correcciones del caso. Para esto los autores proponen que la heterocedasticidad matemáticamente se puede escribir como:  $\sigma_j^2 = \sigma^2 + Z_j\gamma$ , para  $J = 1, 2, \dots, n$ , donde  $E[\varepsilon^2] = \sigma_j^2$ . Es decir una relación determinística, en la cual no se dispone de información para la estimación de  $\gamma$

La caracterización de  $Z_i$  conlleva a una gran variedad de heterocedasticidad, y lo que se pretende es construir pruebas estadísticas para la hipótesis  $H_o : \gamma = 0$ .

Una forma es suponer que se dispone de los errores  $\varepsilon_j$  entonces la expresión anterior se puede escribir como:

$$\begin{aligned}\varepsilon_j^2 &= \sigma^2 + Z_j\gamma + \varepsilon_j^2 - \sigma_j^2 \\ \varepsilon_j^2 &= \sigma^2 + Z_j\gamma + \varepsilon_j^*\end{aligned}$$

Este modelo resulta de una forma clásica del MRL, salvo que la variable de respuesta es no observable.

para el cual  $\varepsilon_j^* = \varepsilon_j^2 - \sigma_j^2$ , y así  $E[\varepsilon_j^*] = 0$



Si no se dispone de los errores  $\varepsilon_j^2$  pues estos son no observables entonces ellos se pueden estimar a través de los residuales  $e_j$ , acudiendo a algunos resultados de convergencia asintótica tenemos:

$$e_j^2 = \sigma^2 + Z_j\gamma + \varepsilon_j^*$$

Lo anterior en forma matricial puede expresarse de la forma

$$\mathbf{e} = \begin{bmatrix} J & Z \end{bmatrix} \begin{bmatrix} \sigma^2 \\ \gamma \end{bmatrix} + \varepsilon^* = \mathbf{D}\mathbf{s} + \varepsilon^*$$

En donde  $e$  es el vector de los residuales al cuadrado,  $J$  es un vector de  $n$  unos y  $Z$  es el vector de los  $n$  valores de  $Z_j$ , entonces  $\mathbf{D} = \begin{bmatrix} J & Z \end{bmatrix}$   $\mathbf{s} = \begin{bmatrix} \sigma^2 \\ \gamma \end{bmatrix}$

Estimando por mínimos cuadrados a  $S$ , tenemos que:

$$\widehat{\mathbf{S}} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{e} = \left( \sum_{i=1}^n d_i d_i \right)^{-1} \sum_{i=1}^n d_i e_i$$

para el cual  $d_i$  es la fila  $i$  de la matriz  $D$ , es decir  $d_i = (1 \ Z_j)$  y  $d_i = \begin{bmatrix} 1 \\ Z_j \end{bmatrix}$ .

Pagan y Hall (1983) demuestran que bajo ciertas restricciones  $\widehat{\mathbf{S}}$  es consistente para  $S$ , bajo el supuesto de homocedasticidad.

El estadístico

$$W = \frac{(\widehat{S} - S)}{\sqrt{n}} \xrightarrow{d} N \left[ 0, (\mu_4 - \sigma^4) \left( p \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n d_t d_t \right)^{-1} \right]$$

Donde  $E[\varepsilon_j^4] = \mu_4$  es decir que el estadístico  $W$  converge distribucionalmente a la Normal con medias 0 y varianza  $(\mu_4 - \sigma^4) \left( p \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n d_t d_t \right)^{-1}$

Esto permite construir pruebas para la hipótesis  $H_0 : \gamma = 0$ . Es muy importante resaltar que aún si la hipótesis nula es falsa se preserva la consistencia de  $\hat{S}$  y la distribución límite.

Hasta el momento no se ha hablado de la naturaleza de  $Z^6$ . Veamos:

Si  $Z$  es no aleatoria entonces:

$$V[W] = (\mu_4 - \sigma^4) \begin{bmatrix} 1 & \hat{Z}_n \\ \hat{Z}_n & \hat{Z}_n^2 \end{bmatrix}$$

Lo cual podrá usarse para realizar cálculos pertinentes a la prueba de hipótesis  $H_0 : \gamma = 0$ , usando como estimadores de  $\mu_4$  y  $\sigma^4$  a:  $\hat{\mu}_4 = \frac{1}{n} \sum_{j=1}^n e_j^4$  y  $\hat{\sigma}^4 = s^4$ .

Y si  $Z$  es aleatoria tenemos que:

$$V_A[W] = \begin{bmatrix} 1 & \mu_2 \\ \mu_2 & b \end{bmatrix} (\mu_4 - \sigma^4)$$

Donde  $\mu_2$  y  $b$  se desconocen, pero puede utilizarse cualquier estimador consistente, con lo que  $V[W]$  puede tener la misma estructura de la estimación para el caso no aleatorio

### 3.7. CONTRASTES PARA DETERMINAR LA FORMA DE LA HETEROCEDASTICIDAD

Comúnmente se desconoce la naturaleza de la heterocedasticidad, para tener alguna información acerca de ella, existen en la bibliografía pruebas con las cuales se contrasta la hipótesis nula de ausencia de heterocedasticidad considerando un modelo particular para la misma. Se mencionarán algunos de estos contrastes, que son especificados como los más utilizados.

---

<sup>6</sup>Este vector debe diferenciarse del estadístico  $Z$ , definido en Preliminares

## Contraste de Goldfeldt y Quandt

Este contraste<sup>7</sup> consiste en dividir las observaciones en dos grupos, un grupo conformado por los valores pequeños de las observaciones y otro formado por los valores grandes, en procura de identificar un cambio de estructura.

Entonces el proceso a seguir es:

- a. Se ordenan las observaciones de forma ascendente y se dividen en dos grupos, digamos de tamaños  $p$  y  $n - p$ , donde  $p$  es un número entero.
- b. Se estiman dos regresiones, una para cada subgrupo de observaciones.
- c. Se calculan las sumas residuales en los dos grupos, digamos  $SCE_1$  para el grupo de observaciones pequeñas y  $SCE_2$  para el grupo de observaciones grandes. Bajo el supuesto de homocedasticidad y la normalidad de los errores se tiene que

$$F = \frac{SCE_2}{SCE_1} = \frac{\sigma_2^2}{\sigma_1^2}$$

Donde  $\sigma_1^2$  y  $\sigma_2^2$  son las varianzas de los dos grupos, así  $F \sim F_{r,s,\alpha}$ , donde  $r = s = \frac{n-p}{2} - (k + 1)$  que es el número de grado de libertad.

Si existe heterocedasticidad del tipo que se ha propuesto entonces la varianza del error será mayor a medida que sea mayor la muestra, como el cuadrado del error está relacionado con la varianza de  $e_i$ , entonces  $SCE_1$  debe ser sensiblemente menor a  $SCE_2$  y  $F$  excedería el valor de las tablas de la distribución  $F$  con el cual se rechazaría  $H_0$ .

Si se sospecha que la relación existente entre la varianza del error y  $Z_i$  es inversa, entonces en el paso (a.) deberá tomarse la ordenación en forma descendente.

---

<sup>7</sup>GOLDFELD, A. QUANT, R.(1965)referido por [Nov]pp200

Si es posible y algunos autores lo recomiendan, excluir observaciones centrales con el fin de identificar con mayor eficacia la diferencia entre los grupos, sin embargo, hacen la aclaración de que si se excluyen demasiadas observaciones, esto conlleva a una pérdida de potencia en el contraste global, pues se pierden muchos grados de libertad al realizar las dos regresiones; la potencia tenderán aumentar, pues las sub-muestras se diferenciarán entre si y se hará mas difícil detectar el problema.

Por otro lado si excluimos pocas observaciones las dos muestras se comportan en forma similar en sus límites próximos y detectar la heterocedasticidad será una tarea difícil, por lo que se decrementa la potencia del contraste.

Si se llega a concluir que el error no presenta heterocedasticidad al caer en uno de las dos situaciones antes presentadas, lo mas probable es que se halla hecho una mala especificación de  $\sigma_i^2$  o que este efecto se deba a otra variable.

Se vuelve este un problema para resolver por medio de ensayo y error tratando de aproximar un análisis de sensibilidad con respecto al número de observaciones.

## Contraste de Breusch y Pagan

Lo que pretende este contraste<sup>8</sup> es comprobar si la varianza de las perturbaciones aleatorias puede explicarse a través de un conjunto de variables  $Z$  estimadas por medio del cuadrado de los errores correspondientes al modelo inicial.

Para esto supongamos que la varianza de las perturbaciones aleatorias depende de un vector de variables  $Z_i$  de dimensión  $p$

$$\sigma_i^2 = h(\alpha_0 + \alpha_1 Z_{i1} + \alpha_2 Z_{i2} + \cdots + \alpha_p Z_{ip})$$

---

<sup>8</sup>BREUSCH, T. PAGAN, A.(1979)referido en [Nov](201)

Si se pudieran encontrar las estimaciones de  $\alpha_1, \alpha_2, \dots, \alpha_n$  una hipótesis nula de homocedasticidad será  $H_0 = \alpha_0 = \alpha_1 = \alpha_n = 0$

Donde el proceso a seguir es:

- a. Estimar el modelo inicial por *MCO* calculando los residuos correspondientes.
- b. Calcular la serie de los cuadrados de los residuos normalizados al cuadrado

$$\hat{e}^2 = \frac{e_i^2}{\hat{\sigma}_i^2}$$

donde  $\hat{\sigma}_i^2 = \frac{SCE^9}{T}$ . Esto se hace con el fin de eliminar las posibles distorsiones ocasionadas por distintas dimensiones del error.

- c. Se estima una regresión de los residuales (calculados en el item anterior) explicada por una constante y el conjunto de variables  $Z_{j1}, Z_{j2}, Z_{2p}$  obteniéndose el  $R^2$  del modelo.
- d. Como el modelo tiene término constante, entonces se valida la relación  $STC = SCE + SCR$ , concluyendo que si el modelo no explica suficientemente la relación cuando  $SCR$  es muy pequeña.
- e. Bajo la hipótesis nula  $SCR/2$  tiene distribución  $\chi^2$  con  $p$  grados de libertad. Si el valor arrojado por este estadístico es mayor que el valor de la distribución  $\chi^2$  tabulada para el nivel de significación especificado, entonces se considera suficientemente grande y en consecuencia se debe rechazar la hipótesis nula.

Este contraste efectivamente sirve para aceptar o descartar la presencia de la heterocedasticidad debido a un conjunto de variables  $Z_i$  y el número de estas variables debe ser pequeño para que se pueda realizar el contraste. Esto permite además cierta flexibilidad ya que no pretende realizar una especificación completa de una función  $h$ .

---

<sup>9</sup>Suma de cuadrados de los errores

## Contraste de Glesjer

Este contraste<sup>10</sup> es un poco similar al anterior, pero este no se limita a una relación lineal si no que propone la variación de las perturbaciones aleatorias en función de una variable  $Z$  que puede estar elevada a una potencia  $h$ , donde  $h$  está entre  $-1$  y  $1$ .

Las etapas de este contraste son:

- a. Estimar el modelo original por *MCO* obteniendo los residuos correspondientes.
- b. Correr cuatro regresiones para los valores absolutos del error del modelo anterior sobre la potencia  $h$  de la variable  $Z_i$ .

$$|e_i| = \alpha_0 + \alpha_1 Z_i^h + \varepsilon_i$$

Donde  $h = \{-1, -0,5, 0,5, 1\}$  se escoge la regreión con parámetro significativo y con mayor  $R^2$

Si el valor es suficientemente grande, se confirmará que hay heterocedaticidad, la cual es producida por  $Z_j$ .

## Contraste de Wite

Lo que pretende este contraste es determinar si las varianzas de las variables explicativas del modelo, sus cuadrados y productos cruzados sirven para determinar la evolución del cuadrado de los residuales cuyas etapas son:

- a. Se estima el modelo por *MCO* y se determina la serie de los residuos.
- b. Se estima el modelo de los valores al cuadrado de los residuales, sobre una constante con todas las variables explicativas del modelo inicial, sus cuadrados y productos cruzados de segundo orden

---

<sup>10</sup>Citado en [Nov](pp203)

$$\hat{e}_i^2 = \alpha_0 + \alpha_1 x_{1i} + \cdots + \alpha_k x_{ki} + \alpha_{k+1} x_{1i}^2 + \cdots + \alpha_k + k + 1 x_{1i} x_{2i} \\ + \alpha_{k+k+2} x_{1i} x_{3i} + \cdots + \alpha_{3k+1} x_{2i} x_{3i} + \cdots + \epsilon_i.$$

- c. El valor  $R_e^2$  nos indicará si las variables escogidas sirven o no para estimar la evolución de los residuales al cuadrado. Si el modelo realmente fuera Heterocedástico entonces  $R_e^2$ , debería ser pequeña y tiene una distribución  $\chi^2$  con  $k - 1$  grados de libertad.

Si se obtiene un valor de  $R_e^2$  que supera el reflejado por las tablas de  $\chi^2$  afirmaremos que existe heterocedasticidad.

White enuncia otro modo de contrastar la existencia de heterocedastidad, a través de la validez de los parámetros, el cual estaría dado por el valor del contraste de significación conjunta  $F$ . Si éste afirmase que el conjunto de variables tienen la suficiente capacidad explicativa sobre  $e_i$ , entonces hay presencia de heterocedasticidad en el modelo.

Otros contrastes que permiten determinar la naturaleza de la heterocedasticidad y que pueden encontrarse en la bibliografía citada son:

- \* Contraste a partir del coeficiente de correlación por rangos de Spearman.
- \* Contraste de Harvey
- \* Contraste Reset de Ramsey
- \* Contrate de Picos
- \* Contraste LM Arch

Además de esto debe tenerse en cuenta que al presentarse una falta de ajuste en modelo, se puede obtener alguna forma de heterocedasticidad. Observación que lleva a entender la importancia del análisis de los residuos, ya que es a través de estos que se puede detectar y contrastar esta problemática.

### 3.8. CÓMO SE CORRIGE LA HETEROCEDASTICIDAD

Hemos dicho ya que la heterocedasticidad está producida casi siempre por las dependencia de la varianza no constante de las perturbaciones aleatorias, inducida por una o más variables presentes o no en el modelo. Se consideraron pruebas para detectarla y corregirla, ahora, si ésta realmente se diese entonces tenemos que esta tomará la forma general  $\sigma_i^2 = f(\sigma^2, Z_i)$  que se denomina como el supuesto simplificador de la heterocedasticidad.

El detectar y conocer la estructura de la heterocedasticidad son aspectos tan importantes como lo es corregirla, tratando de eliminar los efectos que ella conlleva.

Detectado el problema se procede a operar convenientemente las variables del modelo eliminando la fuente de la heterocedasticidad.

Luego se divide el conjunto total de las variables del modelo por la forma estimada de la desviación estándar heterocedástica a la cual ya se le ha encontrado la forma a través de los contrastes, de esta manera se estará corrigiendo el modelo.

En forma resumida, los pasos a seguir en un problema que tenga presencia de heterocedasticidad son:

1. se estiman los parámetros del modelo por medio de *MCO*, ignorando momentáneamente la heterocedasticidad.
2. se formula un supuesto a cerca de la naturaleza de  $\sigma_i^2$  y se emplean los residuales de la regresión a través de mínimos cuadrados ordinarios para contrastar y estimar la forma funcional supuesta
3. Se divide cada observación por la raíz de las varianzas heterocedásticas estimadas después de que cualquiera de los contrastes haya confirmado que el supuesto sim-



plificador es bueno<sup>11</sup>.

4. Se realizan las estimaciones del modelo original, con todas las variables transformadas como se presentó anteriormente, esto es, utilizando la técnica de los mínimos cuadrados generalizados.

---

<sup>11</sup>En forma matricial, esto es equivalente a transformar el modelo por medio de la matriz  $\mathbf{V}^{-1}$

# CONCLUSIONES

A continuación se presentan las conclusiones de este trabajo relacionadas de acuerdo con los aportes teóricos y personales que para las autoras son consideradas como importantes.

## Conclusiones Teóricas

La estadística es una ciencia que tiene gran importancia en la vida práctica debido a que aborda en sus técnicas la resolución de problemas que tienen que ver con los fenómenos estocásticos. Una de las más popularizadas es precisamente la que tiene que ver con la modelación de ese tipo de problemas, modelación que se basa en la técnica de los mínimos cuadrados para la construcción de los estimadores de los parámetros del modelo de interés.

Se ha demostrado que, para el *MGL*, tal técnica produce los mejores estimadores para los parámetros del modelo, siempre que las condiciones (supuestos) sean válidas.

La construcción de un modelo estadístico no es sencilla, se trata de una problemática que incluye, además del conocimiento de las técnicas estadísticas, involucrarse o conocer muy bien el problema que ha de intentar solucionar. Es precisamente la combinación de estos dos conocimientos lo que garantiza el éxito en el trabajo.

Sin embargo, en lo que se refiere a algunos aspectos de la técnica de modelación estadística, hemos evidenciado que a pesar de que la construcción de los estimadores de los parámetros del modelo se resuelve fácilmente mediante la técnica de mínimos cuadrados mencionada, el problema fuerte radica en la verificación del cumplimiento de los supuestos

y su posterior corrección en el caso de que alguno(s) no se cumpla(n).

El problema que abordamos y que dio pie a este documento, problema para el cual en la bibliografía no abunda material de investigación, encontrándose en la mayoría cita de capítulos de libros o temas de un curso de econometría aplicada, permite concluir que es una problemática que se encuentra en vigencia, que aún no se ha establecido la última conclusión ni se ha encontrado la forma exacta para determinar el comportamiento del fenómeno denominado *heterocedasticidad*.

A pesar de que se presentan propuestas para detectar su presencia y contrastes para determinar su forma, así mismo los autores no garantizan la cobertura total de la problemática.

Después de estudiar esta temática y aprender un poco más acerca de la misma, podemos entender por qué a la fecha, el problema de heterocedasticidad junto con el problema de la autocorrelación, son los temas de mayor importancia al momento de la construcción de modelos estadísticos de tipo lineal.

Adicionalmente, y con el ánimo de aportar en el trabajo de investigación, nos atrevemos a formular las siguientes preguntas:

1. Qué pasa con los estimadores de *MCG* cuando además del incumplimiento del supuesto de Homocedasticidad se incumple el supuesto de Incorrelación de los errores, o de otro de los supuestos?.
2. Qué efectos tiene la Heterocedasticidad sobre la construcción de los estimadores del Modelo de Regresión Lineal, cuando las variables de predicción son también aleatorias?
- 3.Cuál es el papel del software estadístico profesional en la solución del problema de

la heterocedasticidad en la práctica?

## Conclusiones Personales

Inicialmente este trabajo se planteó como un trabajo de investigación, pero debido a la reglamentación que surgió nos vimos en la necesidad de encaminarlo a un seminario especial lo que condujo a que se cambiaran los objetivos y metodologías iniciales.

Después de haber presentado esta problemática se deja como inquietud la presentación de ejemplos prácticos y del estudio de esta problemática en aspectos de la vida cotidiana, dentro de nuestro contexto.

Durante la revisión bibliográfica realizada para abordar cada uno de los temas presentados en este seminario, fué necesario unificar la notación matemática que presenta cada uno de los autores para poder llevar una lógica en los desarrollos presentados en este documento.

A pesar de los inconvenientes que surgieron durante el desarrollo de este seminario se logró cumplir con cada uno de los propósitos inicialmente establecidos.

# BIBLIOGRAFÍA

- [Ba] BARBOLLA, R. SANZ, P..Álgebra Lineal y Teoría de Matrices. Prentice Hall. 1990
- [Be] BEHAR, R. Validación de los Supuestos del Modelo de Regresión Lineal, Memorias. Cali. Universidad del Valle. 2002.
- [Be1] BEHAR, R. Supuestos del Modelo de Regresión Lineal. Primer Simposio de Estadística Univarsidad Nacional. Bogotá. 1990.
- [Box] BOX, W. G.. HUNTER J.S.. Estadística parainvestigadores. México. Reverté. 2001.
- [Ca] CANAVOS, G. Probabilidad y Estadística, Aplicaciones y métodos. México. McGraw-Hill. 1988.
- [Co1] COCHRAN. W.. Técnicas de muestreo. México. CECOSA. 1992.
- [Co2] COCHRAN. W.. COX. G.M..Diseños e xperimentales. México. Trillas. 1991.
- [Ar] DE ARCE, R. Conceptos Básicos sobre la Heterocedasticidad en el Modelo Básico de Regresión Lineal, Tratamiento con E-VIEWS. Departamento de Economía Aplicada, Universidad Autónoma de Madrid. 2001.
- [De] DEGROOT, M. Probabilidad y Estadística. México. Addison Wesley Iberoamericana. Segunda Edición. 1988.
- [Dra] DRAPER, N. SMITH, H. Applied Regresión Analysis. USA. Jhon Wiley & Sons. Inc. 1981.
- [Fr] FREUD, J. MILLER, I. MILLER, M. Estadística Matemática con Aplicaciones. México. Prentice Hall. Sexta Edición. 2000.

- [Gr] GRAYBILL, F. A. Theory and Application of the Linear Model. California. Duxbury Press. 1976.
- [Gu] GUJARATI, Damodar N. Econometría. México. McGraw-Hill. Segunda Edición. 1992.
- [He] HERNÁNDEZ, R. FERNÁNDEZ, C. BAPTISTA, P. Metodología de la Investigación. México. McGraw-Hill. Segunda Edición. 1998.
- [Ko] KOROLIUK, V. S. Manual de Teoría de Probabilidades y Estadística Matemática Editorial Mir. Primera Reimpresión 1981. Segunda Edición. 1998.
- [Md] MADDALA, G. S. Econometría. México. McGraw-Hill. Primera Edición. 1985.
- [Mr] MARMOLEJO, M.A. Algebra Lineal II, Estadística. Cali. Universidad del Valle. 1988.
- [Ma] MARTINEZ, A. Métodos Económicos. México. Centro de Estadística y Cálculo del Colegio de Postgraduados, Escuela Nacional de Agricultura. Segunda Edición. 1992.
- [Ma2] MARTINEZ, A.. Diseños Experimentales. México. Trillas. 1988.
- [My] MAYORGA, J., H.. SOTO, O., F.. El análisis de Regresión, Perspectiva Histórica. Validación. Primer Simposio de Estadística Universidad Nacional (Memorias) 1.990. Bogotá. 1990.
- [Md] MENDENHALL, W. SHEAFFER, R. WACKERLY, D. Estadística Matemática con Aplicaciones. México. Grupo Editorial Iberoamericana. 1986
- [MP] MONTGOMERY, D. PECK E. Introduction to Linear Regression Analysis. USA. John Wiley & Sons, Inc. 1.992
- [MR] MONTGOMERY, D. RUGER, G. Probabilidad y Estadística Aplicadas a la Ingeniería. México D.F.. McGraw-Hill. 1.996.

- [MR2] MONTGOMERY, D. Diseño y Análisis de Experimentos. Grupo Editorial Iberoamericana S.A.. 1991.
- [MG] MOOD, M., GRAYBILL, F., BOES, D., C. Introduction to the theory of statistics. McGraw-Hill. Tercera Edición. 1974
- [Nov] NOVALES, Alfonso. Econometría. España. McGraw-Hill. 1993.
- [Pe1] PEÑA, Daniel. Estadística, Modelos y Métodos 1, Fundamentos. España. Alianza Universal. 1991.
- [Pe2] PEÑA, Daniel. Estadística, Modelos y Métodos 2, Modelos Lineales y Series Temporales. España. Alianza Universal. 1991.
- [Qr] QUERUZ, E. L.. Modelo Lineal General (Algunos conceptos Metodológicos). Primer Simposio de Estadística Universidad Nacional (Memorias) 1.990. Bogotá. 1990.
- [Sn] SNEDECOR, G., COCHRAN W., G.. Métodos Estadísticos. México. 1980.